

**Children's development of Quantity, Relevance
and Manner implicature understanding
and the role of the speaker's epistemic state**

Elsbeth Amabel Wilson

Department of Theoretical and Applied Linguistics

Trinity Hall, University of Cambridge

This dissertation is submitted for the degree of

Doctor of Philosophy

August 2017

Children's development of Quantity, Relevance and Manner implicature understanding and the role of the speaker's epistemic state

Elsbeth Amabel Wilson

Summary

In learning language, children have to acquire not only words and constructions, but also the ability to make inferences about a speaker's intended meaning. For instance, if in answer to the question, 'what did you put in the bag?', the speaker says, 'I put in a book', then the hearer infers that the speaker put in *only* a book, by assuming that the speaker is informative. On a Gricean approach to pragmatics, this implicated meaning – a quantity implicature – involves reasoning about the speaker's epistemic state. This thesis examines children's development of implicature understanding. It seeks to address the question of what the relationship is in development between quantity, relevance and manner implicatures; whether word learning by exclusion is a pragmatic forerunner to implicature, or based on a lexical heuristic; and whether reasoning about the speaker's epistemic state is part of children's pragmatic competence.

This thesis contributes to research in experimental and developmental pragmatics by broadening the focus of investigation to include different types of implicatures, the relationship between them, and the contribution of other aspects of children's development, including structural language knowledge. It makes the novel comparison of word learning by exclusion with a clearly pragmatic skill – implicatures – and opens an investigation of manner implicatures in development. It also presents new findings suggesting that children's early competence with quantity implicatures in simple communicative situations belies their ongoing development in more complex ones, particularly where the speaker's epistemic state is at stake.

I present a series of experiments based on a sentence-to-picture-matching task, with children aged 3 to 7 years. In the first study, I identify a developmental trajectory whereby word learning by exclusion inferences emerge first, followed by ad hoc quantity and relevance, and finally scalar quantity inferences, which reflects their increasing complexity in a Gricean model. Then, I explore cognitive and

environmental factors that might be associated with children's pragmatic skills, and show that structural language knowledge – and, associated with it, socioeconomic status – is a main predictor of their implicature understanding. In the second study, I lay out some predictions for the development of manner implicatures, find similar patterns of understanding in children and adults, and highlight the particular challenges of studying manner implicatures experimentally. Finally, I focus on children's ability to take into account the speaker's epistemic state in pragmatic inferencing. While adults do *not* derive a quantity implicature appropriately when the speaker is ignorant, children tend to persist in deriving implicatures regardless of speaker ignorance, suggesting a continuing challenge of integrating contextual with linguistic information in utterance interpretation.

This dissertation is the result of my own work and includes nothing which is the outcome of work done in collaboration except as declared in the Preface and specified in the text.

It is not substantially the same as any that I have submitted, or, is being concurrently submitted for a degree or diploma or other qualification at the University of Cambridge or any other University or similar institution except as declared in the Preface and specified in the text. I further state that no substantial part of my dissertation has already been submitted, or, is being concurrently submitted for any such degree, diploma or other qualification at the University of Cambridge or any other University or similar institution except as declared in the Preface and specified in the text.

It does not exceed the prescribed word limit for the Modern and Medieval Languages Degree Committee (80,000 words excluding bibliography).

Acknowledgements

I am very grateful to my supervisor, Napoleon Katsos, who, many moons ago, rescued me from the brink of an MPhil crisis, introduced me to the wonderful world of experimental pragmatics, and has guided me along the doctoral road with enthusiasm, patience, rigour and care. More importantly, he has constantly sought to open up opportunities, including for some of my more non-conventional aims, and has provided an aspirational example of an academic full of integrity, generosity and graciousness.

Several colleagues have given timely nudges and food for thought which have contributed positively to this thesis. The experimental pragmatics reading group has provided feedback, lively discussion and advice: Jessica Soltys, Kyriakos Antoniou, Alma Veenstra, Özge Öztürk, Draško Kaščelan, Curtis Sharma, Isabelle Lorge, Katie Howard, and especially Dimitra Lazaridou-Chatzigoga (we never could decide on a name for ourselves, so of course the affectionate moniker had to be NapLab). I also benefitted from comments, conversations and correspondence with those further afield, particularly from my doctoral grandfather Richard Breheny, Nausicaa Pouscoulous, Cat Davies, Nicole Gotzner, Uli Sauerland and Dimitrios Skordos.

The third floor crew, who occupied the PhD office with me at various points, provided much appreciated technical support, instruction and companionship on the PhD journey – in order of appearance: Dimitris Alikaniotis and Connor Quinn (for help with R and experimental design), Rowena Bermingham, Carla Pastorino-Campos, Alim Tusun, Giulia Bovolenta, Chris Xia, and Cherry Lam.

I am very grateful to all the schools, nurseries and families who let me test their children, often with cheering willingness, to the ESRC for a studentship, and Trinity Hall for financial support with research and conference costs. Becky Lawrence and Becky Brooks also assisted with some data collection.

The church family at Eden Baptist has encouraged me with prayer and friendship. It's a long road from choosing Russian aged 11 to writing a thesis on developmental pragmatics quite a while later, but that's where it all began, and my mother, Debbie Whitton-Spriggs, was supportive all the way of my sometimes unusual interests, and patient in my ditherings (or were they enriching detours?) *en route*. More recently, she has trod the doctoral path before me – giving me first a wonderful example of how to write a thesis calmly and with a wholesome perspective, and then much guidance and wisdom for navigating PhD life. My husband, Nicholas, not only provided the voice of Bob, technical assistance and much-needed chivvying, but also cheerfully learned a great deal about implicatures; his counsel, friendship and love enabled me to keep going and actually finish this thesis.

Contents

1	The acquisition of implicatures	1
1.1	Grice and some founding principles	2
1.1.1	<i>Basic assumptions and experimental pragmatics</i>	3
1.1.2	<i>Distinct principles and implicatures</i>	6
1.1.3	<i>Inferences and alternatives</i>	8
1.1.4	<i>The Question Under Discussion</i>	10
1.2	Approach to language acquisition	11
1.3	Overview of the thesis	13
2	Quantity, relevance and word learning by exclusion inferences – pragmatics in the early years	16
2.1	Pragmatic theory and development	17
2.2	Learning to understand implicatures	18
2.2.1	<i>Relevance</i>	18
2.2.2	<i>Ad hoc quantity</i>	21
2.2.3	<i>Scalar quantity</i>	21
2.2.4	<i>Comparing implicatures</i>	23
2.2.5	<i>The QUD and relevant alternatives</i>	24
2.3	Word learning by exclusion	27
2.3.1	<i>Theories of WLE</i>	28
2.3.2	<i>Testing WLE theories</i>	31
2.3.3	<i>Summary</i>	33
2.4	Experiment 1A: Quantity, relevance and WLE inferences in children	33
2.4.1	<i>Method</i>	33
2.4.2	<i>Predictions</i>	41

2.4.3	<i>Coding</i>	41
2.4.4	<i>Analysis</i>	42
2.5	<i>Discussion</i>	52
2.5.1	<i>Findings</i>	52
2.5.2	<i>The development of implicatures</i>	54
2.5.3	<i>Improvements to task</i>	65
2.6	<i>Conclusion</i>	66
3	Factors in implicature development	67
3.1	<i>Background</i>	68
3.1.1	<i>Theory of Mind</i>	68
3.1.2	<i>Structural language</i>	74
3.1.3	<i>Socioeconomic Status</i>	76
3.1.4	<i>Monolingualism vs multilingualism</i>	78
3.2	<i>Experiment 1B: the relationship of implicatures with structural language, SES, ToM and languages spoken</i>	80
3.2.1	<i>Method</i>	80
3.2.2	<i>Results</i>	82
3.2.3	<i>Analysis</i>	86
3.3	<i>Discussion</i>	91
3.3.1	<i>Summary of findings</i>	91
3.3.2	<i>Implicatures and structural language knowledge</i>	92
3.3.3	<i>Implicatures and SES</i>	93
3.3.4	<i>Implicatures and monolingualism vs multilingualism</i>	94
3.3.5	<i>Implicatures and ToM</i>	96
3.3.6	<i>Word learning, vocabulary and mono-/multilingualism</i>	98
3.4	<i>Conclusion</i>	100

4	The development of manner implicatures	102
4.1	Background – manner implicatures in theory and in adults	102
4.1.1	<i>Theory of manner implicature</i>	<i>102</i>
4.1.2	<i>A working definition for manner.....</i>	<i>108</i>
4.1.3	<i>Evidence for manner in adults.....</i>	<i>109</i>
4.1.4	<i>Manner in development.....</i>	<i>114</i>
4.2	Experiment 2A: manner implicatures.....	120
4.2.1	<i>Method.....</i>	<i>120</i>
4.2.2	<i>Predictions.....</i>	<i>123</i>
4.2.3	<i>Results</i>	<i>125</i>
4.2.4	<i>Discussion</i>	<i>127</i>
4.3	Experiment 2B: manner implicatures	129
4.3.1	<i>Method.....</i>	<i>129</i>
4.3.2	<i>Results</i>	<i>131</i>
4.3.3	<i>Discussion</i>	<i>133</i>
4.4	General Discussion	134
4.4.1	<i>Manner implicatures in adults.....</i>	<i>134</i>
4.4.2	<i>Manner in development.....</i>	<i>141</i>
4.4.3	<i>Challenges for investigating manner</i>	<i>143</i>
4.5	Conclusion	145
5	Putting the pieces together: epistemic state and implicatures	146
5.1	Reasoning about others' epistemic states.....	147
5.2	In adults.....	148
5.2.1	<i>Perspective-taking.....</i>	<i>148</i>
5.2.2	<i>The epistemic step in implicature</i>	<i>150</i>

5.3	In children's development.....	153
5.3.1	<i>Perspective-taking</i>	153
5.3.2	<i>Perspective-taking and communication</i>	154
5.3.3	<i>Perspective-taking and implicatures</i>	155
5.3.4	<i>Summary</i>	159
5.4	Predictions.....	160
5.4.1	<i>One-step hypothesis</i>	160
5.4.2	<i>Two-step hypothesis</i>	161
5.5	Experiment 3: 'Nosy Neighbours' – ad hocs, scalars and speaker epistemic state.....	161
5.5.1	<i>Method</i>	162
5.5.2	<i>Results and analysis</i>	169
5.5.3	<i>Discussion</i>	174
5.5.4	<i>Summary</i>	181
5.6	Experiment 4: director task – ad hocs and speaker epistemic state.....	181
5.6.1	<i>Method</i>	181
5.6.2	<i>Predictions</i>	188
5.6.3	<i>Results and analysis</i>	188
5.6.4	<i>Discussion</i>	192
5.7	Conclusion.....	207
6	Conclusion	209
6.1	Summary of the thesis.....	209
6.2	Implications for developmental pragmatics.....	211
6.3	Implications for pragmatic theory.....	213
6.4	Implications for the study of developmental pragmatics.....	215
6.5	Implications for educational policy.....	217

6.6	Research outlook	219
6.6.1	<i>Implicatures</i>	220
6.6.2	<i>Structural language</i>	220
6.6.3	<i>Theory of Mind</i>	222
7	References	223
8	Appendices	253
8.1	Stimuli for Experiment 1A	253
8.2	Alternative analysis for Experiment 1A	264
8.3	Background questionnaire for Experiment 1A and 1B.....	266
8.4	Instructions for Theory of Mind tasks	270
8.5	Picture stimuli for Experiment 2A	271
8.6	Picture stimuli for Experiment 2B	273
8.7	Stimuli for Experiment 3	276
8.8	Stimuli for Experiment 4	280

Table of Figures

<i>Figure 2.1</i> Word learning by exclusion	16
<i>Figure 2.2</i> Story book, stand and magnetic board, showing scalar trial pictures	38
<i>Figure 2.3</i> Mean correct responses by utterance type, condition and age group	42
<i>Figure 2.4</i> Mean correct responses by age group, condition and utterance type	43
<i>Figure 2.5</i> Mean correct responses by adults	44
<i>Figure 2.6</i> Mean correct score by age, inference type, condition and block	48
<i>Figure 2.7</i> Mean correct responses by inference type, condition and age group	50
<i>Figure 2.8</i> Proportion correct responses for first block scalar trials only	61
<i>Figure 2.9</i> Histogram of scores by inference type and condition, for 2;8–3;11	63
<i>Figure 2.10</i> Example of improved stimuli for implicatures task	66
<i>Figure 3.1</i> Materials for Sally-Anne test.....	81
<i>Figure 3.2</i> Mean correct score for critical inferences, by age and monolingual vs multilingual	84
<i>Figure 3.3</i> Boxplots showing TROG, BPVS, SES and ToM scores	85
<i>Figure 4.1</i> Example picture cards for Experiment 2A	123
<i>Figure 4.2</i> Mean selection of M picture by condition (utterance) and age group.....	125
<i>Figure 4.3</i> Mean selection M picture by Type (PCI/GCI) and condition (utterance) ..	126
<i>Figure 4.4</i> Mean selection of M picture by item and condition	128
<i>Figure 4.5</i> Mean selection M picture by condition and age group.....	131
<i>Figure 4.6</i> Mean selection M picture by Type, condition and Age group.....	132
<i>Figure 4.7</i> Histogram of participant scores for M condition, for children (left) and adults (right) for Experiment 2B.....	134
<i>Figure 4.8</i> Mean selection of M picture by item, condition and age group	137
<i>Figure 5.1</i> Sketch of experimental stimuli from Kampa & Papafragou (2017).....	158
<i>Figure 5.2</i> Correct responses by condition, speaker knowledge, and age group.....	169
<i>Figure 5.3</i> Correct response by condition, speaker knowledge, scale and age group ..	170
<i>Figure 5.4</i> AQ score by critical (weak/strong) response and speaker knowledge	174
<i>Figure 5.5</i> Mean correct response for those at ceiling in control conditions	176
<i>Figure 5.6</i> Proportion correct responses by item, condition, and age group for scalars	

and ad hocs.....	179
<i>Figure 5.7</i> Histograms of correct picture selection for adults and children.....	180
<i>Figure 5.8</i> Example display: privileged ground ambiguous condition	183
<i>Figure 5.9</i> Example display: common ground unambiguous, common ground ad hoc implicature and privileged ground ad hoc implicature conditions	183
<i>Figure 5.10</i> Apparatus for Experiment 4	185
<i>Figure 5.11</i> Proportion correct responses by condition and age group	189
<i>Figure 5.12</i> Proportion of passers and failers in each condition	190
<i>Figure 5.13</i> Mean correct response by block (with blocks in order of presentation) for both privileged ground conditions, children only	206
<i>Figure 5.14</i> Example of perspective-taking ad hoc implicature task	207
<i>Figure 8.1</i> Pictures of novel items used in Experiment 1A.....	258
<i>Figure 8.2</i> Examples of scalar, ad hoc and relevance pictures in Experiment 1A.....	259

1 The acquisition of implicatures

I go to vote, mummy!

And how did you vote?

Fold paper, and put it in box!

(Hannah, 2 years, personal communication)

Anyone observing a child learning to communicate is at once amazed at their progress, amused at their distinctive use of language, and, perhaps, puzzled at how these young conversationalists can pick up so much, so quickly, and yet sometimes still entirely miss the point. This thesis is about how children develop language, and, in particular, how they learn to understand others in discourse, by making inferences about meaning. In other words, it is about children's pragmatic development, and especially their implicature understanding.

When we humans communicate, we typically mean far more than we literally say. Indeed, we often talk about 'reading between the lines' or 'filling in the gaps'. But this conscious awareness belies the complex inferences that we make all the time as hearers in what, on the surface, seem like trivial cases. Take the following examples – adapted from Experiment 1:

a) Bob came out of the kitchen. His mum asked, 'What did you take from the fridge?' He replied, 'I took a strawberry.'

b) Bob made a crash in the kitchen. His mum asked, 'What did you do with the pile of plates?' He said, 'I broke some of the plates.'

c) It was breakfast time. Bob's mum asked, 'What do you want for breakfast?' He said, 'I'll get the milk.'

[with choice of toast or cereal on the table]

How might Bob's mum understand his reply in each of these cases? Perhaps something like this:

a) I took only a strawberry (and nothing else)

- b) I broke only some of the plates (but not all of them)
- c) I want cereal (not toast)

On closer inspection, these everyday examples turn out to be not so simple. To understand that the speaker means that he took *only* a strawberry, the hearer must have some kind of expectation that the speaker will give a full answer to the question – that he will be maximally informative. And not only that, but the hearer also expects that the speaker himself has the same expectation, that he will be informative, and that his hearer expects him to be so. And so the hearer can infer that if he had taken anything else, he would have said so. Or take the breakfast exchange: to understand that the speaker means that he wants cereal, the hearer must be assuming that the speaker will answer the question, that he will be relevant. And, again, that the speaker himself also expects to make a relevant contribution, and expects his hearer to expect him to do so. That is, the interlocutors must have some expectations about how communication works, and expect each other to share these assumptions. It seems communication is not just a matter of encoding and decoding without reference to context, but reasoning about others' beliefs and intentions in each instance of language use.

At least, this has been one dominant view of communication, following and developing the proposals of Grice, in many ways the father of contemporary pragmatics (Grice, 1975). If this model is along the right lines, and such quotidian conversations involve complex inferencing, this raises a number of questions: Is this what children are doing when they communicate, or not? How do they acquire this skill? What difference to acquisition does the type of implicature make?

1.1 Grice and some founding principles

First, let's take a step back. In this introductory chapter, I outline the theoretical pragmatic foundations of this research, and introduce key concepts: the assumptions made by Grice and the field of experimental pragmatics; the Gricean distinction of types of implicature; the nature of inferences; and the role of the Question Under Discussion. I then turn to the approach to language acquisition within which this thesis sits. Finally, I preview the coming chapters, and the questions about children's

pragmatic development that each seeks to address.

1.1.1 Basic assumptions and experimental pragmatics

This thesis is located within the new and growing programme of research in *experimental pragmatics* (Schwarz, in press), that was catalysed, at least in part, by a study on this very topic – the acquisition of implicatures (Noveck, 2001). It is also firmly *Gricean* in its approach to language (Grice, 1989). Throughout the thesis, as I consider the theoretical background or implications for each experiment, I tend to start with Grice, but also draw significantly on those that have followed him, and in particular those within experimental and probabilistic pragmatics that have clarified or tightened up its model, with a constraint-based approach (e.g. Degen & Tanenhaus, 2014; Franke & Jäger, 2016; Geurts & Rubio-Fernández, 2015).

A first very basic assumption of experimental pragmatics as a field – and the major contribution of Grice himself – is that linguistic meaning has different components: on the one hand, the literal content and, on the other, the pragmatic content. Where the line should be drawn – and, indeed, how many lines – may have varied (e.g. Bach, 2007; Levinson, 2000; D. Wilson & Sperber, 2012), but the distinction is widely acknowledged. And it has immediate consequences for children’s language development, because the challenge for the infant-communicator is to learn to understand both meaning that is cross-situationally stable – semantics – and that which is contextually dependent – pragmatics. The hearer’s job is to infer the speaker’s communicative intention – through *pragmatic* reasoning.¹

Context is clearly a key concept in pragmatics, though often left undefined. Here I do not subscribe to a particular theory of context, but take a broad view that it may include any features of the immediate situation, the preceding discourse, speaker characteristics and shared knowledge. Importantly for empirical studies, this means that there are two aspects of context in any experiment, as Meibauer (2012) points out: ‘first, the context of the experimental setting, second the context evoked by test materials’ (2012: 23). Context is often assumed to be coextensive with common

¹ Note, though, that making this distinction does not commit oneself to a two-step processing model (first semantic, then pragmatic content), which is a different level of analysis, as shall be seen shortly.

ground (H. H. Clark, 1996) – the mutual knowledge that can be presupposed by the speaker and hearer – as opposed to privileged ground – knowledge available only to the speaker or the hearer.

A second basic assumption is that whatever the guiding principles or reasons in a model of communication, they describe or explain communication in general, not particular phenomena. Grice was clear that his principles were based on rational and co-operative human behaviour (Grice, 1989), a claim that has been clearly embraced by probabilistic or computational approaches to pragmatics, which have thus pursued its application to a wide range of pragmatic phenomena, including reference (Franke & Degen, 2016), negation (Nordmeyer & Frank, 2014) and gradable adjectives (Qing & Franke, 2014). This, too, has implications for acquisition, leading us to expect some element of commonality in children’s pragmatic development, broadly construed. This thesis, however, is about implicatures, and I will shortly turn to the distinct categories of implicature proposed by Grice.

It is important to mention at the outset what kind of model the Gricean approach to pragmatics provides – especially as this has been the cause of some worry in recent years. Building on Marr’s (1982) three levels of analysis – computational, algorithmic, and implementational – Franke and Jaeger (2016) helpfully distinguish four different – though not necessarily exhaustive – levels of analysis for pragmatic theory, all at the computational level. Firstly, addressing the question ‘what?’, there are *constraints*, and there are *principles* (e.g. Hurford’s constraint): these are both primarily descriptive, such that the data they are aimed at accounting for is entailed by the abstract structure or element proposed. Secondly, there is the level of *maxims*, like Grice’s own, that are on the cusp of ‘what?’ and ‘why?’ – part description, part explanation. Thirdly, *reasons* – such as speaker rationality – focus only on the ‘why?’. Crucially, these are all distinct from – though potentially related to – *processes*, the ‘how?’ of the ‘what?’ and the ‘why?’. Some have criticised Grice on the basis of being ‘psychologically implausible’, but these typically stem from an expectation that it is about processes: for example, suggesting that recursive mindreading is too costly (e.g. D. Wilson, 2000), or that ‘semantic’ interpretations should be universally less costly (i.e., shorter time course) than ‘pragmatic’ ones, which they are not (e.g. Scott-Phillips, 2014). In other words, they confuse *description* and *reasons* at the computational level, with

processes, at a more algorithmic one, or at least overlook the importance of the former, as Geurts and Rubio-Fernandez (2015) argue.

Like Grice's theory, this thesis hovers between asking 'what?' and asking 'why?'. I am interested in what children do, and why they do it, in terms of rational, co-operative communication and Gricean maxims. I largely try to stay clear of leaping to what this might mean for cognitive processing – although this can be challenging when talking about development in particular, because general cognitive development of course has ramifications for interpretation of behaviour (for instance, arguments such as 'children lack inhibitory control to suppress contextual alternatives'). This might seem particularly to be the case in discussions about the relationship between Theory of Mind and pragmatic reasoning. It, too, is a theory, though, articulated in terms of description of behaviour, heuristics and reasons, just like Gricean pragmatics: people have emotions, desires and beliefs, expect others to have them too, and reason about those mental states, which lays the foundation for our communication.²

However, making predictions from Grice's maxims for children's behaviour does require certain linking hypotheses and assumptions. A key notion in studying children's development is what is *challenging* for them – what they can do at a younger or older age, or in more or less supportive contexts – and this can be married with the complexity found in the model, in this case, of the derivation of different implicatures. One linking hypothesis could be that more complex inferencing translates to more complex processing, which translates to being more challenging to acquire. Until more is known about processing of implicatures in comprehension (and not just scalar ones), this is quite an uncertain step. An attractive alternative, albeit a simplistic one, is that more complex inferencing – involving more components – translates into more skills to acquire, and therefore possible later development. This is my working assumption – and it is an assumption – that provides some traction on formulating expectations for children's pragmatic development from theory.³

² Just as for pragmatics, this is not to say that Theory of Mind cannot be modelled in terms of processes, just that it is not always and does not have to be, and this is the level I operate at here. Some approaches to Theory of Mind, like Apperly's (2011) 'two-systems' approach clearly do try to bridge from reasons to processes.

³ Obviously, one can easily imagine where 10 simple components would be easier than one

1.1.2 Distinct principles and implicatures

As I have already mentioned, Grice's model of communication is grounded on human rationality, and it is captured by an overarching heuristic that describes communicative behaviour: The Co-operative Principle. It states: 'Make your contribution such as is required, at the stage at which it occurs, by the accepted purpose or direction of the talk exchange in which you are engaged' (Grice, 1989: 26). In other words, say the right thing, at the right time, in the right way. And what is 'right' in communication? In addition to the foundational *maxim of quality* – be truthful – three further maxims flesh this out. To paraphrase: the *maxim of quantity* – be informative, but not overly so; the *maxim of relevance* – be relevant; and the *maxim of manner* – be clear and conventional. When speakers appear to not follow these maxims in the literal content of their utterance, hearers can infer their intended meaning, the implicature, on the assumption that they are in fact doing so. The maxims capture expectations of both speakers and hearers, but as I am particularly interested in pragmatic inferencing in comprehension, here I will concentrate more on the hearers' perspective.

Taking the examples from above, with one new case, it can be seen how this might work; each utterance is followed by the implicature, and an indication of the reason for its derivation.

a) Bob came out of the kitchen. His mum asked, 'What did you take from the fridge?' He replied, 'I took a strawberry.'

→ I took only a strawberry (and nothing else)

Otherwise, the speaker would have said so, to be optimally informative:

Quantity (ad hoc)

b) Bob made a crash in the kitchen. His mum asked, 'What did you do with the pile of plates?' He said, 'I broke some of the plates.'

→ I broke only some of the plates (but not all of them)

Otherwise, the speaker would have said so, to be optimally informative:

Quantity (scalar)

complex one, so this working assumption rests on more assumptions, like any other.

c) It was breakfast time. Bob's mum asked, 'What do you want for breakfast?' He said, 'I'll get the milk.'

→ I want cereal (not toast)

Given that the available options are toast or cereal, and that people typically have milk with cereal, this is a relevant answer to the question:

Relevance

d) The man made the door close.

→ in an unusual way

Otherwise, he would have used the briefer expression 'closed the door':

Manner

In this thesis, I adopt Grice's original formulation, and work on the basis of three distinct types of implicature: quantity, relevance and manner. Though neo- and post-Gricean theorists have streamlined his maxims, only Relevance Theory has reduced them to a single principle (Sperber & Wilson, 1995; D. Wilson & Sperber, 2012). Others have maintained a distinction between informativeness and relevance (Horn, 2004), and I will argue in Chapter 4 that there are good reasons to keep manner distinct, too (cf. Levinson, 2000). The distinction between quantity and relevance has also been adopted as uncontroversial in the experimental and developmental literature (e.g. Noveck, 2001; Schulze, Grassmann, & Tomasello, 2013; Stiller, Goodman, & Frank, 2015).

I also gravitate towards a constraint-based version of the Gricean approach, which supposes that a whole number of considerations will contribute to any inference, and that these may be more or less important in different contexts (Degen & Tanenhaus, 2014). In other words, cues to the speaker's intended meaning are given more or less weight, and this will determine the inference. This account supplements Grice's maxims at the levels of 'what' and 'why': 'Under the Constraint-Based account, then, the research program becomes one of identifying the cues that listeners use in service of the broader goal of understanding the representations and processes that underlie generation of implied meanings' (Degen & Tanenhaus, 2014: 672).

Besides these context-driven accounts, there are other approaches to quantity implicatures in particular. Grammatical approaches to scalar – and, more recently, ad

hoc – implicatures involve a covert exhaustifier operator, akin to ‘only’, and alternatives that are provided by the grammar (Chierchia, Crain, Guasti, Gualmini, & Meroni, 2001; Chierchia, Fox, & Spector, 2008). On this view, inserting such an operator or not is a process of disambiguation, in parsing a sentence. Chemla and Singh (2014a, 2014b) point out, though, that the grammatical approach is not only a matter of grammar, just as the context-driven approach is not just a matter of pragmatics; both involve both grammar and pragmatics, but distribute the work differently. This also makes testing the different approaches empirically more challenging. In this thesis, I occasionally refer to this alternative view, but largely work within a context-driven, Gricean approach to implicatures.

1.1.3 Inferences and alternatives

A neo-Gricean model says something not just about motivations for deriving implicatures, but about the nature of the derivation itself. Two characteristics are key: the role of elaborative inference and of alternatives.

A typical relevance inference, such as that triggered by ‘I’ll get the milk’, in example (c) above, could be described in the following steps:

[in the context of choosing breakfast, with toast or cereal as options]

1. The speaker said, ‘I’ll get the milk’.
2. This utterance does not directly address the question.
3. It can be assumed that the speaker intends his utterance to be relevant.
4. His utterance is relevant to the above exchange to the extent that its content relates to the question.
5. The content of his utterance is related to the content of the question by virtue of the fact that the milk he mentions is useful and typically necessary for one of the breakfast options, namely cereal.
6. He means that he would like cereal for breakfast.

(based on Cummings, 2005: 102)

Note, firstly, the kind of reasoning that this is – non-deductive reasoning (Geurts, 2010), where the conclusion is compatible with the evidence so far, but further

evidence may cancel it. Specifically, following Cummings (2005), it has elements of presumptive and elaborative reasoning. Presumptive reasoning depends on arguments from ignorance: the hearer assumes that the speaker is being informative (step 3) and is fully informed (step 6), reasonably based on lack of evidence to the contrary.

Implicatures can thus be cancellable, should evidence to the contrary appear.

Elaborative inferences a) form a cohesive link b) through world knowledge, c) based on what is typically the case. Here, elaborative reasoning is evident in step (5) where, by a cohesive link, the hearer uses world knowledge to establish the relevance of the speaker's utterance.

Cummings argues that it seems clear 'that some form of elaborative inferencing is integral to the recovery of implicatures' (2005: 104), by facilitating the use of communicative knowledge (e.g. of the Co-operative Principle) and world knowledge in utterance interpretation. It seems that they are not just limited to relevance inferences. Take this description of a scalar implicature, for the example 'I broke some of the plates' (cf. Breheny, Ferguson, & Katsos, 2013):

1. The speaker said 'some'.
2. In this situation, propositions with *some* or the stronger and relevant alternative *all* could be true.
3. Assuming that the speaker is being informative (and is expecting me to expect so) ...
4. If he had meant *all*, he would have said 'all' (otherwise he would not be informative).
5. He did not say 'all', therefore he does not know whether he broke all of the plates.
6. Assuming the speaker does know whether he broke all the plates...
7. He means that *he broke some but not all of the plates*⁴.
8. And intends me to reason in this way.

⁴ The assumption in (6) contradicts the conclusion in (5), and therefore by contradiction the assumption in (4), namely, that is the hearer now knows that the speaker did not mean *all*; therefore, the hearer can conclude (7).

Here the elaborative inference is seen in step (2), where the alternative is brought in to the inference, which a) forms a cohesive link between what is said and the discourse or situation context (as shall be seen more clearly below for the Question Under Discussion), b) is accessed as part of linguistic knowledge, and c) is a stronger utterance by linguistic convention. I take it, therefore, that across types of implicature, some kind of elaborative inference, related to the relevance of the utterance, is involved, and is part of the acquisitional challenge for children (see, too, for discussion, Skordos, 2014).

Note, secondly, that an important step in the chain of reasoning for a quantity implicature – and, as shall be seen in Chapter 5, for manner as well – is thinking about alternatives that the speaker could have said but did not. In this case, the alternative is ‘I broke all the plates’ (step 2); for the example, ‘I took a strawberry’, alternatives could be the contextually available ‘strawberry and orange, strawberry and orange and kiwi’, and so on. In both cases, the alternatives are not anything imaginable, but are *relevant* to the discourse and its context.

1.1.4 The Question Under Discussion

But what do I mean when I say that alternatives are *relevant*? Geurts (2010: 47) suggested that ‘the differences between ϕ and ψ [what is said and its alternative] may be (i) relevant to the purposes of the discourse, (ii) of potential interest to the hearer, or (iii) of general interest’, being somewhat reluctant to commit to a more precise definition. Cummings (2005), though, clearly concentrates on the first option of the discourse, and I follow suit here, in line with other work that increasingly recognises the importance of the Question Under Discussion (Roberts, 2012; Van Kuppevelt, 1996). This notion can be traced back to Grice (1975) and has more recently been employed by pragmaticians with an experimental approach (for an overview see Benz & Jasinskaja, 2017; Cummins, 2017); here I adopt the general notion, without committing to a specific theoretical or processing instantiation.

The Question Under Discussion (QUD) does not have to be an explicit question, as in the examples considered so far; it is the topic of discussion, or the subgoal for that discourse that the interlocutors have mutually agreed on to pursue. Relevance can thus be formalised as the degree to which the utterance addresses the QUD. For a

relevance implicature, the elaborative inference derives a speaker meaning that does address the QUD, unlike the literal content. For quantity, the available alternatives – and whether an implicature is derived at all – are constrained by the QUD, by whether they address it more directly than what is literally said. QUDs ‘activate alternative sets, and indicate what kind of information is contextually useful’ (Benz & Jasinskaja, 2017: 182). Empirical evidence has been amassed showing that potential quantity implicatures are not derived if they are not relevant, that is, if they do not address the QUD more fully than the literal content (Breheny, Katsos, & Williams, 2006; Politzer-Ahles & Fiorentino, 2013; A. J. Zondervan, 2010; A. Zondervan, Meroni, & Gualmini, 2008). To illustrate, consider now the ad hoc quantity example in different contexts, with different explicitly stated QUDs, (a–c):

- a. ‘What did you take from the fridge?’
- b. ‘Did you get fruit from the fridge?’
- c. ‘Anyone with some fruit can also have a chocolate bar.’
‘I took a strawberry.’

In cases (b) and (c) the QUD is addressed by the literal content (*I took at least a strawberry*), so no exhaustive inference (*I took only a strawberry*) need be drawn.

This theoretical diversion has consequences for looking at acquisition: on this approach to implicatures, relevance is crucial not just for relevance implicatures, but for quantity as well. Computational models like this ‘represent the task faced by the child rather than as the child solves them’ (Frank, 2014: 211). Hence, the model, in which relevance is an integral part of implicature derivation, is a description of the competence the child is learning. Indeed, as I shall discuss in the following chapter, tracking the QUD and generating relevant alternatives seem to be two factors that play an important role in the course of children’s development of implicatures.

1.2 Approach to language acquisition

This thesis has a narrow focus on implicatures in development, but it is situated in the wider framework of a usage-based approach to language acquisition (Scott-Phillips, 2014; Tomasello, 2003). According to this view, children’s primary goal is not to learn abstract linguistic knowledge per se but to communicate with others, and

their innate tendency to prosociality is a key contributor (Ambridge & Lieven, 2011).

On one such account, Tomasello (2003) identifies three skills that open up the way to communication, emerging from 9–12 months onwards: joint attention, intention reading, and cultural learning or role-reversal imitation (see too Stephens & Matthews, 2014). Firstly, in learning to participate in a joint attentional frame, children move from dyadic to triadic interaction, engaging with another person and an object or event by being aware that both they and their interlocutor are attending to the same thing, and know this. This is supported with gaze-tracking. Secondly, children learn to understand that others have intentions, and, in particular, communicative intentions. For instance, from 9 months, infants react differently depending on whether someone is unwilling or is unable to perform an action (e.g. Behne, Carpenter, Call, & Tomasello, 2005), and later children grasp communicative intention, recognising that in communicating a speaker wishes to direct the hearer's attention or change their beliefs (e.g. Grosse, Behne, Carpenter, & Tomasello, 2010). Thirdly, in role-reversal imitation, children 'must learn to use a symbol toward the adult in the same way the adult used it toward [them]', in order to produce intentional utterances themselves (Tomasello, 2003: 27).

This, then, is in many ways a 'pragmatics first' view of language acquisition. Indeed, Tomasello comments: 'It is interesting to note that the intersubjectivity inherent in socially shared symbols, but not in one-way signals, sets up all kinds of pragmatic 'implicatures' of the type investigated by Grice (1975) concerning expectations that other persons will use the conventional means of expression – that we both know they know – and not others that are more cumbersome or indirect' (2003: 27). An important consequence of this view is that there is no sharp distinction between non-verbal and verbal communication in ontogenetic development. Children's pointing may express intentional meaning in the same way as early utterances, for instance (Tomasello, Carpenter, & Liszkowski, 2007). So, although in this thesis implicatures are viewed as strictly verbal, they are not necessarily the start of children's pragmatic inferencing abilities, as shall be discussed when it comes to word learning by exclusion. Another consequence is that there is no sharp distinction between pragmatic development and lexical or grammatical development, which will be pertinent when looking at the connection between implicature skills, and lexical

and grammatical knowledge. Finally, it ties children's language development closely to their general cognitive development, and especially social cognition, like Theory of Mind, which will also be a topic of investigation.

This thesis is also focussed on children's *understanding* of implicatures. This follows the major trend in recent research, though there are a couple of studies on children's production of scalar implicatures (Eiteljörge, Pouscoulous, & Lieven, 2016; Tieu, 2017) and a much larger body of work on children's referential informativeness (e.g. Davies & Katsos, 2010). Though there are clearly some common prerequisites and skills involved, understanding and production of implicatures are likely to have different developmental paths (Hendriks & Koster, 2010), and understanding will be the sole focus here.

1.3 Overview of the thesis

This thesis investigates children's development of implicature understanding, with three strands running through: the relationship of different types of implicature in development; the contribution of different factors, especially structural language abilities; and the role of Theory of Mind. It is organised into four main chapters (2 to 5), each of which presents a review of the relevant literature for that chapter, together with an empirical study. Although this renders each chapter self-contained to some degree, I assume a basis of preceding chapters, and build on findings and arguments from one chapter to the next.

In Chapter 2, I present a study that examines scalar, ad hoc, relevance and word learning by exclusion inferences, to address the questions:

1. What is the relationship between quantity, relevance and word learning by exclusion in development?
2. What is the nature of word learning by exclusion: pragmatic or lexical?

With its picture-matching study with 3–5-year-olds, this chapter lays the groundwork, in many ways, for the following ones, in being concerned with children's abilities to derive implicatures in a simple, supportive context, and in introducing the methodology used as a basis throughout. The findings suggest early competence with gradual development of implicature understanding over the preschool years, with word learning by exclusion preceding ad hoc and relevance inferences, and scalars

proving the most challenging for children.

Chapter 3 explores additional data collected in the same study as in Chapter 2. It considers the relationship of pragmatic abilities with cognitive factors like Theory of Mind development and structural language competence on the one hand, and life experience factors like socioeconomic status and speaking either one or more than one language, on the other. The main predictor of implicature skills turns out to be structural language abilities.

In Chapter 4, I turn to another type of implicature, manner, on which there is little existing theoretical or empirical work. I therefore ask:

1. Are manner implicatures a distinct category, and is there evidence for them in adults' communication?
2. What are the predictions for their acquisition?
3. Do manner implicatures develop early or late in children's communication?

I lay out some theory behind a working definition of manner implicatures, before spelling out the implications for their acquisition, and presenting a first attempt at investigating children's competence with them. There is some evidence that both adults and children sometimes derive manner inferences, though teasing them apart from quantity inferences remains a challenge.

In Chapter 5, I address the question of the role of Theory of Mind in implicature derivation more directly, by examining children's ability to integrate the speaker's epistemic state into utterance interpretation. My research questions are:

1. Do children engage in perspective-taking in implicature derivation?
2. If not, what does this mean for a Gricean model of pragmatics, and implicature in particular?

I test whether children, like adults, are able to *not* derive an implicature when the speaker is ignorant, through two experiments – a covered-box task, and a referential director task. The results suggest a two-step developmental trajectory: first, children learn to derive pragmatic inferences, and then to integrate the speaker's epistemic state into these inferences, particularly when it differs from theirs.

Finally, in Chapter 6, I offer some conclusions from all four experiments, as well as an outlook to future research and applications of the findings. Deriving implicatures seems to develop as a skill from age 3, with early competence in straightforward

communicative situations. However, there are developmental differences in implicature types, and ongoing development in more complex situations where information from the context must be taken into account.

2 Quantity, relevance and word learning by exclusion inferences – pragmatics in the early years



“Show me the blicket”

Figure 2.1 Word learning by exclusion

Learning words – mapping forms to meanings – is one of the great challenges for children learning language. In a simple scenario, as in Figure 2.1, children may hear a word for the first time, and need to work out its referent. In this case, they are familiar with the label for one of the possible referents, so, from their point of view, the new label could be a second label for the familiar object, or a label for the new object. The latter is children’s preferred strategy – known as word learning by exclusion. In this chapter, I present a study that investigates relatively early competence of quantity, relevance and word learning by exclusion inferences in an ideal context, with the aim of addressing two questions:

1. What is the relationship between quantity, relevance and word learning by exclusion in development?
2. What is the nature of word learning by exclusion: pragmatic or lexical?

In the following section, I ground these research questions in pragmatic theory by drawing out some possible predictions for children’s development, before reviewing the literature on implicature development, and the role of the QUD in children’s implicature comprehension. Then, I turn to word learning by exclusion (henceforth WLE), presenting conflicting accounts and empirical evidence to date, and explaining why it makes sense to examine it along with implicatures.

2.1 Pragmatic theory and development

In Chapter 1, I outlined a Gricean view of pragmatics, in which relevance, quantity and manner implicatures are distinct but share some important commonalities. In particular, relevance, defined as the degree to which the QUD is addressed, is important not only for relevance implicatures, but for quantity and manner implicatures as well (Cummins, 2017; Degen & Tanenhaus, 2014; Roberts, 2012). For one thing, a quantity implicature is derived, or not, depending on the context of utterance, and in particular, on whether stronger alternatives that could have been said are relevant. For another, the kind of elaborative inferencing involved in relevance inferences may also be involved in generating relevant alternatives in quantity inferences. Additionally, quantity implicatures involve negation of alternatives, a feature shared by WLE inferences – as shall be seen below. This means that relevance and quantity inferences involve partially overlapping skills, as do WLE and quantity. Manner inferences fit into this schema, too, with much in common with quantity but with an additional component of utterance *form*, and I consider these on their own in Chapter 4.

	WLE	Quantity	Relevance
Tracking QUD		✓	✓
QUD licenses inference ¹		✓	
Elaborative inference		✓	✓
Negation of alternative	✓	✓	

Table 2.1 WLE, Quantity and Relevance inferences

¹ For quantity, this could be because tracking the QUD leads to generation of relevant alternatives, or it could be that the QUD constrains selection of alternatives. In contrast, for relevance inferences the QUD plays a different role: the apparent gap between the QUD and the utterance triggers a search for an explanatory link.

This schema has implications for children’s pragmatic development, as the learning challenge presented is different in each case. Most basically, quantity implicatures might present more of a challenge than either WLE or relevance, whose features they incorporate. Further, in a typical WLE paradigm, it is arguably not essential to refer to the QUD as the utterance with the novel word is such a strong cue to infer its referent in the context, reasoning by exclusion – although the QUD can of course sometimes help in determining the referent of a novel word (e.g. Tomasello,

Strosberg, & Akhtar, 1996). Different developmental trajectories can therefore be expected for different types of inference, with WLE and relevance emerging before quantity. This prediction does leave open, though, the possibility that there are many factors beyond Grice's maxims that make particular instances of implicatures or subtypes more or less challenging for children (Verbuk & Shultz, 2010). To date, however, no study has brought all types together in one paradigm and a single sample of children, to answer the question: What is the relationship between quantity, relevance and WLE in development?

2.2 Learning to understand implicatures

In this section, I review what is known about children's development of implicatures, through studies that have investigated each in isolation, and, importantly, through those that have looked at more than one type. Then, I discuss the evidence for the crucial contribution of relevance to that development. I highlight the suggestive evidence for an order of acquisition reflecting theoretical predictions already in place, and the mounting evidence for the difficulty children have in tracking and integrating QUD and generating relevant alternatives for quantity implicatures. I also discuss the need for child-friendly methodology that directly taps their implicature-derivation skills.

2.2.1 Relevance

The study of children's understanding of relevance inferences has a relatively long history thanks to studies of a specific type of inference, the indirect request (e.g. Bernicot & Legros, 1987; Elrod, 1987). Many early studies, and even some later ones, suggest that children learn to make relevance inferences relatively late (e.g. aged 8 years and above in de Villiers, de Villiers, Coles-White, & Carpenter, 2009). One probable reason for this finding is the nature of the tasks, which require metalinguistic reasoning, such as explaining why a character said something or what was meant (e.g. Bucciarelli, Colle, & Bara, 2003; de Villiers et al., 2009; Verbuk & Shultz, 2010). That is, they are investigating not only children's comprehension, but their understanding of their comprehension, something that develops later in childhood (Lieven, 2006).

More recent studies have addressed this by trying to directly measure children's

comprehension – their derivation of relevance implicatures – within other experimental paradigms, like action-based or picture-matching tasks (S. Lewis, 2013; Schulze et al., 2013; Tribushinina, 2012). The evidence suggests that children start to derive relevance implicatures at an early age, around 3 years or even younger. This is particularly the case when the world knowledge required to make the elaborative inference, connecting the utterance with its intended meaning, is intuitively closely associated with both.

For instance, Schulze, Grassmann and Tomasello (2013) and Tribushinina (2012) find that German- and Dutch-speaking children as young as 3 years can make inferences where negative emotion is associated with dispreference, or where positive emotion is associated with preference or desire for something:

A: Should [child] give you the elephant?

B: I like elephants. / I don't like elephants.

→ The speaker wants that item / a different item.

(Schulze et al., 2013: 2082)

This is corroborated by Tribushinina (2012), who tests children aged 2;6–3;6 in a game based around shopping and the customer's preferences, and also finds competence in understanding indirect utterances with this age group. In Schulze, Grassmann and Tomasello's study, young children can also make intuitively more complex inferences such as:

A: Do you want the cereal or the roll? [for breakfast]

B: The milk is gone.

→ The speaker wants the roll. (Schulze et al., 2013: 2087)

Here, the relevance of the utterance is understood via the world knowledge that cereal is typically enjoyed with milk for breakfast, and so the precondition for that option is not fulfilled. Children aged 3 (2;10–3;1) find the inverse case, when the precondition is fulfilled, much more challenging, deriving an implicature at chance levels (at a rate of 58%) – compared to above chance for the unfulfilled precondition

inference (at 73%). 4-year-olds (3;10–4;1), in contrast, succeed with this kind of inference:

A: Do you want the cereal or the roll? [for breakfast]

B: I bought milk.

→ I want cereal. (Schulze et al., 2013: 2087)

The authors suggest that the puzzling difference for three-year-olds has to do with the constraints that the utterance places on their action in response: in the precondition unfulfilled case, one of the possible objects is excluded, forcing the selection of the other object. Possibly, the negative utterance may have highlighted the need to access world knowledge to infer relevance. They conclude, therefore, that what children are doing in both cases is a relevance inference; it is just that the precondition fulfilled case includes some more difficult steps.

However, given that in both cases the utterance is, on its literal meaning, irrelevant to the explicit QUD, one wonders why the search for world knowledge is not triggered equally in both cases. Another possibility is that only in the precondition fulfilled case is complex reasoning about the speaker's intended meaning required. When the precondition is not fulfilled, an inference could be made purely through reasoning about world knowledge without reference to speaker intention – along the lines of 'when there is no milk, one cannot have cereal, therefore the speaker cannot have cereal'; this is called a 'material implicature' by Jary (2013; see too Kissine, 2016 and discussion in Chapters 3 and 5). It could even be based on egocentric but benevolent reasoning on behalf of the speaker, such as 'I would not want cereal without milk, so I would better give him the roll'. Neither of these options are available, however, in the precondition fulfilled case; in that case, the hearer has to consider why the speaker has chosen to mention the milk purchase, presumably because he finds it relevant to the discourse (a 'behavioural implicature', on Jary's (2013) typology).

Entertaining this kind of explanation obviously rests on the notion that not all implicatures are created equal – not all are 'fully Gricean' in the sense of requiring reasoning about a speaker's epistemic state and communicative intentions. I will return to this idea in the coming chapters.

2.2.2 Ad hoc quantity

Ad hoc quantity implicatures involve a set of alternatives that are context-specific. In the simplest case, this could be the alternatives <object X, object X and Y>, where an utterance like ‘give me X’ would give rise to an *exhaustive* inference, that the speaker means *only* X. Young children have also been found to be competent with this kind of ad hoc inference. Three-year-olds in Stiller Goodman and Frank’s (2015) study are able to choose appropriately, for example, which smiley face belongs to the puppet when he says ‘my friend has glasses’, selecting the face with only glasses, not glasses and a hat. Children in the 3;6–4-year age group are above chance. Similar results come from other picture-matching tasks (Grosse, Schulze, Noveck, Tomasello, & Katsos, in prep.; Horowitz & Frank, 2015) and eye-tracking studies (Yoon, Wu, & Frank, 2015). Again, these all involve child-friendly experimental designs where the alternatives are presented visually in context, and children are only asked to choose a picture, not to explain their choice. Importantly, they also provide a context where the implicated meaning is relevant, through the combination of the picture-matching task – that is, reference resolution – and the visual display, where the utterance ‘X’, taken to mean ‘at least X’, is not a unique identifier.

Even using such experimental paradigms, though, current findings suggest that these inferences are not yet available to two-year-olds: Stiller, Goodman and Frank (2015) find that they are performing at chance, even in a simplified version of the experiment designed specifically for toddlers, while Yoon, Yu and Frank (2015) find they even prefer the distractor alternative in an eye-tracking task. I return to this apparent ‘gap’ between early pragmatic competence and the development of ad hoc implicatures in the discussion of my own results below.

2.2.3 Scalar quantity

Scalar quantity implicatures are beyond doubt the most studied – and best understood – type of implicature in children, reflecting in part the theoretical debates and adult processing studies centred on them. Most studies concentrate on the scale <all, some>, but a few look at other scales such as <and, or> and <finish, start>.

Those that use Truth Value Judgement Tasks or Felicity Judgement Tasks, which require the participant to express a decision on whether a speaker has described a

situation (to which the participant also has access) correctly or well, find relatively late acquisition, at 5 years or much older (e.g. Foppolo, Guasti, & Chierchia, 2012; Gotzner, Barner, & Crain, 2015; Guasti et al., 2005; Noveck, 2001; Papafragou & Tantalou, 2004; Pouscoulous, Noveck, Politzer, & Bastide, 2007; Scafton & Feeney, 2006; Skordos & Papafragou, 2016; Verbuk, 2012). In a typical case, a speaker describes a picture where a characteristic is true of all items (e.g. all the boys have umbrellas), using the quantifier ‘some’: it is semantically true (the logical meaning of ‘some’ is assumed to be *some and possibly all*), but under-informative. Adults typically reject such utterances, whereas young children are more accepting. However, these kinds of tasks may not be tapping into children’s actual pragmatic competence. Not only are judgements dependent on metalinguistic skills, but they are also potentially measuring only sensitivity to under-informativeness, rather than implicature derivation per se, because only sensitivity to informativeness is needed to reject an under-informative statement (Katsos & Bishop, 2011). Furthermore, the speaker in such tasks is, by necessity, fluctuating in his or her level of cooperativeness, being fully informative in control trials, but under-informative in critical ones. Not much is known about how children respond to varying levels of informativeness or how successfully they track it, though they certainly are sensitive to it (e.g. Chierchia et al., 2001; Pogue, Tanenhaus, & Kurumada, 2017; Siegal, Iozzi, & Surian, 2009). At best, this might create a difference between children and adults’ responses to a varyingly co-operative speaker; at worst, it presents a confound that obscures children’s pragmatic competence. One formulation of such an issue is Pragmatic Tolerance – the notion that children may be more reluctant than adults to say that a speaker is wrong or has said something silly, if the utterance can be accommodated as semantically true (Katsos & Bishop, 2011; Katsos & Smith, 2010).

In contrast, studies that use action-based or picture-matching tasks, like the ad hoc implicature tasks above, where the speaker is behaving in a pragmatically co-operative way across trials, find some emerging competence even at age 4 (Horowitz & Frank, 2015; Pouscoulous et al., 2007), particularly where the quantifier carries focus stress (Miller, Schmitt, Chang, & Munn, 2005)⁵. As the goal of the present study is to

⁵ In addition, a couple of studies have examined whole–part scales, that depend to some extent

compare types of inference in development, a picture-matching task is preferred not only to avoid the potential confounds of judgement tasks, but also to make comparison across inference types more viable (so that cooperativeness is not violated in a different way for different conditions).

Scalar implicatures have been the subject of contrasting theories, with implications for development: on a lexical-scales approach, which is sometimes linked to a grammaticalist theory of scalars, they are challenging for children because there is specific lexical knowledge of scales that children have to acquire, in addition to the terms themselves (e.g. Barner, Brooks, & Bale, 2011). On a contextualist or constraint-based account, scalars are not inherently different from other quantity implicatures, and so there is no extra challenge of learning a lexical scale (although acquiring terms like quantifiers in itself may be challenging). The lexical-scales approach could lead to the prediction that scalars are harder than ad hocs, because of the extra step in acquisition. However, it is in principle possible to make the reverse prediction – that because they ‘only’ require lexical rather than world knowledge, they could be easier – a hypothesis tested by Verbuk (2012). In other words, development is not a simple testing ground for competing theories on scalars, but different theories may be able to better or worse explain developmental patterns.

2.2.4 Comparing implicatures

The diversity in tasks, age groups and languages makes it difficult to compare different types of implicature examined in different studies, or at least making any comparison can only be suggestive. A few studies, however, have combined two kinds of implicature in the same task or experimental session. Katsos and Smith (2010), Katsos and Bishop (2011), Grosse and colleagues (in prep.), and Horowitz and Frank (2015) all consider ad hoc and scalar quantity implicatures together. The overall findings are consistent: ad hoc implicatures are available before scalars in

on contextual relations but to a lesser degree than ad hocs (e.g. I cleaned the kitchen → The speaker didn’t clean the whole house). These suggest a similar development to scalar implicatures, such as those with ‘some’, but are too under-powered in terms of stimuli number or sample size to draw firm conclusions (Papafragou & Tantalou, 2004; Verbuk & Shultz, 2010).

development (where the scale is <all, some>). These studies use a variety of methods: a judgement task with 7-year-olds (Katsos & Smith, 2010); judgement and picture-matching tasks with 5–6-year-olds (Katsos & Bishop, 2011); a picture-matching task with 3- and 5-year-olds (Grosse et al., in prep.); and a picture-matching task with 4- and 5-year-olds (Horowitz & Frank, 2015).

What is of interest, too, is the relationship between relevance and quantity inferences. Only one study, to my knowledge, combines these. Verbuk and Schulz (2010) find no evidence for a difference in children’s performance between quantity inferences – in this case using part–whole scales – and relevance inferences. However, there are a number of issues with their design: the children’s ages range from 5;1 to 8;1 but form one group for analysis; the children are asked what happened and why, and only score as correct if their explanation indicates that they have drawn an implicature (i.e., a heavily metapragmatic task); there is, in addition, a ‘non-verbal’ condition in their design, where children have to draw an inference from a picture rather than utterance, which could affect expectations about the speaker and the task; and both the verbal and non-verbal relevance inferences could arguably be derived without reference to speaker intentions (i.e., material implicature type). So, to date, we have little idea of how relevance and quantity develop together.

2.2.5 The QUD and relevant alternatives

A number of observations on children’s development of quantity implicatures converge to suggest that a key step is learning to track the QUD and generate relevant alternatives, as Papafragou and Skordos (2016) propose.

Firstly, those studies that use Truth Value Judgement Tasks or, to a lesser degree, Felicity Judgement Tasks show lower rates of sensitivity to informativeness in children (and also in adults) than other paradigms, as already mentioned. A further explanation for this discrepancy could concern the shifting QUD (as suggested by, for instance Skordos & Papafragou, 2016). To give a concrete example, consider the first major study on children’s abilities with scalar implicatures: Noveck (2001) asks children to judge six different types of sentence: critical true but under-informative *some* (e.g. ‘some giraffes have long necks’); true and optimally informative *some* (e.g. ‘some birds live in cages’); true *all* (e.g. ‘all elephants have trunks’); false *all* (e.g. ‘all dogs have

spots’); ‘absurd’ *all* (e.g. ‘all chairs tell time’); and ‘absurd’ *some* (e.g. ‘some stores are made of bubbles’). All but the true *some* conditions arguably imply a QUD of quality, not quantity. If children have difficulty tracking the QUD, and stick with one of quality, then they will be more likely to accept true but under-informative statements with ‘some’.

In addition, other studies that include some kind of training component or an element that highlights quantity also find improved performance, as compared with equivalent tasks that do not, which again could be explained in terms of the implicit QUD (Foppolo et al., 2012; Guasti et al., 2005; Papafragou & Musolino, 2003). Interestingly, Guasti and colleagues (2005) repeat Papafragou and Musolino’s training study with a Truth Value Judgement Task, and observe a bimodal distribution in children’s performance: training influences some children’s responses so that they now reject under-informative utterances, but does not influence others’ at all. One possible explanation is that the former group of children are sensitive to informativeness (or able to derive an implicature, depending on the interpretation of this task), but that without the training phase that highlights informativeness they are unaware that it is relevant. In contrast, those studies that present the alternatives, visually or in discourse, render them relevant, and children fare better in these.

Secondly, Barner, Brooks and Bale (2011) set out to test whether children’s particular difficulty with scalar implicatures is a result of difficulty with learning the quantifiers themselves or learning their relationship on a scale. They use a Truth Value Judgement Task that crosses scalar and ad hoc implicature-triggering utterances, with explicitly exhaustive utterances containing ‘only’. A critical difference between their study and others, though, is their use of questions rather than statements, such as ‘are some of the animals sleeping?’. Questions are typically not implicature-triggering contexts, which means that their design cannot test children’s derivation of implicatures. However, the combination with the explicitly exhaustive questions – ‘are only some of the animals sleeping?’ when in fact all are – can identify the problem. Children are able to correctly answer with ‘only’ (in the negative) and without (positively) in the ad hoc case, but not the scalar case. This suggests that they a) know that questions do not typically trigger implicatures, b) know that ‘only’ is an exhaustifier, c) can apply this meaning when alternatives are given in context, d) know

the logical meaning of *some* and *all* (from control trials), but e) cannot apply an exhaustive meaning when the alternatives are scale mates <all, some>.

The authors interpret this as evidence that, for scalar implicatures, children not only have to learn the lexical items themselves, but also the scale between them, in order to derive implicatures. Alternatively, as Papafragou and Skordos (2016) suggest, it could be explained by the challenge of generating relevant alternatives: children are challenged in generating relevant alternatives for a scalar term like *some*, both where this is required by the semantics of *only*, and where it is pragmatically appropriate. Whereas, for ad hoc scales, the naming of two animals depicted ('are the cat and the dog sleeping?') draws more attention to the contrasting alternative ('the cow'). This would mean that the negated alternative could be easier to generate for ad hocs, hence children's adult-like performance with 'only'.

Thirdly, in his thesis, Skordos (2014; Skordos & Papafragou, 2016) sets out to investigate the generation of relevant alternatives directly. Studies like Grosse and colleagues' (Grosse et al., in prep.) find a facilitatory effect of hearing the stronger alternative before the critical quantity inference trial for scalars (and not for ad hocs in their case, as they were approaching ceiling). It could be that the mention of the stronger alternative simply enables it to be generated in the critical trial; the developmental challenge is just learning to access relevant alternatives. Alternatively, it could be that mention of a stronger alternative highlights it as a *relevant* stronger alternative, with a QUD that relates to quantity. Skordos (2014) attempts to tease apart these two options: 5-year-olds provide acceptability judgements on critical *some* sentences in one of two conditions. In the first, they had previously judged sentences with 'all' that were false because of quantity (*some but not all*), i.e., there is a consistent implicit QUD of quantity from *all* trials to *some* trials, but, crucially, still a preceding mention of the stronger alternative. In the second, they had previously judged sentences with 'all' that were false because of quality (*all have Xs not Ys*), i.e., there is a shift in implicit QUD from quality to quantity from *all* trials to *some* trials. The results are enlightening: only in the first condition are children adult-like in their rejection of under-informative *some* sentences. The same is true in another experiment, where the preceding trials contain 'none' and a consistent implicit QUD of quantity. This clearly suggests that children's difficulty with quantity implicatures is

– at least sometimes – down to tracking the QUD and generating relevant alternatives.

Clearly, even stronger evidence for this view would come from a study that combines a direct measure of implicature derivation like that of Grosse et al.'s (in prep.) or Horowitz and Frank (2015) with a manipulation of QUD. However, taken altogether, the studies reviewed here strongly suggest that children's development of quantity implicatures is connected to their development of sensitivity to and inferencing of relevance (where relevance is understood as addressing the QUD). Methodologically, this means that a study that includes mixed conditions across trials should also include a context and, ideally, explicit QUD, in order to avoid obscuring children's pragmatic competence. Empirically, it suggests that a relationship between relevance and quantity may be expected in acquisition – as outlined above – and that an investigation of them would prove fruitful.

2.3 Word learning by exclusion

I now turn to examine word learning by exclusion in more detail, to show how it may fit into the developmental picture of the first research question – concerning the relationship of inferences – and to motivate the second research question – concerning the nature of WLE.

WLE is a much-studied and robust phenomenon. From the first year of life, children tend to choose a novel object as the referent for a novel word (Graham, Poulin-Dubois, & Baker, 1998; Halberda, 2003; Markman, Wasow, & Hansen, 2003), and this tendency strengthens with age (Bion, Borovsky, & Fernald, 2013; Diesendruck & Markson, 2001; Hollich et al., 2000; Merriman, Bowman, & MacWhinney, 1989; Merriman & Schuster, 1991; Scofield & Behrend, 2007). This is typically tested in one of two ways: either a child is presented with one novel and one familiar object, and asked to point to, pick up or look at 'the blicket', or one of two novel objects is labelled (e.g. 'here is a wug'), and then the child is asked for 'the blicket'. Note, though, that the second methodology may be drawing on the blocking effect of being in common ground having been the focus of joint attention (Grassmann, Stracke, & Tomasello, 2009), rather than in common ground given knowledge of conventional labels. Note, too, that, while 'word learning by exclusion' is a commonly used term for this behaviour, what is really at issue is mapping via disambiguation, not learning per se,

though some studies also look at retention of the mapping (e.g. Bion et al., 2013; Horst, Scott, & Pollard, 2010; Spiegel & Halberda, 2011; Zosh, Brinster, & Halberda, 2013). As what I am interested in here is the inferencing process itself – and how it relates to pragmatic inferences like implicatures – looking only at mapping is sufficient.

2.3.1 Theories of WLE

There are two main theories that have been proposed to account for this behaviour, in some ways paralleling the two approaches to scalar implicatures. Firstly, children may have a lexical-specific bias of mutual exclusivity – a heuristic that each (type of) object has only one label (Brosseau-Liard & Hall, 2011; de Marchena, Eigsti, Worek, Ono, & Snedeker, 2011; Haryu & Imai, 1999; Jaswal, 2010; Markman, 1994; Markman & Wachtel, 1988; Markman et al., 2003; Scofield & Behrend, 2007). Using this heuristic, children can reason that if one object already has a known label, the novel label cannot refer to it, so it must refer to another one. This kind of heuristic therefore implies reasoning by exclusion, but no reference to the speaker's intentions.

Secondly, children may be drawing on pragmatic principles, such as those originally formulated by Clark – the Principles of Conventionality and Contrast:

Conventionality: for certain meanings, speakers assume that there is a conventional form that should be used in the language community

Contrast: speakers assume that any difference in form signals a difference in meaning

(E. V. Clark, 2009: 133. See too E. V. Clark, 1988, 1990, 2002, 2004, 2007)

Based on these principles, children can reason that if a speaker was intending to refer to the known object, she would have used the known label (Conventionality); as she did not use that label, but the novel one, she must not intend the known object (Contrast); instead, she must intend to refer to the novel object. Note that here, speaker communicative intentions are key, as argued too by Bloom (2002) and

Tomasello (2000)⁶.

Clark herself points to the similarity between her principles and Grice's Cooperative Principle (E. V. Clark, 1990; E. V. Clark & Clark, 1979), but does not flesh out this connection. The similarities should be clear by now, and have been mentioned, at least in passing, by many developmental pragmaticians (Bale & Barner, 2013; Barner et al., 2011; Katsos & Bishop, 2011; Katsos & Wilson, 2014; Morisseau, Davies, & Matthews, 2013; Stiller et al., 2015; Sullivan & Barner, 2011). Both WLE and implicature inferences require foundational pragmatic abilities like reference to a speaker's intentions and expectations of cooperativeness, but in addition, in the case of quantity and manner, they involve reasoning by exclusion. Indeed, some have suggested that WLE is more comparable to a manner inference than quantity (Brosseau-Liard & Hall, 2011; de Marchena et al., 2011), because of the key role played by the *form* of the utterance (as opposed to the *content*, in the case of quantity). If Grice's manner maxim, 'avoid obscurity', is interpreted as 'be conventional', then WLE is an application of manner to the disambiguation problem: a speaker avoids obscurity by using terms that are conventionally accepted in a linguistic community. However, as shall be seen in Chapter 4, depending how manner implicatures are formally defined, there may still be substantial differences between WLE and manner.

I suggest it is more parsimonious to think of WLE in these Gricean terms, than as a result of the separate principles of Conventionality and Contrast. Gricean maxims do not operate independently but in concert, and sometimes in conflict; relevance and quantity may equally play a role in word learning (as Frank & Goodman, 2014 demonstrate for quantity; and Sullivan & Barner, 2015 for relevance). This provides a

⁶ There are other accounts. Two notable ones are 'N3C' – Novel Name, Nameless Category (Golinkoff, Mervis, & Hirsh-Pasek, 1994), whereby children map novel terms to novel objects, without any reasoning by exclusion, and an associationist account (Smith, 1999), whereby repeated association of cues in the situation – co-present linguistic features or referents – leads to the emergence of attentional biases towards these cues, which helps children learn faster and more accurately. As these have not featured as prominently in the debate on the nature of word learning by exclusion, they will not be considered further here – though, as Frank (2014) argues, associationist and pragmatic accounts can be compatible.

solution to one criticism of a Conventionality and Contrast approach to WLE, namely that one still needs a basic-level bias to end up with the correct referent (Richard Breheny, personal communication). Just as the Mutual Exclusivity bias is assumed to operate alongside others, such as a whole-object bias and a basic-level bias (Markman & Wachtel, 1988), so too Conventionality and Contrast may not be sufficient on their own. The steps in reasoning outlined above crucially assume that the label must be referring to the whole object, and that it cannot be a hypernym or hyponym referring to one of the objects (including the known object).

On a Clarkian account of Conventionality, one would have to appeal to some notion of ‘conventional’ encompassing ‘most frequent’ or ‘most cross-situationally stable’. On a Gricean view, one can straightforwardly appeal to expectations of Cooperativeness in the context, derived from the maxims of quantity. In a situation with two objects of different kinds, using a hyponym is typically over-informative, and implies the presence of another object of the same kind; using a hypernym, on the other hand, implies speaker ignorance about an object. Imagine a context where the possible referents are an apple and an unknown object, and the speaker says ‘give me the *cox*’ – this is over-informative, and implies that the unknown object is also a kind of apple. Then, think of the same context, where the speaker says ‘give me the fruit’ – this is now under-informative, and either implies that the speaker is ignorant of what kind of fruit an apple is, or that the novel object is a kind of fruit (and is the intended referent). In both cases, if the novel object is clearly not fruit, then the utterance is to be avoided. On the same grounds, the listener can infer that a novel word is neither a hyponym nor hypernym of the known object. Of course, this assumes that young children are sensitive to speaker cooperativeness, and, specifically, informativeness, but this may well be justified based on early evidence of these skills (e.g. Schulze & Tomasello, 2015; Tomasello, Strosberg, & Akhtar, 1996).

Beyond these two theories – lexical heuristic or pragmatic principle – there is a third option: a developmental view. It is, of course, possible that children’s WLE strategy changes with development, and the same behaviour may be a product of different processes at different points. Theoretically, though, either direction of change – from pragmatic to lexical or vice versa – is plausible; it could be that a general pragmatic principle can become a more efficient lexical heuristic with

increasing linguistic experience, particularly of the organisation of the lexicon, or it could be that a low-level heuristic turns into more sensitive pragmatic reasoning, as that becomes available with cognitive development (cf. Kalashnikova, Mattock and Monaghan, 2014, for a similar outline).

2.3.2 Testing WLE theories

There have been two main approaches to testing which understanding of WLE is correct. Firstly, some studies have looked at how WLE interacts with other ‘pragmatic’ cues, such as eye-gaze, pointing or speaker characteristics (Diesendruck, Carmel, & Markson, 2010; Grassmann & Tomasello, 2010; Haryu & Imai, 1999; Jaswal & Hansen, 2006). The problem is that there are conflicting findings, and, moreover, conflicting interpretations of the same findings: for example, if a combination of pragmatic cues – gaze and point – can ‘override’ mutual exclusivity, is this evidence of the strength of the mutual exclusivity bias and that it is a lexical heuristic, not a pragmatic principle (Jaswal, 2010)? Or is it evidence that pragmatic interpretation takes into account many sources of information, including an expectation of speaker cooperativeness that may be weightier than single conflicting cues such as eye-gaze?

Secondly, other studies have compared WLE with a ‘domain general’ application of the same strategy – on the assumption that a lexical heuristic would only apply to disambiguation of word meaning, whereas a pragmatic strategy could be applied in other communicative situations. Studies have compared WLE to disambiguation of reference through facts (de Marchena et al., 2011; Diesendruck & Markson, 2001; Kalashnikova et al., 2014; Scofield & Behrend, 2007), and to symbolic gestures (Suanda & Namy, 2013). Again, results have been mixed, finding similar performance for words and facts (Diesendruck & Markson, 2001), or no relationship (e.g. Kalashnikova et al., 2014), depending in part on age, and whether it is a between-subject or within-subject correlational design. However, such comparisons are inherently problematic, because words are by nature conventional, whereas facts are not. This is a distinction that young children have already grasped (Behrend, Scofield, & Kleinknecht, 2001; Childers & Tomasello, 2003). Indeed, using facts referentially is pragmatically uncooperative, unless the facts are already in common ground. For instance, if the speaker has identified one object as ‘the one my uncle gave me’, the

referential expression 'the one my dog likes' could be co-operative by adding information about the object already given in the discourse, or by using a new description to refer to a new object. There is no way of knowing whether a patterning together of WLE and disambiguation of facts used referentially indicates the extension of a pragmatic strategy or a different mechanism with the same outcome, while a difference in behavior between the two tells us nothing about the nature of WLE.

Some studies along these lines have found evidence for a developmental change: Kalashnikova and colleagues (2014) find increasing disambiguation from children (aged 3;7-4;6 and 4;7-5;7) to adults for words but not for facts, and also no significant correlation between performance in the two conditions – though this is not positive evidence for no relationship, but an absence of evidence either way, that could be caused by skew and lack of variation in performance. Even these results are hard to interpret: the authors seem to say that they suggest a shift towards a specific lexical strategy for words, but that this is not a transformation 'from a strictly domain-general to a strictly domain-specific assumption about word learning' (Kalashnikova et al., 2014: 132). Again, there is the confound of what the responses to facts are meant to represent.

What is therefore needed is a comparison between WLE and an indisputably pragmatic phenomenon, just like implicature. On the pragmatic view of WLE, one would expect to find correlation between WLE and implicature across all ages (same strategy for implicature and words); and on the lexical heuristic view of WLE, one would expect no relationship between WLE and implicature, given a baseline of structural language ability. Alternatively, there may be one of two developmental patterns: it may be that WLE starts off as a product of general pragmatic reasoning, but becomes more specialised with age and increasing experience of language and the conventionality of word meaning, predicting decreasing correlation with age (cf. Kalashnikova et al., 2014; Kalashnikova, Mattock, & Monaghan, 2016). On the other hand, a lexical constraint could give way to more useful pragmatic reasoning that combines multiple contextual cues, as pragmatic competence develops, leading to increasing correlation with age.

2.3.3 Summary

Children's comprehension of implicatures seems to emerge around 3 years of age, and there is evidence that ad hoc quantity implicatures precede scalars, and that relevance is also early. WLE is observed still earlier in infants. Despite many observations of the connections between the two, no existing studies have fleshed out a Gricean view of WLE or tested its nature by comparing it to a clearly pragmatic phenomenon like implicatures (particularly ad hoc and relevance inferences). There is also emerging evidence of the role of QUD in children's development, which suggests an interaction of the development of quantity and relevance inferences, also as yet untested.

2.4 Experiment 1A: Quantity, relevance and WLE inferences in children

In the first experiment, I therefore aimed to investigate the relationship between quantity, relevance and WLE inferences in children's development, and the nature of WLE, by testing children aged 3–5 years in their ability to derive scalar, ad hoc, relevance and WLE inferences.

2.4.1 Method

2.4.1.1 Design

The experiment was based on a picture-matching task, which has been successfully employed to investigate children's pragmatic ability in comprehension (Antoniou & Katsos, 2017; Gotzner et al., 2015; Horowitz & Frank, 2015; Hurewitz, Papafragou, Gleitman, & Gelman, 2006; Katsos & Bishop, 2011; Miller et al., 2005; for a methodological review see Schmitt & Miller, 2010). In its basic form, participants hear an utterance, see two pictures, and are asked to choose the picture that goes with the utterance, i.e., depicts the speaker's intended meaning. As I have touched upon while reviewing the context for this study, it has many advantages that make it a valid measure of competence: it does not require metalinguistic reasoning; it provides visual cues to alternative utterances; it is not susceptible to a 'yes' bias or pragmatically tolerant behaviour; and, in general, it is a child-friendly, engaging task.

In this study, I embedded the picture-matching within a story-based task, both because this made it more naturalistic, and because a discourse or visual context tends

to facilitate children's inferencing (e.g. Scafton & Feeney, 2006). Each item consisted of a) a context sentence, b) a question, and c) the critical utterance (an answer to the question). The parts (a) and (b) were uttered by the narrator-experimenter, whereas (c) was said by a puppet (with recorded voice). For (a) and (b) a single image was visible on the page; for (c) two pictures were visible, which the participant had to choose between. Having pre-recorded utterances for the puppet had the advantages a) of making it clearly different from the experimenter's, and b) ensuring all children heard the critical utterances in the same way. Sets of 6 trials formed 'stories', for example 'Bob at the park' or 'Bob at home' (see Table 2.2 for examples).

The study had a $4 \times 2 \times 3$ design. There were four inference types (within-subject condition): relevance, ad hoc quantity, scalar quantity, and WLE. These appeared in two conditions (also within-subject): control and critical. In the control condition, only one of the pictures was a semantic match for the utterance – this was a measure of children's structural language, and understanding of the task in general. In the critical condition, both pictures could be semantic matches for the utterance, but only one matched the implicated meaning intended by the speaker. Thus, the participants' picture choices indicated whether they have derived an implicature or not. Finally, the participants were divided into 3 age-groups for the main analysis: 2;8–3;11, 4;0–4;11, and 5;0–5;11.

In the ad hoc quantity condition, one picture displayed one object, and the other picture displayed the same object and another object. The critical utterance was of the form, 'I V-ed an X', and the control 'I V-ed a Y and an X'. The common item was always second, so that children had to pay attention to the whole utterance, not just the final word, in order to disambiguate. Both pictures therefore matched the literal meaning of the critical utterance, *at least an X*, but only one matched the pragmatically enriched meaning *only an X*. The preceding question was always of the form 'What did you V?'.

In the scalar quantity condition, one picture showed only some items having some property, and the other showed all of them having the same property (i.e., the pictures were minimally different on this dimension). The critical utterance was of the form 'I V-ed some of the Xs', and the control 'I V-ed all of the Xs'. As with ad hocs, both pictures were therefore a match to the literal meaning of the critical utterance, *at*

least some Xs, but only one matched the enriched meaning *only some of the Xs*. The use of ‘some of’ rather than ‘some’ was in line with other developmental studies in English and known to improve children’s performance (e.g. Bill, Romoli, Schwarz, & Crain, 2014; for adults, see too Degen & Tanenhaus, 2014; Y. T. Huang & Snedeker, 2009; Katsos & Bishop, 2011; Verbuk, 2012). In the utterance, the quantifier was not contrast-stressed, even though this also improves children’s performance, as this would make scalar utterances different from ad hoc or relevance utterances. The preceding question was always of the form ‘What did you do with the Xs?’.

For relevance, each picture showed a different item that represented an activity. In the control condition, only one of the pictures depicted the utterance’s meaning; in the critical condition, on the literal meaning, neither picture seemed relevant, so the choice was ambiguous; on the implicated meaning, one of the pictures matched. The question always asked about which activity Bob, the puppet, wanted to do, e.g. ‘What game shall we play?’. Only ‘fulfilled condition’ inferences were used, to avoid the kind of reasoning by exclusion implied by a material implicature that could be the case for ‘unfulfilled condition’.

Finally, the WLE trials were based on one standard version of the WLE task (e.g. Diesendruck et al., 2010; Jaswal, 2010; Markman & Wachtel, 1988; Markman et al., 2003), in which there is one novel object and one known object. This format – rather than the alternative of two novel objects, one of which is first labelled – was chosen to make the WLE trials as similar as possible to the implicature trials, and to avoid having to mention either object in prior discourse. The utterances were always of the form, ‘I V-ed an X’. Word learning trials were somewhat different from quantity and relevance trials, in that there was only a minimal context phrase, such as ‘Bob went into the shop’, and no question. This was so that the discourse context did not provide a conflicting cue as to the intended referent.

	Context sentence	Critical utterance	Control utterance	Picture – critical	Picture – control
Relevance	It was breakfast time. Bob’s dad asked, ‘What would you like for breakfast?’	And I said, ‘I’ll get the milk.’	And I said, ‘I’d like toast.’	Cereal	Toast
Ad hoc	Bob was getting ready for school. His mum asked, ‘What have you packed in your bag?’	And I said, ‘I packed a hat.’	And I said, ‘I packed a book and a hat.’	Hat	Book and hat
Scalar	Bob made a crash in the kitchen. His dad asked, ‘What have you done with the pile of plates?’	And I said, ‘I broke some of the plates.’	And I said, ‘I broke all of the plates.’	Some (not all) plates broken	All plates broken
WLE	He went further inside and...	‘I picked a dax.’	‘I picked a fork.’	Novel object	Fork

Table 2.2 Experiment 1A example items

There were 5 stories (See Appendix 8.1). The first 4 stories included relevance, ad hoc and scalar trials, with 1 critical and 1 control for each, making 6 in total. There were 8 trials per utterance type, with each participant seeing 4 critical and 4 control trials for each type. In total, participants saw 32 trials (plus 4 unambiguous warm-up trials). All the word learning trials were in one block, and made up the final story, in order that the puppet’s use of novel words (and inferred choice of novel items) did not affect the participant’s perception of him as co-operative and rational. All stories were clearly presented as part of the same activity.

Participants saw only the critical or control condition for any one item; items within each story were rotated across participant lists, and arranged such that no two of any utterance type appeared one after the other and no more than two of the critical or control condition appeared together; and the first four stories (blocks) themselves were rotated. This counter-balanced design produced $2 \times 6 \times 4 = 48$ lists. In addition, across lists, the position of the pictures (left or right) was counter-balanced.

2.4.1.2 *Participants*

Participants were recruited from Foundation classes in two local primary schools, from nurseries and preschools in and around Cambridge, and through personal contacts, in the age range 2 years 8 months to 5 years 11 months. Headteachers and nursery managers were contacted with information about the study, and asked whether they would be willing for their educational setting to participate in the project. Parents of the relevant classes or groups were then sent information letters and asked either to opt in or to opt out of the study, in line with the school's policy. The recruitment process was approved by the University of Cambridge Psychology Ethics Committee, following research ethics guidelines (PRE.2014.98).

In total, $N = 135$ children were recruited. Some participants were excluded from analysis on the basis of: environment too noisy for stimuli volume ($N = 2$ in schools), failure to finish the task ($N = 1$ in schools; $N = 7$ in nurseries); or declared developmental disorder ($N = 1$ in schools, $N = 1$ in nurseries). In addition, some children were recruited (given parental consent) but chose not to take part in the study or were absent from school or nursery at time of testing ($N = 4$ in schools, $N = 13$ in nurseries). The responses from 71 monolingual children were included in the final analysis in this chapter; 35 bilingual children also completed the task and were included in the analysis in Chapter 3.

In addition, $N = 28$ children were recruited from Foundation classes in two other local primary schools for the pretests and piloting of this study.

$N = 15$ adults were recruited as a control group via Prolific Academic, an online recruitment platform for research (Prolific Academic Ltd, 2016), and were paid £0.75 for their participation.

2.4.1.3 *Materials*

The task used laminated picture cards fixed with magnetic strips in a 'book', an A5 ring-binder, with pages in plastic pockets so that their order could easily be changed for different lists. The picture cards were photographs, sourced from the BOSS database (Brodeur, Dionne-Dostie, Montreuil, & Lepage, 2010), or online via Pixabay (Braxmeier & Steinberger, 2017), a database of Creative Commons CC0 licensed images, or through an internet image search with filter applied for

noncommercial reuse. Each picture card had a clear black border to highlight that the two pictures are separate quantification domains. Chosen pictures were placed on a magnetic whiteboard by the child.



Figure 2.2 Story book, stand and magnetic board, showing scalar trial pictures

The context sentence and question were narrated by the experimenter, while the critical utterance was prerecorded as the puppet's voice by an adult male with a standard southern British English accent, using Audacity (Audacity Team, 1999). The recorded utterances are played from a laptop computer (Apple MacBook) using VLC (VideoLAN, 2017), either via the computer's inbuilt speakers or small USB-powered speakers (Logitech), depending on the volume required. The puppet was a soft hand puppet of the sort frequently used in classroom settings, sourced from a specialist firm (<http://aspuppets.co.uk/>).

The utterances were written to be child-friendly, in terms of topic and vocabulary. Those for relevance utterances (both critical and control) were devised through pretests: first, adults were asked for objects associated with activities to create a shortlist, via an online questionnaire, using Qualtrics (Qualtrics, 2016). Then, 4- to 5-year-old children ($N = 9$) were asked what a character would need to do an activity. For example, 'Bob wants to go swimming. What does he need to go swimming?'. Those pairs of items where there was no significant overlap of suggestions for the two items and a strong association with a particular object (mentioned frequently by children) were chosen.

For word learning, the novel words themselves were taken from other development studies, with 4 monosyllabic and 4 bisyllabic words with English phonotactics (Barner & Snedeker, 2008; Diesendruck & Markson, 2001; Diesendruck,

Markson, & Bloom, 2003; Halberda, 2003). Photographs of real objects were used, and were tested on adults (via online questionnaires) to find objects that the majority of adults did not recognise (where they responded with 'pass' or a wrong guess). The final list of novel objects consisted of: a dough mixer, a wooden gauge, a climbing chock, an aspergillum, a wooden cooking beater (of a type not usually found in UK kitchens), a mangosteen, an ackee, and an eccentric rod. Known items were also pretested with children to make sure the pictures were clearly identifiable, and problematic items were replaced (Appendix 8.1).

2.4.1.4 Procedure

Participating children were asked by the researcher whether they would like to do an activity with stories. The activity was usually set up in a separate group-work room, away from the noise and distraction of the classroom, although owing to UK child protection guidelines and space limitations, this was not always possible in nurseries. Children recruited through personal contacts were tested in a quiet room at their home.

The child sat at the table opposite the picture-book, which was mounted on a book rest with the magnetic board on the table in front. The experimenter sat to the side, so that the computer, puppet and picture-book could all comfortably be operated.

The experimenter introduced the activity to children, with the following information:

Today we're going to listen to some stories about Bob.

This is Bob.

I'm going to tell you some stories about what Bob did, but sometimes Bob likes to interrupt me; he likes to tell you himself what he did. So sometimes you'll hear Bob speaking and telling part of the story.

In this book, you will see pictures for the story. Sometimes you will see two pictures on one page. Then, choose the picture that goes with the story. You can put the picture that you choose on your board.

Then at the end of each story we will be able to look at the pictures on the board and remember what Bob did.

Remember, when you see two pictures, choose the picture that goes with the story.

Are you ready? Let's practice.

There was then a warm-up story with four trials with the same structure as trials in the main task, except that the critical utterance always unambiguously referred to only one of the pictures. At the end of the warm-up, the child was encouraged ('you've got the hang of that!') and asked whether they could go on to the next story. At the end of each story, the experimenter informed the child that there was another story, so that the child had the chance to stop if desired.

The experimenter looked between the child and the picture during the context sentence and question, to establish joint attention, but looked at the puppet during the critical utterance, and at the child as he or she chose the picture that goes with the story, so that the choice would not be influenced by the experimenter's eye-gaze. If the child was unsure and asked the experimenter for help, the experimenter would continue looking straight at the child, and encourage them to 'choose the picture that goes with the story'. If the child tried to choose both pictures, the experimenter gave a reminder to choose just one.

At the end, children were given a sticker as a thank you. At the end of the second test session (which consisted of language and theory of mind measures, described in Chapter 3), they were given another sticker and certificate of participation. Children recruited through personal contacts were also given a small book. In total, the session took around 15–20 minutes.

The children's picture choices were recorded as photographs of the board, either between each story (in the case of most school children), or after the session had finished (in the case of most nursery children).

On the whole, children seemed to enjoy taking part in the task. Their behaviour was quite varied – some children enjoyed chatting about the pictures and their own experiences (in which case the experimenter would try to gently bring them back to the task); others got on with the activity quietly. Most children very quickly understood the task, but others needed prompting to choose a picture with the critical utterance ('Can you choose the picture that goes with the story?').

2.4.2 Predictions

1. *What is the relationship between quantity, relevance and word learning by exclusion in development?*

Based on the pragmatic theory and extant findings already explored, I make the following tentative predictions:

- WLE and relevance emerge before ad hoc and scalar inferences, across children: WLE and relevance scores are higher than AH and scalar, with this gap closing with age
- Relevance inferences emerge before ad hocs, with higher scores for relevance, and the gap closing with age
- Ad hoc inferences emerge before scalar, with higher scores for ad hocs, and the gap closing with age
- There is a positive relationship between quantity and relevance inferences, when controlling for core language ability

2. *What is the nature of word learning by exclusion: pragmatic or lexical?*

- On a pragmatic view of WLE, WLE scores correlate positively with implicature scores
- On a lexical view of WLE, there is no relationship between WLE and implicature scores
- On a developmental view, the relationship may change over age, either becoming a more positive relationship with increasing age (lexical > pragmatic), or a less positive one (pragmatic > lexical)

2.4.3 Coding

Each child's picture choices (recorded as photographs of each story picture set), were coded as matching the inference or control utterance, and this was then converted into 'correct' or 'incorrect' based on the target for each item. In control trials 'correct' choices are ones that unambiguously match the utterance: in the case of ad hocs, the picture with two objects; for scalars, where objects have the appropriate property; for relevance, the semantic match; and for word learning, the known object. In critical trials, 'correct' choices are ones that match the pragmatically enriched

interpretation: for ad hocs, the picture with only one object; for scalars, where only some of the objects have the appropriate property; for relevance, the relevant picture; and for word learning, the novel object.

Children’s ages were calculated based on their date of birth and date of first testing session, and children categorised into three age groups (see Table 2.3). In addition, based on the background language questionnaires, children were classed as monolingual or multilingual. Only monolingual children’s data is considered in this chapter; I return to the multilingual children in the next chapter.

Age group	Monolinguals	Females	Mean age (months)
2;8-3;11	25	13	40.9
4;0-4;11	25	11	54.0
5;0-5;11	21	10	63.8
Total	71	34	

Table 2.3 Participants in Experiment 1A

2.4.4 Analysis

There is a clear developmental trend for ad hoc, scalar and relevance inferences, but not the WLE inferences, which are already close to ceiling in the youngest age-group tested (Figure 2.3, Figure 2.4 and Table 2.4).

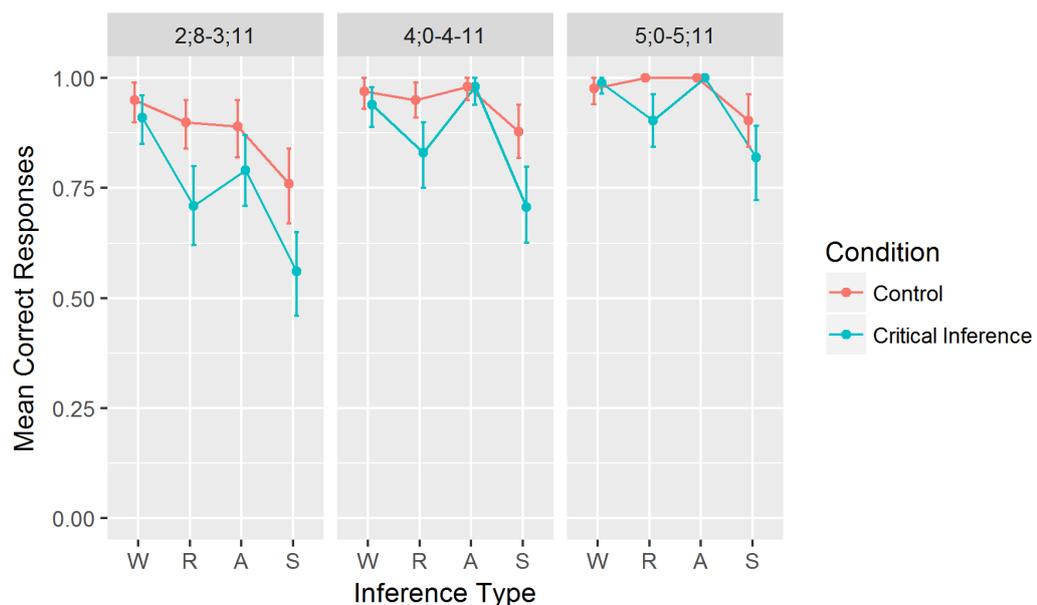


Figure 2.3 Mean correct responses by utterance type, condition and age group
W = Word learning; R = Relevance; A = Ad hoc; S = Scalar;
Error bars show bootstrapped 95% confidence intervals for between-subject comparison

	2;8-3;11		4;0-4;11		5;0-5;11	
	Critical	Control	Critical	Control	Critical	Control
WLE	.91	.95	.94	.97	.99	.98
Relevance	.71	.90	.83	.95	.90	1.00
Ad hoc	.79	.89	.98	.98	1.00	1.00
Scalar	.56	.76	.71	.88	.82	.90

Table 2.4 Proportion of correct responses by condition, inference type and age group

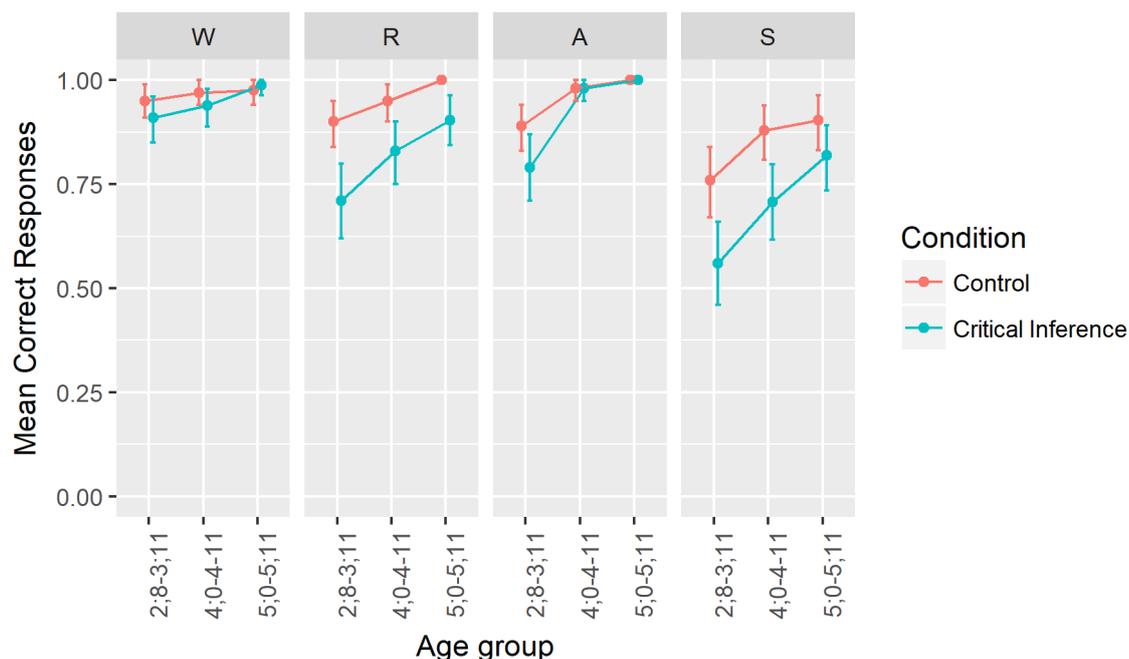


Figure 2.4 Mean correct responses by age group, condition and utterance type

W = Word learning; R = Relevance; A = Ad hoc; S = Scalar;

Error bars show bootstrapped 95% confidence intervals for between-subject comparison

To address the first research question, I use mixed-effects logistic regression models to examine developmental trends across all types of inference.

The second research question requires correlational analyses to specifically compare word learning to implicature inferencing. As across age-groups there is largely ceiling performance and lack of variation in WLE, I compare WLE and ad hoc and relevance inferences, controlling for language, only in the youngest group, for whom scores are lower and there is more variation.

The adult control group perform at ceiling across all conditions and are not included in the analysis (Figure 2.5).

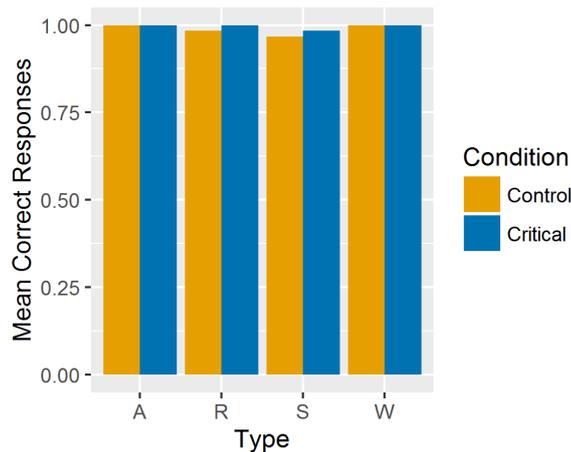


Figure 2.5 Mean correct responses by adults
A = Ad hoc; R = Relevance; S = Scalar; W = Word learning

2.4.4.1 Theory ‘neutral’ – sum coding of contrasts

In the first analysis, the main predictors – condition (critical or control), utterance type (scalar, ad hoc, relevance or WLE), and age group – are entered into the model with sum coding: each factor level is compared to the grand mean, so that main effects can be examined.

A mixed-effects logistic regression model was fitted, using the *lme4* package as part of the R programming language for statistical analysis (Bates, Mächler, Bolker, & Walker, 2015; R Core Team, 2016), using RStudio (RStudio Team, 2016)⁷. The advice of Barr, Levy, Scheepers and Tily (2013) for confirmatory hypothesis testing with mixed models was followed: the maximal model⁸ with all random effects would not converge (which can be a particular problem when using logistic, rather than linear, regression, and where data includes floor or ceiling effects, and possibly small random effects), and so separate models with by-item and by-subject random effects were fitted. Firstly, a model with condition, utterance type and age group as fixed effects, and item by condition, age group and block (or story) number random slopes was fitted, with sum coding of fixed effects. This indicates a main effect of condition, such

⁷ R scripts for all chapters can be found at <https://github.com/elspethwilson>

⁸ Note that the models presented here are not maximal in its proper sense: they do not include the factors of item order, story topic or picture location (left/right for critical item). However, these were counterbalanced across participants. The random effects included are those required by the assumptions of the model (to counteract the assumption of independence of predictors).

that the control condition is higher than the grand mean ($\beta = .53, p < .001$); a main effect of scalar inference type, such that the scalar type is lower than the grand mean ($\beta = -1.25, p < .001$); and an effect of the age group 2;8-3;11, such that it is lower than the grand mean ($\beta = -1.02, p < .001$) – Table 2.5.

	Estimate	SE	z	p
Intercept	2.81	.16	17.1	< .001
Control	.53	.13	4.20	< .001
Ad Hoc	.37	.22	1.72	.086
Relevance	-0.15	.20	-0.78	.44
Scalar	-1.25	.12	-6.56	< .001
2;8-3;11	-1.02	.16	-6.34	< .001
4;0-4;11	.014	.14	.10	.92

Table 2.5 Response ~ Condition + Type + Age group + (1 + Condition + Age group + Block | Item)
Glmer, family = binomial, optimizer = bobyqa, sum coding

Secondly, a model fitted with the same fixed effects and Subject by type random slope was less conservative, indicating in addition small main effects of ad hoc and relevance type, above and below the grand mean respectively. For an alternative approach suggested by Barr et al. (2013), see Appendix 8.2.⁹

⁹ Although neither of these approaches is an ideal implementation of generalised mixed models with maximal random effects, they are arguably still better than the alternative non-parametric variants such as a Friedman’s ANOVA, as they avoid the loss of information through aggregation (per participant/per item), and can easily accommodate multiple predictors.

	Estimate	SE	z	p
Intercept	3.39	.32	10.66	< .001
Control	.57	.08	7.07	< .001
Ad Hoc	2.09	.84	2.50	.013
Relevance	-0.81	.032	-2.56	.011
Scalar	-1.84	.31	-5.91	< .001
2;8-3;11	-0.076	.22	-3.50	.00042
4;0-4;11	.012	.20	.061	.95

Table 2.6 *Response ~ Condition + Type + Age group + (1 + Type | Subject)*
Glmer, family = binomial, optimizer = bobyqa, sum coding

2.4.4.2 Theory dependent – successive difference coding of contrasts

In the second analysis, the factors are coded with successive difference contrasts, so that each level is compared with the previous one. The order of levels within each factor is based on the predictions made in the opening section of this chapter, such that each level is expected to have a lower score than its preceding level. The orders were: control–critical, WLE–relevance–ad hoc–scalar, and decreasing age groups. The intercept is still the grand mean.

As per the approach with sum coding, a mixed effects logistic regression model was fitted with condition, utterance type and age group as fixed effects, and item by condition, age group and block random slopes. This indicates a difference in condition, such that the rate of correct responses for critical trials is lower than for control trials ($\beta = -1.06$, $p < .001$); a difference between relevance and WLE, such that rate of correct response is lower for relevance than WLE ($\beta = -1.18$, $p = .0024$); no difference between relevance and ad hocs; but a difference between ad hocs and scalars, with scalars lower than ad hocs ($\beta = -1.63$, $p < .001$). There is also a difference between age groups: 4-year-olds perform worse overall than 5-year-olds ($\beta = -0.99$, $p = .0024$), and 3-year-olds worse than 4-year-olds ($\beta = -1.04$, $p < .001$) – Table 2.7.

	Estimate	SE	z	p
Intercept	2.80	.16	17.1	< .001
Critical – Control	-1.06	.25	-4.20	< .001
R – WLE	-1.18	.39	-3.03	.0024
AH – R	.052	.32	1.64	.10
S – AH	-1.63	.33	-4.89	< .001
4;0–4;11 – 5;0–5;11	-0.99	.33	-3.04	.0024
2;8–3;11 – 4;0–4;11	-1.04	.20	-5.05	< .001

Table 2.7 *Response ~ Condition + Type + Age group + (1 + Condition + Age group + Block | Item)*
Glmer, family = binomial, optimizer = bobyqa, backward difference coding

The same model with Subject by condition and type random slope (nlminbw optimizer) confirmed these results, but again was less conservative, with an additional significant effect of ad hoc vs relevance, such that ad hoc scores are higher than relevance. In the following discussion, I will base comments on the more conservative models with by-item random slopes.

2.4.4.3 Comparison to chance in three-year-olds

Although three-year-olds perform more poorly than the other two age groups overall, they still might be at above chance levels (in this case 50%). To find out whether this is the case on inference trials, a model with just the intercept, and random intercepts for item and subject was fitted for each type of inference (critical trials only). Only scalar implicatures are not above chance in three-year-olds (Table 2.8).

Type	Estimate	SE	z	p
Ad hoc	3.04	1.34	2.26	.024
Relevance	1.48	.69	2.16	.031
Scalar	.25	.21	1.16	.25
Word learning	3.8	1.46	2.6	.0094

Table 2.8 *Response ~ 1 + (1 | Item.no) + (1 | Subject)*
Glmer, family = binomial, optimizer = bobyqa, contrast coding, critical trials only

2.4.4.4 Relationship between inference types

To examine the relationship between relevance and quantity inferences, I conducted partial correlations (Kendall's τ) for scores in the critical condition

controlling for language (scalar / ad hoc control score) and for age in months (Bonferroni correction applied: α level $.05/2 = .025$). For scalars, there is a significant positive relationship of small to moderate size with relevance when controlling for language and age: $\tau = .21$, $z = 2.5$, $p = .012$. For ad hocs, there is no significantly positive relationship: $\tau = .078$, $z = .94$, $p = .35$.

2.4.4.5 Effect of block number

To examine whether there is a practice effect – that is whether children improve in their performance over the experiment – I added the block number as a predictor into the model with contrast coding, and conducted model comparison (where the blocks are numbered A–D, so that each story appears in each block across lists, excluding block E, which is the WLE story) – Figure 2.6. Adding in block did not significantly improve the model; there is no main effect of block, though visual inspection suggests that different age groups may perform differently for different utterance types across blocks – Table 2.9.

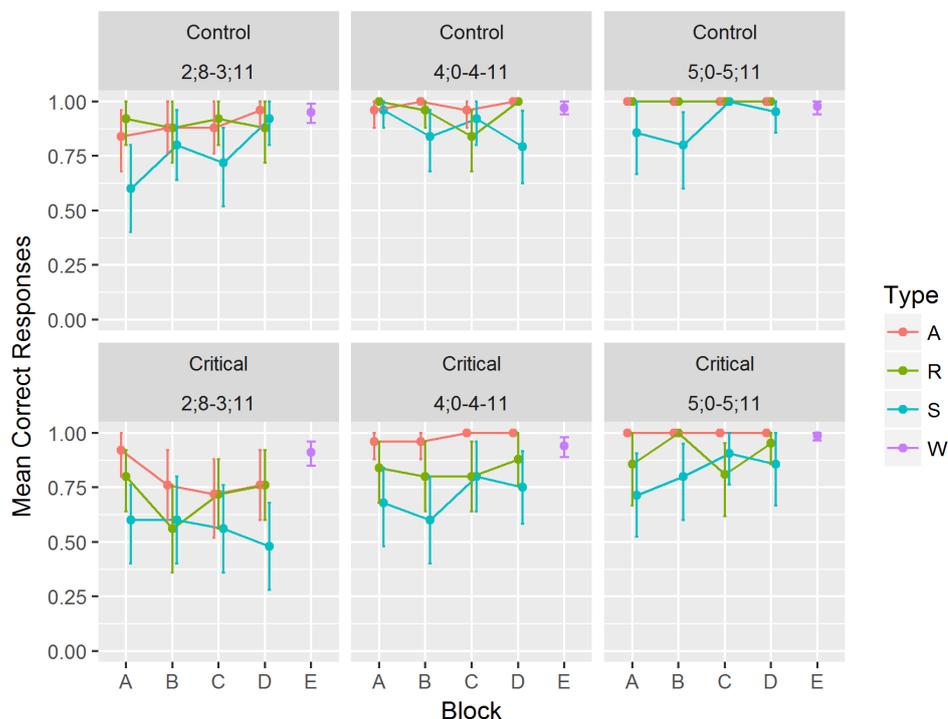


Figure 2.6 Mean correct score by age, inference type, condition and block
Error bars show bootstrapped 95% confidence intervals for between-subject comparison

Model	Df	AIC	Log Lik	Deviance	χ^2	p
Score ~ 1 + (1 + Critical + Age group + Trial_block Item)	37	1476.3	-701.17	1402.3		
Score ~ Critical + (1 + Critical + Age group + trial_block Item)	38	1473.6	-698.81	1397.6	4.72	.03
Score ~ Critical + Type + (1 + Critical + Age group + Trial_block Item)	41	1452.9	-685.45	1370.9	26.71	< .001
Score ~ Critical + Type + Age group + (1 + Critical + Age group + Trial_block Item)	43	1419.5	-666.76	1333.5	37.39	< .001
Score ~ Critical + Type + Age group + Trial_block + (1 + Critical + Age group + Trial_block Item)	46	1421.8	-664.9	1329.8	3.71	.29

Table 2.9 Model comparison for effect of block order

2.4.4.6 WLE and pragmatic inferencing

Overall, WLE scores are close to ceiling, so to examine the relationship between WLE and pragmatic inferencing, I ran partial correlations (Kendall's τ) for WLE scores (critical condition) compared to relevance and ad hoc scores, controlling for language ability (relevance or ad hoc control score), in the youngest age group only. There was no significant correlation between WLE and relevance scores ($\tau = -0.05$, $z = -0.34$, $p = .73$), but a significant, small to moderate one for WLE and ad hocs ($\tau = .34$, $z = 2.3$, $p = .021$), correcting for multiple comparison (Bonferroni, α level at .025).

2.4.4.7 Analysis with 6-month age groups

As other studies suggest that the fourth year of life is a critical one for pragmatic development (e.g. Schulze et al., 2013; Stiller et al., 2015), I also conducted an exploratory analysis with 6 smaller age groups (Figure 2.7). Note, though, that the sample size for each age group is now too small to draw firm conclusions (Table 2.10).

Age group	Monolinguals	Female	Mean age (months)
2;8-3;2	9	5	36.1
3;3-3;8	12	6	42.6
3;9-4;2	11	4	48.6
4;3-4;8	11	4	54.0
4;8-5;2	16	9	60.1
5;3-5;10	12	6	65.6

Table 2.10 Participants in 6-month age groups

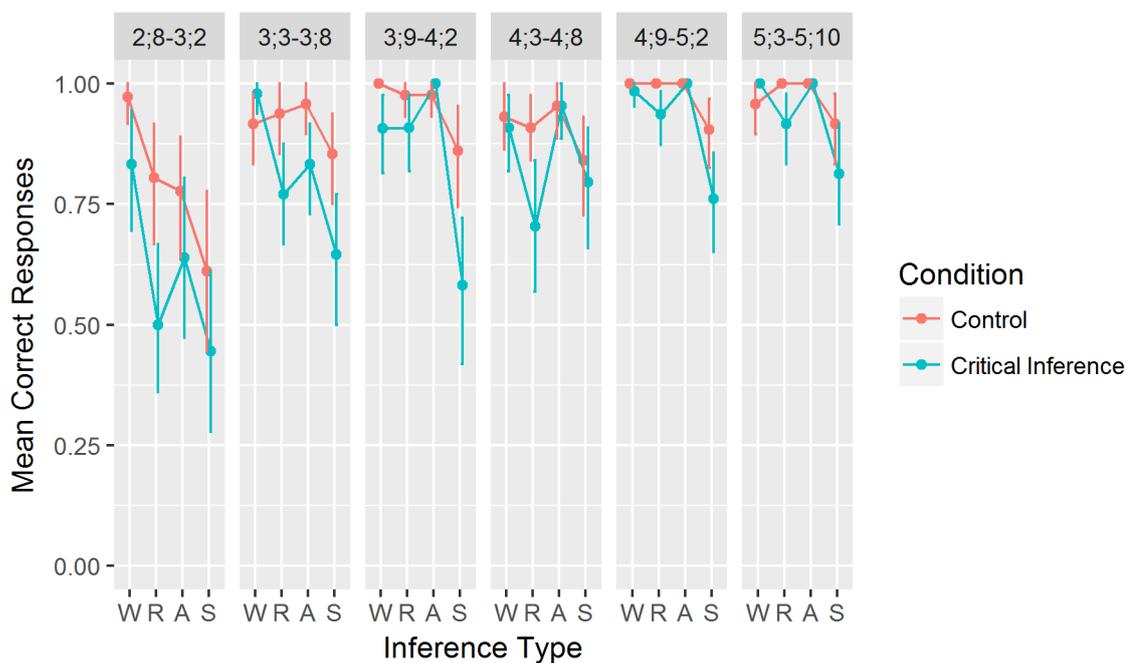


Figure 2.7 Mean correct responses by inference type, condition and age group (small)
Error bars show bootstrapped 95% confidence intervals for between-subject comparison

	2;8-3;2		3;3-3;8		3;9-4;2		4;3-4;8		4;9-5;2		5;3-5;10	
	Crit	Con	Crit	Con	Crit	Con	Crit	Con	Crit	Con	Crit	Con
WLE	.83	.97	.98	.92	.91	1.00	.91	.93	.98	1.00	1.00	.96
R	.50	.81	.77	.94	.91	.98	.70	.91	.94	1.00	.92	1.00
A	.64	.78	.83	.96	1.00	.98	.95	.95	1.00	1.00	1.00	1.00
S	.44	.61	.65	.85	.58	.86	.80	.84	.76	.90	.81	.92

Table 2.11 Proportion of correct responses by condition, inference type and age group
R = Relevance, A = Ad hoc, S = Scalar; Crit = Critical, Con = Control

Visual inspection shows that there is indeed substantial change over the fourth year (over groups aged 2;8–3;2, 3;3–3;8, and 3;9–4;2). I fitted the same models as above, but this time with these smaller age groups as a fixed effect, and corresponding random effect for item (Table 2.12).

	Estimate	SE	z	p
Intercept	2.9	.18	16.6	< .001
Control	.58	.15	3.9	< .001
Ad Hoc	.46	.25	1.88	.061
Relevance	-0.14	.21	-0.66	.51
Scalar	-1.47	.24	-6.21	< .001
2;8–3;2	-1.85	.22	-8.4	< .001
3;3–3;8	-0.53	.22	-3.4	.017
3;9–4;2	.032	.27	.12	.91
4;3–4;8	-0.31	.23	-1.38	.17
4;9–5;2	1.68	.53	3.16	.002

Table 2.12 *Response ~ Condition + Type + Age group (small) + (1 + Condition + Age group (small) + Block | Item)*

Glmer, family = binomial, optimizer = bobyqa, sum coding

This indicates, again, a main effect of Condition, such that control is higher than the grand mean ($\beta = .58$, $p < .001$); a main effect of type, such that scalar trials are below the grand mean ($\beta = -1.47$, $p < .001$); and a main effect of the youngest two age groups, which are below the grand mean ($\beta = -1.85$, $p < .001$; $\beta = -0.53$, $p = .016$), and the 4;9–5;2 group which are significantly above it ($\beta = 1.68$, $p = .002$).

A model with backward differences confirmed that the youngest group performed significantly more poorly than the next youngest ($\beta = -1.31$, $p < .001$) – Table 2.13.

	Estimate	SE	z	p
Intercept	2.9	.18	16.6	< .001
Critical – Control	-1.17	.30	-3.9	< .001
R – WLE	-1.29	.43	-2.98	.0029
AH – R	.60	.35	1.74	.082
S – AH	-1.9	.40	-4.82	< .001
4;9-5;2 – 5;3-5;10	.70	.73	.95	.34
4;3-4; – 4;9-5;2	-2	.68	-2.93	.0034
3;9-4;2 – 4;3-4;8	.35	.37	.93	.35
3;3-3;8 – 3;9-4;2	-0.57	.37	-1.5	.13
2;8-3;2 – 3;3-3;8	-1.31	.25	-5.29	< .001

Table 2.13 *Response ~ Condition + Type + Age group(small) + (1 + Condition + Age group(small) + Block | Item)*

Glmer, family = binomial, optimizer = bobyqa, sum coding

Looking at the very youngest age group (2;8–3;2), a model with just the intercept, and random intercepts for item and subject was fitted for each type of inference (critical trials only), to test performance against chance (50%). Only WLE is above chance (Table 2.14).

Type	Estimate	SE	z	p
Ad hoc	.76	.59	1.29	.197
Relevance	-0.04	.53	-0.083	.93
Scalar	-0.27	.48	-0.56	.58
Word learning	41.7	12.6	3.3	< .001

Table 2.14 *Response ~ 1 + (1 | Item.no) + (1 | Subject)*

Glmer, family = binomial, optimizer = bobyqa, sum coding

2.5 Discussion

2.5.1 Findings

In this study, a novel design was used to examine three different types of implicatures in development in children from 2 years 8 months to 5 years 11 months – relevance, ad hoc quantity and scalar quantity – together with WLE inferences. It is the first study to my knowledge that combines these four inferences in a single study and allows for some comparison between them.

2.5.1.1 Developmental trend for implicatures

The results indicate a developmental trend for implicature comprehension and for semantic comprehension. There is an increasing rate of correct responses with age, whether the correct picture choice requires only a semantic match or the picture selected indicates that the pragmatically enriched meaning has been inferred. Each age group performs better than the younger preceding one, collapsing over condition and inference type.

In addition, it is clear that implicature comprehension in particular is still developing across this age-range, as there is also a main effect of condition, with control trials at higher correct rates than critical trials. Note that this is expected in this experimental design, where control trials are semantically unambiguous, in contrast to some other studies, where the control or baseline condition creates ambiguity or uncertainty such that chance or below chance levels are expected (e.g. Stiller et al., 2015).

There is also a difference between types of pragmatic inference, seen in differences in rate of correct responses across utterance types (though note this includes critical and control conditions): WLE is clearly the easiest; relevance is harder, but ad hoc is not harder than relevance; and scalars are hardest. This aligns with the predictions made about the relative challenge of the different types and a possible developmental order, with the exception of relevance inferences, which were not easier than ad hoc ones – I will discuss below why they may have proved challenging in this study.

2.5.1.2 Quantity and relevance inferences

There is a positive relationship between scalar and relevance inferencing abilities, when partialling out age and language (control scores), i.e., children who do better with scalar inferences also do better with relevance ones, when the contribution of age and language ability (control trials) is also taken into account. There is no evidence for a relationship between ad hocs and relevance.

2.5.1.3 Pragmatic competence at 3 years

The results also show that 3-year-olds as a whole group (2;8–3;11) already have significant pragmatic competence: besides WLE inferences, they are adept at ad hoc

and relevance inferences, scoring above chance level in critical trials. Given the high rate of correct response in the control condition, they are also clearly grasping the task, and it is largely within their linguistic competence. Scalar inferences are more challenging for them, as they are for older children, too.

However, when the group is split into younger and older groups (2;8–3;2, 3;3–3;8 and 3;9–4;2) in an exploratory analysis, it can be seen that it is the older groups that are driving this effect: the youngest of these groups is only above chance in WLE, and there is a main effect of age for the youngest two groups, who are significantly below the grand mean.

2.5.1.4 Word learning by exclusion

Overall, children excelled in WLE inferences, and performed better with these than with any implicature inference. In the youngest age group (2;8–3;11), there was no evidence for a relationship between WLE and relevance inferences, when controlling for language (relevance control scores), but there was a moderate positive relationship between WLE and ad hoc inferences, when controlling for language (ad hoc control scores).

2.5.2 The development of implicatures

The findings confirm the gradual development of pragmatic skills, as children's performance becomes more adult-like with age. They thus corroborate the findings of other studies that employ similar measures for one or two implicature types (e.g. Horowitz & Frank, 2015; Pouscoulous et al., 2007; Schulze et al., 2013; Stiller et al., 2015). While there is evidence for the emergence of the ability to make pragmatic inferences during the fourth year of life, children become more and more adult-like in their implicature derivation across the preschool years.

Many features of this study's design were implemented to support children's inferencing – to make it as easy as possible for them to derive an implicature. For example, a picture-matching task does not require metalinguistic or explicit reasoning, and it does provide relevant alternatives in the visual context, and a co-operative, rather than under-informative, speaker. It was also a naturalistic engaging task, with trials with a question-and-answer structure set within a story. In this way, it was aimed at revealing their maximal pragmatic competence. Even so, children in the age-

range I tested show an improvement towards adult-like pragmatic behaviour. However, I will also discuss below some features that may have actually masked children's pragmatic abilities by making aspects of the inferencing process more challenging.

The findings also indicate that different types of inferences follow different developmental paths, so that children reach adult-like performance earlier or later in each. As I outlined above, this is something that neo-Gricean pragmatic theory would lead us to expect, as different types of implicature involve different inferential processes, and depend on other factors like structural language or world knowledge to different extents in different situations. Relevance implicatures involve an elaborative inference that renders the utterance relevant; quantity implicatures involve, in addition, accessing and negating relevant alternatives. For ad hoc implicatures, alternatives are found in context; for scalar implicatures, some lexical knowledge of alternatives is required. On this basis, I made a tentative prediction that relevance might emerge before ad hoc quantity, which, in turn, might emerge before scalar quantity inferences. The results indicate that, after WLE, ad hoc and relevance emerge before scalar implicatures, but provide no evidence for relevance emerging before ad hocs.

2.5.2.1 Quantity and relevance

Across all age groups, there was also a small-to-moderate positive relationship between scalar and relevance inference, when also taking language (scalar control trials) and age into account. One possible explanation is that there is a common skill underlying development in both, namely an ability to track the QUD and make elaborative inferences. There is already evidence that this is a challenge for children with scalar inferences (Skordos & Papafragou, 2016), and they are aided when the QUD is consistent and clear, rendering alternatives salient. It can therefore be expected that, while these seem and have been treated as very distinct pragmatic inferences, they might develop somewhat in tandem. Of course, a correlation is only indicative, and there are many other plausible explanations, such as general intelligence and attentional capacity which could be facilitating ability to do the task. Further research is needed to isolate the contribution of tracking the QUD and

elaborative inferences, for example, through direct manipulation of the QUD and the availability of alternatives.

There was no evidence for a relationship between ad hoc and relevance inferences, which could be for a number of reasons. Firstly, children were approaching ceiling in ad hocs and there was little variation. Secondly, there were only 4 critical items in each condition. Thirdly, while this study had the great advantage of testing all children with both quantity and relevance implicatures in the same experimental paradigm, there may still have been differing levels of complexity between trials for different implicature types. In particular, while the stronger alternative was visually presented for ad hocs, for the relevance implicatures, the literal meaning of the utterance was not depicted visually, nor was the link in reasoning to the intended meaning via some relevant piece of world knowledge. Take the following example:

Bob's dad asked, 'What game do you want to play?'
And I said, 'I brought a ball'
[Tennis OR Cards]

At the point of the critical utterance, only the intended meaning (tennis) was visually available (depicted by a tennis racket only), together with an alternative (a card game). However, there was no picture of a tennis game in action with the tennis racket and ball together – which was necessary so that the picture was not just chosen as a semantic match. It also meant, though, that the fact that a ball was needed to play tennis was required from world knowledge. This in itself might have made relevance inferences more challenging than ad hocs in this particular study, and obscured any relationship.

2.5.2.2 *Early pragmatic competence*

The findings fit in with other recent studies that find evidence for competence with some implicatures emerging at the age of three years: Stiller, Frank and Goodman (2015) observe ad hoc inferences in American-English-speaking children, and Schulze, Grassmann and Tomasello (2013) observe relevance inferences in German-speaking 3-year-olds. In some ways, this is not surprising. Even younger children seem to have some pragmatic inferencing abilities in understanding the intended meaning of

pointing or requesting (e.g. Grosse, Moll, & Tomasello, 2010; Tomasello et al., 2007). The linguistic and processing demands of this task were also minimised as far as possible, with no metalinguistic reasoning, explicit judgement or tracking a variably co-operative speaker required, as required for Felicity Judgement Tasks. Early competence does, though, present a puzzle of whether and how implicatures can be derived before standard Theory of Mind tests are consistently passed, which I take up in Chapters 3 and 5.

However, the exploratory analysis with narrower age groups corroborates others' findings that before age 3, children do not seem to be able to derive these implicatures, at least in the context of the kind of tasks typically used in experimental pragmatics (Stiller et al., 2015; Yoon et al., 2015). This could be a true reflection of the competence of two-year-olds, and indicative of key differences between verbal pragmatic inferences like implicatures and non-verbal ones, which are available at a younger age (for instance, Schulze and Tomasello (2015) find that even 18-month-olds can interpret non-verbal indirect requests). Alternatively, it could be an artefact of task designs, which thus far have essentially been adapted from those used with older children, rather than designed afresh based on what is known about toddlers' cognitive capacities and experimental tasks that have been successfully employed with them.

Ad hoc quantity implicatures

The youngest age group's (2;8–3;11) high performance with ad hoc inferences suggests that they already possess a number of skills. Besides the prerequisite joint attention and intention reading for communication, they need to be sensitive to informativeness; to be able to conceive of the stronger and weaker alternatives provided visually in context as being related but distinct sets in an entailment relationship; to recognise the stronger utterance as one that the speaker could have said but did not; and to engage in some sort of counterfactual reasoning with negation of the stronger alternative.¹⁰

¹⁰ This rich Gricean interpretation is not the only possibility. The simple ad hoc and relevance inferences in this task could in principle be derived via more basic heuristics (e.g. 'match one uttered label to one object') or associative mappings (e.g. 'milk goes with cereal'). These alternatives raise questions of their own, such as why do 2-year-olds not have these heuristics, when they can do WLE? In Chapter 5, I return to the possibility that pragmatic-like communication is possible without full Gricean reasoning.

If, as suggested by some other studies and by the analysis with smaller age groups here, two-year-olds are unable to derive ad hoc implicatures (Stiller et al., 2015; Yoon et al., 2015), one possibility is that some of these skills (or their integration) develop at around this age. Counterfactual reasoning is already firmly in place in word learning (e.g. Halberda, 2006) and in understanding others' attention to objects (Moll, Koring, Carpenter, & Tomasello, 2006). It is closely linked to the ability to recognise the stronger utterance as one that the speaker could have said but did not, and a related skill is also required in WLE: accessing the label associated with the known object and recognising it as one that the speaker could have said but did not. This requires productive knowledge of that label, which is indeed a predictor of an exclusion inference (Grassmann, Schulze, & Tomasello, 2015). Sensitivity to informativeness is also evident even in two-year-olds, for instance in their ability to distinguish between test and genuine questions based on the situation (Grosse & Tomasello, 2012). It might be, then, that it is the ability to conceive of the alternatives as contrasting but related sets which develops, enabling the inference of ad hoc implicatures.

Another possibility, mentioned by Yoon, Yu and Frank (2015) is a task factor: the salience of the alternative is such that young children struggle to inhibit it as a choice. The tasks used by Stiller, Frank and Goodman (2015), and Yoon, Wu and Frank (2015), as well as in this study, always had a single object in the target picture and two objects in the distractor picture for ad hoc trials. Two objects are unavoidably more visually salient than one, and the picture with two objects includes one that is unique across the whole visual array. This is clearly a difference from the WLE inference, where alternatives are both single objects. What to my knowledge has not yet been tested, is whether carefully controlling the salience of the two pictures would boost performance in toddlers. One way of doing this would be to use several objects of the same type, as used by Horowitz and Frank (2015) with 4- and 5-year-olds, and adopted in Experiment 4 here. The overall number of objects could therefore be the same on both cards or the relative numbers manipulated to vary salience. An alternative would be to vary object size. If, with salience controlled for, children are still at or below chance, then it is more likely some aspect of the inference process itself, yet to be acquired, that is causing children's difficulties.

Relevance implicatures

In this study the youngest age group (2;8–3;11) are also competent with relevance inferences, performing above chance. This corroborates Schulze, Grassmann and Tomasello's (2013) findings with German-speaking children: their younger group, aged 2;10–3;1, were not above chance with relevance inferences based on a fulfilled precondition (mean = 58%), whereas their older group, aged 3;10–4;1, were above chance (mean = 62%). In this study, the age range of the youngest group is wider, encompassing both younger and older three-year-olds who were above chance overall, but in the exploratory analysis with narrower age groups (2;8–3;2, 3;3–3;8 and 3;9–4;2), the same pattern is seen, as the youngest group is at chance. It therefore seems that the fourth year of life is a crucial time for the development of competence with this kind of relevance inference, with other kinds, such as nonverbal indirect requests, emerging even earlier (Schulze & Tomasello, 2015).

One problem with this study is that the associations for the elaborative inference were only pretested with the older children, and so it could plausibly be lack of relevant world knowledge that is restricting the youngest children's performance. This was not a problem in Schulze, Grassmann and Tomasello's study, however. One way to mitigate this could be, for example, a warm-up phase in which the relevant facts are in some way recapped with the child, so that they are, to some extent, part of the common ground and shared experience with the speaker. Alternatively, intuitively easier relevance inferences – such as from preference to desire, as in Schulze et al. (2013) – could be employed.

A further observation about a difference between the stimuli in these two studies is worth noting. In Schulze, Grassmann and Tomasello's task, 4 of the 8 stimuli use the first person (e.g. 'I bought milk', 'I've got a leash') and the other 4 are impersonal constructions (e.g. 'The toothpaste is in the bathroom'), whereas in this study they are all of the first type (e.g. 'I'll get the milk'). It could be that the use of the first person, in referring to an intentional *action* of the speaker, is a clearer cue to the speaker's intended *meaning* than an impersonal description. Perhaps an explanation such as 'a person buys / brings / gets something necessary for what he wants' is easier for children to arrive at than 'a person mentions the existence of something necessary for

what he wants', which might explain why the older three-year-olds do much better in this study (3;9–4;2 score 91% compared to 62% in Schulze and colleagues' study). Of course, this proposal would need testing systematically with a larger sample of children, but it highlights the many factors involved in a relevance implicature that have to be acquired.

2.5.2.3 Development of scalar implicatures

Children aged 4 and 5 years were beginning to be able to derive scalar implicatures, comparable to other studies that use picture-matching or action-based tasks (Horowitz & Frank, 2015; Pouscoulous et al., 2007), with competence with the stronger alternative – in control trials – preceding competence with the implicated meaning of the weaker term – in critical trials. So, while scalar implicatures, in this case with *<all, some>*, present particular challenges in development, the age gap between competence with ad hoc or relevance inferences and scalars is not as great as once thought. As predicted, though, it is the hardest type of inference for children.

Various specific features of the task may have boosted children's performance: the visual availability of the stronger alternative; the absence of numeric alternatives in the subitizing range which would compete as alternatives; and the control trials providing the stronger alternative in the discourse context. Indeed, if one looks only at the scalar trials for the first block (i.e., the first two scalar trials for each participant, one critical and one control), then it seems from the numerical pattern that scalar inferences receive a particular boost from a preceding trial with 'all' for the 4-year-olds (Figure 2.8).

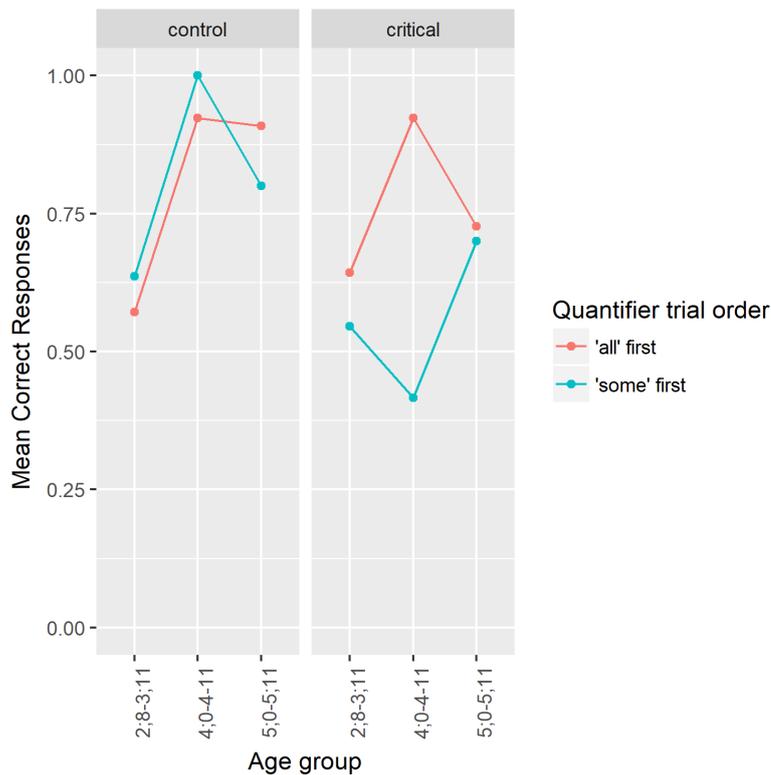


Figure 2.8 Proportion correct responses for first block scalar trials only

There are other aspects of the task, though, that may have made it more challenging than other tasks: the changing QUD across trials (cf. Horowitz & Frank, 2015, who find worse performance when ad hoc and scalar trials are mixed); the more complex pictures containing more items, well above the subitizing range; and the lack of explicit QUD about quantity – questions in scalar trials were of the form, for example ‘What did you do with the plates?’.

An alternative explanation for children’s high performance is that what they are doing here is not a scalar implicature per se – reasoning about informativeness – but purely a contrastive inference, as proposed by Katsos and Wilson (in prep.) and, independently, by Sullivan, Davidson, Wade and Barner (submitted), following Bale and Barner’s (2013) observations about evasive relevance inferences. On this account, when children are given alternatives, like *some* and *all* (either in discourse or visually), they are able to reject under-informative *some* on the basis that it is not *all*, rather than because it is less informative than *all*. In other words, they are performing an exclusion inference, as in WLE. Sullivan and colleagues (submitted) set out to test this possibility, but their results seem inconclusive. The studies they choose to cite imply

that they might expect this to be an issue not only for contrast in discourse (for example, ‘Did you feed all the ducks? I fed some of them’), but also when there is visual contrast only (as in the present study, and other picture-matching tasks). However, their experimental design instead assumes that exclusion is only possible with alternatives contrasted in discourse. 4–7-year-old children do not reward Puppy when he has coloured *all* the stars when he was requested to ‘colour some of them’ (unlike adults, i.e., a scalar entailment relation). They check that this is not due to children deriving a scalar implicature at the point of request through a nonverbal condition, where participants are only shown what Puppy did – and in this case, they now reward him just as adults do. When there is both an utterance and a picture, there is no change in performance from the utterance-only condition. The question, then, is whether a picture-matching task like this is more like the picture-only condition – remembering that, in my task, participants have not heard the stronger term *immediately* before – or more like the picture-and-words condition. Furthermore, the fact that there is not the same effect for ad hoc entailment trials – they do not improve in the picture-only condition – is intriguing (and not discussed by the authors). Could it be that children are indeed deriving an inference at the point of request, but are more willing to cancel it in the case of scalars rather than ad hocs? More work is needed to find out whether children do use a contrastive inference strategy for quantity and relevance inferences, before ascribing it as an explanation to the findings here.

In my study, the youngest age group are not above chance performance when it comes to scalars, and a histogram of child scores (Figure 2.9) reveals that this reflects a modal score of 50%, rather than a bimodal distribution, as in some other studies (Guasti et al., 2005; Skordos & Papafragou, 2016): either children are largely unable to derive the implicated meaning, and so are guessing, given that the literal meaning of *some* matches both pictures, or they are not attending to the quantifier, which also leaves them having to guess.

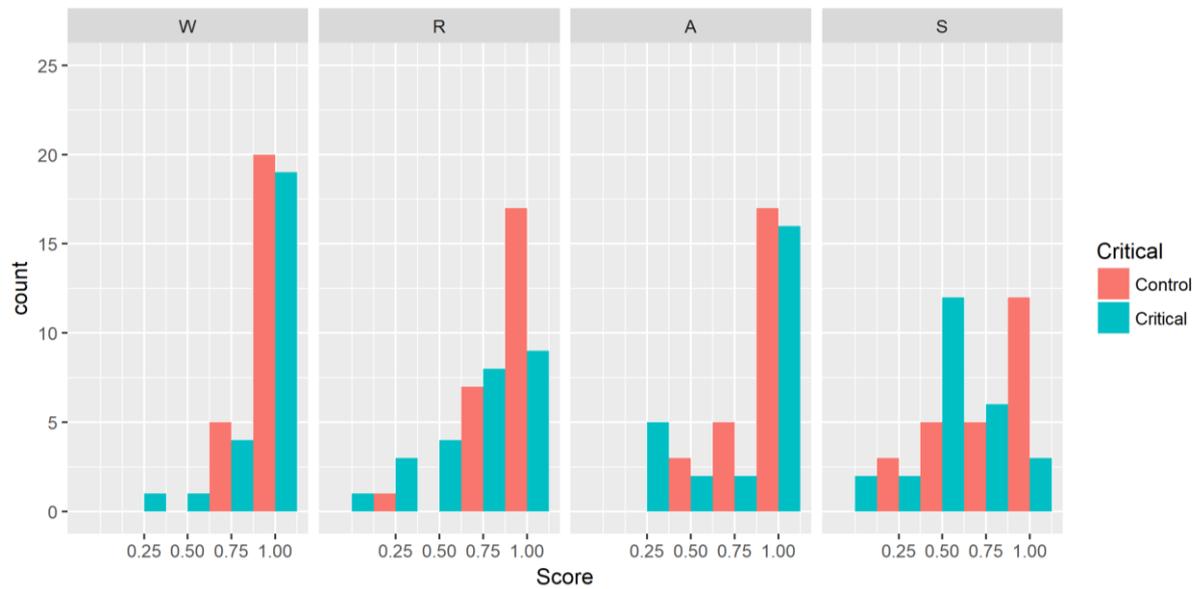


Figure 2.9 Histogram of scores by inference type and condition, for 2;8-3;11

This age group do seem, however, to be largely on the way to acquiring the stronger alternative ‘all’ (with 100% the modal score for control trials), which suggests that not all the children are simply ignoring the quantifier. This pattern could fit in with either of two accounts of scalar implicature acquisition. On the lexical scales account, children here may have learnt the literal meaning of terms on the scale but not how they relate to the scale, so they are not yet able to derive scalar inferences (e.g. Barner et al., 2011). On a more general pragmatic account, they have acquired the scalar terms, but are unable to generate the relevant alternative in this context (Skordos & Papafragou, 2016). This is likely, as the context statement and question preceding the critical utterance do not highlight quantity.

2.5.2.4 Word learning by exclusion

The youngest children in this study certainly excel in WLE inferences, confirming that this is a strategy that children develop very young, perhaps even in the second year of life (e.g. Graham et al., 1998; Halberda, 2003; Markman et al., 2003). Precisely because of this high performance, this study cannot offer much firm evidence for the nature of the WLE inference – whether it is a pragmatic inference, akin to an implicature derivation, or a low-level word learning heuristic. However, examining the youngest age group in the study (2;8-3;11) reveals a positive association between ad hoc and WLE inferences, when language (ad hoc control score) was partialled out, but

provides no evidence for a relationship between relevance and WLE.

On a pragmatic view of WLE, the association between WLE and ad hoc inferences could be explained by the shared inferencing process: both include reasoning with reference to the speaker's intentions about what the speaker could have said, but did not, and excluding these alternatives to derive the speaker's intended meaning. In this case, children could be doing particularly well in WLE in this study overall because the first four stories involving quantity and relevance inferences are effectively pragmatic training that provides relevant practice in inference making and reasoning about speaker intentions. Remember that the word learning trials were placed in a single block which was always presented last, in order that the puppet's use of novel words and choice of novel items did not affect the participant's view of him as co-operative and rational.

On a lexical heuristic view of WLE, one could argue that the association between ad hoc and WLE inferences is not a result of their shared *pragmatic* reasoning, but instead the shared mechanisms of reasoning by exclusion, which, at least in the case of WLE, may not involve reasoning about speaker intentions. The fact that no association with relevance was found – which crucially is pragmatic but does not involve reasoning by exclusion – supports this explanation. Children would then be doing well in WLE overall because of domain general factors like being engaged in a naturalistic task, and because the lexical heuristic is available even in the second year of life – well before any current evidence for implicature derivation.

A future study into the nature of WLE that adopts this promising approach of comparing a clearly pragmatic phenomenon like implicature with word learning inferences therefore needs to a) test children at a younger age where there is more variability in their abilities, and b) consider the influence that each inference type might have on the other. Testing a younger age group, though, requires a type of implicature that at least some children have acquired. One option might be ad hoc implicatures, if indeed balancing visual salience improves performance; another could be the kind of nonverbal indirect request demonstrated by Schulze and Tomasello (2015). Counterbalancing the order of the two conditions would also reveal whether there is more of a training effect from word learning to implicatures or vice versa. Another option could be to compare the training effect on WLE inference rates of a

clearly pragmatic task (like relevance implicature derivation) and an unrelated one.

2.5.3 Improvements to task

I have already raised several issues with the task design in the course of discussion, as well as some possible improvements, and so here I simply summarise and add some improvements to the task which could be used to follow up on my findings.

In order to address the question of the nature of WLE by comparing it with a pragmatic phenomenon, there are two possibilities:

- i. Use an inference that is available at a younger age, such as nonverbal indirect requests (Schulze & Tomasello, 2015) and compare this to WLE, in counterbalanced block order
- ii. Use a training study design to look at difference in performance before and after pragmatic training, compared to an unrelated task

To examine the developmental trajectory of and relationships between different inference types, a new study could ensure that:

- i. Ad hoc trials have equally salient images
- ii. Scalar trials involve a question that more explicitly involves a quantity QUD, and avoids over-informative repetition of the nominal phrase
- iii. Relevance trials are normed for world knowledge at the youngest age of testing
- iv. There is a larger sample and more trials per condition

In addition, inference types could be tested in separate blocks (to avoid shifting the QUD), although this compromises the more naturalistic design. Finally, minimally different trials could be used across inference types, as in Figure 2.10 (inspired by Horowitz & Frank, 2015), although this itself might introduce a new challenge, as a changing QUD is not indicated by the image type.

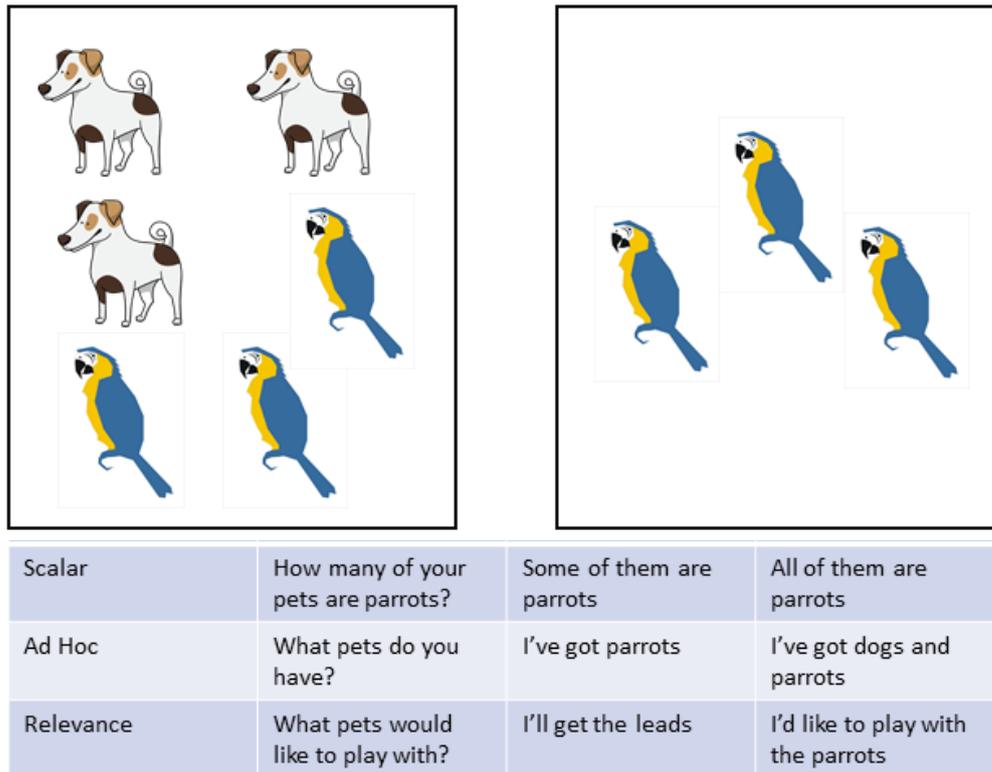


Figure 2.10 Example of improved stimuli for implicatures task

2.6 Conclusion

In a study that combined four inference types in one testing session and in a single experimental paradigm – a picture-matching study with 2;8 to 5-year-olds – I corroborated developmental trends previously seen across different studies: WLE is in place early, followed by ad hoc and relevance inferences, and finally scalars. This pattern also fits in with the computational complexity of each type, viewed within a Gricean framework, although I found no evidence for relevance inferences emerging before quantity ones, as predicted. There is some evidence of patterning together in development across scalar and relevance, as might be expected from their shared features. There is likewise some evidence of relationship between WLE and ad hoc inferences in children aged 2;8–3;11, which could either be indicative of a pragmatic basis for WLE, or due to the reasoning by exclusion evident in both (for WLE see Halberda, 2006). The question of how much pragmatic reasoning is involved in children's word learning strategies, including WLE, therefore invites future research, potentially in the spirit of the study presented here.

3 Factors in implicature development

The development of pragmatic skills – including implicature comprehension – does not, of course, happen in isolation, but in interaction with children’s cognitive and social development, and their environment. To date, those studies looking at children’s development of implicatures within the stream of experimental pragmatics have largely focussed on *when* children acquire certain abilities, possibly to test theory-driven hypotheses. This is certainly useful, but looking, in addition, at pragmatic development in the context of the child might not only reveal something about how children learn to comprehend implicated meaning but, further, eventually have implications in educational or clinical settings.

In this chapter, I examine the relationship between implicature and word learning inferences with aspects of cognitive and linguistic development, namely Theory of Mind and structural language knowledge, and with two aspects of life experience – speaking one or more than one language and socioeconomic status (SES). The latter two have been widely demonstrated to be associated with developmental patterns in other areas of language, like vocabulary and syntax – so what about pragmatics? And how do structural language skills relate to pragmatics in development? And, as we shall see, Theory of Mind is implicated in (post-)Gricean pragmatic theory, so is there empirical evidence?

This chapter constitutes an exploratory study to begin to address such questions. It is exploratory, firstly because existing pragmatic theory and evidence do not give rise to clear-cut predictions, and, secondly, because the data was collected primarily for the study presented in Chapter 2, rather than to target any one of these associations. However, it is amenable to exploring these issues as well, and I follow up one of the interesting observations – on the association between Theory of Mind and implicatures in development – in Chapter 5. I first outline some relevant evidence for each cognitive, linguistic and environmental factor, before presenting the measures, analysis and findings.

3.1 Background

As mentioned, in general there seems to have been little direct focus on these issues. However, various outlines for ‘prerequisites’ for pragmatics have been put forward – that is, those skills that children are supposed to need, and therefore to have acquired previously, to understand something like a quantity implicature. For instance, for Tomasello (2003) word learning is largely a pragmatic phenomenon, and he lists the ‘prerequisite processes’ of segmenting speech and conceptualising referents as well as the ‘foundational processes’ of joint attention, intention-reading, and cultural learning (see, too, Stephens & Matthews, 2014 for pragmatic development in particular). To these can be added for implicature: expectations of cooperativeness, tracking common ground, and, in the case of quantity inferences, generating alternatives which might involve domain restriction and conceptualisation of contrasting sets (Katsos, 2014), as well as structural language and world knowledge (Katsos & Wilson, in prep.). Some of these are taken as self-evident, given the adult experience of implicatures; some are derived from pragmatic theory; and some are supported by empirical evidence from acquisition – though few studies do this directly for implicatures.

In this section, I present an overview of how the four factors that are the focus of this chapter may interact with children’s pragmatic development, and specifically with implicatures. I discuss Theory of Mind, which will be important for Chapter 5 as well, before turning to structural language knowledge, multilingualism and socioeconomic status, in turn.

3.1.1 Theory of Mind

First, I explain briefly what is meant by Theory of Mind (ToM) or mindreading;¹¹ then I show how it is implicated in Gricean approaches to pragmatics; and finally, take

¹¹ Various terms are employed across the psychological, linguistic and philosophical literature, including Theory of Mind, mindreading, folk psychology and social cognition, each reflecting a particular stance (see, for instance, Apperly, 2010, on the use of ‘mindreading’ or Carpendale & Lewis, 2015, for ‘social cognition’). Here, I predominantly use the term ‘Theory of Mind’, reflecting the major use in linguistic pragmatic approaches to child development, and the fact that debate has often centred on acquisition of a ‘theory’ as indicated by passing a False Belief test.

a look at studies that have looked at their relationship in children.

Theory of Mind is a much-researched concept in cognitive psychology as well as philosophy. Simply put, it constitutes ‘the capacity to attribute mental states to oneself and to others, and to reason on the basis of this information in order to interpret and predict others’ behaviors’ (Zufferey, 2010: 6). These mental states may involve beliefs, desires, intentions and affect. Apperly (2010) explains how there have been two basic approaches: ‘theory-theories’, where ToM ‘depends on us having mental state concepts and principles that describe their interactions’, and ‘simulation theories’, where ‘we use our own minds to model (i.e., simulate) those of others’ (2010: 5). From psychological and philosophical perspectives, questions of the nature and development of ToM are very much still open.

In practice, the focus has often been on the development of ToM in children, assuming that at a certain point children ‘get’ the necessary concepts, and attempting to identify when and how. On a modular view of its development, there are several precursors to full ToM, like gaze-tracking, shared attention, and intention detection (Baron-Cohen, 1995; see also Tomasello, 2003). There are even aspects of reasoning about others’ epistemic states that emerge early, like level 1 perspective-taking, which I will return to at much more length in Chapter 5 (San Juan, Khu, & Graham, 2015). Furthermore, second-order ToM tasks, as well as combining conflicting aspects of another’s mind like belief and desire, emerge later in childhood (Apperly, 2010).

However, the gold standard test for Theory of Mind has often been taken to be the ability to reason about others’ beliefs, where an agent’s epistemic state does not match reality, as measured by False Belief tasks. The Sally-Anne, or Change of Location, task (Baron-Cohen, Leslie, & Frith, 1985; Wimmer & Perner, 1983), requires children to track the beliefs of two protagonists about where an object is, and, crucially, realise that one has a false belief about the object’s location, when it has been moved in her absence. In the Unexpected Contents, or Smarties, task (Perner, Leekam, & Wimmer, 1987), children are shown a tube of sweets, only to find out – what a disappointment – that it instead holds pencils or paperclips; they are then asked what they thought was in it, what really is in it, and what a friend will think is in it. Robustly, cross-culturally, children under 4 years tend to fail these tests (Wellman, Cross, & Watson, 2001).

Why does this matter for pragmatic development? Since the dawn of the ‘Gricean era’, Theory of Mind has been implicated, firstly, in pragmatic communication in general, and, secondly, in implicature inferencing, specifically.

Grice (1957) introduced the notion of ‘non-natural meaning’, which, in the very way he elaborates it, assumes that speakers reason about their own beliefs and intentions, and the hearers’ beliefs: ‘I must intend to manipulate or add to my audience’s mental representations in some way... The audience must recognise that I have these intentions... I should intend that my audience believes it, and they should believe it at least in part because they recognize that this was my very intention’ (Scott-Phillips, 2014: 22–23). Scott-Phillips (2014) forcefully argues that for such *ostensive-inferential* communication to take place at all, the hearer has to recognise both the speaker’s informative and communicative intentions, and the speaker has to intend this recognition. He is approaching pragmatics within a Relevance Theory framework (Sperber & Wilson, 1995), but the analysis still stands as an interpretation of Grice’s proposal. This means that *any* communicative act, not just those involving implicatures, irony, reference resolution or other classic ‘pragmatic phenomena’, requires ToM.

Secondly, concentrating on the belief or epistemic state component of ToM, it is clear that it is inherent in Grice’s Co-operative Principle. The maxim of quality – with its supermaxim, ‘try to make your contribution one that is true’ – involves two submaxims – ‘do not say what you believe to be false’, and, importantly here, ‘do not say that for which you lack adequate evidence’ (Grice, 1975). Further, it involves the supposition, on the part of both the speaker and the hearer, that the speaker is adhering to the Conversational Principle, and its ensuing maxims. Under normal circumstances, then, the hearer assumes that the speaker does have adequate evidence for the utterance – that they are knowledgeable (on the relevant matter). This in itself implies an aspect of epistemic state tracking on the part of the hearer. Further, Grice introduces the notion that maxims may clash with one another, so that the speaker is unable to fulfil both, giving an example where the inference itself concerns the epistemic state of the speaker:

A: Where does C live?

B: Somewhere in the South of France.

→ The speaker does not know in which town C lives.

(Grice, 1989: 32)

Here the hearer has to suppose that the speaker cannot be more informative (and therefore follow the maxim of quantity), if she also wants to follow that second submaxim of quality, and therefore there is no quantity implicature from the apparent under-informativeness – the hearer does not conclude that the speaker intends that C does not live in a specific place, but instead that the speaker is not sure. This invites us to think about the inverse case: where hearer and speaker both know that the speaker lacks evidence for the relevant utterance, so there is no need to make an inference about the speaker's epistemic state. What might happen with the quantity implicature?

Theorists following Grice have fleshed out Grice's proposals, concentrating on the two key elements of reasoning involved in implicature derivation: the Competence Assumption (Geurts, 2010) and the Epistemic Step (Sauerland, 2004). Consider the following example:

1. The speaker said that the puppet picked the card with rabbits.
2. Given what the speaker said, the puppet may have picked the card with rabbits only or the card with rabbits and ducks (or other relevant alternatives).
3. Given that the QUD is which card the puppet picked, and that we both assume we are being informative (following the Co-operative Principle), the speaker is giving as much information as is relevant and true.
4. If the speaker knew that the puppet picked the card with rabbits and something else, he would have said so.
5. The speaker did not say so, therefore he does not know whether the puppet picked the card with rabbits and something else.

6. However, assuming that the speaker does know all relevant information, specifically whether the puppet picked the card with rabbits only or the card with rabbits and something else...
 7. Given (6), and that the speaker did not say anything more than ‘the puppet picked the card with rabbits’, the speaker means that the puppet did not pick up the card with rabbits and something else.
 8. The speaker means that the puppet picked the card with only rabbits.
 9. And he intends me to reason in this way.
- (Adapted from Breheny, Ferguson & Katsos, 2013: 424; example based on Kronmüller, Morisseau & Noveck, 2013)

Here, the Competence Assumption is in step (6) – that the speaker is fully knowledgeable with regards to the QUD. The Epistemic Step is then in step (7): from the ‘weak’ inference that the speaker does not know whether X, or X and Y, to the strong inference that the speaker knows that only X and nothing else (to use Geurts’, 2010, term; the ‘weak’ inference is also known as an ignorance inference).

Later pragmatic theorists have diverged on this view of implicature interpretation. Levinson (2000) focusses on so-called Generalised Conversational Implicatures, where the form of the utterance provides a short-cut to its intended meaning. For these, he suggests, an aspect of the context such as the speaker’s epistemic state can cancel an implicature, but it can never stop it being inferred in the first place. On the other hand, a different view comes from theories that treat scalar implicature in particular as a grammatical phenomenon (e.g. Chierchia, 2004; Chierchia et al., 2008): in this case, the context – and particularly semantic information in the utterance, like upward- or downward-entailment – contributes to the disambiguation of the utterance, namely whether it contains an exhaustivity operator or not. This means that the speaker’s epistemic state may not play a role in reasoning about what was said, as above, but it does still come into consideration at the point of interpretation, with a grammatical exhaustivity operator or not (Chemla & Singh, 2014b, 2014b). On the other hand, other Neo-Gricean approaches see the speaker’s epistemic state as licensing or suspending an implicature (e.g. Breheny et al.,

2013: 424) or one of many factors that constrain pragmatic inferencing in a probabilistic model (e.g. Degen & Tanenhaus, 2014).

The key point, therefore, is that – whatever the exact mechanism – all pragmatic theories following from Grice’s original postulations expect that the hearer takes into account the speaker’s epistemic state when deriving Particularised Conversational Implicatures, such as ad hoc quantity implicatures; most agree that this is also the case for Generalised ones, too, such as scalar implicatures, with the exception of Levinson’s (2000) Default view.

This brings me to the conundrum: pragmatic theory has overwhelmingly implicated ToM in communication. ToM, indicated by passing traditional False Belief tasks, emerges around age 4. Some implicatures, like ad hocs and relevance inferences, appear earlier, and other pragmatic inferences – possibly WLE – earlier still. How can this be reconciled? As Scott-Phillips (2014) puts it, it seems that either pragmatics does not involve ToM, or infants have ToM. Actually, there are two flavours to the first solution: a developmental and a life-long one. It could be that children look as if they are doing ostensive-inferential communication, but really they are using some other strategy that they will grow out of (Breheny, 2006). Alternatively, adults have more than one strategy available to them, not all involving full ToM (Jary, 2013, 2013; Katsos & Andrés-Roqueta, 2015; Kissine, 2016). I will return to these options in Chapter 5.

This conundrum is obvious to the pragmatician – leading to the proposal of these possible workaround solutions. I wonder, though, whether cognitive psychologists and philosophers might not be so perplexed: after all, what ToM is and how it develops is still a matter of debate (Apperly, 2010). There are other plausible solutions, based on the signs of *early* emergence of ToM, and its *gradual* development. It could be, as well, that those children who are performing well in implicature tasks at age 3 also happen to be early ToM passers too. This can be tested in a correlational study such as I present here.

A few studies have investigated the link between pragmatic competence and ToM, and have mostly found some sort of association: Veenstra (2010) found a correlation between reaction time for detecting under-informativeness and ToM in 4–8-year-olds (but not with under-informativeness scores themselves); Filippova and

Astington (2008) find that explicit justification for advanced ToM response predicts irony comprehension (together with vocabulary); and Gollek and Doherty (2016) observe an association between performance in a word-learning task with a pragmatic cue (indicating that the referent of a novel word is the familiar object) and ToM, partialling out vocabulary score. No published study to date has looked at ToM with implicature comprehension, however, despite implicature comprehension being a key pragmatic competence.

3.1.2 Structural language

By structural language I mean syntax and vocabulary knowledge. There are two ways this could be related to implicature inferences in development: *specifically* related to utterances that trigger implicatures, and *generally* related to pragmatic development.

Implicatures, as I have defined them in this study, are verbal and linguistic – in the sense that speakers use the conventional code of language in producing an utterance implicating their intended meaning. To infer any given implicature, then, hearers have to comprehend the literal content of the utterance. (In saying this, I do not mean that actual processing has to take place by first interpreting the literal content, then deriving the implicated meaning, as steps in Gricean reasoning are elucidated; instead, top-down anticipation of the implicated meaning may be taking place, but at least at some point or sometimes, the literal content will still play a role.) Moreover, for some types of implicature, rather specific lexical or syntactic knowledge is required. In the previous chapters, I have mentioned diverging accounts of quantity implicatures, with some taking scalar implicatures as a special case for which the trigger word in the utterance (e.g. ‘some’) and its alternatives (‘most’, ‘all’), sit on a scale, which is part of lexical knowledge (e.g. Chierchia et al., 2008). This scale then has to be learned, and children may struggle with scalars because they have learned the alternatives as lexical items but not how they form a scale (Barner et al., 2011).

However, even ad hoc quantity implicatures require lexical knowledge of alternatives – what the speaker could have said but did not, in the context – even if this is simply knowing that a strawberry is a ‘strawberry’, not an ‘orange’, or that ‘mother’ and ‘father’ equate to ‘parents’, in the following example:

A. Did you meet his parents?

B. I met his mother.

→ i. I met only his mother, not his father.

→ ii. I did not meet his parents, instead I met his mother

Without this lexical knowledge, only a contrastive inference is available, as I discussed in relation to scalars in the previous chapter. In this case, though, it would lead to the (actually incorrect) inference (ii). Likewise, for manner implicatures, knowledge of syntactic constructions may be important, for instance, knowing that ‘make the door close’ has the lexicalised alternative ‘close the door’ (an issue I cover in Chapter 4). For WLE inferences, this idea has been tested empirically. Productive knowledge of the label for the known object does predict the rate of exclusion inference: in Grassmann, Schulze and Tomasello’s (2015) study, productive knowledge of the competitor is a better predictor of using a WLE strategy than age in German-speaking 2-, 3- and 4-year-olds.

Secondly, structural language knowledge may be related to pragmatic inferences in development in a more general way – and this is what I am interested in for this study. This means a relationship between total lexical or syntactic knowledge and implicature abilities. This might be an extrapolation of the specific relationship I have just discussed: if a child has a larger vocabulary in general, it is more likely that he will know or have productive knowledge of the relevant alternatives for any given utterance, as well as comprehending its literal content. It might also be a consequence of a child’s language development trajectory in general: one might expect that structural language would facilitate pragmatic development or processing. That is, the more structural language knowledge children have acquired, the more they can access at least some meaning in context, and the more possibility they have to learn how expectations of co-operativeness function in conversation. Alternatively, on a usage-based, pragmatics-first account of language acquisition, better pragmatic skills would facilitate lexical and syntactic acquisition, which in turn make more pragmatic strategies available (Tomasello, 2003). In either case, an association between structural language and pragmatics is predicted.

In studies that have examined this relationship to date, structural language does

turn out to be a predictor of competence deriving implicatures. Antoniou & Katsos (2017) test structural language (in this case only vocabulary, receptive and expressive, plus sentence comprehension scores from control trials) as well as quantity, manner and metaphor inferences, in 6- to 9-year old speakers of Greek, Cypriot Greek and English. They find that language was a predictor of implicature performance for monolinguals and bilinguals (but not bilinguals, for whom only age was a predictor). Similarly, Davies, Andres-Roqueta & Norbury (2016) find that for both children with Specific Language Impairment and typically developing children (Spanish-speaking, aged 5–11 years), language skills predict implicature performance: in the typically developing group, only receptive grammar is a predictor of production of optimally informative referential utterances, whereas in the Specific Language Impairment group sentence recall and vocabulary scores are also predictive. There also turns out to be a relationship between WLE and overall lexical knowledge: vocabulary score is a predictor of WLE, whether receptive or productive. For instance, Kalashnikova, Mattock and Monaghan (2016) find that 17- to -19-month-old English-speaking children with larger receptive vocabularies use a WLE strategy more than their peers with smaller vocabularies (see too Graham et al., 1998; Mervis & Bertrand, 1994; Suanda & Namy, 2013).

The focus of this study is to examine whether this relationship exists for implicature understanding in younger children, as well as to look at the link between WLE and vocabulary in older children than previously tested. It is beyond the scope of a correlational study to establish the directionality of the link – whether structural language affects pragmatics, or vice versa, or indeed whether the influence is mutual.

3.1.3 Socioeconomic Status

Socioeconomic status (SES), as well as multilingual development, has connections to ToM and structural language, as well as – potentially – pragmatic inferencing. Indeed, there is a vast body of research on the relationship between SES and cognitive development, as well as between SES and language acquisition in general, with its important ramifications for education and welfare policy.

SES is a complex concept, referring to ‘one’s access to financial, educational, and social resources, and the social positioning, privileges, and prestige that are derived

from these resources' (Pace, Luo, Hirsh-Pasek, & Golinkoff, 2017: 287). It is typically measured by a proxy, such as income, income–need ratio or maternal education, the latter of which has been the most frequently used as the most reliable predictor for child outcomes (Hoff, Laursen, & Bridges, 2012). Many studies report a connection between SES and child language development, particularly concentrating on vocabulary (for a US focussed overview see Hoff, 2006; for a UK study see Locke, Ginsborg, & Peers, 2002). Striking differences are observed between SES groups, with lower SES children lagging behind their higher SES peers (although note that this effect may be partly a result of the test design favouring middle-class children: E. V. Clark, 2009; Hoff, 2006).

When it comes to pragmatics, the evidence to date is much more scant: social pragmatic abilities like narrative skills may or may not be associated with SES, depending on assessment criteria (Pace et al., 2017: 287), although there is evidence for a difference in early gesture contributing to later SES differences in vocabulary (Rowe & Goldin-Meadow, 2009). In experimental pragmatics studies, SES is typically not mentioned, or only to explain an assumption of middle-class participants (e.g. Diesendruck et al., 2010; Y. T. Huang & Snedeker, 2009). However, Antoniou and Katsos (2017) do measure SES through the Family Affluence Scale (Boyce, Torsheim, Currie, & Zambon, 2006) and parental education, and do not find any evidence for a relationship with implicature comprehension. There is also some indication that SES and ToM may be associated (e.g. Cutting & Dunn, 1999, but Hughes, Deater-Deckard, & Cutting, 1999, did not replicate the effect).

The reasons for the relationship between SES and language acquisition are likely to be multi-faceted and complex, and, as Pace and colleagues (2017) point out, have received less attention from a psycholinguistic approach. They identify three non-mutually exclusive possible factors where more research is needed: differences in processing; differences in quality of input; and differences in learning materials available. Multilingualism adds an extra layer, as these factors may differ across a child's languages (Scheele, Leseman, & Mayo, 2010). This means that the predictions for pragmatics are not yet clear. From the empirical evidence available, it could be that pragmatics patterns with other areas of language acquisition, like vocabulary and grammar, in terms of SES, especially as children's pragmatic performance tends to

correlate with their structural language. However, it could also be the case that SES has no additional effect on pragmatic skills, once structural language is controlled for – much like multilingualism.

3.1.4 Monolingualism vs multilingualism

Another aspect of children's lived experience – and cognitive development – is the number of languages they grow up with. The debate over the 'bilingual advantage' for cognitive function has become complex and nuanced (Bak, 2016). For structural language, and particularly vocabulary, there has been a long-reported effect that bilingual children's vocabulary size or acquisition rate in each language tends to be lower than monolinguals' (e.g. Bialystok & Feng, 2011; Bialystok, Luk, Peets, & Yang, 2010; Hoff, Core, et al., 2012). However, this general finding is now being questioned, with factors such as sequential versus simultaneous bilingualism, or other variations in input, being implicated in any differences, rather than bilingual exposure per se. Whether one or both languages are tested, whether receptive or productive vocabulary is tested, and other variation in testing methods may also contribute to findings of difference. For instance, De Houwer, Bornstein and Putnick (2014) do not find evidence for a difference in matched samples, where bilinguals were exposed to rich input in both languages from birth. Similarly, Bialystok and colleagues (2010) compare vocabulary related to the school context and to the home context, and do not find differences between bilinguals and monolinguals for the school-related words (where they have similar exposure), but did for home-related words (where bilinguals have most exposure in the home language).

For pragmatic abilities, the evidence is similarly mixed, depending on the skill being examined. Some studies employing judgement tasks for sensitivity to Gricean maxims find a bilingual advantage, across an age range of 3–10 years (Foppolo, 2015; Siegal et al., 2010, 2009; Siegal, Matsuo, Pond, & Otsu, 2007; Stateva et al., 2015). For instance, in Siegal and colleagues' (2010, 2009) studies, children listen to two puppets respond to a question, a co-operative and an uncooperative one (in the Gricean sense of being informative and relevant), and have to identify which puppet says something silly. Others testing implicature comprehension directly via picture-selection tasks have not found evidence for a difference between monolingual and bilingual children,

once language and SES are controlled for, in older children aged 6–9 years and 10–13 years (Antoniou & Katsos, 2017; Antoniou, Veenstra, Katsos, & Kissine, 2016). In addition, there is some evidence for an advantage for bilingual children in ToM abilities (Goetz, 2003; Kovács, 2009); repairing miscommunication (e.g. Wermelinger, Gampe, & Daum, 2017); using ‘pragmatic’ cues like eye-gaze or pointing for comprehension (Brojde & Colunga, 2011; Yow & Markman, 2011b); and sensitivity to speaker emotion (Yow & Markman, 2011a). Given the paucity of the studies for any one pragmatic phenomenon, the relationship between multilingual acquisition and implicature skills, though, is still an open question.

Turning to WLE in particular, there are two additional factors that come into play. Firstly, as mentioned, bilinguals tend to have smaller vocabularies in any one language, associated with use of WLE strategy, as I explained above. Secondly, bilinguals have more linguistic experience of overlapping labels – they may know a label in both languages. Some studies have found that bilinguals fare worse on WLE tasks (Bialystok, Barac, Blaye, & Poulin-Dubois, 2010; Davidson, Jergovic, Imami, & Theodos, 1997; Davidson & Tell, 2005; Healey & Skarabela, 2009; Houston-Price, Caloghiris, & Raviglione, 2010), and factors like the precise contents of bilinguals’ vocabularies might also be important (Byers-Heinlein & Werker, 2013). Some studies have also found different developmental pathways for monolingual and bilingual children, with bilingual children using WLE less as they get older (Davidson & Tell, 2005; Kalashnikova, Mattock, & Monaghan, 2015). Kalashnikova and colleagues (2015) used a variation of a word-learning task in which two puppets gave different labels to the same novel object, as well as the standard task: their older bilingual children (mean age 4;11) were worse at WLE but better at accepting lexical overlap than their monolingual peers. Extant studies therefore give a stronger indication for WLE than implicatures – and possibly in the opposite direction – that multilingual acquisition is negatively associated with WLE, whereas it may be positively associated with implicature inferencing.

3.2 Experiment 1B: the relationship of implicatures with structural language, SES, ToM and languages spoken

3.2.1 Method

3.2.1.1 Participants

Participants were a subset of the group recruited for Experiment 1A: of the 71 monolinguals and 35 bilinguals not excluded for other reasons, 58 monolinguals and 26 bilinguals completed all parts of the test sessions and their caregivers returned the background questionnaire, and they were included in the analysis in this chapter.

The unequal numbers of monolingual and bilingual children reflect the predominantly monolingual population of the area: in 2013, 10.6% of primary school children in Cambridgeshire had a first language other than English (Department for Education statistics, 2016).

3.2.1.2 Materials

Implicature task

The data from the implicature task presented in Chapter 2 were used, except that for this analysis, inference types were split into two groups: the first contained scalar, ad hoc and relevance implicatures; and the second, WLE. Also, only the critical trials are included in this analysis.

British Picture Vocabulary Scale (BPVS)

The BPVS-3 (Dunn et al., 2009) was used to test receptive vocabulary. It is a picture-matching task, in which children are asked to point to one of four pictures that matches the word the experimenter says. The pictures are presented in a colour flip-book. The items are arranged in blocks of 12, and the test continues until children have made 8 or more mistakes in a block. Children received one warm-up trial. The raw scores, rather than standardised scores, were used in the analysis, as the comparison of interest is an individual's vocabulary and inferencing ability, rather than vocabulary on a normal distribution for that age.

Mini Test of Receptive Grammar (TROG)

A reduced version of the TROG II (Bishop, 2003) was used, with 20 items (one

from each block of the full TROG, where blocks consist of four items each, with each block testing the same syntactic and semantic competence using different lexical items), plus two warm-up trials. This version was previously used, though not reported, by Reetzke and colleagues (2015). Children were asked to point to the picture that matched what a lady (pre-recorded) says: pictures were presented and recordings played on a laptop computer. Raw scores out of 20 were used. For the analysis, the BPVS and the TROG scores were centred and scaled, and then a mean for each participant calculated, to provide a composite structural language score.

Theory of Mind tests

Two tasks testing False Belief were used: the Change of Location, or Sally-Anne, task (Baron-Cohen et al., 1985; Wimmer & Perner, 1983), and the unexpected contents, or Smarties, task (Perner et al., 1987). The Sally-Anne task was acted out with finger puppets and props (a box, a bucket with a cloth on top, and a marble), and the unexpected contents task was administered using a Smarties tube and three small pencils (Figure 3.1, Appendix 8.4). There was a score of 0 or 1 for the Sally-Anne task, and 0, 1 or 2 for the Unexpected Contents task, making a total score of up to 3.



Figure 3.1 Materials for Sally-Anne test

Languages and SES questionnaire

The background questionnaire (Appendix 8.3) included questions on socioeconomic status and multilingualism. The socioeconomic status questions consisted of a) the Family Affluence Scale (Boyce et al., 2006), and b) parental

education, adapted from the questionnaire used in Antoniou (2015). SES scores for each of the two parts were first centred and scaled, and then a mean calculated for each participant, so that the two components were equally weighted. The multilingualism section included questions on other languages spoken by the child in addition to English, age of exposure, frequency of use, and fluency (scores 0–4), taking inspiration from the Alberta Language Environment Questionnaire (Paradis, 2011) and the Alberta Language Development Questionnaire (see too Antoniou & Katsos, 2017; Paradis, Emmerzael, & Duncan, 2010).

3.2.1.3 Procedure

The BPVS, TROG and ToM tasks were administered in a single session, always in that order. This was the second session with the child (with the implicature task carried out in the first session). The experimenter kept track of BPVS and TROG picture choices on scoresheets, and noted the child's responses in the ToM tasks at the end of the session. The session was also audio-recorded, so recordings could be checked if there was any doubt about the child's ToM response. The child was given a sticker and a thank you certificate at the end of the session. The background questionnaire was sent to parents along with the information and consent form, and returned to the school or nursery.

3.2.2 Results

The implicature scores for bilingual and monolingual groups can be seen in Figure 3.2. Note that they do represent different sample sizes (N = 58 monolinguals, N = 26 bilinguals), and, for the monolinguals, are a subset of the data considered in the last chapter, with scores from only those participants for whom data for all variables is available.

The bilingual group is mixed in terms of the other languages spoken with English (see Table 3.2 Languages spoken by multilingual children), with a range of ages of exposure to English, although less so to the home language (Table 3.1). The exposure and fluency scores suggest overall balanced to English-dominant bilinguals, as would be expected from children in part- or full-time education in UK.

Multilingual participants	21 bilinguals – 5 trilinguals
Fluency of English	3.6/5 mean; 4/5 median; range 1–5
Fluency of Language A	2.4/5 mean; 2.5/5 median; range 0–5
Place of Birth	5/21 born outside UK
Age of Exposure to English	Range 0–52 months, mean 7.5; median 0 (9 children not from birth)
Age of Exposure to Language A	Range 0–24 months (2 children not from birth)
Frequency of Exposure to English	Mean 3.9/5, median 4/5
Frequency of Exposure to Language A	Mean 3.4/5, median 4/5

Table 3.1 Characteristics of multilingual children in complete data subset (N = 26)

Language	Children	Language	Children
Afrikaans	1	Korean	1
Brazilian Portuguese	1	Lithuanian	1
Cantonese	1	Malayalam	1
Czech	1	Marathi	1
Dutch	2	Polish	1
Finnish	1	Portuguese	3
French	3	Spanish	3
Greek	1	Swiss German	1
Hebrew	1	Telugu	2
Hindi	1	Turkish	1
Hungarian	2	Vietnamese	1
Italian	1		

Table 3.2 Languages spoken by multilingual children

The monolingual and bilingual groups do not differ on mean inference score or SES, but they do differ on vocabulary and grammar, such that monolinguals score higher than bilinguals (Table 3.3).

		Mean		Estimate	SE	t	p
Critical score	Monolinguals	85 %	Intercept	.85	.02	47.09	< .001
	Bilinguals	83 %	Bilingual	-0.02	.03	-0.77	.45
SES (centred, scaled)	Monolinguals	-0.016	Intercept	-0.01	.1	-0.16	.88
	Bilinguals	.035	Bilingual	.05	.17	.29	.77
Grammar (TROG raw score)	Monolinguals	12.7	Intercept	12.72	.46	27.87	< .001
	Bilinguals	11.1	Bilingual	-1.65	.821	-2.01	.048
Vocabulary (BPVS raw score)	Monolinguals	76.18	Intercept	76.12	2.35	32.4	< .001
	Bilinguals	66.66	Bilingual	-9.46	4.23	-2.24	.028

Table 3.3 Means and differences between monolingual and multilingual groups for implicature score, SES, Grammar and Vocabulary
lm; dummy coding (monolingual as intercept)

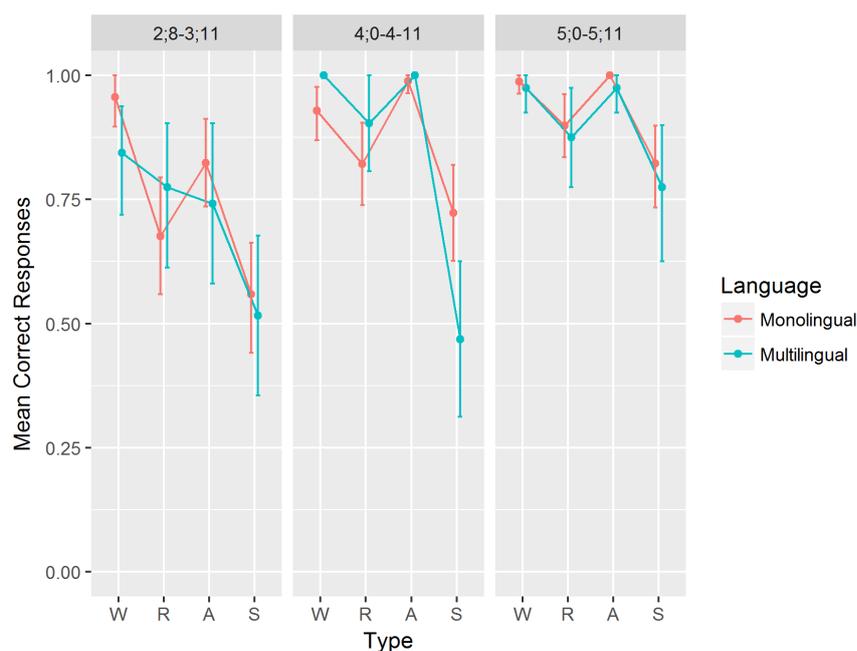


Figure 3.2 Mean correct score for critical inferences, by age and monolingual vs multilingual

	2;8-3;11		4;0-4;11		5;0-5;11	
	Mono-	Multi-	Mono-	Multi-	Mono-	Multi-
WLE	.96	.84	.93	1.00	.99	.98
Relevance	.68	.77	.82	.90	.90	.88
Ad hoc	.82	.74	.99	1.00	1.00	.98
Scalar	.56	.52	.72	.47	.82	.78
N =	17	8	21	8	20	10

Table 3.4 Proportion of correct responses for critical condition only, by inference type, age group and mono- vs multilingual

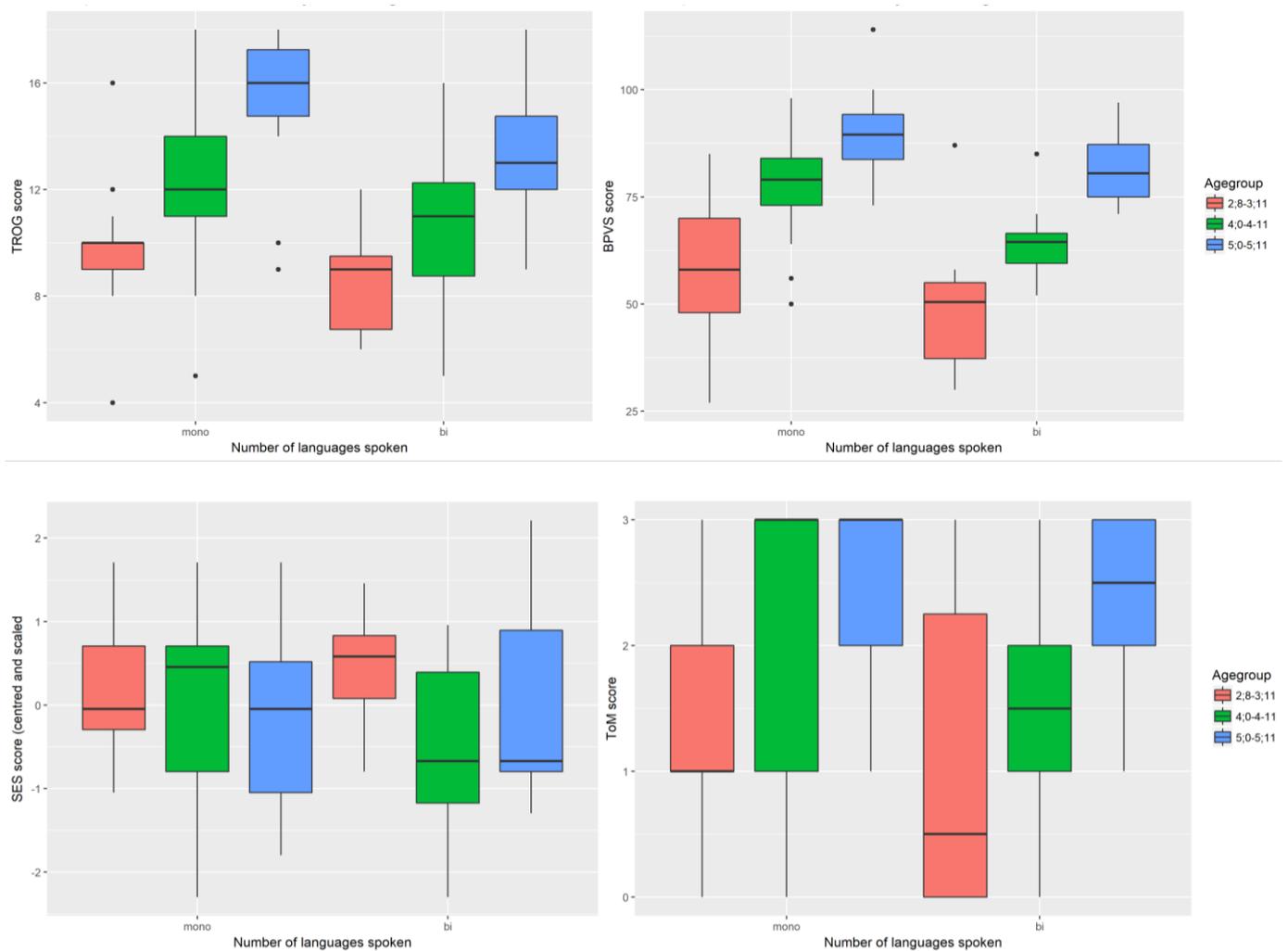


Figure 3.3 Boxplots showing TROG, BPVS, SES and ToM scores (from top, left to right) by age group and monolingual vs multilingual. Note that the age groups are of unequal size – see Table 3.4.

3.2.3 Analysis

To examine the relationship between SES, structural language, ToM, and number of languages spoken, mixed-effects logistic regression models were fitted, with implicature scores in the critical condition (excluding word learning) as the outcome variable. Age, gender, SES, structural language, ToM and number of languages spoken (monolingual/multilingual) were added in turn as fixed effects, and all fixed effects by item random slopes (following Field, Miles and Field, 2012, and Wieling, 2015), using the *lme4* package in R (Bates et al., 2015; R Core Team, 2016; RStudio Team, 2016). Age (in months), structural language, ToM and SES scores were each centred and scaled; Gender and monolingual/multilingual were coded with sum contrasts. The models were built up adding in the fixed effects to an intercept-only model, and then compared using the *anova* function; the fixed effects were added in the order: age, gender, SES, structural language, ToM, monolingual/multilingual.

The factors were added in this order, as I wanted to examine the effect of mono- vs multilingualism, having controlled for the other factors that might vary between the two groups of participants independently of their mono- or multilingualism. Then, I looked at the effect of ToM, accounting for age, gender, SES and language; and finally, language, controlling for Age, Gender and SES.

To look at WLE, structural language and number of languages spoken, I used the BPVS score only (as vocabulary knowledge is clearly associated with WLE, and has been tested in other studies).

3.2.3.1 Language, ToM and number of languages spoken

Age, SES and language significantly improve the model ($\chi^2(1) = 27.93$, $p < .001$; $\chi^2(1) = 4.96$, $p = .03$; $\chi^2(1) = 6.82$, $p = .01$) but not gender (given age), ToM or monolingual/multilingual (Table 3.5).

Model	Df	AIC	Log Lik	Deviance	χ^2	p
Score ~ 1 + (1 + Age + Gender + SES + Language + ToM + Mono/Multilingual Item.no)	29	861.05	-401.52	803.05		
Score ~ Age + (random effects)	30	835.12	-387.56	775.12	27.93	< .001
Score ~ Age + Gender + (random effects)	31	836.79	-387.40	774.79	.33	.57
Score ~ Age + Gender + SES (random effects)	32	833.83	-384.92	769.83	4.96	.03
Score ~ Age + Gender + SES + Language + (random effects)	33	829.01	-381.50	763.01	6.82	.01
Score ~ Age + Gender + SES + Language + ToM + (random effects)	34	829.63	-380.82	761.63	1.38	.24
Score ~ Age + Gender + SES + Language + ToM + Mono/multilingual + (random effects)	35	831.52	-380.76	761.52	.11	.74

Table 3.5 Model comparison for Age, Gender, SES, structural language, ToM and mono- vs multilingualism
glmer, family = binomial, bobyqa optimizer; anova

When the order SES and language are entered in the model comparison is reversed, only language remains a significant predictor of implicature score ($\chi^2(1) = 8.43, p = .004$) – Table 3.6.

Model	Df	AIC	Log Lik	Deviance	χ^2	p
Score ~ 1 + (1 + Age + Gender + SES + Language + ToM + Mono/Multilingual Item.no)	29	861.05	-401.52	803.05		
Score ~ Age + (random effects)	30	835.12	-387.56	775.12	27.93	< .001
Score ~ Age + Gender + (random effects)	31	836.79	-387.40	774.79	.33	.57
Score ~ Age + Gender + Language + (random effects)	32	830.37	-383.18	766.37	8.43	.004
Score ~ Age + Gender + Language + SES + (random effects)	33	829.01	-381.50	763.01	3.36	.07
Score ~ Age + Gender + Language + SES + ToM + (random effects)	34	829.63	-380.82	761.63	1.38	.24
Score ~ Age + Gender + Language + SES + ToM + Mono/multilingual + (random effects)	35	831.52	-380.76	761.52	.11	.74

Table 3.6 Model comparison for Age, Gender, structural language, SES, ToM and mono- vs multilingualism
glmer, family = binomial, bobyqa optimizer; anova

When only monolinguals are considered (with variables centred and scaled for this data subset), the structural language remains the factor which significantly improves the model, once age, gender and SES are taken into account ($\chi^2(1) = 4.53$, $p = .03$) – Table 3.7. For bilinguals, though, the only factor which does so is ToM ($\chi^2(1) = 7.7$, $p = .01$) – Table 3.8.

Model	Df	AIC	Log Lik	Deviance	χ^2	p
Score ~ 1 + (1 + Age + Gender + SES + Language + ToM Item.no)	22	609.62	-282.81	565.62		
Score ~ Age + (random effects)	23	582.00	-268.00	536.00	29.62	< .001
Score ~ Age + Gender + (random effects)	24	583.79	-267.90	535.79	.21	.65
Score ~ Age + Gender + SES (random effects)	25	582.14	-266.07	532.14	3.65	.06
Score ~ Age + Gender + SES + Language + (random effects)	26	579.61	-263.81	527.61	4.53	.03
Score ~ Age + Gender + SES + Language + ToM + (random effects)	27	580.98	-263.49	526.98	.63	.43

Table 3.7 Model comparison for Age, Gender, SES, structural language, and ToM for monolinguals
glmer, family = binomial, bobyqa optimizer; anova

Model	Df	AIC	Log Lik	Deviance	χ^2	p
Score ~ 1 + (1 + Age + Gender + SES + Language + ToM Item.no)	22	308.17	-132.09	264.17		
Score ~ Age + (random effects)	23	305.42	-129.71	259.42	4.75	.03
Score ~ Age + Gender + (random effects)	24	307.24	-129.62	259.24	.17	.68
Score ~ Age + Gender + SES (random effects)	25	308.82	-129.41	258.82	.43	.51
Score ~ Age + Gender + SES + Language + (random effects)	26	307.70	-127.85	255.70	3.12	.08
Score ~ Age + Gender + SES + Language + ToM + (random effects)	27	302.00	-124.00	248.00	7.70	.01

Table 3.8 Model comparison for Age, Gender, SES, structural language, and ToM for bilinguals
glmer, family = binomial, nlminbw optimizer; anova

Indeed, in a partial correlation (Kendall's τ) with SES and structural language, controlling for age, there is a small positive correlation for monolinguals ($\tau = .15$, $p < .001$, $z = 6.02$), but not for bilinguals ($\tau = -0.11$, $p = .002$, $z = -3$).

3.2.3.2 Word learning, vocabulary and number of languages spoken

To examine the relationship between WLE and vocabulary score and speaking more than one language, mixed effects logistic regression models were fitted, with age, gender, SES, vocabulary score (BPVS) and monolingual/multilingual as predictors, and item by age, gender, SES, vocabulary and monolingual/multilingual random slopes, and the models compared with the anova function. Only vocabulary is a significant predictor ($\chi^2(1) = 10.3$, $p = .001$) – Table 3.9.

Model	Df	AIC	Log Lik	Deviance	χ^2	p
Score ~ 1 + (1 + Age + Gender + SES + Vocabulary + Mono/Multilingual Item.no)	22	156.01	-56.00	112.01		
Score ~ Age + (random effects)	23	156.21	-55.10	110.21	1.80	.18
Score ~ Age + Gender + (random effects)	24	158.06	-55.03	110.06	.15	.70
Score ~ Age + Gender + SES+ (random effects)	25	156.63	-53.32	106.63	3.43	.06
Score ~ Age + Gender + SES + Vocabulary + (random effects)	26	148.33	-48.17	96.33	10.30	.001
Score ~ Age + Gender + SES + Vocabulary + Mono/multilingual + (random effects)	27	148.65	-47.33	94.65	1.68	.20

Table 3.9 Model comparison for Age, Gender, SES, structural language, and ToM glmer, family = binomial, optimx.L-BFGS-B optimizer; anova

For monolinguals only, this pattern remains, with vocabulary significantly improving the model ($\chi^2(1) = 12.5$, $p = .0004$) – Table 3.10; for bilinguals, only SES significantly improves the model ($\chi^2(1) = 12.37$, $p = .0004$) – Table 3.11.

Model	Df	AIC	Log Lik	Deviance	χ^2	p
Score ~ 1 + (1 + Age + Gender + SES + Vocabulary + Item.no)	16	106.01	-37.01	74.01		
Score ~ Age + (random effects)	17	106.46	-36.23	72.46	1.55	.21
Score ~ Age + Gender + (random effects)	18	107.93	-35.97	71.93	.53	.47
Score ~ Age + Gender + SES+ (random effects)	19	107.64	-34.82	69.64	2.29	.13
Score ~ Age + Gender + SES + Vocabulary + (random effects)	20	97.14	-28.57	57.14	12.50	.0004

Table 3.10 Model comparison for Age, Gender, SES, structural language, and ToM for monolinguals *glmer*, family = binomial, bobyqa optimizer; anova

Model	Df	AIC	Log Lik	Deviance	χ^2	p
Score ~ 1 + (1 + Age + Gender + SES + Vocabulary + Item.no)	16	71.05	-19.52	39.05		
Score ~ Age + (random effects)	17	68.68	-17.34	34.68	4.37	.04
Score ~ Age + Gender + (random effects)	18	70.68	-17.34	34.68	0	1
Score ~ Age + Gender + SES + (random effects)	19	60.31	-11.16	22.31	12.37	.0004
Score ~ Age + Gender + SES + Vocabulary + (random effects)	20	62.26	-11.13	22.26	.05	.82

Table 3.11 Model comparison for Age, Gender, SES, structural language, and ToM for bilinguals *glmer*, family = binomial, nlminbw optimizer; anova

3.3 Discussion

3.3.1 Summary of findings

Once children’s age was taken into account, the only significant predictors of children’s implicature score were their structural language score and SES, which are correlated. Their gender, ToM and monolingualism/multilingualism did not have a significant association with implicature abilities. For monolinguals only, this pattern remained with structural language as the main predictor of implicature score; for

bilinguals, only ToM improved the model, once age, gender, SES and structural language were taken into account. For WLE inferences, vocabulary was the only factor that improved the model for the whole group, and monolinguals only; for bilinguals, SES did.

Note the lack of effect of gender, which is in keeping with development implicature and WLE studies that report it (e.g. for WLE Akhtar, 2002; Diesendruck et al., 2010; Hoicka & Akhtar, 2011; Kalashnikova et al., 2014; and for pragmatics Katsos et al., 2016; Loukusa, Leinonen, & Ryder, 2007; Schulze & Tomasello, 2015; Sobel, Sedivy, Buchanan, & Hennessy, 2012). Exceptionally, Stiller, Goodman and Frank (2015) do find an effect, such that male children made fewer implicature-correct responses.

3.3.2 Implicatures and structural language knowledge

The results show a positive association between structural-language knowledge and implicature-inferencing skills, and this adds to the findings of Antoniou and Katsos (2017), who find that structural language is a predictor of a composite of scalar, relevance and manner implicature and metaphor comprehension. The present study shows that this association exists with younger children (aged 2;8–5;10 years in this study, compared to 6–9 years in Antoniou and Katsos'), with quantity and relevance implicatures only, and with English-speaking children. Also similarly, this effect seems to be driven by the monolingual children: when they alone are considered, the pattern remains; when only multilingual children are considered, structural language no longer predicts implicature skills, although this could also be due to the small sample size.

These findings join a growing pool of evidence for the relationship between structural language and pragmatic abilities, particularly coming from studies on atypical development. A recent meta-analysis found that the difference between typically developing and Autism Spectrum Disorder groups disappeared for figurative language tasks once they are matched on structural language abilities – though this study concerned a much wider range of phenomena than implicature, including metaphor, irony and metonymy (Kalandadze, Norbury, Nærland, & Næss, 2016). These studies tend to conceptualise the relationship in terms of 'the critical role of

core language skills' for pragmatics (Kalandadze et al., 2016), structural language being 'implicated in the success with pragmatics' (Andrés-Roqueta & Katsos, 2017), or pragmatic ability that 'largely depends on children's language abilities' (Antoniou & Katsos, 2017: 30). In other words, structural language underpins pragmatic skills, like implicature inferencing.

However, as I have already mentioned, this is not the only possible direction of influence, and a cross-sectional correlational study like this cannot determine whether structural language aids pragmatics or vice versa. It could be that better structural language skills help children in their pragmatic processing, at least with processing the literal content and also generating relevant alternatives; but it could be that better pragmatic skills enable children to acquire structural language knowledge at a faster rate. Or, it could be that both of these are the case in development. Of course, a third option is that there is some third factor underlying the relationship, especially given the similarity of the tasks, which all involve picture-matching. Teasing these options apart is a challenge for future research.

3.3.3 Implicatures and SES

The results indicate that SES is a significant predictor of pragmatic scores, once age and gender have been taken into account. There is a significant positive relationship between SES and structural language, indicating that the sample here is typical in reflecting the widespread finding of an association between SES and structural language (Hoff, Laursen, et al., 2012; Pace et al., 2017). However, it seems that SES does not independently contribute to implicature inferencing skills, because when language is also taken into account, the effect disappears. This suggests that the relationship between SES and pragmatic skills is mediated through language knowledge. This finding is different from that of Antoniou and Katsos (2017), who did not find evidence for an effect of SES on implicature skills in development.

Although there was a range of Family Affluence Scale scores (3 – 9 / 9, mean 6.2) and Education (3 – 6 / 6, mean 4.4), the sample was not planned primarily to investigate this factor, and Cambridge, where the research was carried out, has a high level of SES, on average, compared with the national level (Office for National Statistics, 2016), so an independent effect of SES may not have been captured. In

addition, using a measure such as the Family Affluence Scale had the advantage of providing a standardized and valid measure. However, it might have been preferable to update or adapt it to the situation. For example, given the preponderance of tablet computers, it is now more likely that households might have 2 or more devices. Or, again, in Cambridge not having a car or children sharing a bedroom is not necessarily an indicator of lower SES, given the popularity of cycling as the primary mode of transport in the city and some of the highest house prices in the country, respectively. Given that so few studies have addressed this issue, it is clearly one for future research – to find out if this observation is reliable, and, if so, to investigate why pragmatic skills are more robust than vocabulary or syntactic knowledge. For vocabulary, for example, there is a clear link with input, which can vary with SES (Hoff, 2006). Is input such an important factor for development of pragmatic skills like implicatures? Are other factors like prosociality more important? Or are pragmatic tests more ‘robust’ to stylistic differences across SES? A more targeted investigation might yield both an insight into the prerequisites for pragmatic development, and applications in terms of useful interventions.

3.3.4 Implicatures and monolingualism vs multilingualism

There is also no evidence for a difference in competence with implicatures between monolingual and bilingual groups, once age, SES, gender, structural language and ToM have been controlled for. This extends the findings of other recent studies (Antoniou & Katsos, 2017; Antoniou, Veenstra, et al., 2016) to younger children. One possibility is that the high scores overall are masking any difference between monolingual and bilingual groups (scoring a mean of 82% and 79%, respectively, for critical implicature trials). However, even among the youngest age-group, where there is much more variability in scores, which are lower overall, there is no numerical difference (69% vs 68%).

These results suggest that the diverging findings between Antoniou and Katsos’ study (2017), and Siegal and colleagues’ studies (2010, 2009) – which indicate a bilingual advantage – are perhaps not a result of the age difference in their participants, after all (Antoniou and Katsos tested 6–9-year-olds, and Siegal and colleagues, 3–6-year-olds). Instead, it could be due to the different competencies

being tested: implicature comprehension via a picture-selection task, as opposed to sensitivity to cooperativeness via a Conversational Violations Test.

The puzzle here is that, although bilinguals may tend to lag behind their monolingual peers in structural language acquisition for each language, and structural language seems to be related to pragmatic abilities, bilinguals' pragmatic abilities do not seem to be affected (and, indeed, nor does structural language predict implicature skills in bilinguals). One possible reason is that the required lexical and syntactic knowledge for this and other similar tasks is simply not challenging enough to distinguish monolingual and bilingual children; it is within the competence of both groups, so their pragmatic processing is not impacted. Alternatively, it might be that other factors contribute more to pragmatic abilities in bilinguals. That is, the picture presented here is one where several aspects of a child's development are connected to their pragmatic abilities – structural language, social cognition or ToM, world knowledge, linguistic experience – and it is possible that these are weighted differently depending on the child's experience. For example, some have posited that bilinguals' conversational experience plays a role – such that bilinguals compensate for their language deficit by paying more attention to pragmatic cues (e.g. Groba, De Houwer, Mehnert, Rossi, & Obrig, 2017; Siegal et al., 2010) – as well as potentially enhanced ToM. Indeed, ToM is a predictor of implicature scores in bilinguals, on top of structural language, in this study.

However, if structural language is either a proxy for amount of linguistic communicative experience, or itself based on pragmatic skills, then there is no reason to expect a difference between bilinguals and monolinguals per se, as seen here. Although the least interesting option, this is intuitively likely: pragmatic skills are, after all, the skills of *using* language in *context*, and bilinguals may not have any less communication experience across their languages than monolinguals do in their one language. The nature of the communicative interaction that a child receives is more likely to be a factor affecting pragmatic development. SES, which is linked to the kind of communicative interaction that parent-child dyads engage in (e.g. for contingent talk: McGillion, Pine, Herbert, & Matthews, 2017), was not a significant predictor of pragmatic skills in this study, though, once structural language skills were taken into account. Admittedly, this proposal ignores the very experience itself of engaging in

communication in more than one language and the effects that this might have.

However, the reasons for bilinguals' and monolinguals' similar pragmatic development with implicatures must at the moment remain speculative. Given that all studies to date involve relatively small sample sizes, much more research is needed to establish whether there is indeed no effect of the number of languages spoken on pragmatics. Once a better understanding is gained of how multilingual language experience affects pragmatic development, why it does so can also be investigated, including what aspects of their linguistic experience or cognitive skills are involved. Given findings on other cognitive abilities in bilinguals, results might be expected to vary depending on the measures used, the pragmatic skill in question, and the bilinguals' profile (Bak, 2016; Green, 2011).¹²

3.3.5 Implicatures and ToM

There was no evidence for an association between Theory of Mind and pragmatic skills overall, or in the monolingual group, when controlling for age, gender, language and SES. It does not seem to be the case that those doing well in the implicature task are necessarily those passing the ToM tests as well. ToM was a factor, though, for bilinguals' performance, which is a surprising result. On the one hand, one could imagine that ToM might be more important where structural language knowledge is less helpful; on the other, one wonders why ToM should be any more necessary for bilinguals' pragmatic competence than monolinguals'. This requires further investigation, with a larger and more homogenous, matched bilingual sample.

For a Gricean outlook on pragmatic skills, the overall result is surprising: on that view, reasoning about others' intentions and beliefs is seen as a prerequisite for implicatures. Instead, it seems to fit in with new proposals that pragmatic inferencing may not require full ToM; inferential mechanisms and the ability to track common ground may develop separately or not necessarily be always used together (Breheny, 2006; Jary, 2013, 2013; Katsos & Andrés-Roqueta, 2015; Kissine, 2016; Sperber,

¹² For example, I included in the questionnaire here questions about frequency of use and code-switching, but was unable to use this level of detail in the analysis due to the small sample size and also the variable quality of answers to these questions – the questions either need to be made simpler or administered in person by the experimenter.

1994). I will explore these options in much greater detail in Chapter 5. Remember, though, that no strong conclusions can be drawn from this kind of null result.

Here, instead, I consider the issues with this methodological approach to the question of ToM and its role in implicature inferencing. The Sally-Anne and Unexpected Contents tasks are testing a particular instantiation of Theory of Mind, namely explicit reasoning about other's false beliefs about the world. They also have their own linguistic and cognitive demands which may obscure children's actual abilities with false belief. Rubio-Fernández, for instance, argues from a series of experimental demonstrations that it is the attentional challenges which pose a problem for young children in a Change of Location task: tracking the protagonist's beliefs while the test questions increase the saliency of the object, its location and the wrong answer (Rubio-Fernández, 2013; Rubio-Fernández & Geurts, 2013, 2016). On this hypothesis, the Unexpected Contents task would be predicted to be even more challenging, as there is no option of associating the protagonist with the object's first location, and nor is the correct answer physically present (Rubio-Fernández & Geurts, 2016).

The implicature picture-matching task would seem to present far fewer challenges: there are not two conflicting sources of information that participants have to track (such as the protagonist's beliefs, or their former beliefs, and the actual world), as all they know is what the protagonist – in this case Bob the puppet – tells them. That is, they only have to update their beliefs with information from what the puppet says, and this is adding new information, not resolving conflicting information. Further, in this context, there is no reason to think that there is a reason to be cautious in interpretation, for example due to an uninformed or unreliable speaker; the puppet shows himself to be a co-operative speaker, and is talking about what *he* did. Finally, there are no distracting test questions, but simply a picture selection.

In other words, the two tasks potentially, and problematically, differ along too many dimensions, which means that an association between ToM and implicatures may be easily missed. Crucially, no false belief understanding is required for tasks that are typically used to test children's understanding of implicature – only joint attention and intention-reading. What is needed is a test of both implicature comprehension and Theory of Mind at the same time – and this is what I attempt in Chapter 5's study.

3.3.6 Word learning, vocabulary and mono-/multilingualism

There was a significant positive relationship between WLE and overall vocabulary size in monolinguals, in addition to a relationship between WLE and age. This adds to the findings of Kalashnikova, Mattock and Monaghan (2016), who observe a similar relationship in much younger children. They use a slight variation on the WLE paradigm, with two initially novel objects, one of which is then labelled, before another novel term is then introduced as the test item. They suggest that, using this paradigm, any association with vocabulary size cannot be said to be a result of the productive knowledge of alternatives (Grassmann et al., 2015) or general cognitive abilities (Horst, Samuelson, Kucker, & McMurray, 2011). Instead, abstract knowledge about word-referent mappings from vocabulary acquisition in general contributes to the development of WLE as an abstract concept (Graham et al., 1998). They find that looking time at the target object (their outcome measure, given that their participants are infants aged 17–19 months) is positively associated with receptive vocabulary size, and conclude that WLE ‘is not likely to be a necessary precursor of early vocabulary acquisition [but] becomes reliable when infants have acquired more extensive lexical competence’ (2016: 10).

However, a possibility they do not seem to have fully considered is that the infants with larger vocabularies have them because they are better at word learning strategies, like WLE, possibly because of better pragmatic skills. In other words, their design – as well as mine here – cannot identify the direction of causality, if there is one, underlying the association. Based on available evidence so far, it seems that both directions are likely: the better at WLE a child is, the higher the rate of word learning, and larger the vocabulary size; the larger the vocabulary size, the more confidence in WLE as a strategy in general, and in applying it in a particular situation. Further research is required to understand the dynamics of this developing strategy.

There was no evidence for an effect of number of languages spoken on WLE. This is at odds with those studies that find a difference between bilinguals’ and monolinguals’ use of exclusion as a word learning strategy (see Table 3.12 for examples). This has been observed not only in infants (Houston-Price et al., 2010), but also in preschoolers and young children aged 3–6 years, using the same basic methodology (Bialystok, Barac, et al., 2010; Davidson et al., 1997). Also, some of these

other studies have used a mixed bilingual group, as I have done here (Bialystok, Barac, et al., 2010; Houston-Price et al., 2010).

Study	Age of participants	Results monolingual vs bilingual
This study	2;8–3;11 years (younger group) 4;0–4;11 years (middle group) 5;0–5;11 years (older group)	91% vs 80% 94% vs 100% 99 vs 98%
Bialystok et al., 2010 – one familiar and one novel object	Av. 3 years N = 20/40/27 (mono French, Mono English, bilingual) Av. 4.5 years N = 17/29/29	c. 75% vs 20% girls; c. 20% vs 20% boys c. 60% vs 60% girls; c. 50% vs 15 % boys
Davidson et al., 1997 – familiar and novel object	3–4-year-olds N= 16/16/16 5–6-year-olds N= 16/16/16 (mono- / Greek / Urdu)	69 % vs 65 / 60 % 92% vs 67 / 71% (Greek/Urdu)
Kalashnikova et al., 2015 – two novel objects	Younger group mean 4;0 N = 25/13 Older group mean 4;11 N = 24/12	66% vs 80% 93% vs 71%
Davidson & Tell, 2005 – known object with novel part	3–4-year-olds N = 20/20 5–6-year-olds N = 20/20	90% vs 82% 97% vs 65%

Table 3.12 Examples of WLE studies comparing monolinguals and bilinguals

One difference that might account for the discrepancy in findings could be the lower scores in other studies with older children, versus the ceiling effect in this study (overall mean of 95% for monolinguals and 95% for bilinguals), although in other studies, monolinguals are approaching ceiling, ahead of bilinguals. Several of the other studies find a developmental effect, such that the difference between monolinguals and bilinguals emerged with age, either using a classic WLE task, or another test for avoidance of lexical overlap. If anything, though, the pattern seems to be the opposite in my study. Another reason might be different paradigms – some employ two novel objects, one of which is labelled, and others one novel and one known object. However, even those with the same paradigm find different results. This is, again, an area where future research is required, with careful control of bilingual experience and testing language.

3.4 Conclusion

This study contributes an investigation of the effect of SES, structural language skills, ToM and multilingualism on implicature inferencing abilities in children aged 3–5 years, a key time for the development of pragmatic skills. It extends the findings of similar studies for older children (Antoniou & Katsos, 2017; Antoniou, Veenstra, et al., 2016), by showing that structural language knowledge is a predictor of pragmatic skills, controlling for the effect of age, gender and SES, and by not finding any evidence for a difference between monolingual and bilingual children for implicature inferencing, once these factors plus ToM are accounted for. However, the effect of SES, patterning with structural language, and of ToM for bilinguals are new and intriguing, and merit further investigation. In addition, the relationship between WLE and overall receptive vocabulary was found to persist from infancy, although in this study there was no evidence for a difference between monolinguals and bilinguals for WLE, likely due to overall high performance.

In this study, standard and readily available tests were used for the additional factors. However, in future research, they could be better matched. I discussed, in particular, how using standard ToM measures like the Change of Location task may not reveal what competencies are required for implicature understanding. Before giving up on a Gricean model, in which ToM is central to implicatures, other approaches should be taken. I therefore follow this chapter up in a task that combines perspective-taking with implicature derivation in Chapter 5.

To date, experimental pragmatics studies have tended to collect only the minimal information about participants, and measure only the target pragmatic skill. The findings here suggest that on the whole, this may not be problematic if the participants do indeed form a homogenous group in terms of SES, and as long as lexical and syntactic complexity of the stimuli is carefully controlled. However, until we know more about the effects of SES and language experience on pragmatic skills, it would be beneficial to collect this kind of information as well, so that comparisons across studies and meta-analyses are more meaningful. Certainly, for the effects of multilingual experience on cognitive functions or SES on vocabulary acquisition, a vast number of studies in different contexts and different design variants has been

necessary to begin to understand them (e.g. for overviews Bak, 2016; Hoff, 2006), and there is no reason to expect any less is needed for pragmatics.

4 The development of manner implicatures

Since Grice's first enumeration of the maxims within his Co-operative Principle, manner implicatures have been assumed to be part of a speaker's pragmatic toolbox (e.g. Y. Huang, 2017). As shall be seen, they have been taken up in neo-Gricean theories, particularly by Levinson (2000) and Horn (1984, 2004), and more recently by probabilistic approaches to pragmatics in the same spirit of a rational and prosocial speaker and hearer (e.g. Franke, 2009). However, there has been relatively little empirical investigation of manner, or, compared to other implicatures, detailed empirical or theoretical debate. In this chapter, I seek to address the questions:

1. Are manner implicatures a distinct category, and is there evidence for them in adults' communication?
2. What are the predictions for their acquisition?
3. Do manner implicatures develop early or late in children's communication?

I hope to demonstrate that manner implicatures are theoretically plausible, then I spell out some implications for acquisition, and, finally, present some tentative evidence for manner implicature comprehension in both adults and children. However, I also show how in practice they are challenging to investigate empirically and difficult to extricate from quantity inferences.

4.1 Background – manner implicatures in theory and in adults

Despite the preponderance of theoretical and experimental pragmatic research on implicature in recent decades, there has been a dearth of studies focussing on manner implicatures, either with adult speakers or children. In this section, I briefly overview the various theoretical approaches to manner within the Gricean tradition, in order to draw out some minimal criteria that distinguish manner from quantity implicatures, at least in principle. I then review the empirical evidence available for manner implicatures in adults, to suggest that some of the intuitions in the literature are borne out.

4.1.1 Theory of manner implicature

Grice's (1975) formulation of the maxim of manner was rather heterogeneous,

including as it did the submaxims ‘avoid obscurity of expression’, ‘avoid ambiguity’, ‘be brief’, and ‘be orderly’. He thus included literary ambiguity and puns among cases of manner implicatures – as flouting the second submaxim – and iconic aspects of language use – following the fourth submaxim. Subsequent work has focussed on the first and third maxims, and in particular the length of the utterance or its ‘markedness’. Consider Grice’s own example:

Miss X produced a series of sounds that corresponded closely with the score of ‘Home Sweet Home’.

⇒ Miss X’s performance suffered some hideous defect
(contrast with the ‘concise and nearly synonymous sang’)
(Grice, 1975, in Davis, 1991: 313)

The idea is that if the speaker has gone out of his way to use a prolix or marked expression when a less prolix or unmarked alternative is available, the rational hearer can infer that in doing so he *meant* something by choosing that expression, and meant something different from what he would usually mean by the alternative. Grice’s key contribution is that manner inferences involve reference not to ‘to what is said but, rather, to how what is said is to be said’ (in Davis, 1991: 313). It is the form of the utterance, not its informational content, which differs from the alternative of what the speaker could have said. Now this immediately raises the question of what constitutes a *marked* expression, and how it relates to its alternatives, and I return to this issue below.

Following Grice, Horn (1984, 2004) streamlines the maxims into two antinomic principles, Q and R, with R subsuming the second maxim of quantity (‘do not make your contribution more informative than is required’) as well as the second two manner submaxims (‘be brief’ and ‘be orderly’). The R principle states, ‘say no more than you must, modulo Q’, while Q states ‘say as much as you can, modulo Quality and R’. Thus, both the form and content of the utterance depend on the same two principles; Horn’s insight is that both pattern together as part of a cost–benefit approach to communication.

While this is theoretically elegant, Levinson (2000) argues that it is a conflation of two separate characteristics of language use, the content and the form. He instead

suggests that speakers can use the form of the language as a shortcut to the intended meaning, in the case of Generalised Conversational Implicatures (GCIs), separately from the content of an utterance. His version of Grice's manner – the M heuristic – states the following for speakers and hearers:

Speaker's maxim: Indicate an abnormal, nonstereotypical situation by using marked expressions that contrast with those you would use to describe the corresponding normal, stereotypical situation.

Recipient's corollary: What is said in an abnormal way indicates an abnormal situation, or marked messages indicate marked situations.

(Levinson, 2000: 38)

The M heuristic therefore takes the 'formal aspects' of the first and third submaxims ('avoid obscurity of expression' and 'be brief'). Interestingly, this is in contrast to Horn's bifurcation (third and fourth submaxims) but seems intuitively sympathetic to Horn's own minimax approach: an obscure or less frequently used expression is surely 'costly' just as a prolix one is, and therefore within the scope of Horn's Q, as well. Levinson (2000), though, concentrates on *markedness*. This, at his own admission, is taken to consist in a heterogenous set of features, such that a marked expression is 'more morphologically complex and less lexicalized, more prolix or periphrastic, less frequent or usual, and less neutral in register' (Levinson, 2000: 137). In other words, marked expressions are longer than their alternatives and less frequent – both properties of the form of the expression, and the form–meaning pairing in language use – as well as carrying some kind of sociolectal information. Following the weight of examples and discussion in the literature, I will take only the relationship between what is said and its utterance in terms of length and frequency to be at stake here; however, I do return to some interesting parallels with sociolinguistic phenomena in the general discussion below.

Crucially, in Levinson's conception, marked forms sit in contrast to unmarked forms, which give rise to inferences to the stereotypical, or 'I-inferences'. He therefore puts forward a view of manner for which the inferential mechanism is similar to quantity (based on the first maxim of quantity, like scalar implicatures), even if the relationship of the alternatives is different. Both are essentially 'metalinguistic' and

‘negative’, in that they involve generation of alternatives, and negation of some aspect of the alternatives. However, while for quantity, it is the content of the *stronger* alternative that is negated, for manner, it is the stereotypical implication of the *less marked* alternative that is negated. For instance, the implicature *some but not all* is derived by generating the more informative alternative *all* and negating its semantic content; the implicature *closed the door in an unusual way* from the utterance ‘made the door close’ is derived by generating a less marked but semantically similar alternative *closed the door*, and negating its stereotypical implication (not its semantic content). I take this distinction as a clear reason to take manner inferences as a discrete category of implicature at least on theoretical grounds, inviting an empirical examination.

Remember that Levinson’s theory was intended for Generalised implicatures, which are default inferences, or ‘shortcuts’ to meaning.¹³ However, his examples of manner inferences range from those which seem conventionalised, as he comments, to those which intuitively seem to be highly context-dependent, contra his general framework. To illustrate, I have arranged the following examples from intuitively less to more context-dependent, together with their alternatives:

- a) An old, old church
 → a very old church
 (An old church)

¹³ Note that in Levinson’s account of GCIs, he suggests that the heuristics can be instantiated in a type of Default Logic, with extra-logical rules (2000: 47). For scalar implicatures, he proposes a meta-rule such as:

$$\frac{\alpha(\text{WEAK}): M(\alpha(\text{not STRONG}))}{\alpha(\text{not STRONG})}$$

That is, ‘if $\alpha(\text{WEAK})$ is true, and $\alpha(\text{not STRONG})$ is consistent with what is known, then assume $\alpha(\text{not STRONG})$ ’.

One supposes that he might have something similar in mind for manner inferences, such as:

$$\frac{\alpha(\text{MARKED form}): M(\alpha(\text{not UNMARKED}))}{\alpha(\text{not UNMARKED})}$$

This is a key difference of Levinson’s account: the knowledge of specific rule-like heuristics, rather than use of more general reasoning by exclusion and elaborative inference, which restricts Levinson’s account to GCIs. But, I do not set out to test this directly here.

- b) She went to the school
 → not for lessons / as a pupil
 (She went to school)
- c) The Spanish caused the Aztecs to die
 → indirectly
 (The Spanish killed the Aztecs)
- d) A not unreliable service
 → A sort of reliable-ish service
 (A reliable service)
- e) You are permitted to leave
 → but you may stay
 (You may leave → and please do so)
- f) The corners of Sue's lips turned slightly upwards
 → Sue grimaced or smirked
 (Sue smiled)

(Levinson, 2000: 138–152)

It seems that the utterances (d–f) at least might mean rather different things in different contexts. Consider, for example, the following contexts for utterance (e):

- a) The headmaster told off the naughty pupils, then he said, 'You may leave / You are permitted to leave.'
- b) At the party, the host said, 'I'm having such a great time, but you may leave / you are permitted to leave.'

The intuition seems to be that both alternatives have the implication 'and please do so' for (a) and 'but you may stay' for (b). Similarly, the relationship between the semantic content and the implicated meaning seems to vary across contexts: for instance, the negation of the contradictory can convey the mediocre or the extraordinary.

- a) A not unreliable service
 → sort of reliable

b) It took a not inconsiderable effort.

→ a very great effort

In other words, from these examples it is not clear that Levinson's framework and insights are only useful for GCIs. Indeed, it has been influential on theories which do not make such a categorical distinction between GCIs and Particularised Conversational Implicatures (PCIs): there are models of manner implicatures within game-theoretic frameworks which are simpler still in terms of criteria (Bergen, Levy, & Goodman, 2016; Franke, 2009). They suggest that there is a lengthier alternative which is semantically equivalent, equating markedness with length of form. Interestingly, the original and dominant models on this approach – the Rational Speech Act and Iterative Best Response models (Frank & Goodman, 2012; Franke, 2009) – have to be adapted to correctly model manner implicatures, reflecting their more complex nature (Bergen et al., 2016). I, also, draw substantially on Levinson (2000) in this chapter, without ascribing to the default view of some implicatures.

A final precautionary observation on Levinson's examples will become relevant for discussion later: it seems that some examples could plausibly be analysed as quantity, not manner. For (f), where there is no systematic alternation between what is said and its alternative, the alternative could be placed on an ad hoc scale of informativeness, not markedness: <smile, turn lips slightly upwards>. Thus, the inference is that 'the corners of Sue's lips only turned slightly upwards but she did not smile', on the understanding that smiling involves more than turning one's lips *slightly* upwards, and more than just one's lips. Grice's own example of the singer is actually comparable here:

Miss X produced a series of sounds that corresponded closely with the score of 'Home Sweet Home'.

Again, it could be that the presumed singer only produced sounds, but did not sing – or that the sounds corresponded only 'closely', not 'exactly'. In both of these cases, the use of a kind of scalar adverb is particularly problematic, but even without them, some kind of informativeness scale is conceivable. It is hard to tell purely on intuition which maxims should be modelled as playing a role in such cases – or indeed

whether both manner and quantity might.

There is therefore an interesting paradox across the theoretical work on manner implicatures. On the one hand, those like Horn (1984) have noted the close parallel between manner and the second maxim of quantity: both are principles that enjoin speakers of the language to not say more than is necessary in the context. On the other hand, when it comes to examples of utterances that give rise to manner implicatures, it is often unclear whether they are manner or a case of the first maxim of quantity – the potential violation is saying not enough, rather than more than enough. In other words, there is already a hint here that manner implicatures may not be as clear cut in practice as in theory.

4.1.2 A working definition for manner

This survey has therefore shown that theoretical work can provide a basis for treating manner implicatures as an identifiable and distinct category of pragmatic inference, and understandings of markedness have become simpler with iterations of theory. For the purposes of this chapter, I would like to propose a set of minimal criteria for manner inferences, while acknowledging that the actual picture may be far less simple. I take criteria from Levinson's and formal game-theoretic models, but assume, like the latter, that these may not apply only to GCIs but also to PCIs. My working definition of manner is therefore formulated as:

Manner implicatures are derived when what is said contrasts with a less marked alternative that has equivalent semantic content.¹⁴

I suggest thinking of markedness minimally as frequency in context (following Haspelmath, 2006), although this typically has length of form as a corollary. Thus, markedness is rooted in the conventionality of language: linguistic knowledge is social, intersubjective and arbitrary knowledge, such that speakers expect it to be shared by others in the community (Diesendruck, 2011; Kalish & Sabbagh, 2007; D. Lewis, 1969). This is often taken to mean that speakers expect certain meanings to be

¹⁴ What counts as 'equivalent semantic content' is, of course, not straightforward, and a range of interpretations has already been seen in the examples cited. In this chapter I work within the spirit of examples given in the literature, while submitting that future work needs to provide a more measurable implementation of this criterion.

expressed using certain forms (e.g. E. V. Clark, 2007). In one sense, both marked and unmarked forms are obviously conventional, in that the semantic content is accepted by the linguistic community; in another sense, marked forms are *less* conventional – speakers do not expect them to be used for that meaning as much as the unmarked form, because they are less frequent in the context.¹⁵ I also assume that, like quantity implicatures, elaborative inferences play a role, both in generating alternatives and in enriching the implicated meaning in terms of how something is atypical: a manner inference can involve not only an implicated meaning of ‘non-stereotypical’, but also *how* the object or action described is unusual. This conception of manner inferences is minimal and may be simplistic, but does provide a working definition.

If manner implicatures can be defined theoretically, the question is then whether manner implicatures are actually used by speakers and hearers in communication.

4.1.3 Evidence for manner in adults

There are very few experimental studies to date investigating whether adult speakers are sensitive to manner and can derive manner implicatures. Current studies suggest that they do not seem to be as robust as the most robust quantity implicatures, and are certainly more challenging to test. I first briefly consider three sets of studies that are related to manner implicatures, before recapping previous work in which I tried to examine manner inferences directly.

Firstly, Bergen and colleagues (Bergen, Goodman, & Levy, 2012; Bergen et al., 2016) test ‘non-linguistic’ manner inferences, through artificial language paradigms. In these speaker–listener games, there are three possible referents (pictures of shapes) and three possible labels (symbols) which were more or less frequent and more or less costly (literally), respectively. Participants have to work out how to communicate and interpret a reference to one of the objects, given limitations of length (one ‘word’) and cost. The results suggest that speakers and hearers do align frequency and costliness, which is taken by the authors as evidence for sensitivity to manner. However, the

¹⁵ It is necessary to model frequency *in context*, since, for example, what may be marked in the context of a chat by the coffee machine may be completely unmarked in a business meeting of lawyers. A challenge for future research is to make this definition quantifiable, so that it can be submitted to empirical testing.

highly metalinguistic task might not be capturing the same processing as in typical communication (for example, it included literal cost of utterances in dollars). While they also conclude that it is rooted in social reasoning, given that it is evident in nonlinguistic novel situations, the task is completed by linguistically competent people, who may be drawing on their linguistic experience to do so. It is therefore suggestive but by no means firm evidence for manner implicatures.

Secondly, one of the flagship manner examples has been an alternation available in English: the periphrastic causative construction, such as ‘make the plate break’, with a lexicalised causative alternative, ‘broke the plate’ (e.g. Antoniou & Katsos, 2017; Franke, 2009; Levinson, 2000). A set of experimental studies investigating the meaning and use of syntactic constructions expressing causation has found that adults seem to prefer to use a lexical causative (‘break’) for intentional actions, and a periphrastic (‘make break’) for unintentional actions or where the agent is inanimate, in both production and comprehension (Song & Wolff, 2005; Wolff, 2003). They also have better memory for manner of motion or state change when they have heard a periphrastic phrase (e.g. ‘make the truck roll’), and better memory for result when hearing the simple transitive (‘roll the truck’), in a spot-the-difference task without manipulation of intentionality (Kline, Muentener, & Schulz, 2013). These are both different from – though not incompatible with – Levinson’s (2000) proposal that a lexical causative gives rise to a stereotypical interpretation, and a periphrastic one to a marked interpretation – where both can be intentional. The pattern of usage is obviously complex, and sits at the contested boundary between semantics and pragmatics, and, indeed, semantics and syntax. One way of reconciling the two perspectives, for example, is to imagine that if the periphrastic is more often used for unintentional actions, then when it is used for intentional actions (i.e., in a marked way), listeners can pragmatically infer that this usage must be for some particular reason, such as to indicate an unusual or indirect manner of causation – although this requires a richer view of markedness than frequency. This work, though, obviously only applies to one case of purported manner implicatures, but suggests that similar complex usage patterns might be relevant for others, too.

Thirdly, work on referential pacts and speaker-specific adaptation is closely related to manner (e.g. Brennan & Clark, 1996). A referential pact is an instantiation

of conventionality on a local, discourse-specific scale: once two interlocutors have 'agreed' (implicitly) on a referring label in discourse, they persist in doing so, and comprehension is affected if this is deviated from. Following it is like adhering to manner: there is a semantically equivalent alternative that could be used in the context (say, 'settee', when the interlocutors are using the term 'sofa'), and it is more marked in the discourse context, as the already established term is more salient, more frequently used in the discourse, and therefore less marked. Breaking the referential pact, then, is somewhat like flouting manner – the more marked term is used, but, importantly, not necessarily to mean anything different. Metzinger and Brennan (2003) found that adults respond more slowly when the same interlocutor uses a new term for a referent already in common ground in discourse.

So, adult speakers seem to be sensitive to markedness in utterance production and comprehension; use semantically equivalent expressions in different ways; and expect speakers to be conventional on a local as well as global level. But do they actually derive manner implicatures? In previous work, I attempted to adopt experimental paradigms that have been used with scalar implicatures (by Degen, 2015) to test out the kind of manner implicatures cited in the literature, both GCI and PCI (E. Wilson & Katsos, 2016). I devised short scenarios, ending with either a marked or unmarked utterance, which were pretested to check equal naturalness of a marked or unmarked ending (with the stereotypical meaning or manner-implicated meaning explicitly stated) – see Table 4.1 for examples.

Context	Marked utterance → marked meaning	Unmarked utterance → unmarked meaning
Nick and Dan were watching a history programme, but Nick fell asleep. Afterwards he asked, 'What happened at the end?' Dan replied...	'The invaders caused the villagers to die.' → The invaders killed the villagers but indirectly, by introducing disease.	'The invaders killed the villagers.' → The invaders killed the villagers directly.
Jamie got home from his grandmother's. His mum asked, 'Did she give you a drink?' He answered...	'She put a teabag into a cup and poured over boiling water.' → She gave me a cup of tea, only it didn't taste like tea.	'She gave me a cup of tea.' → She gave me a normal cup of tea.

Table 4.1 Example stimuli from Wilson & Katsos (2016)

In the first experiment, participants read each scenario, and chose which sentence was most similar to the highlighted ending – the unmarked or marked meaning. There was a significant difference between the rate of unmarked meaning chosen for unmarked utterances, and marked meaning chosen for marked utterances (92% vs 56%), but no difference between PCIs and GCIs (GCIs were causatives, modals, and negation of contradictory). A similar pattern of results was found in a follow-up rating task. Note that the explicitly stated meaning of unmarked utterances and the utterances themselves tended to be extremely similar, contributing to the near-ceiling score.

The interpretation of these findings depends on what behaviour is expected as default, in the case that no manner inference is derived from the marked utterance, while the unmarked interpretation is derived for unmarked utterances. The most straightforward prediction would be that participants would simply guess, as both options are compatible with what was said (or, in a ratings task, average across trials and participants at the midpoint) – remember that the semantic content of the alternatives is equivalent. This would mean that the results, at chance, suggest that participants are not sensitive to manner.

However, I would argue that chance selection as the null hypothesis is unconvincing in this case; what is instead more likely is that hearers tend to choose the most stereotypical (unmarked) interpretation unless there is reason to think

otherwise. For instance, if no inference is derived on hearing ‘the invaders caused the villagers to die’, then there is little to lead a hearer to pick an explicitly stated implicated meaning like ‘The invaders killed the villagers but indirectly, by introducing disease’. On this interpretation, if they are not sensitive to manner, one would expect similar rates of selection of the marked meaning for both the marked and unmarked utterances – high selection of unmarked meaning. This is not the pattern seen in these studies, which would mean that where the marked interpretation is selected, or rated preferably, this can be taken as an indication that a manner inference has indeed been derived. A third experiment indicated that this is indeed the more plausible interpretation: adults were asked to simply explain what they thought the speaker of the marked or unmarked utterance meant. Around 40% of responses to marked utterances suggested that participants had derived a manner implicature, though there was significant variety in the precise enrichment.

This study constituted the first indication from experimental evidence that manner implicatures are part of adults’ pragmatic competence. They also call into question the distinction between GCIs and PCIs for manner. However, there are a number of drawbacks. While a binary selection or rating task using the ‘spelt out’ implicated meaning might be serviceable for scalar implicatures where there is a clear and minimal difference between what is said and what is implicated (e.g. ‘some’ vs ‘some but not all’), they are more problematic for manner. For manner, the relationship between what is said and what is implicated is more complex – the meaning of ‘not in the usual way’ can be enriched via an elaborative inference in a number of ways. Indeed, the items used in the experiments reflected this, based on those found in the literature. Furthermore, there were notable differences in responses across items. On the one hand, this is enlightening, and suggests that the marked utterances given in the literature (particularly by Levinson, 2000) may differ in interesting ways, such as in context-dependency, or even whether they constitute manner implicatures at all. On the other hand, this makes the comparison between marked and unmarked conditions less reliable and less representative – and is certainly not comparable with, say, scalar or ad hoc experiments, where participants tend to respond more consistently across trials. Finally, neither the experimental design nor the items themselves were at all suitable for use with children, owing to the

written mode, metalinguistic judgements, and complex syntax and vocabulary, driven by examples in the literature. The challenge is therefore to learn from these studies, and design a child-friendly experiment, for which the items are better normed, in order to compare adults' and children's competence with manner.

4.1.4 Manner in development

As there is some – albeit extremely limited – evidence that adults are sometimes sensitive to the maxim of manner in communicating their intended meaning and interpreting others' utterances, the question arises as to how children develop this ability. I now review what research there is, and then present two opposing hypotheses for a developmental trajectory. For the first, I build on the theoretical framework and predictions I put forward in Chapter 2, to hypothesise that manner inferences may develop late, because of their unique features. Secondly, drawing on what is known of children's developing understanding of conventionality, I suggest that, alternatively, some manner inferences may be available early.

4.1.4.1 Existing studies

One exception to the dearth of work specifically on manner is Clark & Kurumada's (2013) overview chapter, in which they concentrate on the submaxim 'be brief' in production. However, in doing so they bring together manner, quantity and relevance – both the form and informational content – by using a rather approximate measure of brevity (mean utterance length). Their chapter therefore opens up the discussion on this topic, but the need remains to bring together theory and existing evidence on manner implicatures, and to spell out the implicatures for acquisition.

Two studies investigate manner implicature comprehension empirically. Firstly, Antoniou and colleagues include manner implicatures in a battery of picture-selection tasks also testing relevance, scalar, metaphor and irony inferences (Antoniou, Grohmann, Kambanaros, & Katsos, 2013; Antoniou & Katsos, 2017). They use only one case of manner – the causative 'make X Y' construction (or rather, its equivalent in Greek – the language of testing). In that study, monolingual Greek-speaking children aged 6;2 to 9 years (mean 7;4) score on average 1.9/3 in manner trials, which is comparable to their scalar implicature scores, but worse than relevance. However, they find no evidence for a correlation with scalar scores, only with relevance, plus

age, IQ and working memory.

Secondly, Okanda, Asada, Moriguchi & Itakura (2015) aim to use a Conversational Violations Test to investigate the developmental trajectory of various types of implicature (relevance, quantity, manner and politeness). They explicitly motivate the inclusion of manner implicatures, citing Siegal's observations (2008) about the possible effect of manner in standard developmental tests, for example in an appearance–reality distinction task where subtly different wording in the experimenter's questions may 'violate the maxim of manner'. However, in the example they give, it is not clear that what is being violated is the maxim of manner – understood in terms of markedness – as the experimenter is not using a marked form when an unmarked one would have been co-operative, but simply obscurely or subtly-phrased questions that children may misunderstand. Okanda and colleagues do not provide a more definite set of criteria for manner maxims than Grice's original catalogue, and their stimuli would only fit within the broadest conception of manner, and certainly not within the criteria I have outlined here. To take two examples of items:

- a) Have you ever taken the train?
I may or may not have / Yes, it was fast.
- b) With whom will you play today?
I am not sure / I will play with Yuko.

For (a), the uncooperative answer is clearly under-informative, with the more informative alternative 'I have'; for (b), there is either no violation – if the speaker really does not know – or violation of quality. Manner is simply not at stake here.

In sum, how and when children learn to comprehend manner implicatures remains an unanswered question.

4.1.4.2 Hypothesis 1: Children learn to derive manner inferences relatively late, after quantity and relevance

Manner implicatures can be seen as the most complex of implicatures, and therefore are expected to emerge late in development. Given the criteria for manner implicatures that I outlined above, manner and quantity share certain features. Like

quantity, manner implicatures involve some sort of inferential mechanism that includes the generation of alternatives, negation of alternatives, and consequent derivation of intended meaning. Like quantity, too, manner implicatures require recognition or generation of *relevant* alternatives in context, which requires tracking aspects of the context, such as the QUD. It could therefore be expected that manner inferences would emerge at least around the same time as quantity implicatures in development.

In addition, though, manner inferences require an appreciation of the conventionality of language at a global level; an ability to detect marked expressions at the local level; and productive knowledge of a semantically equivalent but less-marked expression. For instance, on hearing ‘the man made the door close’, a child has to recognise that this is not the expected less-marked expression, and that this would be the alternative ‘closed the door’, given the context, including the interlocutors’ language usage; to assume that the marked expression has been used intentionally to convey a different meaning; to infer or know from linguistic experience that the unmarked alternative is typically used to mean the stereotypical action; and to negate that alternative implicated meaning to infer that the speaker therefore intends the meaning not of the stereotypical action, but an atypical one. What makes an alternative relevant in the context may for manner not (just) be the QUD, but expectations about language use, including stylistic or sociolinguistic variation, or speaker-specific traits like language-learner status. On Levinson’s model, the derivation is also more complex, involving the negation of the alternative’s implicated meaning, not its semantic content.

It might therefore be expected that manner inferences emerge late – after relevance and quantity – firstly, because of the potentially more complex inference that involves negation of the enriched meaning of the alternative, not its semantic content, and secondly, because of the additional knowledge required of markedness, gained through linguistic experience. The table presented in Chapter 2 could therefore be extended as in Table 4.2.

	WLE	Q	R	M
Tracking QUD		✓	✓	✓
Tracking language use				✓
QUD licenses inference		✓		✓
Elaborative inference		✓	✓	✓
Negation of alternative	✓	✓		
Negation of alternative's implication				✓

Table 4.2 WLE, Quantity, Relevance and Manner inferences

This schema might be surprising, given what I mentioned in Chapter 2, that some have compared WLE inferences to manner (e.g. Brosseau-Liard & Hall, 2011; de Marchena et al., 2011). However, I think there are important differences that justify it. In a typical WLE situation, a novel label has the alternative of the known label for the known object in the context, say ‘strawberry’. The label ‘strawberry’ is the conventional and unmarked label, because it is the most commonly used label for strawberries (in this kind of context) – just like manner. There is no way for the listener to know, though, whether it has the equivalent semantic content to the utterance, given that the uttered label is, of course, novel. Given that in a typical WLE situation, the two available referents are obviously different objects, not more or less typical instantiations of a single type of object, it seems implausible that the hearer might reason that the novel word is a marked alternative with the same semantic content (unless, of course, the reference is resolved as referring to the familiar object). Further, while I have given a minimal definition of markedness as frequency of use in context, it very often co-occurs with length, which also is not met in the WLE scenario. WLE inferences can therefore be seen as certainly related to manner inferences – in that they involve reasoning about alternative forms – but distinct. It is therefore possible that WLE emerges early in children’s development, while manner inferences emerge late.

4.1.4.3 *Hypothesis 2: Children may be able to derive some manner inferences early, or even be more sensitive than adults*

Alternatively, there are three related competences that might suggest a different picture, that manner could emerge early in children's pragmatic development, around the same age as relevance and quantity (3–5 years): appreciation of conventionality, referential pacts, and markedness in gestures.

Studies of children's development of an understanding of the conventional nature of certain phenomena show it to be early and robust – they are able to expect others to share their knowledge appropriately. In his overview, Diesendruck (2011) highlights how young children expect words to be conventional knowledge among speakers of the same language (e.g. Diesendruck & Markson, 2001 and see discussion of WLE in Chapter 2), but do not expect proper names to be (Diesendruck, 2005); how 2- and 3-year-olds expect the rules of games to be shared (Rakoczy, Warneken, & Tomasello, 2008), but not preferences or knowledge of facts (Diesendruck & Markson, 2001; Graham, Stock, & Henderson, 2006); how preschoolers expect object function to be conventional – recognising that the designer of an object does not have a privileged influence on their understanding of an object function, compared with any other user, in contrast to object categorisation (Defeyter, Hearing, & German, 2009). Various theories have been proposed to account for this early development of conventionality (Kalish & Sabbagh, 2007). Sabbagh & Henderson (2007) propose that it is precisely children's *lack* of Theory of Mind that leads them to assume shared knowledge; in pedagogy theory it is suggested that there is a special kind of 'teaching' through ostensive communicative cues such that children assume that knowledge acquired this way is generic (Csibra & Gergely, 2009); and with social-cultural theories it is argued that it is through participation in social practices at a local level that conventionality at a global level emerges (e.g. Rogoff, 2003). All three have in common the possibility that an expectation of conventionality can emerge early in development without full reasoning about others' beliefs, and they therefore propose some kind of 'shortcut' – a cue or cognitive characteristic which allows children to assume conventionality and practise it.

Gricean pragmatics has, by contrast, assumed the need for complex mind-reading for inferencing, although evidence for some implicatures aged 3 calls this into

question (a puzzle I consider at greater length in Chapter 5). It is therefore possible that the extra skills required for manner implicatures, including knowledge of conventionality, are already in place as implicature understanding develops. Further, suggestions that children might even have a heightened sense of conventionality (whether *because* of their lack of Theory of Mind, or as a learning strategy) would allow for the possibility that some manner implicatures could be available for children early, or even where not intended. Of course, this relies on an ability to keep track of the frequency of labels and constructions – their markedness – but cross-situational statistical learning is precisely one strategy argued to be pervasive in language acquisition (e.g. Yu & Smith, 2007).

Appreciation of conventionality is a prerequisite for manner inferences, but more persuasive evidence for potential early competence comes from a referential pact study, in which 3-year-olds are found to be as sensitive to referential pacts as adults, if not more so (Matthews, Lieven, & Tomasello, 2010). This can be interpreted as an appreciation of conventionality at the local level. Like adults, children are found to be slower to respond when the same speaker referred to an object with a different label from that previously. For children, however, unlike for adults, this also extends to a new speaker. In other words, children in this study have a stronger sense of conventionality than adults, with a normative component. Matthews and colleagues suggest that one contributing factor could be children's lesser linguistic experience, resulting in 'more specific memories of previous encounters with referents' (Matthews et al., 2010: 757). As suggested above, one way of viewing referential pacts is as observance of conventionality – and the maxim of manner – at a local level, so in this study children are demonstrating an awareness that there is a 'normal' way of expressing something, a skill that is also required for manner inferences.

Finally, an insightful study finds that children perform a similar kind of inference to manner with gestures: Liebal, Carpenter and Tomasello (2011) investigate children's responses to a pointing gesture in the context of a clearing-up game. It could either be unmarked (a normal, simple pointing gesture, most frequently encountered), or marked (a less frequent pointing gesture involving movement and facial expression). Three-year-olds interpret the marked point as intending an action involving a hidden part of an object (e.g. its contents) more often than they do for the

unmarked point (for which they put the whole object away). One interpretation of this is to see it as akin to a manner implicature: arguably, in the context of a clearing-up game, the two points could share the same ‘semantic content’, of referring to an object that is to be involved in the dyad’s activity, differing in markedness. The marked point is therefore interpreted as not referring to the stereotypical ‘meaning’ of the unmarked point (moving the whole object), but to an atypical meaning. Alternatively, what could be going on here is a contrastive inference, more similar to a referential pact or WLE: the unmarked point ‘means’ ‘put away the whole object’, and so by contrast the marked point must ‘mean’ something different. Either way, it is further evidence for early expectations of conventionality.

These all suggest that some of the additional skills for manner inferences may be in place early, and so manner implicatures may also be available, as long as the prerequisite knowledge of alternative forms is in place for any case, at the same stage as other implicatures.

4.2 Experiment 2A: manner implicatures

To open the investigation of children’s development of manner implicatures, as well as to further our understanding of adults’ comprehension, I conducted two picture-matching experiments, to test whether children make manner inferences at rates similar to or different from inferences to the stereotypical (I-inferences).

4.2.1 Method

4.2.1.1 Design

The experiment was a picture-matching task, similar in principle to the one used in Experiment 1A. As explained there, this kind of task has been successfully employed with children to test pragmatic comprehension skills, and has a number of advantages that make it particularly child-friendly.

This study was a very simple instantiation of such a task: the participants heard an utterance (a single sentence) and had to choose the picture that matched (cf. Horowitz & Frank, 2015, for quantity implicatures). There were only two within-subject conditions – critical and control, or marked and unmarked utterance – and two age groups. In the critical (or M) condition, participants heard a marked

utterance, such as ‘The man made the door close’; in the control condition (or ‘I condition’, adopting Levinson’s terminology), participants heard the unmarked alternative, such as ‘The man closed the door’. In both cases, they saw two pictures, one of which depicted the stereotypical instantiation of the action, and the other an atypical one.

Each participant heard each item in only one condition – critical or control. There were two pseudo-randomly constructed orders (such that no more than two trials in any one condition appeared consecutively), making 4 lists in total, each with 9 trials. Across lists, the position of the pictures (left or right) was counterbalanced. Some items could be viewed as PCIs, and others as GCIs (with the construction ‘make X Y’). This made it a $2 \times 2 (\times 2)$ design: condition \times age group (\times type).

4.2.1.2 *Participants*

Children were recruited from Year 1 classes in two local primary schools around Cambridge (aged 5;10 – 6;10), and through personal contacts.¹⁶ Following headteachers’ consent, parents were sent information letters and asked to opt in to the study. The recruitment process, and study, was approved by the University of Cambridge Psychology Ethics Committee, following research ethics guidelines.

In total, 32 children were recruited. One participant was excluded due to experimenter error, and 6 as they were multilingual, leaving 25 children whose data are included for analysis (N = 9 girls). Given that for manner implicatures, the form of the utterance is important, only monolingual children were included to avoid the effect of language experience or differing competence in English.

Adults (N = 22) were recruited via Prolific Academic, an online recruitment platform for research (Prolific Academic Ltd, 2016), and were paid £0.25 for their participation, which took on average 2 minutes (a rate of £7.50/hour). 4 were excluded as multilingual. In addition, N = 32 adults participated in the pretests, also via Prolific Academic, and were paid £0.40 for their participation. Adult participants were filtered as being born and resident in UK and first language English speakers.

¹⁶ As some parents only returned the consent form and not the background information questionnaire, the exact age of some children was not collected. These children also completed Experiment 3 in a separate testing session.

4.2.1.3 Material

Using the items from my previous work as a starting point (E. Wilson & Katsos, 2016), I devised a list of 12 items all of a PCI nature, which were child-friendly in terms of subject matter and vocabulary, and tested these with adults, informally and online via Prolific Academic. Of these, I selected 6 which had not been ruled out as visually unclear or unlike manner inferences by responses in the pre-test. In addition, I added 3 GCIs, all with the construction ‘make X Y’, following Antoniou and Katsos (Antoniou & Katsos, 2017). See Appendix 8.5 and Table 4.3.

The final stimuli therefore consisted of 9 items: 6 PCIs and 3 GCIs. These were paired with colour picture cards with cartoon-like illustrations sourced from an online free image database (Braxmeier & Steinberger, 2017). There was also one unambiguous warm-up item.

M utterance	I utterance	PCI/GCI
The boy put paint on his face.	The boy did facepainting.	GCI
The girl dipped her brush in paint and put it on the paper.	The girl painted a picture.	PCI
The girl used a knife to make small pieces of apple.	The girl cut the apple.	PCI
The man made the plate break.	The man broke the plate.	GCI
The boy pushed a cloth across the table.	The boy cleaned the table.	PCI
The man picked up some things and moved them around the room.	The man tidied the room.	PCI
The girl made the bag tear.	The girl tore the bag.	GCI
The girl put a duvet on top of the bed.	The girl made the bed.	PCI
The man made the door close.	The man closed the door.	GCI

Table 4.3 Items used in Experiment 2A

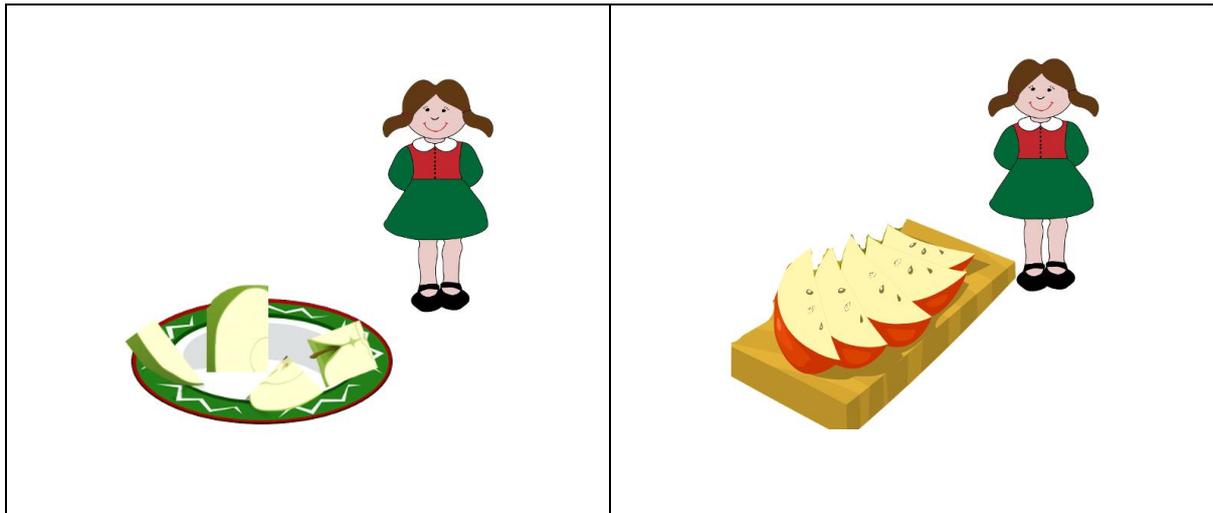


Figure 4.1 Example picture cards for 'The girl used a knife to make small pieces of apple' / 'The girl cut the apple'

4.2.1.4 Procedure

The task was conducted as part of a longer experimental session, and it was always the first task in the session. Children were told that they were going to play a guessing game; the experimenter would describe one of the picture cards, and their job was to choose which one the experimenter was describing or thinking of. (The experimenter was either the author or a summer intern supervised by the author.)

The picture cards were presented in a pile, and turned over by the experimenter for each trial. The picture card chosen by the child was then put in one new pile, and the other card discarded in another new pile.

After the end of the experimental session, the cards chosen by the child were recorded by taking a photograph of the selected picture cards, and the list noted.

Adults completed an online version of the task; they were told that another participant had described one of the pictures, and their job was to choose the picture he or she was describing, by clicking on it. They read the utterances rather than heard them. At the end, adults were asked for some information about their languages, to check whether they were indeed native British English speakers.

4.2.2 Predictions

As in the sentence-selection task (E. Wilson & Katsos, 2016), there are two possible behaviours that might indicate *lack* of M implicatures: guessing (which would manifest itself as chance performance across participants) or preferring the

stereotypical, most frequent meaning. This leads to two alternative sets of predictions, which I cast here only in terms of selection of the marked picture.

On an unbiased interpretation, that no inference leads to random guessing:

- if participants do not make any inferences at all, either marked or unmarked, then chance in both critical or control condition is expected
- if participants only make an inference to the stereotypical for unmarked utterances, but no inference for the marked utterance, then below chance selection of the marked picture is expected for the unmarked utterance, and chance for the marked one
- if both inferences to the stereotypical and manner implicatures are derived, then below chance selection of the marked picture is expected for the unmarked utterance, and above chance for the marked utterance.

If instead, and as already argued is more likely, there is a general bias to the stereotypical, then:

- if participants do not make any inferences or only inferences to the stereotypical, then below chance selection of marked pictures is expected in both conditions
- if participants derive manner implicatures, then a higher rate of marked picture selection is expected for marked utterances than for unmarked ones.

In addition, for the comparison between adults and children:

- if manner implicatures emerge very late, then more marked picture selection for marked utterances is expected for adults than children
- if manner implicatures are comparable to scalar and other implicatures in development, then equivalent performance is expected
- if children have a stronger sense of conventionality than adults, then more marked picture selection for marked utterances is expected for children than adults

The last option is extremely unlikely at the age chosen, given that children already have significant linguistic experience and competence with other types of implicature by 6 years.

4.2.3 Results

Note that there were two minor experimenter errors: for one participant, the left-right orientation of the pictures was swapped; for another, two trials were swapped in the list order. Both participants are included in the present analysis.

Picture selection is coded as M picture (marked) selection or not, in keeping with the predictions. Note that this means that in the unmarked (I) condition, very low rates of M picture selection are expected.

For both children and adults, this is the case; there is also higher rate of M picture selection for marked utterances than unmarked utterances (Figure 4.2, Figure 4.3).

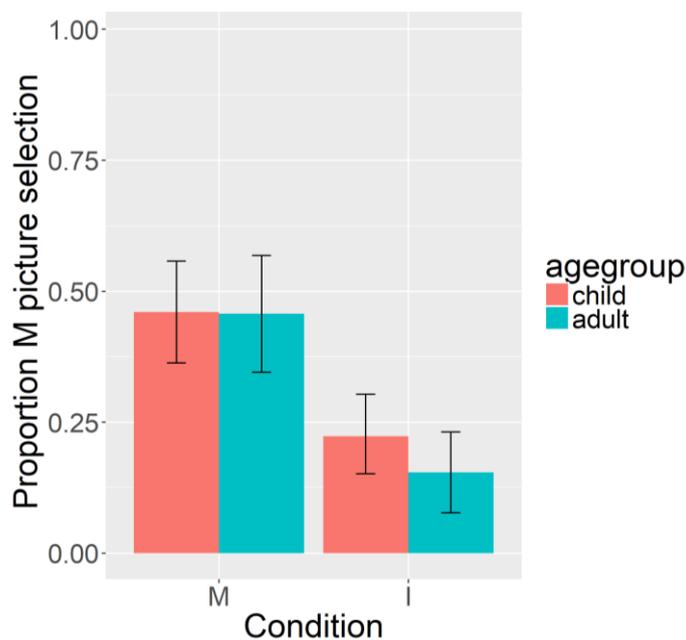


Figure 4.2 Mean selection of M picture by condition (utterance) and age group
Error bars show bootstrapped 95% confidence intervals for between-subject comparison

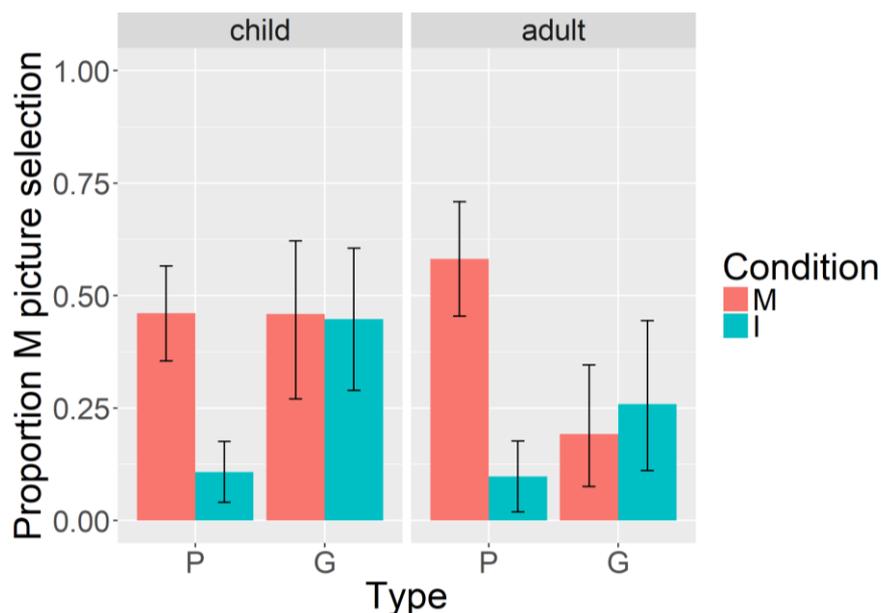


Figure 4.3 Mean selection M picture by Type (PCI/GCI) and condition (utterance)
 Error bars show bootstrapped 95% confidence intervals for between-subject comparison

		child	adult
I	GCI	.45	.26
M	GCI	.46	.19
I	PCI	.11	.10
M	PCI	.46	.58
I	Total	.22	.15
M	Total	.46	.46

Table 4.4 Proportion of M picture selection by type and age group

To examine the effect of condition (marked M vs unmarked I utterance), type (PCI vs GCI), age group (child vs adult), and their interaction, mixed-effects logistic regression models with item (by condition) random slopes and dummy coding were fitted using the *lme4* package in R (Bates et al., 2015; R Core Team, 2016; RStudio Team, 2016). The child age group, manner inference and PCI type were set as baseline.

There was a significant effect of condition, such that for children and PCIs, there are fewer M picture selections in the I condition than in the M condition ($\beta = -2.07$, $p < .001$); an interaction of condition and type, such that for children, the difference of I minus M was less negative for GCI than for PCI ($\beta = 2.01$, $p = .003$); and an

interaction of type and age group, such that for the M condition, the absolute difference between GCI and PCI was less for children than adults ($\beta = -1.92$, $p = .008$) – Table 4.5. There was no effect of age group for M and PCI.

	Estimate	SE	z	p
I	-0.15	.34	-0.45	.65
GCI	-2.07	.46	-4.55	< .001
Adult	-.02	.59	-.03	.98
I:GCI	.54	.37	1.47	.14
I:Adult	2.01	.67	3.02	.003
GCI:Adult	-0.67	.71	-.95	.34
I:GCI:Adult	-1.92	.73	-2.64	.008
I:GCI:Adult	1.10	1.08	1.00	.32

Table 4.5 Condition ~ Type ~ Age group + (1 + Type | Item)

Glmer, family = binomial, optimizer = bobyqa, contrasts = dummy, baseline = M / PCI / Child

4.2.4 Discussion

From the results, there is no evidence of a difference between adults and children in interpretation of marked and unmarked utterances, when all items are examined together. Both groups choose the marked (M) picture more often for marked utterances than for unmarked ones. When the distinction between PCIs and GCIs ('make X Y') is also examined, then age group does play a role: while children and adults pattern together for PCIs, for GCIs they are different, with adults having an overall preference for the unmarked pictures, and children choosing roughly equally between marked and unmarked pictures.

On the assumption that without deriving an implicature, hearers would interpret both unmarked and marked utterances as having an unmarked meaning, this can be taken as evidence that children as well as adults are sometimes sensitive to manner. They are sometimes able to infer that the speaker used a marked utterance intentionally in order to refer to a non-stereotypical performance of an action.

The results also suggest that there is no distinction between so-called GCIs and PCIs. For both adults and children their response rate to GCIs are the same for marked and unmarked utterances, contra predictions extrapolated from Levinson (2000), although for children this rate was much higher than for adults. Note, though, that

there were few GCI items (N=3), and the difference between GCI and PCI was confounded with utterance length for many items.

However, there are a number of problems with this experiment. Visual inspection of by-item responses shows that there is wide variation across items (Figure 4.4). This could reflect something about how marked the utterance is, but it could also indicate that there is some item artefact having an effect, such as irrelevant picture details.

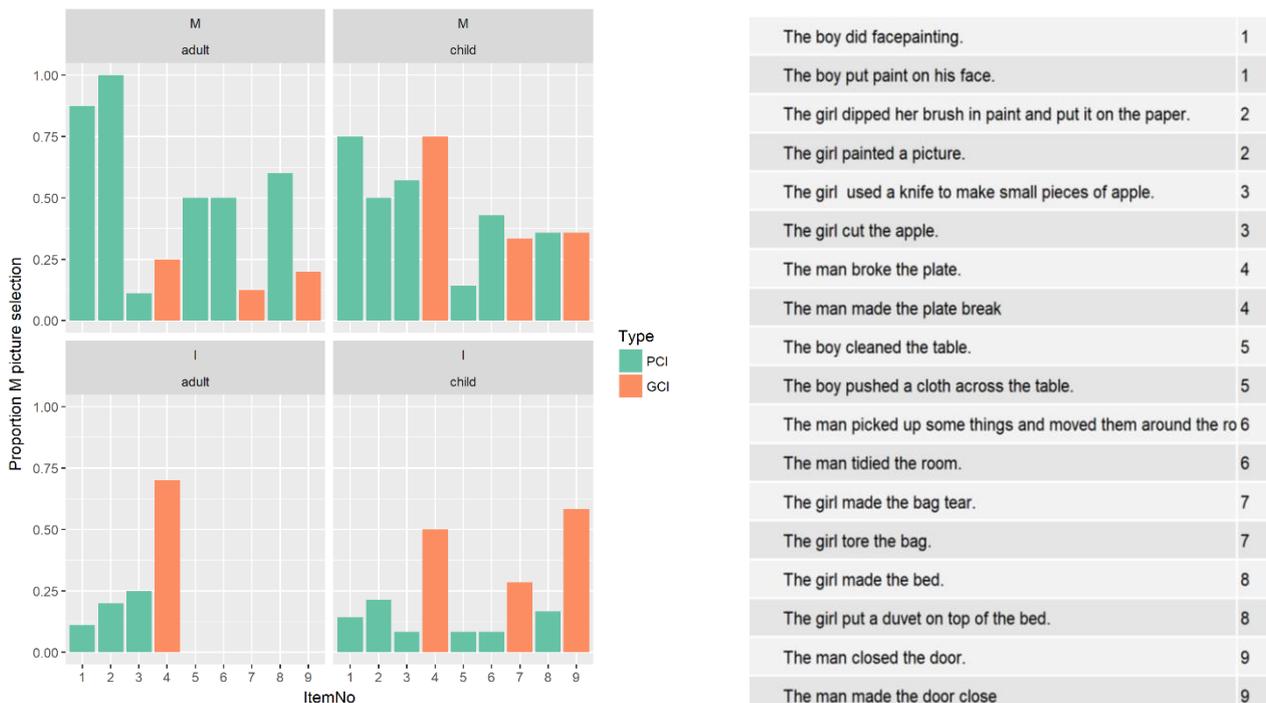


Figure 4.4 Mean selection of M picture by item and condition

Indeed, some comments from child participants as to why they chose particular pictures indicated that it was sometimes for reasons unrelated to manner. One issue might have been that the picture card pairs varied in how similar they were to each other, meaning that the key difference between stereotypical and non-stereotypical meaning depicted in the pictures may have been obscured. Therefore, making the picture pairs as similar as possible and testing the items with adults to see what factors are affecting their choice, would improve the task. This is what I did in the second experiment.

4.3 Experiment 2B: manner implicatures

4.3.1 Method

4.3.1.1 Design

The design was the same as in Experiment 1, with differences only in the materials and procedure, explained below.

4.3.1.2 Participants

Children were recruited from Year 1 and 2 classes in a primary school in Sussex, and through personal contacts (age range 5;8–7;4, mean = 6;8¹⁷). Following headteachers' consent, parents were sent information letters and asked to opt in to the study. Alternatively, parents who were personal contacts of the researchers were sent information letters directly by the researcher. The recruitment process, and study, was approved by the University of Cambridge Social Sciences Research Ethics Committee, following research ethics guidelines.

In total, 17 children were recruited. 2 children were multilingual, but English dominant, and so were included in the analysis for the sake of the sample size; 2 were excluded as they were Chinese-dominant.

Adults (N = 48) were recruited via Prolific Academic, an online recruitment platform for research (Prolific Academic Ltd, 2016), and were paid £0.30 or £0.45 for their participation, which took on average 2 minutes or 5 minutes (depending on the length of the task, with or without justifications, as I explain below). An additional 7 were excluded as multilingual. In addition, N = 184 adults participated in the pretests, also via Prolific Academic.

4.3.1.3 Material

The materials were similar to those used in Experiment 1, except that the items had been pre-tested more extensively. Experiment 1 was run again with (different) adult participants, with participants asked to explain their picture choice, in order to identify features of the pictures themselves that were eliciting selection for irrelevant

¹⁷ Age data for one 7-year-old is missing.

reasons (e.g. size of apple pieces, amount of paint on the paper, age of protagonist). In addition, all the pictures were made as similar as possible (for instance, same person and object where possible). This process was iterated until there was a set of items for which only a minority of responses were coded as due to irrelevant picture details or other irrelevant reasons based on participants' explicit justifications (for the final data, $\leq 36\%$,¹⁸ mean 7%, and $< 20\%$, mean 1%, $N = 24$).

Secondly, the items were included in an utterance-choice task, again with selection justifications, in order to a) identify any aspects of the utterances themselves which were leading to irrelevant inferences (e.g. because of unnatural wording), and b) to ensure that the I utterance is always preferred ($> 70\%$ responses). Again, this task was iterated, in conjunction with the first.

Ultimately, a set of 8 items was created, of which 7 were based on those from Experiment 1, and 1 was a new GCI type ('made the jar open'). In addition, the warm-up item was changed so that it had an implicit QUD of quality rather than quantity (a choice between a blue and orange wall, and a green and pink wall, rather than an all-blue wall). See Table 4.6 and Appendix 8.6.

M utterance	I utterance	PCI/GCI
The woman dipped her brush in paint and put it on the paper.	The woman painted a picture.	PCI
The woman made small pieces of apple.	The woman cut up the apple.	PCI
The man made the plate break.	The man broke the plate.	GCI
The boy pushed a cloth across the table.	The boy cleaned the table.	PCI
The man picked up some things and moved them around the room.	The man tidied the room.	PCI
The man made the jar open.	The man opened the jar.	GCI
The girl put a duvet on top of the bed.	The girl made the bed.	PCI
The man made the door close.	The man closed the door.	GCI

Table 4.6 Items used in Experiment 2B

¹⁸ Although note that this turned out considerably higher than in the final pretest, where it was $\leq 12.5\%$ for any one item.

4.3.1.4 Procedure

The procedure was similar to that of Experiment 1, except that some children were tested by another researcher at the end of an experimental session including other pragmatic and language tasks (N = 12), or as a stand-alone task (N = 2) while others completed the task online under parental supervision (N = 3)¹⁹, as with adults. Social Sciences Research Ethics Committee approval was given for the amended procedure.

In addition, two versions of the online task were carried out with adults: one with justifications, and one without. As there was no difference in the results, the data is collated for analysis.

4.3.2 Results

Note that there was an experimenter error: for six participants, the arrangement of picture pairs as left/right was not counterbalanced as specified in the item lists. They are included in the analysis below.

As in Experiment 2A, picture selection is coded as M picture (marked) selection or not, in keeping with the predictions.

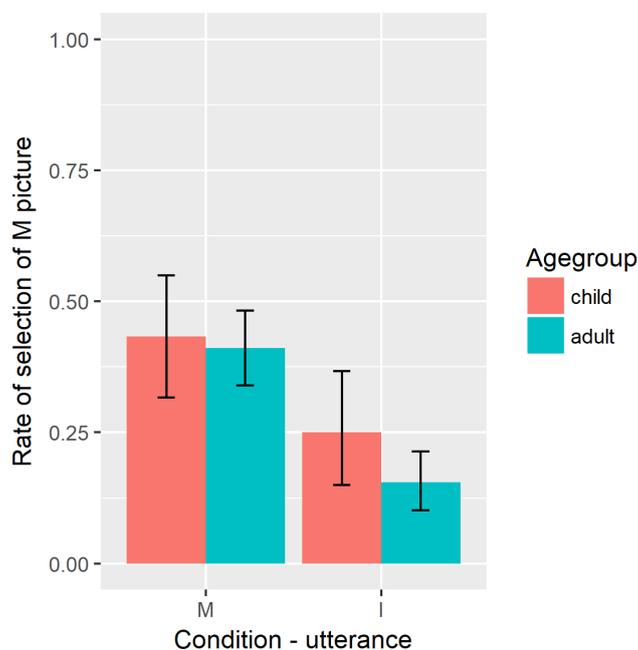


Figure 4.5 Mean selection M picture by condition and age group

¹⁹ Parents were clearly instructed not to help their children; they were all known to the experimenter and had taken part in previous experiments, so were familiar with this idea.

Error bars show bootstrapped 95% confidence intervals for between-subject comparison

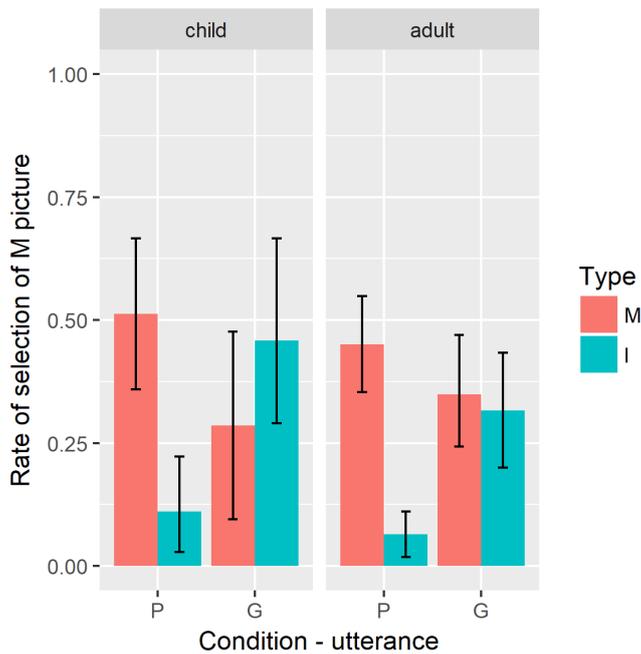


Figure 4.6 Mean selection M picture by Type, condition and Age group
Error bars show bootstrapped 95% confidence intervals for between-subject comparison

		child	adult
I	GCI	.46	.32
M	GCI	.29	.35
I	PCI	.11	.06
M	PCI	.51	.45
I	Total	.25	.15
M	Total	.43	.41

Table 4.7 Proportion of M picture selection by condition, inference type and age group

To examine the effect of condition (marked M vs unmarked I utterance), type (PCI vs GCI), age group (child vs adult), and their interaction, mixed-effects logistic regression models with item (by condition) random slopes and dummy coding were fitted using the *lme4* package in R (Bates et al., 2015; R Core Team, 2016; RStudio Team, 2016). The child age group, manner inference and PCI type were set as baseline.

There was a significant effect of condition, such that for children and PCIs, there are fewer M-picture selections in the I condition than in the M condition ($\beta = -2.47$,

p = .002); and an interaction of condition and type, such that for children, the absolute difference between I and M was greater for PCI than for GCI ($\beta = 3.22$, p = .004). There was no effect of age group, for M and PCI, and no other interactions (Table 4.8).

	Estimate	SE	z	p
	.05	.33	.16	.87
I	-2.47	.78	-3.16	.002
GCI	-0.98	.60	-1.63	.10
Adult	-0.26	.38	-0.67	.50
I:GCI	3.22	1.13	2.85	.004
I:Adult	-0.41	.78	-0.53	.60
GCI:Adult	.55	.67	.82	.41
I:GCI:Adult	-0.53	1.08	-0.49	.62

Table 4.8 Condition ~ Type ~ Age group + (I + Type | Item)
Glmer, family = binomial, optimizer = bobyqa, contrasts = dummy, baseline = M / PCI / Child

4.3.3 Discussion

The pattern of results for Experiment 2B is similar to that of 2A. Both adults and children choose pictures showing a marked interpretation more for marked than for unmarked utterances, and there is no evidence for a difference between adults and children. There is, though, an effect of the type of implicature – PCI vs GCI with the ‘make X Y’ construction. Given the improvements to this task, these findings give more confidence that adults and children do sometimes derive manner inferences. However, I discuss some remaining shortcomings with this methodology below. Given the small sample size for children in this study, the findings must also be taken with caution.

4.4 General Discussion

4.4.1 Manner implicatures in adults

This picture-matching study provides evidence that adult hearers do sometimes derive manner implicatures. However, in comparison to ad hoc and scalar quantity implicatures in similar experimental paradigms, they do so at low rates – for instance, when compared with findings in Chapter 2, or other studies without discourse context (Skordos & Papafragou, 2016; Stiller et al., 2015). This is not as a result of adults responding consistently with either the marked or unmarked picture for marked utterances (‘pragmatic’ and ‘semantic’ responders, Figure 4.7).

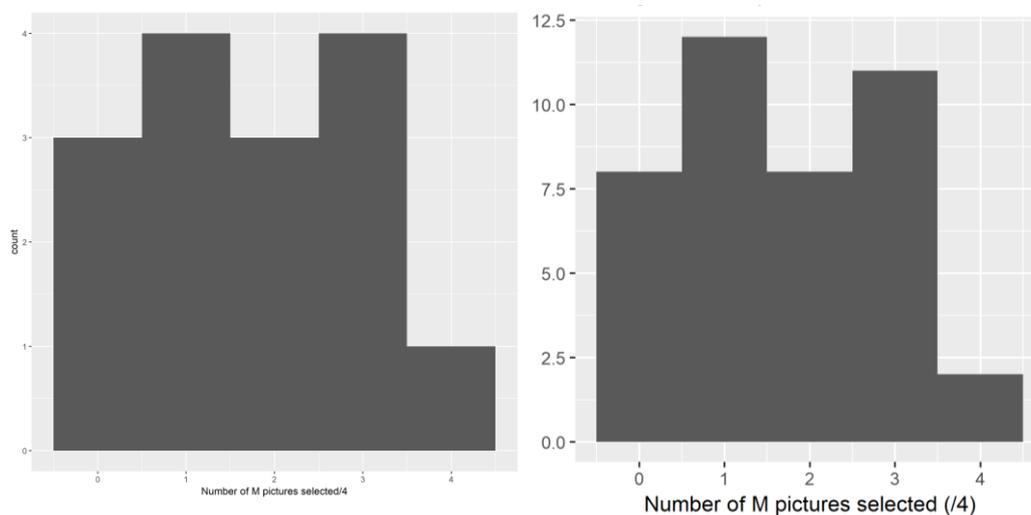


Figure 4.7 Histogram of participant scores for M condition, for children (left) and adults (right) for Experiment 2B

Overall, adults choose a marked picture for a marked utterance around half the time. A chance performance without bimodal distribution in this kind of task would typically be taken as an indication of guessing – of lack of competence. However, I suggest that in the case of manner implicatures, it can instead be taken as representative of the rate of manner implicature derivation. Firstly, it is reasonable to assume that, without reason to think otherwise, hearers will arrive at the stereotypical interpretation (I), rather than choose at random. Secondly, the pretests established that the unmarked (I) utterance was preferred for both pictures – in other words, it is possible to refer to the marked picture with the unmarked utterance, so one can make the assumption that the unmarked alternative is available, when the marked utterance

is heard, and therefore derivation of a manner implicature is, in principle, possible. Thirdly, explicit justifications in the pretests and data collection revealed that choices were largely not based on irrelevant features, and in some cases seem to be based on the form of the utterance leading to a marked interpretation (see Table 4.9).

Utterance	Example justification for M picture
The girl put the duvet on top of the bed.	This is not how you would typically describe the action of making a bed
The woman dipped her brush in paint and put it on the paper.	The painting is more messy, looks more like just paint on paper rather than a nice finished picture
The woman dipped her brush in paint and put it on the paper.	The statement had a random feel about it
The man picked up some things and moved them around the room.	The sentence suggests a rather haphazard action

Table 4.9 Examples of reasons for choosing M picture for M utterance

Generalised vs Particularised implicatures

In this study, there were items that could be called GCIs, with the construction ‘make X Y’, and others that were PCIs. Some have operationalised Levinson’s (2000) default approach to GCIs as predicting higher rates and faster processing of GCIs in non-cancelling contexts because they go through by default, rather than being facilitated by the context (e.g. Breheny et al., 2006; Degen, 2015). On this view, the present results either provide evidence against default inferences, or demand an explanation as to why manner but not quantity implicatures are cancelled in a picture-matching task with single-sentence utterances: adults’ performance was similar in the critical utterance condition for GCI and PCI trials.

However, there was an interaction with condition: for children, the difference in M picture selection between M and I utterance conditions is greater for PCI than GCIs, and the pattern looks similar for adults. This is surprising, but points to the more complex nature of these quasi-conventionalised expressions. Remember that for Levinson, a marked expression such as ‘made the door close’ indicates an unusual or indirect way of doing so. However, studies on the expression and comprehension of causative constructions have indicated that agency and animacy of the agent are

among other factors that affect expression choice (Song & Wolff, 2005; Wolff, 2003). Numerically, the two items that were driving this pattern in Experiment 2B were: ‘the man made the plate break / broke the plate’ and ‘the man made the jar open / opened the jar’ (Figure 4.8). Particularly for the first, an accidental action is likely to be normal or stereotypical, and an intentional one unusual, which might be why a reverse pattern is seen. This indicates that for such cases, extricating the pragmatic inference from semantic and syntactic usage patterns is complex, and depends not only on linguistic but also world knowledge factors.

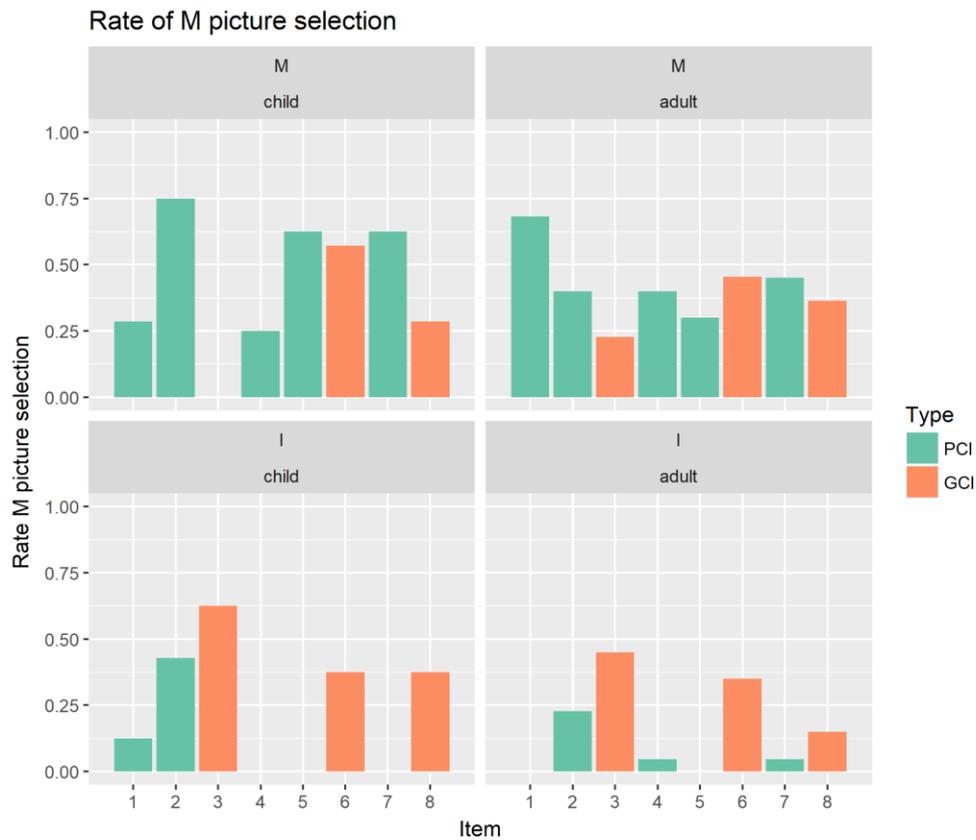


Figure 4.8 Mean selection of M picture by item, condition and age group

Low rates: QUD

Taking the results altogether, though, the question of why the rates of inference are so low still needs to be addressed. One possibility is that if participants are not confident about the interpretation, they are more likely to go with the picture of the stereotypical, more frequent activity. In addition, though, the way different factors contribute to the inference is likely to be more complex than the minimal criteria I set at the outset. In particular, it is worth exploring how manner relates to two aspects of the discourse context: the QUD, and expectations of language use.

In Chapters 1 and 2, I outlined a view of quantity inferences in which the context, and especially the QUD, plays a crucial role – the inferences depend on the relative degree to which the semantic meaning of what is said and its alternative address the QUD (following Degen & Tanenhaus, 2014; Russell, 2012). As a reminder, consider the following contexts for the utterance ‘I broke some of the plates’:

a) Did you break all of the plates?

b) Did you break any of the plates?

(see Degen & Tanenhaus, 2014 p.32)

For (a) the alternative *all* maximally addresses the QUD whereas *some* (*and possibly all*) does not; for (b) both *some* and *all* address the QUD. Typically, in context (a) an implicature is derived, whereas in context (b) it is not. How might manner inferences depend on the QUD? As the alternatives are semantically similar, the relationship must be different from scalars: whatever the QUD, the semantic content of what is said and its alternative can both address the QUD to the same degree. However, the possible implicated meanings – whether to the stereotypical or unusual – may address the QUD to different degrees, so the QUD might constrain utterance interpretation in this way. For example, consider the following, where it seems as if (a) is more likely to lead to a stereotypical, unmarked (I) interpretation, and (b) a marked one:

a) What happened next?

John made the door open.

b) How did you get in without a key?

John made the door open.

In the current task, there was no explicit QUD provided, which could have led to the low rates of manner inferences, especially if an assumed QUD was something like ‘What did she/he do?’.

Furthermore, the QUD could be important for determining the inferred manner implicature in terms of *how* the referent is atypical. For example, consider the following contexts:

a) What was that bang?

The man made the door close.

⇒ The man closed the door by pushing it violently.

b) I thought the automatic door was stuck open.

The man made the door close.

⇒ The man closed the door by overriding the electronic controls and shutting it manually.

Since participants had to infer a QUD in this study, it is plausible that different participants may have done so differently, which then would mean that the stereotypical and, especially, unusual instantiations depicted in the pictures may have matched participants’ own interpretation to a greater or lesser degree – which, again, might contribute to low levels of manner inferences. This is a factor that could be explicitly manipulated in future work, or controlled for by using an alternative methodology, which I will discuss further below.

Low rates: language use

Another aspect of the context needed for manner inferences is the speaker’s language use. Previously, I defined markedness minimally as frequency in context, noting that this would often coincide with other common interpretations of markedness, such as length of form or syntactic complexity. This means that manner inferences are dependent on linguistic experience – both in general, and with the particular interlocutor and discourse. For instance, the speaker may be a first- or second-language speaker; the context may be one where, say, legal or medical jargon is required, or a particular sociolect expected. This may mean that, from the hearer’s point of view, the intentional use of a marked expression may be more or less obvious,

depending on how accurate their expectations about the interlocutors' language use are.

An illuminating parallel here might be psycholinguistic studies relating to saliency and speaker-specific expectations, within a predictive model of language processing. For instance, Jaeger and Weatherholtz (2016) explore a quantitative model of salience for sociolinguistic variants, proposing that 'the salience of a lectal variant is *inversely* related to frequency – specifically to the expected relative contextual frequency of the variant' (2016: 2–3). The greater the salience, the greater the expectation violation, and the greater the surprisal effect – new information gained from the input – that can be observed through behavioural and neural correlates. In the case of a sociolinguistic variant that the listener has not encountered before, expectations about future language use by that speaker are adapted, possibly proportionally to the variant's initial surprisal. Similarly, for manner implicatures, the marked form is one that violates expectations of syntactic or lexical production, given the speaker's language use (including sociolectal or stylistic features from the discourse context, the speaker's own variety, first or second language competence, and so on); the first step in a manner implicature derivation likely consists of the same processes. However, rather than adapting to the speaker's language use, the hearer recognises that this effect could be intentional on the part of the speaker, and looks for the intended meaning.

An open question, therefore, is how hearers decide to treat marked utterances as *intentionally* marked, given that there could be many other plausible explanations, such as the speaker's idiosyncratic usage or a planning failure. Potentially, the use of other cues might help, such as particular prosodic patterns or discourse particles (e.g. 'well, she put the duvet on top of the bed'). As these were lacking in this picture-matching task (as the utterances were either read or spoken with neutral intonation), participants may have been less sure about whether the markedness was intentional, leading to lower rates of manner implicatures. Also, taking expectation violation as part of a manner inference suggests another shortcoming: with few items, and equal numbers of control (unmarked) and marked utterances, it could be that participants did not have enough experience of the speaker to be confident in expectations about the speaker's language use, and therefore attribute intentionality rather than

accommodate a marked utterance. The large proportion of marked utterances might also lead the hearer to accommodate them as normal for the speaker's style.

4.4.2 Manner in development

The results are suggestive of a similar competence with manner implicature for children aged 5–7 and adults, particularly with PCIs. For GCIs, of the periphrastic causative 'make X Y' type, they perform differently.

Firstly, this suggests that though manner might be expected on theoretical grounds to be the most challenging of implicatures, children are approaching adult-like competence not long after they do so with scalar implicatures (see Chapter 2), ruling out a strong version of the first hypothesis – that manner implicatures are acquired *very* late. This finding contributes to the overall trend across experimental pragmatics studies in recent years, of results that reveal pragmatic competence earlier than previously thought, right down to preschool age in the case of relevance and ad hocs (e.g. Papafragou & Skordos, 2016). At the age in this study, children are competent with reasoning by exclusion and contrastive inference (seen in quantity implicatures), so these findings suggest that they are also learning sensitivity to markedness.

However, the present studies cannot clearly distinguish a version of the first hypothesis that simply predicts that manner is acquired after scalar implicatures, on the one hand, and the second hypothesis that children may actually have the ability to derive some manner implicatures early, on the other hand. If the former, then in a study with a wider age range, a progression in which children become competent with manner just after scalars would be expected. If the latter, then some competence with some manner inferences would be expected early, around the same time as relevance inferences, and other manner inferences would be seen later; alternatively, a U-shaped trajectory might be observed, as children relax their expectations of conventionality, and then acquire the ability to derive manner inferences in a more adult-like way.

Like adults, children seem to find it more difficult to detect GCIs than PCIs; the difference between the two conditions was greater for PCIs. This could be indicative of the challenge of detecting the markedness of alternations in syntactic constructions like this in general. Examples of GCI marked utterances in the literature tend to vary

much less from the unmarked alternative than PCI marked utterances. It is therefore plausible that they are less salient, and hence that it is more challenging for a child with less linguistic experience to recognise them as marked or to attribute them to an intentionally meaningful use.

Alternatively, this pattern may be specific to the causative construction used. I noted earlier that there may be several interacting patterns of usage for causative constructions that might contribute to adults' expectations: the periphrastic is associated with unintentional actions, a focus on manner of motion or status change, and describing atypical situations (Kline et al., 2013; Levinson, 2000; Song & Wolff, 2005; Wolff, 2003). There is further evidence that children's expectations may be different again. Muentener and Lakusta (Muentener & Lakusta, 2011) found that children have an 'intention-to-CAUSE' bias: in comprehension and production, children used more causal language (whether lexical or periphrastic, e.g. 'break', 'make break') for intentionally caused motions or changes of state, and more non-causal language (e.g. 'fall') for unintentionally caused changes of state or motion. This would explain nicely why 'make the door close' and 'make the jar open' pattern together numerically in the M condition, in contrast to 'make the plate break': for the former two, both pictures show intentional actions and so children might be expected to choose the marked or unmarked picture more or less evenly; for the latter, only the unmarked picture is intentional. One would, however, therefore expect a preference for the unmarked picture across marked and unmarked utterances, which is not the pattern seen.

Future studies on the development of manner implicatures would therefore benefit, firstly, from taking into account relevant aspects of children's syntactic and semantic development, and, secondly, from a greater number of more similar items, so that any systematic variation between types of implicature can be observed. It may be that, as for scalars with *some*, the causative is a widely cited but atypical example of a marked construction that can trigger a manner inference. Furthermore, it seems likely that factors which boost children's performance with quantity implicatures would have a similar effect on manner inferences. In particular, having a productive knowledge of alternatives is clearly a prerequisite, but additionally making those alternatives salient in context is likely to aid children's inferencing. Note, though, that

simply providing the alternative, unmarked expression in context introduces an ambiguity in interpreting the results: it can then function as a referential pact, so any response could be a reaction to breaking a referential pact, or drawing a contrastive inference, or deriving a manner implicature. In addition, making the QUD explicit about how an action was accomplished – whether in a stereotypical way – might help, just as highlighting quantity boosts quantity implicature rates, again by making the alternatives more salient.

4.4.3 Challenges for investigating manner

The picture-matching study presented here highlights the challenges of investigating manner implicatures experimentally, in terms of both the stimuli used and methodologies employed.

Firstly, while, in principle, manner is a close parallel to the second quantity maxim, in practice it can be challenging to distinguish the contributions of manner and the first quantity maxim. The items in the picture-matching study reflected the kinds of manner implicature examples given in the literature (particularly by Levinson, 2000, following my previous work). However, whether the marked utterance was semantically similar to the unmarked alternative, or was instead less informative in semantic content, was not always clear. Were they therefore treated as manner or as quantity inferences? The justifications given by adult participants suggested that sometimes informativeness was applied, and sometimes manner. Table 4.10 presents the PCI items used in the Experiment 2B; how they could be interpreted in terms of informativeness; and a suggestion of how this might be improved in a future study. For future studies, more norming or some formal test of semantic similarity is required, in order to reduce possible quantity interpretations and ensure that manner inferences are being tested. In addition, with more picture choices, for example of the unmarked, marked and potential quantity-implicated meanings, it might be possible to tease apart hearers' interpretations, without an explicit justification measure.

Secondly, established experimental paradigms that have been successfully employed to investigate adults' and children's understanding and processing of other types of pragmatic inferences may not be well suited to manner. In the case of picture-

matching tasks, the potential issues are twofold: firstly, the implicated meaning and its alternative are not as clearly defined as for quantity implicatures, so, secondly, the pictures available may not be such accurate depictions of their meaning. Also, the picture choice indicates that the hearer has derived an implicature, but cannot rule out some other feature of the utterance or picture causing the choice.

M utterance	I utterance	M meaning	Q meaning	Improved M utterance
The woman dipped her brush in paint and put it on the paper.	The woman painted a picture.	It was unusual painting.	She only put it on the paper, but didn't paint anything; she started to paint.	The woman put paint on the paper to make a picture.
The woman made small pieces of apple.	The woman cut up the apple.	She cut it in an unusual way.	She made small not big pieces of apple.	The woman made pieces of apple.
The boy pushed a cloth across the table.	The boy cleaned the table.	He cleaned it ineffectually.	The boy pushed a cloth across the table only once.	The boy pushed a cloth all around the table.
The man picked up some things and moved them around the room.	The man tidied the room.	Aimlessly, it couldn't be called tidying.	He only moved them but didn't put them away.	The man moved things around the room to their own places.
The girl put a duvet on top of the bed.	The girl made the bed.	She didn't make the bed properly.	She put it on top of the bed but didn't tuck it in.	The girl spread a duvet on top of the bed / tucked a blanket in on the bed.

Table 4.10 Possible quantity and manner inferences for Experiment 2B stimuli

Some kind of act-out task might allow for individual differences in how the manner implicature is enriched, although much more challenging to execute and to code. Another option that would remove some of the ambiguity of pictures would be to use short videos, which could, for example, make it clear that actions are finished. For adults, a reading-time measure might have potential: following the method in Bergen and Grodner's (2012) study on quantity implicatures and epistemic state, a slowdown would be predicted with a follow-on utterance that conflicts with the

implicated meaning, as compared with one that concurs with it. For children, this could be adapted as a story-continuation task, for instance:

a) The man tidied the room / The man moved things around the room to their places. – His wife came in and was happy when she saw it / His wife came in and was unhappy when she saw it.

b) The woman painted a picture / The woman put paint on the paper to make a picture. – Her friend saw it and said, ‘what a beautiful picture.’ / Her friend saw it and said, ‘what an interesting picture.’

Finally, on a constraint-based model of implicature, many factors are expected to contribute to hearers’ inferencing of manner implicatures, including expectations about speaker language use and the QUD as already discussed. More information about the speaker – in the form of more exposure – might help hearers to form expectations about the speaker’s language use, and be more confident about implicated meanings. Other factors that might contribute also need to be controlled for or explicitly manipulated in future research: the use of discourse particles (e.g. ‘well!’) or prosody, and the degree of markedness (infrequency or length of the expression chosen), both of which might cue a marked meaning.

4.5 Conclusion

In this chapter, I argued that manner inferences can be defined as a distinct category of inference theoretically, but found that in practice they can be hard to disassociate from quantity inferences. Based on theory and existing empirical findings, I put forward two hypotheses for children’s development of manner inferences – that they might be the most challenging and acquired late, or, alternatively, that some might be available early. The experiments provided suggestive evidence that both adults and children sometimes derive some manner inferences, and thereby rule out very late development of manner inferences in children. However, much more research and new experimental methods are needed to tease apart manner and quantity inferences.

5 Putting the pieces together: epistemic state and implicatures

Man: I might need one of those.

Woman: Sorry?

Man: I might need one of those.

Woman: One of what?!

Man: The thing you're holding.

Woman: Oh, a pain au chocolat!

(Overheard in campus café)

This reference resolution failure was initially perplexing – until I realised that the woman was blind, and so unable to pick up on the cues like pointing and eye-gaze that her breakfast-partner was using to disambiguate his utterance; he, of course, was failing to take into account this difference in their common ground, in particular with visual perspective-taking.

In this chapter, I turn to the development of a particular aspect of implicature inferences: taking into account the speaker's epistemic state. Remember the puzzle that unfolded in Chapter 3. Young children seem to be pragmatically competent – if not adult-like – at an age where full Theory of Mind, *as measured by* explicit False-Belief tasks, is not consistently demonstrated. Further, in Experiment 1B, Theory of Mind score was not a significant predictor of implicature score, except in the bilingual group. However, an association – or lack of it – between two different tasks is rather a crude instrument for revealing whether ToM development and implicature development are linked, and, if so, how. It is not clear what children are doing in classic implicature tasks in which there is no difference in perspective between the interlocutors. This leaves us with two main questions:

1. Do children engage in perspective-taking in implicature derivation? In other words, even when they look pragmatically competent, are they doing the same thing as adults are thought to do?
2. If not, what does this mean for a Gricean model of pragmatics, and implicature in particular? Do we need a new model for development?

I seek to address these questions in this chapter. I concentrate on tracking the

speaker's epistemic state, and in particular perspective-taking: taking someone else's visual point of view, where the assumption is that what someone else can see is what they know.

Perspective-taking is typically divided into two levels (Flavell, 1977; Flavell, Beilin, & Pufall, 1992): level 1 'assesses what someone else can see', while level 2 'requires participants to adopt someone else's spatial point of view to judge how that person sees a particular visual stimulus' (Ferguson, Apperly, & Cane, 2017: 1646). The first is involved, for example, in reasoning that because of an obstacle, another person cannot see something that you are able to see – imagine how, when moving a large item of furniture, you might tell the person walking backwards to watch out for a step; it would also be needed to successfully play the children's game of hide and seek. The second level is needed, for instance, in Piaget & Inhelder's (1967) famous three-mountain test, where children are asked which pictures show what other people see when they look at a display of three different-sized model mountains from different angles.

In this chapter, I use the term *epistemic state* in its general sense, as typically used in pragmatic theory: the knowledge in question could be acquired through various means, not just visually. For the second experiment, I focus on visual *perspective-taking* as a case of representing and reasoning about *epistemic state*. By *perspective-taking*, I have in mind level 1 perspective-taking, unless otherwise specified.

5.1 Reasoning about others' epistemic states

In Chapter 3, I explained how Theory of Mind is inherent to the Gricean project, a) in general, as necessary for intentional communication, and b) specifically for reasoning about others' epistemic states in implicature derivation, as part of the Competence Assumption which leads to the Epistemic Step. Whether or not an implicature is derived depends partly on whether the hearer thinks the speaker is fully informed about relevant content – whether knowledge or ignorance is assumed. A particular case of this dichotomy is whether relevant information is in common ground – what both the speaker and hearer know that they both know, whether that be through visual access or shared experience – or in privileged ground – if only the

hearer has access to the information (H. H. Clark, 1996).

In this section, I review what is known about perspective-taking in general and in communicative situations in particular, in adults and in children. Adults are able to take into account another's perspective in utterance interpretation – even if it is not yet clear whether they always do so. Children begin to show perspective-taking abilities early, but whether they do so in implicature derivation remains an open question.

5.2 In adults

5.2.1 Perspective-taking

Typically-developing adults demonstrate full Theory of Mind abilities, including perspective-taking: they can pass false belief and other more complex second-order ToM tests (Cummings, 2015; O'Grady, Kliesch, Smith, & Scott-Phillips, 2015; Wellman et al., 2001). One pragmatic phenomenon that has been extensively studied in conjunction with perspective-taking is reference resolution – the production and interpretation of referring expressions in an optimally informative way, which allows both interlocutors to uniquely identify the intended referent (for a brief overview see Heller, Parisien, & Stevenson, 2016). Here I touch on this rich field of study, firstly because one experimental paradigm used will be important for Experiment 4 here, and secondly because the findings on reference inform expectations for implicatures with reasoning about the speaker's epistemic state.

In one typical reference resolution task, commonly known as the director or cubby-hole task – a type of visual-world paradigm (Snedeker & Huang, 2016) – a speaker and hearer are engaged in an activity in which the speaker directs the hearer to pick up or move certain objects in an array – typically a set of shelves or cubby-holes – with the speaker and hearer seated on opposite sides. The interesting catch is that some of these objects can be seen only by the hearer, and not by the speaker, because of a screen between the speaker and those objects. In other words, they are in privileged ground, rather than common ground (H. H. Clark, 1996). This requires the hearer to take into account the speaker's perspective in interpreting the referential utterance, by representing the fact that the speaker cannot see that object, and

reasoning that the speaker therefore cannot know about that object, and so cannot be referring to that object (even if, semantically, the object is a good match).

Findings from offline and online measures (including eye-tracking and reaction time) have varied, and have given birth to a lively and current debate. Eye-tracking studies work on the assumption that eye-gaze mirrors cognitive processes that are not necessarily expressed in language or action: if the participant looks at a particular object, he is considering it as a possible referent. Some such studies have found that participants are sometimes able to use their knowledge of the speaker's epistemic state to anticipate the referent, before the critical part of the utterance is fully complete (e.g. Hanna & Tanenhaus, 2004; Hanna, Tanenhaus, & Trueswell, 2003; Heller, Grodner, & Tanenhaus, 2008). A particularly relevant case comes from Heller, Grodner and Tanenhaus (2008), who found that hearers either did or did not make a contrastive inference online, depending on whether a contrasting object was in common or privileged ground: for instance, upon hearing 'pick up the big...', hearers start to look to the big duck, with the small duck in common ground, rather than the big box, with the small box in privileged ground. In other words, pragmatic expectations of informativeness interact with perspective-taking. Their methodology also overcomes the problem of the privileged ground object being a better match for the utterance than the common ground object, found in some other studies. However, other studies have found that participants may look at both objects, or even at the privileged object first (e.g. Epley, Morewedge, & Keysar, 2004; Keysar, Barr, Balin, & Brauner, 2000; Keysar, Lin, & Barr, 2003). In Keysar and colleagues' studies, participants are even reported to sometimes pick up or point to the privileged ground object when it is a better match for the utterance – 71% of participants in one experiment did this at least once, for instance.

Overall, it seems sometimes participants do not only consider the interlocutor's perspective, but also their own; there are egocentric as well as altercentric biases that may themselves be weighted by a wide variety of contextual factors (Hawkins & Goodman, 2016). However, though the degree to which adults engage in perspective-taking varies across experimental contexts and possibly individuals, by and large, adults can ultimately end up with the intended referent by taking into account others' perspectives, particularly when there are not strongly conflicting cues. Furthermore,

this process can, at least in some circumstances, be incremental, with anticipation of referents based on the speaker's perspective.

5.2.2 The epistemic step in implicature

Given a Gricean model of implicature and the findings from reference resolution studies, there are three basic expectations for adults' implicature interpretation:

1. Hearers will derive an implicature if the Competence Assumption is met.
2. Hearers will not derive an implicature if the Competence Assumption is not met; they may derive some other inference, like an ignorance implicature.
3. Hearers may use the speaker's epistemic state to anticipate implicated meaning ahead of hearing the critical constituent in the utterance.

In this section I review existing studies on adults, in order to show a) that these expectations are borne out, but b) that no existing paradigms can easily be used with children, so a new one has to be developed. As there are only a limited number of studies, I look at any manipulating speaker epistemic state, not just visual perspective.

First, two sets of findings suggest that whether hearers derive a scalar implicature does indeed depend on the speaker's epistemic state. In an offline study, Goodman and Stuhlmüller (2013) look at scalar implicatures using a 'betting measure', in which participants read unfolding scenarios about a speaker who has incomplete or complete knowledge. In the example below, the speaker opens 2 out of 3 or 3 out of 3 envelopes, before making a statement about the envelopes' content:

Letters to Laura's company almost always have checks²⁰ inside.

Today Laura received 3 letters.

How many of the 3 letters do you think have checks inside?

Participant distributes \$100 to bet on 0, 1, 2 or 3 envelopes.

Laura tells you on the phone: 'I have looked at 3 of the 3 letters.

Some of the letters have checks inside.'

²⁰ British English *cheques*

Now how many of the 3 letters do you think have checks inside?

Participant repeats the betting action.

(Goodman & Stuhlmüller, 2013: 178)

The idea is that bets on 2/3 will be higher than bets on 3/3 if the participant has drawn an implicature of *some but not all*. They find that when the speaker has seen inside all the envelopes, as above, bets on 2/3 are indeed higher than on 3/3, and that this pattern is reversed when the speaker has not seen inside all of them. That is, their findings support the theory that hearers take into account a speaker's epistemic state. However, one drawback of their methodology is the fact that participants are told that 'envelopes almost always have cheques inside' in order to make sure that apparent implicatures (betting high on 1 or 2 out of 3) are not due to a prior belief that it was unlikely that many envelopes would have cheques inside. This stacks the situation in favour of them finding a difference between the conditions where the speaker has seen inside 2/3 and 3/3 envelopes: hearers are more likely to guess or bet on 3 out of 3 envelopes containing a cheque when the speaker has only seen inside 2 of them because of this knowledge. This means that it might look as if the implicature has not been derived because of ignorance, but really it could be because of this world knowledge (a different effect of context). In addition, the betting measure itself is one that may tap into participants' conscious or metalinguistic reasoning, rather than purely implicature interpretation. This makes it inappropriate for adaptation for children.

Further evidence comes from Bergen and Grodner's (2012) self-paced reading task. The assumptions here are that a) implicatures and b) accommodation of unexpected contents will both show up as longer reading times. They present participants with scenarios such as:

At my client's request, I [meticulously compiled / skimmed] the investment report.

Some of the real estate investments lost money.

The rest were successful despite the recent economic downturn.

(Bergen & Grodner, 2012: 1452)

They find an effect of speaker knowledge: at the scalar trigger ('some'), reading times are longer when the speaker is fully knowledgeable than partially knowledgeable. But for the continuation ('the rest'), reading times are longer with a partially-knowledgeable speaker – this continuation is unexpected when the anaphor has no established complement set (*some but not all*). Their findings further suggest that this kind of pragmatic processing and integration of contextual information happens in an incremental fashion, as the delay in reading time was observed immediately at the scalar trigger. This also goes against those accounts that might predict automatic generation of scalar implicatures in particular, such as a Default account (Levinson, 2000) or some grammatical accounts (e.g. Chierchia, 2004). Unfortunately, though, the self-paced reading paradigm is also not suitable for young children.

The final study corroborates these findings, but for ad hoc quantity implicatures, and also shows that speaker's epistemic state can be integrated in an anticipatory fashion. Breheny, Ferguson & Katsos (2013) use an eye-tracking paradigm and observe higher rates of anticipatory looks, indicative of an exhaustive inference, in the knowledgeable speaker condition than in the ignorant (or partially knowledgeable) one. In the critical case, both the speaker and hearer watch a video in which a woman places an object, say a spoon, into Box A, then a spoon into Box B, and then a fork into Box B – but the speaker does not see the last action. The speaker then describes the video, although not necessarily in sequential order. In this partially-knowledgeable condition, the utterance 'the woman put a spoon into Box A' is completely ambiguous until the final word, so no anticipatory looks to the target were observed. In contrast, in the condition where the speaker sees the whole video, the participant is able to anticipate the location as soon as 'into' is heard, on the assumption that the speaker is being fully informative, and means *only a spoon* – this is exactly what is seen in the eye-gaze fixations. Note that this study – like Goodman and Stuhlmüller's (2013), but unlike Bergen and Grodner's (2012) – makes explicit and precise what the speaker does and does not know. This makes the interpretation of participants' responses much more transparent – a point I shall return to in discussing the first study below.

Some of these studies – particularly Bergen and Grodner's (2012) – rely on assumptions about the cognitive cost of implicature derivation. There is still some

debate about whether deriving implicatures and integrating contextual information is cognitively costly or not, with evidence largely coming from studies that examine other aspects of context, such as upward- or downward-entailment or the QUD (e.g. Breheny et al., 2006; Politzer-Ahles & Fiorentino, 2013). Remember, however, that it is not straightforward to extrapolate processing predictions from a pragmatic model concerned with the ‘what’ and the ‘why’ – and this kind of model is also the concern of the offline developmental experiments here. I shall not discuss this processing aspect any further at this point, though I will return to it in discussion below.

5.3 In children's development

5.3.1 Perspective-taking

Children develop level 1 visual perspective-taking skills relatively early: Moll and Tomasello (2006) demonstrate that by age 2, children are able to give an adult the toy that only the child but not the adult can see, when the adult appears dissatisfied and is visually searching for ‘the other toy’, and the child and adult have previously experienced both toys together. When the adult simply asks for the toy, they instead give the mutually visible toy. There is a significant difference between these two conditions, although 2-year-olds are not yet at ceiling performance. This is a skill that arguably develops from an earlier capacity for joint-attention; Moll and Meltzoff (2011) argue that the role of social interaction is crucial here.

It seems puzzling that young children are able to respond appropriately based on shared (or non-shared) experience with adults (Moll et al., 2006) before they exhibit level 1 visual perspective-taking. Moll and Meltzoff (2011) point out that in visual perspective-taking tasks, the adult is by necessity present and interacting with the child, while in experiential ones, the adult typically leaves, cutting off social interaction. They postulate that children tend to over-generalise what is shared in interactive situations, meaning that only when the adult stops the social interaction can the child recognise the adult’s ignorance (as confirmed by Moll, Carpenter, & Tomasello, 2011). Crucially, starting from age 2, children are able to recognise an adult’s ignorance in visual perspective-taking if strong cues are given. Level 2 perspective-taking follows later, potentially underpinned by development in executive

functions in the preschool years (Diamond, 2006), and at the same time as epistemic perspective-taking (as in false belief tasks). Thus, 3-year-olds largely fail level 2 tasks, whereas 4½-year-olds largely pass (Moll & Meltzoff, 2011). The understanding that seeing leads to knowing is also evident explicitly, as well as implicitly, from age 3 or 4 years (Robinson, 2011). These abilities are therefore largely in place at the age of children in this study, 5–6 years.

5.3.2 Perspective-taking and communication

Many studies have shown that young children are also able to use their perspective-taking abilities in communicative situations: 18-month-old infants can adapt pointing gestures according to an interlocutor's epistemic state (Diesendruck & Markson, 2001; Liebal, Carpenter, & Tomasello, 2010); 24-month-old infants can use experiential perspective-taking in word learning (Akhtar, Carpenter, & Tomasello, 1996); 2-year-olds can differentiate test questions, where the speaker actually knows the answer, from genuine questions (Grosse & Tomasello, 2012); and children aged 2 and upwards also adjust their production to be more informative when the referent is not visually available (Grigoroglou & Papafragou, 2016; Matthews, Lieven, Theakston, & Tomasello, 2006; O'Neill, 1996). Importantly, two studies have used the director-task paradigm to investigate children's ability to integrate perspective-taking with reference resolution.

Remember that in the director-task paradigm, in the critical condition, there are two objects matching the utterance that are visible to the hearer, one of which is hidden for the speaker; in a baseline condition, the hidden object cannot be a semantic match for the utterance; and in an ambiguous condition, both object matches are in common ground, with the other two items in the display being distractors. The speaker gives instructions such as 'pick up the cup'. Nadig & Sedivy (2002) tested a small group (N = 15) of 5- to 6-year-old English-speaking children using eye-tracking, and found that the critical and the baseline conditions pattern together, with participants looking to the target object more quickly than in the ambiguous condition, although there was some effect of the privileged object in the critical condition.

These findings are corroborated by Nilsen and Graham (2009), who report not

only looking-time results from 5-year-olds (N = 61), but also the object selection results. Children were significantly less likely to pick the alternative target object when it was in privileged ground (mean = 13%), than when it was in common ground with the target object (mean = 56%). This suggests that in an experimental context where the difference in perspective is current and salient, and the objects in common and privileged ground are of similar saliency, young children are able to take into account the speaker's epistemic state in the pragmatic task of reference resolution. This means that this is a suitable experimental paradigm for assessing perspective-taking, and motivates an investigation of the same skills with implicatures.

5.3.3 Perspective-taking and implicatures

Until very recently, no studies have attempted to investigate children's ability to take into account the speaker's perspective in implicature derivation. Two studies, however, have set out to investigate 'ignorance inferences', which are related to, but clearly distinct from, quantity implicatures proper (Hochstein et al., 2016; Papafragou, Friedberg, & Cohen, 2017).

An ignorance inference is derived when the hearer concludes that the speaker does not know whether the more or less informative proposition is the case, reasoning that if the speaker knew that the more informative one is the case, she would have said so, she is being co-operative, and there is no evidence that she *does* know. That is, an ignorance inference is an inference *about* the speaker's epistemic state. An ignorance inference *par excellence* is the disjunctive case, such as 'Bob has a Peugeot or a Ford'. Upon hearing such an utterance, the hearer can reason that if the speaker knew that Bob has a Peugeot, she would have said so, and likewise with the Ford. Since she did not, the hearer can conclude that she is not sure which of the stronger utterances is true (nor can she know that either is false, else this utterance would not be co-operative on the level of quality). Therefore, the hearer infers that the speaker does not know – or is ignorant of – whether Bob has a Peugeot or Bob has a Ford, but does know that one of these is true (cf. Hochstein et al., 2016: 114–115). The same kind of inference occurs with other quantity implicatures as well where the Competence Assumption is not met (as explained with the example 'The puppet picked the card with rabbits' in section 3.1.1).

Hochstein et al. (2016) and Papafragou, Friedberg and Cohen (2017) test 4- and 5-year-olds with a similar experimental design. To take the latter as an example: children watch two videos, which each feature a girl performing an action (e.g. painting a star), and an observer. In one video, the observer watches the entire action; in the other, the observer falls asleep part way through. Participants then hear an utterance, such as ‘The girl coloured some/all of the star’ and are asked, ‘Who said that? Point to the friend who said that.’ In both studies, 5-year-olds, but not 4-year-olds, are able to attribute the under-informative statement (‘some of the star’) to the ignorant speaker (at above chance rates, $M = 63\%$, in Papafragou et al., 2017, and no different from the control trial, $M = 75\%$, in Hochstein et al., 2016).

Papafragou, Friedberg and Cohen (2017) follow up these results with two further studies to try to isolate the 4-year-olds’ difficulty. They found that in the same task without an epistemic state manipulation (i.e., akin to a standard scalar implicature picture-matching task), they are above chance; but in a non-communicative version of the task, in which participants are shown both videos, followed by a picture card taken from one of the videos, and then asked ‘Who saw that?’, they are at chance. This seems to suggest that the younger children’s difficulty lies not in the scalar inference derivation, but instead in the tracking of epistemic states.

This is surprising, however, given the evidence for early development of perspective-taking I reviewed above. One explanation, suggested by Papafragou, Friedberg and Cohen, is that this has to do with the more complex task demands in this particular case. One factor is the need to remember both scenes, as they are videos rather than pictures and neither is available visually at the point the test card is produced (the final frame is visible, but in this the girl has covered up the drawing). This contrasts with other perspective-taking tasks, like the director task, where the difference in perspective is always available and salient to the participant. Another factor they mention is the need for an overt response, comparing this to the difficulties 4-year-olds still have with false-belief tasks (Wellman et al., 2001). An alternative explanation might be that the task in fact does not effectively tap children’s knowledge. Despite the fact that the second follow-up experiment was meant to be non-linguistic, it did rely, crucially, on a question, ‘Who saw this?’. As they point out, like a weak scalar utterance, the ‘weak’ picture card is indeed true of both girls – they

both see the star half-painted at some point. What is needed here is in fact a pragmatic inference to *Who saw only this?* Or, as they explain, a realisation that the weaker picture is not a good representation of everything that the fully-informed observer saw. It is perfectly possible that this particular communicative situation does not provide enough cues for an exhaustive QUD. Note that for these experiments, only 4-year-olds were tested, so it is an open question as to how 5-year-olds would respond, although a control group of adults performed as predicted. In other words, from this study, it cannot be firmly established whether 4-year-olds' difficulty with ignorance inferences is due to the tracking of epistemic knowledge per se, or to the integration of this knowledge into a pragmatic inference.

The conclusion put forward by both studies is that 5-year-olds can derive ignorance inferences, although they are clearly still developing this ability. However, this is not the only interpretation of their findings. At best, the tasks can be said to be a very supportive or scaffolded context for ignorance inferences, as the epistemic state of the (potential) interlocutors is highlighted. Alternatively, what they show is some kind of reasoning about epistemic state, but not an ignorance inference in the traditional Gricean understanding: the epistemic states of the interlocutors are known in these studies, and what is required is to match an utterance to one of these interlocutors. In an ignorance inference, on the other hand, only the utterance is known, and the epistemic state of the interlocutor is inferred based on expectations of informativeness and cooperativeness more generally. At minimum, what might be needed to solve this task is an ability to remember interlocutor's epistemic states, and a sensitivity to informativeness. Hypothetically, one could use reasoning by exclusion along the following lines: I, the participant (and the knowledgeable speaker), would have said that the girl painted all of the star; the utterance was 'the girl painted some of the star'; therefore, the knowledgeable speaker cannot have said that; therefore, it must have been the other speaker. Even if this is the case, these studies are still helpful here, in that they show some sort of reasoning about speakers' beliefs in a communicative context at the age of 5 years.

To emphasise, what is definitely not being tested in these studies is the ability to integrate the speaker's epistemic state into an implicature inference, and take the epistemic step or not, as appropriate, despite claims from one that it 'offers the first

evidence in the literature that preschoolers are able to some degree to consult a speaker's epistemic state when computing a scalar implicature' (Papafragou et al., 2017: 11). In these studies, what is at stake is which speaker said something, given the utterance interpretation; with ignorance inferences, what is at stake is the epistemic state of the speaker, given the utterance; in implicature derivation, what is at stake is being able to derive or not derive the implicature, given the speaker's (assumed) epistemic state. The last of these is what I am investigating in this chapter.

More recently (and after the studies in this chapter were carried out), Kampa and Papafragou (2017) test the interaction of ad hoc implicatures and speaker perspective in 4-year-olds. In the experiment, children always see two photographs, one of a 'speaker' behind a display where both objects are mutually visible and another of the same 'speaker' behind a display where one object is occluded for the speaker (see sketch in Figure 5.1). In one condition – the 'strong' condition – children hear an utterance such as 'I see a spoon and a bowl', and are asked, 'which box is she talking about?' In the other condition – the 'weak' condition – children hear 'I see a spoon'. On the whole, children succeed in both conditions, though they do better in the strong condition. The authors' conclusion is that 4-year-olds are able to take into account speaker perspective in implicature derivation.

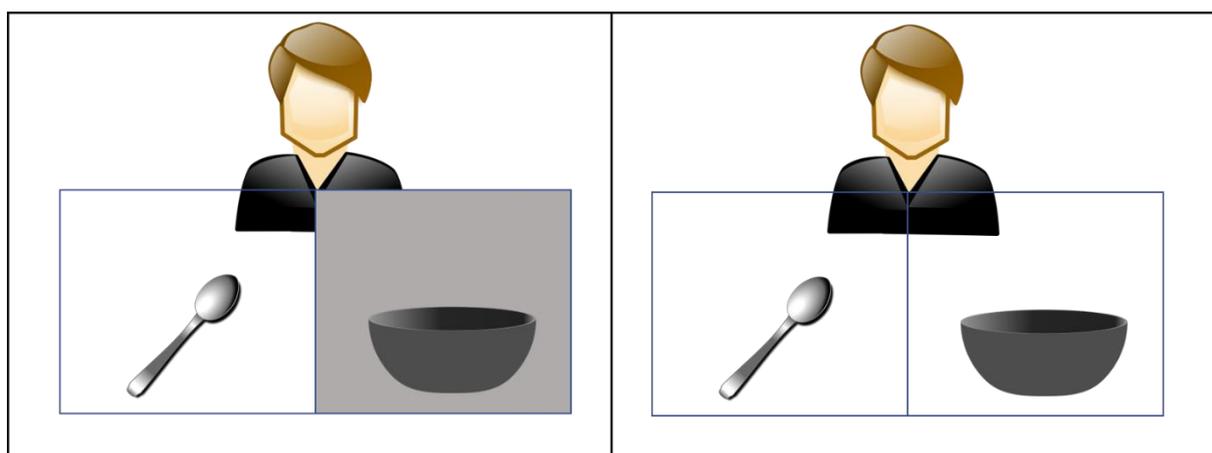


Figure 5.1 Sketch of experimental stimuli from Kampa & Papafragou (2017)

However, what is being tested here is not necessarily the Epistemic Step per se. In the 'strong' condition the utterance could match either box from the hearer's perspective, but only one from the speaker's perspective – so this is straightforwardly

testing perspective-taking. In the ‘weak’ condition, the utterance is again a semantic match for either box from the hearer’s perspective, and a semantic match for either box from the speaker’s perspective, but only a fully informative utterance from the speaker’s perspective for one box (with the occluded object) – or, alternatively, only a match for the exhaustive interpretation for that one box. That is, there are two possible ways to arrive at the correct box selection in the weak condition, along the following lines of reasoning: ‘if the speaker said ‘I see a spoon’ to describe the box where she can see both a spoon and a bowl, she would be under-informative, so it must be the other one’, or ‘if the speaker meant ‘I see only a spoon’ it would not be true for the box where she can see both a spoon and a bowl, so it must be the other one’. Therefore, in this design, both conditions test perspective-taking in referential communication. But on one interpretation neither involve an ad hoc inference, only sensitivity to informativeness, while on the other interpretation, the ‘weak’ condition tests the ability to match an ad hoc inference interpretation to the speaker’s perspective. Note, too, that while the experiment is set up to suggest that the utterance is produced by one speaker about a particular box – it is about reference resolution – the visual display allows for the task to be about matching an utterance (interpretation) to a speaker. In other words, it could be that participants are not answering the explicit question, ‘which box is she talking about?’ but instead ‘which speaker said that?’, just as in Papafragou, Friedberg and Cohen’s and Hochstein et al.’s tasks. Therefore, whether children are able to derive or not derive a quantity implicature depending on the speaker’s perspective is still an open question.

5.3.4 Summary

In sum, there is evidence that adults are able to take into account a speaker’s perspective in utterance interpretation, including in implicature derivation – although the extent to which the process itself combines egocentric and altercentric biases is still a matter of debate. In children, perspective-taking itself emerges early, and in communicative interactions from the second year of life. By age 5, children are competent in level 1 perspective-taking, and are likely to be able to combine this with resolving reference, and matching an utterance to a speaker. They can also derive

quantity implicatures (both ad hoc and scalar), where speaker epistemic state is not at issue. What is not known is whether they can combine all these skills.

5.4 Predictions

By combining expectations from pragmatic theory with what is already known about perspective-taking and implicature understanding in children and adults, two distinct hypotheses can be formulated about children's development of perspective-taking in implicature inferences.

5.4.1 One-step hypothesis

This is the fully Gricean hypothesis: that taking into account the speaker's perspective in implicature derivation develops along with the implicature derivation itself, in one step. The prediction, therefore, is that when children start to make implicature inferences at adult-like rates in situations where perspective is shared between interlocutors – where relevant beliefs are in common ground – they will also make inferences at adult-like rates when this is not the case, where there is privileged ground or possible ignorance on the part of the speaker.

This hypothesis rests on the notion that to be a competent pragmatic communicator requires Theory of Mind abilities such as perspective-taking and reasoning about others' beliefs (Cummings, 2015; Geurts, 2010; Geurts & Rubio-Fernández, 2015; Goodman & Frank, 2016). There is growing evidence that intention-reading abilities do emerge early in development (Mascaro & Sperber, 2015; Tomasello, 2003), and that these are used in early comprehension and production. In their review, Graham, Juan and Khu conclude that 'these findings suggest that as soon as infants begin to reason about the visual perspectives of others, they begin to also use this information to inform their interpretation and production of both nonverbal and verbal communicative behaviours' (Graham, Juan, & Khu, 2016: 4). As an aside, one might also expect that taking the Epistemic Step would be more, not less, complex than not taking it. On this hypothesis, as soon as children develop the ability to make complex inferences such as implicature derivations, they are also able to take into account the speaker's perspective in this inferencing.

5.4.2 Two-step hypothesis

Alternatively, the implicature derivation (assuming full common ground) and reasoning about others' epistemic states first develop separately, and then are later integrated, in two steps. The prediction is that children first perform at adult-like rates for quantity implicatures where common ground can be assumed – where there is no reason to drop the Competence Assumption. Then, later, they do so when the interlocutor's perspective must be taken into account, when the speaker's perspective differs from the hearer's.

This notion assumes that the pragmatic inferencing process itself does not have to include reasoning about others' beliefs; either this element is 'replaced' by own beliefs, or the inference can be separated from the pragmatic strategy adopted in any given situation. For example, one proposal along these lines is that an interpretation norm can be distinguished from pragmatic inference processes (Kissine, 2016). On the one hand, there is an interpretation norm, which could in one context be 'complete recovery of the speaker's communicative intentions', but in another, 'the most relevant given the speaker's perspective', and in yet another, 'simple egocentric relevance'. On the other hand, there is the pragmatic processing, which is 'driven by the selected interpretative norm' (Kissine, 2016). In other words, some pragmatic reasoning is possible without full theory of mind engagement, at least in development and potentially in adults as well (Andrés-Roqueta & Katsos, 2017; Breheny, 2006; Katsos & Andrés-Roqueta, 2015). If children find integration of pragmatic inferences and perspective-taking in processing challenging, as they appear to do with other aspects of the context such as the QUD (Skordos & Papafragou, 2016), then a two-step developmental trajectory is likely.

5.5 Experiment 3: 'Nosy Neighbours' – ad hocs, scalars and speaker epistemic state

In this experiment, I investigated the effect of the speaker's epistemic state (knowledgeable or ignorant) on the rate of scalar and ad hoc quantity implicatures in adults and children. In addition, I looked at the relationship between these rates and social aptitude (representing Theory of Mind abilities) in adults.

5.5.1 Method

5.5.1.1 Design

The design was based on the covered-box paradigm, which has been used with both children (Bill et al., 2014), and adults (Schwarz, Romoli, & Bill, 2015). It is similar to the ‘magic box’ design which has also been used with children (Y. T. Huang, Spelke, & Snedeker, 2013; Li, Barner, & Huang, 2008; Pearson, Khan, & Snedeker, 2011). Schwarz, Romoli and Bill (2015) employ this task to test scalar implicatures by using a detective scenario where it is uncertain whether the speaker was fully informed – this is done to raise the number of no-inference responses, rather than to examine the effect of speaker’s epistemic state per se. However, this makes it a promising paradigm for the present study, as one that is both child-friendly and allows for uncertainty (partial-ignorance) on the part of the speaker.

The covered-box task was similar to a picture-matching task, with a critical difference: participants were presented with a context picture and utterance, followed by the critical utterance and a choice between a visible picture and a covered picture. The utterance and the visible picture conditions were fully crossed (strong utterance/‘strong’ picture, strong/weak, weak/strong, weak/weak – Table 5.1 and Table 5.2). In the critical case, participants heard an utterance with the weaker term (‘some Xs have Ys’, or ‘there is an X’) and saw a picture depicting the stronger term (*all Xs have Ys* or *there is an X and a Y*) alongside a covered picture; if they rejected the visible picture and chose the covered picture as matching the scenario, it could be inferred that they had derived a quantity implicature.

The study had a $2 \times 4 \times 2 \times 2$ design: epistemic state \times condition \times scale \times age. In the analysis, I primarily concentrated on epistemic state, the two comparable conditions (strong/weak and weak/strong), and age: $2 \times 2 \times 2$. Epistemic state was a between-subject condition, and condition and scale within-subject. The speaker was either knowledgeable or only partially knowledgeable (having a good or poor view of the event described); utterance (strong/weak) and pictures (strong/weak) were fully crossed; and critical items triggered either a scalar or an ad hoc quantity implicature. The position of the visible and covered pictures (left/right) were counter-balanced across trials. Each participant saw each item in only one condition (4 lists), and three

pseudo-random orders were produced using a python script, such that no more than two of any one condition occurred together (Python Software Foundation, 2016), so there were 12 lists, each with 16 items (see Appendix 8.7).

Condition	Utterance	Visible picture	Pragmatic choice
Critical weak/strong	Some of the guests drank lemonade	<i>All of the guests drank lemonade</i>	Covered picture
Control strong/weak	All of the guests drank lemonade	<i>Some of the guests drank lemonade</i>	Covered picture
Control strong/strong	All of the guests drank lemonade	<i>All of the guests drank lemonade</i>	Visible picture
Control weak/weak	Some of the guests drank lemonade	<i>Some of the guests drank lemonade</i>	Visible picture

Table 5.1 Nosy Neighbour study conditions – examples with scalar items (Context question: ‘What did the friends drink?’)

Condition	Utterance	Visible picture	Pragmatic choice
Critical weak/strong	There were bananas	<i>There were bananas and grapes</i>	Covered picture
Control strong/weak	There were bananas and grapes	<i>There were bananas</i>	Covered picture
Control strong/strong	There were bananas and grapes	<i>There were bananas and grapes</i>	Visible picture
Control weak/weak	There were bananas	<i>There were bananas</i>	Visible picture

Table 5.2 Nosy Neighbour study conditions – examples with ad hoc items (Context question: ‘What was on the plate?’)

Adults were also given the Autism Spectrum Quotient (ASQ, Baron-Cohen, Wheelwright, Skinner, Martin, & Clubley, 2001), a 50-question survey that measures social aptitude in the general population by asking participants to say to what extent they agree with statements about themselves (e.g. ‘I find it easy to work out what someone is thinking or feeling just by looking at their face’, ‘I enjoy meeting new people’). Some other pragmatic tasks with adults also employ this test: Grodner, Dalini, Pearlstein-Levy and Ward (2012) find that greater social aptitude (indicated by a lower ASQ score) is associated with better perspective-taking in a reference

resolution task, but in the second half of the trials only; Husband (2014) finds that lower ASQ scores show less distinction between implicature cancellations and entailment cancellations in an acceptability rating task – seemingly an opposite effect, although perhaps explainable in the context of the task. Looking at interpretation of ironic utterances, Spotorno & Noveck (2014) find a positive correlation between reading time and ASQ scores, again in the second half of the task, such that those with less social aptitude maintained a difference in reading time between literal and ironic utterances. Goldshtein, Brown-Schmidt and Terkourafi (2017) find no correlation, though, between rejection of under-informative scalar utterances and ASQ scores.

5.5.1.2 *Participants*

Adults (N = 144) were recruited on Prolific Academic, an online recruitment platform for research (Prolific Academic Ltd, 2016). They declared English as their first language, and had been born in and currently lived in UK. 4 participants completed the task twice, and so only their first response was included in the analysis. One additional participant was excluded for failing to complete the task.

Children (N = 30) were recruited from two local schools and local families (age range: 5;2–6;10; N = 14 girls), where parents gave consent for them to participate, following approval from the Psychology Research Ethics Committee (PRE.2014.98). 3 additional children were recruited but excluded due to technical issues (N = 1) or for failing to engage in or understand the task (N = 2). Children were all fluent in English. 18 were monolinguals. 6 were known to be bilingual, but had been exposed to both English and the other language from the first year of life and were exposed to English as much if not more than the other language (i.e. English-dominant). For 6 children, background information including language use was missing (as parents provided consent forms but not background questionnaires). All children's data was included in the present analysis.

5.5.1.3 *Materials*

8 ad hoc and 8 scalar items took the form of a question-answer pair, between two characters called Mr and Mrs Watson. Adults read these mini-dialogues, as in Table 5.1 and Table 5.2, while children heard recordings of a male and female voice saying the utterance only (not the name of the speaker).

The question was always a simple ‘what’ question (with one exception: ‘how did the friends come?’), and the scalar and ad hoc utterances had the same structure: ‘All/some of the friends V-ed X’ or ‘there was an X (and a Y)’.

The text or recordings were accompanied by two pictures – one of which was always covered by a black box in experimental trials (Table 5.3). These were simple colourful cartoon drawings, sourced largely from Pixabay, an online database of photographs and illustrations released free of copyright under Creative Commons CC0 (Braxmeier & Steinberger, 2017), or from an internet search with the ‘labelled for noncommercial reuse’ filter checked. They were manipulated using GIMP – GNU Image Manipulation Program (Kimball, Mattis, & The GIMP Development Team, 2016).

Mrs Watson’s question	Mr Watson’s reply (strong)	Mr Watson’s reply (weak)	‘Strong’ picture	‘Weak’ picture
What was on the table?	There was a basket of apples and a book.	There was a basket of apples.		
What did the friends get?	All of the friends got a balloon.	Some of the friends got a balloon.		

Table 5.3 Examples of ad hoc and scalar stimuli for Experiment 3

Utterances were devised using simple language that would be understandable to children, using Kuperman, Stadthagen-Gonzalez and Brysbaert’s (2012) Age of Acquisition ratings as a guide. In addition, the utterances and pictures went through several rounds of piloting with adults, in order to reduce responses that were based on irrelevant reasons, e.g. irrelevant picture details. As part of this process, other scales, besides <all, some>, were tested: <finish, start>, and part-whole pairs (e.g. <whole, slice>). However, in both cases, the challenge of depicting the intended meaning using

still images was prohibitive – not even the straightforward control items were treated as expected by participants. Thus, these two were abandoned and the items restricted to <all, some> scalar and ad hoc items, which renders the design more comparable to the majority of other implicature studies in the developmental literature.

5.5.1.4 Procedure

Adults completed the task via Qualtrics, an online survey platform, and were paid £1 for their participation (Qualtrics, 2016); children did so on a laptop computer, with the stimuli presented via PowerPoint, and received a sticker and certificate. The children's responses were recorded by the experimenter using pen and paper.

The participants were introduced to the task by reading (or, for children, having read to them), a 'Nosy Neighbours' scenario:

Mr & Mrs Watson were very nosy neighbours. They always liked to know what their neighbours were doing. One day while Mrs Watson was out, their neighbours had a party. When she returned she asked Mr Watson about what had happened that day next door. He told her that some friends had come to a party there.

[However, there was a large fence between their house and the neighbours', so Mr Watson had not had a good view of what was going on; he couldn't see the party very well.]

OR

[Since there was no fence between their house and the neighbours', Mr Watson had a good view of the party; he could see the party really well.]

Your job is to read what Mrs Watson asks about the neighbours' party, and Mr Watson's responses. Below their conversation you will see two pictures, only one of which shows what actually happened at the party. However, one of the pictures is always covered. Click on the picture that shows what actually happened. If it's not the one that you can see, then it will be the covered one.

These instructions were accompanied by pictures of Mr and Mrs Watson, their view of the neighbours' house, and the group of friends who came to the party.

There were then two warm-up trials in which the visible picture was unambiguously correct or incorrect. One of these was designed so that the visible picture was selected, the other the covered picture. In either case, the covered picture was revealed, and feedback was given. This was done to help the participants understand the activity, but also to demonstrate that either the visible or covered picture could be selected and that there was a picture 'under' the black box. At the end of the warm-up, the participant was reminded that that Mr Watson had a good or poor view of the party.

In the experimental trials, the covered picture was never revealed; this was explained to the participants before they started the main task. Between each trial, Mr Watson's view of the neighbours' house (with or without a fence) was displayed for 1 second, before the experiment automatically progressed to the next trial. This was meant to act as a visual reminder of Mr Watson's epistemic state.

Children were tested in a room at their school or home that was as quiet as possible, as part of two test sessions in which they also completed the BPVS, mini-TROG, and a picture-matching task for manner implicatures. The Nosy Neighbour task was always the first presented in the first session. Adults completed the task on their computers (or tablets; 'no mobiles' was specified) wherever they wished – but most likely at home. At the end of the Nosy Neighbour activity, they also completed the Autism Spectrum Quotient questionnaire (Baron-Cohen et al., 2001), and were told: 'In the second part of this task, you will complete a personality questionnaire. Please answer as quickly and honestly as you can.' This was followed by the instructions from the questionnaire itself:

Below is a list of statements. Please read each statement very carefully and rate how strongly you agree or disagree with it by clicking on your answer.

The Nosy Neighbour task took on average 5 minutes to complete, and the ASQ an additional 3 or 4 minutes.

5.5.1.5 Predictions

In line with adults' abilities and other findings on speaker epistemic state and implicatures, the predictions for adult hearers are that:

- The rate of implicature derivation (in the critical weak utterance/strong picture condition) will be higher when the speaker is knowledgeable than when the speaker is ignorant.
- If there is also an effect of speaker knowledge in the control conditions, then: difference in the rate of implicature derivation between knowledgeable speaker and ignorant speaker will be greater for the critical condition than the control conditions.

(While on a straightforward Gricean model, no difference would be expected in the control conditions between knowledgeable and ignorant speaker groups, one could imagine that some participants might adopt some sort of strategy where they do adjust their responses in the control conditions as well – if they suspect that quality as well as quantity is affected by the speaker's partial knowledge. However, the crucial prediction is that the effect is larger for the critical pragmatic condition than any control conditions.)

For children, the predictions are that:

- On the one-step hypothesis, their responses will pattern like those of adults.
- On the two-step hypothesis, there may be a different pattern of responses from adults', such that there is no difference in performance between the knowledgeable speaker and ignorant speaker groups (depending on the age at which the two skills are integrated).

In addition, given the other studies that find some relationship between ASQ scores and pragmatic abilities (e.g. Grodner et al., 2012), it is predicted that those with lower ASQ scores (higher social aptitude) will be better at perspective taking, and therefore will make more adjustment in the ignorant speaker condition, so that their rate of implicature derivation is lower in the critical condition than for those with higher ASQ scores (i.e., a positive correlation between implicature rate and ASQ score). In the knowledgeable speaker condition, the opposite pattern is expected –

those with lower ASQ scores – more social – will derive more implicatures (i.e., a negative correlation).

5.5.2 Results and analysis

Responses were coded as correct if the semantically correct or pragmatically felicitous picture was chosen. In the critical condition, and the strong/weak control condition, this was the covered picture. In the strong/strong and weak/weak control conditions, this was the visible picture. Overall, there were more correct choices in the control conditions than in the critical implicature condition, and there seems to be an effect of speaker knowledge that is different between adults and children.

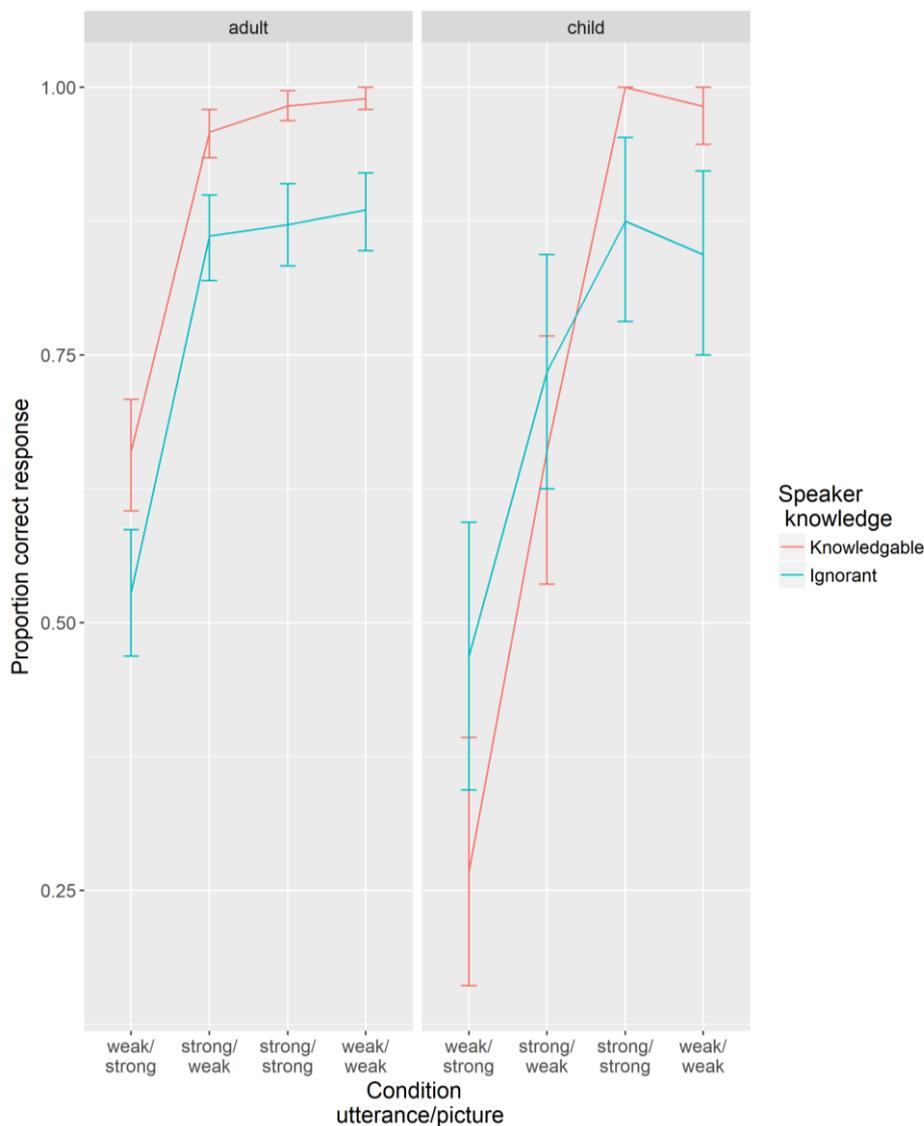


Figure 5.2 Correct responses by condition, speaker knowledge, and age group
Error bars show bootstrapped 95% confidence intervals for between-subject comparison

Speaker:	adult		child	
	knowledgeable	ignorant	knowledgeable	ignorant
Weak/strong	.66	.53	.27	.47
Strong/weak	.96	.86	.66	.73
Strong/strong	.98	.87	1.00	.88
Weak/weak	.99	.89	.98	.84

Table 5.4 Proportion correct responses by condition, speaker knowledge and age group

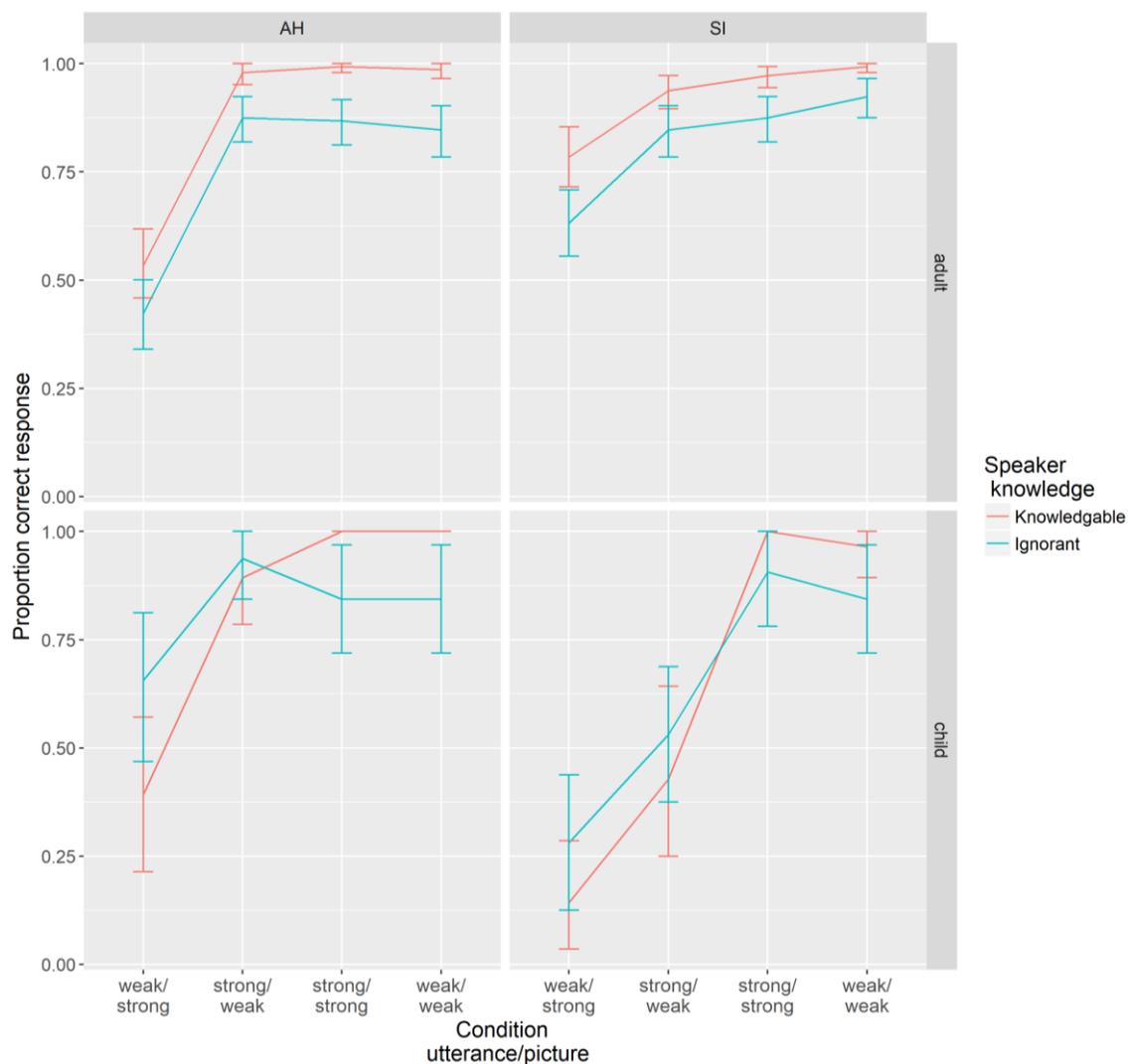


Figure 5.3 Correct response by condition, speaker knowledge, scale and age group
 Error bars show bootstrapped 95% confidence intervals for between-subject comparison

		adult		child	
		knowledgeable	ignorant	knowledgeable	ignorant
Ad hoc	Weak/strong	.53	.42	.39	.66
	Strong/weak	.98	.88	.89	.94
	Strong/strong	.99	.87	1.00	.84
	Weak/weak	.99	.85	1.00	.84
Scalar	Weak/strong	.78	.63	.14	.28
	Strong/weak	.94	.85	.43	.53
	Strong/strong	.97	.88	1.00	.91
	Weak/weak	.99	.92	.96	.84

Table 5.5 Proportion of correct responses by scale, condition, speaker knowledge and age group

5.5.2.1 Implicature rates with knowledgeable and ignorant speakers, in adults and children

To look at the effect of speaker ignorance on implicature derivation in children and adults, a mixed-effects logistic regression model was fitted, examining only the weak/strong control and strong/weak critical conditions – these two conditions are most similar in terms of correct response (rejecting the visible picture). There is little variance in the other two control conditions due to ceiling effects, and high multicollinearity in models fitted with all conditions.

A model with speaker knowledge, condition, scale and age group as fixed effects and their interaction, and subject (by condition) and item (by condition) random slopes was fitted with sum coding for fixed effects and the bobyqa optimizer in R with the *lme4* package (Bates et al., 2015; R Core Team, 2016), following suggested best practice (Barr et al., 2013). There is no main effect of speaker knowledge ($\beta = .04$, $p = .86$) – which is not unexpected, given that this is across both age groups, where the effect of ignorance appears to be opposite for these two conditions. There is a main effect of condition, such that the critical condition is significantly lower than the grand mean ($\beta = -1.86$, $p < .001$); a main effect of age group, such that adults are significantly higher than the grand mean ($\beta = .95$, $p < .001$); and a main effect of scale, such that ad hocs are significantly above the grand mean ($\beta = .84$, $p < .001$). In addition, there is a significant interaction between speaker knowledge and age group,

such that the adult knowledgeable speaker group are overall higher than the grand mean ($\beta = .66$, $p = .01$); an interaction between condition and age group, such that the critical condition for adults is significantly lower than the grand mean ($\beta = -0.62$, $p < .001$); and an interaction between age group and scale, such that adults with ad hocs are significantly below the grand mean ($\beta = -0.93$, $p < .001$). There are no three- or four-way interactions. That is, the analysis confirms what can be seen in Figure 5.3 that adults are performing poorly on ad hocs in the critical condition, compared to scalars, and children poorly overall (for weak/strong and strong/weak conditions) for scalars.

	Estimate	SE	z	p
Intercept	1.71	.38	4.52	< .001
Knowledgeable speaker	.04	.24	.18	.86
Critical condition	-1.86	.30	-6.13	.00
Adult age group	.95	.25	3.78	< .001
Ad hoc scale	.84	.24	3.51	< .001
Knowledgeable speaker: critical	-0.16	.18	-0.92	.36
Knowledgeable speaker: adults	.66	.24	2.70	.01
Critical: adults	-0.22	.19	-1.18	.24
Knowledgeable speaker: ad hoc scale	.02	.14	.15	.88
Critical: ad hoc scale	-0.62	.18	-3.54	< .001
Adult age group: ad hoc scale	-0.93	.15	-6.13	< .001
Knowledgeable speaker: Critical: adults	-0.11	.18	-0.59	.55
Knowledgeable speaker: Critical: ad hoc	-0.09	.14	-0.63	.53
Knowledgeable speaker: adults: ad hoc	.05	.14	.36	.72
Critical: adults: ad hoc	.01	.15	.07	.95
Knowledgeable speaker: critical: adults: ad hoc	-0.07	.14	-0.51	.61

Table 5.6 Response ~ Ignorant * Condition * scale * age group + (1 + Condition | ID) + (1 + Condition | Item)

data: child and adult, weak/strong critical and strong/weak control conditions

Glmer, family = binomial, optimizer = bobyqa, sum coding

When the same model is fitted with treatment coding, with knowledgeable speaker, critical condition, ad hoc inference, and adults age group as the baseline,

there is a significant interaction between speaker knowledge and age group, such that, for the critical condition with ad hocs, the difference of ignorant speaker minus knowledgeable speaker between children and adults is greater for children than adults – as can be seen from Figure 5.3, the pattern is in fact opposite ($\beta = 2.13$, $p = .03$). There is also an interaction of condition and inference, such that for adults and knowledgeable speaker group, the difference of control minus critical condition is smaller for scalars than for ad hocs ($\beta = -3.07$, $p = .001$); and interaction between age group and inference, such that for knowledgeable speaker group in the critical condition, the difference of children minus adult group is more negative for scalars than ad hocs ($\beta = -3.77$, $p < .001$).

	Estimate	SE	z	p
Intercept (Knowledgeable speaker/critical/ad hoc/adult)	.22	.37	.59	.55
Ignorant speaker	-0.69	.38	-1.80	.07
Control condition	6.23	1.10	5.65	< .001
Child age group	-0.66	.71	-0.93	.35
Scalar inference	1.58	.48	3.32	.001
Ignorant speaker: control	-1.70	.96	-1.77	.08
Ignorant speaker: child	2.13	.98	2.17	.03
Control: children	-1.56	1.42	-1.10	.27
Ignorant speaker: scalar	-0.33	.43	-0.77	.44
Control: scalar	-3.07	.93	-3.30	.001
Child age group: scalar	-3.77	.94	-4.00	< .001
Ignorant speaker: control: children	1.41	1.94	.73	.47
Ignorant speaker: control: scalar	1.25	1.03	1.22	.22
Ignorant speaker: child: scalar	.15	1.18	.13	.90
Control: child: scalar	.48	1.74	.28	.78
Ignorant speaker: control: child: scalar	-1.11	2.20	-0.51	.61

Table 5.7 Response ~ Ignorant * condition * scale * age group + (1 + Condition | Item) + (1 + Condition | ID)

data: child and adult, weak/strong critical and strong/weak control conditions

Glmer, family = binomial, optimizer = bobyqa, dummy (treatment) coding

5.5.2.2 Autism Spectrum Quotient scores and implicature rates

I consider the responses in the critical (weak/strong) condition only. This is the condition for which a relationship is expected, as it requires a pragmatic inference,

and furthermore the rate of correct choices in the other, control conditions is mostly approaching ceiling, obscuring any relationship. On visual inspection, there appears to be a very small difference between the knowledgeable and ignorance conditions: a positive relationship in the knowledgeable condition, and no relationship or a negative one in the ignorance condition.

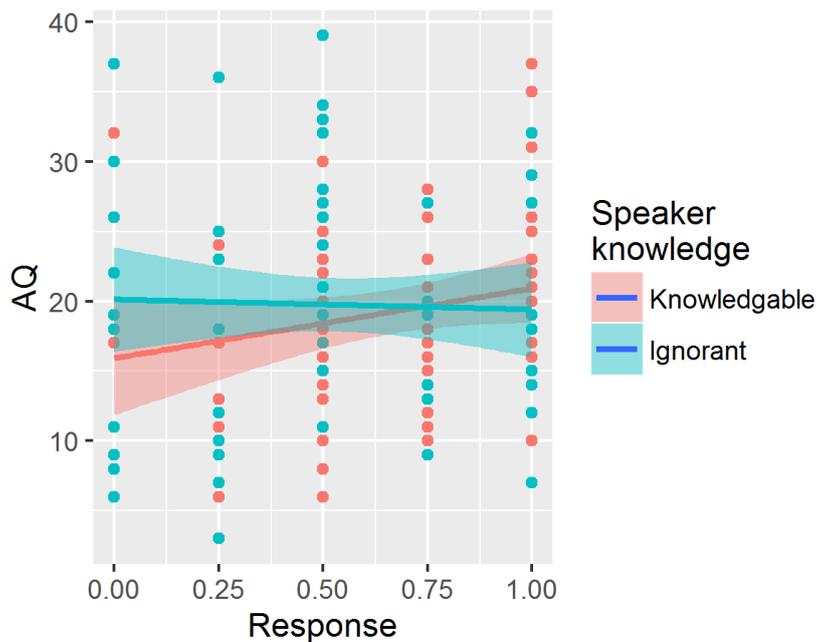


Figure 5.4 AQ score by critical (weak/strong) response and speaker knowledge

Neither of these relationships is statistically significant, however: for the ignorant group, $z = -0.55$, $p = .58$, $\tau = -0.049$ (Kendall's τ , two-tailed, using R Stats Package *cor.test*). For the knowledgeable group, $z = 1.867$, $p = .0619$, $\tau = .172$. When this correlation is bootstrapped to get the 95% confidence intervals, the intervals cross 0 (-0.003, .352, Normal), indicating that the direction of the relationship that was tending towards significance in the sample may not be the same in the population.

5.5.3 Discussion

5.5.3.1 Findings

The prediction was for an interaction between the critical and control conditions between the knowledgeable speaker and ignorant speaker group, such that the drop in implicature rates in the critical condition from knowledgeable to ignorant speaker groups would be greater than in the other conditions. Comparing the weak/strong

critical and strong/weak control conditions, I found a main effect of age group and condition: adults choose the target picture more often than children, indicating a developmental trend, and fewer implicatures are derived in the critical condition than correct semantic interpretations in the control conditions. The interaction between age group and speaker epistemic state indicates that for adults there are more correct responses when the speaker is knowledgeable across both conditions together. In the critical condition, children and adults display opposite patterns: while adults make more inferences when the speaker is knowledgeable, children make more inferences when the speaker is ignorant. For children, it looks as if the critical and control (weak/strong, strong/weak) conditions pattern together numerically, and the other two control conditions (strong/strong, weak/weak) also pattern together, with lower rates of correct choice in the knowledgeable speaker group in the first pair, and higher rates in the second pair.

5.5.3.2 Epistemic state and implicatures in adults

There was a main effect of ignorance for adults: across all conditions, there were fewer correct choices in the ignorant group. This is unexpected, given that there is nothing in the experimental context that is meant to indicate that the speaker is violating the maxim of quality, as well as being only partially knowledgeable and therefore not maximally informative.

One possible explanation might be that in this experimental context at least some speakers are applying a meta-strategy as a response to the speaker's ignorance: choosing the opposite of the semantically or pragmatically (in normal circumstances) correct option, across all conditions. This could be a result of the nature of the task encouraging participants to engage in higher level reasoning, possibly on account of its game-like setting or the fact that the participants are in effect observers or eavesdroppers rather than interlocutors. However, when we look at the subset of participants who score 100% in the strong/strong and weak/weak control conditions (N = 63 in the knowledgeable speaker group; N = 53 in the ignorant speaker group, for adults only), the same pattern is seen – a main effect of condition, but no effect of speaker epistemic state. There is no difference in the drop from the strong/weak control condition to the weak/strong critical between the knowledgeable and ignorant

speaker groups (98% to 66% and 87% to 53% respectively – Figure 5.5). That is, even for those who are correctly accepting the visible picture in the two control conditions where the picture is a semantic match of the utterance (and therefore who are assuming that what the speaker says must be true), there is still an overall effect of the speaker’s epistemic state on utterance interpretation in the other two conditions, rather than just an effect on implicature derivation. This suggests that this is not the only explanation for the findings.

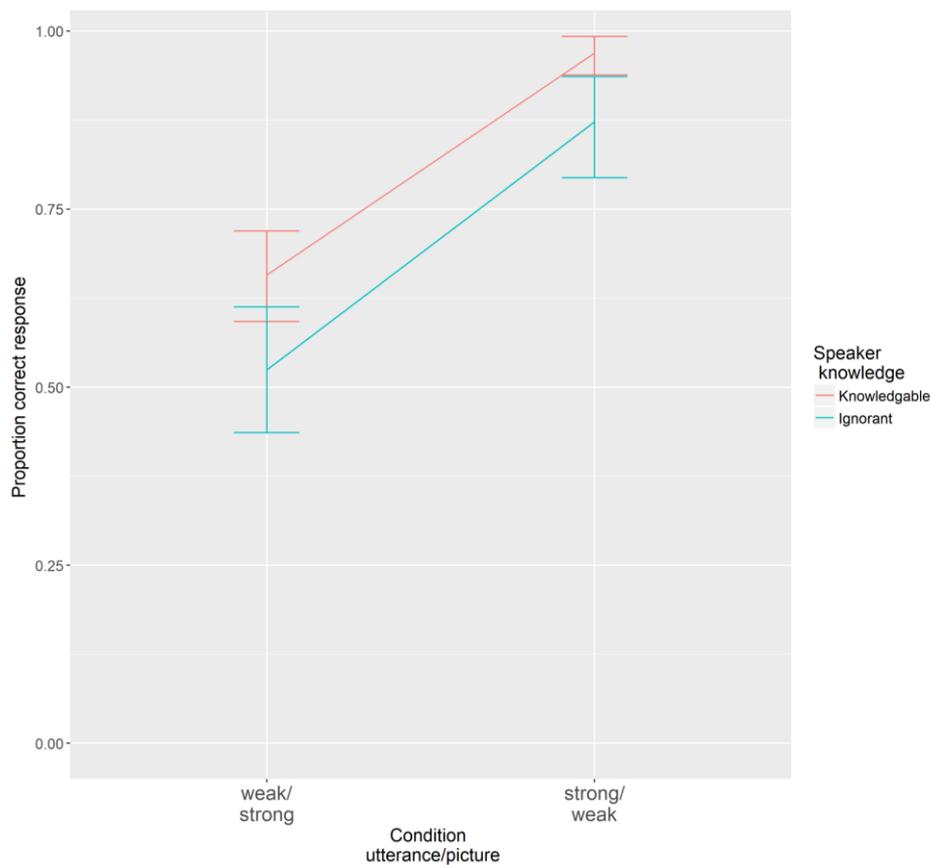


Figure 5.5 Mean correct response for those at ceiling in control conditions (weak/weak, strong/strong)

Another possibility might be that the nature of the task means that hearers have diverging understandings of the goal. Remember that they were asked to pick the picture that showed what actually happened at the party. It is possible that some participants are trying to choose the picture that reflects the real world, based on the speaker’s utterance, and others are trying to choose the picture that reflects the speaker’s world. This could lead to different behaviour in some conditions. For example, in the critical condition, participants might reject the visible picture (showing *all* of the friends with balloons) on an enriched interpretation for either the

actual world or for the speaker's world; alternatively, they may accept it for the actual world on the reasoning that for all the friends the speaker saw, each had the characteristic in the utterance, but in the actual world there are more friends who did not. This difference in domain restriction applies too for the strong/strong control condition.

The larger issue here is that it is uncertain what Mr Watson, the speaker, has seen. In this respect, this study is like Bergen and Grodner's (2012) study, where the extent of the speaker's knowledge or ignorance is left vague (e.g. 'she skimmed the report'). However, it is unlike Goodman and Stuhlmüller's (2013) betting design, or Breheny, Ferguson and Katsos' (2013) eye-tracking study, where it is clear to the hearer what the speaker does and does not know. It could be that while imprecise information about the speaker's epistemic state is sufficient to suspend implicatures in a self-paced reading task, it is not in a picture-selection task. In the latter, by its very nature, a precise interpretation is required. This could lead participants to switch to some higher-level strategy that affects both the critical and control conditions, or to be more uncertain about their choice. Furthermore, what is required here is more than level 1 perspective-taking: it is not the case that the participants have to understand that the speaker cannot see and know about something that they can. Instead, the participant has to imagine that the participant might have seen something different from what is presented in the visible picture. This might be an additional challenge for adults, and certainly for children.

5.5.3.3 Epistemic state and implicatures in children

Children do not show such a clear pattern as adults, but if anything, they tend towards choosing the correct picture more in the weak/strong critical and strong/weak control conditions when the speaker is ignorant, and in the strong/strong and weak/weak conditions when the speaker is knowledgeable. In fact, this means that when the speaker is ignorant, they have a preference for choosing the covered picture. This could be the result of a heuristic such as: if the speaker's knowledge is not certain, choose the uncertain picture. That is, like the adults, they may be sensitive to the speaker's knowledge state, but applying an interpretation strategy in all conditions, whether they could trigger a quantity implicature or not. Task factors as

discussed for adults above could be contributing to this response. In addition, as already mentioned, in this study, the role of the participant is a third-party observer, rather than active interlocutor. While the ability to learn through observing or overhearing has already developed by this age, in production studies, children's performance becomes much more adult-like if the task is interactive and with a clear communicative goal (Grigoroglou & Papafragou, 2016). This means that the Nosy Neighbour kind of task could have been particularly challenging for them.

5.5.3.4 *Rates of ad hoc and scalar implicatures*

For both adults and children, the rate of ad hoc implicature is surprisingly low, across both ignorant and knowledgeable speaker conditions, given that there is good evidence that children are able to make ad hoc inferences from 3 years (as seen in Chapter 2), and adults are also expected to be at rates approaching ceiling. Note, though, that Bott and Chemla (2016) also find low rates (below 50%) in adults with the same experimental paradigm.

One explanation might have to do with the perceived QUD and domain restriction. It may be that the discourse context did not bias the participant to expect an exhaustive answer – in other words, the implied QUD was not, for example, 'what were all the things on the rug?', but instead, 'what were some things on the rug?'. This could particularly be the case for items where the 'container' (e.g. 'on the table', 'on the rug', 'by the door') is not very distinct or salient. Again, note that here the design is different from that of Breheny, Ferguson and Katsos (2013), where there is a clear container that is an appropriate size for the objects placed in it, creating a clear domain for quantification.

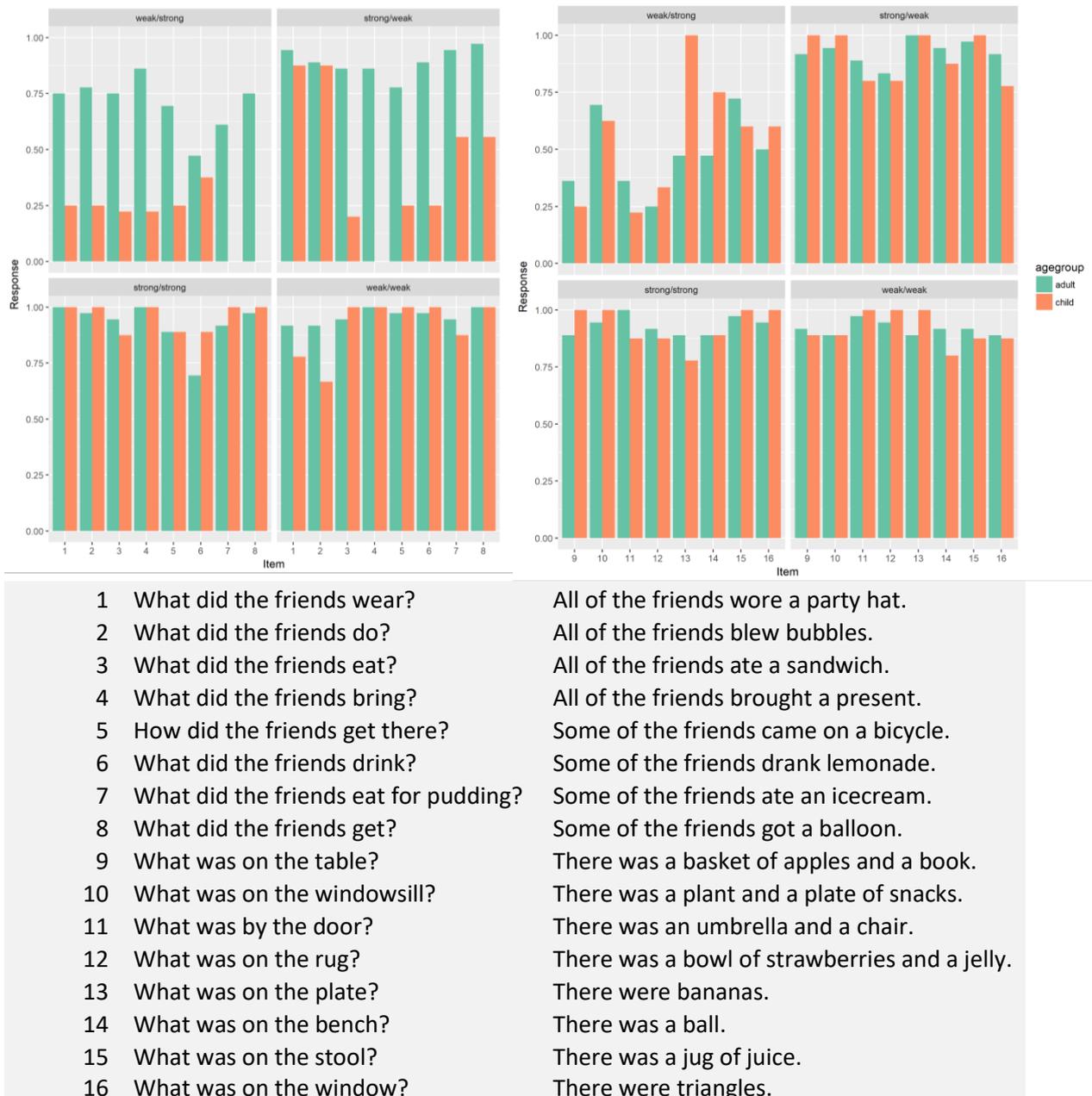


Figure 5.6 Proportion correct responses by item, condition, and age group for scalars (left) and ad hocs (right)

A second possible explanation has to do with task factors. The covered-box task was chosen as one that had been used successfully with children, and in designs where some uncertainty about the speaker's epistemic state was required (Bill et al., 2014; Schwarz et al., 2015). However, it has more potential layers of complexity than a typical binary-choice picture-selection task. As one picture is always covered, there is

an added element of risk for participants: they need to cross a threshold of confidence that the visible picture is not the intended one in order to choose the covered one. Potentially for adults, the more conventionalised nature of scalar implicatures means that they can draw on their linguistic experience of the intended use of ‘some’ in these kinds of contexts, and this might lead them to be more confident about the scalar than ad hoc implicatures. Children perform poorly for scalar implicatures as well, potentially because of their lack of linguistic experience giving them such a priori expectations, or because the explicit question did not highlight quantity. Even in the control condition, there is a bimodal distribution of children’s responses for scalars (Figure 5.7). In Bill, Romoli, Schwarz and Crain’s study (Bill et al., 2014), 4–5-year-olds also rejected the visible picture only around 20% of the time in the critical scalar condition, although at a slightly younger age.

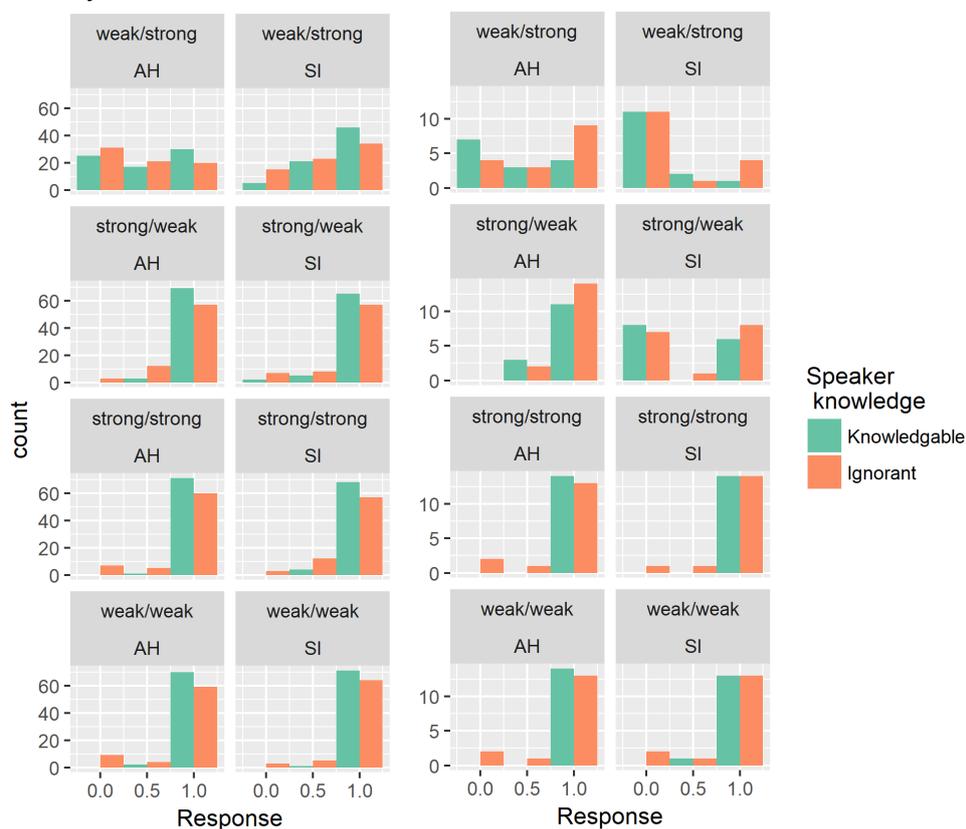


Figure 5.7 Histograms of correct picture selection for adults (left) and children (right)

5.5.3.5 Implicature rates and social skills

There was no evidence for a relationship between implicature rates and social aptitude as measured by the Autism Spectrum Quotient. This could be simply due to

the issues with the task that I discussed above, or the task design that means that each participant sees only 4 critical trials, which is not very suitable for looking at these sorts of correlations. It does, however, accord with Antoniou, Cummins and Katsos' (2016) study in which no relationship was found between the ASQ and rates of rejection of under-informative utterances, once age, gender, IQ and verbal abilities had been controlled for.

5.5.4 Summary

The positive contribution of this study is its indication that children as well as adults are sensitive to a speaker's epistemic state in communication, as measured by an offline picture-selection (covered-box) task. However, the results suggested that this task was measuring responses to different implied QUDs (in the case of adults), and that uncertain speaker epistemic state was associated with the 'uncertain' choice of the covered picture (in the case of children). The task could therefore be improved upon by: reducing uncertainty about the degree of the speaker's ignorance; removing ambiguity in the interpretation of the task and its instructions; and increasing interactivity and making the participant part of the dialogue with a clear communicative goal. These are addressed in the next study.

5.6 Experiment 4: director task – ad hocs and speaker epistemic state

5.6.1 Method

5.6.1.1 Design

The task combined the director-task paradigm (described above) and a picture-matching task testing ad hoc implicatures (e.g. Horowitz & Frank, 2015). Horowitz and Frank's task is not unlike that used in Experiment 1A here, except that the pictures contain more than one of each object: in their task, children have to pick the book cover that the speaker is referring to, and a book cover might display, for example, 4 cats, or 2 cats and 2 birds. The utterance is of the form, 'there are cats' – so there is the possibility of an ad hoc inference, for example *there are only cats*. Similarly, in this task, children played a game with a puppet that involved following his instructions to collect picture cards from an array of four picture cards.

Both experimental paradigms have been used successfully with 5–6-year-old children. With the director task, Nadig and Sedivy (2002) and Nilsen and Graham (2009) found evidence that children take into account the speaker's perspective in reference resolution (also San Juan, Morra, Gibbard, Khu, & Graham, 2017); for the implicature picture-matching task, Horowitz and Frank (2015) found high levels of ad hoc implicature understanding in 4- to 5-year-olds. Together with the evidence for the early development of level 1 perspective-taking reviewed above, this meant that children in the present study were expected to succeed in a condition that tests straightforward ad hoc inferences, and in a condition that tests perspective-taking; what the task had the potential to isolate was the ability to combine the two in a critical condition with implicatures plus perspective-taking.

In addition, the director task provides multiple cues to the speaker's perspective: what is in common ground and privileged ground is visually available to the participant; the difference in perspective between the speaker and hearer is physically manifest and salient in the experimental set-up (e.g. Nadig & Sedivy, 2002); and the speaker gives verbal cues to their ignorance (e.g. Matthews et al., 2006). The design does involve attributing an epistemic state to a puppet, but other studies using puppets suggest that there is no reason to think that this renders the task a less valid measure than if the speaker were a real human interlocutor (e.g. Diesendruck & Markson, 2001; Hochstein et al., 2016; Siegal et al., 2010).

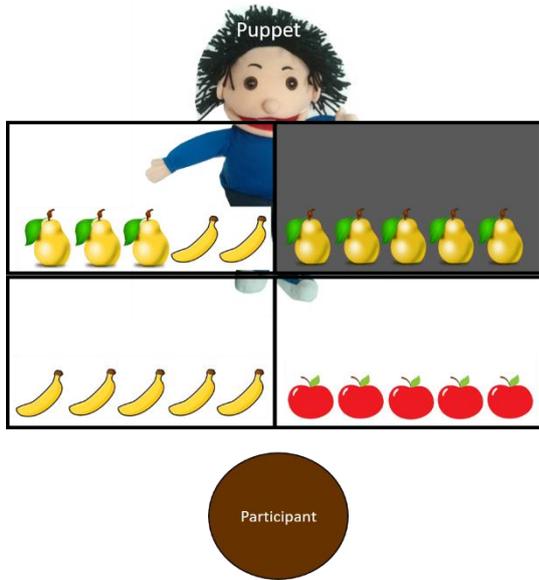


Figure 5.9 Example display: common ground unambiguous, common ground ad hoc implicature and privileged ground ad hoc implicature conditions

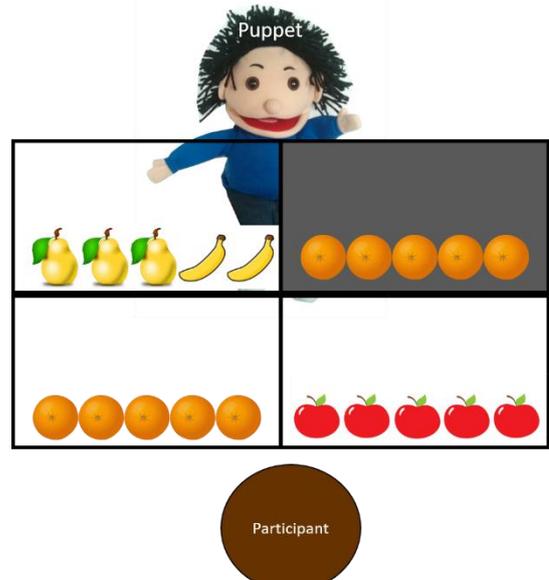


Figure 5.8 Example display: privileged ground ambiguous condition

Conditions	Utterance	Target
Common ground unambiguous	Pick the card with apples	Card with apples
Common ground ad hoc implicature	Pick the card with bananas	Card with only bananas
Privileged ground ambiguous	Pick the card with oranges	Card with oranges in common ground
Privileged ground ad hoc implicature	Pick the card with pears	Card with bananas and pears

Table 5.8 Utterance and target card for each condition

The experiment had a 4×2 design: the same task with four within-subject conditions was given to two age groups, children and adults. The four conditions were common ground unambiguous, common ground ad hoc, privileged ground ad hoc, and privileged ground ambiguous. In the common ground unambiguous condition, only one card, visible to both the puppet and participant, matched the description.

This condition acted as a check of engagement in the game, and as a filler. In the common ground ad hoc implicature condition, two cards, both visible to both the puppet and participant, were semantic matches for the utterance (the card with only Xs, and the cards with Xs and Ys), but only one matched an exhaustive interpretation (the card with only Xs); this condition checked children's ability to make ad hoc inferences when the speaker's epistemic state is not at stake. In the privileged ground semantic condition, two cards were matches for the utterance (both cards with Xs), but one was in common ground and the other in privileged ground; this condition was designed to check children's perspective-taking in this paradigm. Finally, in the privileged ground ad hoc implicature condition, one matching card (the card with Xs and Ys) was in common ground while another (the card with only Xs) was in privileged ground; this tested children's ability to suspend the quantity implicature and pick the card in common ground. Importantly, from the puppet's point of view 'Pick the card with Xs' was the most informative way of describing the card with Xs and Ys given what he could see – it uniquely identified the target card. In each condition, the remaining two cards in the display were distractors (or remaining three, in the case of common ground unambiguous).

Children were also given the Sally-Anne Change of Location task to measure Theory of Mind reasoning about false beliefs (Baron-Cohen et al., 1985).

5.6.1.2 Participants

33 children were recruited from two local primary schools in Cambridge, UK, aged 5;3–6;4 (N = 16 girls) where parents gave consent for them to participate (through an opt-out or opt-in approach, as dictated by the school, following approval from the Psychology Research Ethics Committee). A further 5 children were excluded due to experimenter error (N = 2), little knowledge of English (N = 1) or not cooperating (N = 2). Children were all fluent in English. 16 of those recruited were known to be monolinguals; 3 were known to be bilingual. For the remaining 14 children, background information including language use was missing (as parents provided consent, via not opting out, but not background questionnaires) – but they are still included in the analysis.

Adults (N = 36) were recruited via Prolific Academic, an online recruitment

platform for research, and were paid £0.60 for completing the task (£6/hour). They declared English as their first language, and had been born in and currently lived in UK.

5.6.1.3 Materials

The array of picture cards was displayed in a wooden frame, divided into four boxes (two by two), that was placed on a table at eye-height for the participants. One of the boxes was obscured with a piece of black foam, stuck to the frame on the side that the puppet sat. Each box had a clear Perspex card-holder, which picture cards could easily be placed into or removed from. To one side of the array, there was a cardboard box for placing the collected cards in (Figure 5.10).

The picture cards themselves were double-sided, with the top half of the card cut out in silhouette, as a cue that both the puppet and the participant could see the card. The pictures were simple and colourful cartoon illustrations of objects known to children, sourced largely from Pixabay, an online database of photographs and illustrations released free of copyright under Creative Commons CC0 (Braxmeier & Steinberger, 2017), or from an internet search with the 'labelled for noncommercial reuse' filter checked. They were manipulated using GIMP (GNU Image Manipulation Program, Kimball et al., 2016). Each picture card showed 5 items, either 5 of the same item (e.g. 5 bananas) or 2 of one item and 3 of another (e.g. 2 bananas and 3 pears). In each display, 3 of the cards showed 5 of the same item, and 1 showed two types of item. There were 6 sets of 5 picture cards, each with a theme (e.g. fruit, vegetables, insects, see Appendix 8.8).

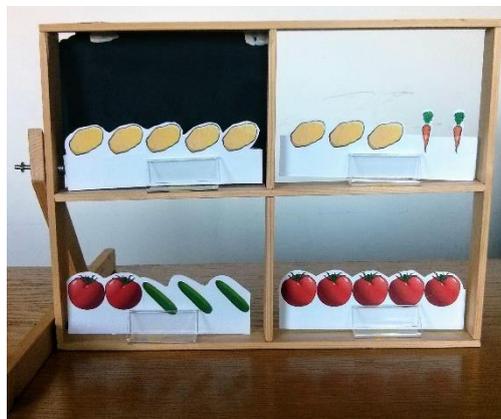


Figure 5.10 Apparatus for Experiment 4

The puppet's voice was pre-recorded by an adult male with a standard southern British English accent, using Audacity (Audacity Team, 1999) and played from a computer, using VLC (VideoLAN, 2017).

In addition, the Sally-Anne task was acted out using puppets, a bucket and a box, as in Experiment 1.

The adult version of the task was conducted online, via Qualtrics (Qualtrics, 2016). Instead of a puppet, participants saw a cartoon avatar. The cards appeared in the same 2×2 array as for children, with a white background if they were in common ground, and a grey background for privileged ground (indicating that the speaker could not see them).

5.6.1.4 Procedure

The procedure was based on previous studies that employ the director task with children (Nadig & Sedivy, 2002; Nilsen & Graham, 2009).

Participants were told that they were going to play a game with a puppet called Bob. Bob sat on the other side of the display from the child, and gave them instructions. The experimenter operated the puppet and the computer, which played the pre-recorded voice of the puppet.

For the warm-up phase, based on Nilsen and Graham (2009), the puppet explained that he wanted to play a guessing game with the child: he could see three of the items, but not the fourth. He asked the child to describe it, so that he could guess what it was. In the warm-up phase, each card had only one item on, and the warm-up items were all different from those used in the test phase. After each trial, the puppet guessed (correctly) what the item was and thanked the child. There were three warm-up trials. The aim of the warm-up was to highlight the difference in perspective between the puppet-speaker and participant-hearer. The puppet also explained after the first warm-up trial that he would turn around so that he could not see as the experimenter put out the new cards each time. Between each trial thereafter, he thanked the child and said, 'now I'm turning around'. This was to ensure that it was clear that he could not have seen what was on the privileged card.

The puppet then explained that they were going to play a different game: in this

game, the child had to collect cards and put them in a 'card box'. He would tell them which card to pick. He also explained that each time the child collected four cards, he or she would receive a sticker for their sticker chart. This was to motivate the child to keep playing the game and to make it a more enjoyable experience. At this point the experimenter showed the sticker chart and stickers to the child, but kept the stickers out of the way during the trials to avoid distraction.

The game proceeded one set of cards at a time: the order of the sets stayed the same across participants (fruit – farm animals – clothes – vegetables – animals – insects). For each set of cards, children saw all four conditions. The order of presentation of conditions within each set was counterbalanced across 6 lists (that minimised the number of cards the experimenter had to replace), and the position of the privileged ground card was also rotated around the display across the sets. Each time the child chose a card and put it in the card box, the experimenter would replace that card, and change any other cards in the display as the next trial required. There were no more than two card changes between each trial, to minimise the gap between each trial.

Before each set, the experimenter asked the child: 'Which cards can Bob see? And which can he not see? Does he know what's on that card?'. The experimenter emphasised the child's invariably correct response that Bob could not see the privileged ground card.

If the child was uncertain during the game and looked for a clue or reassurance from the experimenter about their choice of card, the experimenter remained neutral, looking straight at the child rather than the display, and said something like, 'Pick the card Bob wants for the card box', or 'Do you want to put it in the card box?'.

Adults completed an online version of the task. Instead of the warm-up production task, they answered questions to check that they had understood the set-up of the task correctly, particularly which cards they and the interlocutor represented by the avatar could and could not see and know about. In addition, they were asked only twice which cards the speaker could see, at the beginning and half way through the trials.

5.6.2 Predictions

Both adults and children are straightforwardly predicted to:

- Select the correct picture card in the unambiguous condition – indicating successful comprehension of the task
- Select the pragmatically felicitous card indicating an ad hoc implicature in the common ground ad hoc condition – indicating competence with ad hoc implicatures where perspective-taking is not at issue
- Select the common ground card in the privileged ground ambiguous condition – indicating perspective-taking where pragmatic inferencing is not at issue

In addition, adults are predicted to select the common ground card in the privileged ground ad hoc implicature condition – indicating not deriving an implicature, as the epistemic step cannot be taken, taking into account the speaker's perspective.

There are good reasons – reviewed above – to predict that children will also succeed in the privileged ground ad hoc condition. If they do, this would not be conclusive evidence of either the one-step or two-step hypothesis, as it could be that the age-range in this study is too old to capture development in progress. Alternatively, if children do not succeed in the critical condition, but in all other conditions, this is good evidence for the two-step hypothesis that children first acquire the two skills of pragmatic inferencing and perspective-taking separately, and then learn to combine them.

5.6.3 Results and analysis

5.6.3.1 *Sally-Anne Task*

All children passed the Sally-Anne task, except for one who was therefore excluded from the analysis. They also invariably answered the questions about which cards the puppet could see and know about correctly.

5.6.3.2 Children and adults in director task

Adults were at ceiling in all conditions except privileged ad hoc; children, on the other hand, were at ceiling only in the unambiguous and common ground ad hoc conditions.

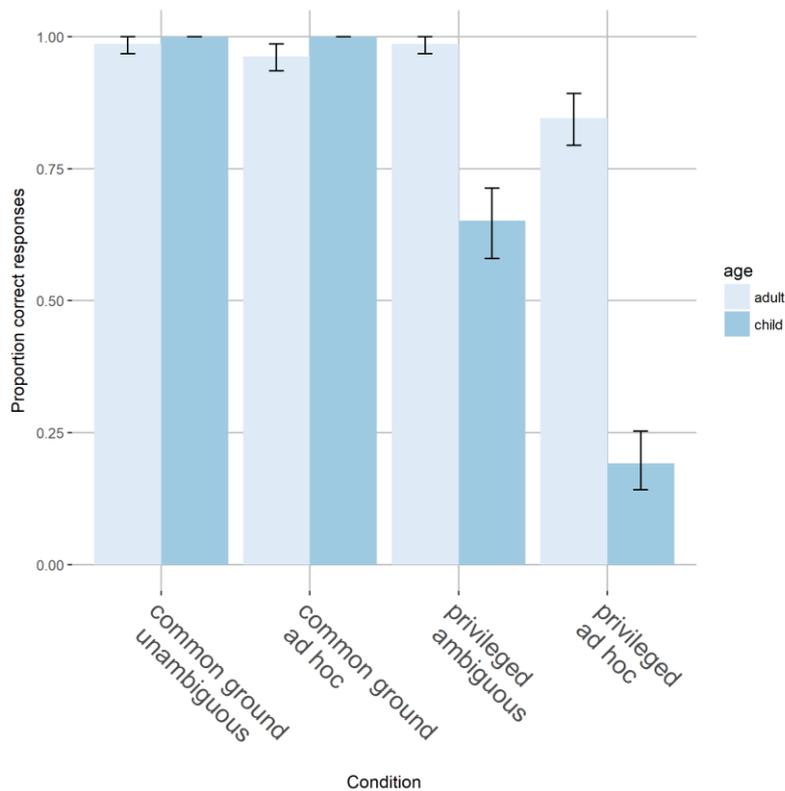


Figure 5.11 Proportion correct responses by condition and age group
Error bars show bootstrapped 95% confidence intervals for between-subject comparison

As the data was largely bimodally distributed (79% of children and 64% of adults scored either 6/6 or 0/6 in the privileged ground ad hoc condition; Hartigan's $D = .23$, $p < .001$), participants were coded as 'passers' (scoring 5/6 or 6/6) or 'failers' (otherwise), and χ^2 based analyses were conducted.

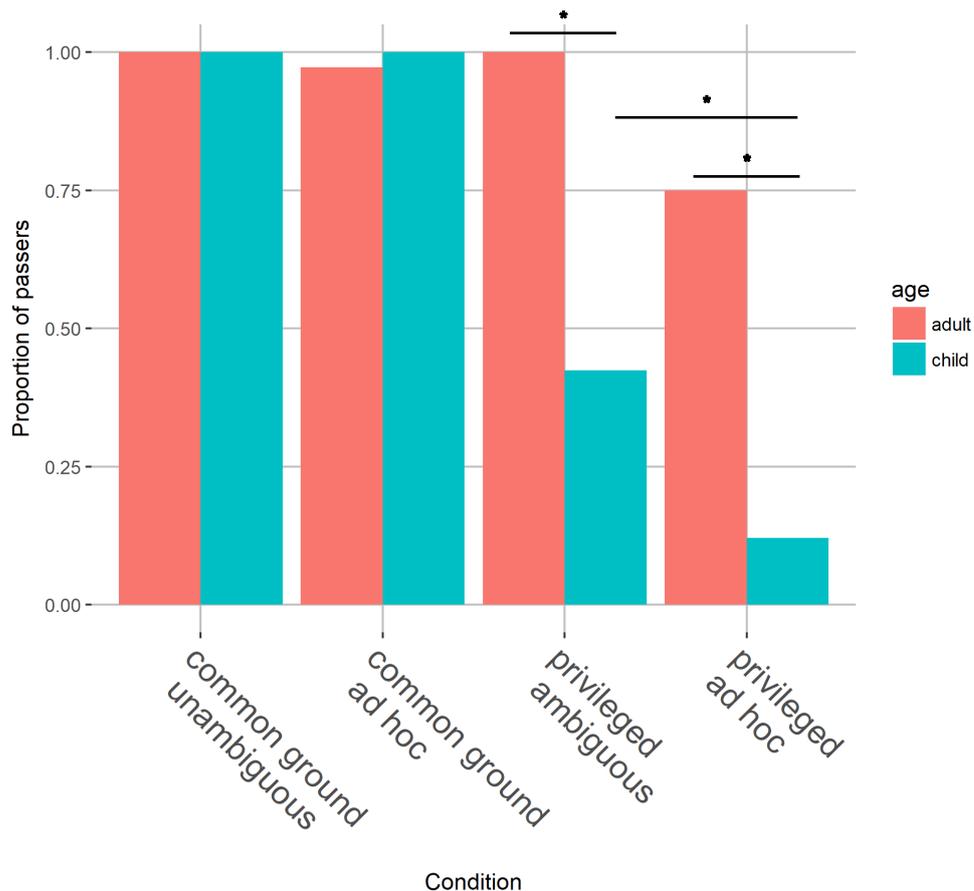


Figure 5.12 Proportion of passers and failers in each condition

To examine whether children’s performance in the two privileged ground conditions was the same, McNemar’s χ^2 test was used (for related groups, Table 5.9): there were significantly more passers in the privileged ground semantic, than privileged ground ad hoc condition (McNemar’s $\chi^2 = 8.5$, $p = .003$). To look at the difference between children and adults, Fisher’s exact test was used (Table 5.10 and Table 5.11): there was a significant association of age and performance, with more adult passers than child passers in the both the privileged ground semantic and privileged ground ad hoc implicature conditions (both $p < .001$).²¹

²¹ The maximal mixed effects logistic regression model (Barr, et al., 2013) failed to converge due to ceiling/floor performances and small random effect sizes (*lme4* in R: R Core Team, 2016; Bates, et al., 2015). A model with condition and age as fixed effects (sum coding), by-item (list) random slope, and by-subject random intercept, indicated a main effect of age ($\beta = 1.99$, $p < .01$) – children performed worse than adults – and condition (common ground ad hoc $\beta = 1.93$, $p < .001$; privileged ad hoc $\beta = -3.88$, $p < .001$; privileged semantic $\beta = -1.08$, $p < .001$).

Children		Privileged ground ad hoc	
Privileged ground ambiguous		Passer	Failer
	Passer	4	10
	Failer	0	19
(McNemar's $\chi^2 = 8.5$, $p = .003$; $N = 33$)			

Table 5.9 Number of child passers vs failers in privileged ground ad hoc and privileged ground ambiguous conditions

Privileged ground ad hoc		
	Passer	Failer
Adult	27	9
Child	4	29
(Fisher's exact test, $p < .001$)		

Table 5.10 Comparison of number of adult vs child passers in privileged ground ad hoc condition

Privileged ground ambiguous		
	Passer	Failer
Adult	36	0
Child	14	19
(Fisher's exact test, $p < .001$)		

Table 5.11 Comparison of number of adult vs child passers in privileged ground ambiguous condition

In other words, there are three groups of children: those that do not seem to take into account the speaker's perspective at all, although they excel with ad hoc inferences when all the relevant information is in common ground; those that are able to take into account the speaker's perspective to resolve semantic ambiguity, but are not able to not derive an ad hoc inference; and finally those who are adult-like and are able to take into account the speaker's perspective in both semantic and pragmatic interpretation. For adults, in contrast, there are only two groups: those that do and those that do not take into account the speaker's perspective in implicature derivation, with the latter in the minority in this experimental context.

Child#1	Child#2	Child#3	Adult	
✓	✓	✓	✓	Common ground unambiguous
✓	✓	✓	✓	Common ground ad hoc
-	✓	✓	✓	Privileged ground ambiguous
-	-	✓	✓	Privileged ground ad hoc
19	10	4		= N

Table 5.12 Numbers of passers in each group

5.6.4 Discussion

The results show that children at age 5–6 years have not yet developed an adult-like ability in combining perspective-taking with ad hoc implicature derivation in this kind of task. Adults in this task were able to derive an ad hoc quantity implicature when the speaker’s perspective converged with theirs; to take the speaker’s perspective into account in resolving a semantic ambiguity; and to take the speaker’s perspective into account to not derive an ad hoc implicature, when the speaker’s perspective differs from the hearer’s. In contrast, children excelled with ad hoc implicatures when the speaker’s perspective was not at issue, like adults; they also were able to make explicit judgements about other’s actions involving false beliefs, as they overwhelmingly passed the Sally-Anne test. However, more than half of the children appeared to fail to take into account the speaker’s perspective to resolve a semantically ambiguous utterance, and the vast majority persisted in deriving an ad hoc implicature, even when the speaker could not see the picture card associated with the implicated meaning. These findings suggest that the ability to integrate knowledge of a speaker’s epistemic state with pragmatic inferences develops gradually.

5.6.4.1 Perspective-taking and implicatures in adults: Gricean reasoning and individual differences

This is the first offline task with ad hoc implicatures with adults to demonstrate that adults can engage in full Gricean reasoning; previous studies either used online measures (Breheny et al., 2013), or looked at scalar implicatures (Goodman & Stuhlmüller, 2013). It therefore lends support to existing evidence that adults are able to take into account the speaker’s perspective in their pragmatic inferencing, in this case with quantity implicatures. As an offline study, the findings cannot arbitrate in

the debate about whether such altercentric perspective-taking is integrated early in processing (e.g. Hanna et al., 2003), or only as a late, effortful process (e.g. Epley et al., 2004), or is based on both perspectives considered simultaneously (Heller et al., 2016). Nor can these results contribute to answering the question of whether utterance interpretation is always fully Gricean – with intention-reading, belief tracking, and so on – or whether it can be accomplished with different strategies and differing extents of Theory of Mind involvement in different contexts. In this task, there are very strong contextual cues that perspective-taking is important for understanding the utterance, and so it seems reasonable to assume that even if adults use different strategies in different circumstances, they would consult the speaker’s perspective in this context. An alternative explanation would be that adults extrapolate a rule such as ‘never choose the card on the grey background’ and are largely able to follow this (as in control conditions of director ToM tasks, e.g. Symeonidou, Dumontheil, Chow, & Breheny, 2016). The current data cannot distinguish between these two possible interpretations²².

Adults are not at ceiling, though, in the critical privileged ground ad hoc condition, in contrast to the other three conditions. This is consistent with the individual variability observed in pragmatic inferencing tasks with adults (e.g. Franke & Degen, 2016) – not all adults are ‘perfect pragmaticians’. In many other studies, adults are not at ceiling in implicature inferencing conditions, even when there is no explicitly manipulated factor such as epistemic state. This could be perhaps because of other demands of the experimental context, even a lack of context and need for accommodation (e.g. Antoniou, Cummins, et al., 2016; Guasti et al., 2005; Scafton & Feeney, 2006). In situations where there *are* additional contextual factors to integrate, adults also seem to show differing degrees of integration of these into utterance interpretation. For example, Pogue, Kurumada and Tanenhaus (2016) investigated

²² Indeed, this is a potential problem with any version of the director task. However, there are at least two pieces of evidence to suggest that adults do not routinely ‘screen out’ the privileged object: a) in eye-tracking studies, participants are found to check the privileged object when it is a distractor (e.g. Epley, Morewedge, & Keysar, 2004), and b) Rubio-Fernandez (2016) investigated whether director tasks are testing ToM or merely selective attention, and in a new adaptation of the task found that ToM was indeed involved.

whether listeners are sensitive to a speaker's under-informative use of pronominal adjectives, and found that, while most participants were indeed adapting to the speaker's informativeness, a few-participants (5/32) persisted with a form-based generalisation even when the instructions were changed to cue attention to speaker differences in 'clarity' and 'naturalness'. Similarly, Dulcinati and Pouscoulous (2016) manipulated the speaker's cooperativeness, and found reduced levels of ad hoc implicatures when the speaker is uncooperative, but only at a drop from 90% to 72% in the co-operative speaker condition. Most relevantly, the findings from reference director tasks have been mixed, as discussed above, at the very least suggesting that this can be a challenging task.

The insights from Hawkins and Goodman (2016) and Heller, Parisien and Stevenson (2016), both within a constraint-based view of reference resolution, might be able to account for the difference between the two privileged ground conditions in adults. Hawkins and Goodman (2016) argue that in the kind of discourse context created by a director task experiment, hearers expect speakers to actually be over-informative, because of the uncertainty created by the privileged ground; the 'mistakes' hearers make in tasks like Keysar, Lin and Barr's (2003) are due to violations of these expectations. In their production and comprehension experiment, Hawkins and Goodman found that speakers did indeed use 'over'-informative referential expressions. Heller, Parisien and Stevenson (2016), meanwhile, propose that the conflicting results in the literature on reference resolution can be accounted for by a single model, in which hearers resolve the reference of the expression by considering both common-ground and privileged-ground domains simultaneously, but with different weightings, together with expectations about use of referring expressions. They do not consider pragmatic expectations of informativeness per se, but it might be informally incorporated to model this experiment's task.

On their model, for the privileged ground ambiguous condition, as the hearer considers only common ground, the utterance 'oranges' is a good match for the one card with oranges, on either an unenriched 'at least oranges' or an exhaustive 'only oranges' reading. The speaker is also optimally informative in using such a description, on either interpretation, as it is the only card with oranges in common ground. Secondly, as the hearer considers both common and privileged ground together,

'oranges' is now an equally good match for either card with oranges, again on either the 'at least' or 'only' readings. But now the speaker would be under-informative, given that the utterance is therefore ambiguous. Overall, therefore, the common ground card is the best interpretation, as it preserves the speaker's informativeness – and in my experiment, adults are at ceiling in this condition.

For the privileged ground ad hoc condition, though, the pattern is different. Considering only common ground, the utterance 'pears' is a match for the card with pears and bananas on an 'at least' reading – and strictly speaking it is an informative description as a unique identifier of the only card with pears, from the speaker's point of view. However, on an exhaustive interpretation, it is not a good match; and, further, in this referential-communicative context, the hearer might expect the speaker to be more informative, e.g. 'the card with pears and bananas'. In either case, the speaker would be uncooperative. Considering the common and privileged grounds together, 'pears' is a match for either card with pears (and bananas) on an 'at least' reading, but only a good match for the card with only pears on an exhaustive reading. So, on no enrichment, it is ambiguous, and the speaker is under-informative, but, on the exhaustive inference, it is not ambiguous and the speaker is optimally informative. Therefore, overall, there could be more of a tendency for hearers to consider the card in privileged ground in this condition than in the privileged ground ambiguous condition, as it fulfills expectations of informativeness of the speaker.

5.6.4.2 Perspective-taking and implicatures in children: a two-step development

The findings suggest three groups of children: those who fail to take into account the speaker's perspective in a straightforward case of semantic ambiguity let alone pragmatic inference; those who can take the speaker's perspective in the semantic ambiguity condition only, but not in the implicature condition; and a minority who are able to take into account the speaker's perspective in both the semantic and pragmatic privileged ground conditions, in an adult-like way. However, children were overwhelmingly able to pass the Sally-Anne Theory of Mind test, which arguably represents more complex Theory of Mind skills including level 2 perspective-taking. They were also able to explicitly state which picture cards the puppet could and could not see and know about.

Taken as indicating a developmental progression, these findings support the two-step development hypothesis: first children acquire the ability to reason about others' epistemic states and make pragmatic inferences separately – seen in this study in the group of children who are able to derive ad hoc implicatures and reason about different perspective and false beliefs, but who do not take into account the speaker's perspective in deriving implicatures. Then, later, they learn to integrate the two skills – seen in the small group who take the speaker's perspective into account to resolve semantic ambiguity, and derive an implicature or not as appropriate for the speaker's epistemic state. Anecdotally, some of the children's comments suggest that this interpretation is on the right track. In the critical condition, a few children hesitated or expressed doubt as to which picture card they should choose, indicating that they were aware of the conflicting cues (linguistic and contextual). Others made comments such as 'Bob has x-ray eyes' or thought that Bob could see the privileged card, again suggesting an attempt to reconcile their interpretation of the utterance with the contextual information. These children are aware of both sources of information, but not yet able to integrate them in pragmatic processing.

This two-step account accords with other studies that find that integrating relevant contextual information into implicature inferences is challenging for young children. For instance, children at this age seem to struggle to track the QUD and recognise relevant alternatives when deriving scalar implicatures. Skordos and Papafragou (2016) found that when the implicit QUD alternates between quality and quantity – when a first statement is rejected because it is false, and a second statement is to be rejected because it is under-informative – children perform worse compared with when the QUD is consistently a matter of quantity. A similar finding comes from Horowitz and Frank (2015): children's success in scalar implicatures in a picture-matching task is lower when scalar and ad hoc trials are mixed together than when there are only scalar trials. While both implicatures in this study are quantity, it could be the subtly changing QUD that is challenging for young children – from an implicit 'How many...?' to 'What is everything...?'.

Further, over the course of development from age 5 years to adult, Scafton & Feeney (2006) observe a rise and then fall in the proportion of scalar inference responses in a judgement task with little supportive context. While the authors

suggest an explanation in terms of dual processes, this pattern can also be simply explained in terms of integration of contextual information: the youngest children perform poorly as they are still acquiring scalar implicatures (which are acquired later than ad hocs); the older children have acquired the ability to derive implicatures but do so regardless of the context, which they have not yet learned to integrate into the inferencing process; the oldest children and adults take into account the context, which in this case does not highlight quantity as a QUD as it consists of ‘out of the blue’ statements to judge, and therefore do not derive implicatures more often.

In Experiment 4, I also found that children struggle even to integrate the speaker’s perspective to resolve a semantically ambiguous utterance, in contrast to adults. This is more puzzling. This condition was designed as a check that children can indeed take the speaker’s perspective, when no pragmatic inference is required, as would be expected given the previous findings that children are able to integrate speaker perspective in reference resolution (Nadig & Sedivy, 2002; Nilsen & Graham, 2009). The findings suggest, however, that even in this case, integrating contextual information with linguistic comprehension is challenging, as indicated by the inconsistent responses of some children (picking both the common ground and privileged picture card across trials). Indeed, not only pragmatic, but also semantic and syntactic processing is sensitive to contextual factors, and this may be more fragile in children, especially where available visual stimuli conflict with common ground (De Cat, 2015; Pomper & Saffran, 2016; Trueswell, Sekerina, Hill, & Logrip, 1999).

The pattern of children’s performance here fits in well with Papafragou and Skordos’ (2016) proposal that across different pragmatic phenomena, the ability to integrate linguistic and non-linguistic sources of information develops gradually. The children in Experiment 4 are at an age when this integration is only starting to happen for ad hoc quantity inferences in a task of this complexity. More work is therefore needed to chart this development with increasing age, to confirm this hypothesis. At a different level of analysis, a similar theory is that children have difficulties with ‘top-down’ pragmatic processing, where contextual cues – such as an array of pictures in common ground – are used to predict a possible utterance and its meaning (Snedeker, 2015; Snedeker & Huang, 2016). Instead they tend towards ‘bottom-up’ processing, starting with the literal meaning of the sentence, and then deriving the implicated

meaning, possibly taking contextual information into account. One piece of evidence for this hypothesis comes from Rabagliati and Robertson's (2017) study on referential production, which found that adults but not children pro-actively monitor non-linguistic ambiguity – noticing two of the same type of object in a visual array before knowing the referential expression they are required to produce.

As Katsos and Wilson (in prep.) suggest, these two theories can easily be combined: the top-down route requires early integration of multiple sources of information, including non-linguistic sources of information like world knowledge stored in long-term memory, and the immediate discourse and visual context in short-term or working memory. The bottom-up route, on the other hand, can rely on the literal meaning of the utterance to cue the search for other relevant pragmatic information, although this is likely to be more effortful, and, in the case of children, potentially not successful at all. This theory could explain the findings of this study: while adults are able to take in the visual array and the speaker's different perspective, and anticipate likely utterances (including the fact that the speaker will not refer to the picture card in privileged ground, especially in the non-linguistic ambiguity case of the semantic privileged ground condition), children start with the utterance and then struggle to integrate the contextual information, including the speaker's perspective, to appropriately infer the intended meaning.

How does this relate to the constraint-based view presented above? One option is that children weight the privileged ground domain more than adults do. Another is that they weight speaker informativeness more than adults do, and so choose the option for privileged ground ad hoc that renders the speaker maximally informative, over considerations of common versus privileged ground. A related idea is that the description in the privileged ground ad hoc condition is a preferred match for the privileged ground picture – on both an 'at least' and an exhaustive reading – and that this is weighted by children more than speaker perspective. In contrast, for the privileged ground ambiguous condition, weighting the privileged ground domain more than adults does not change expectations as much – it makes the utterance ambiguous, rather than favouring the privileged ground picture card. And indeed, we see a flatter distribution of correct choices for this condition than the privileged ground ad hoc one (which is bimodal across children and adults). Pragmatic

development, then, involves learning to appropriately weight conflicting cues to speaker meaning.

What might contribute to this shifting weighting of contextual cues? While some research has demonstrated a link between children's Theory of Mind and referential communication (Khu et al., unpublished research, cited in Graham et al., 2016), this is unlikely to be a contributing factor here, as children overwhelmingly passed the Sally-Anne test of False Belief reasoning, as expected at this age, and level 1 perspective-taking is available even earlier. More likely, the task itself might have proved too challenging for some children, given their developing Executive Function skills, especially inhibition²³. The preschool years are a key time of change in Executive Function abilities, in terms of both the components and their integration (De Cat, 2015; Diamond, 2006). Some studies have found a relationship between children's inhibitory control and their perspective-taking in referential communication (Nilsen & Graham, 2009), as well as their sensitivity to communicative ambiguity (Gillis & Nilsen, 2014; Nilsen & Graham, 2012), although other studies have failed to find this association (e.g. Nilsen, Mangal, & MacDonald, 2013). The better the hidden object matches the utterance, the more inhibition is required to choose the common ground object instead. In Experiment 4, the privileged ground picture in the privileged ground ad hoc condition was arguably a better match, so might require more inhibition to suppress. The common and privileged ground pictures in the privileged ground ambiguous condition are identical – and therefore, semantically, an equal match for the utterance – which might be a reason for better performance in this condition, although other challenges might also prevent adult-like performance, which I discuss below.

5.6.4.3 Theoretical implications: different pragmatic strategies

The starting point for this study was the Gricean model, in which Theory of Mind, and particularly reasoning about the speaker's epistemic state, play an important role in pragmatic inferencing, which presents a puzzle for development. I now take a step back, therefore, and consider what the findings from this study mean

²³ Executive Function is typically divided into working memory, cognitive flexibility, and inhibitory control.

for pragmatic theory.

In comparison to Experiment 1B, this study was designed more specifically to investigate a component of ToM in inferencing, namely perspective-taking. The findings therefore have more relevance for the debate on the nature of pragmatic inferencing, namely whether speakers are always ‘fully Gricean’, employing their ToM for utterance interpretation, or, instead, are able to engage in what are typically thought of as pragmatic phenomena without mind-reading. The findings suggest that, at the very least in development if not across the lifespan, pragmatic reasoning does not always have to involve taking into account another’s epistemic state, as a simple psychological instantiation of Grice’s model would imply. As in many other studies where there was no difference in perspective between interlocutors, children excelled in deriving ad hoc quantity implicatures. But, they could not take into account the speaker’s perspective when this was different from their own, despite salient cues – suggesting that even in the shared perspective case, they were perhaps simply assuming common ground rather than actively taking this into consideration in the pragmatic inference.

This puzzle can be approached in at least three different ways. In the first, it can simply be argued that children are still developing in their cognitive capacities, but once they do, they surely employ Theory of Mind in their pragmatics. Thus, child data is not problematic for pragmatic theory (for typically developed adults); that can be kept intact, and a separate model proposed for child pragmatic development. A second, related, approach is to distinguish social and visual perspective-taking, in communication and in development, again leaving standard pragmatic theory largely intact. On a third approach, one can aim for a single unified theory that can account for adult and child data, at least on the grounds of parsimony: the child data presented in this and other studies then becomes highly problematic for pragmatic theories that assume mind-reading. I now discuss these three approaches in more detail.

1. Developing ToM and pragmatic skills

One kind of possible solution is put forward by Breheny (2006), whose idea is that ‘basic communication’ – which can look very much like more sophisticated communication – is actually possible without full ToM, but instead only with joint-

attention, action concepts and benevolent tendencies that emerge around 9–12 months of age (Tomasello, 2003). On his ‘minimalist relevance-theoretic’ account, ‘the prototype for the concept of communication includes one agent drawing another agent’s attention to a situation in a shared situation’ (Breheny, 2006: 96), where situation is defined technically as a set of properties, individuals and locations, in the spirit of Barwise and Perry (1983). What is useful is the notion that children might be merely drawing on their ability to jointly attend in a shared situation and on their understanding of goal-directed actions to engage in a kind of ‘basic communication’, which adults can nevertheless interpret intentionally. Children then ‘grow out’ of this strategy – when they develop ToM, they develop adult-like pragmatic skills.

However, Breheny’s (2006) argumentation rests on taking scalar implicatures as an example of a phenomenon that uncontroversially requires ToM and, drawing on earlier studies that found competency with scalars only at 5 years and above, he cites this as evidence for ‘basic communication’ before this age. Ad hoc and relevance inferences are arguably even surer cases of full pragmatic phenomenon (given more recent grammatical approaches to scalar implicatures), but are acquired even younger, before children are reliably passing standard False Belief tests – as in Experiment 1A. Furthermore, Breheny’s theory as it stands could not account for the two-step developmental trajectory seen here: some linking explanation is required for the fact that acquiring Theory of Mind enables pragmatic processing whilst this pragmatic processing does not immediately draw on all acquired components of Theory of Mind. In other words, how can ‘basic communication’ include the capacity for deriving implicatures in some situations?

2. Developing different perspective-taking abilities

A related but more promising suggestion comes from the observation that – counter-intuitive as it may seem – visual perspective-taking is in fact more challenging than other seemingly more complex aspects of social cognition. In Experiment 4, and to some extent in Experiment 3, visual perspective-taking was taken as a proxy for mind-reading abilities. However, Moll and Kadipasaoglu (2013) argue that social perspective-taking and visual perspective-taking are distinct, and that the first is ontogenetically primary: young infants succeed with perspective-taking when it is

based on common experience – i.e. shared action and shared discourse – but visual perspective-taking develops later (aged 2–3 years), and patchily at that. Crucially, visual perspective-taking differs in that it relies on static in-the-moment cues to a difference in perspective, rather than dynamic, relational ones. They suggest that ‘joint attention thus directly paves the way to early forms of social, but not visual, perspective-taking’. Similarly, Bishop and Adams (1991) question what is being tested in referential communication tasks (where privileged ground is not at stake): in children with SLI, their informativeness in such a task did not match their sensitivity to listener needs and knowledge in a conversational setting, suggesting task factors such as scanning the array are an issue (supported in more recent work by Davies & Kreysa, 2017; Rabagliati & Robertson, 2017).

If visual perspective-taking is additional to or indirectly related to ToM, then one might expect that integrating it into pragmatic inferencing is a skill that emerges later in development. In other words, if the question of interest is how and when children use their ToM abilities in a pragmatic skill, such as implicature derivation, visual perspective-taking tasks may not actually be the best place to start. Instead, social perspective-taking, where common and privileged ground is established over interaction and discourse, would be a better indicator. This insight allows two seemingly opposing views and findings to be integrated: language and communication can still be fundamentally about intention-reading – being ostensive-inferential – with communication development going hand-in-hand with joint attention, intention-reading and eventually full ToM development. Meanwhile, some cognitive abilities, such as visual perspective-taking, present serious challenges when the need to integrate them with linguistic processing arises. It also leaves open the possibility that a standard approach to pragmatic theory – about speakers’ communicative intentions – can apply throughout the lifespan: in effect, development follows the one-step hypothesis, with only some edge cases providing exceptions to the trend, such as combining inferencing with visual perspective-taking.

3. Developing pragmatic skills without ToM

Alternatively, it could be the case that pragmatic communication does not have to be so complex, not only in development but also across the lifespan. Kissine (2016)

argues that pragmatic processes must be kept separate from pragmatic norms in modelling pragmatic competence (see, too, Jary, 2013; Sperber, 1994). The context determines the pragmatic norm – which is like a strategy – which in turn determines the pragmatic processes involved in interpretation. The norms can be purely egocentric, allocentric, or fully Gricean, where allocentric norms require at least implicit first-order ToM and enable interpretations that are at odds with the speaker’s perspective to be ruled out, while fully Gricean norms involve full ToM. The idea is that, while adults can switch between interpretation strategies, children develop these strategies consecutively. Crucially, on Kissine’s model, some kinds of implicatures can be calculated via the pragmatic processes determined by an allocentric norm, while others require full Gricean processing. The first kind only requires expectations of relevance in the situation rather than reference to speaker intention. In the following example, the utterance itself makes an assumption based on background knowledge available (having had breakfast is a good reason for not wanting coffee and croissant) which allows the hearer to arrive at the implicated meaning:

I have already had breakfast.

→ The speaker does not want coffee and croissant.

(Kissine, 2016)

The ad hoc quantity implicatures in the present study could potentially also be analysed in a similar way: given the visual context, ‘the card with pears’ is a relevant (and in this case informative) way to describe a card with only pears – from the speaker’s or hearer’s perspective. In other words, this model challenges the received view that an implicature derivation involves taking the Epistemic Step based on the Competence Assumption, which makes it easier to explain why taking the Epistemic Step could be easier than *not* taking it – because an implicature *can* be derived without it.

As it stands, though, Kissine’s model also does not straightforwardly explain a two-step development, as implicature interpretation happens only with at least allocentric interpretation, which *does* take the speaker’s perspective into account, so a situation where hearers can derive implicatures but not take the speaker’s perspective into account in communication would not be predicted. More recently, though, he has suggested that some inferences, like scalar implicatures, may be available via

egocentric reasoning strategies, in the context of the pragmatic abilities of people with Autism Spectrum Disorder (Kissine, 2017). Extrapolating from those comments, it thus seems best not to take types of pragmatic phenomena and categorise them by the amount of perspective-taking required. Rather, an inference may be available on egocentric strategies in one context, but not in another, depending on what contextual information is needed – for example just world knowledge, or speaker perspective as well. Thus, children might succeed with quantity implicatures where the speaker’s perspective is not at stake, but not succeed where it is, while adults may employ different strategies context-dependently.

Note that this suggestion and the previous explanation, about the distinct challenges of visual perspective-taking, are not mutually exclusive. It remains an open question whether the integration of speaker perspective with pragmatic inferencing is inconsistent only in development or across the lifespan. The challenge for models like Kissine’s, as well as similar ones such as Andres-Roqueta and Katsos’ (2017), is to flesh out which aspects of context trigger which interpretive strategies, how these are monitored, and whether there is a default strategy.²⁴

5.6.4.4 *Improvements to the task*

This novel experimental paradigm, which combined the classic reference resolution director task with an ad hoc picture-matching task, enabled an assessment of perspective-taking in ad hoc implicature derivation. There are a number of ways that this task could be improved upon, as well as ways in which it could be extended.

Firstly, the design copied that of Horowitz and Frank’s (2015) picture-matching task, in which the picture cards always displayed the same number of items. For the ‘only Xs’ card this was, for example, 4 cats, and for the ‘Xs and Ys’ card, 2 cats and 2 birds. I used 5 items, so that the alternative ‘half’ was slightly less salient. However, a potential problem remains: this means that the number of Xs differs between the ‘pragmatic’ choice card, with 3 cats, and the distractor card with 5 (or vice versa, in the privileged ground ad hoc condition in Experiment 4). For those children who are not integrating perspective at all, therefore, it could be that they choose the card with

²⁴ These are just examples of more minimal models of communication, without full ToM; there are, of course, others, such as Richard Moore’s (e.g. Moore, 2016a, 2016b) or Liz Irvine’s (2017).

only Xs because they have derived an ad hoc implicature, or because it is simply more visually salient and more cat-ish, for example. This could easily be checked by varying the number of items on each card, and making sure that they are matched between 'pragmatic' and distractor choices.

Secondly, following Nilsen and Graham (2009), there was a warm-up production task that was meant to draw attention to the speaker's different perspective, as the puppet asked the participant to describe the card he could not see. However, there is a risk that this might suggest to children that the aim of the game is to show the puppet what is on the hidden card – even though the puppet announces after the warm-up: 'Now let's play a different game'. Children can struggle to switch between tasks and QUDs so there is the possibility that some perseverated with the initial game, and therefore were more inclined to choose the privileged picture card as the one the puppet wanted. This could have been reinforced by the lack of (negative) feedback to card choice: the puppet always responded, 'Okay, now I'm turning around'. Children are able to learn about expectations of informativeness from feedback (Matthews, Butcher, Lieven, & Tomasello, 2012; Matthews, Lieven, & Tomasello, 2007), and so this could have reinforced the child's interpretation strategy if interpreted as a positive response. Another factor that may have caused the high rate of selection of the privileged card in the semantic disambiguation condition is the mixing of trial types: if children are unable to integrate speaker perspective in implicature derivation, this forces them to choose the privileged picture card in the privileged ground ad hoc condition, which, in turn, licenses selection of the privileged card for the semantic condition. This would lead us to expect decreasing performance in the privileged ground ambiguous condition over the experiment, which, however, is not exactly the pattern observed (Figure 5.13). Separating trial types into blocks would be one solution, or making sure the privileged ambiguous always precedes the privileged ad hoc condition in any block.

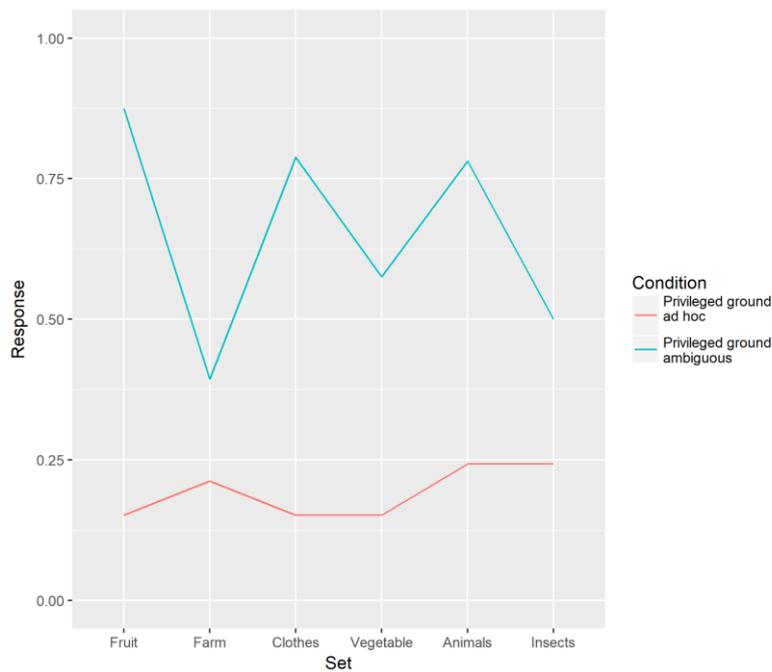


Figure 5.13 Mean correct response by block (with blocks in order of presentation) for both privileged ground conditions, children only

More generally, I have suggested that this task in particular may carry challenges that are not directly related to children’s pragmatic competence. It may be that they are aware of the conflicting information, but unable to resolve it: online methods such as eye-tracking and reaction-time measures may reveal such patterns of behaviour. For instance, one might predict that of those children who fail in the critical privileged ground ad hoc condition, those that can do perspective-taking, as shown in the privileged ground ambiguous condition, might be slower to react than those who cannot, as they try to resolve the conflict. Furthermore, a task that combines social, instead of visual, perspective-taking with implicature derivation might be less challenging for children, if indeed social precedes visual perspective-taking in some senses, as Moll and Kadipasaoglu (2013) argue. For instance, imagine a scenario as in Figure 5.14, somewhat akin to the Change of Location task: here the hearer has to *not* derive an exhaustive implicature, choosing the box with only dogs from their

perspective, but instead choose the box with dogs and pigs, which, from the speaker's perspective, is the box with *only* dogs.

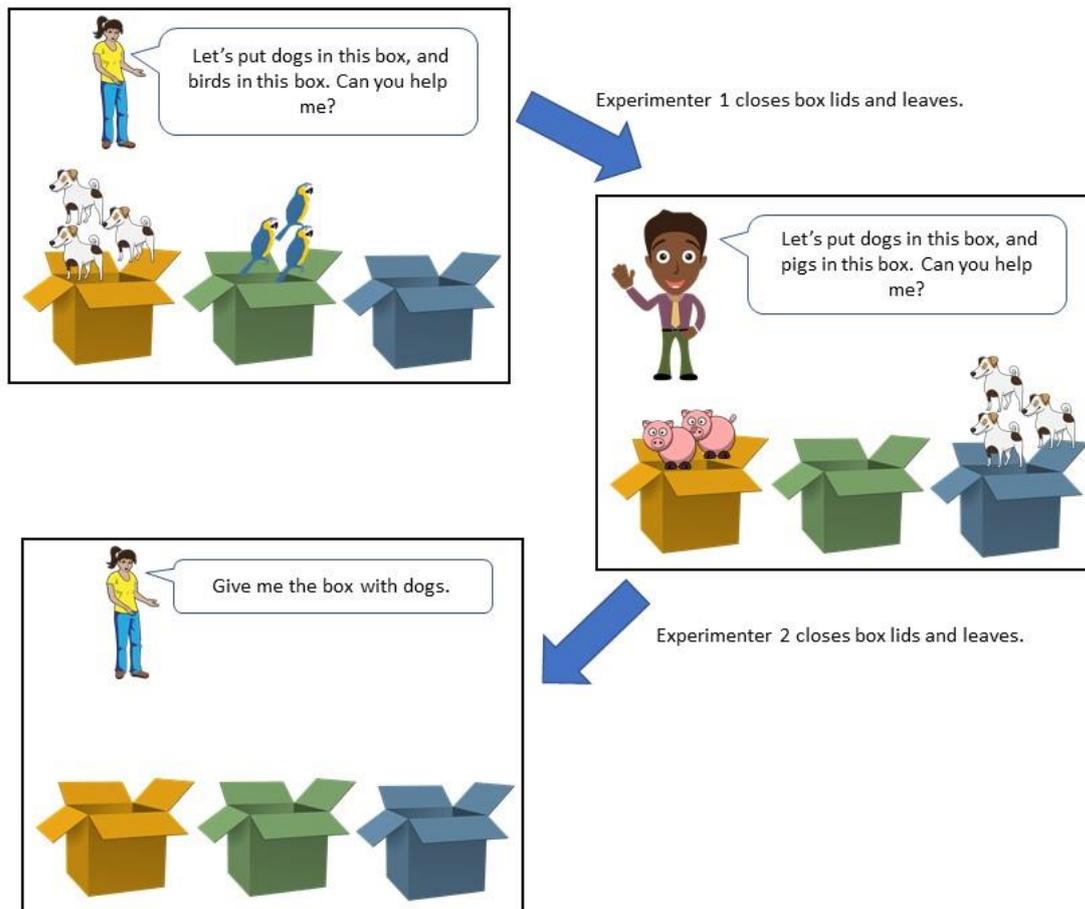


Figure 5.14 Example of perspective-taking ad hoc implicature task

5.7 Conclusion

In this chapter, I set out to address two questions. Firstly, do children engage in perspective-taking in implicature derivation? And secondly, what does this mean for a Gricean model of implicature? In two studies, I concentrated on children's quantity implicature understanding.

The findings confirm that young children *are* sensitive to a speaker's epistemic state or perspective in communication, as seen in Experiment 3. However, they are not always able to integrate this with implicature derivation. While they may look very much like adults in straightforward communicative contexts – such as have predominated in developmental studies thus far – they are found to be still developing in implicature skills when the speaker's epistemic state must be taken into account.

Experiment 4 provided evidence for a two-step development, with implicature derivation and reasoning about speaker perspective developing separately, before being combined as speaker perspective is taken into account in pragmatic processing. I further suggested that this is a result of the difficulty of monitoring contextual cues and integrating them with pragmatic processes.

These findings present a challenge for a Gricean model of implicature in which reasoning about the speaker's epistemic state is integral. I identified two promising avenues for solutions to this puzzle. On the one hand, it could be that children can actually take into account the speaker's perspective for implicatures from the outset, but only when it is social perspective-taking, rather than visual perspective-taking, which presents its own idiosyncratic challenges. On the other hand, it could be that children and adults have a variety of strategies open to them to 'do pragmatics', not all of which involve complex mind-reading. These options are, of course, not mutually exclusive, but both require further theoretical development and empirical investigation.

6 Conclusion

6.1 Summary of the thesis

This thesis set out to investigate the development of young children's understanding of implicatures. In particular, it looked at a range of implicature types, and the role of other factors in their development, such as structural language knowledge and Theory of Mind. It presented a series of studies that used child-friendly behavioural measures, all based on a picture-matching task, to examine children's competence. It contributes to the experimental pragmatics literature on children's development, particularly by showing that children's ability to engage in deriving implicatures develops in the preschool years, from age 3, and is associated with their structural language knowledge. In addition, though, children's early competence in straightforward communicative situations is paired with ongoing development in more complex ones, where sources of contextual information, such as the speaker's epistemic state, have to be integrated into the pragmatic inference. Therefore, at an age where children may show adult-like competence with implicatures in some situations, they may in fact not attain it in others, depending on the cognitive skills required and still to be acquired.

In Chapter 2, Experiment 1A aimed to address two research questions: what is the relationship between different implicature types in development? And, what is the nature of WLE? The findings from the novel picture-matching study, which combined quantity, relevance and WLE inferences, added to a growing body of evidence that the preschool years, age 3–5 years, are a crucial time for learning to understand implicatures. 5-year-olds are approaching adult-like competency in all inference types, but the youngest of the 2- to 3-year-olds excel only with a potentially related inference, WLE. They also showed that word learning exclusion inferences are the first to develop, followed by relevance and ad hoc quantity, and finally by scalar inferences. The correlation of children's performance on relevance and scalar implicatures was tentatively suggestive of the important role played by sensitivity to relevance and elaborative inferences for both implicature types. This was a first study to compare WLE inferences with a clearly pragmatic skill like implicatures, but, given children's

high performance in WLE, the potential to answer the second research question was limited. In the youngest age-group a correlation between ad hoc and word learning inferences could be attributed to a common pragmatic component, as predicted by a pragmatic theory of WLE, or to their common exclusion mechanism.

In Chapter 3, I explored cognitive and environmental factors that might be associated with children's pragmatic development, using standard tests of vocabulary and grammar, Theory of Mind, socioeconomic status and number of languages spoken, with a subset of the data collected for Experiment 1A. The main finding was that structural language knowledge was a key predictor of children's implicature understanding at this young age. Socioeconomic status was also associated through correlation with structural language. There was no evidence that growing up bilingual or monolingual makes a difference for implicature understanding. There was also no evidence for an association with Theory of Mind, and, given that this was so surprising, I followed this up with a more direct investigation in Chapter 5.

In Chapter 4, I turned to manner implicatures, asking whether adults and children derive manner inferences, as indicated by pragmatic theories. Given their relative neglect in the theoretical and empirical literature, more groundwork was required, and the results are more tentative than for quantity and relevance implicatures. I set out a working definition of manner implicatures, and spelt out some predictions of acquisition, suggesting that on theoretical grounds they might be predicted to be the last of the implicatures to emerge in development, while given empirical work on comparable skills they might, alternatively, be available relatively early. The results of experiments 2A and 2B with 5–7-year-olds suggested that children and adults may have similar understanding for clearly contextually-dependent inferences. However, investigating manner implicatures poses particular challenges, as discussed.

Finally, in Chapter 5, the role of Theory of Mind in implicature inferences was addressed more directly, by asking whether children can take into account the speaker's perspective in utterance interpretation. I presented two experiments, both with quantity implicatures only. The results of Experiment 3 suggested that children, as well as adults, are sensitive to the speaker's epistemic state, but the design was unable to tease apart the contribution for pragmatic inferences in particular. In

Experiment 4, I combined an ad hoc picture-matching task with the referential communication director task, and found evidence for a two-step development in children: first they learn to take a person's perspective into account outside utterance interpretation and for semantic disambiguation, and to derive implicatures; then they learn to integrate both skills, of implicature derivation and perspective-taking. This indicated that children's adult-like performance in tasks with a simple context, as Experiment 1A, can belie their ongoing development, particularly in learning to integrate different sources of relevant information.

6.2 Implications for developmental pragmatics

This thesis contributes to the growing body of evidence to suggest that implicature understanding develops relatively early as a pragmatic skill, beginning from around 3 years. It may emerge before some pragmatic skills like irony (Filippova, 2014), but alongside others like presupposition (Bill et al., 2014; Pouscoulous, 2013) or some metaphors (Pouscoulous, 2014). It may be preceded by related inference-making abilities that share some components, such as WLE, which involves negation of alternatives and which, in Experiment 1A, was found to be associated with ad hoc implicatures.

However, the ability to understand implicatures does not emerge at a single point, but is instead ongoing in development throughout childhood, depending, firstly, on the type of implicature, and, secondly, on the context. Firstly, the ability to derive different types of implicature may be learned at different times. Due to the nature of their derivation, relevance, ad hoc and scalar implicatures may be learned to some extent sequentially, as seen in Experiment 1A. Furthermore, even once children are sensitive to informativeness, relevance and manner, and can engage in pragmatic reasoning, particular instances of implicatures may be more or less challenging, depending on the linguistic and real-world knowledge that they demand. For instance, quantity implicatures might depend on the particular scalar relations between quantifiers or real-world sets; relevance implicatures on the linking fact from world-knowledge in an elaborative inference; or manner implicatures on the linguistic knowledge of the unmarked alternative. In this thesis, it was seen, for example, that scalar implicatures with 'some' seem to be harder than ad hocs or relevance inferences,

owing, at least in part, to the challenge of learning quantifiers – remember that 3-year-olds succeeded with control trials with the stronger alternative ‘all’ only 76% of the time in Experiment 1A. Similarly, the results of Experiment 1B indicated that structural language knowledge was a main predictor of implicature skill (though, of course, the directionality of this association could not be established). And again, in Experiments 2A and 2B on manner inferences, those GCIs that depend on particular linguistic knowledge of alternatives – causatives – seemed to pattern differently for children and adults.

Secondly, the communicative context may make it more or less challenging for children to arrive at the implicated meaning. In Experiment 1A, children performed well in comparison to other similar studies, most likely because of the experimental context, in which alternatives were provided visually and trials included a context question. In Experiment 4, children were sensitive to the speaker’s epistemic state, but unable to integrate it in the pragmatic inference – they continued to derive an ad hoc implicature even when the speaker was ignorant, unlike adults. This contributes to an overall picture in which children may find it challenging to integrate linguistic and non-linguistic information in utterance interpretation, as suggested by Papafragou and Skordos (2016). Towards the algorithmic level of explanation, this in turn might be due to limitations on children’s capacity to predict upcoming utterances and their meaning, given a discourse and visual context (Snedeker, 2016).

In other words, there are many aspects of implicature understanding, which develop over time in interdependence and contribute to the gradual progress towards adult-like competence: knowledge about communication (including expectations of cooperativeness), structural language knowledge, and world knowledge; pragmatic reasoning skills, including elaborative inferences and reasoning by exclusion; and social cognition, including mindreading and, especially, reasoning about others’ beliefs. Once these knowledge and skills are in place, children must also learn to integrate them online for utterance interpretation. Our understanding of the way these factors work together in adults’ pragmatic understanding is being greatly enriched in current research, particularly within a constraint-based approach to implicatures (e.g. Degen & Tanenhaus, 2014; Pogue, Kurumada & Tanenhaus, 2016; van Tiel, van Miltenburg, Zevakhina & Geurts, 2014); the agenda for developmental

pragmatics is to build up such a fine-grained model for children's changing understanding, too.

6.3 Implications for pragmatic theory

Three implications for pragmatic theory can be highlighted. Firstly, Chapter 4 sought to investigate manner implicatures in development, but against a dearth of theoretical and empirical research. While they can be categorised theoretically as a distinct class to some extent, it proved challenging to find examples that were unambiguously manner, rather than quantity as well. This stems from an essential property of manner inferences – they are based on the *form* of the utterance, not the content, and so the alternative must have the same semantic content. However, in line with the Principle of Contrast, discussed in Chapter 2 in the context of word learning (E. V. Clark, 1988), some, like Clark, would argue that true synonymy is never found in a linguistic system: there is always some difference in meaning, even if it is in the sociolectal information attached to the word, for instance. One exception might be alternatives in English such as 'open' – 'cause to open'. However, it was seen that even with these there are complex usage patterns in terms of agency, direct causality, and animacy (e.g. Song & Wolff, 2005), that go well beyond the simple predictions made by neo-Gricean pragmatics (e.g. Franke, 2009; Levinson, 2000). And, needless to say, this example is language-specific, so more cross-linguistic cases would need to be documented before it can be taken as a robust phenomenon.

Yet, there were some indications in Experiments 2A and 2B, and from previous work (E. Wilson & Katsos, 2016), that adult speakers do sometimes derive manner implicatures, particularly from the more qualitative data of justifications or free responses: hearers inferred something about the non-stereotypical nature of the act described, presumably via the marked form. Perhaps what is needed, therefore, is a new category, where informativeness and manner interact – where the form of the utterance is an additional, intentional, trigger to the speaker's meaning, on top of the over-informative content. If so, then a further open question is whether the degree of markedness contributes not just to recognising a marked form, triggering a manner implicature, but also to the inferred meaning. In other words, do speakers use a greater discrepancy in markedness – or frequency – between the marked form and its

alternative to indicate a more atypical situation? Consider, for instance, the following cases:

- a) Bill is meeting a woman this evening.
→ Bill is not meeting his wife. (*Via quantity*)
- b) Bill is meeting a female adult this evening.
→ ? Bill is not meeting his wife this evening, and it is unusual for him to meet up with women. (*Via quantity and manner*)
- c) Bill is meeting a female human being this evening.
- d) Bill is meeting a member of the species *homo sapiens* of female gender this evening.²⁵

On an information-theoretic approach to language, frequency in context and informativeness are negatively correlated – such that the lower the frequency in context, the more informative. This could be a promising way to conceptualise the relationship between quantity and manner, and to test predictions in processing and in development (cf. Bannard, Rosner & Matthews, 2017, for an information-theoretic approach to children’s production). When compared with the rich understanding of quantity, and especially scalar, inferences, there is much work still to be done on manner.

Secondly, while the focus within developmental – and to some extent, theoretical – research has been on single types of implicature, the way that maxims apply together and simultaneously must not be neglected. Grice (1989) himself considers cases of clashes between maxims, and Levinson (2000) proposes a hierarchy for his three heuristics, Q, M and I, to resolve conflicts. However, hearers also understand the speaker’s meaning by making assumptions about the maxims in concert. In Chapters 1 and 2, I argued, following Skordos and Papafragou (2016), that sensitivity to and reasoning about relevance is also crucial for quantity implicatures. The finding in Experiment 1A that scalars and relevance inferences are correlated in acquisition hints

²⁵ These examples are inspired by those in Cummings (2005: 16). However, my interpretation differs from hers, in implicating the maxim of manner and the *varying* interpretations that may arise.

at this connection – quantity implicatures cannot be derived in an adult-like way without also having acquired the ability to track the QUD and generative *relevant* alternatives. Then, I suggested in Chapter 4 that relevance to the QUD is equally important for manner implicatures, and, further, above, that manner may often serve to intensify or elaborate a quantity implicature. In other words, a comprehensive model of implicature understanding has to account for both distinct instances of quantity, manner, or relevance inferences, and also inferences that involve reasoning about all three. This, then, has implications for acquisition, as the question becomes: how do children learn to identify flouting or exploitation of a combination of maxims, and weight expectations of cooperativeness appropriately?

Thirdly, the Gricean understanding that communication – and implicature derivation – inherently involves reasoning about the interlocutor’s intentions and beliefs may have to be revised. Although to be taken with some caution, the findings of Chapter 5 add to other studies with adults and with atypical populations which suggest that some pragmatic reasoning is available without engaging in full reasoning about the others’ beliefs, or that both perspectives are considered (e.g. Andrés-Roqueta & Katsos, 2017; Heller et al., 2016; Kissine et al., 2015). If this is the case, then a new and more diverse model would be required, in which heuristics can replace the Competence Assumption and Epistemic Step, in some circumstances. It may be that *more*, or different, distinctions within pragmatic competences are required to model these findings: for example, between quantity implicatures where reasoning about others’ beliefs is vital, and those where heuristics suffice. A constraint-based model has the potential to capture these many dimensions to utterance interpretation, and the changing weighting of different sources of information over development (e.g. Degen & Tanenhaus, 2014; Heller et al., 2016). Again, this is an area where more work is needed, especially with *social* perspective-taking and pragmatic inferences.

6.4 Implications for the study of developmental pragmatics

Research on children’s pragmatic development has burgeoned within the experimental pragmatics programme over the last two decades. When it comes to implicatures, the focus has been predominantly on scalar implicatures (e.g., Katsos, 2014; Papafragou & Skordos, 2016). This thesis adds to the small but growing

collection of studies on relevance and ad hoc implicatures, and breaks new ground for investigation of manner implicatures. This broader focus is needed not only to give us a more comprehensive picture of children's development of implicatures, but also because the differences and similarities between types of implicature may help us to better understand which cognitive skills children have to learn – or, to put it another way, what the challenges for acquisition are. This has already proved a fruitful avenue in the case of ad hoc and scalar quantity implicatures, where studies comparing them have indicated that acquiring quantifiers or generating quantifier alternatives is a particular challenge in the case of <all, some>, as children find scalars harder than ad hocs, all else being equal (Barner et al., 2011; Horowitz & Frank, 2015; Horowitz, Schneider, & Frank, under review; Katsos & Bishop, 2011; Katsos & Smith, 2010). Examining a range of scalar, ad hoc, relevance and manner implicatures has the potential to highlight the role of linguistic, social or world knowledge and of components of inferencing, and how children learn to integrate these.

Similarly, many studies to date have been driven, at least in part, by a 'how young can we go' research question, and have thus used supportive, child-friendly tasks to demonstrate that children younger than previously thought are able to derive implicatures. In such tasks, the speaker is co-operative (or there is no reason to think otherwise, except for the confounding critical under-informative condition in acceptability judgement tasks), and there is little or no contextual information to suggest that an implicature is not relevant. Experiment 1A here was in this spirit. However, the findings presented in Chapter 5 suggest that children's adult-like competence in these optimal settings do not represent the whole story; instead, preschoolers are still very much developing in their ability to derive implicatures appropriately, when more information from the context must be taken into account. Developmental pragmatics therefore needs a variety of tasks and experimental contexts to provide a full picture of children's development, which is important in forming realistic expectations, for example in clinical settings.

The experiments in this thesis were all versions of a picture-matching task (including the covered-box and director tasks). I suggested that this had advantages over Acceptability or Truth Value Judgement Tasks, particularly in avoiding metalinguistic reasoning and in testing implicature comprehension, rather than just

sensitivity to speaker cooperativeness. However, reflecting these methods now, I can see that picture-matching tasks also have limitations. In particular, as I suggested in Chapter 2, the visual presence of alternatives not only supports an implicature inference, but allows for a slightly different contrastive inference, in which the alternative is excluded with no reference to informativeness. For instance, reasoning could be along the following lines: ‘It’s not the picture with *all* because the speaker did not say ‘all’, so it is the other one’, or ‘the speaker did not say ‘strawberry [and orange]’, so it’s the other one’. Another potential problem with picture-matching is the introduction of a new challenge: inhibition of the alternative when it is so salient in context (and arguably more salient, in the case of quantity). This means that there is still a need for developing methods that are capable of testing a variety of inferences in young children. One option, used by Miller, Schmitt, Chang and Munn (2005) but not widely adopted, is an act-out task, in which children have to follow instructions, such as ‘make some faces happy’. Devising and coding such scenarios for relevance and manner would be more challenging, but worthwhile. Another avenue, as in Experiment 4, is to borrow existing methodological paradigms employed for other pragmatic or communicative phenomena in development. For instance, interactive tasks where the experimenter and child are jointly engaged in an activity used to investigate indirect requests, in typically and atypically developing children, might be promising for implicatures as well (e.g. in Kissine et al., 2015; Schulze & Tomasello, 2015). Finally, the implication of Experiment 1B is that, until we know more about what is driving the association of structural language knowledge and implicature abilities, structural language should be carefully controlled – in terms of using vocabulary and constructions that are known to be well within the grasp of the children tested – or children’s language skills should be measured separately.

6.5 Implications for educational policy

Looking slightly further afield, the kind of research presented in this thesis may eventually have implications for educational policy and practice. In UK, the National Curriculum for Key Stages 1 and 2 (primary school, aged 4–10 years), introduced a requirement of inference-making as part of literacy. For Key Stage 1, the Curriculum states that children should be able to ‘understand both the books they can already

read accurately and fluently and those they listen to by making inferences on the basis of what is being said and done' (Department for Education, 2013: 11). Obviously, this pertains to reading skills in particular. However, while reading and listening to books demands many additional skills, it clearly also has much in common with spoken language comprehension, including the typically pragmatic skills of implicature, presupposition, irony, and so on, as well as higher level critical inferences and analysis. Indeed, pragmatic theory has been applied widely to texts, reading and writing (e.g. B. Clark, 2014; and, for children's literature, Meibauer, 2017).

As Williams (2014) points out, in the Curriculum there is no break-down of different kinds of inference, despite Kispal's (2008) earlier Department for Children, Schools and Families report on reading, which outlines several different types of inference. The report draws on a large literature on reading (e.g. Kamil, Pearson, Moje, & Afflerbach, 2011), and in particular Graesser, Singer and Trabasso's (1994) typology, but there is no indication that these are based on or influenced by theories within linguistics, such as a Gricean approach to pragmatics. The examples of inference put forward in the report, however, include not only those made consciously on reflection about a text, but also many inferences made online in comprehension, of the sort which are typically studied in psycholinguistic and pragmatic approaches to language. Examples include anaphora resolution, cohesion inferences, and relevance inferences (to use standard pragmatic terminology). Furthermore, some of the factors identified as being involved in reading inferences are strikingly similar to those observed as playing a role in developmental pragmatics studies: word-level knowledge (including vocabulary), shared background knowledge of writer and reader (i.e. common ground), and accessibility of background knowledge. Kispal cites Bowyer-Crane and Snowling, who found that the children in their study who were less adept at reading were 'in possession of the knowledge... but are unaware of the need to draw on that knowledge' (2005: 199); this could be a parallel of children's struggle to generate relevant alternatives in quantity implicatures (Skordos & Papafragou, 2016), or of the relationship between relevance inference skills and general knowledge (Anagnostopoulou et al., 2017).

Kispal then turns to possible practices to improve inferencing skills, although her review is focussed on Key Stages 2 and 3 (ages 8–14): the main concepts of overt

'why?' questioning are unlikely to be able to be extended to younger ages, because of ongoing metalinguistic and production development (Lieven, 2006). Studies and paradigms from developmental pragmatics and psychology could have a contribution to make here. For instance, looking at children's production, Matthews and colleagues (2012) find that feedback involving questions that label alternatives helps children to produce informative requests more than general feedback asking for clarification. Likewise, providing alternatives in context and highlighting their relevance facilitates children's derivation of quantity implicatures, although it is not yet known whether doing so can improve children's derivation skills longer-term.

In other words, connecting research on reading inferences with research on developmental pragmatics, and using these together as a basis for intervention studies could then better inform recommendations for policy and practice in schools. In particular, teachers and educators might be helped by clearer typologies of inferences, detailing which inferences are likely to develop as part of general communication and oracy skills, and which may be more specific to reading in a reflective context. In addition, pragmatics could provide theoretical frameworks for approaching conversational and reading inferences, for understanding their commonalities and differences, or for making predictions about strategies that might aid children's inferencing skills. However, it is an open question whether pragmatic skills like implicature require or can be improved by intervention, or whether only higher-level inferences are amenable to such strategies. In the meantime, developmental pragmatics research, including the present studies, can at least contribute to realistic expectations for children's achievement, including the fact that skills such as implicature understanding in more complex situations are still very much developing at the point children enter school in UK. Conversely, the rich literature on inferences in reading may have insights to offer for developmental pragmatics, too.

6.6 Research outlook

The studies presented in this thesis point to questions for future research. I highlight some of them here, based on the three strands running through the thesis – the relationship of implicatures in development (Chapters 2 and 4), the role of structural language knowledge (Chapters 3 and 4), and of Theory of Mind (Chapters 3

and 5).

6.6.1 Implicatures

As outlined above, many questions remain about manner implicatures in general, but there is also the issue of how children learn to understand them. Here, manner implicatures were investigated separately from quantity and relevance, due to their unique challenges as well as lack of established experimental paradigms and items to test them. In future, examining their relationship with quantity and relevance could shed light on the common challenges in acquisition, or on the differences. To what extent is structural language knowledge, including knowledge of conventional usage, required on top of quantity-like inferences? What cues, like degree of markedness or prosody, help or challenge children in deriving manner inferences? What kind of sources of information in the context, especially speaker-specific characteristics, affect manner inferences, and how and when do children learn to integrate these?

The results of Experiment 1A hinted at a ‘floor’ to some implicature skills, which corroborate the findings of other studies that children under 3 years are not able to derive simple ad hoc implicatures (Stiller et al., 2015; Yoon et al., 2015). Why is there this gap between early WLE inferences and similar word learning inferences, or non-verbal inferences? Is it the structural language burden, the nature of the inference, or a by-product of task challenges? More studies that control and manipulate the complexity of the language, the salience of the stimuli, and the demands of the experimental task could begin to answer these questions.

6.6.2 Structural language

There is mounting evidence for an association between structural language knowledge – vocabulary and grammar skills – and implicature understanding, but which direction does any causality in this relationship go, and why? And, is any contribution of structural language to pragmatic inferencing – at the ‘local’ or the ‘global’ level – to do with the processing of any given utterance, or with cumulative linguistic experience? Establishing, for instance, to what extent familiarity with vocabulary and constructions used in the task aids implicature understanding, or whether an association persists with non-verbal inferences, could start to address these questions. Furthermore, does it make a difference whether this knowledge is

receptive or expressive? Remember that for word learning, *productive* knowledge of alternatives is a predictor of an exclusion inference (Grassmann et al., 2015).

One avenue for addressing these issues could be longitudinal studies, which have been scarcely employed in the recent experimental pragmatics programme but have long enriched understanding of other aspects of children's linguistic and cognitive development. For instance, Brooks and Meltzoff (2015) are able to find a developmental association between different cognitive skills. They find that gaze-following at 10.5 months predicts mental-state word knowledge at 2.5 years, which in turn predicts Theory of Mind skills at 4.5 years, even controlling for general language, maternal education and nonsocial attention skills. Such studies offer the potential to join the dots between early and late pragmatic abilities, and related or prerequisite skills, or reveal discontinuities.

Another avenue could be cross-linguistic studies. Katsos and colleagues observed striking discrepancies as to when children were able to reject an under-informative utterance with 'some' in their study of the acquisition of quantifiers across 31 languages (Katsos et al., 2016). They note that linguistic factors such as negative concord and partitive markers seem to have some effect in children's performance overall (not just with under-informative 'some'), and that it is conceivable that other features like agreement or the semantic field of the quantifiers could also play a role in the variation. Cross-linguistic studies not only of very specific inferences, like scalars with 'some', but also ad hoc quantity implicatures, for example, might be able to tease apart contributions of language knowledge from general cognitive development.

A related issue is the contribution of formal education. Unfortunately, in Experiments 1A and 1B, there were not enough age-matched children at nursery and school to perform a comparison. Formal education and the literacy that accompanies it can have a significant effect on the rate of children's word learning (Bloom, 2002). Does this change in input also have an effect on children's pragmatic development – either indirectly through structural language knowledge or directly? In Katsos and colleagues' (2016) cross-linguistic study, whether children were in formal schooling turned out to be a predictive factor for performance with quantifiers (both semantic and pragmatic skills) in exploratory analyses. Similarly, the effect of SES in Experiment 1B was associated with structural language knowledge, but there has been little

previous work on how input differences associated with SES affect pragmatic skills such as implicature understanding. Knowing more about these associations could have important implications for testing or intervention.

6.6.3 Theory of Mind

In Chapter 5, the integration of visual perspective-taking and pragmatic inferencing was shown to be challenging for children, and further studies were suggested to investigate the generality of this effect with social perspective-taking, where common and privileged ground is established through discourse and interaction. Experiments 3 and 4 added to a small number of studies that consider the effect of speaker epistemic state, informativeness, deception and the QUD for children's understanding (e.g. Pogue et al., 2017; Scrafton, 2009). However, other aspects of context which involve reasoning about the speaker have been examined in implicature studies with adults, and in word learning studies with children: for example, cooperativeness (Dulcinati & Pouscoulous, 2017; Pouscoulous & Dulcinati, 2017), reliability (Grodner & Sedivy, 2011; Sobel et al., 2012), expertise (Sobel & Corriveau, 2010), honesty (Feeney & Bonnefon, 2013), face-threatening contexts (Bonnefon, Feeney, & Villejoubert, 2009), and speaker-specific usage patterns (Yildirim, Degen, Tanenhaus, & Jaeger, 2016). More studies like this with both age groups may lead us to a more nuanced account of pragmatic development, in which children and adults differ in their use or weighting of different cues in communication, and children become adult-like over many years. For example, is there a difference between speaker-specific traits that are particular to the context (for instance, deception or politeness) and those that are persistent (perhaps informativeness, or features connected to L2 language use)? To what extent is Theory of Mind required for tracking these different characteristics, reasoning about them, and integrating them into pragmatic inferences? And is this the same for WLE inferences and implicatures? The role of Theory of Mind in children's pragmatic development – as well as in inferencing across the life-span – is part of a current live debate. It connects implicature and word learning strategies to other pragmatic and cognitive skills, and invites further research on children's development.

7 References

- Akhtar, N. (2002). Relevance and early word learning. *Journal of Child Language*, 29(3), 677–686.
- Akhtar, N., Carpenter, M., & Tomasello, M. (1996). The role of discourse novelty in early word learning. *Child Development*, 635–645.
- Ambridge, B., & Lieven, E. V. (2011). *Child Language Acquisition: Contrasting Theoretical Approaches*. Cambridge: Cambridge University Press.
- Anagnostopoulou, N., Abbot-Smith, K., Schulze, C., & Matthews, Danielle. (2017). *Relevance Inferencing in 3-year-olds: Real World Knowledge Matters*. Poster presented at the International Congress for the Study of Child Language, Lyon, France.
- Andrés-Roqueta, C., & Katsos, N. (2017). The contribution of grammar, vocabulary and theory of mind in pragmatic language competence in children with Autistic Spectrum Disorders. *Frontiers in Psychology*, (8), 996.
- Antoniou, K. (2015). *The Effects of Childhood Bilingualism and Bilectalism on Executive Control and Implicature Understanding* (PhD dissertation). University of Cambridge.
- Antoniou, K., Cummins, C., & Katsos, N. (2016). Why only some adults reject uninformative utterances. *Journal of Pragmatics*, 99, 78–95.
- Antoniou, K., Grohmann, K., Kambanaros, M., & Katsos, N. (2013). Does multilingualism confer an advantage for pragmatic abilities? In *Online Proceedings Supplement of the 37th Annual Boston University Conference on Language Development*.
- Antoniou, K., & Katsos, N. (2017). The effect of childhood multilingualism and bilectalism on implicature understanding. *Applied Psycholinguistics*, 1–47.
- Antoniou, K., Veenstra, A., Katsos, N., & Kissine, M. (2016, May). *Bi-dialectal pragmatics: An investigation into Dutch-Flemish bi-dialectal children's implicature understanding skills*. Paper presented at the 37th Annual Meeting of the Department of Linguistics, School of Philology, Aristotle University of Thessaloniki, Greece.

- Apperly, I. (2010). *Mindreaders: the cognitive basis of 'theory of mind'*. Hove: Psychology Press / Taylor & Francis Group.
- Audacity Team. (1999). Audacity(R) (Version 2.1.2). The name Audacity(R) is a registered trademark of Dominic Mazzoni.
- Bach, K. (2007). Regressions in pragmatics (and semantics). In N. Burton-Roberts (Ed.), *Pragmatics* (pp. 24–44). Basingstoke: Palgrave MacMillan.
- Bak, T. H. (2016). Cooking pasta in La Paz. *Linguistic Approaches to Bilingualism*, 6(5), 699–717.
- Bale, A., & Barner, D. (2013). Grammatical alternatives and pragmatic development. In A. Fălăuș (Ed.), *Alternatives in Semantics*. (pp. 238–66). Basingstoke: Palgrave Macmillan.
- Bannard, C., Rosner, M., & Matthews, D. (2017). What's Worth Talking About? Information Theory Reveals How Children Balance Informativeness and Ease of Production. *Psychological Science*, 28(7), 954–966.
- Barner, D., Brooks, N., & Bale, A. (2011). Accessing the unsaid: The role of scalar alternatives in children's pragmatic inference. *Cognition*, 118(1), 87–96.
- Barner, D., & Snedeker, J. (2008). Compositionality and Statistics in Adjective Acquisition: 4-Year-Olds Interpret Tall and Short Based on the Size Distributions of Novel Noun Referents. *Child Development*, 79(3), 594–608.
- Baron-Cohen, S., Leslie, A. M., & Frith, U. (1985). Does the autistic child have a 'theory of mind'? *Cognition*, 21(1), 37–46.
- Baron-Cohen, S., Wheelwright, S., Skinner, R., Martin, J., & Clubley, E. (2001). The autism-spectrum quotient (AQ): Evidence from Asperger syndrome/high-functioning autism, males and females, scientists and mathematicians. *Journal of Autism and Developmental Disorders*, 31(1), 5–17.
- Barr, D. J., Levy, R., Scheepers, C., & Tily, H. J. (2013). Random effects structure for confirmatory hypothesis testing: Keep it maximal. *Journal of Memory and Language*, 68(3), 255–278.
- Barwise, J., & Perry, J. (1983). *Situations and attitudes*. Cambridge, MA: MIT Press.

- Bates, D., Mächler, M., Bolker, B., & Walker, S. (2015). Fitting Linear Mixed-Effects Models Using lme4. *Journal of Statistical Software*, 67(1), 1–48.
- Behne, T., Carpenter, M., Call, J., & Tomasello, M. (2005). Unwilling versus unable: infants' understanding of intentional action. *Developmental Psychology*, 41(2), 328–337.
- Benz, A., & Jasinskaja, K. (2017). Questions Under Discussion: From Sentence to Discourse. *Discourse Processes*, 54(3), 177–186.
- Bergen, L., Goodman, N. D., & Levy, R. (2012). That's what she (could have) said: How alternative utterances affect language use. In *Proceedings of the 34th Annual Conference of the Cognitive Science Society* (pp. 120–125). Austin, TX: Cognitive Science Society.
- Bergen, L., & Grodner, D. J. (2012). Speaker knowledge influences the comprehension of pragmatic inferences. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 38(5), 1450–1460.
- Bergen, L., Levy, R., & Goodman, N. D. (2016). Pragmatic reasoning through semantic inference. *Semantics and Pragmatics*, 9(20).
- Bernicot, J., & Legros, S. (1987). Direct and indirect directives: What do young children understand? *Journal of Experimental Child Psychology*, 43(3), 346–358.
- Bialystok, E., Barac, R., Blaye, A., & Poulin-Dubois, D. (2010). Word mapping and executive functioning in young monolingual and bilingual children. *Journal of Cognition and Development*, 11(4), 485–508.
- Bialystok, E., & Feng, X. (2011). Language proficiency and its implications for monolingual and bilingual children. In A. Y. Durgunoglu & C. Goldenberg (Eds.), *Language and Literacy Development in Bilingual Settings* (pp. 121–138). New York: The Guildford Press.
- Bialystok, E., Luk, G., Peets, K. F., & Yang, S. (2010). Receptive vocabulary differences in monolingual and bilingual children. *Bilingualism: Language and Cognition*, 13(4), 525–531.
- Bill, C., Romoli, J., Schwarz, F., & Crain, S. (2014). Scalar Implicatures Versus Presuppositions: The View from Acquisition. *Topoi*, 1–15.

- Bion, R. A. H., Borovsky, A., & Fernald, A. (2013). Fast mapping, slow learning: Disambiguation of novel word-object mappings in relation to vocabulary learning at 18, 24, and 30 months. *Cognition*, *126*(1), 39–53.
- Bishop, D. V. M. (2003). *TROG-2 – Test for reception of grammar-2* (2nd ed.). London: Harcourt.
- Bloom, P. (2002). *How children learn the meanings of words*. Cambridge, MA: MIT press.
- Bonnefon, J.-F., Feeney, A., & Villejoubert, G. (2009). When some is actually all: Scalar inferences in face-threatening contexts. *Cognition*, *112*(2), 249–258.
- Bott, L., & Chemla, E. (2016). Shared and distinct mechanisms in deriving linguistic enrichment. *Journal of Memory and Language*, *91*, 117–140.
- Bowyer-Crane, C., & Snowling, M. J. (2005). Assessing children’s inference generation: What do tests of reading comprehension measure? *British Journal of Educational Psychology*, *75*(2), 189–201.
- Boyce, W., Torsheim, T., Currie, C., & Zambon, A. (2006). The Family Affluence Scale as a measure of national wealth: validation of an adolescent self-report measure. *Social Indicators Research*, *78*(3), 473–487.
- Braxmeier, H., & Steinberger, S. (2017). Pixabay. Retrieved from www.pixabay.com
- Breheny, R. (2006). Communication and Folk Psychology. *Mind & Language*, *21*(1), 74–107.
- Breheny, R., Ferguson, H. J., & Katsos, N. (2013). Taking the epistemic step: Toward a model of on-line access to conversational implicatures. *Cognition*, *126*(3), 423–440.
- Breheny, R., Katsos, N., & Williams, J. (2006). Are generalised scalar implicatures generated by default? An on-line investigation into the role of context in generating pragmatic inferences. *Cognition*, *100*(3), 434–463.
- Brennan, S. E., & Clark, H. H. (1996). Conceptual pacts and lexical choice in conversation. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *22*(6), 1482–1493.
- Brodeur, M. B., Dionne-Dostie, E., Montreuil, T., & Lepage, M. (2010). The Bank of Standardized Stimuli (BOSS), a New Set of 480 Normative Photos of Objects to Be Used as Visual Stimuli in Cognitive Research. *PLoS ONE*, *5*(5), e10773.

- Brojde, C., & Colunga, E. (2011). Bilingual and Monolingual Children Attend to Object Properties Differently in a Word Learning Task. In *Proceedings of the Annual Boston University Conference on Language Development* (pp. 73–84). Somerville, MA: Cascadilla Press.
- Brooks, R., & Meltzoff, A. N. (2015). Connecting the dots from infancy to childhood: A longitudinal study connecting gaze following, language, and explicit theory of mind. *Journal of Experimental Child Psychology*, 130(0), 67–78.
- Brosseau-Liard, P. E., & Hall, D. G. (2011). Explaining the Disambiguation Effect in Lexical Development: New Evidence Favoring Lexical Constraints over Social-Pragmatic Knowledge. In *Proceedings of the 35th Annual Boston University Conference on Language Development* (Vol. 35, pp. 85–96). Somerville, MA: Cascadilla Press.
- Bucciarelli, M., Colle, L., & Bara, B. G. (2003). How children comprehend speech acts and communicative gestures. *Journal of Pragmatics*, 35(2), 207–241.
- Byers-Heinlein, K., & Werker, J. F. (2013). Lexicon structure and the disambiguation of novel words: Evidence from bilingual infants. *Cognition*, 128(3), 407–416.
- Carpendale, J. I. M., & Lewis, C. (2015). The Development of Social Understanding. In *Handbook of Child Psychology and Developmental Science* (Vol. 2(10), pp. 1–44). New York: Wiley.
- Chemla, E., & Singh, R. (2014a). Remarks on the Experimental Turn in the Study of Scalar Implicature, Part I. *Language and Linguistics Compass*, 8(9), 373–386.
- Chemla, E., & Singh, R. (2014b). Remarks on the Experimental Turn in the Study of Scalar Implicature, Part II. *Language and Linguistics Compass*, 8(9), 387–399.
- Chierchia, G. (2004). Scalar implicatures, polarity phenomena, and the syntax/pragmatics interface. In A. Belletii (Ed.), *Structures and beyond* (Vol. 3, pp. 39–103). Oxford: Oxford University Press.
- 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 *Proceedings of the 25th Boston University Conference on Language Development* (pp. 157–168). Somerville, MA: Cascadilla Press.

- Chierchia, G., Fox, D., & Spector, B. (2008). The Grammatical View of Scalar Implicatures and the Relationship between Semantics and Pragmatics. In P. Portner, C. Maeinborn, & K. von Stechow (Eds.), *Semantics: An International Handbook of Natural Language Meaning* (Vol 3, pp. 2297–2331). New York, NY: Mouton de Gruyter.
- Clark, B. (2014). Pragmatics and inference. In P. Stockwell & S. Whiteley (Eds.), *Cambridge Handbook of Stylistics* (pp. 300–314). Cambridge: Cambridge University Press.
- Clark, E. V. (1988). On the logic of contrast. *Journal of Child Language*, 15(2), 317–335.
- Clark, E. V. (1990). On the pragmatics of contrast. *Journal of Child Language*, 17(2), 417–431.
- Clark, E. V. (2002). Making use of pragmatic inferences in the acquisition of meaning. In D. Beaver (Ed.), *The construction of meaning* (pp. 45–58). Stanford, CA.: CSLI Publications.
- Clark, E. V. (2004). Pragmatics and Language Acquisition. In L. Horn & G. Ward (Eds.), *Handbook of Pragmatics* (pp. 562–577). Oxford: Oxford University Press.
- Clark, E. V. (2007). Conventionality and contrast in language acquisition. In M. A. Sabbagh & C. Kalish (Eds.), *Right thinking: The development of conventionality* (pp. 11–23). San Francisco: Jossey-Bass.
- Clark, E. V. (2009). *First language acquisition*. Cambridge: Cambridge University Press.
- Clark, E. V., & Clark, H. H. (1979). When Nouns Surface as Verbs. *Language*, 55(4), 767–811.
- Clark, E. V., & Kurumada, C. (2013). ‘Be brief’: from necessity to choice. In L. Goldstein (Ed.), *Brevity* (pp. 233–248). Oxford: Oxford University Press.
- Clark, H. H. (1996). *Using language*. Cambridge: Cambridge University Press.
- Csibra, G., & Gergely, G. (2009). Natural pedagogy. *Trends in Cognitive Sciences*, 13(4), 148–153.
- Cummings, L. (2005). *Pragmatics: A multidisciplinary perspective*. Edinburgh: Edinburgh University Press.

- Cummings, L. (2015). Theory of mind in utterance interpretation: the case from clinical pragmatics. *Frontiers in Psychology*, *6*, 1286.
<http://doi.org/10.3389/fpsyg.2015.01286>
- Cummins, C. (2017). Contextual Causes of Implicature Failure. *Discourse Processes*, *54*(3), 207–218.
- Cutting, A. L., & Dunn, J. (1999). Theory of Mind, Emotion Understanding, Language, and Family Background: Individual Differences and Interrelations. *Child Development*, *70*(4), 853–865.
- Davidson, D., Jergovic, D., Imami, Z., & Theodos, V. (1997). Monolingual and bilingual children's use of the mutual exclusivity constraint. *Journal of Child Language*, *24*(1), 3–24.
- Davidson, D., & Tell, D. (2005). Monolingual and bilingual children's use of mutual exclusivity in the naming of whole objects. *Journal of Experimental Child Psychology*, *92*(1), 25–45.
- Davies, C., Andrés-Roqueta, C., & Norbury, C. F. (2016). Referring expressions and structural language abilities in children with specific language impairment: A pragmatic tolerance account. *Journal of Experimental Child Psychology*, *144*, 98–113.
- Davies, C., & Katsos, N. (2010). Over-informative children: Production/comprehension asymmetry or tolerance to pragmatic violations? *Lingua*, *120*(8), 1956–1972.
- Davies, C., & Kreysa, H. (2017, July). *Is children's referential informativeness driven by their visual scanning behaviour?* Paper presented at the 14th International Congress for the Study of Child Language, Lyon, France.
- Davis, S. (Ed.). (1991). *Pragmatics: A reader*. Oxford: Oxford University Press.
- De Cat, C. (2015). The cognitive underpinnings of referential abilities. In L. Serratrice & S. E. M. Allen (Eds.) *The Acquisition of Reference* (pp. 263–283). Amsterdam: John Benjamins.
- De Houwer, A., Bornstein, M. H., & Putnick, D. L. (2014). A bilingual–monolingual comparison of young children's vocabulary size: Evidence from comprehension and production. *Applied Psycholinguistics*, *35*(6), 1189–1211.

- 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.
- de Villiers, P. A., de Villiers, J. G., Coles-White, D., & Carpenter, L. (2009). Acquisition of Relevance Implicatures in Typically-Developing Children and Children with Autism. *Proceedings of the Annual Boston University Conference on Language Development*, (pp. 121–132). Somerville, MA: Cascadilla Press.
- Defeyter, M. A., Hearing, J., & German, T. C. (2009). A developmental dissociation between category and function judgments about novel artifacts. *Cognition*, *110*(2), 260–264.
- Degen, J. (2015). Investigating the distribution of some (but not all) implicatures using corpora and web-based methods. *Semantics and Pragmatics*, *8*(11), 1–55.
- Degen, J., & Tanenhaus, M. K. (2014). Processing Scalar Implicature: A Constraint-Based Approach. *Cognitive Science*, 667–710.
- Department for Education. (2013). *English Programmes of Study: Key Stages 1 and 2 National Curriculum in England*. Department for Education.
- Department for Education statistics. (2016). EAL Learners in Schools. Retrieved 3 June 2017, from <https://naldic.org.uk/the-eal-learner/research-and-statistics/>
- Diamond, A. (2006). The early development of executive functions. In E. Bialystok & F. I. M. Craik (Eds.), *Lifespan cognition: Mechanisms of change*. (pp. 70–95). Oxford: Oxford University Press.
- Diesendruck, G. (2005). The principles of conventionality and contrast in word learning: An empirical examination. *Developmental Psychology*, *41*(3), 451–463.
- Diesendruck, G. (2011). The scope and origins of children’s assumption of conventionality. In M. Siegal & L. Surian (Eds.), *Access to Language and Cognitive Development* (pp. 116–134). Oxford: Oxford University Press.
- Diesendruck, G., Carmel, N., & Markson, L. (2010). Children’s sensitivity to the conventionality of sources. *Child Development*, *81*(2), 652–668.
- Diesendruck, G., & Markson, L. (2001). Children’s avoidance of lexical overlap: a pragmatic account. *Developmental Psychology*, *37*(5), 630–641.

- Diesendruck, G., Markson, L., & Bloom, P. (2003). Children's Reliance on Creator's Intent in Extending Names for Artifacts. *Psychological Science, 14*(2), 164–168.
- Dulcinati, G., & Pouscoulous, N. (2016). Cooperation and exhaustification. In F. Salfner & U. Sauerland (Eds.), *Pre-proceedings of Trends in Experimental Pragmatics* (pp. 39–45). Berlin.
- Dulcinati, G., & Pouscoulous, N. (2017, June). *Scalar implicatures in non-cooperative contexts*. Paper presented at XPrag 2017, Cologne, Germany.
- Dunn, L., Dunn, L., Sewell, J., Styles, B., Brzyska, B., Shamsan, Y., & Burge, B. (2009). *The British Picture Vocabulary Scale* (3rd ed.). London: GL Assessment.
- Eiteljörge, S., Pouscoulous, N., & Lieven, E. (2016). Some pieces are missing: Scalar implicatures in children. In F. Salfner & U. Sauerland (Eds.), *Pre-proceedings of Trends in Experimental Pragmatics* (pp. 46–52). Berlin.
- Elrod, M. M. (1987). Children's understanding of indirect requests. *The Journal of Genetic Psychology, 148*(1), 63–70.
- 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.
- Feeney, A., & Bonnefon, J.-F. (2013). Politeness and honesty contribute additively to the interpretation of scalar expressions. *Journal of Language and Social Psychology, 32*(2), 181–190.
- Ferguson, H. J., Apperly, I., & Cane, J. E. (2017). Eye tracking reveals the cost of switching between self and other perspectives in a visual perspective-taking task. *The Quarterly Journal of Experimental Psychology, 70*(8), 1646–1660.
- Field, A., Miles, J., & Field., Z. (2012). *Discovering statistics using R*. London: SAGE.
- Filippova, E. (2014). Irony production and understanding. In *Pragmatic Development in First Language Acquisition* (pp. 261–278). Amsterdam: John Benjamins.
- Filippova, E., & Astington, J. W. (2008). Further development in social reasoning revealed in discourse irony understanding. *Child Development, 79*(1), 126–138.
- Flavell, J. H. (1977). The development of knowledge about visual perception. *Nebraska Symposium on Motivation, 25*, 43–76.

- Flavell, J. H., Beilin, H., & Pufall, P. (1992). Perspectives on perspective-taking. In H. Beilin & P. B. Pufall (Eds.) *Piaget's Theory: Prospects and Possibilities*, (pp. 107–139). Hove and London: Lawrence Erlbaum Associates.
- Foppolo, F. (2015, February). *The bilingual advantage revealed in monolinguals that are early exposed to L2*. Paper presented at the 41st Incontro di Grammatica Generativa, Università per Stranieri di Perugia, Italy.
- 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.
- Frank, M. C. (2014). Learning words through probabilistic inferences about speakers' communicative intentions. In I. Arnon, M. Casillas, C. Kurumada, & B. Estigarribia (Eds.), *Language in Interaction: Studies in honor of Eve V. Clark* (pp. 207–230). Amsterdam: John Benjamins.
- Frank, M. C., & Goodman, N. D. (2012). Predicting pragmatic reasoning in language games. *Science*, 336(6084), 998–998.
- Frank, M. C., & Goodman, N. D. (2014). Inferring word meanings by assuming that speakers are informative. *Cognitive Psychology*, 75, 80–96.
- Franke, M. (2009). *Signal to act: Game theory in pragmatics*. (PhD thesis). University of Amsterdam.
- Franke, M., & Degen, J. (2016). Reasoning in reference games: Individual-vs. population-level probabilistic modeling. *PloS One*, 11(5), e0154854.
- Franke, M., & Jäger, G. (2016). Probabilistic pragmatics, or why Bayes' rule is probably important for pragmatics. *Zeitschrift Für Sprachwissenschaft*, 35(1), 3–44.
- Geurts, B. (2010). *Quantity implicatures*. Cambridge: Cambridge University Press.
- Geurts, B., & Rubio-Fernández, P. (2015). Pragmatics and processing. *Ratio*, 28(4), 446–469.
- Gillis, R., & Nilsen, E. S. (2014). Cognitive flexibility supports preschoolers' detection of communicative ambiguity. *First Language*, 34(1), 58–71.
- Goetz, P. J. (2003). The effects of bilingualism on theory of mind development. *Bilingualism Language and Cognition*, 6(1), 1–15.

- Goldshtein, M., Brown-Schmidt, S., & Terkourafi, M. (2017, February). *Logical response to 'some' scalar implicatures and ASD-like traits: The limitations of the binary task*. Poster presented at the CSLI Workshop: Bridging computational and psycholinguistic approaches to the study of meaning, Stanford, CA.
- 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.
- Gollek, C., & Doherty, M. J. (2016). Metacognitive developments in word learning: Mutual exclusivity and theory of mind. *Journal of Experimental Child Psychology*, 148, 51–69.
- Goodman, N. D., & Frank, M. C. (2016). Pragmatic language interpretation as probabilistic inference. *Trends in Cognitive Sciences*, 20(11), 818–829.
- Goodman, N. D., & Stuhlmüller, A. (2013). Knowledge and implicature: Modeling language understanding as social cognition. *Topics in Cognitive Science*, 5(1), 173–184.
- Gotzner, N., Barner, D., & Crain, S. (2015, June). *What's the alternative? How children compute implicatures with different scales*. Paper presented at XPrag 2015, Chicago.
- Graesser, A., Singer, M., & Trabasso, T. (1994). Constructing inferences during narrative text representation. *Psychological Review*, 101, 371–395.
- Graham, S. A., Juan, V. S., & Khu, M. (2016). Words are not enough: how preschoolers' integration of perspective and emotion informs their referential understanding. *Journal of Child Language*, 1–27.
- Graham, S. A., Poulin-Dubois, D., & Baker, R. K. (1998). Infants' disambiguation of novel object words. *First Language*, 18(53), 149–164.
- Graham, S. A., Stock, H., & Henderson, A. M. (2006). Nineteen-month-olds' understanding of the conventionality of object labels versus desires. *Infancy*, 9(3), 341–350.
- Grassmann, S., Schulze, C., & Tomasello, M. (2015). Children's level of word knowledge predicts their exclusion of familiar objects as referents of novel words. *Developmental Psychology*, 1200. <http://doi.org/10.3389/fpsyg.2015.01200>
- Grassmann, S., Stracke, M., & Tomasello, M. (2009). Two-year-olds exclude novel objects as potential referents of novel words based on pragmatics. *Cognition*, 112(3), 488–493.

- Grassmann, S., & Tomasello, M. (2010). Young children follow pointing over words in interpreting acts of reference. *Developmental Science*, *13*(1), 252–263.
- Green, D. W. (2011). Language control in different contexts: the behavioral ecology of bilingual speakers. *Frontiers in Psychology*, *2*.
- Grice, H. P. (1957). Meaning. *The Philosophical Review*, 377–388.
- Grice, H. P. (1975). Logic and conversation. In R. Stainton (Ed.), *Perspectives in the Philosophy of Language* (pp. 41–58). Peterborough Ontario, Canada: Broadview Press.
- Grice, H. P. (1989). *Studies in the Way of Words*. Cambridge, MA: Harvard University Press.
- Grigoroglou, M., & Papafragou, A. (2016). Are children flexible speakers? Effects of typicality and listener needs in children’s event descriptions. In *Proceedings from the 38th Annual Meeting of the Cognitive Science Society* (pp. 782–787). Austin, TX: Cognitive Science Society.
- Groba, A., De Houwer, A., Mehnert, J., Rossi, S., & Obrig, H. (2017). Bilingual and monolingual children process pragmatic cues differently when learning novel adjectives. *Bilingualism: Language and Cognition*, 1–19.
- Grodner, D., Dalini, M., Pearlstein-Levy, S., & Ward, A. (2012). Factors that contribute to the use of perspective in referent identification. Paper presented at the 25th Annual CUNY Conference on Human Sentence Processing.
- Grodner, D., & Sedivy, J. (2011). The effect of speaker-specific information on pragmatic inferences. *The Processing and Acquisition of Reference*, 2327, 239–72.
- Grosse, G., Behne, T., Carpenter, M., & Tomasello, M. (2010). Infants communicate in order to be understood. *Developmental Psychology*, *46*(6), 1710–1722.
- Grosse, G., Moll, H., & Tomasello, M. (2010). 21-Month-olds understand the co-operative logic of requests. *Journal of Pragmatics*, *42*(12), 3377–3383.
- Grosse, G., Schulze, C., Noveck, I., Tomasello, M., & Katsos, N. (in prep.). Three-year-olds make some, but not all inferences based on informativeness.
- Grosse, G., & Tomasello, M. (2012). Two-year-old children differentiate test questions from genuine questions. *Journal of Child Language*, *39*(1), 192–204.

- 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.
- Halberda, J. (2003). The development of a word-learning strategy. *Cognition*, 87(1), B23–B34.
- Halberda, J. (2006). Is this a dax which I see before me? Use of the logical argument disjunctive syllogism supports word-learning in children and adults. *Cognitive Psychology*, 53(4), 310–344.
- Hanna, J. E., & Tanenhaus, M. K. (2004). Pragmatic effects on reference resolution in a collaborative task: evidence from eye movements. *Cognitive Science*, 28(1), 105–115.
- Hanna, J. E., Tanenhaus, M. K., & Trueswell, J. C. (2003). The effects of common ground and perspective on domains of referential interpretation. *Journal of Memory and Language*, 49(1), 43–61.
- Haryu, E., & Imai, M. (1999). Controlling the application of the mutual exclusivity assumption in the acquisition of lexical hierarchies. *Japanese Psychological Research*, 41(1), 21–34.
- Haspelmath, M. (2006). Against markedness (and what to replace it with). *Journal of Linguistics*, 42(1), 25–70.
- Hawkins, R. X., & Goodman, N. D. (2016). Conversational expectations account for apparent limits on theory of mind use. In *Proceedings of the Thirty-Eighth Annual Conference of the Cognitive Science Society* (pp. 1889–1894). Austin, TX: Cognitive Science Society.
- Healey, E., & Skarabela, B. (2009). Are children willing to accept two labels for one object? A comparative study of mutual exclusivity in bilingual and monolingual children. In Marinis, T., Papangeli, A. & Stojanovik, V. (Eds.) *Proceedings of the Child Language Seminar 2007* (pp. 49–58). University of Reading.
- Heller, D., Grodner, D., & Tanenhaus, M. K. (2008). The role of perspective in identifying domains of reference. *Cognition*, 108(3), 831–836.
- Heller, D., Parisien, C., & Stevenson, S. (2016). Perspective-taking behavior as the probabilistic weighing of multiple domains. *Cognition*, 149, 104–120.

- Hendriks, P., & Koster, C. (2010). Production/comprehension asymmetries in language acquisition. *Lingua*, *120*(8), 1887–1897.
- 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*(1), 107–135.
- Hoff, E. (2006). How social contexts support and shape language development. *Developmental Review*, *26*(1), 55–88.
- Hoff, E., Core, C., Place, S., Rumiche, R., Señor, M., & Parra, M. (2012). Dual language exposure and early bilingual development. *Journal of Child Language*, *39*(1), 1–27.
- Hoff, E., Laursen, B., & Bridges, K. (2012). Measurement and model building in studying the influence of socioeconomic status on child development. In L. Mayes & M. Lewis (Eds.) *The Cambridge Handbook of Environment in Human Development* (pp. 590–606). Cambridge: Cambridge University Press.
- Hoicka, E., & Akhtar, N. (2011). Preschoolers joke with jokers, but correct foreigners. *Developmental Science*, *14*(4), 848–858.
- Hollich, G. J., Hirsh-Pasek, K., Golinkoff, R. M., Brand, R. J., Brown, E., Chung, H. L., ... Rocroi, C. (2000). Breaking the language barrier: An Emergentist coalition model of the origins of word learning. *Monographs of the Society for Research in Child Development*, *65*(3), 1–122.
- Horn, L. (1984). Toward a new taxonomy for pragmatic inference: Q-based and R-based implicature. In D. Schriften (Ed.), *Meaning, form, and use in context: Linguistic applications* (pp. 11–42). Washington, DC: Georgetown University Press.
- Horn, L. (2004). Implicature. In G. Ward & L. Horn (Eds.), *The Handbook of Pragmatics* (pp. 3–28). Oxford: Blackwell Publishing Ltd.
- Horowitz, A., & Frank, M. C. (2015). Sources of developmental change in pragmatic inferences about scalar terms. In *Proceedings of the 37th Annual Conference of the Cognitive Science Society* (pp. 926–931). Austin, TX: Cognitive Science Society.
- Horowitz, A., Schneider, R., & Frank, M. C. (under review). The trouble with quantifiers: Explaining children’s deficits in scalar implicature.

- Horst, J. S., Samuelson, L. K., Kucker, S. C., & McMurray, B. (2011). What's new? Children prefer novelty in referent selection. *Cognition*, *118*(2), 234–244.
- Horst, J. S., Scott, E. J., & Pollard, J. A. (2010). The role of competition in word learning via referent selection. *Developmental Science*, *13*(5), 706–713.
- Houston-Price, C., Caloghiris, Z., & Raviglione, E. (2010). Language experience shapes the development of the mutual exclusivity bias. *Infancy*, *15*(2), 125–150.
- Huang, Y. (2017). Implicature. In Y. Huang (Ed.), *The Oxford Handbook of Pragmatics* (pp. 155–179). Oxford: Oxford University Press.
- Huang, Y. T., & Snedeker, J. (2009). Semantic Meaning and Pragmatic Interpretation in 5-Year-Olds: Evidence From Real-Time Spoken Language Comprehension. *Developmental Psychology*, *45*(6), 1723–1739.
- Huang, Y. T., Spelke, E., & Snedeker, J. (2013). What Exactly do Numbers Mean? *Language Learning and Development*, *9*(2), 105–129.
- Hughes, C., Deater-Deckard, K., & Cutting, A. L. (1999). 'Speak roughly to your little boy'? Sex Differences in the Relations Between Parenting and Preschoolers' Understanding of Mind. *Social Development*, *8*(2), 143–160.
- Hurewitz, F., Papafragou, A., Gleitman, L., & Gelman, R. (2006). Asymmetries in the acquisition of numbers and quantifiers. *Language Learning and Development*, *2*(2), 77–96.
- Husband, E. M. (2014). A subclinical study of the cognitive resources underlying scalar implicature: A focus on scalar adjectives. In C. T. Schutze & L. Stockall (Eds.), *Connectedness: Papers by and for Sarah Van Wagenen* (pp. 189–211). UCLA Working Papers in Linguistics.
- Irvine, E. (2017, August). 'Language origins: When is a code not a code?' Paper presented at the ESPP Conference, University of Hertfordshire, UK.
- Jaeger, T. F., & Weatherholtz, K. (2016). What the heck is salience? How predictive language processing contributes to sociolinguistic perception. *Frontiers in Psychology*, *7*.
- Jary, M. (2013). Two types of implicature: Material and behavioural. *Mind and Language*, *28*(5), 638–660.

- Jaswal, V. K. (2010). Explaining the disambiguation effect: Don't exclude mutual exclusivity. *Journal of Child Language*, 37(1), 95–113.
- Jaswal, V. K., & Hansen, M. B. (2006). Learning words: Children disregard some pragmatic information that conflicts with mutual exclusivity. *Developmental Science*, 9(2), 158–165.
- Kalandadze, T., Norbury, C., Nærland, T., & Næss, K.-A. B. (2016). Figurative language comprehension in individuals with autism spectrum disorder: A meta-analytic review. *Autism*. [Epub ahead of print]. 10.1177/13623613166668652
- Kalashnikova, M., Mattock, K., & Monaghan, P. (2014). Disambiguation of novel labels and referential facts: A developmental perspective. *First Language*, 34(2), 125–135.
- Kalashnikova, M., Mattock, K., & Monaghan, P. (2015). The effects of linguistic experience on the flexible use of mutual exclusivity in word learning. *Bilingualism: Language and Cognition*, 18(4), 626–638.
- Kalashnikova, M., Mattock, K., & Monaghan, P. (2016). Mutual exclusivity develops as a consequence of abstract rather than particular vocabulary knowledge. *First Language*, 451–464.
- Kalish, C. W., & Sabbagh, M. A. (2007). Conventionality and cognitive development: Learning to think the right way. *New Directions for Child and Adolescent Development*, 2007(115), 1–9.
- Kamil, M. L., Pearson, P. D., Moje, E. B., & Afflerbach, P. (2011). *Handbook of Reading Research* (Vol. 4). New York: Routledge.
- Kampa, A., & Papafragou, A. (2017). Gricean epistemic reasoning in 4-year-olds. In *Proceedings of the 39th Annual Meeting of the Cognitive Science Society* (pp. 2362–2365). Austin, TX: Cognitive Science Society.
- Katsos, N. (2014). Scalar Implicatures. In D. Matthews (Ed.), *Pragmatic Development* (pp. 183–198). John Benjamins.
- Katsos, N., & Andrés-Roqueta, C. (2015). For which pragmatic phenomena is Theory of Mind necessary? Taking a different perspective In F. Salfner & U. Sauerland (Eds.), *Pre-proceedings of Trends in Experimental Pragmatics* (pp. 76–81). Berlin.

- Katsos, N., & Bishop, D. V. M. (2011). Pragmatic tolerance: Implications for the acquisition of informativeness and implicature. *Cognition*, *120*(1), 67–81.
- Katsos, N., Cummins, C., Ezeizabarrena, M.-J., Gavarró, A., Kraljević, J. K., Hrzica, G., ... Noveck, I. (2016). Cross-linguistic patterns in the acquisition of quantifiers. *Proceedings of the National Academy of Sciences*, *113*(33), 9244–9249.
- Katsos, N., & Smith, N. (2010). Pragmatic tolerance and speaker/comprehender asymmetries. In *Proceedings of the 34th Annual Boston Conference in Language Development* (pp. 221–232). Somerville, MA: Cascadilla Press.
- Katsos, N., & Wilson, E. (in prep.). Acquiring implicatures. In K. Schneider & E. Ifantidou (Eds.), *Developmental and Clinical Pragmatics*. Berlin: de Gruyter Mouton.
- Katsos, N., & Wilson, E. (2014). Convergence and divergence between word learning and pragmatic inferencing. In J. Degen, M. Franke, & N. D. Goodman (Eds.), *Proceedings – Formal and Experimental Pragmatics Workshop* (pp. 14–20). Tübingen.
- Keysar, B., Barr, D. J., Balin, J. A., & Brauner, J. S. (2000). Taking perspective in conversation: The role of mutual knowledge in comprehension. *Psychological Science*, *11*(1), 32–38.
- Keysar, B., Lin, S., & Barr, D. J. (2003). Limits on theory of mind use in adults. *Cognition*, *89*(1), 25–41.
- Kimball, S., Mattis, P., & The GIMP Development Team. (2016). GNU Image Manipulation Program (Version 2.8.18).
- Kispal, A. (2008). Effective Teaching of Inference Skills for Reading. Literature Review. Research Report DCSF-RR031. *National Foundation for Educational Research*.
- Kissine, M. (2016). Pragmatics as Metacognitive Control. *Frontiers in Psychology*, *6*. 2057. <http://doi.org/10.3389/fpsyg.2015.02057>
- Kissine, M. (2017, June). *What Autism Spectrum Disorder can teach us about pragmatics*. Paper presented at XPrag 2017, Cologne, Germany.
- Kissine, M., Cano-Chervel, J., Carlier, S., Brabanter, P. D., Ducenne, L., Pairon, M.-C., ... Leybaert, J. (2015). Children with Autism Understand Indirect Speech Acts: Evidence from a Semi-Structured Act-Out Task. *PLOS ONE*, *10*(11), e0142191.

- Kline, M., Muentener, P., & Schulz, L. (2013). Transitive and periphrastic sentences affect memory for simple causal scenes. In *Proceedings of the 35th Annual Conference of the Cognitive Science Society* (pp. 794–798). Austin, TX: Cognitive Science Society.
- Kovács, Á. M. (2009). Early bilingualism enhances mechanisms of false-belief reasoning. *Developmental Science*, *12*(1), 48–54.
- Kronmüller, E., Morisseau, T., & Noveck, I. A. (2013). Show me the pragmatic contribution: a developmental investigation of contrastive inference. *Journal of Child Language*, 1–30.
- Kuperman, V., Stadthagen-Gonzalez, H., & Brysbaert, M. (2012). Age-of-acquisition ratings for 30,000 English words. *Behavior Research Methods*, *44*(4), 978–990.
- Levinson, S. C. (2000). *Presumptive meanings: The theory of generalized conversational implicature*. Cambridge, MA: MIT Press.
- Lewis, D. (1969). *Convention: a philosophical study*. Cambridge, MA: Harvard University Press.
- Lewis, S. (2013). *Pragmatic enrichment in language processing and development* (PhD dissertation). University of Maryland.
- Li, P., Barner, D., & Huang, B. H. (2008). Classifiers as count syntax: Individuation and measurement in the acquisition of Mandarin Chinese. *Language Learning and Development*, *4*(4), 249–290.
- Liebal, K., Carpenter, M., & Tomasello, M. (2010). Infants' use of shared experience in declarative pointing. *Infancy*, *15*(5), 545–556.
- Liebal, K., Carpenter, M., & Tomasello, M. (2011). Young children's understanding of markedness in non-verbal communication. *Journal of Child Language*, *38*(4), 888–903.
- Lieven, E. (2006). Language Development: Overview. In K. Brown (Ed.), *Encyclopedia of Language & Linguistics (Second Edition)* (pp. 376–391). Oxford: Elsevier.
- Locke, A., Ginsborg, J., & Peers, I. (2002). Development and disadvantage: implications for the early years and beyond. *International Journal of Language & Communication Disorders*, *37*(1), 3–15.

- Loukusa, S., Leinonen, E., & Ryder, N. (2007). Development of pragmatic language comprehension in Finnish-speaking children. *First Language*, 27(3), 279–296.
- Markman, E. M. (1994). Constraints on word meaning in early language acquisition. *Lingua*, 92, 199–227.
- Markman, E. M., & Wachtel, G. F. (1988). Children's use of mutual exclusivity to constrain the meanings of words. *Cognitive Psychology*, 20(2), 121–157.
- Markman, E. M., Wasow, J. L., & Hansen, M. B. (2003). Use of the mutual exclusivity assumption by young word learners. *Cognitive Psychology*, 47(3), 241–275.
- Marr, D. (1982). *Vision: A computational investigation into the human representation and processing of visual information*. NY: New York: Henry Holt & Co.
- Mascaro, O., & Sperber, D. (2015). Pragmatic Inference In Infancy. In F. Salfner & U. Sauerland (Eds.), *Pre-proceedings of Trends in Experimental Pragmatics* (pp. 95–102). Berlin.
- Matthews, D., Butcher, J., Lieven, E., & Tomasello, M. (2012). Two-and Four-Year-Olds Learn to Adapt Referring Expressions to Context: Effects of Distracters and Feedback on Referential Communication. *Topics in Cognitive Science*, 4(2), 184–210.
- 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(3), 403–422.
- Matthews, D., Lieven, E., & Tomasello, M. (2007). How toddlers and preschoolers learn to uniquely identify referents for others: A training study. *Child Development*, 78(6), 1744–1759.
- Matthews, D., Lieven, E., & Tomasello, M. (2010). What's in a manner of speaking? Children's sensitivity to partner-specific referential precedents. *Developmental Psychology*, 46(4), 749–760.
- McGillion, M., Pine, J. M., Herbert, J. S., & Matthews, D. (2017). A randomised controlled trial to test the effect of promoting caregiver contingent talk on language development in infants from diverse socioeconomic status backgrounds. *Journal of Child Psychology and Psychiatry*. [Epub ahead of print]. 10.1111/jcpp.12725

- Meibauer, J. (2012). What is a context? Theoretical and empirical evidence. In R. Finkbeiner, J. Meibauer, & P. Schumacher (Eds.), *What is a Context? Linguistic approaches and challenges* (Vol. 196, pp. 9–32). Amsterdam: John Benjamins.
- Meibauer, J. (2017). Pragmatics and children's literature. In R. Giora & M. Haugh (Eds.), *Applications of Intercultural Pragmatics: Cognitive and Sociopragmatic Perspectives on Language Use* (pp. 371–388). Berlin/Boston: Walter de Gruyter.
- Merriman, W. E., Bowman, L. L., & MacWhinney, B. (1989). The mutual exclusivity bias in children's word learning. *Monographs of the Society for Research in Child Development*, *i*–129.
- Merriman, W. E., & Schuster, J. M. (1991). Young children's disambiguation of object name reference. *Child Development*, *62*(6), 1288–1301.
- Mervis, C. B., & Bertrand, J. (1994). Acquisition of the Novel Name-Nameless Category (N3C) Principle. *Child Development*, *65*(6), 1646–1662.
- Metzing, C., & Brennan, S. E. (2003). When conceptual pacts are broken: Partner-specific effects on the comprehension of referring expressions. *Journal of Memory and Language*, *49*(2), 201–213.
- Miller, K., Schmitt, C., Chang, H.-H., & Munn, A. (2005). Young children understand some implicatures. In *Proceedings of the 29th Annual Boston University Conference on Language Development* (pp. 389–400). Somerville, MA: Cascadilla Press.
- Moll, H., Carpenter, M., & Tomasello, M. (2011). Social engagement leads 2-year-olds to overestimate others' knowledge. *Infancy*, *16*(3), 248–265.
- Moll, H., & Kadipasaoglu, D. (2013). The primacy of social over visual perspective-taking. *Frontiers in Human Neuroscience*, *7*, 558. <http://doi.org/10.3389/fnhum.2013.00558>
- Moll, H., Koring, C., Carpenter, M., & Tomasello, M. (2006). Infants determine others' focus of attention by pragmatics and exclusion. *Journal of Cognition and Development*, *7*(3), 411–430.
- Moll, H., & Meltzoff, A. N. (2011). Perspective-Taking and its Foundation in Joint Attention. In J. Roessler, H. Lerman, & N. Eilan (Eds.), *Perception, causation, and objectivity* (pp. 286–304). Oxford: Oxford University Press.

- Moll, H., & Tomasello, M. (2006). Level 1 perspective-taking at 24 months of age. *British Journal of Developmental Psychology*, 24(3), 603–613.
- Moore, R. (2016a). Gricean communication and cognitive development. *The Philosophical Quarterly*, 67(267), 303–326.
- Moore, R. (2016b). Gricean communication, joint action, and the evolution of cooperation. *Topoi*, 1–13.
- Morisseau, T., Davies, C., & Matthews, D. (2013). How do 3- and 5-year-olds respond to under- and over-informative utterances? *Journal of Pragmatics*, 59, Part A(0), 26–39.
- Muentener, P., & Lakusta, L. (2011). The intention-to-CAUSE bias: Evidence from children's causal language. *Cognition*, 119(3), 341–355.
- Nadig, A. S., & Sedivy, J. C. (2002). Evidence of perspective-taking constraints in children's on-line reference resolution. *Psychological Science*, 13(4), 329–336.
- Nilsen, E. S., & Graham, S. A. (2009). The relations between children's communicative perspective-taking and executive functioning. *Cognitive Psychology*, 58(2), 220–249.
- Nilsen, E. S., & Graham, S. A. (2012). The development of preschoolers' appreciation of communicative ambiguity. *Child Development*, 83(4), 1400–1415.
- Nilsen, E. S., Mangal, L., & MacDonald, K. (2013). Referential communication in children with ADHD: challenges in the role of a listener. *Journal of Speech, Language, and Hearing Research*, 56(2), 590–603.
- Nordmeyer, A. E., & Frank, M. C. (2014). A pragmatic account of the processing of negative sentences. In *Proceedings of the 36th Annual Conference of the Cognitive Science Society* (pp. 2699–2704). Austin, TX: Cognitive Science Society.
- Noveck, I. A. (2001). When children are more logical than adults: Experimental investigations of scalar implicature. *Cognition*, 78(2), 165–188.
- Office for National Statistics. (2016). Towns and cities analysis, England and Wales, March 2016. Retrieved 31 May 2017, from <https://www.ons.gov.uk/peoplepopulationandcommunity/housing/articles/townsandcitiesanalysisenglandandwalesmarch2016/2016-03-18>

- O'Grady, C., Kliesch, C., Smith, K., & Scott-Phillips, T. C. (2015). The ease and extent of recursive mindreading, across implicit and explicit tasks. *Evolution and Human Behavior*, 36(4), 313–322.
- Okanda, M., Asada, K., Moriguchi, Y., & Itakura, S. (2015). Understanding violations of Gricean maxims in preschoolers and adults. *Developmental Psychology*, 6, 901.
- O'Neill, D. K. (1996). Two-year-old children's sensitivity to a parent's knowledge state when making requests. *Child Development*, 659–677.
- Pace, A., Luo, R., Hirsh-Pasek, K., & Golinkoff, R. M. (2017). Identifying Pathways Between Socioeconomic Status and Language Development. *Annual Review of Linguistics*, 3(1), 285–308.
- Papafragou, A., Friedberg, C., & Cohen, M. (2017). The role of speaker knowledge in children's pragmatic inferences. *Child Development*. [Epub ahead of print].
10.1111/cdev.12841
- Papafragou, A., & Musolino, J. (2003). Scalar implicatures: experiments at the semantics–pragmatics interface. *Cognition*, 86(3), 253–282.
- Papafragou, A., & Skordos, D. (2016). Scalar Implicature. In J. Lidz, W. Snyder, & J. Pater (Eds.), *Oxford Handbook of Developmental Linguistics* (pp. 611–632). Oxford: Oxford University Press.
- Papafragou, A., & Tantalou, N. (2004). Children's computation of implicatures. *Language Acquisition*, 12(1), 71–82.
- Paradis, J. (2011). Individual differences in child English second language acquisition: Comparing child-internal and child-external factors. *Linguistic Approaches to Bilingualism*, 1(3), 213–237.
- Paradis, J., Emmerzael, K., & Duncan, T. S. (2010). Assessment of English language learners: Using parent report on first language development. *Journal of Communication Disorders*, 43(6), 474–497.
- Pearson, H. A., Khan, M., & Snedeker, J. (2011). Even more evidence for the emptiness of plurality: An experimental investigation of plural interpretation as a species of implicature. In *Proceedings of SALT* (Vol. 20, pp. 489–507).

- Perner, J., Leekam, S. R., & Wimmer, H. (1987). Three-year-olds' difficulty with false belief: The case for a conceptual deficit. *British Journal of Developmental Psychology*, 5(2), 125–137.
- Piaget, J., & Inhelder, B. (1967). The coordination of perspectives. *The Child's Conception of Space*, 8, 209–246.
- Pogue, A., Kurumada, C., & Tanenhaus, M. K. (2016). Talker-Specific Generalization of Pragmatic Inferences based on Under- and Over-Informative Prenominal Adjective Use. *Language Sciences*, 2035. <http://doi.org/10.3389/fpsyg.2015.02035>
- Pogue, A., Tanenhaus, M. K., & Kurumada, C. (2017, February). *Preschoolers' pragmatic generalizations based on scalar adjective use*. Poster presented at the CSLI Workshop: Bridging computational and psycholinguistic approaches to the study of meaning, Stanford, CA.
- Politzer-Ahles, S., & Fiorentino, R. (2013). The Realization of Scalar Inferences: Context Sensitivity without Processing Cost. *PLOS ONE*, 8(5), e63943.
- Pomper, R., & Saffran, J. R. (2016). Roses Are Red, Socks Are Blue: Switching Dimensions Disrupts Young Children's Language Comprehension. *PLOS ONE*, 11(6), e0158459.
- Pouscoulous, N. (2013). Early pragmatics with words. In C. Schulze & F. Liedtke (Eds.), *Beyond Words: Content, Context, and Inference* (pp. 121–146). Berlin: Walter de Gruyter.
- Pouscoulous, N. (2014). 'The elevator's buttocks': Metaphorical abilities in children. In D. Matthews (Ed.), *Pragmatic Development in First Language Acquisition* (pp. 239–260). Amsterdam: John Benjamins.
- Pouscoulous, N., & Dulcinati, G. (2017). *Quantity implicature in a competitive game*. Paper presented at XPrag 2017, Cologne, Germany.
- 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.
- Prolific Academic Ltd. (2016). Prolific Academic. Retrieved from www.prolific.ac
- Python Software Foundation. (2016). Python (Version 3.6). Python Software Foundation. Retrieved from www.python.org

- Qing, C., & Franke, M. (2014). Gradable Adjectives, Vagueness, and Optimal Language Use: A Speaker-Oriented Model. *Proceedings of SALT 2014*, 23–41.
- Qualtrics. (2016). (Version 2016). Provo, Utah, USA: Qualtrics. Retrieved from www.qualtrics.com
- Rabagliati, H., & Robertson, A. (2017). How do children learn to avoid referential ambiguity? Insights from eye-tracking. *Journal of Memory and Language*, 94, 15–27.
- R Core Team. (2016). R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. 2015. www.R-Project.org
- Rakoczy, H., Warneken, F., & Tomasello, M. (2008). The sources of normativity: young children's awareness of the normative structure of games. *Developmental Psychology*, 44(3), 875– 881.
- Reetzke, R., Zou, X., Sheng, L., & Katsos, N. (2015). Communicative development in bilingually exposed Chinese children with autism spectrum disorders. *Journal of Speech, Language, and Hearing Research*, 58(3), 813–825.
- Roberts, C. (2012). Information structure: Towards an integrated formal theory of pragmatics. *Semantics and Pragmatics*, 5, 6–1.
- Robinson, E. J. (2011). Development of understanding of the causal connection between perceptual access and knowledge state. In J. Roessler, H. Lerman, & N. Eilan (Eds.), *Perception, causation, and objectivity* (pp. 324–341). Oxford: Oxford University Press.
- Rogoff, B. (2003). *The cultural nature of human development*. Oxford: Oxford University Press.
- Rowe, M. L., & Goldin-Meadow, S. (2009). Differences in early gesture explain SES disparities in child vocabulary size at school entry. *Science*, 323(5916), 951–953.
- RStudio Team. (2016). RStudio: Integrated Development for R (Version 1.0.136). Boston, MA.: RStudio, Inc. Retrieved from www.rstudio.com
- Rubio-Fernández, P. (2013). Perspective tracking in progress: Do not disturb. *Cognition*, 129(2), 264–272.
- Rubio-Fernández, P. (2016). The director task: A test of Theory-of-Mind use or selective attention? *Psychonomic Bulletin & Review*, 1–8.

- Rubio-Fernández, P., & Geurts, B. (2013). How to Pass the False-Belief Task Before Your Fourth Birthday. *Psychological Science*, 24(1), 27–33.
- Rubio-Fernández, P., & Geurts, B. (2016). Don't mention the marble! The role of attentional processes in false-belief tasks. *Review of Philosophy and Psychology*, 7(4), 835–850.
- Russell, B. (2012). *Probabilistic reasoning and the computation of scalar implicatures*. (PhD thesis). Brown University, Providence, RI.
- Sabbagh, M. A., & Henderson, A. M. E. (2007). How an appreciation of conventionality shapes early word learning. *New Directions for Child and Adolescent Development*, 2007(115), 25–37.
- San Juan, V., Morra, S., Gibbard, K., Khu, M., & Graham, S. A. (2017, July). *Three- and four-year-olds rapidly integrate a communicative partner's perspective*. Paper presented at the International Congress for the Study of Child Language, Lyon, France.
- Sauerland, U. (2004). Scalar implicatures in complex sentences. *Linguistics and Philosophy*, 27(3), 367–391.
- Scheele, A. F., Leseman, P. P., & Mayo, A. Y. (2010). The home language environment of monolingual and bilingual children and their language proficiency. *Applied Psycholinguistics*, 31(1), 117–140.
- Schmitt, C., & Miller, K. (2010). Using comprehension methods in language acquisition research. *Experimental Methods in Language Acquisition Research*, 35–56.
- Schulze, C., Grassmann, S., & Tomasello, M. (2013). 3-Year-Old Children Make Relevance Inferences in Indirect Verbal Communication. *Child Development*, 84(6), 2079–2093.
- Schulze, C., & Tomasello, M. (2015). 18-month-olds comprehend indirect communicative acts. *Cognition*, 136, 91–98.
- Schwarz, F. (in press). Experimental Pragmatics. In M. Aronoff (Ed.), *Oxford Research Encyclopedia of Linguistics*.
- Schwarz, F., Romoli, J., & Bill, C. (2015). Scalar implicatures in processing: Slowly accepting the truth (literally). In *Proceedings of SuB 19, Georg August University, Göttingen*. Göttingen.
- Scofield, J., & Behrend, D. A. (2007). Two-year-olds differentially disambiguate novel words and facts. *Journal of Child Language*, 34(4), 875–889.

- Scott-Phillips, T. (2014). *Speaking Our Minds: Why human communication is different, and how language evolved to make it special*. Basingstoke: Palgrave MacMillan.
- Scrafton, S. (2009). *Experimental Investigations into how Children and Adults Process the Implicature Associated with the Scalar Term Some*. (PhD thesis). Durham University.
- Scrafton, S., & Feeney, A. (2006). Dual processes, development, and scalar implicature. In *Proceedings of the 28th Annual Conference of the Cognitive Science Society* (pp. 774–779). Austin, TX: Cognitive Science Society.
- Siegal, M. (2008). *Marvelous Minds: The Discovery of what Children Know*. Oxford: Oxford University Press.
- Siegal, M., Iozzi, L., & Surian, L. (2009). Bilingualism and conversational understanding in young children. *Cognition*, *110*(1), 115–122.
- Siegal, M., Matsuo, A., Pond, C., & Otsu, Y. (2007). Bilingualism and cognitive development: Evidence from scalar implicatures. In *Proceedings of the Eighth Tokyo Conference on Psycholinguistics* (pp. 265–280). Hituzi Syobo Tokyo, Japan.
- Siegal, M., Surian, L., Matsuo, A., Geraci, A., Iozzi, L., Okumura, Y., & Itakura, S. (2010). Bilingualism accentuates children's conversational understanding. *PLoS One*, *5*(2), e9004.
- Skordos, D. (2014). *Scalar implicatures in children: Alternatives and relevance*. (PhD thesis) University of Delaware.
- Skordos, D., & Papafragou, A. (2016). Children's derivation of scalar implicatures: alternatives and relevance. *Cognition*, *153*, 6–18.
- Smith, L. B. (1999). Children's noun learning: How general learning processes make specialized learning mechanisms. In B. MacWhinney (Ed.), *The emergence of language* (pp. 277–303). Mahwah, New Jersey: Lawrence Erlbaum Associates.
- Snedeker, J. (2015, July). *Scalar implicature: a whirlwind tour with stops in processing, development and disorder*. Talk, Tübingen.
- Snedeker, J., & Huang, Y. T. (2016). Sentence Processing. In E. L. Bavin & L. Naigles (Eds.), *The Handbook of Child Language* (2nd ed., pp. 409–437). Cambridge: Cambridge University Press.

- Sobel, D. M., & Corriveau, K. H. (2010). Children monitor individuals' expertise for word learning. *Child Development, 81*(2), 669–679.
- Sobel, D. M., Sedivy, J., Buchanan, D. W., & Hennessy, R. (2012). Speaker reliability in preschoolers' inferences about the meanings of novel words. *Journal of Child Language, 39*(1), 90–104.
- Song, G., & Wolff, P. (2005). Linking Perceptual Properties to Linguistic Expressions of Causation. In M. Achard & S. Kemmer (Eds.), *Language, culture and mind* (pp. 237–250). Stanford, CA.: CSLI Publications.
- Sperber, D. (1994). Understanding verbal understanding. In J. Khalfa (Ed.) *What Is Intelligence?* (pp. 179–198). Cambridge: Cambridge University Press.
- Sperber, D., & Wilson, D. (1995). *Relevance: Communication and Cognition*. Oxford: Blackwell.
- Spiegel, C., & Halberda, J. (2011). Rapid fast-mapping abilities in 2-year-olds. *Journal of Experimental Child Psychology, 109*(1), 132–140.
- Spotorno, N., & Noveck, I. A. (2014). When is irony effortful? *Journal of Experimental Psychology, 143*(4), 1649–1665.
- Stateva, P., Andreetta, S., Dupuy, L. E., Cheylus, A., Deprez, V., Van Der Henst, J.-B., ... Reboul, A. (2015, September). *Pragmatic abilities in bilinguals: the case of scalar implicatures*. Paper presented at the 8th Syntax, Phonology and Language Analysis conference (SinFonJA 8), University of Ljubljana, Slovenia.
- Stephens, G., & Matthews, D. (2014). The communicative infant from 0–18 months: The social-cognitive foundations of pragmatic development. In Matthews, Danielle (Ed.), *Pragmatic Development in First Language Acquisition* (Vol. 10, pp. 13–36). Amsterdam: John Benjamins Publishing Company.
- Stiller, A. J., Goodman, N. D., & Frank, M. C. (2015). Ad-hoc Implicature in Preschool Children. *Language Learning and Development, 11*(2), 176–190.
- Suanda, S. H., & Namy, L. L. (2013). Young word learners' interpretations of words and symbolic gestures within the context of ambiguous reference. *Child Development, 84*(1), 143–153.

- Sullivan, J., & Barner, D. (2011). Number words, quantifiers, and principles of word learning. *Wiley Interdisciplinary Reviews: Cognitive Science*, 2(6), 639–645.
- Sullivan, J., & Barner, D. (2015). Discourse bootstrapping: preschoolers use linguistic discourse to learn new words. *Developmental Science*, 19(1), 63–75.
- Sullivan, J., Davidson, K., Wade, S., & Barner, D. (submitted). Differentiating scalar implicature from mutual exclusivity in language acquisition.
- Symeonidou, I., Dumontheil, I., Chow, W.-Y., & Breheny, R. (2016). Development of online use of theory of mind during adolescence: An eye-tracking study. *Journal of Experimental Child Psychology*, 149, 81–97.
- Tieu, L. (2017). *Children's scalar implicatures: comparing comprehension and production*. Poster presented at the International Congress for the Study of Child Language, Lyon, France.
- Tomasello, M. (2000). The social-pragmatic theory of word learning. *Pragmatics*, 10(4), 401–413
- Tomasello, M. (2003). *Constructing a language: A usage-based theory of language acquisition*. Cambridge, MA: Harvard University Press.
- Tomasello, M., Carpenter, M., & Liszkowski, U. (2007). A new look at infant pointing. *Child Development*, 78(3), 705–722.
- Tomasello, M., Strosberg, R., & Akhtar, N. (1996). Eighteen-month-old children learn words in non-ostensive contexts. *Journal of Child Language*, 23(1), 157–176.
- Tribushinina, E. (2012). Comprehension of relevance implicatures by pre-schoolers: The case of adjectives. *Journal of Pragmatics*, 44(14), 2035–2044.
- Trueswell, J. C., Sekerina, I., Hill, N. M., & Logrip, M. L. (1999). The kindergarten-path effect: Studying on-line sentence processing in young children. *Cognition*, 73(2), 89–134.
- Van Kuppevelt, J. (1996). Inferring from Topics: Scalar Implicatures as Topic-Dependent Inferences. *Linguistics and Philosophy*, 19(4), 393–443.
- Van Tiel, B., Van Miltenburg, E., Zevakhina, N., & Geurts, B. (2014). Scalar diversity. *Journal of Semantics*, 1–39.

- Veenstra, A. (2010). *Pragmatic Tolerance and Theory of Mind in Children's Comprehension of Under-Informativeness* (Masters thesis). University of Groningen.
- Verbuk, A. (2012). Developmental evidence against the theoretical distinction between Horn and pragmatic scales. *Journal of Pragmatics*, 44(12), 1680–1700.
- Verbuk, A., & Shultz, T. (2010). Acquisition of Relevance implicatures: A case against a Rationality-based account of conversational implicatures. *Journal of Pragmatics*, 42(8), 2297–2313.
- VideoLAN. (2017). VLC media player (Version 2.2.4). Retrieved from videolan.org/vlc/
- 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.
- Wermelinger, S., Gampe, A., & Daum, M. M. (2017). Bilingual toddlers have advanced abilities to repair communication failure. *Journal of Experimental Child Psychology*, 155, 84–94.
- Wieling, M. (2015, January). Mixed-effects regression and generalized additive modeling. Workshop at the Department of Theoretical and Applied Linguistics, University of Cambridge. Resources retrieved from <http://www.let.rug.nl/~wieling/statscourse/>
- Williams, J. C. (2014). Recent official policy and concepts of reading comprehension and inference: the case of England's primary curriculum. *Literacy*, 48(2), 95–102.
- Wilson, D. (2000). Metarepresentation in linguistic communication. In D. Sperber (Ed.), *Metarepresentations: A multidisciplinary perspective* (Vol. 10, pp. 411–448). Oxford: Oxford University Press.
- Wilson, D., & Sperber, D. (2012). *Meaning and relevance*. Cambridge: Cambridge University Press.
- Wilson, E., & Katsos, N. (2016). In a manner of speaking: an empirical investigation of Manner Implicatures. In F. Salfner & U. Sauerland (Eds.), *Pre-proceedings of Trends in Experimental Pragmatics* (pp. 170–176). Berlin.
- 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.

- Wolff, P. (2003). Direct causation in the linguistic coding and individuation of causal events. *Cognition*, 88(1), 1–48.
- Yildirim, I., Degen, J., Tanenhaus, M. K., & Jaeger, T. F. (2016). Talker-specificity and adaptation in quantifier interpretation. *Journal of Memory and Language*, 87, 128–143.
- Yoon, E. J., Wu, Y. C., & Frank, M. C. (2015). Children's Online Processing of Ad-Hoc Implicatures. In *Proceedings of the 37th Annual Conference of the Cognitive Science Society* (pp. 2757–2762). Austin, TX: Cognitive Science Society.
- Yow, W. Q., & Markman, E. M. (2011a). Bilingualism and children's use of paralinguistic cues to interpret emotion in speech. *Bilingualism: Language and Cognition*, 14(4), 562–569.
- Yow, W. Q., & Markman, E. M. (2011b). Young bilingual children's heightened sensitivity to referential cues. *Journal of Cognition and Development*, 12(1), 12–31.
- Yu, C., & Smith, L. B. (2007). Rapid word learning under uncertainty via cross-situational statistics. *Psychological Science*, 18(5), 414–420.
- Zondervan, A. J. (2010). Scalar implicatures or focus: an experimental approach. *LOT Dissertation Series*, 249.
- Zondervan, A., Meroni, L., & Gualmini, A. (2008). Experiments on the role of the question under discussion for ambiguity resolution and implicature computation in adults. In *Proceedings of SALT* (Vol. 18, pp. 765–777).
- Zosh, J. M., Brinster, M., & Halberda, J. (2013). Optimal Contrast: Competition Between Two Referents Improves Word Learning. *Applied Developmental Science*, 17(1), 20–28.
- Zufferey, S. (2010). *Lexical pragmatics and theory of mind: the acquisition of connectives* (Vol. 201). Amsterdam/Philadelphia: John Benjamins Publishing.

8 Appendices

8.1 Stimuli for Experiment 1A

	Context	Critical Utterance	Control Utterance	Critical target	Control target
	Warm up:				
	It was time for a snack.	Bob's mum asked, 'Do you want the big apple or the little one?'	And I said, 'I want the little apple.'	little apple	big apple
	Bob was drawing a picture.	His dad asked, 'What have you drawn?'	And I said, 'I've drawn a teddy bear.'	teddy bear	flowers
	Next they watched TV.	His dad asked, 'Do you want to sit on the black chair or the brown chair?'	And I said, 'I want to sit on the black chair.'	black chair	brown chair
	Bob was doing some baking.	His dad asked, 'What have you made?'	And I said, 'I've baked a cake.'	cake	bread
	This is a story about Bob at home...				
R1	It was breakfast time.	Bob's dad asked, 'What would you like for breakfast?'	And I said, 'I'll get the milk.'	And I said, 'I'd like toast.'	cereal toast
A1	Bob was getting ready for school.	His mum asked, 'What have you packed in your bag?'	And I said, 'I've packed a hat.'	And I said, 'I've packed a book and a hat.'	hat book and hat
S1	Bob was laying the table.	His dad asked, 'What did you do with the rows of cups?'	And I said, 'I filled some of the cups with juice.'	And I said, 'I filled all of the cups with juice.'	some of the cups all of the cups
A2	Bob came out of the kitchen.	His dad asked, 'What did you take from the fridge?'	And I said, 'I took a strawberry.'	And I said, 'I took an orange and a strawberry.'	strawberry orange and strawberry
R2	Bob had a new toy.	His mum asked, 'Where do you want to use it?'	And I said, 'I'll get my coat.'	And I said, 'I want to play inside.'	outside inside

S2	Bob made a crash in the kitchen	His dad asked, 'What did you do with the pile of plates?'	And I said, 'I broke some of the plates.'	And I said, I broke all of the plates.	some	all
This is a story about Bob at school...						
A3	Bob was eating his packed lunch.	A friend asked, 'What did you put in your lunchbox?'	And I said, 'I put in a sandwich.'	And I said, 'I put in a biscuit and a sandwich.'	sandwich	biscuit and sandwich
S3	Bob was doing some building.	His teacher asked, 'What did you do with the pile of blocks?'	And I said, 'I used some of the blocks.'	And I said, 'I used all of the blocks.'	some	all
R3	Bob was making a picture.	His teacher asked, 'How would you like to make your picture?'	And I said, 'I'll fetch a brush.'	And I said, 'I'd like to use pencils.'	paints	crayons
A4	Bob was doing some drawing.	His teacher asked, 'What have you drawn?'	And I said, 'I've drawn a picture of my mum.'	And I said, 'I've drawn a picture of my dad and my mum.'	mum	dad and mum
S4	Bob was doing some baking	His teacher asked, 'What did you do with the gingerbread men on the tray?'	And I said, 'I gave smiley faces to some of the gingerbread men.'	And I said, 'I gave smiley faces to all of the gingerbread men.'	some	all
R4	It was time for art.	His teacher asked, 'What would you like to do in art?'	And I said, 'I'll get the glue.'	And I said, 'I'd like to draw with crayons.'	collage	crayons
This is a story about Bob in the garden...						
S5	Bob went to see how his plant had grown.	His mum asked, 'What did you do with the flowers on your plant?'	And I said, 'I picked some of the flowers.'	And I said, I picked all of the flowers.	all	some
R5	There was a stream in the garden.	His dad asked, 'What do you want to do by the stream?'	And I said, 'I brought a net.'	And I said, 'I want to draw the stream.'	drawing	fishing
A5	Bob went to the vegetable patch.	His mum asked, 'What did you find?'	And I said, 'I've found a carrot.'	And I said, 'I've found a potato and a carrot.'	carrot	potato and carrot

R6	There was lots to do in the garden.	His mum asked, 'What fruit do you want to pick?'	And I said, 'I'll get a ladder.'	And I said, 'I want to pick strawberries'	strawberries	apples
S6	Bob emptied the washing machine.	His mum asked, 'What did you do with the pile of T-shirts?'	And I said, 'I hung up some of the T-shirts.'	And I said, 'I hung up all of the T-shirts.'	all	some
A6	Bob felt a bit cold in the garden.	His dad asked, 'What have you put on to keep warm?'	And I said, 'I've put on a jumper.'	And I said, 'I've put on a scarf and a jumper.'	jumper	scarf and jumper

This is a story about Bob in the park...

R7	It was hot in the park.	Bob's dad asked, 'How do you want to cool down?'	And I said, 'I've got my trunks.'	And I said, 'I want a drink.'	pool	drink
A7	Bob went over to the ice cream van.	His dad asked 'What did you buy?'	And I said, 'I got a lolly.'	And I said, 'I got a drink and a lolly.'	can and lolly	lolly
S7	Bob was enjoying the park.	His dad asked, 'What did you do to the ducks on the pond?'	And I said, 'I fed some of the ducks.'	And I said, 'I fed all of the ducks.'	all	some
A8	Bob was looking in the pond.	His mum asked, 'What have you seen?'	And I said, 'I've seen a frog.'	And I said, 'I've seen a fish and a frog.'	frog	fish and frog
R8	They wanted to have some fun.	Bob's dad asked, 'What game shall we play?'	And I said, 'I've brought a ball.'	And I said, 'Lets play cards.'	tennis	cards
S8	Bob liked collecting things from the park.	His mum asked, 'What did you do with the leaves under the tree?'	And I said, 'I picked up some of the leaves.'	And I said, 'I picked up all of the leaves.'	some	all

This is a story about Bob at the shop. Bob went to a big shop one day and to buy a lot of things. He went around the shop and put different things in his basket.

Near the door...		'I picked a dax.'	'I picked a fork.'	dax	fork
He went further inside and...		'I chose a wug.'	'I chose some scissors.'	wug	scissors
Around the next corner...		'I got a jop.'	'I got an apple.'	jop	apple

Bob went up the stairs and...	'I got a zev.'	'I got a knife.'	zev	cutlery knife
He looked on a top shelf and...	'I picked up a blicket.'	'I picked up a spoon.'	blicket	spoon
He looked on the bottom shelf and...	'I picked up a pimwit.'	'I picked up a banana.'	pimwit	banana
He went to the back of the shop	'I chose a fendle.'	'I chose a pen.'	fendle	pen
His basket was almost full but...	'I got a chatten.'	'I got a bucket.'	chatten	bucket

Table 8.1 Items for Experiment 1A

Blicket



Pimwit



Fendle



Chatten



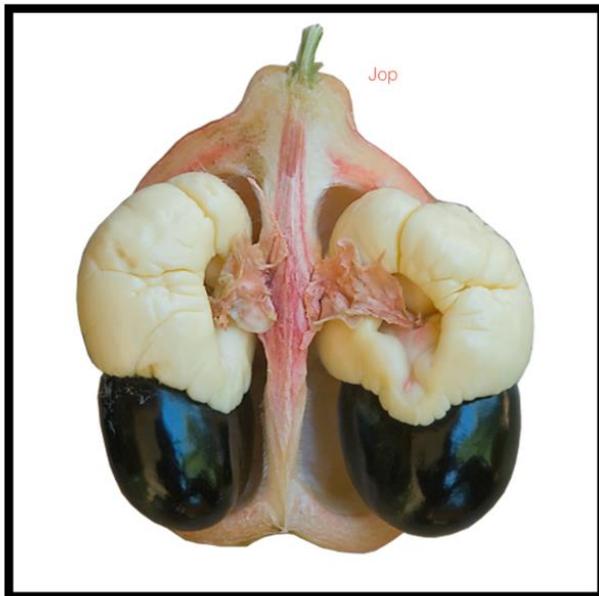


Figure 8.1 Pictures of novel items used in Experiment 1A (From top, left to right: a beater, a mangosteen, an aspergillum, a gauge, an eccentric rod, a dough mixer, an ackee, a climbing chock)

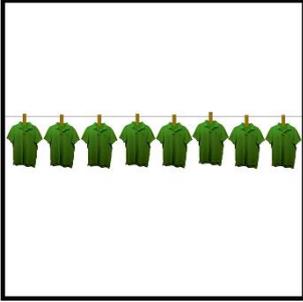
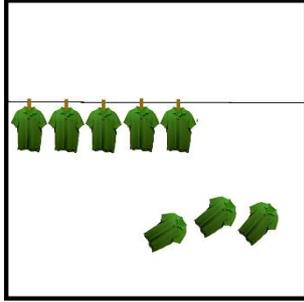
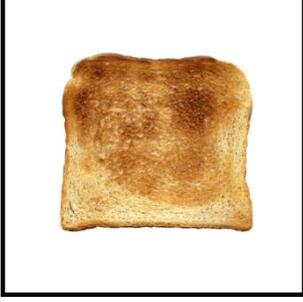
Context picture	Control picture	Critical picture
		
		
		

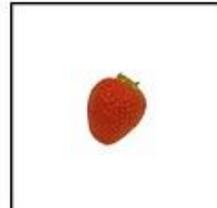
Figure 8.2 Examples of scalar, ad hoc and relevance pictures in Experiment 1A

Ad hoc pictures

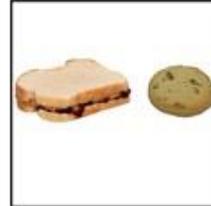
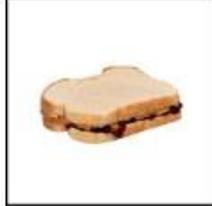
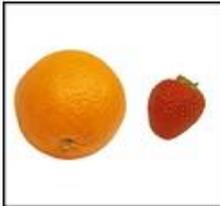
A1



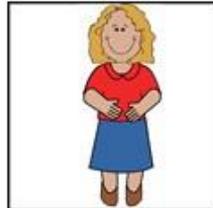
A2



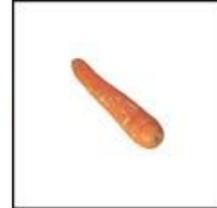
A3



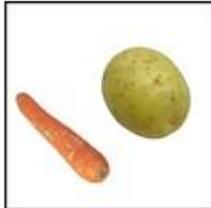
A4



A5



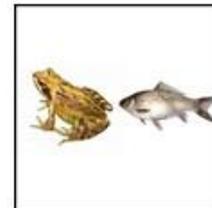
A6



A7

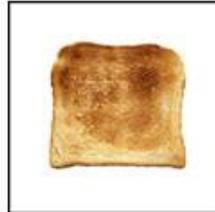


A8



Relevance pictures

R1



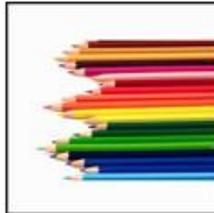
R2



R3



R4



R5



R6



R7

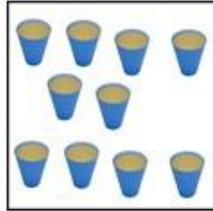


R8



Scalar pictures

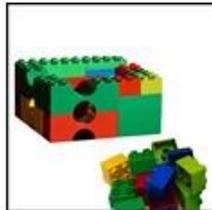
S1



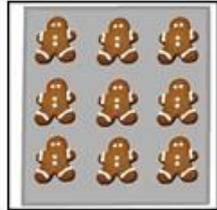
S2



S3



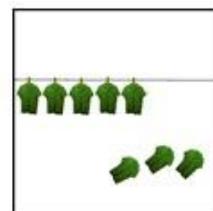
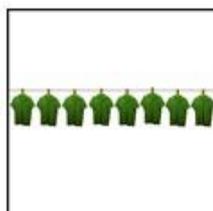
S4



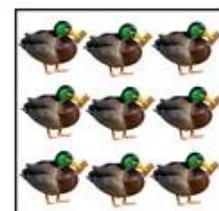
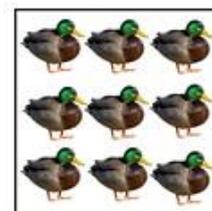
S5



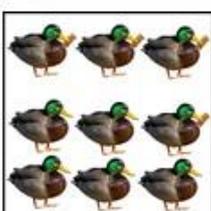
S6



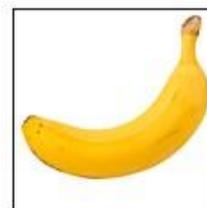
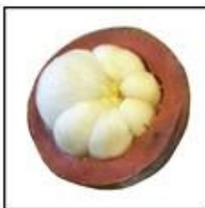
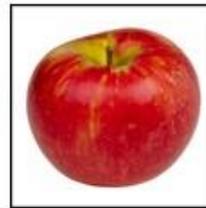
S7



S8



WLE pictures



8.2 Alternative analysis for Experiment 1A

An alternative strategy suggested by Barr et al. (2013) in the face of nonconvergence, is to do a separate analysis for each fixed effect with the corresponding random slope. For instance, here to test the effect of condition (control vs critical) a model is fitted with all fixed effects (condition, type and age group), but only random slopes for condition (item by condition, and subject by condition). Repeating this for type and age group as well (see Tables 2.6, 2.7, 2.8), the above results are confirmed: there are main effects of condition, of the youngest age group, and of scalars, and marginal main effects of ad hoc and relevance inferences as well.

	Estimate	SE	z	p
Intercept	3.18	.23	13.92	< .001
Control	.66	.16	4.05	< .001
Ad Hoc	.54	.24	2.25	.024
Relevance	-0.25	.23	-1.06	.29
Scalar	-1.42	.22	-6.48	< .001
2;8-3;11	-1.00	.21	-4.67	< .001
4;0-4;11	.06	.22	.26	.79

Table 8.2 *Response ~ Condition + Type + Age group + (1 + Critical | Item) + (1 + Critical | Subject)*
Glmer, family = binomial, optimizer = bobyqa, sum coding

	Estimate	SE	z	p
Intercept	3.20	.25	12.61	< .001
Control	.58	.08	7.09	< .001
Ad Hoc	.39	.29	1.37	.17
Relevance	-0.26	.23	-1.16	.25
Scalar	-1.20	.25	-4.82	< .001
2;8-3;11	-1.16	.26	-4.44	< .001
4;0-4;11	.03	.24	-0.14	.89

Table 8.3 *Response ~ Condition + Type + Age group + (1 + Age group | Item) + (1 + Subject)*
Glmer, family = binomial, optimizer = bobyqa, sum coding

	Estimate	SE	z	p
Intercept	3.63	.37	9.91	< .001
Control	.59	.08	7.14	< .001
Ad Hoc	2.25	.93	2.41	.016
Relevance	-0.83	.39	-2.10	.036
Scalar	-2.03	.39	-5.18	< .001
2;8-3;11	-0.80	.22	3.55	.00038
4;0-4;11	.01	.21	.06	.96

Table 8.4 *Response ~ Condition + Type + Age group + (1 + Item) + (1 + Type | Subject)*
Glmer, family = binomial, optimizer = bobyqa, sum coding

8.3 Background questionnaire for Experiment 1A and 1B

Pragmatic development study

Background Questionnaire



1. General information

a. Please can you provide:

NB question 1a is just used to link this questionnaire to the data from your child's study and the consent form; all data will be held anonymously and securely.

Your child's name	
Your child's date of birth	
Your name	
Your relationship to child	

b. Where was your child born? _____

c. If outside UK, when did the child move to UK? _____

d. Does your child speak any language(s) other than English? YES / NO

If NO, please proceed to question 3. If YES, please continue to question 2.

2. Language information

a. Which language(s) can your child speak other than English?

Language A	
Language B	
Language C	

b. How often is your child exposed to each language(s)? please tick

	Never	Rarely	Sometimes	Usually	Always
English					
Language A					
Language B					
Language C					

c. At what age did your child start receiving consistent and significant exposure to each of his/her languages?

	Age in months
English	
Language A	
Language B	
Language C	

consistent and significant: the child started attending a school (e.g., primary school / kindergarten / daycare) where instruction was held in that language, or one of the main caregivers (e.g., babysitter, parents) of the child started to consistently use that language with the child.

d. How does your child use their language(s) with you, their parents or guardians?

Languages used with and by the child's MOTHER or GUARDIAN 1, if applicable

Language MOTHER / GUARDIAN 1 uses with CHILD						Language CHILD uses with MOTHER / GUARDIAN 1				
	Never	Rarely	Some-times	Usually	Always	Never	Rarely	Some-times	Usually	Always
En										
A										
B										
C										

Languages used with and by the child's FATHER or GUARDIAN 2, if applicable

Language FATHER / GUARDIAN 2 uses with CHILD						Language CHILD uses with FATHER / GUARDIAN 2				
	Never	Rarely	Some-times	Usually	Always	Never	Rarely	Some-times	Usually	Always
En										
A										
B										
C										

e. For each language in any given situation, do you tend to use that language alone, or mix it with another language? (i.e., would you use English and Language A/B/C words in the same sentence? NB this 'code-switching' is a normal aspect of being multilingual)

		Amount of code-switching MOTHER / GUARDIAN 1 uses with CHILD					Amount of code-switching CHILD uses with MOTHER / GUARDIAN 1				
	with language (En/A/B/C):	Never	Rarely	Some-times	Usually	Always	Never	Rarely	Some-times	Usually	Always
En											
A											
B											
C											

		Amount of code-switching FATHER / GUARDIAN 2 uses with CHILD					Amount of code-switching CHILD uses with FATHER / GUARDIAN 2				
	with language (En/A/B/C):	Never	Rarely	Some-times	Usually	Always	Never	Rarely	Some-times	Usually	Always
En											
A											
B											
C											

f. Does another adult, besides the child's parents, regularly take care of your child at home (e.g. grandparent, babysitter)? YES / NO

If yes, does this person speak any other languages with your child? YES / NO

Languages used with and by the child's CAREGIVER

Use additional tables in Appendix if more adults regularly take care of the child.

Language OTHER REGULAR CAREGIVER uses with CHILD						Language CHILD uses with OTHER REGULAR CAREGIVER				
	Never	Rarely	Sometimes	Usually	Always	Never	Rarely	Sometimes	Usually	Always
En										
A										
B										
C										

g. If the child has any siblings, which language(s) do they speak together? For each child in the family, complete a separate table. Use additional tables in Appendix if necessary.

Language SIBLING 1 uses with CHILD						Language CHILD uses with SIBLING 1				
	Never	Rarely	Sometimes	Usually	Always	Never	Rarely	Sometimes	Usually	Always
En										
A										
B										
C										

Language SIBLING 2 uses with CHILD						Language CHILD uses with SIBLING 2				
	Never	Rarely	Sometimes	Usually	Always	Never	Rarely	Sometimes	Usually	Always
En										
A										
B										
C										

h. Does your child attend a complementary school / club in their language(s)?

Language A	YES / NO
Language B	YES / NO
Language C	YES / NO

i. How fluent would you say your child is in his/her languages?

	Not fluent	Limited fluency	Somewhat fluent	Quite fluent	Very fluent
	Little understanding or speaking ability	Some understanding; can say short, simple sentences e.g. answer the phone, or greet a neighbour	Good understanding; can express him/herself on many topics e.g. explain what he/she wants; give instructions	Can understand and use the language adequately for most situations e.g. follow films or TV shows	Understands almost everything. Very comfortable expressing him/herself in the language in all situations
En					
A					
B					
C					

3. Language learning

a. Has your child been diagnosed with any language impairments or learning difficulties?
YES / NO

4. General Information

Please circle your answer.

a. Does your family have a car?	No	Yes, just one	Yes, more than one	
b. During the past year, how many times did you travel away on holiday as a family?	None	One	Two	More than two
c. How many computers (laptops or desktop) does your family own?	None	One	Two	More than two
d. Does your child have his/her own bedroom?	Yes	No		

e. What is your highest level of education? Please circle your answer.

YOUR EDUCATION	ANY OTHER PARENT OR GUARDIAN'S EDUCATION (if applicable) Please state relation to child _____
Primary School	Primary School
Secondary School	Secondary School
College	College
University – degree	University – degree
University – masters	University – masters
University – PhD	University – PhD

Thank you very much for taking the time to fill in this questionnaire!

Please return it to your child's school.

If you have any questions, please do not hesitate to be in contact with Elspeth Wilson
ep321@cam.ac.uk 07792 354633
 Department of Theoretical and Applied Linguistics, English Faculty, 9 West Rd, CB3 9DP

All information provided in this questionnaire will be treated confidentially, held securely, and, once linked to your child's study data, stored only with an anonymised ID number.

Appendix

	Language CAREGIVER/SIBLING (delete as appropriate) uses with CHILD					Language CHILD uses with CAREGIVER/SIBLING				
	Never	Rarely	Some-times	Usually	Always	Never	Rarely	Some-times	Usually	Always
En										
A										
B										
C										

8.4 Instructions for Theory of Mind tasks

Sally-Anne task

I'm going to tell you a story. Listen and watch carefully.

This is Sally. This is Anne.

Sally and Anne are playing with their marble.

Sally puts the marble in the box. Then she goes outside, where she can't see or hear anything that's going on here.

Then Anne gets the marble out of the box and puts it in the bucket. Then she takes it out of the bucket and out of the room.

Where did Sally put the marble at the beginning?

Where is the marble now?

Here comes Sally. Sally wants to play with the marble.

Where will she look first for the marble?

Smarties task

Look what I've got here.

What do you think is in the box?

Let's have a look. [open]

What's in there? Oh! Pencils!

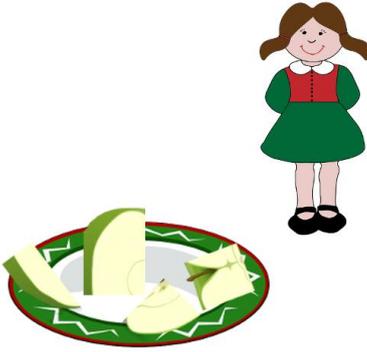
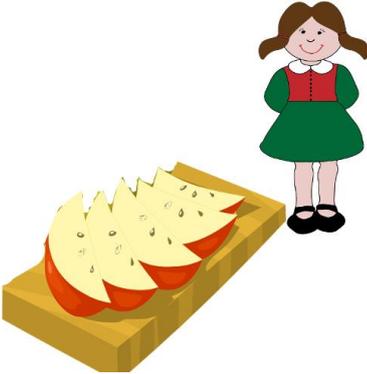
Let's close it up.

At the beginning, when you first saw the box closed up, what did you think was inside the box?

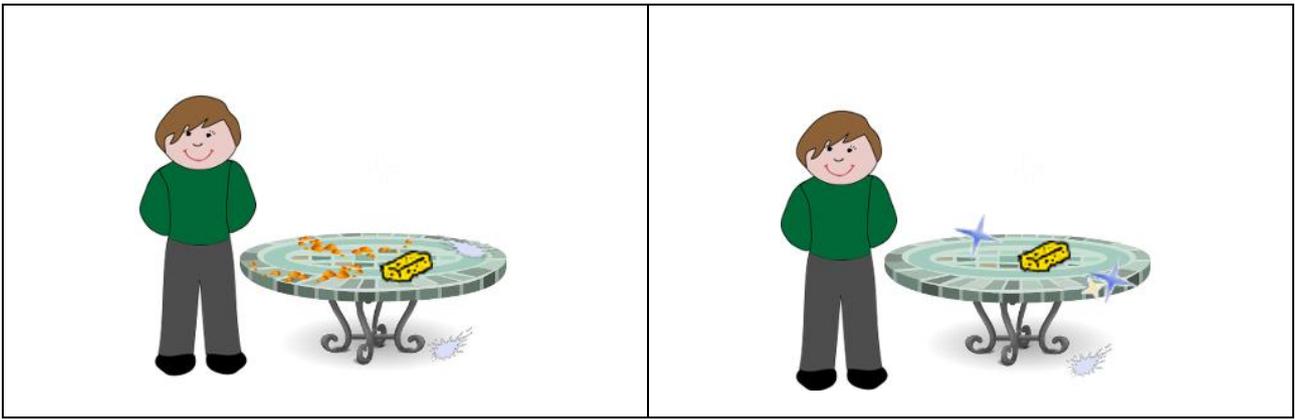
What's really inside the box?

Next I'm going to show this box to your friend X. What do you think X will think is inside the box?

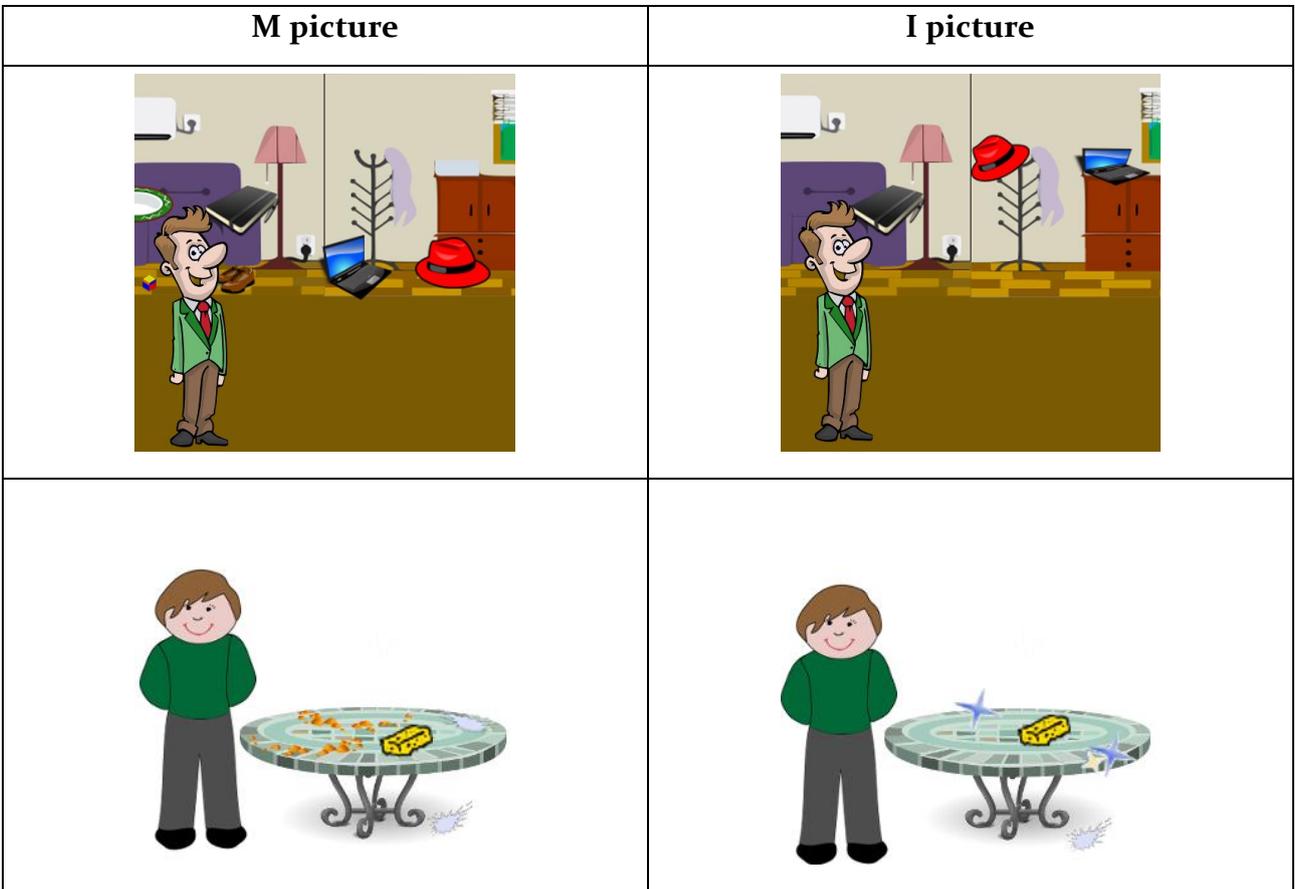
8.5 Picture stimuli for Experiment 2A

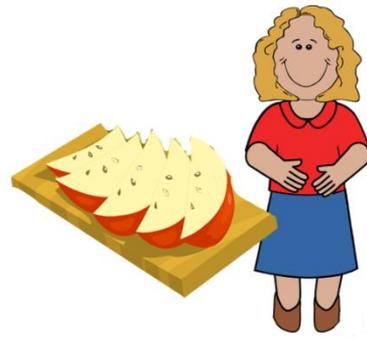
M picture	I picture
	
	
	
	





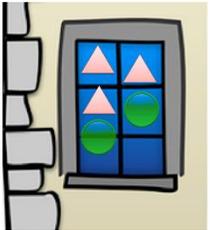
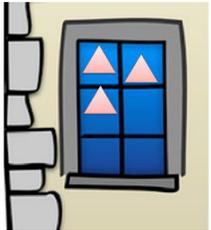
8.6 Picture stimuli for Experiment 2B







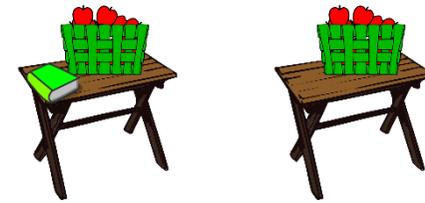
8.7 Stimuli for Experiment 3

Question	Utterance – ‘strong’	Utterance – weak	Visible picture – ‘strong’	Visible picture – ‘weak’
What was on the rug?	There was a bowl of strawberries and a jelly.	There was a bowl of strawberries.		
What did the friends drink?	All of the friends drank lemonade.	Some of the friends drank lemonade.		
What was on the window?	There were triangles and circles.	There were triangles.		
What did the friends eat?	All of the friends ate a sandwich.	Some of the friends ate a sandwich.		

What was on the table?

There was a basket of apples and a book.

There was a basket of apples.



What did the friends get?

All of the friends got a balloon.

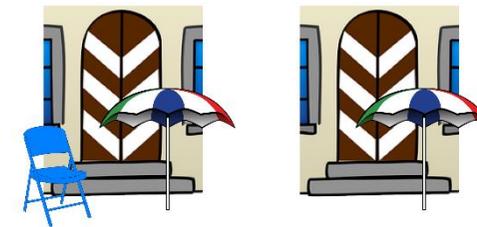
Some of the friends got a balloon.



What was by the door?

There was an umbrella and a chair.

There was an umbrella.



How did the friends get there?

All of the friends came on a bicycle.

Some of the friends came on a bicycle.



What did the friends wear?

All of the friends wore a party hat.

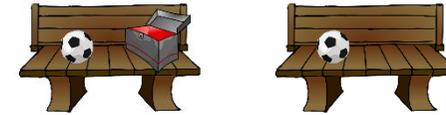
Some of the friends wore a party hat.



What was on the bench?

There was a ball and a box.

There was a ball.



What did the friends bring?

All of the friends brought a present.

Some of the friends brought a present.



What was on the windowsill?

There was a plant and a plate of snacks.

There was a plant.



What did the friends eat for pudding?

All of the friends ate an icecream.

Some of the friends ate an icecream.



What was on the stool?

There was a jug of juice
and a bag of crisps.

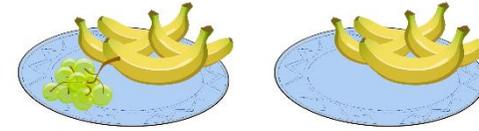
There was a jug of juice.



What was on the plate?

There were bananas and
grapes.

There were bananas.



What did the friends do?

All of the friends blew
bubbles.

Some of the friends blew
bubbles.



8.8 Stimuli for Experiment 4

Utterance	Display	
Pick the card with bananas	pears	bananas
	pears/bananas	oranges
Pick the card with pears	pears	bananas
	pears/bananas	apples
Pick the card with apples	pears	bananas
	pears/bananas	apples
Pick the card with oranges	oranges	bananas
	pears/bananas	oranges
Pick the card with dogs	pigs	hens
	dogs	pig/hen
Pick the card with hens	pigs	hens
	dogs	pig/hen
Pick the card with pigs	pigs	hens
	horses	pig/hen
Pick the card with horses	pigs	horses
	horses	pig/hen
Pick the card with hats	hat / shoe	jumper
	hat	shoe
Pick the card with jumpers	hat / shoe	jumper
	hat	shoe
Pick the card with t-shirts	hat / shoe	tshirt
	tshirt	shoe
Pick the card with shoes	hat / shoe	tshirt
	hat	shoe
Pick the card with tomatoes	tomatoes	carrots
	carrot/potato	tomatoes
Pick the card with cucumbers	cucumber	carrots
	carrot/potato	potato
Pick the card with potatoes	cucumber	carrots
	carrot/potato	potato
Pick the card with carrots	tomato	carrots

	carrot/potato	potato
Pick the card with cats	rabbit	rabbit/cat
	cat	cow
Pick the card with cows	cow	rabbit/cat
	cat	cow
Pick the card with rabbits	rabbit	rabbit/cat
	cat	duck
Pick the card with ducks	rabbit	rabbit/cat
	cat	duck
Pick the card with butterflies	caterpillars	bees
	butterfly	bee/caterpillar
Pick the card with bees	caterpillars	bees
	butterfly	bee/caterpillar
Pick the card with spiders	caterpillars	spider
	spider	bee/caterpillar
Pick the card with caterpillars	caterpillars	bees
	spider	bee/caterpillar