ULSTER UNIVERSITY

Ambiguous questions and perfectible conditionals: the perspective from language acquisition

by

Federica Di Bacco (BA, MSc)

A thesis submitted for the
degree of Doctor of Philosophy

in the
School of Communication
Faculty of Arts, Humanities and Social Sciences
Ulster University

October 2018

I confirm that the word count of this thesis is less than 100.000 words.

Contents

Acknowledgements **iv**

Abstract **v**

1 Introduction **1**
1.1 Preliminaries: the semantics of questions . . . 3
1.1.1 Questions as sets of propositions: Hamblin/Karttunen approach . 4
1.1.2 Questions as partitions: Groenendijk & Stokhof . . . 6
1.1.3 Summary . . . 8
1.2 Questions and discourse coherence . . . 8
1.2.1 Implications for focus . . . 8
1.2.2 Questions organising discourse . . . 10
1.3 The role and effect of the QUD . . . 18
1.3.1 The role of the QUD for scope interaction . . . 19
1.3.2 The role of the QUD for Conditional Perfection . . . 22
1.4 Scalar Implicatures and their derivation . . . 24
1.4.1 Grammatical approach . . . 27
1.5 Contribution of this thesis . . . 29

2 Scope literature review **32**
2.1 An introduction on scope ambiguity . . . 33
2.2 Childen and scope ambiguity: Quantifier - Negation studies . . . 35
2.2.1 The role of the context for scope assignment . . . 37
2.2.2 The Question and Answer Requirement . . . 40
2.2.3 Pragmatic factors and the effect of priming . . . 42
2.2.4 Modal verbs and scope ambiguity . . . 44
2.2.5 Modal verbs and the QUD . . . 46
2.2.6 Summary . . . 49
2.3 The scope of doubly quantified sentences . . . 50
2.3.1 A hierarchy for scope taking operators . . . 51
2.3.2 Processing and scope assignment . . . 53
2.3.3 A mental model for quantifiers . . . 57
2.3.4 Summary . . . 59
2.4 Adults' interpretation of scope ambiguity: experimental studies . . . 59
2.4.1 Sentence disambiguation and effects on the reading times . . . 60
2.4.2 The effect of the context for doubly quantified sentences . . . 64

2.4.3 Summary . . . 69
2.5 Summary of chapter . . . 70

3 Scope experiments **72**
3.1 Experimental studies . . . 73
3.2 Experiment 1: Ambiguous doubly quantified sentences . . . 76
3.2.1 Predictions . . . 77
3.2.2 Methodology . . . 79
3.2.2.1 Participants . . . 79
3.2.2.2 Materials and procedure . . . 79
3.2.3 Results and discussion . . . 83
3.3 Experiment 2: Ambiguous questions . . . 86
3.3.1 Predictions . . . 87
3.3.2 Methodology . . . 88
3.3.2.1 Participants . . . 88
3.3.2.2 Materials and procedure . . . 88
3.3.3 Results and discussion . . . 88
3.4 Summary of chapter . . . 94

4 Conditional Perfection literature **96**
4.1 Conditional meaning . . . 98
4.1.1 The classic account . . . 98
4.1.2 Strict conditionals and possible world semantics . . . 100
4.1.3 *If* as an adverbial restrictor . . . 101
4.1.4 Summary . . . 103
4.2 Conditional perfection . . . 103
4.2.1 The nature of Conditional Perfection . . . 108
4.2.2 CP as a Scalar Implicature . . . 108
4.2.3 CP and exhaustification . . . 114
4.2.4 Summary . . . 116
4.3 Reasoning on conditionals: Mental logic vs Mental model . . . 117
4.3.1 The Mental logic program . . . 117
4.3.2 Mental model conditionals . . . 119
4.3.3 Summary . . . 121
4.4 Experiments on the comprehension of conditionals . . . 121
4.4.1 Truth conditional experiments . . . 122
4.4.2 Experiments on conditional inferences . . . 123
4.4.3 Experiments on the role of the QUD for CP . . . 128
4.4.4 Summary . . . 135
4.5 The development of conditional reasoning . . . 136
4.5.1 Experimental studies on children . . . 137
4.5.2 Children's behaviour with disjunction and SIs . . . 140
4.5.3 Conjunctive interpretation of *or*: Alternative based approach . . . 147
4.5.4 Summary . . . 150
4.6 Summary of chapter . . . 151

5 Conditional Perfection experiments **153**

5.1 Experimental studies . . . 157
5.2 Experiment 1 . . . 157
5.2.1 Predictions . . . 157
5.2.2 Methodology . . . 159
5.2.2.1 Participants . . . 159
5.2.2.2 Materials and procedure . . . 159
5.2.3 Results and discussion . . . 165
5.3 Experiment 2 . . . 167
5.3.1 Predictions . . . 167
5.3.2 Methodology . . . 167
5.3.2.1 Participants . . . 167
5.3.2.2 Materials and procedure . . . 168
5.3.3 Results and discussion . . . 168
5.4 Summary of chapter . . . 180

6 Conclusions **182**

A Scope experiment items **189**

B Conditional experiment items **194**

Acknowledgements

It would not have been possible to complete this study without the help and support of the people around me.

First and foremost, I am extremely grateful to my supervisory team, Jacopo Romoli, Raffaella Folli and Christina Sevdali, for their guidance, advice and support. Without them this thesis would not have been possible.

I would also like to thank Lyn Tieu and Vincenzo Moscati for their help and discussion at various stages of this project.

I am lucky to have a number of friends and colleagues who have helped me with a variety of aspects of my experimental work: in particular, I would like to thank Lynda Kennedy, Frances Kane, Athanasia Asyllogistou, Sofja Volkova, Sean Fox and all the people in 2D02. Thanks to Dorothy Ahn for allowing me to use her drawings for the stimuli of my experiments. I would also like to thank the children, parents and personnel of the schools that have been involved in this research project for their time and helpfulness in facilitating my data collection. Last but not least, I would like to thank my family for their continuous encouragement and support throughout my studies. And to Paolo, thanks for always being there for me and keeping me together through this!

Abstract

This thesis investigates two phenomena - scope ambiguity resolution and conditional perfection, from the point of view of language acquisition. An empirical study was conducted to determine how scopally ambiguous sentences and perfectible conditionals are interpreted by children and adults, and if there are any differences between the two groups. In relation to the first, two experiments conducted on children's and adults' interpretation of scopally ambiguous declarative sentences and questions have shown that children are adult-like in their ability to access inverse scope. This is in contrast with the view that sees children as unable to obtain inverse scope, and it is instead compatible with the QUD approach. I propose an extension of the QUD approach to include questions. In the case of conditionals, two experiments were conducted to test the theory that conditional perfection is a scalar implicature: if it is so, children should obtain this inference less often than adults. However, the results show that both children and adults obtain a conjunctive-like reading for these kind of statements. This reading is widely reported in children, and one theory attributes it to their inability to construct the meaning of a conditional in their mind. The conjunctive-like reading is less frequent in adults, but it is possible that pragmatic factors have been reported for the high rate of occurrence observed in this study, as adults' interpretation of conditionals is reported to be influenced by the task. I also sketch another explanation of the results, based on the idea that both the conjunctive inference and conditional perfection can be derived as scalar implicatures.

I hereby declare that with effect from the date on which the thesis is deposited in Ulster University Doctoral College, I permit

1. the Librarian of the University to allow the thesis to be copied in whole or in part without reference to me on the understanding that such authority applies to the provision of single copies made for study purposes or for inclusion within the stock of another library.

2. the thesis to be made available through the Ulster Institutional Repository and/or EThOS under the terms of the Ulster eTheses Deposit Agreement which I have signed.

Chapter 1

Introduction

The research presented in this dissertation consists of a set of language acquisition experiments aimed at investigating the interaction of grammar and pragmatic abilities in two domains: scope ambiguity resolution, and the inferences associated with conditional sentences.

Both domains are widely thought to be affected by pragmatic factors, and in particular the effect of the Question Under Discussion (QUD), a notion introduced in formal pragmatics by Roberts (1996). The discourse model she proposed, building on Lewis (1979) and Stalnaker (1978), is based on the idea that discourse is a sort of game based on the ultimate aim of sharing information between the interlocutors, who are at any time committed to answer an implicit QUD. For this reason, in Roberts's model, every assertion is always to be understood as an answer to the current QUD. This idea became prominent in the field of language acquisition, and in particular for what relates to scope ambiguity resolution: it is well known that children often differ in their interpretation of scopally ambiguous sentences when compared to adults (Musolino, 1998), and initial explanations of this difference had focused on the possibility that children and adults might differ in their grammatical abilities (Musolino 1998, see also Musolino, Crain, and Thornton 2000, Lidz and Musolino 2002). Subsequenft research showed how children were in fact able to access adult-like interpretation, provided that the right QUD is made relevant by the context (Gualmini, 2004a, Krämer, 2000, Gualmini, Hulsey, Hacquard, and Fox, 2008). Proponents of this approach claimed that previous results can be explained in light of the fact that children are only able to access the interpretation that

answers the QUD, while adults can overcome its effect and chose the interpretation that makes the sentence true: it follows from this that, under this approach, the management of the QUD is thought to be the only difference between the two groups (Gualmini et al., 2008).

Similarly, other phenomena have been shown to be sensitive to the influence of the QUD: for example, Scalar Implicatures (SIs) (Zondervan, Meroni, and Gualmini, 2008, Zondervan, 2011) and Conditional Perfection (CP) (Von Fintel, 2001, Farr, 2011), itself considered a SI by some scholars (Horn, 2000, Atlas and Levinson, 1981, van der Auwera, 1997). While CP is recognised as a very common inference, it has been noted (Von Fintel, 2001) and demonstrated experimentally (Farr, 2011) that it occurs much more frequently when the conditional sentence is understood as an answer to a question on the consequent of the conditional. The effect of the QUD on conditionals, however, has not been investigated in children yet: children's interpretation of conditionals has mostly been investigated in psychology, and several experimental studies have evidenced that children generally obtain a conjunctive-like interpretation of conditional sentences, considering them true only when both propositions are true (Paris, 1973, Barrouillet and Lecas, 1998, Barrouillet, Grosset, and Lecas, 2000, Barrouillet et al., 2000, Barrouillet, Gauffroy, and Lecas, 2008), similarly to what is observed for sentences with disjunction (Paris, 1973, Braine and Rumain, 1981). Disjunction sentences have received a great deal of attention in language acquisition, being the focus of several experimental studies aiming at investigating this phenomenon (Gualmini, Crain, Meroni, Chierchia, and Guasti, 2001, Chierchia, Guasti, Gualmini, Meroni, Crain, and Foppolo, 2004) . Children's interpretation of conditionals on the other hand has not received an equal degree of attention in language acquisition literature. Moreover, as mentioned before, a notable theory sees CP as a scalar implicature (Horn, 2000, Atlas and Levinson, 1981, van der Auwera, 1997), which in turn it might mean that perfectible conditionals are not interpreted in the same way by adults and children, as it is often observed that the latter compute SIs at a much lower rate (Noveck, 2001, Papafragou and Musolino, 2003, Guasti, Chierchia, Crain, Foppolo, Gualmini, and Meroni, 2005, Foppolo, Guasti, and Chierchia, 2012, a.o.). Therefore, as the nature of CP is debated, the language acquisition approach might prove helpful in investigating the claim that this phenomenon is indeed a SI. Another point of interested is to investigate whether a parallel can be drawn between the conjunctive-like interpretation of disjunctions and that of the corresponding

interpretation of conditionals.

This thesis has two main aims: in relation to scope ambiguity, the experimental study aims at investigating the claims made by the Question and Answer Requirement (QAR) for scope assignment by testing the interpretation of an ambiguous question directly, as the effect of the QUD has until now only been investigated through the interpretation of declarative sentences. In relation to CP, this thesis aims to investigate the phenomenon through the lens of language acquisition. While the ultimate goal is to investigate the effect of the QUD on CP on children, there are currently a limited number of language acquisition studies on conditionals, as this topic has mostly been investigated in psychology, as researchers have been concerned with conditional reasoning (Barrouillet and Lecas, 1998, Barrouillet et al., 2000, 2008, Gauffroy and Barrouillet, 2009). The experimental studies presented in the next chapters therefore focus on establishing if children's interpretation of a conditional differ from adults, and in particular if the latter obtain more biconditional readings, as it would be predicted if CP is indeed to be regarded as a scalar inference.

The rest of this chapter introduces some of the concepts that will be relevant for the literature review and experimental work presented in the next section, and it is organised as follows: section 1.1 briefly introduces an account of the semantics of questions, section 1.2 focuses on the role of questions for discourse coherence, drawing on Roberts's (1996) and Büring's (2003) discourse models. Section 1.3 aims to present the role of the QUD for the resolution of scope ambiguity and for the occurrence of CP. The next section 1.4 moves onto the derivation of Scalar Implicatures, providing an account of how they are derived, as this will be particularly relevant for the study presented in Chapter 5.

1.1 Preliminaries: the semantics of questions

This section presents a brief summary of some influential accounts on the semantics of questions, which will be needed to understand the model of discourse that is relevant for the QUD approach. QUD-based theories have their foundation in question-answer relation accounts that take the meaning of a question to be the set of its answers, starting with Hamblin (1973).

1.1.1 Questions as sets of propositions: Hamblin/Karttunen approach

Hamblin's (1973) account was one of the first attempts at formalising the semantics of question. His approach is based on his previous work (Hamblin, 1958), and in particular on some postulates on the notions of questions and answers he had proposed.

(1) a. An answer to a question is a statement.
b. Knowing what counts as an answer is equivalent to knowing the question.
(Hamblin, 1958, p. 162)

The second postulate in (1-b) is the foundation of his approach to questions. Indeed, the idea that knowing what counts as an answer counts as knowing the question was later refined into the main notion of Hamblin's (1973) semantics of questions: the notion that the meaning of a question is defined by their set of possible answers. As Hamblin puts it, "pragmatically speaking, a question sets up a choice situation between a set of propositions, namely, those propositions that count as answers to it"(Hamblin, 1973, p. 48). That is to say, each question defines a set: the set of the propositions which constitute the possible answers to that question, irrespective of whether they are true or false. For example, in the case of (2), a *wh*-question, the set of its possible answers corresponds to the set of propositions in the form *x speaks German*. In the case of (3), a *yes-no* question, the set of its possible answers contains only two members, the proposition that Bill speaks German, and the proposition that Bill does not speak German.

(2) *Wh-questions*:
Who speaks German? =
{Bill speaks German, Sam speaks German, Mary speaks german, ... } =
λp[$\exists$x(p = speaks(x, German))] =
{speaks(b,g), speaks(s,g), speaks(m,g), ... }

(3) *Yes-no questions*:
Does Bill speak German? =
{Bill speaks German, Bill doesn't speak German} =
λp[(p = speaks(b,g)) $\vee$ (p = $\neg$speaks(b,g))] =
{speaks(b,g), $\neg$speaks(b,g)}

Hamblin's analysis dealt with root questions, and it is silent in relation to embedded questions. Karttunen (1977) proposed a slightly amended version of Hamblin's model, and he suggested analysing direct questions as containing a cover performative verb: therefore, the question in (2) would be analysed as *I ask you to tell me who speaks German.* Another point that Karttunen raises, concerning question embedding, regards questions that are embedded under verbs such as *tell*, *know*, *indicate* etc. Considering the example in (4), it can be observed how (4-b), but not (4-a), presupposes that John told the truth.

(4) a. John told Mary that Bill and Susan passed the test.
b. John told Mary who passed the test.

Karttunen claims that having the embedded question *who passed the test* denote the set of all possible answers, true and false, it is not possible to explicate the meaning of *tell* in a straightforward way. On the other hand, if a question is taken to denote the set of all true answers, instead of the set of all possible answers, greatly simplifies the analysis of such sentences. For this reason, Karttunen departs from Hamblin in considering the meaning of a question to be the set of all its true answers, instead of the set of all possible answers. That is to say, supposing the set of speakers of German to include Bill, Frank and Sam, we obtain the following meaning for the question in (3), repeated below.

(5) Who speaks German? =
{Bill speaks German, Frank speaks German, Sam speaks German}

Karttunen also notes that each of the propositions in the set is a partial true answer, while the conjunction of all propositions in the set constitutes a complete true answer to the question. However, while a possible semantic ambiguity between questions requiring a partial answers and questions requiring a complete answer had been proposed in earlier literature (Hintikka, 1976), Karttunen claims that the distinction between the two possibilities is purely pragmatic: questions are interpreted as requiring a complete answers, and the evidence is in the fact that a sentence like (6) is regarded as contradictory.

(6) John remembers who came to the party, although he doesn't remember that Mary came.

Partial answers arise from pragmatics, in circumstances in which an exhaustive answers is not expected - for example, as signalled by the expression *for instance* in (7). The concepts of complete and partial answer has been further refined by Groenendijk and Stokhof (1984), as it will be explained in the next section.

(7) Who, for instance, came to the party last night?

1.1.2 Questions as partitions: Groenendijk & Stokhof

Groenendijk and Stokhof (1984, 1982) take another step towards separating the semantics and the pragmatics of questions, and their approach is regarded as particularly influential for what concerns the distinction between a partial and a complete answer to a question. In particular, they claim that the amendment proposed by Karttunen (1977) to Hamblin (1973) is insufficient: knowing the answer to (5) means knowing, of the people that speak German, that is true that they speak German. However, Groenendijk & Stokhof observed that our intuitions are generally stronger than that: for one person to the answer to *who speaks German?* we would expect that person to know that Bill, Frank and Sam speak German; and not believe that anyone else does. Therefore, they propose a stronger degree of exhaustivity compared to Karttunen. To account for these intuitions, they propose an analysis based on the idea that a question identifies a partition of the set of possible worlds into subset of worlds where the answer to that question is the same.

Yes-no questions divide the set of possible worlds into two partitions:

(8) Did John come?

John came.
John didn't come.

For a wh-question, each partition contains a complete answer to the question: supposing that our universe comprises 3 individuals (Sam, Mary, Bill), the cells of the partitions for the question *who came to the party?*

(9) Who came to the party?

No one came to the party.
Sam came to the party.
Mary came to the party.
Bill came to the party.
Sam and Mary came to the party.
Bill and Mary came to the party.
Sam and Bill came to the party.
Mary and Bill came to the party.
Sam, Mary and Bill came to the party.

It should be noted that, since each cell of the partition represents a complete answer to the question, the content of the cells should be interpreted as "Sam and no one else came" and so on. In this model, giving a complete answer to a question means selecting the cell of the partition which contains the actual world. Partial answers are still possible: an answer like *Sam came, but I am not sure about Mary and Bill* does not select any partition, however it provides pertinent information, and it rules out some of the partitions. A subsequent account, proposed by Heim (1994), solves the issue of strong vs. weak exhaustivity outlined before by proposing that questions are ambiguous between their strongly exhaustive and weakly exhaustive reading: the weakly exhaustive one corresponds to Karttunen's (1977) semantics (knowing who came to the party corresponds to knowing that Sam, Mary and Bill came to the party), and the strongly exhaustive one corresponds to Groenendijk and Stokhof's (1984) semantics (knowing who came to the party corresponds to knowing that Sam, Mary, Bill and no one else came to the party). In the case of embedded questions, each different verb might then select a different reading between the two. Much subsequent research has been devoted to explaining the nature of this ambiguity: for example, a number of recent theories have proposed that strong readings are not part of the meaning of a question, but they arise as an implicature (Berman, 1991, Klinedinst and Rothschild, 2011, Uegaki et al., 2015, Nicolae, 2013, George, 2011).

1.1.3 Summary

The semantics of question is a complex topic, and an exhaustive account of it is beyond the scope of this dissertation: the main relevant notion for the rest of this chapter is that the idea, initially proposed by Hamblin (1973), that the meaning of a question corresponds to the set of their answers is a well established one, and it has been the basis for several subsequent theories, which have aimed at investigating how embedded questions combine with verbs (Karttunen, 1977) and what constitutes an exhaustive answer (Groenendijk and Stokhof, 1984, Heim, 1994). For what is relevant for this dissertation, I will assume that questions denote the set of their true answers. The next section moves onto the relevance of question-answer relations for discourse.

1.2 Questions and discourse coherence

Question-answer relations have often been identified as part of the mechanism regulating discourse coherence, determining the set of alternatives under consideration, which in turn has an implication for focus and for relevance. In particular, the QUD model (Roberts, 1996, Büring, 2003) sees discourse as being entirely organised around overt or covert questions which are continually raised by interlocutors. This section discusses the role of questions for discourse coherence, focusing in particular on the relevance of the QUD.

1.2.1 Implications for focus

Rooth (1985, 1992) adopts a model similar to Hamblin's for questions in his work on focus. His main assumption is that each expression has two different semantic values, an ordinary one (referred to as $[\![a]\!]^{\circ}$) and a focus semantic value, $[\![a]\!]^{F}$. This has en effect on a variety of phenomena, including scalar implicatures and question-answer congruence. In relation to the latter, Rooth (1992) considers the following question-answer pairs.

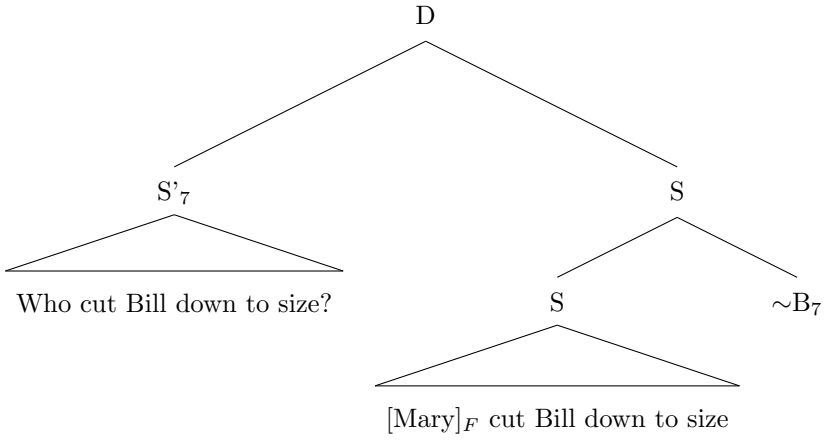
(10) a. Who cut Bill down to size?
b. $[\text{Mary}]_F$ cut Bill down to size.

(11) a. Who did Mary cut down to size?

b. Mary cut [Bill]$_F$ down to size.

While answers (10-b) and (11-b) are identical, apart from focus, they are not interchangeable in conversation; and answering one question with the other's answer would result in an incongruent exchange. Similarly to Hamblin (1973), Rooth (1992) notes that a question selects a set of potential answers, and the role of focus is to signal that the alternatives in the set are relevant. The mismatch between (11-b) as an answer of (10-a) and vice versa is due to the fact that the two questions denote different sets: the propositions in the first would be in the form *x cut Bill down to size*, while the ones in the other set are in the form *Mary cut x down to size*. These alternatives are evoked by the presence of a covert focus interpretation operator, referred to as ∼, as it can be seen in the representation below for (10).

(12)



The presence of focus introduces the variable B_7 in (12), which is identified with the ordinary semantic value of the question, i.e. the set of its possible answers, which includes the proposition *Mary cut Bill down to size* and other propositions in the form *x cut Bill down to size*. Trying to answer (11-a) with (12-b) would result in a mismatch, as it is a sentence in the form *Mary cut y down to size*, and therefore not a member of the set of salient alternatives denoted by B_7.

1.2.2 Questions organising discourse

Roberts (1996) devised a model of discourse organised around questions and answers. She borrows from Stalnaker (1978) the idea that discourse is akin to a game, organised around one main goal - discovering and sharing with other interlocutors information about our world. Using Stalnaker's notions of *common ground* (the set of propositions taken for granted by participants at any point in a discourse) and *context set* (the set of possible worlds compatible with the propositions in the common ground), Roberts claims that participants work together to reduce the context set to a singleton set: the actual world.

To reach this goal, interlocutors make a number of moves, which can be divided into two types, *questions* and *answers* (also called respectively *set up moves* and *payoff moves*, using terminology from Carlson (1983)). The interpretation of each move is composed of two aspects: *proffered content* and *presupposed content*. Therefore, in Roberts model, it is assumed that discourse progresses by participants continually raising and answering questions. While the questions are often implicit, every sentence is nonetheless to be taken as the answer to the specific question that participants are committed to answer in that particular moment in conversation: the Question Under Discussion (QUD).

The main goal of the discourse can itself be seen as providing an answer to what Roberts calls "the Big Question": *what's the way things are?*. Since it would be impossible to answer such a question directly, it is indeed possible that the strategy participants adopt is to develop sub-goals which are easier to achieve, and work through them to facilitate achieving the main goal: that is to say, the main goal (the Big Question), which would be too difficult to answer at once, gets broken down into *subquestions*, which are more manageable and easier for participants to answer. Subquestions are such because they are entailed by the superquestion(s) above them - question entailment is defined in Groenendijk and Stokhof (1984, p. 16) as follows: "one interrogative Q1 entails another Q2 iff every proposition that answers Q1 answers Q2 as well". The other side of the coin of this entailment relationship is that every answer to the entailed subquestion provides a partial answers (or eliminates some cells in the partition, in Groenendijk and Stokhof's (1984) words) for the superquestion above it. Hence, the questions participants raise as part of their strategy towards the goal of discourse are not casual, they are always logically related by this relationship of entailment to the Big Question (which entails any

other possible questions) and, more importantly, to any other question that participants have already agreed to answer: a question can only be considered by interlocutors if it is a part of a strategy to answer a superquestion.

In (13), there is an example of *question stack*, indicating a strategy, with its scheme of question relations. Here, (13-a) is a broader superquestion compared to (13-a-i) and (13-a-ii) that sit below it: answering it would also provide an answer to (13-a) and (13-b) and the questions below them. The Big Question *what's the way things are?* is the ultimate superquestion, and it entails every other question.

(13) What's the way things are?

a. Who ate what?

(i) What did Hilary eat?

Did Hilary eat tofu?

Did Hilary eat beans?

(ii) What did John eat?

Did John eat tofu?

Did John eat beans?

From this structure, it can be see intuitively how answering one of the questions that sit above would also provide an answer to any subquestion below it: knowing what Hilary ate would also mean knowing whether she ate tofu or not, and so on. At the same time, it can be seen how answering (13-a-i) or (13-a-ii) provides a partial answer to (13-a) Adopting Hamblin's (1973) semantics for questions, Roberts considers the possible answers which constitute the meaning of a question as proffered alternatives. Roberts calls the set containing all possible answer to a question the set of *Q-alternatives*, and partial answers and complete answers are defined in Roberts (1996) as follows:

(14) A **partial answer** to a question q is a proposition which contextually entails the evaluation – either true or false – of at least one element of Q-alt(q). A **complete answer** is a proposition which contextually entails an evaluation for each element of Q-alt(q).

In the case of our question(13-a-i), repeated below as (15), if we know that Hilary was invited to a party and there was nothing but tofu and beans to eat, we must assume that the meaning of the question are the alternatives in (15-a), and a complete answer would be one that entails one of the possibilities in (15-b).

(15) What did Hilary eat?

a. {Hilary ate tofu, Hilary ate beans}

b. {Hilary ate tofu and beans, Hilary ate tofu and not beans, Hilary ate beans and not tofu, Hilary ate nothing}

Now, we can see how providing an answer to the subquestions *did Hilary eat tofu?* and *did Hilary eat beans?* does address at least one of the possible elements of the Q-alternatives of the superquestion (13-a-i) and therefore answering one of the subquestions does provide a partial answer to its superquestion. Answering both subquestions in turn addresses both elements, and therefore provides a complete answers to (13-a-i).

Roberts claims that entailment relations between questions are exploited as part of the strategy that participants adopt in order to reach the goal of sharing information; that is to say, conversation is organised around a stack of questions like the one in (13-a), but it should also be noted that participants are committed to answering one question at the time (the immediate QUD). A question is removed from the stack once it is resolved (or deemed unanswerable), and new questions can be accepted by participants and enter the stack, provided that the new questions will help answer these that are already present in the stack. Non-questions moves (the assertions) are also influenced by this consideration: if the participants in conversation want to be cooperative, they will strive to make their utterances relevant, i.e. they will make assertions that provide at least a partial answer to the current QUD.

It should be noted however that the notion of answerhood can sometimes be more complicated than a simple entailment relation. Consider the example in (17), and the CG information in (16).

(16) CG(17-a) $\supseteq$ {John is allergic to clams, one will not eat anything one is allergic to, one will eat something unless one has a reason not to}

(17) a. What kind of seafood will John eat?

b. Isn't John allergic to clams?

Answering (17-b) doesn't necessarily yields an answer to (17-a): while it is true that an affirmative answer to (17-b), in conjunction with the information contained in the CG, will entail a partial answer to (17-a), a negative answer to (17-b) won't entail an answer to (17-a), since there might be other possible reasons for John to avoid clams other than being allergic. In cases such as this, Roberts assumes the presence of an implicit bridging question, like the one in (18).

(18) What reason would John have not to eat clams?

Given the assumption present in the CG that one will eat something unless one has a reason not to, an answer to (18) does entail a partial answer to the question above, (17-a), restoring the notion that every move in the game must be *relevant* in relation to the QUD, as formalised in (19) (Roberts, 1996, p. 21).

(19) "A move m is *Relevant* to the question under discussion q, i.e., to last(QUD(m)), iff m either introduces a partial answer to q (m is an assertion) or is part of a strategy to answer q (m is a question)".

Roberts's (1996) discourse model serves as a foundation for her analysis of the presuppositions of prosodic focus in English. She claims that, since discourse is cooperative, and as such, participants are always trying to make sure that all the interlocutors are aware of what the CG contains, what the current QUD is, and what strategy they are following. Focus has a role in aiding the cooperation between participants, as it gives information about the type of information structure the utterance occurs in; and what is its role in it. Following Rooth (1992), Roberts argues that questions and assertions are associated with a set of possible alternatives, and the alternative of a declarative statement are presupposed by its prosody. Assuming that there is always at least one intonation phrase for sentential utterance, and that there is at least one focused constituent per intonation phrase, Roberts proposes the prosodic factors listed in (20) as pragmatically relevant (Roberts, 1996, p. 29).

(20) a. The choice of intonation phrase constituent(s) (which I tentatively assume are always correlated with syntactic constituents); and

b. within the intonation phrase,

(i) the choice of focused constituent,

(ii) the placement of pitch accent(s),

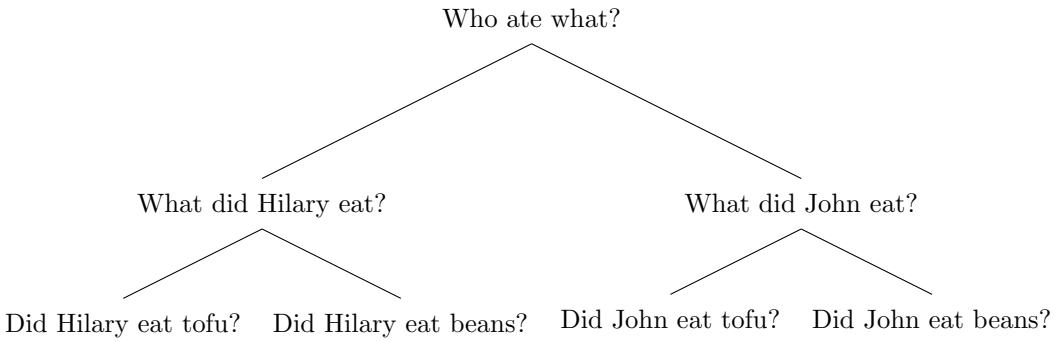
(iii) the choice of pitch accent(s), and of phrase accent and boundary tone.

c. the relative prominence of different intonation phrases, both within an utterance and across utterances.

The notion that assertions, as well as questions, have focal alternatives can be used to shed light on the mechanism that discriminates between congruent and incongruent question-answer pairs: "A conversational move β is defined as congruent to a question *?a* if to β is identical to the set of Q-alternatives that constitutes the semantics of *?a*" (p. 31). In other words, the question and the answer must evoke the same set of alternatives.

Büring (2003), borrows for his model some of the notions defined by Roberts, such as the notions of *Question Under Discussion* and *relevance*, albeit amending them slightly; and more importantly, the notion of discourse strategy. One relevant introduction he makes is a new kind of notation for a discourse strategy, a d(discourse) tree. For example, the strategy in (13) can be represented using the d-tree in (21), allowing us to see intuitively question relations.

(21)



D-trees are in every aspect similar to Roberts's notion of discourse strategy: each node of the tree is a move, moves are ordered hierarchically, and for a d-tree to be well-formed, each move must obey relevance. Now, with this new notation, the QUD is defined, for any move M, as the move M' immediately dominating it. Relevance, still based on the QUD, is defined as follows in the d-trees approach.

(22) a. An assertion a is relevant in a d-tree iff A is an answer to the QUD for A in DT.

b. A question Q is relevant in a d-tree iff at least one answer to Q is an answer to the QUD for Q in DT.

Büring (2003)'s aim is to redefine *congruence* in relation to *Contrastive Focus*. It had already been noted that English has two kind of accents, phonologically different and not interchangeable, called *A-focus* and *B-focus* in previous literature (Jackendoff, 1972). Büring, adopts the name *Contrastive Topic* (CT) for B-focus, and calls A-focus just *Focus* (F).

(23) a. What about Fred? What did he eat?

b. $[\text{Fred}]_{CT}$ ate the $[\text{beans.}]_{F}$

(24) a. What about the beans? Who ate them?

b. $[\text{Fred}]_{F}$ ate the $[\text{beans.}]_{CT}$

Answering (24-a) with the intonation of (23-b), or vice versa, would not be congruent. Büring argues that, while Roberts (1996) had identified the different prosodic pattern, she had failed to differentiate between CT and Focus, treating them both as foci in her model. For this reason, she had assumed both sentences (23-b) and (24-b) receive the same appropriateness conditions, and both presuppose the question *who ate what?*. However, the fact that they cannot be used interchangeably to answer (23-a) and (24-b) should argue against the fact that the two accent patterns can be reduced to the same thing.

Büring proposal is that CT marking indicates a strategy, and therefore, its occurrence and its position in a sentence is related to the strategy said sentence is a part of. The first step to this process is identifying how CT works. Firstly, the function CT is applied to a sentence to hold a *CT value*, i.e. for a declarative sentence a, $[\![a]\!]^{CT}$ yields a CT value - a set of question meanings derived by the *CT value formation algorithm* in the way described below.

(25) a. **Step 1:** Replace the focus by a *wh*-word and front the latter; if focus marks the finite verb or negation, front the verb instead.

b. **Step 2:** Form a set of questions from the results of Step 1 by replacing the contrastive topic by some alternative to it.

(Büring, 2003, p. 6)

Therefore, the CT value formation for (23-a) yields the results in (26).

(26) a. **Step 1:** What did Fred eat?

b. **Step 2:** What did Fred eat?

What did Mary eat?

What did ... eat?

This set of questions corresponds to the set of question meaning below, the actual CT-value of (24-a).

(27) $[\![[\text{Fred}]_{CT} \text{ ate the } [\text{beans}]_F]\!]^{CT} = \{\{\text{x ate y} \mid \text{y} \in \text{D}_e \mid \text{x} \in \text{D}_e\}$

On the other hand, the CT value formation for (24-a) would look like the one in (28), and its CT value is provided in (29).

(28) a. **Step 1:** Who ate the beans?

b. **Step 2:** Who ate the beans?

Who ate the eggplants?

Who ate ...?

(29) $[\![[\text{Fred}]_F \text{ ate the } [\text{beans}]_{CT}]\!]^{CT} = \{\{\text{x ate y} \mid \text{x} \in \text{D}_e \mid \text{y} \in \text{D}_e\}$

In line with Rooth (1992, 1996), Büring concludes that these examples allow us to see how (23-a) can only be part of a strategy consisting of questions in the form *what did x eat?*, like the one listed in (26-b), and (24-a) can only be part of a strategy consisting of questions in the form *who ate x?*, listed in (28-b).

In summary, both Roberts's (1996) and Büring's (2003)'s accounts rely on questions for their theory of how discourse coherence is achieved, as questions have a crucial part in the strategy participants adopt to exchange information.

The QUD approach to context has received a lot of attention in recent years: in particular, question-based model of discourse have been employed extensively in the analysis of a variety of linguistic phenomena. In recent years however, the QUD approach as formulated by Roberts has also received some criticism, often based on its perceived restrictiveness. For example, as Büring (2003) and Van Rooy (2003) observed, one of the problems is that a yes-no question like (8), repeated below as (30-a), is appropriately answered by B in (30-b), although the answer in question does not eliminate any cell in the partition.

(30) a. A: Did John come?

b. B: Presumably.

Other scholars have argued that the QUD framework might be too restrictive to handle truly naturalistic text (Riester, Forthcoming): for example, Roberts (1996) claims that every incoming discourse move has to be relevant and contribute to answering the QUDs which are already in the stack. However, Riester, Brunetti, and De Kuthy (2017) observed that people often seem to introduce irrelevant material in conversation, which nevertheless have a topical connection to the ongoing discourse. Similar observation have been made by Onea (2016) and Velleman and Beaver (2015). A solution to this is to loosen the requirement on the QUD side that QUDs be ordered by the sub-question relation and allow strategies for achieving discourse goals to be more complex than sequences of sub-questions. Other scholars have argued that defining relevance entirely in terms of question-answer relations might be too restrictive: reporting a recent set of experiments on the processing of polar questions (Tian and Breheny, 2016) and negative sentences (Tian, Ferguson, and Breheny, 2016), Breheny (to appear) contrasts the QUD model against Relevance Theory (RT), a framework developed by (Sperber and Wilson, 1986). RT si founded on the *Principle of Relevance*, that states that human cognition is aimed at optimising relevance. Optimising relevance is in turn defined as obtaining a positive cognitive effect - that is to say, making a worthwhile difference to the individual's representation of the world (Wilson and Sperber, 2002). In an eye-tracking

experiment, Tian and Breheny (2016) report that, when reading a polar question like the one in (31). Such question should correspond to a partition {*Bill has opened the window, Bill has not opened the window*}. Participants were asked to read (31), and showed visual stimuli to represent the two propositions.

(31) a. Has Bill opened the window?
b. Has Bill not opened the window?
c. Hasn't Bill opened the window?

It was found that, while the attention of the subjects is initially divided between the representation of the two propositions, the researchers found a significant bias to the positive image towards the end of the sentence. The data suggests that, when processing a polar question, both representations are active initially, however the positive one becomes more active once the sentence processing is underway. While the initial situation is coherent with the QUD framework (e.g. participants are considering both partitions, in line with Groenendijk and Stokhof's (1984)'s view of questions as partitions), Breheny (to appear) claims that the bias to the positive image suggests the hearer is reasoning about relevance: that is to say, the hearer might have inferred the typical manner in which an answer would achieve relevance for the speaker.

In spite of these objections, the QUD framework is recognised as a powerful tool for the analysis of a wide range of linguistic phenomena: for example, QUD-based explanations have been developed for phenomena such as scalar implicatures (Zondervan, 2009), cleft sentences (Tian, Breheny, and Ferguson, 2010), ellipsis (Clifton Jr and Frazier, 2018), discourse-level implicatures (Jasinskaja, 2009), and several others. The next section introduces the relevance of the QUD for scope ambiguity resolution and CP, which will be the main focus of this thesis.

1.3 The role and effect of the QUD

This section introduces some of the research that has been conducted on the effect of the QUD in two domains, the resolution of scope ambiguity and the occurrence of CP, which will be the main focus of this dissertation. Both domains are widely thought to be affected by the influence of the QUD, and the topic has attracted much interest in

experimental research, which will be presented in more detail in the next chapters of this dissertation.

1.3.1 The role of the QUD for scope interaction

The interpretation of scope bearing elements has been widely investigated in relation to children's language acquisition. In particular, much research on this topic has been focused on scopally ambiguous sentences containing a quantifier and a negation, like the ones in (32) and (33).

(32) Every horse didn't jump over the fence.

(33) The detective didn't find some guys.

Such sentences are ambiguous between two possible interpretations: one in which the quantified NP takes scope over the negation, and one in which the negation takes scope over the quantified NP. For example, the sentence in (32) is ambiguous between the interpretation in (34-a) and the one in (34-b).

(34) a. Every horse is such that it didn't jump over the fence.
$\forall$x [horse (x) $\rightarrow$ $\neg$ jump over the fence (x)]

b. Not every hose jumped over the fence.
$\neg\forall$x [horse (x) $\rightarrow$ jump over the fence (x)].

Interpretation (34-a) is usually called *surface scope*, as the two scope bearing elements are interpreted in the same order as they appear in the linear order of the sentence. Interpretation (34-b) is referred to as *inverse scope* as the order is reversed. Over the years, researchers have been concerned with finding out whether children are able to access both interpretations of (32) and similar sentences. Musolino (1998) observed that preschoolers (aged 4 to 6) would often reject the inverse scope of ambiguous sentences, i.e. they would describe (32) as false when presented as description of a situation in which two horses jumped over the fence and a third horse did not. Similarly, they rejected (33) in a context in which a detective found two guys out of four. Adults on the

other hand accepted these sentences as true. These results led to the idea that children are limited to surface scope because of a difference in their grammar which prevents them from obtaining inverse scope interpretations. This observation was challenged by subsequent studies on children's resolution of scope ambiguity, as it was observed that it was in fact possible for them to access inverse scope interpretation (Gualmini, 2004a, Krämer, 2000).

In particular, researchers noted that the context in which an ambiguous sentence was presented had an effect on whether the inverse scope interpretation is obtained by children: Gualmini (2004a) reported that children were much more likely to accept (35) rather than (36) as a description of a story in which the Troll only manages to deliver two of the four pizzas a customer has ordered, even if (36) is ungrammatical in adult grammar due to the PPI properties of *some*.

(35) The troll didn't deliver some pizzas.

a. The troll didn't deliver any pizzas. (Surface scope)

b. Some pizzas were such that they weren't delivered. (Inverse scope)

(36) The troll didn't lose some pizzas.

a. The troll didn't lose any pizzas. (Surface scope)

b. Some pizzas were such that they weren't lost. (Inverse scope)

The two sentences are both true on their inverse scope interpretation, but children were able to access the inverse scope interpretation of (35) and similar sentences in the vast majority of cases (90%), while acceptance rates for (36) and similar sentences were much lower (50%). The initial explanation proposed for this relates to the felicity conditions of the negation: as Gualmini (2004a,b) points out, it has been proposed in literature that negated sentences are felicitous when used to point out a discrepancy between the expected outcome of an action and its actual outcome (Wason, 1965, De Villiers and Flusberg, 1975): the context provided for (35) and (36) had raised an expectation that the Troll would deliver all four pizzas, so (35) is felicitous in this context as it does point out a discrepancy between what was expected to happen and what actually happened; while (36) does not.

This proposal was later refined to a more general one that could hold for all sentences and not only the ones containing a negation: the Question and Answer Requirement (Gualmini et al., 2008). Building on Roberts's (1996) idea that a sentence is always understood as an answer to a question, Gualmini et al. claimed that previous results obtained could be explained in light of this consideration. Supposing that the context provided for (32) (repeated below with its two possible interpretations) had given rise to the implicit question *did any horse jump over the fence?*, it can be noted how only the surface scope interpretation (37-a) is informative, while (37-b) is uninformative in relation to such question.

(37) Every horse didn't jump over the fence.

a. Every horse is such that it didn't jump over the fence (i.e. no horse jumped). (Surface scope)

b. Not every horse jumped over the fence. (Inverse scope)

Gualmini et al. (2008) go on to demonstrate this hypothesis using a sentence that is in every respect similar to (37), presented in a context that should make relevant a QUD to be answered by the inverse scope interpretation of the test sentence.

(38) **Context story:** Caillou is supposed to deliver four letters to Rosie, but he drops two of them on the way to Rosie's house and ends up only delivering two. **Supposed QUD:** Was every letter delivered?

(39) **Test sentence:** Every letter was not delivered.

While (39) is in every way similar to (37) (both sentences contain the same scope bearing elements, in the same order; and they are both true on their inverse scope interpretation), the results obtained by Gualmini et al. (2008) differ dramatically from the ones obtained by Musolino (1998): while (37) has be accepted only 7.5% of the time, (39) had an acceptance rate of 80% - an effect attributed to the different QUD.

The results obtained in the studies discussed here provide solid evidence of the fact that the QUD has an effect on which interpretation gets selected for a scopally ambiguous sentences, and indeed Gualmini et al. (2008) argued that the management of the QUD

should be considered the only factor of difference between children and adults. However, further research on scope ambiguity with elements different from quantifiers and negation has obtained different results: in particular, studies on sentences containing modal verbs and negation have shown that children sometimes seem insensitive to the effect of the QUD (Moscati, 2011, Moscati, Romoli, Demarie, and Crain, 2016). This finding points towards the possibility that there other factors at play that need to be investigated. The next section moves onto the effect of the QUD for another phenomenon, Conditional Perfection.

1.3.2 The role of the QUD for Conditional Perfection

Conditional Perfection (CP) is the name used by Geis and Zwicky (1971) to refer to the phenomenon for which, in certain circumstances, conditionals are interpreted as biconditionals; that is to say, a sentence in the form *if p, then q* is thought to also convey the meaning *if not p, not q*, as exemplified below.

(40) a. If you mow the lawn, I'll give you five dollars.
b. If you don't mow the lawn, I won't give you five dollars.

The nature of CP is debated, and various explanation have been put forward over the years as to how it arises. The majority of these consider CP to arise as an implicature of some sort: for example, one of the most prominent approaches considers it a scalar implicature (Horn, 2000, Atlas and Levinson, 1981, van der Auwera, 1997), and this view will be presented in more detail in the next section.

Another popular explanation is that CP is a general quantity impicature that arises as a product of exhaustivity when answering questions, explicit or otherwise (Von Fintel, 2001). Using examples from Groenendijk and Stokhof (1984), Von Fintel points out how answers to questions are generally interpreted as exhaustive: when hearing (41-b) the hearer assumes that Robin and Hilary only left the party early. It can be noted that this observation holds for both sentential answers and constituent answers (in bold).

(41) a. Q: Who left the party early?
b. A: **Robin and Hilary** left the party early.

Again following Groenendijk and Stokhof (1984), the same kind of reasoning can be applied to conditionals, as exemplified in (42).

(42) a. Q: Will Robin come to the party?

b. A: **If there is vegetarian food** Robin will come to the party[1].

Following this idea, Von Fintel (2001) assumed that a conditional gets perfected only when they answer a question inquiring on sufficient conditions for the consequent. Farr (2011) tested the influence of different questions on CP experimentally: in her study, adult participants would read a short story that served as a context for the conditional test sentence. The test sentence itself was preceded by one of two different kind of questions: either a question on the antecedent of the conditional (*what-if-p question*) or one on the consequent (*when-q question*), as exemplified in (43).

(43) Monika sells seafood on the market. She gets 1 euro for a crab, 2.50 euros for an eel, 15 euros for a lobster and 2.50 euros for a pike. Kerstin, an employee of Monika, cannot remember the prices. Since she does not want to ask Monika again, she asks Sahra, who also works for Monika. Sahra knows the prices exactly.

What-if-p question:	*When-q question:*
What happens if I sell an eel?	When do I get 2.50 euros?

Answer: If you sell an eel, you get 2.50 euros.

As questions are supposed to elicit an exhaustive answer, participants were expected to accept the conditional as an answer to the *when-q* question, but to reject it in the case of the *what-if-p* question. These predictions were borne out. Participants on both conditions were then asked whether the conditional test sentence was a sufficient answer to the question (first experiment) or whether it was true answer (follow up experiment).

[1]Von Fintel (2001) notes that (42-a) is a yes/no question, and therefore it should receive a yes/no answer. However, as the question that would be needed to elicit an exhaustive list of all sufficient conditions (i.e. *What are all the antecedents p such that in all p worlds/p-situations the consequent q is true?*) is regarded as implausible in natural conversation, he assumes that (42-a) might be answered by saying "That depends", followed by the interlocutor asking "On what?". (42) represents a shortened version of this conversation.

Both studies showed that, for both experiments, participants in the *what-if-p question* condition were much more likely to perfect the conditional compared to the participants in the *when-p question* condition (89.2% vs. 16.3%, 56.6% vs. 11.1% for the follow up).

However, as Von Fintel (2001) notes, conditional get perfected much too often to suppose its occurrence can only be justified by the presence of an *explicit* question inquiring on the sufficient conditions for the consequent. Adopting Roberts's (1996) model, however, it can be supposed that *implicit* QUDs of this kind are much more frequent (Von Fintel, 2001). While the results obtained by Farr (2011) provide promising evidence that CP might indeed arise as a product of the QUD, other experimental studies have obtained conflicting results on the same topic (Cariani and Rips, 2016), hence more research would be needed to further investigate this claim. No studies testing the effect of the QUD for CP have been conducted on children so far.

The next section moves on to the account that sees CP as a SI, which is also relevant for language acquisition, since SIs have often been reported to be derived at a lower rate from children compared to adults (Noveck, 2001, Papafragou and Musolino, 2003, Guasti et al., 2005, Foppolo et al., 2012, Chierchia, Crain, Guasti, Gualmini, Meroni, et al., 2001, among many others).

1.4 Scalar Implicatures and their derivation

As discussed in section 1.3.2, one of the most prominent explanation of CP is the one that considers it a SI: this section introduces the notion of SI, and briefly discusses two theories on how they are derived, which will be relevant for the second part of this dissertation.

Firstly, it has long been known that a sentence like (44-a) derives the implicature in (44-c): indeed, experimental work on adults has shown that people often consider (44-a) false in a situation in which (44-b) is also true (Noveck, 2001, Papafragou and Musolino, 2003, Noveck and Posada, 2003, Bott and Noveck, 2004).

(44) a. Some of my friends speak French.
b. All of my friends speak French.
c. → Some but not all of my friends speak French.

SIs share several properties with other kind of implicatures, for example, they are cancellable (Grice, 1989); that is to say, they can be contradicted by the speaker without resulting in a contradiction, which is taken to show that they are not part of the lexical content of the utterance.

(45) a. Some of my friends speak French.
b. ...In fact, all of them do.

There are various approaches that try to explain how SIs arise: one of the most common approaches sees SIs as a pragmatic phenomenon, claiming that they derived from Grice's cooperation principle, and in particular from three maxims, the maxims of Quantity (make your contribution as informative as possible), Quality (contribute only what you know to be true) and Relevance (make your contribution relevant). This approach claims that this kind of implicature that arises from reasoning about the unsaid - more specifically, SIs arise from the hearer comparing the uttered sentence against a stronger competitor, and reasoning that if this more informative alternative had been true and relevant, a cooperative speaker would have uttered it instead. Thus, the sentence in (44-a), which contains the weaker quantifier *some*, gets compared against (44-b), which contains the stronger term *all*. This observation leads to the assumption that if the speaker had believed the stronger sentence to be true she would have uttered it instead of (44-a), this results in the inference (44-c).

SIs can arise from a variety of lexical categories as well as quantifiers: for example verbs (e.g. *might* - *must*), adjectives (e.g. *warm* - *hot*) and connectives (e.g. *or* - *and*). Thus, the same reasoning is thought to give rise to CP from a conditional statement, and different proposals have been made as to the nature of the alternatives against which the conditional sentence would be compared. Atlas and Levinson (1981) proposed the scale in (46), which has however been criticised for not holding a satisfactory explanation: the stronger alternative proposed is nothing else but the conditional plus the CP inference. As scalar implicatures arise from the idea that the speaker does **not** believe the stronger sentence to be true, a scale like (46) is implying that the perfected conditional arises from its own negation (Horn, 2000).

(46) if and only if p, q

↑ if p, q.

Other possibilities have been explored. For example, van der Auwera (1997) proposed the scale in (47). Such a scale would incur in the problem that the expression at its very top would have to be infinitely large to incorporate all possible antecedents, and it seems impossible that the hearer would be able to consider an infinitely long list of possible alternatives (Von Fintel, 2001, Horn, 2000).

(47) ...

if p, q, and if r, q and if s, q

if p, q and if r, q

↑ if p, q.

The derivation of SIs is a widely investigated phenomenon in language acquisition: the majority of studies have found that children derive this kind of inference at a lower rate compared to adults (Noveck, 2001, Papafragou and Musolino, 2003, Guasti et al., 2005, Foppolo et al., 2012, Chierchia et al., 2001, among many others), which in turn has been attributed to a variety of reasons, which will be explored in more detail in section 4.5.2. One critique directed at the pragmatic theory of SIs is that it can only derive implicatures for the whole utterance, as the Maxim of Quality cannot apply to embedded sentences: for example,(48-a) (from Geurts and Pouscoulous (2009)) does indeed seem to give rise to the inference that Bob believes that Anna ate *some but not all* of the cookies, but the pragmatic approach seemingly derives a weaker inference, namely that Bob has no opinion as to whether Anna ate all of the cookies, as expressed by (48-b).

(48) a. Bob believes that Anna ate some of the cookies.

b. It is not the case that Bob believes that Anna ate all the cookies.

A solution to this is offered by the theory according to which the computation of SIs is a grammatical phenomenon, which does allow for SIs to arise in embedded position. An overview of this account is provided in the next section.

1.4.1 Grammatical approach

The observation that the pragmatic account of SI is unable to derive SIs in embedded sentences has been used to argue against this account. Recently, a grammatical approach to SIs has been proposed, which provides a solution to this problem. The grammatical approach, developed by Chierchia (2006) and Fox (2007), is a radical departure from Grice's original idea in that SIs are derived from reasoning about what the speaker could have said, arguing instead that SIs are derived from the application of a grammatical operator, *exh*, which is equivalent to a covert version of *only*. This operator strengthens the utterance by conjoining it with the negation of some of its alternatives. Following the definition given by Fox (2007), given in (49), *exh* takes a sentence (S) and the set of its alternatives (A), *c* represents its context and *IE* represents the set of alternatives that are innocently excludable[2].

(49) $[\![\text{exh}(\text{A}(\text{S}, \text{c}))(\text{S})]\!] = [\![\text{S}]\!] \wedge \bigwedge \{\neg[\![\text{S}_i]\!] : \text{S}_i \in \text{IE } (\text{A}(\text{S},\text{c}))\}$

The derivation process then takes the sentence and its set of alternatives, and gives back as a result the uttered sentence conjoined with the negation of its alternatives. For example, in the case of (44-a), the set of alternatives to *some* includes the stronger quantifier *all*.

(50) a. $\text{A}(\exists, \text{c}) = \{\forall\}$

b. $\text{exh}\{\text{A}(\exists, \text{c})\} = \exists \wedge \neg\forall$

From the notion of innocently excludable alternatives, it follows that *exh* cannot negate alternatives that are symmetric with one another: two alternatives A and B are defined as symmetric if the negation of one entails the truth of the other. Symmetric alternatives cannot be negated by *exh*: negating both A and B would result in a contradiction. This can be observed in the case of a sentence with disjunction, like (51-a).

[2] *Exh* aims to negate as many alternatives in ALT(p) as possible, while preserving consistency with the original sentences, and for this reason alternatives are not negated arbitrarily. Therefore, to see if a proposition is innocently excludable, one has to look at every maximal set of alternative propositions such that its exclusion is consistent with the perjacent: every such set can be excluded consistently as long as nothing else in A is excluded. Hence, only propositions that are in every one of these sets are innocently excludable. All other propositions would be an arbitrary exclusion, since excluding it would force us to include a proposition from the other maximal exclusion sets.

(51) a. A ∨ B

b. Alternatives: {A, B, A ∧ B}

The disjunction has three stronger alternatives, which correspond to each one of the two disjuncts and the conjunction. Here, to see which alternatives are innocently excludable, one has to look at every maximal set of alternatives such that its exclusion is consistent with (51-a): these are the two sets, {A, A ∧ B} and {B, A ∧ B}. The innocently excludable alternatives are the ones that are in every set: the set of innocently excludable alternatives for (51), therefore contains only one proposition, $A \wedge B$.

This approach provides a solution to the problem of embedded SIs exemplified in (48): *exh* can apply to the subordinate clause, and this generates the correct SI.

(52) a. Bob believes that exh[Anna ate some of the cookies.]

b. Bob believes [that Anna ate some of the cookies] ∧ [that ¬(Anna ate all of the cookies.)]

Experimental evidence has evidenced that SIs are indeed computed in embedded positions: for example Chemla and Spector (2011) found that participants often obtained the reading in (53-b) for (53): this reading is computed by deriving the scalar implicature for the embedded sentence (every letter is connected with exh[some of its circles]), unlike (53-b) which is computed by deriving the SI for the whole sentence.

(53) Every letter is connected with some of its circles.

a. Every letter is connected with at least one of its circles, and no letter is connected with all its circles.

b. Every letter is connected with at least one of its circles, and it is not the case that each letter is connected with all its circles.

Both approaches will be relevant for the rest of this dissertation, as they are central to much of the experimental literature on SIs discussed in Chapter 4, and both pragmatics or grammar-based explanations have been attempted justify the results obtained in children.

1.5 Contribution of this thesis

This thesis aims to investigate two domains from the point of view of language acquisition: scope ambiguity resolution, focusing in particular on the role of the QUD; and CP. The role of the QUD has been the focus of much research in recent years, but in spite of this, the role of questions for scope assignment had not been tested directly, and the QAR, in its current formulation, does not make any prediction in relation to the disambiguation of scopally ambiguous questions. This is probably due to the fact that it is not clear whether a question can be interpreted by the hearer as the answer to another question. For this reason, two experiments have been conducted on adults and children on scopally ambiguous questions. One tested the ambiguous question in combination with a declarative sentence (as the interpretation of the declarative sentence is dependent on the disambiguation of the question), and the other testing the interpretation of the ambiguous question directly. The main finding of the two experiments is that children are able to access inverse scope in the vast majority of cases, just like adults do. In line with the QAR, these results can be interpreted in two ways: firstly, it is possible that questions are not interpreted by the hearer as answers to a QUD, and therefore its influence is removed. Since the influence of the QUD is argued to be the only factor of difference between children and adults, its removal would also have removed any difference between the two groups. The second possible explanation is tied in with the concept of *superquestion*, introduced in section 1.2.2: the salient QUD participants are trying to answer at any given moment stands in relation with broader questions by which it is entailed - its superquestions. It is therefore possible that, like a (possibly implicit) QUD determines which interpretation gets selected for a sentence, in the same way the underlying superquestion influences the interpretation of an explicit question. It can be supposes that context provided for the test question in our experiment has made relevant a superquestion that only entails the inverse scope interpretation of the test question. In light of this, I propose the QAR be amended as follows.

Extended QAR: The interpretations available for a scopally ambiguous question are the one(s) entailed by its superquestion.

The second part of this thesis focuses on the interpretation of conditionals and the inferences arising from them (in particular conditional perfection and the conjunctive-like interpretation often observed in children) in light of work on disjunction. While it has

long been known that both conditionals and sentences containing a disjunction tend to receive a conjunctive-like reading in children (Paris, 1973), language acquisition research has mostly focused on the latter (Braine and Rumain, 1981, Chierchia et al., 2004, Gualmini, Meroni, and Crain, 2000). Two experiments were conducted on children's and adults' interpretation of conditionals, using a Truth Value Judgment Task, in order to investigate what reading of a conditional children and adults obtain, and if there is any difference between the two groups. The results have revealed that both groups seem to favour a conjunctive interpretation, accepting the conditional as true only when both antecedent and consequent are true. As for children's performance, their results are in line with previous results which have shown that children consistently endorse this interpretation, possibly because they are only able to construct in their mind a representation in which both propositions are true, and therefore judge the conditional false in every other circumstance (Barrouillet and Lecas, 1998). The results obtained in adults are more surprising: conjunctive like interpretations are common in adults as well, but their rate is generally lower (Paris, 1973, Oberauer and Wilhelm, 2003, Evans, Handley, and Over, 2003). As the interpretation obtained for conditionals is reported to vary widely depending on the task (see Barrouillet and Lecas, 1998, Marcus and Rips, 1979), I argue that pragmatic consideration might have pushed participants towards this interpretation. There is however another possibility that could offer a partial explanation for the results obtain in this study: building on the observation that, under the material implication account, a conditional *if p then q* is equivalent to $\neg p \vee q$, one could draw a parallel with the theory proposed by Singh, Wexler, Astle-Rahim, Kamawar, and Fox (2016) for children's conjunctive interpretations of disjunction: said interpretation is thought to arise as a SI, due to children having access to a different set of alternatives. My proposal is that CP could potentially be derived, in adults at least, by the same mechanism, with children deriving a different reading which is at least in part compatible with our results.

The rest of this thesis is organised as follows: Chapter 2 discusses some of the literature relevant to scope ambiguity resolution in adults and children, focusing in particular on experimental studies which have shown that there is sometimes a difference between the two groups, and the reasons that have been proposed for this. Chapter 3 presents two experiments on scope ambiguity resolution that have been conducted on adults an

children, on their disambiguation of doubly quantified declarative sentences and questions, and discusses how the results for this study relate to previous literature. The two following chapter move onto Conditional Perfection: Chapter 4 discusses some of the literature on conditionals: reviewing the various theories that have been proposed on the nature of CP and focusing on particular on the one that sees CP as a SI. Attention is also devoted to children's interpretation of disjunction, in order to draw a parallel with conditionals. Chapter 5 presents the two experiments that have been conducted on children's and adults' interpretation of conditionals. And finally, Chapter 6 summarises the findings of the this thesis, their implication for previous studies and reports some considerations for future research.

Chapter 2

Scope literature review

The aim of the following section is to present a review of relevant literature on scope ambiguity, including experimental studies on how the ambiguity is resolved by adults and children. Studies on children have mostly focused on the fact that, albeit generally linguistically savvy, preschoolers 4 to 6 years old often differ from adults in the extent to which they can access some of the readings available to adults. Various possible reasons for this asymmetry have been identified and tested experimentally, including processing factors and the role of discourse. Studies on adults have been aimed at finding out whether some interpretations might be more difficult to obtain than others. It is believed that the observed scope preferences arise from lexical properties of the specific quantifiers involved, from the complexity of the mental models associated with the competing interpretations, or again from discourse factors, including the plausibility of the possible interpretations in relation to the context.

The following chapter is organised as follows: section 2.1 presents an explanation on the phenomenon of scope ambiguity and how it is resolved. Section 2.2 details some of the experimental studies on children's interpretation of scopally ambiguous sentences containing a quantifier and a negation, or a modal verb and a negation. Section 2.3 introduces some of the research on the topic of doubly quantified sentences, which has focused on trying to uncover what factors determine scope assignment and how the different readings are constructed in the mind of the speaker. Section 2.4 expands on this topic, discussing some experimental studies conducted on adults' interpretation of doubly quantified sentences.

2.1 An introduction on scope ambiguity

Sentences with more than one scope-bearing elements, like nominal quantifiers, modals or negation, are often associated with more than one interpretation. For example, a sentence with a universal quantifier and an existential quantifier, such as (1), is ambiguous between the interpretation in (1-a), in which one single doctor took each pencil (*a* > *each*); and the one in (1-b), in which each pencil was taken by a (possibly different) doctor (*each* > *a*).

(1) A doctor took each pencil.

a. One single doctor took each pencil.
$\exists$x [doctor (x) & $\forall$ y [pencil (y) $\rightarrow$ took (x, y)]]

b. Each pencil was taken by a (possibly different) doctor.
$\forall$x [pencil (x) $\rightarrow$ $\exists$ y [doctor (y) & took (y, x)]]

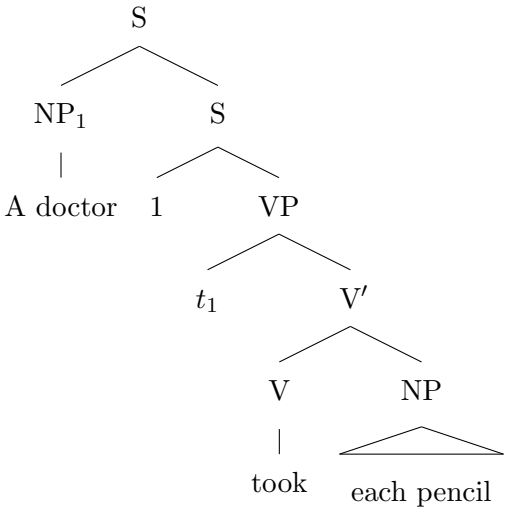
The first interpretation is typically called *surface scope*, because the two quantifiers are interpreted in the same order they appear in the linear order of the sentence, while the other is called *inverse scope*.

One prominent approach to the ambiguity of sentences like (1) is associating the two readings with two different syntactic representations at Logical Form (LF): since LF is the level of syntactic representation that determines the semantics of a sentence, having two different representation at this level is what gives rise to the two meanings associated with the same sentence, with one quantifier moving above the subject. The two configurations therefore differ with respect to which of the two quantifiers c-command the other (May, 1977, Fox, 2000)[1].

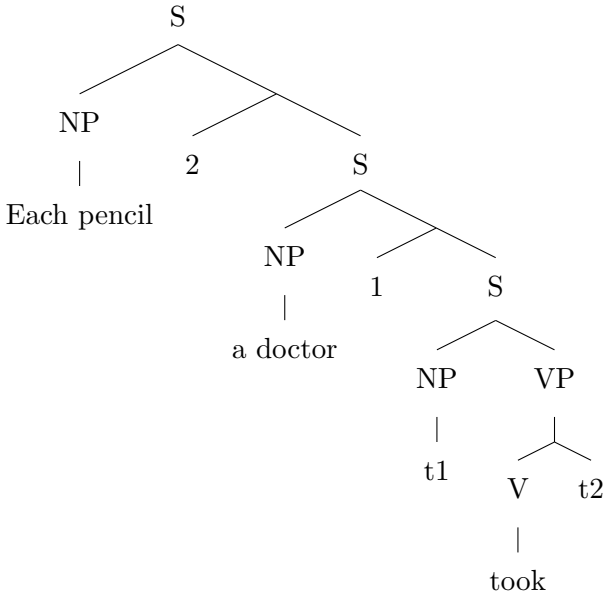
[1] C-command is defined in Reinhart (1976): A node A commands a node B iff:

1. A does not dominate B and B does not dominate A;
2. The first branching node dominating A also dominates B.

(2) a.



b.



The study of scope ambiguity has attracted considerable attention over the years, as it allows to uncover some of the underlying processes of language comprehension and acquisition; in particular, if scopally ambiguous sentences involve two or more possible representations, one might find a reflection of that in their processing and more indirectly in the way children acquire their possible interpretations. Indeed, a large number of studies have found an asymmetry between surface and inverse scope in sentence comprehension. The surface scope interpretation (i.e. *a* > *each* in the case of (1)) appears to be generally preferred over the inverse scope one. This asymmetry has been widely investigated, and several factors have been proposed over the years as the reason for its occurrence. In addition to this, it seems that children differ from adults in the interpretation of such sentences: most of the studies that have investigated children's comprehension of scopally ambiguous sentences focused on the interaction between quantifiers

and negation, using sentences like (3).

(3) Every horse didn't jump over the fence.

These studies have found evidence that children struggle to get inverse scope readings. The initial proposal linked this difficulty to their developing grammatical abilities (Musolino, 1998, Lidz and Musolino, 2002). However, subsequent studies have later shown that contextual influence also plays a role in the resolution of the ambiguity: given enough contextual manipulation, children appear to access the inverse scope reading more, if not at adult-like level (Gualmini, 2004a, Krämer, 2000, Gualmini et al., 2008, Gualmini, 2004b). The results of several experimental studies on the topic of children's acquisition of scope ambiguity resolution are presented in the next section.

2.2 Childen and scope ambiguity: Quantifier - Negation studies

From the late 90s, a number of studies have aimed at investigating children's comprehension of scope interaction in sentences containing a quantifier and negation. Before then, there had been studies on the interaction between quantifiers (for example Bucci (1978), Brooks and Braine (1996), Philip (1994)), but these studies were mostly aimed at uncovering other peculiarities of children's behaviour in relation to quantifiers. One of these peculiarities, noted by Philip (1994), is that children tend to give what he calls "symmetric responses" for sentences with a universal quantifier.

(4) Is every farmer feeding a donkey?

Children would give a negative answer to (4) if the picture associated with the question showed a donkey remaining unfed; even if all the farmers depicted were in fact feeding a donkey each.

The first systematic study of sentences with a quantifier and a negation is Musolino (1998), who tested who tested pre-schoolers (aged 4 to 6) on sentences containing a variety of quantifiers (including *some* and *every*) appearing either in subject or object

position. The possible interpretations of sentences containing these elements vary, depending both on the nature of the quantifier and on the position of the elements in the sentence. For this reason, he manipulated both quantifier type and word order, aiming to uncover what routes children take to get to an adult-like comprehension of language. The methodology used was a Truth Value Judgment Task (TVTJ) (Crain and Thornton, 1998): participants would hear a short story narrated by one experimenter and acted out using small toys as props, and then would have to say whether a test sentence was true in the context provided by the story. The test sentences contained either the existential quantifier *some* or the universal *every*, with the quantifier appearing in subject or object position. An example for each condition is provided below.

(5) a. Every horse didn't jump over the fence.
b. The Smurf didn't buy every orange.
c. Some girls won't ride on the merry-go-round.
d. The detective didn't find some guys.

The stories associated with sentences with a quantifier in subject position always involved a set of three characters aiming to fulfil an action, with only two of them succeeding (in the case of (5-a), two horses jumping over a fence, while a third horse did not). The ones associated with sentences with a quantifier in object position depicted a character fulfilling an action on one object out of a set of three (in the case of (5-b), the Smurf buys one out of three oranges he had been considering).

These contexts make the sentences with a quantifier in subject position true on their inverse scope interpretation, and the ones with a quantifier in object position true on the surface scope interpretation: as for (5-a), the inverse scope interpretation (6-b) is the one that corresponds to what happened in the story.

(6) a. No horse jumped over the fence. (Surface scope, *every* > *not*)
b. Not every horse jumped over the fence. (Inverse scope, *not* > *every*)

Musolino found that the sentences that were true on their inverse scope interpretation, like (5-a) and (5-d), were often rejected by children: (5-a) and similar sentences had an acceptance rate of 7,5% only, and (5-d) and similar sentences were accepted 50% of the

time. On the other hand, sentences that were true on their surface scope, like (5-b) and (5-c) had higher chances of being accepted (85% and 100% respectively).

The results of this experiment show that children had difficulties accessing the inverse scope interpretation of sentences with a quantifier and a negation. Adults accepted the sentences nearly 100% of the time, regardless of whether they were true on their surface scope or inverse scope interpretation. Based on this difference, Musolino concluded that children are simply unable to interpret scope bearing elements in a position different than the one they occupy in the linear order of the sentence; a claim known as the Observation of Isomorphism[2]. He attributed this inability to an incompleteness of children's grammar, that only makes available to them a subset of the possible options available in adult grammar. Musolino's conclusions had important implications: a difference in grammar would imply that children are unable to access inverse scope, regardless of other factors. However, subsequent studies brought forward the possibility that the results obtained by Musolino could be due to the influence of the context, and resolved to investigate how discourse factors might play a role in influencing the resolution of scope ambiguity.

2.2.1 The role of the context for scope assignment

Following Musolino (1998), other researchers resolved to further investigate children's difficulty in accessing inverse scope. In particular, Gualmini (2004a) focused on the role of the context children's interpretation of sentences with a negation and the quantifier *some*, like the ones in (7-a) and (7-b). Both of these sentences were amongst the ones used by Musolino in his study and, in spite of the fact that they contain the same elements, (7-a) had a significantly higher acceptance rate.

(7) a. The detective didn't find some guys.
b. The old man didn't hurt some guys.

(7-a) was presented as description of a story in which a detective was playing hide-and-seek with some friends, while the context for (7-b) involved an old man accidentally

[2]However it should be noted that, as emphasised by Musolino (1998), the Observation of Isomorphism is intended as a descriptive generalisation, and not as a principle.

running over some of his friends with a lawnmower. While both sentences were syntactically identical, and true on their inverse scope interpretation, Gualmini claims they differed in their felicity given the expectations in the context. In the case of (6), the real world knowledge gives rise to the expectation that the detective would find all of his friends as required by the rules of the game. For (7), real world knowledge pointed in the opposite direction: one is not expected to run over (some or) all people with a lawnmower. This observation prompted Gualmini to assume that context might play a role in the resolution of scope ambiguity. Previous studies on negative sentences had evidenced how these are subject to a specific set of felicity conditions: they are normally used to point out a discrepancy between the expected outcome and the actual outcome of a situation (Gualmini 2004a, Russell 1949; see also Gualmini et al. 2008).

Gualmini (2004b) resolved to test this experimentally. He constructed contexts structurally similar to the ones used by Musolino (1998), but that could also trigger an expectation to influence the interpretation of the test sentences. In one of the stories used, a Troll is in charge of delivering four pizzas to Grover, but he drives too fast and ends up losing two. This story was associated to two possible test sentences.

(8) a. The troll didn't deliver some pizzas.
b. The troll didn't lose some pizzas.

Both (8-a) and (8-b) are true on their inverse scope interpretation, but only (8-a) is felicitous as it represents the denial of an expectation – namely that the Troll would deliver all four pizzas. As predicted by Gualmini (2004a,b), children were able to accept (8-a) 90% of the time, while they only scored 50% for (8-b), a result identical to the one observed in Musolino (1998) for similar sentences. From this evidence, Gualmini concluded that children do have an adult-like ability to access inverse scope of negative sentences containing the indefinite some, provided the felicity conditions of the target sentence are satisfied.

A similar experiment conducted with the quantifier *every* showed the same pattern. The story proposed for (9) was similar to the troll story from the previous experiment: Caillou is supposed to deliver four letter to Rosie, but he drops one on the way to her house, and therefore only delivers three.

(9) Every letter was not delivered.

This sentence in an inverse scope context was accepted by children 80% of the time, a great improvement compared to the analogous (5-a) used by Musolino (1998), which was only accepted 7,5% of the time. The two sentences differ in the fact that (5-a) is active and (9) is passive, but voice is not thought to influence the scope ambiguity resolution[3].

The results presented in this section show that children are indeed able to access inverse scope. However, it was observed that the idea that a negative sentence has to represent the denial of an expectation presents some difficulties. Firstly, it is hard to determine what exactly counts as a negative sentence: if this category is defined to include every sentence containing a negation, the definition might be too broad. For example, in the case of a sentence like the one in (10), it could be hard to define what expectation is being denied.

(10) John said that Mary will not come to the party.

(10) is the denial of the claim that John did not say that Mary will not come to the party, but it does not seem correct to assume that such an expectation must be relevant to make (10) felicitous. Gualmini et al. tried revising the claim, to include not all sentences that contain a negation but only the ones in which said negation is in an unembedded clause. Still, they claim the requirement is still insufficient, considering the example in (11).

(11) Let me tell you something about how you should behave when you meet Mary. You should smile and shake her hand. Most importantly, you should not say anything impolite.

While (11) is to be considered a negative sentence by the new definition, it does no seem to express the negation of any salient expectation.

For this reason, researchers resolved to find a more general pragmatic requirement that could account for the felicitousness of declarative sentences as well as negative ones: this

[3]Gualmini et al. (2008) tested the passive version of sentences (8-a) and (8-b) without finding any significant difference in their acceptance rate, as it will be detailed in the next subsection.

led to the development of the Question and Answer Requirement for scope assignment, which is presented in the next section.

2.2.2 The Question and Answer Requirement

Gualmini et al. (2008) looked into the assumption that every sentence is always understood as an answer to a question, as proposed by Roberts (1996). As discussed in Chapter 1 of this dissertation, Roberts's model is based on the assumption that participants are committed to the goal of discovering and sharing information about the world, with the ultimate aim of answering the Big Question "what's the way things are?". As this question is too broad to be answered directly, the strategy participants apply is to break it down into sub-questions. Participants address each one of these sub-questions at a time, and when one is answered (or deemed unanswerable), it is replaced by another question, which then become the (immediate) Question Under Discussion (QUD). These questions are normally implicit, and inferred from the context: in this sense, every sentence in a conversation is the answer to whatever QUD the context has made relevant in that moment.

Gualmini et al. claim that the interpretation of a scopally ambiguous sentence is always driven by the immediate QUD: as the sentence needs to be understood as an answer to it, the interpretation that gets selected is the one which effectively answers the QUD (i.e. it entails either a yes or a no to it). They call this phenomenon the Question and Answer Requirement (QAR). The hypothesis is that the QAR applies both to adults and children but, upon hearing an infelicitous sentence, the former are able to disregard the QUD made relevant by the context and accommodate a new one to be answered by the true reading of the ambiguous sentence, adhering to the Principle of Charity (Davidson, 1975). In natural conversation, it can be difficult to know exactly what the QUD is in a given moment, as the topic tends to jump around, but it should be relatively easy in a TVTJ, which normally employs short stories only a few lines long, with a small set of characters and one clear goal. Gualmini et al. claim that in Musolino's (1998) study, children were unable to access inverse scope because his stories didn't follow a clear pattern: in the horse story associated to (5-a) for example, the three horses first considered jumping over a barn, but then rejected the idea as they thought the barn too high to jump over. After this they consider jumping over a fence, with two out

of the three horses fulfilling the actions. Gualmini et al. claim this change of plans described in the story did not make a single possible outcome clear in the story, and therefore children could not recognize a clear QUD: this confusion led them to reject the test sentences. Adults on the other hand are better at dealing with pragmatic infelicity, and were therefore able to accommodate another QUD on the basis of the test sentence they heard. Gualmini et al. tested this prediction against another possible explanation of children's behaviour in Musolino's experiment: namely, that they simply have a preference for surface scope and tend to select this interpretation when the context is not clear. To verify this, they resort to the passive versions of (8-a) and (8-b), associated to the same story about a Troll having to deliver four pizzas but losing two. They assumed that such a context gives rise to the QUD as to whether the troll deliver all pizzas.

(12) a. Some pizzas were not delivered.

b. Some pizzas were not lost.

In this case, both sentences are true on their inverse scope interpretation, but only (12-a) fulfils the QAR as it entails a negative answer to the QUD. (12-b), on the other hand, is uninformative unless it is interpreted on its inverse scope (*not* > *some*, i.e. no pizzas were lost), which makes it a good answer to the QUD, albeit false in context. If children have a preference for surface scope, there is no reason why they should not interpret (12-b) on its surface scope. If on the other hand they are driven in their interpretation by the QUD, they should select the inverse scope interpretation of (12-b), even though it is not normally available in adult grammar. The results showed that children have no problem accepting (12-a) (which was accepted 88% of the time, similarly to its active counterpart). (12-b) was accepted 50% of the time, like the active (8-b), and Gualmini et al. noted that when rejecting the sentence children gave justifications coherent with its inverse scope interpretation. They used this evidence to claim that children have no preference whatsoever for surface scope; and that the interpretation they select is always the one that is coherent with the immediate QUD, even in circumstances when the selected reading would normally be unavailable in adult grammar. The results of these last experiments have shown that children are in fact able to access the inverse scope interpretation of sentences with a quantifier and a negation. Gualmini et al. propose that the QAR might suffice as the only factor to explain the difference observed between

children and adults, without requiring any additional explanation. In their view, an appropriate context is all that is required to allow children to access inverse scope like adults. Other researchers however do not support this conclusions in full: while it is recognised that the contextual manipulations implemented in the experiments related in this section had an important impact on the results obtained for children, subsequent research has found evidence that the context alone cannot be regarded as the only explanation for the difference between children and adults found in early studies. One of the factors that have been considered is the processing load associated with inverse scope, as there is evidence that inverse scope interpretations might be more costly for the parser, as it will be discussed in the next section.

2.2.3 Pragmatic factors and the effect of priming

Another possible explanation for children's behaviour in relation to scope comes from the idea that obtaining an inverse scope interpretation is subject to processing difficulties: an idea that is widely supported by experimental research on adults, some of which is discussed in section 2.3. Ambiguous sentences have received a great deal of attention from researchers, who have been concerned with trying to understand whether the two possible readings are constructed at the same time in the mind of the reader/hearer, or only one of them is constructed, and possibly later revised if contextual clues signal that such interpretation is unlikely or implausible.

Viau, Lidz, and Musolino (2010) conducted a study on children, testing the possible role of priming in making inverse scope interpretation more frequent. In particular, they used a "comprehension-to-comprehension" priming technique, exposing children to unambiguous sentences with the same truth condition as the ambiguous ones, before testing them on the latter.

The study by Viau et al. (2010) builds on a previous one conducted by Musolino and Lidz (2006), which had put forward the idea that children's poor performance in disambiguating sentences to their inverse scope might be due to their immature processing abilities, which make it harder for them to compute the more costly inverse scope interpretation. Viau et al. tested this hypothesis in an experiment consisting of two condition: Condition 1 was a replica of Musolino's (1998) experiment, using the same materials, including the horse story and its associated test sentence reported below as

(13). In Condition 2, a slightly different version of the same story was used, with the three horses jumping over a log; and then two of them jumping over a fence. The test sentence was also different, aiming at setting up a context of denial: as it had already been noted for the previous studies, negative sentences are thought to be easier to process when used to point out the discrepancy between the expected outcome of a situation and its actual outcome (Russell, 1949), as this make their use felicitous. In line with this, another claim is that contrasting a negative sentence with a declarative one might make the first easier to accept (Horn, 1989), therefore this manipulation was integrated in the test sentence (14).

(13) Every horse didn't jump over the fence.

(14) Every horse jumped over the log, but every horse didn't jump over the fence.

While it was initially observed that the presence of the prefaced declarative sentence made a significant difference in acceptance rate (60%, compared to only 15% for (13)), it needs to be observed that the difference could be due to either the different test sentence, or the different context: while in the story provided for (13), three horses consider jumping over a barn, then give up the idea (Viau et al. dub this type of context *Early Failure*, EF), in the story for (14), the three horses successfully manage to jump over a log before two of them jump over a fence (*Early Success*, ES). It is therefore possible that the ES context made the inverse scope interpretation of (14) more plausible and therefore easier to obtain. However, in subsequent experiments, they also proved that children were able to access the inverse scope interpretation of (13) and similar sentences when presented in the unsupportive Early Failure context, if they had been exposed to the more felicitous Early Success stories in previous trials. They argue that this priming effect has to be taken as evidence against the "pragmatic only" view, that sees contextual factors as the only thing relevant for children's scope assignment: if Early Success stories induce children to accept inverse scope because they generate the right QUD, it is not clear how they could have an effect on the following sentences, which are instead presented in an unsupportive context. A similar effect was found by exposing children first to unambiguous sentences equivalent to the inverse scope of the ambiguous sentences (e.g. *not every horse jumped over the fence* for (13)), confirming the existence of a priming effect. Viau et al. took this large improvement as showing that pragmatics

cannot be the only explanation behind children's behaviour, and that processing factors must also be playing a role - they propose a new "pragmatics + processing" approach, opposed to Gualmini et al.'s (2008) "pragmatics only" approach. As mentioned before, the idea had already been introduced by Musolino and Lidz (2006), who claimed that the surface scope interpretation might be the one preferred by the parser, and therefore get constructed early in the mind of the speaker. The inverse scope would only be constructed later, when the sentence material is reanalysed. This whole process might be too costly for children's immature parsers explaining why children, unlike adults, are stuck with the surface scope interpretation.

In sum, the proposal that discourse factors and in particular the management of the QUD are the only reason behind the difference observed between children and adults in their ability to obtain inverse scope has been challenged by evidence that children's immature processing abilities might also play a role. Neverheless, there seems to be clear evidence that having a context that makes salient a QUD that can be answered by the inverse scope of an ambiguous sentence greatly improves the chances that children will accept said inverse scope interpretation. The next section however discusses some experimental studies on sentences containing a modal verb and a negation, which have evidenced a different pattern of response: specifically, the effect of the QUD seems less marked for modal verb-negation interaction than it is for quantifier-negation interaction.

2.2.4 Modal verbs and scope ambiguity

Modal verbs, like quantifiers, can generate a scope ambiguity when they appear in a negated sentence. This phenomenon is rarely observed in English, but other languages, like Italian, have more freedom for modal verbs scope. For example, the sentence in (15) is ambiguous between the two readings (15-a) and (15-b).

(15) Lo studente non deve andare a scuola.
The student not must go to school.

a. It is not the case that the student must go to school.

b. It is necessary that the student does not go to school.

Studies have been conducted in both languages, to determine if children's behaviour with scope extends to modal verbs, however these new experiments uncovered some differences of another nature.

Moscati and Gualmini (2008) conducted a study in English, testing children and adults on sentences containing the negative modal *cannot*, as in (16).

(16) The lion cannot be in the same cage as the tiger.

Even though *cannot* is composed of the elements can and not, in this order, its only interpretation (at least in adult grammar) is an inverse scope one, i.e. in the case of (16), is not the case that the lion can be in the same cage as the tiger. Children tested on (16) and similar sentences proved to be able to access this adult inverse scope interpretation of (16) (91% of the time). A similar experiment in Italian gave the same result (93.3% of acceptance), proving that children are able to access inverse scope for sentences containing modal verbs. Another study (Gualmini and Moscati, 2009) tested children's understanding of other sentences that are normally interpreted in their inverse scope configuration by adults, like the ones containing the verb *need* and a negation, in this order.

(17) To be a good farmer, you need not feed the zebra.

(17) is unambiguous for adults, as it can only be interpreted as saying that it is not necessary that you feed the zebra, to be considered a good farmer. Unlike what happened in the previous studies, children rejected (17) in a context that made it true in 68,3% of cases, where their motivation showed that they were accessing the surface scope interpretation (*not* > *need*, implying a prohibition) instead. A similar experiment was replicated in Italian (Gualmini and Moscati, 2009), using sentences containing the modal *può* and negation.

(18) Il contadino può non dare le carote all'elefante.
The farmer MOD not give the carrots to-the elephant.

(18) has two possible interpretations, one expressing possibility (the surface scope one, *può* > *non*) and one expressing a prohibition (the inverse scope *non* > *può*), but the only available interpretation in adult grammar is the surface scope one. In spite of this, children consistently rejected (18) in a situation that made this reading true (69,3% of cases).

The conclusion proposed by these studies was that, regardless of scope relations, children have a tendency to select the stronger interpretation, i.e. the one that entails the other: this can be observed in both (17) and (18): in the case of (17), if one is prohibited to feed the zebra, then it must be the case that one does not need to feed it. However, as Gualmini and Moscati point out, it is possible that a more thorough contextual manipulation could bring children closer to an adult-like behaviour, in the same way that was observed for sentences with a quantifier and a negation. The next section details the results of several experiments conducted in order to test this claim.

2.2.5 Modal verbs and the QUD

The studies on the effect of the QUD on children presented in the previous sections have uncovered that the interpretation of a sentence that contains more than one scope bearing elements seems to correlate with the immediate QUD made relevant by the context, to the point that children are sometimes able to access readings that are not normally available to adults provided they are answers to the QUD. For this reason, researchers became concerned that children might be failing to access the correct interpretation of sentences with modal verbs because the context stories in the experiments related so far might not have made the right QUD relevant, hence generating a confusion similar to the one children had experienced in the first quantifier + negation experiments.

The role of the QUD for sentences containing modal verbs was investigated more systematically by Moscati (2011), who ran a study in which test sentences would be preceded by an explicit QUD, to have a better control of its influence. The study featured sentences similar to (18) and was divided into two conditions: in Condition 1 the explicit QUD would make both readings of the test sentence felicitous, while in Condition 2 it would only make the felicitous the "prohibitive" reading that is not available in adult grammar. The two possible questions associated with (18) are presented below, together with the test sentence and its two possible readings.

(19) Condition 1 QUD:
Il contadino deve dare le carote all'elefante?
Must the farmer give carrots to the elephant?

(20) Condition 2 QUD:
Il contadino può dare le carote all'elefante?
Can the farmer give carrots to the elephant?

(21) Test sentence:
Il contadino può non dare le carote all'elefante,
The farmer MOD not give the carrots to-the elephant.
a. It is possible that the farmer does not give carrots to the elephant.
b. * It is not possible that the farmer gives carrots to the elephant.

In Condition 1, the QUD (19) is answered by both (21-a) and (21-b), as they both entail a negative answer. However, the Condition 2 QUD (20) is only answered by (21-b), while (21-a) is uninformative and does not yield an answer. The prediction would be that for Condition 1, children behave in an adult-like fashion as both the QUD and grammatical correctness are influencing them towards the same reading. In Condition 2, there is a competition of these two factors, therefore one of them should prevail. Remember that Gualmini (2004a) had found that, at least for sentences with the quantifier *some* and a negation, children might select a reading that is not available in adult grammar as long as it does answer the QUD; but this does not seem to be the case with modal verbs: in this instance, children showed little sensitivity to the manipulation, accepting the target sentence 30% and 40% of the time for the two QUDs respectively. Moscati concludes that the evidence suggests that the adult interpretation of similar sentences is not available to children.

The effect of the QUD was further investigated comparing the effect it has both on "quantifier + negation" sentences and "modal verb + negation" sentences. Moscati et al. (2016) tested Italian children on their comprehension of both types of sentences when preceded by an explicit question. Quantifier target sentences were similar to the ones used in the previous experiments illustrated so far.

(22) QUD:
Il fotografo ha fotografato tutti gli animali?
Did the photographer shoot all of the animals?

(23) Il fotografo non ha fotografato qualche animale.
The photographer not has shoot some animal.

Both the surface scope reading (i.e. the photographer didn't shoot any animal) and the inverse scope one (i.e. there are some animals the photographer didn't shoot) are good answers to the QUD, but (23), like its English equivalent, can only receive an inverse scope interpretation in adult grammar. In addition to this, only this reading was true in context.

As for modal target sentences, they were similar in structure to (19).

(24) Modal QUD:
Fred deve guidare la moto?
Must Fred drive the bike?

(25) Fred può non guidare la moto.
Fred mod not drive the bike.
a. It is possible for Fred not to drive the bike.
b. *It is not possible for Fred to drive the bike.

Again, both the surface scope reading (i.e. "Fred might not ride the bike") and inverse scope reading (i.e. "Fred cannot ride the bike") answer the QUD but, as mentioned before, in this case the surface scope reading is the only one available to adults. For both conditions, the "pragmatic only" view proposed by Gualmini et al. (2008) predicts that children will be able to access the true interpretation of the sentences similarly to adults. However, Moscati's (2011) results with modal verbs contradict this expectation. The results obtained by Moscati et al. (2016) confirmed the predictions for quantifiers: the acceptance rate for sentences containing these elements was close to 100% for both children and adult controls. With modal verbs, however, children's results mirrored the one obtained in the previous study: children only accepted (25) and similar sentences

43,1% of the time, while adults generally accepted the same sentences. Another experiment was conducted using a different QUD, as it could be objected that (22) and (24) did not contain the same quantifier as the test sentences associated with them, making it harder to identify them as the right QUDs. The questions therefore revised to include the elements present in the test sentences, which remained the same. However, this manipulation did not produce any significant improvement in children's acceptance of the test sentences: again, children performed well with quantifiers, but they were still unable to accept the modal verb sentences in the majority of cases. From this evidence, Moscati et al. concluded that the adult readings of sentences containing a negation and a modal verb are unavailable to children. The predictions made by the QAR do not seem to hold in this case: not even a supportive context and an explicit QUD appear to bring children to an adult-like behaviour. Gualmini et al. (2008) had proposed that adults prefer to be lead in their selection of interpretation by the Principle of Charity (Davidson, 1975) while for children this is overridden by pragmatic felicity. Moscati et al. propose that there is an additional stage: children shift from relying on pragmatics to having a preference for the true interpretation of a sentence, but initially children move from a stage in which they are only able to access one scope interpretation. They call this stage Unique Scope Assignment (USA). These last studies have revived the possibility that some readings are simply unaccessible to children. In the case of quantifier + negation sentences, the issue had been resolved with a more thorough control of the context these sentences appeared in, but the same contextual manipulations did not work for modal verb + negation sentences.

2.2.6 Summary

To sum up the studies reviewed so far, in contrast with what was evidenced by the early studies on children's ambiguity resolution - which had given rise to the idea that children are unable to compute inverse scope - it is now acknowledged that children are in fact able to access inverse scope interpretations. Different explanations have been proposed to justify the fact that a difference between children and adults had shown a difference in the first place: some researchers have referred to contextual factors, and in particular to the role played by the context determined Question Under Discussion: Gualmini et al. (2008) claim that, for an ambiguous sentence, children are only able to accept interpretation that answer the current QUD. It has indeed been shown that

children and adults can access the same interpretations once the effect of the QUD has been accounted for. Sentences containing a modal verb and a negation, however, seem to challenge this hypothesis, as children are generally unable to obtain the adult-like interpretation of an ambiguous sentence, even when it represents a good answer to the QUD (Moscati et al., 2016, Moscati and Gualmini, 2008, Gualmini and Moscati, 2009). Processing factors are also thought to play a role: it is possible that the surface scope interpretation is preferred by the parser, with the inverse scope one coming necessarily at the cost of a reanalysis, a process which might be easier for adults than for children (Viau et al., 2010, Musolino and Lidz, 2006).

The next section moves onto doubly quantified sentences, which have been an object to study in relation to scope ambiguity resolution in several studies that have aimed to shed light on which factors contribute to the selection of one interpretation instead of the other.

2.3 The scope of doubly quantified sentences

Literature on children's scope ambiguity comprehension has mostly focused on sentences containing a quantifier and a negation or, more recently, sentences with a modal verb and a negation. However, sentences with more than one quantifier present the same ambiguity.

(26) A doctor took each pencil.

a. (Surface scope reading, *a* > *each*):
A single doctor took each pencil.

b. (Inverse scope reading, *each* > *a*):
Each pencil was taken by a (possibly different) doctor.

As briefly mentioned in section 2.2, children's quantifier comprehension has been investigated in several studies (for example Bucci, 1978, Brooks and Braine, 1996, Philip, 1994), but none of them has been specifically concerned with scope relations. In addition to this, adults' interpretation of sentences with a quantifier and a negation seems to be uncontroversial from the studies reviewed so far: they are able to access inverse scope and surface scope alike. This is not always the case for doubly quantified sentences:

sentences similar to (26) have received a great deal of attention from researchers, and experimental studies have consistently evidenced that adults have a preference for the surface scope interpretation. Its inverse scope interpretation seems to be harder to obtain, even in presence of a supportive context, and even when it is obtained, its interpretation appears to be associated to a processing cost (Gillen 1991, Filik, Paterson, and Liversedge 2004, Anderson 2004 among others).

Over the years, researchers have been trying to uncover why one interpretation is preferred over the other. Various factors have been proposed: early studies have focused on the different lexical nature of quantifiers, claiming that some of them are more prone to wide scope (Ioup, 1975, VanLehn, 1978). Other studies have focused on how quantifiers are represented in the mind of the speaker, exploring the possibility that some configurations are easier to represent than others (Fodor, 1982, Altmann and Steedman, 1988). Several experimental studies have investigated this topic experimentally (Gillen 1991, Anderson 2004, Kurtzman and MacDonald 1993, among others), focusing in particular on the role of processing. Recall that it has been proposed that children tend to select to the surface scope interpretation because the inverse scope one might be harder to process (Viau et al., 2010, Musolino and Lidz, 2006): a similar explanation could hold for adults' behaviour with doubly quantified sentences. Another factor that has received attention in recent times is the role of the context that, while thoroughly considered for the experiments related previously, had not been investigated systematically for doubly quantified sentences until recently. This section provides an overview of the studies on doubly quantified sentence comprehension. Investigating different approaches to the resolution of scope ambiguity in doubly quantified sentences is particularly relevant for the goal of the present study, in order to decide which factors are worth taking into account and the choice of methodology to employ.

2.3.1 A hierarchy for scope taking operators

One possibile hypothesis regarding quantifier scope ambiguity is that the resolution is affected by the lexical nature of the quantifier used: in particular, some quantifiers might be more likely to take scope over the others. Ioup (1975) collected intuitive judgments on a variety of ambiguous doubly quantified sentences, some of which are exemplified in (27).

(27) a. I saw a picture of each child.
b. She knows a solution to every problem.
c. Ethel has a dress for every occasion.

Informants were asked to rate the sentences on a scale according on whether they perceived each sentences leaned towards one interpretation or the other. Relying on her own intuitions as well as the ones of her informants, Ioup proposed the hierarchy in (28) for quantifiers, based on their tendency to take wide scope over other quantifiers. The hierarchy only includes plural quantifiers, but she also proposes that *a* and $some_{sg}$ would have to be placed between *every* and *all*.

(28) Each > every > all > most > many > several > $some_{pl}$ > a few.

Another factor Ioup thought to be affecting the likelihood of a quantified NP taking scope over another is its grammatical function. In particular, she observed that subjects are in general likely to take wide scope over the rest of the material. She ordered grammatical functions in a hierarchy, similarly to what was done for quantifiers.

(29) Deep and surface subject > deep subject/surface object > indirect object > preposition object > direct object.

The idea that quantifiers may be ordered according to their likeliness to take wide scope was adopted by VanLehn (1978), who used two different methodologies to determine which quantifiers are more wide scope prone. The first methodology consisted of conducting a corpus analysis of doubly quantified sentences, resorting to the context (and sometimes directly asking the author of the passage) to determine the intended scoping. The second part of his study employed a sentence comprehension task. Participants were presented with a set of doubly quantified sentences written on small cards to avoid any effect of intonation, and asked to read them silently. After this, participants would be asked to provide a paraphrasis for the sentence just read. If their paraphrasis did not disambiguate the scope relations assigned, they were asked additional questions by the experimenter that would help clarify their interpretation. One example of a sentence with possible questions associated to it is reported below.

(30) Every guy kissed a girl.

(31) a. Did they all kiss the same girl?
b. If there are 5 guys, how many girls does this imply got kissed?
c. Is there a different girl per guy?

VanLehn reports that participants found answering these questions rather difficult, which in itself might suggest that participants had not initially computed the scope relations of the sentence, leaving the ambiguity unresolved[4]. As the responses received for each sentence varied greatly depending on the quantifier used (e.g. the universal a had a stronger tendency to take wide scope when paired with *all* rather than *each*), he proposed that lexical properties of quantifiers must play a role in scope assignment, reaching the same conclusion as Ioup (1975). VanLehn proposed the hierarchy in (32) for quantifiers.

(32) Each > every > all of the > all the > other plural quantifiers.

The claim that lexical properties of the individual quantifiers might affect scope assignment has proven relevant for subsequent experimental studies, including the present one, as researchers have been concerned with finding out whether supporting factors might make the dispreferred inverse scope interpretation more likely. The next section moves onto sentence processing, as processing costs and limitations are often believed to play an important role in ambiguity resolution.

2.3.2 Processing and scope assignment

Over the years, different models of sentence comprehension have been developed, and they account for ambiguity resolution in different ways. One of the major distinctions that can be identified across different models is in the way they predict multiple readings are built in the mind of the hearer: for our purposes, we can divide them into two broad categories. One of them includes models that propose that a number of syntactic

[4]The idea that the ambiguity resolution might be left suspended by the hearer until prompted by the context or a question has been explored in various subsequent studies. In particular, among the scholars that supported this view that scope is not computed unless it is necessary are Fodor (1982) and Gillen (1991).

principles guides the processor into building a single interpretation for a sentence, even an ambiguous one. One such proposal is the Garden Path model (Frazier and Fodor, 1978, Frazier and Rayner, 1982, Rayner, Carlson, and Frazier, 1983, a.o). The interpretation selected for an ambiguous sentence is always the one that corresponds to the simplest possible structure; and other factors, like contextual plausibility, are only taken into account at a later stage. These factors are in some case able to trigger a reanalysis of processed material, and therefore generate another interpretation of the sentence in analysis, but this process comes at a cost. On the other end of the spectrum are those theories that claim that the processor is able to take into account multiple types of linguistic/extra linguistic information simultaneously, including syntactic structure, morphological cues, lexical frequency and contextual plausibility (Altmann and Steedman 1988, Crain and Steedman 1985 a.o.). The processor therefore constructs several possible analyses of the sentences in parallel, and evaluates them as it goes along to then select one interpretation. The task is relatively easy when all the factors point towards one analysis, but becomes more costly when multiple reading of a sentences are equally supported. Researchers have tested these theories experimentally. One of the claims made by the Garden Path theory is that the parser prefers to structure incoming lexical material together with material already received (Frazier and Rayner, 1982, Frazier, 1978) (*Late Closure* strategy). According to this claim, in sentences like (33-a) and (33-b), the temporarily ambiguous NP *a mile* will initially be analysed as the direct object of the verb *jog* as in (33-a). When the parser encounters subsequent disambiguating material, like in sentence (33-b), it will be forced to go back to the ambiguous material and try a reanalysis to make sense of the sentence. Therefore, the Garden Path theory predicts more processing complexity for (33-b).

(33) a. Since Jay always jogs a mile this seems like a short distance to him.
b. Since Jay always jogs a mile seems like a short distance to him.

Another claim is that any incoming material will be attached by the processor into the phrase-marker being constructed using the fewest nodes (*Minimal Attachment* strategy). This is exemplified by a reduced relative clause like (34).

(34) The defendant examined by the lawyer turned out to be unreliable.

The verb *examined* is temporarily ambiguous, as it could either be a past tense (therefore making the preceding NP an agent, and the sentence a main clause), or a participial verb taking the preceding NP as theme in a reduced relative clause. Ferreira and Clifton (1986) monitored eye movements while subjects read sentences like (34), or (35) which begins with an inanimate noun.

(35) The evidence examined by the lawyer turned out to be unreliable.

They found no difference in the interpretation of (35) and (34): reading times were longer for the reduced relative sentences compared to the unreduced ones similar to (33-a), regardless of whether the first noun was animate or inanimate. The fact that the inanimate NP the evidence is a poor fit for the agent role had no effect on the syntactic analysis. They take this to show that contextual plausibility does not intervene in the parsing of the sentence, at least during the first pass reading.

Among the studies that favour parallel processing, some have placed emphasis on the role of referential context for the resolution of the ambiguity (Altmann and Steedman, 1988, Crain and Steedman, 1985). Altmann & Steedman suggest that, while different analyses are evaluated in parallel for plausibility against the discourse context, the one that best satisfies the "Principle of Parsimony" will be selected: "a reading that carries fewer unsupported presuppositions will be favoured over one that carries more" (Altmann and Steedman, 1988, p. 203). This notion is extremely relevant for doubly quantified sentences, as the Principle of Parsimony predicts that a sentence like (36) below is more parsimonious when interpreted on its surface scope, at least in absence of context, as it only requires one individual (one doctor) to be added to the discourse model; as opposite to the inverse scope interpretation, which would require the addition of several more entities.

(36) A doctor visited every patient.

The studies on sentence comprehension mentioned so far were conducted on adults, with methodologies (measuring participants' eye movement while reading a sentence) which are not suited for conducting the same kind of research on children, who might not have acquired the ability to read yet. To investigate whether the same observation made for

adults could hold for children, a different kind of experimental task was designed, requiring children to act out the instruction provided by a sentence containing a temporary syntactic ambiguity, like (37) (Trueswell, Sekerina, Hill, and Logrip, 1999).

(37) Put the frog on the napkin in the box.

(37) is temporarily ambiguous at the "on the napkin" region because the PP can be a Destination, or a Modifier of the previous phrase. The aim of Trueswell et al.'s (1999) study was to investigate whether children are subject to the Garden Path effect; and whether they are able to generate a new interpretation for the sentence once the disambiguating material is encountered. This experiment would also provide information regarding whether children are able to make use of pragmatic information, and in particular the Referential Principle formulated by Crain and Steedman (1985), which claims that when presented with a sentence, adults tend to look for referents in the immediate context. Tanenhaus, Spivey-Knowlton, Eberhard, and Sedivy (1995) had already noted that adults are quickly able to make sense of the phrase "in the box" when the context provides a single possible referent for the preceding material (i.e. there is a single frog, and it is placed on a napkin). In a context with an additional frog not placed on a napkin, and an empty napkin, participants often mistook the phrase "on the napkin" as a Destination, as proved by their eye-movements; but were able to revise it.

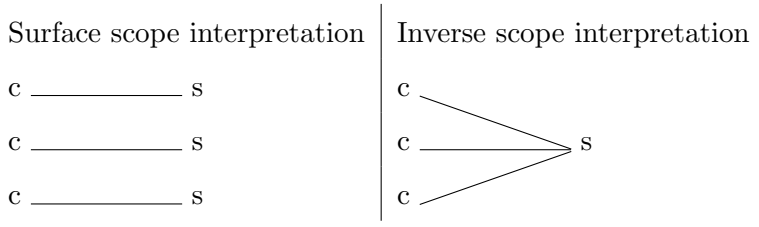
However, Trueswell et al. found no significant difference across different contexts for children: they seemed insensitive to the Referential Principle as they preferred the Destination interpretation for the phrase "on the napkin", and were generally unable to revise this commitment. Trueswell et al. dub this phenomenon the "Kindergarten Path effect". The Kindergarten Path effect has been held responsible for children's difficulties with inverse scope uncovered by the experiments illustrated in the previous sections, and processing factors have been proposed as an explanation alongside to the Question and Answer Requirement. While the QAR places all emphasis on the effect of contextual felicity, researchers have pointed out that it might be the case that contextual information interacts with processing: pragmatic felicity might work as a trigger to help the reanalysis of the material already processed (Viau et al., 2010). The next section moves onto the processing of quantifiers, focusing on the model proposed by Fodor (1982), as these elements are crucial for the work of this thesis.

2.3.3 A mental model for quantifiers

Amongst the research that tried to find an explanation for processing difficulties, Fodor's (1982) work tried to capture the nature of the mental representation of doubly quantified sentences containing a universal and an existential quantifier. She proposed that the interpretation that is generated for an ambiguous sentence could be determined from the amount of mental work required to get it. Building on linear order theories (Johnson-Laird and Tagart 1969, Bunt 1985, Kroch 1974 amongst others), which claim that quantified NPs are likely to take scope over one another according to their left-to-right order in the surface structure of the sentence they appear in, Fodor's model predicts that the earlier a quantifier appears in a sentence, the easiest it is for it to take wide scope. For example, in the case of a universal quantifier such as *every* it would require no particular effort to extend its scope over the rest of the sentence once it is encountered; while extending it backwards over material encountered before would require special mental computations. She also allows that there are differences between quantifiers, with some quantifiers (e.g. *each*) more prone to take wide scope, as already proposed in previous literature (Ioup, 1975, VanLehn, 1978). Linear order and quantifier type would therefore interact to determine scope preference.

Fodor further claims that doubly quantified sentences are associated to a non-linguistic mental model of representation. So for example, a sentence like (38) will generate two possible representations, corresponding to its two different scope readings, represented below.

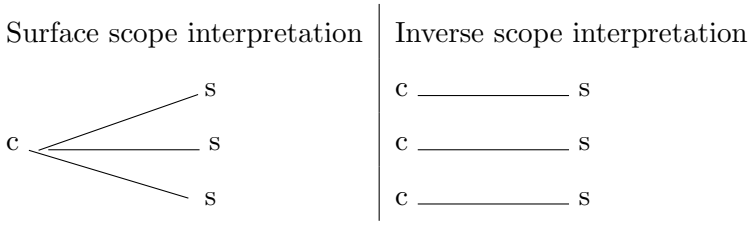
(38) Every child saw a squirrel.



Although Fodor claims these kinds of diagrams are not necessarily present as a mental image in the mind of the hearer, he argues that they are nevertheless the real form of the semantic representation of quantified sentences. In this light, the stronger difficulty

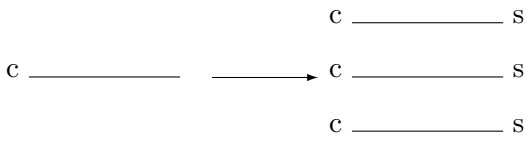
associated with the inverse scope reading of an sentence with an existential quantifier in subject position can be explained with the necessity of one intermediate step.

(39) A child saw every squirrel.



During the on-line computation of (39) and similar sentences, the existential quantifier would at first be given a singular interpretation, which then needs to be revised when the universal quantifier is encountered later in the sentence in order to obtain an inverse scope interpretation.

(40)



The surface scope interpretation is exempt from this problem, therefore less costly and preferred. This observation can also account for the fact that inverse scope seems easier to get for sentences with a universal quantifier in subject position, as in that case a multiple referent has already been established. Fodor's model is in some ways similar to the model of sentence comprehension proposed by Altmann and Steedman (1988) and Crain and Steedman (1985) briefly mentioned in the previous section, as they both predict difficulties accommodating more than one entity in the discourse model when a universal quantifier is encountered after an existential. These difficulties should be mitigated if the context has already introduced multiple referents (which are therefore already established in the mind of the hearer), however studies aimed at testing this hypothesis (Anderson, 2004) did not provide a confirmation for it.

2.3.4 Summary

The studies so far have shown that a variety of factors might concur in determining which interpretation of an ambiguous sentence gets selected. For scopally ambiguous sentences containing two quantifiers, one of the major factors at play seems to be the lexical properties of the quantifiers involved (Ioup, 1975, VanLehn, 1978), with some quantifiers more likely to take wide scope over others. Other possibilities that have been considered relate to how sentences are processed. According to some theories, the surface scope interpretation of an ambiguous sentence is formed first into the mind of the hearer, and then possibly discarded when contextual clues and other factors make this first reading impossible or unlikely, making it easier to obtain. The inverse scope interpretation on the other hand might require a reanalysis of material already computed after discarding the previous interpretation, making it more costly and harder to get. This hypothesis has been proposed as a reason for children's failure to obtain inverse scope in some of the studies detailed in section 2.2 (see Viau et al. 2010, Musolino and Lidz 2006), but several studies on adults have reached the same conclusion, as it will be explained in the next section. Another possibility for processing sentence ambiguity is that more than one reading is constructed at the same time in the mind of the hearer, and therefore more than one possible explanation is evaluated against contextual clues. This hypothesis also predicts that surface scope is more easily computed than inverse scope for sentences like *a doctor took each pencil*, as the latter requires multiple entities (e.g. more than one doctor) to be added to the mental model present in the mind of the hearer.

The section that follows reviews a few studies aimed to investigate the comprehension of scopally ambiguous questions containing two quantifiers.

2.4 Adults' interpretation of scope ambiguity: experimental studies

The studies conducted by Ioup (1975) and VanLehn (1978) on quantifier scope had relied significantly on intuitions, coming either from informants or from the researchers themselves. More recent studies, however, implemented different experimental techniques, often aimed at collecting judgements from naive participants through questionnaires or

other forms of comprehension tasks. The nature of the task varies: some aimed to elicit from the participant an acceptance or a rejection for a given interpretation (presented in the form of a diagram or paraphrase for the ambiguous sentences), while others require participants to choose amongst multiple options (e.g. participants are shown two possible diagrams or paraphrases, and asked to say which one is pertinent to the ambiguous sentence). Other methodologies do not require participant to give a conscious judgment about scope relations: a number of studies have employed methodologies such as measuring reading times or tracking eye movement to asses the grade of complexity associated to a reading: more complex interpretation should be harder to compute, and therefore require a longer reading time. Detailed below are some experiments that aimed to test the findings presented in the previous sections: what elements concur in the selection of one reading for an ambiguous sentence, how the selected interpretation is obtained, and whether inverse scope is linked with increased processing costs.

2.4.1 Sentence disambiguation and effects on the reading times

Most studies on ambiguity resolution for doubly quantified sentences have been aimed at investigating which reading of doubly quantified sentence is preferred by adults, whether one of the two readings is more complicated than the other to obtain, and what factors concur to its complexity. Many of these studies have reached the conclusion that multiple factors, including contextual information, general world knowledge and lexical properties of the words used, all influence which interpretation gets preferred by the hearer.

In one such study, Kurtzman and MacDonald (1993) tested participants on their interpretation of ambiguous *a...every* and *every...a* sentences, by having participants read first an ambiguous sentence like (41), and then one of two possible continuations, which disambiguated the first sentences towards its surface scope or inverse scope, like (41-a) and (41-b) respectively. Participants were then asked to judge whether the second sentence was a natural continuation of the first.

(41) A kid climbed every tree.

a. The kid was full of energy.

b. The kids were full of energy.

(42) Every kid climbed a tree.

a. The tree was full of apples.

b. The trees were full of apples.

They found participants were more likely to accept the surface scope continuations as natural, with an acceptance rate of around 80%. The difference with inverse scope continuation was more evident for the *a...every* sentences, for which inverse scope continuations were only accepted around 20% of the time; compared to *every...a* sentences, for which inverse scope continuations had an acceptance rate of 40%-60%, depending on the verb used. While these results support the idea that there is a preference for surface scope, Kurtzman & McDonald argued that the inverse scope interpretation is nonetheless considered in parallel by the hearer, as proposed by Crain and Steedman (1985) and Altmann and Steedman (1988). Amongst the factors influencing the ambiguity resolutions are the Single Referent Principle (which claims that the existential quantifier *a* is preferentially interpreted as referring to a single entity if encountered before the universal quantifier, similarly to what claimed by Fodor (1982)), and the Thematic Hierarchy Principle, which claims that agentive NPs have a stronger preference for taking wide scope. When multiple factors favour one reading, a clear preference will be shown towards that interpretation. However if the factors conflict, the preference will be weaker.

Gillen (1991) reached a similar conclusion, arguing that multiple factors are considered for scope ambiguity resolution. Based on her data, obtained across different tasks (e.g. self-paced reading, picture selecting, continuation sentence evaluation) and from sentences containing a variety of quantifiers, she argues that linear order, choice of quantifiers and kind of task all have an effect. Gillen also supports the possibility, originally introduced by VanLehn (1978), that participants might initially not compute scope beyond what's necessary, leaving the ambiguity open unless prompted to resolve it. She also found that singular continuations for universal first sentences, similar to (42-a) were read faster than plural continuations like (42-b), in contrast to what is predicted by Fodor's view and the linear order hypothesis.

Several other studies investigated the processing load associated to inverse scope by means of self-paced reading task involving an ambiguous sentences and a disambiguating continuation similarly to the task adopted by Kurtzman and MacDonald (1993). In

general, these studies observed slower reading times for the disambiguating elements in the inverse scope continuation sentences. For example, Tunstall (1998) reports slower reading times for the critical region (subject - verb) of sentence (43-b), that is to say, the inverse scope continuation of an existential first sentence (the vertical bars indicate the regions of texts into which it was divided).

(43) Kelly showed | a photo | to every critic | last month.

a. The photo was | of a run-down | building.

b. The photos were | of a run-down | building.

Inverse scope continuations for universal first sentences resulted in slower reading times compared to inverse scope continuations of existential first sentences like (44-b). Tunstall ascribes this to the "vagueness" of (44), which suggests, but does not require, a different critic for every photo - a view compatible with the idea that scope resolution is left open until it is necessary to resolve the ambiguity. Other self-paced reading studies (Filik et al., 2004, Anderson, 2004) show the same pattern.

(44) Kelly showed | every photo | to a critic | last month.

a. The critic was | from a major | art gallery.

b. The critics were | from a major | art gallery.

However, Filik et al. (2004) argued against the hypothesis that scope relations are possibly left unresolved, claiming that the ambiguity is resolved on-line. They tested this claim by testing reading times for the portion of an ambiguous sentence occurring before the disambiguating material that should prompt the reader to select one interpretation or the other. Kurtzman and MacDonald (1993) had proposed that multiple factors affect the ambiguity resolution, with sentences for which all factors point towards one interpretation being resolved more easily.

For example, based on the hierarchies proposed by (Ioup, 1975), *a* is expected take scope over *every*; and a Direct Object (DO) is expected to scope over an Indirect Object (IO). Fodor's (1982) observations also predict that the existential is likely to take wide scope when occurring first in the linear order of the sentence.

Fillik et al. conducted a self-paced reading experiment, dividing the test sentences into several sections, with the quantified NPs occurring in the second region (in bold in the example below). According to Gillen (1991), the ambiguity resolution should not have been triggered at the quantified region of (45-a,b) and similar sentences (critical regions in bold), as the disambiguating material has not been encountered yet, and therefore reading time for that part of the sentence should not be affected by the lexical and syntactical factors driving the resolution. If this is the case, the critical regions of sentences in which these factors conflict (like (45-b) and (46-a)) should not be read any slower than the critical regions of sentences in which the factors favour the same interpretation.

(45) DO-first sentences:

a. The celebrity gave | **an in-depth interview to every reporter from the newspaper, but** | the interview(s) was/were | not very | interesting.

b. The celebrity gave | **a reporter from the newspaper every in-depth interview, but** | the reporter(s) was/were | not very | interesting.

(46) IO-first sentences:

a. The celebrity gave| **a reporter from the newspaper every in depth interview, but the reporter(s) was/were** | not very | interested.

b. The celebrity gave | **every reporter from the newspaper an in depth interview, but— the interview(s) was/were** | not very | interesting.

In contrast with Gillen's hypothesis, Filik et al. report that reading times for the quantified region of (45-a,b) and (46-a,b) were slower when the scope-influencing factors conflicted, i.e. for sentences like (46-b) and (45-a).

In addition to this, there is evidence that inverse scope sentences might be harder to process, regardless of whether one interpretation is "forced" on the hearer by means of a disambiguating continuation. Anderson (2004) asked participants to read the sentence in (47) and answer the following question.

(47) A climber scaled every cliff.

(48) How many climbers scaled cliffs?

a. One

b. Several

(47) was read slower by participants who answered *several*. The results obtained in the self-paced reading studies mentioned above could be explained by presuming participants had committed early to the surface scope interpretation, and then had been forced to a reanalysis upon encountering the disambiguating continuation, as predicted by the Garden Path model (Frazier 1978 and others). On the other hand, the slower reading times obtained for (47) offered without disambiguating seem to signal an increase in processing cost intrinsic for inverse scope.

In summary, it is now widely believed that different factors influence scope ambiguity resolution; including choice of quantifiers, thematic role of the quantified NP, processing costs associated to the selected interpretation. It is still debated whether the ambiguity is resolved as soon as it is encountered, or whether its resolution might be suspended until the listener is prompted to commit to one interpretation or the other. Finally, while the processing difficulties associated with inverse scope interpretation have so far been ascribed to the necessity to revise the "default" surface scope interpretation initially obtained (Garden Path model); or to the necessity of assigning multiple referent to the existential quantifier (Parallel processing model), it is also possible that inverse scope interpretations are per se harder to get, due to an intrinsic increase of processing costs required to obtain them.

While some of the studies mentioned so far have made reference to a possible effect of the context on the interpretation of an ambiguous sentence, some recent studies have introduced a more systematic investigation of the role of the context, aimed at determining which elements might make the inverse scope interpretation more natural and easier to obtain. The next section discusses two such studies.

2.4.2 The effect of the context for doubly quantified sentences

Anderson (2004) conducted a study on how the presence of a context might influence the resolution of the ambiguity in *a...every* sentences, as well as the processing load associated with obtaining inverse scope (measured through reading times). Ambiguous

sentences were embedded in a short story which would bias participants either towards surface scope or inverse scope, and followed by a disambiguating continuation which might be in accordance or not the intended reading (sections in bold in (49) and following examples). Participants were asked to read the story, presented to them one sentence at the time (each line corresponds to one presentation region in the self-paced reading task). Once read the story, participants would answer question (53). The combined design of the experiment allowed Anderson to investigate multiple factors, as well as providing a measure of the processing cost associated to inverse scope.

(49) Surface scope biasing context, singular continuation:
With the increased popularity of adventure sports,
the cliffs outside Campbellton were becoming a popular destination.
One weekend, the climbing equipment shop sponsored
a show to demonstrate the sport.
While an announcer described the techniques,
an experienced climber scaled every cliff.
The climber was very skilled.
The shop's sales increased substantially the next weekend.

(50) Surface scope biasing context, plural continuation:
With the increased popularity of adventure sports,
the cliffs outside Campbellton were becoming a popular destination.
One weekend, the climbing equipment shop sponsored
a show to demonstrate the sport.
While an announcer described the techniques,
an experienced climber scaled every cliff.
The climbers were very skilled.
The shop's sales increased substantially the next weekend.

(51) Inverse scope biasing context, singular continuation:
With the increased popularity of adventure sports,
the cliffs outside Campbellton were becoming a popular destination.
One weekend, the climbing equipment shop sponsored
a race between climbing enthusiasts.

While an official timed the event,

an experienced climber scaled every cliff.

The climber was very skilled.

The shop's sales increased substantially the next weekend.

(52) Inverse scope biasing context, plural continuation:

With the increased popularity of adventure sports,

the cliffs outside Campbellton were becoming a popular destination.

One weekend, the climbing equipment shop sponsored

a race between climbing enthusiasts.

While an official timed the event,

an experienced climber scaled every cliff.

The climbers were very skilled.

The shop's sales increased substantially the next weekend.

(53) How many climbers scaled cliffs?

a. One.

b. Several.

Analysing the proportion of responses to the comprehension question (53), Anderson (2004) found that, for both contexts, participants chose the inverse scope response more often following the plural continuation (55%), while the singular continuation proved even more effective, leading to a surface scope response more than 70% of the time for both types of contexts. The context on the other hand did not have an effect on participants' answers. In line with previous experiments, plural continuations, which disambiguated to the inverse scope, were read significantly more slowly compared to singular ones. This seems to prove that the difficulty of assigning inverse scope is not mitigated by a supportive context. Recall that Altmann and Steedman (1988) had predicted that the inverse scope reading of an *a...every* sentence is difficult to get because it requires accommodating several entities – in this case, several climbers – into the discourse model, while the surface scope interpretation only requires accommodating one. They also predicted that such difficulty would be mitigated if the context had

already introduced the relevant entities so that they are already established in the mind of the hearer, but the data obtained by Anderson do not seem to confirm this as having a context which explicit mentions multiple climbers did not decrease reading times.

Kim (2013) aimed to conduct a more systematic study on the effect of discourse for doubly quantified sentences, and resolved to manipulate two factors: expectation and focus. The first kind of manipulation is achieved by influencing what the reader expects: if the reader is expecting an inverse scope reading, that reading should become more accessible. The latter involves making certain aspects of the inverse scope more salient in the context, to increase the likelihood that inverse scope is accessed. Below is an example of how these manipulations were achieved: the expectation context explicitly mentions the expectation of the narrator of the story, while the focal context introduces in the narration another possible set of objects (sentences highlighted in bold).

(54) Expectation context:
An environmental activity was held at a school today, and participants were making big houses out of cardboard. At last, I saw three big playhouses finished and displayed. Surprisingly, the igloo was the work of a 6th-grade boy, and the rocket was the work of a 5th-grade boy. When I saw the doll house, **I thought it could not be a work of a child**. However, it was the work of a little child, this time a 4th-grade girl!

(55) Focal context:
An environmental activity was held at a school today, and participants were making big houses out of cardboard. At last, I saw three big playhouses finished and displayed. Surprisingly, the igloo was the work of a 6th-grade boy, and the rocket was the work of a 5th-grade boy. When I saw the doll house, **I thought it could not be a work of a child**. However, it was the work of a little child, this time a 4th-grade girl! After I left, **I heard that four more houses were displayed, but none of them was a work of a child**.

In an off-line experiment, participants read one of the context stories, and then were asked to judge whether sentence (56) was a possible description of the story.

(56) Some child built every playhouse that I saw today.

Kim found that the focal context was more effective in making (56) and similar sentences acceptable (62% of the time, compared to 27% for the expectation context). She calls this effect *focal advantage*. However, she ran a replica of this experiment on a larger number of participants (80, against around 30 that had participated in the previous ones), and the addition of a "baseline" condition, not featuring any of the contextual manipulation highlighted in (54) and (55). The results were only confirmed for the expectation context, for which the acceptance rate scored around 30%. The acceptance rate of the test sentence for the focal condition dropped to 35% from the previous 62%.

Kim also tested the effect of priming on inverse scope acceptance. In a study similar to the one conducted by Viau et al. (2010) on children (discussed in section 2.2.3), participants were exposed to unambiguous sentences with the same truth conditions as the inverse scope interpretation of the ambiguous ones, like (57), before receiving the ambiguous trials.

(57) Some child or other built every playhouse that I saw today.

Viau et al. (2010) had found an effect of priming on children for quantifier + negation sentences: after being exposed to an unambiguous sentence with the same truth values as the inverse scope reading of the ambiguous one, they were much more likely to accept the ambiguous sentence, even when presented in a context which is thought to be unsupportive and had previously caused them to reject the same sentence. In Kim's experiment, though, the priming effect did not show: participants' acceptance of the the inverse scope sentences ranged between 26% and 35%, depending on which context was used (baseline, expectation or focal). Even the unambiguous prime items like (57) were not universally accepted: their acceptance rate was between 57% and 65% across the three experiments.

From the evidence found by the studies presented in this section it can be concluded that it is not proved that a supportive context has an effect on making inverse scope more available to adults for doubly quantified sentences, in contrast to what had been observed for children in the quantifier + negation studies related in section 2.2. While there is some evidence that the focal context devised by Kim (2013) increased the acceptance

of inverse scope, a replica experiment does not seem to confirm this; which has led to the hypothesis that the results of the first experiments might have been due to the fact that the majority of participants in that group belonged to a more "inverse scope prone" population, who is able to access inverse scope without difficulties regardless of contextual factors. A larger proportion of people seems to find inverse scope difficult to get, and the supportive context does not mitigate this. Finally, a small group of people fluctuate in their scope assignment, and are sensitive to the influence of discourse factors. This claim would need to be more thoroughly investigated to determine if it is true that there are such individual differences across people. What is clear from the work here presented is that, in line with with many of the previous studies, a large number of people find inverse scope harder to access for doubly quantified sentences, even in a context perceived to be supportive of this interpretation. Kim did not investigate processing factors, but claims the difference across participants might be due to "mental flexibility", that is to say, "how well participants can switch" (Kim, 2013, p. 200) between the two interpretations in a way similar to what is proposed by the Garden Path theory.

2.4.3 Summary

The studies reviewed in this section have been aimed at investigating adults' interpretation of scopally ambiguous sentences, focusing in particular on what elements concur in their disambiguation. It is generally agreed upon that various factors influence the choice of interpretation (Gillen, 1991, Kurtzman and MacDonald, 1993). Moreover, while some researchers had claimed that ambiguity resolution might only be triggered when necessary, leaving the sentence ambiguous when committing to one of the two interpretation is not necessary (Gillen, 1991, VanLehn, 1978), evidence obtained in self-paced reading studies seem to point away from this hypothesisf (Filik et al., 2004). Among the factors that play a role in scope ambiguity resolution, the context might have an effect in biasing participants towards one reading or the other (Anderson, 2004, Kim, 2013), but it does not seem sufficient to mitigate processing costs associated to inverse scope (Anderson, 2004).

2.5 Summary of chapter

The studies presented in this section investigated the processing of ambiguous sentences, with the goal of building a model of sentence representation, and to tease apart the contribution of different lexical, grammatical and contextual factors. In general, scholars agree that for doubly-quantified sentences the inverse scope interpretation is harder than the surface one to obtain and it is associated with a higher processing cost. While there have been various proposals on why this happens, the asymmetry between the two readings is not fully understood yet. Experiments on children have evidenced that they differ from adults in how they interpret sentences containing a quantifier and negation: children appear to struggle with obtaining the inverse scope interpretation. The reason of this has been ascribed both to processing and contextual factors. Experimental evidence provided confirmation for both, consistently with the idea that more than one factor concurs to scope resolution and in general to the interpretation of any ambiguous sentence. On-line experiments on adults, and in particular the ones conducted with the "continuation" method, have the merit of allowing researchers to observe details such as eye movement and reading times for inverse scope sentences, and have therefore been very valuable to reach the conclusion that inverse scope seems to come at a cost. However even such methodology is not free from criticism, as some linguists have argued that a sentence containing the definite article *the* does not necessarily disambiguate a doubly quantified sentence (Pupa, 2008). Consider, for instance, the case of the sentence in (58).

(58) Every doctor employs a receptionist. The receptionist has many responsibilities.

In (58), contextual plausibility makes it impossible that there is a single receptionist for all of the doctors, yet the sentence makes perfect sense, as the definite article can be interpreted as referring to the receptionist of each doctor.

In sum, however, a robust result in the literature is that in the case of doubly quantified sentences inverse scope is dispreferred, even when supported by contextual evidence. On the other hand, it is reported that adults are able to obtain this interpretation for sentences with a quantifier and a negation, while children need additional pragmatic

support. What appears largely unexplored in this literature is how exactly children interpret doubly quantified sentences. Researchers have been greatly concerned with their interpretation of quantifier + negation sentences and have uncovered some interesting features of children's behaviour in relation to scope, including the fact that pragmatic felicity seems to play a very important role for their interpretation. Further research is required to establish whether the difference observed between adults and children in experiments involving negative sentences is still present for sentences with two quantifiers, and whether context can be determinant for children for these sentences as well. The next section presents two experiments that have been conducted with this aim.

Chapter 3

Scope experiments

This chapter presents the results of two experiments on children's and adults' interpretation of scopally ambiguous sentences and questions containing two quantifiers, focusing in particular on the effect of the Question Under Discussion (QUD) for scope ambiguity resolution. Most of the literature on children's and adults' interpretation of scopally ambiguous sentences reviewed in Chapter 2 has investigated sentences containing a quantifier and a negation: the results obtained in these studies had initially shown that children are often unable to access the inverse scope of such sentences, while adults are generally capable of accessing both surface scope and inverse scope readings, leading to the conclusion that children's grammar differ from adults' in a way that makes it impossible for them to access inverse scope (Musolino, 1998).

Further investigation on the same kind of sentences later disproved this conclusion by showing that children are able to access the inverse scope, provided the sentences are presented in a more supportive context. Among the context based approaches, the Question and Answer Requirement came to prominence: this model is based on the notion that, to get selected, the interpretation of an ambiguous sentence must represent a good answer to a question made salient by the context. Gualmini et al. (2008) demonstrated that implementing this kind of contextual manipulation (i.e. presenting the sentences in a context that made relevant a QUD that would be answered by the inverse scope interpretation of the ambiguous sentence) made children capable of accessing the inverse scope of sentences including a quantifier and a negation. Having shown that children are able to access inverse scope at the same rate as adults, they justified previous results

that found a difference between children and adults by proposing that the management of the QUD was indeed the only factor responsible, with children only being able to access the interpretation that answers the QUD. However, this effect was not found for sentences containing modal verbs, as it has been shown that children tend to interpret these differently from adults, sometimes favouring interpretations altogether unavailable in adult grammar, apparently disregarding the QUD (Moscati, 2011, Moscati et al., 2016). These observations led to some doubts as to whether the QUD can be considered the only factor behind the difference initially observed between adults and children.

In order to further investigate these claims, two experiments were conducted on children and adults as part of this study, aiming to investigate how scope ambiguity is resolved in a situation in which supposedly the Question Under Discussion does not play a role. Such situation is the disambiguation of a scopally ambiguous *question* instead of a declarative sentence: while Gualmini et al. (2008) claim that the QAR is always in place, and therefore (following Roberts (1996)), every sentence in conversation is always understood as the answer to a QUD, they make no prediction as to what would happen with questions, since it is difficult to suppose that a question could represent the answer to another question.

3.1 Experimental studies

The two studies on scopally ambiguous sentences and questions presented in this section were conducted with the aim of investigating how scope ambiguity is resolved in a situation in which the Question Under Discussion does not play a role.

It needs to be remember, however, that the QUD approach is based on Roberts's (1996) discourse model, which in turn moves from the notion that discourse is entirely organised around question, with participants in conversation trying to reach the goal of sharing information - and answering the Big Question "what's the way things are" - by committing to answering the much narrower current QUD. Every sentence is therefore always offered, and understood, as an answer to the current QUD. In relation to scope assignment, the QAR claims that the effect of the QUD is always in place and there is no other difference between children and adults but in their management of the QUD:

children are always bound to answer it, rejecting any interpretation that does not provide an answer (i.e. entail either a *yes* or a *no*) The effect is so strong that they are also able to obtain interpretations that are unavailable in adult grammar, provided they are a good answer to an understood question (Gualmini, 2004a). The QAR holds for for adults as well, but for them other factors concur in the process of disambiguation: adults rely on the QUD to select an interpretation, but also show a preference for the true reading, influenced by the Principle of Charity (Davidson, 1975). In virtue of their more sophisticated pragmatic abilities, adults are able to set aside the question made relevant by the context to accommodate one that is answered by the true reading of the ambiguous sentence. In summary, the claim made by the QAR is that there is no other difference between children and adults but their management of the QUD, with children being entirely reliant on the one made salient by the context for their interpretation.

The premises for this study are based on the fact that the QAR makes no prediction as what would happen in a scenario in which the QUD does not play a role. One such domain is represented by ambiguous *questions*, since it is not obvious that a question could be received by the hearer as the answer to a QUD: therefore, it must be assumed that the interpretation of such question must rely on some other factor not yet investigated. The interpretation of ambiguous questions could therefore shed light on what these factors are.

Previous studies on the subject of scope ambiguity resolution in children and adults have generally used sentences containing a quantifier and a negation, like (1). For this reason, the interrogative form of such sentences, (1-a) or (1-b) were initially considered for this study.

(1) Every horse didn't jump over the fence.

   a. Didn't every horse jump over the fence?
   b. Did every horse not jump over the fence?

Interpreting such sentences in question forms, however, presented some difficulties: negative questions are subject to a bias that could influence the way in which they are answered: Romero (2005) observed that yes-no question with a preposed negation, like (1-a) convey the idea that the speaker thinks the positive answer is the true one. Sentences with a non-preposed negation like (1-b) do not carry the same bias, but both

negative questions however could generate a confound with their answer, since it is not clear what a *no* would mean - negating the content of the sentence, or reiterating the negation present in the negative questions (2-a) and (2-b).

(2) Did every horse not jump over the fence?

a. No, (every horse didn't jump over the fence).

b. No, (it's not true that every horse didn't jump over the fence).

In light of this, sentences with two quantifiers were preferred for this study. Doubly quantified sentences present the same kind of ambiguity as sentences containing a quantifier and a negation, and their interpretation has been investigated extensively, though mostly with adult participants. Most of these studies paired an existential quantifier (*a* or *some*) with the universal quantifier *every*, and it was consistently found that participants have a strong preference for surface scope, with the inverse scope interpretation being selected infrequently even in supportive contexts (Anderson, 2004, Kim, 2013), as well as being generally associated with longer reading times, signalling processing difficulties (Filik et al., 2004, Tunstall, 1998). Initial pilot trials for this study, conducted on a limited number of participants, showed that, following the same pattern as previous studies, participants failed to access the inverse scope interpretation of (3) and similar questions.

(3) Did a doctor take every pencil?

a. **Surface scope reading:** Is it true that there is a single doctor who took every pencil?

b. **Inverse scope reading:** Is it true that, for every pencil, a (possibly different) doctor took it?

For this reason, the quantifier *each* was preferred over *every* in the final version of the experiment, as *each* is considered more wide scope prone because of its lexical properties (Ioup, 1975, VanLehn, 1978).

(4) A doctor took each pencil.

Two experiments were conducted: Experiment 1 was aimed at testing children and adult participants on their interpretation of scopally ambiguous declarative sentences as well as questions. The ambiguous sentences were presented to participants as answers to an explicit question containing the same quantifiers (which was, therefore, also ambiguous), as exemplified in (5).

(5) a. Did a doctor take each pencil?
b. A doctor took each pencil.

While in this experiment participants are asked to judge the truthiness of the ambiguous declarative sentence provided as an answer, their interpretation of such sentence is dependent on the interpretation received by the question, providing therefore an indirect way of looking at scopally ambiguous questions.

Section 3.3 presents the results of a follow-up experiment in which participants were asked to answer the ambiguous question directly.

3.2 Experiment 1: Ambiguous doubly quantified sentences

Experiment 1 tested children's and adults' interpretation of scopally ambiguous sentences containing quantifiers *a* and *each* presented in an inverse scope-supportive context and preceded by an explicit QUD, in order to investigate whether adults and children are able to access inverse scope at the same rate.

The methodology chosen was a Truth Value Judgment Task (Crain and Thornton, 1998), featuring a total of 16 items, 4 of which were targets. The sentences were presented to participants as a description of a short story (4-5 sentences long), and participants were expected to judge whether the sentence was a true description of the story. The order of the quantifiers in the crucial sentences was held constant across trials (existential first): all target sentences had the structure in (6).

(6) A boy ate each fruit.

The order of quantifiers was kept constant for one main reason: this word order is considered the only one that truly allows to be sure that an inverse scope interpretation has been accessed. For the sentences in which the order of the elements is reversed, the inverse scope reading entails the surface scope one (cfr. Scontras, Tsai, Mai, and Polinsky, 2014).

(7) Every shark attacked a pirate

a. **Surface scope reading:** For every shark there is a (possibly different) pirate that it attacked.

b. **Inverse scope reading:** There is one pirate that every shark attacked.

If there is a single pirate that every shark attacked, as the inverse scope reading presupposes, it is also true that every shark attacked a pirate, albeit the same one. Therefore, a participant hearing (7) might accept the sentence without necessarily accessing inverse scope.

The stories were presented to participants with the aid of an animated PowerPoint presentation, shown to them on a laptop while the experimenter read through the script. Prior to the beginning, participants were given a brief explanation about the task, being told that they would have to judge whether what the puppet said was true or false based on what happened in each story. The stories always followed the same structure, with a small set of characters and one clear goal, to avoid any confusion, as having a more complex plot is thought to hinder the selection of a clear QUD (Gualmini et al., 2008).

3.2.1 Predictions

Following the QAR as proposed by Gualmini et al. (2008), we can predict that children and adults will not differ in their interpretation of scopally ambiguous sentences containing two quantifiers, provided that the context has made relevant a QUD that is answered by the inverse scope interpretation of the ambiguous sentence. Another claim, linked to the first, is that children and adults do not differ in their grammatical abilities - as shown by the fact that they can in some circumstances access inverse scope at the same rate as adults - they only differ in their pragmatic abilities, being unable to set

aside the QUD and accept an interpretation of the test sentence that does not answer it.

The aim of this experiment is to further investigate the claims made by the QAR, based on two assumptions: the first one is that children and adults only differ in how they receive and accommodate the QUD, and therefore they should be able to access inverse scope at the same rate in a situation in which the QUD has been removed. The second assumption is that the QUD can be removed by the use of a question instead of a declarative; since it is dubious that a question could be interpreted as the answer to another question.

Therefore, having established a controlled linguistic environment in which the QUD and other possible contextual factors are accounted for, we expect to find no difference between adults and children on their interpretation of sentences containing two quantifiers. However, it should be noted that the explicit QUDs used in this experiment contain the same two quantifiers as the following test sentences, and therefore they are also ambiguous between a surface scope reading an an inverse scope reading, as exemplified in (8).

(8) Did a boy eat each fruit?

a. **Inverse scope reading:** is it true that there is a single boy who ate each fruit?

b. **Surface scope reading:** Is it true that for each fruit there is a boy who ate it?

Given that the QUD is itself ambiguous, the interpretation of the test sentence should depend on which of the two interpretations is selected for the QUD: the judgment obtained for the ambiguous test sentence should also reflect the scope assignment for the ambiguous question.

3.2.2 Methodology

3.2.2.1 Participants

23 adults and 27 children (age range 4;11-6;06, mean 5;07) participated in the study. Children participants were recruited from primary schools of the Greater Belfast area. Participants who did not answer at least 75% of control and filler items correctly were excluded from the analysis. After exclusion, data from 21 adults and 22 children were analysed.

3.2.2.2 Materials and procedure

The first 4 trials were training items, designed to make sure participants had an understanding of the task. The training items did not contain any of the crucial elements, but otherwise followed the same structure as test and control items. Two of the sentences used in the training items were true, and two were false. An example is provided below.

(9) The bear and the dog are having a picnic. They want to eat the muffin, but there is only one. In the end, the dog decides to eat a cookie, so the bear eats the muffin.
Experimenter: Did the bear eat the muffin?
Puppet: The bear ate the muffin.

(a) Example of a training item for Experiment 1: The bear ate the muffin.

(b) Outcome of training item.

FIGURE 3.1: Training item: The bear ate the muffin.

The 4 training items were followed by the 12 main trials, of which 4 targets, 4 fillers and 4 control items containing only one of the quantifiers used in the targets (i.e. *a* or *each*).

The items were arranged in pseudo-random order, but each test item was followed by a filler to balance the number of *true* and *false* answers. Moreover, the *a* controls were displayed at the end of the presentation, as it will be discussed below. The 4 controls were designed to test participants understanding of the quantifiers when they appeared in isolation: for this purpose, the other quantified NP was replaced by a proper name. An example of a control story with its associated QUD and test sentence is given for each quantifier.

(10) ***Each* control story:** Anna is a doctor. She is going to visit some patients: a boy, a man and an old woman. Anna visits the boy and gives him some medicine. Then she visits the man and gives him the medicine too. However she realises that there is no time to visit the last patient, the old woman, so Anna does not visit the old woman, and the old woman has to leave.
Experimenter: Did Anna visit each patient?
Puppet: Anna visited each patient.

(a) Example of a control item (*Each*) for Experiment 1: Anna visited each patient.

(b) Outcome of control story.

FIGURE 3.2: Control *Each* (false): Anna visited each patient.

(11) ***A* control story**: Lucy goes to a restaurant. She is waiting for someone to come and speak to her. There are some waiters in the room. Finally, a waiter comes and greets Lucy.
Experimenter: Did a waiter greet Lucy?
Puppet: A waiter greeted Lucy.

In the *a* control story, the quantified NP refers to a single individual (one waiter). This is incompatible with the reading we were trying to get for the test sentences. Recent

(a) Example of a control item (*A*) for Experiment 1: A waiter greeted Lucy.

(b) Outcome of control story.

FIGURE 3.3: Control *A* (true): A waiter greeted Lucy

studies (Viau et al., 2010, Kim, 2013) have evidenced how it is in some cases possible to prime a certain interpretation of a scopally ambiguous sentences using unambiguous sentences with the same truth values as the ambiguous one. In this case, it is possible that having sentences in which the quantifier *a* is taken to refer to a single individual would have biased participants against its narrow scope interpretation when it appears in conjunction with *each*, and is therefore referring to more than one entity. For this reason the two *a* controls were always the last items participants would see.

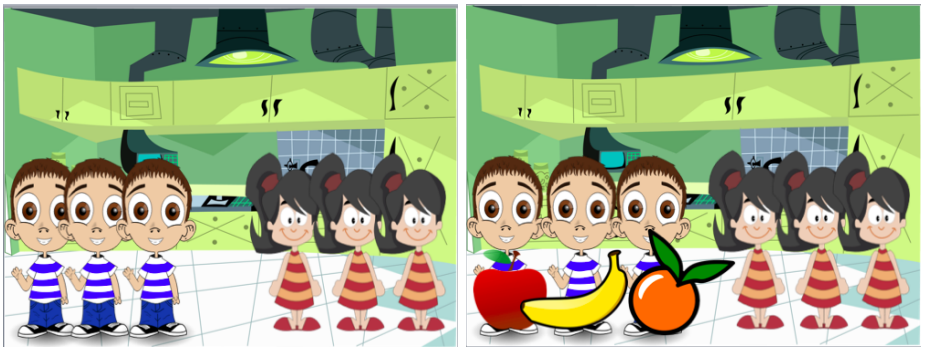
As for the 4 targets, each one of them was associated to a story designed to make the inverse scope interpretation of the following sentences plausible. The stories always introduced more than one character (precisely, two sets of three characters each) from the beginning of the story, to avoid possible problems presented by the Principle of Parsimony (Altmann and Steedman, 1988), which claims that the processing difficulties associated to the inverse scope of doubly quantified sentences are due to the fact that multiple entities need to be added to the mental model of the hearer (e.g. one boy for each of the fruits) compared to the surface scope interpretation that only requires adding one entity (e.g. one single boy). Having multiple characters represented from the beginning should mitigate this difficulty: in the stories used for this study, the two sets of characters considered performing an action on a set of objects, but in the end only one group of characters performs the action.

(12) The boys and girls want to have a snack. They go looking for something to eat. In the kitchen there are three pieces of fruit: an apple, a banana and an orange. Both the boys and the girls want to eat something. The boys get to the kitchen

first. One boy eats the apple, another boy eats the banana. The third boy eats the orange. There is no fruit left for the girls so the girls don't eat any fruit.

Experimenter: Did a boy eat each fruit?

Puppet: A boy ate each fruit.



(a) Example of a target item for Experiment 1: A boy ate each fruit.

(b) Outcome of target item.

FIGURE 3.4: Target item: A boy ate each fruit.

Another precaution taken to make sure that the context would make the right QUD plausible was the statement of a clear goal for the characters. As noted by Gualmini et al. (2008), in previous studies on children's ability to access inverse scope, some of the stories used as context for the target sentence had a "shifting goal": instead of committing to one single goal and maintaining it throughout the story, the characters would consider one initial goal, but then settled for another. One example of this can be found the stories used in Musolino's (1998) study of children comprehension of sentences containing a quantifier and a negation.

(13) This story is about three horses who have decided to practice jumping. There is a barn and a fence over which the horses will try to jump. The horses first consider jumping over the barn. They start running toward it but as they get closer, realize that the barn is too tall for them to jump over. Since they do not want to hurt themselves, they decide not to jump. Next, the horses consider jumping over the fence which doesn't look as tall as the barn, and they line up in front of it. The first horse clears the fence. Great jump! The second horse also clears the fence. Nice jump! The third horse gets ready to go but remembers that he hurt his leg the day before and judges it wise not to take any risks: his

leg needs to rest. He therefore decides not to jump.

Test sentence: Every horse didn't jump over the fence

As the three horses in Musolino's story considered first jumping over a barn, and then over a fence, Gualmini et al. claim that this change made it more complicated to establish a clear QUD, hindering children's access to inverse scope. In order to avoid this, the target story for this study, (12), mentions an explicit goal ("the boys and the girls want to have a snack") which is then maintained for the rest of the story, with no mention of other possible actions characters might undertake.

Another feature of the context stories used in this study was to have two groups of characters concurring towards the same goal ("both the boys and the girls want to eat something"). In her study of the role of context in the interpretation of doubly quantified sentences, Kim (2013) used the same strategy in some of her trials (which she calls *focal context* condition). Compared to a baseline condition with only one set of characters, this contextual manipulation seems to have had a significant effect in improving the rate of acceptance for inverse scope sentences, although a repetition of the experiment did not hold the same results.

3.2.3 Results and discussion

Data analysis were conducted on the proportion of true answers (which signal an inverse scope interpretation of the test sentence) and accuracy for controls and fillers.

No significant difference was found between adults and children in their rate of inverse scope responses to the test sentence: both groups judged the sentences true at the same rate (91,7% for adults and 92% for children, Mann-Whitney U = 3682, p = 0.928).

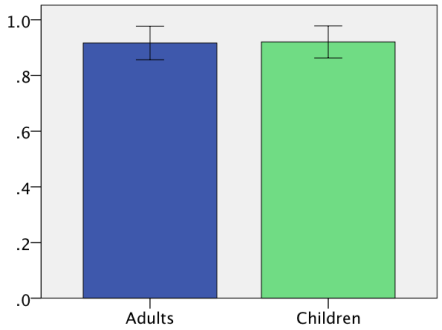


FIGURE 3.5: Experiment 1: Acceptance rate for test sentences.

Filler and control items, analysed jointly, had an accuracy rate of 94.6% for adults and 95,5% for children, proving that participants were able to give both true and false answer, and had a good understanding of quantifiers in isolation.

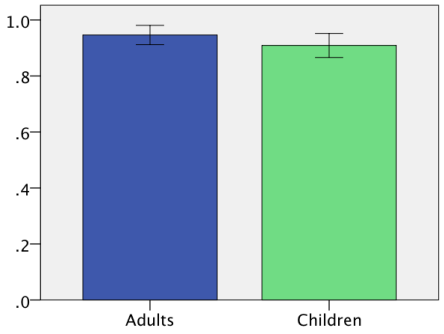


FIGURE 3.6: Experiment 1: Accuracy rate for control and filler items.

The prediction made by the QUD approach as proposed by Gualmini et al. (2008) was that children and adults would not differ in their interpretation of scopally ambiguous sentences once the right QUD is provided. The results for this experiment confirm this prediction, as children were adult-like in their ability to access inverse scope for ambiguous sentences containing two quantifiers when these are presented in a context making salient their inverse scope interpretation. The explicit QUDs used in this experiment contain the same two quantifiers as the following test sentences, and therefore they were also ambiguous between a surface scope reading an an inverse scope reading, as exemplified in (14).

(14) Did a boy eat each fruit?

a. **Surface scope reading:** is it true that there is a single boy who ate each fruit?

b. **Inverse scope reading:** Is it true that for each fruit there is a (possibly different) boy who ate it?

Accepting the test sentence is subordinate to the QUD receiving an inverse scope interpretation: therefore, it can be inferred that participants interpreted the ambiguous QUD, as well as the test sentence, on its inverse scope interpretation; as only asking (14-a) can lead to the inverse scope interpretation of the test sentence (paraphrase in (15)) to be a true and relevant answer.

(15) Each fruit was eaten by a (possibly different) boy.

This detail is of particular interest: the inverse scope interpretation obtained for the declarative sentence can be justified by appealing to the disambiguating effect of the QUD, but how the interpretation of an ambiguous *question* is selected is not clear under the present assumptions. Building on Roberts's (1996) model of discourse, the QAR predicts that the declarative sentence is understood as an answer to the current QUD, and, in the case of an ambiguous sentence, this translates to the fact that the interpretation that gets selected is the one that provides a good answer to the QUD (Gualmini et al., 2008). In the QAR model, any difference found between children and adults is attributed to the QUD and its management.

However, ambiguous questions are currently unaccounted for by the QAR, which makes no prediction as to how they will be interpreted, or as to what factors drive their interpretation. Effectively, it is not sure that the interpretation of a question could be explained by a similar model. While it seems sensible to assume that a declarative sentence can be taken to be the answer to a question, it seems counterintuitive to suppose that a question could itself be interpreted as the answer to some other QUD. In addition to this, studies on modal verb - negation interaction have shown that children are unable to access some readings even when these are good answers to the QUD (Moscati, 2011, Moscati et al., 2016), casting some doubts over the hypothesis that the management of the QUD might be the only explanation behind the difference observed in previous studies.

No difference between adults and children was found in the current study for the interpretation of scopally ambiguous declarative sentences, and indirectly, questions. While the current formulation of the QAR does not provide an explanation to the latter, a proposal can be attempted based on this framework, building on two main assumptions. The first one is that a question cannot be interpreted as the answer to another question, and therefore an interrogative sentence is not influenced by a QUD. The second assumption, taken from the "QUD only" approach to scope assignment is that the effect of the QUD is the only factor influencing the disambiguation of a scopally ambiguous sentence and hence relevant to the difference between children and adults. The implication of this last assumption is that the lack of a QUD caused by having an ambiguous question

instead of a declarative sentence erased this difference entirely, making the inverse scope reading accessible to both groups at the same rate.

The next section presents the results of a follow-up experiment designed to test participants on their interpretation of scopally ambiguous questions more directly.

3.3 Experiment 2: Ambiguous questions

Experiment 2 tested children's and adults interpretation of the ambiguous questions used as QUDs in Experiment 1, in order to investigate more directly how scope ambiguity is resolved in ambiguous questions as opposed as declarative sentences.

The results obtained for Experiment 1 have proven that children are able to access the inverse scope interpretation of scopally ambiguous sentences containing the quantifiers *a* and *each*. Following the QAR approach, it must be assumed that the disambiguation of said sentences relied on their context, and in particular on the explicit Question Under Discussion that preceded each sentence. However, said QUD was also ambiguous, containing the same two quantifiers as the declarative test sentence: as it has already been discussed in section 3.2.3, holding the assumption that participants relied on the QUD to disambiguate the test sentences, it must also be presumed that the preceding QUDs were themselves interpreted on their inverse scope interpretation, because if it were otherwise the test sentences couldn't have provided a coherent and true answer.

(16) a. Question: Did a boy eat each fruit?

b. **Inverse scope reading:** Is it true that, for each fruit, a (possibly different) boy ate it?

(17) a. A boy ate each fruit.

b. **Inverse scope reading:** Each fruit was eaten by a (possibly different) boy.

(16-a), on its inverse scope interpretation (16-b) is answered by (17-a) on its inverse scope interpretation (17-b), as the latter counts as a good answer under Roberts (1996) - i.e. it entails either a *yes* or a *no*, in this case a positive answer. So, under the QAR the

interpretation obtained for the declarative test sentence can easily be explained by the effect of the explicit QUD used. The QAR model, however, does not make a prediction on how ambiguous *questions* are interpreted by adults and, crucially, by children since the effect of the QUD is supposed to be the only factor determining which interpretation gets selected by the latter. It is not intuitive that a question could itself be taken as the answer to another question in Roberts's model, yet both adults and children were able to access its inverse scope interpretation in the vast majority of cases: these results are unaccounted for by the QAR in its current formulation.

Experiment 2 is aimed at testing children and adults' interpretation of ambiguous questions directly, in order to further investigate how their interpretation is obtained.

3.3.1 Predictions

In section 3.2.3, a possible explanation for the results of Experiment 1 was discussed, based on the assumption that the interpretation of an interrogative is not influenced by a QUD, and therefore ambiguous question would be a domain in which the QUD does not play a role. Based on this, a possible explanation for the results is that once the QUD is removed, so is the difference between children and adult, making the two groups equally likely to access inverse scope.

These assumptions serve as a starting point for Experiment 2: if this experiment were to yield the same results as Experiment 1, providing a further confirmation that adults and children are able to access the inverse scope of scopally ambiguous questions, this must be taken as a confirmation that the interpretation of an ambiguous question is indeed not driven by the QUD, and therefore the only point to difference between adults and children is removed. Alternatively, it is possible that there are other factors affecting the interpretation of an ambiguous question, and therefore QAR might be extended to include interrogatives.

3.3.2 Methodology

3.3.2.1 Participants

27 adults and 18 children (age range 4;03-6;09, mean 5;05) participated in the study. Participants who did not answer at least 75% of control and filler items correctly were excluded from the analysis. Data from 27 adults and 15 children were analysed after exclusion.

3.3.2.2 Materials and procedure

Test, control and filler items were the same as Experiment 1, with the only difference that instead of hearing a question - answer pair and being asked to judge the latter in terms of truthiness and falsehood, participants would only hear the ambiguous question, asked to them by the puppet appearing on screen at the end of the story. A *yes* answer corresponded to the inverse scope interpretation of the ambiguous question, while a *no* answer corresponded to the surface scope one.

(18) The boys and girls want to have a snack. They go looking for something to eat. In the kitchen there are three pieces of fruit: an apple, a banana and an orange. Both the boys and the girls want to eat something. The boys get to the kitchen first. One boy eats the apple, another boy eats the banana. The third boy eats the orange. There is no fruit left for the girls so the girls don't eat any fruit. Puppet: Did a boy eat each fruit?

The procedure and instructions participants received were the same as for Experiment 1.

3.3.3 Results and discussion

No significant difference was found between adults and children, as they gave *yes* answers in 78% and 83% of cases respectively, (Mann-Whitney U = 360, p = 0.392). proving that they were obtaining the inverse scope interpretation of the ambiguous test question.

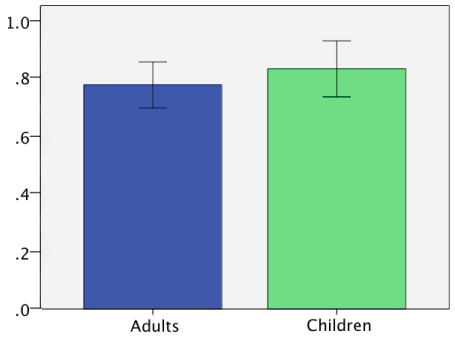


FIGURE 3.7: Experiment 2: Yes responses for test questions.

Control items were answered correctly in over 90% of instances by participants in both groups, similarly to what happened in Experiment 1.

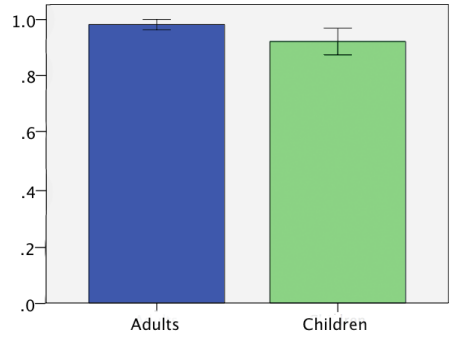


FIGURE 3.8: Experiment 2: Accuracy rate for control and filler items.

The prediction made for this experiment, based on the QUD approach as formulated in Gualmini et al. (2008) was that children and adults should show no difference in accessing inverse scope for scopally ambiguous questions, based on the assumption that using a question instead of a declarative sentence effectively removes the QUD and its influence. Another necessary assumption, made by the QAR, is that children and adults only differ in how they receive and accommodate the QUD, and therefore they should be able to access inverse scope at the same rate in a situation in which the QUD has been removed. Indeed, in line with several previous experimental studies (Gualmini, 2004a, Krämer, 2000, Gualmini et al., 2008), the results obtained for Experiment 1 and Experiment 2 show that children and adults are able to access inverse scope for ambiguous questions at the same rate: this evidence is in contrast with the Observation of Isomorphism (Musolino, 1998).

It needs to be observed however that the rates of inverse scope obtained for adults and children in this study were higher than what generally reported in several of the studies reviewed in Chapter 2: among the factors that have contributed to the high

rate of inverse scope responses observed in adults and children, the high rate of inverse scope is probably due to several factors, including the choice of quantifiers: early studies concerned with the lexical nature of quantifiers have identified *each* as more prone to take wide scope compared to *every*, which has been used more in previous experiments. Moreover, the context was studied to be supportive of the inverse scope interpretation: this was achieved by making the set of objects more salient compared to the set of characters: recall that the experiment presented three sets, two sets of characters (in the case of the example in (12), three boys and three girls) and one set of objects. Both sets of characters aim to perform an action with, or in relation to, the set of objects, which have previously been shown on screen in a preceding slide. As discussed by Unsworth, Gualmini, and Helder (2008), it is possible that having the objects on which the set of characters is expected to perform an action belong to a pre-defined set had a facilitating effect for the inverse scope interpretation. This contextual evidence provided should insure that a QUD suitable to be answered by the inverse scope interpretation of the sentence is selected. The role of parsing for scope ambiguity resolution in adults and children, addressed for example by Musolino and Lidz (2006) and Viau et al. (2010), was not explicitly addressed in the study here related.

As it is evident from previous literature, the effect of the QUD is well documented, nevertheless it is not universally agreed that the notion of QUD alone explain children's behaviour and the difference compared to adults which has been sometimes observed in previous studies. In particular, assuming that every sentence that an individual hears needs to be evaluated against the QUD, the next issue to be addressed is which criteria are relevant for this evaluation. The QAR model, as initially proposed (see Gualmini et al., 2008, Hulsey, Hacquard, Fox, and Gualmini, 2004), posits that the parser filters the relevant interpretations, making available those that answer the QUD, that is to say those that entail either a *yes* or a *no* to said question. This entailment requirement, however, is not the only mechanism at play, as shown by the fact that adults can sometimes prefer the true interpretation of the sentence even when it does not abide said requirement, possibly because they are able to discard the previous QUD to accommodate a different one. Some of the literature reviewed in Chapter 2 has evidenced that obtaining an inverse scope interpretation is usually associated to a processing cost: several studies conducted with a self-paced reading methodology have shown that inverse scope interpretations are associated to longer reading times (Filik et al., 2004, Anderson,

2004, Tunstall, 1998). While it was initially supposed that the observed processing cost might be due to the disambiguating context forcing participants to revise their initial surface scope interpretation to obtain the inverse scope one, Anderson (2004) later observed that such interpretation incurs in a cost even when no context is provided and therefore no revision has been triggered. This evidence points toward the possibility that IS is in itself more costly to process. However the picture seems more nuanced for what concerns this children: due to the limitations of the methodology, there are currently no self-paced reading studies that have been conducted on children, but there is evidence that children might be subject to the same processing constraint (Viau et al., 2010), and possibly be more sensible to the effect of the context, as per the QUD approach. The two scope experiments presented in this chapter have however shown that children can easily access the inverse scope interpretation of ambiguous declarative sentences and questions, therefore, given the previous results, we do not expect to find an increased cost for adults: it is indeed possible that the cares taken to make the contest as supportive as possible for such interpretation have mitigated this cost. Further experimental testing would be needed to ascertain this.

Going back to the implications of the results presented in this chapter for the QAR model, there are two options at this point compatible with the results obtained in this last experiment: one can assume the QAR is not relevant for questions, or an extended version of it must play a role in this domain as well. I will discuss the second possibility in the rest of this section. In Roberts's (1996) model, discussed in section 1.2.2: discourse is organised around a series of questions, with the questions in turn being organised by entailment relations, defined as follows: "one interrogative Q1 entails another Q2 iff every proposition that answers Q1 answers Q2 as well" (Groenendijk and Stokhof, 1984, p. 16). For example, every one of the question in (19) is entailed by anything above it. The broader question in (19-a) entails everything that appears below it: providing an answer to (19-a) would also answer the narrower questions (19-a-i) and (19-a-ii).

(19) What's the way things are?

a. Who ate what?

(i) What did Hilary eat?

Did Hilary eat tofu?

Did Hilary eat beans?

(ii) What did John eat?
Did John eat tofu?
Did John eat beans?

Therefore, it can be supposed that the immediate QUD generated by the context is related to other broader questions, by which it is entailed: assuming the same mechanism that holds for declarative sentences is at play, it is possible that the only interpretation available for ambiguous questions is the one that is entailed by the super question made relevant by the context.

In the case of the short stories used in this study, it is possible that the context provided had made salient a superquestion that only entails the inverse scope interpretation of the target sentence. See for example the context story below.

(20) The boys and girls want to have a snack. They go looking for something to eat. In the kitchen there are three pieces of fruit: an apple, a banana and an orange. Both the boys and the girls want to eat something. The boys get to the kitchen first. One boy eats the apple, another boy eats the banana. The third boy eats the orange. There is no fruit left for the girls so the girls don't eat any fruit.
Experimenter: Did a boy eat each fruit?

For (20), for example, we must assume that, since we were interested in knowing whether all fruits were eaten by some boy or other, the context generated question (21-b), which is equivalent to, and therefore entails, the inverse scope interpretation of the test sentence, (22-b).

(21) a. Is it true that there is a single boy who ate each fruit?
b. Is it true that, for each fruit, there is a boy who ate it?

(22) Did a boy eat each fruit?
a. Is it true that there is a boy that ate each fruit?
b. Is it true that for each fruit, there is a boy who ate it?

In Roberts model, there is no clear indication on how the superquestion of a specific QUD is to be identified: indeed, one of the points of criticism moved to some previous studies was that participants were presented with a scopally ambiguous declarative sentence, and the QUD that said sentence was (supposedly) understood to answer was only guessed. Later studies however have resolved to make the question explicit (Moscati et al., 2016). For what concerns the present study, (22) seems a plausible option, part of a broader strategy: recall that the goal of the stories used in the experiments reported in this chapter was to provide a supportive context for the inverse scope interpretation of the ambiguous question and sentence. The experiment presents 3 sets, 2 sets of characters the set of boys, the set of girls, and the set of fruit, and emphasis is placed on the fact that both sets of characters intend to perform an action with, or in relation to the set of objects. Following the idea, discussed in literature by (Unsworth et al., 2008) that having objects belonging to a predefined set, and requiring the protagonist to perform an action on all the objects, might generate a QUD that calls for an answer that specifies whether all objects have been involved. In the case of the context story provided for (22), it might have generated a superquestion (or multiple superquestions) that call for information on what happened to all the members of the set of objects and the set of characters. In keeping with Roberts proposal, (21-a) would represent the current QUD by the end of the story.

If this is the case, the Question and Answer Requirement would need to be extended to include questions, as in (23) below.

(23) **Extended QAR:** The interpretations available for a scopally ambiguous question are the one(s) entailed by its superquestion.

Empirical evidence for this proposal could be obtained by investigating how the same ambiguous questions are interpreted in different contexts: in particular, one could test the interpretation of the same ambiguous question both in a context which makes relevant a surface scope biasing super-question, and in one that makes relevant an inverse scope biasing super-question. A possible example of this is provided in (24).

(24) a. **Surface scope QUD context:** The leader of the afterschool club hopes to find a boy who practices all sports: if there is a single boy who enters the

football tournament, the basketball tournament and the rugby tournament, he will give him a medal.

b. **Inverse scope QUD context:** The leader of the afterschool club hopes to run three tournaments, a football tournament, a basketball tournament and a rugby tournament. He needs one more person to complete the teams, so he hopes that someone will join each one of the tournaments.

(25) Did a boy enter each tournament?

a. **Surface scope reading:** Is it true that there is a single boy who entered each tournament?

b. **Inverse scope reading:** Is it true that, for each tournament, a (possibly different) boy entered it?

(25-a) explicitly mentions that the goal of the character in the story is to find one single boy, who will be awarded a medal for his sport merits. (25-b), on the other hand, should bias participants toward the inverse scope "some boy or other" reading. If the QAR can indeed be extended to include questions, one could expect (25) to be interpreted on its surface reading more often when associated with the corresponding surface scope QUD context, and vice versa. I leave investigating the potential role of an extended QAR for further research.

3.4 Summary of chapter

This section presented two experiments aimed at investigating the interpretation of scopally ambiguous questions containing two quantifiers. Experiment 1 tested children and adults' interpretation of ambiguous declarative sentences preceded by an ambiguous QUD containing the same quantifiers, proving that children and adults can access inverse scope at the same rate for these kind of sentences. Since the interpretation of the declarative sentence was dependent on the one obtained by the question, it can be inferred that the ambiguous questions have also been disambiguated towards their inverse scope. Experiment 2 tested ambiguous questions directly, and again adults and children were able to access the inverse scope interpretation in the vast majority of cases. The results provide another confirmation of the fact that children are able to obtain inverse

scope interpretations, contrary to initial proposals that they might be unable to obtain these interpretations altogether. While the results for declarative sentences are in line with the QAR approach, the status of ambiguous question goes beyond the current formulation of the QAR. Two possibilities that compatible with the current state of affair have been discussed: the first relies on the assumption that the QUD does not play a role in the interpretation of questions, therefore its influence was removed. The other possibility is to amend the QAR to account for the role of the QUD in this domain: I have sketched an extension of the QAR that would extend to ambiguous questions, based on the assumption that the reading selected for an ambiguous question is the one entailed by its superquestion. Further experimental research will be needed to test this prediction.

Chapter 4

Conditional Perfection literature

The aim of this chapter is to present a literature background on conditionals, and in particular on Conditional Perfection (CP), the phenomenon through which conditional sentences receive a biconditional interpretation in some circumstances. For example, the sentence in (1-a) (from Geis and Zwicky 1971), tends to give rise to the inference in (1-b).

(1) a. If you mow the lawn, I'll give you five dollars.
    b. → If you don't mown the lawn, I won't give you five dollars.

This chapter focuses in particular on the nature of CP and the circumstances in which it occurs, in order to establish the relevance of this phenomenon for language acquisition. Said relevance is twofold: firstly, CP is widely regarded as a pragmatic inference, therefore the biconditional interpretation is generally not taken to be a part of the semantic meaning of the conditional. While its pragmatic nature is generally agreed upon by the scholars who have investigated its occurrence, its exact nature is debated. One prominent account sees CP as a Scalar Implicature (SI)[1] (Horn, 2000, van der Auwera, 1997, López Astorga, 2014a,b), although there is not agreement as to what competitor could give rise to such implicature.

Another approach sees CP as a more general quantity implicature arising as a result of the effect of the Question Under Discussion, and more precisely, as an effect of the

[1]Not considering, however, possible accounts of Scalar Implicatures as a grammatical phenomenon, which would make in turn conditional perfection also semantic in nature (see Chierchia, Fox, and Spector, 2012).

fact that answers to questions tend to be intepreteted exhaustively: as observed by Groenendijk and Stokhof (1984), the sentence in (2-b) tends to be interpreted as an exhaustive list of people who came to the party, when received as an answer to (2-a).

(2) a. Who came to the party?
b. John, Mary and Bill came to the party.

In the same way, if a question like (3-a) is made salient by the context, one would draw the inference that mowing the lawn is the only way for obtaining five dollars, therefore obtaining the perfected interpretation (Von Fintel, 2001).

(3) a. How can I earn five dollars?
b. If you mow the lawn, I'll give you five dollars.

Children and adults have often been shown to differ in their pragmatic competence, as it has already been discussed in Chapter 2 and Chapter 3 of this dissertation: indeed, the QUD approach for scope ambiguity resolution is based on the notion that children sometimes fail to access readings of a scopally ambiguous sentence that are available to adults entirely because of their not yet fully developed pragmatic abilities. Differences between children and adults with respect to their pragmatic competence have also been observed in connection to other phenomena, amongst which one of the most widely investigated is the occurrence of SIs: it has long been known that children and adults do not derive SIs at the same rate (Noveck, 2001, Papafragou and Musolino, 2003, Guasti et al., 2005, Foppolo et al., 2012, a.o.). While the status of CP as a SI is debated, looking at CP from an acquisition perspective could help determine whether this approach is valid, by comparing children's and adults' interpretation of conditionals and sentences that give rise to standard SIs. Moreover, experimental studies that have been conducted on children and adults on conditional sentences have evidenced that sometimes the two groups differ in their interpretation, with children obtaining a conjunctive-like interpretation of such sentences (Paris, 1973, Barrouillet and Lecas, 1998, Barrouillet et al., 2000, Barrouillet and Lecas, 1999), in a way similar to what is observed for sentences with disjunction (Paris, 1973, Braine and Rumain, 1981, Gualmini et al., 2001, Chierchia et al., 2004), and there are various hypotheses as to how this inference arises.

The rest of this chapter is organised as follows: section 4.1 briefly discusses some of the most prominent accounts of the meaning of conditional sentences. Section 4.2 introduces the notion of Conditional Perfection, and how it can be derived, focusing on particular on its nature as a pragmatic phenomenon. Of particular relevance for this dissertation are the account that see CP as a scalar implicature, and the one that that treats it as occurring because of the QUD. Section 4.3 discusses the two main theories of conditional reasoning: the Mental logic program and the Mental model program. The following section, 4.4 presents some of the relevant experimental studies that have been conducted on adults. Section 4.5 discusses some of the experimental studies conducted on children on their interpretation of conditionals, drawing a parallel with the ones conducted on disjunction.

4.1 Conditional meaning

The topic of conditional meaning has received much attention from philosophers, psychologists and logicians. While an exhaustive account of the topic is beyond the scope of this dissertation, this section briefly presents some of the most influential proposals. Over the years, different accounts have been put forward, moving from material implication account, to accounts that invoke possible world semantics (Stalnaker, 1968, Lewis, 1973). In addition to this, arguments have been put forward that reject the nature of *if...then* as a connective altogether (Kratzer, 1986).

4.1.1 The classic account

The classic account of indicative conditionals is the *material conditional* one, based on the idea that *if...then* is a two valued truth-functional connective.

(4) $p \rightarrow q$.

Therefore, conditionals have the truth conditions of the material conditional truth function of classic propositional logic, represented in the truth table below.

p	q	if p then q
T	T	T
T	F	F
F	T	T
F	F	T

TABLE 4.1: Truth conditions of the material conditional *if p then q*.

As it can be observed from the table, a material conditional is false only in the situation in which its antecedent is true, and the consequent is false. This seems to hold for certain cases: for example, if one utters the statement in (5), the promise seems only broken (and therefore, the statement is false) if one is healthy and does not come to class (Suber, 1997).

(5) If I am healthy, I will come to class.

However, the premises of this account result in the fact that false antecedents make the conditional sentence true, no matter what the consequent; and the same happens with true consequents, giving rise to the two paradoxes (6-a) and (6-b) (Evans and Over, 2004).

(6) a. Given not-p, it follows that if p then q.
b. Given q, it follows that if p then q.

Following the truth conditions in table 4.1, the conditional sentence can be asserted as true once one has the belief that the antecedent is false. Applying (6-a) to (7) allows us to see more clearly why the paradox arises.

(7) If it rains then the plants will die.
a. It will not rain.
b. It will not rain, therefore if it rains the plants will die.

Here Evans and Over (2004) propose this reasoning: supposing the plants are dying from the absence of rain, and suppose the speaker to hold the belief that the drought

will continue, as in (7-a). This validates the statement in (7-b) according to the truth function table, even if there are no good reasons for asserting (7-b) from these premises. Following the same drought scenario, the paradox is easily identified: if one holds the belief that the plants are dying (from a lack of water) as asserted in (8-a), it makes no sense to infer (8-b) from that, even if the premises of the inferences are true.

(8) a. The plants will die.

b. The plants will die, therefore if it rains the plants will die.

For these and other reasons, it is now generally assumed that the material conditional account does not hold for natural language indicative conditionals, and that these conditionals should not be treated as truth funcional. Nevertheless, there have been theories of conditional meaning based on this approach (Grice 1989, Jackson 1987 a.o.).

An amended version of this truth table has been proposed by Wason (1966), who dubbed it *defective* truth table: in this version, a false antecedent makes the conditional neither true or false, but irrelevant (I), faring better with the paradoxes in (6).

p	q	if p then q
T	T	T
T	F	F
F	T	I
F	F	I

TABLE 4.2: Truth conditions of a defective conditional *if p then q*.

In spite of the paradoxes highlighted so far, the material conditional account has been defended by various scholars that have attempted to find an explanation for the paradoxes mentioned in this section (Jackson, 1987, Grice, 1967, Jackson, 1979, Rieger, 2006).

4.1.2 Strict conditionals and possible world semantics

C.I. Lewis (1918) proposed an account of conditionals that pairs the truth conditional approach with the notion of necessity: a conditional is true when the antecedent is necessarily true; that is to say, in a conditional in the form *if p then q*, the role of

the antecedent is restricting the worlds under analysis to the ones where p is true. A strict conditional therefore has the form in (9), where $\Box$ is a necessity modal universally quantifying over some modal base (von Fintel, 2011, Kratzer, 1981, 1979).

(9) $\Box$ (p $\rightarrow$ q)

This proposal has the merit of solving the paradoxes highlighted above for the material conditional: under this account, a false antecedent does not necessarily make the whole sentence true. The *variably strict conditional* account, developed independently by Stalnaker (1968) and Lewis (1973) departs from C.I. Lewis's account in what the relevant worlds taken into consideration are the ones that differ minimally from the world under consideration and in which the antecedent of the conditional is true. Stalnaker (1968) semantics uses a selection function, called *f*. For any proposition p and each world w, *f* selects a world w', which is the most similar world to w in which p is true. *If p, then q* then is true if and only if q is true in the nearest p world selected by *f*.

(10) f(w,p).

Under this approach, the truth conditions for the first two situations in 4.1 remain the same: for the "T - T" situation, its closest possibility in which p is true is nothing but "T - T" itself. The same is true for "T - F": in this case p is true, and q is false.

Referring to the truth table, the "F - T" and "F - F" situations are the ones which the strict conditional differs from the material one. Here *if p then q* is still true in some circumstances in which p is false, but it is false in others: its truth or falsehood will be determined by whether "T - T" is the closest possibility to "F - T" in which p is true or not. Similarly for "F - F". The conditionals for which antecedent is false will therefore be true in some circumstances and false in others.

4.1.3 *If* as an adverbial restrictor

Breaking from classical accounts of indicative conditionals, Kratzer (1986) claims that the analysis of *if...then* as a two place connective is not a reality of the logical form of natural languages. The idea comes from an observation of Lewis (1973), who had

proposed that, in some cases, the role of *if* is to restrict adverbs. He provides the following examples.

(11) Often if it is raining, my roof leaks.

Lewis argues that in (11), the if-clause has no meaning of its own. Instead, its only role is to introduce a restriction on the adverb *often*.

Kratzer (1986) generalised this account to extend to all if-clauses. Breaking with previous tradition, she rejected the notion of *if...then* as a two place connective, claiming that this kind of connective is not part of the logical form of natural languages. Instead, the if-clause combines with a modal or quantificational operator. For example, in a sentence like (12), the role of *if* is to restrict the universe of discourse necessary for the interoperation of the modal *will*. Therefore, (12) has the logical form in (13).

(12) If my hen has laid eggs today, the Cologne Cathedral will collapse tomorrow.

(13) [must: my hen has laid eggs today] the Cologne Cathedral will collapse tomorrow.

Furthermore, Kratzer argues that even in bare conditionals, conditionals without any explicit modal operators, there is silent modal necessity operator that is being restricted by the antecedent. However, while this view has been particularly influential in linguistics, it is also controversial. For instance, Edgington (2014) observed that, if the unpronounced modal operator is an epistemic *must*, bare conditionals resemble strict conditionals in what they are asserting that all p possibilities are q possibilities. Moreover, Von Fintel and Iatridou (2002) observed that the assumption that *if* has no meaning does not seem intuitively correct when one compares (14-a) and (14-b).

(14) a. If Caesar woke up, he had tea.
b. When Caesar woke up, he had tea.

(14-a) and (14-b) are not equivalent: (14-a) seem to suggest uncertainty as whether Cesar woke up. Replacing the covert operator *must* with *probably* does not seem to

represent a solution, either, as this would imply that *if p, q* is true once the probability of *q* once *p* is given has reached a certain degree of probability (Edgington, 2005), but believing the statement "it is probable that *if p then q*" to be true is not equivalent to believing that *if p then q*.

4.1.4 Summary

This section has provided a brief summary of some of the most relevant accounts of the meaning of conditionals that have been developed over the years. The initial proposal of equating the conditional construction to the material conditional of logic has been criticised on grounds that it gives rise to various paradoxes, and other theories have been developed. Nevertheless, the material conditional has been the starting point of many theories of conditional meaning, and it is still defended as a plausible theory by a number of influential scholars (Jackson, 1987, Grice, 1967, Jackson, 1979, Rieger, 2006). The following section presents the phenomenon of Conditional Perfection, discussing two accounts on its nature.

4.2 Conditional perfection

Geis and Zwicky (1971) were the first in the literature to draw attention to the fact that conditionals often get "perfected" into biconditionals: a sentence in the form *if p then q* invites an inference in the form "if not p, not q", a phenomenon they named Conditional Perfection[2]. This can be observed in the classic example (15).

(15) a. If you mow the lawn, I'll give you five dollars.
b. → If you don't mown the lawn, I won't give you five dollars.

Geis and Zwicky (1971) take CP to be an invited inference, hence a pragmatic phenomenon, discarding the view that the biconditional interpretation might be part of the semantic meaning of the conditional: they base this claim on the observation that perfection is restricted. In particular, CP applies when conditionals are used to express

[2]As van der Auwera (1997) and Horn (2000) point out, the phenomenon had already been observed by Ducrot (1969), who did not give a name to it.

predictions, promises, threats, lawlike statements, commands, and counterfactuals: they provide the following list to exemplify each type of statement.

(16) a. Prediction: If John leans out that window any further, he'll fall.
b. Threat: If you disturb me tonight, I won't let you go to the movies tomorrow.
c. Lawlike statement: If you heat iron in a fire, it turns red.
d. Command: If you see a white panther, shout "Wasserstoff" three times.
e. Counterfactual: If Andrew were here, Barbara would be happy.

Much subsequent literature has indeed focused on the cases in which CP does not arise, which are often called "biscuit conditionals" after the example in (17), used by Austin (1961), and which will be briefly discussed below.

(17) There are biscuits on the sideboard if you want them.

In their squib, Geis and Zwicky (1971) did not go into details as to how CP is derived, other than arguing that it does not seem to be derived by Grice's (1975) cooperative principles and observing that there exists "a tendency of the human mind to perfect conditionals to biconditionals".

Other scholars have argued that Geis and Zwicky's view of this phenomenon is incorrect: among the objections that have been raised, Lilje (1972) argues that the claim that the human mind has a tendency to perfect conditionals is too strong, and not coherent with another claim made in the same squib, weaker than the first - that there is a quasi-regular association between the biconditional interpretation and a conditional. In Lilje's view, to understand a conditional as a biconditional is simply to misunderstand it. Strictly speaking, this observation is true if a conditional is taken to have the meaning in table 4.1, nevertheless, this should not be taken as proof that CP is merely a fallacy of reasoning and not a linguistic phenomenon. Boër and Lycan (1973) also argued in favour of the notion that CP is non-linguistic, arising instead as a product of contextual information and real world knowledge.

In spite of these objections, most scholars have however accepted the fundamentals of Geis and Zwicky's account, namely that CP is a form of inference, and as such, a

linguistic phenomenon. While it is debated what exactly the nature of the inference is, the fact that it is arising as a product of pragmatics is well recognised.

One of the most convincing arguments for the pragmatic nature of CP is that the inference is cancellable[3]: as van Canegem-Ardijns and Van Belle (2008) observe, they differ in this respect from entailments and logical inferences, which are not cancellable by additional information.

(18) If you mow the lawn, I'll give you five pounds.
...But also if you paint the garage.

Moreover, as Geis and Zwicky had already observed, CP does not arise from all kind of conditionals: while they had provided a list of specific types of sentences that invite this inference, van Canegem-Ardijns and Van Belle (2008) have tried to provide a more general account of what conditionals invite perfections and which ones do not. In their analysis, the conditionals that invite CP are the ones which have characteristics of desirability and control. One example of such sentences are conditional promises, like the one in (19-a), which readily invites the inference in (19-b).

(19) a. If you get me some coffee, I'll give you a cookie.
b. If you don't get me some coffee, I won't give you a cookie.

The authors argue that conditional promises invite CP as a product of the following characteristics: "A[ntecedent] is desirable for S[peaker]; H[earer] has control over A; S wants H to do A; S considers C[onsequent] desirable for H because S assumes H wants C; S has control over C" (van Canegem-Ardijns and Van Belle, 2008, p. 353). Another class

[3]In their analysis, van Canegem-Ardijns and Van Belle (2008) distinguish between two different inferences: the *if not p, not q* one and the *only if p, q* one: they claim that (i) gives rise to one or both of (i-a) and (i-b).

(i) If you mow the lawn, I'll give you five dollars.
a. If you don't mow the lawn, I won't give you 5 dollars.
b. Only if you mow the lawn, will I give you 5 dollars.

More specifically, they claim that the *only if p, q* inference is more easily cancelled compared to the other. The *if not p, not q* one on the other hand, is harder to cancel if relevant, which depends on the content of the proposition. In the case of (15-a), they consider the addition of (ii) as peculiar on the ground that it clashes with the the (presumed) illocutionary goal of the utterance - getting the hearer to mow the lawn.

of perfectible conditionals, related to this, are what they call "preconditional conditionals", due to the fact that the antecedent serves as a precondition for the performance of the speech act about the consequent.

(20) a. If the weather is fine tomorrow, I'll go for a swim.
b. If you mow the lawn, you do not need to pay me back.
c. If you pay your contribution, you may participate in the barbecue.
d. If I get a company car, I'll accept the job.
e. If you sell your car, I'll sell mine.

The examples in (20) share those feature of desirability and control with the conditional promise in (19-a): in the latter, speaker has control over the action described, and they think it desirable for either themselves or the speaker. It should be noted however that experimental studies have evidenced that not only "desirable" conditionals get perfected: in the case of the example in (21), from Cummins, Lubart, Alksnis, and Rist (1991), it can hardly be argued that the consequent is desirable for hearer or speaker, yet Cummins et al.'s study showed that (21) tends to get perfected.

(21) If my finger is cut, then it bleeds.

Another common feature across the two conditions is that both types of conditionals seem to answer an implicit question *on what condition q?*. This intuition has given rise to one prominent account of CP, which will be discussed in more detail in section 4.2.3.

(22) Promise conditionals:
a. On what condition will you give me a cookie?
b. If you get me some coffee (I will give you a cookie).

(23) Preconditional conditionals:
a. On what condition will you go for a swim?
b. If the weather is fine tomorrow (I'll go for a swim).

Other conditionals, on the other hand, are not perfectible. One well known example are the so called "biscuit conditionals".

(24) If you are hungry, there are biscuits in the cupboard.

Although there have been attempts to explain the non-perfectibility of biscuit conditionals assuming these have a different semantics compared to standard conditionals, for example by Siegel (2006), it is generally assumed that their non-perfectibility arises from pragmatics. Dancygier and Sweetser (2005) argue that such conditionals escape perfection because they are not part of the same speech act class as the perfectible conditionals discussed above: conditionals that express promises, threats etc. are offering an alternative scenario: their function is that of building a conditional mental space, which in turns involves "a simultaneous set-up of an alternative mental space in which P does not hold, and therefore not-Q is expected to hold as a result" (p. 41).

On the other hand, in conditionals like (24), the speaker is not setting up alternatives but simply trying to provide a context for the consequent, similarly to what happens in the example in (25).

(25) If you need help, my name is Ann. (Dancygier and Sweetser, 2005)

Here, they claim the function of the antecedent is simply clarifying the nature of the consequent or, in Horn's (2000) words, "offer grounds for the relevance of the statement that q" (p. 347). In the case of (25), the antecedent is signalling that the consequent should not be taken as a personal introduction but as one made for the purposes of customer service interaction. The context is provided to make sure that the sentence *my name is Ann* is taken in the right way by the hearer, not to imply that the sentence is subject to the conditions expressed in the antecedent. Similarly, Franke (2007) argues that the role of the antecedent in biscuit conditionals is to add its contents to the Common Ground, as they might not be felicitous, or even intelligible, otherwise.

In summary, although some objections were initially raised to Geis and Zwicky's account of Conditional Perfection, there is general agreement that CP should be treated as a pragmatic inference arising from certain types of conditional statements. However, there is no general consensus as to its exact nature, as it will be discussed in the next section.

4.2.1 The nature of Conditional Perfection

In the previous section it has been discussed how CP is not considered to be a part of the semantics of conditionals, but rather an inference arising from conversational clues and real world knowledge. This goes to explain why it only arises in a limited set of cases, while other "classes" of conditionals are not generally perfected: in particular, van Canegem-Ardijns and Van Belle (2008) have identified desirability and control as characteristics of perfectible conditionals. The desirability requirement however seems debatable, as non-desirable conditionals get perfected as well: this can be observed in many of the sentences used in in the experimental studies described in section 4.4, including the one in (21), for which Cummins et al. (1991) found high rates of biconditional interpretations. The exact nature of CP, however, is debated: while Geis and Zwicky (1971) did not go into details as to what they believed the nature of CP to be, they observed that it does not seem to be derived from Grice's cooperative principles. In spite of this, the two major accounts of the phenomenon sees it as a quantity implicature: either a SI (Horn, 2000, Atlas and Levinson, 1981, van der Auwera, 1997) or a more general quantity implicature arising from the influence of the QUD (Von Fintel, 2001). This section expands on these two proposals.

4.2.2 CP as a Scalar Implicature

While there is general consensus that the perfected reading of the conditional is a pragmatic phenomenon, the details of how this implicature arises are debated. Grice (1975) proposed that conversational implicatures come in two kinds: Particularised Conversational Implicatures (PCI) and Generalised Conversational Implicatures (GCI). PCIs arise as an effect of the circumstances: the implicature does not simply arise from uttering p, but it is a product of contextual considerations. GCIs, on the other hand do not need any specific contextual circumstances to arise. Levinson (2000) argues that Grice's discussion lacks details, and proposes the definitions in (26) to capture Grice's intentions.

(26) a. An implicature i from utterance U is *particularized* iff U implicates i only by virtue of specific contextual assumptions that would not invariably or even normally obtain.

b. An implicatures i is *generalized* iff U implicates i unless there are unusual specific contextual assumptions that defeat it. (p.16)

The dialogue in (27) to provide an example that evidences the difference between PGI and GCI.

(27) A: What time is it?
B: Some of the guests are already leaving.

In (27), B's response gives rise to the implicature that it is late - this implicature is a particularised one, arising from the contextual circumstances. However, the utterance "some guests are already leaving" also implicates that *not all guests are leaving* - in this case, the implicature is arising from the choice of *some* as opposite to *all*, and as such it is a generalised implicature, that occurs independently of the context. Levinson argues that Conditional Perfection is a class of generalised implicature - it arises as a product of the conditional form, and not from features of the context.

Therefore, CP arises as a GCI from the need to avoid communicating unnecessary information explicitly (which Levinson calls I-heuristic): in turn, this strategy is needed to communicate more complex information without being hindered by the slow speed rate of human communication, relying on the hearer to enrich the content of a sentence by making some assumptions about the communicative intent of the speaker. CP, then, is an I-implicature in the sense that it allows the speaker to maximise informational load of a conditional utterance on the assumption that the hearer will move to the more specific biconditional unless circumstances dictate otherwise. Because of this, the speaker does not need to make biconditionality explicit.

In this sense, in Levinson's (2000) view CP is a conversational implicature arising from *informativeness*. In particular, he refers to his Principle of Informativeness (or "I-Principle"), which in turn is derived from Grice's second maxim of quantity.

(28) Grice's Quantity Maxim 2: Do not make your contribution more informative than it's required.

On the basis of (28), the hearer might in some cases amplify the informational content of the speaker's utterance by enriching it, as explained by the I-Principle in (29).

(29) Levinson's I-Principle:

a. Speaker's maxim: the maxim of minimization. "Say as little as necessary"; - produce the minimal linguistic information sufficient to achieve your communicational end (bearing Q in mind).

b. Recipient's corollary: Amplify the informational content of the speaker's utterance by finding the most *specific* interpretation, up to what you judge to be the speaker's m-intended point. (Levinson, 2000, p. 114)

A classic example of the I-Principle in action is the phenomenon known as *conjunction buttressing* (Atlas and Levinson, 1981, Levinson, 2000), as in (30) (in which the symbol "I++>" is used by Levinson to mean "communicates" - what is communicated is taken to be the sum of what is said and what is implied).

(30) a. Hans wrote a novel and he sold the rights to Macmillan.

b. I++> Hans wrote a novel and then sold the rights to Macmillan.

Levinson (2000) claims that the enriched sentenced (30-b) is implicated via the I-Principle: the conjunct events (writing a novel, selling the rights) are interpreted as occurring in temporal succession. A similar process is at play with conditionals: (31-a) gives rise to the implicature in (31-b). The combination of what explicitly said plus the implicature leads to (31-c), which is what is communicated.

(31) a. If you mow the lawn, I'll give you five dollars.

b. I+> If you don't mow the lawn, I will not give you five dollars.

c. I++> (If and) only if you mow the lawn, will I give you five dollars.

Assuming both the conditional and biconditional interpretation of (31) are available, the biconditional one is more specific, therefore by the I-principle, the hearer will assume that this is the interpretation the speaker had intended.

While this model provides an explanation as to how the biconditional interpretation is selected, it does not go into detail as to how said interpretation becomes available in the first place: since the biconditional interpretation is not thought to be part of the semantics of a conditional, there is no explanation as to how it becomes available as an "enriched meaning" version of the speaker's utterance for the hearer.

A solution to this problem is offered by the Scalar Implicature (SI) account of CP. SIs are a particular form of GCIs arising from Grice's first maxim of Quantity and involving two or more utterances, which are ranked according to their informativeness, based on the idea that some terms can be ordered in a scale from the strongest (most informative) to the weakest (least informative). Amongst the terms that are generally considered to form a scale and give rise to this kind of implicature are quantifiers: consider the example in (32).

(32) a. Some of my friends speak French.
b. All of my friends speak French.

Here, because of the ranking between the two quantifiers ($\{all > some\}$), (32-a) is weaker and conveys less information when compared to the stronger (32-b), by which is entailed, giving rise to the intuition that the speaker would not utter (32-a) in a situation in which they could have uttered (32-b), giving rise to the inference in (33).

(33) Not all of my friends speak French.

This intuition had already been grasped by early scholars: Mill (1889) had already observed that "if I say to any one, I saw some of your children today, he might be justified in inferring that I did not see them all, not because the words mean it, but because, if I had seen them all, it is most likely that I should have said so" (p. 501). In later years, Grice (1975) put forward a formal approach to this reasoning, based on the notion that conversation is a cooperative effort that interlocutors undertake in order to share information one with another. In order to achieve this goal, interlocutors abide a set of Cooperative Principles, committing to offering statements that are true, relevant and

informative, and to the assumption that the other interlocutors are also being cooperative. Scalar implicatures arise from this need to informativeness and completeness, captured in the first Maxim of Quantity.

(34) Quantity Maxim 1: Make your contribution as informative as it is required.

Assuming that participants are choosing to be collaborative and respect (34), the choice of the weaker term *some* should signal to the hearer that the speaker does not believe the stronger statement (32-b) to be true, therefore leading to the inference in (33).

A more detailed, step by step formalisation of this reasoning is offered by Gamut (1991):

(35) A speaker S makes correct use of a sentence A in order to make a statement before a listener L just in case:
(i) S believes that A is true;
(ii) S believes that A is relevant to the subject of the conversation;
For all sentences B of which A is a logical consequence (and which are not equivalent to A), (i) and (ii) do not both hold with respect to B.

Therefore, to make a correct use of (32-a), the speaker has to believe that the statement *some of my friends speak French* is true and relevant. Moreover, since this sentence logically follows from the stronger (32-b), the speaker must either believe (32-b) not true, or believe it not relevant. Therefore, the hearer comes to the conclusion that some but not all of my friends speak French.

Quantifiers are not the only elements that can be ranked according to their informativeness: verbs (e.g. *might* - *must*), adverbs (*possibly* - *necessarily*) and adjectives (*warm* - *hot*) are among the terms that can be ordered in a scale and therefore give rise to a SI (Horn, 1972). Scales are restricted, and there is general consensus amongst researchers that, to group the various elements together as member of a scale, they must present some requisite, although it is debated exactly what this requisite would be: for example, Atlas and Levinson (1981) argue that the elements must be lexicalised to the same degree, and belong to the same semantic field, while Matsumoto (1995) claims the selection is restricted to elements that have the same monotonicity. This requirements stem from

the necessity of avoiding the symmetry problem (see Breheny, Klinedinst, Romoli, and Sudo, 2017). Consider the example in (36).

(36) a. Some of my friends speak French.
b. Some but not all of my friends speak French.

(36-b) is stronger than (36-a), but, if it were considered an alternative, the implicature that *all of my friends speak French* would arise from conjoining (36-a) with its negation. This implicature is clearly incompatible with (36-b). Therefore, the set of possible alternatives needs to be restricted in order not to incur in this paradox. More recently, Katzir (2007) and Fox and Katzir (2011) proposed a general theory of alternatives not based on the notion of scales altogether, imposing instead a structural condition on alternatives: all structures that are at most as complex as the one under consideration can serve as alternatives, provided they are also stronger. This removes the necessity for the assumption that alternatives have to be restricted lexically in order not to give rise to the symmetry problem (Horn, 1972).

Some scholars have argued that the same mechanism that derives SIs could be responsible for CP, and different proposals have been made as to what the competitor sentence that generate the inference should look like. One first hypothesis could be to consider "if and only if" as a stronger scalar term to compete with *if* (Atlas and Levinson, 1981). However, as van der Auwera (1997) points out, if we take *iff* to be the stronger term, the assertion of the weaker statement "if p then q" should result in the negation of the stronger "if and only if p then q", which is the opposite of the observed effect. For this reason, van der Auwera (1997) proposed the following alternative for a scale.

(37) ...
if p, q, and if r, q and if s, q
if p, q and if r, q
↑ if p, q

The scale in (37) is based on the fact that sentences enumerating more alternatives that allow q will always be stronger than just "if p, q", therefore under this theory asserting the latter corresponds to negating all the alternative conditions for q. However, this

proposal has attracted some criticism, base on the fact that a scale like (37) would require an infinite list of possible alternatives (Horn, 2000). Since SIs are computed by the hearer comparing the weaker statement to the stronger alternative(s), this would require the hearer to keep in their mind an infinite list of possible competitors. In addition to this, Von Fintel (2001) argues that it is generally assumed that relevant stronger statements should not be more complex forms of the weaker statement, otherwise it could be supposed that the speaker has chosen the weaker statement for the sake of brevity and not because they believe the stronger statement to be false. A similar idea is supported by Katzir (2007).

Horn (2000) suggested the scale in (38), in which the competitor for the conditional is given by the statement q whitout any further condition.

(38) q / Whatever the case, q
↑ if p, q

Horn claims that the conditional is not only weaker, but also contains extra information (i.e. a sufficient condition for q) and, following the assumption that speakers are committed to offer relevant information (Grice, 1975), this must mean that the condition expressed by the conditional statement is necessary, since nothing could be more relevant than necessity. Von Fintel (2001) argues that a similar reasoning does not derive CP, but a weaker implicature: negating the statement q does not imply that p is the necessary and sufficient condition for p, but it only that q is not unconditionally true.

4.2.3 CP and exhaustification

As briefly mentioned in the last section, the most recent account of CP relies on question exhaustification. The account was initially proposed by Von Fintel (2001), who argues that CP should be regarded as a quantity implicature, arising from the effect of a (possibly implicit) QUD. Groenendijk and Stokhof (1984) have observed that answers to questions are generally interpreted to be exhaustive: for example in the following examples (39) and (40) it is implied that Robin and Hilary and no one else left the party early; and that Sandy only ate a tossed salad for lunch. The observation holds for constituent answers (in bold in the examples) as well as sentential ones.

(39) a. Q: Who left the party early?
b. A: **Robin and Hilary** left the party early.

(40) a. Q: What did Sandy eat for lunch?
b. A: Sandy ate **a tossed salad** for lunch.

This interpretation occurs when a question is answered by a conditional statement as well, as in the example in (41): again the answer is understood exhaustively, therefore giving rise to the implication that the availability of vegetarian food is the only condition under which Robin will come to the party.

(41) a. Q: Will Robin come to the party?
b. A: **If there is vegetarian food** Robin will come to the party[4].

Based on this observation, Von Fintel proposes that the interpretation of a conditional in its perfected form depends on whether the conditional is received as the answer to a question which is eliciting an exhaustive answer. Therefore, different QUDs could give rise to different interpretations of the conditional, in a way similar to what the Question and Answer Requirement predicts for scopally ambiguous sentences. Von Fintel discusses different situations in which CP does not arise depending on the relevant question: for example, conditionals in a situation in which the QUD is whether p is a sufficient condition for q do not get perfected.

(42) a. Q: Will you give me five dollars if I mow the lawn?
b. A: Sure, (I will give you five dollars if you mow the lawn).

Another example is a situation in which the QUD is on the antecedent, like in (43).

(43) a. A: John is in Amherst today.

[4](41) is a yes-no question, that should receive a yes-no answer. Groenendijk and Stokhof (1984) had considered conditionals under their material implication account, treating them as truth value functions, which does not raise any problem with using the conditional as an answer to a yes/no question. Von Fintel (2001) subscribes to an account based on a (covert) universal quantifier, and takes the dialogue in (41) is taken to be a shortened version of one in which the first speaker has asked under which condition Robin will come to the party.

b. Implicit QUD: What (of current interest) follows from John's being in Amherst today?

c. B: If he is in Amherst, he'll be home late tonight.

Here, the answer is not taken to mean that if John is not in Amherst, he will not be late - for example, he could be home late every time he takes a trip. Another possibility are mention-some answers: Groenendijk and Stokhof (1984) had already noted that not all questions require exhaustive answers. In some contexts, a partial answer can suffice. This is evident in the example in (44): the hearer would not require, nor expect, an exhaustive list of options where one acceptable option can suffice.

(44) a. Q: Where around here can I buy Italian newspapers?

b. A: You can get them at Out of Town News in Harvard Square.

Following Lilje (1972), Von Fintel also reconsiders the classic "five dollars" example.

(45) a. If you mow the lawn, I'll give you five dollars.

b. → Only if you mow the lawn, I'll give you five dollars.

Here, he claims, the implication is much weaker than (45-b): namely, that the five dollars are not free for taking, and that the hearer must do something to earn them. Mentioning one option for obtaining them that is acceptable both to speaker and hearer might be all that is required.

4.2.4 Summary

This section has provided an outline of the phenomenon of Conditional Perfection, focusing on the fact that CP is considered a pragmatic inferences and providing an account of the condition under which it arises. Two major proposals as to its nature have been discussed: the GCI/SI approach, and a recent approach based on the QUD. It needs to be reminded however, that the two accounts are not incompatible, since the QUD affects the computation of SIs, as demonstrated experimentally by Zondervan et al. (2008). The next section moves onto the issue of how conditional meaning is constructed in the mind of the speaker, discussing the two major theories on the topic.

4.3 Reasoning on conditionals: Mental logic vs Mental model

This section presents the two main accounts on conditional reasoning: one based on mental logic theories and one based on mental models. This literature is relevant for this dissertation as theories on conditional reasoning have often served as a background to justify some of the results obtained in experiments on the interpretation of conditionals - and in particular, the mental model program provides a possible account of the differences observed between children and adults in this regard, as it will be discussed below.

4.3.1 The Mental logic program

The mental logic program, proposed by Braine and O'Brien (1991) is a mental natural deduction system, and it is based on describing inference rules in order to establish how a logical particle (including conjunction, disjunction and *if*) is interpreted. Braine and O'Brien claim that a valid theory should predict which inferences involving a logical particle will arise consistently and be considered valid, as well as explain how people judge the truth of sentences containing the logical particle in question. Indeed, in this model the meaning of an indicative conditional is entirely dependent on its inferences: judging the truth table approach problematic because of the paradoxes arising from it, the authors propose an account of the meaning of conditionals based on two primary inferences: Modus Ponens and the Schema for Conditional Proof.

Modus Ponens (MP) is considered one of the standard inferences arising from conditionals, together with Modus Tollens, and research has focused extensively on their occurrence (the findings obtained by some of these studies are discussed in section 4.4.2).

(46) Conditional statement: If there is vegetarian food, Robin will come to the party.
MP: If p then q; p, therefore q.
There is vegetarian food, therefore Robin will come to the party.

The second inference, called Conditional Proof, states that when the consequent of a conditional can be derived from a set of premises and the antecedent of the conditional

as a hypothetical assumption, then the veridicality of the conditional can be asserted on the premises alone.

(47) Conditional Proof: first suppose p; for any proposition q that follows from the supposition of p taken together with other information assumed, one may assert *if p then q*.

According to the schema for Conditional Proof, to evaluate whether a conditional is true in a given state of affair, the reasoner initially supposes that the antecedent is true, then tries to derive the consequent on the basis of the information known about the state of affair. If the operation is successful then the reasoner judges the conditional true. If on the other hand the negation of the consequent is derived, then the conditional is judged false.

Braine and O'Brien's model does not posit that lexical entries of connectives correspond to truth tables: instead, the inference rules provide instructions as to how truth follows from premises to conclusions. However they also claim the reasoning schema is subject to some restrictions: for example, it has to be stipulated that nothing follows from a contradiction (except that one of the initial assumption must be wrong). In addition to this, they claim that no evaluation of the conditional statement is possible when the antecedent is false, as the constraints blocks the derivation, leaving the meaning of the sentence undetermined. While the Mental logic program offers an explanation for some of the intuitions offered by reasoners when they are asked to evaluate conditionals (in particular in relation to "defective" response, corresponding to the truth table in 4.2), one issue with this proposal is that, by denying that connectives are associated to truth table, it is not clear how they compose with other sentences, for example in the case of embedded conditionals like the one in (48).

(48) John knows that if it rains, Bill will cancel the class.

The next section moves onto the Mental model theory, which has provided the background for a large number of experimental studies conducted on conditionals (Barrouillet and Lecas, 1998, Barrouillet et al., 2000, 2008, Gauffroy and Barrouillet, 2009), and provides a possible account of children's interpretation of conditionals.

4.3.2 Mental model conditionals

The mental model theory, proposed by Johnson-Laird and Byrne (1991), is an approach based on the notion that people generate mental models that represent entities, events, processes and relations iconically: that is to say, mental models have a structure that corresponds to the structure of what they represent, and contain a token for each referent, property and relations between referents. For example, this theory predicts that from the premises in (49), one would build a mental model like the one in (50) (Johnson-Laird, 2010), where the left-to-right axis is time, and the vertical axis represents contemporaneous events.

(49) A before B.
B before C.
D while A.
E while C.

(50)

A	B	C
D		E

Once the model in (50) had been extablished, the reasoner is able to infer a new relation from it: D before E. In this sense, the process of reasoning is given by the construction and manipulation of mental models. Crucially, Johnson-Laird and Byrne (1991) claim that people only construct mental models to represent what is true, and not to represent what is false, in order to reduce the load placed on our working memory. For example, in this model the representation of a sentence containing a connective, like (51), would proceed in the steps represented in (52), with each line representing a possibility.

(51) Either there is a blue circle, or there is a green triangle.

(52)

There is a blue circle.
There is a green triangle.

In regard to conditionals, the authors are assuming a material conditional truth table, therefore the mental model constructed for a conditional represents the three possible

states in (53), which correspond to the rows of the truth table in which the material conditional is true.

(53)
p q
¬p q
¬p ¬q

However, the model predicts that not all the possibilities represented in (53) are immediately constructed upon encountering the conditional: since models which contain more possibilities are thought to be harder to process, initially only the situation in which the antecedent is satisfied are constructed, while a "mental footnote", represented by "...", is made for the other possibilities. In other words, individual explicitly think about the possibility in which both p and q are true, and they use the footnote to flesh out the model at a later stage should it be required.

(54)
[p] q
...

In (54), the first row represents the co-occurrence of the propositions p and q, while the second represents an implicit model which contains the alternative circumstances in which p does not hold (the square brackets around p indicate its exhaustivity). Only when required (e.g. when the premises are not satisfied, that is to say, when the antecedent is false) the model is then fully fleshed out to represent the three rows of the truth table in which the material conditional is true, therefore obtaining the representation in (53).

Therefore this model is based around the idea that the construction of the mental model proceeds in steps, with the various steps of representing the different possibilities. It follows that the construction of each additional step incurs in a cost for the reasoner. These assumptions have some implications for the occurrence of the inferences associated to conditionals, including CP; as well as for language acquisition. It has indeed been observed that children's and adults' interpretation of a conditional statement often differ, and an influential account on why this happened has been developed from a slightly

modified version of the mental model theory (Barrouillet and Lecas, 1998, Barrouillet et al., 2000, 2008, Barrouillet and Lecas, 1999), as it will be discussed in section 4.5.1.

4.3.3 Summary

This section has discussed two accounts of conditional reasoning: one is the Mental logic program (Braine and O'Brien, 1991), which is based on the notion that the meaning of logical particles, including the connective *if...then*, is entirely based on their inferences. The second account, the Mental model theory, argues that relations between objects and events are represented in the mind of the speaker by iconic models: in the case of conditionals, this involves constructing mental models for all three situations in which a material conditional is true. The mental model theory therefore predicts that obtaining the material implication meaning would require more developed congnitive abilities compared to a biconditional or conjunction, since these are true in a subset of situation compared to the former: these predictions are borne out, as it will be discussed in more detail in section 4.5. The next section moves onto experimental studies conducted on adult participants on their comprehension of conditionals and the derivation of the inferences that arise from such sentences.

4.4 Experiments on the comprehension of conditionals

Over several decades, researchers in psychology have been concerned with investigating empirically how conditional statements are interpreted. Experimenters initially focused on testing participants comprehension of conditionals in order to establish what the truth conditions of these sentences are, as different proposals had been put forward, including the ones summarised in section 4.1. Other experiments have aimed to test the occurrence of the fallacies of Affirming the Consequent and Denying the Antecedent. Although these experiments did not test CP explicitly, these two fallacies turn into valid reasonings if a conditional is interpreted as a biconditional; therefore, their endorsement or lack of endorsement by participants could shed light on whether CP has occurred. More recently, experiments have been conducted to test the occurrence of CP, and the role of questions.

4.4.1 Truth conditional experiments

Initial experiments, conducted by psychologists, were aimed at finding out what the truth conditions of a conditional are by testing the interpretation of participants. Johnson-Laird and Tagart (1969) conducted an experiment in which participants were asked to sort a set of cards, each of which had a capital letter written on the left and a number written on the right, on the basis of a sentence referring to them. (55) is an example of such sentences.

(55) If there is an A on the left, then there is a 7 on the right.

It was found that participants would often sort the cards according to the defective conditional truth table (table 4.2): that is to say, participants judged (55) and similar statements to be irrelevant when the antecedent was false.

Legrenzi (1970), on the other hand, found different results: in his experiment, participants received a demonstration on the operation of a machine with a ball that could roll in two different directions, and some flashing lights. He reports that participants more often interpreted (56) and similar sentences as a material biconditional (73.3% of responses).

(56) If the ball rolls left, the red light flashes.

While these results do not definitely confirm any of the two hypothesis, it is possible that the variation observed in participants' interpretation was due to the different methodology used in the two experiments: for example, one thing to notice is that Johnson-Laird and Tagart (1969) had given participants the option of judging the sentence true, false or irrelevant, while Legrenzi (1970) only gave them a choice between true and false. Moreover, in Legrenzi's experiment people were asked to assess (56) as a description of the functioning of a machine: now, (56) would be true in a situation in which the red light on the machine was always flashing regardless of the position of the ball, yet, as Marcus and Rips (1979) observed, it is doubtful whether such sentences would be considered true in those circumstances. It is therefore possible that participants' answers were influenced by current knowledge on the workings of the machine, relevance or confounding factors.

Still, these results are especially relevant in what they served as a baseline for further studies aimed to test the frequency of the occurrence of the biconditional interpretation, in particular by Marcus and Rips (1979), as detailed in the next section.

4.4.2 Experiments on conditional inferences

It has long been known that conditional statements give rise to two consistent logical argument constructions: Modus Ponens (MP) and Modus Tollens (MT), and two related fallacious form of reasoning: Affirming the Consequent (AC) and Denying the Antecedent (DA). An example of each inference is provided below.

(57) Conditional statement: If there is vegetarian food, Robin will come to the party.

a. MP: If p then q; p, therefore q.
There is vegetarian food.
Therefore, Robin will come to the party.

b. MT: If p then q; not-q, therefore not-p.
Robin did not come to the party.
Therefore, there is no vegetarian food.

c. AC: If p then q; q, therefore p.
Robin came to the party.
Therefore, there is vegetarian food.

d. DA: If p then q; not-p therefore not-q.
There is no vegetarian food.
Therefore, Robin will not come to the party.

For an argument to be logically valid, there must be no logical possibility in which the premises of the argument are true and its conclusion are false. A fallacy instead is not logically valid, as the conclusion is not logically necessary given the premises: in (57-c), there could be other ways q came about, as it is not specified whether there are more conditions that would entice Robin into coming to the party. A similar reason holds for

(57-d), since it was nowhere specified that the availability of vegetarian food was the only condition for Robin to come. For these reasons, AC and DA are considered logical fallacies under the theories of meaning of conditional in section 4.1, however they turn into logically valid forms of reasoning if the conditional is perfected into a biconditional.

The occurrence of AC and DA has been tested extensively in psychological research and, more recently, these two inferences have been used as a means to test the occurrence of CP (Cariani and Rips, 2016). However, the rates of endorsement for the two fallacies vary widely from one study to another, possibly also because of the pragmatic requirements associated with the occurrence of CP. For example, Taplin (1971) conducted a study on the endorsement of DA and AC by using a Syllogism Evaluation Task. Participants were asked to judge whether a sentence would logically follow from the premises that were supplied. The syllogisms participants were asked to judge were similar in the form to (57-a-d). Taplin found that almost half of the participants (48%) endorsed DA consistently, and a slightly smaller percentage (42%) endorsed AC consistently. Taplin argues that participants were not answering in accordance to a common truth table, but instead their interpretation varied, with 37.5% of participants preferring the biconditional interpretation. In other experiments conducted with the same methodology by Taplin and Staudenmayer (1973, Staudenmayer 1975) the proportion of biconditional responses obtained ranged between 19% and 75%.

Marcus and Rips (1979) aimed at conducting an experiment to investigate a possible effect of the task and instructions received by participants that could account for the heterogeneity of previous results. In particular, they were starting from two considerations: the first one was that participants seemed to adopt a different truth table for the conditionals in Legrenzi (1970) (biconditional results) compared to Johnson-Laird and Tagart (1969) (defective conditional results). The second consideration is that previous experiments had employed one of two tasks: either conditional evaluation or syllogism evaluation. It is possible that the two tasks prompted different forms of reasoning, causing different results in the comprehension of conditionals. To avoid this, they tested participants in both tasks in order to obtain comparable results. They also used sentences similar to the ones of previous experiments, (58) and (59) are examples of Syllogism Evaluation Task problems that borrow from Johnson-Laird and Tagart (1969) and Legrenzi (1970) respectively.

(58) If there's a B on the left side of the card card, then there's a 1 on the right side.
There's a B on the left side.
Test sentence: There's a 1 on the right side.

(59) If the ball rolls left, then the red light flashes.
The ball rolled left.
Test sentence: The red light flashed.

The results for the evaluation tasks showed that responses were mostly in line with the material conditional interpretation (69.8%), with a minority of biconditional responses (7.4%). The Syllogism Evaluation Task produced 32.1% of material conditional responses, and 12.3% of biconditional responses. However, for this second task the majority of responses (52.5%) were logically contradictory, as no single truth function could account for them. In both of the tasks, a small proportion of responses signalled a conjunction interpretation of the conditional (i.e. the conditionals were accepted when both the antecedent and the consequent were true and rejected otherwise). The results obtained by Marcus and Rips differ substantially from the ones obtained by previous experiments, particularly in the rate of material conditional versus biconditional interpretations. One explanation offered in relation to the Evaluation task is that in Johnson-Laird and Tagart's (1969) study, participants had the option of judging the conditional irrelevant, while Marcus and Rips asked their participants to judge it either "consistent" or "inconsistent". They claim once this difference is taken into account, the discrepancy is explained, as they argue those participants who gave *incomplete conditional* answer in the previous study would have given material conditional ones in this one. The difference in the proportions of biconditional responses obtained in comparison with Legrenzi (1970) in turn, might be due to a feature of the experimental context that made participants more inclined towards the biconditional interpretation: recall that Legrenzi used a machine in which a ball-bearing could roll through one of two channels, thereby lighting either a red or a green lamp. Since he demonstrated the operation of the machine to the participants before they were asked to judge the conditional statements, the demonstration may have convinced them that each channel was connected to a single lamp (e.g., the left channel to the green lamp and the right channel to the red one). As for the different results obtained in the Syllogism conditions compared to previous studies with the same methodology, Marcus and Rips argue there might be an

effect of the number of trials participants were exposed to: in previous experiments (e.g. Taplin and Staudenmayer 1973) participants were asked to judge a bigger number of syllogisms, which might have helped "stabilise" their answers towards the biconditional truth function.

The studies reviewed so far do not provide a definite answer as to how conditionals are interpreted. Indeed, it has been reported in various experimental studies that participants do not seem to stick to one truth table, but instead they might be relying on other factors to derive their interpretation, and researchers have investigated the possibility that pragmatic reasoning might play a determining role on which interpretation gets selected. For this reason, further studies have been oriented towards testing the role of contextual cues and real world knowledge for the interpretation of conditionals.

In one of these studies, Cummins et al. (1991) observed that the occurrence of AC and DA inferences seems to be governed by the possible alternatives to the antecedent: for example, in the case of (60), people would be more inclined to incur in the AC fallacy in (60-a) rather than (60-b), as there are fewer alternatives as to what would cause a finger to bleed, rather than causes for cavities.

(60) a. If my finger is cut, then it bleeds. My finger is bleeding. Therefore, my finger is cut.

b. If I eat candy often, then I have cavities. I have cavities. Therefore, I eat candy often.

In their study, participants were asked to judge the conditionals once they were embedded under the AC argument in (60), and the MP, MT, and DA provided below. Participants were asked to judge whether the conclusion could be drawn from these premises.

(61) MP: If my finger is cut, then it bleeds. My finger is cut. Therefore, it bleeds.

(62) MT: If my finger is cut, then it bleeds. My finger is not bleeding. Therefore, my finger is not cut.

(63) DA: If my finger is cut, then it bleeds. My finger is not cut. Therefore, it is not bleeding.

What they found was that participants were more likely to accept AC and DA as valid form of reasoning when the causal relation between antecedent and consequent was perceived as strong, and when there were few *disabling conditions* for the causal relationship: for example, one such condition for (60-b) would be "brushing after eating a candy"[5].

Cummins et al. observed that a person that has interpreted the conditional "if my finger is cut, then it bleeds" as a biconditional is likely to endorse all of the possible arguments (MP, MT, AC and DA). From this, they assume that the participants that have shown this pattern of acceptance are likely to have perfected the conditional into a biconditional.

In another experiment on the role of context on the occurrence of AC and DA, Thompson (1994) tested adults on their interpretation of conditionals, to investigate the influence of perceived necessity and sufficiency for conditional inferences. In particular, her study was aimed at investigating whether *contextual cueing* might have an influence on the occurrence of the Affirming the Consequent (AC) and Denying the Antecedent (DA) fallacies. According to contextual cueing theories, people should be more likely to commit these fallacies when the relationship is perceived as necessary[6], as in the example below.

(64) Necessary relation: If the TV is plugged in, then it works.

a. AC: The TV is working. Is it plugged in? (Expected answer: yes)

b. DA: The TV is not plugged in. Is it working? (Expected answer: no)

(65) Non-necessary relation: If the car runs out of gas, then it stalls.

a. AC: The car has stalled. Did it run out of gas? (Expected answer: maybe)

b. DA: The car has not run out of gas. Did it stall? (Expected answer: maybe)

[5] Judgments on causal relations and possible disabling conditions were collected by participants as a part of the same experiment.

[6] As with Cummins et al.'s (1991) experiment, judgments regarding the necessity or sufficiency of the relationship between antecedent and consequent were collected from participants.

As mentioned before, Thompson (1994) claims that this is due to the perceived alternatives for the antecedent - for example, in the example above, alternative antecedents would be other reasons for a car to stall without running out of gas (e.g., broken fanbelt, overheated engine, etc.). In her experiment, participants were exposed to conditionals expressing either a necessary or a non-necessary relation, and then asked to answer questions like (64-a,b) and (65-a,b). As predicted, Thompson found that people gave very few *maybe* answers in the necessary condition with the percentage of this kind of answers never rising above 20%.

The data collected in the studies presented so far has not held a definite answer as to what are the truth conditions of conditionals, as the rate of endorsement in each condition varies. However, the results obtained seem to evidence a clear effect of pragmatic factors: first of all, the task used seem to have an effect on the results obtained. Secondly, it is evident that people are influenced by real world knowledge and prior belief in their interpretation of conditionals: the availability of alternative antecedents and contextual cues play a part in determining whether the biconditional interpretation arises. Recall that research on scope ambiguity resolution also found that the interpretation selected is affected by pragmatic factors (Gualmini, 2004b, Viau et al., 2010), and in particular by the effect of the QUD (Gualmini et al., 2008), with the interpretation that provides an answer to the salient QUD more likely to be selected. A similar approach has been proposed by Von Fintel (2001) in relation to CP: this approach, related in section 4.2.3, predicts that the bicondtional interpretation does indeed arise from pragmatics, as an effect of exhaustivity in question answers, and is therefore dependent on the QUD. The next section discusses two experiments conducted to investigate the QUD approach for CP in adults.

4.4.3 Experiments on the role of the QUD for CP

This section presents two empirical studies on the role of the QUD for CP, based on Von Fintel's (2001) proposal that sees CP as a product of exhaustivity in question answers. While the studies reviewed so far have aimed at testing the truth conditions of a conditional or the occurrence of the AC and DA inferences, and made no explicit reference to CP, the studies are of particular interest as they tested the occurrence of this phenomenon directly.

The first of these studies, conducted by Farr (2011), consisted of an offline experiment on German speakers, in which adult participants were asked to read a story which was followed from one of two questions: either an antecedent based question (*what happens if p?*) or a consequent based one (*when q?*). Participants then had to evaluate whether the answer to the question - the conditional sentence - was sufficient by answering either yes or no to (67).

(66) Monika sells seafood on the market. She gets 1 euro for a crab, 2.50 euros for an eel, 15 euros for a lobster and 2.50 euros for a pike. Kerstin, an employee of Monika, cannot remember the prices. Since she does not want to ask Monika again, she asks Sahra, who also works for Monika. Sahra knows the prices exactly.

What-if-p question:	*When-q* question:
Kerstin: What happens if I sell an eel?	Kerstin: When do I get 2.50 euros?

Target sentence: If you sell an eel, you get 2.50 euros.

(67) Did Sahra answer Kerstin's question sufficiently? [Yes] [No]

In the context provided, participants were expected to regard the conditional answer as insufficient if CP occurred. Following Von Fintel (2001), this occurrence is expected to be much more frequent following the *when-q* question rather than the *what-if-p* one, since participants know from the story that selling an eel results in nothing else than receiving 2.50 euros. The antecedent based question, on the other hand, should give rise to CP less often.

These predictions were borne out: participants gave negative answers, signalling that CP had occurred, 89.2% of the time in the *when-q* context. On the other hand, *what-if-p* questions only resulted in 16.3% of negative answers. From these results, Farr concluded that Von Fintel's theory is adequate to account for CP, as the effect of the different kind of question seems to be the decisive factor for the occurrence of this inference. However, Farr observed that the task chosen might have had an influence on the results.

A replica experiment was conducted using a TVTJ, as the latter is a methodology widely used on studies on various inferences, including SIs, which are thought to be sensitive to focus (Zondervan, 2009). In this second set of experiments, participants were asked whether Sahra's answer was true (as in (68)) instead of being asked whether it was sufficient.

(68) Is the answer of Sahra true?

Compared to the previous experiment, CP occurred less frequently in the *when-q* condition (36%). The *what-if-p* condition gave rise to CP 11.1% of the time. The difference is still more marked than what observed for SIs by Zondervan (2009), who found a less marked difference between focused and non-focused cases, with implicatures arising in 77% of the case in the first circumstances, and in 51% of the second. Farr concludes that the reason why CP occurs at a different rate compared to SIs is to be found in its nature: under Von Fintel's account, CP is a Quantity Implicature (QI). SI arise from the comparison between a scalar term and its stronger competitor, which are part of a finite list. For QIs, on the other hand, a list of potentially infinite alternatives has to be considered: in the example (39-a), repeated below as (69-a), there is no defined set of potential answers.

(69) a. Who left the party early?
b. Robert and Hilary left the party early.

Therefore, Farr takes this difference as a confirmation of the different nature of CP compared to SIs: while for *or* SIs, participants have only one stronger alternative (*and*), the comparison in more readily available, leading to a more frequent occurrence of the SI even in unfocused conditions, this is not the case for conditionals. In the CP experiments there are no concrete alternatives for participants to keep in mind which could lead to a large number of no answers in the non-focus condition.

Farr's study found concrete evidence that antecedent based questions and consequent based questions affect the interpretation of conditionals, as predicted by the QUD approach to CP proposed by Von Fintel (2001). However, it is not clear whether all questions trigger an exhaustive answer: in some contexts, a *mention-some* answer might be

sufficient.

Cariani and Rips (2016) conducted an experimental study to test the influence of different questions on the occurrence of CP. Their first aim was to provide a formal way of deriving perfection, using the *strict conditional* account of conditional meaning as a starting point.

(70) Strict: *If A, B* is true in context C at world w iff for all $v \in C(w)$, either A is false at v or B is true at v.

The next thing required is a question that can be answered by a conditional, and the assumption that the conditional will be interpreted exhaustively in this circumstance (Groenendijk and Stokhof, 1984). While a conditional could provide an answer to both antecedent based questions, and consequent based ones. Cariani and Rips (2016) provide a few examples for each category, referring to antecedent based questions as "[ANT?]" and consequent based question as "[CONS?]".

(71) [ANT?]

a. What follows from p?

b. What happens if p?

(72) [CONS?]

a. What are all the ways in which q?

b. How might q happen?

c. Is q true?

As Von Fintel (2001) postulated, consequent based questions are the ones that give rise to CP more frequently. Upon hearing a conditional that is understood as an answer to a (possibly implicit) question on the consequent Q_{cons}, the reasoning of the hearer would then follow steps.

(73) a. Step 1: provided that *If A, B* is understood as an exhaustive answer to Q_{cons}, the speaker is not in a position to assert *If D, B* for any D that competes with A.

b. Step 2: provided that the speaker is informed about the truth-values of these conditionals, she must believe all conditionals of the form *If D, B* are false.

c. Step 3: if all conditionals of the form *If D, B* are false for every antecedent that competes with A, then *if not A, not B* must be true.

Building on the previous experiments that have tested experimentally the occurrence of the DA and AC fallacies, Cariani and Rips (2016) resolved to use these inferences as a tool to test the occurrence of CP. As these two invalid forms of reasoning become valid once a conditional is perfected, they are expected to be endorsed more often by participants if they are obtaining the perfected reading of the conditional. However, previous studies that tested participants' endorsement of AC and DA were conducted with the goal to investigate causal and necessity relations. Cariani and Rips (2016) aimed instead to test the effect of explicit question for the interpretation of the conditional. For this reason, they chose sentences that did not immediately suggest a necessity relation between the antecedent and the consequent.

In their first experiment, the authors tested participants on polar questions: participants were assigned to two different conditions, one which contained a consequent based question (in square brackets below) and one in which said question was omitted.

(74) John has taken a test on Chapters 4-6 that has not been graded yet.
[You ask Mary, "Did John do well on the test?"]
Mary says, "If John understood Chapter 5, then John did well on the test."
Assume that John did not do well on the test.
Given this information, does Mary's statement imply that John did not understand Chapter 5?

The rate of *yes* responses for (74) however showed no difference between the [CONS?] and the "no question" condition. The presence of the question does not seem to have had an effect. The same results were found in an experiment that compared the effect of two different questions for the same conditional: participants were again divided into two conditions, one in which the story contained a question about the antecedent, ana

one with a question about the consequent. Moreover, the requirement for exhaustivity was made explicit in the questions used.

(75) John has taken a test on Chapters 4-6 that has not been graded yet.
[CONS?] Condition: You ask Mary: "What are all the ways John could manage to do well on the test?"
[ANT?] Condition: You ask Mary: "What are all the things that could happen in case John understood Chapter 5?"
Mary responds, "If John understood Chapter 5, then John did well on the test."
Assume that John did well on the test.
Given this information, does Mary's statement then imply that John understood Chapter 5?

These contextual manipulation did not have an effect, as they did not lead participants to perfect conditionals: the rate of endorsement for AC and DA compared to MP and MT was in line with the baseline provided by previous experiments. The different question did not have an effect either, with [CONS?] only marginally raising the acceptance rate for AC compared to [ANT?].

Based on the results obtained in these experiments, Cariani and Rips (2016) argue that, while CP can be linked to exhaustivity, there is some evidence that questions alone are not sufficient to obtain exhaustivity, but other factors are necessary to trigger the exhaustive reading of conditionals. They claimed that the results could be due to two possible reasons: either the participants failed to infer a perfected conditional from an exhaustive answer, or participants did not interpret the conditional as exhaustive in the first place.

Different results were obtained in two experiments in which the conditional was either followed or preceded by an explicit statement that entails its perfected reading, underlined in (76).

(76) John has taken a test on Chapters 4-6 that has not been graded yet.
[CONS?] Condition: You ask Mary, "What are all the ways John could manage to do well on the test?"
Mary responds, "If John understood Chapter 5, then John did well on the test

That is the ONLY way John could have done well on the test."

[ANT?] Condition: "Mary responds, If John understood Chapter 5, then John did well on the test.

That is the ONLY thing that could have happened if John understood Chapter 5."

Assume that John did well on the test.

Given this information, does Mary's statement imply that John understood Chapter 5?

A similar effect was obtained by adding a sentence that explicitly informs the participant that the speaker knows all the ways the consequent could come about, and is willing to relate them.

(77) a. [ANT?] condition: You ask Mary, "What are all the things that could happen in case John understood Chapter 5?"

b. [CONS?] condition: You ask Mary, "What are all the ways John could manage to do well on the test?"

Mary responds, "If John understood Chapter 5, then John did well on the test."

In both these experiments where the exhaustivity of the speaker's answer was made explicit, participants were likely to endorse the AC and DA inferences when exposed to a question on the consequent. The rate of endorsement decreased significantly when the speaker was answering an antecedent based question.

One of the possible explanations provided by Cariani and Rips for the results of the first experiments was that one of the reasons for CP not arising could be that the conditional answer is not interpreted exhaustively. In light of the different results obtained in the last two experiments, in which the exhaustivity was made explicit, this explanation seems likely. It is possible that the contexts used in the experiments encourage participants to prefer a *mention some* answer instead of a *mention all* one, required to interpret the conditional exhaustively. This argues against the premises of Von Fintel's account, namely that questions trigger an exhaustive response.

Consider a dialogue like (78).

(78) a. Who are all the people who came to the party?

b. John came.

(78-a) is phrased as demanding a *mention all* answer. Still, the *mention some* seems acceptable: one could infer that the speaker is rejecting the responsibility of naming a long list of participants. Similarly to what proposed by Farr (2011) in her observation regarding the difference between SIs and CP, Cariani and Rips argue that the questions used in the study ("what are all the ways John could manage to do well on the test" and similar questions) placed an impossible demand on the addressee, as the list of possible alternative is potentially endless.

4.4.4 Summary

This section has presented some of the experimental studies that have been conducted over the course of several decades to investigate the interpretation of conditionals. Initial experiments conducted in psychological research were aimed at determining the truth conditions of conditional sentences. Such experiments have often obtained conflicting results, showing that adult participants tends to shift between interpretation, not sticking to one version of the truth table. Children on the other hand often interpret the conditional as if it were a conjunction, judging the sentence true only when both antecedent and consequent are true. Experiments conducted on conditional inferences also had variable results, which can most likely be ascribed to to the fact that said inferences are influenced by pragmatic reasoning. While the studies conducted on conditional inferences provided a way of testing CP indirectly, only recently researchers have turned to testing its occurrence directly by investigating the effect of the QUD for this inference, following the theory proposed by Von Fintel (2001). The results obtained in the two studies presented in this section, however, are conflicting: while Farr (2011) found a clear effect of the QUD, Cariani and Rips's (2016) results are less clear, as they found that participants were not always prone to interpret questions exhaustively. QUD approach to CP offers an interesting path of research, especially from a language acquisition perspective: as the scope studies related in the first part of this dissertation have evidenced, children and adults seem to differ in their management of the QUD, and this could have

an effect for the interpretation of conditionals as well. On the other hand, the SI account of CP has not been tested experimentally so far, in spite of the fact that SIs are a widely investigated phenomenon that has been studied extensively in language acquisition. In light of this, next section moves onto the acquisition of conditional reasoning, discussing the results of some of the experiments conducted on children's interpretation of conditionals and drawing a parallel with disjunction, which has been shown to receive a similar interpretation.

4.5 The development of conditional reasoning

It has often been reported in experimental literature on the comprehension of conditionals that children interpret conditionals differently compared to adults: the first study to report this difference was the one conducted by Paris (1973), who used a truth value methodology: children and adult participants were asked to to judge the veridicality of a conditional in the four possible state of affairs: *p, q*; *not-p, q*; *p, not-q*; and *not-p, not-q*. Paris found that 8-year-old children would only accept the conditional as true in the first situation, in which both antecedent and consequent are true, therefore obtaining a conjunctive-like interpretation of the conditional. Interestingly, a similar behaviour was observed by Paris for sentences with the disjunction. Several subsequent studies have reported similar results for conditionals, and different explanations have been attempted as to why children display this behaviour: the Mental logic program for conditionals sees it as arising from pragmatics, with children initially obtaining a *defective* interpretation for the conditional, under which the interpretation of a conditional with a false antecedent is undetermined (Braine and O'Brien, 1991). Braine and O'Brien (1991) further claim that, when one is in doubt as to whether consider a proposition true or false, it is better to consider it false if a judgment has to be made: therefore, when prompted to make a judgment on a sentence with an undetermined truth value, people prefer to judge it false. Another possibility is that children at this stage of development might lack a facility for hypothetical thinking, and therefore they might be unable to correctly assess the truth conditions of a conditional. A number of recent experimental studies however have attempted an explanation based on the mental model approach for children's conjunctive interpretations, leading to the claim that a slightly amended version of the mental model theory can justify the results obtained for children (Barrouillet and

Lecas, 1998, Barrouillet et al., 2000, Barrouillet and Lecas, 1999). As mentioned before, conjunctive-like readings are reported to arise for disjunction sentences as well (Paris, 1973, Braine and Rumain, 1981, Gualmini et al., 2001, Chierchia et al., 2004): a recent theory claims that these interpretation arise as an SI, being obtained by children and not by adults due to the different alternatives available to the two groups (Singh et al., 2016).

This section is organised as follows: section 4.5.1 reviews some of the recent studies on children's interpretation of conditionals. Section 4.5.2 moves onto children's interpretation of sentences containing a disjunction, as several studies have evidenced that children give a conjunctive inerpretation for such sentences. This section also discusses some of the studies conducted on children's behaviour in relation to SIs, as this phenomenon is associated to both disjunction and CP. Section 4.5.3 discusses an account of children's conjunctive interpretation of disjunction based on SIs.

4.5.1 Experimental studies on children

Several empirical studies have been concerned with comparing children's and adults' interpretation of conditionals, in order to investigate whether the two groups differ, and how the interpretation changes with age. The majority of these studies have used either the truth table or the conditional inferences as a methodology: for example, Paris (1973) observed that children aged about 7 tend to judge a sentence *if p then q* true when both p and q are true and false in every other case, therefore effectively treating it as a conjunction.

In a similar study, Barrouillet and Lecas (1998) asked participants to judge which cases of the truth table violated a rule in the form *if p then q*, finding comparable results: third graders (aged 7-8) were likely to give a conjunctive interpretation for the conditional (64%), with this percentage decreasing with age (32% and 9% for 11-12 year-olds and 14-15 year-olds respectively). The authors claim results can be justified by the mental model approach: recall that in this theory, presented in section 4.3.2, deductive reasoning is a process based on the construction and manipulation of mental representations. Since the task of constructing such models is demanding, reasoners would initially construct a simpler initial model: in the case of a conditional, such model would only contain the

possibility in which the antecedent is satisfied, and an implicit representation of other possibilities, as shown in (79).

(79) [p] q
. . .

Only when required (e.g. when the premises are not satisfied, that is to say, when the antecedent is false) the model is then fully fleshed out to represent the three rows of the truth table in which the material conditional is true, like in (80).

(80)
p	q
¬p	q
¬p	¬q

According to the theory, having to construct a higher number of models makes the deduction harder to obtain, due to the fact that they are more tasking for the subject's working memory. As children are reported to have lower working memory capacities compare to adults, Barrouillet and Lecas (1998) claim that rate of conjunctive interpretation for conditionals obtained by younger participants can be justified by making a small change to the mental model theory. They suggest that, initially, a single model representing the co-occurrence of p and q, with no additional implicit model and no supposition of exhaustivity of p, is constructed. This representation would therefore be equivalent to a conjunctive one. Young children, with limited working memories, would be unable to construct any further model, hence sticking to a conjunctive interpretation of the conditional. On the other hand, older children seem to favour biconditional interpretations, which might be due to the addition at this developmental stage of a second model, in the form $\neg p\ \neg q$. Finally, the conditional interpretation which requires the full three models is achieved at an older age.

To test this hypothesis, Barrouillet and Lecas (1999) conducted a double experiment in which the same participants took part in two tasks, one on their interpretation of conditionals and one aimed to test their working memory capacity. The conditional task produced the same pattern observed in previous experiments: participants were given one conditional rule (e.g. "if you put on a white shirt then you must put on green trousers") and then were asked to represent all the cases that complied with that

rule (e.g. pair a blue shirt with green trousers, a blue shirt with red trousers, etc.). Again 8-year-olds gave a high rate of conjunctive responses (49%), with 11-year-olds preferring biconditional responses instead (72%) and 14-year-old generally interpreting the conditional according to the truth table. The second experiment was a counting span task aimed at determining the working memory capacity of participants, who were given a set of card each printed with a varying number of dots. Participants were asked to count the dots on each card, and then to recall the number of dots present on each card in the correct order in which they were presented. Barrouillet and Lecas (1999) found that the participants who consistently gave conjunctive responses also displayed lower working memory span rates compared to participants that had given biconditional and conditional interpretations, with the latter group of participants showing the highest rate of the three. The authors argue that these results confirm their hypothesis: the three interpretations (conjunctive, biconditional and conditional) appear in succession because of the improving working memory capacity, which allows for the construction of an increasing number of mental models.

The conjunctive interpretation for conditional, however, has been observed in adults as well as children, with adults often showing a near equal mixture of conjunctive and conditional interpretations (Barrouillet et al., 2000, Oberauer and Wilhelm, 2003, Evans et al., 2003). A further confirmation of the possibility that conjunctive interpretation may be related to working memory capacity seems to come from experiments conducted on adults: Evans, Handley, Neilens, and Over (2007) found that adults who performed better in a test of general intelligence were also more likely to prefer the conditional interpretation over the conjunctive one: in particular, higher ability participants gave responses in line with the defective truth table for conditionals (table 4.2). Similar results were obtained by Fugard, Pfeifer, Mayerhofer, and Kleiter (2011), who also report that conditional interpretations take longer to process compared to conjunctive ones. They adopted a probabilistic semantics for *if*, following Ramsey (1990), who argued that, to infer their degree of belief in *if* p *then* q, people initially assume p and then fix their degree of belief in q. If p turns out to be false, these degree of belief are rendered void. Under this proposal, one explanation could be that the participants that select a conjunctive interpretation are executing this process only partially.

In sum, there is some evidence that children go through a series of stages in the development of conditional thinking, starting with a conjunctive interpretation in which the

conditional is deemed true only when both conjuncts are true (therefore equating it to *and*), then moving upon a biconditional interpretation and finally reaching a conditional one with adult age. However, there are plenty of studies conducted on adults that show that neither conjunctive nor biconditional interpretations disappear as people age, as these are still reported to be frequent in adults.

Interestingly, the same kind of conjunctive interpretation is widely reported for children for sentences with disjunction as well: Paris (1973) has already observed that this interpretation is frequent in young children. The topic attracted considerable attention in the field of language acquisition, with several subsequent studies (Braine and Rumain, 1981, Chierchia et al., 2004, Gualmini et al., 2000) obtaining similar results, as it will be discussed in the next section. Moreover, sentences containing a disjunction often give rise to a SI: the derivation of this kind of inference is also often reported to differ between children and adults (Noveck, 2001, Papafragou and Musolino, 2003, Guasti et al., 2005, Foppolo et al., 2012, Chierchia et al., 2001, among others). The next section reviews some of the studies that have been conducted on this topic.

4.5.2 Children's behaviour with disjunction and SIs

As discussed in section 4.2.2, one of the most prominent accounts of CP sees it as a scalar implicature, arising from comparing the conditional against a more informative alternative: according to Horn (2000), the more informative alternative is simply the statement that *q* (or that *whatever the case, q*), therefore the scale to derive CP would look like the one in (81).

(81) (Whatever the case,) q.
↑ if p, q.

The derivation of SIs, like the interpretation of conditionals, is another area in which the behaviour of children and adults is reported to differ, with children deriving SIs at a lower rate compared to adults, at least with scalar terms like the quantifier *some* and the disjunction. In addition to the scalar implicature angle, sentences containing the disjunction have also received a great deal of attention in language acquisition due to the fact that they sometimes receive a conjunctive interpretation by children (Paris,

1973, Braine and Rumain, 1981, Chierchia et al., 2004, Gualmini et al., 2000), which in turn has been explained as an effect of children's different derivation of SIs (Singh et al., 2016).

At a semantic level, a sentence containing the disjunction *or* like (82) is true when either one of the disjuncts are true, as for the table in 4.3.

(82) John ate chicken or beef.

p	q	p or q
T	T	T
T	F	T
F	T	T
F	F	F

Table 4.3: Truth conditions of disjunction.

However, (82) is often associated to a SI, arising from the comparison with the stronger alternative containing the conjunction, (83). Because of the implicature, the disjunction is usually understood as incompatible with (83) and therefore interpreted as an exclusive disjunction.

(83) John ate chicken and beef.

It has often been reported that children's interpretation of sentences containing a disjunction is different from adults': Paris (1973) reported that children give a conjunctive-like interpretation for sentences containing a disjunction, an observation that was later confirmed by Braine and Rumain (1981). Paris's explanation of the results is that children are adopting what he calls a "mismatch strategy": comparing auditory information (the sentences containing the connective, which were read to them by the experimenter) to the visual information (the picture representing the various situations corresponding to different lines of the truth table) without any regard for the logical connection expressed

by the disjunction[7]. Braine and Rumain (1981) offered a similar explanation claiming that children 5-6 years old assess the truth conditions of the two propositions, but disregarding the connective.

Moreover, conjunctive interpretations are not the only point of difference between children and adults in relation to disjunction. As mentioned before, a sentence with disjunction is often interpreted by adults on its exclusive reading due to the scalar inference arising from the comparison with the stronger sentence containing a conjunction, in the same way as it happens with other scalar terms: however, empirical research has often shown that children derive SIs at a much lower rate compared to adults: for example, Smith (1980) had observed that children treat a sentences containing *some* as compatible with *all*. Several subsequent studies have investigated children's behaviour by using a variety of methodology and different scalar terms. In one of the first studies, Noveck (2001) tested participants on their interpretation of sentences containing modal verbs.

As discussed in the section 4.2.2, modal verbs can give rise to scalar implicatures. The verb *might*, for example, is part of a scale which also includes *must*. This gives rise to the inference that (84-b) is false.

(84) a. There might be a parrot in the box.
b. There must be a parrot in the box.

Noveck (2001) conducted an experiment aimed at testing whether children 5 to 9 years old get the same pragmatic interpretation (the scalar implicature) with the modal verb *might* as adults: if this is the case, they should reject a sentence containing this verb as false in a context in which the *must* sentence is true. In this study, participants were presented with three boxes, one of which was closed, and the other two were open so that participants could see their content. One of the open boxes contained a parrot and the bear, and the second contained only a parrot. The experimenter then told participants that the content of the closed box was unknown, but they were the same as either the "parrot + bear" box or the "parrot only" box. After receiving this explanation, participants were then asked to judge sentence (84-a). Noveck found that children from

[7] A similar explanation was offered by Taplin, Staudenmayer, and Taddonio (1974) in relation to children's interpretation of conditionals: they observed that the conjunctive interpretation might be due to the fact that children are unable to grasp the logical connection between the two proposition, instead regarding the sentence *if p, then q* as a simple juxtaposition of the events expressed by *p* and *q*.

7 years of age had sufficient mastery of the task. Of this group, 80% of the 7 year old and 69% of the 9 year old accepted the test sentence as a description of a situation in which the stronger scalar term is also true. In contrast to the children, only 35% of adults accepted the same sentences. From this evidence, Noveck inferred that children, differently from adults, interpreted *might* in its logical (semantic) interpretation, treating it as compatible with *must* due to the fact that the pragmatic interpretation is less accessible to children. In another experiment, Noveck tested the quantifier *some*. Each participant was presented with 5 sentences similar to (85).

(85) Some giraffes have long necks.

Similarly to the first experiment, the sentences were presented to participants in a situation that also made true the stronger alternative (i.e. "all of the giraffes have long necks."). The results were not different form the ones of the modal verbs experiment, as Noveck found that 8- and 10-year old children were likely to agree with all the 5 test sentences (this happened in 68% and 70% of cases respectively for the two age groups). Adults showed the same pattern of acceptance only 33% of the time.

In the same way, studies on SIs arising from disjunction sentences revealed a similar pattern: children do not usually derive scalar inferences at the same rate adults do, but instead they frequently judge (82) and similar sentences as true in a situation in which the stronger competitor (83) is also true. Among the researchers that have investigated this phenomenon, Gualmini et al. (2000) tested children and adults' interpretation of a sentence containing a disjunction in a conjunctive context. For example, in the case of (86) participants were told a story in which Tigger and Pooh Bear are driving, and they need to change a flat tyre, but neither of them can take the wheel off. Finally they join forces and try to take the wheel off together, and they succeed. This makes the test sentence (86-a) false, unless interpreted as compatible with (86-b).

(86) a. Tigger or Pooh Bear can take off the wheel.
b. Tigger and Pooh Bear can take off the wheel.

Gualmini et al. (2000) found that only 4 out of 13 children rejected (86-a) and similar sentences, while adults always rejected them. Similar results were obtained by Chierchia

et al. (2004): in one of the tasks for this study, participants were presented with a puppet, who made a bet on the outcome of a story that they were about to be told. The bet had the form in (87).

(87) I bet that Batman will take a cake or an apple.

In the story, it turned out that Batman took both a cake and an apple, therefore making both disjuncts true: in this situation, children accepted the puppet's bet as true 95% of the time, while adults acceptance was lower at 60%.

In sum, the results obtained for both quantifiers, modal verbs and disjunction seem to confirm that children interpret sentences with scalar elements such as disjunction differently from adults. As it is generally agreed that children do posses knowledge of the semantic meaning of those terms, the different behaviour supposedly arises from their inability to derive SIs, which in turn is generally attributed to reasons arising from pragmatics: for example, Noveck (2001) claimed that, as the scalar inference arises when relevant, the difference between the two groups might be due to the fact that children's relevance conditions are more easily satisfied than adults, and therefore they tend to stick to the literal meaning. Chierchia et al. (2004, 2001) offered an explanation based on processing factors: they argue that, while children posses the necessary pragmatic and semantic knowledge required to obtain the implicature, they are not able to use it because of processing limitations. Another possible explanation is that children might be more pragmatically tolerant, accepting the weaker sentence in spite of the pragmatic violation (Katsos and Bishop, 2011).

However, other studies have shown children to access SIs at a higher rate compared to the studies reviewed so far: for example, Papafragou and Musolino (2003) investigated children's performance by using a variety of scalar terms, in order to compare whether some of them give rise to SIs more often than others. In their study, children and adult speakers of Greek were tested on the occurrence of scalar implicatures with the numeral *two*, the quantifier *some* and the verb *start*, in a situation in which sentences containing their stronger competitors (three, all, finished) were also true. For example, in the case of the quantifier *some*, participants were told the story in which three horses decide to jump over a fence, and all of them succeed in doing so. A puppet then offered (88-a)

as a description of the story, with the context making it true on its literal meaning, but pragmatically infelicitous due to the stronger alternative (88-b) also being true.

(88) a. Some of the horses jumped over the fence.
b. All of the horses jumped over the fence.

Similarly, for both the *two / three* and *start / finish* condition, the context would make true the stronger alternatives (89-b) and (90-b).

(89) a. Two horses jumped over the fence.
b. Three hoses jumped over the fence

(90) a. The girl started making the puzzle.
b. The girl finished making the puzzle.

Participants were assigned to one of the three conditions exemplified above, and after being presented with the stories, they were asked to judge whether the puppet "answered well" or not. Papafragou and Musolino found different results across the conditions: in line with previous findings, children almost never rejected the puppet's statement in the *some / all* condition (12.5%) and the *start / finish* condition (10%). However, they did reject it in the majority of cases in the *two / three* condition (65%). Adults on the other hand almost always rejected the puppet's statement across all three conditions (above 90%). Papafragou and Musolino's results cast some doubts over the fact that the behaviour observed in children for modal verbs, quantifiers and the disjunction can be generalised to all other elements that give rise to SIs.

This consideration triggered further research on children's computation of SIs with elements other than the ones used in the studies presented so far. One example of this is Free Choice Inferences (FCIs). FCIs are a form of quantity implicature, which are thought to be derived as scalar implicatures (Fox, 2007, Franke, 2011, Chemla, 2009, Van Rooij, 2010): under this account, (91-a) receives the meaning in (91-b) because it is compared against the stronger sentences (92-a) and (92-b).

(91) a. Tom may drink tea or coffee.

b. Tom may drink tea and Tom may drink coffee.

(92) a. Tom may drink tea.

b. Tom may drink coffee.

In a TVJT conducted on Mandarin speaking children, Tieu, Romoli, Zhou, and Crain (2015) compared children's derivation of SI on (91-a) and similar sentence to their derivation of SIs associated with the disjunction. The study found that, in line with previous results, children generally interpreted the disjunction on its inclusive reading, deriving the exclusive inference only in 18% of cases. On the other hand, they were able to derive the FCI for (91-a) and similar sentences 91% of the time.

These results cast a new light on the previous ones, and in particular on the idea that children are not able to derive SIs. It was initially proposed that children are unable to derive said inferences, proposed on the basis of the results obtained with disjunction, quantifier and modal verb sentences (Chierchia et al., 2004, Noveck, 2001). However, the reason for children inability to compute SIs for those sentences could instead be found in the fact that children fail to construct alternatives in the same way as adults, as they are unable to derive those alternatives that require lexical replacement.

In relation to perfectible conditionals, it is interesting to notice that, under its SI analysis proposed by Horn (2000), CP does not require lexical substitution: the alternative q is already contained in the conditional sentence. In this sense, CP should be more akin to FCIs than to SIs with disjunction or quantifiers.

(93) (Whatever the case,) q.

↑ if p, q.

Indeed, a study conducted by van Tiel and Schaeken (2017) found evidence in favour of this view, as well as further evidence in favour of the alternative based approach to SIs: van Tiel and Schaeken measured adults' reading times for implicature triggering sentences with *some*, like (88); as well as FCIs and perfectible conditionals. The results showed that computing scalar inferences with *some* was associated to a processing cost for participants. FCIs and perfected conditionals, on the other hand, did not incur in the additional processing cost. This alternative based approach is particularly relevant, as

it has been the basis of a recent theory on children's interpretation of sentences containing a disjunction, for which they give conjunctive-like interpretation, similarly to what happens for conditionals. Indeed, Singh et al. (2016) argue that the conjunctive interpretations observed in children for such sentences might be explained by their inability to construct the same alternatives as adults, as will be discussed in the next section.

4.5.3 Conjunctive interpretation of *or*: Alternative based approach

A recent proposal, developed by Singh et al. (2016), builds on the Alternative based account to provide an explanation for children's conjunctive interpretation of sentences containing a disjunction: under this account, said conjunctive interpretation is indeed a SI, which can be explained by the assumption that children have access to a different set of alternatives compared to adults.

It is thought that there are two steps that are crucial to computation of implicatures for a sentence S: the derivation of alternatives, and the strengthening process, called *ALT*, and *STR* in Singh et al. (2016) respectively (Horn, 1972, Gazdar, 1979).

(94) a. ALT: generates a set of alternatives for S, *ALT(S)*.
b. STR: strengthens the basic meaning of S by negating specific elements of *ALT(S)* and conjoining the result with S.

The third element necessary to the computation of SI for sentences with disjunction is the knowledge of the meaning of the logical operator involved: it is generally agreed upon that children know the meaning of the logical operator, Singh et al. (2016) argues that the difference between children and adults must be due to one of the two other elements, the functioning of which is discussed in more detail below.

STR is identified by Singh et al. (2016) as equivalent to the operator *exh* in Fox (2007): a grammatical exhaustivity operator, equivalent to a covert form of *only*, which takes as arguments a sentence p and its set of alternatives. Singh et al.'s (2016) definition of *exh*, borrowed from Fox (2007), is given in (95), where A represents the set of alternatives, S represents the sentence, c its context and *IE* represents the set of alternatives that are innocently excludable.

(95) $[\![\text{exh}(\text{A}(\text{S, c}))(\text{S})]\!] = [\![\text{S}]\!] \wedge \bigwedge \{\neg[\![\text{S}_i]\!] : \text{S}_i \in \text{IE } (\text{A}(\text{S,c}))\}$

ALT is the mechanism that generates a set of alternatives for the sentence under consideration: there are two ways in which alternatives are generated, which include lexical substitution (for example, in the case of SIs arising from quantifiers like (85) and similar sentences, where the alternative is derives by substituting *some* for *all*) and deletion (FCIs like the ones derived from (91-a) and similar sentences arise from alternatives derived by deletion).

The strengthening of a sentence s happens by exhaustifying the alternatives of s, ALT(s), by conjoining s with the negation of various elements of ALT(s), while maintaining the coherence of s. The sentence is therefore parsed as *exh*(ALT(s),s), and the resultant is the strengthened form of the sentence s, which corresponds to s conjoined with the negation of various elements of ALT(s). Singh et al. (2016) claim these assumptions are sufficient to derive both the adult strengthened meaning of disjunction (exclusive disjunction) and the child one (conjunctive interpretation). Consider a sentence s, corresponding to $p \vee q$. Recall that there are two mechanisms for deriving alternatives: the first one, lexical replacement, requires lexical access and is not available to children under the Alternative based approach. Alternatives however can also be derived by deleting lexical material (Katzir, 2007). The set of alternatives to s for adults and children are presented in (96-a) and (96-b) respectively.

(96) a. $\text{ALT}_{adult} = \{\text{p, q, p} \wedge \text{q}\}$

b. $\text{ALT}_{child} = \{\text{p, q}\}$

As it can be observed in (96), the adult set of alternatives, p and q are both in the set, and so is their conjunction $p \wedge q$. This makes p and q unavailable to *exh*, as they are both closed under conjunction: neither p nor q are innocently excludable, because each is symmetric with the other. $p \wedge q$ on the other hand is innocently excludable, and remains the only alternative available in ALT(s). The exhaustification of these alternatives returns the exclusive disjunction $(p \vee q) \wedge \neg (p \wedge q)$ as the strengthened meaning of s, s^{+}. This is exemplified in (97).

(97) Processing of disjunctive sentence in adults:

a. EXH$(ALT_{adult})(p \vee q)$

b. $ALT_{adult} = \{p,\ q,\ p \wedge q\}$

c. EXH$\{\text{ALT}_{adult}\} = (\text{p} \vee \text{q})\ \neg(\text{q} \wedge \text{q})$.

In the case of children, however, the process involves two parsings: as q and r are not closed under conjunction, they are available and *exh* can apply recursively. The sentence will therefore be parsed as in (98), where C represents the set of alternatives available to children ({p, q}).

(98) EXH$_2(C_2)($EXH$_1(C_1)(p \vee q))$

The first application of *exh*, (99-a), is vacuous, as p and q are the only two alternatives in the set, and they are symmetric with each other, hence neither is innocently excludable. As nothing can be negated at this stage, the meaning obtained by the parse in this application of *exh* is simply the inclusive disjunction: $p \vee q$.

(99) First layer of exhaustification:

a. EXH$(ALT_{child1})(p \vee q)$

b. $ALT_{child1} = \{\text{p, q}\}$

c. EXH$\{\text{ALT}_{child1}\} = (\text{p} \vee \text{q})$.

On the second application of *exh*, (100-b) however, the symmetry gets broken: the new set of alternatives contains $\{p \wedge\neg q,\ q \wedge\neg\ p\}$, which are no more symmetric and can be negated.

(100) a. EXH$(\text{ALT}_{child2})((\text{EXH}(\text{ALT}_{child1})\text{p} \vee \text{q})$

b. $ALT_{child2} = \{$EXH$(ALT_{child1})(s) :\ s \in ALT_{child1}\} =$ EXH$(\text{ALT}_{child1})\text{p},$ EXH$(\text{ALT}_{child1})\text{q}, = \{\text{p} \wedge\neg\ \text{q},\ \text{q} \wedge\neg\ \text{p}\}$

c. EXH$\{\text{ALT}_{child}\} = \text{p} \wedge \text{q}$.

After the second application of *exh*, the negated alternatives are now conjoined with $p \vee q$, therefore the new enriched meaning of the sentence is $(p \vee q) \wedge\neg(p \wedge\neg\ q) \wedge\neg(q \wedge\neg\ p)$. This negates all possible situations except for the one in which both p or q are true, deriving the conjunctive inference.

Singh et al. (2016)'s proposal has the merit of providing an explanation for children's conjunctive interpretation of sentences with disjunction while maintaining previous assumptions, including that children know the meaning of logical operator (in contrast with the claim made by Paris (1973) and Braine and Rumain (1981) that the conjunctive interpretations observed in children for disjunction sentences is due to the fact that participants ignore the logical operator, and simply match the contact of the two disjuncts to the situation they are shown), and the idea that children might only be able to derive those inferences that do not require lexical substitution.

4.5.4 Summary

This section has discussed some of the language acquisition studies relevant for CP: firstly, studies on the comprehension of conditionals conducted on children have generally evidenced that they tend to give a conjunctive interpretation for these kind of sentences. Several different explanations have been proposed for this behaviour: children might be unable to comprehend the logical relation between the two propositions (Taplin et al., 1974) and simply interpret the conditional as a juxtaposition of the events expressed by the two propositions. Another explanation is that the conjunctive interpretation might arise for pragmatic reasons, due to the consideration that a sentence with a false antecedent cannot be definitively determined to be true or false, and when in doubt, it is better to consider the sentence false (Braine and O'Brien, 1991). Building on the Mental model theory, it has also been proposed children might lack the cognitive abilities necessary to construct all mental models required for the meaning of conditionals, finding themselves only abel to generate the one that represents the co-occurrence of antecedent and consequent (Barrouillet and Lecas, 1998, Barrouillet et al., 2000, Barrouillet and Lecas, 1999). Conjunctive-like interpretations have been observed in children for sentences containing *or* (Paris, 1973, Braine and Rumain, 1981): a recent proposal claims that the reason for this might be due to children's different derivation of scalar inferences (Singh et al., 2016). It has indeed been observed that children often derive these kind of inferences at a much lower rate compared to adults (Noveck 2001, Papafragou and Musolino 2003, Foppolo et al. 2012, Chierchia et al. 2001 among others). One possible explanation of this is that while children are able to perform the strengthening operation necessary to derive a SI, they might be unable to access the lexicon to retrieve the scalemate(s) of the scalar term encountered, and therefore to obtain the

same alternatives that adults do. Indeed, children have been shown to be adult like in their derivation of FCIs, which do not require lexical access (Tieu et al., 2015). In the same way, under Singh's (2016) account, children's conjunctive interpretations are derived as a SI, due to the fact that the alternatives available to children differ from the ones available to adults.

4.6 Summary of chapter

This chapter has presented an overview on conditional perfection, taking different accounts for the semantics of conditionals as a starting point, and focusing on particular on identifying the characteristics of conditional perfection, and the circumstances that lead to its occurrence. It is indeed established in the literature that CP is a pragmatic phenomenon, as researchers have observed that its occurrence is limited (van Canegem-Ardijns and Van Belle, 2008, Dancygier and Sweetser, 2005), and, similarly to other pragmatic inferences, it is defeasible (van Canegem-Ardijns and Van Belle, 2008). Moreover, experimental studies have confirmed that pragmatic factors have an effect on the occurrence of the AC and DA inferences, the endorsement of which is linked to a biconditional interpretation of the conditional statement. While its nature as a pragmatic phenomenon is recognised, the exact nature of CP is debated, with two prominent hypotheses seeing it either as a Scalar Implicature (Horn, 2000, van der Auwera, 1997) or a Generalised Conversational Implicature arising from informativeness (Levinson, 2000). More recently, it has been proposed that the QUD might play a role in the occurrence of CP, with the biconditional interpretation arising when the conditional is understood as a response to a question about its consequent (Von Fintel, 2001). However, empirical studies that have investigated this phenomenon have found conflicting results (Farr, 2011, Cariani and Rips, 2016). While the Scalar Implicature account of CP has received some criticism, mostly due to the difficulty of defining a viable scale that could give rise to the biconditional inference, it is still considered a viable hypothesis (López Astorga, 2014a,b, Moldovan, 2009). In addition to this, experimental research on children's comprehension of conditionals has shown that they do not obtain the same interpretation that adults do, as children tend to judge conditionals true only when both antecedent and consequent as true, therefore giving a conjunctive-like interpretation of such sentences (Paris, 1973, Barrouillet and Lecas, 1998, Barrouillet et al., 2008, Gauffroy and

Barrouillet, 2009, Barrouillet and Lecas, 1999). A similar behaviour was observed in children for *or*: children tend to interpret sentences containing the disjunction as if it were a conjunction (Paris, 1973, Braine and Rumain, 1981). A recent proposal sees this conjunctive-like interpretation of disjunction as a SI (Singh et al., 2016): while it has been observed that children generally derive SIs at a lower rate compared to adults (Chierchia et al. 2004, Noveck 2001, Papafragou and Musolino 2003, Foppolo et al. 2012 among many others), recent research has shown that children's behaviour is adult-like for those inferences in which the alternative use lexical material already present in the SI-triggering sentences, such as FCIs (Tieu et al., 2015), leading to the claim that the difference observed in previous studies (which mostly tested sentences containing *or* and *some*) was due to the fact that deriving an inference with these items require a lexical substitution. Similarly, according to Singh et al. (2016), conjunctive interpretation in children can be explained in the same terms, with children deriving this interpretation as a SI from a different set of alternatives.

Chapter 5

Conditional Perfection experiments

The aim of this chapter is to present an empirical study conducted on adults and children in order to investigate their interpretation of conditionals. More specifically, the aim of this study is to test experimentally the SI approach to CP through the language acquisition angle, employing a methodology similar to the one commonly used in experiments on SIs. As it has been discussed in the previous section, the interpretation of conditionals has been the subject of much research: empirical studies have generally evidenced that adults and children provide different interpretation of conditional statements (Paris, 1973, Barrouillet and Lecas, 1998, Barrouillet et al., 2000, Gauffroy and Barrouillet, 2009, Barrouillet and Lecas, 1999). In particular, it has been noted that children often consider the conditional true only when both antecedent and consequent are true, obtaining a conjunctive-like interpretation. Adults on the other hand display a variety of possible interpretations, ranging from what seems to be a simple material conditional interpretation (Johnson-Laird and Tagart, 1969), biconditional interpretations (Legrenzi, 1970), and in some cases showing significant rates of endorsement for the same conjunctive-like interpretation displayed by children (Paris, 1973, Oberauer and Wilhelm, 2003, Evans et al., 2003). This variation is thought to be due to the experimental design (Barrouillet and Lecas, 1998).

One of the most prominent account of conditional perfection is that it is to be considered a SI, arising from the negation of a stronger alternative: various possible alternatives

have been proposed, for instance the one developed by Horn (2000).

(1) a. If p then q.
b. (Whatever the case,) q.

Experimental studies on SIs have often revealed that children derive SIs at a lower rate compared to adults (Noveck, 2001, Papafragou and Musolino, 2003, Guasti et al., 2005, Foppolo et al., 2012, Chierchia et al., 2001, among others). One of the theories as to why SIs are harder to derive for children relies on the hypothesis that the difficulties in deriving them stems from the necessity for counterfactual reasoning (Breheny, 2006, Proust, 2002, Riggs, Peterson, Robinson, and Mitchell, 1998): upon hearing (1-a) and other SI-inducing sentences, the hearer is required to reason about what the speaker could have said - the stronger statement (1-b) - in order to infer that the latter is not true. It needs to be noted that the majority of studies that have investigated this difference between children and adults however have used a limited set of scalar terms, the most widely investigated being quantifiers (*all* - *some*) and the disjunction (*and* - *or*), and the results obtained for these two elements have been generalised to all scalar terms. Research conducted on other terms has not always confirmed this view: a study conducted by Van Tiel, Van Miltenburg, Zevakhina, and Geurts (2016) has shown that adults' derivation of SIs varies depending on the scale, with some terms giving rise to SIs more often, possibly due to the perceived distinctness of the scalar terms involved: scales in which the used word is perceived as more semantically distant from the stronger alternative give rise to more SIs. This variation is not limited to adults: indeed, Papafragou and Musolino (2003) had already observe that while children computed SIs at a very low rate for *some* sentences, they were more likely to compute them for numerals, albeit still at a significantly lower rate compared to adults. While both groups display considerable variation across scalar terms, the results are generally consistent in showing that there is a significant difference between adults and children. Based on this observation, one could expect to observe the same pattern of behaviour for conditionals, with adults deriving CP more often.

However, another hypothesis has been proposed on the derivation of SIs, which claims that children do not derive these inferences in situations in which adults do because they are unable to construct the required alternatives when these require lexical access,

but they are indeed able to derive SIs when the alternatives are made available to them (Gualmini et al., 2001, Tieu et al., 2015, Barner, Brooks, and Bale, 2011, among others). One way for the alternative to be available is to be contained in the uttered sentence, as it is the case for conditional under the scale in (1), in which the consequent q serves as the competitor and is obviously a subconstituent of the conditional statement itself.

Based on the results obtained by van Tiel and Schaeken (2017), who compared reading times for SIs for sentences like (2-a) to the ones for CP and other inference-triggering sentences, it does indeed seem the case that the derivation of CP is differs from the derivation of the SI in (2-c) and other inferences that require lexical access. When a lexical substitution is required to derive the stronger alternative, like in the case of (2-b), reading times increased, signalling an increased processing cost. The derivation of CP on the other hand did not seem to incur in such cost.

(2) a. Some of my friends speak French.
b. All of my friends speak French.
c. → Some but not all of my friends speak French.

(3) a. If you mow the lawn, I'll give you five dollars.
b. (Whatever the case,) I'll give you five dollars.
c. → If you mow the lawn, I'll give you five dollars, and it's not the case that, whatever the case, I'll give you five dollars.
(→ if and only if you mow the lawn, I'll give you five dollars).

As well as evidencing children's conjunctive interpretation for conditionals as discussed above, Paris's study also revealed that children behave similarly with the disjunction *or*, again treating it as if it were a conjunction. Starting with these findings, children's interpretation of disjunction has received much attention in subsequent research, which has focused on investigating the reasons behind the difference uncovered between the two groups: given that children seem to have a knowledge of the semantic meaning of *or*, experimental studies have generally concluded that the difference observed in comparison to adults must arise from other causes (Braine and Rumain, 1981, Gualmini et al., 2001, Chierchia et al., 2001). Moreover, as children usually derive SIs at a much lower rate compared to adults, they sometimes treat *or* as an inclusive disjunction, compatible with

and (Gualmini et al., 2001, Chierchia et al., 2001). Adults on the other hand derive the implicature in (4-b) from sentence (4-a), interpreting the disjunction exclusively.

(4) a. John ate chicken or beef.
b. John ate chicken or beef but not both.

Singh et al. (2016) claims that this is how conjunctive inferences are derived in children: his proposal, discussed in section 4.5.3, provides an explanation for children's conjunctive interpretation of disjunction observed in several experimental study (Paris, 1973, Braine and Rumain, 1981, Chierchia et al., 2004, Gualmini et al., 2000).

As mentioned before, conditionals are also reported to receive a conjunctive interpretation by children (Paris, 1973, Barrouillet and Lecas, 1998, Barrouillet et al., 2000, 2008, Gauffroy and Barrouillet, 2009, Barrouillet and Lecas, 1999), and sometimes adults as well (Paris, 1973, Oberauer and Wilhelm, 2003, Evans et al., 2003), but in spite of this they have not received the same degree of attention in language acquisition research. While children's interpretation of conditionals has been investigated extensively, the studies on this topic have mostly been aimed at researching how conditional statements are represented in the mind of the hearer: of particular relevance for language acquisition is the Mental Model program, proposed by Johnson-Laird and Byrne (1991), which has provided the basis for the account of development of conditional reasoning developed by Barrouillet and his colleagues (Barrouillet and Lecas 1998, Barrouillet et al. 2000, 2008, Gauffroy and Barrouillet 2009, Barrouillet and Lecas 1999 a.o.). Under this proposal, a model representing the co-occurrence of the antecedent and the consequent is the first to be constructed in the mind of the hearer upon coming across a conditional and, as children do not yet posses the same degree of cognitive capacities compared to adults, they are unable to add the further models representing situations in which the antecedent is false. This section instead presents the results of two experiments conducted on children's and adults' comprehension of conditionals from the perspective of language acquisition, using a methodology similar to the ones employed in research on scalar inferences in order to establish what interpretation of conditionals is obtained by children in such circumstances.

5.1 Experimental studies

Two experiments were conducted on children's and adults interpretation of conditional sentences, and particularly on the occurrence of CP. These studies were conducted with a Truth Value Judgment Task methodology (Crain and Thornton, 1998), in which participants were asked to judge the truth of a conditional sentence used to describe a short story - this type of task is widely employed in studies on SIs in children and adults. A detailed description of the procedures and materials used is provided in the next section.

5.2 Experiment 1

Experiment 1 tested children and adults on interpretation of conditionals using a TVTJ: participants were asked to judge the veridicality of a set of sentences, each one of which was provided as description of a short story. The short story was presented to participants with the aid of an animated PowerPoint presentation shown on a laptop while the experimenter read through the script. Prior to the beginning, participants were given a brief explanation of the task. After this, the PowerPoint presentation began with some introductory slides: participants were told that they were about to hear some stories about a character called Mr Pig, who likes to sell things in the market and brings something different to sell every day (picture 5.1(a)). The next slide showed the other characters appearing in the presentation, Mr Pig's friends who buy things from him when he is in the market (picture 5.1(b)). The last of these slides introduced the puppet, Lucy, and participants were told that she would be watching the story and trying to guess what was is going to happen (picture 5.1(c)).

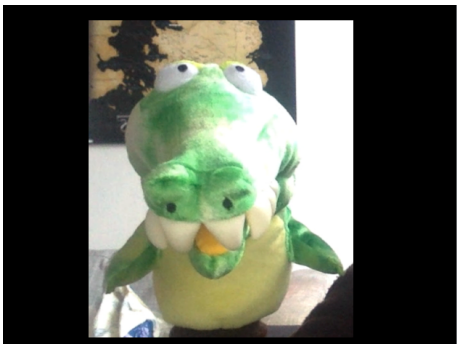
Participants were then told that they would have to judge whether what the puppet said was true or false based on what happened at the end of the story. While the stories were very simple, with minimal changes in wording, the scenario differ in terms of whether the antecedent, the consequent or both were true in the context.

5.2.1 Predictions

The aim of this experiment is to test how conditionals are interpreted by children and adults, and in particular to test the occurrence of CP.

(a) Introduction of Experiment 1, slide 1.

(b) Introduction of Experiment 1, slide 2.



(c) Introduction of Experiment 1, slide 3.

FIGURE 5.1: Introduction of Experiment 1

One of the most prominent proposals on the nature of CP is that it is to be regarded as a SI: in this case, one would expect to observe the same kind of behaviour commonly observed for this kind of inferences, with children obtaining the inference at a lower rate compared to adults, as reported in many experimental studies conducted on SIs. It has been observed however that the rate at which implicatures are computed, by both children and adults, varies depending on the specific scalar item: for example Van Tiel et al. (2016) noted that these inferences occur more frequently in adults when the scalar items are perceived as more distinct. Similarly, Papafragou and Musolino (2003) found that children derive more SIs with numerals and verbs than they do with quantifiers, although their performance remains different from adults'. That is to say, even if within population we might find differences across different inferences, we expect to see the same pattern across populations. This *uniformity prediction* has indeed already been used in the literature to determine the SI-status of other inferences, namely multiplicity inferences (Tieu, Bill, Romoli, and Crain, 2017), under the formulation in (5).

(5) **Uniformity prediction of the implicature approach:** If multiplicity inferences and scalar implicatures are of the same nature, we expect to observe the same pattern of between-group differences (or between-group similarities) across the two kinds of implicatures.

The Alternative based approach, however, makes different predictions: children have been shown to be adult-like for those implicatures that do not require lexical substitution, such as FCIs (Tieu et al., 2015). Under Horn's (2000) scale, which we are adopting here, CP is one such inference, as the stronger alternative *q* is contained in the continditional *if p then q*, children should therefore be adult like and obtain CP at the same rate adults do, or at least their performance should be closer to adults' compared to what is observed for scalar implicatures involving lexical substitutions .

5.2.2 Methodology

5.2.2.1 Participants

35 adults and 19 children (age range 4;01 - 7;08, mean age 5;06) participated in the study. Children participants were recruited from primary schools of the Greater Belfast area. Participants who did not answer at least 75% of control and filler items correctly were excluded from the analysis. After exclusion, data from 31 adults and 17 children were analysed.

5.2.2.2 Materials and procedure

The experiment consisted of 16 short stories associated to a sentence, of which 4 were targets, 8 controls (4 conditional controls, 2 *and* controls and 2 "FF-controls", an example of each is provided below) and 4 fillers. Prior to the beginning, participants received two training items, one of which was true and the other false, which were not conditionals and did not contain any connective. Participants were give feedback for their answer on these items.

(6) a. **Training item true:**

Today Mr Pig has brought strawberries to the market. Puppet, what do you

think will happen?

Puppet: Mr Pig will sell the strawberries to Cat.

Oh look, today Cat came to the market! Cat decides to buy the strawberries from Mr Pig. Now the day is over, the market is closed and Mr Pig goes home.

So, Puppet had said "Mr Pig will sell the strawberries to Cat", was Puppet right or wrong?

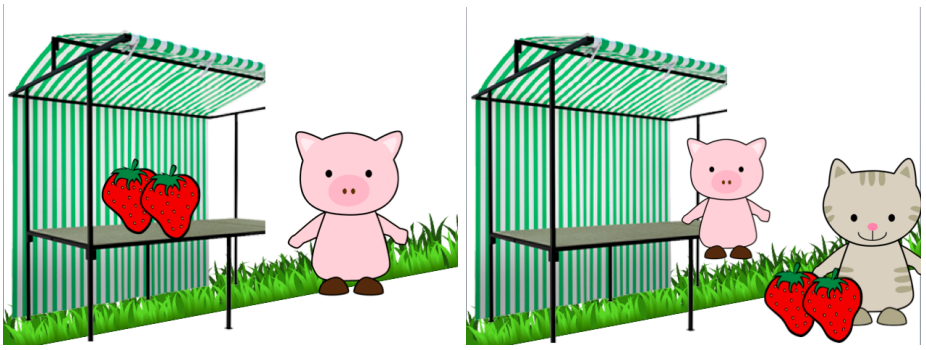
b. **Training item false:**

Today Mr Pig has brought sweets to the market. Puppet, what do you think will happen today?

Puppet: Mr Pig will sell the sweets to bear.

Oh look, today Cat came to the market. Cat decides to buy the sweets from Mr Pig. Now the day is over, the market is closed and Mr Pig goes home.

So, Puppet had said: "Mr Pig will sell the sweets to bear", did she say the right thing or the wrong thing?



(a) Example of a training item for Experiment 1: Mr Pig will sell the strawberries to Cat.

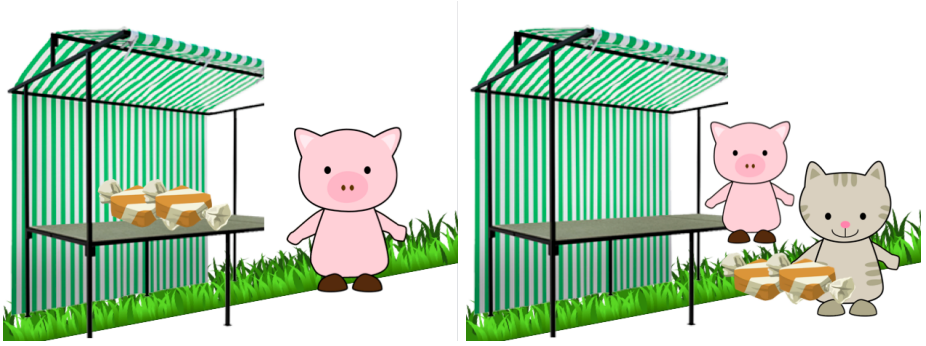
(b) Outcome of training item

FIGURE 5.2: Training item true.

The two training items were followed by the 16 main trials, which were arranged in a pseudo-random order, with each target item followed by a filler to balance the number of *yes* and *no* answer. Target sentences were aimed at testing the occurrence of a biconditional interpretation, and to do so, they consisted of a conditional presented in a "F-T" scenario, like in (7).

(7) **Target item:**

Today Mr Pig has brought cheese to the market. Puppet, what do you think it



(a) Example of a training item for Experiment 1: Mr Pig will sell the sweets to Bear.

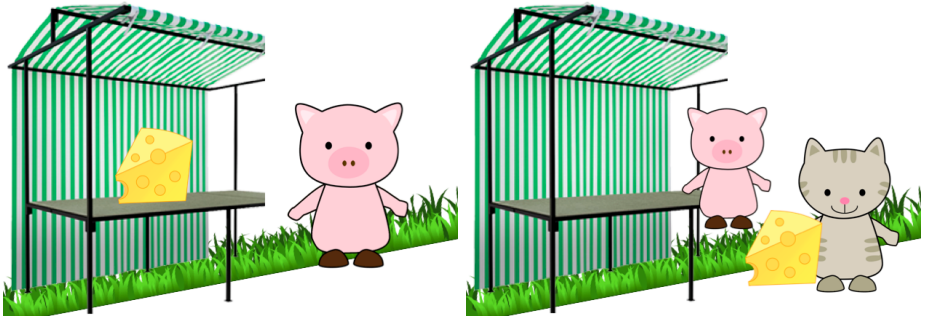
(b) Outcome of training item

FIGURE 5.3: Training item false.

will happen? Puppet: If Elephant is coming to the market today, Mr Pig will sell his cheese.

Oh look, today Cat came to the market. Cat decides to buy the cheese from Mr Pig.

So, Puppet had said "If Elephant is coming to the market today, Mr Pig will sell his cheese", did she say the right thing or the wrong thing?



(a) Example of a test item for Experiment 1: If Elephant is coming to the market today, Mr Pig will sell his cheese.

(b) Outcome of test item.

FIGURE 5.4: Test item for Experiment 1.

While the antecedent of the conditional is false, the consequent is true, making the sentence as a whole true under the material conditional truth table. However, the sentence is false on its biconditional interpretation, so the participant's answer should disambiguate between these two readings. The target sentence is also false under the conjunctive interpretation of the conditional, therefore a negative answer would also be compatible with this reading. As it can be observed from the script in (7), the context stories were kept very simple to be sure that the effect of the context would be kept to

a minimum: for example, the generic question "what will happen?" was used instead of more specific ones (for example, questions based on the antecedent or consequent of the test sentence) to avoid a possible effect of the QUD.

Two types of control items were used to make sure that participants could distinguish between a conditional sentences and conjunction: *and* controls and FF-controls. The two *and* controls were used to ascertain that participants interpreted sentences containing the conjunction correctly. These two control were both false: in one of them both conjuncts were false, and in the other one of the conjuncts was true. This last case is exemplified in (8).

(8) ***And* control:**

Today, Mr Pig has brought cake to the market. Puppet, what do you think will happen today?

Puppet: Chicken is coming to the market today, and Mr Pig will sell his cake.

Oh look, today Chicken came to the market! Chicken doesn't like cake, so he leaves without buying the cake.

Now the day is over, the market is closed and Mr Pig goes home. So, Puppet had said "Chicken is coming to the market today, and Mr Pig will sell his cake", did she say the right thing or the wrong thing?

FF-controls, as their name suggests, where conditionals presented in a "F-F" situation in which both antecedent and consequent are false, like the one in (9).

(9) **FF-control:**

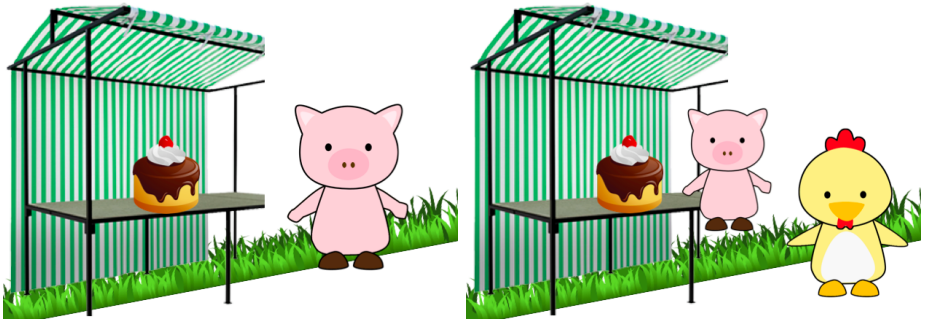
Today Mr Pig has brought balloons to the market. Puppet, what do you think will happen today?

Puppet: If Sheep is coming to the market today, Mr Pig will sell his balloons.

Sheep couldn't come today to the market. Instead, today Chicken came to the market! Chicken doesn't like balloons, so he leaves without buying the balloons.

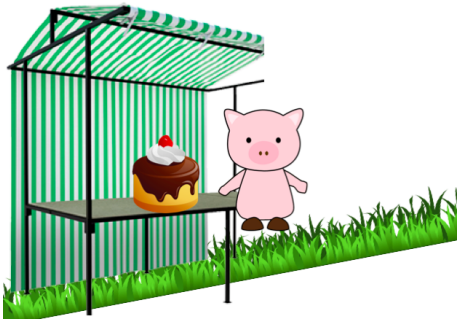
Now the day is over, the market is closed and Mr Pig goes home.

So, Puppet had said: "If Sheep is coming to the market today, Mr Pig will sell his balloons", did she say the right thing or the wrong thing?



(a) Example of *and* control for Experiment 1: Today Chicken is coming to the market today, and Mr Pig will sell his cake.

(b) *And* control for Experiment 1: Chicken came to the market



(c) Outcome of *and* control

FIGURE 5.5: *And* control for Experiment 1

Each one of the target items was followed by a filler which was associated to either a true sentence or a false one, in order to balance the number of positive and negative answers. The filler items were not conditionals nor did they contain any connective, but were instead similar to the training items used for the task.

(10) **Filler item:**

Today Mr Pig has brought pencils to the market. Puppet, what do you think will happen today?

Puppet: Mr Pig will sell the pencils to Elephant/Sheep.

Oh look, today Sheep came to the market! Sheep buys the hats. Now the day is over, the market is closed and Mr Pig goes home.

So, Puppet had said "Mr Pig will sell the pencils to Elephant/Sheep", did she say the right thing or the wrong thing?

4 more control items were used to test that participants had a sufficient understanding of conditionals to reject them in a false situation ("T-F") and accept them in a true one

("T - T"). Examples of these are provided below.

(11) **True control:**

Today Mr Pig has brought oranges to the market. Puppet, what do you think it will happen today?

Puppet: If Giraffe is coming to the market today, Mr Pig will sell his oranges.

Oh look, today Giraffe came to the market. Giraffe decides to buy the oranges from Mr Pig.

Now the day is over, the market is closed and Mr Pig goes home. So, Puppet had said: "If Giraffe is coming to the market today, Mr Pig will sell his oranges", did she say the right thing or the wrong thing?

(12) **False control:**

Today Mr Pig has brought eggs to the market. Puppet, what do you think will happen today?

Puppet: If Giraffe is coming to the market today, Mr Pig will sell his eggs.

Oh look, today Giraffe came to the market! Giraffe doesn't like eggs, so he leaves without buying the eggs.

Now the day is over, the market is closed and Mr Pig goes home. So, Puppet had said: "If Giraffe is coming to the market today, Mr Pig will sell his eggs", did she say the right thing or the wrong thing?

In summary, there were 4 types of conditional items, to represent each possibility of the truth table, with two types of non-conditional items (fillers and *and* sentences) making up the remaining trials.

Item type	S1	S2
Control true	T	T
Target item	F	T
Control false	T	F
FF-Control	F	F

Table 5.1: Summary of experimental item types

5.2.3 Results and discussion

Children and adults generally performed well on control and filler items, answering them correctly in the vast majority of cases: 96.5% and 90.6% respectively for the two groups.

Rejection of the target items occurred in 72.1% of cases for children, and in 92.7% of cases for adults: the difference between the two groups was significant (Mann-Whitney U = 3344, p <0.001).

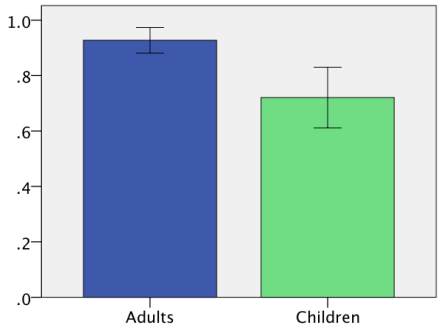


FIGURE 5.6: Experiment 1: *No* responses for test sentences.

Recall that one of the two predictions made by the SI account of CP were that children would be unable to derive CP, or derive it at a lower rate compared to adults: this prediction seems to be borne out. While children were comparable to adults in their interpretation of the test sentences, nevertheless the difference was significant: this is compatible with a *uniformity prediction*. For example, the results obtained here are roughly compared to the ones obtained by Papafragou and Musolino (2003) for numerals, for which it was found that children derived the implicature and rejected the test sentence in the majority of cases (65%), but still significantly less than adults. Furthermore, in some trials participants were prompted to provide a reason for their rejection: when doing so, they usually referred to the false antecedent, which provides further evidence in favour of the fact they had derived CP. The Alternative based approach predicted that children would derive this kind of implicature at a higher rate compared to those that require lexical substitution, and indeed it is possible that the fact that this process of substitution was not required in this case helped children in accessing this inference.

The results obtained for FF-controls, however, points towards a different direction. These items were also frequently answered negatively by both adults (87.1%) and children (82.4%), and no significant difference was found in this case between the two groups (Mann-Whitney U = 1004, p = 0.531). These results point away from a CP

interpretation for the test items, and towards the possibility that both groups obtained a conjunctive interpretation: indeed it is possible that the majority of the participants were simply rejecting the "F-T" situation test items because they had obtained a conjunctive interpretation of the conditional sentence, as this behaviour has been reported for both groups in previous literature.

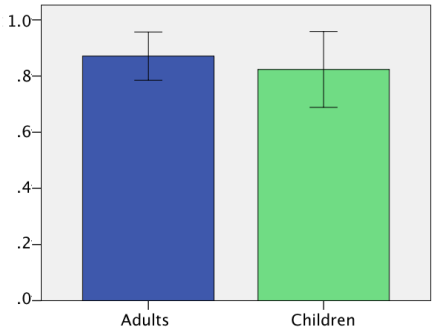


FIGURE 5.7: Experiment 1: *No* responses for FF-controls.

The results obtained for children however mirror more closely those obtained in previous experiments: several studies have indeed shown that children tend to reject conditionals as false unless both antecedent and consequent are true; a possible explanation for these results is provided by the Mental model theory, according to which this interpretation is thought to arise because only the mental model in which both antecedent and consequent are true is formed in the child's mind (Barrouillet and Lecas, 1998, Gauffroy and Barrouillet, 2009, Barrouillet and Lecas, 1999). The other major account of conditional reasoning - the Mental logic program, on the other hand, relies on pragmatics to justify these results: according to Braine and O'Brien (1991), their model makes the same predictions as a defective truth table in table 5.2.

p	q	if p then q
T	T	T
T	F	F
F	T	I
F	F	I

TABLE 5.2: Truth conditions of a defective conditional.

Therefore, the truth value of a conditional with a false antecedent is indeterminate, and the conjunctive and biconditional pattern of response would arise as the result of the application of two different pragmatic principles: the conjunctive responses would result

from the fact that, since *not-p* cases are deemed to be irrelevant for the conditional statement, the question asked by the experimenter is deemed inappropriate. The conjunctive response would result from a pragmatic judgment according to which it is better to consider this proposition to be false rather than true if a choice has to be made. This possibility will be discussed in more detail in the next section, in relation to the result of a follow-up experiment.

The next section presents the results of a follow-up experiment conducted in order to test the effect of priming on the interpretation of conditionals.

5.3 Experiment 2

The results obtained for Experiment 1 evidenced that adults almost always reject a conditional in a "F-T" situation, which could be attributed to the occurrence of CP. However, participants also frequently rejected control items in which both antecedent and consequent were false: from this, it can be inferred that both groups obtained a conjunctive interpretation of the conditional statements. A follow-up experiment was conducted, which consisted of a replica of Experiment 1 using the same materials and methodology, but adding one training item that could bias participants towards giving a positive answer to the FF-control items.

5.3.1 Predictions

Using a "F-F" situation as a training item, and instructing participants that a conditional in such situation is to be considered true, should have an effect on the responses received for FF-control items, reducing the number of negative answers and bringing participants to accept these more.

5.3.2 Methodology

5.3.2.1 Participants

34 adults and 27 children (age range 4;11 - 6;06, mean age 5;07) participated in the experiment. Children participants were recruited from primary schools of the Greater

Belfast area. Participants who did not answer at least 75% of control and filler items correctly were excluded from the analysis. After exclusion, data from 28 adults and 21 children were analysed.

5.3.2.2 Materials and procedure

All test, controls and filler items were the same as for Experiment 1 - with the exception of two different training items before the main trials. The two training items of the previous experiments were replaced with two conditional sentences, one presented in a situation equivalent to a false control, like (12) and the other equivalent to a FF-control like the one in (9). Participants received feedback for these two items: while the vast majority of people correctly gave a false response to the first one, the majority of them also rejected the FF-control type training item. These participants were invited to reflect on such sentences, and invited to treat such item as true.

5.3.3 Results and discussion

Again, children and adults performed well on controls and fillers, with an accuracy of 91.4% for adults and 92.3% for children.

Compared to the previous experiment, rejection for the test items decreased slightly: children gave a negative answer 69.3% of the time, and adults 56.3% of the time. Unlike the previous experiment, this time the difference between the two groups was not significant (Mann-Whitney U = 4284, p = 0.059).

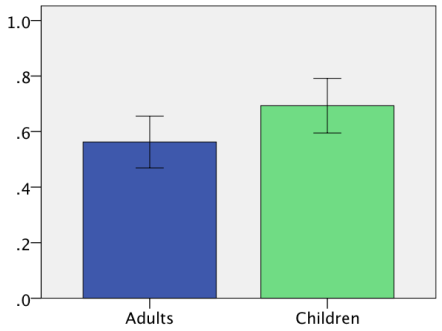
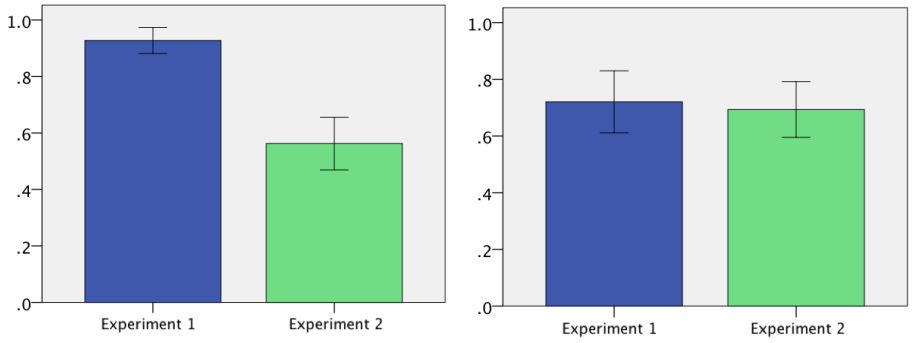


FIGURE 5.8: Experiment 2: *No* responses for test sentences.

Across the two experiments, the decrease is only significant for adults (Mann-Whitney U = 4410, p <0.001), while children's decrease in acceptance is not significant (Mann-Whitney U = 2910, p = 0.711).



(a) Experiments 1 and 2: rejection rate of test items for adults.

(b) Experiments 1 and 2: rejection rate of test items for children.

FIGURE 5.9: Comparison of rejection rate for test items between the two experiments.

Adults' rejection of FF-control items decreased significantly, but in spite of the training item, it remained above 60%. While this result is lower than the one obtained in Experiment 1 and closer to what found in previous studies, adult participants still gave a response consistent with a conjunctive interpretation, even after being instructed that such items were to be considered true in the training items phase. Similarly to the first experiment, children rejected these controls over 90% of the time. The difference between the two groups was significant in this case (Mann-Whitney U = 848, p <0.001).

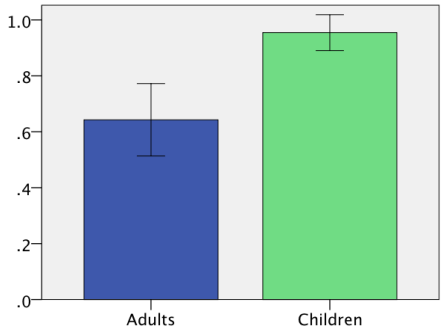
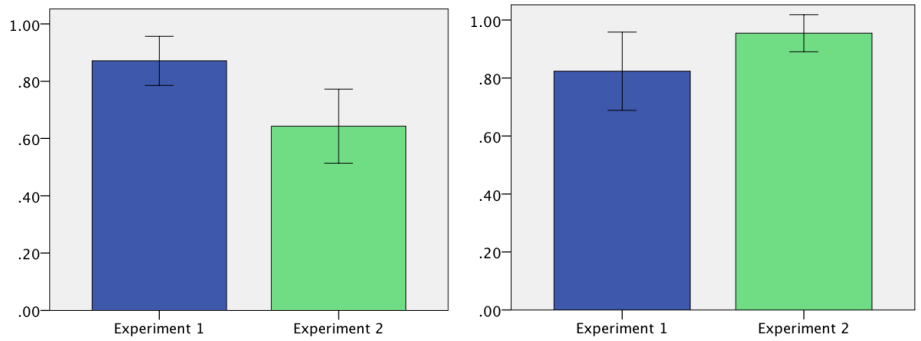


FIGURE 5.10: Experiment 2: *No* responses for FF-controls.

Comparing adults' performance across the two experiments (figure 5.11(a)), we found a significant difference, as the percentage of rejection significantly increased compared to Experiment 1 (Mann-Whitney U = 1340, p <0.008). On the other hand, children's percentage of rejection for the test items varied only slightly (figure 5.11(b)), and no

significant difference was found between the two performances (Mann-Whitney U = 650, p = 0.060).



(a) Experiments 1 and 2: rejection rate of FF-control items for adults.

(b) Experiments 1 and 2: rejection rate of FF-control items for children.

FIGURE 5.11: Comparison of rejection rate for FF-control items between the two experiments.

Adults display an increase in acceptance across the two experiments for both test items and FF-controls. While the latter was expected since participants were explicitly instructed to consider FF-control items true in the context provided, the increase in acceptance for test items is surprising. One possible explanation might be found in the priming effect of the FF-control type training item. Participants were instructed to consider this item true, and this effect might have carried over to test items, acting as a "reminder" of the fact that a conditional can be true if its antecedent is false, therefore also making participants more prone to accept the F-T situations. This factor only affected adults, as children seem less sensitive to the feedback received in the training phase of the experiment: neither their answers to the test items nor the ones to the FF-controls changed significantly.

Going back to the predictions made for Experiment 1 in section 5.2.1, these results do not offer a conclusive answer as to whether CP is to be considered a SI: adults percentage of rejection for FF-items remains high in spite of the training offered, and children show no change in their pattern of behaviour. The data obtained for both experiments shows that both groups obtain a conjunctive-like interpretation in both cases. Based on previous literature, there are a few possibility that could explain the results obtained in this study, which will be explored in turn in the rest of this section. Firstly, the slightly amended version of the Mental model program proposed by Barrouillet and Lecas (1998) might be sufficient to explain the results obtained in children, even though

this model makes different prediction for adults. Another possibility to consider is that our results might have been influenced by the task: indeed, Braine and O'Brien (1991) argued that the conjunctive-like interpretations observed in children and adults are due to pragmatics. A third and more tentative explanation of our results could come from the same mechanism that derives conjunctive-like readings for sentences with disjunction in children, as proposed by Singh et al. (2016). These three possibilities are discussed in more detail below.

As discussed in Chapter 4, several experimental studies conducted on children have reported that they often select a conjunctive interpretation for a conditional statement (Paris, 1973, Barrouillet and Lecas, 1998, Barrouillet et al., 2000, Barrouillet and Lecas, 1999). These results have generally be interpreted in light of the Mental Model theory for conditional reasoning: this theory proposes that people construct mental models for the possibilities that make the sentence true under the material conditional truth table, and moreover only the situation in which the antecedent is true is initially represented, leaving the other two possibilities ("F-T"and "F-F") to be fleshed out at a later stage if necessary.

(13) p q
. . .

Young children however are unable to proceed and make explicit the other true possibilities, and therefore they find themselves with only one model available - the one in which both antecedent and consequent are true, making the conditional equivalent to a sentence with a conjunction. Unlike what happened in the experiments presented in this chapter, previous studies on children and adults have also consistently proven that the conjunctive interpretation is selected at a much lower rate by adults (typically between 10-30%), being replaced initially by a biconditional one (around 15 years of age) and later by the expected material conditional interpretation. In light of this, the Mental model theory does not seem sufficient to explain the results obtained in this study, as adults percentage of rejection for FF-controls was roughly the same as children's in Experiment 1, and it remained above 60% even in Experiment 2, after participants were explicitly invited to consider the conditional statement true in those situations.

There are other studies that have report a higher rate of conjunctive interpretations for adults: for example, Evans et al. (2003) and Oberauer and Wilhelm (2003) report a rate of around 40% for conjunctive interpretation. These studies, however, have used a different methodology: participants were asked to estimate the probability of a conditional statement. For example, after being shown a deck of cards with with a combination of letters and colours, they would be asked to establish the likeliness of (14). The authors found that participants would often reject "F-F" cases.

(14) If a card has an A, then it is red.

Paris (1973), who asked participants to judge the truth of a conditional in various situations, also observed that adults would display this behaviour 45% of the time. On the other hand, it has been observed that when participants are asked not to assess a conditional, but to *evoke* the possibilities that make true a conditional statement, they evoke "F-F" situations amongst these (Barrouillet and Lecas, 1998). These differences can be explained by an effect of the task: as Barrouillet et al. (2008) points out, it is probable that these two experimental methodologies require different forms of reasoning. The mental model theory is based on the idea that each mental model represents a possibility, while in turn entries in a truth table represent the truthiness or falseness of an assertion given a particular possibilities (Johnson-Laird and Byrne, 2002). Therefore, it can be assumed that possibilities and truth values require two different ways of reasoning: the first (reasoning about possibilities) requires people to reason about assertions they consider true, and try to find out what must, can or cannot occur in the world described by these assertions. When reasoning about truth values on the other hand, people are not required to construct a state of affairs, but instead they reason about one that has been given to them. Since the mental model initially constructed looks like the one in (13), when evaluating a conditional statement, said statement is initially deemed true only for those cases that match the initial model. The models in which the antecedent is false, on the other hand, are fleshed out at a later stage and as such they are not part of the core meaning of the sentence, but instead they are perceived just as possibilities that are compatible with the core meaning. Therefore, asking people to evoke all possibilities compatible with the conditional statement elicits a fully fleshed out model that includes all three possibilities under which the conditional is true, while asking people to assess the truthiness of a conditional, those implicit cases represented by the ellipsis in (13)

are not considered, leaving their truth value indeterminate. Sill, the rate of conjunctive interpretations obtained in the study presented in this section are significantly higher than what found by Oberauer and Wilhelm (2003) and Evans et al. (2003), so even this proposal does not provide a full explanation.

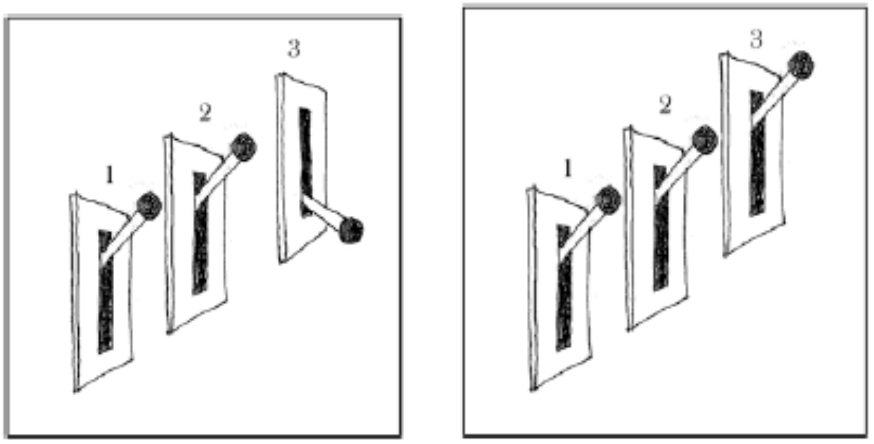
As introduced in section 5.2.3, the Mental logic model provides another explanation for conjunctive interpretations: following this approach, it is possible that that the participants of the two experiments presented in this section initially assigned an indeterminate truth value to the test and FF-control sentences and, when faced with a binary choice between right or wrong and no middle option, chose the latter option - in this case, the pattern of response would be due to pragmatic factors depending on the task used.

The consideration that the methodology used to elicit a judgment from participants might play a role in which interpretation is selected is not new: indeed, as Barrouillet and Lecas (1998) observed, the interpretation of conditional statements in experimental studies tends to vary widely, and this is thought to be caused by the variation in the type of task used, as well as by particular features within a task. For example, reviewing past experiments on the occurrence of AC and DA fallacies, Marcus and Rips (1979) noted that participants are more inclined to give material conditional answers when, in a syllogism evaluation task, they are provided with several choices (for example "always true/false", "sometimes true/false", "never true/false" as in Taplin and Staudenmayer (1973), Experiment 2) than when they are faced with a choice between true and false. It is possible that a similar effect holds for obtaining a conjunctive interpretation against a conditional or biconditional one: as discussed in the previous section, the Mental logic program (Braine and O'Brien, 1991) offers a similar explanation for conjunctive interpretations: the authors claim that, when a sentence is irrelevant to a picture, it is more appropriate to consider it false rather than true. Indeed, in the present study participants were explicitly asked whether what the puppet said was right or wrong, and no third option was given, so it is possible that participants had initially deemed the sentence irrelevant, in line with the defective truth table in 5.2 but, when prompted to evaluate the puppet's statement in terms of right or wrong, they were pushed towards a negative response.

In addition to this, another contextual factor that might have influenced the interpretation obtained by participants in our experiment is the use of a different antecedent.

Recall that for both test sentences and FF-controls the puppet utters a sentence in the form *if p, q*, and the story presents a situation *not-p, not-q*, that is to say, the puppet's guess is that *If Sheep is coming to the market today, Mr Pig will sell his balloons*, and not only is the antecedent not verified, but an alternative event takes place ("Today Bear couldn't come to the market. Instead, today Chicken came to the market"). It is possible that a true *not-p, not-q* situation (e.g. no animal comes to the market, and the balloons are unsold) would have obtained different results. However, one experiment conducted by (Gauffroy and Barrouillet, 2009) on the comprehension of conditionals incorporated a similar condition, which they call "other event" condition: for example, participants were asked to judge the causal conditional in (15) in a *not-p not-q* situation and one of two possible drawings was shown as a representation of the antecedent: either one in which all levers were up (no other event), or one in which lever 3 was down (other event).

(15) If the lever 2 is down, the rabbit case is open.



(a) Other event condition. (b) No other event condition.

FIGURE 5.12: Representation of the antecedent of (15) in Gauffroy and Barrouillet (2009).

While the authors had predicted participants could be more inclined towards negative answers by the other event, they found no effect of this manipulation. Nevertheless, it can be argued that the other event was more salient in our FF-controls than it was in Gauffroy and Barrouillet's experiment: while in their representation of the other event condition the false antecedent is represented (figure 5.13(a) does depict lever 2 down), in the FF-controls used for this experiment, the antecedent of the sentence was not

depicted at all but only mentioned in the story (i.e. "Sheep couldn't come today..."). This could have led participants to make the assumption that the occurrence of an event different to the one mentioned in the antecedent of the conditional constitutes a violation of the prediction made by the Puppet. Both explanations provided for the results (i.e. that adult participants had left the truth value of the test sentence undetermined and gave a wrong answer after being presented with a binary choice, and that the other event might have influenced the interpretation) could possibly be solved with a slightly amended version of this experiment. First of all, instead of asking participants whether the puppet was right or wrong, the question could be rephrased in order to make sure that participants have the option of mentioning if they consider the sentence irrelevant. Secondly, the slides for the target and FF-control items could be amended, so that the *not-p* event is represented visually, and not only mentioned to the participant to increase its salience.

(16) Amended FF-control:

Today Mr Pig has brought balloons to the market. Puppet, what do you think will happen today?

Puppet: If Sheep is coming to the market today, Mr Pig will sell his balloons.

Sheep decided to stay home today, and not go to the market. Instead, today Chicken came to the market! Chicken doesn't like balloons, so he leaves without buying the balloons.

Now the day is over, the market is closed and Mr Pig goes home.

So, Puppet had said: "If Sheep is coming to the market today, Mr Pig will sell his balloons", **can you say from what happened if Lucy was right or wrong?**.

If participants are indeed accessing the defective truth table, giving them the possibility to say that one cannot know whether the puppet's prediction was right or wrong in the context should lead them to chose this option. One last point to be addressed is that that there is a possibility that the omission of of the Indirect Object from the test sentence was implicitly filled in by the participant as a specific: the test sentence uttered by the Puppet in (16) above could lend itself to be interpreted as "If sheep is coming to the market today, then Mr Pig will sell the apples **to sheep**". Such interpretation would obviously have led to a rejection, as it is false in context. A further amendment might be

(a) Amended FF-control: Sheep decided to stay home today, and not go to the market.

(b) Outcome of amended FF control.

FIGURE 5.13: Possible amended version of FF-control.

necessary to weaken the expectation that the animal mentioned in the Puppet's sentence is the only possible buyer. This could be achieved for example by depicting a small set of characters at the beginning of each story, setting an expectation that any of them might be the potential buyer.

I mentioned above that another possible explanation can be attempted for the results of this study, on the basis of the behaviour observed in children for disjunction sentences. In sections 4.5.2 and 4.5.3 we discussed some literature showing that (17-a) and similar sentences are consistently been reported to receive a conjunctive-like interpretation by children (corresponding to interpretation (17-b)), while adults tend to interpret these kind of sentences under their exclusive interpretation, a reading that is thought to arise as a SI (interpretation (17-c)).

(17) a. John ate chicken or beef.

b. John ate chicken and beef.

c. John ate chicken or beef, but not both.

The conjunctive interpretation is thought to arise as a scalar inference, available only to children because they have a different set of competitors available compared to adults. Adults' set of possible alternatives include *p and q*, which is derived from the uttered sentence *p or q* by a process of lexical substitution. As children are unable of performing this process, the only alternatives available to them are those constructed by deleting lexical material from the original sentence. Said alternatives correspond to the two disjuncts, *p* and *q*. Therefore the strengthening process, by which the alternatives are

negated and joined with the original statement to yield the implicature, yields (18-a) and (18-b) as results for children and adults respectively.

(18) a. Adult implicature: p or q, and not p and q.
b. Child implicature: Both p and q.

Similarly to this, CP is thought to arise from the comparison between the conditional statement and an alternative obtained by deletion, at least under Horn's (2000) proposal - in the case of CP, the alternative statement corresponds to the consequent, *q*.

(19) a. If p then q.
b. q.

The one in (19) is not the only scale that has been proposed to derive CP as a scalar inference: for example another proposal is the one in (20), put forward by van der Auwera (1997).

(20) ...
if p, q, and if f, q and if s, q
if p, q and if r, q
↑ if p, q

Both (19) and (20) have been criticised, albeit for two different reasons: Horn's (2000) scale is considered insufficient to derive perfection (Von Fintel, 2001, Farr, 2011, Cariani and Rips, 2016). Instead, critics claim it derives a much weaker implicature, namely that *q* does not hold unconditionally, but it does not rule out that there could be other possibilities under which *q* holds. The scale proposed by van der Auwera (1997), on the other hand, has been criticised on grounds that one has to assume an infinitely large expression as the alternative, as for CP to arise from *if p then q* it is necessary to negate all possible antecedents. It seems unlikely to assume that interlocutors can derive and hold in their mind such infinite list of possible antecedents. It is however possible that this inference might be derived in another way.
Going back to the notion of conditional meaning, it should be noted that, under the material implication account, a conditional *if p then q* is equivalent to $\neg p \vee q$. Based on

this very simple assumption, we could have a theory of conditional perfection for adults and the corresponding conjunctive inference for children, along the lines of what Singh et al. (2016) have proposed for disjunctions. Applying Singh et al.'s (2016) approach to $\neg p \vee q$ does indeed derive the biconditional inference. Under this approach, the alternatives available to adults are {¬p, q, (¬p ∧ q)}. The negation of the latter, conjoined with the original sentence, gives rise to a biconditional inference, as it can be observed in table 5.3.

(21) Processing of $\neg p \vee q$ in adults:

a. EXH$(ALT_{adult})(\neg p \vee q)$

b. $ALT_{adult} = \{\neg p, q, \neg p \wedge q\}$

c. EXH$\{\text{ALT}_{adult}\}$ = (¬p ∨ q) ∧¬(¬p ∧ q).

p	q	EXH(¬p ∨ q)
1	1	1
1	0	0
0	1	0
0	0	1

TABLE 5.3: Truth conditions of (¬p ∨ q) ∧¬(¬p ∧ q).

Moving onto children, this proposal makes different predictions: as children should not be able to construct $\neg p \wedge q$ as an alternative, it predicts that children will derive a different inference, namely one that entails $\neg p \wedge q$. This can be observed below: the exhaustification mechanism, *exh*, applies recursively, with the first application being vacuous as no alternative is innocently excludable due to their symmetry.

(22) First layer of exhaustification:

a. EXH$(ALT_{child1})(\neg p \vee q)$

b. $ALT_{child1} = \{\neg p, q\}$

c. EXH$\{\text{ALT}_{child1}\}$ = (¬p ∨ q).

This changes in the second layer of exhaustification, with the alternatives now exhaustified and the symmetry broken.

(23) Second layer of exhaustification:

a. $\text{EXH}(\text{ALT}_{child2})((\text{EXH}(\text{ALT}_{child1})\neg \text{p} \vee \text{q})$

b. $ALT_{child2} = \{\text{EXH}(ALT_{child1})(s) : s \in ALT_{child1}\} = \text{EXH}(ALT_{child1})p, \text{EXH}(ALT_{child1})q,$
$= \{\neg p \wedge \neg q,\ q \wedge p\}$

c. $\text{EXH}\{\text{ALT}_{child}\} = (\neg \text{p} \vee \text{q}) \wedge \neg(\neg \text{p} \wedge \neg \text{q}) \wedge \neg(\text{p} \wedge \text{q})$ = entails $\neg \text{p} \wedge \text{q}$.

The second application of the exhaustification mechanism derives an inference under which the sentence is true when the first disjunct is true and the second is false, and false in every other circumstance, as it can be observed in table 5.4.

p	q	EXH(EXH($\neg$p $\vee$ q))
1	1	0
1	0	0
0	1	1
0	0	0

TABLE 5.4: Truth conditions of $(\neg \text{p} \vee \text{q}) \wedge \neg(\neg \text{p} \wedge \neg \text{q}) \wedge \neg(\text{p} \wedge \text{q})$.

As for the present study, this makes the prediction that children should have accepted the test sentences, while rejecting all other possibilities. While this prediction is not entirely borne out, it should indeed be noted that children's acceptance of these items was significantly higher than adults' in Experiment 1, while their acceptance for FF-control items was lower. Interestingly, the rates for this study are similar to the ones obtained by Paris (1973), who also reported a higher rate of rejection for situations in which both antecedent and consequent are false (92%) compared to "F-T" situations, equivalent to our target sentences.

Moreover, as Singh et al. (2016) observed, under this proposal is not strictly necessary to assume a difference between adults and children: any population that lacks a conjunctive alternative will obtain the same conjunctive reading. This consideration might be pertinent to the case of conditionals, since adults are often reported to obtain the conjunctive-like inference at significant rates (Paris, 1973, Oberauer and Wilhelm, 2003, Evans et al., 2003). On the other hand, results reported for conditionals are considerably more complicated than the ones observed for disjunction: while in the first case it is generally reported that adults derive the $\neg p \wedge q$ for a sentence $p \vee q$, conditionals have proven more unpredictable, with participants shifting from one interpretation to

the other. Another possible problem to take into account in relation to the possibility here discussed is that this approach rests on the equivalence between a conditional *if p then q* as *not-p or q*, coherently with its material implication meaning. However, the reconstrual of the conditional *if p then q* as *not-p or q* does involves generating a disjunct (*not-p*) which is not part of the original sentence: this is apparently in contrast with the notion, part of the alternative based approach, that children are unable to access the lexicon, and therefore can only consider material already available in the sentence. It is possible however that the context provided for the test sentence in the experiments detailed in this chapter made the disjunct *not-p* more available, as *p* was indeed false, and its falsehood critical for the test sentence. While the proposal here outlined represents a first step to address a possible extension of the alternative based approach to conditionals, the SI angle here introduced could provide an interesting path for further research.

5.4 Summary of chapter

This chapter has presented the results of an empirical study conducted to investigate adults' and children's interpretation of conditional sentences. Children's performance evidenced that, in line with previous experiments, they judge the conditional true if both antecedent and consequent are true, and judge it false in the majority of cases in other circumstances. Nevertheless, children gave significantly more positive answers compared to adults for test items in Experiment 1, signalling that they sometimes considered the conditional true when the antecedent is false and the consequent true. Adults on the other hand consistently rejected test items as well as FF-controls in Experiment 1, although Experiment 2 has shown that once they are primed to accept FF-controls, they are also more prone to accept test items, possibly because the priming effect carries over to other conditionals with a false antecedent. The results obtained in previous experiments have often been explained in light of the Mental model program, which claims that the model in which both antecedent and consequent are true (*p q*) is the first one to be constructed in the mind of the hearer. While older individuals are able to add further models to their representation, and derive the biconditional and material conditional interpretations (*not-p, q* and *not-p, not-q*), children are unable to perform this operation, and therefore they only accept the conditional statement as true when

both propositions are true. While promising, this explanation does not account for the high rate of conjunctive-like answers obtained for adults. A pragmatic explanation was also considered: it is indeed possible that giving participants the option of judging the conditional irrelevant would have prompted them to do so, resulting in lower rejection. Another feature of the context that might have influenced participants is that an alternative antecedent was featured in the short stories and pictures: it is possible that removing it or making it less salient might diminish the rate of rejection, for FF-controls at least. Moreover, it has been observed that sentences containing a disjunction are also often given a conjunctive interpretation by children (Paris, 1973, Braine and Rumain, 1981, Chierchia et al., 2004): in this case, the conjunctive interpretation is thought to arise as a SI due to the fact that children derive alternatives differently compared to adults, taking the two disjuncts as their set of alternatives (Singh et al., 2016). The same might be true for conditionals: indeed, we have shown how the perfected reading of a conditional can be derived in the same way as the exclusive reading of *or*. Under the same theory, children should derive a different implicature, namely one that entails $\neg p \wedge q$, which corresponds to our test sentences. Interestingly, in Experiment 1 children were more likely to accept these items compared to FF-controls. The predictions made by this model are not entirely borne out in this study, nevertheless they provide an interesting starting point. As a wide range of interpretation has been observed for conditionals in previous literature, it is not only possible that there exist different populations that are able to access different readings, but also that the readings obtained by the same subject might vary.

Chapter 6

Conclusions

In this thesis I have investigated the resolution of scope ambiguity and the occurrence of Conditional Perfection, by testing children's and adults' comprehension of scopally ambiguous questions and conditional statements, in order to investigate the development of grammatical and pragmatic abilities in children.

The first part of this thesis investigated scope ambiguity, a well-researched phenomenon in both adults and children: in Chapter 2, I reviewed several studies that have have generally evidenced that inverse scope interpretations are less frequent and more costly compared to the surface scope ones (Filik et al., 2004, Anderson, 2004, Tunstall, 1998, Kim, 2013), possibly because the sentence is initially analysed on its surface scope, and the inverse scope interpretation requires a reanalysis, as predicted for example by the Garden Path theory (Frazier and Fodor, 1978, Frazier and Rayner, 1982, Rayner et al., 1983, a.o). In line with this, early studies on children had found that they access inverse scope at a much lower rate compared to adults (Musolino, 1998), leading to the hypothesis that children might not yet posses the necessary grammatical abilities to obtain the aforementioned reading. However, subsequent studies attributed the difference found between the two groups to pragmatics: Gualmini (2004b) noted that children had possibly rejected some of the inverse scope sentences in Musolino's study due to their implausibility in the context, but they fared much better when sentences of the same structure where provided in a more supportive context. Later, this pragmatic felicity requirement was further refined to claim that, for the inverse scope interpretation to be accessible for children, said reading has to represent a good answer to the current

Question Under Discussion (Gualmini et al., 2008). The QUD approach gained much popularity as it was shown experimentally that children are indeed able to access inverse scope readings, even in situations in which adults would not (Gualmini, 2004b). According to the proponents of this theory, the presumed inability of children to access inverse scope observed in previous experiments was entirely due to their different management of the QUD, as they are adult-like in their grammatical abilities, and no other difference subsist between them and adults.

However there are still a number of open questions in relation to the effect of the QUD. For example, children's sensitivity to the effect of the QUD seems to falter in the case of modal verbs (Moscati, 2011, Moscati et al., 2016): for sentences containing these elements, children show a preference for the strongest reading in spite of thef effect of the QUD. Against this background, part of the research reported in this thesis has been aimed at investigating the effect of the QUD more directly: while previous experiments have tested children and adults interpretation of a declarative sentence (taken to be the answer to an explicit or implicit QUD), the experiments presented in Chapter 3 used a QUD which was also ambiguous. In Experiment 1, the ambiguous QUD (1-a) was associated to a corresponding declarative sentence, which presented a corresponding ambiguity (1-b).

(1) a. Did a boy eat each fruit?
b. A boy ate each fruit.

Participants' responses showed that they interpreted (1-b) on its inverse scope, which in turn suggests that they also interpreted the ambiguous question (1-a) on its inverse scope. Experiment 2 tested the interpretation of (1-a) directly, by asking participants to answer it: again, both children and adults appeared to have interpreted the question in its inverse scope in the majority of cases. The results obtained for the scope experiments in this study have shown that (1) both adults and children are able to access the inverse scope of the declarative (1-b) when presented as an answer to (1-a) - which in turn also suggests they have interpreted (1-a) on its inverse scope and (2) children and adults are able to access the inverse scope of (1-a) when asked to answer it directly. In relation to previous literature, there are a number of observations to be drawn. Firstly, our results confirm the finding, obtained by several previous studies, that children are able

to access inverse scope (Krämer, 2000, Gualmini et al., 2008, Gualmini, 2004b), contrary to the initial assumptions made by Musolino (1998) - in fact, in our results children were adult-like, at least when a supportive context is provided for the ambiguous sentence. Therefore, in line with much previous literature, the results obtained in this study go against the claim that a grammatical difference is responsible for the difference between children and adults observed in early studies, and they instead support the notion that pragmatic factors (and possibly other influences) must have been responsible. As the majority of the studies cited so far have been conducted on sentences containing a quantifier and the negation, our results also go to show that children can access inverse scope for doubly quantified sentences. In relation to the QUD approach, these results are compatible with the QAR as formulated by Gualmini et al. (2008). Recall that, according to this approach, children are adult-like in their ability to access inverse scope, and they only differ from adults in their management of the QUD: since it is not clear whether a question can be interpreted as the answer to the QUD, it is possible that having participants disambiguate an interrogative and not a declarative sentence removed the influence of the QUD, also removing any possible difference between the two groups. However, analogously to what is proposed for declarative sentences, one might wonder whether there is any pragmatic factor driving the interpretation of ambiguous questions. As the current formulation of the QAR does not provide an explanation as to what factors concur in the interpretation of an ambiguous question, I suggest that there is a possible extension of the QAR compatible with our data on questions, which also makes predictions about the role of pragmatics in the interpretation of ambiguous questions, based on the notion of superquestion. In this study, participants heard an ambiguous QUD which was presented to them as a question about a short story they had just heard, thus it can be assumed that the context played a role in disambiguating the QUD towards its inverse scope. Indeed, in Roberts's (1996) model of discourse, on which the QAR for scope assignment is based, discourse is organised around a series of questions, which are related to each other by entailment relations: therefore a QUD may be entailed by its superquestion(s). Question entailment is such that "one interrogative Q1 entails another Q2 iff every proposition that answers Q1 answers Q2 as well" (Groenendijk and Stokhof, 1984, p. 16). These entailment relations can be responsible for which interpretation of an ambiguous question gets selected: like for an ambiguous declarative sentence the effect of the QUD leads to the selection of the reading that is a good answer (i.e. entails either yes or no to it); in the case of questions, one can presume that the context makes

relevant one specific superquestion, and upon hearing an ambiguous question, this leads the hearer to select the reading that is entailed by said superquestion.

(2) **Extended QAR:** The interpretations available for a scopally ambiguous question are the one(s) entailed by its superquestion.

Therefore under this hypothesis, it can be assumed that the context provided for (2-b) made relevant a superquestion which entails the inverse scope of (2-b), pushing participants towards this interpretation. Testing this intuition experimentally would provide an interesting path of future research: one could test whether questions are affected by the context by manipulating the context in which they are presented to make relevant different subquestion. This would be achieved by associating it to two different contexts, each designed to make relevant a different superquestion.

For example, in the case of (1-a), we predict that that a context which makes salient the expectation that there should be one specific boy eating each fruit should make relevant the superquestion *is there a single boy who ate each fruit?*, which entails the surface scope interpretation of (2-a). In the same way, a context which gives rise to the superquestion *is it true that, for each fruit, a (possibly different) boy ate it?* should bias participants towards inverse scope. I leave investigating this prediction experimentally for further research.

The second part of this dissertation is concerned with Conditional Perfection, the phenomenon through which conditionals sometimes receive a biconditional interpretation. In chapter 4, I have presented some of the relevant literature, focusing in particular on the various theories on how this phenomenon comes to occur. One prominent account of CP is based on the effect of the QUD, and in particular on exhaustivity in question-answer pairs: according to Von Fintel (2001), the biconditional interpretation arises as a product of a (possibly implicit) question on the consequent of the conditional, like (3-a).

(3) a. How do I get five dollars?
b. If you mow the lawn, I'll give you five dollars.

While there has been some experimental research, conducted on adults on the effect of the QUD for CP (Farr, 2011, Cariani and Rips, 2016), this effect has not yet been

explored in language acquisition. While the effect of the QUD on the interpretation of conditionals provides a very promising path for future research, this work has taken a first step in determining how children interpret conditionals. There is in fact reason to doubt that the behaviour of children in relation to CP could be similar to the one observed in adults, as another prominent account on the nature of CP considers it a scalar implicature, and children are often reported to derive these kind of inferences at a much lower rate compared to adults, possibly because they struggle with the counterfactual reasoning required to derive this inference (Breheny, 2006, Proust, 2002, Riggs et al., 1998, among many others). Two experiments, presented in Chapter 5, were conducted on children's and adults' interpretation of conditionals: we expected to see a lower rate of endorsement for the biconditional inference in children if CP is indeed to be considered a scalar inference (cf. Tieu et al., 2017). Participants' judgements were collected as part of a TVTJ, in which they were asked to judge the veridicality of a conditional in various situations corresponding to different lines of the truth table. The two experiments however evidenced that both groups often give a conjunctive-like interpretation of conditional sentences, accepting it as true only when both propositions are true. Indeed, this kind of interpretation has often been reported in children (Paris, 1973, Barrouillet and Lecas, 1998, Barrouillet et al., 2000, 2008, Gauffroy and Barrouillet, 2009), but it is much less frequent in adults, with percentages ranging from 20% to 50% (Paris, 1973, Oberauer and Wilhelm, 2003, Evans et al., 2003, Barrouillet and Lecas, 1999, Fugard et al., 2011): a much lower rate than what observed in this study. In this sense, the results obtained for children are in line with the ones found in previous literature and as such, they are consistent with the amended version of the Mental model proposed by Barrouillet and Lecas (1998). Adults' results on the other hand are more surprising, as the percentage of conjunctive interpretations obtained in this study was higher than previously reported. Adults' interpretation of conditionals however is reported to display a wide variation in function of the task used (Barrouillet and Lecas, 1998, Marcus and Rips, 1979), so one possibility is that participants were influenced towards a conjunctive-like reading by some feature of the task used in our experiments. For example, participants are reported to differ when they are asked to give a binary judgment on the truthiness or falseness of a conditional compared to when they are given more options (Marcus and Rips, 1979): as the experiments conducted as part of this study did indeed ask participants to judge whether the conditional was true or false, without giving them any other option, this factor might be responsible for the high

proportions of conjunctive interpretations found. This is in line with the explanation given by proponents of the Mental logic approach for conjunctive interpretation (Braine and O'Brien, 1991): under this model such interpretations are a product of pragmatics, arising because the subject has initially interpreted the sentence as irrelevant, and from the consideration that when one is unsure but has to make a choice, it is best to consider the statement false rather than true. Therefore, one possible path for future research would be to present participants with a ternary choice, prompting them to judge the conditional as true, false or irrelevant (or asking whether the truthiness/falseness can be inferred from the situation). This change might yield a difference in children as well as adults as, following the Mental logic program, conjunctive interpretations are attributed to pragmatics for both groups.

Another point of interest is that conjunctive interpretations for conditionals were reported early in the literature, as were conjunctive interpretations of disjunction (Paris, 1973), but the latter have been the object of much recent experimental investigation (Braine and Rumain, 1981, Chierchia et al., 2004, Gualmini et al., 2000, Singh et al., 2016), while conditionals have received a lesser degree of attention in language acquisition. A recent theory, developed by Singh et al. (2016), claims that the conjunctive-like interpretation of disjunction is a SI. In Chapter 5, I have proposed an analogue theory for conditionals that could derive CP in adults, as well as offer a partial explanation for children's conjunctive inferences: the starting point of this is the consideration that, under the material implication account, a conditional (*if p then q*) is equivalent to a sentence with disjunction ($\neg p \vee q$), in the sense that the two sentences have the same truth conditions. I have shown how applying the model proposed by Singh et al. (2016) derives a biconditional inference in adults, and one that entails $\neg p \wedge q$ in children.

In summary, this thesis has been concerned with the resolution of scope ambiguity, and the interpretation of conditionals in adults and children. In relation to scope ambiguity resolution, I have proposed a possible extension of the QAR for scope assignment, formulating the proposal that entailment relations between questions are responsible for the disambiguation of an ambiguous question. While this possibility would need to be investigated experimentally, it is also possible that the QAR in its current formulation suffices to account for our results, by assuming that the QUD is the only factor of difference between children and adults, and that this factor was removed in our experiment. For what concerns the interpretation of conditionals, the results obtained

seem to confirm that pragmatic considerations are at least in part responsible for the interpretation obtained. In light of previous literature, in which the interpretation of conditionals is reported to "shift" from one set of truth conditions to another, it seems likely that pragmatic factors do play a role in which interpretation is ultimately selected. The question as to whether CP can be consider a scalar implicature remains open, and I have proposed a possible new derivation for it.

More in general, this thesis has been concerned with investigating children's pragmatic competence, in particular for what relates to the influence of the QUD: the scope experiments provide and advancement in comprehension on the effect of the QUD and scope ambiguity resolution, and the conditional experiments represent a first step towards the goal of investigating this phenomenon in light of the QUD approach.

Appendix A

Scope experiment items

Training items:

Item 1: The bear and the dog are having a picnic. They want to eat the muffin, but there is only one. In the end, the dog decides to eat a cookie, so the bear eats the muffin.

Experiment 1: Experimenter: Did the bear eat the muffin?
Puppet: The bear ate the muffin.
Experiment 2: Puppet: Did the bear eat the muffin?

Item 2: The monkey and the zebra want to do something fun. They want to read the book, but only one of them can. In the end, the monkey decides to listen to music. instead, so the zebra reads the book.

Experiment 1: Experimenter: Did the monkey read the book?
Puppet: The monkey read the book.
Experiment 2: Puppet: Did the monkey read the book

Item 3: The frog and the lion want to play with the toy car, but only one of them can play. In the end, the lion decides to play with a ball instead, so the frog plays with the toy car.

Experiment 1: Experimenter: Did the lion play with the toy car

Puppet: The lion played with the toy car.

Experiment 2: Puppet: Did the lion play with the toy car?

Item 4: The elephant and the cat go to the market to buy food. They want to buy milk, but there is only one bottle left. In the end, the cat decides to buy fruit juice instead, so the elephant buys the milk.

Experiment 1: Experimenter: Did the elephant buy milk?

Puppet: The elephant bought milk.

Experiment 2: Puppet: Did the elephant buy milk?

Test items:

Item 1: The boys and girls want to have a snack. They go looking for something to eat. In the kitchen there are three pieces of fruit: an apple, a banana and an orange. Both the boys and the girls want to eat something. The boys get to the kitchen first. One boy eats the apple, another boy eats the banana. The third boy eats the orange. There is no fruit left for the girls so the girls don't eat any fruit.

Experiment 1: Experimenter: Did a boy eat each fruit?

Puppet: A boy ate each fruit.

Experiment 2: Did a boy eat each fruit?

Item 2: The police and the fire-fighters are called because three neighbourhood cats need to be rescued from trees. To reach the cats, they will need ladders: there is a small ladder, a medium ladder and a big ladder. Both the policemen and the firefighters want to use the ladders. One policeman uses the small ladder, another policeman uses the medium ladder. The third policeman uses the big ladder. The firefighters don?t use any ladders.

Experiment 1: Experimenter: Did a policeman use each ladder?

Puppet: A policeman used each ladder.

Experiment 2: Puppet: Did a policeman use each ladder?

Item 3: The doctors and the nurses are looking for pencils to write with. They see some pencils on a desk. There is a green pencil, a blue pencil and a red pencil. Both the doctors and the nurses have a look at the pencils.?The doctors get first choice, so they go to the desk first. One doctor takes the green pencil, another doctor takes the blue pencil. The third doctor takes the red pencil. The nurses don?t take any pencils.

Experiment 1: Experimenter: Did a doctor take each pencil?
Puppet: A doctor took each pencil.
Experiment 2: Did a doctor take each pencil?

Item 4: The schoolboys and the teachers go on a trip to a farm. There are animals to feed: a rabbit, a pig and a sheep. Both the boys and the teachers want to feed the animals. The boys take some carrots to feed the animals. One boy feeds the rabbit, another boy feeds the pig. The third boy feeds the sheep. The teachers don't feed any animals.

Experiment 1: Experimenter: Did a boy feed each animal? Puppet: A boy fed each animal.
Experiment 2: Puppet: Did a boy fed each animal?

Control items:

Item 1 (Each): John is a fierce pirate. He is going to sink these boats: the brown boat, the blue boat and the red boat. John sinks the brown boat, and then he sinks the blue boat. Finally he also sinks the red boat.
Experiment 1: Experimenter: Did John sink each boat?
Puppet: John sank each boat.
Experiment 2: Puppet: Did John sink each boat?

Item 2 (Each):): Anna is a doctor. She is going to visit some patients: a boy, a man and an old woman. Anna visits the boy and gives him some medicine. Then she visits the man and gives him the medicine too. However she realises that there is no time to visit the last patient, the old woman, so she doesn't visit the old woman and she has to leave.

Experiment 1: Experimenter: Did Anna visit each patient?

Puppet: Anna visited each patient.

Experiment 2: Puppet: Did Anna visit each patient?

Item 3 (A): Lucy goes to a restaurant. She is waiting for someone to come and speak to her. There are some waiters in the room. Finally, a waiter comes and greets Lucy.

Experiment 1: Experimenter: Did a waiter greet Lucy?

Puppet: A waiter greeted Lucy.

Experiment 2: Puppet: Did a waiter greet Lucy?

Item 4 (A): Mary is playing her guitar in a park. She hopes someone will like her music. There are some boys listening to her, and they will give her flowers if she plays well. In the end, they don?t like her music, so they decide not to give Mary any flowers.

Experiment 1: Experimenter: Did a boy give flowers to Mary?

Puppet: A boy gave flowers to Mary.

Experiment 2: Puppet: Did a boy give flowers to Mary?

Filler items:

Item 1:: The giraffe and the cow are playing a game and they want to pick up the rock. In the end, the cow thinks the rock is too heavy for her, so the giraffe picks up the rock.

Experiment 1: Experimenter: Did the cow/giraffe pick up the rock?

Puppet: The cow/giraffe picked up the rock.

Experiment 2: Puppet: Did the cow/giraffe pick up the rock?

Item 2: The zebra and the lion are walking through a garden, looking for their friend, the penguin, who is hiding. In the end, the zebra sees the penguin. The lion doesn't see him because he is behind the tree.

Experiment 1: Experimenter: Did the lion/zebra see the penguin?

Puppet: The lion/zebra saw the penguin.

Experiment 2: Puppet: Did the lion/zebra see the penguin?

Item 3: The pig and the horse need to speak to the guard dog to ask him if they can play in the garden, but they are both afraid of the guard god, so they can?t decide who

should go. In the end, the pig is too scared so he hides behind the tree. The horse speaks to the dog instead.
Experiment 1: Experimenter: Did the pig/horse speak to the dog?
Puppet: The pig/horse spoke to the dog.
Experiment 2: Puppet: Did the pig/horse speak to the dog?

Item 4: The squirrel and the fox are playing in a park. They see their friend, the cat, and want to say hello to him. In the end, the fox says hello to the cat. The squirrel is busy eating his acorn so he doesn't say hello to the cat. Experiment 1: Experimenter: Did the squirrel/fox say hello to the cat?
Puppet: The squirrel/fox said hello to the cat.
Experiment 2: Puppet: Did the squirrel/fox say hello to the cat?

Appendix B

Conditional experiment items

Introduction: This is Mr Pig. He likes to sell things in the market. Everyday he brings something different to sell in the market. Mr Pig has some friends who come to the market to buy things from him. Elephant, Cat, Bear, Chicken, Mouse, Sheep and Giraffe all like to buy things from him when he is in the market.

Training items:

Item 1: Today Mr Pig has brought strawberries to the market. Puppet, what do you think will happen?

Puppet: Mr Pig will sell the strawberries to Cat.

Oh look, today Cat came to the market! Cat decides to buy the strawberries from Mr Pig. Now the day is over, the market is closed and Mr Pig goes home.

So, Puppet had said "Mr Pig will sell the strawberries to Cat", was Puppet right or wrong?

Item 2 (Experiment 1): Today Mr Pig has brought sweets to the market. Puppet, what do you think will happen today?

Puppet: Mr Pig will sell the sweets to bear. Oh look, today Cat came to the market. Cat decides to buy the sweets from Mr Pig. Now the day is over, the market is closed and Mr Pig goes home.

So, Puppet had said: "Mr Pig will sell the sweets to bear", did she say the right thing or the wrong thing?

Item 2 (Experiment 2): Today Mr Pig has brought sweets to the market. Puppet, what do you think will happen today?

Puppet: If Bear is coming to the market today, Mr Pig will sell his sweets.
Bear wanted to come to the market today, but in the end he could not come. Instead, today Sheep came to the market! Sheep doesn't like sweets, so she leaves without buying the sweets. Now the day is over, the market is closed and Mr Pig goes home.
So, Puppet had said "If Bear is coming to the market today, Mr Pig will sell his sweets", was Puppet right or wrong?

Test items:

Item 1: Today Mr Pig has brought cheese to the market. Puppet, what do you think will happen today?
Puppet: If elephant is coming to the market today, Mr Pig will sell his cheese.
Oh look, today cat came to the market! Cat decides to buy the cheese. Now the day is over, the market is closed and Mr Pig goes home.
So, Puppet had said "If If elephant is coming to the market today, Mr Pig will sell his cheese", was Puppet right or wrong?

Item 2: Today, Mr Pig has brought bananas to the market. Puppet, what do you think will happen today?
Puppet: If chicken is coming to the market today, Mr Pig will sell his bananas.
Oh look, today bear came to the market! Bear decides to buy the bananas. Now the day is over, the market is closed and Mr Pig goes home.
So, Puppet had said "If If chicken is coming to the market today, Mr Pig will sell his bananas", was Puppet right or wrong?

Item 3: Today, Mr Pig has brought cookies to the market. Puppet, what do you think will happen today?
Puppet: If Sheep is coming to the market today, Mr Pig will sell his cookies.
Oh look, today Mouse came to the market! Mouse decides to buy the cookies. Now the day is over, the market is closed and Mr Pig goes home.
So, Puppet had said "If Sheep is coming to the market today, Mr Pig will sell his cookies", was Puppet right or wrong?

Item 4: Today, Mr Pig has brought carrots to the market. Puppet, what do you think will happen today? Puppet: If Bear is coming to the market today, Mr Pig will sell his carrots.
Oh look, today Elephant came to the market! Elephant decides to buy the carrots. Now

the day is over, the market is closed and Mr Pig goes home.

So, Puppet had said: "If Bear is coming to the market today, Mr Pig will sell his carrots", was Puppet right or wrong?

Control items:

Item 1: Today, Mr Pig has brought eggs to the market. Puppet, what do you think will happen today?

Puppet: If Giraffe is coming to the market today, Mr Pig will sell his eggs. Oh look, today Giraffe came to the market! Giraffe doesn't like eggs, so she leaves without buying the eggs. Now the day is over, the market is closed and Mr Pig goes home.

So, Puppet had said "If If Giraffe is coming to the market today, Mr Pig will sell his eggs", was right or wrong?

Item 2: Today, Mr Pig has brought apples to the market. Puppet, what do you think will happen today?

Puppet: If Cat is coming to the market today, Mr Pig will sell his apples.

Oh look, today Cat came to the market! Cat doesn't like apples, so he leaves without buying the apples. Now the day is over, the market is closed and Mr Pig goes home.

So, Puppet had said "If Cat is coming to the market today, Mr Pig will sell his apples", was Puppet right or wrong?

Item 3: Today, Mr Pig has brought flowers to the market. Puppet, what do you think will happen today?

Puppet: If Mouse is coming to the market today, Mr Pig will sell his flowers. Oh look, today Mouse came to the market! Mouse decides to buy the flowers. Now the day is over, the market is closed and Mr Pig goes home.

So, Puppet had said "If Mouse is coming to the market today, Mr Pig will sell his flowers", was Puppet right or wrong?

Item 4: Today, Mr Pig has brought oranges to the market. Puppet, what do you think will happen today?

Puppet: If Giraffe is coming to the market today, Mr Pig will sell his oranges.

Oh look, today Giraffe came to the market! Giraffe decides to buy the oranges. Now the day is over, the market is closed and Mr Pig goes home.

So, Puppet had said "If Giraffe is coming to the market today, Mr Pig will sell his oranges", was Puppet right or wrong?

Item 5 (FF-control): Today, Mr Pig has brought milk to the market. Puppet, what do you think will happen today?
Puppet: If Elephant is coming to the market today, Mr Pig will sell his milk. Elephant wanted to come to the market today, but in the end he could not come. Instead, today Bear came to the market! Bear doesn't like milk, so he leaves without buying the milk. Now the day is over, the market is closed and Mr Pig goes home.
So, Puppet had said "If Elephant is coming to the market today, Mr Pig will sell his milk", was Puppet right or wrong?

Item 6 (FF-control): Today, Mr Pig has brought balloons to the market. Puppet, what do you think will happen today?
Puppet: If Sheep is coming to the market today, Mr Pig will sell his balloons.
Sheep wanted to come to the market today, but in the end he could not come. Instead, today Chicken came to the market! Chicken doesn't like balloons, so he leaves without buying the balloons. Now the day is over, the market is closed and Mr Pig goes home.
So, Puppet had said "If Sheep is coming to the market today, Mr Pig will sell his balloons", was Puppet right or wrong?

Item 7: Today, Mr Pig has brought orange juice to the market. Puppet, what do you think will happen today?
Puppet: Bear is coming to the market today, and Mr Pig will sell his orange juice.
Bear wanted to come to the market today, but in the end he could not come. Instead, today Sheep came to the market! Sheep doesn't like orange juice, so she leaves without buying the orange juice. Now the day is over, the market is closed and Mr Pig goes home.
So, Puppet had said "Bear is coming to the market today, and Mr Pig will sell his orange juice", was Puppet right or wrong?

Item 8: Today, Mr Pig has brought cake to the market. Puppet, what do you think will happen today?
Puppet: Chicken is coming to the market today, and Mr Pig will sell his pencils.
Oh look, today Chicken came to the market! Chicken doesn't like cake, so he leaves without buying the cake. Now the day is over, the market is closed and Mr Pig goes home.

So, Puppet had said "Chicken is coming to the market today, and Mr Pig will sell his pencils", was Puppet right or wrong?

Filler items:

Item 1: Today, Mr Pig has brought jam to the market. Puppet, what do you think will happen today?

Puppet: Mr Pig will sell the jam to the Giraffe / Sheep. Oh look, today Giraffe came to the market! Giraffe buys the jam. Now the day is over, the market is closed and Mr Pig goes home.

So, Puppet had said "Mr Pig will sell the jam to the Giraffe / Sheep", was Puppet right or wrong?

Item 2: Today, Mr Pig has brought tomatoes to the market. Puppet, what do you think will happen today?

Puppet: Mr Pig will sell the tomatoes to the Bear / Elephant.

Oh look, today Bear came to the market! Bear buys the tomatoes. Now the day is over, the market is closed and Mr Pig goes home.

So, Puppet had said "Mr Pig will sell the tomatoes to the Bear / Elephant", was Puppet right or wrong?

Item 3: Today, Mr Pig has brought hats to the market. Puppet, what do you think will happen today?

Puppet: Mr Pig will sell the hats to Cat / Mouse.

Oh look, today Cat came to the market! Cat buys the hats. Now the day is over, the market is closed and Mr Pig goes home.

So, Puppet had said "Mr Pig will sell the hats to the Cat / Mouse", was Puppet right or wrong?

Item 4: Today, Mr Pig has brought pencils to the market. Puppet, what do you think will happen today?

Puppet: Mr Pig will sell the pencils to Sheep / Elephant. Oh look, today Sheep came to the market! Sheep buys the hats. Now the day is over, the market is closed and Mr Pig goes home.

So, Puppet had said "Mr Pig will sell the pencils to Sheep / Elephant", was Puppet right or wrong?

Bibliography

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

David Lewis. Scorekeeping in a language game. *Journal of philosophical logic*, 8(1): 339–359, 1979.

Robert C Stalnaker. *Assertion*. Wiley Online Library, 1978.

Julien Musolino. *Universal Grammar and the Acquisition Of Semantic Knowledge: An Experimental Investigation into the Acquisition of Quantifier-Negation Interaction in English*. PhD thesis, University of Maryland, 1998.

Julien Musolino, Stephen Crain, and Rosalind Thornton. Navigating negative quantificational space. *Linguistics*, 38(1):1–32, 2000.

Jeffrey Lidz and Julien Musolino. Children's command of quantification. *Cognition*, 84 (2):113–154, 2002.

Andrea Gualmini. Some knowledge children don't lack. *Linguistics*, pages 957–982, 2004a.

Irene Krämer. *Interpreting indefinites: An experimental study of children's language comprehension*. PhD thesis, Radboud University Nijmegen Nijmegen, 2000.

Andrea Gualmini, Sarah Hulsey, Valentine Hacquard, and Danny Fox. The question–answer requirement for scope assignment. *Natural language semantics*, 16(3):205, 2008.

Arjen Zondervan, Luisa Meroni, and Andrea Gualmini. Experiments on the role of the question under discussion for ambiguity resolution and implicature computation in adults. In *Semantics and linguistic theory*, volume 18, pages 765–777, 2008.

Arjen Zondervan. The role of qud and focus on the scalar implicature of most. *Experimental Pragmatics/semantics*, 175:221, 2011.

Kai Von Fintel. Conditional strengthening, 2001.

M Farr. Focus influences the presence of conditional perfection: Experimental evidence. In *Proceedings of Sinn & Bedeutung*, volume 15, pages 225–239. Citeseer, 2011.

Laurence R Horn. From if to iff: Conditional perfection as pragmatic strengthening. *Journal of pragmatics*, 32(3):289–326, 2000.

Jay David Atlas and Stephen C Levinson. It-clefts, informativeness and logical form: radical pragmatics (revised standard version). In *Radical pragmatics*, pages 1–62. Academic Press, 1981.

Johan van der Auwera. Conditional perfection. *On Conditionals Again*, 143:169, 1997.

Scott G Paris. Comprehension of language connectives and propositional logical relationships. *Journal of experimental child psychology*, 16(2):278–291, 1973.

Pierre Barrouillet and Jean-François Lecas. How can mental models theory account for content effects in conditional reasoning? a developmental perspective. *Cognition*, 67 (3):209–253, 1998.

Pierre Barrouillet, Nelly Grosset, and Jean-François Lecas. Conditional reasoning by mental models: Chronometric and developmental evidence. *Cognition*, 75(3):237–266, 2000.

Pierre Barrouillet, Caroline Gauffroy, and Jean-Francois Lecas. Mental models and the suppositional account of conditionals. *Psychological Review*, 115(3):760, 2008.

Martin DS Braine and Barbara Rumain. Development of comprehension of "or": Evidence for a sequence of competencies. *Journal of experimental child psychology*, 31 (1):46–70, 1981.

Andrea Gualmini, Stephen Crain, Luisa Meroni, Gennaro Chierchia, and Maria Teresa Guasti. At the semantics/pragmatics interface in child language. In *Semantics and Linguistic Theory*, volume 11, pages 231–247, 2001.

Gennaro Chierchia, Maria Teresa Guasti, Andrea Gualmini, Luisa Meroni, Stephen Crain, and Francesca Foppolo. Semantic and pragmatic competence in children's and adults' comprehension of or. In *Experimental pragmatics*, pages 283–300. Springer, 2004.

Ira A Noveck. When children are more logical than adults: Experimental investigations of scalar implicature. *Cognition*, 78(2):165–188, 2001.

Anna Papafragou and Julien Musolino. Scalar implicatures: experiments at the semantics–pragmatics interface. *Cognition*, 86(3):253–282, 2003.

Teresa Guasti, Gennaro Chierchia, Stephen Crain, Francesca Foppolo, Andrea Gualmini, and Luisa Meroni. Why children and adults sometimes (but not always) compute implicatures. *Language and cognitive processes*, 20(5):667–696, 2005.

Francesca Foppolo, Maria Teresa Guasti, and Gennaro Chierchia. Scalar implicatures in child language: Give children a chance. *Language learning and development*, 8(4): 365–394, 2012.

Caroline Gauffroy and Pierre Barrouillet. Heuristic and analytic processes in mental models for conditionals: An integrative developmental theory. *Developmental Review*, 29(4):249–282, 2009.

Daniel Büring. On d-trees, beans, and b-accents. *Linguistics and philosophy*, 26(5): 511–545, 2003.

Charles L Hamblin. Questions in montague english. *Foundations of language*, 10(1): 41–53, 1973.

Charles L Hamblin. Questions. *The Australasian journal of philosophy*, 36:159–168, 1958.

Lauri Karttunen. Syntax and semantics of questions. *Linguistics and philosophy*, 1(1): 3–44, 1977.

Jaakko Hintikka. *The Semantics of Questions and the Questions of Semantics: Case Studies in the Interrelations of Logic, Semantics, and Syntax, Jaakko Hintikka*, volume 28. North-Holland Publishing Company, 1976.

Jeroen Groenendijk and Martin Stokhof. *On the semantics of questions and the pragmatics of answers*. Foris, 1984.

Jeroen Groenendijk and Martin Stokhof. Semantic analysis of wh-complements. *Linguistics and Philosophy*, 5(2):175–233, 1982.

Irene Heim. Interrogative semantics and karttunen's semantics for know. In *Proceedings of IATL*, volume 1, pages 128–144, 1994.

Stephen Robert Berman. *On the semantics and logical form of {\it wh\/}-clauses*. PhD thesis, 1991.

Nathan Klinedinst and Daniel Rothschild. Exhaustivity in questions with non-factives. *Semantics and Pragmatics*, 4(2):1–23, 2011.

Wataru Uegaki et al. *Interpreting questions under attitudes*. PhD thesis, Massachusetts Institute of Technology, 2015.

Andreea Cristina Nicolae. *Any questions? Polarity as a window into the structure of questions*. PhD thesis, 2013.

Benjamin Ross George. *Question embedding and the semantics of answers*. University of California, Los Angeles, 2011.

Mats Rooth. *Association with focus*. PhD thesis, GLSA, Dept. of Linguistics, University of Massachusetts, Amherst, 1985.

Mats Rooth. A theory of focus interpretation. *Natural language semantics*, 1(1):75–116, 1992.

L Carlson. Dialogue games: An approach to discourse analysis. *Reidel*, 1983.

Ray S Jackendoff. Semantic interpretation in generative grammar. 1972.

Mats Rooth. Focus. the handbook of contemporary semantic theory, ed. by shalom lappin, 271-97, 1996.

Robert Van Rooy. Questioning to resolve decision problems. *Linguistics and Philosophy*, 26(6):727–763, 2003.

Arndt Riester. *Joint information structure and discourse structure analysis in an underspecified drt framework*. Questions in Discourse, CRiSPI Series. Brill, Forthcoming.

Arndt Riester, Lisa Brunetti, and Kordula De Kuthy. Annotation guidelines for questions under discussion and information structure. In *Information structure in lesser-described languages: Studies in prosody and syntax, Studies in Language Companion Series*. John Benjamins Publishing, 2017.

Edgar Onea. *Potential Questions at the Semantics-Pragmatics Interface*. Brill, Leiden, 2016.

Leah Velleman and David Beaver. Question-based models of information structure. In *The Oxford Handbook of Information Structure*. 2015.

Ye Tian and Richard Breheny. Dynamic pragmatic view of negation processing. In *Negation and polarity: Experimental perspectives*, pages 21–43. Springer, 2016.

Ye Tian, Heather Ferguson, and Richard Breheny. Processing negation without context–why and when we represent the positive argument. *Language, Cognition and Neuroscience*, 31(5):683–698, 2016.

Richard Breheny. *elevance: Pragmatics and Interpretation*, chapter Language processing, relevance and questions. CUP, to appear.

Dan Sperber and Deirdre Wilson. *Relevance: Communication and Cognition*. Harvard University Press, 1986.

Deirdre Wilson and Dan Sperber. Relevance theory. In *Handbook of pragmatics*. Blackwell, 2002.

Arjen Zondervan. Experiments on qud and focus as a contextual constraint on scalar implicature calculation. *From experiment to theory, presuppositions, negation and scalar implicatures. Basingstoke: Palgrave*, 2009.

Ye Tian, Richard Breheny, and Heather J Ferguson. Why we simulate negated information: A dynamic pragmatic account. *The Quarterly Journal of Experimental Psychology*, 63(12):2305–2312, 2010.

Charles Clifton Jr and Lyn Frazier. Context effects in discourse: The question under discussion. *Discourse Processes*, 55(2):105–112, 2018.

Ekaterina Jasinskaja. *Pragmatics and prosody of implicit discourse relations: The case of restatement*. PhD thesis, 2009.

Andrea Gualmini. *The ups and downs of child language*. Outstanding Dissertations in Linguistics series. Routledge, 2004b.

Peter C Wason. The contexts of plausible denial. *Journal of verbal learning and verbal behavior*, 4(1):7–11, 1965.

Jill G De Villiers and Helen B Tager Flusberg. Some facts one simply cannot deny. *Journal of Child Language*, 2(2):279–286, 1975.

Vincenzo Moscati. Discourse under control in ambiguous sentences. *Experimental pragmatics and semantics*, pages 63–78, 2011.

Vincenzo Moscati, Jacopo Romoli, Tommaso Federico Demarie, and Stephen Crain. Born in the usa: a comparison of modals and nominal quantifiers in child language. *Natural Language Semantics*, 24(1):79–115, 2016.

Michael L Geis and Arnold M Zwicky. On invited inferences. *Linguistic inquiry*, 2(4): 561–566, 1971.

Fabrizio Cariani and Lance J Rips. Experimenting with (conditional) perfection. 2016.

Gennaro Chierchia, Stephen Crain, Maria Teresa Guasti, Andrea Gualmini, Luisa Meroni, et al. The acquisition of disjunction: Evidence for a grammatical view of scalar implicatures. In *Proceedings of the 25th Boston University conference on language development*, pages 157–168, 2001.

Ira A Noveck and Andres Posada. Characterizing the time course of an implicature: An evoked potentials study. *Brain and language*, 85(2):203–210, 2003.

Lewis Bott and Ira A Noveck. Some utterances are underinformative: The onset and time course of scalar inferences. *Journal of memory and language*, 51(3):437–457, 2004.

H Paul Grice. *Studies in the Way of Words*. Harvard University Press, 1989.

Bart Geurts and Nausicaa Pouscoulous. Embedded implicatures?!? *Semantics and pragmatics*, 2:4–1, 2009.

Gennaro Chierchia. Broaden your views: Implicatures of domain widening and the 'logicality' of language. *Linguistic inquiry*, 37(4):535–590, 2006.

Danny Fox. Free choice and the theory of scalar implicatures. In *Presupposition and implicature in compositional semantics*, pages 71–120. Springer, 2007.

Emmanuel Chemla and Benjamin Spector. Experimental evidence for embedded scalar implicatures. *Journal of semantics*, 28(3):359–400, 2011.

Andrea Gualmini, Luisa Meroni, and Stephen Crain. The inclusion of disjunction in child grammar: Evidence from modal verbs. 01 2000.

Klaus Oberauer and Oliver Wilhelm. The meaning (s) of conditionals: Conditional probabilities, mental models, and personal utilities. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 29(4):680, 2003.

Jonathan St BT Evans, Simon J Handley, and David E Over. Conditionals and conditional probability. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 29(2):321, 2003.

Sandra L Marcus and Lance J Rips. Conditional reasoning. *Journal of Verbal Learning and Verbal Behavior*, 18(2):199–223, 1979.

Raj Singh, Ken Wexler, Andrea Astle-Rahim, Deepthi Kamawar, and Danny Fox. Children interpret disjunction as conjunction: Consequences for theories of implicature and child development. *Natural Language Semantics*, 24(4):305–352, 2016.

Robert May. The grammar of quantification. *Unpublished doctoral dissertation, MIT*, 1977.

Danny Fox. *Economy and semantic interpretation*, volume 35. MIT press, 2000.

Tanya Miriam Reinhart. *The syntactic domain of anaphora*. PhD thesis, Massachusetts Institute of Technology, 1976.

Wilma Bucci. The interpretation of universal affirmative propositions. *Cognition*, 6(1): 55–77, 1978.

Patricia J Brooks and Martin DS Braine. What do children know about the universal quantifiers all and each? *Cognition*, 60(3):235–268, 1996.

William Churchill Houston Philip. *Event quantification in the acquisition of universal quantification.* PhD thesis, University of Massachusetts, 1994.

Stephen Crain and Rosalind Thornton. The truth value judgment task: Fundamentals of design. *University of Maryland working papers in linguistics*, 6:61–70, 1998.

LJ Russell. Human knowledge—its scope and limits. *Philosophy*, 24(90):253–260, 1949.

Donald Davidson. Thought and talk. *Mind and language*, 1975:7–23, 1975.

Joshua Viau, Jeffrey Lidz, and Julien Musolino. Priming of abstract logical representations in 4-year-olds. *Language Acquisition*, 17(1-2):26–50, 2010.

Julien Musolino and Jeffrey Lidz. Why children aren't universally successful with quantification. *Linguistics*, 44(4):817–852, 2006.

Laurence Horn. A natural history of negation. 1989.

Vincenzo Moscati and Andrea Gualmini. More facts that isomorphism cannot explain. In *Proceedings of SALT*, volume 17, pages 202–209, 2008.

Andrea Gualmini and Vincenzo Moscati. The early steps of modal and negation interactions: Evidence from child italian. 2009.

Kathryn Gillen. *The comprehension of doubly quantified sentences.* PhD thesis, Durham University, 1991.

Ruth Filik, Kevin B Paterson, and Simon P Liversedge. Processing doubly quantified sentences: Evidence from eye movements. *Psychonomic Bulletin & Review*, 11 (5):953–959, 2004.

Catherine Anderson. *The structure and real-time comprehension of quantifier scope ambiguity.* PhD thesis, Northwestern University, 2004.

Georgette Ioup. Some universals for quantifier scope. *Syntax and semantics*, 4:37–58, 1975.

Kurt A VanLehn. Determining the scope of english quantifiers. Master's thesis, MIT, 1978.

Janet Dean Fodor. The mental representation of quantifiers. In *Processes, beliefs, and questions*, pages 129–164. Springer, 1982.

Gerry Altmann and Mark Steedman. Interaction with context during human sentence processing. *Cognition*, 30(3):191–238, 1988.

Howard S Kurtzman and Maryellen C MacDonald. Resolution of quantifier scope ambiguities. *Cognition*, 48(3):243–279, 1993.

Lyn Frazier and Janet Dean Fodor. The sausage machine: A new two-stage parsing model. *Cognition*, 6(4):291–325, 1978.

Lyn Frazier and Keith Rayner. Making and correcting errors during sentence comprehension: Eye movements in the analysis of structurally ambiguous sentences. *Cognitive psychology*, 14(2):178–210, 1982.

Keith Rayner, Marcia Carlson, and Lyn Frazier. The interaction of syntax and semantics during sentence processing: Eye movements in the analysis of semantically biased sentences. *Journal of verbal learning and verbal behavior*, 22(3):358–374, 1983.

Stephen Crain and Mark Steedman. On not being led up the garden path: The use of context by the psychological parser. *Natural language parsing*, pages 320–358, 1985.

Lyn Frazier. *On comprehending sentences: Syntactic parsing strategies.* PhD thesis, University of Connecticut, 1978.

Fernanda Ferreira and Charles Clifton. The independence of syntactic processing. *Journal of memory and language*, 25(3):348–368, 1986.

John C Trueswell, Irina Sekerina, Nicole M Hill, and Marian L Logrip. The kindergarten-path effect: Studying on-line sentence processing in young children. *Cognition*, 73(2): 89–134, 1999.

Michael K Tanenhaus, Michael J Spivey-Knowlton, Kathleen M Eberhard, and Julie C Sedivy. Integration of visual and linguistic information in spoken language comprehension. *Science*, 268(5217):1632, 1995.

Philip N Johnson-Laird and Joanna Tagart. How implication is understood. *The American Journal of Psychology*, 82(3):367–373, 1969.

Harry C Bunt. Mass terms and model-theoretic semantics. 1985.

Anthony S Kroch. *The semantics of scope in English.* PhD thesis, Massachusetts Institute of Technology, 1974.

Susanne Tunstall. *The interpretation of quantifiers: semantics and processing.* PhD thesis, University of Massachusetts-Amherst, 1998.

Jung-Hee Kim. *The intransigence of inverse scope: the effect of discourse, priming and population differences.* PhD thesis, UNIVERSITY OF HAWAI 'I, 2013.

Francesco Pupa. *Ambiguous articles: An essay on the theory of* descriptions.* PhD thesis, City University of New York, 2008.

H Paul Grice. Logic and conversation. *1975*, pages 41–58, 1975.

Maribel Romero. Two approaches to biased yes/no questions. In *Proceedings of the 24th West Coast Conference on Formal Linguistics*, pages 352–360, 2005.

Gregory Scontras, Cheng-Yu Edwin Tsai, Kenneth Mai, and Maria Polinsky. Chinese scope: An experimental investigation. In *Proceedings of Sinn und Bedeutung*, volume 18, pages 396–414, 2014.

Sharon Unsworth, Andrea Gualmini, and Christina Helder. Children's interpretation of indefinites in sentences containing negation: A reassessment of the cross-linguistic picture. *Language Acquisition*, 15(4):315–328, 2008.

Sarah Hulsey, Valentine Hacquard, Danny Fox, and Andrea Gualmini. The question-answer requirement and scope assignment. *MIT working papers in Linguistics*, 48: 71–90, 2004.

Chierchia, Fox, and Spector. *Scalar implicature as a grammatical phenomenon*, volume 3, pages 2297–2331. Mouton de Gruyter, edited by maienborn, von heusinger & portner edition, 2012.

Miguel López Astorga. ¿ podemos evitar la perfección del condicional enfocando el antecedente o son necesarios antecedentes alternativos? *Revista signos*, 47(85):267–292, 2014a.

Miguel López Astorga. La interpretación de enunciados condicionales como bicondicionales y su problemática en la tarea de selección. *Diversitas: Perspectivas en Psicología*, 10(1), 2014b.

Pierre Barrouillet and Jean-Francois Lecas. Mental models in conditional reasoning and working memory. *Thinking & Reasoning*, 5(4):289–302, 1999.

Robert C Stalnaker. A theory of conditionals. In *Ifs*, pages 41–55. Springer, 1968.

David Lewis. *Counterfactuals*. Harvard University Press, 1973.

Angelika Kratzer. Conditionals. 1986.

Peter Suber. Paradoxes of material implication. *URL http://www. earlham. edu/˜ peters/courses/log/matimp. htm*, 1997.

Jonathan St BT Evans and DE Over. *If: Supposition, pragmatics, and dual processes*. Oxford Cognitive Science (Pape, 2004.

Frank Jackson. *Conditionals*. Blackwell, 1987.

Peter C. Wason. Reasoning. In *New Horizons in Psychology*, pages 135–151. Penguin Books, 1966.

H Paul Grice. William james lectures. *Harvard University*, pages 0041–58, 1967.

Frank Jackson. On assertion and indicative conditionals. *The Philosophical Review*, 88 (4):565–589, 1979.

Adam Rieger. A simple theory of conditionals. *Analysis*, 66(291):233–240, 2006.

Clarence Irving Lewis. A survey of symbolic logic. 1918.

Kai von Fintel. Conditionals. 2011.

Angelika Kratzer. The notional category of modality. *Words, worlds, and contexts*, pages 38–74, 1981.

Angelika Kratzer. Conditional necessity and possibility. In *Semantics from different points of view*, pages 117–147. Springer, 1979.

Dorothy Edgington. Indicative conditionals. In Edward N. Zalta, editor, *The Stanford Encyclopedia of Philosophy*. Metaphysics Research Lab, Stanford University, winter 2014 edition, 2014.

Kai Von Fintel and Sabine Iatridou. If and when if-clauses can restrict quantifiers. In *URL http://web. mit. edu/fintel/www/lpw. mich. pdf. Unpublished ms, MIT, contribution to the Workshop in Linguistics & Philosophy at the University of Michigan*, 2002.

Dorothy Edgington. Ramsey's legacies on conditionals and truth. In Hallvard Lillehammer and D. H. Mellor, editors, *Ramsey's Legacy*. Oxford University Press, 2005.

Oswald Ducrot. Présupposés et sous-entendus. *Langue française*, (4):30–43, 1969.

J. L. Austin. *Philosophical Papers*. Oxford University Press, 1961.

Gerald W Lilje. Uninvited inferences. *Linguistic Inquiry*, pages 540–542, 1972.

Steven E Boër and William G Lycan. Invited inferences and other unwelcome guests. *Research on Language & Social Interaction*, 6(1-4):483–505, 1973.

Ingrid van Canegem-Ardijns and William Van Belle. Conditionals and types of conditional perfection. *Journal of Pragmatics*, 40(2):349–376, 2008.

Denise D Cummins, Todd Lubart, Olaf Alksnis, and Robert Rist. Conditional reasoning and causation. *Memory & cognition*, 19(3):274–282, 1991.

Muffy EA Siegel. Biscuit conditionals: Quantification over potential literal acts. *Linguistics and Philosophy*, 29(2):167–203, 2006.

Barbara Dancygier and Eve Sweetser. *Mental spaces in grammar: Conditional constructions*, volume 108. Cambridge University Press, 2005.

Michael Franke. *The pragmatics of biscuit conditionals*. Inst. for Logic, Language and Computation, 2007.

Stephen C Levinson. *Presumptive meanings: The theory of generalized conversational implicature*. MIT Press, 2000.

John Stuart Mill. *An Examination of Sir William Hamilton's Philosophy*. Longmans, Green, and Company, 1889.

LTF Gamut. *Logic, Language, and Meaning, volume 1: Introduction to Logic*, volume 1. University of Chicago Press, 1991.

Laurence Horn. *On the Semantic Properties of Logical Operators in English*. PhD thesis, University of California, 1972.

Yo Matsumoto. The conversational condition on horn scales. *Linguistics and philosophy*, 18(1):21–60, 1995.

Richard Breheny, Nathan Klinedinst, Jacopo Romoli, and Yasutada Sudo. The symmetry problem: current theories and prospects. *Natural Language Semantics*, Nov 2017. ISSN 1572-865X. doi: 10.1007/s11050-017-9141-z. URL `https://doi.org/10.1007/s11050-017-9141-z`.

Roni Katzir. Structurally-defined alternatives. *Linguistics and Philosophy*, 30(6):669–690, 2007.

Danny Fox and Roni Katzir. On the characterization of alternatives. *Natural Language Semantics*, 19(1):87–107, 2011.

Martin D Braine and David P O'Brien. A theory of if: A lexical entry, reasoning program, and pragmatic principles. *Psychological review*, 98(2):182, 1991.

PN Johnson-Laird and RMJ Byrne. Essays in cognitive psychology. deduction, 1991.

Philip N Johnson-Laird. Mental models and human reasoning. *Proceedings of the National Academy of Sciences*, 107(43):18243–18250, 2010.

Paolo Legrenzi. *Relations between language and reasoning about deductive rules*. na, 1970.

John E Taplin. Reasoning with conditional sentences. *Journal of Verbal Learning and Verbal Behavior*, 10(3):219–225, 1971.

John E Taplin and Herman Staudenmayer. Interpretation of abstract conditional sentences in deductive reasoning. *Journal of Verbal Learning and Verbal Behavior*, 12(5): 530–542, 1973.

Herman Staudenmayer. Understanding conditional reasoning with meaningful propositions. *Reasoning: Representation and process in children and adults*, pages 55–79, 1975.

Valerie A Thompson. Interpretational factors in conditional reasoning. *Memory & cognition*, 22(6):742–758, 1994.

Jonathan St BT Evans, Simon J Handley, Helen Neilens, and David E Over. Thinking about conditionals: A study of individual differences. *Memory & Cognition*, 35(7): 1772–1784, 2007.

Andrew JB Fugard, Niki Pfeifer, Bastian Mayerhofer, and Gernot D Kleiter. How people interpret conditionals: shifts toward the conditional event. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 37(3):635, 2011.

Frank P Ramsey. *Philosophical Papers*, chapter General Propositions and Causality, pages 145–163. Cambridge University Press, 1990.

John E Taplin, Herman Staudenmayer, and Judith L Taddonio. Developmental changes in conditional reasoning: Linguistic or logical? *Journal of Experimental Child Psychology*, 17(2):360–373, 1974.

Carol L Smith. Quantifiers and question answering in young children. *Journal of Experimental Child Psychology*, 30(2):191–205, 1980.

Napoleon Katsos and Dorothy VM Bishop. Pragmatic tolerance: Implications for the acquisition of informativeness and implicature. *Cognition*, 120(1):67–81, 2011.

Michael Franke. Quantity implicatures, exhaustive interpretation, and rational conversation. *Semantics and Pragmatics*, 4(1):1–82, June 2011. doi: 10.3765/sp.4.1.

Emmanuel Chemla. Similarity: Towards a unified account of scalar implicatures, free choice permission and presupposition projection. *Under revision for Semantics and Pragmatics*, 2009.

Robert Van Rooij. Conjunctive interpretations of disjunctions. *Semantics and Pragmatics*, 3:11–1, 2010.

Lyn Tieu, Jacopo Romoli, Peng Zhou, and Stephen Crain. Children's knowledge of free choice inferences and scalar implicatures. *Journal of Semantics*, 33(2):269–298, 2015.

Bob van Tiel and Walter Schaeken. Processing conversational implicatures: alternatives and counterfactual reasoning. *Cognitive science*, 41(S5):1119–1154, 2017.

Gerald Gazdar. Pragmatics: Implicature, presupposition, and logical form, 1979.

Andrei Moldovan. Pragmatic considerations in the interpretation of denying the antecedent. *Informal Logic*, 29(3):309–326, 2009.

Richard Breheny. Communication and folk psychology. *Mind & Language*, 21(1):74–107, 2006.

Joëlle Proust. Can "radical" simulation theories explain psychological concept acquisition? *Advances in consciousness research*, 45:201–228, 2002.

Kevin J Riggs, Donald M Peterson, Elizabeth J Robinson, and Peter Mitchell. Are errors in false belief tasks symptomatic of a broader difficulty with counterfactuality? *Cognitive Development*, 13(1):73–90, 1998.

Bob Van Tiel, Emiel Van Miltenburg, Natalia Zevakhina, and Bart Geurts. Scalar diversity. *Journal of Semantics*, 33(1):137–175, 2016.

David Barner, Neon Brooks, and Alan Bale. Accessing the unsaid: The role of scalar alternatives in children?s pragmatic inference. *Cognition*, 118(1):84–93, 2011.

Lyn Tieu, Cory Bill, Jacopo Romoli, and Stephen Crain. Testing theories of plural meanings: Insights from child language. 2017.

Philip N Johnson-Laird and Ruth MJ Byrne. Conditionals: a theory of meaning, pragmatics, and inference. *Psychological review*, 109(4):646, 2002.