THE ROLE OF EPISTEMIC REASONING IN PRAGMATIC DEVELOPMENT

by

Alyssa Kampa

A dissertation submitted to the Faculty of the University of Delaware in partial fulfillment of the requirements for the degree of Doctor of Philosophy in Linguistics

Fall 2021

THE ROLE OF EPISTEMIC REASONING IN PRAGMATIC DEVELOPMENT

by

Alyssa Kampa

Approved: ____
Benjamin Bruening, Ph.D.
Chair of the Department of Linguistics and Cognitive Science

Approved: ____
John A. Pelesko, Ph.D.
Dean of the College of Arts and Sciences

Approved: ____
Louis F. Rossi, Ph.D.
Vice Provost for Graduate and Professional Education and
Dean of the Graduate College

I certify that I have read this dissertation and that in my opinion it meets the academic and professional standard required by the University as a dissertation for the degree of Doctor of Philosophy.

Signed: ______________________________

Anna Papafragou, Ph.D.
Professor in charge of dissertation

I certify that I have read this dissertation and that in my opinion it meets the academic and professional standard required by the University as a dissertation for the degree of Doctor of Philosophy.

Signed: ______________________________

Kaja K. Jasińska, Ph.D.
Member of dissertation committee

I certify that I have read this dissertation and that in my opinion it meets the academic and professional standard required by the University as a dissertation for the degree of Doctor of Philosophy.

Signed: ______________________________

Roberta M. Golinkoff, Ph.D.
Member of dissertation committee

I certify that I have read this dissertation and that in my opinion it meets the academic and professional standard required by the University as a dissertation for the degree of Doctor of Philosophy.

Signed: ______________________________

Satoshi Tomioka, Ph.D.
Member of dissertation committee

ACKNOWLEDGMENTS

First and foremost, I would like to thank my advisor, Anna Papafragou. I was able to write 40,000 words for this dissertation, but yet I still don't have enough words to express my gratitude to her all of the time and effort she has invested in me. From answering emails and looking over drafts at all hours of the day and night (I'm still not entirely convinced she sleeps) to finding polite ways to say "that research design makes no sense" and helping me to come up with a better one, I feel extremely privileged to be able to call Anna my mentor. I truly admire her as a scholar and as a role model.

I would also like to express my gratitude to Kaja Jasińska for helping me to develop the fNIRS neuroimaging study in this dissertation. Teaching someone from the ground up how to design, run, and analyze a neuroimaging study is no mean feat and to do so in an area of research that is not in her comfort zone is an excellent example of the admirable way she always challenges herself to grow intellectually. I am grateful to her for everything that she has invested in seeing this project come to fruition as well as for helping me to grow in the same way. I also want to thank the other members of my committee, Roberta Michnick Golinkoff and Satoshi Tomioka, for always being so helpful, especially as this dissertation process has taken some unexpected turns, as well as for contributing their time and expertise to improving this dissertation.

I would like to thank my labmates from the Language and Cognition lab, Katie Richards, Myrto Grigoroglou, Sarah Fairchild, Yue Ji, Dionysia Saratsli, and Ariel Mathis, for all of their help and insightful feedback over the years on everything from experimental design to conference presentations to how to take glamour shots of stuffed animals. Also worthy of acknowledgment are the undergrads (Gabby Taubenfeld, Bryce Lamers, Julia Kahane, Danielle Longo, Charlotte Donoho, Merri Smith, Erin Diviney) who spent countless hours, including giving up many of their Sundays, to help me test child participants at the museum and preschools. I also want to thank my honorary lab mates from the BOLD Lab (Jo Hannon, Kelsey Mulford, Violet Kozloff, Sarah Wang, Ben Zinszer) for their help, companionship, and snacks, as well as the undergrads who helped me test participants as I could not have done it alone; special shout-out to Kelsey for cajoling her friends and family members to participate in the fNIRS study and to her family for hosting me while we tested them.

I also owe my gratitude to the hundreds of parents and children who participated in these studies and others, with special thanks to Early Language Center at the University of Delaware, the Methodist Preschool in Newark, and the Delaware Children's Museum for allowing us to conduct research onsite. Thanks to the National Science Foundation (#1632849) for funding this research.

I would like to thank my fellow grad students in Linguistics who have made Monday night trivia an institution for helping me to destress and remember that there's more to life than data and manuscripts (there's also burrata cheese boards).

To my parents, Jim and Laurie Kampa (aka Dad and Mom), throughout my life and over the course of this degree, even though at times you haven't really understood

what I'm doing or why, you've always given me your full support and a seemingly endless supply of enthusiasm and for that I am very grateful.

Lastly I would like to thank my husband, best friend, partner, gentleman caller, hype man, general favorite, and banana clipper finder, Greg. There aren't enough silly nouns to describe what you are to me, and I'm glad we've been able to go through this process and through life together.

TABLE OF CONTENTS

LIST OF TABLES xi
LIST OF FIGURES xiii
ABSTRACT xvi

Chapter

1 INTRODUCTION 1

1.1 Scalar Implicatures 3
1.2 Epistemic Reasoning in Scalar Inference 4
1.3 Epistemic Reasoning in Scalar Inference in Non-Linguistic Communication 6
1.4 Epistemic Reasoning in Scalar Inference From a Neural Perspective 7
1.5 Overview 8

2 EPISTEMIC REASONING IN SCALAR INFERENCE IN LINGUISTIC COMMUNICATION 10

2.1 Experiment 1 16

2.1.1 Methods 16

2.1.1.1 Participants 16
2.1.1.2 Procedure 16

2.1.2 Predictions 22
2.1.3 Coding 22
2.1.4 Results 23
2.1.5 Discussion 24

2.2 Experiment 2 26

2.2.1 Methods 26

2.2.1.1 Participants 26
2.2.1.2 Procedure 27

2.2.2 Coding 29
2.2.3 Results 29
2.2.4 Discussion.. 30

2.3 Experiment 3 31

2.3.1 Methods 32

2.3.1.1 Participants 32
2.3.1.2 Procedure 33

2.3.2 Coding 34
2.3.3 Results 35
2.3.4 Discussion.. 36

2.4 General Discussion 38

3 EPISTEMIC REASONING IN SCALAR INFERENCE IN NON-LINGUISTIC COMMUNICATION.. 42

3.1 Introduction 42

3.1.1 Drawings as Communicative Symbols.. 45
3.1.2 Current Study.. 48

3.2 Experiment 4 50

3.2.1 Methods 50

3.2.1.1 Participants 50
3.2.1.2 Procedure 51

3.2.2 Coding 58
3.2.3 Results 58
3.2.4 Discussion.. 61

3.3 Experiment 5 62

3.3.1 Methods 62

3.3.1.1 Participants 62
3.3.1.2 Procedure 63

3.3.2 Coding 63
3.3.3 Results 63

3.3.4 Discussion 66

3.4 Experiment 6 66

3.4.1 Methods 66

3.4.1.1 Participants 66
3.4.1.2 Procedure 67

3.4.2 Coding 67
3.4.3 Results 67
3.4.4 Discussion 70

3.5 General Discussion 71

3.5.1 Developmental Limitations and Growth 75
3.5.2 Final Thoughts 78

4 EPISTEMIC REASONING IN SCALAR INFERENCE FROM A NEURAL PERSPECTIVE 80

4.1 Introduction 80

4.1.1 Neural Systems and Scalar Implicature 85
4.1.2 Current Study 88

4.2 Methods 89

4.2.1 Participants 89
4.2.2 Behavioral Individual Differences Battery (Children) 90

4.2.2.1 Theory of Mind Task (Sally-Anne) 90
4.2.2.2 Vocabulary Task (Picture Vocabulary Test) 91
4.2.2.3 Executive Functions Task (Dimensional Change Card Sorting Task) 91

4.2.3 Neuroimaging Tasks (Adults and Children) 92

4.2.3.1 Scalar Implicature Task 93
4.2.3.2 Theory of Mind Task 97

4.3 Hypotheses and Analysis Plan 99

4.3.1 Behavioral Tasks: Individual Differences and Scalar Implicature Performance (Children) 99

4.3.2 Neuroimaging Tasks (Children and Adults) 99

4.3.2.1 Scalar Implicature 99
4.3.2.2 Theory of Mind 100
4.3.2.3 Scalar Implicature and Theory of Mind (Conjunction Analysis) 101

4.4 Results and Discussion 101

4.4.1 Behavioral Tasks: Individual Differences and Scalar Implicature Performance (Children) 101
4.4.2 Neuroimaging Tasks (Adults and Children) 104

4.4.2.1 Preliminaries: fNIRS Data Preprocessing, Spatial Registration, and Statistical Parametric Mapping 104
4.4.2.2 Scalar Implicature 105
4.4.2.3 Theory of Mind 114
4.4.2.4 Scalar Implicature and Theory of Mind (Conjunction analysis) 117

4.5 General Discussion 119

4.5.1 The Cognitive Signature of Scalar Inference 119
4.5.2 Limitations and Extensions 122
4.5.3 Final Remarks 124

5 SUMMARY & CONCLUSIONS 125

REFERENCES 131

Appendix

A STIMULI LIST FOR EXPERIMENTS 1-6 151
B SUPPLEMENTARY MATERIALS FOR CHAPTER 4 152
C IRB APPROVAL FOR HUMAN SUBJECTS RESEARCH 159
D IRB APPROVAL FOR HUMAN SUBJECTS RESEARCH 161
E PERMISSIONS 163

LIST OF TABLES

Table 1: Number of Passers and Failers in the More-Informative and Less-Informative Conditions for Experiment 1 for Adults, 5-year-olds, and 4-year-olds. 24

Table 2: Number of Passers and Failers in the More-Informative and Less-Informative Conditions for Experiment 2 for Adults, Older 4-year-olds (4;6-4;11), and Younger 4-year-olds (4;0-4;5). 30

Table 3: Number of Passers and Failers in the More-Informative and Less-Informative Conditions for Experiment 3 for Adults, 5-year-olds, and 4-year-olds. 35

Table 4: Parameter Estimates for Accuracy in Experiment 4. 59

Table 5: Parameter Estimates for Accuracy in Experiment 5. 64

Table 6: Parameter Estimates for Accuracy in Experiments 4-5. 65

Table 7: Parameter Estimates for Accuracy in Experiment 6. 68

Table 8: Parameter Estimates for Accuracy in Experiments 4, 5, & 6. 70

Table 9: Example Stimuli for fNIRS Theory of Mind Task 98

Table 10: Parameter Estimates for Behavioral Accuracy in the Scalar Implicature Task 103

Table 11: Regions with Significant HbO Activation for Scalar Inference for Adults 106

Table 12: Regions with Significant HbR Activation for Scalar Inference for Child Passers 109

Table 13: Regions with Significant HbR Activation for Scalar Inference for Child Non-passers 110

Table 14: Regions with Significant HbR Activation for Scalar Inference for Child Passers – Group 1 112

Table 15: Regions with Significant HbO Activation for Mentalizing for Adults . 116

Table 16: Regions with Significant HbR Activation for Mentalizing for Child Scalar-Implicature Passers........ 116

Table B.1: Additional Regions with Significant HbO Activation for Scalar Inference for Adults........ 157

Table B.2: Additional Regions with Significant HbR Activation for Scalar Inference for Child Passers........ 158

LIST OF FIGURES

Figure 1: Visual displays for the introductory phase of Experiment 1. The faces have been blurred in this image per University of Delaware policy but were not blurred for the experiment. 17

Figure 2: A schematic view of the visual and verbal stimuli for a test trial in Experiments 1, 2, and 3. Across all studies, the photo on the left depicts the "full access" box (and the full-knowledge person). The photo on the right depicts the "limited access box" (and the limited-knowledge person). In the More-Informative condition, participants heard a statement mentioning two objects (Experiments 1 and 2) or saw the action of circling two objects (Experiment 3), and were expected to select the full access box. In the Less-Informative Condition, participants heard a statement mentioning one object (Experiments 1 and 2) or saw the action of circling one object (Experiment 3), and were expected to select the limited access box. The faces have been blurred in this image per University of Delaware policy but were not blurred for the experiment. 20

Figure 3: Visual displays for the introductory phase of Experiment 3. At the end of each pre-test trial, participants heard "I see this" as an animation of a red circle appeared around two of the objects in the array above the photos in succession. In pre-test trial 1, the picture of the book was circled first, followed by the giraffe. In pre-test trial 2, the picture of the book was circled first, followed by the cup. The faces have been blurred in this image per University of Delaware policy but were not blurred for the experiment. 34

Figure 4: Visual displays for the introductory phase of Experiment 4. The faces have been blurred in this image per University of Delaware policy but were not blurred for the experiment. 52

Figure 5: An example visual display from a test trial in Experiment 4. The photo on the left depicts the “full access” box (and the full-knowledge agent). The photo on the right depicts the “limited access box” (and the limited-knowledge agent). The photo in the center depicts her drawing of one of the boxes. Photos were presented successively from left to right with a 2 second delay between the appearance of each photo, with the top photo presented last. All photos remained on screen until the end of the trial. The faces have been blurred in this image per University of Delaware policy but were not blurred for the experiment. 56

Figure 6: Results for Experiments 4, 5, and 6. The horizontal line indicates chance performance (0.5). Error bars indicate +/- 1 standard error. An example drawing from the Less-Informative condition is given for each experiment. 59

Figure 7: The arrangement of the 3x10 array (Panel 1a) on a sample adult participant. 93

Figure 8: An example trial from each of the two conditions: Perspective-Taking, Scalar-Inference. In each trial, participants saw each photo on the screen appear (with a 2s delay in between) then heard the sentence indicated and were expected to select one photo (correct answer highlighted here). The photo on the left within each image depicts the “full-access” box (and the full-knowledge speaker). The photo on the right depicts the “limited-access box” (and the limited-knowledge speaker). The faces have been blurred in this image per University of Delaware policy but were not blurred for the experiment. 95

Figure 9: Performance in the behavioral individual differences battery tasks for passer and non-passer scalar implicature groups. Error bars display +/- 1 SE. 103

Figure 10 Activation patterns in the Scalar Implicature task for adults, child passers, and child non-passers. Activation is shown for HbO for adults and HbR for children (HbR for adults and HbO for children can be found in supplementary materials). 107

Figure 11: Comparison of child passer subgroups in the Scalar Implicature task. Group 1 displayed activation in the dorsomedial PFC and dorsolateral PFC, similar to adults. Group 2 displayed no cohesive group patterns of activation. 111

Figure 12: Neural activation patterns for adults, child scalar-implicature passers and child scalar-implicature non-passers in the Theory of Mind neuroimaging task with an HbO contrast. Adults displayed significant activation in the ventral pre-frontal cortex. Child scalar-implicature passers displayed significant activation in the dmPFC. Child scalar-implicature non-passers did not display any significant group patterns of activation (HbO or HbR). 115

Figure 13: Conjunction of activation for the Scalar-Inference > Perspective-Taking contrast (scalar inference derivation) in the Scalar Implicature task and activation for the Mental > Physical contrast (mentalizing) in the Theory of Mind task for adult participants. 118

Figure B.1: An example trial from the Control condition. In each trial, participants saw each photo on the screen appear (with a 2s delay in between) then heard the sentence indicated and were expected to select one photo (correct answer highlighted here). The photo on the left within each image depicts the “full-access” box (and the full-knowledge speaker). The photo on the right depicts the “limited-access box” (and the limited-knowledge speaker). The faces have been blurred in this image per University of Delaware policy but were not blurred for the experiment. 153

Figure B.2: Timing for presentation of stimuli. Participants completed 2 runs, consisting of 6 blocks (2 Perspective-Taking, 2 Scalar-Inference, 2 Control) with a 15 second rest before each block. 154

Figure B.3: Timing for presentation of Theory of Mind stimuli for one story. Participants completed 2 (children) or 4 (adults) runs, consisting of 4 stories (2 Mental, 2 Physical) with a 12 second fixation before each story. Each story was followed by a probe sentence after which participants indicated their response. 154

Figure B.4: Activation patterns in the Scalar Implicature task for adults, child passers, and child non-passers. Activation is shown for HbR for adults and HbO for children. 155

Figure B.5: Neural activation patterns for adults, child scalar-implicature passers and child scalar-implicature non-passers in the Theory of Mind neuroimaging task with an HbR contrast 156

ABSTRACT

Human communication relies on the ability to assess what others intend to communicate beyond what they directly say. Under classic theories of communication, a critical component of this process is the evaluation of what another person knows (epistemic reasoning) and how that knowledge will affect the value of what they communicate. For example, a sentence such as "Some tires need to be replaced after 50,000 miles" can be interpreted as giving rise to the inference "Some but not all tires need to be replaced after 50,000 miles" when uttered by an expert. However, the same inference does not arise when the sentence is spoken by a 16-year-old driver with limited knowledge about tires. Adults have been shown to adjust the inferences they make during communication in accordance with the knowledge state of the speaker, but in young children this ability is debated. This dissertation explores the role that epistemic reasoning plays in the development of pragmatic inferences in both linguistic and non-linguistic communicative contexts using behavioral and neural approaches.

In a series of studies, we investigate whether young children are capable of making sensitive adaptions to the epistemic state of a communicator during pragmatic inference. We use variations on a novel visual world paradigm in a carefully chosen set of contexts to explore how young children develop scalar inference competence and how that development interacts with the development of epistemic reasoning. First, we focus our attention on 4- and 5-year-old children to address existing debates over the role of epistemic reasoning in early derivation of scalar implicatures. Next,

we examine how expectations of informativeness and other communicative principles extend to forms of ostensive communication using non-linguistic symbols. Finally, we employ neuroimaging methods to look through a different lens at the relationship between pragmatic development and epistemic development. These finding contribute not only to our understanding of scalar implicatures, but also to the broader fields of pragmatics, theory of mind, symbolic development, and cognitive neuroscience.

Chapter 1

INTRODUCTION

Language provides a direct medium for people to interact and communicate with others, but just as often, language allows people to communicate through what is left *unsaid*. If one goes to a zoo, and the zookeeper says, “Some of the animals are safe to pet,” one must be able to interpret this correctly as “Some but not all of the animals are safe to pet” in order to enter and leave the zoo with the same number of limbs. These intended meanings are rarely, if ever, spoken out loud, and require extensive computations about the speaker’s characteristics and motivations, yet even adults with wildly different developmental backgrounds interpret them successfully and consistently. Given that children are unlikely to ever hear implied meanings spoken out loud, how do children develop the ability to make these inferences?

According to an influential model of communication (Grice, 1975), a listener is able to infer a speaker’s intended meaning by utilizing mutually available principles of communication that stem from expectations of rational behavior. Rational communicators are expected to be cooperative to further the goals of a conversation (the Cooperative Principle), and this requires adherence to certain communicative principles, or maxims, such as the Maxim of Quantity. The Maxim of Quantity states that a communicator should make their contribution as informative as required by the purposes of the exchange (and no more and no less). In order for communication to be successful, both participants need to cooperatively give as much information as is required of them, and they assume that their conversational partner will do the same.

Thus, when a speaker appears to violate this communicative principle, such as by failing to give sufficient information, a listener might interpret their statement as having a non-literal intended meaning (to repair an otherwise unsuccessful contribution). For example, given that the statement “All of the animals are safe to pet” would have given greater relevant information if true (information about all of the animals instead of just some of them), the listener draws the conclusion that either 1) (the speaker knows that) “All of the animals are safe to pet” is not true, or 2) the speaker does not know whether this more informative statement is true. If the listener is able to conclude that the speaker is knowledgeable about the animals’ docility, as presumably a zookeeper would be, then the listener will interpret “Some of the animals are safe to pet” to mean “Not all of the animals are safe to pet.” This inference is referred to as a scalar implicature.

This dissertation will focus specifically on the processes and developmental phenomena that characterize scalar implicature derivation, but the scope of these findings and the questions they raise extend to other pragmatic phenomena and communication much more broadly. These studies will explore not only the inferences we make when communicating with others, but also how our other cognitive abilities, such as theory of mind, inform and interact with those processes. We will go beyond language as a communicative medium to explore how children and adults interpret and derive inferences from symbols, such as pictures or drawings, and the contributions that epistemic reasoning makes to that process. Finally, we will look at how these cognitive systems interact online during the derivation process as well as how that interaction develops in early childhood from both a behavioral and neural perspective.

1.1 Scalar Implicatures

As outlined above, scalar implicatures are a type of pragmatic inference in which a listener computes a non-literal meaning for a statement such as "Some of the animals are safe to pet" and derives the inference that **not all** of the animals are safe to pet. Scalar implicatures are so called because they rely on the evaluation of linguistic alternatives that form a scale ordered in terms of logical strength (Horn, 1972, 1989). In this context, "All of the animals are safe to pet" is the 'stronger' alternative because it asymmetrically entails "Some of the animals are safe to pet." If "All of the animals are safe to pet" is true, then "Some of the animals are safe to pet" is necessarily true, but not vice versa. Scalar implicatures are derived through the generation and negation of stronger scalemates and in some cases require the listener to have lexical knowledge of a given scale (e.g., <some, all>). Much of the literature on scalar implicatures has focused on scalar implicatures derived from quantifiers, but scalar implicatures can also be derived from other types of logically ordered scales, such as modals or numbers (Horn, 1989; Levinson, 2000). Scalar implicatures can even be derived from contextually-determined alternatives; these implicatures are called "particularized" or "ad-hoc" scalar implicatures (Hirschberg, 1985). For example, if a tiger, a llama, and a goat are in pens, and a zoo visitor is told, "The llama and the goat are safe to pet," the implication arises that the tiger is not safe to pet. Following the previous logic, given that the statement "The tiger, the llama, and the goat are safe to pet" would have given greater relevant information if true, the listener infers through the choice of the less-informative statement that the tiger is not safe to pet.

Adults successfully derive scalar implicatures, although it has been argued that scalar implicature derivation requires cognitive processing resources (de Neys & Schaeken, 2007; Marty, Chemla, & Spector, 2013) and might involve lengthier processing time compared to literal sentences (Huang & Snedeker, 2009a; although see Grodner, Klein, Carbary, & Tanenhaus, 2010; Foppolo & Marelli, 2017; cd. also Degen & Tanenhaus, 2015). Children, however, often struggle to grasp the subtleties of these hidden intended meanings (Noveck, 2001; Chierchia, Crain, Guasti, Gualmini, & Meroni, 2001; Papafragou & Musolino, 2003; Guasti, Chierchia, Crain, Foppolo, Gualmini & Meroni, 2005; Huang & Snedeker, 2009b; Foppolo, Guasti & Chierchia, 2012; among others): for instance, 5-year-olds – unlike adults – judge that the sentence "Some of the horses jumped over the fence" is a good description of an event in which all of the horses jumped over the fence, even though the sentence is underinformative (Papafragou & Musolino, 2003). Children's performance improves in paradigms in which they are given more nuanced response choices beyond a binary judgement (Katsos & Bishop, 2011), when relevant scalar alternatives are made more easily accessible (Barner, Brooks, & Bale, 2011; Stiller, Goodman, & Frank, 2015; Skordos & Papafragou, 2016), or when task demands are simplified (Pouscoulous, Noveck, Politzer, & Bastide, 2007).

1.2 Epistemic Reasoning in Scalar Inference

A prominent debate in the field centers on the nature of the computational process required to derive a scalar implicature. According to classic Gricean and other inferential accounts, scalar implicature derivation is a complex social process during which the listener relies on their understanding of the speaker's goals, characteristics, and motivations to interpret what the speaker intended to communicate (see, e.g., Horn, 1972; Sperber & Wilson, 1986; Carston, 1998; Sauerland, 2004, 2012; Geurts, 2010). Under this approach, assessing the speaker's epistemic state and the information to which they have access is a requisite component of this process. In contrast, other formal linguistic accounts have proposed that scalar implicature derivation could occur through entirely linguistic means wherein the literal interpretation of an utterance is strengthened through a grammatical operator without necessarily appealing to the characteristics of the speaker (see e.g., Chierchia 2004; Chierchia, Fox, & Spector, 2009, 2012; Hochstein, Bale, Fox, & Barner, 2016; Barner, Hochstein, Rubenson, & Bale, 2018; Hochstein, Bale, & Barner, 2018).

While there is some evidence to indicate that adults do consult the epistemic state of the speaker while interpreting scalar implicatures (Bergen & Grodner, 2012; Breheny, Ferguson & Katsos, 2013; Fairchild & Papafragou, 2018), existing evidence would suggest that children below the age of 5 are not capable of incorporating the speaker's epistemic state into their interpretation of scalar implicatures, and even 5-year-olds sometimes struggle (Hochstein, Bale, Fox, & Barner, 2016; Barner, Hochstein, Rubenson, & Bale, 2018; Papafragou, Friedberg & Cohen, 2018). Under a Gricean approach, listeners must evaluate the knowledge state of the speaker to derive a pragmatic inference, thus children should adapt to a speaker's epistemic state from their earliest demonstrations of scalar inference. Yet, in existing research, the ability to sensitively adapt to the epistemic state of a speaker appears to not emerge until roughly a year and a half after children's earliest successes with scalar implicature (Stiller, 2015). Is this gap due to a lack of suitable, sufficiently non-demanding paradigms testing children this age, or is this gap an important deficiency that chips away at the consequence of epistemic reasoning in scalar implicature derivation? Critical evidence for Gricean theories that young children consider and adapt to a speaker's epistemic state is missing, and this discrepancy needs to be reconciled. In

Chapter Two of this dissertation, we investigate 4- and 5-year-olds' epistemic sensitivity during scalar inference using a novel paradigm to address these questions.

1.3 Epistemic Reasoning in Scalar Inference in Non-Linguistic Communication

While Grice's body of work has been highly influential in the field of pragmatics, the role that scalar implicatures play in linguistic communication is only one piece of the puzzle. Grice theorized that the communicative principles and maxims that underlie linguistic communication should in fact extend to all forms of ostensive and intentional communication (see also Sperber & Wilson, 1986). He proposed that the Maxim of Quantity should apply to non-linguistic exchanges; for example, an assistant helping with the repair of a car is expected to hand the mechanic exactly the quantity of screws that he needs. Although Grice's discussion centered on the interpretation of non-linguistic actions, in theory the same principles should apply to non-linguistic symbols, such as pictures and drawings. However, this direction remains relatively unexplored, and relevant empirical evidence is needed to support these claims.

The first question of interest is whether children and adults do expect pictures, like utterances, to be informative in accordance with the expectations of the exchange. Pictures are an appropriate form of non-linguistic symbol to test because children both create and interpret pictures from an early age, and there is evidence to suggest that even young children understand that pictures can be used communicatively (DeLoache 1987, 1991; Preissler & Bloom, 2008; Allen, Bloom, & Hodgson, 2010). If it can be established that expectations of informativeness apply to pictures, the next question is whether adults and children consider the epistemic state of a picture's creator while evaluating the picture's informativeness, as has been theorized for linguistic forms of communication. We explore these questions in Chapter 3.

1.4 Epistemic Reasoning in Scalar Inference From a Neural Perspective

Scalar implicatures have been extensively studied in the developmental literature, but debates continue over the exact developmental timeline for scalar inference. Early studies in the literature reported failures to derive scalar implicatures until late in development (Noveck, 2001) while more recent studies have reported fairly early successes (Stiller, Goodman, & Frank, 2015), but even recent studies have shown that minor shifts in behavioral paradigms can often result in dramatic effects on children's performance depending on the type of scalar implicature, the methodology used, and the context (Pouscoulous, Noveck, Politzer, & Bastide, 2007; Katsos & Bishop, 2011; Barner, Brooks, & Bale, 2011; Foppolo, Guasti & Chierchia, 2012; Stiller, Goodman, & Frank, 2015; Skordos & Papafragou, 2016).

In addition to scalar implicatures, a number of other cognitive processes (the most relevant being theory of mind, executive functions, and language), are undergoing critical development within the same time period (between the ages of 3 to 5). It has been theorized that each of these systems could play a role in scalar implicature derivation, yet existing studies have often failed to find a correlation between skills in these areas with early scalar implicature abilities. Behavioral approaches to the study of scalar implicature development can only go so far in settling these debates. In Chapter 4, we present a novel approach to address these issues: using spatial neuroimaging, we investigate the role that theory of mind, executive functions, and language play in the derivation of scalar inferences for adults and young children.

1.5 Overview

The primary focus of this dissertation is to assess the role of epistemic reasoning in scalar implicature derivation and development. The body of this dissertation contains 4 sets of experiments designed to look at how young children incorporate epistemic reasoning into scalar inference derivation (Ch. 2), how the role of epistemic reasoning extends to scalar inference in non-linguistic communication (Ch. 3), and how theory of mind and other systems interact with scalar inference online and during development from a neural perspective (Ch. 4). Each set of experiments was prepared independently for journal publication, and to a large extent each manuscript remains intact (this can account for some redundancy between the three chapters, although this has been minimized). First, in Chapter 2, we test 4- and 5-year-olds' sensitivity to epistemic state during scalar inference using a novel visual world paradigm[1]. Then, in Chapter 3, we adapt this paradigm to investigate whether 4-year-olds, 5-year-olds, and adults display this same epistemic sensitivity to scalar inferences that rely on non-linguistic symbols. Next, in Chapter 4, we adapt our paradigm from Chapter 2 to spatial neuroimaging methods to ask whether theory of mind networks are active during online comprehension of scalar inference in young children and adults. We also explore how theory of mind development interacts with scalar inference development from a neural perspective. Finally, in Chapter 5, we

---

[1] This manuscript was recently published in Developmental Science (Kampa & Papafragou, 2020) and will be referred to as such in Chapters 3 and 4.

summarize and discuss the findings of these studies on scalar implicature development to clarify the insights we've gained and identify the areas that still have yet to be explored.

Chapter 2

EPISTEMIC REASONING IN SCALAR INFERENCE IN LINGUISTIC COMMUNICATION

Having a conversation with someone often requires the ability to go beyond what is said to interpret what is meant. For example, if at a garage sale someone says, “Some toys on this table are free,” an adult knows to interpret this as “Some but not all toys on this table are free.” According to Grice (1975; 1989), adults can compute these non-literal meanings because they assume that conversational partners must work together to further the goals of the conversation (the Cooperative Principle). To do so, communicators must obey a set of conversational principles, or maxims. The above example relies on the Maxim of Quantity, which states that a speaker in a conversation should give as much information as needed. Given that the statement “All toys on this table are free” would have given greater relevant information if true (and since the exchange is presumed to be cooperative), the listener infers that the choice of the less informative statement implies that the speaker is not in a position to offer the more informative statement - either because she does not know for a fact that such a statement is true, or more likely (if she is the person in charge of the garage sale) because she knows that it is not true. This inference is known as a scalar implicature.

Scalar implicatures rely on a comparison of a lexical item to a stronger alternative that the speaker could have used but did not (Grice, 1989). In the garage sale example, the lexical alternatives “some” and “all” form a scale ordered in terms of logical strength (Horn, 1972, 1989). In order to compute an implicature, the listener must be able to access this logically ordered scale (in addition to quantifiers, modals and numerals are other examples of logical scales; Horn, 1989; Levinson, 2000). In other cases, scalar implicatures rely on ad-hoc scales that depend upon context-specific information (Hirschberg, 1985). If at the garage sale of the previous example there is a teddy bear, a yo-yo, and a doll on the table, and the seller says “The teddy bear and the yo-yo are free,” the listener derives the inference that the doll is not free, following the same reasoning process about alternatives outlined above.

An open issue in the field is the extent to which children are able to both appraise and sensitively adapt to information about the speaker during scalar pragmatic reasoning. Under broadly Gricean accounts, the hearer needs to engage in a rich computational process that takes into account the speaker’s communicative intentions given his/her goals and knowledge state (see, e.g., Horn, 1972; Sperber & Wilson, 1986; Carston, 1998; Sauerland, 2004, 2012; Geurts, 2010; see also Chierchia 2006; Chierchia, Fox, & Spector, 2012 for alternative accounts). Accordingly, Gricean approaches predict that hearers filter the interpretation of scalar terms through the speaker’s perspective: the use of a weaker scalar term such as *some* may lead to the inference that the speaker lacks stronger knowledge or (if the speaker is an expert, as in the earlier zookeeper example), the inference that the stronger term does not hold, as far as the speaker knows (Sauerland, 2004; Geurts, 2010). Evidence from both eye tracking (Breheny, Ferguson & Katsos, 2013) and reading times (Bergen & Grodner, 2012) during online comprehension suggests that adults take into account speaker knowledge during the derivation of scalar implicatures. In a recent demonstration, adult comprehenders judged underinformative sentences more leniently when the sentences belonged to non-native compared to native speakers, presumably because non-native speakers know less about the lexical resources in a language (Fairchild & Papafragou, 2018). For children, however, sensitivity to speaker knowledge is still an unsettled topic.

Available experimental evidence suggests that children are not able to incorporate information about a speaker's epistemic state into the derivation of scalar implicatures before the age of five, and even then performance is highly fragile (Hochstein, Bale, Fox, & Barner, 2016; Barner, Hochstein, Rubenson, & Bale, 2018; Papafragou, Friedberg & Cohen, 2018). In a study by Papafragou and colleagues (2018), children were presented with videos of two observers, one with full knowledge of an event that occurred (e.g., a girl coloring a star) and one who fell asleep part way through. They then heard either a more informative or less informative statement describing the event (e.g., "The girl colored some/all of the star") and had to attribute the statement to either the limited-knowledge or the full-knowledge observer. According to the maxim of Quantity, only a knowledgeable observer had sufficient evidence to assert the more-informative statement, and only a speaker who lacked knowledge was justified in uttering a less-informative statement. Both 4- and 5-year-olds succeeded with more-informative statements, but only 5-year-olds performed above chance on less-informative statements (and even then, accuracy hovered around 60%). These results persisted when the quantifiers were replaced with ad-hoc scales. In a later manipulation which removed the epistemic component, 4-year-olds' performance improved. Similarly, in two recent studies (Hochstein et al., 2016; Barner et al., 2018), children were presented with two action figures, one blindfolded and one with full visual access. The action figures watched as an animal stole an object and announced “Look, an orange and a banana! Look what I’m taking!” and were asked what happened. Barner and colleagues (2018) found that 4-year-old children presented with disjunctive sentences (e.g., “He took an orange or a banana”) failed to compute non-scalar “ignorance” implicatures (i.e., they failed to match the disjunctive statement to the blindfolded speaker and thus to reason that only a speaker who lacks knowledge is pragmatically justified in producing a disjunctive statement), despite succeeding in a matched ad-hoc implicature task in which the epistemic aspect was removed. Barner et al. (2018) concluded that 4-year-olds have substantial difficulty with the epistemic aspects of implicature derivation and even potentially compute scalar implicatures in the absence of epistemic reasoning (cf. Hochstein, Bale, & Barner, 2018).

However, there may be another explanation for these failures. In the studies just reviewed, children had to watch lengthy scenarios while tracking the epistemic state of multiple speakers and the actions of multiple agents. These high task-specific cognitive processing demands may have masked 4-year-olds’ ability to incorporate epistemic state into scalar implicature derivation (see Papafragou et al., 2018 for discussion of this possibility, also Pouscoulous et al., 2007). Additionally, Barner and colleagues’ (2018) argument that 4-year-olds compute scalar implicatures in the absence of epistemic reasoning hinges on the assumption that the ignorance implicature task in which children fail is minimally different (and only in terms of epistemic computations) from the ad hoc scalar implicature task that children pass. However, the differences may not be minimal; children have been shown to struggle with disjunction (Singh, Wexler, Astle-Rahim, Kamawar, & Fox, 2016; although see Tieu, Romoli, Zhou, & Crain, 2015) and it may not be clear to children that the ignorant speaker (who is blindfolded) would have enough information to produce even a disjunctive statement ("He took a carrot or a banana"). Thus the question of whether children can incorporate speaker perspective into scalar implicature derivation needs to be examined further.

Children's sensitivity to speaker knowledge during pragmatic comprehension bears on broader debates over the extent to which children are able to flexibly adapt to the perspective of a communicative partner, with some commentators claiming that children remain egocentric and inflexible until late in development (Piaget, 1970; Shatz, 1980; Perner & Leekam, 1986; Girbau, 2001; Epley, Morewedge, & Keysar, 2004; Davies & Katsos, 2010), and others arguing that children display sophisticated adaptations to others at early ages (O'Neill, 1996; Nadig & Sedivy, 2002; Matthews, Lieven, Theakston, & Tomasello, 2006; Bahtiyar & Küntay, 2009; see Grigoroglou & Papafragou, 2018, 2019, for a review). Similarly, some researchers have argued that children rely on social reasoning of a Gricean character in word-learning (Clark, 1990; Baldwin, 1993a, 1993b; Akhtar, Carpenter, & Tomasello, 1996; Carpenter, Nagell, Tomasello, Butterworth, & Moore, 1998; Diesendruck, Markson, Akhtar, & Reudor, 2004; Southgate, Chevallier, & Csibra, 2010), while others have challenged the proposal that word learning requires Gricean reasoning (Preissler & Carey, 2005; Regier, 2005; de Marchena, Eigsti, Worek, Ono, & Snedeker, 2011).

Here, across three experiments, we asked whether 4- and 5-year-olds are able to assess and sensitively adapt to a speaker's knowledge state during pragmatic reasoning using a novel task. Our task, inspired by paradigms developed to study reference and other communicative phenomena in the visual world (e.g., Nadig & Sedivy, 2002), was designed to simplify the cognitive demands of previous tasks: it established the speaker's knowledge state through visual access to a simple, static display as opposed to a complex, dynamic event, and provided a clear conversational goal. Children were presented with either a more informative or a less informative statement about the contents of a box. The statements formed an ad hoc scale ordered in terms of logical strength: the more-informative statement (stronger alternative) mentioned both of the objects in a box, while the less-informative statement (weaker alternative) mentioned only one of them, and children had to decide which one of two possible boxes a speaker was describing. The speaker had complete visual access to the contents of one box (and hence complete knowledge) but only partial visual access to the contents of the second box (hence partial knowledge); crucially, the boxes did not differ from the participant's perspective. We expected, following the logic of Papafragou et al. (2018), that, if children and adults derive pragmatic inferences using a Gricean approach, they should use epistemic reasoning to link the more informative statement to the speaker under conditions of complete knowledge and the less informative statement to the speaker under conditions of partial knowledge. If children are inflexible or insensitive to the speaker's perspective, they would be expected to choose between the boxes at chance or rely on other potential biases (perhaps a preference for the fully-knowledgeable speaker; see, e.g., Hochstein et al., 2016; Ozturk & Papafragou, 2016). Evidence that young children and adults could both assess a speaker's epistemic state and use that assessment to interpret scalar statements would support a mentalistic model of implicature-computation and contribute to the broader conclusion that children incorporate the speaker's perspective when interpreting utterances in context.

In Experiment 1, we used this paradigm to ask whether 4- and 5-year old children consult a speaker's epistemic state in interpreting scalar statements. In Experiment 2, we extended this method to explore how explicit presentation of alternatives impacts the youngest children in our sample. Finally, in Experiment 3, we adapted this paradigm into a non-linguistic task to probe the mechanisms underlying epistemic pragmatic inference in children.

2.1 Experiment 1

2.1.1 Methods

2.1.1.1 Participants

Twenty-six 4-year-olds (M_{age} = 4;6, *range:* 4;0-4;11, 13 female) and twenty-six 5-year-olds (M_{age} = 5;5, *range:* 5;0-5;11, 17 female) participated. Children were recruited from Newark (DE) preschools and the Delaware Children's Museum in Wilmington, DE. An additional ten children contributed data but were excluded for failing preliminary tests (see Procedure for criteria). A control group of twenty-six adult participants was also tested. Adult participants were recruited with a HIT (Human Intelligence Task) posting on MTurk. No age or gender data were collected for this sample.

2.1.1.2 Procedure

For child participants, the experiment began with an introductory phase designed to familiarize them with the types of displays used in the main task. Participants were first presented with a single photo on a computer screen (Figure 1, panel 1). The photo depicted a girl sitting across from and facing out towards the participant. In front of her was a cardboard box with two vertical compartments. The compartment to the left (from the participant's viewpoint) was see-through and empty. The compartment to the right contained an object (a teacup) but was blocked so that only the participant but not the girl could see its contents.

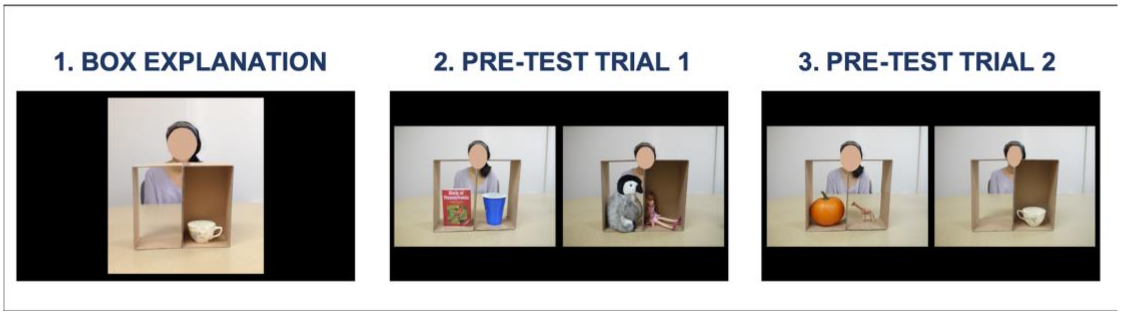

Figure 1: Visual displays for the introductory phase of Experiment 1. The faces have been blurred in this image per University of Delaware policy but were not blurred for the experiment.

The children were told that the girl in the photo was the experimenter's friend Danielle, and she had a special box in front of her. The children were then asked whether they themselves could look through the two parts of the box ("Can you look through this side?"), and if not, why not. (For the blocked compartment, the children typically said no and mentioned that it was closed or blocked.) Children were asked if they thought Danielle could look through the blocked side ("So do you think she can look through that one?" <pointing>). The children were then asked whether Danielle knew about the item in the blocked compartment ("Does she know there's a teacup there?"). No feedback was given, but the response was recorded and used for exclusions, as will be explained later.

All participants then completed two pre-test trials (Figure 1, panels 2 and 3). In each trial, two photos were shown side by side. One photo showed Danielle and a limited-access box similar to the introduction phase; the other showed Danielle and a full-access box where both compartments were unobstructed. One photo appeared on screen, followed by the other one after 2 seconds, as the experimenter explained the task (presentation order/side of the screen for the limited access box photo was counterbalanced across participants). Participants were told that they were going to play a game with the boxes. The experimenter explained that Danielle would look at each box and, pointing to each side of the boxes, reminded children whether it was open so that Danielle could look through it or closed so that she could not. Participants were told that Danielle was going to talk about just one of the boxes, and they had to decide which box she was talking about. After this explanation ended (and both photos had stayed up for 2 seconds), children heard a recorded sentence from a female speaker and were asked, "Which box is she talking about?" Both photos remained on screen until children responded. The two boxes had different contents, so the sentence unambiguously described one of them. For instance, in pre-test trial 1, the full-access box had a book and a cup, the limited-access box had a penguin and a doll, and the sentence was "I see a book and a cup". In pre-test trial 2, the full access box had two objects, but the limited-access box only had an object in the blocked compartment, and nothing in the open compartment. The sentence was: "I see nothing". One 4-year-old and two 5-year-olds failed both pre-test trials and were excluded.

Participants then completed 8 test trials. The set-up was the same as the pre-test trials. However, the contents of the boxes within a trial were identical (e.g., a spoon and a bowl; Figure 2), and only one of the objects in the limited-access box was visible to the experimenter's friend (here, the spoon). There were two within-subject conditions (More-Informative, Less-Informative) depending on whether children heard a more-informative statement mentioning both objects (e.g., "I see a spoon and a bowl") or a less-informative statement mentioning only one object – the one that remained visible to the speaker in the limited-access box but was actually present in the other box too (e.g., "I see a spoon"). Participants were again asked, "Which box is she talking about?"

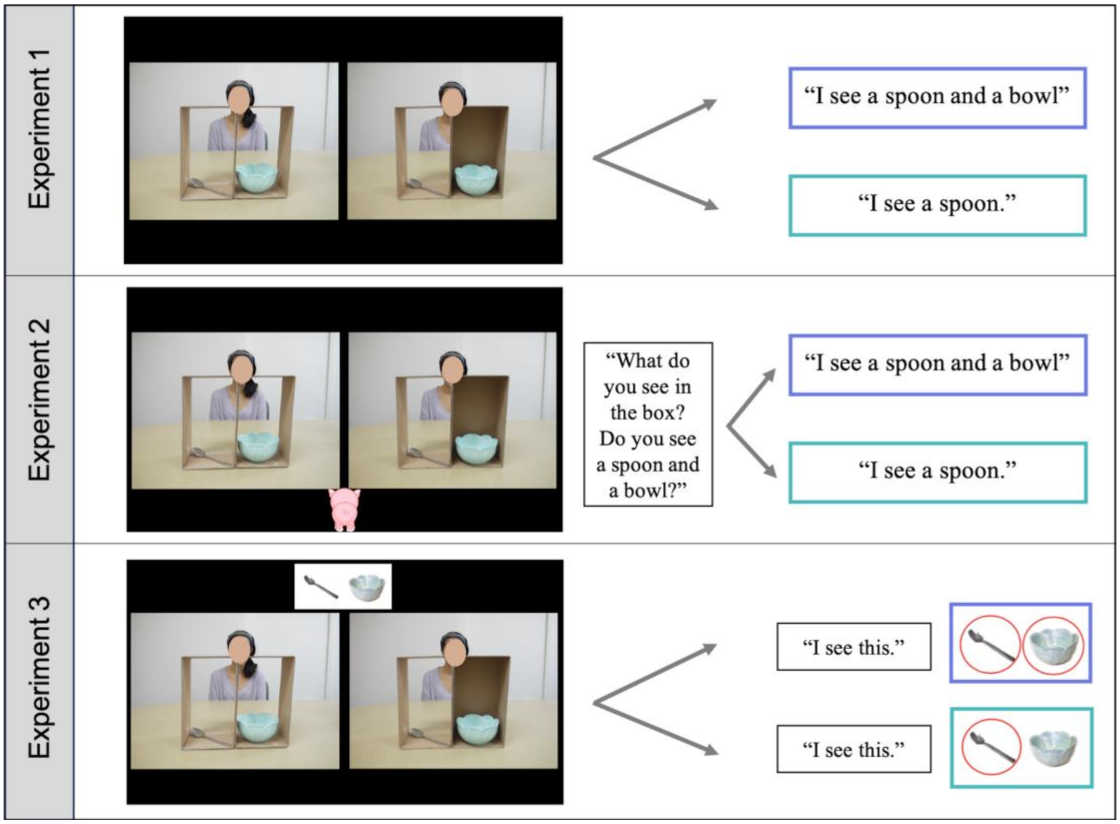

Figure 2: A schematic view of the visual and verbal stimuli for a test trial in Experiments 1, 2, and 3. Across all studies, the photo on the left depicts the “full access” box (and the full-knowledge person). The photo on the right depicts the “limited access box” (and the limited-knowledge person). In the More-Informative condition, participants heard a statement mentioning two objects (Experiments 1 and 2) or saw the action of circling two objects (Experiment 3), and were expected to select the full access box. In the Less-Informative Condition, participants heard a statement mentioning one object (Experiments 1 and 2) or saw the action of circling one object (Experiment 3), and were expected to select the limited access box. The faces have been blurred in this image per University of Delaware policy but were not blurred for the experiment.

For test trials 1-4, the girl in the photos was Danielle; for trials 5-8, the girl in the photos was Julia, another friend of the experimenter’s that was introduced for variety (the pre-recorded statements had a different female voice for Julia). Participants were given 4 more-informative and 4 less-informative test sentences in a mixed order, always beginning with a more-informative one. This trial order was chosen given previous evidence that children benefit from trial orders which present stronger (more-informative) relevant alternatives first (Skordos & Papafragou, 2016; see also Barner et al., 2011; Stiller et al., 2015). Two presentation lists were created; assignment of statements corresponding to each condition (More-Informative, Less-Informative) to trials was counterbalanced across lists. The position of the limited access box photo across trials was also counterbalanced within each list. Statements had neutral stress (i.e., “I see a spoon” and not “I see a SPOON”).

To introduce Julia, and to remind children of the critical properties of the boxes, after the first 4 test trials, children were presented with a reminder trial and questions modeled after the initial box introduction phase. The trial included a photo of a new friend, Julia, and a limited-access box with only one object (a blue plastic cup in the blocked compartment). As in the box introduction phase, children were asked whether they could look through each side of the limited-access box, and whether Julia could look through the blocked side (for the importance of such reminders of the visual properties of the display, see Nadig & Sedivy, 2002). No feedback was given. We also asked whether Julia knew about the contents of the blocked compartment (“Does she know there’s a cup there?”). Children who replied “yes” to this question in both the box-introduction and this reminder phase were excluded on the basis that they did not understand the nature of the limited-access box (N = 6 4-year-olds, 1 5-year-old).

2.1.2 Predictions

Two patterns should emerge if participants could successfully incorporate the perspective and knowledge of the speaker into their responses on test trials, in line with a Gricean model of pragmatic inference. In the More-Informative condition, they should take the statement (e.g., "I see a spoon and a bowl") to describe the full access box (and thus belong to the fully-knowledgeable speaker), because the girl could not see the bowl in the limited-access box. In the Less-Informative condition, they should take the statement (e.g., "I see a spoon") to describe the limited-access box (and thus belong to the limited-knowledge speaker) because it would be under-informative under these circumstances for the full-knowledge speaker to only mention one object instead of both. Notice that these responses on both conditions presuppose sensitivity to speaker knowledge since from the participant's perspective both boxes have identical, visible contents. If children failed to consider the speaker's perspective, they were expected to respond at chance or rely on other potential biases (e.g., a preference for knowledgeable speakers; see also Hochstein et al., 2016; Ozturk & Papafragou, 2016).

2.1.3 Coding

Correct responses involved selecting the full-access box in the More-Informative condition and the limited-access box in the Less-Informative condition. Participants were given a mean accuracy score between 0 and 1 for each condition. In the critical Less-Informative condition, most participants (72 out of 78; 92.3%) had scores of either 0 or 1, and an additional 4 (5.1%; combined 97.4%) had scores of either .25 or .75. Thus, for each condition, participants with a mean score equal to or greater than 0.75 were designated as passers and those with a mean score equal to or lower than 0.25 as failers. Any participant with a score of .50 was also designated as a failer, even though this response type was vanishingly low across all studies.

2.1.4 Results

A logistic mixed effects model with Condition (More-Informative, Less-Informative) and Age (4, 5, Adult) as fixed variables and with subject and item as random factors could not converge because of the almost total lack of variance in the data (Eager & Roy, 2017). Non-parametric tests were thus conducted on the data.

Adults and 5-year-olds performed at ceiling in this task (see Table 1). Fisher's exact test revealed no significant differences in performance between adults and 5-year-olds in either condition. Even in 4-year-olds, an exact test of goodness of fit showed that the number of passers in both the More-Informative and the Less-Informative condition was significantly different from the expected ratio due to chance (More-Informative: p = .002, Less-Informative: p = .029). Furthermore, Fisher's exact test revealed no significant differences between 4-year-olds and either adults or 5-year-olds in the More-Informative condition (p = 1 for all comparisons; Bonferroni correction). The same test did indicate a significant difference between 4-year-olds and adults (p = .01), and a marginally significant difference between 4-year-olds and 5-year-olds (p = .05) in the Less-Informative condition. For 4-year-olds, performance on the More-Informative condition was marginally better compared to the Less-Informative condition (Fisher's exact test, p = .05, 2-tailed).

Table 1: Number of Passers and Failers in the More-Informative and Less-Informative Conditions for Experiment 1 for Adults, 5-year-olds, and 4-year-olds.

	Classification	Condition	
		More-informative	Less-Informative
Adults	Passers	26***	26***
	Failers	0	0
5-year-olds	Passers	26***	25***
	Failers	0	1
4-year-olds	Passers	25***	19*
	Failers	1	7

Note. Asterisks denote performance significantly different from the expected ratio of passers/failers due to chance ($p < 0.05$ = *, $p < 0.01$ = **, $p < 0.001$ = ***)

2.1.5 Discussion

Experiment 1 investigated whether 4- and 5-year-old children are able to incorporate epistemic reasoning into the derivation of scalar inferences. Results suggest that children are capable of this type of complex pragmatic reasoning; in a simple task using a novel paradigm inspired by referential communication tasks in the visual world, both 4- and 5-year-olds were able to successfully interpret statements of different informational strength in accordance with the knowledge state of the speaker. In fact, 5-year-olds performed at an adult-like level in this new task. The findings of the present study lower prior estimates of the age at which children display the ability to take the epistemic step during scalar inference derivation (Hochstein et al., 2016;

Papafragou et al., 2018) – and are consistent with other studies with very simple paradigms that show an ability to compute scalar inferences (without special attention to the epistemic component) at young ages (Stiller et al., 2015).

The present data raise two possible issues. First, there are different varieties of scalar implicature depending on the specificity of the alternative that is being negated (see Geurts, 2010). For example, a statement such as "I see a spoon" could be interpreted to mean either "I see a spoon and nothing else" or "I see a spoon and not a bowl", depending on whether the speaker had access to the more-informative alternative (among other possible factors). In our task, the limited-knowledge speaker did not know the context-specific more-informative alternative (although the child did), thus the less-informative statement could only be interpreted as conveying a "nothing else" implicature (see the adult data in Breheny et al., 2013 for this type of inference). However, in most prior developmental studies where 4-year-old children derived ad-hoc scalar implicatures successfully (Barner et al., 2011; Barner et al., 2018; Papafragou & Tantalou, 2004), the speaker had full access to the contextual set of alternatives and thus the more-informative alternative could be taken to be negated. It would be valuable to set up a context with explicitly presented contextual alternatives to better compare 4-year-olds' inferences to previous tasks. Second, while adults and 5-year-olds performed at ceiling, 4-year-olds did not; the majority displayed very successful pragmatic inference, but some 4-year-olds relied on some other strategy (e.g., choosing the full-knowledge speaker for all trials). Further examination of performance within this age group could provide a clearer picture of pragmatic development. In Experiment 2, we addressed both of these issues.

2.2 Experiment 2

In Experiment 2, we replicated Experiment 1 but introduced a conversational context such that the test sentences ("I see a spoon and a bowl", "I see a spoon") were now the response to a question from a conversational partner ("What do you see in the box? Do you see a spoon and a bowl?"). By providing a response to an overt question under discussion (QUD; see Roberts, 1996, 2004; cf. Stalnaker, 1979) that explicitly invoked the stronger alternative, we increased the likelihood that the speaker would be taken to negate a specific stronger alternative (i.e., "I see a spoon and not a bowl"). Furthermore, we focused exclusively on 4-year-old participants and divided them into younger and older sub-groups to gain a clearer picture of scalar implicature development at this age.

2.2.1 Methods

2.2.1.1 Participants

Forty 4-year-olds contributed data; they were divided into a younger group (N = 20, M_{age} = 4;3, *range:* 4;0-4:5, 13 female) and an older group (N = 20, M_{age} = 4;9, *range*: 4;6-4:11, 15 female). Children were recruited from Newark (DE) preschools and the Delaware Children's Museum in Wilmington, DE. An additional six children participated but were excluded from data analyses for failing to meet inclusionary criteria (see Procedure). A control group of twenty adult participants (M_{age} = 29.1, *range:* 22-35, 12 female) was also tested. Adult participants were recruited with a HIT (Human Intelligence Task) posting on MTurk.

2.2.1.2 Procedure

Experiment 2 used the same stimuli and basic procedure as Experiment 1 but with the addition of a new character (as well as some smaller changes detailed below). As before, during the box introduction phase, participants were familiarized with the limited-access box, and were told that they would play a game with boxes. For the pre-test trials, they were shown two photos with Danielle and the limited-access vs. full-access box. They were told that Danielle would look at both boxes, "pick" one of them, and they would have to "figure out which box she picked." At that point, the experimenter said, "Let's see if someone can help us!" A small clipart pig appeared below the photos in the middle of the screen. The experimenter greeted the character ("Hello, Wilbur!") and Wilbur (through a pre-recorded message) responded "Hello." The experimenter then asked: "Wilbur, can you ask Danielle some questions to help us figure out which box she picked?" Wilbur agreed, then turned to face the pictures, as if talking to Danielle, and asked her, "Danielle, what do you see in the box? Do you see an X [name of object]?" followed by Danielle's response. Participants were asked, "Which box did she pick?" and had to point at a box.

For test trials, the procedure was the same, but Wilbur's question always mentioned two objects (e.g., for the trial in Figure 2: "Danielle, what do you see in the box? Do you see a spoon and a bowl?"). The original sentences from Experiment 1 with less-informative and more-informative test sentences (uttered in 'Danielle's' and 'Julia's' voices) were used.

The pre-test trials differed slightly from Experiment 1. As before, the girl's responses unambiguously described one of the boxes. However, unlike Experiment 1, there was only one object in each box (not two), always appearing in the open compartment. In pre-test trial 1, the full-access box had a spoon, the limited-access box had a penguin, and the exchange was: “Danielle, what do you see in the box? Do you see a penguin?” - “I see a penguin”. In pre-test trial 2, the full-access box had a book, the limited-access box had a box, and the exchange was: “Danielle, what do you see in the box? Do you see a basket?” - “I see a book”. The structure of these pre-test trials, and Wilbur’s incorrect guess in pre-test trial 2, was set up to prevent participants from associating an answer that matched or mismatched the wording of the question with a specific box. Given that in the test trials, participants should select the *full-access box* when the answer matched the question (e.g., “…Do you see a spoon and a bowl?” “I see a spoon and a bowl”), in pre-test trial 1, the correct answer was to select the *limited-access box* when the answer matched the question. Inversely, given that in test trials, participants should select the *limited-access box* when the answer mismatched the question (e.g., “…Do you see a spoon and a bowl?” “I see a spoon”), in pre-test trial 2, the correct answer was to select the *full-access box* when the answer mismatched the question. Participants were excluded for failing the pre-test trials (N = 1 younger 4-year-old) or not completing the task (N = 1 younger 4-year-old).

We also made a slight change to the exclusion criteria in Experiment 2. The children who incorrectly replied during the box introduction phase that the girl knew the identity of the object in the blocked compartment were now given corrective feedback (“No, I don’t think she knows there’s a teacup there because she can’t look through that side; it’s blocked”), and were only excluded if they replied incorrectly again at the reminder trial midway through the test phase (N = 4 younger 4-year-olds). The majority of 4-year-olds (38 out of 46) understood the perspective and knowledge requirements of the box in the photo without needing feedback.

2.2.2 Coding

Again, the majority of participants (49 out of 60; 82%) had scores of either 0 or 1 in the critical Less-Informative condition, and an additional 9 (15%; combined 97%) had a score of .75, so coding was kept identical to Experiment 1.

2.2.3 Results

Adults and both groups of 4-year-olds performed well in both the More-Informative and Less-Informative conditions (see Table 2). Fisher's exact test comparisons did not reveal any significant differences on individual conditions between adults and older 4-year-olds (More-Informative: $p = 1$, Less-Informative: $p = .49$), or adults and younger 4-year-olds (More-Informative: $p = .60$, Less-Informative: $p = .23$). Additionally, there were no significant differences between the two sub-groups of 4-year-olds per condition (More-Informative: $p = .61$, Less-Informative: $p = .41$). Further analyses (exact test of goodness of fit) revealed that the number of passers per condition was significantly different from the expected ratio due to chance for all age groups (adults and older 4s: $p < .001$ for all comparisons; younger 4s: More-Informative: $p = .002$, Less-Informative: $p = .041$). Finally, Fisher's exact test showed that performance did not differ between the More-Informative and Less-Informative condition for either the older ($p = 1$) or the younger 4-year-olds ($p = .70$).

Table 2: Number of Passers and Failers in the More-Informative and Less-Informative Conditions for Experiment 2 for Adults, Older 4-year-olds (4;6-4;11), and Younger 4-year-olds (4;0-4;5).

	Classification	Condition	
		More-Informative	Less-Informative
Adults	Passers	20***	20***
	Failers	0	0
Older 4s	Passers	19***	18***
	Failers	1	2
Younger 4s	Passers	17**	15*
	Failers	3	5

Note. Asterisks denote performance significantly different from the expected ratio of passers/failers due to chance ($p < 0.05$ = *, $p < 0.01$ = **, $p < 0.001$ = ***)

2.2.4 Discussion

Experiment 2 replicated the finding of Experiment 1: 4-year-olds can incorporate speaker knowledge state into the derivation of scalar inferences. By recruiting greater numbers within this age group, we could see that even young 4-year-olds successfully use pragmatic principles to attribute statements of varying logical strength to speakers in accordance with their knowledge state. These results suggest that preschool children are able to derive scalar inferences with epistemic sensitivity both in cases where the speaker's statement ("I see a spoon") is taken as negating a specific stronger alternative ("I see a spoon and not a bowl"; Experiment 2), and in cases where the speaker's statement is taken as conveying a "nothing else" implicature ("I see a spoon and nothing else"; Experiment 1).

Could young children be relying on other strategies to succeed in Experiments 1 and 2 instead of computing genuine scalar inferences? One possibility could be that children used a superficial mapping strategy to succeed: given children's strong performance in the More-Informative condition, they could have correctly reasoned about the full-access box for the More-Informative trials, and then just selected the other option for the Less-Informative trials. Another possibility is that children could select the correct boxes in the Less-Informative condition on the basis of probability: there is a higher probability that the speaker who sees only one object will mention that object compared to the other speaker. We believe that these possibilities are unlikely explanations for our findings, and we return to them in the discussion of Experiment 3.

2.3 Experiment 3

In Experiments 1 and 2, we found robust evidence that 4-year-olds are able to adapt to speaker knowledge during pragmatic inference. This raises questions about whether this ability extends to other forms of pragmatic inference. Grice (1989) proposed that pragmatic principles, including the Maxim of Quantity, extend to other types of co-operative communication (see also Sperber & Wilson, 1985), but relevant evidence so far in children and adults is limited (Bass, Bonawitz, & Gweon, 2017; Papafragou et al., 2018; Gweon & Asaba, 2018).   In a previous study, 4-year-old children failed to incorporate speaker knowledge during scalar inference in a non-linguistic paradigm adapted from a linguistic paradigm; however, they also failed in the linguistic version (Papafragou et al., 2018). In Experiment 3, we revisit the question of whether children (and adults) adapt to speaker knowledge during pragmatic inference in non-linguistic exchanges using a simple task.

Experiment 3 was identical to Experiment 1, but instead of sentences, the speaker indicated which box she had chosen by circling the objects she saw in that box from a selection of pictures. Thus, instead of verbally mentioning either two objects ("I see a spoon and a bowl") or one ("I see a spoon"), she indicated what she saw by circling either two objects (a spoon and a bowl) or one (a spoon) from a display with pictures of possible objects. As in our previous experiments, if children (and adults) incorporated the perspective and knowledge state of the communicator, they should recognize that only the full-knowledge person had sufficient evidence to circle both a spoon and a bowl. Similarly, a pragmatic responder should recognize that only the limited-knowledge person was justified in circling only a spoon, as the full-knowledge person who saw both should circle both to disambiguate and be informative in accordance with pragmatic principles of communication.

2.3.1 Methods

2.3.1.1 Participants

Twenty-three 4-year-olds (M_{age} = 4;4, *range:* 4;0-4;11, 15 female) and twenty-three 5-year-olds (M_{age} =5;4, *range:* 5;0-5;11, 12 female) contributed data. Children were recruited from Newark (DE) preschools and the Delaware Children's Museum in Wilmington, DE. An additional group of 7 children participated but were excluded from data analyses for failing to meet inclusionary criteria (see Procedure). A control group of twenty-three adult participants (M_{age} = 28.7, *range:* 20-34, 10 female) was also tested. Adult participants were recruited with a HIT (Human Intelligence Task) posting on MTurk.

2.3.1.2 Procedure

Experiment 3 modified the stimuli and procedure of Experiment 1 as follows. Within each trial, the original pairs of photos were presented as in Experiment 1. As before, participants were told that they were going to play a game with the boxes where Danielle would look at both boxes, then select one; participants were told that they had to figure out which box she picked. Participants were told that the toys from the boxes would appear on the top of the screen, and Danielle would circle what she sees "in just one of the boxes." This was accomplished by presenting pictures of the objects in the boxes at the top of the screen placed against a white background (Figure 2). Participants heard an audio recording of Danielle saying, "I see this" while a red circle was slowly drawn around one or both of the objects. Participants were asked, "Which box did she choose?" For the More-Informative test trials, the girl circled both objects in succession. For the Less-Informative test trials, she circled only one (Figure 2).

The pre-test trials in Experiment 3 were slightly different from Experiment 1 in order to prevent participants from associating the spatial position (left/right) of the toys on top of the screen with the spatial position of the box photos. The photos for pre-test trial 1 were similar to Experiment 1; both boxes contained two unique objects. All four objects appeared as images in the white box above the photos (Figure 3, panels 2 and 3). The objects in the white box for both pre-test trials were arranged in a random order to prevent participants from associating the objects on the left of the white box with the photo on the left. In pre-test trial 2, one object (a cup) appeared in both boxes. Although the object appeared in both boxes, there was only one picture of a cup in the white box (to set the precedent for the test trials; see Figure 2).

Figure 3: Visual displays for the introductory phase of Experiment 3. At the end of each pre-test trial, participants heard "I see this" as an animation of a red circle appeared around two of the objects in the array above the photos in succession. In pre-test trial 1, the picture of the book was circled first, followed by the giraffe. In pre-test trial 2, the picture of the book was circled first, followed by the cup. The faces have been blurred in this image per University of Delaware policy but were not blurred for the experiment.

With the exceptions above, Experiment 3 used the same counterbalancing, trial order, and photos as Experiment 1. Participants were excluded for not completing the task (*N* = 2 4-year-olds), for saying "yes" when asked whether Danielle knows the identity of the object in the blocked compartment during both box-introduction and reminder trials (*N* = 2 4-year-olds, 1 5-year-old), or because of parental interference during the task (*N* = 1 4-year-old, 1 5-year-old).

2.3.2 Coding

A majority of participants (47 out of 69; 68%) performed at ceiling (1) or floor (0) in the Less-Informative trials, and an additional 17 participants (25%; combined total 93%) performed near ceiling (.75) or floor (.25). As a result, coding was kept identical to Experiments 1 and 2.

2.3.3 Results

Comparisons between age groups using Fisher's exact test did not reveal any significant differences in the More-Informative condition (see Table 3; adults vs. 4s: $p = 1$, adults vs. 5s: $p = .224$, 4s vs. 5s: $p = .489$). In the Less-Informative condition, the same test revealed a significant difference in performance only between adults and 4-year-olds ($p = .029$); there was no significant difference between adults and 5-year-olds ($p = .284$) or between 4-year-olds and 5-year-olds ($p = .231$). Fisher's exact test revealed a significant difference between the Less-Informative and More-Informative condition for 4-year-olds ($p < .001$) but not for 5-year-olds ($p = .135$) or adults ($p = .235$).

Table 3: Number of Passers and Failers in the More-Informative and Less-Informative Conditions for Experiment 3 for Adults, 5-year-olds, and 4-year-olds.

	Classification	Condition	
		More-Informative	Less-Informative
Adults	Passers	23***	21***
	Failers	0	2
5-year-olds	Passers	21***	16
	Failers	2	7
4-year-olds	Passers	23***	11
	Failers	3	12

Note. Asterisks denote performance significantly different from the expected ratio of passers/failers due to chance ($p < 0.05$ = *, $p < 0.01$ = **, $p < 0.001$ = ***)

Adults performed significantly differently from the expected ratio due to chance for both conditions in analyses using exact goodness of fit (More-Informative: $p < .001$, Less-Informative: $p < .001$). However, 5-year-olds and 4-year-olds did so only in the More-Informative condition (5s: More-Informative: $p < .001$, Less-Informative: $p = .093$; 4s: More-Informative: $p < .001$, Less-Informative: $p = 1$).

Performance on each condition was compared between Experiments 1 and 3 using Fisher's exact test (we collapsed across age groups in children for simplicity). For adults there was no significant difference (More-Informative: $p = 1$, Less-Informative: $p = .215$). For children performance was significantly better in Experiment 1 compared to Experiment 3 for the Less-Informative condition ($p = .006$) but was similar across experiments for the More-Informative condition ($p = .559$).

2.3.4 Discussion

As predicted by pragmatic theories (cf. Grice, 1989; Sperber & Wilson, 1986), adults in Experiment 3 used epistemic reasoning to compute scalar inferences from non-linguistic (pictorial) stimuli. Children, however, reliably matched a more-informative communicative act (circling two objects) to a speaker with full knowledge, but failed to reliably match a less-informative communicative act (circling one object) to a speaker with limited knowledge. Given children's earlier linguistic success, and the fact that performance on More-Informative trials (that require perspective-taking but not a scalar inference) remained unaffected from Experiment 1, we do not wish to conclude that children cannot apply Gricean epistemic reasoning to non-linguistic stimuli. It seems more likely that features of the task impacted children's performance in the Less-Informative trials in Experiment 3.

One possibility lies with children's developing understanding of symbolic representation (DeLoache, 1991; DeLoache, 2000; Uttal, O'Doherty, Newland, Hand & DeLoache, 2009; Allen, Bloom, & Hodgson, 2010). Children have been shown to struggle with the idea that symbolic artifacts (such as pictures) have a *dual nature*: they are both physical objects in themselves as well as abstract representations of something else (DeLoache, 2000; Uttal, et al., 2009). The high degree of visual interest (color and detail) in the pictorial representation of the circled objects may have impacted children's ability to overcome their physical nature and assess them as abstract representations meant to represent fully the items in the box. This difficulty may have also impacted children's ability to connect the act of circling objects to a speaker's visual access. Alternatively, given that circling may be less familiar as a communicative act, children may have been unwilling to interpret circling one object as excluding another. We are currently pursuing these topics in on-going work.

Regardless of the specific explanation for children's behavior in Experiment 3, the selective failure of 4- and 5-year-olds in a paradigm with the same features as Experiment 1 makes alternative explanations for success in Experiments 1 and 2 unlikely. Recall that one possibility raised at the end of Experiment 2 was that children might successfully match a more-informative statement to a full-knowledge speaker, and then simply select the other box in the Less-Informative condition without full pragmatic understanding. Experiment 3 shows that this strategy does not apply: even though children quite successfully chose the full-knowledge communicator for More-Informative trials, they did not match the other type of box to the other communicator. A different possibility was that children could select the correct boxes in the Less-Informative condition on the basis of probability. In

Experiment 3, even though the speaker who saw only one object had a higher probability of circling the spoon in the Less-Informative trials, children still did not reliably select that answer. We conclude that simple strategies involving contrast or probability are unlikely to explain children's success in the linguistic versions of this task.

2.4 General Discussion

In conversation, listeners typically assume that the speaker is providing as much relevant information as is required by the purpose of the exchange (and no more than the speaker has evidence for; Grice, 1975). Thus when a speaker offers a less-informative statement ("Some of the animals are safe to pet"), the listener infers that the speaker is not in a position to offer a more-informative statement ("All of the animals are safe to pet") – either because the speaker does not know whether the more-informative statement is true or (in case the speaker is presumed to be knowledgeable on the topic) because the speaker knows that the more-informative statement is false. In a set of three experiments, we investigated whether 4- and 5-year-olds are able to incorporate speaker knowledge into the computation of ad-hoc scalar implicatures. In these studies, we defined being able to compute epistemic scalar inferences as being able to link a more informative statement with a full-knowledge speaker and a less informative statement with a limited-knowledge speaker.

Our paradigm was based on a simple, clear way of establishing that someone's knowledge differs from the child's; speaker knowledge state was established through visual access and could be determined instantly by looking at a photo instead of watching a set of videos or a scenario unfold (Hochstein et al., 2016; Barner et al., 2018; Papafragou et al., 2018). Furthermore, the present paradigm included a clear conversational goal (the identification of the box that the speaker is talking about). In both of these respects, the current study was simpler than past attempts to link the informativeness of a sentence to a speaker's mental state (Hochstein et al., 2016; Barner et al., 2018; Papafragou et al., 2018). Our results provide the first evidence that children as young as 4 can flexibly adjust to a speaker's epistemic state during the derivation of scalar implicatures (Experiments 1 and 2). Furthermore, unlike prior studies, 5-year-olds are as good as adults in consulting the speaker's epistemic state (Experiment 1). Our results lower prior estimates of when young children display sophisticated adaptions to others' epistemic states during implicature derivation. They further suggest that children's failures in previous tasks are better explained by task-specific processing difficulties rather than insufficient development of epistemic reasoning ability.

From a more general perspective, the present data offer one of the earliest demonstrations of children's success with scalar implicature (cf. also Barner et al., 2011; Stiller et al., 2011; Papafragou et al., 2018, Exp.2, none of which included an epistemic component). Recall that this area has attracted a lot of experimental attention and is typically thought to pose challenges to 5-year-olds and sometimes even older learners (Noveck, 2001; Chierchia et al., 2001; Papafragou & Musolino, 2003; Guasti et al., 2005, Huang & Snedeker, 2009b, Horowitz, Schneider, & Frank, 2018). Our findings confirm the conclusion that the ability to compute scalar implicatures is fairly task-dependent (Papafragou & Musolino, 2003; Papafragou & Tantalou, 2004; Guasti et al., 2005; Pouscoulous et al., 2007; Barner et al., 2011; Foppolo et al., 2012; Skordos & Papafragou, 2016). Furthermore, they leave open the possibility that a version of the present paradigm could be used to uncover sensitivity to speaker knowledge during scalar implicatures in even younger children.

Viewed most broadly, our results are consistent with studies reporting that sensitivity to others' mental states appears even in infancy (Baldwin, 1991; Onishi & Baillargeon, 2005; Liszkowski, et al., 2008; Kovács, Téglás, & Endress, 2010; Southgate, et al., 2010), and characterizes both speech production and comprehension at early ages (Shatz & Gelman, 1973; Clark, 1990; Baldwin, 1993a, 1993b; O'Neill, 1996; Carpenter et al., 1998; Nadig & Sedivy, 2002; Matthews et al., 2006; Southgate, et al., 2010). From a theoretical perspective, our results comport with – and offer support to - broader pragmatic theories according to which scalar implicature derivation (alongside other pragmatic phenomena) requires rich computations about the speaker's perspective and knowledge state (Horn, 1972; Sperber & Wilson, 1986; Carston, 1998; Sauerland, 2004, 2012; Geurts, 2010). Furthermore, they offer grounds for assuming continuity between the mechanisms responsible for the computations of these inferences in adults (see Bergen & Grodner, 2012; Breheny et al., 2013; cf. Fairchild & Papafragou, 2018) and young learners.

The findings and paradigm from this set of experiments provide fertile ground for a continued investigation into the pragmatic ability of young children. Notice that in the present set of studies children's scalar inference computation was highly supported by design factors such as providing relevant alternatives, selecting between two photos in a forced-choice paradigm, etc. A promising direction for future research would involve manipulating or reducing this support to better identify which factors make the greatest contribution to children's success.

A second direction for future work concerns the extension of pragmatic principles to non-linguistic communication. Our results show that, as predicted by Gricean and later accounts (Grice, 1975; Sperber & Wilson, 1986), adults generalize pragmatic principles to other forms of purposive exchanges, for instance, by applying the Maxim of Quantity to pictorial symbols (Experiment 3). However, children do not immediately do the same, perhaps because of the features of our non-linguistic task. It remains to be seen whether different instances of ostensive communication could reveal children's ability to reason about what is contained vs. omitted in pictures to infer what the creator of the picture wanted to convey, as predicted by pragmatic accounts.

Finally, the current paradigm could be adapted to test whether children are sensitive to calculations involving epistemic state in pragmatic domains beyond implicature. For instance, methods similar to the present ones could be extended to the domain of word learning or referential communication to adjudicate between differing accounts over whether children engage in rich, Gricean pragmatic reasoning during these processes (Clark, 1990; O'Neill, 1996; Diesendruck & Markson, 2001; Nadig & Sedivy, 2002; Matthews, et al., 2006) or not (Glucksberg et al., 1975; Conti & Camras, 1984; Preissler & Carey, 2005; de Marchena et al., 2011). Ultimately, the current results and methods should be integrated within a unified account of how children use social-pragmatic reasoning to interpret a variety of linguistic expressions (and also non-linguistic communicative symbols) in accordance with broad principles of communication.

Chapter 3

EPISTEMIC REASONING IN SCALAR INFERENCE IN NON-LINGUISTIC COMMUNICATION

In Chapter 2, we introduced a novel visual world paradigm to ask whether children and adults apply pragmatic principles to linguistic communication. We found the first evidence that 4-year-olds (as well as 5-year-olds) successfully incorporate speaker knowledge into the derivation of pragmatic inferences. However, this ability did not generalize immediately to non-linguistic communicative contexts. We now explore the application of pragmatic maxims to non-linguistic symbols further in Chapter 3 through a new version of this paradigm.

3.1 Introduction

In a set of studies, we revisit the question of whether children and adults apply pragmatic principles to forms of ostensive communication beyond language. Grice (1989) proposed that other forms of purposive exchanges involve participants working towards a common goal and outlined non-linguistic applications of the conversational maxims. To take the Maxim of Quantity, just as participants in a conversational exchange are expected to provide as much information as required by their partner, participants in non-linguistic exchanges are expected to provide a contribution that is neither more or less than required. For example, if someone is assisting with the repair of a car and the mechanic needs four screws, they are expected to hand him four, not two or six (Grice, 1989). Other pragmatic frameworks such as Relevance Theory

(Sperber & Wilson, 1986) share the basic assumption that the principles underlying pragmatic interpretation in language can be applied to other intentional signals.

More generally, the idea that human communication is a species of human rational behavior has gained wide acceptance in the literature (see Tomasello, 2010; Frank & Goodman, 2012), and there is evidence that even infants are sensitive to properties of rational agents (Gergely, Nádasdy, Csibra, & Bíró, 1995; Meltzoff, 1995; Gergely & Csibra, 1997; Csibra & Gergely, 1998; Csibra, Gergely, Bíró, Koós, & Brockbank, 1999; Gergely, Bekkering, & Király, 2002). For example, children as young as 18 months can recognize the intentions of an agent attempting an action, even when the agent fails to complete the action, and do so with sensitivity to whether the agent is human vs. mechanical (Meltzoff, 1995). Existing literature has focused on the ability of children to interpret actions (including gestures) as intentional and communicative (Baker, Saxe, & Tenenbaum, 2009; Leekam, Solomon, & Teoh, 2010; Vivanti et al., 2011; Moore, Mueller, Kaminski, & Tomasello, 2015), with some reference to the nature of the inferences children are making (Tomasello, Call, & Gluckman, 1997; Bohn, Call, & Tomasello, 2019). However, the implications of this perspective for how children and adults derive pragmatic inferences from non-linguistic communicative stimuli such as actions, gestures, or pictures have not yet been systematically tested. For instance, it remains unknown whether children and adults can draw pragmatic inferences from underinformative non-linguistic symbols (e.g., pictures or drawings).

The mechanisms underlying non-linguistic communication tie into broader debates over the nature of pragmatic inference. Under Gricean, Relevance-theoretic and many subsequent approaches, scalar implicatures and other pragmatic computations are presumed to be the result of an extensive, rich social reasoning process in which the addressee interprets a communicative signal in line with assumptions about the communicator's knowledge state, goals, etc. (see, e.g., Grice, 1972, 1989; Horn, 1972; Sperber & Wilson, 1986; Carston, 1998; Sauerland, 2004, 2012; Geurts, 2010). These approaches contrast with certain formal linguistic theories which maintain that pragmatic inferences could be derived through linguistic mechanisms alone, and need not rely on mental perspective-taking (see e.g., Chierchia 2004; Chierchia, Fox, & Spector, 2009, 2012; Hochstein, Bale, Fox, & Barner, 2016; Barner, Hochstein, Rubenson, & Bale, 2018; Hochstein, Bale, & Barner, 2018). Existing experimental evidence primarily supports the first class of accounts: both adults (Bergen & Grodner, 2012; Breheny, Ferguson, & Katsos, 2013; cf. Fairchild & Papafragou, 2018) and 5-year-old children (Hochstein, Bale, Fox, & Barner, 2016; Papafragou, Friedberg & Cohen, 2018) compute scalar implicatures in accordance with the epistemic state of the speaker. Furthermore, in simple contexts, even younger children do so: for instance, 4-year-olds reliably recognize that a more informative statement ("I see a spoon and a bowl") belongs to a speaker who can see both objects in a display and a less informative statement ("I see a spoon") belongs to a speaker with partial visual access to the display (even when they themselves can see both objects; Kampa & Papafragou, accepted). Within this debate (see Barner, Hochstein, Rubenson, & Bale, 2018; Hochstein, Bale, & Barner, 2018), the question of whether children and adults consult a communicator's epistemic state in bringing pragmatic principles to bear on the interpretation of non-linguistic stimuli presents a valuable test of a rich, mentalistic model of pragmatic interpretation. On this model, children and adults are expected to apply pragmatic principles to non-linguistic (as well as linguistic) communicative stimuli in line with the communicator's epistemic state, and evidence that they do not do so would be problematic. Alternatively, on the view that linguistic mechanisms play a specialized role in the derivation of pragmatic inferences, evidence that children and adults apply pragmatic principles to non-linguistic (as well as linguistic) tokens would not be easily accounted for without additional theoretical machinery.

More broadly, results could contribute to a broader discussion of whether adults and especially young children routinely take into account the mental perspective of others in communication: some authors have argued that children have difficulty evaluating and editing communicative content (Glucksberg, Krauss & Higgins, 1975; Conti & Camras, 1984) and others have offered evidence for children's flexible adjustments to other people's perspectives in both speech production and comprehension (Shatz & Gelman, 1973; Baldwin, 1991; O'Neill, 1996; Nadig & Sedivy, 2002; Matthews, Lieven, Theakston & Tomasello, 2006).

3.1.1 Drawings as Communicative Symbols

Drawings provide an excellent domain to test whether pragmatic principles such as an expectation of informativeness extend to non-linguistic symbols. Adults regularly use drawings and other non-linguistic symbols and actions communicatively (Uttal, 1994; Montepare, Koff, Zaitchik, & Albert, 1999; Goldstone & Sakamoto, 2003; Rentschler, Jüttner, Osman, Müller, & Caelli, 2004; Uttal, O'Doherty, Newland, Hand Liu, DeLoache, 2009). Drawings differ in terms of visual detail and other types of information they contain (as anyone who has used visual instructions to assemble furniture can confirm).

Children interact with drawings and pictures from early ages and can recognize faces and objects in pictures even in infancy (Dirks & Gibson, 1977; Rose, 1977; DeLoache, Strauss, & Maynard, 1979; Barrera & Maurer, 1981; DeLoache & Burns, 1994), even if they have not been exposed to pictures previously (Hochberg & Brooks, 1962). They begin to create their own drawings and pictures shortly afterwards, and by 30 months can reliably use pictures to identify real-world referents (Deloache, 1987, 1991). For example, in a study by DeLoache (1987), an experimenter pointed to where a toy was hidden in a picture of a room and asked children to retrieve the toy. Thirty-month-old children successfully retrieved the toy, indicating an understanding that pictures can be used communicatively, although younger (24-month-old) children failed. By age 3, children can recognize that drawings can differ in their level of informativeness and can evaluate them with respect to other individuals' goals. For instance, 3-year-olds recognize that if someone's goal is to identify a drawing as a house, then a more detailed drawing of a house will be more useful, and a less detailed prototypical drawing will be more useful than a more detailed non-prototypical drawing (Allen, Bloom, & Hodgson, 2010). Furthermore, children from an early age reason about drawings in terms of the intentions of their creator. In one study (Preissler & Bloom, 2008), when an adult artist drew an object, 2-year-olds later mapped a novel name for the drawing to the object that the adult had been looking at; further manipulations showed that this response was not merely because there was more attention given to that object. This sensitivity to creator's intent extends to other artefacts (see Gelman & Bloom, 2000; Jaswal, 2006; but see Richert & Lillard, 2002).

Despite these early successes, interpreting pictorial symbols requires an understanding of dual representation, the idea that symbols are both physical entities and abstract representations of something else (DeLoache, 1987, 1995, 2000, 2002; Uttal, Schreiber, & DeLoache, 1995; Tomasello, Striano, & Rochat, 1999; Uttal et al., 2009), and children sometimes struggle to overcome the physical nature of a symbol to access its abstract representational role. For example, in one study by DeLoache (2000), 3-year-olds were initially able to use a 3D scale model to retrieve a hidden object in a room that was identical to the scale model. However, when the physical nature of the 3D model was emphasized by having children play with it first, 3-year-olds subsequently failed to use the model abstractly to retrieve the object. The opposite was also true: 2.5-year-olds initially failed to retrieve the object hidden in the room when presented with a 3D scale model, but when the abstract nature of the 3D model was emphasized by putting it behind a screen, they succeeded.

So far, this work has not considered the inferences children might derive from what drawings *omit* in addition to what they contain (i.e., scalar inferences from drawings). Given the inherently selective nature of symbols such as drawings (i.e., an artist rarely depicts everything they see), it is an open question whether children recognize that the choice to include some visual details but not others depends on one's wish to be informative, given one's knowledge and the goals of the exchange. Understanding drawings pragmatically involves generating (and negating) relevant, more-informative drawings – a process that may be challenging for young learners. Consistent with this possibility, in a study in which children were asked to draw a display, 7- to 10-year-olds adapted their drawings to a partner when told that their drawings would be used in a communicative game, but 4- to 6-year-olds could not readily adapt their drawings to the context (Light & Simmons, 1983). Additionally, children's ability to produce recognizable drawings is tied to their ability to identify the distinctive features of categories, which is undergoing continued development during these years (Long, Fan, Chai, & Frank, 2019).

3.1.2 Current Study

In three experiments, we investigated whether children and adults are able to use sensitivity to informativeness to derive scalar inferences from drawings. We adapted a simple paradigm from a previous study (Kampa & Papafragou, 2020), in which 4-year-old children successfully calculated scalar inference from a speaker's sentence in accordance with the speaker's epistemic state (cf. also Papafragou et al., 2018). The paradigm itself was inspired by visual-world studies of referential communication that manipulate whether participants and another agent share visual access to a display or not to assess sensitivity to others' knowledge (e.g., Nadig & Sedivy, 2002).

In the present version of the paradigm, participants (4-year-olds, 5-year-olds and adults) were presented with a pair of photos, each depicting a girl ("Danielle") facing the participant behind a two-compartment box. The boxes' compartments contained identical toys across the two photos. However, in one photo, both the participant and the girl had full visual access to the two compartments of the box (e.g., they could both see that the box contained a penguin and a pumpkin; full knowledge). In the other photo, the participant had full visual access to both compartments of the box, but the girl could only see one compartment because the other one was blocked from her view (e.g., Danielle could see that the box contained a penguin but could not see the pumpkin; limited knowledge). Participants were told that Danielle would look at each box, then draw what she sees in one of the boxes. Their task would be to decide which box she chose. At the end of the trial, participants saw the girl hold up a line drawing of either one object (a penguin) or two objects (a penguin and a pumpkin), and heard a female voice say: “I see this!” Next, they were asked: “Which box did she draw?”

We expected adults to match a more informative (two-object) drawing with a full-knowledge communicator, given that the limited-knowledge communicator could not see and communicate about both objects. Success on this type of trial required visual perspective-taking, since both boxes had identical contents from the participant’s viewpoint. We also expected adults to match a less informative (single-object) drawing with a limited-knowledge communicator, given the pragmatic expectation that a full-knowledge communicator should not fail to present complete and relevant information (i.e., the full contents of the box). Because the object depicted in the one-object drawing was present in both boxes (e.g., both boxes contained a penguin), participants could not choose the correct box on the basis of perspective-taking alone without considering pragmatic expectations about the amount of information a communicative drawing should contain. Of interest was whether children would also succeed on both of these two trial types, or, alternatively, whether there would be developmental differences between 4- and 5-year-olds (as in many previous studies of scalar implicature with linguistic stimuli; e.g., Hochstein et al., 2016; Papafragou et al., 2018). Furthermore, since the less informative trials offered a more stringent test of sensitivity to the principle of informativeness, we were interested in comparing success with less vs. more informative drawings across age groups: lower accuracy with less informative compared to more informative drawings would point to specific difficulties with scalar implicature and not with perspective taking more generally (perhaps especially so for children).

In Experiment 4, we introduced a basic version of this communicative interaction to assess how children (and adults) interpreted the use of non-linguistic symbols such as drawings. In Experiments 5 and 6, we extended our main findings by changing properties of the drawings in connection to their symbolic function. Notice that even though Danielle's communicative contribution contained some language ("I see this!"), the scalar alternatives being assessed throughout were pictorial (specifically, other drawings that Danielle could have produced but did not).

3.2 Experiment 4

3.2.1 Methods

3.2.1.1 Participants

Twenty-five 4-year-olds (M_{age} = 4;7, *range:* 4;0 to 4;11, 15 female, 10 male) and twenty-five 5-year-olds (M_{age} = 5;5, *range:* 5;0-5;11, 14 female, 11 male) participated. Children were recruited from Newark (DE) preschools and the Delaware Children's Museum in Wilmington, DE. An additional three 4-year-olds contributed data, but were excluded for failing preliminary tests (see Procedure for criteria). Twenty-five adult participants (M_{age} = 27.4, *range:* 18-35, 9 female, 16 male) were also tested. Adult participants were recruited with a HIT (Human Intelligence Task) posting on MTurk. All participants were monolingual speakers of English.

3.2.1.2 Procedure

For child participants, the experiment began with a warm-up phase designed to introduce the types of photographs used in the main task. Participants were first presented with a single photo on a computer screen (Figure 4). The photo depicted a girl sitting across from and facing out towards the participant. In front of her was a cardboard box with two vertical compartments. The compartment to the left (from the participant's viewpoint) was see-through and empty. The compartment to the right contained an object (a teacup) but was blocked so that only the participant but not the girl could see its contents.

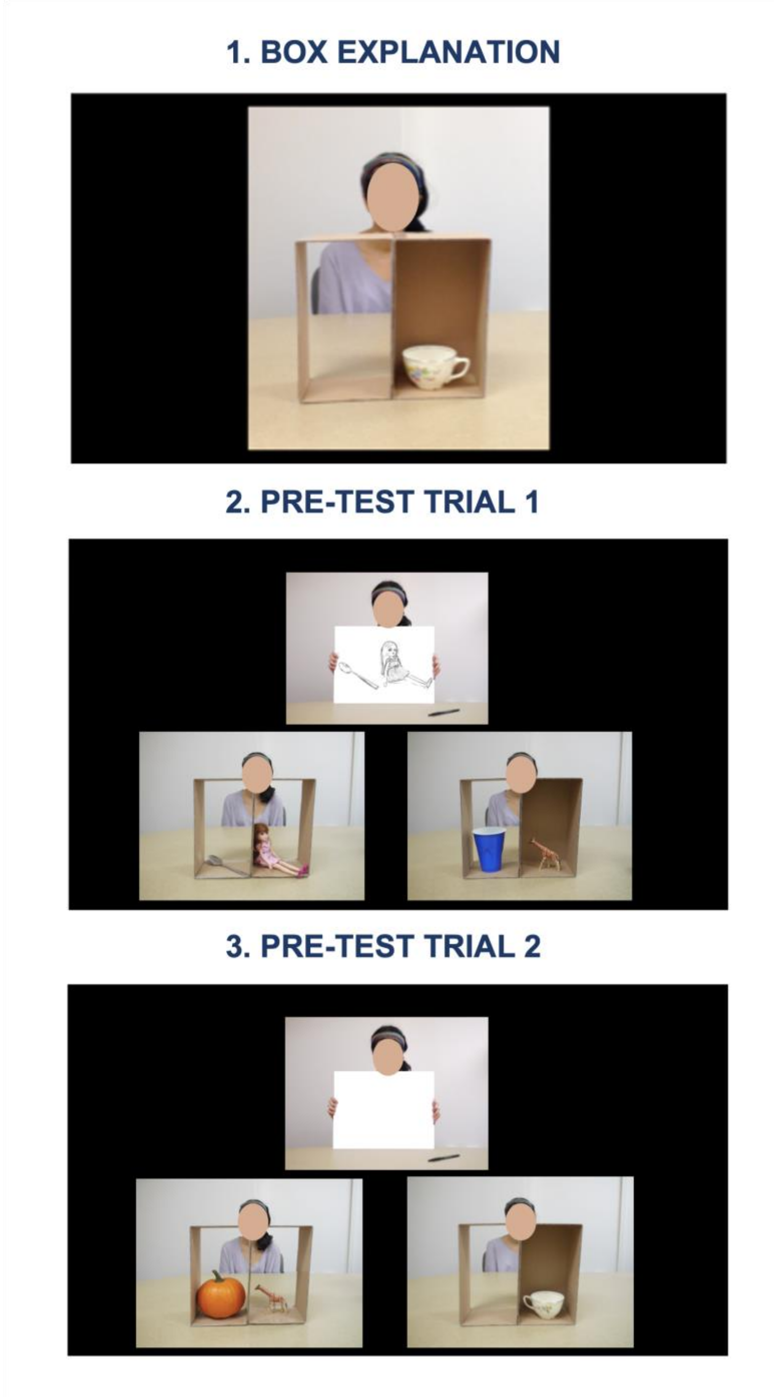

Figure 4: Visual displays for the introductory phase of Experiment 4. The faces have been blurred in this image per University of Delaware policy but were not blurred for the experiment.

The children were told that the girl in the photo was the experimenter’s friend, Danielle, and that she had a special box in front of her. The children were then asked whether they themselves could look through the two parts of the box (“Can you look through this side?”), and if not, why not. (For the blocked compartment, the children typically said “no” and mentioned that it was closed or blocked.) Children were asked if they thought Danielle could look through the blocked side (“So do you think she can look through that one?” <pointing>). The children were then asked whether Danielle knew about the item in the blocked compartment (“Does she know there’s a teacup there?”). The children who incorrectly replied that the girl knew the identity of the object in the blocked compartment were given corrective feedback ($N = 4$; “No, I don’t think she knows there’s a teacup there because she can’t look through that side; it’s blocked”).

All participants then completed two pre-test trials. In each trial, two photos were shown side by side. One photo showed Danielle and a limited-access box similar to the warm-up phase; the other showed Danielle and a full-access box where both compartments were unobstructed. One photo appeared on screen, followed by the other one after 2 seconds, as the experimenter explained the task (presentation order/side of the screen for the limited access box photo was counterbalanced across participants). Participants were told that they were going to play a game with the boxes. The experimenter explained that Danielle would look at each box and, pointing to each side of the boxes, reminded children whether it was open so that Danielle could look through it or closed so that she could not. Participants were told, “[Danielle] is going to draw what she sees in just one of the boxes, and you have to decide which box she chose.” After this explanation ended (and both photos had stayed up for 2 seconds), a third photo appeared above the other photos which depicted Danielle holding up a drawing on a piece of white paper (see, for example, Figure 4; in subsequent trials, each photo automatically appeared 2 seconds after the previous photo); when this picture appeared, children heard a generic recorded sentence from a female speaker that said "I see this!" Participants were asked, "Which box did she draw?" Both photos remained on screen until children responded. The two boxes had different contents, so the drawing unambiguously depicted one of them. For instance, in pre-test trial 1, the full-access box had a spoon and a doll, the limited-access box had a cup and a giraffe, and the girl held up a drawing of a spoon and a doll (depicted as two separate objects on a single piece of white paper; Figure 1). In pre-test trial 2, the full access box had two objects, but the limited-access box only had an object in the blocked compartment, and nothing in the open compartment. The drawing canvas was empty to indicate that there was nothing to see in the limited-access box. Because the trials in the critical testing condition present drawings of only one object, the pre-test trials were specifically designed to avoid presenting a drawing of one object when the contents in the boxes unambiguously differed, so as not to influence children's performance on critical test trials. No feedback was given during pre-test trials, and all children passed at least one of the pre-test trials and continued with the study.

The drawings were created using an online photo-to-line-drawing converter. Each object from the boxes was photographed individually against a white background, then that photo was converted into a black-and-white drawing and presented as described (see Figure 4). The drawings were slightly smaller than their corresponding objects in the photos, at a ratio of approximately 7:10.

Participants then completed 8 test trials. The set-up was the same as the pre-test trials; three photos appeared on the screen in succession from left to right with a 2 second delay in between each photo (Figure 5). As the third photo appeared, participants heard an audio recording from a female speaker that said "I see this!" Critically, for the test trials, the contents of the boxes within a trial were identical (e.g., a spoon and a bowl) but only one of the objects in the limited-access box (here, the spoon) was visible to the experimenter's friend. There were two within-subject conditions (More-Informative, Less-Informative) depending on whether children saw a more informative drawing depicting both objects (e.g., a drawing of a spoon and a drawing of a bowl placed on the same canvas) or a less informative drawing depicting only a single object – the one that remained visible to the speaker in the limited-access box but was present in the other box too (e.g., a spoon). Participants were again asked, "Which box did she draw?"

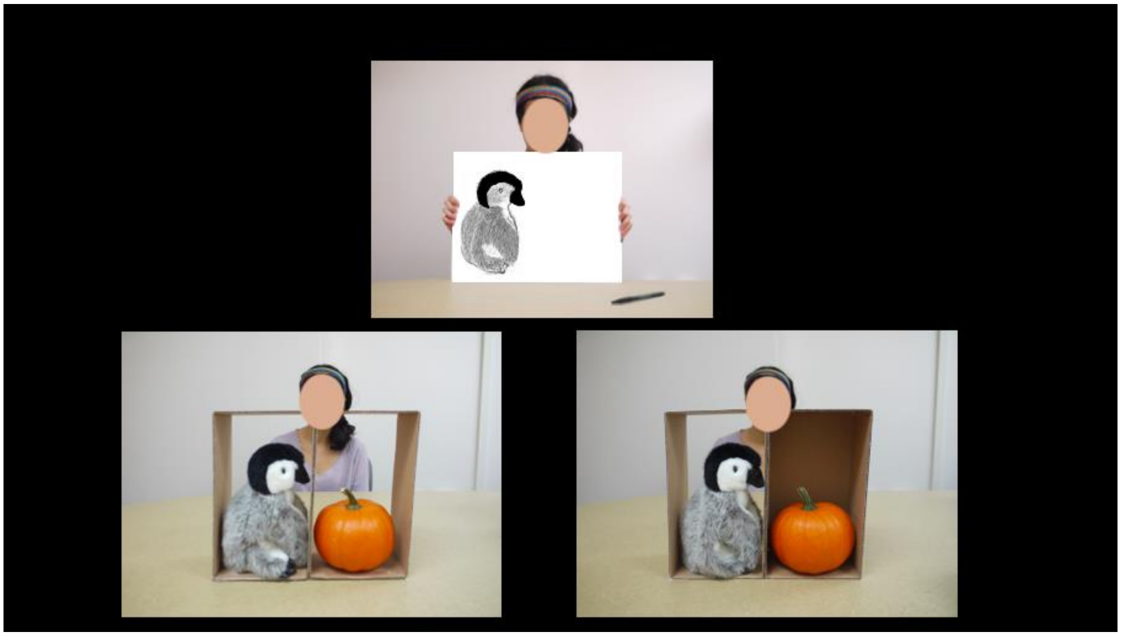

Figure 5: An example visual display from a test trial in Experiment 4. The photo on the left depicts the “full access” box (and the full-knowledge agent). The photo on the right depicts the “limited access box” (and the limited-knowledge agent). The photo in the center depicts her drawing of one of the boxes. Photos were presented successively from left to right with a 2 second delay between the appearance of each photo, with the top photo presented last. All photos remained on screen until the end of the trial. The faces have been blurred in this image per University of Delaware policy but were not blurred for the experiment.

For test trials 1-4, the girl in the photos was Danielle; for trials 5-8, the girl in the photos was Julia, another friend of the experimenter’s (the pre-recorded statements had a different female voice for Julia). Participants were given 4 more-informative and 4 less-informative test drawings in a mixed order, always beginning with a more-informative one (for the role of the availability of a logically stronger relevant alternative, see Skordos & Papafragou, 2016). Two presentation lists were created; assignment of statements corresponding to each condition (More-Informative, Less-

Informative) to trials was counterbalanced across lists. The position of the limited access box photo across trials was also counterbalanced within each list.

To introduce Julia, and to remind children of the critical properties of the boxes, after the first 4 test trials, children were presented with a reminder trial and questions that were modeled after the warm-up phase. The trial included a photo of a new friend, Julia, and a limited-access box with only one object (a blue plastic cup in the blocked compartment). As in the warm-up phase, children were asked whether they could look through each side of the limited-access box, and whether Julia could look through the blocked side (for the importance of such reminders of the visual properties of the display, see Nadig & Sedivy, 2002). No feedback was given. We also asked whether Julia knew about the contents of the blocked compartment ("Does she know there's a cup there?"). Children who replied "yes" to this question in both the warm-up and this reminder phase were excluded since they did not understand the nature of the limited-access box (N = 3 4-year-olds).

To reiterate, two patterns should emerge if participants could successfully extend pragmatic principles and incorporate the perspective and knowledge of the creator of the drawings into their responses on test trials. In the More-Informative condition, they should take the more informative drawing (both objects) to describe the full access box and thus belong to the fully knowledgeable agent, because the girl could not see the bowl in the limited-access box. In the Less-Informative condition, they should take the less informative drawing (e.g., only a penguin) to describe the limited-access box (and thus belong to the limited-knowledge agent) because it would be under-informative (i.e., violate pragmatic expectations) under these circumstances for the full-knowledge agent to only draw one object when they see two, especially when that second object is needed to disambiguate. Notice that these responses on both conditions presuppose sensitivity to agent knowledge since from the participant's perspective both boxes have identical, visible contents. However, agent knowledge alone is not sufficient to succeed in the critical Less-Informative condition, since the object depicted in the less informative drawing is visible to the agent in both boxes (e.g., she draws a penguin, and both boxes contain a penguin); it is only through considering the amount of information that the agent chose to present (maxim of Quantity) that the participant has a basis to reject the full-access box (for which a drawing of one object would be insufficient) and select the limited-access box.

3.2.2 Coding

A correct response was defined as selecting the full-access box in the More-Informative condition and the limited-access box in the Less-Informative condition. Participants were given a score of 0 or 1 for each trial accordingly.

3.2.3 Results

We analyzed the data using multi-level logistic mixed effects modeling. All models were fit using the glmer function of the lme4 package (Bates, 2005; Bates, Maechler, & Bolker, 2011; Bates, Maechler, Bolker, & Walker, 2015; Bates & Sarkar, 2007) in the R Project for Statistical Computing (R Development Core Team, 2012). Interactions that did not significantly improve model fit (assessed via a model comparison) were removed from the final model. Accuracy (by trial) was analyzed with a model that included Condition (More-Informative, Less-Informative) and Age (Adults, 5-year-olds, 4-year-olds) as fixed predictors, and Subjects and Items as random intercepts. Figure 6 summarizes the data.

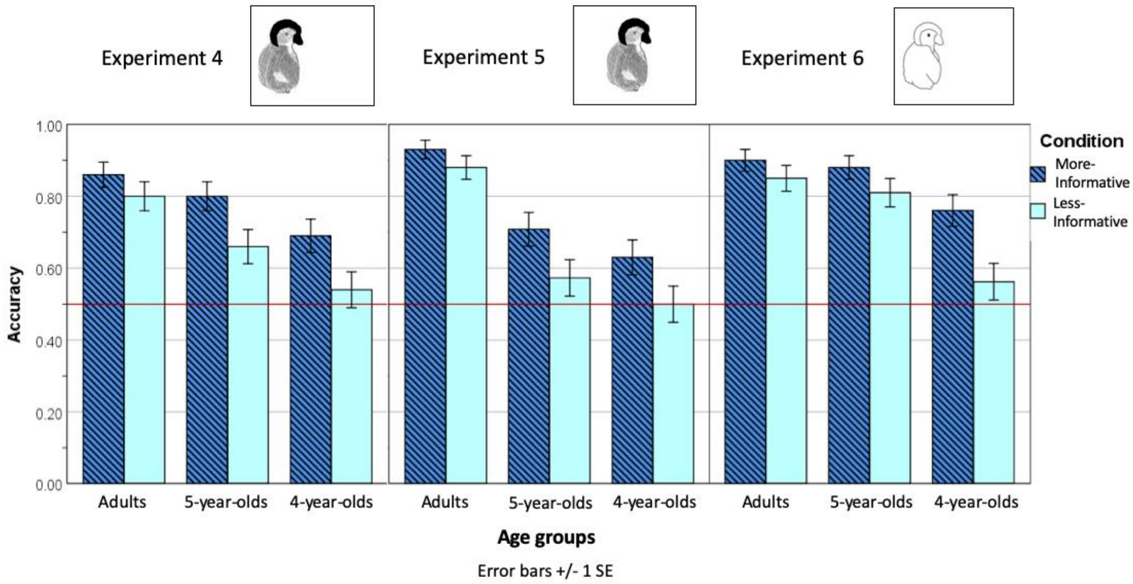

Figure 6: Results for Experiments 4, 5, and 6. The horizontal line indicates chance performance (0.5). Error bars indicate +/- 1 standard error. An example drawing from the Less-Informative condition is given for each experiment.

Table 4: Parameter Estimates for Accuracy in Experiment 4.

Effects	Estimate	SE	*z*
Intercept	1.8534	0.2583	7.175***
Condition (More-Informative, Less-Informative)	-0.8021	0.2197	-3.651***
Age (Adults vs. children)	-1.3404	0.4642	-2.888**
Age (4-year-olds vs. 5-year-olds)	0.6741	0.4717	1.429
Condition (MI, LI): Age (Adults vs. children)	-0.2722	0.5205	-0.523
Condition (MI, LI): Age (4-y vs. 5-year-olds)	-0.0749	0.4984	-0.150

Significance levels: * $p < .05$, ** $p < .01$, *** $p < .001$; MI is More-Informative, LI is Less-Informative

Both fixed predictors were included in the final model: Condition (More-Informative, Less-Informative) as a first-level predictor, and Age (Adults, 5-year-olds, 4-year-olds) as a second-level predictor. The fixed effect of Condition was coded with centered contrasts (-.5, .5). The fixed effect of Age was analyzed with using a Helmert coding scheme with two simple contrasts, first, comparing adults to children (c_1: -.66, .33, .33), then 5-year-olds to 4-year-olds (c_2: 0, -.5, .5). The same coding strategy was followed in all subsequent experiments. Table 4 presents the parameter estimates for the multi-level model.

The model revealed an effect of Condition: participants performed better in the More-Informative condition than the Less-Informative condition (M_{More} = .78, M_{Less} = .66; β = -0.80, z = 7.18, p < .001). The model also revealed an effect of Age when adults were compared to children: unsurprisingly, adults performed better (M_{Adult} = .83, M_{Child} = .67; β = -1.34, z = -2.89, p = .003). There was no effect of Age when 4-year-olds and 5-year-olds were compared to each other (M_{4yos} = .62, M_{5yos} = .73; β = 0.67, z = 1.43, p = .15). Finally, there was no significant interaction between Condition and Age (regardless of how Age was coded).

In order to determine whether each age group's performance in the More-Informative and Less-Informative conditions significantly differed from performance expected by chance, we conducted a one-sample t-test with a test value set at 0.5 (two referent choices) for each age group and condition. All age groups performed significantly differently from chance in the More-Informative condition (Adults: M = .86, t(24) =6.65, p < .001; 5-year-olds: M = .80, t(24) = 6.00, p < .001; 4-year-olds: M = .69, t(24) = 2.57, p = .017), but only adults and 5-year-olds performed significantly differently from chance in the Less-Informative condition (Adults: M = .80, t(24) =

5.04, $p < .001$; 5-year-olds: $M = .66$, $t(24) = 2.22$, $p = .036$; 4-year-olds: $M = .54$, $t(24) = .49$, $p = .62$).

3.2.4 Discussion

Experiment 4 found that adults applied the principle of informativeness to communicative exchanges with non-linguistic symbols: they correctly matched both a less informative drawing to a creator with limited knowledge and a more informative drawing to a creator with full knowledge, in accord with pragmatic principles. Thus, adults generated and negated relevant alternatives for non-linguistic symbols: when a creator presented a drawing (e.g., a penguin) that was less informative than an alternative (e.g., a penguin and a pumpkin), they inferred that the more informative alternative was not true. However, 4- and 5-year-olds performed significantly worse than adults in this task; furthermore, the two child groups did not differ from one another (even though in 5-year-olds, unlike 4-year-olds, performance was reliably different from chance in both conditions). Additionally, for both adults and children, the Less-Informative condition was more difficult, presumably because it introduced scalar reasoning to a task that already involved taking someone else's visual (and epistemic) perspective.

The present findings contrast with 4- and 5-year-olds' high level of success in a previous, otherwise identical study of scalar inference that used linguistic utterances as stimuli (Kampa & Papafragou, 2020). What could be responsible for children's difficulties? One possibility stems from the fact that children sometimes struggle with the idea that symbols are both physical entities and abstract representations of something else (DeLoache, 1987, 1995, 2000, 2002; Uttal, Schreiber, & DeLoache, 1995; Uttal et al., 2009). Inspired by this work, we identified an aspect of our design that could have increased the pull of the physical dimension of the drawings, thereby making it harder for children to treat them as abstract representations of the objects in the boxes, especially when these representations were meant to be intentionally selective as in the Less-Informative trials. Recall that the placement of the drawings in the Less-Informative trials was faithful to the actual display: the drawings were presented on the left side of the canvas (with blank space on the other side.) We hypothesized that disrupting the symmetry between the drawing and the display configuration in the Less-Informative condition (e.g., by placing the drawing in the middle of the canvas) might make it clearer that the image is meant as a symbol that is intentionally selective. Experiment 5 tested this possibility.

3.3 Experiment 5

3.3.1 Methods

3.3.1.1 Participants

Twenty-five 4-year-olds (M_{age} = 4;8, *range:* 4;2 to 4;11, 13 female, 12 male) and twenty-four 5-year-olds (M_{age} = 5;4, *range:* 5;0-5;11, 12 female, 13 male) participated. Children were recruited from Newark (DE) preschools and the Delaware Children's Museum in Wilmington, DE. An additional seven children contributed data, but were excluded for failing preliminary tests (see Procedure for criteria). Twenty-five adults (M_{age} = 29.8, *range:* 21-35, 11 female, 14 male) also participated. Adult participants were recruited with a HIT (Human Intelligence Task) posting on MTurk. All participants were monolingual speakers of English.

3.3.1.2 Procedure

Experiment 5 used the same photos, drawings, procedure, etc. as Experiment 4 but placed the drawings for the Less-Informative condition in the middle of the canvas instead of on the left side. As before, participants were excluded for failing both of the pre-test trials (*N* = 1 4-year-old, 1 5-year-old) or for saying "yes" when asked whether Danielle knows the identity of the object in the blocked compartment during both warm-up and reminder trials (*N* = 5 4-year-olds).

3.3.2 Coding

Coding was identical to Experiment 4.

3.3.3 Results

The same data analytic strategy as Experiment 4 was used. Accuracy (by trial) was analyzed with a model that included Condition (More-Informative, Less-Informative) and Age (Adults, 5-year-olds, 4-year-olds) as fixed predictors and Subjects and Items as random intercepts (see Figure 6 and Table 5). The model showed an effect of Condition: participants performed better in the More-Informative condition than the Less-Informative condition (M_{More} = .76, M_{Less} = .65; β = -0.67, z = -3.16, p = .002). The model also revealed an effect of Age when adults were compared to children (M_{Adult} = .91, M_{Child} = .60); β = -0.80, z = 7.18, p < .001). There was no effect of Age when 4-year-olds were compared to 5-year-olds (M_{4yos} = .57, M_{5yos} = .64; β = 0.38, z = 1.03, p = .30). Finally, there was no significant interaction between Condition and Age (regardless of how Age was coded).

A series of one-sample t-tests showed that performance on More-Informative trials was different from chance (0.5) for adults and 5-year-olds, but not 4-year-olds (Adults: M = .93, t(24) = 14.01, p < .001; 5-year-olds: M = .70, t(23) = 3.83, p = .001;

4-year-olds: *M* = .63, *t*(24) = 2.01, *p* = .06). A separate series of one-sample t-tests showed that performance on Less-Informative trials was different from chance for adults but not for children in either age group (Adults: *M* = .88, *t*(24) = 7.56, *p* < .001; 5-year-olds: *M* = .58, *t*(23) = 1.28, *p* = .21; 4-year-olds: *M* = .51, *t*(24) = .14, *p* = .89).

Table 5: Parameter Estimates for Accuracy in Experiment 5.

Effects	Estimate	SE	*z*
Intercept	1.6133	0.2240	7.203***
Condition (More-Informative, Less-Informative)	-0.6658	0.2109	-3.157**
Age (Adults vs. children)	-2.3119	0.4279	-5.403***
Age (4-year-olds vs. 5-year-olds)	0.3829	0.3726	1.028
Condition (MI, LI): Age (Adults vs. children)	-0.0130	0.5847	0.022
Condition (MI, LI): Age (4- vs. 5-year-olds)	-0.0618	0.4566	-0.135

Significance levels: * *p* < .05, ** *p* < .01, *** *p* < .001; MI is More-Informative, LI is Less-Informative

Table 6: Parameter Estimates for Accuracy in Experiments 4-5.

Effects	Estimate	SE	*z*
Intercept	1.7938	0.2175	8.248***
Condition (More-Informative, Less-Informative)	-0.7328	0.1522	-4.815***
Age (Adults vs. children)	-1.8510	0.3200	-5.784***
Age (4-year-olds vs. 5-year-olds)	0.5273	0.3052	1.728
Experiment	-0.1174	0.2631	-0.446
Condition (MI, LI): Age (Adults vs. children)	-0.2769	1.5094	-0.544
Condition (MI, LI): Age (4- vs. 5-year-olds)	-0.0749	0.4916	-0.152
Condition (MI, LI): Experiment (4, 5)	0.0590	0.3243	0.184
Age (Adults vs. children): Experiment (4, 5)	-1.3822	0.7653	-1.806
Age (4-year-olds vs. 5-year-olds): Experiment (4, 5)	-0.2680	0.7050	-0.381
Condition (MI, LI): Age (A vs C): Experiment (4, 5)	0.2997	0.7848	0.382
Condition (MI, LI): Age (4 vs 5): Experiment (4, 5)	0.0127	0.6759	0.019

Significance levels: * $p < .05$, ** $p < .01$, *** $p < .001$; MI is More-Informative, LI is Less-Informative, A is for Adults, C is for Children

To compare across experiments, we ran a model that included Condition (More-Informative, Less-Informative), Age (Adults, 5-year-olds, 4-year-olds), and Experiment (Experiment 4, Experiment 5) as fixed predictors, and Subjects and Items as random intercepts (Table 6). As in previous analyses for Experiments 4 and 5, the model revealed an effect of Condition ($M_{More} = .77$, $M_{Less} = .66$; $\beta = -0.73$, $z = -4.82$, $p < .001$). There was also an effect of Age only when adults were compared to children ($M_{Adult} = .87$, $M_{Child} = .64$; $\beta = -1.85$, $z = -5.78$, $p < .001$) but not when 4-year-olds and 5-year-olds were compared to each other ($M_{4yos} = .59$, $M_{5yos} = .69$; $\beta = 0.53$, $z = 1.73$, $p = .08$). There was no significant effect of Experiment, and no interactions between factors were significant (regardless of how Age was coded).

3.3.4 Discussion

Experiment 5 showed that the spatial positioning of the drawings did not appear to be a critical component in children's interpretation of symbols in Experiment 4: performance did not improve when the drawings were presented in the center of the canvas.

An alternative explanation for children's persistent difficulties could be that, in both Experiment 4 and 5, the drawings were highly detailed and visually interesting. As dual representation theory suggests (DeLoache, 2000; Uttal et al., 2009), those features might have made it difficult for children to resist the physical dimension of the drawings. If so, reducing the level of detail in the drawings might help children overcome the limitations observed in Experiments 4 and 5 and treat drawings as abstract, often selective representations of the objects in the boxes. Experiment 6 pursued this possibility.

3.4 Experiment 6

3.4.1 Methods

3.4.1.1 Participants

Twenty-four 4-year-olds (M_{age} = 4;5, *range:* 4;0 to 4;10, 14 female, 10 male) and twenty-five 5-year-olds (M_{age} =5;7, *range:* 5;0-5;10, 15 female, 10 male) contributed data. Children were recruited from Newark (DE) preschools and the Delaware Children's Museum in Wilmington, DE. An additional three children participated, but were excluded from data analyses for failing to meet inclusionary criteria (see Procedure). Twenty-five adults (M_{age} = 27.6, *range:* 19-34, 6 female, 19 male) also participated. Adult participants were recruited with a HIT (Human Intelligence Task) posting on MTurk. All participants were monolingual speakers of English.

3.4.1.2 Procedure

Experiment 6 used the same photos, procedure, etc. as Experiment 4, but the drawings were made more schematic (see example drawing in Figure 6) by removing unnecessary lines, shading, and other details using Photoshop. As before, participants were excluded for saying “yes” when asked whether Danielle knew the identity of the object in the blocked compartment during both warm-up and reminder trials (N = 2 4-year-olds, 1 5-year-old).

3.4.2 Coding

Coding was identical to Experiments 4 and 5.

3.4.3 Results

The same data analytic strategy as Experiments 4 and 5 was used. Accuracy (by trial) was analyzed with a model that included Condition (More-Informative, Less-Informative) and Age (Adults, 5-year-olds, 4-year-olds) as fixed predictors and Subjects and Items as random intercepts (see Figure 6 and Table 7). The model revealed an effect of Condition: participants performed better in the More-Informative than the Less-Informative condition (M_{More} = .85, M_{Less} = .74; β = -0.86, z = -3.55, p < .001). The model also revealed an effect of Age: adults performed significantly better than children (M_{Adult} = .88, M_{Child} = .75; β = -1.14, z = -2.38, p = .02) but in addition, 5-year-olds performed significantly better than 4-year-olds (M_{5yos} = .85, M_{4yos} = .66; β

= -1.35, z = -2.70, p = .007). There was no significant interaction between Age and Condition (regardless of how Age was coded in the model).

A series of one-sample t-tests showed that both adults' and 5-year-olds' performance was significantly different from chance (0.5) on both More-Informative and Less-Informative trials, but 4-year-olds performed differently from chance only on More-Informative trials (More-Informative - Adults: M = .91, $t(24)$ = 11.36, p < .001; 5-year-olds: M = .88, $t(24)$ = 12.97, p < .001; 4-year-olds: M = .77, $t(23)$ = 4.66, p < .001; Less-Informative - Adults: M = .86, $t(24)$ = 7.36, p < .001; 5-year-olds: M = .81, $t(24)$ = 5.80, p < .001; 4-year-olds: M = .55, $t(23)$ = .50, p = .222).

Table 7: Parameter Estimates for Accuracy in Experiment 6.

Effects	Estimate	SE	*z*
Intercept	2.4113	0.2936	8.213***
Condition (More-Informative, Less-Informative)	-0.8647	0.2438	-3.547***
Age (Adults vs. children)	-1.1401	0.4790	-2.380*
Age (4-year-olds vs. 5-year-olds)	-1.3533	0.5007	-2.703**
Condition (MI, LI): Age (Adults vs. children)	-0.3166	0.5714	-0.554
Condition (MI, LI): Age (4- vs. 5-year-olds)	0.4990	0.5655	0.883

Significance levels: * p < .05, ** p < .01, *** p < .001. MI is More-Informative, LI is Less-Informative

To compare across all experiments, we ran a model that included Condition (More-Informative, Less-Informative), Age (Adults, 5-year-olds, 4-year-olds), and Experiment (Experiment 4, 5, 6) as fixed predictors and Subjects and Items as random intercepts (Table 8). To investigate whether the features of the drawings impacted performance, the fixed effect of Experiment was analyzed with a contrast (c_1: -.66, .33, .33) comparing Experiment 6 (schematic drawings) to Experiments 4 and 5 (complex drawings). As in previous analyses for Experiments 4 and 5, the model revealed an effect of Condition: participants performed significantly better in the More-Informative condition than the Less-Informative condition (M_{More} = .80, M_{Less} = .69; β = -0.77, z = -5.72, p < .001)., There was also an effect of Age: across experiments, adults performed significantly better than children (M_{Adult} = .87, M_{Child} = .68; β = -1.64, z = -6.14, p < .001), and 5-year-olds performed significantly better than 4-year-olds (M_{4yos} = .61, M_{5yos} = .74; β = 0.77, z = 2.94, p = .003). Importantly, performance was significantly better in Experiment 6, which used schematic drawings, than in Experiments 4 and 5, which used abstract drawings (M_{Exp1} = .73, M_{Exp2} = .70, M_{Exp3} = .79; β = -0.59, z = -2.37, p = .02). There were no significant interactions (regardless of how Age was coded).

Table 8: Parameter Estimates for Accuracy in Experiments 4, 5, & 6.

Effects	Estimate	SE	*z*
Intercept	1.9535	0.1527	12.794***
Condition (More-Informative, Less-Informative)	-0.7706	0.1348	-5.718***
Age (Adults vs. children)	-1.6438	0.2678	-6.137***
Age (4-year-olds vs. 5-year-olds)	0.7713	0.2625	2.938**
Experiment (Abstract vs. Complex; 6 vs. 4 & 5)	-0.5827	0.2456	-2.372*
Condition (MI, LI): Age (Adults vs. children)	-0.1885	0.3237	-0.582
Condition (MI, LI): Age (4- vs. 5-year-olds)	0.1178	0.2934	0.402
Condition (MI, LI): Experiment (4, 5, 6)	-0.0534	0.2990	0.858
Age (Adults vs. children): Experiment (4, 5, 6)	-0.9134	0.6847	-1.334
Age (4vs5): Experiment (4, 5, 6)	1.3982	0.7783	1.796
Condition (MI, LI): Age (AvsC): Experiment (4, 5, 6)	0.1920	0.6967	0.276
Condition (MI, LI): Age (4vs5): Experiment (4, 5, 6)	-0.3014	0.7888	-0.382

Significance levels: * $p < .05$, ** $p < .01$, *** $p < .001$; MI is More-Informative, LI is Less-Informative, A is for Adults, C is for Children

3.4.4 Discussion

In Experiment 3, we found that decreasing the level of detail in the drawings, thereby emphasizing their abstract nature, significantly improved overall performance. Adults continue to perform better than children; furthermore, as in previous experiments, the Less-Informative condition was more challenging than the More-Informative condition for all age groups, presumably because of its added demands (scalar reasoning in addition to perspective taking).

However, 5-year-olds performed better than 4-year olds; furthermore, comparisons to chance showed that both adults and 5-year-olds reliably followed expectations of informativeness for schematic drawings in both the More-Informative and the Less-Informative condition. However, 4-year-olds still struggled to apply pragmatic principles in the Less-Informative condition. This pattern of results, therefore, suggests that, at least for the older children in our sample, abstract drawings can be interpreted as symbols that are subject to pragmatic principles and can give rise to the equivalent of a scalar inference. We return to the significance of these patterns in the General Discussion.

3.5 General Discussion

According to Grice (1975, 1989), communication relies on inference, which is governed by expectations of rational behavior. Under this approach, communicators are expected to be cooperative in order to further the goal of a conversation. One expectation is that communicators should provide as much information as required by their communicative partner (the maxim of Quantity). Listeners draw inferences about what a speaker intended based on these principles through a process of reasoning about the speaker and their goals and characteristics, such as their epistemic state. Grice and others (e.g., Sperber & Wilson, 1986) have proposed that the same principles underlie purposive exchanges beyond language, but relevant evidence so far has been limited. In a set of studies, we tested this proposal by asking whether children and adults use the pragmatic maxim of Quantity in non-linguistic communication. We adapted a simple visual-world paradigm that has been used recently with 4- and 5-year-old children to elicit successful Quantity (scalar) inferences from utterances in line with the speaker's epistemic state (Kampa & Papafragou, 2020) to ask how drawings are understood as communicative symbols. Specifically, we tested whether children and adults could use the relative informational content of drawings to compute what kind of message the communicator had in mind.

Our findings show that, for adults, pragmatic principles apply beyond linguistic communication to non-linguistic symbols: throughout all our experiments, adults successfully interpreted drawings as representing objects, recognized that a creator would not produce a drawing of an object they could not see, and drew inferences from the omission of an object in a drawing about the scene ('box') that the creator intended to pick out. In the same series of experiments, 4- to 5-year-olds' performance was systematically lower than that of adults. Nevertheless, in all experiments, both adults and children performed worse when they had to reason about drawings that potentially violated the principle of informativeness and therefore involved scalar reasoning (Less-Informative condition) compared to drawings that could be interpreted simply via perspective-taking (More-Informative condition), suggesting a common underlying mechanism. Furthermore, both adults' and children's performance improved when the drawings were abstract (Experiment 3): under those circumstances, 5-year-olds performed better than younger children in our sample and could reliably use the degree of informativeness of a drawing to pick out a referent in a scene in accordance to the communicator's mental state (cf. also Experiment 1). To our knowledge, the data from adults and 5-year-olds provide one of the first pieces of evidence in support of the much cited but rarely tested claim that principles of communication such as the need to be informative underlie all forms of purposive exchanges (cf. Grice, 1972, 1989; Sperber & Wilson, 1986).

Our results have implications for the development of children's symbol understanding. As alluded to earlier, children have shown success at interpreting the abstract representation of pictures earlier than age five in simple retrieval and referent selection tasks (DeLoache, 1987, 1991, 2000; Marzolf & DeLoache, 1994; Preissler &

Bloom, 2008; see also Allen et al., 2010). However, when they have to use that abstract representation for more complex processes (e.g., to judge differences between appearance and reality; Zaitchik, 1990; Thomas, Nye, & Robinson, 1994), they are less successful, and challenges with dual representation continue into adulthood (Goldstone & Sakamoto, 2003; Kaminski, Sloutsky, & Heckler, 2009). In our task, children had to not only understand that the drawings represented a real item in the boxes, but had to be able to use that abstract representation to reason about a person's intentions in producing a drawing, the set of relevant alternative drawings they could have produced, and the relative informativeness of those drawings. Thus our data are a novel test of dual representation theory and suggest that 5-year-olds at least understand that drawings carry information not only by virtue of what they contain but also of what they leave out, although their ability to assess a drawing's informational content may be impacted by the characteristics of the drawing.

From a different perspective, our data have theoretical implications for current research on children's computation of scalar inferences and other types of pragmatic meaning in language. Five-year-olds' performance in Experiment 6 differs from several prior studies showing that children of this age have difficulty with pragmatic inferences, especially the computation of scalar implicatures. In many of these studies, children's sensitivity to pragmatic inference was measured by the ability to evaluate and reject underinformative utterances offered either by the experimenter (Noveck, 2001; Guasti et al., 2005) or by a speaker known to be incompetent (e.g., a 'silly puppet'; Papafragou & Musolino, 2003; Katsos & Bishop, 2011). Our paradigm differed from these earlier studies in two key respects. First, we did not ask children to explicitly assess the informational content of drawings but simply to match drawings to the display that a person was facing (and presumably intended to depict). Second, and most importantly, the content of the drawing was motivated by the visual perspective of the creator: when the drawing omitted an object in the display, this occurred because the creator's perspective did not give access to that object. The success of the 5-year-olds in Experiment 6 confirms the conclusion that early sensitivity to scalar inference is context- and task-dependent (Noveck, 2001; Papafragou & Musolino, 2003; Papafragou & Tantalou, 2004; Guasti et al., 2005; Pouscoulous et al., 2005; Katsos & Bishop, 2011).

Finally, notice that, for 5-year-olds to succeed on this task, they had to not only derive pragmatic inferences from pictorial symbols but do so from the epistemic perspective of a communicator that was sometimes different from their own. Thus our results add to the small but growing body of evidence in the literature showing that preschoolers are able to consult a speaker's epistemic state when computing a scalar implicature (Kampa & Papafragou, accepted; Papafragou et al., 2018; Hochstein et al., 2016), as predicted by rich, Gricean and post-Gricean models of implicature computation (cf. Horn, 1972; Sperber & Wilson, 1986; Grice, 1989; Carston, 1998; Sauerland, 2004, 2012; Geurts, 2010, among others). More broadly, our data are consistent with the position that children are able to adapt to a communicative partner's characteristics and needs at young ages (Baldwin, 1991; O'Neill, 1996; Nadig & Sedivy, 2002; Matthews, Lieven, Theakston, & Tomasello, 2006; Bahtiyar & Küntay, 2009; see Grigoroglou & Papafragou, 2018, 2019, for a review), but are harder to reconcile with proposals according to which children are largely egocentric and fail to adjust to other's needs until late in development (Piaget, 1970; Shatz, 1980; Perner & Leekam, 1986; Girbau, 2001; Epley, Morewedge, & Keysar, 2004; Davies &

Katsos, 2010). Relatedly, the present results are in line with reports that young children can consider a creator's intent when interpreting drawings (Gelman & Ebeling, 1998; Gelman & Bloom, 2000; Richert & Lillard, 2002; Jaswal, 2006; Preissler & Bloom, 2008; cf. also Allen et al., 2010), and further suggest that 5-year-olds can consider a creator's intent even when interpreting what a drawing is meant to communicate by omitting (as opposed to simply including) information.

3.5.1 Developmental Limitations and Growth

A recurring finding across Experiments 1, 2 and 3 was that 4-to-5-year-olds performed significantly worse than adults and, with the exception of Experiment 3, there were no differences between child groups. This finding is striking given that both 4- and 5-year-olds had great success in a previous study involving a matched linguistic task (Kampa & Papafragou, 2020, esp. Exp.1). In that study, participants were told that Danielle would look at each box, then tell us what she sees in one of the boxes, and their task would be to decide which box she was describing. At the end of the trial, participants heard a female voice say either "I see a penguin" or "I see a penguin and a pumpkin" (compare to hearing "I see this" with a drawing of a penguin and/or pumpkin in the present set of studies), and were asked: "Which box is she talking about?" Adults and 5-year-old children performed at ceiling and even 4-year-olds successfully chose the limited knowledge box when Danielle was less informative (i.e., mentioned only one object) and the full knowledge box when Danielle was more informative (i.e., mentioned both objects). Against this background, how can the present difficulties be explained?

One possibility relates to children's progressive development of dual representation during the preschool years. Most of the literature on dual representation has centered on 2- and 3-year-olds and their ability to interpret symbols in simple retrieval tasks (DeLoache, 1987, 1995, 2000; Uttal, Schreiber, & DeLoache, 1995; Uttal et al., 2009), but - as mentioned already - there is evidence to suggest that the construct of dual representation and its challenges continue in later childhood and even into adulthood (Goldstone & Sakamoto, 2003; Kaminski, Sloutsky, & Heckler, 2009). Since the detailed drawings are more visually interesting than the schematic ones, it could be that children are more focused on their physical nature as drawings rather than their abstract nature as representations of the toys in the box; concreteness has been shown to impede children's ability to link symbols to their referent in a task that requires abstract reasoning (see discussion of similar cases in Uttal et al., 2009). A similar phenomenon is also seen in tasks such as the false photograph task (Zaitchik, 1990; cf. also Slaughter, 1998; Leslie, 2000), in which children fail to reject a photograph that differs from the current reality because they cannot overcome the "pull of the real" of the photograph. In Experiment 6, where the demands of dual representation were reduced through fairly small alterations to the drawings to reduce the pull of their physical representation, performance significantly improved. One possible piece of evidence against this account is that, typically, pictorial symbols which look more visually similar to their referents aid rather than impede dual representation (DeLoache, Kolstad, & Anderson, 1991; DeLoache, Miller, & Rosengren, 1997). Because the level of detail in the detailed drawings makes them more visually similar to the items in the boxes compared to the schematic drawings, this account should predict improved performance with detailed drawings, contrary to what we observe.

A second, related explanation for children's difficulty using non-linguistic symbols abstractly is that children may be more familiar with using language as a communicative and referential medium compared to drawings. Language is used to refer to concrete objects from the earliest stages of development (Markman, 1987; Gleitman, & Gleitman, 1992; Tomasello, 1992, 2000; Golinkoff, Mervis, & Hirsh-Pasek, 1994; Gillette, Gleitman, Gleitman, & Lederer, 1999; Waxman & Booth, 2003; Tardif et al., 2008) but children's first drawings are typically scribbles that do not obviously refer to real-world entities (Luquet, 1927; Piaget & Inhelder, 1948, 1967; Krampen, 1991; Eng, 2013; Long, Fan, Chai, & Frank, 2019; although see Matthews, 1984; Yamagata, 1997). It may be that children perform well in the previous linguistic version of the task (Kampa & Papafragou, 2020) but struggle here because they are more familiar with language as a way of referring to objects in the world. There are two reasons to doubt that this explanation is viable. First, the evidence suggests that even if there may be a delay initially, by age 2.5 children already know that drawings are referential (DeLoache, 1987, 1991, 2000; Preissler & Bloom, 2008; Salsa & Vivaldi, 2016). Second, as Experiment 6 shows, children improve when given schematic drawings, even though such drawings are presumably less familiar to children compared to the rich drawings meant for children in books and other visual media.

A more plausible alternative, in our view, is that 4- and 5-year-olds' difficulty with non-linguistic symbols could reflect a difference in expectations for language vs. pictorial symbols. For example, language and pictorial symbols differ in terms of their level of effort; it is relatively low effort to produce a few more words (e.g. "I see a penguin… and a pumpkin"), thus reasonable to expect a speaker to be exhaustive in mentioning everything relevant that they see. However, the same is not true of pictures. It is relatively high effort to produce drawings, particularly the highly detailed drawings in Experiments 4 and 5, and there is rarely an expectation that an artist will draw everything that they see. Evidence suggests that preschoolers (and potentially even toddlers) can make inferences about an agent's motivations given his or her costs (Jara-Ettinger, Gweon, Tenenbaum, & Schulz, 2015, 2016; Jara-Ettinger, Tenenbaum, & Schulz, 2015). For example, when crackers are placed on a low box (low-effort) and cookies on a high box (high-effort), if a cookie-loving puppet selects crackers, 5-year-old children will say that he cannot climb (Jara-Ettinger, Gweon, Tenenbaum, & Schulz, 2015). Relatedly, if it is low effort for a communicator to provide more information (e.g., in language), and they chose not to, it is easier to interpret that omission as intentional. If it is high effort to provide more information (e.g., a second detailed drawing), children may neglect to interpret its omission as intentional. This could provide an additional or alternative explanation for why pragmatic performance improved in Experiment 6 that used abstract line drawings; because it would be relatively low effort to produce a second drawing, its omission is more likely to be correctly interpreted as communicative. Four-year-olds' continued failure in Experiment 6 could be due to a still-developing understanding of drawings as communicative rather than depictive tools.

3.5.2 Final Thoughts

Our findings show that adults and even young children apply the pragmatic principle of informativeness to forms of communication beyond language, but the characteristics of non-linguistic symbols impact young children's ability to draw inferences about what the symbols are meant to communicate. These data contribute to our understanding of how communicative inferences develop across different domains, and throw light on widely accepted but rarely tested assumptions from classic pragmatic theorizing. Furthermore, these results are likely to generalize to other non-linguistic domains beyond drawings, such as communicative actions (Bass, Bonawitz, & Gweon, 2017; Gweon & Asaba, 2018), as well as to more intensely collaborative contexts (cf. Nadig & Sedivy, 2002).

Viewed most broadly, our results open up the possibility that other maxims beyond Quantity should be used to interpret ostensive behavior. For instance, the Maxim of Quality states that a communicator should not say something they believe is false or that for which they lack evidence. Applied to non-linguistic purposive exchanges, if someone is helping to bake a cake, they are expected to hand the baker sugar and not salt (see Grice, 1989). According to the Maxim of Relation, a communicator is expected to be relevant: if someone is mixing ingredients for a cake, they do not expect to be handed a good book (ibid.). Finally, according to the maxim of Manner, communicators are expected to be brief and orderly and to avoid ambiguity or obscurity of expression: if someone needs a tool, they expect it to be handed to them in a straightforward way, not through a complex hand motion (see also the framework of Relevance Theory, where a single pragmatic principle is considered to be not only necessary, but sufficient to account for pragmatic inference in both the linguistic and the non-linguistic domain; Sperber & Wilson, 1986). In all of these cases, expectations about rational behavior are expected to guide how people interpret both non-linguistic stimuli and actions (e.g., what someone did) and omissions from these stimuli (e.g., what someone did not do but could have).

Chapter 4

EPISTEMIC REASONING IN SCALAR INFERENCE FROM A NEURAL PERSPECTIVE

In Chapters 2 and 3, we found evidence that young children do consult the epistemic state of the speaker during scalar inference in both linguistic and non-linguistic communication, although ongoing development of dual representation impacted children's success with nonlinguistic symbols. Here we adapt the same paradigm to neuroimaging methods to probe how broader theory of mind networks interact with scalar implicature derivation online and in development from a novel perspective. We also explore how other supporting mechanisms (executive functions, language development more broadly) interface with this process.

4.1 Introduction

Experimental evidence for children's early competency with scalar inferences can at times appear conflicting, and debates continue over the exact developmental timeline for scalar implicature development as well as the factors that contribute to this development. Early studies report failure as late as 9-years (Noveck, 2001), while some recent studies report scalar implicature reasoning ability as early as 3.5-years (Stiller, Goodman, & Frank, 2015; Kampa, Richards, & Papafragou, 2019). We know that modifying the cognitive burden of a given task improves children's ability to successfully derive scalar implicatures, whether through simplifying task demands (Pouscoulous, Noveck, Politzer, & Bastide, 2007) or providing children with relevant alternatives (Stiller et al., 2015; Skordos & Papafragou, 2016). However, there is no clear consensus on the role that other contributing factors, such as Theory of Mind reasoning, play in children's derivation of scalar implicatures.

We take the position, along with many studies, that the period between the ages of 3 and 5 often forms a critical juncture in the development of scalar implicatures (Noveck, 2001; Chierchia, Crain, Guasti, Gualmini, & Meroni, 2001; Papafragou & Musolino, 2003; Guasti, Chierchia, Crain, Foppolo, Gualmini & Meroni, 2005; Huang & Snedeker, 2009b; Barner, Brooks, & Bale, 2011; Foppolo, Guasti & Chierchia, 2012; Stiller et al., 2015 among others). This period of development is also an important time for the development of other relevant cognitive skills and functions such as language (Gleason, 2005), executive functions (Carlson, 2005; Garon, Bryson, & Smith, 2008; Zelazo & Müller, 2011), and Theory of Mind (Baron-Cohen, Leslie & Frith, 1985; cf. Wimmer & Perner, 1983; Perner & Wimmer, 1985; Wellman, Cross, & Watson, 2001; Hollebrandse, van Hout, & Hendriks, 2014), all of which may be recruited during the derivation of scalar implicatures in children and adults. Language development may be expected to interact with scalar implicature development not only because pragmatic inferences require the ability to produce and interpret language, but also because certain types of scalar implicatures, such as those arising from quantifiers, modals, and numerals, require knowledge of alternatives on a lexical scale. Indeed, there is evidence that linguistic factors, such as vocabulary and morpho-syntax specifically, predict pragmatic ability (Rollins, 1999; Foppolo, Mazzaggio, Panzeri, & Surian, 2020). Executive function skills also may be expected to play a role in scalar implicature derivation due to the heavy cognitive processing burden of the derivation process, including generating alternative statements the speaker could have said and rejecting them in light of the current discourse state, and there is some evidence that executive functions impact scalar implicature computation in adults (DeNeys & Schaeken, 2007; Apperly et al., 2008; Marty & Chemla, 2013; Antoniou, Cummins, & Katsos, 2016), although other studies have found conflicting evidence (Antoniou et al., 2014; Heyman & Schaeken, 2015; Fairchild & Papafragou, 2021).

There is considerably more controversy over the role that Theory of Mind plays in scalar implicature derivation. Under Gricean and several post-Gricean accounts, scalar implicature derivation is a complex, social, inferential process that requires the listener to incorporate information about the speaker's knowledge state and other characteristics (Grice, 1975, 1989; Sperber & Wilson, 1986, 2002; Sauerland, 2004). However, under a different class of lexical and grammatical accounts, it has been proposed that scalar implicature derivation can be achieved through linguistic mechanisms alone without the need for epistemic reasoning (Levinson, 2000; Chierchia, 2004, 2006; Fox 2007, 2012; Chierchia, Fox, & Spector, 2008). The proposed differences in the nature of scalar implicature derivation between these two classes of theories necessarily give rise to different predictions about the requisite mechanisms and abilities involved. Evidence that Theory of Mind is recruited during scalar implicature derivation would not be directly accounted for by grammatical accounts. Conversely, evidence that Theory of Mind is not recruited during scalar implicature derivation in adults would pose serious problems for Gricean and neo-Gricean theories under which epistemic reasoning is thought to be a critical component of such a derivation. Relatedly, evidence that Theory of Mind is not recruited during scalar implicature derivation in children if it is for adults might indicate that children rely on different cognitive skills while developing scalar implicature derivation skills.

There is evidence that adults adapt to a speaker's epistemic state in studies measuring online comprehension (Breheny, Ferguson, & Katsos, 2013) and reading times of scalar interpretation (Bergen & Grodner, 2012). Additionally, in a recent study by Fairchild and Papafragou (2021), Theory of Mind was found to predict scalar implicature ability; in fact, executive functions, when controlled for Theory of Mind, did not. Moreover, there is evidence that even preschool-aged children make sensitive adaptations to a speaker's epistemic state during scalar inference (Papafragou, Friedberg, & Cohen, 2018; Kampa & Papafragou, 2020). However, in other studies, researchers have often failed to find a correlation between Theory of Mind tasks and pragmatic tasks (Antoniou, Grohmann, Kambanaros, & Katsos, 2013; see Matthews et al., 2018 for a review) or have pointed to discrepancies between Theory of Mind and scalar implicature abilities, leading some to claim that young children compute scalar implicatures in the absence of epistemic reasoning (Barner, Hochstein, Rubenson, & Bale, 2018; Hochstein, Bale, & Barner, 2018). Additionally, individuals with Autism Spectrum Disorder, a disorder characterized by deficits in Theory of Mind, are still able to compute scalar implicatures, although in many of the studies showing success, Theory of Mind ability was either not measured or reported to be relatively high (Pijnacker, Hagoort, Buitelaar, Teunisse, & Geurts, 2009; Chevallier, Wilson, Happé, & Noveck 2010; Hochstein, Bale, & Barner, 2018; Van Tiel & Kissine 2018; although see also Noveck, Guelminger, Georgieff, & Labruyere, 2007). Few studies have tested scalar implicatures, Theory of Mind, language, and executive functions with the same participants in order to directly compare these variables.

Resolving the role that various supporting mechanisms play in children's development of scalar implicatures is critical to the pragmatics literature because children provide a unique opportunity to gain insight into the mechanisms underlying scalar implicature derivation that adults cannot provide. Unlike adults, who have fully-developed executive functions, Theory of Mind skills, and language abilities, children in their preschool years are undergoing development in all three of these areas, as well as in the development of scalar implicatures. Studying children in this age group can provide insight into the extent to which these abilities develop in tandem but independently or, alternatively, scaffold each other (Saxe, 2006).

We propose a novel approach to investigating the mechanisms employed during preschool children's derivation of scalar implicatures: spatial neuroimaging. Behavioral research on scalar implicature, language, executive function, and Theory of Mind development can only go so far in evaluating the separate contributions of these skills. Behavioral research can identify a correlation between two abilities (e.g., scalar implicature derivation and Theory of Mind) but cannot assess whether one of those processes (Theory of Mind) is engaged during the other (scalar inference). Some behavioral response types for scalar implicature, such as binary truth value judgement tasks, have been criticized for being too metalinguistic for small children, making it difficult to assess their true abilities (Katsos & Bishop, 2011); similar criticisms have been made of Theory of Mind tasks for being too short (only one or two questions) and too focused on false beliefs (Apperly, 2012; Matthews, Lieven, Theakston, & Tomasello, 2018). Neuroimaging methods can directly look at online neural activation during these processes without influence from response type or length. Additionally, the existing literature on children's scalar implicatures ability has been criticized for not being able to separate out whether children who succeed at young ages on simple tasks are truly undergoing the full process of scalar implicature derivation (as adults in the task presumably are), or whether they are able to give the correct response through less complex reasoning processes (Katsos & Bishop, 2011). This debate is unlikely to reach a resolution because behavioral studies can only indirectly reveal the mechanisms underlying pragmatic processing in preschool-aged children. Investigating the neural systems that underlie the development of these abilities in preschool-aged children has the potential to answer some of these questions directly.

4.1.1 Neural Systems and Scalar Implicature

Neuroimaging studies of scalar implicature and other types of pragmatic reasoning in adults have indicated the recruitment of frontal areas such as the left inferior frontal gyrus (LIFG), a region associated with higher level linguistic processing, as well as the left anterior Middle Frontal Gyrus (MFG) and the Medial Frontal Gyrus (MeFG), two areas associated with executive functions (Shetreet, Chierchia, & Gaab, 2014; Politzer-Ahles & Gwilliams, 2015). To our knowledge, one study so far has studied the neural activation of scalar implicature derivation in children using spatial neuroimaging methods (Shetreet et al., 2014). In this study, 6-year-olds viewed a scene and heard a sentence describing it; the sentence contained a quantifier (*some*, *none*, *all*) and either matched the picture or did not (e.g., “Some elephants are dancing” when the picture either depicted SOME or ALL of the animals dancing). Like adults, 6-year-olds displayed activation in the LIFG in contexts that had the quantifier *some*. Nonetheless, it is valuable to study children who are still in the critical stages of development for scalar implicatures (ages 3 to 5), as well as the abilities that support this development. There have been additional studies on online scalar implicature derivation using EEG/ERP (electroencephalogram/event-related potential) recording in both adults and children (Noveck & Posada, 2003; Politzer-Ahles, Fiorentino, Jiang, & Zhou, 2013; Spychalska, Kontinen, & Werning, 2014; Hartshorne, Snedeker, Liem Azar, & Kim, 2015); while useful, these studies cannot provide the detailed spatial information required to assess which specific networks (Theory of Mind, executive functions, etc.) are active during the derivation process given the limitation of EEG/ERP to localize neural activity.

Concerning the relationship between scalar inference and Theory of Mind, recruitment of the dorsomedial prefrontal cortex (dmPFC) and the temporo-parietal junction (TPJ), which are involved in perspective taking and Theory of Mind, has been indicated during communicative processes (Sassa et al., 2007; Willems et al. 2010), although not specifically with scalar implicature derivation. For Theory of Mind more broadly, spatial neuroimaging studies investigating children's competency are very limited in number. An fMRI study by Gweon and colleagues (2012) investigated the activation of Theory of Mind neural networks in children between the ages of 5-11 while children listened to stories that had 'mental' descriptions (e.g., "Mary thought that Sarah…"), 'physical' descriptions ("Mary walked to the store…"), or 'social' descriptions (e.g., "Mary played the flute, and everybody at the party danced…"). Children displayed comparable activation to adults in the right temporo-parietal junction (RTPJ), left temporo-parietal junction (LTPJ), prefrontal cortex, and the medial prefrontal cortex (mPFC) for the mental stories compared to the physical stories. Additionally, participants displayed increasing selectivity in bilateral TPJ to mental state information with age; children with more advanced development of Theory of Mind displayed more specialized recruitment of Theory of Mind networks.

There are considerably more neuroimaging studies investigating executive functions in children; the literature becomes extensive beyond the age of 7 (Cantlon, Brannon, Carter, & Pelphrey, 2006; Ciesielski, Lesnik, Savoy, Grant, & Ahlfors, 2006; Dibbets, Bakker, & Jolles, 2006; O'Hare, Lu, Houston, Bookheimer, & Sowell, 2008; see Houdé, Rossi, Lubin, & Joliot, 2010 for a meta-analysis). Studies using adolescent participants have found activation in the prefrontal cortex, consistent with the neuroimaging literature on executive functions in adults (Schulz et al., 2004; Smith et al., 2006; Carrion, Garrett, Menon, Weems, & Reiss 2007; Olesen, Macoveanu, Tegner, & Klingberg, 2007; Crone, Zanolie, Van Leijenhorst, Westenberg, & Rombouts, 2008; O'Hare et al., 2008). Children also display recruitment of frontal networks, but with less consistency and in more limited areas (Bunge, Dudukovic, Thomason, Vaidya, & Gabrieli, 2002; Crone, Donohue, Honomichl, Wendelken, & Bunge, 2006a; Crone, Wendelken, Donohue, Van Leijenhorst, & Bunge, 2006b; Dibbets et al., 2006; Crone et al., 2008; O'Hare et al., 2008), for example, activation specifically in the left ventral prefrontal cortex instead of the full network including other frontal and parietal regions for working memory in 7- to 9-year-olds (O'Hare et al., 2008).

In sum, existing neuroscience literature on the development of scalar implicature, Theory of Mind, and executive functions systems studies provides insight into the abilities of older children. However, research is needed on children who are still in the primary developmental stages of these cognitive areas (ages 3 to 5). Work with this target age range, however, has been limited by the challenges of using fMRI methods, which can be intimidating and too restrictive of movement for young children. This problem can be resolved by using functional Near Infrared Spectroscopy (fNIRS) neuroimaging (Gallagher, Béland, & Lassonde, 2012; Quaresima, Bisconti, & Ferrari, 2012; Fishburn, Norr, Medvedev, & Vaidya 2014; Nishiyori, 2016; Walsh et al., 2017; Jasińska et al., 2021). fNIRS is a neuroimaging method that utilizes light absorption and reflection to measure hemodynamic activity. Participants wear a cap on their head which connects to a set of light emitters and detectors. Given that oxygenated and deoxygenated blood reflect and absorb light differently, by measuring the amount of light that is reflected at a given wavelength, it is possible to detect the location of neural activity. This method is child-friendly because it can compensate for movement, and it is less intimidating for young children to wear a cap on their head than to be placed in an MRI chamber (Gallagher, Béland, & Lassonde, 2012; Quaresima, Bisconti, & Ferrari, 2012; Nishiyori, 2016).

4.1.2 Current Study

In the present study, we used fNIRS methods to investigate the contributions of Theory of Mind, executive functions, and language systems to scalar implicature derivation during development and adulthood. Child participants completed a behavioral individual differences task battery (with Theory of Mind, executive functions, and language tasks) as well as two neuroimaging tasks using fNIRS; adult participants completed only the neuroimaging tasks. The neuroimaging tasks consisted of a scalar implicature task and a Theory of Mind task (to directly compare neural activation for mentalizing to activation in the scalar implicature task). To our knowledge, this is the first time pragmatic inference in children between the ages of 3.5 to 5 has been studied using spatial neuroimaging techniques.

Our primary focus was to investigate which supporting networks (Theory of Mind, executive functions, language) are active during scalar implicature derivation for both young children and adults. Of particular theoretical interest was the contribution of Theory of Mind to scalar implicature derivation, since this component allowed us to test the diverging predictions of prominent pragmatic theories (see section 1.0). Given that pragmatic development is still ongoing in the age group that we tested, we were also interested in identifying differences in children who successfully derive scalar implicatures compared to children who do not. The novel approach of using neuroimaging methods to address questions raised in previous behavioral research provides sa unique perspective into existing debates on the nature of scalar implicature derivation.

4.2 Methods

4.2.1 Participants

Thirty-five preschool children (M_{age} = 4;10, *range:* 3;9 to 6;3, 19 female) and 27 adults (undergraduate students; 20 female) participated. Three children who participated dropped out of the study before completing any of the neuroimaging tasks and were excluded from analysis. Children were recruited from the Newark (DE) and Laurel (DE) areas through local postings and targeted ads on Facebook. They received a small gift for participating, and their parents received monetary compensation for their travel. Adult participants were undergraduate students at the University of Delaware and received course credit for their participation in the study. All participants were right-handed, monolingual English speakers.

G*Power software, version 3.1.9.6, was used to compute power (Faul, Erdfelder, Buchner & Lang, 2009). Power calculations assuming α=.05, power (1-β)=.8, one predictor (condition), indicate a sample of 34 is sufficient to detect an effect size of 0.25, which is considered small (Cohen, 1992). A recent review of fNIRS studies found that the median sample size of published studies (included in the review) was 20 (Kohl et al., 2020).

We collected socioeconomic status, demographic, and language background information for our child participants. The majority of children were white and had college-educated mothers. Parents completed an SES ladder survey (MacArthur Scale of Subjective Social Status; Adler et al., 2000) in which they reported how they viewed their standing (income, education, respect) compared to 1) other members of their community, and 2) other people in the United States of America. Our participants' parents reported a slighter higher than average SES ladder score (*Range:* 0-10; $M_{community}$ = 6.5, $SD_{community}$ = 1.77; M_{USA} = 6.13, SD_{USA} = 1.31).

4.2.2 Behavioral Individual Differences Battery (Children)

For child participants, the study began with a short play period, intended to help children feel more comfortable in the lab space, followed by a behavioral test battery, which consisted of a Theory of Mind task (Baron-Cohen et al., 1985), a vocabulary task (Gershon et al., 2014), and an executive functions task (Zelazo, Frye, & Rapus, 1996).

4.2.2.1 Theory of Mind Task (Sally-Anne)

First, children completed the Sally-Anne task (Baron-Cohen et al., 1985), a classic task designed to test Theory of Mind ability. In this task, the child was told a story with dolls in which one of the characters has a false belief about where an item is located, and the child was asked to indicate where the character will look for that item.

4.2.2.2 Vocabulary Task (Picture Vocabulary Test)

Next, participants completed two tasks from the NIH Toolbox Cognition Battery using the NIH Toolbox iPad app (Weintraub et al., 2013). Participants first completed the Picture Vocabulary Test (Gershon et al., 2014). In this 5-minute task, children were presented with four images on a screen, heard an audio-recording of a word, and had to select which image the word was describing; one child did not finish this task.

4.2.2.3 Executive Functions Task (Dimensional Change Card Sorting Task)

Participants then completed the Dimensional Change Card Sorting task (DCCS; Zelazo et al., 1996), also using the NIH Toolbox app. In this 10 minute task, children had to learn to match pictures based on color, then switch to matching them based on shape, then match them either based on color or shape after hearing an audio-recording of either the word “color” or “shape” before each trial. The DCCS task has been shown to have high convergent validity with other executive functions tasks (Zelazo et al., 2014), such as the Flanker task (Eriksen & Eriksen, 1974), as well as more comprehensive measures of executive functions, such as the D-KEFS (Delis, Kaplan, & Kramer, 2001). Four children lost patience with the length of the behavioral battery and DCCS task specifically and did not complete the final DCCS task but were still included in analyses for the tasks they did complete.

4.2.3 Neuroimaging Tasks (Adults and Children)

Participants completed two neuroimaging tasks: a scalar implicature task, adapted for neuroimaging (Kampa & Papafragou, 2020), and a Theory of Mind task (Gweon et al., 2012). Children always performed the tasks in this order, in the interest of prioritizing the scalar implicature task if children did not complete both measures ($N = 4$). The order of tasks for adults was counterbalanced. Participant responses in the neuroimaging tasks were collected using a button-box (Cedrus Inc). Prior to beginning the neuroimaging tasks, children completed training on how to press the buttons in response to stimuli on the screen.

During the fNIRS testing, the hemodynamic response was measured with a Shimadzu LightNIRS Near Infrared Spectroscopy system acquiring data at 7.4 Hz. The lasers and detectors were arranged in a 3x10 array resulting in 47 channels (see Figure 7). This array was chosen to allow for coverage of frontal regions previously indicated to be active during pragmatic inference (e.g., LIFG; see section 1.1) as well as to allow for the most general coverage of relevant executive functions, Theory of Mind, and language areas possible given that fNIRS has restrictions on the amount of total surface area that can be covered by the array. The lasers were factory set to 780, 805, and 830 nm. The 15 lasers and 15 detectors were segregated into alternating grid placement. Positioning of the array was accomplished using the 10–20 system (Jasper, 1958) to maximally overlay the key regions of interest (for additional details, see Shalinsky, Kovelman, Berens, & Petitto, 2009; Jasińska & Petitto, 2013, 2014; Jasinska, Berens, Kovelman, & Petitto, 2017). The depth of recording in the cortex is approximately 3 cm. Prior to recording, every channel was tested for optimal signal to noise ratio using LightNIRS inbuilt software.

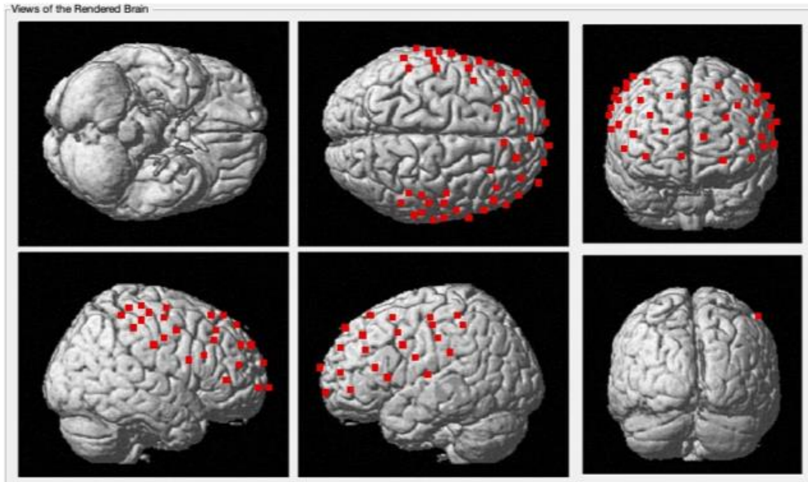

Figure 7: The arrangement of the 3x10 array (Panel 1a) on a sample adult participant.

4.2.3.1 Scalar Implicature Task

After completing the behavioral individual differences battery but before NIRS cap placement, child participants completed an introductory phase in order to become familiar with the types of displays used in the test trials for the task. This introductory phase was identical to the "box explanation" and pretest trials in Kampa & Papafragou, 2020 (see that paper for details). After completing the two pretest trials, children were told that they would continue the game later (after NIRS cap placement). After the fNIRS cap was placed on the participants' head and the probes were attached, children participated in a practice session during which they were trained on how to use a button-box to indicate their answers (the button on the left = the box on the left; the button on the right = the box on the right). Adults received written instructions only and did not complete the practice session.

The scalar implicature task was adapted from a paradigm used previously to test 4- and 5-year-olds' scalar inference derivation (Kampa & Papafragou, 2020). For this task, participants viewed two photos placed side by side on a computer screen (see Figure 8, Panel A for an example). Each photo depicted a person ("the experimenter's friend") sitting across from and facing out towards the participant (for example, the girl in Panel A). In front of her was a cardboard box with two vertical compartments. In the "full-access" box photo, both vertical compartments of the box were open and see-through, thus the participant and the girl both had full visual access to the box. In the "limited-access" box photo, the compartment to the left (from the participant's viewpoint) was see-through, like the compartments in the full-access box. However, the compartment to the right was blocked such that only the participant but not the girl could see its contents. The contents of the two boxes were identical (e.g., a penguin and a pumpkin), although one of the objects in the limited-access box was not visible to the experimenter's friend (here, the pumpkin; Figure 8). After both photos were presented, children heard a recorded sentence from a female speaker ("I see…") and were asked, "Which box is she talking about?" and had to select one of the photos.

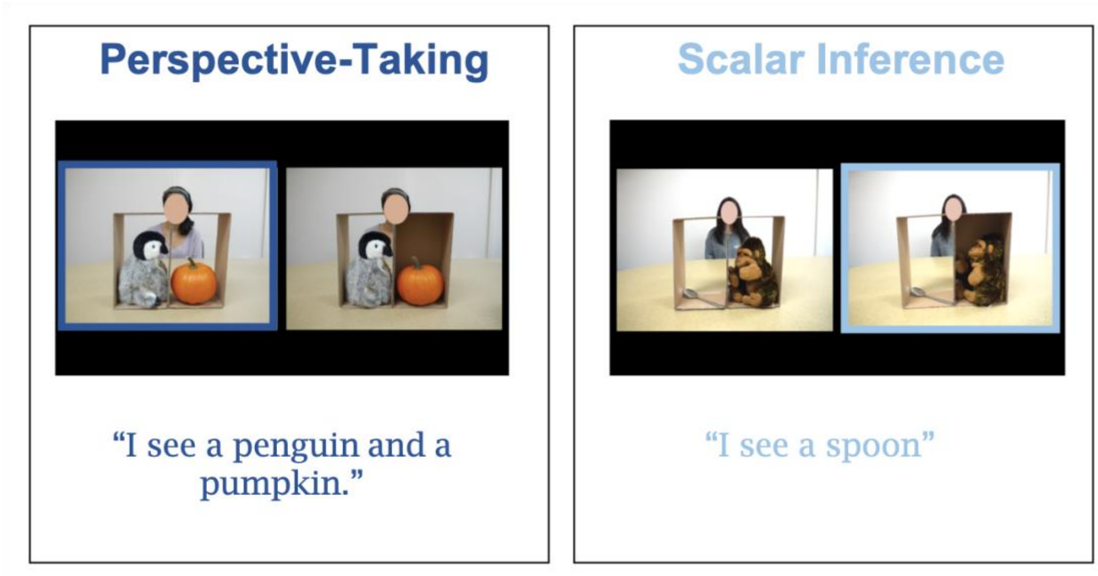

Figure 8: An example trial from each of the two conditions: Perspective-Taking, Scalar-Inference. In each trial, participants saw each photo on the screen appear (with a 2s delay in between) then heard the sentence indicated and were expected to select one photo (correct answer highlighted here). The photo on the left within each image depicts the "full-access" box (and the full-knowledge speaker). The photo on the right depicts the "limited-access box" (and the limited-knowledge speaker). The faces have been blurred in this image per University of Delaware policy but were not blurred for the experiment.

There were two within-subjects conditions: Perspective-Taking and Scalar-Inference. (These conditions correspond to the conditions named More-Informative and Less-Informative, respectively, in Kampa & Papafragou, 2020). The Perspective-Taking and Scalar-Inference conditions were designed to be minimally different from each other. Both conditions used the same photos (counterbalanced between participants) and required the participant to evaluate the visual perspective of the speaker. The target Scalar-Inference condition differed from the Perspective-Taking condition only in that the speaker's statement uttered provided less information ("I see a penguin" vs. "I see a penguin and a pumpkin" respectively). A successful responder in the Perspective-Taking condition should recognize that only the speaker with full visual access could produce the more informative statement ("I see a penguin and a pumpkin"), since the speaker with limited knowledge could not see one of the objects; the responder should thus take the full-access box to be what the speaker is talking about. A successful responder in the Scalar-Inference condition should recognize that a communicator with full knowledge should not fail to present complete and relevant information (i.e., the full contents of the box), and thus match the less informative statement ("I see a penguin") to the speaker with limited knowledge, hence picking the limited-access box as the one being talked about (see Papafragou, Friedberg, & Cohen, 2018 or Kampa & Papafragou, 2020 for a more detailed explanation of this logic).[2]

The task used a block-design. Each participant completed 2 runs lasting approximately 5 minutes each. Each run consisted of 6 blocks of test trials (2 Perspective-Taking, 2 Scalar-Inference, 2 Control blocks) in a counterbalanced order. Each block contained 4 trials (12 seconds each) of the same type. For each trial, the photo on the left appeared on screen first, followed by the photo on the right after 2 seconds (presentation order for the limited access box photo was counterbalanced across participants). After both photos were presented, children heard a recorded sentence from a speaker matched to the person in the photo ("I see…"), then a recording from a distinct narrator voice asking, "Which box is she talking about?" Participants were given 6 seconds after the audio ended to indicate their response by

[2] Child participants also completed a control condition which did not produce any results of note and will not be discussed in this paper (see supplementary materials for details).

pressing the appropriate button; both photos remained on screen for those 6 seconds. For variety, the people in the pictures (and their corresponding recorded voices) differed across blocks; each "friend" was depicted in two blocks total (always different types) across the experiment. The interstimulus interval (ISI) between each block was 15 seconds long; adults saw a black screen with a fixation cross, children saw a video of jellyfish moving (Run 1) or small, colorful dots moving (Run 2).

4.2.3.2 Theory of Mind Task

Children completed a neuroimaging Theory of Mind task adapted from a study by Gweon and colleagues (2012; see section 1.1). There were two within-subjects conditions: Mental and Physical. In the Mental condition, participants listened to stories that described a protagonist's mental state using mental state verbs; in the Physical condition, participants listened to stories that described physical events, objects, and states (see Table 9). The stories were 20 seconds in length and matched for number of words (*M* = 51.6 words), number of sentences (4.7), and Flesch Reading Ease Level (M = 90.4). After each story, participants were asked, "Does this come next?" followed by a probe sentence that either matched the story or came from an unrelated story (4 seconds). Participants indicated their judgement by pressing either a green button for "yes" or a red button for "no" (6 seconds). The question was designed to verify attention; of interest was the neural activation while listening to the two types of stories, not the participants' behavioral responses.

Table 9: Example Stimuli for fNIRS Theory of Mind Task

Condition	Example story	Example probe
Mental	"One day a pirate told Jimmy about a hidden treasure. The pirate **thought** that the treasure was buried behind Jimmy's house. Jimmy **believed** him. So Jimmy dug a big hole behind his house, but he didn't find a treasure. Jimmy soon **realized** the pirate didn't **know** where the treasure was."	"Jimmy was mad. He had done all that work for nothing."
Physical	"One day, a little robin landed on a huge, strange-looking tree, and ate one of its berries. This was a magical tree that had special powers. In the spring, the robin laid three eggs. Soon, two of the eggs cracked and little robins came out. But the last egg did not crack for a long time."	"When the last egg cracked a beautiful firebird came out."

Each run consisted of 4 trials in a palindromic, counterbalanced order (Mental-Physical-Physical-Mental and vice versa) with a 12 second rest before each trial and lasted approximately 2 minutes and 48 seconds. Adults completed 4 runs; children were only able to consistently complete 2 runs.

4.3 Hypotheses and Analysis Plan

4.3.1 Behavioral Tasks: Individual Differences and Scalar Implicature Performance (Children)

As discussed previously (see section 2.3.1) and based on the results of past findings (Kampa & Papafragou, 2020), we predicted that pragmatically-sophisticated children in the scalar implicature task would match a more informative statement with a full-access box and a less informative statement with a limited-access box, in accordance with the expectation that a speaker should be as informative as their knowledge allows them to be. Of interest was whether children's performance on the Language, ToM, and EF components of the individual differences behavioral battery would predict their behavioral performance on the scalar implicature task.

4.3.2 Neuroimaging Tasks (Children and Adults)

4.3.2.1 Scalar Implicature

The goal of the neural measures of the scalar implicature task was to assess which networks are active during scalar inference in children and in adults (see section 1.2). Group activation maps were generated comparing the target Scalar-Inference condition to the Perspective-Taking condition. Note that these conditions were carefully matched in terms of perspective taking and visual detail such that removing activation for the Perspective-Taking condition from the Scalar-Inference condition (i.e., through standard neuroimaging subtraction analysis) yielded activation specific only to the process of scalar inference. This task was specifically chosen so that any activation present after the subtraction could not be attributed to perspective taking or other mentalizing processes independent of the scalar inference.

Of particular theoretical interest was whether Theory of Mind networks would be active during scalar inference in both age groups. If Theory of Mind is involved in the scalar inference derivation process, we would expect to find activation of neural networks associated with Theory of Mind, specifically activation in the dorsomedial prefrontal cortex (dmPFC; other regions associated with Theory of Mind, such as the rostral anterior cingulate cortex, the medial posterior parietal cortices, and the bilateral temporal parietal junction would not be not covered by the depth and surface area of the fNIRS array chosen for this study).

We also predicted that executive functions and language networks would be active during scalar inference. If executive functions are involved in scalar inference derivation, we would expect to find activation in the prefrontal cortex (PFC), possibly the dorsolateral PFC specifically (see Diamond, 2002 for a review). However, it is worth mentioning that some studies have found that young children still developing executive functions rely on sensory-motor pathways, including regions such as the premotor and parietal cortex (Ciesielski et al., 2006; Simmonds et al., 2007; Suskauer et al., 2008). If language is involved in scalar inference derivation, we expect to find activation in the Left Inferior Frontal Gyrus (LIFG). This is a region associated with higher level linguistic processing (particularly Broca's area, a subregion of the LIFG) that was found to be active during scalar inference in prior studies (Sheetreet et al., 2014).

4.3.2.2 Theory of Mind

Group activation maps were generated for "mentalizing" (equivalent to the subtraction of the Physical condition from the Mental condition in this task). Given previous results (Gweon et al., 2012), we expected that adult participants would display activation in the dorso-medial pre-frontal cortex (dmPFC), the only relevant Theory of Mind region covered by the array. Of interest was whether activation of Theory of Mind networks for children would correspond with their scalar inference development. Recall that previous research indicates that, as children develop Theory of Mind abilities, they display more selective recruitment of Theory of Mind regions (Saxe et al., 2009; Gweon et al., 2012). If Theory of Mind is a prerequisite of scalar implicature derivation, children with more developed scalar implicature ability should display more developed Theory of Mind, indicated through specialized recruitment of Theory of Mind networks.

4.3.2.3 Scalar Implicature and Theory of Mind (Conjunction Analysis)

If there is a strong relationship between mentalizing and scalar inference, we expect at least some overlap in the neural activity for these two processes, particularly in the dmPFC or other known Theory of Mind regions.

4.4 Results and Discussion

4.4.1 Behavioral Tasks: Individual Differences and Scalar Implicature Performance (Children)

Given that the behavioral study from which this paradigm was adapted found a bimodal distribution of responses, with children either succeeding or failing across the board (Kampa & Papafragou, 2020), we used a data-driven approach (k-means cluster; Wagstaff, Cardie, Rogers, & Schrödl, 2001) to analyze behavioral scalar implicature results and determine if there were distinct patterns of responding (e.g., pass vs. fail). This analysis identified two distinct groups of children: 20 scalar implicature passers (who achieved at least 75% accuracy in the Scalar Inference Condition; $M_{passers}$ =

95.9%, *range:* 82.1% to 100%; M_{age} = 4.85) and 15 scalar implicature non-passers ($M_{failers}$ = 55.8%, *range:* 30% to 73.8%; M_{age} = 4.89).[3]

To analyze the relationship between the behavioral individual differences tasks and performance on the scalar implicature task, we first conducted a set of basic analyses (two-sample t-tests for scale data; Fisher's exact test for categorical data) to compare between the two groups identified by the k-means cluster analysis (passers and non-passers); we found no significant differences between passers and non-passers in any of the behavioral battery tasks (*Language* (*Vocabulary)*: $M_{Passers}$ = 63.42, $M_{Non\text{-}passers}$ = 60.87; $t(32) = 1.43$, $p = .16$; *Executive Functions* (*DCCS)*: $M_{Passers}$ = 56.71, $M_{Non\text{-}passers}$ = 53.07; $t(29) = .56$, $p = .58$; *Theory of Mind* (*Sally-Anne)*: Fisher's exact test, $p = .56$; see Figure 9).

In order to assess whether performance in any of the behavioral individual differences tasks predicted behavioral scalar implicature task performance, the data were analyzed using multi-level logistic mixed effects modeling. Accuracy in the target condition (Scalar-Inference) of the scalar implicatures task was analyzed with a model that included Age (continuous), Language (Vocabulary), Executive Functions (DCCS), and Theory of Mind (Sally-Anne) scores as fixed predictors and random intercepts for subject and item. Interactions that did not significantly improve model fit were removed from the final model, see Table 10. The model revealed no significant effects or interactions.

The finding that behavioral individual differences measures failed to predict children's scalar implicature performance is reminiscent of previous studies (see

---

[3] The results do not change if we define passers in terms of accuracy in both the Scalar Inference and the Perspective-Taking Conditions.

Matthews et al., 2018), and motivates the need to look at the neural signatures of scalar implicature derivation to gain additional evidence for the mechanisms underlying this pragmatic phenomenon.

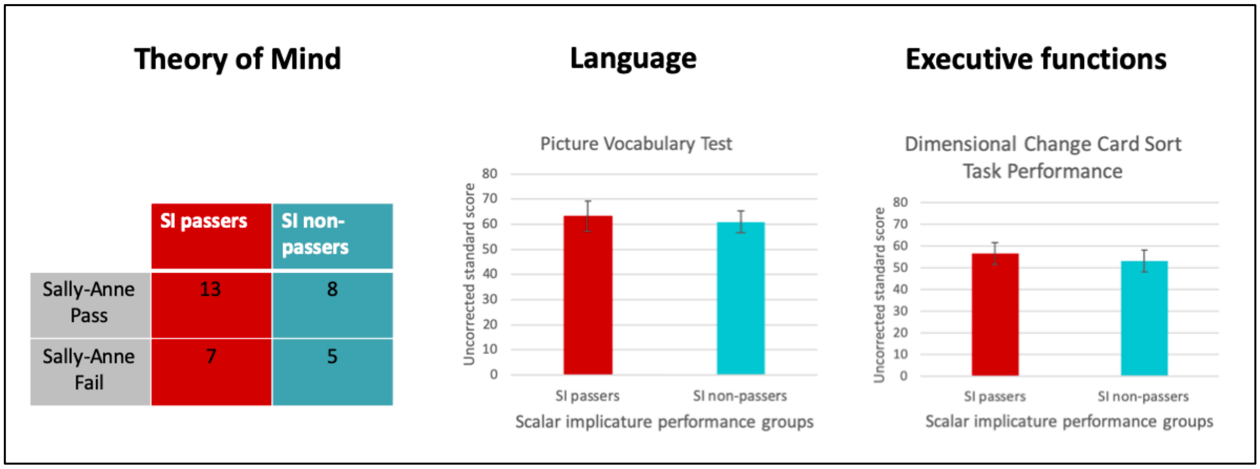

Figure 9: Performance in the behavioral individual differences battery tasks for passer and non-passer scalar implicature groups. Error bars display +/- 1 SE.

Table 10: Parameter Estimates for Behavioral Accuracy in the Scalar Implicature Task

Effects	Estimate	SE	*z*
Intercept	6.010	5.235	1.148
Age (continuous)	-0.824	1.064	-0.775
Language	0.207	0.128	-1.614
Executive Functions	0.0282	0.032	0.884
Theory of Mind	0.508	1.136	0.448
Age (continuous):Language:Theory of Mind	0.036	0.050	0.722
Language:Executive Functions:Theory of Mind	-0.014	0.015	-0.923

Significance levels: * $p < .05$, ** $p < .01$, *** $p < .001$.

4.4.2 Neuroimaging Tasks (Adults and Children)

4.4.2.1 Preliminaries: fNIRS Data Preprocessing, Spatial Registration, and Statistical Parametric Mapping

Data were analyzed using a Matlab-based NIRS-SPM Version 4 (Jang et al., 2009; Ye, Tak, Jang, Jung, & Jang, 2009), which uses the neuroimaging suite SPM12 (Nichols, 2013; Flandin & Novak, 2020). Using the modified Beer–Lambert equation, NIRS-SPM converts optical density values into concentration changes in oxygenated and deoxygenated hemoglobin response (HbO and HbR, respectively). Changes in HbO and HbR concentrations were filtered with a HRF filter and decomposed using a Wavelet-Minimum Description Length (MDL) detrending algorithm in order to remove global trends resulting from breathing, blood pressure variation, vasomotion, or participant movement artifacts and improve the signal-to-noise ratio (Jang et al., 2009). We filtered out step functions that were identified in each timeseries. This allowed us to correct for motion artifacts as well as drift in signal related to respiration.

NIRS channels were registered to MNI space with the Haskins Pediatric Brain Atlas (Molfese, Glen, Mesite, Pugh, & Cox, 2015) in NIRS-SPM's stand-alone registration function (Singh, Okamoto, Dan, Jurcak, & Dan, 2005) by using a three dimensional digitizer (Polhemus Corp.). Registration was done individually for each child. The spatial registration function yielded MNI coordinates represented by each channel with corresponding labels for anatomical regions, including Brodmann labels, maximally located at each channel position. Specifically, the function provides a coverage percentage for a given anatomical region at each channel.

We used a general linear model based analysis approach that allows for the creation of activation maps with super-resolution localization and applied Sun's tube correction for multiple comparisons. Models for HbO and HbR contain experimental regressors convolved with the corresponding hemodynamic response function with time derivatives. NIRS-SPM (Statistical Parametric Mapping) creates the models for HbO and HbR with opposing polarity so that a significant model fit for HbO indicates increased concentration and for HbR decreased concentration.

4.4.2.2 Scalar Implicature

Beginning with the adult data, regions that were found to have significant HbO activation during scalar inference for adults are listed in Table 11 (see also Figure 10). Regions of note include the left inferior frontal gyrus (LIFG; specifically, Broca's area), the dorsolateral pre-frontal cortex (dlPFC), and the dorso-medial prefrontal cortex (dmPFC). Other regions of activation include Wernicke's area, the primary somatosensory cortex, and the pre-motor and supplementary motor cortex.

Table 11: Regions with Significant HbO Activation for Scalar Inference for Adults

Contrast	Region	Brodmann Area	X	Y	Z	β value	p value
SI > PT	Broca's area	45	-16.3	52.3	44.3	0.00259	<0.05
SI > PT	Primary Somatosensory Cortex	1/2/3	64.7	-19	46	0.00238	<0.05
SI > PT	Left dorsolateral PFC	9	-47.3	4.3	55.7	0.00230	<0.05
SI > PT	Dorsomedial PFC	9	-10.3	40.3	56.7	0.00222	<0.05
SI > PT	Pre-motor and supplementary motor cortex	6	64.6	-19	46	0.00189	<0.05
SI > PT	Pre-motor and supplementary motor cortex	6	-53	9.7	44.7	0.00165	<0.05
SI > PT	Dorsomedial PFC	9	-16.3	52.3	44.3	0.00162	<0.05
SI > PT	Right Dorsolateral PFC	46	49	53.3	-5.7	0.00155	<0.05
SI > PT	Wernicke's area	40	-47.3	-37.3	63.3	0.00153	<0.05
SI > PT	Pre-motor and supplementary motor cortex	6	-47.3	4.3	55.7	0.00153	<0.05

* This table displays the ten channels with the highest level of significance; the rest can be found in supplementary materials. PFC refers to the prefrontal cortex.

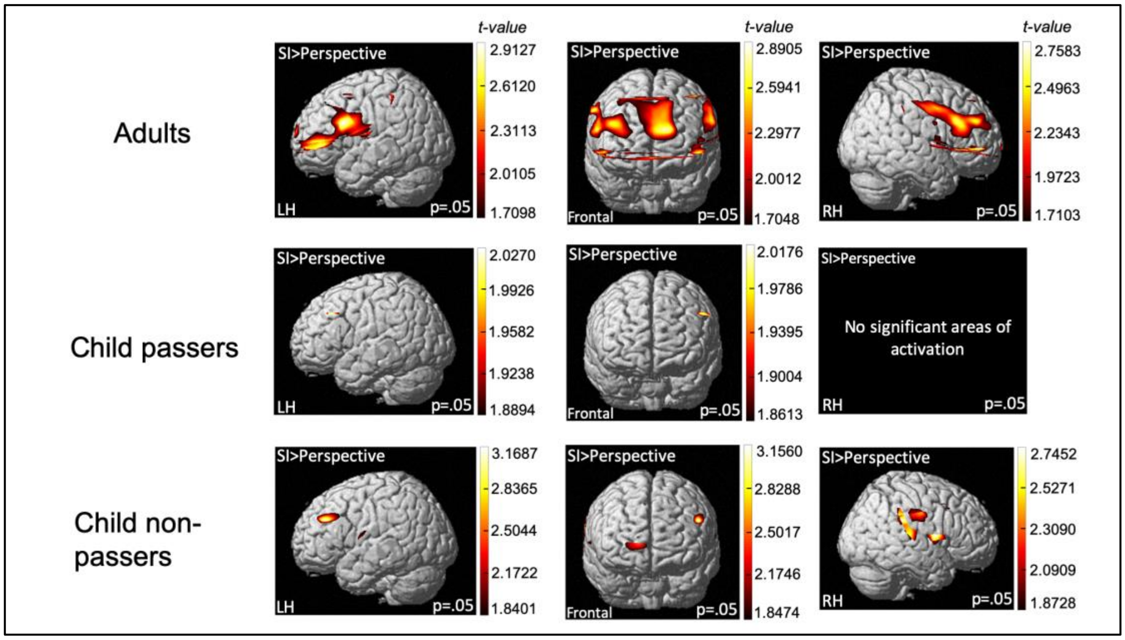

Figure 10 Activation patterns in the Scalar Implicature task for adults, child passers, and child non-passers. Activation is shown for HbO for adults and HbR for children (HbR for adults and HbO for children can be found in supplementary materials).

For children, nine passers and nine non-passers had usable fNIRS data for the scalar implicature task. While adults showed canonical patterns of HbO and HbR concentrations (increased HbO concentrations relative to HbR), children showed the inverse patterns (increased HbR relative to HbO). This phenomenon has been observed frequently with infants and very young children; evidence suggests that more complex stimuli and cognitive processes trigger this inverse shape of HbO and HbR responses (Taga et al., 2003a, 2003b; Grossmann et al., 2010a, 2010b; Watanabe et al., 2012; Issard and Gervain, 2017; see Issard and Gervain, 2018 for a review). It has been proposed that children display inverse HbO and HbR patterns when stimuli are particularly difficult for the tested developmental stage (Issard and Gervain, 2018).

Typically, inverse responses give way to canonical responses over the course of development (Emberson et al., 2017a, 2017b) as children develop more familiarity with and facility for a given ability.

Regions that were found to have significant HbR activation during scalar inference for child passers and child non-passers are listed in Table 12 and Table 13, respectively. Child passers displayed activation in key regions such as the left inferior frontal gyrus (specifically Broca's area), the dorsolateral pre-frontal cortex (dlPFC), and the dorso-medial prefrontal cortex (dmPFC), similar to adults. Child non-passers displayed the highest level of activation in the superior temporal gyrus (STG) and middle temporal gyrus (MTG). There was some overlap in significantly active regions for child non-passers with the child passers and adults (dlPFC, Wernicke's area, and the pre-motor and supplementary motor cortex). However, child non-passers did not display significant levels of activation in the LIFG or the dmPFC.

Table 12: Regions with Significant HbR Activation for Scalar Inference for Child Passers

Contrast	Region	Brodmann Area	X	Y	Z	β value	*p* value
SI > PT	Broca's area	45	48.7	48.3	20.3	0.0028	<0.05
SI > PT	Broca's area	45	57	34.3	-4.3	0.0025	<0.05
SI > PT	Dorsomedial PFC	9	3.33	57	39.7	0.0020	<0.05
SI > PT	Broca's area	45	49.3	38.3	31.3	0.0011	<0.05
SI > PT	Wernicke's area	40	-63	-46.7	42.7	0.0010	<0.05
SI > PT	Left Dorsolateral PFC	46	-50	48	-3	0.0006	<0.05
SI > PT	Right Dorsolateral PFC	46	47	56.3	-4.7	0.0005	<0.05
SI > PT	Pre-Motor and Supplementary Motor Cortex	6	-64.7	2.3	25.7	0.0005	<0.05
SI > PT	Broca's area	44/45	48.7	28.7	42.7	0.0004	<0.05
SI > PT	Wernicke's area	40	67	-29.7	45.7	0.0004	<0.05

*This table displays the ten channels with the highest level of significance; the rest can be found in supplementary materials. PFC refers to the prefrontal cortex.

Table 13: Regions with Significant HbR Activation for Scalar Inference for Child Non-passers

Contrast	Region	Brodmann Area	X	Y	Z	β value	*p* value
SI > PT	Superior temporal gyrus	22	71	-34.3	23.7	0.0034	<0.05
SI > PT	Middle temporal gyrus	21	73	-26.3	-7.7	0.0029	<0.05
SI > PT	Right Dorsolateral PFC	46	46	50.7	22.7	0.0025	<0.05
SI > PT	Wernicke's	40	68	-42.3	32.7	0.0025	<0.05
SI > PT	Pre-Motor and Supplementary Motor Cortex	6	-64	4.3	25.3	0.0022	<0.05
SI > PT	Medial temporal area	48	-57.3	16.7	1.7	0.0021	<0.05
SI > PT	Frontopolar area	10	-11.3	74	5.3	0.0020	<0.05

*PFC refers to the prefrontal cortex

In order to probe further into the profile of the children who succeed behaviorally on the scalar implicature task, we used a data-driven approach (k-means cluster; Bonomini et al., 2015) based on the neural activation beta statistics (HbO and HbR for Left, Right, and Center) to determine whether there were distinct patterns of responding within the passers group. We found that the best fit for the data was to divide the child passers into two distinct groups. Although behaviorally identical, only some of the child passers displayed a distinct and congruent group neural profile (Figure 11; Table 14). That group of child passers (Group 1) displayed similar patterns of neural activation to adults during scalar implicature derivation (dmPFC and dlPFC), but the other group (Group 2) displayed no cohesive significant neural patterns during scalar implicature. It should be noted that even the Group 1 child passers, who displayed activation in adult-like regions, displayed inverse HbR and HbO responses, potentially indicating that these stimuli are still difficult for this group at their developmental level. A set of simple tests (two-sample t-tests for scale data; Fisher's exact test for categorical data) revealed no significant differences in performance on behavioral tasks between the two subgroups (*Age:* M_{G1} = 4.71, $Range_{G1}$ = 4;1 to 5;10 ; M_{G2} = 5.12, $Range_{G2}$ = 4;2 to 6;1 ; *Language (Vocabulary)*: M_{G1} = 67, M_{G2} = 63; $t(7) = -0.75$, $p = 0.48$; *DCCS*: M_{G1} = 69, M_{G2} = 50; $t(7) = 1.90$, $p = 0.11$; *Theory of Mind* (*Sally-Anne)*: Fisher's exact test, $p = .41$).

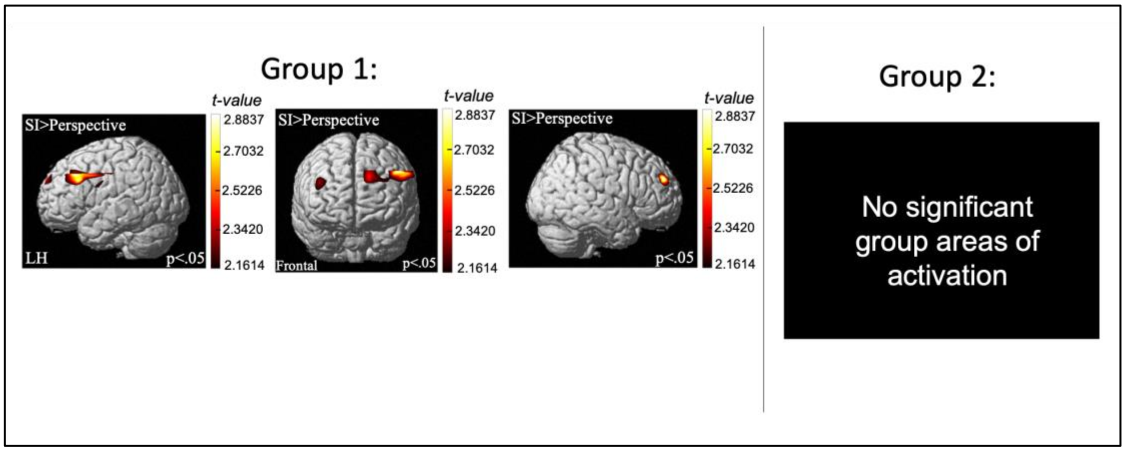

Figure 11: Comparison of child passer subgroups in the Scalar Implicature task. Group 1 displayed activation in the dorsomedial PFC and dorsolateral PFC, similar to adults. Group 2 displayed no cohesive group patterns of activation.

Table 14: Regions with Significant HbR Activation for Scalar Inference for Child Passers – Group 1

Contrast	Region	Brodmann Area	X	Y	Z	β value	p value
SI > PT	Broca's area	45	50	47.7	17.7	0.0034	0.05
SI > PT	Medial temporal area	48	-59.6	8.3	-2.6	0.0031	0.05
SI > PT	Dorsolateral PFC	9	41	36	44	0.0021	0.05
SI > PT	Dorsomedial PFC	45	-2.3	58.7	38.7	0.0017	0.05
SI > PT	Primary somatosensory cortex	11	-69	-30.3	23.7	0.0016	0.05

*PFC refers to the prefrontal cortex.

In sum, although behavioral results from the Language (Vocabulary), Executive Functions, and Theory of Mind tasks failed to predict children's scalar implicature performance, neural activation patterns provided insight into the role that these networks play in scalar implicature derivation. Adults displayed activation in regions associated with language (LIFG), executive functions (dorsolateral PFC), and, critically, Theory of Mind (dorsomedial PFC). This provides support for the Gricean and other inference-driven pragmatic theories, according to which Theory of Mind plays a critical role in scalar implicature derivation (Grice, 1989; Sperber & Wilson, 1986, 2002; Sauerland, 2004). Note that this activation of epistemic reasoning during a scalar task is above and beyond that required for visual perspective taking in the same task; activation for visual perspective taking is subtracted out from this analysis.

Child passers in the scalar implicature task displayed activation in similar regions to adults. Critically, child passers displayed activation in the dmPFC, indicating active recruitment of Theory of Mind regions during scalar implicature derivation. Follow-up analyses revealed that there are distinct neural profiles within the passers group (even though these results are only suggestive because of low Ns). One group of child passers displayed comparable neural activity to adults, including activation of networks involved in executive functions (dlPFC) and Theory of Mind (dmPFC) during scalar inference, supporting the idea that preschool children are capable of deriving scalar inferences with sensitivity to a speaker's epistemic state (Papafragou et al., 2018; Kampa & Papafragou, 2020). However, the other child passers failed to display a cohesive group neural pattern. This provides potential insight into the high variability of success in previous scalar implicature studies: even children who display a cohesive behavioral group pattern (passers) in some cases display distinct neural profiles.

Child non-passers in the scalar implicature task seemed to rely predominantly on other networks, specifically the superior temporal gyrus and middle temporal gyrus, although they did display activation in some of the same regions as adults (dorsolateral prefrontal cortex, Wernicke's area, and the pre-motor and supplementary motor cortex). The superior temporal gyrus plays a role in spoken word recognition, and the middle temporal gyrus has been indicated in cognitive processes such as semantic memory, language, and visual perception. Notably, the child non-passers group did not cohesively display significant activation in Theory of Mind regions, indicating a possible connection between Theory of Mind development and scalar implicature success. They also did not display significant specialized activation in the

LIFG, indicating a possible connection between the development of higher-level linguistic abilities and scalar implicature success.

4.4.2.3 Theory of Mind

For adults, greater HbO activation was observed in the dorsolateral prefrontal cortex, the supramarginal gyrus, and the primary somatosensory cortex (Figure 12, Table 15). Given that the motivation for adding this task was to assess the relationship between scalar implicature derivation and Theory of Mind, we decided to keep children in the passers and non-passers groups to analyze the results of this task and look for differences in Theory of Mind development between these two groups. Child passers from the scalar implicature task displayed activation in the dorsomedial prefrontal cortex and supramarginal gyrus during mentalizing in this task (Table 16). Child non-passers in the scalar implicature task did not display any significant group patterns of activation.

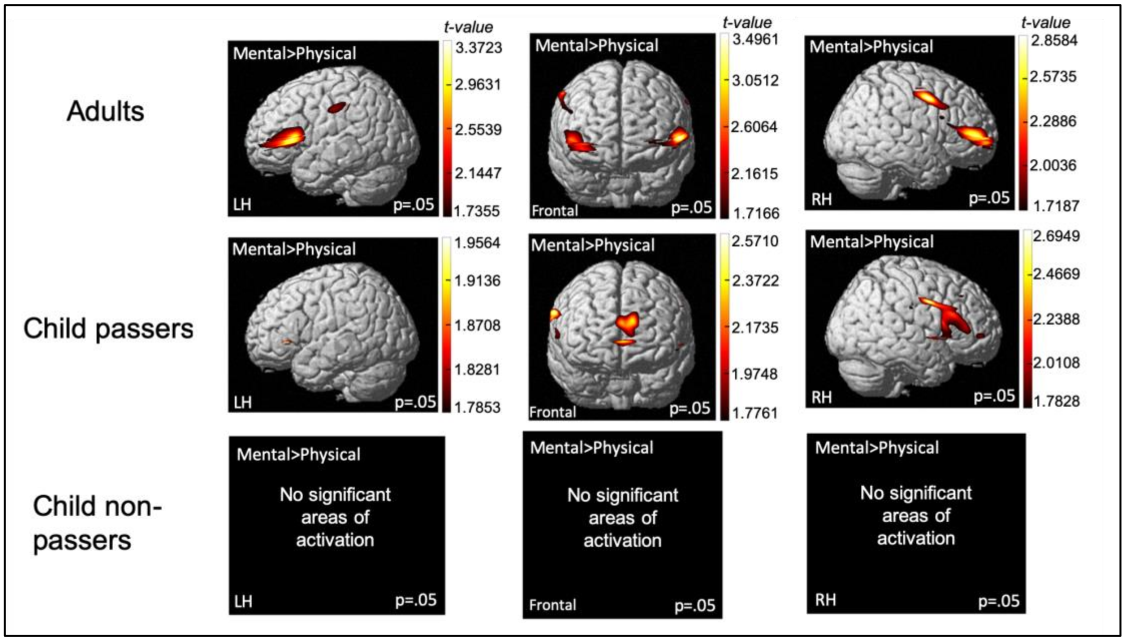

Figure 12: Neural activation patterns for adults, child scalar-implicature passers and child scalar-implicature non-passers in the Theory of Mind neuroimaging task with an HbO contrast. Adults displayed significant activation in the ventral pre-frontal cortex. Child scalar-implicature passers displayed significant activation in the dmPFC. Child scalar-implicature non-passers did not display any significant group patterns of activation (HbO or HbR).

Table 15: Regions with Significant HbO Activation for Mentalizing for Adults

Contrast	Region	Brodmann Area	X	Y	Z	β value	*p* value
Mental > Physical	Dorsolateral PFC	46	-39.6	51.6	26.6	0.0016	0.05
Mental > Physical	Supramarginal gyrus	40	-54.3	-46.6	55.6	0.0011	0.05

*PFC refers to the prefrontal cortex.

Table 16: Regions with Significant HbR Activation for Mentalizing for Child Scalar-Implicature Passers

Contrast	Region	Brodmann Area	X	Y	Z	β value	*p* value
Mental > Physical	Dorsomedial PFC	9	25.6	43.6	48.6	0.0006	0.05
Mental > Physical	Supramarginal gyrus	40	70	-38.6	24.3	0.0005	0.05

*IFG refers to the inferior frontal gyrus. PFC refers to the prefrontal cortex.

Summarizing, for adults, we found activation in the Dorsolateral PFC and the Supramarginal Gyrus. There was no significant HbO activation in the dmPFC during mentalizing, even though the stimuli were identical to those used in a previous fMRI study that found activation in the dmPFC (Gweon et al., 2012).[4] As predicted, children who did well behaviorally on the scalar implicatures task (passers) displayed a cohesive group neural profile in the Theory of Mind task, namely activation in the

[4] There was significant HbR activation in the dmPFC (see supplementary materials).

dmPFC, indicating specialized Theory of Mind regions. Child non-passers in the scalar implicature task did not display any significant group patterns of activation in the Theory of Mind task. This provides further evidence of a connection between Theory of Mind and scalar inference; children who successfully derive scalar inferences display more developed Theory of Mind networks.

4.4.2.4 Scalar Implicature and Theory of Mind (Conjunction analysis)

Finally, after analyzing each neuroimaging task individually, we generated conjunction maps of the scalar implicature and Theory of Mind tasks, which show overlap in patterns of neural activity for the Scalar-Inference > Perspective-Taking contrast (scalar inference derivation) and the activation Mental > Physical contrast (mentalizing). For adults, there were regions of significant activation for scalar inference derivation and mentalizing that overlapped, specifically in the LIFG and dlPFC (Figure 13). For child scalar-implicature passers and child scalar-implicature non-passers, there were no significant regions of activation that overlapped.

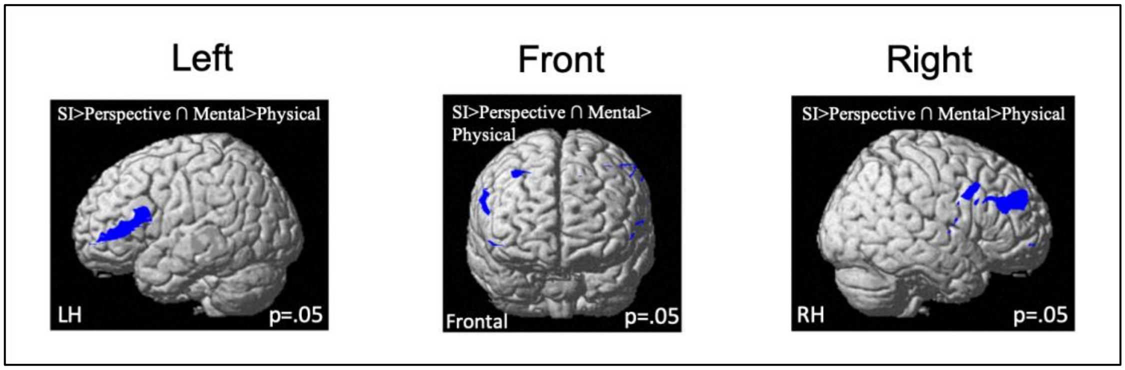

Figure 13: Conjunction of activation for the Scalar-Inference > Perspective-Taking contrast (scalar inference derivation) in the Scalar Implicature task and activation for the Mental > Physical contrast (mentalizing) in the Theory of Mind task for adult participants.

If Theory of Mind is an active part of scalar implicature derivation in adults as predicted by classic social-inferential pragmatic theories (Grice, 1989; Sperber & Wilson, 1986, 2002; Sauerland, 2004), we would expect there to be some overlap between networks active during scalar inference and networks active during mentalizing. The conjunction analysis revealed that there were regions active during mentalizing that were also active during scalar inference derivation for adults, specifically in the LIFG and dlPFC. This indicates a relationship between Theory of Mind reasoning and scalar inference, as both processes rely on overlapping regions. Conjunction analyses for child scalar-implicature passer and scalar-implicature non-passer groups did not return any significant regions of overlap between these tasks. This may be because neural specialization for children is still on-going in this age range or due to a high degree of individual variation (the conjunction analysis compares group level activation).

4.5 General Discussion

4.5.1 The Cognitive Signature of Scalar Inference

In conversation, when speakers provide statements that give limited information ("Some toys on the table are free"), listeners typically infer that the speaker is not in a position to offer a more-informative statement ("All toys on the table are free"). This type of inference, a *scalar implicature,* has been extensively studied, yet debates continue over the contributions of various mechanisms to scalar implicature derivation and the way those mechanisms interact during development. Using a novel fNIRS neuroimaging approach combined with more familiar behavioral methods, we investigated the contributions of cognitive processes including Theory of Mind, executive functions, and language to the derivation of scalar implicatures during development as well as in adulthood.

Although behavioral measures of Theory of Mind, executive functions, and language failed to predict children's behavioral scalar implicature performance in our study (consistent with previous studies; see Matthews et al., 2018 for a review), fNIRS technology allowed us to directly test for activation of these networks during scalar inference to gain insight into the role that each mechanism plays in the derivation process, something not possible using behavioral methods alone. Our neuroimaging results indicated that adults' scalar inference derivation is supported by all three target mechanisms (Theory of Mind, executive functions, and language). Specifically, we found significant activation for adults and some children in the dorsomedial PFC (Theory of Mind), the dorsolateral PFC (executive functions), and the LIFG (language) during scalar inference. Moreover, we demonstrated that adults recruit the same neural networks during both scalar inference derivation and mentalizing; this provides additional evidence of a relationship between Theory of Mind and scalar inference.

We provide the first evidence from spatial neuroimaging techniques that some 4- to 5-year-old children display similar patterns of neural activity to adults during scalar inference, specifically recruitment of Theory of Mind (dmPFC), executive functions (dlPFC), and language (LIFG) networks. The period between age 3 to 5 is a crucial period for development in all of these areas, and this study provides evidence that these areas of development are interacting. While both child passers and child non-passers in the scalar implicature task displayed activation in the dorsolateral PFC, only passers displayed activation in the dorsomedial PFC and the LIFG. Moreover, children who succeeded on the scalar implicature task displayed specialized activation of Theory of Mind networks in the Theory of Mind task, an indicator of more developed Theory of Mind ability; children who struggled with the scalar implicature task as a group did not display any significant patterns of activation in the Theory of Mind task. Thus, there appears to be a relationship between more developed Theory of Mind networks with more adult-like patterns of neural activity during scalar inference.

Taken together, these results provide novel evidence bearing on theoretical debates over the nature of pragmatic inference, specifically the role that Theory of Mind plays during this process. Under Gricean and several newer inferential accounts, scalar implicature derivation is a complex, social-inferential process that requires the listener to incorporate information about the speaker's knowledge state and other characteristics (Grice, 1975, 1989; Sperber & Wilson, 1986, 2002; Sauerland, 2004). However, lexical and grammatical accounts have proposed that scalar implicature derivation can be achieved through linguistic mechanisms alone without the need for epistemic reasoning (Chierchia et al., 2008; Fox, 2007; Chierchia, 2004, 2006; Levinson, 2000) and that preschool children do not necessarily engage in epistemic reasoning during scalar inference (Hochstein et al., 2016; Barner et al., 2018; Hochstein et al., 2018). This study provides evidence that networks for mental reasoning are implicated in scalar implicature computation in adults and even some young children. This evidence is critical for Gricean and other inferential theories and is not easily accounted for by grammatical or lexical approaches to scalar implicature.

These results also provide novel insight into the variable and context-dependent performance of young children found in previous scalar implicature studies. Although children who were designated as "passers" all consistently performed near ceiling on the behavioral portion of the scalar implicature tasks (as adults typically do; see also Kampa & Papafragou, 2020), the neuroimaging analyses revealed a different picture: only a subgroup of these children displayed a coherent group neural profile similar to that of the adults. The other subgroup of child passers, who appeared behaviorally identical to the other child passers and adults, displayed distinct individual neural patterns. Thus, it appears that during development, there may be different pathways to success for children whose pragmatic skills are still in flux, although, given the consistency of adult responses, it would appear that typically developing children do converge on a reliance on ToM networks during scalar inference by adulthood. An interesting follow-up line of inquiry would be to test a wider age range with a focus on individual differences, possibly with a longitudinal component, to gain insight into how and when this change is brought about in children. This discrepancy between behavioral success and atypical neural patterns may also help account for why children with autism have shown success with scalar implicatures (Pijnacker et al., 2009; Chevallier et al., 2010; Hochstein et al., 2018; Van Tiel & Kissine 2018) despite having impaired Theory of Mind reasoning; they may be using alternative strategies or networks like the second subgroup of child passers in this task. Additional experimentation and theorizing are needed to assess the exact nature of those strategies.

4.5.2 Limitations and Extensions

Although this study provides a vital foray into the mechanisms supporting scalar inference in adults and young children, it has some limitations. The most apparent limitation is the low number of participants in the child groups. Despite our best efforts, it was challenging for 4- and 5-year-old children to complete a multi-task neuroimaging study. The children had to wear a cap with electrodes attached while trying to sit as still as possible in a dark room during the tasks (fNIRS data collection is sensitive to light and movement), separate from their parents to avoid distractions. We also faced recruitment challenges for this type of study given the relatively small local pool and COVID-related difficulties impacting in-person lab testing.

Another limitation is, that while we found activation of regions associated with executive functions (notably the dorsolateral PFC), the relationship between executive functions, scalar implicature, and Theory of Mind may be more complicated. In a recent study, Fairchild and Papafragou (2020) found that executive functions ability was correlated with scalar implicature derivation ability, but not when Theory of Mind was controlled for. They proposed that the correlation between executive functions and scalar implicatures could be accounted for by the engagement of executive functions during the Theory of Mind reasoning process itself (separate from the scalar implicature). This introduces the possibility that executive functions regions may be active during scalar inference in our task as a by-product of the Theory of Mind reasoning involved, rather than through a direct reliance on executive functions for scalar implicature derivation. At the very least, the dorsolateral PFC activation displayed by adults in the Theory of Mind task provides evidence that executive functions are active during mentalizing. More comprehensive measures of Theory of Mind and executive functions with these neuroimaging techniques might be helpful in further teasing out the contributions of executive functions to the scalar implicature process separately from the contributions of Theory of Mind.

Finally, this study opens a number of possible directions for continued inquiry into the development of pragmatic inference in preschool children. First, the paradigm we used relies on an ad hoc (entirely context-dependent) scale, which might increase the salience of relevant scalar alternatives through the presence of context; it would be interesting to see whether these results generalize to quantifier scales (e.g., involving *some* and *all*) within the same age group (see Papafragou & Musolino, 2003; Barner et al., 2011; Katsos & Bishop, 2011). Second, the present approach could also be applied to other populations, such as children with autism, to assess whether they rely on the same systems to achieve behavioral success with scalar implicature derivation as neurotypical children. Third, our findings provide evidence of a relationship between Theory of Mind and language development with scalar implicature development, but the scope of this paper precludes us from making stronger claims as to whether a causal relationship exists between these systems. Future work should continue to elucidate the nature of the connections between these mechanisms.

4.5.3 Final Remarks

Here we presented the first spatial neuroimaging study to look at scalar inference in 3.5- to 5-year-old children and one of the very few studies to study early pragmatic development using neural measures more broadly. We chose to investigate scalar implicatures because they have been extensively studied using behavioral measures; however, the present neuroimaging approach could easily be applied to other types of conversational inferences or other pragmatic phenomena that have not been as extensively explored. The regions that we found to be active during scalar inference (dmPFC, dlPFC, LIFG) are likely to also be recruited during other types of pragmatic inference, and the role that supporting mechanisms such as Theory of Mind, executive functions, and language play in this process plausibly generalize to other types of pragmatic inference, at least according to Gricean models of inferential communication (see Fairchild & Papafragou, 2021). Thus the current study provides not only a crucial look below the surface into children's scalar implicature derivation, but also a novel contribution to the broader developmental study of pragmatics and conversational implicature, bridging the gap between behavioral patterns for conversational inference and the neural substrates that underlie young learners' pragmatic calculations.

Chapter 5

SUMMARY & CONCLUSIONS

In conversation, participants are expected to provide as much relevant information as is required by the purpose of the exchange (Grice, 1975). Thus when a speaker offers a statement that appears to be less than fully informative ("Some of the animals are safe to pet"), the listener either interprets the statement to mean 1) not all of the animals are safe to pet (if the speaker is presumed to be knowledgeable) or 2) the speaker doesn't know whether all of the animals are safe to pet (but at least some are). The goal of this dissertation was to assess the role of epistemic reasoning in scalar implicature derivation and development. Under a Gricean approach, listeners must evaluate the knowledge state of the speaker to derive a pragmatic inference, however previous research seemed to indicate that young children fail to display epistemic sensitivity during scalar inference, despite early successes in simpler contexts. Other proposals by Grice, such as the idea that communicative principles and maxims extend to non-linguistic communication, also lacked sufficient empirical support. Another concern was the highly task- and context-dependent performance of young children in scalar implicature studies, as well as a lack of a consensus on how potential supporting factors (theory of mind, executive functions, and language) interact with scalar implicature development. We investigated all of these areas of controversy through a set of variations on a novel visual world paradigm. These findings shed new light not only on existing debates in the scalar implicature literature, but also on phenomena in the broader field of pragmatics and other disciplines (theory of mind, symbolic development, and cognitive neuroscience).

As stated, the primary focus of this dissertation was to improve our understanding of the role that epistemic reasoning plays in scalar implicature derivation and development. In Chapter 2, we tested 4- and 5-year-olds’ sensitivity to epistemic state during scalar inference using a novel visual world paradigm and found the first evidence that children as young as 4 can flexibly adjust to a speaker’s epistemic state during the derivation of scalar implicatures. Then, in Chapter 3, we adapted this paradigm to non-linguistic symbols (drawings) and found empirical evidence that 5-year-olds and adults display the same epistemic sensitivity to scalar inferences in non-linguistic communicative contexts (although success was impacted by the features of the symbols). Finally, in Chapter 4, we used fNIRS technology to conduct the first spatial neuroimaging study on scalar inference in 3.5- to 5-year-olds and found that theory of mind (as well as executive functions and language) networks are active during scalar inference as well as initial evidence that theory of mind and scalar implicature systems interact during development.

Although this dissertation focuses on scalar implicatures specifically, these findings make broad contributions to the field of pragmatics. From a theoretical perspective, our data accord with inferential theories of communication proposed by Grice and others (Grice, 1972, 1989; Horn, 1972; Sperber & Wilson, 1986; Carston, 1998; Sauerland, 2004, 2012; Geurts, 2010), wherein pragmatic inferences are thought to be derived through a rich social reasoning process in which the addressee filters their interpretation of an utterance through the lens of the communicator’s knowledge state, goals, etc. Epistemic reasoning is held to be a critical component of this class of theories, and these findings provide empirical support that scalar inferences are derived in line with pragmatic expectations about the speaker and their knowledge state and are done so from early in development. We even find that this epistemic reasoning process extends to pragmatic inferences derived from other forms of ostensive intentional communication beyond language. Additionally, we find important evidence that theory of mind networks in the brain are engaged during scalar inference in online comprehension as well as connections between young children's scalar inference and theory of mind development.

Our data also provide evidence to support the widely accepted but rarely tested claim that principles of communication such as the need to be informative underlie all forms of purposive exchanges (cf. Grice, 1972, 1989; Sperber & Wilson, 1986). We find that pictorial non-linguistic symbols are subject to the same expectations of informativeness as linguistic utterances, and that these expectations begin in early childhood. These data comport with the idea that human communication is a species of human rational behavior (Tomasello, 2010; Frank & Goodman, 2012) and that both children and adults evaluate the presentation of pictorial symbols in an interactive context in line with rational expectations of how a communicator in an exchange will behave.

These findings also contribute to our understanding of children's development of communicative competence. Our results show that young children not only display a high degree of pragmatic sophistication, but also show sensitive early adaptions to others during communication and are consistent with the position that children are able to adapt to a communicative partner's characteristics and needs at young ages (Baldwin, 1991; O'Neill, 1996; Nadig & Sedivy, 2002; Matthews, Lieven, Theakston,

& Tomasello, 2006; Bahtiyar & Küntay, 2009; see Grigoroglou & Papafragou, 2018, 2019, for a review). We also provide evidence that young children are able to evaluate a creator's intent when interpreting drawings (Gelman & Ebeling, 1998; Gelman & Bloom, 2000; Richert & Lillard, 2002; Jaswal, 2006; Preissler & Bloom, 2008; cf. also Allen et al., 2010). This evidence contrasts with reports that children are largely egocentric fail to adjust to other's needs until late in development (Piaget, 1970; Shatz, 1980; Perner & Leekam, 1986; Girbau, 2001; Epley, Morewedge, & Keysar, 2004; Davies & Katsos, 2010.

This line of research breaks valuable new ground in a number of pragmatic directions, but, in turn, each new finding opens up another set of paths to explore, each with questions to be answered. Although children interpret non-linguistic symbols in line with pragmatic expectations, we see some additional challenges compared to linguistic symbols. While we theorize that these additional difficulties are largely due to burden of dual representation (which is still undergoing development in this age range), there is still the possibility that there could be differences in expectations for how we use language as a communicative medium as opposed to how we use non-linguistics symbols like drawings, which are typically meant to be depictive and are not often part of a direct communicative exchange, to express communicative intentions. Additionally, while our neuroimaging study in Chapter 4 contributes an impactful initial investigation into the realm of the neural substrates that underlie pragmatic inference in developing children, there are still quite a number of questions left unanswered. We observe a connection between Theory of Mind and Scalar Inference systems, but more answers are needed on exact nature of that connection and the direction of causality between the interactions we observed. Our data also raise questions about individual patterns in children's neural activity and how those patterns change over the course of development. In addition, while we focused on theory of mind networks as the area of greatest theoretical interest to the field of pragmatics, future studies could place more focus on probing the contributions of executive functions or language.

The paradigm developed for this set of studies was designed to address complex questions while maintaining a simple, child-friendly interface, and thus could easily be adapted to explore the development of other types of implicatures and pragmatic phenomena. We chose to investigate scalar implicatures because there is already an established literature off of which to build, but these studies now provide a jumping off point to study the less explored maxims and areas of pragmatics. In this vein, while we explored pictorial nonlinguistic symbols, Grice's proposal stated that communicative principles and maxims should apply to all forms of communication. This opens a door into a vast range of other pragmatic directions, including other maxims to test with pictorial symbols as well as exploring other forms of ostensive and intentional communication through non-linguistic means. In addition, while we focused on epistemic state as a speaker characteristic, presumably the inferential process recruited during pragmatic inference includes an appeal to other characteristics besides knowledge state. One promising characteristic to explore would be speaker motivations and how the interpretation of underinformative statements changes when someone is not a cooperative communicator and purposely leaves out relevant information.

The interdisciplinary nature of this dissertation gives these findings broad appeal. Children have complex minds and with so many cognitive processes developing in the same period of time, it’s critical to bridge the gaps between these disciplines and build our understanding of how these processes interact. These findings go some way in tying threads between different areas of development, such as pragmatics, theory of mind, symbolic understanding, executive functions, and cognitive neuroscience, but there are still so many connections to be made and questions to be explored. We believe that this research makes some head way towards that goal and hope that it will inspire further progress in that direction.

REFERENCES

Adler, N. E., Epel, E. S., Castellazzo, G., & Ickovics, J. R. (2000). Relationship of subjective and objective social status with psychological and physiological functioning: Preliminary data in healthy, White women. *Health psychology*, *19*(6), 586.

Akhtar, N., Carpenter, M., & Tomasello, M. (1996). The role of discourse novelty in early word learning. *Child Development*, *67*(2), 635-645.

Allen, M. L., Bloom, P., and Hodgson, E. (2010). Do young children know what makes a picture useful to other people? *Journal of Cognition and Culture*, *10*, 27-37.

Antoniou, K., Kambanaros, M., Grohmann, K. K., & Katsos, N. (2014). Is bilectalism similar to bilingualism? An investigation into children's vocabulary and executive control skills. In *BUCLD 38: Proceedings of the 38th Annual Boston University Conference on Language Development* (Vol. 1, pp. 12-24). Cascadilla Press Somerville, MA.

Antoniou, K., Cummins, C., & Katsos, N. (2016). Why only some adults reject under-informative utterances. *Journal of Pragmatics*, *99*, 78-95.

Antoniou, K., Grohmann, K. K., Kambanaros, M., & Katsos, N. (2013). Does multilingualism confer an advantage for pragmatic abilities? http://ktisis.cut.ac.cy/handle/10488/4102 Accessed June 2020.

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

Apperly, I. A., Back, E., Samson, D., & France, L. (2008). The cost of thinking about false beliefs: Evidence from adults' performance on a non-inferential theory of mind task. *Cognition*, *106*(3), 1093-1108.

Bahtiyar, S., & Küntay, A. C. (2009). Integration of communicative partner's visual perspective in patterns of referential requests. *Journal of Child Language*, *36*(3), 529-555.

Baker, C. L., Saxe, R., & Tenenbaum, J. B. (2009). Action understanding as inverse planning. *Cognition*, *113*(3), 329-349.

Baldwin, D. A. (1993). Infants' ability to consult the speaker for clues to word reference. *Journal of Child Language*, *20*(2), 395-418.

Baldwin, D. A. (1993). Early referential understanding: Infants' ability to recognize referential acts for what they are. *Developmental Psychology*, *29*(5), 832.

Barner, D., Brooks, N., & Bale, A. (2011). Accessing the unsaid: The role of scalar alternatives in children's pragmatic inference. *Cognition*, *188*, 87-96.

Barner, D., Hochstein, L. K., Rubenson, M. P., & Bale, A. (2018). Four-year-old children compute scalar implicatures in absence of epistemic reasoning. In K. Syrett and A. Arunachalam, (Eds.), *Semantics in Acquisition*. Amsterdam: Benjamins, 326-349.

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

Barrera, M. E., & Maurer, D. (1981). Recognition of mother's photographed face by the three-month-old infant. *Child Development*, 714-716

Bass, L., Bonawitz, E., & Gweon, H. (2017). Didn't know, or didn't show? Preschoolers consider epistemic state and degree of omission when evaluating teachers. *Proceedings from the 39th Meeting of the Cognitive Science Society*.

Bates, D. M. (2005). Fitting linear mixed models in R. R News, 5, 27–30. Available at: http://cran.rproject.org/doc/Rnews/RnewsfZO05-1.pdf>.

Bates, D. M., Maechler, M., & Bolker, B. (2011). *lme4: Linear mixed-effects models using S4 classes.* R package version 0.99375-39. Available at: http://CRAN.Rproject.org/package-lme4>.

Bates, D. M., Maechler, M., Bolker, B., & Walker, S. (2015). Fitting linear mixed-effects models using lme4. *Journal of Statistical Software*, 67, 1–48.

Bates, D. M., & Sarkar, D. (2007). *lme4: Linear mixed-effects models using S4 classes.* R package version 0.99875-6.

Bergen, L., & Grodner, D. J. (2012). Speaker knowledge influences the comprehension of pragmatic inferences. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *38*, 1450.

Bohn, M., Call, J., & Tomasello, M. (2019). Natural reference: A phylo-and ontogenetic perspective on the comprehension of iconic gestures and vocalizations. *Developmental science*, *22*(2), e12757.

Bonomini, V., Zucchelli, L., Re, R., Ieva, F., Spinelli, L., Contini, D., Paganoni, A., & Torricelli, A. (2015). Linear regression models and k-means clustering for statistical analysis of fNIRS data. *Biomedical Optics Express*, *6*(2), 615-630.

Breheny, R., Ferguson, H., & Katsos, N. (2013). Taking the epistemic step: Toward a model of on-line access to conversational implicatures. *Cognition*, *126*, 423-440.

Bunge, S. A., Dudukovic, N. M., Thomason, M. E., Vaidya, C. J., & Gabrieli, J. D. (2002). Immature frontal lobe contributions to cognitive control in children: evidence from fMRI. *Neuron*, *33*(2), 301-311.

Cantlon, J. F., Brannon, E. M., Carter, E. J., & Pelphrey, K. A. (2006). Functional imaging of numerical processing in adults and 4-y-old children. *PLoS Biol*, *4*(5), e125.

Carlson, S. M. (2005). Developmentally sensitive measures of executive function in preschool children. *Developmental Neuropsychology*, *28*(2), 595-616.

Carpenter, M., Nagell, K., Tomasello, M., Butterworth, G., & Moore, C. (1998). Social cognition, joint attention, and communicative competence from 9 to 15 months of age. *Monographs of the Society for Research in Child Development*, i-vi, 1-143

Carrion, V. G., Garrett, A., Menon, V., Weems, C. F., & Reiss, A. L. (2008). Posttraumatic stress symptoms and brain function during a response-inhibition task: an fMRI study in youth. *Depression and Anxiety*, *25*(6), 514-526.

Carston, R. (1998). Negation, ‘presupposition’ and the semantics/pragmatics distinction. *Journal of Linguistics*, *34*(2), 309-350.

Chemla, E., & Singh, R. (2014). Remarks on the experimental turn in the study of scalar implicature, Part I. *Language and Linguistics Compass*, *8*(9), 373-386.

Chierchia, G. (2004). Scalar implicatures, polarity phenomena, and the syntax/pragmatics interface. In Belletti (Ed.), *Structures and Beyond* (pp. 39-103). Oxford University Press.

Chierchia, G. (2006). Broaden your views: Implicatures of domain widening and the "logicality" of language. *Linguistic Inquiry, 37*(4), 535-590.

Chierchia, G., Crain, S., Guasti, M. T., Gualmini, A., & Meroni, L. (2001). The acquisition of disjunction: Evidence for a grammatical view of scalar implicatures. In A. H.-J. Do et al. (Eds.), *BUCLD 25 Proceedings*, 157-168. Somerville, MA: Cascadilla Press.

Chierchia, G., Fox, D., & Spector, B. (2009). Hurford's constraint and the theory of scalar implicatures. In P. Egré & G. Magri (Eds.), *Presuppositions and Implicatures - Proceedings of the MIT-Paris workshop*, 47–62. Cambridge, MA: MIT Working Papers in Linguistics (MITWPL).

Chierchia, G., Fox, D., & Spector, B. (2012). The grammatical view of scalar implicatures and the relationship between syntax and semantics. In C. Maienborn, K. von Heusinger, & P. Portner (Eds.), *Semantics: An International Handbook of Natural Language Meaning;* Volume 1. New York: Mouton de Gruyter.

Ciesielski, K. T., Lesnik, P. G., Savoy, R. L., Grant, E. P., & Ahlfors, S. P. (2006). Developmental neural networks in children performing a Categorical N-Back Task. *Neuroimage*, *33*(3), 980-990.

Clark, E. V. (1990). On the pragmatics of contrast. *Journal of Child Language*, *17*(2), 417-431.

Conti, D. J., & Camras, L. A. (1984). Children's understanding of conversational principles. *Journal of Experimental Child Psychology*, *38*(3), 456-463.

Crone, E. A., Donohue, S. E., Honomichl, R., Wendelken, C., & Bunge, S. A. (2006a). Brain regions mediating flexible rule use during development. *Journal of Neuroscience*, *26*(43), 11239-11247.

Crone, E.A., Wendelken, C., Donohue, S., Van Leijenhorst, L., & Bunge, S.A. (2006b). Neurocognitive development of the ability to manipulate information in working memory. *Proceedings of the National Academy of Sciences of the United States of America, 103*, 9315–9320.

Crone, E.A., Zanolie, K., Van Leijenhorst, L., Westenberg, M., & Rombouts, S.A.R.B. (2008). Neural mechanisms supporting flexible performance adjustment during development. *Cognitive, Affective, and Behavioral Neuroscience*, *8*, 165–177.

Csibra, G., & Gergely, G. (1998). The teleological origins of mentalistic action explanations: A developmental hypothesis. *Developmental Science*, *1*(2), 255-259.

Csibra, G., Gergely, G., Bíró, S., Koos, O., & Brockbank, M. (1999). Goal attribution without agency cues: the perception of 'pure reason' in infancy. *Cognition*, *72*(3), 237-267.

Davies, C., & Katsos, N. (2010). Over-informative children: Production/comprehension asymmetry or tolerance to pragmatic violations? *Lingua*, *120*(8), 1956-1972.

Degen, J., & Tanenhaus, M. K. (2015). Processing scalar implicature: A constraint-based approach. *Cognitive Science, 39*(4), 667-710.

Delis, D. C., Kaplan, E., & Kramer, J. H. (2001). *Delis-Kaplan Executive Function System®(D-KEFS®): Examiner's Manual: Flexibility of Thinking, Concept Formation, Problem Solving, Planning, Creativity, Impulse Control, Inhibition*. Pearson.

DeLoache, J. S. (1987). Rapid change in the symbolic functioning of very young children. *Science*, *238*(4833), 1556-1557.

DeLoache, J. S. (1991). Symbolic functioning in very young children: Understanding of pictures and models. *Child Development*, *62*(4), 736-752.

DeLoache, J. S. (1995). Early symbol understanding and use. *Psychology of Learning and Motivation*, *33*, 65-116.

DeLoache, J. S. (2000). Dual representation and young children's use of scale models. *Child Development*, *71*(2), 329-338.

DeLoache, J. S. (2002). Symbolic artifacts: Understanding and use. *Blackwell handbook of childhood cognitive development*, 206-226.

DeLoache, J. S., & Burns, N. M. (1994). Early understanding of the representational function of pictures. *Cognition*, *52*(2), 83-110.

DeLoache, J. S., Kolstad, V., & Anderson, K. N. (1991). Physical similarity and young children's understanding of scale models. *Child Development*, *62*(1), 111-126.

DeLoache, J. S., Miller, K. F., & Rosengren, K. S. (1997). The credible shrinking room: Very young children's performance with symbolic and nonsymbolic relations. *Psychological Science*, *8*(4), 308-313.

DeLoache, J. S., Strauss, M. S., & Maynard, J. (1979). Picture perception in infancy. *Infant Behavior and Development*, *2*, 77-89.

De Neys, W., & Schaeken, W. (2007). When people are more logical under cognitive load: Dual task impact on scalar implicature. *Experimental Psychology*, *54*(2), 128-133.

De Marchena, A., Eigsti, I. M., Worek, A., Ono, K. E., & Snedeker, J. (2011). Mutual exclusivity in autism spectrum disorders: Testing the pragmatic hypothesis. *Cognition*, *119*(1), 96-113.

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

Dibbets, P., Bakker, K., & Jolles, J. (2006). Functional MRI of task switching in children with specific language impairment (SLI). *Neurocase*, *12*(1), 71-79.

Diesendruck, G., & Markson, L. (2001). Children's avoidance of lexical overlap: A pragmatic account. *Developmental Psychology*, *37*(5), 630.

Dirks, J., & Gibson, E. (1977). Infants' perception of similarity between live people and their photographs. *Child Development*, 124-130.

Eager, C., & Roy, J. (2017). Mixed effects models are sometimes terrible. *arXiv preprint arXiv:1701.04858*.

Emberson, L. L., Boldin, A. M., Riccio, J. E., Guillet, R., & Aslin, R. N. (2017). Deficits in top-down sensory prediction in infants at risk due to premature birth. *Current Biology*, *27*(3), 431-436.

Emberson, L. L., Cannon, G., Palmeri, H., Richards, J. E., & Aslin, R. N. (2017b). Using fNIRS to examine occipital and temporal responses to stimulus repetition in young infants: Evidence of selective frontal cortex involvement. *Developmental Cognitive Neuroscience*, *23*, 26-38.

Eng, H. (2013). *The psychology of children's drawings: From the first stroke to the coloured drawing*. Routledge.

Engelhardt, P. E., Bailey, K. G., & Ferreira, F. (2006). Do speakers and listeners observe the Gricean Maxim of Quantity? *Journal of Memory and Language*, *54*(4), 554-573.

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

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

Fairchild, S., & Papafragou, A. (2018). Sins of omission are more likely to be forgiven in non-native speakers. *Cognition*, *181*, 80-92.

Fairchild, S., & Papafragou, A. (2021). The role of executive function and Theory of Mind in pragmatic computations. *Cognitive Science*, 45, e12938.

Fishburn, F. A., Norr, M. E., Medvedev, A. V., & Vaidya, C. J. (2014). Sensitivity of fNIRS to cognitive state and load. *Frontiers in human neuroscience*, *8*, 76.

Flandin, G., & Novak, M. J. (2020). fMRI data analysis using SPM. In *fMRI* (pp. 89-116). Springer, Cham.

Foppolo, F., Guasti, M. T., & Chierchia, G. (2012). Scalar implicatures in child language: Give children a chance. *Language Learning and Development*, *8*(4), 365-394.

Foppolo, F., & Marelli, M. (2017). No delay for some inferences. *Journal of Semantics*, *34*(4), 659-681.

Foppolo, F., Mazzaggio, G., Panzeri, F., & Surian, L. (2020). Scalar and Ad-Hoc pragmatic inferences in children: guess which one is easier. *Journal of Child Language*, 1-23.

Fox, D. (2007). Free choice and the theory of scalar implicatures. In *Presupposition and implicature in compositional semantics* (pp. 71-120). Palgrave Macmillan, London.

Frank, M. C., & Goodman, N. D. (2012). Predicting pragmatic reasoning in language games. *Science*, *336*(6084), 998-998.

Gallagher, A., Béland, R., & Lassonde, M. (2012). The contribution of functional near-infrared spectroscopy (fNIRS) to the presurgical assessment of language function in children. *Brain and Language*, *121*(2), 124-129.

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

Gelman, S. A., & Bloom, P. (2000). Young children are sensitive to how an object was created when deciding what to name it. *Cognition*, *76*(2), 91-103.

Gelman, S. A., & Ebeling, K. S. (1998). Shape and representational status in children's early naming. *Cognition*, *66*(2), B35-B47.

Gergely, G., Bekkering, H., & Király, I. (2002). Developmental psychology: Rational imitation in preverbal infants. *Nature*, *415*(6873), 755.

Gergely, G., & Csibra, G. (1997). Teleological reasoning in infancy: The infant's naive theory of rational action: A reply to Premack and Premack. *Cognition*, *63*(2), 227-233.

Gergely, G., Nádasdy, Z., Csibra, G., & Bíró, S. (1995). Taking the intentional stance at 12 months of age. *Cognition*, *56*(2), 165-193.

Gershon, R. C., Cook, K. F., Mungas, D., Manly, J. J., Slotkin, J., Beaumont, J. L., & Weintraub, S. (2014). Language measures of the NIH toolbox cognition battery. *Journal of the International Neuropsychological Society, 20(6),* 642-651.

Geurts, B. (2010). *Quantity implicatures*. Cambridge: Cambridge University Press.

Girbau, D. (2001). Children's referential communication failure: The ambiguity and abbreviation of message. *Journal of Language and Social Psychology*, *20*(1-2), 81-89.

Gleason, J. B. (2005). The development of language (6th ed.). Pearson.

Gleitman, L. R., & Gleitman, H. (1992). A picture is worth a thousand words, but that's the problem: The role of syntax in vocabulary acquisition. *Current Directions in Psychological Science*, *1*(1), 31-35.

Glucksberg, S., Krauss, R. M., & Higgins, E. T. (1975). The development of referential communication skills. *Review of Child Development Research*, *4*, 305-345.

Goldstone, R. L., & Sakamoto, Y. (2003). The transfer of abstract principles governing complex adaptive systems. *Cognitive Psychology*, *46*(4), 414-466.

Golinkoff, R. M., Mervis, C. B., & Hirsh-Pasek, K. (1994). Early object labels: the case for a developmental lexical principles framework. *Journal of child language*, *21*(1), 125-155.

Grice, H. P. (1975). Logic and conversation. In P. Cole & J. Morgan (Eds.), *Syntax and Semantics,* Volume 3. New York: Academic Press.

Grice, P. (1989). *Studies in the way of words*. Cambridge, MA: Harvard University Press.

Grigoroglou, M., & Papafragou, A. (2018). Acquisition of Pragmatics. In R. Clark & M. Aronoff (Eds.), *Oxford Research Encyclopedia of Linguistics*. Online edition: Oxford University Press.

Grigoroglou, M., & Papafragou, A. (2019). Interactive contexts increase informativeness in children's referential communication. In press, *Developmental Psychology*.

Guasti, M. T., Chierchia, G., Crain, S., Foppolo, F., Gualmini, A., & Meroni, L. (2005). Why children and adults sometimes (but not always) compute implicatures. *Language and Cognitive Processes*, *20*(5), 667-696.

Grodner, D. J., Klein, N. M., Carbary, K. M., & Tanenhaus, M. K. (2010). "Some," and possibly all, scalar inferences are not delayed: Evidence for immediate pragmatic enrichment. *Cognition*, *116*(1), 42-55.

Grossmann, T., Oberecker, R., Koch, S. P., & Friederici, A. D. (2010). The developmental origins of voice processing in the human brain. *Neuron*, *65*(6), 852-858.

Grossmann, T., Parise, E., & Friederici, A. (2010). The detection of communicative signals directed at the self in infant prefrontal cortex. *Frontiers in Human Neuroscience*, *4*, 201.

Guasti, M. T., Chierchia, G., Crain, S., Foppolo, F., Gualmini, A., & Meroni, L. (2005). Why children and adults sometimes (but not always) compute implicatures. *Language and Cognitive Processes*, *20*(5), 667-696.

Gweon, H., & Asaba, M. (2018). Order matters: Children's evaluation of underinformative teachers depends on context. *Child Development*, *89*(3), e278-e292.

Hartshorne, J. K., Snedeker, J., Liem Azar, S. Y. M., & Kim, A. E. (2015). The neural computation of scalar implicature. *Language, Cognition and Neuroscience*, *30*(5), 620-634.

Heyman, T., & Schaeken, W. (2015). Some differences in some: examining variability in the interpretation of scalars using latent class analysis. *Psychologica Belgica*, *55*(1), 1.

Hirschberg, J. L. B. (1985). *A theory of scalar implicature*. Ph.D. thesis, University of Pennsylvania.

Hochstein, L., Bale, A., & Barner, D. (2018). Scalar implicature in absence of epistemic reasoning? The case of autism spectrum disorder. *Language Learning and Development*, *14*(3), 224-240.

Hochstein, L., Bale, A., Fox, D., & Barner, D. (2016). Ignorance and inference: do problems with Gricean epistemic reasoning explain children's difficulty with scalar implicature? *Journal of Semantics, 33*, 107-135.

Hollebrandse, B., van Hout, A., & Hendriks, P. (2014). Children's first and second-order false-belief reasoning in a verbal and a low-verbal task. *Synthese*, *191*(3), 321-333.

Horn, L. (1972). *On the semantic properties of the logical operators in English*. Doctoral diss. UCLA Distributed by IULC, Indiana University

Horn, L. (1989). *A natural history of negation*. Chicago: University of Chicago Press.

Horn, L. (2013). Implicature. In *Routledge Companion to Philosophy of Language* (pp. 75-88). Routledge.

Horowitz, A. C., Schneider, R. M., & Frank, M. C. (2018). The trouble with quantifiers: exploring children's deficits in scalar implicature. *Child development*, *89*(6), 572-593.

Houdé, O., Rossi, S., Lubin, A., & Joliot, M. (2010). Mapping numerical processing, reading, and executive functions in the developing brain: an fMRI meta-analysis of 52 studies including 842 children. *Developmental Science*, *13*(6), 876-885.

Huang, Y. T., & Snedeker, J. (2009a). Online interpretation of scalar quantifiers: Insight into the semantics–pragmatics interface. *Cognitive Psychology*, *58*(3), 376-415.

Huang, Y. T., & Snedeker, J. (2009b). Semantic meaning and pragmatic interpretation in 5-year-olds: Evidence from real-time spoken language comprehension. *Developmental Psychology, 45*, 1723-1739.

Issard, C., & Gervain, J. (2017). Adult-like processing of time-compressed speech by newborns: A NIRS study. *Developmental Cognitive Neuroscience*, *25*, 176-184.

Issard, C., & Gervain, J. (2018). Variability of the hemodynamic response in infants: Influence of experimental design and stimulus complexity. *Developmental Cognitive Neuroscience*, *33*, 182-193.

Jang, K. E., Tak, S., Jung, J., Jang, J., Jeong, Y., and Ye, J. C. (2009). Wavelet-MDL detrending for near infrared spectroscopy (NIRS). Journal of Biomedical Optics, 14(3), 1-13.

Jara-Ettinger, J., Gweon, H., Tenenbaum, J. B., & Schulz, L. E. (2015). Children's understanding of the costs and rewards underlying rational action. *Cognition*, *140*, 14-23.

Jara-Ettinger, J., Tenenbaum, J. B., & Schulz, L. E. (2015). Not so innocent: Toddlers' inferences about costs and culpability. *Psychological Science*, *26*(5), 633-640.

Jara-Ettinger, J., Gweon, H., Schulz, L. E., & Tenenbaum, J. B. (2016). The naïve utility calculus: Computational principles underlying commonsense psychology. *Trends in Cognitive Sciences*, *20*(8), 589-604.

Jasińska, K. K., & Petitto, L. A. (2013). How age of bilingual exposure can change the neural systems for language in the developing brain: A functional near infrared spectroscopy investigation of syntactic processing in monolingual and bilingual children. *Developmental Cognitive Neuroscience*, *6*, 87-101.

Jasińska, K. K., & Petitto, L. A. (2014). Development of neural systems for reading in the monolingual and bilingual brain: New insights from functional near infrared spectroscopy neuroimaging. *Developmental Neuropsychology*, *39*(6), 421-439.

Jasińska, K. K., Berens, M. S., Kovelman, I., & Petitto, L. A. (2017). Bilingualism yields language-specific plasticity in left hemisphere's circuitry for learning to read in young children. *Neuropsychologia*, *98*, 34-45.

Jasińska, K. K., Shuai, L., Lau, A. N., Frost, S., Landi, N., & Pugh, K. R. (2021). Functional connectivity in the developing language network in 4-year-old children predicts future reading ability. *Developmental Science*, *24*(2), e13041.

Jasper, H. H. (1958). The ten-twenty electrode system of the International Federation. *Electroencephalogr. Clin. Neurophysiol.*, *10*, 370-375.

Jaswal, V. K. (2006). Preschoolers favor the creator's label when reasoning about an artifact's function. *Cognition*, *99*(3), B83-B92.

Kaminski, J. A., Sloutsky, V. M., & Heckler, A. F. (2009). Concrete instantiations of mathematics: A double-edged sword. *Journal for Research in Mathematics Education,* 90-93.

Kampa, A., Richards, C. & Papafragou, A. (2019, November 7-10). *Preschool children generate pragmatic inferences from both words and pictures.* The 44th Boston University Conference on Language and Development, Boston, MA, United States.

Kampa, A., & Papafragou, A. (2020). Four-year-olds incorporate speaker knowledge into pragmatic inferences. *Developmental science, 23(3),* e12920.

Katsos, N., & Bishop, D. V. (2011). Pragmatic tolerance: Implications for the acquisition of informativeness and implicature. *Cognition, 120,* 67-81.

Kovács, Á. M., Téglás, E., & Endress, A. D. (2010). The social sense: Susceptibility to others' beliefs in human infants and adults. *Science*, *330*(6012), 1830-1834.

Leekam, S. R., Solomon, T. L., & Teoh, Y. S. (2010). Adults' social cues facilitate young children's use of signs and symbols. *Developmental Science*, *13*(1), 108-119.

Leslie, A. M. (2000). Theory of mind as a mechanism of selective attention. In M. Gazzaniga (Ed.), *The New Cognitive Neurosciences* (pp. 1235-1247). Cambridge, MA: MIT Press.

Levinson, S. (2000). *Presumptive meanings*. Cambridge, MA: MIT Press.

Light, P., & Simmons, B. (1983). The effects of a communication task upon the representation of depth relationships in young children's drawings. *Journal of Experimental Child Psychology*, *35*(1), 81-92.

Liszkowski, U., Carpenter, M., & Tomasello, M. (2008). Twelve-month-olds communicate helpfully and appropriately for knowledgeable and ignorant partners. *Cognition*, *108*(3), 732-739.

Long, B., Fan, J., Chai, Z., & Frank, M. C. (2019). Developmental changes in the ability to draw distinctive features of object categories. https://psyarxiv.com/8rzku Accessed Feb. 2020.

Luquet, G. H. (1927). Le dessin enfantin. (Bibliothèque de psychologie de l" enfant et de pédagogie.).

Markman, E. M. (1994). Constraints on word meaning in early language acquisition. *Lingua*, *92*(1-4), 199-227.

Marty, P., Chemla, E., & Spector, B. (2013). Interpreting numerals and scalar items under memory load. *Lingua*, *133*, 152-163.

Marzolf, D. P., & DeLoache, J. S. (1994). Transfer in young children's understanding of spatial representations. *Child Development*, *65*(1), 1-15.

Matthews, J. (1984). Children drawing: Are young children really scribbling?. *Early child development and care*, *18*(1-2), 1-39.

Matthews, D., Biney, H., & Abbot-Smith, K. (2018). Individual differences in children's pragmatic ability: a review of associations with formal language, social cognition, and executive functions. *Language Learning and Development*, *14*(3), 186-223.

Matthews, D., Lieven, E., Theakston, A., & Tomasello, M. (2006). The effect of perceptual availability and prior discourse on young children's use of referring expressions. *Applied Psycholinguistics*, *27*, 403-422.

Meltzoff, A. N. (1995). Understanding the intentions of others: re-enactment of intended acts by 18-month-old children. *Developmental psychology*, *31*(5), 838.

Moore, R., Mueller, B., Kaminski, J., & Tomasello, M. (2015). Two-year-old children but not domestic dogs understand communicative intentions without language, gestures, or gaze. *Developmental Science*, *18*(2), 232-242.

Molfese, P., Glen, D., Mesite, L., Pugh, K., & Cox, R. (2015). 21st annual meeting of the Organization for Human Brain Mapping (OHBM).

Montepare, J., Koff, E., Zaitchik, D., & Albert, M. (1999). The use of body movements and gestures as cues to emotions in younger and older adults. *Journal of Nonverbal Behavior*, *23*(2), 133-152.

Nadig, A.S., & Sedivy, J.C. (2002). Evidence of perspective-taking constraints in children’s on-line reference resolution. *Psychological Science*, *13*, 329–336.

Nishiyori, R. (2016). fNIRS: An emergent method to document functional cortical activity during infant movements. *Frontiers in Psychology, 7,* 533.

Noveck, I. (2001). When children are more logical than adults: Experimental investigations of scalar implicature. *Cognition*, *78*, 165-188.

Noveck, I. A., Guelminger, R., Georgieff, N., & Labruyere, N. (2007). What autism can reveal about every… not sentences. *Journal of Semantics*, *24*(1), 73-90.

Noveck, I. A., & Posada, A. (2003). Characterizing the time course of an implicature: An evoked potentials study. *Brain and Language*, *85*(2), 203-210.

O’Hare, E.D., Lu, L.H., Houston, S.M., Bookheimer, S.Y., & Sowell, E.R. (2008). Neurodevelopmental changes in verbal working memory load-dependency: an fMRI investigation. *NeuroImage*, *42*, 1678–1685.

Olesen, P.J., Macoveanu, J., Tegner, J., & Klingberg, T. (2007). Brain activity related to working memory and distraction in children and adults. *Cerebral Cortex*, *17*, 1047–1054.

O'Neill, D. K. (1996). Two‐year‐old children's sensitivity to a parent's knowledge state when making requests. *Child Development*, *67*(2), 659-677.

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

Ozturk, O., & Papafragou, A. (2016). The acquisition of evidentiality and source monitoring. *Language Learning and Development*, *12*(2), 199-230.

Papafragou, A., Friedberg, C., & Cohen, M. (2018). The role of speaker knowledge in children’s pragmatic inferences. *Child Development, 89,* 1642-1656.

Papafragou, A. & Musolino, J. (2003), Scalar implicatures: experiments at the semantics-pragmatics interface. *Cognition, 86,* 253-282.

Papafragou, A., & Tantalou, N. (2004). Children’s computation of implicatures. *Language Acquisition, 12*, 71-82.

Perner, J., & Leekam, S. R. (1986). Belief and quantity: Three-year olds' adaptation to listener's knowledge. *Journal of Child Language*, *13*(2), 305-315.

Perner, J., & Wimmer, H. (1985). “John thinks that Mary thinks that…” attribution of second-order beliefs by 5-to 10-year-old children. *Journal of Experimental Child Psychology*, *39*(3), 437-471.

Piaget, J. (1970). *Science of Education and the Psychology of the Child*. New York: Viking.

Piaget, J., & Inhelder, B. (1948). *La représentation de l'espace chez l'enfant*.

Piaget, J., & Inhelder, B. (1967). In: *The Child's Conception of Space*, translated by FJ Langdon and JL Lunzer, New York: WW Norton.

Pijnacker, J., Hagoort, P., Buitelaar, J., Teunisse, J. P., & Geurts, B. (2009). Pragmatic inferences in high-functioning adults with autism and Asperger syndrome. *Journal of Autism and Developmental Disorders*, *39*(4), 607.

Politzer-Ahles, S., Fiorentino, R., Jiang, X., & Zhou, X. (2013). Distinct neural correlates for pragmatic and semantic meaning processing: An event-related potential investigation of scalar implicature processing using picture-sentence verification. *Brain Research*, *1490*, 134-152.

Politzer-Ahles, S., & Gwilliams, L. (2015). Involvement of prefrontal cortex in scalar implicatures: evidence from magnetoencephalography. *Language, Cognition and Neuroscience*, *30*(7), 853-866.

Pouscoulous, N., Noveck, I. A., Politzer, G., & Bastide, A. (2007). A developmental investigation of processing costs in implicature production. *Language Acquisition*, *14*(4), 347-375.

Preissler, M. A., & Carey, S. (2005). The role of inferences about referential intent in word learning: Evidence from autism. *Cognition*, *97*(1), B13-B23.

Quaresima, V., Bisconti, S., & Ferrari, M. (2012). A brief review on the use of functional near-infrared spectroscopy (fNIRS) for language imaging studies in human newborns and adults. *Brain and Language*, *121*(2), 79-89.

R-Core-Team (2012). R: A language and environment for statistical computing. Vienna, Austria.

Regier, T. (2005). The emergence of words: Attentional learning in form and meaning. *Cognitive Science*, *29*(6), 819-865.

Rentschler, I., Jüttner, M., Osman, E., Müller, A., & Caelli, T. (2004). Development of configural 3D object recognition. *Behavioural Brain Research*, *149*(1), 107-111.

Richert, R. A., & Lillard, A. S. (2002). Children's understanding of the knowledge prerequisites of drawing and pretending. *Developmental Psychology*, *38*(6), 1004.

Roberts, C. (1996). Information structure in discourse: Towards an integrated formal theory of pragmatics. *Working Papers in Linguistics-Ohio State University Department of Linguistics*, 91-136.

Roberts, C. (2004). Context in dynamic interpretation. In L. Horn & G. Ward (Eds.), *The Handbook of Pragmatics*, (pp. 197-220).

Rollins, P. R. (1999). Early pragmatic accomplishments and vocabulary development in preschool children with autism. *American Journal of Speech-Language Pathology*, *8*(2), 181-190.

Rose, S.A. (1977). Infants' transfer of response between two-dimensional and three-dimensional stimuli. *Child Development*, *48*, 1086-1091.

Sassa, Y., Sugiura, M., Jeong, H., Horie, K., Sato, S., & Kawashima, R. (2007). Cortical mechanism of communicative speech production. *Neuroimage*, *37*(3), 985-992.

Sauerland, U. (2004), Scalar implicatures in complex sentences. *Linguistics and Philosophy, 27*, 367–91.

Sauerland, U. (2012). The computation of scalar implicatures: Pragmatic, lexical or grammatical? *Language and Linguistics Compass, 6,* 36-49.

Saxe, R., Moran, J. M., Scholz, J., & Gabrieli, J. (2006). Overlapping and non-overlapping brain regions for Theory of Mind and self reflection in individual subjects. *Social, Cognitive and Affective Neuroscience*, *1*(3), 229-234.

Saxe, R. R., Whitfield-Gabrieli, S., Scholz, J., & Pelphrey, K. A. (2009). Brain regions for perceiving and reasoning about other people in school-aged children. *Child Development*, *80*(4), 1197-1209.

Schulz, K. P., Fan, J., Tang, C. Y., Newcorn, J. H., Buchsbaum, M. S., Cheung, A. M., & Halperin, J. M. (2004). Response inhibition in adolescents diagnosed with attention deficit hyperactivity disorder during childhood: an event-related FMRI study. *American Journal of Psychiatry*, *161*(9), 1650-1657.

Shalinsky, M. H., Kovelman, I., Berens, M. S., & Petitto, L. A. (2009). Exploring cognitive functions in babies, children & adults with near infrared spectroscopy. *JoVE (Journal of Visualized Experiments)*, (29), e1268.

Shatz, M. (1980). Communication. In P. Mussen (Ed.), *Handbook of Child Psychology: Vol. 3: Cognitive Development* (Vol. Eds. J. Flavell & E. Markman), 841–889. New York, NY: Wiley.

Shatz, M., & Gelman, R. (1973). The development of communication skills: Modifications in the speech of young children as a function of listener. *Monographs of the Society for Research in Child Development*, 1-38.

Shetreet, E., Chierchia, G., & Gaab, N. (2014a). When some is not every: Dissociating scalar implicature generation and mismatch. *Human Brain Mapping*, *35*(4), 1503-1514.

Shetreet, E., Chierchia, G., & Gaab, N. (2014b). Linguistic inability or poor performance: Dissociating scalar implicature generation and mismatch in the developing brain. *Developmental Psychology*, *50*(9), 2264.

Simmonds, D.J., Fotedar, S.G., Suskauer, S.J., Pekar, J.J., Denckla, M.B., & Mostofsky, S.H. (2007). Functional brain correlates of response time variability in children. *Neuropsychologia*, *45*, 2147– 2157.

Singh, A. K., Okamoto, M., Dan, H., Jurcak, V., & Dan, I. (2005). Spatial registration of multichannel multi-subject fNIRS data to MNI space without MRI. *Neuroimage*, *27*(4), 842-851.

Singh, R., Wexler, K., Astle-Rahim, A., Kamawar, D., & Fox, D. (2016). Children interpret disjunction as conjunction: Consequences for theories of implicature and child development. *Natural Language Semantics*, *24*(4), 305-352.

Skordos, D., & Papafragou, A. (2016). Children's derivation of scalar implicatures: Alternatives and relevance. *Cognition, 153*, 6-18.

Slaughter, V. (1998). Children's understanding of pictorial and mental representations. *Child Development*, *69*(2), 321-332.

Southgate, V., Chevallier, C., & Csibra, G. (2010). Seventeen-month-olds appeal to false beliefs to interpret others' referential communication. *Developmental Science*, *13*(6), 907-912.

Sperber, D., & Wilson, D. (1986). *Relevance: Communication and Cognition*. Cambridge, MA: Harvard University Press. 2nd ed. 1995.

Spychalska, M., Kontinen, J., & Werning, M. (2014). Electrophysiology of pragmatic processing: exploring the processing cost of the scalar implicature in the truth-value judgment task. In *Proceedings of the Annual Meeting of the Cognitive Science Society* (Vol. 36, No. 36).

Stalnaker, R. (1979). Assertion. In P. Cole (Ed.), Syntax and semantics 9 (pp. 315–332). New York: Academic Press.

Stiller, A., Goodman, N. D., & Frank, M. C. (2015). Ad-hoc implicature in preschool children. *Language, Learning, and Development*, *11*, 176-190.

Suskauer, S. J., Simmonds, D. J., Caffo, B. S., Denckla, M. B., Pekar, J. J., & Mostofsky, S. H. (2008). fMRI of intrasubject variability in ADHD: anomalous premotor activity with prefrontal compensation. *Journal of the American Academy of Child & Adolescent Psychiatry*, *47*(10), 1141-1150.

Taga, G., Asakawa, K., Hirasawa, K., & Konishi, Y. (2003). Hemodynamic responses to visual stimulation in occipital and frontal cortex of newborn infants: a near-infrared optical topography study. *Early Human Development*, *75*, 203-210.

Taga, G., Asakawa, K., Maki, A., Konishi, Y., & Koizumi, H. (2003). Brain imaging in awake infants by near-infrared optical topography. *Proceedings of the National Academy of Sciences*, *100*(19), 10722-10727.

Tardif, T., Fletcher, P., Liang, W., Zhang, Z., Kaciroti, N., & Marchman, V. A. (2008). Baby's first 10 words. *Developmental Psychology*, *44*(4), 929.

Thomas, G. V., Nye, R., & Robinson, E. J. (1994). How children view pictures: Children's responses to pictures as things in themselves and as representations of something else. *Cognitive Development*, *9*(2), 141-164.

Tieu, L., Romoli, J., Zhou, P., & Crain, S. (2015). Children's knowledge of free choice inferences and scalar implicatures. *Journal of Semantics*, *33*(2), 269-298.

Tomasello, M. (1992). *First verbs: A case study of early grammatical development*. Cambridge University Press.

Tomasello, M. (2000). The item-based nature of children's early syntactic development. *Trends in cognitive sciences*, *4*(4), 156-163.

Tomasello, M. (2010). *Origins of human communication.* Cambridge, MA: MIT press.

Tomasello, M., Call, J., & Gluckman, A. (1997). Comprehension of novel communicative signs by apes and human children. *Child development*, *68*(6), 1067-1080.

Tomasello, M., Striano, T., & Rochat, P. (1999). Do young children use objects as symbols? *British Journal of Developmental Psychology*, *17*(4), 563-584.

Uttal, D. H. (1994). Preschoolers' and adults' scale translation and reconstruction of spatial information acquired from maps. *British Journal of Developmental Psychology*, *12*(3), 259-275.

Uttal, D. H., O'Doherty, K., Newland, R., Hand, L. L., & DeLoache, J. (2009). Dual representation and the linking of concrete and symbolic representations. *Child Development Perspectives*, *3*(3), 156-159.

Uttal, D., Schreiber, J. C., & DeLoache, J. S. (1995). Waiting to use a symbol: The effects of delay on children's use of models. *Child Development*, *66*(6), 1875-1889.

Vivanti, G., & Rogers, S.J. (2011). Action understanding and social learning in autism: A developmental perspective. *Life Span and Disability*, *14* (1), 7-29.

Van Tiel, B., & Kissine, M. (2018). Quantity-based reasoning in the broader autism phenotype: A web-based study. *Applied Psycholinguistics*, *39*(6), 1373-1403.

Wagstaff, K., Cardie, C., Rogers, S., & Schrödl, S. (2001, June). Constrained k-means clustering with background knowledge. In *Icml* (Vol. 1, pp. 577-584).

Walsh, B., Tian, F., Tourville, J. A., Yücel, M. A., Kuczek, T., & Bostian, A. J. (2017). Hemodynamics of speech production: an fNIRS investigation of children who stutter. *Scientific Reports*, *7*(1), 1-13.

Watanabe, H., Homae, F., & Taga, G. (2012). Activation and deactivation in response to visual stimulation in the occipital cortex of 6-month-old human infants. *Developmental Psychobiology*, *54*(1), 1-15.

Waxman, S., & Booth, A. (2003). The origins and evolution of links between word learning and conceptual organization: New evidence from 11-month-olds. *Developmental Science*, *6*(2), 128-135.

Weintraub, S., Dikmen, S. S., Heaton, R. K., Tulsky, D. S., Zelazo, P. D., Bauer, P. J., Carlozzi, N. E., Slotkin, J., Blitz, D., Wallner-Allen, K., Fox, N. A., Beaumont, J. L., Mungas, D., Nowinski, C. J., Richler, J., Deocampo, J. A., Anderson, J. E. , Manly, J. J., Borosh, B., Havlik, R., Conway, K., Edwards, E., Freund, L., King, J.W., Moy, C., Witt, E., and Gershon, R. C. (2013). Cognition assessment using the NIH Toolbox. *Neurology*, *80*(11 Supplement 3), S54-S64.

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

Willems, R. M., & Varley, R. A. (2010). Neural insights into the relation between language and communication. *Frontiers in Human Neuroscience*, *4,* 203.

Wilson, D., & Sperber, D. (2002). Relevance theory. In G. Ward & L. Horn (Eds.), *Handbook of Pragmatics*. Blackwell.

Wimmer, H., & Perner, J. (1983). Beliefs about beliefs: Representation and constraining function of wrong beliefs in young children's understanding of deception. *Cognition*, *13*(1), 103-128.

Ye, J. C., Tak, S., Jang, K. E., Jung, J., & Jang, J. (2009). NIRS-SPM: statistical parametric mapping for near-infrared spectroscopy. *Neuroimage*, *44*(2), 428-447.

Yamagata, K. (1997). Representational activity during mother-child interaction: The scribbling stage of drawing. *British Journal of Developmental Psychology*, *15*(3), 355-366.

Zaitchik, D. (1990). When representations conflict with reality: The preschooler's problem with false beliefs and "false" photographs. *Cognition*, *35*(1), 41-68.

Zelazo, P. D., & Müller, U. (2011). *Executive function in typical and atypical development.* In U. Goswami (Ed.), *The Wiley-Blackwell handbook of childhood cognitive development* (p. 574–603). Wiley-Blackwell.

Zelazo, P. D., Frye, D., & Rapus, T. (1996). An age-related dissociation between knowing rules and using them. *Cognitive Development*, *11*(1), 37-63.

Zelazo, P. D., Anderson, J. E., Richler, J., Wallner-Allen, K., Beaumont, J. L., Conway, K. P., Gershon, R., & Weintraub, S. (2014). NIH Toolbox Cognition Battery (CB): Validation of executive function measures in adults. *Journal of the International Neuropsychological Society: JINS*, *20*(6), 620.

Appendix A

STIMULI LIST FOR EXPERIMENTS 1-6

Trial number	Full access box		Limited access box		Sentence or Drawing
	Left compartment	Right compartment	Left compartment	Right compartment*	
Pretest 1	Spoon	Doll	Cup	Giraffe	Spoon, Doll
Pretest 2	Pumpkin	Giraffe	Empty	Teacup	Nothing
Test 1	Spoon	Bowl	Spoon	Bowl	Spoon, Bowl
Test 2	Penguin	Pumpkin	Penguin	Pumpkin	Penguin
Test 3	Basket	Monkey	Basket	Monkey	Basket, Monkey
Test 4	Bear	Book	Bear	Book	Bear
Test 5	Teacup	Plate	Teacup	Plate	Teacup
Test 6	Doll	Pumpkin	Doll	Pumpkin	Doll, Pumpkin
Test 7	Spoon	Monkey	Spoon	Monkey	Spoon
Test 8	Bowl	Penguin	Bowl	Penguin	Bowl, Penguin

* Compartment is visible to participant, but not to “Danielle,” the girl who is creating the drawings

Appendix B

SUPPLEMENTARY MATERIALS FOR CHAPTER 4

Materials and Methods

Control condition

In the Control trials, the visuals were identical to the other conditions except that the items in the two boxes differed (see Figure S1). The sentences were identical to the Scalar-Inference condition (indicating the object that is visible to the speaker in the limited-access box; e.g., "I see a spoon"), but in this case the referent was unambiguous, so participants did not need to engage in implicature reasoning or perspective taking to select the appropriate box.

Results

Control condition

Children who were designated Passers in the Scalar-Inference condition performed well on the Control trials (*M* = 96.9%, *range:* 81.3% to 100%), while children designated Non-Passers had more difficulty (*M* = 62.4%, *range:* 18.8% to 87.5%).

Figures

Figure B.1: An example trial from the Control condition. In each trial, participants saw each photo on the screen appear (with a 2s delay in between) then heard the sentence indicated and were expected to select one photo (correct answer highlighted here). The photo on the left within each image depicts the “full-access” box (and the full-knowledge speaker). The photo on the right depicts the “limited-access box” (and the limited-knowledge speaker). The faces have been blurred in this image per University of Delaware policy but were not blurred for the experiment.

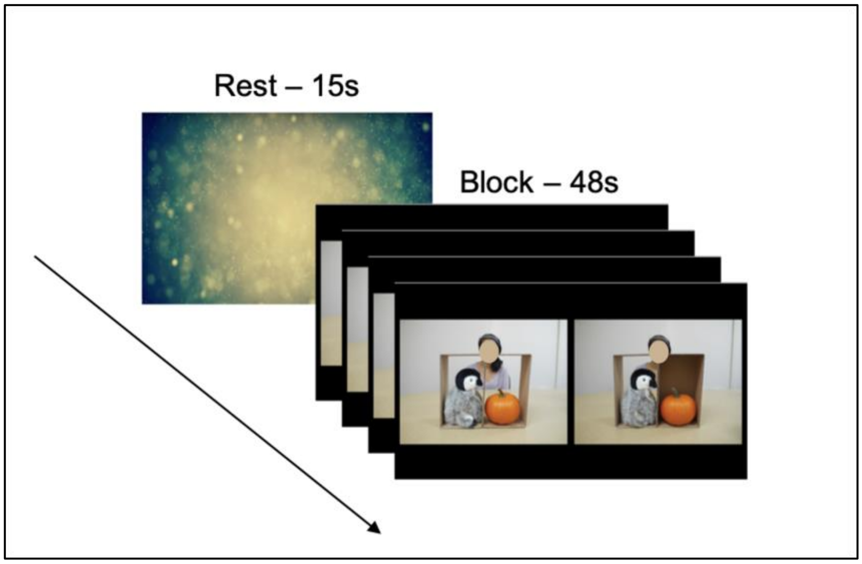

Figure B.2: Timing for presentation of stimuli. Participants completed 2 runs, consisting of 6 blocks (2 Perspective-Taking, 2 Scalar-Inference, 2 Control) with a 15 second rest before each block.

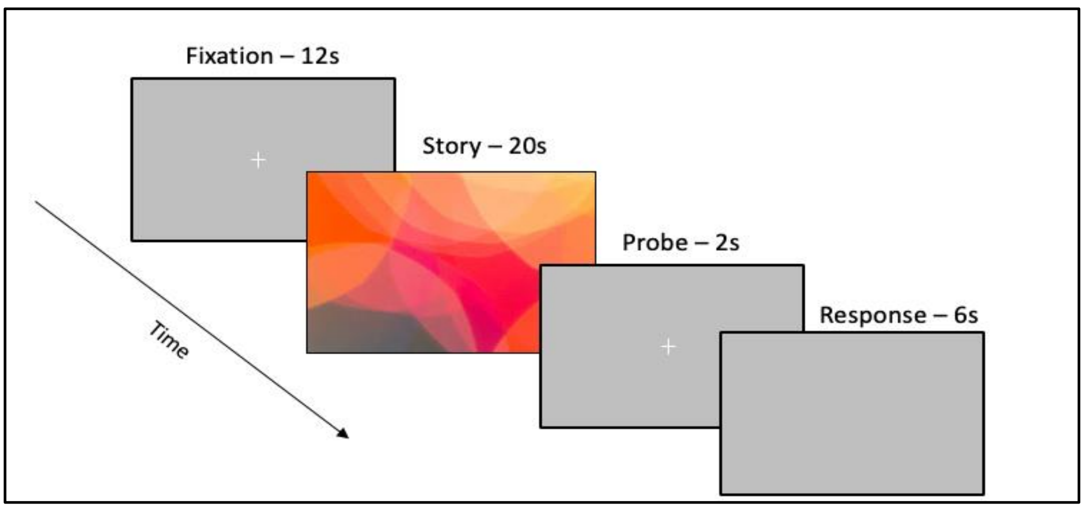

Figure B.3: Timing for presentation of Theory of Mind stimuli for one story. Participants completed 2 (children) or 4 (adults) runs, consisting of 4 stories (2 Mental, 2 Physical) with a 12 second fixation before each story. Each story was followed by a probe sentence after which participants indicated their response.

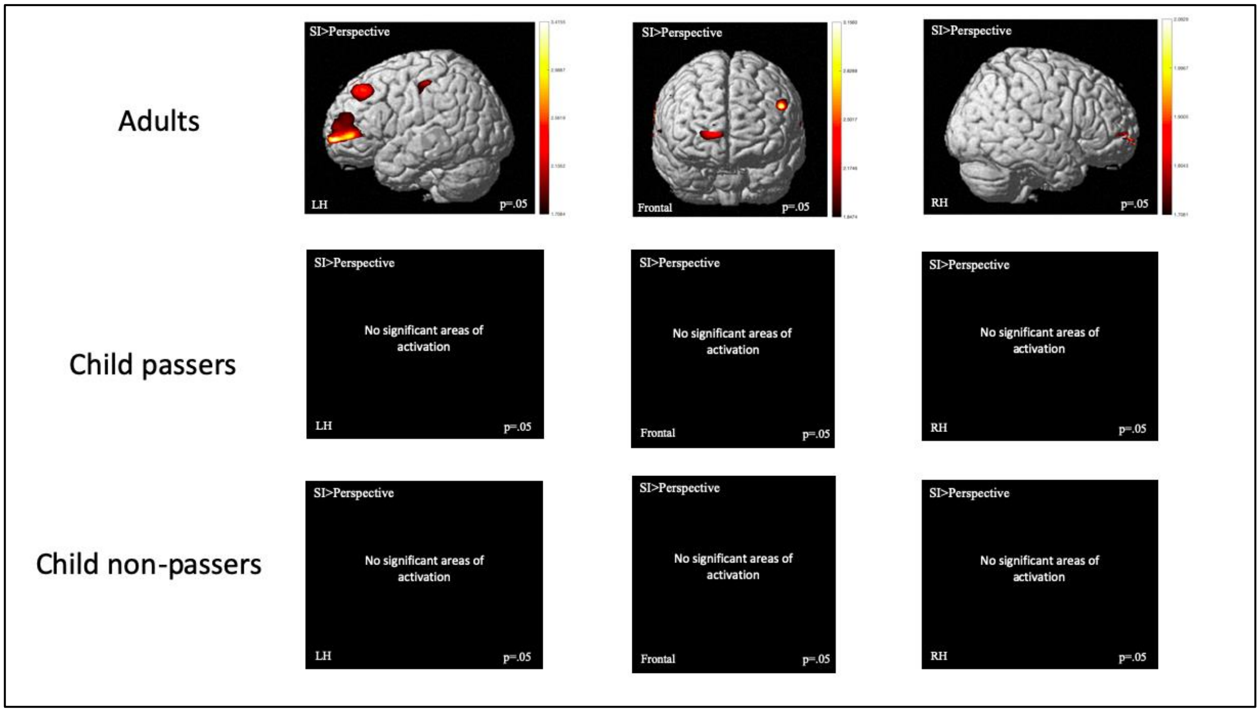

Figure B.4: Activation patterns in the Scalar Implicature task for adults, child passers, and child non-passers. Activation is shown for HbR for adults and HbO for children.

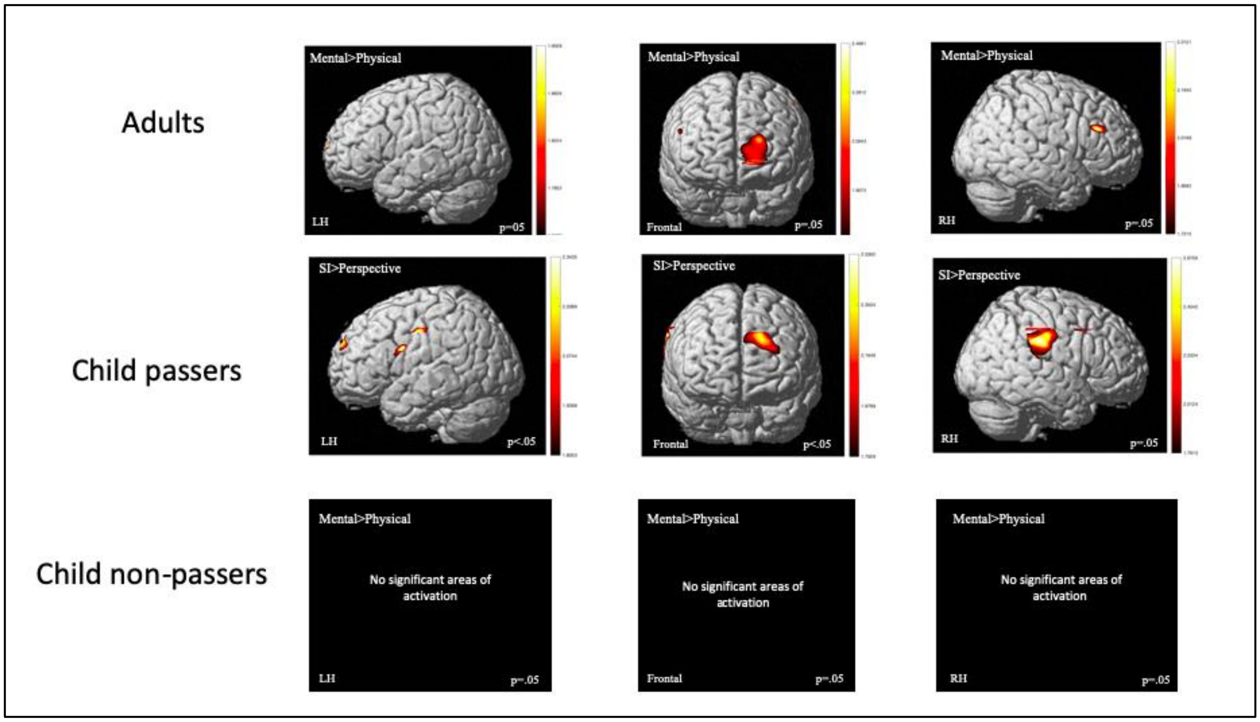

Figure B.5: Neural activation patterns for adults, child scalar-implicature passers and child scalar-implicature non-passers in the Theory of Mind neuroimaging task with an HbR contrast

Tables

Table B.1: Additional Regions with Significant HbO Activation for Scalar Inference for Adults.

Contrast	Region	Brodmann Area	X	Y	Z	β value	*p* value
SI > PT	Wernicke's area	40	-52.3	-46.3	57	0.00153	<0.05
SI > PT	Frontopolar area	10	-13.7	64.7	29.3	0.00152	<0.05
SI > PT	Supramarginal gyrus	40	-65	-35	43.7	0.00152	<0.05
SI > PT	Broca's area	45	-54.7	39	0	0.00149	<0.05
SI > PT	Somatosensory cortex	1,3	-54	-23.3	59.3	0.00145	<0.05
SI > PT	Dorsolateral PFC	9	42.3	38.3	41.3	0.00143	<0.05
SI > PT	Dorsolateral PFC	46	-39.7	51.7	26.7	0.00141	<0.05
SI > PT	Wernicke's area	40	56	50	54.3	0.00137	<0.05
SI > PT	Frontopolar area	40	-39	62	-3	0.00136	<0.05
SI > PT	Dorsolateral PFC	9	26.3	33.3	56.3	0.00133	<0.05

* This table displays the ten channels with the highest level of significance; the rest can be found in supplementary materials. PFC refers to the prefrontal cortex.

Table B.2: Additional Regions with Significant HbR Activation for Scalar Inference for Child Passers.

Contrast	Region	Brodmann Area	X	Y	Z	β value	*p* value
SI > PT	Subcentral Area	43	70	-9.3	26.3	0.0003	<0.05
SI > PT	Dorsolateral PFC	46	-25.7	62.7	24.7	0.0003	<0.05

* PFC refers to the prefrontal cortex

Appendix C

IRB APPROVAL FOR HUMAN SUBJECTS RESEARCH

Institutional Review Board
210H Hullihen Hall
Newark, DE 19716
Phone: 302-831-2137
Fax: 302-831-2828

DATE:	February 25, 2020
TO:	Anna Papafragou, PhD
FROM:	University of Delaware IRB
STUDY TITLE:	[312739-21] Language Acquisition: Word learning and pragmatic inference
SUBMISSION TYPE:	Continuing Review/Progress Report
ACTION:	APPROVED
APPROVAL DATE:	February 25, 2020
EXPIRATION DATE:	March 1, 2021
REVIEW TYPE:	Expedited Review
REVIEW CATEGORY:	Expedited review category # *(7)*

Thank you for your Continuing Review/Progress Report submission to the University of Delaware Institutional Review Board (UD IRB). The UD IRB has reviewed and APPROVED the proposed research and submitted documents via Expedited Review in compliance with the pertinent federal regulations.

As the Principal Investigator for this study, you are responsible for and agree that:

- All research must be conducted in accordance with the protocol and all other study forms as approved in this submission. Any revisions to the approved study procedures or documents must be reviewed and approved by the IRB prior to their implementation. Please use the UD amendment form to request the review of any changes to approved study procedures or documents.
- Informed consent is a process that must allow prospective participants sufficient opportunity to discuss and consider whether to participate. IRB-approved and stamped consent documents must be used when enrolling participants and a written copy shall be given to the person signing the informed consent form.
- Unanticipated problems, serious adverse events involving risk to participants, and all non-compliance issues must be reported to this office in a timely fashion according with the UD requirements for reportable events. All sponsor reporting requirements must also be followed.

Oversight of this study by the UD IRB REQUIRES the submission of a CONTINUING REVIEW seeking the renewal of this IRB approval, which will expire on March 1, 2021. A continuing review/progress report form and up-to-date copies of the protocol form and all other approved study materials must be submitted to the UD IRB at least 45 days prior to the expiration date to allow for the required IRB review of that report.

If you have any questions, please contact the UD IRB Office at (302) 831-2137 or via email at hsrb-research@udel.edu. Please include the study title and reference number in all correspondence with this office.

Appendix D

IRB APPROVAL FOR HUMAN SUBJECTS RESEARCH

Institutional Review Board
210H Hullihen Hall
Newark, DE 19716
Phone: 302-831-2137
Fax: 302-831-2828

DATE:	November 11, 2020
TO:	Kaja Jasinska, PhD
FROM:	University of Delaware IRB
STUDY TITLE:	[1139108-11] Neural Basis of Linguistic and Cognitive Processing and Reading in Children and Adults
SUBMISSION TYPE:	Continuing Review/Progress Report
ACTION:	APPROVED
APPROVAL DATE:	November 11, 2020
EXPIRATION DATE:	November 12, 2021
REVIEW TYPE:	Expedited Review
REVIEW CATEGORY:	Expedited review category # *(4,6,7)*

Thank you for your Continuing Review/Progress Report submission to the University of Delaware Institutional Review Board (UD IRB). The UD IRB has reviewed and APPROVED the proposed research and submitted documents via Expedited Review in compliance with the pertinent federal regulations.

As the Principal Investigator for this study, you are responsible for and agree that:

- All research must be conducted in accordance with the protocol and all other study forms as approved in this submission. Any revisions to the approved study procedures or documents must be reviewed and approved by the IRB prior to their implementation. Please use the UD amendment form to request the review of any changes to approved study procedures or documents.
- Informed consent is a process that must allow prospective participants sufficient opportunity to discuss and consider whether to participate. IRB-approved and stamped consent documents must be used when enrolling participants and a written copy shall be given to the person signing the informed consent form.
- Unanticipated problems, serious adverse events involving risk to participants, and all non-compliance issues must be reported to this office in a timely fashion according with the UD requirements for reportable events. All sponsor reporting requirements must also be followed.

Oversight of this study by the UD IRB REQUIRES the submission of a CONTINUING REVIEW seeking the renewal of this IRB approval, which will expire on November 12, 2021. A continuing review/progress report form and up-to-date copies of the protocol form and all other approved study materials must be submitted to the UD IRB at least 45 days prior to the expiration date to allow for the required IRB review of that report.

If you have any questions, please contact the UD IRB Office at (302) 831-2137 or via email at hsrb-research@udel.edu. Please include the study title and reference number in all correspondence with this office.

Appendix E

PERMISSIONS

Chapter 2 of this dissertation was published in the journal Developmental Science.

Kampa, A. & Papafragou, A. (2020). Four-year-olds incorporate speaker knowledge into pragmatic inferences. *Developmental Science*, *23*, e12920. https://doi.org/10.1111/desc.12920

This re-use of the publication is in accordance with the copyright agreement, which states:

> The Owner hereby licenses back to the Contributor the following rights with respect to the final published version of the Contribution (the "Final Published Version")… the right to re-use the Final Published Version or parts thereof for any publication authored or edited by the Contributor (excluding journal articles) where such re-used material constitutes less than half of the total material in such publication. In such case, any modifications must be accurately noted.

More information on licensing for Wiley journals can be found here: https://authorservices.wiley.com/author-resources/Journal-Authors/licensing/licensing-info-faqs.html