THE LOGIC OF PRAGMATICS
An experimental investigation of children and adults

Tutor: Chiar. ma Prof.ssa Maria Teresa Guasti (Facoltà di Psicologia)
Coordinatore: Chiar. mo Prof. Giuseppe Vallar

Tesi di Dottorato di:
Dott.ssa Francesca FOPPOLO
Matricola n° R00253

Ciclo XVIII
Anno Accademico 2005/2006
To my husband Alessandro…
(with the promise to never write a dissertation again!)
CONTENTS

INTRODUCTION pag. 1

CHAPTER 1
The phenomenon of scalar implicature computation:
theoretical accounts pag. 5

1. The phenomenon pag. 5
2. Different accounts of Scalar Implicatures pag. 15
   2.1 Levinson (2000): a theory of Generalized Conversational Implicatures pag. 15
   2.2 Relevance Theory: the balance of cost and benefit against a default account of Sis pag. 21
   2.3 A grammar-driven approach: seeking for spontaneous logicality pag. 27
3. Conclusive remarks pag. 44
CHAPTER 2
Scalar implicature computation and logicality:
the case of adults

1. Monotonicity and SI computation: an off-line study on or pag. 48
   1.1 Experiment 1: a Truth Value Judgment Task pag. 49
      1.1.1 Subjects pag. 49
      1.1.2 Material and Procedure pag. 50
      1.1.3 Results pag. 52
      1.1.4 Discussion pag. 56
   1.2 Conclusive remarks on Experiment 1 pag. 63
2. A reaction-time study on the interpretation of or pag. 67
   2.1 Experiment 2: exclusive or inclusive OR? pag. 67
      2.1.1 Subjects pag. 67
      2.1.2 Material pag. 68
      2.1.3 Procedure pag. 73
      2.1.4 Predictions pag. 74
      2.1.5 Results pag. 80
      2.1.6 Discussion pag. 85
   2.2 Conclusive remarks on Experiment 2 pag. 90

CHAPTER 3
Beyond scalar implicatures:
introduction to the phenomenon of “donkey-sentences”

1. The puzzle of donkey anaphora resolution pag. 94
2. Alternative interpretations of “donkey sentences” pag. 97
3. The debate on ambiguity pag. 101
   3.1 Kanazawa (1994): preserving inferential patterns pag. 101
   3.2 Lappin and Francez (1994): an E-type account of “donkey sentences” pag. 104
   3.4 Krifka (1996): the rule of Pragmatic Strengthening pag. 108
   3.5 Chierchia (1992/1995): the availability of two alternative readings pag. 111
# CHAPTER 4

*Defaults and logicality in “donkey anaphora” resolution: an experimental investigation*

1. Experiment 3A - Between universal and existential interpretation of the anaphora: is there a default?  
   1.1 Subjects  
   1.2 Procedure  
   1.3 Material  
   1.4 Predictions  
   1.5 Results  
   1.6 Discussion  

2. Experiment 4 - Checking for preference  
   2.1 Subjects  
   2.2 Procedure  
   2.3 Material and Predictions  
   2.4 Results and Discussion  

3. Experiment 3B - Degrees of accessibility of the alternative readings  
   3.1 Material  
   3.2 Predictions  
   3.3 Results  
   3.4 Discussion  

4. Conclusive remarks on donkey anaphora resolution

---

# CHAPTER 5

*Scalar implicature computation and logicality: the case of children*

1. Children’s performance on SIs: a survey on previous experimental findings  
2. Experiment 5: is there a developmental effect in the computation of SIs?  
   2.1 Subjects  
   2.2 Material and Procedure  
   2.3 Results  
   2.4 Discussion
3. Experiment 6: generalizing performance across similar items pag. 171
   3.1 Subjects pag. 171
   3.2 Material and Procedure pag. 172
   3.3 Results pag. 172
   3.4 Discussion pag. 172

4. Taking stock pag. 173

5. Experiment 7: knowledge of some and all in felicitous contexts pag. 174
   5.1 Subjects pag. 174
   5.2 Material and Procedure pag. 174
   5.3 Results pag. 175
   5.4 Discussion pag. 175

6. Experiment 8: investigating children’s knowledge of informational strength pag. 175
   6.1 Subjects pag. 176
   6.2 Material and Procedure pag. 176
   6.3 Results pag. 178
   6.4 Discussion pag. 178

7. Experiment 9: priming the scale pag. 180
   7.1 Subjects pag. 180
   7.2 Material and Procedure pag. 180
   7.3 Results pag. 182
   7.4 Discussion pag. 182

8. Taking stock pag. 183

9. Experiment 10: a novel experimental design pag. 186
   9.1 Subjects pag. 186
   9.2 Material and Procedure pag. 187
   9.3 Prediction pag. 189
   9.4 Results pag. 190
   9.5 Discussion pag. 194

10. Conclusive remarks on SI computation in children pag. 200
<table>
<thead>
<tr>
<th>Appendix A</th>
<th>pag. 203</th>
</tr>
</thead>
<tbody>
<tr>
<td>Appendix B</td>
<td>pag. 204</td>
</tr>
<tr>
<td>Appendix C</td>
<td>pag. 205</td>
</tr>
<tr>
<td>Appendix D</td>
<td>pag. 206</td>
</tr>
<tr>
<td>Appendix E</td>
<td>pag. 207</td>
</tr>
<tr>
<td>Appendix F</td>
<td>pag. 208</td>
</tr>
<tr>
<td>Appendix G</td>
<td>pag. 209</td>
</tr>
<tr>
<td>Appendix H</td>
<td>pag. 210</td>
</tr>
<tr>
<td>Appendix I</td>
<td>pag. 211</td>
</tr>
<tr>
<td>Appendix L</td>
<td>pag. 212</td>
</tr>
<tr>
<td>Appendix M</td>
<td>pag. 213</td>
</tr>
</tbody>
</table>

**REFERENCES**

pag. 215
In this dissertation I will present some experimental works that investigate how we interpret sentences containing scalar items such as or or some and ambiguous sentences containing a special kind of anaphora. These phenomena have extensively been studied in the semantic tradition, and are known in the literature respectively with the name of Scalar Implicatures and Donkey Anaphora. The former have recently attracted the attention of many psycholinguists (cf. a.o. Noveck and Posada, 2003, Bott and Noveck, 2004, Breheny et al., 2005), while the latter have mostly been investigated in a purely theoretical perspective. Interestingly, though, the debate on the theoretical status of both phenomena is still vivid and not yet concluded (cf. a.o. Sperber and Wilson, 1986, Levinson, 2000, Carston, 2002, Chierchia, 2004/2006, Fox, 2003 for a discussion of Scalar Implicatures and, a.o., Kanazawa, 1994, Chierchia, 1992/1995, Krifka, 1996 for a discussion of Donkey Anaphora). I believe that experimental works such as the one I'm proposing may help disentangle the factors at play and ultimately decide among competing theories.

This is the core of the claim I intend to make here: the way we deal with the phenomena under investigation testify the fact that we (quite unconsciously and spontaneously) appeal to “logicality” in performing tasks as natural as interpreting sentences or interacting with an interlocutor. As emerges from the studies that I will present, logic properties such as entailment patterns and monotonicity properties of the context seem in fact to determine the interpretation we assign to pronouns appearing in sentences of a special kind, such as donkey sentences, and to affect the pragmatic inferences that we derive in our conversational exchanges. Only appealing to these logic notions, for example, a generalization on the interpretation of donkey pronouns introduced by different quantifiers has been proposed (cf. Kanazawa, 1994) and the hypotheses that this interpretation constitutes the default one has been advocated. This neat and simple claim is at the same time pregnant of significance for a theory that aims at

Abstract
investigating the mechanisms that underlie sentence processing and sentence comprehension. Analogously, I propose an explanation of scalar implicature computation in terms of the semantic/pragmatic notion of informativeness, which ultimately resides on the logic notion of monotonicity.

Children themselves have been found to be sensitive to logic properties of the syntactic context, such as Downward Entailingness (cf. Boster and Crain, 1993 and Gualmini, 2003). At the same time, however, it has been claimed that they behave poorly with Scalar Implicatures (cf. a.o. Chierchia et al., 2001/2004, Noveck, 2001, Papafragou and Musolino, 2003, Guasti et al., 2005). On the basis of the experimental works that I conducted on children, it seems that they in fact posses most of the knowledge necessary to derive pragmatic inferences. And, in particular, that they are aware that a semantic/pragmatic notion of informativeness is involved in talk exchanges. What they seem to lack in order to reach adultlike performance is the full lexical trace associated to scalar terms, which presumably takes time to be internalized and to become automatized.

In 2001, the question of scalar implicature computation in children has been brought back by Noveck in a paper entitled “When children are more logical than adults” (cf. Noveck, 2001). In the perspective adopted here, adults showed a high level of spontaneous logicality in solving pragmatic tasks. It seems therefore that a logic of pragmatics indeed emerged, and permeates the mechanisms to which we, as cooperative speakers, seem to conform effortlessly and spontaneously. In the same spirit, thus, we can conclude that, in fact, children are as logical as adults, resulting to be in the end more competent than we thought in solving “pragmatic” tasks.
Acknowledgments

On a foggy November 29th, 1999, I was discussing my MA thesis in a quite anonymous room at the University of Milan, Department of Philosophy. Its title was (translated) “Aspects of the semantics and pragmatics of disjunction”. By my side, to face the examining board, Gennaro Chierchia and Maria Teresa Guasti, my advisors then. Behind me, my family: my husband (of two-weeks, at that time) Alessandro, my parents, my grandparents Giacomo and Ines, Alessandro’s mother Iva…

Today it’s a foggy November 29th, 2006.

Tomorrow, the final draft of my PhD dissertation is due. By my side, Teresa, as always, and Gennaro, only “physically” more distant now. They have been my advisors in this long academic journey. Constantly present and encouraging. And they have been much more than that. I am extremely grateful to them for all their (multidimensional) support in all these years. I will miss Gennaro a lot. Without his charismatic presence, and his placid manners, Bicocca looks much emptier…and much less reassuring, in a sense. I hope to have the chance to work with him again in the future. And with Teresa, too, of course!

Behind me, today as then, my whole family, again, constantly. Only larger, after the most important addition to its members, Lorenzo.

On the table today, my PhD dissertation, my last one (hopefully!). Semantics and pragmatics again, but “higher level” (…again, hopefully!).

Exactly 7 years later, the story seems to repeat itself in a cycle. In a way, nothing seems to have changed. But in another, it actually has. And a lot, indeed!

During the four years of my PhD program, I spent a Fall semester in Amherst. I remember every single moment, place, situation, conversation, impression, sensation…that I’ve experienced in those four months. Each instant spent there was significant. And I’m not (only) talking about the amazing nature that I found there: October maples, squirrels rushing up and down across the
multicolour foliage, November snow, even an opossum in a bush in campus… I’m talking about the people I met there. They all will have such a long-lasting influence on me. Not only on me as a linguistic, but mostly on me as a feeling/thinking being. Any time I talk about UMass people and Amherst I’m conscious to be so rhetorical, so much dramatic in my enthusiasm that it may sound as “typical Italian exaggeration”. Well, no! And I’m sure all the people who had the lucky to have been there for a while know what I’m talking about.

For all her time, and precious teaching, and reassuring comments (and also for her delicious pumpkin soup!) I want to deserve a special thank to Lyn Frazier. And another special thank is deserved to Tom Roeper, the best interlocutor ever: he’s never tired to listen to you and has always insightful and original comments to make and interesting stories to tell (and also, he also organizes amusing Halloween dinners!) For their welcome and friendship, I can’t forget all the other people in South College, especially Angelika and Lisa, because time spent with them is always amusing, and Chris Potts, for his sparkling semantic proseminar, his enthusiasm (and pizza dough). For their generous and patient help with programming and running eye-tracker, I want to kindly thank all the people in the Tobin Hall psych-lab, in particular: Chuck Clifton, Keith Rayner, Barb Juhasz, Becca Johnson, Tim Slattery, Jane Ashby, Adrian Staub. For their friendship and nice conversations, I want to thank all UMass grad students, especially Liane Jeschull and Anna Verbuck. For her help with bureaucracy and faxing and for her contagious laugh, thanks also to Kathy.

I believe that my UMass experience was so unique and perfect especially because I lived at Hobart Lane. I want to express my gratitude to Barbara Partee and Volodja Borschev for being such marvellous landlords. I’m always moved when I think of how special they are and how special and unique that sharing was. And I also thank my kind and amusing house-mate Yoshi Yamada, for adding his delicate exquisite Japanese touch to the house (even if this meant fish-smell at 7 a.m. sometimes!). I know that he remembers that period with the same nostalgia.

Another abroad experience had a great influence on me. That was the summer school in St. Petersburg, July 2004. Living at the 18th floor in the dorm in

x
Васильеского with an impressive view on the gulf of Finland changed my point of view on many things. Especially, I learnt not to take everything for granted (because most of the time it is not). I thank my room-mate Lena Anishchenko for trying to make me understand the logic of the reality I found there (even if I partly failed in that). I also thank John Bailyn, for his amusing (and tough!) Russian classes (and soccer games), and Chris Potts again, for lots of conversations (and discussions!) and never-ending walks around the city with Lena and me.

Back to Milan…in the first place I want to thank all my colleagues that shared at least some part of this experience with me, in particular: Lola De Hevia, Giorgio Magri, Simona Sacchi, Leonor Romero, Federica Castellini, Davide Crepaldi. A special thank to my former 435-room-mates Claudia Caprin for her friendship and encouragement, Luca Vanin for sharing tea and confidences and Mirco Fasolo especially, for always finding an excuse for paying me compliments (and giving me a hug!). To my present 1002-room mates Fabrizio Arosio and Dana Kuefner who always refill the coffee-and-chocolate corner and Flavia Adani especially, because she was absolutely convincing when she kept saying that I was beautiful (despite the fact that I grew 20 kg during pregnancy!).

A special thank goes to my dear friend and colleague Carlo Geraci, the most good-hearted guy I know (and, maybe because of this, even probably the messiest!) and to Silvia, for making him happy.

Also, I’m grateful to all the people in Bicocca-Psychology department, especially to Emanuela Bricolo for her help with E-Prime programming and Marcello Gallucci and Claudia Caprin for their help in solving some of my “statistic” doubts.

I’m very grateful to all members of the “linguist” family in Milan, especially to Francesca Panzeri and Fabio Del Prete for their comments on an early draft of chapters 2 and 3 and Carlo Cecchetto, because he’s always encouraging and reassuring. For her help with statistics and a lot more (her wonderful oriental dinners included!) I want to thank Maria Nella Carminati. For his help, comments, friendship, delicate paintings (and sage&thyme tea sharing!), I want to thank Orin Percus.
Of course, I cannot forget Joke De Lange and Chiara Ioghà who helped me in carrying out some of the experiments with children, and all the participants in my studies (my friends and relatives used as guinea pigs for pilots especially), and the teachers and directors of the schools in which the experiments were carried out, for their willingness to help and their understanding.

Among my non–linguist friends, I want to thank Rudy and Lilly, because they always worry for my emotional/psychological stability under stress (and Alessandro’s especially!) and for always letting me giving vent to my feelings.

But the people I think I owe most and felt more close in this adventure have been the people in my family. Without their (even material) help all this could not have happened. This dissertation is dedicated to them. To my grandparents Giacomo and Ines, who always worry for my academic (and spiritual) life. Having been a real failure in the latter, I hope they’re proud of me for the former! To my nonna Agnese and my nonno Angelo and also to nonna Martina. To my delicious nieces Chiara and Irene and to my lively nephews Simone, Davide and Luca, who showed their enthusiasm towards my testing stories. And to my sister Silvana and Bruno, and Stefania and Marco for having such terrific children.

An immense thank goes to my parents Vito and Giovanna, and to Alessandro’s parents Iva and Ugo, for being such wonderful grandparents to our child. Without their precious help in baby-sitting, house-works and cooking (and a lot more!), I’m sure I would have never made it (and probably we wouldn’t have even physically survived!).

In the end, I want to say I’M SORRY to Lorenzo, my (nearly) 15 month child, and to Alessandro, my partner of (more than) 15 years, for having been such a busy mother and such an absent wife for the whole last year. It was hard for everybody. The only thing I can say, now that everything is over, is the simples but most sincere and effective…THANK YOU!

(...for being my whole life)

Milan, 29th November 2006
In the very first page of an undoubtedly influent book in the linguistic tradition, this passage is reported:

“and what is the use of a book” thought Alice “without pictures or conversations?”
[from Lewis Carroll’s “Alice’s Adventures in Wonderland”, 1865]

Luckily, in this book, we’ll have some pictures indeed, which actually were presented as stimuli in some of the experiments I will report, and certainly a lot of conversations, which actually constitute the central topic of my whole dissertation. My aim here is, in fact, to investigate the “logic” that permeates the way we behave when we participate as addressee in talk exchanges, and which normally falls in the realm of what is labeled “pragmatics”.1

Two linguistic phenomena will be specifically investigated in this dissertation: Scalar Implicatures and Donkey Sentences. The first is a “pragmatic” phenomenon that has recently attracted the interest of many people working in the field of semantics and psycholinguistics. Scalar Implicatures are inferences that

---

1 The term “pragmatic” always appears in quotes to signal the fact that the way this term is used here is different from the way it is commonly used, such as, for example, opposed to “semantic”.
normally arise in conversational exchanges, in which expressions of the form “A or B” or “some Ys”, by virtue of the scalar term they contain, tend to convey the inference than “not both A and B” and “not all Ys” is what the speaker intended to communicate.

The second, instead, is a “semantic” phenomenon that has extensively been studied in a theoretical, but rarely in an experimental, perspective. Donkey sentences are represented by examples like (1), in which an indefinite expression embedded in the restrictor of a quantifier functions as the antecedent of a pronoun that follows in the nuclear scope, giving rise to semantic ambiguity:

(1) Every man who has a son likes to play with him
    \rightarrow Every father likes to play with his son/s (= with all his sons, if he has more than one);
    \rightarrow Every father likes to play with (only) one of his sons (for example, his favorite one, or the youngest).

At first glance, the ones just presented seem two distant, unrelated phenomena. At a closer look, however, they reveal some similarities that I intend to specifically address and highlight in this dissertation. In the first place, both constitute an intriguing challenge for those semanticists who intend to provide a theoretical account of such phenomena within a formal framework. Secondly, the current “theoretical status” of both is still debated. More interestingly, however, I believe that what these phenomena have more in common is their potential in shedding new light on the way we understand and interpret linguistic material like the one investigated here which, apparently, has nothing to share with logics. In particular, the experimental findings of the works that investigated these topics separately converge on the revelation of a general interpretative strategy. Namely, as the title of this dissertation suggests, the fact that we consistently appeal to “logical” notions in solving tasks that seem to share little with logic, such as anaphora resolution or “pragmatic” inferencing.
The dissertation is organized as follows. Chapter 1 provides a theoretical introduction to the phenomena of Scalar Implicatures, in which different accounts are presented and compared. Among these is the grammar-driven approach proposed by Chierchia (2001/2004) that will constitute the theoretical framework of the experimental works I will be reporting in other chapters.

Chapter 2 is devoted to experimental works on Scalar Implicature computation in adults: two experiments are presented in which the same phenomenon is investigated by means of different experimental techniques.

Chapter 5, instead, addresses the same experimental questions from the point of view of acquisition, and is focused on the investigation of “pragmatic” inferences in children. All the experimental works presented in this last chapter aim at attesting which factors may influence children’s computation of Scalar Implicatures and eventually contribute to their failure in this task. Or, else, contribute to improve their performance.

Between these chapters dedicated to the investigation of Scalar Implicature computation in adults and children, there is a section fully dedicated to the second of the phenomena studied, i.e., “donkey anaphora” resolution. This section comprises an introductory chapter on the theoretical accounts proposed in the literature (Chapter 3) and a chapter dedicated to an experimental study that specifically addresses the question of the existence of a default (preferred) interpretation of “donkey sentence” such as (1) in adults (Chapter 4).
CHAPTER 1

THE PHENOMENON OF SCALAR IMPLICATURE COMPUTATION: THEORETICAL ACCOUNTS

Since Grice (1957), a lot of theoretical works have been developed by semanticists to account for the phenomenon of Scalar Implicatures, which has more recently attracted the interests of linguists and psycholinguists who are developing the question further on experimental grounds. In this chapter, I will introduce the phenomenon of conversational implicatures and will discuss some aspects of three major families of theoretical approaches to account for it, in order to give a sense of the complexity and liveliness of the current debate. Further on, I will tackle the question in more detail in an experimental perspective, reporting some works that I conducted on adults and children (cf. Chapter 2 and 5 respectively).

1. The phenomenon

The term *implicature* is due to Grice, who originally discussed examples like (1) below to illustrate a common phenomenon that occurs in our talk exchanges:
(1)  *On the phone*

    Alex: What are you doing?
    Francesca: I’m playing with a baby in the garden

What I intend to suggest to my interlocutor in (1), is that the baby I’m playing with is not my son Lorenzo (nor, for example, one of my nephews). The use of the indefinite article suggests, actually, *implicates*, such a conclusion, which my addressee is entitled to derive from my answer. Analogously, from utterance (2) the *implicature* (3) is normally derived:

(2)  Lorenzo drawn my attention *and* started eating his book

(3)  Lorenzo drawn my attention *and then* started eating his book

According to Grice, the added implicature is not part of the literal meaning of the sentence (what is said) but is part of what is implicated by the use of such a sentence in normal circumstances: a speaker S chooses to produce a certain utterance U iff S “intended the utterance of U to produce some effect in an audience by means of the recognition of this intention” (1957: 385). Actually, speakers and hearers are supposedly interacting in a rational and cooperative way, following a general rule formulated by Grice as the Cooperative Principle, fully reported below:

(4)  “Make your conversational contribution such as is required, at the stage at which it occurs, by the accepted purpose or direction of the talk exchange.” [Grice, 1967]

This general principle is instantiated by four general maxims of conversation governing rationale interchange, namely: Quality (say things that you believe/know to be true); Quantity (1. make your contribution as informative as is required; 2. do not make your contribution more informative than is required); Relevance (be relevant); Manner (be perspicuous). Considering the
examples given above, we can think the implicature in (1) to be linked mainly to the first maxim of Quantity, and the one in (3) to the maxim of Manner. To exemplify the effect of the other two maxims, consider the exchange in (5), in which the Quality maxim is at play, and the one in (6), in which the communicative effect is due to a violation of the maxim of Relevance:

(5) Lorenzo (in 6 months, maybe!): Mummy, where’s old Mac Donald?
Francesca: You probably have thrown it under the sofa.

(6) Grandpa: Did you sleep well tonight?
Francesca: Well, Lorenzo had a toothache.

By means of the answers in the examples above, what I intend to communicate is, respectively, that I genuinely don’t know where old Mac Donald is, I can only make a guess (case 5) and that I didn’t sleep that much (case 6).

The examples that we have encountered so far are instances of what Grice called particularized conversational implicatures (PCIs henceforth), providing the following definition: an implicature $\phi$ from an utterance $U$ is particularized iff $U$ implicates $\phi$ only by virtue of specific contextual assumptions that would not invariably or even normally obtain. Specifically, PCIs depend on particular contextual features. Opposed to this class of implicatures are the generalized conversational implicatures (GCIs henceforth), i.e. implicatures that are always (independently from features of the context) derived by uttering $U$ unless there are specific contextual assumptions that defeat it. Grice considered this last class of implicatures “more valuable for philosophical purposes”: these are in fact the implicatures that are hard to distinguish from semantic or conventional content, and in particular are the inferences that arise from the use of the logical connectives.

To have a broader picture of the classes of meanings studied by Grice, we can insert the above distinction in a more general schema:
As regards the first branching under the notion “what is implicated”, a line is drawn between *conventional* and *conversational implicatures*. This in order to distinguish between two classes of implicatures: those that are not calculable from the cooperative principle and the cooperative maxims, like the “conventional” implicatures arising from the use of certain terms, such as “but” or “therefore” (and thus linked to the conventional meaning of words) and those that instead can be derived from that pragmatic framework. I will only deal here with this second type of implicatures while, for what concerns the first type, I refer the reader to Christopher Pott’s dissertation (2003) “The logic of conventional implicatures”, in which he gives an extended and systematic account of such, too frequently unconsidered in the past, family of implicatures. As for the second branching, under the notion “conversational implicatures”, the distinction between *generalized* and *particularized implicatures* is reported, the one from which we started our discussion. As for the class of GCIs, a further specification can be added after Horn (1972), who was the first to introduce the definition of *scalar implicature*, the sub-class of implicatures that I will focus on throughout my dissertation. Let’s consider some examples:

(8) Lorenzo put the cow or the horse on old Mac Donald’s tractor

(9) Some of the books on the shelf have missing pages
Sentences in (8) and (9) contain two different scalar items, namely *or* and *some*, and normally give rise to the *scalar implicatures* below (SIs henceforth):

(10) Lorenzo put the cow *or* the horse on old Mac Donald’s tractor, *but not both*

(11) *Some* of the books on the shelf have missing pages, *but not all*\(^1\)

The question here is: how do we get from *or* and *some* to the SI *not both* and *not all*? The basic idea is the following. Uttering something involves choosing from a range of reasonable alternatives. In particular, sentences (8) and (9) contain two scalar items, which belong to the following informational scales:

(12) *or* < and

(13) *some* < *many* < *most* < *all*

Using a generalized notion of entailment, we can view the elements in a scale as in a subset/superset relationship (*all* \(\subseteq\) *some*; *and* \(\subseteq\) *or*) given that a statement including the upper elements in the scale (as for example *all* or *and*) asymmetrically entails a sentence containing the lower elements (*some* and *or* in our examples). Thus stronger elements in a scale are more informative (i.e. true in a smaller set of circumstances) than weaker ones and this fact is crucial to understand the reason why implicatures are derived. Because of the presence of a scalar term, sentences in (8) and (9) typically evoke the following alternative possibilities:

(14) Lorenzo put the cow *and* the horse on old Mac Donald’s tractor

---

\(^1\) The inference in (11) is actually what Grice would have called the GCI related to the use of a sentence like (9). If, however, this same sentence were used in the following interchange, as the answer to mummy’s question, then the inference which is likely to be derived (by the maxim of Relevance) would be the one reported below:

- Mummy: “Lorenzo is chewing something…what is it?”
- Daddy: “Some of the books on the shelf have missing pages”
- PCI: what Lorenzo is chewing are the missing pages of the books (sob!)
(15) All the books on the shelf have missing pages

Sentences (14) and (15) not only are semantically compatible with the sentences uttered in (8) and (9), they also entail them: in a situation where, for example, a sentence of the form “A and B” is true also the alternative “A or B” containing the weaker element would have to be, given the basic (logical) meaning of or. As a consequence, choosing to utter (8) and (9) over the stronger statements (14) and (15) that could have been uttered instead is taken to convey that these alternative statements do not hold. This intent is taken over by the addressee, who interprets these utterances as (10) and (11) respectively, in which the implicature is explicitly factored in. The actual steps of this inference remain controversial, as we will see below when discussing different approaches. But the distinction, that ensues from this characterization of the facts, between a first level of propositional or basic meaning (in which a sentence containing a weaker element in the scale is compatible with the corresponding alternative in which a stronger element is used) and a second level of pragmatically enriched meaning (in which a sentence containing a weaker element in the scale is interpreted as excluding the corresponding alternative containing a stronger element, as stated for example in sentences (10) and (11)) is widely agreed upon. In Gricean terms, this is how SIs come about when a hearer encounters, for example, a sentence like (9), containing a weak term like some:

(16)

i. the speaker said (9) and not (15), which would have been also relevant;

ii. (15) entails (9) \[\text{some and all are part of a scale and all } \subseteq \text{ some}\];

iii. if the speaker had evidence that all the books on the shelf had missing pages, then she would have uttered statement (15) \[\text{by Quantity maxim}\];
iv. the speaker has no evidence that all the books on the shelf have missing pages;

v. the hearer assumes that the speaker is well informed and is being cooperative;

vi. thus the hearer concludes that sentence (15) does not hold, explicitly factoring in the implicature as shown in (11).

In the classical view just sketched, SIs are computed “globally”, i.e. after the grammar has completed its job. It is in fact assumed that one first computes the plain meaning of a sentence (by means of the semantic module) and then, taking into account the relevant alternatives, one eventually strengthens that meaning by adding the implicature (by means of the pragmatic module). As it is evident in this account, the two modules operate in sequential order: the pragmatic module makes its entry on the scene only after the semantic module has derived (compositionally) the meaning of the sentence, which constitutes the input for the pragmatic module. Following recent discussion on the semantic of focus (cf. Rooth, 1985), Krifka (1995) proposes an explicit model of SI computation according to which the recursive part of the semantic module is set up in such a way as to compute, next to the truth-conditional content of a sentence $\llbracket S \rrbracket$, also its relevant alternatives $\llbracket S \rrbracket^{ALT}$. Back to example (9) above, containing the weak term some, the semantic module would derive its truth-conditional content (i.e. $\exists x [\text{book}(x) \land \text{have-missing-pages}(x)]$) against the background of a relevant set of alternatives that might have been uttered instead of it. In the case at hand, given the (positive) quantifier scale in (13), the relevant set of scalar alternatives would be the following: $\llbracket \text{some} \rrbracket^{ALT} = \{ \text{some} < \text{many} < \text{most} < \text{all} \}$. The shift from sentence (9) to (11) is thus made by following the steps outlined in (16), that Krifka formalizes by defining a notion of scalar context incrementation whereby adding sentence (9) to a context $c$ amounts to adding to $c$ the following (which corresponds to (11)): $\exists x [\text{book}(x) \land \text{have-missing-pages}(x)] \land \neg \forall x [\text{book}(x) \rightarrow \text{have-missing-pages}(x)]$. 
A further note on implicature derivation is worth taking into consideration. We have seen that, crucially, the alternative statements that one could utter (for example (9) and (15)) are not independent of one another, given that the latter (which is the stronger one) entails the former. However, the existence of lexical entailments such as those from sentences (14) and (15) above to sentences (8) and (9) is a necessary but not sufficient condition for SIs to come about (a point duly emphasized in Horn, 1972). To make this point clearer, consider for example the following pairs:

(17)
(a) Lorenzo had lunch
(b) Lorenzo had lunch in his high chair

(18)
(a) Lorenzo moved quickly through the room
(b) Lorenzo run quickly through the room

The (b)-sentences in (17) and (18) actually entail the (a)-sentences. Yet, uttering the (a)-sentences typically does not convey any particular intention about the (b)-sentences: for example, stating that Lorenzo had lunch does not tell us anything as to how, where or when he did it and therefore we have no reason to infer from (17a) that Lorenzo had lunch, but not in his high chair, i.e., to deny (17b). Examples of this sort can be multiplied ad libitum. In particular, these considerations will prove useful when discussing the experimental works on children in Chapter 5.

To conclude our description of the phenomenon of SI, let’s consider two distinctive properties that characterize implicatures in general: in the first place, differently from entailments, SIs are cancellable (i.e. defeasible) without giving rise to contradiction, as shown in (19):
(19) (After a clash from the kitchen)
Lorenzo dropped the plate or the glass from his high chair, or both

Moreover, they are nondetachable, given that any other expression with the same coded content will tend to carry the same implicature.

Let’s now turn to a more controversial discussion on the generation and status of implicatures in general, and SIs in particular, given that the experiments I conducted were focused on this particular sub-class of inferences. Of the various theoretical perspectives on SIs that have been originally inspired by Grice’s work discussed in this section, it’s worth considering three different families of approaches in more detail. The first, that I will define Canonical Neo-Gricean, might be characterized by a set of assumptions of the following tenor:

(20) Canonical Neo-Gricean approaches:
(a) SIs are added to a fully specified basic meaning by default in a cost free way;
(b) SIs may be cancelled or suspended;
(c) although SIs are computed “globally” at the level of whole utterances, they may also intrude (i.e. occur in embedded position).

These (oversimplified) assumptions are common to, e.g., Horn (1972), Levinson (2000) or also Chierchia and Mc. Connell Ginet (2000). These approaches are often contrasted with Relevance Theoretic approaches that are designed to derive the following assumptions (from Sperber & Wilson, 1986/1995):

(21) Relevance Theoretic approaches:
(a) SIs are really explicatures, i.e., part of a pragmatically driven process of enrichment of an underspecified semantic representation;
(b) SIs are not derived by default, but they are derived at a cost to meet expectations of *optimal relevance*;

(c) SIs are computed globally, and the phenomenon of intrusion must be accounted for as part of the *explicature* process.

More recently, there have been proposals that revive the idea that SIs are computed compositionally as part of a grammatically driven process, an idea that hadn’t been pursued after Gazdar (1979). A characteristic common to such approaches is that the alternatives relevant to SIs computation are processed and possibly factored in as part of the compositional semantics, by analogy with what has been proposed for focus by Rooth (1985). An example of an approach of this sort is Chierchia (2004) (see also i.a. Fox (2004) who shares this general take). More recently, Chierchia (2006) has proposed a refined version of his theory of SI computation and polarity phenomena in which he reaches the following “tentative conclusions: maybe SIs and NPI licensing are REALLY the same thing” [handout 2: 5]. If compositional processes that derive the whole meaning of sentences are generally considered as part of the grammar, under this approach that can be labeled “grammar-driven”, also the derivation of SIs and thus of the (pragmatically) enriched interpretation assigned to, e.g. (9), is yielded by a compositional process similar to the one involved in deriving the meaning of a simple sentence like *Lorenzo is eating my newspaper*. The two compositional processes only differ in their complexity, as the derivation of SIs requires the computation of alternatives, a process that is similar to what happens if an element in the sentence gets focused, like *Lorenzo is EATING_F my newspaper* (alt: *reading*). To summarize, these are the assumptions of this grammar-driven approach:

(22) **Grammar-driven approaches:**

(a) simultaneous computation of alternatives (by default);

(b) factoring in of alternatives into meaning at various stages of the computation (locally).
At this point, it’s time to turn to a more detailed description of each of the approaches sketched above, starting from one representative of the Neo-Gricean family, namely Stephen Levinson (2000).

2. Different accounts of Scalar Implicatures

2.1 Levinson (2000): a theory of Generalized Conversational Implicatures

In his 2000 book “Presumptive Meanings”, Levinson develops a theory of generalized conversational implicatures within the Gricean tradition in which he revisits the traditional maxims and proposes some heuristics to explain how GCIs are generated. His declared aim, as reported at the beginning of Chapter 1 of his book, is to “defend the notion of a generalized conversational implicature (...) where what is at stake is the generalized nature of such inferences” [pg. 11: 3-5]. He then continues by anticipating the main claim of his theory: “A generalized implicature is, in effect, a default inference, one that captures our intuitions about a preferred or normal interpretation” [pg. 11: 3-5]. His focus on a sub-class of implicatures, namely GCIs, is mainly motivated by the recognition that these inferences, because of their generalized, stable nature, provide an unparalleled import to linguistic theory. As he explicitly states: “they have much closer interactions with the stable body of rules of formation and construal that form a grammatical system” [pg. 22: 7-9].

Under this perspective, he intends to introduce a new way of thinking about the maxims proposed by Grice: instead of viewing them as rules or behavioural norms, he re-thinks them as inferential heuristics that lay beyond behavioural norms, constituting the background for inferential principles. The first of these heuristic is related to Grice’s first maxim of Quantity and it is simply formulated as: “What isn’t said, isn’t”. This Q-heuristic is central to our discussion on SI, given that this special kind of GCIs are generated for the effect of the Q-heuristic in particular, to which both speaker and addressee are mutually oriented. To complete the picture, we will briefly consider the other two
heuristics, related to Gricean second maxim of Quantity and to the maxim of Manner respectively. To begin with, I-heuristic is formulated as the following principle (where the “I” notation stands for “Informativeness”, cf. Atlas and Levinson, 1981): “What is expressed simply is stereotypically exemplified”. By this heuristic we can explain a great number of interpretative tendencies such as:

(23) generality-narrowing: I need a friend => a sincere one;
(24) conditional perfection: if you help shovelling the ramp I’ll drive you home => if and only if you help shovelling the ramp I’ll drive you home;
(25) conjunction buttressing: Volodja turned the kettle on to make tea and the water started boiling => Volodja turned the kettle on to make tea and as a result/ causing/ and then the water started boiling;
(26) together implications: Carlo and Silvia had a baby => they had it together;
(27) local co-reference: I was at Barbara’s for Thanksgiving. She prepared a delicious cranberry-pie and the turkey was impressive! => she = Barbara;
(28) bridging inference: The dinner was superb. Lyn’s pumpkin soup was memorable => the soup was part of the dinner.

This is how M-heuristic, instead, is formulated: “What’s said in an abnormal way isn’t normal”. For example, using elaborate paraphrases such as “Angelika’s lips turned slightly upward” implicates that she didn’t exactly smile.

The above-mentioned heuristics are generally in force, and this gives them their communicative efficacy. The implicatures that are drawn following these principles are thus generalized, in the sense that they have the status of preferred interpretations and are derived by default. This claim is central to the purposes of our discussion on SIs, and will prove crucial in the comparison with alternative accounts and in the interpretation of experimental data.

---

2 In Levinson’s notation, the symbol => stands for implicates.
The status of default inferences recognised to GCIs cannot be separated from a second crucial property of these inferences that sets them apart from entailments, namely, their defeasibility. As already seen in the introductory section, implicatures can be cancelled if and when additional information to the contrary is provided in the context. This property is related to the logic notion of *nonmonotonicity*, which is predicated of an inference or argument when it may be defeated by the addition of further premises (versus deductive systems of reasoning, which are by definition monotonic or nondefeasible). For example, consider the asymmetry between the nonmonotonic behaviour of SIs, represented in (29), and the monotonic nature of entailments, which accounts for the ungrammaticality of example (30):

\[(29)\]  Yoshi prepared *three* Japanese traditional dishes and, in fact, a total of *four*

\[(30)\]  *Yoshi prepared three* Japanese traditional dishes and, in fact, a total of *two*

There are different nonmonotonic reasoning systems\(^3\) that could prove useful in understanding how GCIs are defeated, none of which alone would suffice in explaining implicature cancellation. Reducing GCIs generation to default logics can prove useful in some but not all cases of implicature generation/cancellation. For these reasons, Levinson reversed the perspective on the argument: instead of seeking to understand implicatures in the framework of

\(^3\) For example, consider the following syllogisms, all of which are instances of nonmonotonic reasoning systems, given that their conclusion is not valid when a certain (controversial) premise is added to the context:

\[\text{Induction} \quad \text{Abduction}\]
\[
P(a) \quad A(x) (P(x) \rightarrow Q(x)) \]
\[
Q(a) \quad Q(a) \]
\[
A(x) (P(x) \rightarrow Q(x)) \quad P(a) \]

\[\text{Practical Reasoning} \quad \text{Default Logics}\]
\[
A(x) (P(x) \rightarrow Q(x)) \quad A(x) ([P(x) \& M(Q(x))] \rightarrow Q(x)) \]
\[
\text{Goal } Q(a) \quad P(a) \]
\[
\text{Goal } P(a) \quad Q(a) \]
theories of default inferences he proposes to think of GCIs as the prototype of
default inferences. These inferences are actually generated under the mutual
assumption (of speaker and recipient) of tacit coordination through specific
heuristics such as the ones outlined above. They thus have the force of strong
presumptions, which, according to Levinson, cannot be assimilated to the two
already recognised layers of meaning, namely sentence-meaning and speaker-
meaning. Conversely, they belong to a broad third layer of meaning, the
presumptive-meaning, which lays midway between the other two and, he claims,
is irreducible to those.

The relevance of GCIs to linguist theory, thus the need of a systematic
theory of such phenomena, is also attested by Horn’s generalization of universal
patterns of constraints on lexicalization (cf. Horn, 1989). Horn’s argument is
based on the traditional square of opposition, in which the corners of the square
vary on Quantity between universals and particulars on the high-low axes and on
Quality between affirmations and negations on the right-left axes. The elements
that occupy the four corners stand in precise relation to the others, in particular:
the A and E corners are contraries, i.e. they cannot be simultaneously true (though
they may be simultaneously false); the I and O corners are sub-contraries, i.e. they
cannot be simultaneously false (though they may be simultaneously true); A and
O, and E and I respectively are contradictories, i.e. they cannot be simultaneously
true or false. All these relations are represented in the diagram below:

(31)

Square of opposition (applied to quantifiers)
Different logical operators can occupy the vertices of the square. If we take quantifiers as example, this is how they are distributed in the square: A = All, E = No, I = Some, O = Not all. As for the relations that hold vertically between corners, A entails I and E entails O. On the other hand, the nature of the sub-contrary relation, the one that holds horizontally between the I and O vertices of the square, has been controversial since the introduction of the square of opposition two and a half millennia ago. The debate was centred on the status of this relation, namely on the question whether this relation could be considered logical (with the I vertex implying the O one) or non-logical. The theory of GCIs provides a ready solution to the puzzle that would not be solved otherwise. Building on that theory, we can define the relationship between the vertices of the square as follows: the I corner (corresponding for example to some) carries a generalized scalar implicature that the other vertex O (not all) also holds. As we have seen, some strongly suggests not all and not all strongly suggests some. However, this relation cannot be a logical relation such as an entailment relation given that these inferences can be suspended: one can negate, without contradiction, the normally arising inference by saying some, in fact all or not all, indeed none. By the nature of SIs and the way negation reverts the scale, one can view the vertices of the square as forming the following scales: <A, I> and <E, O>. The assertion of I (some) will carry the implicature \(\neg A (\neg \text{all})\), which in turn corresponds to O (not all), while the assertion of O (not all) will carry the implicature \(\neg E (\neg \text{No})\), which in turn corresponds to I (some). These inferences are so regular and predictable that many philosophers over the centuries were misled into analysing these relations as semantic relations. This fact constitutes also an indication of the fact that it’s not easy to separate the content of GCIs from semantic content.

Moreover, the necessity of a theory of GCIs is also called for to explain an important empirical fact, once again observed by Horn (1989). Namely, the fact that there seems to be a universal lexicalisation constraint on the O corner of the square: taking English as an example (but the observation can be extended virtually to any language), there’s not a lexical item such as *nall to express not
all, or *nalways to express not always, or *nand to express not both (and the list can be extended to all forms occupying the O corner of the square of opposites). The theory of GCIs provides an immediate account of this universal lexicalisation constraint, given that what it is conveyed by a GCIs is generally not lexicalized. This fact builds on a Redundancy Constraint on lexicalization, of the following tenor: whenever a lexical item P carries a generalized implicature W, there will be no fully lexicalized counterpart P’ that lexicalises the content of W. For this reason, given that I items always implicate O items on the square, then there will be no lexicalization of the O items. The fact that the O items and not the I items lack the lexicalization (providing that the implication between these corners is mutual) is explained by a general tendency to avoid more complex solutions when a simpler alternative is available: given a choice between a positive and a negative term carrying the same informational load, the positive term (corresponding to the I vertex) will be chosen as the basis for the redundancy constraint.

The attempt, specifically declared by Levinson, to defend a theory of GCIs that posit them as default inferences is opposed to other accounts, Sperber & Wilson’s Relevance approach in particular, that reduce GCIs to nonce inference, i.e. one-off inferences to a specific speaker-meaning that arise in a particular occasion given the specific context in which the participants to the conversation happen to be. According to Levinson, a theory, like Relevance Theory, that lacks a notion of preferred default interpretations (such as the GCIs in his approach), is unable to account for the strength and generality of the pragmatic inferences associated with the logical operators that operate in natural languages, such as the SIs associated to the use of some and or that I have tested experimentally.

In this connection, to better understand the ongoing debate and before introducing the new perspective I will be assuming in my experiments, it’s useful to briefly give an account of the Relevance Theory approach, highlighting its major topics and the disagreement points with the Neo-Gricean approach we have just encountered.
2.2 Relevance Theory: the balance of cost and benefit against a default account of SIs

Relevance Theory constitutes an attempt to work out in detail an inferential model of communication as an alternative to a classical code model according to which a communicator encoded her intended message into a signal that is then decoded by the audience using an identical copy of the code. On the contrary, according to an inferential model, a communicator provides evidence of her intention to convey a certain meaning, which is in turn inferred by the audience on the basis of the evidence provided. Goal of inferential pragmatics is to explain how the audience gets to the intended speaker’s meaning on the basis of the evidence provided, through a process that employs more than a simple decoding of the signal. The backbone of the pragmatic theory formulated by Dan Sperber and Deirdre Wilson (1986 and 1995) is the Cognitive Principle of Relevance, which is formulated as follows: “Human cognition tends to be geared to the maximization of relevance”. Inferential communication takes places against this cognitive background. From this principle it follows that ostensive stimuli, designed to attract the audience’s attention, create expectations of relevance: if a communicator produces an ostensive stimulus, she encourages her audience to presume that it is relevant enough to be worth processing. This behaviour is not captured by the Gricean Cooperative Principle, it’s based on the Cognitive Principle of Relevance, together with a second Communicative Principle of Relevance, according to which “every ostensive stimulus conveys a presumption of its own optimal relevance”, that in turn can be spelled out as follows:

(32) Presumption of Optimal Relevance:
   a. the ostensive stimulus is relevant enough to be worth the audience’s processing effort;
   b. it is the most relevant one compatible with communicator’s abilities and preferences.
As one can note, the recourse to the notion of “processing effort” is central to this approach. As its exponents put it, “the greater the processing effort required, the less relevant the input will be”; on the other hand, “the greater the positive cognitive effects achieved by processing an input, the greater the relevance of the input for the individual at that time” [2004]. According to these assumptions, relevance may be assessed in terms of cognitive effects (or benefits) and processing effort (or cost). A cognitive effect is defined as positive when it bears a worthwhile difference to the individual’s representation of the world. For example, true conclusions, revisions, and, especially, contextual implications (i.e. conclusions deducible from input and context together but not from input or context alone) constitute positive cognitive effects. And cognitive effects, thus relevance of the input, can be scaled on various grades, according to which an input is picked out from a mass of competing stimuli because it is the MOST relevant (again, in term of costs and benefits) in that given situation.

In Relevance Theory, the comprehension procedure, i.e. the identification of the explicit content of an utterance, is seen as an inferential process: the hearer’s goal is that to construct a hypothesis about the speaker’s intended meaning that satisfies the presumption of optimal relevance conveyed by her utterance. The overall comprehension process involves three subtasks, which can be schematized as follows:

(33)

a. constructing an appropriate hypotheses about explicit content (i.e. explicatures, cf. below for a definition of this term) by means of a pragmatic enrichment process (like for example decoding, disambiguation, reference resolution, etc.);

b. constructing an appropriate hypotheses about the intended contextual assumptions (i.e. implicated premises);

c. constructing an appropriate hypotheses about the intended contextual implications (i.e. implicated conclusions).
These three subtasks should not be seen as sequentially ordered, but as parallel aspects of the on-line process of comprehension, which happens against a background of expectations that may be revised or elaborated as the utterance unfolds. Wilson and Sperber provide the following example, here abbreviated (cf. Wilson and Sperber, 2000: 615-718):

(34)

Peter: Did John pay back the money he owned you?
Mary: No. He forgot to go to the bank.

Using the relevance-theoretic comprehension procedure, Peter can get to an interpretation of Mary’s utterance along the following lines. First of all, by considering the presumption of relevance that Mary’s ostensive behaviour conveys and analysing the expectations of optimal relevance, he will solve the lexical ambiguities (choosing one of the possible meanings of the word “bank”, interpreting it as “bank_f” = financial institution) and interpret the indexical expressions in the clause (in this case, interpreting “he” as referring to John). Referring back to the subtasks outlined in (33) above, Peter will undergo the following steps in reasoning, identifying the implicated premises and conclusion of Mary’s utterance and its explicit content (and, possibly, deriving weak implicatures):

(35)

i. implicit premise:
   forgetting to go to the bank_f may make one unable to repay the money one owes;

ii. implicit conclusion:
   John was unable to repay Mary the money he owes because he forgot to go to the bank_f;

iii. explicature:
   John was unable to go to the bank_f;
iv. **possible (weak) implicature:**

John may repay Mary when he next goes to the bank.

What’s interesting to observe at this point, is that the use of the notion of implicature in Relevance Theoretic approach is different from the one encountered in the Gricean tradition. Here it denotes the set of implicit premises, conclusions and implicatures that pertain to an utterance, and which follow directly from the principle of relevance and the presumption of optimal relevance considered above. In the Gricean and Neo-Gricean account, instead, it denotes the content that is implicated by a speaker’s utterance, assuming that he is cooperative and obeying the maxims. Moreover, the notion of implicature in Relevance Theory is opposed to the notion of *explicature*, which denotes the explicit content of an utterance and which, in Gricean terms (cf. schema of meanings in (7) above), corresponds to the “what-is-said” branch (application field for semantics), as opposed to the “what-is-implicated” branch (application field for pragmatics), which itself divides into two more classes, corresponding to the PCI and the GCI respectively, the former dependent from context and the latter arising as default interpretations. Crucially, this distinction is absent in Relevance Theory, given that all “particularized” implicatures (in Gricean terms) arise in dependence of the situation (by considerations of relevance). Instead, all the implicatures that are called “generalized” (in Gricean terms), are assimilated in Relevance Theory to *explicatures* and derived from the linguistically encoded logical form by pragmatical enrichment. As a consequence, all the examples given in the preceding section as examples of GCIs (cf. examples 23-28) are classified in Relevance Theory as *explicatures*, i.e. as content considered being explicitly communicated and not simply inferred. Taking the case of scalar implicatures, which are of main interest throughout this dissertation, we can compare the two approaches by analysing the following scalar example (cf. Carston, 2000: 646):

(36) *Some* of the children were sick
What is communicated, by means of (36), is expressed below:

(37) *Some but not all* the children were sick

According to the Gricean and Neo-Gricean approach, we get to (37), which corresponds to “what is communicated”, by combining the semantic/literal content of (36), corresponding to “what is said” (i.e. what contributes truth-conditionally) reported in (38) and the SI corresponding to “what is implicated” reported in (39), by following the steps indicated in (16):

(38) *At least some (perhaps all)* of the children were sick
(39) *Not all* the children were sick

On the contrary, according to Relevance Theoretic approach, we gain an enrichment (or strengthening) of the encoded content of (36) as the one expressed in (37) by the direct application of the unique communicative principle of relevance that derives (37) as the *explicature* of (36). What in Levinson’s terms would be defined as Q or I enrichments deriving a GCI, in this formulation are not considered implicatures at all, but pragmatically derived aspects of what is said, i.e. explicit truth-conditional content of the utterance. Related to this, another difference between the two accounts emerges from the consideration of examples like (40)-(42) below:

(40) If *some* of the teeth have already been cut, the others will be cut shortly
(41) Lorenzo doesn’t have *six* teeth, he has *seven*!
(42) When he displays his seven teeth grin, Lorenzo isn’t *cute*, he’s *terrific*!

According to Levinson, the examples above are to be interpreted as cases of “pragmatic intrusion on truth-conditional content”, in which the SI associated
to the scalar terms (respectively part of the scales: <some, all>, <n, n+1>, <cute, terrific>) necessarily becomes part of the “what-is-said” content otherwise the sentence would contain a contradiction \(^4\) (in this sense, it “intrudes”). On the Relevance Theoretic approach, instead, the “some but not all” (like the “exactly six” and the “not terrific”) enrichments are not to be considered implicatures at all but part of the explicit content of the utterance, thus explicatures. According to this view, cases like (40)-(42) can be assimilated to the phenomenon of free enrichment in which pragmatics contributes to the proposition explicitly communicated by an utterance even in the absence of an overt linguistic element in the utterance (in this sense is “free”, out of linguistic control), like in the example below:

\[(43)\] **Hobart Lane, Nov. 13\(^{th}\), 6.50 a.m.:**

It’s snowing. [HERE and NOW]

The bracketed element is provided on pragmatic grounds alone. Without such an element, however, the proposition expressed would be too vague and general to yield sufficient cognitive effects. The process described above finds independent support from philosophers of language outside Relevance Theory that are convinced of its psychological reality (a.o. Recanati, 1993, 2001 and Bach, 1994, 2000).\(^5\) On the other hand, some others (e.g. Stanley, 2000 and 2002) deny a concept of free enrichment on the grounds that any contextually supplied

---

\(^4\) A different position is being held by Horn (1985/1989) who defends the pragmatic account of the scalar inferences by assimilating cases like (41) and (42) to a non logical use of negation (namely, metalinguistic negation). However, Horn’s account cannot explain all the “pragmatic intrusion” examples and for this reason Levinson departs from this view.

\(^5\) We can consider a third class between “what is said” and “what is implicated”. As observed by Bach (1994/2001), some aspects of speaker’s meaning are to be considered neither part of what is said nor part of what is implicated. According to his view, not all implicit components of an utterance are to be defined as implicatures. Consider the examples below, in which pragmatic enrichments are represented in brackets:

- I haven’t had breakfast \{today\}
- Rudy and Lilly are married \{together\}
- They had a baby and got married \{in this order\}

These enrichments are truth-conditionally relevant on the one hand but at the same time they cannot be part of “what is said”, given that they are cancellable without contradiction. In Bach’s terms, these are to be called implicitures, to be distinguished from the class of implicatures.
constituent that appears in the explicit content of an utterance must be articulated in its logical form (hence, multiplication of covert variables to be saturated at the level of LF). This same take is shared by those Neo-Griceans, such as Levinson and Horn, who consider those aspects of utterance meaning which are viewed by relevance-theorists like pragmatic components of explicatures (derived by a process of free enrichment) like GCIs. The conviction underlying this position is always the fact that natural language semantics is essentially truth-conditional, hence minimally propositional, so that any pragmatic process other than disambiguation and saturation is confined in the realm of implicature. We won’t pursue further the debate on this argument that, though undoubtedly interesting, would push us too far from our current purposes. What it’s important to observe, to the purposes of our discussion, are the different theoretical assumptions made and the different positions held by these alternative approaches in tackling the same phenomenon. In this perspective, we will now turn to a third type of approach, the one that I have previously defined as “grammar-driven”. This approach departs from the two theories just considered in many respect, though reconciling some aspects of both in its final architecture, as we shall see soon.

2.3 A grammar-driven approach: seeking for spontaneous logicality

In this section, I will focus on the work put forward by Gennaro Chierchia since 1999 (considering its final version, cf. Chierchia, 2004) trying to implement the original formulation of the theory outlined there with some recent developments, that he presented in a cycle of lessons he held together with Danny Fox at Harvard and MIT in spring semester 2006 and are currently in press.

I will start my discussion from the original paper, in which Chierchia addresses the issue of SIs and their connection with polarity phenomena in order to better understand the grammatical mechanisms that underlie these twin (in many respect) phenomena, and the way in which syntax, semantics and pragmatics interact with each other. Taking the standard Gricean and Neo-Gricean

---

6 I want to warmly thank Danny Fox for kindly (and quickly!) sending me all the class material and of course Gennaro Chierchia for insightful (overseas) discussion over it.
accounts of SI presented above in section 1 and 2.1 as starting points, Chierchia presents some data that cast some doubts on the strictly modular view that those accounts assume. As we have seen, the dominant view maintains that implicature are computed “globally”, i.e. after the semantics of the root sentence has been computed and the truth-conditional content derived. However, this view cannot account for example (44), directly readapted from Chierchia:

(44) Lorenzo believes that some books are good to eat

If we keep the globalist account, then sentence (44) would have (45) as its alternative (given the scale <some, all>) and (46) as its (root) implicature:

(45) Lorenzo believes that all books are good to eat
(46) Lorenzo doesn’t believe that all books are good to eat

However, (46) is too weak to express the meaning conveyed by (44), given that what (46) expresses is the following: it is compatible with Lorenzo’s belief that not every book is good to eat. What sentence (44) seems to convey instead is the stronger assertion reported in (47):

(47) Lorenzo believes that not all books are good to eat

Sentence (47) better captures the meaning expressed in (44), which better describes the fact that there is a (well defined) set of books (namely, his own books, let’s call it \(B_L\)) that Lorenzo never eats, while, of course, he eats all the other books in the bookcase. Of this defined set of books \(B_L\) one can assert that Lorenzo thinks \(\neg B_L\) are not good to eat. This strong assertion cannot be derived from (46). Analogously, the globalist account fails to derive the right implications from sentences like (48), in which sentential connectives interact:
(48) Lorenzo is either messing up the bookcase or pushing some of my CDs in the VHS recorder slot

If we follow the steps of a globalist account, the negation (associated to the SI related to some) would wind up in the wrong place, taking scope over the whole disjunction (as in (49)), while we would want it to negate only the NP in the second disjunct of the alternative, as in (50):

(49) It is not the case that [Lorenzo is either messing up the bookcase or pushing all of my CDs in the VHS recorder slot]
(50) Lorenzo is [either messing up the bookcase] or [he is pushing NOT all of my CDs in the VHS recorder slot]

The implicature derived in (49) is actually too strong given that, in connection with (48), it would derive the unwarranted (and, knowing Lorenzo well, utopian) conclusion in (51):

(51) Lorenzo is not messing up the bookcase

The way out of this impasse is to be found in a different approach to SI computation. The guiding idea of such a proposal is that implicatures are introduced locally, as soon as their trigger (the scalar term) is encountered, as exemplified in the syntactic representation of sentence (48) that is reported below (imagining a bottom up semantic computation):
In the same spirit as Krifka’s proposal (cf. section 1 above), Chierchia assumes that any expression $\alpha$ comes with its plain value $\llbracket \alpha \rrbracket$ and its strengthened (scalar) value $\llbracket \alpha \rrbracket^S$, recursively assigned by the grammar. Such a scalar value is computed by exploiting the stronger alternatives to the plain value, which constitute $\llbracket \alpha \rrbracket^{S-ALT}$, i.e. the set of the stronger alternatives that are lexically determined given the scalar lexical entry. For example, in (53) the plain value, the set of alternatives and the scalar value associated to the lexical entry *some* are represented:

\[
\begin{align*}
\llbracket \text{some} \rrbracket &= \lambda P \lambda Q \exists x [P(x) \land Q(x)] \\
\llbracket \text{some} \rrbracket^{S-ALT} &= \{ \text{some} < \text{many} < \text{most} < \text{all} \} \\
\llbracket \text{some} \rrbracket^S &= \lambda P \lambda Q \exists x [P(x) \land Q(x)] \land \neg \forall x [P(x) \rightarrow Q(x)]
\end{align*}
\]

The negation of the strongest alternative in the set, that constitutes part of the strengthened value of $\alpha$, is the result of the application of Krifka’s general pragmatic rule: “if grammar allows for a stronger or a weaker interpretation of a structure, choose the one that results in the stronger interpretation of the sentence,
if consistent with general background assumptions!” [Krifka, 1996: 12]. In Chierchia’s proposal, as we have mentioned, the landing site of implicatures is different from the globalist account explored so far, given that the implicature is not computed at the root level but it is added locally, as soon as the scalar term is encountered. This proposal has interesting consequences for the overall architecture of grammar at the syntax-semantic-pragmatic interfaces. In particular, semantics and pragmatics are no more to be intended as separate modules, given that “pragmatic computations and grammar-driven one are “interspersed”. Implicatures are not computed after truth-conditions of (root) sentences have been figured out; they are computed phrase by phrase in tandem with truth-conditions (or whatever compositional semantics computes).” [Chierchia, 2004] This compositional, grammar-driven treatment of SIs seems actually the one that gets us to a better understanding of the phenomenon, following a path already successfully beaten for presupposition.\(^7\) As we will see extensively in the chapter dedicated to experimental works on SI (cf. Chapter 2), one can derive clear-cut predictions about the process of SI computation in Chierchia’s theory, given his own assumptions that distinctly sets him apart from the Relevant Theory approach, at least with what concerns the cost of SI computations: “One would expect that speakers will use the values grammar provides them with cooperatively. This means that they will tend to use the strongest interpretation (consistent with the context) for which they have evidence. So the default interpretation in a concrete communicative situation will be the one provided by strengthened values” [Chierchia, 2004].

In the localist account of SI computation just presented, once implicatures are introduced (by default), they are projected upwards and filtered out or adjusted. At each step in the derivation, a Strength Condition must be checked locally, in which the potential strong values and the plain value are compared so that the strong value never gets weaker than the plain one. According to Chierchia, the SI projection algorithm is actually affected by certain kind of

\(^7\) Historically, presuppositions were at first considered a purely pragmatic phenomenon, not amenable to a grammar driven compositional treatment (cf. e.g. Kempson, 1975).
contexts, as captured in the generalization reported below, to which our discussion will now turn:

(54)  
Generalization on SIs:  
(Ordinary) scalar implicatures are suspended in the contexts that license *any* (as a Neg Pol or as Free Choice Item)  
[Chierchia, 2004]

The relation between SIs and Negative Polarity Items (NPIs henceforth) remains a central topic in Chierchia’s work, also in its more recent developments. This connection is actually the most intriguing part of the theory, given that it allows us to define the grammatical mechanisms that are common to these, apparently distant, phenomena. Moreover, pushing the question to reach its core significance, “the codependence/interaction of SIs with NPI/FCIs\(^8\) reveals important aspects of the spontaneous logicality of language” [Chierchia & Fox, class handout 1:1], a topic that is central to the work I’m presenting here. I will not explore the details of Chierchia’s account on the isomorphism between the semantics of SIs and that of *any*, since this would push us too far from the purposes of the current discussion. I refer the reader to his most recent work on the matter. To not shift our attention from SIs, what we will therefore consider here will be the distributional facts that pertain to SI computation and/or cancellation, according to the generalization reported in (54), only mentioning, though not exploring further, the interactions with *any* licensing.

The generalization of the distributional properties of SIs rests on the contrast between examples (55)-(a-b-c) and examples (56)-(a-b-c) below (cf. Chierchia, 2004 for a more complete set of examples and references):

\(^8\) For those not familiar with the topic in question, NPI is the abbreviation commonly used for Negative Polarity Item, while FCI stands for Free Choice Item. These notations will be adopted henceforth.
Chapter 1

(55)
(a) No toy that makes animal cries or jingles is silent when Lorenzo’s around;
(b) Every toy that makes animal cries or jingles is switched on when Lorenzo’s around;
(c) If a toy makes animal cries or jingles, then Lorenzo will love it!

(56)
(a) Some toy that makes animal cries or jingles is always switched on when Lorenzo’s around;
(b) Every toy in Lorenzo’s basket makes animal cries or jingles;
(c) If you switch on a toy in Lorenzo’s basket, then you’ll hear animal cries or jingles.

The scalar term or in examples (55) appears, respectively, in the following contexts: the restrictor (or 1st argument) of the negative quantifier; the restrictor (or 1st argument) of the universal quantifier; the antecedent of conditional. In examples (56), instead, it appears, respectively, in the following contexts: the restrictor (or 1st argument) of the existential quantifier; the nuclear scope (or 2nd argument) of the universal quantifier; the consequent of conditional. Three crucial observations are in order here, two based on empirical distributional facts and a third based on theoretical considerations. First of all, in the first set of examples the SI associated to or seems cancelled or suspended. In fact, the interpretation normally derived from sentences (55)-a-b-c is compatible with the existence of a toy, such as Old Mac Donald tractor, which both makes animal cries and jingles, that Lorenzo loves and is always switched on. On the contrary, by hearing sentences (56)-a-b-c the SI associated to or seems likely to arise: namely, I expect all Lorenzo’s toys to make animal cries or jingles but presumably not both (the existence of a toy such as Old Mac Donald tractor is not expected, and it would force us to subsequently accommodate, by removing the implicature derived to meet context requirements). In second place, another empirical consideration on
the examples above is due: while sentences in (55) license any (be it free-choice or NPI), this is not so for sentences in (56), in which the presence of any renders the sentence ungrammatical. In this respect, consider the following examples:

(57)

(a) No toy makes any sound: the pile must be flat;
(b) Every toy that makes any sound is switched on when Lorenzo’s around;
(c) If a toy makes any sound then Lorenzo will love it!

(58)

(a) #Some toy that makes any sound is always switched on when Lorenzo’s around;
(b) #Every toy in Lorenzo’s basket makes any sound;
(c) If you switch on a toy in Lorenzo’s basket, then you’ll hear any sound.

In third place, in a theoretical perspective, these two sets of examples display different logical properties. In particular, the property that characterize examples (55) and distinguish them from the other set is the logical property of Downward Entailingness, which is defined as follows: a context (or function) is Downward Entailing (DE henceforth) when it allows inferences from a set to its subsets (i.e., iff \( f(A) \models f(B) \) whenever \( B \subseteq A \)). For example, the inference from sentence (59) to (60) is valid because the negative quantifier No represents a DE context:

(59) No student is reading a book
(60) No student is reading L’étranger by Camus

\[ A \models f(B) \text{ whenever } B \subseteq A \]
Conversely, a context is defined Upward Entailing when the inferences go the other way round, from the subsets to their superset (i.e., iff $f(B) \Rightarrow f(A)$ whenever $B \subseteq A$), as attested in the inference below, where the existential quantifier is used in the examples:

(61) Some student is reading a book
(62) Some student is reading L’étranger by Camus

Differently from examples (55), all the contexts in which the scalar term appeared in (56) above (in which, as we saw, any was not licensed) are actually Upward Entailing contexts.

Taking stock, we have observed that sentences (55) share the following properties (both empirical and theoretical):

(63)
- the SIs associated to the scalar term is absent;
- these contexts license any;
- the contexts in which the scalar term appears are Downward Entailing.

These are to be contrasted with the properties that characterize examples (56) instead, which are summarized below:

(64)
- the SIs associated to the scalar term is normally present;
- these contexts don’t license any;
- the contexts in which the scalar term appears are Upward Entailing.

Crucially, the contrast observed between the two sets of examples holds cross-linguistically. Chierchia’s generalization in (54) above actually captures the
connection between the empirical facts observed: any-licensing and SI cancellation. The next step is theoretical: providing a characterization of the logic properties of the contexts to explain the empirical distribution of the two phenomena. This is done by appealing to the notion of DE-ness, which, though reductive in some respects, at present gives us the best characterization of the context/function that removes/suspends implicatures on the one hand and licenses any on the other. Precisely, in Downward Entailing contexts the direction of entailments is reverse, so that what constitutes the stronger value of an expression \(\alpha\) in a NON-DE context becomes the weakest one when inserted in a DE context. To illustrate this point, consider a scalar sentence like (65) which gets inserted under a negative operator in (66):

\begin{align*}
(65) & \text{ Under the prune-tree, Lorenzo usually tears the leaves or the flowers off the branches} \\
(66) & \text{ Under the prune-tree, Lorenzo never tears the leaves or the flowers off the branches}
\end{align*}

The strongest meaning associated to sentence (65) is the one in which or is assigned the exclusive interpretation, according to which Lorenzo usually either tears the leaves or the flowers off the branches of the prune-tree, but not both. This interpretation renders the sentence more informative, thus true in a smaller sets of circumstances compared to the case in which or is assigned the inclusive interpretation (corresponding to its logic, basic meaning according to which Lorenzo usually tears the leaves or the flowers off the branches of the prune-tree, or even/possibly both). This happens when or is embedded in positive, NON-DE contexts, such as the one represented in sentence (65). Under negation, thus in a DE context such as the one represent in sentence (66), the entailment relation is reversed: in DE contexts, the inclusive interpretation of or becomes the most informative. A comparison across the different truth-conditions is summarized in the following chart:
(67)

Truth-conditions and entailment relations in positive (NON-DE) and negative (DE) environments

<table>
<thead>
<tr>
<th></th>
<th></th>
<th>L or inc</th>
<th>L or exc</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>1</td>
<td>1</td>
<td>0</td>
</tr>
<tr>
<td>1</td>
<td>0</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>0</td>
<td>1</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
</tbody>
</table>

NON-DE
(or exc ⊆ or inc)

Legend:

\[ L = 1^{\text{st}} \text{ disjunct (= tearing off leaves); } F = 2^{\text{nd}} \text{ disjunct (=tearing off flowers); } 1 \text{ and } 0\]

stands for TRUE and FALSE respectively; \( \subseteq \) represents the set-inclusion relation, where \( A \subseteq B \) is to be read as “ \( A \) is a subset of \( B \)”; the arrows indicate the direction of the entailment relation. The shadowed cells represent the situations in which the sentence is “false”: the more these situations are, the more informative the sentence is.

As it is evident from the schema above, the set of circumstances which renders the sentence true in positive contexts is narrower in the exclusive than the inclusive interpretation of or (two instead of three situations). Thus, the exclusive or is more informative than the inclusive if the context in NON-DE. On the contrary, when the sentence is embedded under negative operators like not or never, i.e. in a DE context, the set of circumstances that renders the sentence true is narrower when or is assigned the inclusive interpretation, thus this is the most informative interpretation in such contexts. Given the conversational tendency to be maximally informative, we will always tend to interpret or accordingly, as
exemplified in the examples above: i.e. assigning *or* an *inclusive* interpretation in contexts like those in (55) (i.e. in DE contexts) and assigning it the *exclusive* interpretation in NON-DE contexts like those represented by examples (56). These observations will reveal crucial to understand the mechanisms that govern implicature derivation/suspension and the link with the empirical distribution of NPI/FCIs like *any* to which we now turn.

Let’s consider some crucial aspects of the semantics of NPIs such as *any*, the ones that are more closely related to the topic of SIs under discussion. Consider that my intent is only to give a hint of how Chierchia argues for a parallel treatment of these two phenomena. For this reason, semantics of NPIs will be presented here in a very simplified form. Let’s say that quantifiers are normally associated with a certain quantificational domain. For example, when I claim that “Every book has missing pages” I typically don’t mean that every book in the world has missing pages, but, rather, that every book in the reach of Lorenzo has, i.e. every book which belongs to a specific domain $D_L$ (where $D_L$ corresponds, for example, to the bookcase in my living-room). Appealing to this notion of domain of quantification, Kadmon and Landman (1993) proposed that the NPI *any* is to be treated as an existential quantifier that widens the domain of quantification that would otherwise be assumed. As for SIs, however, the operation of domain widening must lead to a global strengthening of the assertion otherwise it results in ungrammaticality. However, inserting a NPI in a positive context would lead to a loss of information. For this reason, the use of *any* in a positive sentence like “Lorenzo is eating *any* book” is ungrammatical: it would extend the domain $D_L$ to a wider domain $D_W$ containing, ideally, all the books in the world. This move would result in a global weakening of the assertion: in positive contexts, the more specifically constrained the quantificational domain of an existential will be, the more informative a sentence we get. The opposite holds for negative contexts. This fact explains why items such as *any* are only grammatical in negative environments, thus their status of NPIs and their distribution that patterns SIs cancellation. As we have seen in our examples, *any* is actually licensed only in
negative, DE, contexts like (55)-a-b-c but it’s not grammatical if inserted in positive, NON-DE, contexts such as (56)-a-b-c.

In Chierchia’s most recent work (cf. Chierchia, 2006), a unified account of polarity sensitive items, like any and scalar implicatures, is being considered. In this new formulation, a feature \( \sigma \) (mnemonic for “strong”) is introduced as regulating the activation of scalar alternatives associated to scalar and negative polarity items. This feature can be assigned two values: \([\pm \sigma]\). Selecting \([+\sigma]\) results in the activation of the scalar alternatives (ALTs henceforth); selecting \([-\sigma]\) results in the selection of the plain meaning in which ALTs are not active. The crucial point is that, whenever the feature \([+\sigma]\) is encountered, then the constraint on strengthening applies and an exhaustivization operator \(O\) (which has a meaning akin to that of only) must be used.\(^{10}\) For our purposes, it suffices saying that the result of this mandatory operation always results in the selection of the strongest – most informative – interpretation of the sentence containing the scalar item. Very informally, \(O\) applied to a sentence of the form “Lorenzo did (A or B)” in which the ALTs are active will results in the derivation of the scalar implicature associated to or: Lorenzo did \(O\) (A or\([+\sigma]\) B) = Lorenzo did only (A or B) = Lorenzo did only (A or B) and not (A and B).

The choice between the binary values of the feature \(\sigma\) is optional in case of scalar terms (and the analysis of which factors determine/influence this choice goes well beyond the purposes of the present discussion, so this questions won’t be addressed here); the selection of its positive value (i.e. \([+\sigma]\)) is instead mandatory in case of NPIs. Let’s explain why. NPIs are defined as scalar Determiners (i.e. weakest elements on the quantifier scale) with active domain variants (in the feature notation, \([+D]\)). Remember that the classical analysis of NPIs and FCIs (cf. for example Kadmon and Landman, 1993) assumes that their function is that to widen the domain of quantification that would otherwise be assumed. As we have seen, this domain extension (thus the use of NPIs), is limited to negative contexts because the operation of extending the domain in

\(^{10}\) We are not entering here the formal details of the application of \(O\). Very briefly, this is how \(Only\) operator applies: \(\text{ONLY}(C, p) = p \land \forall q \in C \ [q \rightarrow p \subseteq q]\).
positive contexts would lead to a loss of information. For this reason, NPIs are assigned the feature \([+\sigma, +D]\), in which the D feature activates domain variants and the \(\sigma\) feature activates scalar alternatives. Crucially, this feature can only be set on its positive pole because, differently from SIs, the use of an NPI in positive contexts leads to ungrammaticality of the sentence, from which we can’t recover.

On the contrary, scalar items can be assigned the feature \([\pm \sigma]\), which signals the fact that the alternatives on the scale can or cannot be active (i.e. the scalar term is assigned the feature \([+\sigma]\) or \([-\sigma]\) respectively). This does not amount to say, however, that scalar items are lexically ambiguous, for example by saying that there are two lexical entries associated to some: some\(_1\) which means “at least one” and some\(_2\) which means “some but not all”. Beyond classical arguments based on a principle of economy, there is a simple argument against the lexical ambiguity hypotheses: this ambiguity cannot be extended to all elements on a scale given that stronger elements don’t have an alternative to be negated. For example, there isn’t a lexical alternative to (every\(_D\)\(x = \) all the individuals in D) which excludes, for example most or many or some individuals in D. This would lead us to exclude that we can draw implicatures from stronger elements of a scale even if, in fact, implicatures arise from these elements in negative contexts. For example, let’s consider every in a negative statement like (68):

(68) Lorenzo cannot mimic every animal

This sentence, though being compatible with a situation in which there is no animal that Lorenzo can mimic, it would normally convey (69), which is more informative than its negative counterpart:

(69) There is some animal which Lorenzo can mimic

This implicature couldn’t be accounted for under an ambiguity thesis, because there being no lexical alternative to stronger elements in the scale. Resorting to the option of feature selection proposed by Chierchia, we can instead
say that every, like all scalar items, can select the feature [+\(\sigma\)], thus rendering the other elements on the scale relevant in certain situations and deriving implicatures like (69) to render the sentence more informative.

Once the feature has been selected, the operation of adding implicatures, as we have discussed at the beginning of this section (cf. examples (49)-(50)), can apply at the level of the root sentence or in embedded position. This leads to results that differ in their informational strength. Consider the following derivation, applied to a scalar term embedded in a NON-DE context:

(70)
Sentence S:

*If there is a C on the back, then there is an A or a B on the front*

No implicature:

(i) (C on the back \(\rightarrow\) (A or B or possibly both) on the front)

Root implicature:

(ii) O (C on the back \(\rightarrow\) (A or B) on the front)

(iii) (C on the back \(\rightarrow\) (A or B)) AND NOT (C on the back \(\rightarrow\) (A and B) on the front)

Embedded implicature:

(iv) (C on the back \(\rightarrow\) O (A or B) on the front)

(v) (C on the back \(\rightarrow\) (A or B) AND NOT (A and B) on the front)

In particular, the output of the implicature computed at the global level (i.e. root implicature) is compatible with a situation in which there is a C on the back without there being both an A and a B on the front. This reflects the fact that the speaker is considered ignorant as to whether the stronger alternatives hold. On the contrary, assertion (v) is stronger than the root implicature (iii) above, in fact (v) entails (iii). This operation resembles a lot the epistemic step, according to which the hearer often concludes that the speakers believes that the stronger alternatives to the scalar terms she used do NOT hold (inferring, for example,
from the use of *some* that the speaker believes that *not all* holds, not simply that she *doesn’t know* if *all* does hold). The same reasoning, applied to a scalar term embedded in a DE context would lead to the following derivations:

(71)
Sentence S’:

*If there is an A or a B on the front, then there is a C on the back*

No implicature:

(i)  (A or B or possibly both) on the front $\rightarrow$ C on the back

Root implicature:

(ii)  $O (\text{(A or B) on the front } \rightarrow C \text{ on the back})$

(iii)  $((\text{A or B)})$ on the front $\rightarrow$ C on the back

$\text{AND NOT } ((\text{A and B) on the front } \rightarrow C \text{ on the back})$

Embedded implicature:

(iv)  $(O (\text{A or B) on the front } \rightarrow C \text{ on the back})$

(v)  $((\text{A or B}) \text{ AND NOT } (\text{A and B) on the front } \rightarrow C \text{ on the back})$

Again, the root implicature in (iii) signals that the speaker is considered ignorant as to whether the inference holds in case there is both an A and a B on the front. By embedding the implicature as in (v), instead, the hearer can conclude that the speaker believes that, whenever there is an A and a B on the front then there isn’t a C on the back. This conclusion, however, is unwarranted in most cases when the scalar term is embedded in a DE context because it leads to a weakening of the assertion. On the distribution of SIs with respect to the monotonicity properties of the context, these are the generalization reported in Chierchia-Fox, 2006:
(72) **Distributional generalizations**

(a) The exhaustive interpretation\(^{11}\) of a scalar term is easier in a NON-DE than in a DE context
\[
\rightarrow \text{the exclusive interpretation of } or \text{ is easier in the consequent than in the antecedent of a conditional} \text{ (compare (c)-sentences (55) and (56) above);}
\]

(b) Having an implicature embedded in a DE context is way harder than having it embedded in a NON-DE context
\[
\rightarrow (71-v) \text{ is harder than (70-v);}
\]

(c) The flip between having an implicature and not having it is relatively easy in non DE contexts
\[
\rightarrow \text{flipping between (70-i) and (70-iii/v) is relatively easy;}
\]

(d) The flip between having an implicature and not having it is hard in a DE context
\[
\rightarrow \text{flipping between (71-v) and (71-i/iii) is hard;}
\]

(e) All of the above generalizations ought to be restated replacing ‘DE’ with ‘\textit{any}-licensing’.

[adapted from class Handout 1:4]

The fact that SI computation/suspension (and NPI licensing) is affected by the kind of (syntactic) context in which the scalar term appears is related to the strengthening issue: adding an implicature in a NON-DE context strengthens, adding it in a DE context leads to weakening. This can be formalized as follows, using notations already introduced above:

(73)

(i) \[ [C \Rightarrow O (A \text{ or } B)] \subseteq [C \Rightarrow (A \text{ or } B)] \]

(ii) \[ [(A \text{ or } B) \Rightarrow C] \subseteq [(O (A \text{ or } B)) \Rightarrow C] \]

\(^{11}\) Via application of \(O\).
Crucially, both for NPI licensing and SI computation, what’s important is that the result of the operation must be stronger than the starting point. This constraint explains why these phenomena are sensitive to the logical properties of the contexts in which they appear, such as monotonicity properties. To conclude, a difference between the two phenomena is found in the fact that, differently from NPIs, SIs are subject to a so called “flip-flop effect”: being inserted and removed in a “piston-like” fashion (given that scalar requirements are checked at each step in the derivation), they can be added and then cancelled across DE operators or whenever contextual requirements are incompatible with the adding of the implicature. In on-line experiments on processing, backtracking should signal this effect.

3. Conclusive remarks

In this introductory chapter, we have analyzed three theoretical approaches to the phenomena of implicatures in general and of scalar implicatures in particular. Starting from the analysis of the classical work on implicatures proposed by Grice, we have in turn discussed the theory of Generalized Scalar Implicature formulated by Stephen Levinson, the Relevance Theory advocated by Sperber and Wilson and the grammar-driven approach proposed by Gennaro Chierchia.

All these authors put forward a different understanding of the phenomena, each proposing a different analysis of how we actually derive pragmatic inferences and how the different modules subsuming sentence comprehension interact. The major topics of divergence and discussion that we encountered can be summarized as follows:

(74)
- global vs. local computation of alternatives;
- default vs. context-driven implicature derivation;
- grammatical vs. extra-linguistic mechanisms of SI derivation;
- sequential vs. parallel intervention of the different modules;
- implicatures vs. explicatures;
- conversational maxims vs. principle of relevance.

The comparison among these different approaches that we have drawn in this chapter rests on purely theoretical grounds. Our intent does not exhaust itself here, though. The contrast amongst these theories is actually better captured on experimental grounds. For this reason, in Chapter 2 I will present some experiments that were conducted on adults testing their interpretation of sentences containing scalar terms and in Chapter 5 I will shift the discussion on acquisition presenting some experiments on pragmatic inferences in children.
CHAPTER 2

SCALAR IMPLICATURE COMPUTATION AND LOGICALITY:
THE CASE OF ADULTS

In the previous chapter, we introduced the phenomenon of scalar implicature and different approaches to it developing our discussion in a purely theoretical perspective. In this chapter, I’m going to shift the focus on experimental investigation, presenting two different studies that were conducted on adults using different experimental techniques. The general aim of these studies was to investigate the interpretation that subjects assign to one particular scalar item, or, manipulating the semantic/syntactic context in which it appears. What seems to emerge from these studies is that adults are extremely sensitive to logical distinctions (such as Downward Entailment) and logic notions (such as entailment relations), even in performing tasks which - apparently - have much more to share with pragmatic rules and principles than with logic. In the end, adults that participated in these experiments show a high rate of spontaneous logicality, which, I am convinced, they would themselves even doubt to master which such ease. The theoretical framework that inspired all the experimental works that will be presented here is the grammar-driven approach proposed by Gennaro Chierchia that I introduced in the previous chapter (cf. Chapter 1, section 2.3).
The predictions derived from this approach will be compared on experimental grounds with the predictions that ensue from the alternative approaches to SI there discussed.

1. Monotonicity and SI computation: an off-line study on or

In this section I will present a study in which the interpretation of the scalar item or was tested in different contexts displaying different entailment patterns: in the first and second arguments of the universal quantifier every, of the negative quantifier no, and of the Italian indefinite article un. These quantifiers differ in the entailment properties of their arguments, that can either be Upward (↑) or Downward (↓) Entailing (cf. Chapter 1, for a definition of these terms). Following Chierchia (2006), the lexical entry associated to scalar terms such as or can select the feature [±σ] depending on the fact that its scalar alternatives are activated or not. The activation of ALTs is actually optional but, in case ALTs becomes active, then the strongest interpretation of the scalar term must be selected. For this reason, if the scalar term is assigned the feature [+σ] then the mandatory application of the exhaustivization operator O leads to different interpretations of the scalar term depending on the entailing properties of the context in which it is embedded, as explicated in the chart below:

<table>
<thead>
<tr>
<th>Type of Quantifier</th>
<th>Monotonicity 1st argument</th>
<th>if [+σ]</th>
<th>Monotonicity 2nd argument</th>
<th>if [+σ]</th>
</tr>
</thead>
<tbody>
<tr>
<td>Every</td>
<td>↓</td>
<td>inclusive</td>
<td>↑</td>
<td>exclusive</td>
</tr>
<tr>
<td>No</td>
<td>↓</td>
<td>inclusive</td>
<td>↓</td>
<td>inclusive</td>
</tr>
<tr>
<td>Un</td>
<td>↑</td>
<td>exclusive</td>
<td>↑</td>
<td>exclusive</td>
</tr>
</tbody>
</table>

Table 1. Chierchia’s system. Predicted interpretation of or (if ALTs are active) depending on the monotonicity properties of the two arguments of each quantifier tested.

According to the chart above, whenever the scalar term or is embedded in a DE context, such as the 1st argument of the universal quantifier or the 1st and 2nd
arguments of the negative quantifier, then it will get the inclusive interpretation given that, in DE contexts, this is the strongest – most informative – interpretation. On the contrary, when it is embedded in NON-DE contexts, such as the 2nd argument of the universal quantifier, or the 1st and 2nd arguments of the Italian indefinite article un, it’ll get the exclusive interpretation given that this is the strongest interpretation in such contexts. These predictions functioned as the guiding lines for the design of the experiment that I’m going to present in detail in the next section.

1.1 Experiment 1: a Truth Value Judgment Task

An experiment was conducted on a group of Italian students employing a methodology normally used to test children, known in the literature as the Truth Value Judgment Task (Crain and Thornton, 1998). This methodology is considered particularly apt to test children, given that the task and the experiment requirements are of immediate understanding. For this reason it can be successfully employed to test adults too. The peculiarity of this task is that subjects are shown a series of stories acted out with props and toys. At the end of each story they are asked to evaluate a statement uttered by a puppet with the role of providing a good one-sentence-summary of what has happened in the story. I actually used a modified version of this task in which the critical sentences were in fact predictions made by the puppet Merlin the Magician, who had to guess what had happened in the story, instead of simply describing the facts. This modification was considered necessary to render the use of disjunction more plausible: uttering a disjunctive clause such as A or B seems to be more felicitous when the speaker ignores – thus guess – which between the disjuncts (A or B or possibly both) is true.

1.1.1 Subjects

A total of 75 subjects participated in this study (13 of which were discarded from further analysis because they didn’t complete the task). Subjects
were recruited in a high school near Bergamo (Lombardia) and were students attending their fourth year (mean age: 18.5 years).

1.1.2 Material and Procedure

Three different quantifiers were tested in this experiment: the universal quantifier *every* (it. *ogni*), the negative quantifier *no* (it. *nessuno*), the Italian indefinite article *un*. Let’s use the abbreviation Q (for “quantifier”) to name this condition, which varies on three levels depending on the quantifier tested (let’s call these Q1, Q2 and Q3 for *every*, *no* and *un* respectively). Moreover, each quantifier was presented in two conditions in which the position of *or* was varied, appearing in the restrictor or in the nuclear scope of the quantifier. Let’s use A (for “argument”) to classify this condition, which varies on two levels: A1 and A2, corresponding to 1st and 2nd argument of the quantifiers. As we have discussed above, depending on the type of quantifier, the two arguments have different entailing properties that may affect the interpretation of the scalar term embedded (namely, they can be DE or NON-DE, as reported in Table 1 above). Thus, considering the interaction of conditions Q and A, the effect of the monotonicity properties of the context on the interpretation of the scalar term they embed is revealed.

A total of twelve stories were created, comprising six critical sentences containing *or* (treated as 3(Q)×2(A) within subject factor) and six fillers and controls (cf. Appendix A). The critical sentences were actually the predictions that the puppet Merlin the Magician made at the end of each story in the attempt of guessing what has happened. Each scenario was the same for all the stories: first Merlin the Magician came on the screen and the situation (for example: a party with food) and all the characters involved in the story (for example: four boys invited at the party) were presented to him. Then Merlin was blindfolded and his ears were covered so that he could not hear nor see what happened. In the scenario that we are considering as example, this is how the story goes:
Chapter 2

(1)

*Four boys were at a party with lots of interesting food to eat: sandwiches, cookies, candies and fruits. There were also chairs to sit down and toys. The boys were hungry and they decided to take something to eat: in turn, three of them took a sandwich AND a cookie and sat down on a chair to eat their snacks. The fourth boy took a pear and a candy instead and ate them while playing with a toy-car.*

At the end of the story, Merlin is asked to guess what has happened and, in the case of our example, this is the sentence he ends up with (Condition Q1-A1):

(2)  *Every boy that took a sandwich OR a cookie sat down*

Crucially, Merlin’s predictions always contained disjunction, either in the 1st (Condition A1) or in the 2nd (Condition A2) argument of the quantifiers. To see also an example of Condition A2, let’s consider the story in (3), correlated to Merlin’s critical sentence in (4):

(3)  *

*Four girls went to the dog pound to visit the poor homeless animals. All the girls caressed and played with the pets. When it was time to go home, one of the girls decided that she could not leave the pets there. So she took a dog AND a cat home.*

This is Merlin’s prediction (Condition Q3-A2):

(4)  *A girl took a dog OR a cat home*
In all cases, given the outcome of the stories (crucially both disjuncts are always realized), subjects’ acceptance of the critical statements is only compatible with an inclusive interpretation of *or*.

The short stories were previously acted out using props and toys, recorded with a digital camera and then transferred on a VHS to be played on a TV screen. Subjects were divided in small groups and then were shown a video in a quiet room in their school. They were instructed to watch the stories and then judge if Merlin’s prediction was good or bad with respect to what happened in the story. After each story, they had to record their judgments on Merlin’s sentence on a score sheet (on which the sentence was repeated for convenience) using the scale 1 to 5, where 1 corresponded to “very good” and 5 to “very bad”.

1.1.3 Results

The distribution of subjects’ answers to the critical sentences is plotted in the graph below. Each bar reports the interpretation associated to the scalar term *or* for each quantifier tested. Data are obtained by analysing the percentage of subjects that respectively: accepted the critical statements in a situation in which both disjuncts were true (calculated by adding together answers of type 1 and 2 for positive contexts and 3 and 4 for negative ones); rejected them (calculated by adding together answers of type 3 and 4 for positive contexts and 1 and 2 for negative ones); expressed uncertainty (total of answers 3 on the scale).

---

1 Please note that the scale is reverted under negation.
Table below, instead, reports the proportions of subjects that accepted the critical statement calculated over the total of those that made a decision, thus excluding answers of type 3 from the total. On its rows, the results for the three different quantifiers (condition Q) are listed separately for each level of condition A, in which the position of or is varied between the 1st and 2nd argument of the quantifier. Depending on the interaction between the type of quantifier (condition Q) and the position of or (condition A), a different type of context (DE vs. NON-DE) embeds the scalar term, as indicated in the third column. In the fourth column, the percentage of subjects that found the sentence “good enough” (thus accepting an inclusive interpretation of or) is reported, obtained by adding together answers 1 and 2 on the scale (for every and un) and 4 and 5 (for no), excluding answers of type 3 from the total, as we said. These are reported separately (in percentage over the total number of participants) in the fifth column. Finally, the last column reports the mean rate assigned on the scale.
Table 2. Results. Legend: column 4, 5 and 6 report respectively: the proportion of acceptance rate of the critical sentences (corresponding to the inclusive interpretation of or calculated on subjects that made a decision, thus excluding answers 3); the proportion of uncertainty; the mean grade assigned on the scale with respect to the three quantifiers (Condition Q, column 1) and the position of the scalar term (Condition A, column 2), ultimately with respect of the type of context (DE vs. NON-DE, column 3) that embeds the scalar term. Note: the symbol * in the last column signals that the scale is reverted under negation.

<table>
<thead>
<tr>
<th>Cond. Q: type of Quantifier</th>
<th>Cond. A: position of or</th>
<th>Type of context</th>
<th>% yes answers ((\alpha_{incl}))</th>
<th>% answers of type 3</th>
<th>Mean Rate</th>
</tr>
</thead>
<tbody>
<tr>
<td>Q1 (Every)</td>
<td>A1</td>
<td>DE</td>
<td>89%</td>
<td>11%</td>
<td>1.87</td>
</tr>
<tr>
<td></td>
<td>A2</td>
<td>NON-DE</td>
<td>44%</td>
<td>23%</td>
<td>3.32</td>
</tr>
<tr>
<td>Q2 (No)</td>
<td>A1</td>
<td>DE</td>
<td>93%</td>
<td>5%</td>
<td>4.52*</td>
</tr>
<tr>
<td></td>
<td>A2</td>
<td>DE</td>
<td>84%</td>
<td>6%</td>
<td>4.13*</td>
</tr>
<tr>
<td>Q3 (Un)</td>
<td>A1</td>
<td>NON-DE</td>
<td>77%</td>
<td>18%</td>
<td>2.52</td>
</tr>
<tr>
<td></td>
<td>A2</td>
<td>NON-DE</td>
<td>41%</td>
<td>45%</td>
<td>3.19</td>
</tr>
</tbody>
</table>

Data will be analysed taking three aspects into consideration: first of all, the mean rate assigned by subjects on the scale, indicative of how much they judged Merlin’s prediction in line with what has happened in the story. Secondly, a comparison will be drawn between the percentages of subjects that considered the critical sentence good enough (i.e. those assigning it grade 1 or 2) vs. those that considered it not sufficiently in line with the facts (i.e. those assigning it grade 4 or 5), excluding answers of type 3 from the total. Finally, the rate of uncertain judgements will be taken into account.

I will analyse subjects’ distribution with respect to the effect of position (Condition A) for each quantifier in turn. Starting from the universal quantifier every (Q1), a significant difference between conditions A1 and A2 was found, both comparing the mean rate assigned on the scale by means of a paired sample t-test (1.87 vs. 3.32, t(61)=15.01, p<.001) and the proportions of acceptance of the critical sentence excluding answers 3 (89% vs. 44%, t(101)= 5.57, p<.001). These results are indicative of the fact that subjects accepted the inclusive interpretation...
of or significantly more often when it appeared in the 1st vs. 2nd argument of the universal quantifier, thus in a DE vs. NON-DE context. Moreover, considering the percentage of subjects that assigned grade 3 on the scale, this was pretty high in condition A2 (23% of overall participants), and this uncertainty clearly emerges also in the mean acceptance rate, which is very close to 3. As regards the negative quantifier no (Q2), the proportions of yes/no answers did not differ between conditions A1 and A2 in which or position was varied between the 1st vs. 2nd argument of the quantifier (93% vs. 84%, t(115)=1.51, p=.13, n.s.). Considering the rate assigned on the scale, it’s interesting to note that in condition Q2 subjects seem very sure about their judgements, as revealed by the low level of uncertainty expressed in both conditions A1 and A2 on the scale (rate is above 4 in both cases), and by the very low percentage of subjects that resorted to answer “3” in these cases (only 5% and 6% overall participants).

Finally, we will consider the case of subjects’ interpretation of or in the scope of the indefinite article un (condition Q3), the only one that doesn’t seem to conform to the predictions stated in the introductory section (cf. Table 1). A comparison between the two conditions A1 and A2 (where or appears in the 1st vs. 2nd argument of the indefinite article) revealed a significant difference. First of all, the comparison of the mean scale rate turned out statistically significant (2.52 vs. 3.19, t(61)=11.32, p<.0001). Secondly, the percentage of subjects that accepted the sentence in Condition A1 was significantly higher than the percentage of those that rejected it in Condition A2 (77% vs. 41%, t(79)=3.43, p<.001). Considering the level of uncertainty that emerges from subjects’ judgements, we can observe that the percentage of subjects that chose grade “3” on the scale in Condition A2 is very high, actually the highest of all conditions, being 45% overall participants. To conclude our scrutiny on the results obtained in this experiment, we can make a comparison between the data on un and those on every. The distance between conditions A1 and A2 (significant in both cases) is just as wide for un (Condition Q3) than it is for every (Condition Q1). Moreover, statistical analysis reveals that the mean rates on the scale do not differ between Q1 and Q3 with respect to Condition A2 (3.32 vs. 3.19, t(61)=1.59,
p=.12, n.s.). Interestingly, however, Q1 and Q3 do differ with respect to the mean rate assigned on the scale to the sentences in Condition A1, where subjects revealed to be much more certain about their judgements in case of the universal quantifier *every* than in case of *Un* (1.87 vs. 2.52, t(61)=-10.53, p<.0001).

Finally, we can evaluate the effect of syntactic context on the interpretation assigned to *or* by comparing the mean scale rate in case of DE vs. NON-DE contexts without differentiating between types of quantifier. Crucially, this difference turned out highly statistically significant (t(184)=3.97, p<.001).

1.1.4 Discussion

Aim of the experiment just described was to investigate which interpretation adults assign to disjunction by means of a truth value judgment task. The design of this experiment was a 3×2 within subject condition design in which the scalar term *or* was tested in the scope of three different determiners (namely, the universal and the negative quantifiers and the Italian indefinite article) varying the position in which it appeared (1\textsuperscript{st} or 2\textsuperscript{nd} argument of these determiners). This in order to test the effect of the monotonicity properties of the contexts (which vary across determiners and arguments as schematized in Table 1 above) on the interpretation of the scalar term they embed. Following Chierchia (2006), whenever feature has been selected the scalar ALTs are activated, then the interpretation assigned to the scalar term must be the strongest one, by the mandatory application of the exhaustivization operator \( \text{O} \). Given that in Downward Entailing contexts (such as the 1\textsuperscript{st} argument of the universal quantifier *every* or the 1\textsuperscript{st} and 2\textsuperscript{nd} arguments of the negative quantifier *no*) entailments relations are reverted, then the strongest meaning of a scalar term such as *or* ends up in being the inclusive one, which coincides with the basic (literal, logical) meaning of disjunction. On the contrary, when *or* is embedded in a NON-DE context (such as the 2\textsuperscript{nd} argument of the universal quantifier *every* or the 1\textsuperscript{st} and 2\textsuperscript{nd} arguments of the indefinite *un*) then its strongest meaning corresponds to the exclusive one, thus the implicature should be added. This is what is predicted for any occurrences of a scalar term in which the ALTs are actually active. If ALTs
are not active, then no special prediction is made for NON-DE contexts, given that the activation of the alternatives depends on contextual, extra-linguistic factors whose interplay is not under study here. It ultimately resides in a subjective choice that we could call a strategy. On the contrary, for DE context, inclusive or is predicted in any case, given that there’s no way to disentangle the plain and the strengthened value in such contexts. Bearing such predictions in mind, I will discuss the general findings of the first experiment presented by analysing the interpretation that subjects assigned to or and its correlation with the monotonicity properties of the context in which it appeared.

I will start my discussion considering subject’s responses in DE contexts. The preference that emerges when or is embedded under negation or when it appears in the restrictor of the universal quantifier is clear: subjects overwhelmingly accept the critical sentences, compatible with the inclusive interpretation of or. If we take the results obtained in the three conditions characterized by DE-ness together (namely, condition Q1-A1 and Q2-A1/2), this is the picture that actually emerges:

(5)

- in DE contexts, 89% of the subjects² accept the critical sentence containing the scalar item or in a situation in which both disjuncts are true, and this answer is compatible with inclusive interpretation of or;
- as for the level of uncertainty that emerges in case of DE contexts, we can conclude that subjects seem overall pretty sure about their answers given that the percentage of subjects that resort to answer 3 was only 7% overall.

We can contrast this distribution with the case of NON-DE contexts, represented in our experiment by three conditions respectively: the nuclear scope

² This percentage was calculated on the subjects that made a decision about the sentence, thus excluding those that resorted to answer “3” on the scale.
of the universal quantifier (condition Q1-A2) and the contexts introduced by the indefinite article (conditions Q3-A1/2). If we consider subjects’ distribution across these conditions overall, this is the picture that emerges, which clearly contrasts with the one obtained in case of DE contexts schematized above:

(6)

- in NON-DE contexts, only 54% of the subjects\(^3\) accept the critical sentence containing the scalar item *or* in a situation in which both disjuncts are true; half of the subjects that made a decision actually reject these sentences, showing their preference for the *exclusive* interpretation of *or* in those cases;

- the level of uncertainty that emerges in case of NON-DE contexts is much higher than before, given that the percentage of subjects that resort to answer 3 in these contexts reaches 31% overall.

If we plot the percentage of subjects that reject vs. those that accept the critical sentences (without excluding those that resort to answer “3” whose percentage is plotted in a separate bar), the contrast between DE and NON-DE contexts (that turned out statistically significant, as we have seen at the end of the section dedicated to results) is well captured, as shown in the graph below:

\(^3\) Again, this proportion was calculated excluding answers “3” from the total.
The pattern of distribution of subjects’ responses is the one predicted by the theory: in DE contexts, inclusive *or* is predicted by default, being this interpretation the strongest one in such contexts in which the entailment relations are reverted. On the contrary, no prediction is made for NON-DE context, in which the interpretation assigned to *or* depends on the choice made: subjects that select *or*\([-\sigma]\) will accept the critical sentence, those selecting *or*\([+\sigma]\) will reject it. As a matter of fact, if *or*\([+\sigma]\) is selected then the scale <or, and> is accessed and ALTs are activated. In this case, the alternative statement “A *and* B” will become relevant and thus the statement containing *or* will be rejected on the basis of the constraint on strengthening according to which the strongest alternative must be selected. This amounts to saying that, whenever the critical statement is rejected, *or* is assigned the exclusive interpretation, corresponding to “A *or* B and *not* (A *and* B)”. The fact that the activation of SI is optional and that the feature selection arbitrary is also attested by the fact that nearly one subject out of five didn’t actually take a decision, providing a vague response such as grade “3” on
the scale. I believe that this distribution reflects the conflict between a
“pragmatical” and a “logical” choice, which invests the single participant as well
as the whole group. I’m not going to discuss here which factors intervene in the
actual selection of the feature associated to scalar items, an argument that could
constitute alone the object of an entire dissertation. I think it is interesting to note,
however, that the percentage of subjects that consider “relevant” to derive the SI
associated to or is in a sense low: only one subject out of three judges the sentence
containing or inappropriate in a situation in which both disjuncts are true (while a
second accepts it and a third doesn’t make a choice). My guess is that this
distribution may be due to the particular version of the TVJT that was used.
Differently from the standard procedure normally adopted, here the puppet tries to
guess what has happened, instead of providing a pure description of the facts
(remember that Merlin was blindfolded, so he couldn’t watch the stories).
Moreover, in this experiment subjects are required “pragmatic” judgements: they
are asked if the critical statement is “appropriate”, not if it is “true”, as in the
classical TVJT. This may induce some subjects to be more indulgent than they
would normally be in ordinary conversations, accepting a “guess” like “A or B
will happen” as sufficiently close to the facts when A and B happened. This
doesn’t exclude the possibility that these same subjects, when asked about truth-
value judgements, could reject the corresponding statement in form of assertion,
like “A or B has happened” in case this were used as a description of the same
situation. In any case, as I said above, the manipulation on the standard version
was considered necessary in this case in order to render the use of or felicitous.

In what follows I will discuss the results obtained for the third of the
determiner tested, namely the Italian indefinite article un. As we have seen, these
are in a way problematic with respect to the theoretical framework I’m assuming.
According to the predictions made above (cf. Table 1), subjects’ distribution
should not vary depending on the position of the scalar item: being this determiner
NON-DE in both of its arguments, no difference between conditions A1 and A2
for Q3 was expected. Instead, a significant difference was found, and I will now
propose a tentative explanation of this fact.\textsuperscript{4} This unwelcome result is essentially due to the fact that a lower percentage of subjects than expected rejected the critical statement in condition A1 (thus assigning the inclusive interpretation to or when it appears in the restrictor of the indefinite article), compared to those that rejected it in condition A2. I believe this effect to depend on the form of the critical sentence, namely to the fact that the DP-subject contains a relative clause. It’s a well-known fact in the linguistic literature that the adding of a relative clause has some effects on the overall interpretation of the sentence. I’m specifically referring here to the phenomenon of subtrigging, extensively studied by Dayal (1996) in connection with the use of free-choice any in episodic constructions. One of the examples she discusses is the “rescue” of ungrammatical sentences such as (7) by adding a relative construction in the first DP, as in (8):

(7) \#Any girl took a pet home

(8) Any girl that went to the dog pound took a pet home

The use of free-choice any is ungrammatical in (7) because the sentence is episodic. Note, in fact, that normally only generic statements such as (9) licenses FC any:

(9) Any girl would love to have a pet

Furthermore, consider that English has the option of bare plurals to express generic statements like (9):

(10) Girls love pets

In Italian there’s not such an option. Crucially, however, one can add a relative construction to the DP subject in an episodic sentence to gain a generic flavour. Consider the pair below:

\textsuperscript{4} I want to thank Francesca Panzeri for the suggestion.
Sentence (11) can express both an episodic and a generic statement in Italian. Sentence (12), in which a relative construction is added, is instead preferably interpreted as expressing the generic statement: kids that have been to the dog-pound love pets.

Let’s consider the critical sentence used to test or in the restrictor of the indefinite article in my experiment. This is reported below:

(13) Una bambina che ha preso la fragola o il gelato è salita in barca
     [= a girl that took the strawberry or the ice-cream got on the boat]

The form of this sentence resembles a lot the Italian example (12) and the English example (8). As we have seen, the effect gained in those examples was that of a generic statement (this fact licensing the use of FC any in (8)). For this reason, (8) and (12) can be paraphrased in form of the “rules” below:

(14) If a girl went to the dog pound, she took a pet home
(15) Se un bambino è stato al canile, allora ama gli animali
     [= if a kid has been to the dog pound, then he loves the animals]

Drawing a parallel with the phenomenon of subtrigging, I believe that the presence of the relative clause inside the DP-subject in the critical statement is responsible for the fact that more subjects than expected assigned the inclusive interpretation to or when it appeared in the 1st argument of the indefinite article.

---

5 Except that in (13) the predicate in the matrix clause appears in the past tense, and this fact can shadow the generic effect gained by the presence of the relative clause.
Along the lines of (14) and (15), the critical statement used in my experiment for condition Q3-A1, can be reformulated in form of the “rule” (16):

\[(16) \text{ If a girl took the strawberry or the ice-cream, she got on the boat}\]

Crucially, the disjunction in (16) appears in the antecedent of a conditional, thus in a DE context. Thus, rule (16) is satisfied if, for each girl that took the strawberry or the ice-cream or even both, it is true that she got on a boat. Given that in the story presented there were 4 girls and crucially only one of them got on a boat, namely the girl that took the strawberry and the ice-cream, rule (16) is compatible with what happened in the story. It is thus plausible to assume that subjects that accepted the critical sentence (13) in such a situation, thus assigning or the inclusive interpretation, interpreted Merlin’s prediction not as an episodic statement (thus not simply as a description of what has happened in the story) but as a sort of a generic “rule” on the lines of (16). This possibility is even more plausible if we consider the fact that Merlin didn’t physically assist to the stories (he was blindfolded) so he could only guess (or predict) what happened. And guesses are normally expressed in form of a rule indeed.

1.2 Conclusive remarks on Experiment 1

The general idea behind the off-line study presented in this chapter was the investigation of the effect of the logical property of monotonicity on adults’ interpretation of a scalar term such as or. In a broader perspective, being Downward Entailing a purely logical classification, the results obtained in this study can give us a first hint of the extent of spontaneous logicality in the way adults understand and interpret sentences.

Once the subjective choice is made, by activating the scalar alternatives or not, then SI computation depends on their effect on assertion’s informativeness: if the adding of the implicature leads to a general strengthening of the assertion, then it is added, otherwise it is not (according to the exhaustivization operator O). What rules this effect of global strengthening/weakening are exactly the
monotonicity properties of the context in which the scalar term is embedded: in a DE context, given that entailment relations are reverted, the adding of the implicature leads to a weakening of the assertion; in NON-DE contexts, instead, it leads to a general strengthening. This is the explanation/prediction at the level of the theory. Do adults really behave this way? Are they sensitive to purely logical distinctions such as being DE or not? If we consider the data emerged from the study just described then the answer seems “yes, in fact, they do”: in DE contexts, such as the 1st and 2nd argument of the negative quantifier and the 1st argument of the universal quantifier the interpretation assigned to or by the large majority of the subjects is the inclusive one; in NON-DE contexts, such as the 2nd argument of the universal quantifier, subjects split (and are much more unsure about their choice).

By means of a completely different task, Noveck, Chierchia and colleagues (2002) addressed the same experimental question. They tested the interpretation of or using a reasoning task in which two premises and a conclusion were presented to the subjects, who, given the premises, had to accept or reject the conclusion provided. Critical material comprised a first premise in which or appeared in the antecedent or in the consequent of a conditional, which display different entailing properties: the former is DE while the latter is NON-DE. This first premise was followed by a second one and a conclusion in order to test the interpretation assigned to the disjunction appearing in the sentences. Interestingly, the material that was used contained only simple arbitrary material, such as capital letters, lacking any reference to realistic content that can be influenced by extra world-knowledge. Moreover, only one occurrence per item type was presented to the same subject, in order to encourage spontaneous reactions and block the emergence of a strategy across similar items. Specifically, the table below reports the two critical items that are of much interest for the purposes of our discussion about downward entaillingness and or interpretation, and the results that Noveck, Chierchia and colleagues obtained:
Antecedent of conditional (DE) | Consequent of conditional (NON-DE)
---|---
1st premise | If there is a P or a Q then there is an R | If there is a P then there is a Q and an R
2nd premise | There is a P | There is a P
 | There is a Q | There is a Q or an R
Conclusion | There is an R | There is a Q or an R
Expectation | Yes (= inclusive or) | No (= exclusive or)
Results (%) | 85% Yes | 75% No

Table 3. Noveck et al., 2002, Experiment 1. Some of the critical material used and the results obtained in a reasoning task on the interpretation of or in DE and NON-DE contexts. Both results are significantly different from chance (p<.05).

As it is evident from the percentages reported in the chart above, their results are perfectly in line with the results obtained in the experiment I conducted by means of a completely different task: or is preferably treated inclusively in DE contexts and exclusively in NON-DE contexts. The results obtained are in fact significantly different from chance for both contexts.

In a second series of experiments presented in the same paper, the authors address one possible objection to the results obtained in the first experiment just discussed: namely, the fact that the rejection of the conclusion in NON-DE contexts may not be effectively due to the actual computation of the scalar implicature not both but simply to the fact that the conclusion provided is underininformative in a more general sense (given that and would be a more appropriate conjunction in the context provided). To address this point, they investigated whether or not underinformativeness is ground for rejection in general presenting material of this sort:

Table 4. Noveck et al., 2002, Experiment 1. Control on Experiment 1 (cf. Table 1): the conclusion provided is incomplete but subjects accept it despite its general underinformativeness.
In the example presented above, the conclusion is valid but incomplete because it doesn’t provide the whole piece of information, namely that Q and R are expected. If underinformativeness were a sufficient reason for rejection, then one would expect a significant rejection rate in examples of this sort, mirroring the results obtained in the previous experiment for NON-DE contexts. But clearly this objection is not tenable, given that 92% of the subjects actually accepted the “underinformative” conclusion provided. Thus, the rejection of the disjunctive conclusion in the first experiment is motivated by the fact that or is interpreted exclusively.

Finally, the distinction between DE and NON-DE contexts in SI computation was also tested in the Noveck et al. study using the same material of Experiment 1 (cf. Table 1) in which they turned the conclusion into a question, which represents a DE environment. For example, for the NON-DE context, the assertion that constituted the conclusion in the first experiment, namely, *There is a Q or an R* was converted into the question *Is there a Q or an R?* After this manipulation, the percentage of “yes” answers (corresponding to an inclusive interpretation of or) rose from 25 to 80%, and the difference between the assertion and the question condition is highly significant (p < .001). This result is expected given the difference in the monotonicity properties of the two environments, and in lines with the results previously found.

In summary, as clearly emerges from the studies presented in this chapter so far (i.e. my Experiment 1 and the experiments in the Noveck et al. study), adults seem to model the interpretation of the sentences they encounter on the basis of the logical properties of the syntactic context in which the critical term appears. Probably they do it unconsciously, but they do it consistently. Under the theoretical perspective I’m adopting, once they (more or less arbitrarily) have chosen the relevant option by selecting the feature [+σ], activating the scalar alternatives or not (the way they do this is beyond the purposes of our discussion here), the interpretation they assign to or is strictly determined by the logical properties of the context in which the scalar term appears, and their effect on informativeness of the whole sentence: whenever or appears in a DE context then
the implicature is never added because it leads to a general weakening, while in a NON-DE context it always is (if the scalar alternatives are activated), because it leads to a general strengthening. The reason why this is so ultimately resides in the logic notion of entailment, in tandem with the pragmatic notion of strengthening, as we have discussed.

2. **A reaction-time study on the interpretation of or**

In a second study, I investigated adults’ interpretation of disjunction by means of an experiment in which subjects were asked to evaluate sentences containing *or* in an on-line task in which their answering times were measured. On-line tasks are actually more apt to capture the time-course of scalar implicature computation, and can help in deciding among alternative accounts of their derivation. In particular, from the different theoretical accounts presented in Chapter 1, pretty sharp predictions were derived and tested in the experiment I’m going to present in the remaining part of this chapter.

2.1 **Experiment 2: exclusive or inclusive OR?**

2.1.1 **Subjects**

A total of 30 subjects participated in this experiment. Participants were mainly 1st year students at the Psychological Faculty of the University of Milano-Bicocca, and received credits for their participation. They were randomly divided into three lists, in which the same type of critical sentence was presented using different items, in order to have a control on the material used.

2.1.2 **Material**

To avoid interferences from extra-linguistic factors on the interpretation of sentences, all the material presented in this experiment contained only fantasy names for characters and objects (cf. Appendix B). Characters were in fact introduced as inhabitants of weird planets with their bizarre objects, unfamiliar to
inhabitants from the Earth. To have a better idea of the material used, consider this sample:

(17)

This is a Glimp
This is a curp
This is a dorf

The first picture represents a Glimp, which lives on planet Glimp and is one of the strange aliens that appear in the experiment. The other two pictures represent some of the bizarre objects with their fantasy names written below.

The experiment presented a 2×2 condition design, in which two conditions were created as a within subject factor, each displaying 2 different levels. The type of syntactic environment in which the disjunction appeared corresponded to the sentence types in Condition I, which varied between the two levels reported in (18):

(18) Condition I: Sentence-type
(a) If a P has an A or a B, then he also has a C
    [DE context]
(b) If a P has a C, then he also has an A or a B
    [NON-DE context]

In sentence of type (a) or is embedded in the antecedent of conditional, which crucially constitutes a DE environment; in sentences of type (b), instead, or is embedded in the consequent of the conditional, which constitutes a NON-DE
environment. Each sentence variable was presented to each subject in two different situations (corresponding to levels 1 and 2 of Condition II described below), which modulated the interpretation associated to the scalar term contained in the sentences by means of a scenario represented by a set of pictures. Given the presence of *or*, each sentence can have two possible interpretations, depending on which reading is assigned to the scalar item. By means of a concrete example taken from the material actually used in the experiment (remember that fantasy names for characters and objects were used), let’s consider the two interpretations (*inclusive* vs. *exclusive*) associated to *or* in sentences of type (a) and (b) (Condition I, cf. (18)) and their mutual entailment relations:

(19)

(a) *If a Glimp has a curp or a dorf, then he also has a pencil*

\[
\text{Or}_{\text{inc}}: \forall x \left[ \text{Glimp}(x) \land \exists y \text{ curp}(y) \land \text{has}(x,y) \lor \exists z \text{ dorf}(z) \land \text{has}(x,z) \right] \rightarrow \exists k \left[ \text{pencil}(k) \land \text{has}(x,k) \right]
\]

\[
\text{Or}_{\text{exc}}: \forall x \left[ \text{Glimp}(x) \land \exists y \text{ curp}(y) \land \text{has}(x,y) \lor \exists z \text{ dorf}(z) \land \text{has}(x,z) \land \neg \text{has}(x,y \land z) \right] \rightarrow \exists k \left[ \text{pencil}(k) \land \text{has}(x,k) \right]
\]

**direction of entailment:** (Or_{inc}) $\Rightarrow$ (Or_{exc})

(b) *If a Glimp has a pencil, then he also has a curp or a dorf*

\[
\text{Or}_{\text{inc}}: \forall x \left[ \text{Glimp}(x) \land \exists k \text{ pencil}(k) \land \text{has}(x,k) \right] \rightarrow \exists y \left[ \text{curp}(y) \land \text{has}(x,y) \lor \exists z \text{ dorf}(z) \land \text{has}(x,z) \right]
\]

\[
\text{Or}_{\text{exc}}: \forall x \left[ \text{Glimp}(x) \land \exists k \text{ pencil}(k) \land \text{has}(x,k) \right] \rightarrow \exists y \left[ \text{curp}(y) \land \text{has}(x,y) \lor \exists z \text{ dorf}(z) \land \text{has}(x,z) \land \neg \text{has}(x,y \land z) \right]
\]

**direction of entailment:** (Or_{exc}) $\Rightarrow$ (Or_{inc})

The two possible interpretations associated to *or* are not logically independent of one another, as expressed by the fact that always one interpretation entails the other. As one can note, the direction of the entailment follows a
different pattern depending on the properties of the context in which the scalar term is embedded: in sentences of type (a), representing a DE context, *inclusive or* is more informative, while the revert holds for sentences of type (b), which represent a NON-DE context. Giving these entailment relation patterns, two different situations were created to test the critical sentences. These correspond to the two levels of Condition II reported below:

(20) Condition II: situation-types

S1 (EXC): a situation representing the *exclusive* interpretation of *or*;
S2 (INC): a situation representing the *inclusive* interpretation of *or*.

Notice that, since the two interpretations are not logically independent of one another, sentence of type (a) turn out to be false in situation 1 if *or* is assigned the *inclusive* interpretation, while sentence of type (b) turn out to be false in situation 2 if *or* is assigned the *exclusive* interpretation. A control (false) condition was added, in which the scenario presented rendered the sentence “false” under any interpretation of *or*.

To represent conditions S1 and S2, four pictures appeared on the screen, each containing one P (for example, a Glimp) with an object A (for example, a curp) and/or an object B (for example, a dorf) and/or an earthly object C (for example, a pencil); the fourth picture always functioned as a distractor and contained one alien P with an earthly object instead of an alien object. The chart below reports all the combinations of characters and objects that appeared in the different situations in which both sentences of type (a) and (b) were tested:

<table>
<thead>
<tr>
<th>Situation</th>
<th>PICT. 1</th>
<th>PICT. 2</th>
<th>PICT. 3</th>
<th>PICT. 4</th>
</tr>
</thead>
<tbody>
<tr>
<td>1 Glimp with …</td>
<td>1 Glimp with …</td>
<td>1 Glimp with …</td>
<td>1 Glimp with …</td>
<td>1 Glimp with …</td>
</tr>
<tr>
<td>S1=EXC</td>
<td>curp + pencil</td>
<td>dorf + pencil</td>
<td>curp + dorf</td>
<td>a car</td>
</tr>
<tr>
<td>S2=INC</td>
<td>curp + pencil</td>
<td>dorf + pencil</td>
<td>curp + dorf + PENCIL</td>
<td>a car</td>
</tr>
</tbody>
</table>

*Table 5. Experimental design.* Different combinations of objects in scenarios S1 and S2.
For example, consider the following sentences:

(21)
(a) If a Glimp has a curp or a dorf, then he also has a pencil
(b) If a Glimp has a pencil, then he also has a curp or a dorf

These were tested (on different subjects) in the following scenarios, representing conditions S1 and S2 respectively:

(22)
S1: only compatible with exclusive interpretation of or.

(23)
S2: only compatible with inclusive interpretation of or.

Please note that the only crucial difference between the two scenarios is represented by the third picture in the sequence. Of course, during the experiment,
the order of presentation of the four pictures was completely randomized across items and subjects, especially because reaction times were recorded.

As we said, scenario S1 is only compatible with the exclusive interpretation of or, which is the most informative in sentence of type (b) but not in sentence of type (a). On the contrary, scenario S2 is only compatible with the inclusive interpretation of or, which is the most informative in sentence of type (a) but not in sentence of type (b). This is, in summary, the structure of the experimental design:

<table>
<thead>
<tr>
<th>Cond. I Sentence</th>
<th>Monotonicity of context</th>
<th>Entailment relations</th>
<th>Cond. II S1-exc</th>
<th>Cond. II S2-inc</th>
</tr>
</thead>
<tbody>
<tr>
<td>(a) antecedent of conditional</td>
<td>DE</td>
<td>inclusive ↓ exclusive</td>
<td>False (if inclusive)</td>
<td>True (if inclusive)</td>
</tr>
<tr>
<td>(b) consequent of conditional</td>
<td>NON-DE</td>
<td>exclusive ↓ inclusive</td>
<td>True (if exclusive)</td>
<td>False (if exclusive)</td>
</tr>
</tbody>
</table>

Table 6. Schema of the experimental design.

2.1.3 Procedure

The experiment was realised using E-Studio and run using E-Run, both applications of the software E-Prime. Responses were coded by E-Merge and then submitted to statistical analysis by means of E-Data Aid and SPSS 12. Subjects were tested in a quite room using a laptop. They were also given headphones to hear the oral instructions and the presentation of the objects. All the material was presented in written and oral form except for critical sentences, which only appeared in a written form on the screen as to avoid any influence on spontaneous. Subjects’ task was to evaluate the critical sentences as “true” or “false” with respect to a scenario constituted by four pictures that appeared on the screen. Given that all the sentences contained fantasy names for characters and objects, subjects were told at the beginning that they would be shown some strange characters from three different planets that did something with bizarre objects that can only be found on those planets. They were reassured that the
experiment was not a test to measure their ability to recognise and/or memorize names of strange objects. It was made clear that to evaluate the sentence, in fact, this ability was not required at all. Subjects could decide their pace: before each critical trial, an introductory screen was shown to present the alien and the objects that they will encounter immediately after, with their fantasy names (similar to the presentation in (17) above). When ready to start, they had to press the space bar to move on to the critical trial; otherwise, they could take a rest before pressing the bar. After the introductory screen, critical material was presented as follows (after the subject pressed the space bar): at the top of a black screen the sentence appeared (in white). Subjects were instructed to read (silently) the sentence and then press the space bar key to see the four pictures describing the situation against which they had to evaluate the sentence as “true” or “false”. By pressing the bar, the four pictures appeared on the screen in the space below the sentence (in a random order). To answer, subjects had to press one of two highlighted keyboard keys: a green key for “true” (that was on the left margin of the keyboard, corresponding to the position of the key for the letter “a”) and a red key for “false” (that was on the right margin, corresponding to the key for the letter “l”). After pressing it, they were either asked to move on by pressing the space bar (whenever their answer was “false”) or, in case they answered “true”, they had to answer another question that appeared in the middle of the screen (the four pictures remained there): “How much do you think the sentence is a good description of the situation represented in the pictures?” They were given a scale of response varying from 1 (bad) to 5 (good). Only time to answer the True/False question was recorded, starting from the moment they pressed the bar to make the pictures appear on the screen till they pressed the True/False key.

Critical conditions were treated as a within subject factor: each subject was shown the complete battery of the material but saw only one occurrence per each critical item-type, for a total of 17 test items, 4 of which were critical test sentences containing or and the others were fillers.
2.1.4 Predictions

As we have seen, given the presence of or, each sentence of the type (a) and (b) can have two possible interpretations, depending on which is assigned to the scalar item (cf. (18) above). Two questions are addressed in this experiment:

(24)

(i) whether one of the candidate interpretations constitutes the preferred one depending on the syntactic environment it appears in. To this purpose, the rate of acceptance/rejection of the sentences across conditions will be considered;

(ii) whether the derivation of the implicature is a costly process. To this purpose, the analysis of reading times (RTs henceforth) will be analysed.

One of the approaches that have been discussed in Chapter 1 is the grammar-driven approach proposed by Chierchia (2004/2006). According to this approach, the process of computing implicatures is virtually costless and, once the feature associated to the scalar term has been selected, it depends on mechanisms operating by default: whenever [+σ] is chosen, the strongest interpretation must be derived (by default). As we have seen, depending on the monotonicity properties of the syntactic environment that embeds the scalar term, the inclusive interpretation of or is the strongest in DE contexts and the exclusive in NON-DE ones. Crucially, when a scalar term is embedded in a DE contexts (as in the (a)-sentences in my experiment), the fact of activating the ALTs associated to it or not leads to the same outcome: if or is assigned the feature [+σ], then the strongest interpretation must be derived. Being this the inclusive one in case of DE contexts, this coincides with the basic/logical meaning of or, hence the same one that would be assigned if the ALTs were not activated. Under this formulation, we don’t know which mechanism leads to the final interpretation assigned to the (a)-sentence in my experiment, unless we derive a parallelism with the interpretation assigned to or in NON-DE contexts, such as (b)-sentences.
Crucially, in those contexts, the selection of [+σ] or [-σ] leads to a different result: if ALTs are activated, in fact, the strongest interpretation must be assigned to *or*, corresponding to the *exclusive* one in NON-DE contexts. Otherwise, the basic meaning – without the implicature – is derived. My prediction is that this last option may apply in situation S2, where the context is compatible with the *inclusive* interpretation of *or*: given that a flip between having or not having the implicature in NON-DE contexts is relatively easy, some subjects may choose to go for the plain meaning of the scalar term (selecting [-σ]), thus accepting the sentence in such a situation. In this respect, I’m not drawing any specific prediction on the rate of acceptance of sentence (b) in Condition S2. My only guess is that the percentage of subjects that accept the (b)-sentence in this condition is higher than those who accept the (a)-sentence in Condition S1 (even if no clear prediction about the range of this difference can be made). On the contrary, I believe that the first option (i.e. *or* is assigned the feature [+σ]) is the most plausible in Condition S1: when facing a scenario that is only compatible with the *exclusive* interpretation of *or*, I expect ALTs to be activated for disjunction in sentences of type (b). Hence a high rate of acceptance of this type of sentence should emerge, given that the adding of the implicature in such a context (which is NON-DE) leads to a general strengthening of the assertion\(^6\). If this prediction is correct, then a difference between the two sentence types in this condition is to be expected. Building on the considerations just made and on the distributional generalization listed in Chapter 1 (cf. (72), section 2.3), this is how the predictions compatible with a grammar-driven approach can be summarized:

---

\(^6\) Note, however, that a low rate of acceptance of the (b)-sentence in Condition 1 is not in principle incompatible with the theory because it would depend on feature selection about which the theory doesn’t make predictions.
With respect to feature selection, as I said, I’m not making any predictions. The factors that can influence subject’s choice in this respect are not under investigation here, given that their analysis alone could constitute the topic of a whole dissertation. Crucially, however, one cannot exclude a priori the possibility that ALTs remain inactive in both conditions for both types of sentence, presumably as a consequence of subjects’ interpretation of the task demands. If this is the case, then no difference is expected between the two sentence types, but an opposite result is predicted in terms of acceptance/rejection rate between Condition II-S1/S2, each representing a different interpretation of disjunction. The table below summarizes the predictions derived by the theory under discussion as a function of feature selection:

<table>
<thead>
<tr>
<th>Cond. I Sentence</th>
<th>Monotonicity of context</th>
<th>Cond. II S1 (EXC)</th>
<th>Cond. II S2 (INC)</th>
</tr>
</thead>
<tbody>
<tr>
<td>(a) antecedent of conditional</td>
<td>DE</td>
<td>HARD</td>
<td>EASY</td>
</tr>
<tr>
<td></td>
<td></td>
<td>(\Rightarrow) high rejection</td>
<td>(\Rightarrow) high acceptance</td>
</tr>
<tr>
<td></td>
<td></td>
<td>(\Rightarrow) high RTs if accepting</td>
<td>(\Rightarrow) low RTs if accepting</td>
</tr>
<tr>
<td>(b) consequent of conditional</td>
<td>NON-DE</td>
<td>EASY</td>
<td>relatively EASY</td>
</tr>
<tr>
<td></td>
<td></td>
<td>(\Rightarrow) high acceptance</td>
<td>(\Rightarrow) acceptance/split</td>
</tr>
<tr>
<td></td>
<td></td>
<td>(\Rightarrow) low RTs if accepting</td>
<td></td>
</tr>
</tbody>
</table>

Table 7. Predictions on subjects’ distribution across conditions.

<table>
<thead>
<tr>
<th>Cond. II</th>
<th>if ([-\sigma])</th>
<th>if ([+\sigma])</th>
</tr>
</thead>
<tbody>
<tr>
<td>S1 (EXC)</td>
<td>(a) = (b) rejection</td>
<td>Acceptance rate: (b): high (a) &lt;&lt; (b)</td>
</tr>
<tr>
<td>S2 (INC)</td>
<td>(a) = (b) acceptance</td>
<td>Acceptance rate: (a): high (b) &lt; (a-S2) &amp; &gt; (a-S1)</td>
</tr>
</tbody>
</table>

Table 8. Predictions as a function of feature selection.
With respect to RTs, the framework I am adopting doesn’t make specific predictions about a presumptive “cost” of implicature computation. Given that this approach considers the mechanism responsible for SI computation (once the feature has been selected) as a mechanism that operates by default and thus (virtually) costless, no significant difference across conditions and sentence types in dependence of implicature derivation or non derivation is in principle predicted. However, considering the fact that flipping between having and not having an implicature is expected to be harder in DE than in NON-DE contexts, one specific prediction can be formulated. In fact, I expect higher RTs for subjects that accept the (a)-sentence in situation S1 (they have to add the implicature, despite the fact that it leads to a weakening of the assertion) with respect to those that reject it (do not accommodate and stick to the strongest interpretation of or in DE contexts). In my opinion, however, the difference in RTs eventually found, would have nothing to do with a “cost” of implicature computation/cancellation but would have to be intended as the result of a process of accommodation of some sort.

The crucial observation to make at this point is that, even if a general “cost” of implicature computation eventually emerged from the experimental data, this would not mine the theoretical architecture I am proposing: the mechanism which SI computation is based on in a grammar-driven approach is independent from “cost and benefit” evaluations. On the contrary, these notions are crucial for the whole Relevance Theoretic explanatory architecture, for which the process of computing implicatures cannot be but effortful. The main claim of this approach is the following: implicatures are costly, they are only derived when required by (extra-linguistic) context in pursuit of optimal relevance. Starting from this claim, the following predictions can be, in my opinion, drawn with respect to the experiment under discussion, depending on which, between the specific scenario represented time by time in the pictures or the task in general influence subjects’ strategy. If it is the case that subjects are influenced by the experimental situation in general, then they should perform equally across conditions: depending on their interpretation of the experimental requirements, they will always or never derive the implicature, independently from the specific scenario or sentence-type
presented and only following their personal interpretation of the task demands. If instead they are influenced by the specific scenario presented each time during the task, then a different behavior is expected between the two levels of Condition II, given that these are compatible with opposite interpretation of the scalar term. Hence, in situation S2, the interpretation of or which is made explicit by the context is the inclusive one, thus no implicature is required to accept the sentence. In situation S1, conversely, exclusive or is realized in the scenario, thus the computation of the implicature is required to render the sentence true. As far as I understand this approach, no clear predictions can be formulated with respect to which context features can favor derivation: it seems to me that much depends on how subjects interpret the task demands and/or the experimental situation. Moreover, it doesn’t seem to me that specific logic properties of the syntactic environment (such as Downward Entailment) are taken into consideration as factors affecting implicature derivation, or at least, I haven’t find any specific mention of this contrast in the literature.7 Following these suggestions, the unique clear-cut prediction that I can derive from such approaches is the following: with respect to Condition I, no difference is in principle expected between the two types of sentences varying the type of scenario presented (S1 or S2), given that implicature derivation depends on overall consideration of relevance (which in turn depends on how participants interpret the situation) and not from logic properties of the syntactic environment, such as DE-ness. In any case, a more clear-cut prediction can be made on reading times. Given that implicature derivation is univocally considered a costly process in the Relevance Theoretical approach, a significant difference in RTs is predicted between the two levels of Condition II, independently from sentence-type: higher RTs are predicted for subjects that accept critical sentences in S1, given that this scenario is only compatible with exclusive or (hence they have to derive the implicature to accept the sentence) compared to those who reject them in the same condition. An opposite trend is instead predicted for S2: higher RTs for subjects rejecting the

---

7 It should be noted, however, that the eventuality that these syntactic/logical factors influence subjects’ behavior with respect to SI computation is not in principle incompatible with Relevance Theory.
sentence with respect to those accepting it, given that this scenario is only compatible with *inclusive or*, hence a “true” judgment comes for free. This, in summary, is how the two hypotheses differentiate with respect to their predictions on RTs:

<table>
<thead>
<tr>
<th></th>
<th>Relevance Theory</th>
<th>Grammar-driven theory</th>
</tr>
</thead>
<tbody>
<tr>
<td>RTs</td>
<td>(a) = (b)</td>
<td>(a)-sentence in S1: true &gt;&gt; false</td>
</tr>
<tr>
<td></td>
<td>S1 (true) &gt;&gt; S1 (false) &amp; (S2) false &gt;&gt; S2 (true)</td>
<td></td>
</tr>
<tr>
<td></td>
<td>&amp; S1 (true) &gt;&gt; S2 (true) &amp; S1 (false) &gt;&gt; S2 (false)</td>
<td></td>
</tr>
</tbody>
</table>

Table 9. *Relevance Theory and Grammar-driven approach in comparison.* Legend: sentences of type (a) represent a DE context, sentences of type (b) a NON-DE context; Condition S1 represents a scenario compatible with *exclusive interpretation of or*; Condition S2 a scenario compatible with *inclusive* interpretation.

As for control condition, “false” answers at ceiling are expected by both theories.

Let’s now turn to the analysis of the results obtained, to see whether they conform to the predictions put forward in this section and help deciding between the two competing theories under discussion.

### 2.1.5 Results

Before entering the details of the analysis of the data obtained in this experiment, some considerations are worthy of remark. First of all, a comparison across the three lists was made, to verify that they were not different from one other. As expected, no significant difference was found ($F(2)=.643$, $p=.53$). Secondly, it’s important to notice that accuracy on controls is attested around 95% overall, given that only three subjects made a mistake on one of the control items.

In the third place, a remark on reaction times is due. Before analysing RT data, anomalous effects were curtailed in two steps. First, excessively long reading
times, those exceeding 2.5 times the mean item time overall participants were excluded from the analysis because it was assumed that these could be due to processes other than that of interest (overall, RTs on eight items were excluded, corresponding to 3.33% on the total). Successively, individual cut-off values were calculated for each participant, corresponding to the mean (calculated on the total response times per subject on all items) + 2 SD over all critical items (disregarding conditions); values above this cut-off were smoothed to the upper cut-off value (overall, RTs on nine items were smoothed, corresponding to 3.75% on the total).

Let’s now turn to the analysis of the results obtained. These are reported in the table below, divided per type of sentences (Condition I) which crucially differ in their monotonicity properties: the 2nd column reports the type of scenario in which the sentence is evaluated (recall that S1 corresponds to the exclusive interpretation of or while S2 corresponds to the inclusive); the 3rd column reports the percentage of answers “true”, followed by the rate assigned to the scale that appears in the 4th column; the last three columns report respectively: the response times to answer “true”, to answer “false” and the mean total response time per condition.

<table>
<thead>
<tr>
<th>Condition I: sentence type</th>
<th>Cond. II Situation</th>
<th>% True</th>
<th>Scale rate</th>
<th>RT for True</th>
<th>RT for False</th>
<th>Mean RT</th>
</tr>
</thead>
<tbody>
<tr>
<td>(a) = DE</td>
<td>If a P has an A or a B, then he also has a C</td>
<td>S1 (exc)</td>
<td>57%</td>
<td>3.47</td>
<td>11320</td>
<td>7167</td>
</tr>
<tr>
<td></td>
<td></td>
<td>S2 (inc)</td>
<td>90%</td>
<td>3.81</td>
<td>8937</td>
<td>12362</td>
</tr>
<tr>
<td>(b) = NON-DE</td>
<td>If a P has a C, then he also has an A or a B</td>
<td>S1 (exc)</td>
<td>87%</td>
<td>4.38</td>
<td>9734</td>
<td>8341</td>
</tr>
<tr>
<td></td>
<td></td>
<td>S2 (inc)</td>
<td>77%</td>
<td>4.04</td>
<td>10183</td>
<td>11754</td>
</tr>
</tbody>
</table>

Table 10. Results of Experiment 2.

Data on critical items can be analyzed with respect to different parameters: percentage of “true” and “false” answers; time taken to make a decision between
“true” and “false”; grade assigned to the scale. Within subject comparisons on acceptance rate were made between the two levels of Condition II (S1 and S2), taking the type of sentence (Condition I (a) vs. (b). representing DE vs. NON-DE context) as independent variables. Let’s start from considering the percentages of “true” and “false” responses for the two sentence types in the critical conditions, which are plotted in the graphs below:

As it is evident from the graphs above, the pattern of subjects’ answers in the two conditions clearly differentiate depending on the types of sentence,
ultimately on the monotonicity properties of the context in which *or* appears. First of all, sentences of type (b) are more widely accepted than sentences of type (a) in both conditions. As a matter of fact, 87% of the subjects accept (b)-sentence Condition II-S1, representing a scenario only compatible with the *exclusive* interpretation of *or*, and 77% of them accept it in Condition II-S2, when the situation presented is only compatible with the *inclusive* interpretation of *or*. The difference between these two proportions showed by the a paired sample t-test is not significant (87% vs. 77%, t(29)=1.00, p=.33, n.s).

Conversely, subjects display a different behaviour with respect to sentences of type (a), representing a DE context. Precisely, a large majority of them (90%) accept this sentence in Condition II-S2, compatible with *inclusive* interpretation of *or*, while only half of them (57%) accept it in S1, where *exclusive* interpretation of *or* is represented. The difference between these percentages is highly significant (90% vs. 57%, t(29)=-3.34, p<.01). Crucially, the rate of acceptance of the (a)-sentence in S1 (EXC) is also significantly different from the rate of acceptance of the (b)-sentence in the same condition (57% vs. 87%, t(29)=-3.07, p<.01). Conversely, a comparison between the two types of sentences in S2, representing the *inclusive* interpretation of *or*, is not statistically significant, even if a numerical difference is attested in the proportions of acceptance rate (90% vs. 76%, t(29)=-1.44, p=16, n.s.).

Analysing subjects’ judgements on the scale in case of “true” answers, an interesting comparison emerges: a highly significant difference in the mean scale rate in Condition II-S1 is attested between the two types of sentences, being only 3.47 for sentence of type (a) and 4.38 for sentence of type (b) (3.47 vs. 4.38, t(41)=-2.59, p<.01). Hence, subjects that accepted sentence of type (b) in a scenario only compatible with the *exclusive* interpretation of disjunction considered the (b)-sentence more appropriate in that condition than those who accepted sentence of type (a) in the same condition.

Data reported in Table 9 above are also interesting in another respect: reaction times to evaluate critical items in different conditions can be compared, considering overall mean RTs per sentence-type or distinguishing between RT to
answer “true” and “false” separately. The graphs below plots respectively: the mean RTs for both sentence types in the two conditions (Graph E) and the mean RTs differentiating between “true” and “false” answers (Graph F).

Considering the data plotted in the top graph first, a point worth of remark is the fact that no significant difference emerges from the comparison between the mean RTs for sentences of type (a) and (b) across Condition II ((a) vs. (b), t(29)=.084, p=.93 for S1 and t(29)=-1.27, p=.22 for S2, n.s.), nor from the comparison between the two levels of Condition I (S1 vs. S2, t(29)=.31, p=.76 for sentence of type (a) and t(29)=-1.01, p=.32 for sentence of type (b), n.s.). These results seem to indicate that the processing load required to evaluate both types of sentences in both conditions was almost identical, at least if we consider mean RT overall. However, this consideration should be handled with care, given that one need to integrate the overall picture with the data plotted in Graph F, showing RTs for both sentence types and situations differentiated this time between “true” and “false” type of answer. Taking predictions on RTs derived by Relevance Theory into consideration (cf. Table (8)), I first drawn a comparison on RTs for answering “true” between situations S1 and S2 and a comparison on RTs for answering “false” between the same conditions. Note that I didn’t differentiate

---

Graph E

**Graphs E and F. Mean RTs per sentence-type across conditions.**

Legend: (a)=sentence containing or in a DE context; (b)=sentence containing or in a NON-DE context; S1=scenario compatible with exclusive or; S2=scenario compatible with inclusive or.

Graph F

Mean RTs for answers “true” and “false”

---

83
data between the two sentence-types, on the basis that no difference due to the monotonicity properties of the two contexts is in principle expected in this approach. None of the comparisons turned out statistically significant (t(88)=.73, p=.47 for “true” and t(23)=-1.86, p=.08 for “false”, n.s.). In second place, always following the suggestions derived form Relevance Theory, I have drawn another comparison between the RTs for answering “true” and the RTs for answering “false” within the same condition. Statistical analysis doesn’t reveal any difference between the RTs for answering “true” vs. the RTs for answering “false” in Condition II-S2 (t(56)=-1.11, p=.27, n.s.). However, a significant difference in RTs between the two types of answers is found in Condition II-S1 (t(55)=2.14, p=.04). Let’s take a closer look at this crucial finding. This result seems in fact to be the only one consistent with the predictions made within the Relevance Theory approach. If we take a closer look at the data, however, we can observe that this result is in fact mostly due to the high RTs to answer “true” in case of sentence (a) in the situation considered. Crucially, if we compare this measure to the mean time to answer “false” when evaluating this sentence in this condition, the comparison turns out statistically significant (t(25)=2.21, p<.05). Most interestingly, none of the other comparisons that were drawn within the same sentence types of answers (“true” vs. “false”) in the same condition turned out statistically significant.

2.1.6 Discussion

Before starting our general discussion on the results obtained in this experiment, let’s review the predictions made above to see whether and to what extent they conform to the results obtained. One of the questions addressed in Experiment 2 (cf. (23) in section 2.1.4) is the existence of a default interpretation of the scalar terms in accordance with the syntactic context in which they are embedded. To answer this question empirically, two conditions were created: Condition I, which varies between two sentence-types that crucially display different entailing properties: (a)-sentences, representing DE contexts and (b)-sentences, representing NON-DE contexts; Condition II, instead, varies between
situation-types: scenario S1 is only compatible with exclusive interpretation of or while S2 in only compatible with the inclusive one. If the monotonicity properties of the context had an effect on the default interpretation of or, then one would expect a different rate of acceptance/rejection of the critical sentence-types across conditions. This rate would depend on the entailment relations that occur between the two interpretations of or in the context in which it appears, ultimately on the monotonicity properties of such a context: in DE contexts (i.e. sentences of type (a)), the inclusive is the most informative interpretation, while in NON-DE contexts (i.e. sentences of type (b)), the exclusive is the most informative one. Given these entailment relations, if subjects selected their interpretation of the scalar term by default obeying a general rule of strengthening, they should always prefer the most informative interpretation (once ALTs are active). In particular, they should assign or the inclusive interpretation when it is embedded in a DE context (i.e. rejecting the (a)-sentence in Condition S1 and accepting it in Condition S2) and assign it the exclusive interpretation when it is embedded in a NON-DE context (i.e. rejecting the (b)-sentence in Condition S2 and accepting it in Condition S1). In terms of a comparison between sentence-types across conditions, they should reject sentence (a) but accept sentence (b) in condition S1; conversely, they should reject sentence (b) and accept sentence (a) in condition S2. As we saw when presenting the results, these predictions were partially confirmed by the data collected. First of all, a different pattern of subjects’ responses was found between the two types of context (cf. Graphs C and D): sentences of type (a) representing a DE context are in fact significantly more accepted in a scenario compatible with inclusive or (i.e. S2) than in a scenario compatible with exclusive or (i.e. S1), where the acceptance rate is actually the lowest of all conditions (we’ll get back on this below). Secondly, a significant difference was found between sentence-types in Condition II-S1: subjects split in evaluating the (a)-sentence but overwhelmingly accept the (b)-sentence in such a condition, thus deriving the SI in NON-DE contexts. This result was corroborated also by the mean rate assigned on the scale: as we saw, subjects that accepted
sentences of type (b) in S1 selected significantly higher scores on the scale than those who accepted sentences of type (a) in the same condition.

Conversely, predictions are only partially confirmed for sentences of type (b), representing a NON-DE context. In particular, these sentences are overwhelmingly accepted in both situation types. In this respect, it seems that deriving the implicature or not doesn’t depend exclusively on the monotonicity properties of the context in which the scalar term appears. Moreover, the comparison between sentences of type (a) and (b) in S2, i.e. in a situation only compatible with the inclusive interpretation of or, is only partially in line with the predictions made: as expected, subjects overwhelmingly accept the (a)-sentence in such a condition, but also they tend to accept the (b)-sentence more than expected. As a consequence, the proportion of acceptance rate between the two sentence-types is only numerically but not significantly different. This comparison is shown by the green arrow in the graph below, which plots the proportions of subjects’ acceptance and rejection of the critical sentences across conditions.

Graph G. Proportions of subject’s acceptance and rejection of critical sentences across conditions. Legend: (a) and (b) = sentence-types; S1 and S2 = levels of Condition II (situation-types): S1 = scenario only compatible with exclusive or; S2 = scenario only compatible with inclusive or. Red arrows represent critical comparisons.
The results discussed above, however, are to be integrated with other considerations. In the first place, we said that flipping between having or not having the implicature is predicted to be much harder in DE than NON-DE contexts. On the basis of this consideration, I expected more subjects to “flip” when evaluating sentences of type (b) than when evaluating sentences of type (a). Building on this consideration, we can provide an explanation of the fact that more subjects than expected accepted the critical (b)-sentence in a condition only compatible with the inclusive interpretation of or (i.e. S2, cf. right-most bar in the graph), a result that seemed contrary to predictions at first glance. Precisely, given that flipping between having or not having the implicature is predicted to be relatively easy in NON-DE contexts, when facing a scenario that is only compatible with the inclusive interpretation of or, subjects may reset their initial selection from [+σ] to [-σ]. The same operation is predicted to be much more costly in case of DE contexts, because in this case the adding of the implicature leads to a general weakening of the assertion (failing to meet O requirements) and for this reason only half of the subjects add it, and, when they do, they assign a pretty low score to the sentence (as emerges from the mean scale rate assigned in this condition, cf. Table 9). For this reason I expected condition (a)-S1 to be the hardest condition of all and in fact this was the only condition in which subject split (cf. the left-most bar in the graph above).

Considering the predictions outlined in Table 7 with respect to feature selection, I believe that the only way to individuate subjects’ preference is to analyse the proportion of acceptance rate of sentences (b) in Condition II-S1. In this condition, the scenario is only compatible with exclusive interpretation of or. To accept the critical sentence in this condition subjects should unequivocally derive the SI: thus, they should activate the scalar ALT associated to or by selecting the feature [+σ]. Only in this case they would accept the critical (b) sentence. If, conversely, they followed a general strategy of selecting the feature [-σ], then they should unequivocally reject the (b)-sentence in situation S1. But, in fact, the results obtained for sentence (b) in this critical condition show that subjects overwhelmingly accept it, thus select the positive value of σ feature. It is
also conceivable to assume that the same strategy is maintained across different items during the same experimental setting.

In the remaining part of this discussion session I will take into consideration the second of the questions raised in (23) above (cf. section 2.1.4), namely whether the process of computing implicatures is costly. To this purpose, I will consider the RTs recorded in this experiment. As I said, measuring the time that subjects spend in evaluating a scalar sentence can be a way of quantifying the “cost” of implicature computation. Given that in this respect we are comparing two different approaches that crucially differ on the “cost” attributed to SI computation (cf. Table (8)), RT data can ultimately provide arguments for discarding one of these competing theories. Specifically, the crucial data are the times that subjects took to accept a sentence in a scenario compatible with exclusive interpretation of or (i.e. S1) and to reject it in a scenario only compatible with inclusive interpretation (i.e. S2): in both cases, in fact, their response signals the fact that the implicature has been derived. If, as Relevance Theory claims, the process of computing implicatures is costly, then we should find significantly higher RTs for these subjects. This, at first glance, seems partially confirmed by the data recorded: one of the crucial comparisons, namely, the one between RTs for answering “true” in Condition S1 and the RTs for answering “false” in the same condition (not differentiating between the two types of sentence) revealed a significant difference, as reported in the preceding section.

However, this does not suffice to support the claim that implicatures are costly, at least for two reasons. In the first place, none of the other comparisons listed in Table (8) to support Relevance Theory turned out statistically significant. In the second place, the result obtained in the above mentioned comparison is to be handled with care. As we have seen, the difference emerged is in fact mostly due to the high RTs to accept the (a)-sentence in condition S1, which are significantly different from the RTs to reject the same type of sentence in that condition. As reported in Table (8), this was the only comparison predicted as crucial by the grammar-driven approach. According to the framework I am adopting, no cost is to be associated to scalar implicature computation per se. A
cost is instead to be expected when the implicature is derived despite the fact that the scalar term is embedded in a DE context, in which the adding of the implicature would result in a weakening of the overall assertion. This is, in my opinion, the reason why only those subjects that accept the (a)-sentence in S1, thus deriving the implicature in a DE context, took significantly longer than the subjects that reject that sentence in the same condition. If the cost were to be attributed to implicature computation, then the same contrast should be found in case of sentence (b), but this is not so. To account for the data obtained in this on-line task, the claim that implicature computation is costly is, in my opinion, to be rejected.

At the same time, the claim that the default interpretation of the scalar term depends on the monotonicity properties of the context in which the scalar term is embedded is largely supported by the data obtained in this experiment: without such a claim, it would be difficult to account for the fact that sentence (a) in which or appears in a DE context, when evaluated in scenario S1 representing the exclusive interpretation, is the hardest condition of all, both in terms of subjects’ distribution and in terms of RTs.

2.2 Conclusive remarks on Experiment 2

The results obtained in this experiment seem to be in contrast with recent works on SI computation realized within the Relevance Theoretic tradition. In particular, I’m referring to the works by Noveck and Posada (2003), Bott and Noveck (2004), Breheny et al. (2005) and Katsos et al. (2005). By means of different techniques, these authors conducted on-line experiments on adults while evaluating underinformative sentences containing scalar terms such as some and or in different experimental “situations” (as for, e.g., presence or absence of a preceding biasing context, or different instructions/suggestions given to participants to fulfil the task). Very generally, the results of these studies seem to point to the same direction, namely: whenever subjects compute the implicature associated to a scalar term, they do it at a cost. This is reflected by a slow down in correspondence of the scalar trigger when measuring reading times (like in the
studies presented by Breheny and colleagues and Katsos and colleagues), or by an increased time to process the whole sentence (when measuring reaction times, like in the Bott and Noveck’s study or in the ERP study conducted by Noveck and Posada). These results were uniformly interpreted by these authors as evidence of the fact that scalar implicature computation is a costly process.

In this respect, I would like to point out that the “cost” of implicature computation has been the centre of the most recent debates between supporters of Relevance Theory on the one hand and of the Default approaches (as, e.g., that of Levinson, cf. Chapter 1, section 2.1) on the other. The claim that implicature is added at a cost by our processing system is necessary to differentiate Relevance Theory from Default approaches. On the latter approach, a default process is something that our computational/processing system performs automatically, thus it is by definition virtually costless. On the former, every operation imposed to our processing system must be evaluated in terms of “costs and benefits”, ultimately in terms of “relevance” to contextual assumptions, as we have discussed in Chapter 1 (cf. section 2.2): only those stimuli that are relevant enough are worth a processing effort. From this assumption, the claim that implicatures are costly necessarily follows: implicatures are only derived when explicitly required by the context, i.e. when the benefits that one gains from their computation overcome the processing effort required to derive them. If implicatures were costless, then the principle of optimal relevance would loose its foundation. This is the reason why all the experimental works on scalar implicatures within the Relevance Theoretic tradition have been focused on finding evidence of such a “cost”.

Without entering too much in the details of each single study, I would like to make some general considerations about the findings of the works mentioned above. In the first place, let’s consider subject’s distribution. It’s interesting to note that in most (if not all) cases subjects split when they have to judge an underinformative sentence, even when the sentence is given “out of the blue”, i.e. in the absence of a preceding context. This is a clear indication, according to my view, that subjects are adopting a strategy to which they stick when solving the experimental task: half of the subjects consider the computation of the implicature
“relevant enough” (to borrow from Relevance Theory terminology) and thus add the implicature; the other half, instead, keep to the plain meaning of the scalar term, and do not derive the implicature. I believe that the solution proposed by Chierchia (2006) well explains these facts, being feature selection the result of a subjective choice, thus a personal strategy. On the contrary, it’s more difficult to find a ready explanation of this split in subjects’ distribution within the Relevance Theory given that the presumption of optimal relevance of a given stimuli should in principle be the same across all participants.

On the other hand, RT data seems at first glance to be better explained by Relevance Theory. The crucial comparison, according to this approach, is between the RTs of the two groups: subjects that derive the implicature always take longer than the rest. This result suffice, according to them, to claim that the process of computing SI is costly and thus subjects only derive SIs when the benefits obtained by the adding of a SI exceed the processing effort required for its derivation. However, it’s not that clear that this overload is effectively due to the adding of the implicature per se. As discussed by Noveck and Posada, their results clearly show that subjects that behave “pragmatically”, thus adding the implicature (in their experiment, a total of 12 out of 19) were overall slower on all items (even on controls) compared to the other group of subjects. They explicitly talk about “two kinds of responders” that reflect “two sort of strategies” (cf. Noveck and Posada, 2003: 209). I believe in the end that this result doesn’t go against the theoretical framework I am adopting and that it doesn’t prove that implicature computation is a costly process per se. Quite the contrary, indeed.

In the end, I believe that the intriguing debate on pragmatic inference, which has very recently attracted the interest of psycholinguists, is far from being solved. To begin with, pragmatic inferences are a hard topic to investigate experimentally. Also, experimental work on this has just got under way. Lastly, the majority of the studies have been focused on measuring the “cost” of implicature derivation. Though interesting, I think that this is not the only question to be solved within a pragmatic theory. To thoroughly understand the phenomenon of SI computation one should first provide a systematic account of
which factors can favour or block its emergence. The experiments I have presented in this chapter were designed in this perspective and I believe they have provided interesting cues to pursue further.
CHAPTER 3

BEYOND SCALAR IMPLICATURES: INTRODUCTION TO THE PHENOMENON OF “DONKEY-SENTENCES”

As we will see in this chapter, scalar implicatures are not the only phenomenon to be linked to the strictly logical distinction of monotonicity. Sensitivity to the entailment properties of the context seems to emerge also in the interpretation we associate to a particular class of ambiguous sentences, known in the literature as “donkey sentences”. These offer an unusual window on the way we understand and solve ambiguities, taking part by right in our general discussion on spontaneous logicality in adults. The name “donkey sentence” is due to the original examples proposed by Geach,¹ which specifically were sentences of this form:

(1) Every man who owns a donkey beats it

In a sentence like (1), an indefinite DP (a donkey) is embedded in a relative subordinate clause introduced by a quantificational NP (every man) and is

¹The original examples proposed by Geach (1962) were of this form:
(i) If any man owns a donkey, he beats it
(ii) Any man who owns a donkey beats it
followed by the anaphoric pronoun *it* which calls for an interpretation by some preceding element in the clause. The only clause-internal possible candidate to serve this role is the indefinite *a donkey*. And to the way we interpret this indefinite as a referent to the anaphora, a long series of studies has been devoted by many contemporary linguists and philosophers in which the intricate questions related to three major topics in the semantic tradition, i.e., anaphora, binding and interpretation of indefinites conflated together in the attempt of solving the ambiguity associated to the Geachean examples. I will, very briefly, discuss some of the theories developed in the field of semantics in order to give the reader an idea of the theoretical problems raised, before introducing an experimental study designed to investigate how subjects interpret donkey sentences in neutral and biased situations, which will fill the remaining part of this chapter.

1. The puzzle of donkey anaphora resolution

Given the structure of the sentence in (1), a unique semantic representation was initially considered by Geach, the one given in (2):

\[
\forall x \left[ \text{man}(x) \land \exists y \text{ donkey}(y) \land \text{has}(x,y) \right] \rightarrow \forall z \left[ \text{donkey}(z) \land \text{has}(x,z) \rightarrow \text{beats}(x,z) \right]
\]

The formula in (2) can be paraphrased as (2'):

\[
\text{(2') Every farmer who owns a donkey beats all the donkeys he owns}
\]

This seemed to be the best approximation of the meaning of the original example in (1). However, this is not the only interpretation that can be associated to the pronoun. As observed by some authors at the beginning of the ‘90s, especially Chierchia (1992), there is a second interpretation that can be derived by the original donkey example proposed by Geach, namely, the one reported in (3) and paraphrased as (3'):
(3) \( \forall x \ (\text{farmer}(x) \land \exists y \ \text{donkey}(y) \land \text{has}(x,y)) \rightarrow \exists z \ (\text{donkey}(z) \land \text{has}(x,z) \land \text{beats}(x,z)) \)

(3') Every farmer who owns a donkey beats one of the donkeys he owns

The fact that this interpretation was more difficult to detect at first glance does not suffice to conclude that this reading of the anaphora is ruled out compositionally. Actually, this interpretation is even the most natural if one introduces the donkey sentence in a context such as the one reported below (on the lines of a scenario proposed by P. Casalegno and reported in Chierchia (1992), pg. 116) which seems to induce the alternative interpretation of the anaphora in a natural way:

(4) Suppose there is a spreading insomnia problem amongst farmers in Massachusetts so that all the farmers in the region contact the doctor to solve their problem. Detecting repressed anger as the cause of the insomnia, the doctor suggests the following remedy to vent their anger: choose one of their donkeys and beat it every day. This in order to eliminate the hidden cause of the insomnia. After the colloquium with their doctor…every farmer that owns a donkey beat it.

In a scenario such as the one reported in (4), it seems relatively easy to get to the interpretation in (3), according to which it suffices to beat one of the donkeys owned to vent anger and recover from insomnia (not ultimately, this alternative interpretation is even warranted in the sake of all the other poor donkeys owned by such terrible farmers!). Moreover, there are some sentences which even seem to lack the interpretation in (2) and only have the interpretation in (3), like examples (5) and (6) below, which are borrowed from the literature:
Every person who had a credit card paid his bill with it [Cooper]

Every person who has a dime will put it in the meter [Pelletier e Schubert]

As a matter of fact, it seems that the most natural interpretation of these sentences is the one reported below as (5') and (6'), which parallels interpretation (3) above:

(5') Every person who had a credit card paid his bill with one of his/her credit cards

(6') Every person who has a dime will put some of his/her dimes in the meter

and not the one that would be obtained by assuming the interpretation in (2), which would end up with the bizarre interpretations (5'') and (6''):

(5'') Every person who had a credit card paid his bill with each of his/her credit cards

(6'') Every person who has a dime will put each of his/her dimes in the meter

Cases such as (5) and (6), which supposedly are not unique in ordinary conversations, seem to confirm the fact that an interpretation of the pronoun alternative to the one traditionally assigned to it should be considered. This alternative interpretation, nevertheless, seems not to be available in all cases, given that examples such as (7), proposed by Heim, are clearly false in a situation in which a man (x) owns a slave (y) but not y’s offspring, which is a situation that parallels interpretation (3'):

(7) Every man who owned a slave, owned his offspring
At this point, I believe to have sketched the intriguing puzzle posed by donkey sentences introduced by the universal quantifier every in all its complexity: we have encountered the original Geachean example (1) which is usually interpreted as in (2), but can easily get the interpretation in (3) in the appropriate context, as shown in (4); examples (5) and (6) which seem to lack the Geachean interpretation proposed in (2); finally example (7) which conversely seems to lack the alternative interpretation (3) only being felicitous under interpretation (2). The picture complicates itself further if we add to our repertoire of donkey sentences some examples that are introduced by a head determiner different from the universal quantifier. To better understand the question and the proposed semantic solutions, we will focus our discussion on three particular constructions, which, following Chierchia (1992) in particular, are associated two alternative semantic interpretations of the pronoun they contain: the universal (∀ henceforth) and the existential (∃ henceforth) interpretation, which parallel interpretations represented in (2) and (3) above respectively.

2. Alternative interpretations of “donkey sentences”

We have already encountered in the preceding section one of the constructions that I’m going to discuss further, which is repeated here as (8) for convenience, followed by its two possible interpretations:

(8) **Every** farmer that owns a donkey beats *it*

(a) **∀-reading:**
\[ \forall x \left[ \left( \text{farmer}(x) \land \exists y \text{ donkey}(y) \land \text{has}(x,y) \right) \rightarrow \forall z \left[ \text{donkey}(z) \land \text{has}(x,z) \rightarrow \text{beats}(x,z) \right] \right] \]

(b) **∃-reading:**
\[ \forall x \left[ \left( \text{farmer}(x) \land \exists y \text{ donkey}(y) \land \text{has}(x,y) \right) \rightarrow \exists z \left[ \text{donkey}(z) \land \text{has}(x,z) \land \text{beats}(x,z) \right] \right] \]

direction of entailment: (a) \(\Rightarrow\) (b)
Paraphrasing, sentence (8) is true under interpretation (a), corresponding to the *universal* interpretation, when all the farmers that own some donkeys beat *all* the donkeys they own (and false if they beat only some of them); on the contrary, under interpretation (b), corresponding to the *existential* interpretation, sentence (8) is true when all the farmers that own some donkeys beat *at least one* of the donkeys they own (and supposedly not all of them). As one can observe, (8a) entails (8b): if every farmer beat *all* the donkeys they own, it will certainly follow that they beat *at least one* of the donkeys they have, while the reverse does not hold. When encountering a sentence such as (9), the interpretation that seems to be more frequently accessed in the ∀-interpretation. However, there are cases in which the other reading seems more natural, as we have seen when discussing examples (5) and (6) above, to which the ∃-interpretation is preferably assigned.

Let’s now turn to the second type of donkey sentence we are going to consider, the one introduced by the negative quantifier, which is reported in (9) together with its candidate interpretations:

(9) **No farmer that owns a donkey beats it**

(a) **∃-reading:**
\[ \neg \exists x \ [ \text{farmer}(x) \land \exists y \ \text{donkey}(y) \land \text{has}(x,y) \land \text{beats}(x,y) ] \]

(b) **∀-reading:**
\[ \neg \exists x \ [ \text{farmer}(x) \land \exists y \ \text{donkey}(y) \land \text{has}(x,y) \land \forall z \ [ \text{donkey}(z) \land \text{has}(x,z) \rightarrow \text{beats}(x,z) ] ] \]
**direction of entailment:** (a) \(\Rightarrow\) (b)

Under interpretation (a), which corresponds to the *existential* interpretation of the pronoun, the sentence in (9) is true if there is no man that beats the donkeys he owns (crucially, for the sentence to be true, he shouldn’t beat any of the donkeys he owns, *not even one*); under interpretation (b), instead, sentence (9) turns out to be true if no farmer beats all the donkeys he has (crucially, the sentence is true in a situation in which there is a farmer who beats *one or some*.
but not all, the donkeys he has). In this case, the \( \exists \)-interpretation is the strongest, entailing the \( \forall \)-interpretation. The former seems also the interpretation that is more frequently accessed when encountering sentences such as (10). However, there are cases in which it leaves place to the alternative \( \forall \)-reading, which seems to be the one assigned to examples of the sort of (10), borrowed from the literature:

(10)  No man that has an umbrella leaves it at home in a rainy day

Surprisingly, the most natural interpretation of a sentence like (10) is the universal interpretation (represented by (b)) according to which, if one has some umbrellas doesn’t leave all of his umbrella at home when it rains but, supposedly, will take at least one of the umbrellas he has with him.

Finally, let’s consider a third type of sentence, the one introduced by the existential quantifier, which is represented in (11), together with its alternative interpretations:

(11)  Some farmer that owns a donkey beats it

(a)  \( \exists \)-reading:

\[
\exists x \ [\text{farmer}(x) \land \exists y \ \text{donkey}(y) \land \text{has}(x,y) \land \text{beats}(x,z)]
\]

(b)  \( \forall \)-reading:

\[
\exists x \ [\text{farmer}(x) \land \exists y \ \text{donkey}(y) \land \text{has}(x,y) \land \forall z \ [\text{donkey}(z) \land \\
\text{has}(x,z) \rightarrow \text{beats}(x,z)]]
\]

\text{direction of entailment}: (b) \( \Rightarrow \) (a)

Here interpretation (b), which corresponds to the universal interpretation, entails (a), which represents the existential interpretation. As a matter of fact, under interpretation (b), the sentence in (11) is true when there is a farmer that owns some donkeys and beats all the donkeys he owns, while interpretation (a) requires that a farmer beats one or some of the donkeys he owns, not necessarily
all of them. In this case, the interpretation that is normally assigned to sentences like (11) seems to be the $\exists$-interpretation, reported in (a). Conversely from what observed for the other quantifiers, in case of some the most frequent reading is the weakest one, being entailed by the alternative $\forall$-reading. This $\forall$-interpretation turns out to be the most natural in examples like (12):

\begin{equation}
\text{(12) Some boy that has an apple doesn’t share it with his school-mates}
\end{equation}

Intuitively, what it seems reasonable to expect is that, if a boy has some apples, he won’t share any of the apples he has with his school-mates, which is exactly the interpretation we obtain by applying the reading in (b), considering the fact that (12) contains a negation, thus $\forall z [\text{apple}(z) \land \text{has}(x,z) \rightarrow \neg \text{share}(x,z)] \equiv \neg \exists z [\text{apple}(z) \land \text{has}(x,z) \land \text{share}(x,z)]$.

We have seen that for each type of donkey sentence, two alternative interpretations can be derived. Of these, one seems normally more readily accessed than the other, which is nevertheless attested in some examples. In case of the donkey sentences introduced by the universal and the negative quantifier, the preferred interpretation is the strongest one. Conversely, in case the existential quantifier introduces the donkey sentence, then the supposedly preferred interpretation is the weakest of the alternatives.

If it’s possible to construct two logical interpretations for each sentence type, are they both (always) available to the Grammar? Is it possible to talk about defaults in donkey anaphora resolution? And which, if any, is the reason for this default? How is the access to the alternative reading modulated, by which pragmatic factors/circumstances? Or else, can the availability of the two interpretations be reduced to lexical ambiguity or vagueness? To all these questions the rich semantic literature on this topic has not provided a conclusive answer yet. On experimental grounds, on the other hand, not many studies have been developed to deal with this intriguing thematic, except for some isolated works (as far as I know, three are the experimental works published on donkey anaphora so far: Conway and Crain (1995), presenting a study on children, and
the studies by Yoon (1996) and Geurts (2002) on adults). I will review some of the theoretical and experimental works in the next section.

3. The debate on ambiguity

It has been claimed by many parts that the possibility of two alternative resolutions of the donkey anaphora is the result of accidental, more or less vaguely defined, pragmatic factors and for this reason such a shifting in meaning has not been considered by some authors in a purely compositional perspective. Under these assumptions, donkey sentences are not judged ambiguous in a genuine sense, given that there’s only one logical representation of their meaning attested. The undisguised fact they can sometime display a different interpretation of the anaphora from the one predicted by the theory is explained by appealing to pragmatic/extra-linguistic interferences in the derivation of their interpretation. Along these lines are the solutions to the “donkey” debate proposed, for example, by Kanazawa (1994) and Lappin and Francez (1994).

Conversely, some semantic theories build on the consideration that two alternative interpretations are made available by the Grammar in the derivation of the meaning of donkey sentences, and the pattern of distribution of universal vs. existential readings is explained by general principles modulating the choice between the two alternatives. In this framework we will consider the contributions by Krifka (1996), who starts form the experimental work proposed by Yoon (1994/1996), and especially Chierchia (1992/1995).

3.1 Kanazawa (1994): preserving inferential patterns

Kanazawa (1994) starts his discussion on donkey anaphora from an observation found in Rooth (1987), which he formulates in form of a principle as “Rooth’s generalization”, according to which every determiner whose nuclear scope (i.e. 2\textsuperscript{nd} argument) is Downward Entailing imposes an existential
interpretation of the donkey sentence it introduces (cf. Kanazawa, 1994: 118). Kanazawa individuates the limits of this generalization in the fact that it builds only on the monotonicity properties of the 2nd argument of the determiners, while, according to his view, the relevant condition that determines the preferred reading of the anaphora is to be found in the monotonicity properties of the restrictor (i.e. 1st argument) of the determiners that introduce the donkey sentence. On purely semantic grounds, the explanation builds on the argument of the ‘preservation of inferential patterns’ due to the left monotonicity of the determiner and systematically predicts which, between a universal or existential interpretation of the pronoun, should be preferred depending on the semantic properties of the quantifier that embeds the antecedent of the anaphora. His generalization can be summarised as follows:

<table>
<thead>
<tr>
<th>Type of QUANTIFIER</th>
<th>Monotonicity (1st argument)</th>
<th>Expected Interpretation</th>
</tr>
</thead>
<tbody>
<tr>
<td>Every</td>
<td>↓</td>
<td>universal</td>
</tr>
<tr>
<td>No</td>
<td>↓</td>
<td>existential</td>
</tr>
<tr>
<td>Some</td>
<td>↑</td>
<td>existential</td>
</tr>
<tr>
<td>Most</td>
<td>↑↓</td>
<td>no preference predicted</td>
</tr>
</tbody>
</table>

Table 1: Kanazawa’s generalization. Expected preferred readings associated to the donkey anaphora depending on the left monotonicity properties of the quantifier that embeds it.

Depending on the monotonicity properties of the restrictor of the quantifier that introduces the donkey sentence, the interpretation of the anaphora is expected to be the following: universal in case of the universal quantifier and existential in case of the existential and negative quantifiers. In case of a determiner like most, which is not Upward nor Downward Entailing in its first argument, no preference is predicted. The explanation of why the expected distribution of the two interpretations patterns like this resides in the above mentioned principle of preservation of inferential patterns. This principle starts from the idea that the

---

2 cf. Rooth (1987). Please note that the original examples examined by Rooth contained always symmetric determiners, which were DE in both their arguments.
validity of the inferences based on the left monotonicity properties of the head
determiner, which are absolutely natural to derive, should be preserved in the
derivation from sentences devoid of anaphoric dependencies, like example (13), to
sentences that display such dependencies, like the donkey-example (14):

(13)
(a) Every farmer who owns a donkey is rich
(b) Every farmer who owns a female donkey is a farmer who
owns a donkey
(c) Every farmer who owns a female donkey is rich

(14)
(a) Every farmer who owns a donkey feeds it
(b) Every farmer who owns a female donkey is a farmer who
owns a donkey
(c) Every farmer who owns a female donkey feeds it

The derivation in (14) is valid only if the universal interpretation of the
pronoun in (14a) is adopted. On the contrary, if the existential interpretation were
adopted, then sentence (14a) would be true in a situation in which a farmer owned
two donkeys, a male and a female, and only fed the male. In this case, from two
ture premises (i.e. (14a) and (b)) the false conclusion (14c) would be derived and
the inferential pattern in (13) would not be preserved. Resorting to this
explanation, Kanazawa can account for the fact the weakest interpretation is the
one preferred in case of the donkey sentences introduced by some. To preserve the
natural inference in (13) and derive a valid conclusion, in fact, the existential
(weakest) reading of the anaphora ought to be selected.

In his systematic and simple account of donkey sentences, Kanazawa
deserves no (or little) place to the factors that can intervene in the selection of the

3 For the monotonicity properties of the quantifier that, being Downward Entailing in its 1st
argument, allows only inferences from a set to its subsets (cf. Chapter 1 section 2.3).
other alternative, such as peculiar lexical properties of the clause components (for examples, the semantic properties of the predicates) or extra-linguistic, pragmatic factors such as world-knowledge or conversational expectations (except for, maybe, plausibility judgements). In any case, his generalization on the distributions of the expected readings of donkey sentences seems widely agreed upon. And its underlying logic is remarkable for its neatness and simplicity.

3.2  *Lappin and Francez (1994): an E-type account of “donkey sentences”*

Lappin and Francez believe that an adequate theory of donkey anaphora should outline a unique compositional procedure to derive a semantic representation capable of supporting both readings. In their *E-type* approach, donkey pronouns are treated as functional expressions $f(x)$ in which $f$ denotes a function from individuals $x$, belonging to a certain contextually determined set, to sum of individuals, each of which stands in a certain (contextually varied) relation with $x$. In case of the pronoun *it* in the donkey sentences, the function that represents the anaphoric reading is a function that associates, to each argument $a$ (for example, farmers), a sum of individuals $i$ (for example, donkeys) that stand in a certain relation with $a$ (for example, are owned). The resulting function would be something of the following form: $f = \lambda a \text{[THE DONKEY(S) THAT } a \text{ OWNS]}$

This function is claimed to be maximal by default, in the sense that it associates each farmer with the maximal sum of the donkeys he owns, thus generating (by default) the strongest interpretation of the pronoun. However, this is not the only possibility, given that, depending on contextual factors, the function can also be intended as non-maximal, in the sense of being a choice function that associates each farmer with a (contextually) appropriate sum of the donkeys he owns (not necessarily the total amount of them). This solution is actually criticized by Chierchia (1995) who claims that the ambiguity problem is not solved but only shifted in their theory: here ambiguity doesn’t affect the determiners but affects the function associated to the donkey pronoun which should in turn be considered ambiguous to account for the occurrence of both readings. As for their distribution, Lappin and Francez’s proposal builds on the
consideration that some predicates impose certain restrictions on the cardinality of sums of individuals denoted by their plural arguments. For example, predicates of the form “a puts his dimes in the meter” (cf. example (6) above), in which the plural description “his dimes” denotes a certain sum $S$, are intended to impose certain restriction on the cardinality of $S$, excluding in the specific case that this equals the total number of the individuals in the set, given the property of the predicate “put-in-the-meter ($a$, dimes)” to generally exclude the proposition that “a puts all his dimes in the meters”. This restriction on the cardinality imposed by the specific relational verb under consideration is responsible for the $\exists$-reading of the pronoun in sentences like (6). On the other hand, these authors are forced to assume that such a restriction imposed by the predicate is cancelled under the effect of negation, building on the consideration that there isn’t any limit, pragmatically justified, on the number of dimes that one is supposed not to put in the meter. This step is necessary to account for the fact that the most natural interpretation of the negative counterpart to (6), reported below as (15), is what we have labelled as the $\exists$-reading of the anaphora, paraphrased as (15’), in which the function $f$ that associates to each argument $a$ (in this case, person) a sum of individuals $i$ (in this case, dimes) seems again maximal by default, in the sense that it associates each person with the maximal sum of the dimes he owns:

\[
(15) \quad \text{No person who has a dime will put it in the meter}
\]

\[
(15') \quad \text{No person who has a dime will put any of his dimes in the meter}
\]

3.3 Yoon (1994/1996): an experimental investigation on alternative readings’ distribution

Along the lines of a grammatical account of the distribution of the alternative readings is the contribution by Yoon (1994 and 1996). Starting from the consideration that donkey pronouns can be paraphrased as plural definite descriptions, Yoon explains the alternation of the two readings in interpreting donkey sentences by appealing to the same mechanism that is responsible for the alternation of strong (= universal) vs. weak (= existential) readings of plural
predications. Such a mechanism, she claims, is affected by specific semantic properties of the predicate (in tandem with the monotonicity properties of the matrix clause). In particular, she focuses her attention on two classes of predicates that are semantically classified as total vs. partial and stative vs. episodic.

With respect to the first distinction, these are two paradigmatic examples she proposes with their candidate interpretations:

(16) The windows are closed \(\approx\) all the windows are closed
(17) The windows are open \(\approx\) at least one of the windows is open

The different interpretation assigned to these otherwise similar sentences depends on the semantic property of the predicate they contain: being closed is in fact a total predicate, thus requires a strong reading of the plural predication, i.e. the predicate applies to every part of the individual predicated, while being open is a partial predicate, thus requires a weak reading according to which, to render the sentence true, it suffices that the predicate applies only to a/some (contextually determined) part/s of the plural NP.\(^4\)

The second distinction advocated by Yoon to explain the alternation between strong vs. weak readings is exemplified by the examples below, in which static vs. episodic predicates\(^5\) are inserted in the sentences:

(18) The children are five years old \(\approx\) all the children are five years old
(19) The children are making a toy plane \(\approx\) at least some of the children are making a toy plane

\(^4\) Consider the pair clean (= total) vs. dirty (= partial): intuitively, to claim that a glass is clean it must be clean in all its parts; conversely, to claim that a glass is dirty it suffices that it has a tiny spot on it (namely, that not all its parts are clean). If CLEAN\(x\) indicates that all parts of \(x\) are clean and \(\subseteq\) stands for the part relation, then we can translate the predicates dirty/clean as follows: i. dirty = \(\lambda x \exists y [y \subseteq x \land \text{DIRTY}(y)]\); ii. clean = \(\lambda x \neg \exists y [y \subseteq x \land \text{DIRTY}(y)]\).

\(^5\) This distinction is analogous to the one proposed by Kratzer (1995) between individual vs. stage level predicates.
Responsible for the universal (strong) reading of example (18) is the static predicate “being five years old” which has to be satisfied by all members of the set denoted by the NP, while the existential (weak) reading of example (19) is due the presence of an episodic predicate according to which, for the sentence to be true, it suffices that only a (contextually determined) part of the plural NP has the property predicated.

Analogously, donkey pronouns alternate between universal and existential interpretation depending on the type of predicate contained in the matrix clause. For example, taking into consideration the first distinction proposed by Yoon, this is how the two readings come about:

(20) Every man who has a garage with a window keeps it closed while he is away
    ≈ Every man who has a garage with a window keeps every window closed while he is away [∀-reading]

(21) Every man who has a garage with a window keeps it open while he is away
    ≈ Every man who has a garage with a window keeps at least one window open while he away [∃-reading]

Yoon conducted an experiment to test her predictions. She presented 50 subjects with a situation description and then asked them to judge if a sentence correctly described it. All the target sentences contained a restrictive relative clause and were varied among three conditions: they were introduced by the definite article the (like “The children are sick/healthy”), by the universal quantifier every or by the negative quantifier no (these were in fact donkey sentences like “Every/No farmer who owned a donkey kept it healthy/let it get sick during the rainy season”). Predicates varied with respect to the total vs. partial (like being healthy/sick) and the static vs. episodic (like being five years old/flying a kite) distinction. What she found is that, in case of simple sentences introduced by The and in donkey sentences introduced by every, the percentage of
a strong \((universal)\) reading of the pronoun correlates with the semantic properties of the predicate, being significantly higher for \textit{total} (vs. \textit{partial}) and for \textit{static} (vs. \textit{episodic}) predicates. The pattern of responses for the sentences introduced by the negative quantifier, instead, is different, given that a general preference for the \textit{existential} (weak) reading seems to emerge overall. However, as discussed by Geurts (2002, see his comments on page 133-134, which I won’t address here) the results obtained for \textit{no} can be reduced to an experimental artifact.

The interpretation associated to the anaphora in the donkey sentences in Yoon’s experiment came as a surprise because the variable in such sentences is a singular pronoun \((it)\), thus in such a situation the notion of \textit{total} vs. \textit{partial} predicate should not apply. The explanation she proposes is that the pronoun \textit{it} appearing in donkey sentences, though singular in form, should be interpreted as the sum of individuals which its antecedent can be anchored to. The predicate on this variable then is predicated on this sum of individuals and leads to different interpretations depending on the semantic properties of the predicate itself: if it is \textit{total} or \textit{static} then the interpretation of the pronoun is \textit{universal}, if the predicate is \textit{partial} or \textit{episodic} then the interpretation of the pronoun becomes \textit{existential}.

\section*{3.4 Krifka (1996): the rule of Pragmatic Strengthening}

Taking Rooth’s generalization as a starting point (cf. section 3.1), and building on the empirical data presented by Yoon (cf. section 3.3), Krifka (1996) proposes a theory of the distribution of \textit{universal} vs. \textit{existential} readings of donkey pronouns (and plural predications in general) that accounts both for the correlation captured by Rooth between donkey anaphora resolution and entailing properties of the quantifier and for a fact that remained unexplained in Yoon’s grammatical approach. Regarding this latter point, he observes, for example, that reducing the distinction between \textit{total} vs. \textit{partial} predicates on purely lexical/semantic grounds, as Yoon did, can be problematic. This, in fact, doesn’t leave any place to pragmatic considerations such as which state of affairs are expected or preferred in a given situation that, in his view, can affect the interpretation of the predicates. For example, in a context in which it is of vital
importance to leave all the windows of a certain room open to guarantee the necessary amount of oxygen, the interpretation of the predicate \textit{open/closed} seems reversed with respect to the one associated to examples (16) and (17) above. In such a context, in fact, a sentence like (17) above, repeated here as (22):

(22) The windows are \textit{open}

would certainly be interpreted as stating that “we’re ok, \textit{all} the windows are open”, because this is how the windows in that room are expected to be. On the other hand, a sentence like (16) above, repeated here as (23):

(23) The windows are \textit{closed}

would immediately cause panic, because it would be interpreted as “warning! There’s \textit{at least one} window closed” (i.e. “\textit{not all} the windows are open”).

Besides, Krifka also observes that other types of predicates that fall outside the \textit{total/partial} distinction investigated by Yoon normally get a strong reading when inserted in an affirmative plural predication (i.e. in Upward-Entailing environments), while they get a weak interpretation when inserted in a negative plural predication (i.e. in Downward-Entailing environments)\(^6\), as exemplified below:

(24) The windows are made of security glass \approx \textit{all of them} are made of security glass

(25) The windows are NOT made of security glass \approx \textit{not even one} of them is made of security glass

\(^6\) Actually, Krifka attributes this observation to Lappin (1989), formulating what he calls the “Lappin’s generalization”: a non-collective predication \(P(x)\) on a sum individual \(x\) is preferably interpreted as follows:  
  \begin{enumerate}
    \item \(\forall y \ [y \subseteq x \rightarrow P(y)]\) if \(P(x)\) is in an UE environment
    \item \(\exists y [y \subseteq x \land P(y)]\) if \(P(x)\) is in an DE environment
  \end{enumerate}
Comparing the preferred and the dispreferred interpretations in terms of their logical strength, he observes that the preferred ones are normally the strongest. Thus, he proposes the following two rules to explain the preference for universal readings in Upward Entailing environments (and in Donkey sentences introduced by UE operators in their nuclear scope, as attested by Rooth’s generalization) and existential readings in Downward Entailing environments (and in Donkey sentences introduced by DE operators in their nuclear scope):

(26) “If a predicate P applies to a sum of individuals x, grammar does not fix whether the predication is universal (\(\forall y \ [y \subseteq x \rightarrow P(y)]\)) or rather existential (\(\exists y \ [y \subseteq x \land P(y)]\)), except if there is explicit information that enforces one or the other interpretation.”
[Krifka, 1996: 13]

(27) “If grammar allows for a stronger or weaker interpretation of a structure, choose the one that results in the stronger interpretation of the sentence, if consistent with general background assumptions!”
[Krifka, 1996: 13]

As for rule (26), explicit information that can enforce one interpretation over the other can for example be represented by a predicate that is not neuter with respect to the total/partial distinction, following Yoon. As a matter of fact, this distinction, though not purely grammatical as claimed by Yoon, can intervene in the alternation of universal vs. existential readings of donkey pronouns, as the examples below confirm:

(28) No man who has a garage with a window keeps it closed while he is away \(\approx\) No man who has a garage with a window keeps every window closed while he is away       \([\forall\text{-reading}]\)
(29) No man who has a garage with a window keeps it open while he is at home \(\approx\) No man who has a garage with a window keeps any window open while he is at home       \([\exists\text{-reading}]\)
Despite the presence of a Downward Entailing operator (which, following Rooth, would select an existential interpretation) the pronoun in (28) is interpreted universally, and this shift in reading is attributed by Krifka to the presence of a total predicate like “being closed” in the donkey sentence. Conversely, the presence of the partial predicate “being open” results in a preference for the existential interpretation of the anaphora in (29).

In terms of the question under discussion, about the availability of two different readings of donkey sentences as input for semantic interpretation, the theory proposed by Krifka seems to support the hypotheses that two equally plausible readings are generated by the Grammar (cf. (26)) and then selected according to pragmatic rules such as (27), or influenced by semantic/pragmatic factors such as the type of predicate used (like in example (28)), or by the context of utterance (like in examples (22)-(23) above).

3.5 Chierchia (1992/1995): the availability of two alternative readings

If one probes Chierchia’s bibliography with the aim of finding a work specifically devoted to donkey sentences, he won’t find any. However, the problematic associated to this topic has constituted an important issue in the development of the theory of dynamic binding presented in his 1992 and 1995 works. Crucially, his 1995 book opens with a trio of sentences to exemplify the anaphora and binding issue in which the third is the canonical Geachean donkey sentence reported at the beginning of this chapter. Before discussing this example, consider sentence (30), taken from Chierchia and reported below:

\[(30) \quad [\text{No student}]_i \text{ believes that he}_i \text{ is treated fairly}\]

In this sentence, the anaphoric link (expressed by co-indexation) between the pronoun he and its antecedent no student is well-formed, because the quantified NP precedes (and c-commands) the pronoun it binds. In sentences such as (31), instead, the only possible binder of the anaphora cannot c-command the
pronoun because it is inserted in a restrictive relative clause, thus the sentence is ungrammatical:

(31) *Every professor that met [no student], will soon meet him,

This same problem should be found in donkey sentences such as (1), repeated here as (32) for convenience, in which the c-command relation, which is supposed to be necessary for the well-formed anaphoric links, is absent:

(32) Every farmer who owns [a donkey], beats it,

In analogy to what happens in ungrammatical sentences such as (31), the indefinite *a donkey* that binds the anaphora appears in the restrictive relative clause, not c-commanding it. Yet, the sentence in (32) is felicitous and interpretable. Analysed in classical DRT approach, such a sentence would only get the $\forall$-interpretation, given that indefinites in the restriction of the universal quantifier systematically get universal force (rather informally: $\forall_{x,y} [\text{farmer}(x) \land \text{donkey}(y) \land \text{owns}(x,y)] \rightarrow [\text{beats}(x,y)]$. As we have seen in the introductory section to this chapter, however, this interpretation is not the only one that can be assigned to sentences like (32). To cope with facts that remained unexplained by classical DRT approaches, Chierchia proposes a variant of it, namely the Dynamic Binding approach, which constituted the core of his 1995 project. A consequence of this approach is that it derives only $\exists$-readings for donkey sentences. To integrate $\forall$-readings into this picture, a step considered necessary to account for the interpretation of donkey sentences introduced by *every*, for example, Chierchia draws an analogy with the interpretations of unbound pronouns in classical examples such as the following, due to B. H. Partee:

(33) Either Morrill Hall doesn’t have [a bathroom], or it is in a funny place
This is an example in which anaphora across inaccessible domain seem nevertheless to function, given that we can make sense of such a sentence despite the fact that the pronoun is not properly (semantically) bound. Resorting to the E-type strategy, in fact, we can retrieve the content of the unbound pronoun through context. Translating it as an E-type pronoun means to consider it as a function from individuals (in this case, places) to individuals (in this case, bathroom located in those places), where the nature of this function is contextually specified (we are assuming, in fact, that every building has at least one bathroom).\footnote{Note that the sentence is perfectly acceptable even if it turns out in the end that Morrill Hall has in fact two bathrooms. As a matter of fact, no uniqueness presupposition applies here.} According to Chierchia, this same strategy (which is generally and independently available) is responsible for $\forall$-readings of donkey sentences, which couldn’t be derived in the dynamic binding framework alone. Hence, to interpret the donkey pronoun in sentence (32) we can resort to an E-type strategy parallel to what used for sentence (33): the subject NP brings to salience in the context a function from farmers into the donkey(s) they own, which delivers us the intended $\forall$-readings without further ad hoc postulations.

In terms of simplicity of the formal architecture, the solution proposed by Chierchia to the question of the alternative readings of donkey sentences is remarkable given that it makes use of mechanisms inherited from existing theories that apply independently to broader phenomenon than donkey sentences. Moreover, his solution prevents the resort to any type of lexical ambiguity, which is viewed suspiciously by Chierchia on the basis of a simple empirical consideration: no language in the world seem to overtly realise a lexical ambiguity between, for example, $\text{every}_\exists$ and $\text{every}_\forall$, showing different words or morphemes to express the two readings of the same quantifier. This fact is indeed surprising if it the two interpretations associated to the donkey pronouns really rested on lexical ambiguity of the determiners that embed them.

To sum up, the position held by Chierchia with respect to the question under discussion is pretty sharp: in case of donkey sentences, two possible
readings are made available by the Grammar, and then one pops out as favoured in certain constructions, as clearly stated in the following passage:

(34) “I will assume that both ∃-readings and ∀-readings are generally available, with some sentences favoring one over the other.”
[Chierchia, 1995: 65]

This is exactly what happens in case of other ambiguous structures like, for example, quantifier scope ambiguities in which multiple readings are generated:

(35) “For sentences that contain multiple quantifiers, the grammar generates all possible scopings, but some such scopings may be hard or impossible to get, for a variety of reasons (such as the nature of the common ground, processing factors, etc.”
[Chierchia, 1995: 65]

We have sketched above how the two readings are generated in the Dynamic Binding system proposed, which was the main goal of the work under analysis. About the distribution issue, Chierchia himself admits that he is not going to say “anything as to why certain contexts favour certain readings” [Chierchia, 1995: 65].

At this point of our discussion, I believe it is time to face the question experimentally, given that only experimental investigation can shed light on the issue of alternative interpretations of donkey sentences and ultimately decide among competing theories about their availability and distribution.
Rethinking the donkey issue more in an experimental perspective, we can build on some of the semantic considerations presented in the preceding chapter and test different hypotheses that can in principle be derived from their semantic apparatus. Two questions will constitute the guiding lines of the experimental investigation that I will present in this chapter:

(1)

i. which, if any, between the existential and the universal readings of the donkey anaphora is derived (or preferred) depending on the monotonicity properties of the context?

ii. can extra-linguistic context affect this (default) interpretation?

Actually, both questions can be reformulated in these terms:
iii. are both $\exists$-readings and $\forall$-readings made available by the computational system?

As far as this last question is concerned, two alternative hypotheses can be derived, as in (2):

(2)

$H_{i}$ two readings are actually generated by the computational system and then one is selected (by default, as the result of pragmatic rules, or for semantic properties of some clause components, or else for the interference of extra-linguistic factors);

$H_{ii}$ only one interpretation is actually made available by the computational system (for example on the basis of the monotonicity properties of the syntactic environment in which the anaphora appears) the other one being ruled out.

Crucially, different answers to the interrogatives outlined above will provide a cue as to which hypothesis in (2) is to be rejected. For this reason, I addressed the questions in (1) by means of a reaction-time study which comprises two parts: a first experiment, labeled Experiment 3A, specifically addresses question (i), investigating which interpretation, if any, emerges as preferred in the evaluation of donkey sentences in the absence of an extra-linguistic context. A second experiment, labeled Experiment 3B, addresses question (ii) instead, investigating the effect of a preceding extra-linguistic context on donkey anaphora resolution. The comparison and mutual integration of the results obtained in these two experiments will provide a first attempt to answer question (iii) and thus an indication as to which, between $H_{i}$ and $H_{ii}$, is more plausible given the empirical facts. As for methodology, Experiment 3 actually exploits a design similar to the reaction-time study used to test the effect of monotonicity on the interpretation of
or presented in the preceding chapter (cf. Chapter 2, section 2.1). Analogous material to that experiment was also used, as we will see in the description.

1. **Experiment 3A - Between universal and existential interpretation of the anaphora: is there a default?**

1.1 **Subjects**

A total of 60 subjects participated in the whole study: 30 of them were randomly assigned to Experiment 3A and 30 to Experiment 3B. Participants were mainly 1st year students at the Psychological Faculty of the University of Milano-Bicocca, and received credits for their participation. For Experiment 3A only (but note that the material used in Experiment 3B was the same) subjects were randomly assigned to one of three lists, in which the same type of sentences was presented by means of different material, in order to have a control on the type of the material used.

1.2 **Procedure**

The same procedure used for Experiment 2 on scalar implicatures was adopted for both parts of Experiment 3, which were designed and run using E-Prime (cf. Chapter 2, section 2.1.3 for more details on the procedure adopted). Stimuli, comprising a critical sentence and four figures, were presented on a laptop screen. Subjects were instructed to evaluate the sentence by pressing different keyboard keys, corresponding to answers “true” or “false”. In case they answered “true”, they had to express a second judgement about the “goodness” of the sentence using a scale 1 to 5. Time to answer “true” or “false” was recorded by the program. Differently from Experiment 3A (and Experiment 2), in Experiment 3B the screen with the critical sentence and the pictures was preceded by another black screen displaying a paragraph that appeared in white in the middle of it. The paragraph was also presented in auditory mode: all the oral parts were previously recorded and then automatically played during the execution of
the experiment. Subjects were provided with headphones to better hear the audio parts. Aim of the paragraphs was to bias the sentence towards one of the two possible readings of the ambiguous critical sentence displayed in the successive screen.

Before entering the details of the experimental design, an important note on the material employed in both experiments is due. As described for Experiment 2, all the sentences contained fantasy names for characters and objects to avoid any interference from extra-linguistic sources. For donkey anaphora, in fact, names and predicates have revealed to be particularly influent in the resolution of the ambiguity (see, for example, Yoon’s account in Chapter 3, section 3.3). Thus, to avoid any kind of interference, an accurate choice of the predicate has been made, in order to remain as neutral as possible with regards of the total/partial or stative/eventive distinctions: only verbs of the same form (conservare in un barattolo (lit. preserving in a jar), tenere dentro un secchio (lit. keeping in a bin), proteggere con una foglia (lit. protecting with a leaf)) were used, all conjugated in the Simple Present tense. The choice of using “real” verbs instead of fantasy predicates was made in the end because it seemed that using imaginary names for actions constituted an extra element of possible misunderstanding of the pictures. In fact, it was impossible to recreate an “imaginary” action, given that every action that one could invent always resembled a real “earthly” action, no matter which creative name one invented to refer to it. For characters and objects, instead, this problem didn’t emerge given that the objects chosen for the pictures did not resembled anything one could encounter in ordinary situations, rendering the recourse to a fantasy name extremely natural. Given that time to read and evaluate the sentence was recorded, all the sentences had exactly the same length in their Italian version, and all the pictures had the same complexity.

1.3 Material

This experiment displayed a 3×3 condition design in which three conditions were created as a within subject factor, each displaying three different levels (cf. Appendix C). As first condition, I considered the type of sentence
containing the anaphora, which could either be introduced by the universal quantifier *every* (it. “ogni”), the negative quantifier *no* (it. “nessuno”) or the existential quantifier *some* (it. “qualche”). Condition I is exemplified below by abstract examples:

(3) Condition I: Sentence-type

(a) Every P [that has an A] VP
(b) No P [that has an A] VP
(c) Some P [that has an A] VP

In second place, these sentences were presented to each subject in three different situations (corresponding to levels (1)-(2)-(3) of Condition II that will be described shortly) modulating the interpretation associated to the anaphora by means of a scenario represented by a set of pictures. As we have discussed in Chapter 3, not only two interpretations can be associated to the anaphora but also these alternative interpretations are not logically unrelated. To better understand the design of the experiment, let’s review the two interpretations (*universal* vs. *existential*) associated to sentences of types (a)-(c) given in (3) and their mutual entailment relations, by means of a concrete example taken from the material actually used in the experiment:

(4)

(a) Every Flont that has a vilp keeps it in a bin

\[\forall\text{-reading: } (\bullet)\]

\[\forall x[[\text{Flont}(x) \land \exists y \text{ vilp}(y) \land \text{has}(x,y)] \rightarrow \forall z[\text{vilp}(z) \land \text{has}(x,z) \rightarrow \text{keeps it in a bin}(x,z)]]\]

\[\exists\text{-reading: }\]

\[\forall x[[\text{Flont}(x) \land \exists y \text{ vilp}(y) \land \text{has}(x,y)] \rightarrow \exists z[\text{vilp}(z) \land \text{has}(x,z) \land \text{keeps it in a bin}(x,z)]]\]

direction of entailment: \((\bullet a_\forall) \Rightarrow (a_3)\)
(b) No Flont that has a vilp keeps it in a bin

∃-reading: (●)
\[\neg \exists x[\text{Flont}(x) \land \exists y \text{ vilp}(y) \land \text{has}(x,y) \land \text{keeps it in a bin}(x,y)]\]

∀-reading:
\[\neg \exists x[[\text{Flont}(x) \land \exists y \text{ vilp}(y) \land \text{has}(x,y) \land \forall z[\text{vilp}(z) \land \text{has}(x,z) \rightarrow \text{keeps it in a bin}(x,z)]]]\]

direction of entailment: (●b) \(\Rightarrow\) (b∀)

(c) Some Flont that has a vilp keeps it in a bin

∃-reading: (●)
\[\exists x[\text{Flont}(x) \land \exists y \text{ vilp}(y) \land \text{has}(x,y) \land \text{keeps it in a bin}(x,z)]\]

∀-reading:
\[\exists x[[\text{Flont}(x) \land \exists y \text{ vilp}(y) \land \text{has}(x,y) \land \forall z[\text{vilp}(z) \land \text{has}(x,z) \rightarrow \text{keeps it in a bin}(x,z)]]]\]

direction of entailment: (c∀) \(\Rightarrow\) (●c∃)

The predicted preferred reading (according for example to Kanazawa’s generalization) is marked with (●): it corresponds to the universal (∀) interpretation of the anaphora in case its antecedent is embedded in a relative clause introduced by the universal quantifier (sentence of type (a)); it corresponds to the existential (∃) interpretation if it is embedded in a relative clause introduced by the negative or the existential quantifier instead (sentences of type (b) and (c) respectively). Most importantly, one interpretation always entails the other, but not vice-versa, as shown below each pair. As one can note, in case of sentences of type (a) and (b), the preferred interpretation coincides with the strongest, the entailing one; on the contrary, in case of sentences of type (c), the preferred interpretation is the weakest, thus the entailed one.

These observations on entailment relations will reveal crucial in understanding how the three levels for Condition II were created. These are described below:
(5) Condition II: Situation

S1 (NDT):
a situation which is compatible with both candidate interpretations of the anaphora, thus rendering the sentence true on both readings (I will refer to this condition as the non-differentiating true situation = NDT);

S2 (NDF):
a situation which is incompatible with both candidate interpretations of the anaphora, thus rendering the sentence false on both readings (I will refer to this control condition as the non-differentiating false situation = NDF);

S3 (DC):
a situation which is compatible only with one of the two possible resolutions of the anaphora, thus rendering the sentence true on one interpretation but crucially false on the other. Notice that, since the interpretations are not logically independent of one another, there can be only one kind of differentiating situation per sentence (I will refer to this condition as the differentiating-critical situation = DC).

To represent these conditions, four pictures appeared on the screen: three pictures contained an alien P (for example, a Flont) with two, three or four As (for example, vilps) and some other objects to keep or protect these As, such as bins, leaves or jars; the fourth picture always functioned as a distractor and contained one alien P with two earthly objects instead of alien objects. Different combinations of characters and objects were composed to represent the different situations of Condition II for all sentence-types, as summarised in the chart below:
Table 1. Experimental design. Description of the scenarios associated to each situation-type (Condition II) and sentence-type (Condition I).

<table>
<thead>
<tr>
<th>Cond. I</th>
<th>Cond. II</th>
<th>PICTURE 1</th>
<th>PICTURE 2</th>
<th>PICTURE 3</th>
<th>PICTURE 4</th>
</tr>
</thead>
<tbody>
<tr>
<td>S1=NDT</td>
<td>all vilps are in a bin</td>
<td>all vilps are in a bin</td>
<td>all vilps are in a bin</td>
<td>1 object is in a bin</td>
<td></td>
</tr>
<tr>
<td>a</td>
<td>S2=NDF</td>
<td>1 vilp is in a bin</td>
<td>1 vilp is in a bin</td>
<td>no vilp is in a bin</td>
<td>1 object is in a bin</td>
</tr>
<tr>
<td></td>
<td>S3=DC</td>
<td>1 vilp is in a bin</td>
<td>1 vilp is in a bin</td>
<td>1 vilp is in a bin</td>
<td>1 object is in a bin</td>
</tr>
<tr>
<td>S1=NDT</td>
<td>no vilp is in a bin</td>
<td>no vilp is in a bin</td>
<td>no vilp is in a bin</td>
<td>1 object is in a bin</td>
<td></td>
</tr>
<tr>
<td>b</td>
<td>S2=NDF</td>
<td>1 vilp is in a bin</td>
<td>1 vilp is in a bin</td>
<td>2 vilps are in a bin</td>
<td>1 object is in a bin</td>
</tr>
<tr>
<td></td>
<td>S3=DC</td>
<td>1 vilp is in a bin</td>
<td>2 vilps are in a bin</td>
<td>3 vilps are in a bin</td>
<td>1 object is in a bin</td>
</tr>
<tr>
<td>c</td>
<td>S1=NDT</td>
<td>no vilp is in a bin</td>
<td>all vilps are in a bin</td>
<td>no vilp is in a bin</td>
<td>1 object is in a bin</td>
</tr>
<tr>
<td></td>
<td>S2=NDF</td>
<td>no vilp is in a bin</td>
<td>no vilp is in a bin</td>
<td>no vilp is in a bin</td>
<td>1 object is in a bin</td>
</tr>
<tr>
<td></td>
<td>S3=DC</td>
<td>1 vilp is in a bin</td>
<td>no vilp is in a bin</td>
<td>2 vilps are in a bin</td>
<td>1 object is in a bin</td>
</tr>
</tbody>
</table>

In order to give a clearer idea of the experimental material, I report some of the scenarios presented for each sentence type below. For example, (6) represents the differentiating critical condition (DC) for sentence of type (a) (to abbreviate, let’s label it condition a-S3). As one can see, every alien in the pictures keeps only one of the objects he owns in a bin. Hence this scenario is only compatible with the existential interpretation of the anaphora, which is the
dispreferred one according to Kanazawa’s generalization. If the $\forall$-reading happens to be the preferred one, subjects should reject the critical sentence in such a condition.

(6)

Scenario (7), instead, represents condition S1 (NDT) for the same type of sentence (i.e. condition a-S1), which was presented to a different list of subjects. This scenario actually represents the universal interpretation of the anaphora, given that all the aliens in the pictures keep all their objects in a bin. As seen in (3), however, this interpretation is the strongest in case of sentences introduced by every and thus it entails the other interpretation. For this reason, the scenario is actually compatible with both interpretations given that, whenever the sentence is true under the strongest ($\forall$) interpretation, it will also be true under the alternative ($\exists$) interpretation. In this case, true answers at ceiling are expected, independently from the interpretation assigned to the donkey anaphora:
Below are the critical condition DC (i.e. S3) and NDT (i.e. S1) for sentence of type (b). Precisely, (8) represents the critical condition for the (b)-sentence. As one can note, the scenario is only compatible with the universal interpretation of the pronoun given that all the aliens in the pictures keep at least one of their objects in a jar (while they should keep none). False answers are thus expected in condition b-S3 represented in (8) if the preferred reading of the donkey sentence introduced by no happens to be the existential.
According to the preferred interpretation, as we said, the aliens should keep none of their objects in a jar. This is actually what happens in the pictures in (9), representing the NDT condition, which is actually compatible with both interpretations given that, in case of no, the existential reading entails the universal. For this reason, true at ceiling is expected in condition b-S1 below.
Analyzing the scenarios presented for some, (10) reports the DC condition (S3) in which the situation represented in the pictures is only compatible with the existential reading of the anaphora. As expected, the aliens that keep their objects in a bin keep only some of the objects they own in a bin, leaving at least one outside. If this interpretation happens to be the preferred one, as expected, then a high acceptance rate is predicted when evaluating (10).

(10) Some Glimp that has a dorf keeps it in a bin

Condition c-S3: only compatible with ∃-reading
(Expected answer = TRUE if ∃-reading preferred)

Please note that, given the entailment relations in (3), different types of answers are expected across Condition I in case of the critical condition S3. As we have seen, for sentences introduced by the universal and the negative quantifier (i.e. sentences of type (a) and (b)), the strongest interpretation coincides with the interpretation predicted as preferred. Thus, scenario S3 will necessarily represents the weakest interpretation (the one predicted as dispreferred in these cases) and for this reason a “false” answer is expected in the DC condition, as indicated in (6) and (8). On the contrary, for sentences introduced by the existential quantifier (i.e. sentences of type (c)), the strongest interpretation coincides with the interpretation predicted as dispreferred. Thus, in case of some, S3 will represent the weakest interpretation (which in this case is the one predicted as preferred)
and thus “true” answers are expected for this sentence type in the DC condition, as indicated in (10). Moreover, to test the NDT condition (S1) for sentence of type (c), the universal reading is to be represented in the pictures because in this case this is the strongest interpretation. For this reason, in scenario (11) below representing the NDT condition, the alien that keeps his objects in a jar, doesn’t leave any outside:

(11)

<table>
<thead>
<tr>
<th>Some Plosc that has a fert preserves it in a jar</th>
</tr>
</thead>
<tbody>
<tr>
<td><img src="image1.jpg" alt="Jars with objects inside and outside" /></td>
</tr>
<tr>
<td><img src="image2.jpg" alt="Jars with objects only inside" /></td>
</tr>
</tbody>
</table>

Condition c-S1: representing ∀-reading (entailing ∃-reading \(\Rightarrow\) expected answer = True)

What happens in (11) seems contrary to what would be expected under the preferred existential interpretation of the anaphora according to which the aliens that keep their objects in a jar leave at least one outside. In any case, given that this preferred interpretation is entailed by the other, true answers at ceiling are in principle to be expected. However, the fact that for (c)-sentence the NDT condition actually represents the dispreferred reading may have an impact on the rate of acceptance of the sentence in this condition, as we will discuss later.
1.4 Predictions

As we have seen in (3), each sentence (a)-(c) in Condition I is associated with two possible interpretations depending on which interpretation is assigned to the donkey anaphora they contain. One first question addressed in this experiment is whether one of the candidate interpretations constitutes the default one. To answer this question, data from the DC condition (i.e. S3) for each sentence type will prove crucial. With respect to this condition, two hypotheses can be made, each one deriving contrasting predictions that can be tested experimentally:

(12)
- $H_0$: there is no default, thus both interpretations are generated and are equally accessible
  $\Rightarrow$ subjects should split;

- $H_1$: there is a default interpretation, which can be compatible or not with the scenario presented in the pictures for the DC condition
  $\Rightarrow$ subjects should go for it, accepting or rejecting the sentence accordingly (no split).

No difference instead is expected between $H_1$ and $H_0$ with respect to the other conditions (NDT and NDF) for which, respectively, true and false answers at ceiling are predicted. In particular, the hypothesis $H_1$ that I’m assuming will derive the predictions already mentioned in (6)-(11) above, which are summarized in the chart below for convenience and compared with the predictions derived from $H_0$:
Table 2. Predictions. Predicted results across conditions according to the alternative hypothesis $H_0$ and $H_1$. Results obtained in Condition S3 (cf. last column) will prove crucial in rejecting one of the alternative hypotheses.

<table>
<thead>
<tr>
<th>Cond. I Sentence</th>
<th>Cond.S1-NDT $H_1 = H_0$</th>
<th>Cond.S2-NDF $H_1 = H_0$</th>
<th>Cond.S3-DC $H_1 \neq H_0$</th>
</tr>
</thead>
<tbody>
<tr>
<td>(a)</td>
<td>True</td>
<td>False</td>
<td>False, Split</td>
</tr>
<tr>
<td>(b)</td>
<td>True</td>
<td>False</td>
<td>False, Split</td>
</tr>
<tr>
<td>(c)</td>
<td>True</td>
<td>False</td>
<td>True, Split</td>
</tr>
</tbody>
</table>

Let’s now turn to analyze the results obtained in the experiment just described.

1.5 Results

Data obtained in this experiment are summarized in the table below, that reports the percentages of “true” and “false” answers and subject’s judgments on sentence appropriateness (only assigned in case of positive answer) for each level of Conditions I and II (sentence types and situations): 

<table>
<thead>
<tr>
<th>Cond. I Sentence</th>
<th>Cond.S1 NDT</th>
<th>Cond.S2 NDF</th>
<th>Cond.S3 DC</th>
<th>Preferred reading</th>
</tr>
</thead>
<tbody>
<tr>
<td>(a)</td>
<td>100% True (4.57)</td>
<td>93% False</td>
<td>57% True (3.41)</td>
<td><em>split</em></td>
</tr>
<tr>
<td>(b)</td>
<td>100% True (4.83)</td>
<td>97% False</td>
<td>93% False</td>
<td>Ξ</td>
</tr>
<tr>
<td>(c)</td>
<td>80% True (3.38)</td>
<td>97% False</td>
<td>87% True (3.58)</td>
<td>Ξ</td>
</tr>
</tbody>
</table>

Table 3. Results. Subjects’ distribution in evaluating the 3 sentence types (Condition I) in the 3 situations (Condition II).
These data were submitted to statistical analysis using the program for data analysis SPSS 12. Please note that I will not consider data from the three lists separately because a comparison across lists revealed that they are not significantly different from one another, as expected \( F(2)=0.039, p=.96, \text{n.s.} \). Thus, I first analysed the data obtained for the control-false condition (Condition S2-NDF) and the overall percentage of correct responses was 96%, indicating that subjects understood the task and performed correctly. A repeated measure ANOVA showed no significant difference among sentence types in this condition, as expected \( F(2)=0.244, p=.78, \text{n.s.} \). A similar result was obtained for the control-true condition (Condition S1-NDT), where the percentage of correct responses overall was also attested around 96%. In this case, however, an ANOVA revealed a significant difference among the sentence-types in this condition \( F(2)=7.25, p<.01 \). If we analyse the responses given in this condition for the three sentence types separately, it turns out that, while no mistake was made for sentences of type (a) and (b) (percentage of correct responses being 100% in both cases), a lower correct response rate (i.e. 80%) was found for sentence of type (c) introduced by the existential quantifier some. An analysis of proportions of yes/no answers, by means of a related sample t-test, revealed a significant difference between this type of sentence and the other two in this condition \( 80\% \text{ vs. } 100\%, t(29)=2.69, p<.01 \). This difference, however, can be explained by the fact that the interpretation of the pronoun which is represented in the scenario is the universal interpretation, which is also the one predicted as the disfavoured one. For a better understanding of this point, consider again the scenario reported in (11) above. As already observed, only in case of some the “dispreferred” \( \text{universal} \) reading is the strongest one, the one that entails the favoured \( \text{existential} \) interpretation and for this reason was the one to be represented in the NDT condition. However, despite the fact that the sentence was “true” under the \( \text{existential} \) interpretation (because this was actually entailed by the \( \text{universal} \) interpretation represented in the scenario), it’s conceivable that some of the subjects noted this “lack” of the preferred interpretation, thus judging the sentence “false”. This explanation is also supported by the observation that the mean score on the scale attributed to the
subjects who accepted the sentence as “true” (i.e. 3.38) is significantly lower if compared to the mean score given by subjects in the same NDT condition to sentences of type (a) and (b), which was 4.57 and 4.83 respectively (t(52)=4.02 and t(52)=5.34, p<.0001).

I will now discuss the analysis on the results obtained in the critical DC for each quantifier. In case of the negative and the existential quantifier (thus Condition I-b and I-c), subjects show a preference for the existential interpretation: 93% rejected the critical sentence with no in the scenario representing the \(\forall\)-interpretation of the anaphora (cf. (8) above), while 87% of subjects accepted the critical sentence containing some in the scenario compatible with the \(\exists\)-interpretation of the anaphora (cf. (10) above). Subjects’ distributions were compared with chance performance and in both cases the proportions revealed significantly different from chance (p<.0001). On the contrary, in case of the sentences introduced by the universal quantifier every, subjects split: 43% of them actually rejected the critical statement in the scenario only compatible with the \(\exists\)-interpretation of the anaphora (cf. (6) above), while 57% accepted it, even if they didn’t express a truly positive judgment on its adequacy (mean rate on the scale for those who accepted the sentence was 3.41). The proportion of acceptance of (a) sentence in Condition S3 reveals not different from chance (p<.05).

However, if we analyse the scores on the scale in case of affirmative answers, an interesting comparison emerges between S1 and S3 for sentences of type (a): subjects that accepted the critical (a)-statements in the NDT condition (S1) gave a high score on the scale when judging sentence’s adequacy to the scenario (mean scale rate in Condition a-S1=4.57). Conversely, when judging the adequacy of the sentence with respect to the scenario in the DC condition, subjects that accepted the sentence (roughly half of them, as we said) judged it much less adequate (mean scale rate in Condition a-S3=3.41). Statistical comparison revealed that the two means are significantly different from one another (t(45)=4.28, p<.001).

We have seen that in the analogous experiment on scalar implicatures presented in the previous chapter, analysis of the reaction-time measures revealed crucial in the comparison between alternative hypotheses on the processes
involved in implicature computation. In the same spirit, we can examine some of the reaction times recorded during this experiment to better understand the behaviour of subjects in evaluating the sentences introduced by the universal quantifier in the critical situation, with respect to which subjects split. In this perspective, these are, according to me, the relevant RTs to be taken into consideration:

<table>
<thead>
<tr>
<th>Cond. I Sentence</th>
<th>DC (S3) true</th>
<th>NDT (S1) true</th>
<th>DC (S3) false</th>
<th>NDF (S2) false</th>
</tr>
</thead>
<tbody>
<tr>
<td>(a)</td>
<td>12003</td>
<td>7486</td>
<td>11042</td>
<td>8605</td>
</tr>
<tr>
<td>(c)</td>
<td>7942</td>
<td>8760</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Table 4. Comparison on RTs. Crucial RTs in analysing subjects’ distribution in the evaluation of sentences introduced by every. RTs are presented separately according to the type of answers (true or false).

Considering the data in the table above, two comparisons would reveal crucial: in first place, a crossed comparison between the RTs of subjects that in S3 (DC) accepted the critical (a)-sentence vs. those that accepted the critical (c)-sentence in the same condition and the same comparison in S1 (NDT) (as control); secondly, a crossed comparison between the RTs of subjects that accepted the critical (a)-sentence in condition S1 (NDT) vs. those that accepted the same type of sentence in condition S3 (DC) and between the RTs of subjects that rejected the critical (a)-sentence in condition S2 (NDF) vs. those that rejected the same type of sentence in condition S3 (DC). This is what emerged from statistical analysis with respect to the comparisons just outlined (for convenience, I marked with an * those comparisons that turned out statistically significant):

1 Given that differences in RT may depend on the response-type (true vs. false responses), I will only consider direct comparisons of RTs across subjects who gave the same (true or false) response across different situations.
1.6 Discussion

Data on the critical sentences introduced by negative and existential quantifiers conform to the predictions reported in Table 2 above: in both cases, the \( \exists \)-reading seems to be preferred over the alternative interpretation. The preference emerged is actually the one predicted by Kanazawa’s generalization, who proposes a model based on the monotonicity properties of the quantifiers that embed the anaphora. Statistical analysis conducted on the results obtained on these two quantifiers support the existence of a default interpretation (against chance performance), so that the conflict between hypotheses \( H_1 \) and \( H_0 \) sketched in (2) seems to be solved against \( H_0 \), that is to be rejected, at least in case of donkey sentences introduced by some and no.

On the contrary, the result obtained for the sentences introduced by the universal quantifier every do not conform to predictions. In the critical situation, subjects split: roughly half of them reject the critical (a)-sentence, as predicted, but the other half accepts it. Subjects’ distribution in evaluating the critical statement can have two opposite explanations: either \( H_0 \) is correct in case of donkey anaphora introduced by the universal quantifier, there being no default interpretation in this case; otherwise, we can hypothesise that there actually is a preferred interpretation (as stated by \( H_1 \)) but the other interpretation is for some reason more easily accessed in these sentences than in sentences introduced by other quantifiers. As we will see, this possibility will be explicitly tested in Experiment 3B.
In the aim of defending $H_1$ and claim that we can talk about default interpretations for any kind of donkey sentence despite the unattended results obtained in case of *every*, let’s take a closer look at the data we got for the sentences introduced by this quantifier. First of all, if we compare the mean scale rate associated to the “true” responses in case of the DC (S3) condition and the NDT (S1) condition we find a significant difference between the two: subjects that accept the (a)-sentence in a scenario compatible with the $\exists$-reading of the anaphora (S3) are much less sure about their judgments than when accepting the same type of sentence in a scenario representing the $\forall$-reading (S1). This contrast may actually signal the presence of a conflict between two competing interpretations: whenever the contrast is solved in favour of the $\exists$-reading in condition S3, this is done at a cost that is reflected by the lower scores on the scale.

The same conclusion can be drawn by analysing RT measures in the relevant comparisons (cf. (13)). These results are particularly interesting, because they reveal that accepting the donkey sentence introduced by the universal quantifier in the DC condition, which is only compatible with $\exists$-reading, is particularly costly if compared with the cost (in terms of RT measures) of accepting the donkey sentence introduced by the existential quantifier in an analogous condition representing the $\exists$-reading of the donkey pronoun. Moreover, the difference between these two sentence types only emerges in the critical S3 situation and not, for example in the NDT condition (S1), where the cost of processing sentences of type (a) and (c) is just the same. This last result indicates that the cost (always in terms of RT measures) of processing the two types of sentences is comparable, and when a difference is found, as in the S3 condition, then this is not due to the mere presence of a different quantifier per se at the beginning of the sentence.

Moreover, if we take into account only the RTs for the sentences introduced by the universal quantifier in the two conditions (S1 vs. S3), we find a significant difference (in terms of RTs) between subjects that accept the (a)-sentence in S1 and those that accepted it in S3. Subjects were significantly much
slower in accepting the sentence introduced by \textit{every} in a scenario compatible with the $\exists$-reading than in accepting it in a scenario representing the $\forall$-reading of the donkey pronoun. This result gets its full significance if completed with the comparison between subjects that instead reject the (a)-sentence in a scenario compatible with the $\exists$-reading (S3) and those that reject it in a condition that renders the sentence unequivocally false (S2): the difference between the RTs obtained in these two cases is not statistically significant, reflecting the fact that rejecting a donkey sentence introduced by the universal quantifier in a scenario only compatible with the $\exists$-reading is just as costly as rejecting it in a condition that renders the sentence false. Considered together, these results seem to indicate that, despite the fact that half of the subjects accepted the donkey sentence introduced by \textit{every} in a situation compatible with the $\exists$-reading, they do it at a cost: in first place, they are not so convinced about the choice made, as the mean rate on the scale signals, and, secondly, they have to work harder to access this reading and accept the sentence in the end, as emerges by the analysis of the RT measures.

These results constitute undoubtedly interesting indications, but they cannot suffice in discarding $H_0$ in case of donkey sentences introduced by \textit{every}. If one wants to ultimately defend the default hypotheses for any kind of donkey sentences (discarding $H_0$ once and for all), then her claim must be justified by independent reasons. In order to shed some light on this question, a simple follow-up to this experiment was conducted, in the attempt to make the results obtained here in a way clearer.

2. \textbf{Experiment 4 - Checking for preference}$^2$

In order to check if a preference actually emerged between the two alternative interpretations of donkey anaphora embedded in the restrictor of the universal quantifier, a very simple experiment, in form of a questionnaire, was

---

$^2$ This idea was suggested by Lyn Frazier, who I warmly thank.
submitted to a new group of subjects. Sets of sentences of different types were inserted in the questionnaire, functioning as filler for the critical items that corresponded to the same sentences used in the reaction-time study just presented. These sentences were re-tested in a new form to see if a preference emerged between the two readings of the pronoun. The form and modality of this study will be described in detail in the following sections.

2.1 Subjects

A total of 20 subjects participated in the experiment. Subjects were mainly 1st year students at the Psychological Faculty of the University of Milano-Bicocca, and received credits for their participation.

2.2 Procedure

Subjects were tested in a classroom in the university building. They received a block of ten small sheets pinned together. They were given oral and written instructions on how to fulfil the test. The first page contained an example about which they could ask clarification questions. They were reminded neither to consult their classmates nor to go back and correct already solved items. Time taken to complete the whole task was about 16 minutes. The experiment presented a series of sentences followed by two possible paraphrases. Subjects were instructed to read the target sentence first and then choose the paraphrase that, in their opinion, better expressed the meaning of the test sentence.

2.3 Material and Predictions

Testing material comprised 3 critical donkey sentences of the same type as the ones used in the reaction time study discussed in the preceding section. They were followed by two paraphrases, each representing one of the two candidate interpretations of the donkey pronoun, as in the example below (cf. Appendix E):
Every Flont that has a vilp keeps it in a bin

(a) Every Flont that has some vilps keeps all the vilps he has in a bin

(b) Every Flont that has some vilps keeps one of the vilps he has in a bin

Subjects were asked to indicate which paraphrases between (a) and (b) better expressed the meaning of the sentence in italics. Please note that the underlining is put here for convenience but original testing material did not contain any underlined part. Moreover, the order of presentation of the two alternative paraphrases was randomly varied across subjects.

As we have discussed above, the interpretation that (depending on the left monotonicity properties of the universal quantifier) is predicted as favoured in case of donkey pronouns in the scope of universal quantifier is the universal interpretation, which correspond to option (a) in (14). If hypothesis $H_1$ on the existence of a default interpretation of the pronoun is tenable, then subjects should consistently choose one of the two alternative paraphrases. Moreover, if they comply with Kanazawa’s generalization, basing their judgments on the left monotonicity properties of the quantifier, they should consistently choose the paraphrases in (a). Otherwise, subjects should split, as predicted by $H_0$.

2.4 Results and Discussion

The result obtained in this questionnaire study is pretty sharp: in 82% of the cases subjects chose the paraphrases corresponding to the $\forall$-interpretation of the pronoun. Precisely, 17 out of 20 subjects (corresponding to 78% of subjects) chose this paraphrase at least two out of three times. All these percentages are significantly different from chance ($p<.05$).

The conclusion that can be drawn from this follow-up experiment is clear-cut: it seems that we can talk about a preferred/default interpretation of the anaphora in absence of a context even for sentences introduced by the universal
quantifier. If we accept this conclusion, then one should explain why in the reaction-time study this preference didn’t emerge. One possible explanation can build upon the consideration that the two alternative readings may have different “degrees” of accessibility, depending on the type of head determiner that introduces the clauses and the (extra-linguistic) context in which they are used. In a situation in which the sentences are to be evaluated out of the blue in the absence of a context or situation, like in the questionnaire study, it is conceivable to expect subject to select the default interpretation. In tasks in which a situation is presented, as it was in the reaction-time study, the extra-linguistic context could influence the interpretation of the donkey pronoun rendering the alternative reading more salient in the context/situation provided. In the case at hand, the alternative reading rendered the sentence true in the scenario presented (cf. 6), and this could be the reason why half of the subjects override the default in that situation. Such a possible line of explanation has been specifically investigated in the second part of the reaction-time study presented, i.e. Experiment 3B, in which donkey sentences were proposed in a scenario introduced by a biasing context. I will now turn to the description of this experiment, which will add interesting topics to the ongoing discussion.

3.  Experiment 3B - Degrees of accessibility of the alternative readings

If we want to talk about default interpretations, then we need to attest experimentally the access to the alternative readings of the donkey pronoun in situations that specifically induce such readings. Remember that in the preceding chapter, we discussed several examples presented in the literature that testify the effective ambiguity of the donkey pronoun. As it was made clear, the accessibility of these alternative interpretations depends much by the specific example used and by the quantifier that precedes the anaphora. Intuitive judgments on this point actually show that the alternative $\exists$-reading in case of sentences introduced by the universal quantifier are more easily accessed than in case of sentences introduced by the negative and the existential quantifiers. For these last quantifiers, as a
matter of fact, very marked examples should be construed to access the ∀-reading over the ∃-reading, while for the universal quantifier a simple biasing context could suffice in accessing the ∃-reading over the ∀-reading (think, for example, at the context proposed by P. Casalegno in Chapter 3, section 1).

Aim of the experiment I’m going to present is twofold: in the first place, it aims at supporting the hypotheses that alternative readings are actually available, a necessary step to uphold the default-hypotheses advocated in Experiment 3A; secondly, it aims at testing the hypotheses that the degree of accessibility of these alternative readings may vary according to the type of quantifier that introduces the donkey sentence. Before continue, please refer to sections 1.1 and 1.2 for the description of the subjects that participated in this experiment as well as for a description of the procedure here employed. This experiment, in fact, comes as a completion to Experiment 3A and is to be analyzed in the comparison with the results obtained there.

3.1 Material

Exactly as in Experiment 3A, three types of donkey sentences were presented, created as a within subject factor, in which I varied the type of quantifier that precedes the anaphora (cf. Appendix D). This could either be the universal quantifier every (it. “ogni”), the negative quantifier no (it. “nessuno”) or the existential quantifier some (it. “qualche”). Actually, the sentences that were used in this experiment were the same used in Experiment 3A and are repeated below for convenience:

(15) Condition I: sentence-type
   (a)  \textit{Every} \ P [that has an \ A] \ VP
   (b)  \textit{No} \ P [that has an \ A] \ VP
   (c)  \textit{Some} \ P [that has an \ A] \ VP

Differently from Experiment 3A, however, these sentences were only presented in two situations, paralleling two of the levels of Condition II already
described for Experiment 3A. The control true condition (NDT) was in fact absent in this case.

(16) Condition II: Situation

S1 (DC):

a situation which is compatible only with one of the two possible resolutions of the anaphora, thus rendering the sentence true on one interpretation and false on the other. Notice that, since the interpretations are not logically independent of one another, there can be only one kind of differentiating situation per sentence (I will refer to this condition as the differentiating-critical situation = DC);

S2 (NDF):

a situation which is incompatible with both candidate interpretations of the anaphora, thus rendering the sentence false on both readings (I will refer to this control condition as the non-differentiating false situation = NDF).

The crucial difference with the preceding experiment is represented by a biasing context that was added before the presentation of each critical sentence. Most importantly, the sentences and the scenario used in Experiment 3B were exactly the same as the ones used in the preceding experiment in the same conditions. Thus, in case of the universal and the negative quantifier, the scenario represented in the four pictures was only compatible with the reading predicted as dispreferred (i.e. the $\forall$-reading), while in case of the existential quantifier the scenario was compatible with the reading predicted as preferred (i.e. the $\exists$-reading). As already observed, in case of some, the choice of the scenario couldn’t parallel the one made for the other two quantifiers because in this case the universal interpretation, which is the one predicted as dispreferred, is also the strongest one, thus entailing the existential. This is the reason why in case of sentence of type (c) I used a scenario representing the (weakest) $\exists$-reading
preceded by a context inducing the (strongest) $\forall$-reading in the attempt to bias such an interpretation and thus making subjects reject the sentence in the scenario presented. Of course, in analysing the results, one should keep in mind this added difficulty due to the peculiar entailing properties between the readings associated to the anaphora in the scope of this quantifier.

To better give an idea of the manipulation introduced in the design, let’s consider the DC condition (S1) per each sentence type, preceded by its biasing context. In (17) I present an example of the DC condition for sentence of type (a) (Condition a-S1). As one can see, the interpretation that is biased by the context is the existential interpretation of the donkey pronoun, which is also the only one compatible with the scenario. Precisely, none of the aliens in the pictures keep all of his objects in a bin. This state of affairs is also the one induced by the biasing context, according to which each alien is expected to put only one (and not all) of his objects in a bin.

(17)

On planet Flont there’s the plague of the termites. To face it, Flonts use special traps, the vilps. Fighting termites causes a lot of stress to the Flonts. To improve the situation, they consulted a famous magician who suggests them a remedy: they have to sacrifice a vilp putting it into a bin full of water; a vilp in the water doesn’t function as a trap anymore, but releases an anti-stress substance. Let’s see if Flonts are less stressed given that…

<table>
<thead>
<tr>
<th>Every Flont that has a vilp keeps it in a bin</th>
</tr>
</thead>
<tbody>
<tr>
<td><img src="image1.png" alt="Image" /> <img src="image2.png" alt="Image" /></td>
</tr>
<tr>
<td><img src="image3.png" alt="Image" /> <img src="image4.png" alt="Image" /></td>
</tr>
</tbody>
</table>

**Condition a-S1:** situation & context: $\exists$-reading
In (18) and (19) examples for no and some are reported instead: in case of these quantifiers, the reading that emerged as the preferred one in the preceding experiment was the existential. For this reason the alternative universal reading is the one biased by the paragraph: in case of the negative quantifier, this interpretation is also the one compatible with the scenario presented in the pictures. As one can see in (18), none of the aliens put all of their objects in a bin. In fact, all of them leave one object outside, as suggested in the paragraph.

(18)

<table>
<thead>
<tr>
<th>No Plosc that has a gurf keeps it in a bin</th>
</tr>
</thead>
<tbody>
<tr>
<td><img src="image1.png" alt="Image" /></td>
</tr>
<tr>
<td><img src="image2.png" alt="Image" /></td>
</tr>
<tr>
<td><img src="image3.png" alt="Image" /></td>
</tr>
<tr>
<td><img src="image4.png" alt="Image" /></td>
</tr>
</tbody>
</table>

**Condition b-S1**: situation & context: \( \forall \)-reading

In case of the existential quantifier reported below, the situation represented in the pictures is not compatible with the interpretation biased by the context, given that, as we said, this is the strongest one thus would render the sentence true under any interpretation, shadowing the actual choice made by subjects. As one can see in (19), the paragraph induces the universal interpretation of the pronoun, requiring that at least some aliens keep all their objects in a bin.
Conversely, the aliens in the pictures that keep their objects in a bin keep only some of them, always leaving at least one outside.

(19)

The radio news at 11.00 urged the population to put all their gurfs into bins made of lead because if they’re kept outside they are radioactive. Let’s see if, after the bulletin…

<table>
<thead>
<tr>
<th>Some Glimp that has a dorf keeps it in a bin</th>
</tr>
</thead>
<tbody>
<tr>
<td><img src="image1.png" alt="Image" /> <img src="image2.png" alt="Image" /></td>
</tr>
<tr>
<td><img src="image3.png" alt="Image" /> <img src="image4.png" alt="Image" /></td>
</tr>
</tbody>
</table>

**Condition c-S1**: situation ∃-reading; context: ∀-reading

As for the control NDF condition (Condition II-S2), a biasing context was added in the attempt to induce subjects to accept the sentence despite the fact that the situation represented in the pictures rendered it unequivocally “false”. For example, this is one of the controls used in the experiment:
(20)

In this planet, many Plosc own samps, objects extremely sensitive to the sunlight. For this reason, they always have to be protected under a leaf when the sun shines. Thus, in sunny days…

<table>
<thead>
<tr>
<th>Every Plosc that has a samp protects it with a leaf</th>
</tr>
</thead>
<tbody>
<tr>
<td>![Images of Plosc protecting samps with leaves]</td>
</tr>
</tbody>
</table>

**Condition a-S2 (control):** situation: false; context: true

As one can note, the sentence, under its favoured interpretation, is totally in line with the context that precedes it which exhorts every alien to protect all her objects under a leaf. If one were to answer on the basis of the sentence only, without evaluating it with respect to a scenario, she would presumably evaluate the sentence “true”. On the contrary, if one, as required, were to evaluate the sentence with respect to the situation represented by the pictures, then she would be forced to reject it.

### 3.2 Predictions

In the control condition (S2), I expected subjects to answer “false”, despite the presence of the biasing context designed to induce a “true” answer. As a matter of fact, subjects were ultimately instructed to evaluate the sentences with respect to the scenario represented in the pictures, keeping in mind what they have read/heard in the preceding context and trying to find, where possible, an interpretation of the sentence compatible with both the scenario and the context. For this reason, if they understood the task correctly, they had to reject the
sentence because it was incompatible with the scenario presented under any possible interpretation of the sentence.

As for the critical conditions (S1), these have to be evaluated in comparison with their counterparts in Experiment 3A. Without such a comparison, results from this experiment are indeed not interpretable. Remember that the critical sentences and the scenarios presented were the same in the two experiments, but they were preceded by a biasing context in Experiment 3B in the attempt of inducing one interpretation of the ambiguous sentence over the other. In particular, the context always biased subjects towards the interpretation that emerged (or was predicted) as dispreferred in Experiment 3A. Given that the adding of the preceding context was the only manipulation made in the design, any difference in the evaluation of sentences in the critical conditions between the two groups of subjects that participated in the two experiments is to be intended as the effect of this manipulation, namely, the presentation of a biasing paragraph before each critical sentence. Given these premises, three are the hypotheses that one could investigate in the comparison between Experiment 3A and 3B, none of which is free from problems in interpreting the data obtained in the previous experiment:

(21)

- $H_0$: only one interpretation is made available by the computational system, thus extra-linguistic context cannot have any influence whatsoever on the way we interpret sentences. Specifically, being there no alternative, one cannot talk about default interpretations

problematic data: split on evaluating every in Exp. 3A; examples on alternative readings in the literature
predictions: no difference in the DC conditions between Exp. 3A and 3B for any of the quantifiers

---

3 This same outcome could be also obtained if the contexts used were not apt to the purpose. In this eventuality, the fact that no effect is detected would mostly be due to a flaw in the
- H₁: alternative interpretations are equally available in the computational system and none constitutes the default

  problematic data: strong preferences emerged for no & some in Exp. 3A; strong preferences emerged for every in Exp.4

  predictions: significant difference in the DC conditions between Exp. 3A and 3B for all quantifiers

- H₂: alternative interpretations are made available by the computational system; one constitutes the default; extra-linguistic context can influence the choice between the two, overriding the default

  problematic data: split on evaluating every in Exp. 3A (partly rescued by Exp. 4)

  predictions: accessibility of these alternatives and effect of context over it may depend on the type of quantifier that embeds the anaphora. Thus, a difference in the DC conditions between Exp. 3A and 3B should be found AND a difference amongst quantifiers is possibly found

It will be clear by now that my intent here is that to defend the last of the hypotheses presented above. According to this hypothesis, these are, in sum, the predictions derived:

experimental design. Unfortunately, we do not have independent control for this, given that the only controls were negative conditions and context was not expected to influence a negative judgment turning it into a positive one. The only corroboration of the design adopted comes from the results obtained - internally - for the DC conditions for sentences of type (a), for which the effect of context is remarkable.
Table 5. Predictions. Predictions derived by H₂, and their comparison with alternative hypotheses.

<table>
<thead>
<tr>
<th>Cond. I Sentence</th>
<th>Cond. II Scenario</th>
<th>Biasing context</th>
<th>subjects’ answers</th>
<th>how much should be &gt;?</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
<td>Exp. 3A</td>
<td>Exp. 3B</td>
</tr>
<tr>
<td>(a)</td>
<td>∃</td>
<td>∃</td>
<td>57% T</td>
<td>&gt; 57% T</td>
</tr>
<tr>
<td>(b)</td>
<td>∀</td>
<td>∀</td>
<td>7% T</td>
<td>&gt; 7% T</td>
</tr>
<tr>
<td>(c)</td>
<td>∃</td>
<td>∀</td>
<td>13% F</td>
<td>&gt; 13% F</td>
</tr>
</tbody>
</table>

extent of this difference may vary across different quantifiers

results compatible with H₁ & against H₀

results against H₁ & H₀

In order to decide amongst competing hypotheses, we should analyze the results first, to see how they conform to the predictions outlined above.

3.3 Results

I will start my discussion on the analysis of the control conditions. As expected, context had no effect on the percentages of “false” answers provided by subjects overall, which were attested about 88% in Experiment 3B and above 90% in Experiment 3A (t(196)=−1.91, p=.06, n.s.).

Results on the critical DC conditions for the three quantifiers in the two experiments are summarised in the chart below, and will be analysed in detail, considering the case of each quantifier separately.

Table 6. Results. Rate of acceptance of critical sentences in the DC condition (S1): comparison between subjects’ acceptance rate in Experiment 3A and 3B (* indicates that the difference is statistically significant).
I will start from considering sentences of type (a) first, the ones introduced by the universal quantifier *every*. In the critical condition, context induced the $\exists$-reading of the anaphora, which is the one predicted as dispreferred and is also the only one compatible with the scenario represented in the pictures (cf. example (17) above). Percentage of acceptance of this type of sentence in the same situation in Experiment 3A was 57%; in this experiment, in which the biasing context was added, the percentage of acceptance rose at 87%, and the statistical comparison between the proportions of acceptance rate in the two experiments revealed a significant difference ($t(58)=-2.69, p<.01$).

A different result was obtained for the other two quantifiers. In particular, sentence of type (b) introduced by the negative quantifier *no* were accepted only by 7% of the subjects in situation S3 in Experiment 3A, and this percentages slightly increased to 17% in Experiment 3B after the adding of the biasing context which should favour the access to the alternative $\forall$-reading of the donkey sentence. Though this result goes in the right direction, the difference between subjects’ answers in the two experiments is not statistical significant ($t(64)=-1.23, p=.22, \text{n.s.}$). Interestingly, there is another hint that suggests that subjects treated the two sentences differently in the two experiments: a pairwise comparison of overall mean RTs on sentences of type (b) revealed a significant difference between the two conditions. Namely, though not shifting their final interpretation, subjects took longer to make a decision in case of Experiment 3B than Experiment 3A ($t(58)=-2.00, p<.05$). This difference cannot be considered a direct effect of the manipulation made in Experiment B (namely, the adding of the context), given that none of the analogous comparisons made on overall RTs for the other sentence types (after the same manipulation was made) turned out statistically significant.

A different perspective should be adopted in considering the results for sentences of type (c), introduced by the existential quantifier *some*. Recall that the design for this condition differed from the other two quantifiers, given that the entailment relations between the two interpretations follow a different pattern. These sentences, in fact, were to be evaluated in a scenario compatible with the
∃-interpretation in condition S3 in Experiment 3A, which is the reading predicted as the favoured one. Percentage of acceptance was expected to be high in that experiment, and in fact it was about 87%. In Experiment 3B, the same scenario was presented, but the sentences were introduced by a context in which the other, dispreferred, interpretation was biased. The effect of such a context was predicted to induce more “false” answers than Experiment 3A, given that the interpretation biased by the context was incompatible with the scenario presented. However, no effect of this manipulation was detected, given that subjects stoke to their preferred interpretation and accepted the critical sentence virtually at the same rate of Experiment 3A, thus not shifting their interpretation under the effect of the preceding context (t(64)=.37, p=.71, n.s.).

3.4 Discussion

First of all, a remark is due on the results obtained for the sentences introduced by the existential quantifier that we have just presented. I think that this result should be handled with care with respect to the predictions outlined above. In particular, it is conceivable that this result depends more on the configuration of the entailment patterns between the alternative readings associated with the pronoun in the scope of some than to the actual effect of the biasing context on its interpretation. Remember that in this case the paragraph induced the universal interpretation, which is the one predicted as dispreferred. My guess is that, however effective the context may have been on the immediate interpretation of the sentence, this interpretation is put aside once the subject sees that the situation represented in the pictures is compatible with the most preferred alternative reading. Mostly for this reason, in my opinion, subjects keep accepting the sentence in the scenario provided, compatible with the ∃-reading, thus overriding the effect of the context towards the ∀-reading.

More generally, taking stock from the results obtained in Experiment 3B, these are the interesting points worth of discussion, especially in the comparison among the alternative hypotheses presented in (21). In the first place, a significant difference is found for sentences introduced by the universal quantifier (and a
trend in the right direction is also attested for sentences introduced by the negative quantifier). This fact, together with the results already obtained in Experiment 3A for the sentences introduced by the universal quantifier, pushes in the direction of a rejection of $H_0$, according to which only one interpretation is made available by the computational system. Unless one decides to argue for a non homogeneous treatment of donkey anaphora with respect to this issue.

With respect to the comparison between the other two alternative hypotheses, the same predictions were made for what concerns the effect of the context on subject’s responses: in particular, both hypotheses predicted an increment of subject’s acceptance rate in Experiment 3B with respect to Experiment 3A. In this respect, data obtained are in line with both hypotheses. The result that instead reveals crucial in choosing one hypotheses over the other is the different effect of the biasing context observed across sentence types. As we have seen, the range of this difference varies a lot depending on the quantifier that embeds the anaphora: in case of the universal quantifier, context has a great effect on the interpretation of the donkey pronoun, while this effect is much less visible for the sentences introduced by the negative quantifier, being virtually null in case of sentences introduced by the existential quantifier (but, as we said, this last result may have a different explanation). Considering these results at the light of the predictions outlined in Table 5, I believe that the hypotheses that better conforms to the experimental data is $H_2$, in the form repeated here: both interpretations are made available by the computational system; one constitutes the default (following the generalization proposed by Kanazawa, which i ultimately based on the left monotonicity properties of the determiner in which the anaphora is embedded in order to preserve inferential patterns); the degree of accessibility of the alternative (dispreferred) reading may vary according to the quantifier that precedes the anaphora; context can influence the interpretation of the donkey pronoun overriding the default, but this effect depends on the accessibility of the alternative reading (which varies according to the type of quantifier that introduces the donkey sentence). Under this perspective, this is how we can summarize the results obtained:
(22)
- a difference in the DC conditions between Exp. 3A and 3B was found
  \(\Rightarrow\) two interpretations are made available by the computational system (pro \(H_1\) and \(H_2\), against \(H_0\));
- AND a difference amongst quantifiers was found
  \(\Rightarrow\) the availability of these alternatives and the effect of context over it may depend on the type of quantifier that embeds the anaphora (pro \(H_2\), against \(H_1\)).

4. Conclusive remarks on donkey anaphora resolution

Considering the data just presented, the interpretation that emerges as the default one in the experiments presented in this chapter is the one predicted by Kanazawa’s generalization, which ultimately resides on the logic notion of monotonicity and is justified by appealing to the notion of preservation of inferential patterns, which is again logically determined. As already discussed in the previous chapter, logicality seem to provide the explanatory tools to understand some of the processes that affect sentence comprehension and determine the reason why we interpret sentences the way we do. The case of donkey anaphora is paradigmatic in this respect, given that it provides us with an example of ambiguity resolution governed by default rules based on logical properties of the clause components. As we have seen, donkey sentences can be considered genuinely ambiguous (at the interpretative level) between two readings, namely the *universal* and the *existential*. These interpretations are not logically independent of one another, given that one reading always entails the other, as in the schema provided in (3) above. Differently from the solution proposed in case of scalar implicatures, in case of donkey anaphora we cannot explain the distribution of the preferred reading by appealing to the logic notion of informativeness (or strength of the assertion). In the first place because, for what concerns donkey sentences, the predicted preferred reading does not always
coincides with the most informative interpretation. Paradigmatic in this respect is the case of donkey sentences introduced by the existential quantifier, in which the reading predicted as the default one is the existential, despite the fact that this is the less informative interpretation, entailed by the universal one. In second place, there is another crucial test to verify that informativeness cannot provide the right cues in explaining the distribution of defaults for the interpretation of donkey pronouns. As pointed out to me by Chierchia (p.c.), crucially the notion of informativeness is linked to the monotonicity properties of the context. If the interpretation assigned to the donkey pronoun were justified in terms of strength of assertion, then such an interpretation would depend on the monotonicity properties of the context in which the pronoun appeared. For example, a donkey sentence introduced by the universal quantifier, when embedded under negation (thus in a DE context) would be assigned the existential reading, given that, in a negative context, this interpretation would result the strongest, i.e. the most informative one. But this is not so, quite the contrary, as exemplified by the donkey negative sentence below:

(23) Every farmer who owns a donkey doesn’t beat the donkey he owns

This sentence is preferably assigned the universal reading according to which the farmers doesn’t beat any of the donkeys they own, and not the more informative existential interpretation according to which there is at least one donkey that each farmer doesn’t beat (but he could beat the others). This simple test excludes informativeness as the crucial property in determining the default in donkey anaphora resolution. This is in fact determined appealing to other logical notions, such as the monotonicity properties of the quantifier that embeds the anaphora and the preservation of inferential patterns. Following Kanazawa (cf. Chapter 3, section 3.1), the interpretation predicted to be the default one (and was confirmed as such in the study presented here) strictly depends on the monotonicity properties of the 1st argument of the quantifier that embeds the anaphora (i.e. the fact of being Downward or Upward Entailing, see Chapter 1 for
details on this distinction). After having determined the distribution of default interpretations by appealing to strictly logical notions, the second necessary step that has been made was that of testing the accessibility of the alternative interpretation and the factors that can eventually force its emergence. The claim that has been made taking the experimental data into consideration is that the alternative interpretations are accessible in various degrees depending of the type of quantifier that introduces the sentence. For example, it’s easily accessed in case of the universal quantifier, and less in case of other quantifiers. I believe that this solution is quite plausible and much in line with the facts emerged from experimental investigation.\(^4\)

\(^4\) It is interesting to observe that this claim is also compatible with other experimental results reported in the literature. In the first place, an experiment conducted on 3- to 5-year-old children showed that they overwhelmingly (86%) accepted donkey sentences introduced by the universal quantifier in a situation corresponding to its weak interpretation. (cf. Conway and Crain, 1995 and Crain and Thornton, 1998). Analogously, an experimental study conducted on adults in Dutch showed that, while subjects do not access the universal interpretation of donkey sentences introduced by weak determiners, such as some and no, they seem to access both interpretations in case of strong determiners such as every (cf. Geurts, 2002). I refer the reader to the interesting discussions and questions raised by these studies, which I believe sets interesting basis for further investigation.
In Chapter 1 I have introduced the phenomenon of scalar implicatures (SIs) and discussed different theoretical approaches to it. In Chapter 2 I reported some experiments that I conducted on adults to investigate how they preferably interpret the scalar item or and which factors can influence this default. What emerged from the two studies there reported is that adults display a spontaneous logicality in dealing with sentences that contain a scalar item, always selecting the most informative interpretation (when the context supports the activation of the scalar alternatives). Crucially, the fact of being the “most informative” interpretation is not an absolute property of one interpretation over the other but it depends on the monotonicity properties of the context that embeds the scalar term. In this respect, the inclusive interpretation of or is the most informative one in DE contexts and the exclusive interpretation is the most informative one in NON-DE contexts. Few studies have specifically investigated the connection between downward entailing properties and scalar implicature computation in children (cf. Boster and Crain, 1993, Chierchia et al. 2001/2004 and Gualmini, 2003). In particular, the study that Chierchia and colleagues was conducted in the same
period and sharing the same theoretical background assumptions as the first experiment presented in Chapter 2 of this dissertation. Taking disjunction as a case study, they investigated children’s sensitivity to the DE/NON-DE distinction by means of the same task already described for my Experiment 1, i.e. the Truth Value Judgment Task (Crain and Thornton, 1998). In particular, they conducted two experiments, both on children and adults (as control group), in which two linguistic contexts were investigated, namely the two arguments of the universal quantifier *every*. As we already saw (cf. Chapter 2, section 1) these display different entailing properties: the restrictor (or 1st argument) of the universal quantifier is DE and its nuclear scope (or 2nd argument) is NON-DE. Each experiment comprised four target trials (interspersed with fillers) of the same kind. Experiment 1 presented items like (1) in which *or* appears in a DE context, while Experiment 2 presented items like (2) in which *or* appears in a NON-DE context:

(1) *Every dwarf who chose a banana or a strawberry received a jewel*

(2) *Every boy chose a skate-board or a bike*

Both sentences were uttered by a puppet as a description of a story in which the relevant characters did both the things mentioned, thus in a scenario which was always consistent with the *inclusive* interpretation of *or*. Percentages of acceptance of sentences of type (1) were significantly different from chance for both groups of subjects: 91.6% and 95.5% of the critical statements were accepted by children and adults respectively. This result is consistent with the expectation that SIs are not added in DE contexts. On the contrary, data from the second experiment showed a crucial difference between children and adults: 100% of adults rejected sentences of type (2) but only 50% of the children did. A comparison between the two experimental conditions revealed that children are sensitive to the DE/NON-DE distinction given that they pattern differently in the two environments with respect to the interpretation assigned to *or*: they overwhelmingly (91.6%) accept sentence (1) (which represents a DE context), but
only half of them (50%) reject sentence (2) (which represents a NON-DE context).

On the one hand, this experiment satisfies our curiosity about children’s sensitivity to monotonicity properties. On the other, however, it raises general questions about scalar implicature computation in children that have not yet been answered: are children able to derive SIs in the end? If they are, why only half of them derive the SI in sentences (2)?

1. Children’s performance on SIs: a survey on previous experimental findings

In recent years, different experimental works on the acquisition of SIs have been carried out with children of different ages (from 5 to 11 years) in different languages (English, French, German, Greek, Italian) and by means of different experimental methods. In general, all these studies seem to agree on the observation that children fail to derive pragmatic inference, their performance being different from adults’. Authors close to the Relevance Theory approach attributed this fact to the limited resources of children’s processing system, which is not yet mature enough to bear the load of implicature computation (as observed in Chapter 1, implicature derivation is considered costly in this approach, cf. section 2.2). Others explained children’s inability to compute SIs by appealing to a general difficulty in understanding task requirements (cf. for example Papafragou and Musolino, 2003).

Back in the eighties, investigating the logical competence of children from 4 to 7 year of age, Carol Smith (1980) found that children consistently interpreted some as meaning some and possibly all. In the same period, Braine and Rumain (1981) concluded that 7- to 9-year-olds tend to assign or an inclusive reading, taking sentences of the kind A or B to mean A or B and possibly both. More recently, Noveck (2001) presented a first systematic investigation of the emergence of SIs in children, modeled on the previous work by Smith (1980). Using a Statement Evaluation Task (SET) in which the subject had to evaluate a
series of sentences containing *some* and *all*, he showed that otherwise logically competent children do not derive the SI associated to *some* up to age 11. For example, 7-year-old children rejected the underinformative statements *some elephants have trunks* only 11% of the time. These same children did not have any problem in rejecting false statement including *all* and in accepting true statements including *some* or *all*. However, as discussed in Guasti et al. (2005) and Papafragou & Musolino (2003), children’s general failure in deriving SIs in the Noveck’s study seems to depend on features of the experimental design. In particular, SET is a task that requires contextual knowledge not readily available to the experimental subjects; moreover, the question that was asked to subjects, posited in form of agreement/disagreement with the statement being made, remains rather vague in ways that is likely to affect their performance. Using a different task, the Truth Value Judgment Task (TVJT henceforth, Crain and Thornton, 1998) Guasti et al. (2005) tested the youngest age group investigated by Noveck, i.e., 7-year-olds, and showed that these children were as competent as adults in deriving SIs. In a typical trial using the TVJT, children are presented with stories acted out using toys and props in front of the children. At the end of the stories, a puppet is asked by the experimenter to report what happened. For example, the puppet’s statement could be a sentence like *some monkeys are eating a cookie* uttered in a situation in which *all* the relevant monkeys were eating a cookie. Children had to say whether the puppet was right or wrong (or said it well or badly) with respect to what happened in the story presented. In Guasti et al. (2005), 7-year-olds rejected the puppet’s underinformative statements as much as adults did. Thus, the use of a different task like the TVJT, which seems arguably more appropriate to test pragmatic inferences than the SET for the reasons just mentioned, does indeed make a difference. However, SIs remain problematic for 5-year-olds, following Chierchia et al. (2001/2004) and Papafragou and Musolino (2003). Always using the TVJT, these last mentioned authors investigated the interpretation of three different scalar items (*some, two, start*) in 5-year-old Greek speaking children and found a low percentage of derivation of the SIs associated to *some* and *start* (12.5% and 10%, respectively) and a better performance on the
numeral scale (65% of derivation of the SI). Papafragou and Musolino conjectured that children’s poor performance was due to a failure in understanding that they were not being asked about the truth or falsity of statements (which were in fact true, given that *some* is in a subset relation with *all*), but about the pragmatic appropriateness of the statements. To test this conjecture, Papafragou and Musolino designed a second experiment in which the experimental goals were made as explicit as possible, by adding a training session before the test started. During the training, children were made aware of pragmatic anomalies and instructed to correct the puppet when she provided a true but underinformative description of some objects in the scene. For example, when she called a dog a “little animal with four legs” instead of using the more appropriate term “dog”. After the training, children’s performance improved, and the computation of SIs rose to 52.5% in case of *some*, 47.5% in case of *start* and 90% in case of numerals. However, these results have to be taken with some caution, as children’s performance with *some* and *start* remains well below adults’ performance (which is around 90% of derivation of SIs) and is likely to be not different from chance. Thus, although the explicitness of the experimental goals seems to matter, the impact of explicit training remains limited. In a similar vein, Papafragou (2003), using a novel task which she claimed to be more naturalistic, still found that 5-year-olds were reluctant to reject underinformative statements including the aspectual verb *start*, its synonymous *begin* (50% rejections), but were more willing to reject underinformative statements including *half*.1 Poor

---

1 In this task, children were asked to reward a puppet or not depending on whether she had completed a certain task. For example, the puppet had to color a star. She first goes away to accomplish the task, then she comes back (without the star) and the experimenter asks her whether she has colored the star. The puppet answers were of two types: *I began/started to color the star* or *I colored half of the star*. In these conditions, it was found that children refused to award the prize to the puppet around 50% of the time on the ground that she did not complete her job. However, it is not clear whether we can interpret children’s refusal as an indication that they have derived the SI related to the verb, for the reasons that follow. Consider that the basic meaning of *I began to color the star* is (i) while (ii) gives its enriched meaning:

(i) I began to color the star, and perhaps I finished to color it
(ii) I began to color the star, but did not finish to color it

Given that the task is to reward the puppet if it is assumed that she completed the task, I expect children who interpret the sentence as in (ii), thus deriving the SI, not to reward the puppet. However, even the children who interpreted the sentence as in (i), thus not deriving the SI, may not want to reward the puppet for the reason that she did not say explicitly that she completed the
performance with another scalar item, *or*, is reported in Chierchia et al. (2001/2004). As we saw always using a TVJT, they found that *or* tends to be interpreted inclusively by children (about 50% of the time without any training) even in a context in which the exclusive interpretation is more appropriate and thus favoured by adults. Finally, Doitchinov (2003) reported that 8-year-olds fail in the derivation of the pragmatic inferences associated to the use of epistemic modals (see also Noveck, 2001), although they were behaving like adults with respect to the SI associated to *some*.

Summarizing the previous findings on SIs, four general conclusions seem to emerge from the acquisition literature so far:

\[(3)\]

- children are less competent than adults in computing Scalar Implicatures (Smith (1980), Brain and Rumain (1981), Noveck (2001));
- children at 7 seem to be as competent as adults in the derivation of SIs if the task is appropriate (Guasti et al. (2005));
- children at 5 still have difficulties in deriving pragmatic inferences even when the task is designed to favour their derivation (Chierchia et al. (2001/2004), Papafragou & Musolino (2003), Papafragou (2003));
- children seem to respond differently to different scalar items.

Taking these considerations into account, the following questions naturally emerge:

task, but merely that she “started and perhaps finished”. As a matter of fact, given that the puppet is explicitly asked if she colored the star, a simple “yes” answer would suffice to get the prize. Answers others than “yes” may suggest that the task was not completed, independently from the calculation of the SI related to the scalar term used in their response. Therefore, it is not clear if we can interpret these results as an indication that children derive SIs more readily in this task.
(4)

(i) at what age do children manifest an adult-like behaviour in the generation of scalar implicatures?
(ii) which factors are responsible for children’s poor performance?
(iii) conversely, which factors can contribute in improving their performance?
(iv) do children answer randomly, as Papafragou and Musolino’s results seem to suggest?
(v) why do some scalar items elicit more rejections than others?

I addressed these questions through a series of experiments with different age groups of Italian speaking children, throughout which I kept the methodology constant. Firstly, I conducted a developmental study on the scalar item some (Experiment 5), specifically to address the question of which is the critical age for the emergence of an adult-like behavior in the derivation of SIs. This study is important to properly evaluate previous results and to set the ground for further experimentation. Then, by means of different experiments, I investigated the possible factors that might be responsible for the non-adult like behavior displayed by children with respect to SIs in the previous studies. These factors can be identified in the following terms:

(5)

(i) difficulty linked to the particular version of the scalar term used in the test (this factor will be tested in Experiment 6);
(ii) insufficient or incorrect general knowledge of the items involved in the scale and their uses in felicitous contexts (tested in Experiment 7);
(iii) lack of awareness that items differ with respect of informational strength or failure to meet relevance expectations and/or working memory limitations (addressed in Experiment 8);
(iv) difficulties in the actual recovery of the scale (addressed in Experiment in 9);
(v) flaws of the experimental design (addressed in Experiment 10).

Analyzing children’s distribution in previous experiments, I conjectured that a strategy was involved in children’s responses and, consequently, that the general tendency to accept underinformative statements could be reverted by exploiting a different experimental design (this was in fact the specific aim in Experiment 10).

2. **Experiment 5: is there a developmental effect in the computation of SIs?**

Although there is general agreement on the fact that children are less prone than adults in generating SIs, different results are attested depending on the age tested and on the task and experimental methodology employed. To establish whether there is a developmental effect, I carried out an experiment in which children aged between 4 to 7 years were tested by means of a TVJT on the quantifier *some*. Aim of this experiment was to determine to what extent children of different ages, which were tested using the same method and the same material, were able to infer SIs. This assessment is important for properly evaluating the results of previous studies and is a prerequisite for further investigations on the factors responsible for children’s failure.

2.1 **Subjects**

Sixty-two Italian children ranging in age from 4- to 7-years-old and twelve adults participated in the study. Eleven children were discarded from further analysis since they did not conclude the task, had language problems or were not cooperative during the experimental session (respectively for a total of: five children aged 4, five aged 5 and one aged 7). Subjects were divided by age in five
different groups: thirteen 4-year-olds (Age Range: 4;1-4;10, Mean Age: 4;5, SD=.31), twelve 5-year-olds (Age Range: 5;0-5;8, Mean Age: 5;3, SD=.31), twelve 6-year-olds (Age Range: 6;1-6;9, Mean Age: 6;2, SD=.12 ), fifteen 7-year-olds (Age Range: 7;0-7;9, Mean Age: 7;5, SD=.28), twelve adults. Children were recruited from the nursery schools Giustizia (Milan municipality), Milano Due and the primary school in San Pellegrino Terme (Bergamo municipality). Adults were undergraduate and graduate students at the University of Milano-Bicocca who volunteered to participate.

2.2 Materials and Procedure

I used a videotaped version of the TVJT to test children and adults about statements including *some.* The experiment consisted of five critical trials containing the scalar operator *some*, which were true but pragmatically inappropriate in the context used. On a typical trial, there were five characters, e.g., five Smurfs who were on holiday, that could choose between two alternatives, e.g. to go for a boat trip or a car trip. To place emphasis on quantity the narrator commented that there were many characters in the story and that it would have been interesting indeed to find out how many of them would choose option A and how many would choose option B. Then the story went on, with each character expressing his/her preference for one or the other solution. In the end, all the characters opted for the same thing: in our example they all decided for a boat trip. At this point, the narrator asked a puppet, Carolina, to say what was happening in the story in the end. She always provided a description by means of critical statements containing *some*, like *Some Smurfs are going on a boat*, which could be judged “true” only if the SI associated to *some* was not

---

2 Papafragou and Musolino suggested that the TVJT might not be appropriate to measure children’s communicative abilities, since it is designed to elicit judgments about truth or falsity that pertain to the domain of semantics. This view is only compatible with theories in which the processes that derive the basic and the enriched meaning are carried out by distinct modules (semantics and pragmatics) and are of a distinct nature. Under a grammar-driven approach, instead, SIs are factored in during the compositional computation of meaning and thus the enriched meaning, like the basic meaning, can be evaluated in terms of truth value judgments. Of course, for the task to be adequate, it is necessary that intentions of the speakers are clear to the hearer, as they always have to be in any successful communicative situation.
derived (see Appendix F for the complete set of material used). Below are the clips of the beginning and ending of the target trial taken as an example:

(6)

At the beginning...

This is a story about some Smurfs on holidays. They have a car and a boat. They are deciding what to do: going on a trip by boat or have tour by car? Let’s see what happens!

In the end...

Experimenter:

Carolina, would you tell us what is happening in the end?

Carolina:

Some of the Smurfs are going on the boat

Subjects were then asked to judge if Carolina said it “well or badly” with respect of what had really happened in the story. If they did not agree with Carolina’s statement, subjects were asked to say what was wrong with her statement and how she could have said it better, so that the experimenter could take note of the suggestion and inform the puppet (that was not present during the experiment) about how she could improve her Italian. To ensure that subjects could reject false statements and didn’t give their answers at chance, critical stories were interspersed with fillers (a total of four), two of which were clearly true and two clearly false. Moreover, two warm-up stories (one true and one false) preceded the test session. If a subject performed poorly on the two warm ups and on two of the fillers, her answers were excluded from data analysis (this was the case for 6 children overall who never rejected false statements). All the stories were acted out in front of the puppet using props and toys and were recorded with a digital video camera. They were then transferred on a CD to be played on a
laptop in front of the subject (total recording time: approx. 15 minutes). I used the digital version of the stories with the older children and adults, while I decided to use a live version of the same experiment with the group of the 4-year-olds, acting out all the stories in front of the children to better capture their attention. Children were tested individually in a quiet room, where they watched the video together with an experimenter. In case of the 4-year-olds, two experiments were present, one acting out the stories live and the other animating the puppet.\(^3\) Children were invited to indicate their answers and comments to the experimenter, who filled a score sheet and took note of the explanations. Some days before the proper experiment started, the group of children was acquainted with the experimenters who spent a couple of hours in the daycare playing with the children and with the snail puppet Carolina. She was introduced to the children as a baby snail who has just moved to Italy and was willing to improve her Italian with the help of children, given that she thought adults to be annoying and boring. Children were asked to help Carolina in this task, by participating in a game of story-telling one of the following days. Nearly all the children eagerly volunteered to help.\(^4\) The test sessions were tape-recorded for further control. Adults were shown the video in small groups and were given a score sheet to write their comments to the puppet’s statements.

### 2.3 Results

The distribution of subjects’ answers with respect to the mean acceptance rate of critical statements for each age group is plotted in the graph below.

---

\(^3\) I want to thank Joke De Lange for patiently helping me in carrying out the experiments “live” with the 4-year-olds.

\(^4\) We previously asked the parents’ consensus to include their children in the study.
It is immediately evident from the graph that the group of older children (6- and 7-years-olds) patterns like the group of adults but differently form the group of younger children (4- and 5-years-olds). As a matter of fact, less than 20% of the population of older children and adults accepted the critical statements. Respectively, 16% of the 6-year-olds, 18% of the 7-year-olds and 13% of adults didn’t derive the implicature associated with the scalar term. Conversely, the rate of acceptance of the same statements was above 50% for the group of younger children. Respectively, 53% of the 4-year-olds and 58% of the 5-year-olds didn’t derive the SI. On control items, instead, no difference was found between adults and children of all age groups: adults responded correctly 97% of the time; 5-, 6- and 7-year-olds gave correct answers 100% of the time; 4-year-olds 96% of the time.

As in previous works (Noveck, 2001, Papafragou and Musolino, 2003), I submitted the proportions obtained to the analysis of variance ANOVA using the proportion of “No” responses as the dependent measure and age as a factor, but keeping in mind that the data did not yield a normal distribution (see also Guasti et al., 2005). A main effect of age was found (F(4, 58)=3.91, p <.01). Post hoc Duncan test showed that the responses of 4- and 5-year-olds differ significantly from the responses of 6-, 7-year-olds and adults (p<.05).
A further analysis was conducted, in which I considered subject’s distribution with respect to the number of times they accepted the critical statements (from 0 up to 5 times, given that each subject was presented with five different critical trials). The distribution of each age group is plotted in Graph B below:

Graph B. Subject’s distribution with respect to age and number of times they rejected the critical statements (note that all age groups comprised 12 subjects except for the 4-year-olds who were 13 and the 7-year-olds who were 15).

To establish whether the developmental effect brought up by the ANOVA in the preceding analysis was reliable, subjects were then divided in two groups: one group was comprised of the subjects who accepted the critical statements 3 or more times, and the other group was comprised of subjects who accepted the critical statements less than 3 times (cf. Guasti et al., 2005). It turned out that 58% of the 4- and of the 5-year-olds, respectively, accepted the critical statements three or more times, while only 17%, 13% and 8% of the 6- and 7-year-olds and adults, respectively, accepted them. An analysis of proportions applied to these data revealed a significant difference between the group of the 4- and 5-year-olds on the one hand and the group of the 6- and 7-year-olds and adults on the other (p<.05).
2.4 Discussion

This first experiment shows a clear developmental effect in the emergence of the ability of computing SIs in which the age of 6 seems to be the critical turning point. From 6 years on, in fact, children seem to behave like adults in mostly rejecting the critical statements with *some* when a parallel description with *all* would be more appropriate. This is taken as evidence of the fact that children (and adults) are computing the SI associated to *some*. By contrast, children younger than 6 tend to split: only half of them behave like older children and adults, mostly rejecting the critical statements, while the other half mostly accepted the critical statements. This result seems to show that at the age of 4 and 5 not all children properly master the ability of deriving the SI associated to *some*, given that 50% of them accept a sentence like “*Some Xs did Y*” as a description of a story in which *all Xs did Y*.

By analyzing individual distributions, we observed that younger children were not answering randomly but were in fact very consistent with their answers, as their bimodal distribution reveals (cf. Graph B above). As we will see, this observation will reveal crucial.

With respect to previous findings (cf. for example Noveck, 2001), my experiment shows that, by the age of 6, children can already derive the SI associated to *some* as much as adults when the appropriate evidence for the evaluation of the statements is available. This result confirms (and extends) what was previously found in Guasti et al. (2005) with 7-year-old children. It also corroborates Papafragou and Musolino’s (2003) decision to include 5-year-old children in their studies on the comprehension of scalar terms, since this age seemed somehow critical for the derivation of SIs.

A final observation is worth mentioning at this point: children who rejected the puppet statements always motivated their rejection by invoking the strongest term in the scale, saying, for example, that *Some Smurfs are going on a boat* was a bad description of the story because *All Smurfs are going on a boat*. This not only indicates that children understand the task (this fact being also shown by the appropriateness of the corrections in case of control–false
conditions). It also clearly indicates that children rejected the critical underinformative sentence for the right reason, recognizing that a scale and a scalar comparison was involved in the sentence.

Although these results provide an answer to the first of the questions raised in (4), they do not help us in understanding what prevents children at 4- and 5-year-olds to derive pragmatic inferences - if it’s a matter of inability at all. In their 2003 paper, Papafragou and Musolino propose that children have problems with SIs because they do not understand the experimental demands, i.e., children are not aware that they are asked to judge the appropriateness of the sentences and not their truth. These authors show that when it’s made clear that they are being asked about the pragmatic appropriateness of a sentence through a training session and by manipulating the story, children’s performance improves. For example, children’s rejection rate of critical statements including some raises from 12% in their Experiment 1 (without training) to 52% in their Experiment 2 in which a training session was added. However, training and awareness of the experimental goals might not be the only or even the effective reasons of the observed improvement in Papafragou and Musolino’s second experiment. In fact, some other (not explicitly mentioned) manipulations were made in this second task, in addition to the adding of a training session. Firstly, the age range in this second experiment was wider, given that some 6-year-olds were included (age range of children varied between 4;11 and 5;11 (Mean Age=5;3) in Experiment 1 and between 5;1 and 6;5 (Mean Age=5;7) in Experiment 2). Secondly, the scalar items in the sentences were tested in object rather than in subject position (where they appeared in Experiment 1); thirdly, the puppet was not asked to say what has happened in the story (as it was in Experiment 1), but to comment on how well the main character did; fourthly, the focus of the whole story was explicitly directed on the main character’s performance. Given that all these manipulations were introduced at once, it is not clear whether the improvement is due to a single factor or to a constellation of factors. The first of the manipulations listed above is particularly relevant to our discussion given that we know, from the results of my own experiment, that the age of 6 is critical with respect to SI computation.
Hence, the possibility that a higher rejection rate in Papafragou and Musolino’s experiment with training is effectively due to the age of the population (at least to some extent) is to be taken into account. This conjecture is in fact all the more reasonable considering that, even after training, the rejection rate remains overall low (52%). According to me, this could signal the fact that only some of the children benefited from training and generated SIs in the second experiment.

Another point is worth of remark, I believe. It is surprising that 5-year-olds rejection rate in the experiment I conducted (i.e. 42%) is much higher than the rate of rejections in Papafragou and Musolino’s experiment that functioned as a baseline (12%), especially because the task used in these two experiments is the same. One possible explanation for this divergence in the results obtained is the number of characters or objects on the scene to which the narrator referred to when using some or all. This factor constitutes a crucial difference between the two experiments: the relevant characters/objects in the Papafragou and Musolino’s study were only three, while I used five characters/objects in my study. In fact, I consider somehow infelicitous to use some or all when there are only three characters or objects, because the most appropriate way to describe such a situation would be, in my opinion, to report the exact number of the characters and objects involved in the action. To report small numerosities, in fact, one can rely on the automatic process of subitizing.

Let’s go back to our initial questions about which factors that can influence children’s performance: if it is not, or not only, awareness of the experimental goals that matters, what else might be responsible for children is apparent failure to infer SIs? And why, ultimately, do subjects’ responses give rise to a bimodal distribution? In the attempt to answer these questions, I started by testing the first of the factors listed in (5), namely, if the difficulty emerged with some in the experiment just discussed could be linked to the particular version of the scalar term used in that study. Please note that, given that children were found to be competent with SIs from the age of 6 onward, all the further experiments I’m going to present in this chapter will involve only 4- and 5-year old children.
3. Experiment 6: generalizing performance across similar items

This experiment investigates whether younger children’s performance in Experiment 5 is linked to the particular term used to test them, namely “qualche” (translated as *some*) or else if the results obtained in this first experiment can be extended to other semantically equivalent scalar terms, maintaining all the other conditions equal.

In Italian, there are at least two words that correspond to English *some*: *qualche* and *alcuni dei* (*some of*). While *qualche* is morphologically singular, even if it can refer to a plurality, *alcuni dei* (*some of*) is morphologically plural. Given the property of non-detachability of SIs, no difference is in principle to be expected between the two lexical variants of the same scalar term. However, it is also conceivable that, depending on the frequency of use of the different variants, one version is easier for children at 4 and 5, this also in consideration of the fact that these Italian operators, though semantically equivalent, actually differ a lot in their syntax and distribution. If children’s difficulty were linked to the particular operator used, then the use of a different operator should have an effect on their performance. On the other hand, if poor performance were due to the inability of computing SIs itself, then the use of a different operator should have no effect.

Thus, to verify that children’s performance in my first experiment was not to be attributed to the particular item chosen, I carried out a second experiment in all identical to the first one except for the fact that the critical statements included the variant *alcuni dei* instead of *qualche*.

3.1 Subjects

A new group of twelve 5-year-olds (Age Range: 5;4-5;11; Mean Age: 5;7) and twelve adult native speakers of Italian participated in this experiment. Children were recruited from the nursery schools Giustizia (Milano Greco) and Milano Due. Adults were undergraduate and graduate students at the University of

---

3 I used the partitive version of *alcuni* (i.e. *alcuni dei*) because this is more felicitous in the experimental setting used.
Milano-Bicocca. With the same criteria adopted in Experiment 5, three subjects were excluded from the analysis in this experiment.

3.2 Materials and Procedure

The test material was the same used in the first experiment, with the crucial difference that the critical statement at the end of each story contained the operator *alcuni dei* (*some of*) instead of *qualche* (*some*) (cf. Appendix G). Also the procedure was identical to the one previously adopted.

3.3 Results

Rate of acceptance of the critical statements containing *some of* was 50% for children and only 8% for adults. With respect to fillers, both groups performed correctly 100% of times. I thus drawn a comparison between Experiments 5 and 6, building on the fact that the two experiments employed exactly the same material and procedure except for the critical manipulation mentioned. In particular, I compared the acceptance rate of the group of the 5-year-olds and adults in Experiment 5 and 6 carrying out a 2 (Condition: *some* versus *some of*) × 2 (Age: 5-year-olds, adults) ANOVA with percentages of “yes” responses as dependent measure. I found a main effect of age, $F(1, 44)=12.67, p<.001$, but no effect of condition nor any interaction. As in the previous experiment, we divided each age group with respect of the number of “yes” responses: subjects that accepted the critical statements three or more times, and subjects that did it less than three times. As in Experiment 5, it turned out that children were split in two groups: out of the twelve 5-year-old participants, six always accepted the critical statements and six never did. With the exception of one, all the adults always rejected the critical statements instead.

3.4 Discussion

The main finding is that no difference emerged in subjects’ responses to critical statements between the two experiments: as previously found in Experiments 5, the 5-year-old participants in this experiments only accepted the
critical *some* of statements 50% of the time. By contrast, adults’ rejection rate was above 90% in both cases. As we discussed above, if children’s low rejection rate of the underinformative statements in the first experiment were due to the particular scalar item used (namely, *qualche*), by using a lexical variant of the same item one should find different results. However, no difference was found.

This is conceivable, given the property of non-detachability of SIs. Both the variants tested should in fact give rise to the same implicature, namely, *not all*.

For this reason, the fact that only 50% of the children accept the critical underinformative statements (thus not deriving the SI associated to some) seems to be related with the process of deriving SIs itself, not with the lexical variants of the scalar term used.

### 4. Taking stock

The results emerged from the experiments presented so far unequivocally show that children perform differently from adults with respect to SI computation. This results, as we said, corroborates the results presented by other authors. The first guess that is reasonable to make to explain children’s performance is that children at 5 are not yet “mature” enough to carry out the whole process of SI computation, which requires specific knowledge and resources. In the introduction to this chapter (cf. (5)), I identified a series of controls to check children’s (in)ability to perform each step that I consider necessary to derive SIs. So far, we have tested a first possible factor that could have affected children’s performance in Experiment 5, i.e. a difficulty linked to the variant of the scalar term used in that test. As Experiment 6 clearly shows, however, the same difficulty emerges by using another variant of the same scalar term. Thus, the reason why only half of the children population seems able to derive SIs should be found somewhere else in the derivation. In this respect, insufficient or incorrect general knowledge of the items involved in the scale and their uses in felicitous contexts may affect children’s performance (cf. (5)-ii). This factor is explicitly tested in the experiment that follows.
5. **Experiment 7: knowledge of some and all in felicitous contexts**

As a first step in the investigation of the factors that may prevent children from deriving SIs, I think it’s important to establish whether children understand the meaning of the two items involved in the scale in felicitous context. In this particular case, we need to check if children accept *some* and *all* when they are used to describe, respectively, a situation in which only some (and not all) characters are performing an action (and thus a situation that would render a statement with *some* true and one with *all* false) and a situation in which all characters are doing the same action (that would make a statement including *all* true).

In their 2003 paper, Papafragou and Musolino tested children’s comprehension of *all* in contests that made the statements true or false, but they did not test sentences with *some* in contexts that made them true and felicitous. We cannot discard *a priori* the possibility that, although children know the truth conditions associated to *all*, they are confused about *some*. This factor may be responsible for children’s low derivation of the SI associated to this item found in children.

5.1 **Subjects**

A new group of fifteen 4- to 5-year-olds participated in the experiment. They were recruited from the nursery school Pallanza in the Milan municipality. Three children were discarded from further analysis because they did not want to complete the test. Thus, final analysis is based on twelve children (Age Range: 3;11-5;9; Mean Age: 4;7, SD=.76).

5.2 **Materials and Procedure**

Test materials consisted of a total of twelve critical stories, eight of which were described by using statements containing *all* and four by statements containing *some*. Five of the eight statements containing *all* were false descriptions of the story presented and three were true, while the four statements...

---

6 I would like to thank Chiara Iogha for her help in submitting this test.
with *some* were all true in the situation given (see Appendix H for the complete set of sentences used). In addition, critical trials were interspersed with fillers, for a total of sixteen stories overall. As in the previous experiments, a digital version of the TVJT was used. The experiment was administrated on a laptop screen in a quiet room in the day care center after a familiarization phase with the children.

5.3 Results

Children accepted true statements containing *some* and *all* 92% and 100% of the time respectively and rejected false statements containing *all* 96% of the time. No statistical difference among these means is found (F(2, 22)=1.00, p=.38).

5.4 Discussion

Data obtained in this experiment univocally show that 4- and 5-year-old children do not have any difficulty in evaluating statements including the quantifiers *some* and *all* in felicitous contexts. This result corroborates and extends what previously found by Papafragou and Musolino (2003). From these results, I believe one can exclude the hypothesis that children’s difficulties with SIs are due to poor understanding of the quantifiers *some* and *all* or to confusion between these quantifiers. What is at issue, here, seems to be some difficulty related to some other step specifically linked to the process of SI derivation. The next experiment is devoted to test another aspect of this hypothesis.

6. Experiment 8: investigating children’s knowledge of informational strength

As we discussed in Chapter 1, some logical words, like the quantifier *some* I’ve been testing in my experiments, are ordered in a scale with other items with respect to their informational strength (cf. Chapter 1, section 1). In the scale we are considering (i.e. *some<all*), *some* is the lowest element, thus the less informative, while *all* is the most informative one. One might conjecture that young children do not know, or are not sensitive to, the fact that scalar items
differ in terms of their informational strength, or are less prone than adults to meet expectations of relevance based on quantity of information of a given statement. Therefore, they may not be able to choose which element of a scale is more informative and thus more relevant in a given context.

6.1 Subjects

Nineteen Italian-speaking children participated in the study. Two of them were discarded from further analysis since they didn’t complete the task. The seventeen children that were included in the analysis ranged in age from 5;1 to 5;11 (Mean Age: 5;7, SD=.29). All participants were recruited from the nursery school Giustizia (Milan municipality). Crucially, these were the same children that were previously tested in Experiments 5 and 6 reported above, and that failed to derive the pragmatic inferences in those tasks. They were tested on this new task some weeks after the first one.

6.2 Materials and Procedure

For this experiment, I used a task devised by Chierchia et al. (2001/2004) already successfully employed in the study of implicatures associated with *or*, namely the Felicity Judgment Task (FJT). In this task, subjects are asked to judge which one of two given statements constitutes a better description of a picture or situation. As in the other experiments, children were tested individually in a quiet room of the nursery school, but the experimental setting was different. This time, two experimenters were present, each manipulating a puppet, the seal Silvana and the raccoon Alberto.\footnote{I again want to thank Joke De Lange for her help.} Children were told that these new puppets were friends of Carolina, and, like Carolina, were trying to improve their Italian. Given that they were always quarrelling about which of them was the best, they decided to test their competence in a competition in which the child was asked to be the judge. The competition consisted in describing some pictures, which were displayed on the table in front of the child and the puppets. For each picture the two puppets in turn gave the better description they could and the child had to judge each time.
which puppet said it better. One of the two experimenters recorded the puppets’ score for each trial on a score sheet and the winner was declared in the end. Correct and appropriate responses were counterbalanced through "puppets” and sessions so that the final score of the winner was always 6 out of 11 and each puppet in turn resulted to be the winner of the competition.

The test material consisted of a series of eleven pictures, five of which were critical items described by the puppets by using statements including some and all. On a typical target trial, the picture displayed five characters performing the same action, for example five chipmunks going on a bus, as in the picture below. One of the puppets (in turn) described the picture with a statement containing some, which is true but informationally weak in the context, while the other puppet described it with the more appropriate (informationally stronger) statement containing all. A sample of the test material is given below:

(7)

Interspersed were 6 fillers, i.e. pictures that were described by a clearly true statement by one puppet and by a clearly false statement by the other in turn,
to check that children paid attention to the task. The complete set of the sentences used is given in Appendix I.

6.3 Results

In the critical conditions, nearly all children chose the statement with *all* when the choice had to be made against the informationally weaker alternative containing *some*. The percentage of correct answers on critical items was actually 95%, while on fillers it was 99%.

6.4 Discussion

I believe that the results obtained in this experiment unequivocally show that children know that a statement containing *all* is more informative, thus more appropriate, than its counterpart with *some* in a situation in which all characters are performing the same action. This amounts to saying that they know the maxim of quantity and also they know that it is relevant to solve the task. This, on the one hand, allows us to reject the hypothesis that children lack knowledge of the informational strength associated to scalar items; on the other, it also shows that children can meet the expectations of relevance by choosing the most informative statement. This outcome is remarkable if we consider that the knowledge of an ordering in informational strength and the ability to meet relevance expectation are a prerequisite for process of computing SIs. It is even more if we consider the fact that the children tested in this experiment were the same subjects that participated in the previous experiments and that crucially failed to derive the implicature. Or, at least, didn’t consider the implicature relevant to that particular situation in that specific task.

It’s also interesting to note that my findings are in line with those reported by Chierchia et al. (2001/2004) on *or* using the same methodology. In that study too, they found that children tested with the TVJT were equivocal in deriving the implicature associated with *or*, accepting statements including *or* in situations in which *and* would have been more appropriate. Nevertheless, when children were asked to choose between alternative descriptions, as in the FJT, no child preferred
a statement including or over a statement including and when this was more appropriate in the situation provided.

It’s worth noticing, however, that the results obtained with the FJT do not provide evidence that children can derive SIs. They merely prove that some piece of knowledge that is necessary but not sufficient to derive SIs (i.e. the knowledge of an ordering in informational strength) is available to them. This does not prove that they master the whole knowledge necessary for SI computation, considering the fact that the process of deriving SIs is far more complex than choosing the most informative between two given alternatives. It requires the hearer to understand that the utterance with some conveys the meaning that not all holds for the speaker. In turn, this is possible if subjects know the scale, can recover it by themselves, are able to generate the alternative description containing the stronger item in the scale and thus negate it for the fact that the weaker term has been used. Knowing that two items are different in informational strength when asked to compare them once they are provided does not imply knowing that these items are ordered in a scale, nor that one is automatically activated upon hearing the other. Moreover, it doesn’t prove that children can construct and/or recover the relevant alternative by themselves, given that in the FJT the two alternatives are given.

However, the present experiment is critical in another respect. Reinhart (1999) observes that children behave poorly in tasks in which they are asked to evaluate two alternative descriptions, as it is the case in sentences involving association with focus, in sentences including non-reflexive pronouns or in sentences involving SIs. Recall that, upon hearing an underinformative statement with some, children have to recover the scale, build the alternative representation containing all, compare the two sentences and reject the first on the ground that the second is informationally stronger, thus more appropriate. In such situations,

---

8 It is well-known that English speaking children sometimes take (i) as meaning (ii)
(i) Mama Bear scratches her
(ii) Mama Bear scratches herself
According to Reinhart (1999) this happens because, in order to reject (i) as a false description of a situation in which Mama Bear is scratching herself, children have to recover the alternative description in (ii), compare it to (i) and reject (i) on the ground that there is an informationally stronger, and thus a more appropriate way to express the property of self-scratching, i.e., (ii).
Reinhart claimed, children know what they have to do to, i.e., they know that they have to reject the informationally weaker statement, but they fail to do so because they are unable to hold two representations in their memory store. My results clearly show that children’s memory is not limited in this way, given that they can successfully compare and choose between alternatives. In my view, this makes it very unlikely that children’s poor performance with SIs is due to memory deficiencies.

7. **Experiment 9: priming the scale**

In the Felicity Judgment Task used in Experiment 8 the two alternative descriptions were readily given to the children, whose task was simply to choose which was the best one. In normal conversational exchanges, though, the two alternatives are not given: by hearing *some* one should recall the scale which *some* belongs to, and then consider that, since the speaker used *some*, which is low in the scale, she knows/thinks that every higher elements in the scale do no hold. In particular, that is the case that *not all*. I considered the possibility that the activation of the scale is less automatic in children than adults, and tested this hypothesis in a new experiment specifically designed to facilitate the priming of the scale *some*<all.

7.1 **Subjects**

A new group of twelve 5-year-olds (Age Range: 5;4-5;10, Mean Age: 5;7, SD=1.83) participated in this experiment. Children were recruited from the nursery school Giustizia (Milano Greco).

7.2 **Materials and Procedure**

The test material was the same used in Experiment 6, with the exception that a new set of characters was added to each critical story (cf. Appendix L). So, for example, in the story of the Smurfs on holiday that I have described in section 2.2, a bunch of dwarfs and some candies were added to the same set. Contrary to
what I did in previous experiments employing the TVJT, I decided to act out the stories live in front of the child because the toys used in these new stories were too many to be properly captured in the video. In this “live” version, two experimenters were involved, one narrating the stories and the other manipulating the puppet. Differently to what happened in Experiment 6, here each story comprised two parts: the first part focused on the added group of character (e.g. the dwarfs) and their objects (e.g. candies) while the second was basically the same story used in the previous experiment. On a critical trial, first the experimenter described the situation and introduced the characters to the puppet and the child. For example, she introduced the stories by saying:

(8) _This is story about a group of Smurfs and a group of dwarfs that are on holidays. Look how many of them we have! They can do a lot of interesting things here. For example, they have a boat, so they can go for a trip on the river by boat. They also have a car and they can drive their car in the forest. Oh, look! There are also a lot of candies here! If they are lazy, they can even relax and eat candies. But let’s see what happens…_

Then the proper story started. At first, the attention of the child was drawn on the group of added characters, in our example, the dwarfs, that were discussing about eating candies. In the end every dwarf took a candy and sat down to eat it. At this point the experimenter stopped narrating the story and addressed the puppet, asking her what was happening in the story so far. The puppet always described the first part of the story by using a true and appropriate statement containing _all_, for example _All dwarfs are eating candies_. The child was immediately asked to evaluate if the puppet said it “well or badly” with respect to the situation. Then the story went on focusing on the second set of character. Crucially this part was identical to the one provided in the previous experiment (in our examples, this part ended with all the Smurfs going on a boat). Again, the puppet was asked to tell what was happening and this time she described the story

---

9 Let me thank again Joke De Lange for her help in carrying out this experiment.
with a true but underinformative statement containing *some* (the same statement used in Experiment 6). Again, the child was asked to evaluate the puppet’s statement. Every time she judged it “bad”, the child was asked to correct it so that the puppet could learn the correct or better way to say things.

As for previous experiments, the child was tested in a quiet room and was told that she will watch some stories together with Carolina (previously introduced to the children in the class as a baby snail learning Italian) to help her to improve her language. The whole experimental session was audio-recorded.

All subjects who gave non-adult responses in this task were submitted to the FJT used in Experiment 8 ten-fifteen days after the first session.

### 7.3 Results

The acceptance rate of correct *all*-statements and fillers was 100%. No children in fact had any hesitation in accepting the true and appropriate description with *all*. On the contrary, the acceptance rate of the underinformative *some*-statements was lower, being 58%. The proportions of acceptance rate of the *all* vs. *some* statement are significantly different from each other (p<.001). Moreover, the rate of acceptance of the underinformative statements in this experiment was compared to the rate of acceptance of the same sentences in Experiment 6. An ANOVA was carried out with percentages of “Yes” responses in the two experiments as a dependent measure: no effect of condition was actually found (F (1, 22)=.006, p=.93, n.s.).

### 7.4 Discussion

The manipulation made in this experiment aimed at favouring the activation of the scale in virtue of the fact that hearing an *all*-statement before a *some*-statement should prime the scale. This should make the child aware of the fact that the second statement is less appropriate than the first in the given situation with the effect of having more subjects rejecting the critical statement with *some*, in comparison to Experiment 6. If the difficulty that emerged in the previous experiment were effectively due to a difficulty in the spontaneous
retrieval of the scale, then the mentioning of the stronger scalar alternative should result in an improved performance by children. However, the low percentage of rejection of critical statements seem to disconfirm this fact, as it seems that the manipulation introduced in this experiment had no effect whatsoever in improving children’s performance with respect to SI computation.

It’s interesting to note, in this respect, that half of the subjects treat the two types of statement (*some* vs. *all*) differently. And that those that do so, thus rejecting the underinformative statement with *some*, do it for the right reason: when asked to say what was wrong with the puppet’s statements, all the children correct her by substituting *all* to *some* (*but not all*). Moreover, the children that “failed” in deriving the SI in this experiment were submitted to the FJT: all performed 100% correctly in this task except for one.

8. **Tacking stock**

On the basis of the experiments I have presented so far, we assessed that: (a) children know what *some* and *all* mean; (b) they can compare statements that differ in quantity of information; (c) they can meet expectations of relevance; (d) they seem not to have memory limitations that prevent them from handling two representations in mind. Given this rich system of knowledge and the ability to meet communicative expectations, it is all the more surprising that children derive SIs less than adults. But in fact this is what emerges from all previous studies. In my opinion, this statement needs further qualifications.

To explore the question further, let us return to one intriguing result of Experiment 5. Remember that younger children were not uniformly less prone than adults in deriving SIs, but were split into two groups: one group that, like adults, always derived SIs, and one that never did. I claim that this result comes about because children persevere in the answers they offer the first time. In fact, the response given to the first target statements (and the explanation provided) is reiterated for all the other target statements of the same form. Chierchia et al. (2001/2004) also found such a bimodal distribution and so did Papafragou and
Musolino in their second experiment (as one can infer from their footnote 9). This outcome is rather puzzling and is a hint that children may have adopted some sort of strategy to which they stick during the whole experimental session. This conjecture seems plausible if compared to adults’ behavior in other studies on SIs. In particular, the groups of adults studied by Noveck (2001) and Guasti et al. (2005) by means of a Sentence Evaluation Task (SET) displayed a bimodal distribution: roughly half of them rejected the critical underinformative statements with some and half didn’t. As Guasti et al. (2005) discuss, when tested by a different task such as the TVJT in which relevance expectations are made clearer, adults improve and overwhelmingly get to reject underinformative statements. The conjecture put forward by Guasti and colleagues to account for adults’ behavior in the SET is that a strategy was adopted: half of the adults (those that rejected the critical statements) interpreted the experimental instructions to be about the informativeness of the given description (in an imagined conversational background) and followed the same strategy commonly used in normal conversational exchanges. The other half, instead, is thought to have adopted a second strategy, with two possible variants: they might have suspended the usual conversational norms because engaged in an experimental task of some sort in which these norms were not considered to hold, given that their application in the particular situation rendered all the statements proposed patently false. Thus, they might have conjectured that the experimenter was not asking them to agree or not with false statements and for this reason they might have concluded that either the experimenter had in mind the other meaning of some (some or even all) or she was asking subjects to find a context that made the statement not trivially false (e.g. a context which included baby giraffes in evaluating “Some giraffes have a long neck”).

The possibility that a strategy is involved in adults response pattern is further supported by Noveck and Posada’s (2003) findings. In their SET study (cf. also the discussion in Chapter 2, section 2.2), the group of adults who always rejected the critical underinformative statements with some took longer to evaluate all the statements, even those that did not involve a SI. Conversely,
adults who always accepted the critical statements were overall faster to respond to all kinds of statements. The split observed in adults performing the SET may be interpreted as a hint that they specifically approach this task appealing to different strategies that they maintain during the whole experimental session. When adults perform a more naturalistic task, like the TVJT, the strategy adopted is instead univocal and resembles the one that they normally adopt in ordinary conversational exchanges.

Turning to children, on the other hand, we observe that there is a group that always patterns differently from adults. The 7-year-olds tested by Guasti et al. (2005) using the SET overwhelmingly accepted the critical underinformative statements, behaving differently from adults that split in this task instead. Analogously, the 5-year-olds tested in my experiment using the TVJT (as in others study that used the same methodology) split, while adults overwhelmingly rejected the underinformative statements. Crucially, however, when 7-year-old children were tested in the SET but after a training session was added (in which the child was instructed to select the more specific description of an object, like grape instead of the more general term fruit), they patterned like adults, yielding a bimodal distribution. A similar improvement in children’s answers was observed in Papafragou and Musolino’s Experiment 2, as we discussed above, where a training session was added in which experimental goals and expectations were made as explicit as possible to evoke a pragmatic response in children.

In consideration of these facts pertaining to children’s and adults’ performance in different tasks, it seems all the more plausible to view subject’s performance as the result of a response strategy which basically depends on the type of task and on the age of participants. In terms of Chierchia’s formulation I am adopting here, this would amount to saying that feature selection associated to scalar terms (i.e. $\pm \sigma$, cf. Chapter 1, section 2.3) is ultimately affected by the type of task (namely, extra-linguistic context) and by the age of subjects. In these terms, then, why should younger children in general prefer $[-\sigma]$ over $[+\sigma]$? My guess is that the younger the subjects are, the less they “dare” in making their choice. If the context is “neuter”, such as in the SET, children tend to follow a
cautious strategy, overwhelmingly selecting $[-\sigma]$. At the same time, however, young children appear to be sensible to more or less explicit manipulations of the task demand: when it is made clear that pragmatic appropriateness, and not simple falsity, can be a basis for rejection (for example, by the adding of a training session before the SET or by using a more naturalistic setting like the TVJT), then they “dare” more, i.e., shift their strategy and select $[+\sigma]$. Yet, readiness to understand task demands remains in my opinion age-dependant, given that only older children and adults overwhelmingly shift their strategy depending on task (SET vs. TVJT). The fact that they shift proves, in my opinion, that nothing prevents children to derive SIs. If this hypothesis is tenable, then the fact that half of the 5-year-olds don’t compute SIs in the TVJT (Experiments 5 and 6) has little to do with objective impediments (such as memory limitations) in deriving such implicatures.

At this point, a further attempt could be done to enhance the probability that 5-year-olds revert their strategy and “dare”. This was the guiding line of the last experiment I’m going to present, in which the manipulation is made at the level of experimental design and not simply in terms of instructions.

9. **Experiment 10: a novel experimental design**

In this experiment my aim was twofold. In the first place, I wanted to design an experiment which favored the emergence of SIs even in children as young as 5, in order to finally assess their competence in pragmatic inferencing. In second place, I wanted to investigate the SI associated with different scalar terms, in order to test the hypotheses that a lexical component is involved in SI computation.

9.1 **Subjects**

A group of forty-seven 5-year-olds and forty adult native speakers of Italian participated in the experiment. Children were recruited from different nursery schools in the Milan Area: nursery schools Cesalpino and Soffredini
(Milan municipality) and the nursery school in Cornate D’Adda (Milan). Adults were undergraduate and graduate students at the University of Milano-Bicocca who volunteered to participate. As in previous experiments, those subjects that, for any reason, didn’t complete the task were excluded from the analysis. This was the case for seven children, so that final data comprise a total of forty 5-year-olds (Age Range: 4;11-5;11; Mean Age: 5;4; SD: 0.15).

9.2 Materials and Procedure

In their 2003 study, Papafragou and Musolino tested different scalar items as a between-subject factor. In this experiment I tested three different scales (numerals, *some* < *all* and *a piece of* < *whole*) but treating these as within-subject factors. Thus, each subject was presented with all scalar items but only saw one critical sentence of each scale. In addition, each subject was presented with control items for all different scalar items. The material consisted of 48 statements based on the three scales: *<alcuni dei, tutti>* (some of, all), *<due, tre>* (two, three), and *<un pezzo, tutto>* (a piece of, whole). It included 16 critical statements, 8 containing *some of* (4 in subject and 4 in object position), 4 containing *two* and 4 containing *a piece of*. In addition, there were 32 control statements; of these, 20 were true statements, 4 for each of the items involved in the scales: *all, some of, two, a piece of, whole* and 12 were false statements, 4 for each of the following scalar items: *all, two* and *whole*. The complete set of the sentences used is given in Appendix M. Below I report an example for one of the critical stories used to test the scale *<a piece of, whole>* (cf. also example (6) above, which was also used here):
This is a story about Gargamel, who is considering building a paddock where to cage the Smurfs when he captures them... Let’s see what happens!

Experimenter:
Silvana, would you tell us what is happening in the end?
Silvana:
*Gargamel built a piece of the paddock*

The 48 statements were divided in 4 lists of 12 statements each, so that each list included one statement of each type (for a total of twelve). Lists were treated as a between-subject factor when performing statistical analysis. Subjects were randomly assigned to one of the four lists, so that each subject was presented with only one sentence of each type, but all subjects were shown with statements from all the three scales. Items were presented semi-randomly, to avoid that two consecutive trials contained the same type of scalar item.

Some considerations about the material I used are in order here. First, I used the same material of Experiment 5 to test *some* and *two* in subject position. Differently from Papafragou and Musolino, I used three objects to test the scale *two*<more than two* and five objects for the scale *some*<all (cf. discussion section at the end of Experiment 5). Secondly, as described above, I tested *some* both in subject and object position to ensure that syntactic position itself didn’t affect the computation of SIs. Beyond what I already observed above (cf. section 2.4), this decision was motivated by the need to test a *piece of* in object position where it is more natural, and in this way to have a control for this position.
As in previous experiments, I used a digital version of the TVJT and played the stories on a laptop in a quiet room in the nursery school. As in previous experiments, children were first familiarized with the experimenter and the puppet, which in this case was a baby seal, Silvana, who was learning to speak Italian. Stories were previously recorded with a digital video camera (total recording time = 15 minutes approx.). Children answers were directly reported on a score sheet by the experimenter and the whole test session was also tape-recorded for further control. Adults were shown the video in small groups and were given a score sheet to write their comments to the puppet’s statement.

9.3 Predictions

As I said above, by adopting this mixed design, I wanted to test two hypotheses:

(10)

(a) that children performed according to a strategy and in this sense their non-adult like behaviour in previous experiment is not to be interpreted as objective “failure” in deriving SIs;
(b) that a lexical component is involved in the generation of SIs.

As far as the first argument is concerned, I predicted that the alternation of different scales within the same experimental session could enhance the probability that children activated the scalar component, blocking at the same time
a tendency to persevere which could have affected children’s performance in between-subjects designs previously adopted.\footnote{At first, I interpreted the results obtained in experiments using a between-subject design only as the effect of a tendency to persevere that children displayed in consequence of the design. Thus, the explanation of children’s performance that I initially proposed was exclusively in terms of adoption of a strategy. Given that the answer provided to the first item tested was reiterated though similar items (normally 4 or 5 in between-subject designs) and that the distribution of subjects reported in those experiments was just what one would expect from a guessing situation, I considered difficult to interpret previous results. Children may have simply guessed the answer to the first scalar statement and then maintained the same response through the whole session. As a consequence, all previous experiments supposedly were either underestimating or overestimating children’s competence and readiness in deriving SIs. (cf. Foppolo, 2005).}

As regards the second question, if children have the general ability to draw SIs (i.e., they know the maxim of quantity, have no memory limitations, can meet relevance expectations, as shown in the experimental works reported so far), but do not have a complete/stable representation of the lexical entry of a given scalar item (or the memory traces are not well established), I expect to find different behaviours on the part of the same subject depending on the scalar item tested, and thus differences in performance between children and adults on some, but not necessarily all, scales.

\section*{9.4 Results}

On control items, adults gave 100\% correct responses for all items, except for true statements containing \textit{a piece of}, for which the percentage of correct answers was 97.5\%. Children gave correct responses 100\% of the time to numerals (true) and \textit{all} (false); 97.5\% of the time to \textit{all} (true), to numerals (false) and to \textit{whole} (true and false); 92.5\% to \textit{some} (true); 80\% to \textit{a piece of} (true). Statistical analysis, by means of test Q of Cochran revealed a significant difference across items in children (Q=27.64, df=7, \(p<.001\)). Pair wise comparisons show that a reliable difference is found between \textit{a piece of} (true) and all the other control items. In particular, children’s performance on \textit{a piece of} (true) is different from \textit{all} (true), \textit{whole} (true and false), numbers (false) (\(\chi^2=6.13\), df=1, \(p<.05\)) and \textit{all} (false), number (true) (\(\chi^2=8.89\), df=1, \(p<.05\)). In fact, some children rejected true statements including \textit{a piece of} because they interpreted this item in a numerical sense. For example, in a situation in which a character had 10...
built a piece of the puzzle (and not the whole puzzle), some children said that the puppet was wrong because the character had used three pieces of the puzzle and not just one. In Italian, in fact, there’s no phonological distinction between un pezzo (lit. a piece of) in which un is used as an indefinite article, and un pezzo (lit. one piece of) in which un is used as the numeral “one”. Given that in some of the stories the pieces of the objects used could effectively be counted (like in the stories where the character managed a puzzle or a paddock), this fact could lead some children to take the expression un pezzo in its numerical sense, thus making her rejecting the otherwise true proposition. However, leaving for the moment the data on a piece of aside, we can conclude that children performed in an adult-like way on all the other control items, confirming that they have no problems in evaluating sentences containing scalar items in contexts where these are felicitous.

Moving to critical items, in which the weak scalar term on the scale was used in a situation compatible with the strongest one, this are the results obtained: children rejected critical statements 97.5% of the time for numerals, 62.5% of the time for a piece of, 70% for some in subject position and 75% for some in object position. Adults, instead, rejected critical statements between 97% and 100% for all scales. These results are reported in the table below for convenience:

<table>
<thead>
<tr>
<th></th>
<th>SOME (subj)</th>
<th>SOME (obj)</th>
<th>TWO</th>
<th>A PIECE OF</th>
</tr>
</thead>
<tbody>
<tr>
<td>ADULTS (N=40)</td>
<td>N=39</td>
<td>N=40</td>
<td>N=39</td>
<td>N=40</td>
</tr>
<tr>
<td></td>
<td>97.5%</td>
<td>100%</td>
<td>97.5%</td>
<td>100%</td>
</tr>
<tr>
<td>CHILDREN (N=40)</td>
<td>N=28</td>
<td>N=30</td>
<td>N=39</td>
<td>N=25</td>
</tr>
<tr>
<td></td>
<td>70%</td>
<td>75%</td>
<td>97.5%</td>
<td>62.5%</td>
</tr>
</tbody>
</table>

Table 1. Number of subjects that derived SIs and percentage of SI derivation for all scalar items

The graph below, instead, reports children’s distribution with respect to SI computation for the three different scalar items (data for some are considered together, not distinguishing between the two positions):
It’s important to observe that, when children rejected the underinformative statements, they always explained their rejection by invoking the stronger term in the scale, saying, for example, that *Gargamel has built a piece of the paddock* was a bad description of the story presented because *Gargamel has built the whole paddock*.

Statistical analysis, by means of test Q of Cochran, revealed a significant difference across scalar items in children (Q=17.67, df=3, p<.001); pair wise comparisons across items in children data by means of Pearson Chi-Square Test revealed that numbers differ from all the other scalar items tested, in particular: they differ from *a piece of* ($\chi^2=15.31$, df=1, p<.001), from *some* in subject position ($\chi^2=11.11$, df=1, p<.001) and from *some* in object position ($\chi^2=8.54$, df=1, p<.01). The proportion of children who generated the SI for *some* (both in object and subject position) and for numerals is significantly different than the one predicted by chance based on the binomial distribution (p<.05), while it is not so for *a piece of*.

With respect to the comparison between children and adults, this is what statistical analysis revealed. Interestingly, no difference is found between children
and adults with respect to the number scale. By contrast, differences still exist between children and adults with respect to the other items, in particular, with respect to *a piece of* \( (\chi^2=15.60, \text{ df}=1, p<.0001) \), to *some* in subject position \( (\chi^2=11.11, \text{ df}=1, p<.001) \) and to *some* in object position \( (\chi^2=9.72, \text{ df}=1, p<.01) \).

As I did in previous experiments, I considered the number of times that SI was derived overall. Given that in this experiment four different underinformative items were given, subjects had one to four possibilities to derive a SI. Crucially, in this case the items tested were always different, while in experiments using a between-subject design all the possibilities given to subjects always contained the same scalar terms. The graph below illustrates the distribution of children as a function of the number of times they rejected critical statements (i.e., derived the SI):

Overall, most of the children, i.e., 75%, computed SIs 3 or 4 times; 87.5% did so at least 2 times. The observed distribution of children is different than one predicted by chance \( (\chi^2=10.80, \text{ df}=3, p<.01) \). Crucially, statistical analysis revealed that there is a significant difference between the number of 5-year-old subjects who generated SIs 3 or 4 times in Experiment 5 and those who did so in
this experiment ($\chi^2=4.66$, df=1, p<.05). In any case, considering the proportions of subjects that generated an implicature 3 or 4 times in this experiment, a difference between children (i.e. 75%) and adults (i.e. 95%) is still found ($\chi^2=6.27$, df=1, p<.05).

9.5 Discussion

Four major results of this experiment are worth taking into consideration:

(11)

(a) children’s distribution is not bimodal anymore;
(b) children’s rejection of critical statements is much higher than in previous experiments;
(c) differences exist across the three scales in children (but not in adults);
(d) differences between children and adults remains, except for numerals.

More specifically, the change in the experimental design was very effective in eliminating the tendency to persevere and thus the ensuing bimodal distribution observed in other experiments in which a between-subject design was adopted. As we said, most of the children in this experiment (i.e. 75%) computed SIs 3 or 4 times. This is remarkable, if we compare this result with what was found in previous experiments: for example, in the first study presented in this chapter, only 58% of the 4- and 5-year-old children computed the SI associated to some at least three times. The contrast observed between these two experiments (in which analogous material was tested by adopting the same procedure) suggests that the improvement in children’s performance is an effect of the design adopted. In this connection, it is worth pointing out that in reasoning tasks involving SIs, such as the one carried out by Noveck et al. (2002), the repetition of the same problem was avoided. Quoting directly from these authors “In order to encourage spontaneous reactions, no problem is presented twice”.

194
Specifically, mixing different scalar items seems very effective overall in eliciting more rejections of critical statements, favoring in some way the retrieval of the scale. This effect can be explained under the hypotheses I am assuming, namely, that a lexical component is involved in the derivation of SIs. I call this the “Enriched Lexicon Hypotheses” (cf. Foppolo and Guasti, 2005 and Chierchia et al., 2005), according to which the lexical entry of scalar items consists of two pieces of information: the representation of the basic meaning and the scalar component. For example, (12) constitutes the lexical entry associated to *some*:

\[(12)\]

\[
\text{Representation of the basic meaning: } \ll some\rr = \lambda P \lambda Q \exists x[P(x) \land Q(x)]
\]

\[
\text{Scale: } \ll some\rr^{S-ALT} = \{ \text{some}<\text{many}<\text{most}<\text{all} \}
\]

In order to derive implicatures, it is not sufficient to activate the representation of the basic meaning; it is also necessary to have and retrieve the scale component. This entails that scalar elements need to be linked to form a scale (a sort of paradigm) that should be automatically retrieved together with the representation of the basic meaning. In other words, the scale needs to be part of the lexical entry of a scalar term, as, for example, grammatical features like gender or argument structure are associated to nouns and verbs. The process of firstly establishing which items are to be instantiated in the scale and then making their memory traces strong enough to be automatically retrieved is something that can take a while to be completed, as for any other aspect of the lexicon.

This hypothesis suggests also the possibility that the retrieval of the scale can be primed, at least for those children that have the scale already available in their lexicon but do not usually automatically retrieve it. In this respect, the fact of seeing more than one scalar item can help children to activate the scalar component besides the basic meaning that is normally activated when the scalar term is evaluated in felicitous contexts.\(^{11}\)

\(^{11}\) In this experiment, target and control statements were mixed. One might object that subjects’ improvement is due to the fact that in the same session children saw the weak and the strong...
The Enriched Lexicon Hypothesis proposed above also predicts that different scalar items may be lexicalized at different stages in the development, depending on the kind of scale and on how frequently the basic or the scalar meaning are used. This provides a ready explanation to the third remarkable result obtained in this experiment (cf. (11)c), namely that a difference among different scalar items was found in children (but not in adults). In particular, this result corroborates Papafragou and Musolino’s conclusion that numerals are treated differently from other scales. Crucially, however, children’s rejection rate in case of numbers was not different from adult’s in my experiment: here, 97.5% of children derived the SI associated to numbers, while only 65% did so in the Papafragou and Musolino’s study. This contrast provides an indication that the design adopted in the experiment under discussion is more apt to capture the ability to compute SI in children, even considering children’s performance on the other scales. If it is true, in fact, that their performance on the some<all scale was less good than the number scale, it was in any case much better than the one obtained in previous works. For example, the SI associated to some was only inferred by 12.5% of children in Papafragou and Musolino’s Experiment 1, and by 42% in my Experiment 5, while in this experiment the proportion of children that rejected underinformative statements containing some in subject position significantly rose up to 70%.

Finally, I will take into consideration the fourth result mentioned above (cf. (11)d), namely, the fact that, although an improvement is observed, I still find that children are overall less prone than adults to generate SIs. In particular we observed that, while 75% of the children generated SIs 3 or 4 times in Experiment 10, a significantly higher proportion of adults (i.e. 95%) did so. This fact suggests that there still is a group of children who is uncertain about which elements are element of the scale (e.g., some and all). I can reject this hypothesis on the basis of the results obtained in Experiment 9 reported in section 7. In that experiment, which was a manipulation of Experiment 5, for each story children were presented with two sentences, one including all and one including some to test if the hearing of the strongest scalar term would induce more pragmatic responses in children when evaluating the weakest term. In fact, no effect of manipulation was found: hearing all before some within the same story doesn’t enhance the probability that children reject underinformative some. Thus, it seems to be an effect of mixing different scales together (and supposedly, of showing number scale that has a special status, as we will discuss.
parts of scales and that do not benefit from this new experimental design. For these children the problem seems not one of simple retrieval: supposedly, they have not completed the lexical representation of the scale yet. Taking different scales into consideration and the fact that these gave rise to different rejection rates by children, one might conjecture that the scale associated to *a piece of* is the most difficult one (among the scales tested). In fact, even if we discard those children that performed badly on control items for *a piece of* (which, as we saw, was more problematic independently from SI computation), the proportion of children who still fail to generate the SI for this item remains quite low, compared to the others (about 67%). This suggests that this scalar item is related with some inherent difficulty and that a greater number of children are still unsure about which elements its scale includes.

On the contrary, the number scale seems the easiest of all. At this point, one may wonder why numbers are special in eliciting SIs more easily than other items, thus being more frequently assigned an “exactly” interpretation. For example, a sentence like *Lyn paid three coins for the pizza* is generally taken to mean *Lyn paid three coins and no more for the pizza* or equivalently that *Lyn paid exactly three coins for the pizza*. So far, I have been assuming that numbers are not different from the other scalar items, and in particular that the “exactly” interpretation is derived via SI. This assumption is challenged by, for example, Musolino (2004). In particular, he rejects this hypothesis on the ground that numbers seem to behave differently from other scalar items given that they also have an “at most” reading. However, it seems to me that there are neither strong theoretical arguments nor empirical arguments in favour of the view that numbers deserve a different semantic treatment with respect to other scalar items.

---

12 A further widespread assumption is that numbers qua lexical entries (and possibly in predicate position) have an “exact” reading. For example *three men* is a property true of groups constituted by exactly three men. However, in clauses, such properties are subject to existential closure, so that the truth conditions of, e.g., (a) are those in (b):

(a) three men walked in
(b) there is a group of exactly three men that walked in

Such truth conditions are of course compatible with, say, four men walking in. Thus, in clauses, numbers receive an “at least” interpretation. The “exact” interpretation at a clausal level is generally held to be derived via a pragmatic enrichment process (cf. a.o. Chierchia 2004 for a discussion on this).
Taking this for granted, the interesting question turns to be the following: why do children generate SIs associated to the numeral scale numbers more easily than in case of other scalar items? Or else, what is special about numbers?

Numbers are special in two respects. First, they are related to extra-linguistic scales. Dehane (1997) and Feigenson, Dehaene and Spelke (2004) propose that human (infants) and animal species are endowed with two numerical systems for representing numerosities: one system serves to represent approximate numerical quantities and the other serves to represent numerically distinct individuals. Language helps in putting these two representations together to yield a representation of sets of individuals whose cardinality increases by adding one individual to the set (Spelke & Tsivikin, 2001). Gelman and Butterworth (2005) suggest that, although language may facilitate the use of numerical concepts, a numerical scale is available independently of language. Among other things, this explains the rapid introduction of a generative counting rule by population whose language does not possess number words or only has number words up to three. Secondly, the linguistic number line is generated through a universal/parametric device (Ionin and Matushanski, 2004). Essentially, the grammar of numbers is made up of:

(13) i. a finite (and small) set of basic lexical entries (typically, adjectives or nouns, depending on the language): one, two,….ten, hundred, thousand;
ii. plus two combinatorial devices for forming complex numbers, namely:
   - a head-complement schema (with a multiplicative semantics)
   - a coordination schema (with an additive semantics)

The head-complement schema is recognizable via case assignment. In languages with rich case like Slavic languages one observes recursive assignment of case in complex numbers (where in expressions like three thousand men the
outermost numeral (*three*) assigns case to the internal numeral (*thousand*) which in turn assigns case to the noun, cf. Ionin and Matushansky, 2004). The semantics that naturally goes with this construction is multiplication: *three thousand men* is $3 \times 1000 \times$ (a unit of man). Other numbers are formed via coordination, which is overtly visible in many languages, such as for example German. Consider (14), whose representation is given in (15):

\begin{align*}
(14) & \text{zwei und zwanzig Männer} \\
& \text{two and twenty men} \\
(15) & \text{[NP [NP zwei Männer] (und) [NP zwanzig Männer]]}
\end{align*}

Here, the first occurrence of the noun *Männer* (men) is elided (i.e., marked as non pronounced) under identity with the second occurrence. As is it usually the case with coordination, structures like (14) have an additive semantics. Just like *Alex and Lorenzo* is interpreted as the set/group made up of Alex and Lorenzo, *zwei und zwanzig Männer* is the union of a group of two and a group of twenty men (which is disjoint from the first group). In this way, a simple number like *two* has the same meaning in *two men, twenty-two men, or two thousand men*. The factors that give rise to parametric variation in this schema are the lexical choices of basic entries and the specific points where a language uses addition vs. multiplication (cf., e.g. *eighty* = $8 \times 10$ vs. *quattro-vent* = $(4 \times 2 \times 10)$). This mechanism, described here in its barest outline, constitutes a powerful generative device capable of producing the linguistic number line with what appears to be wholly grammar internal means. Thus, on the one hand, the acquisition of number

---

13 In Russian, lower cardinals (2,3,4) assign paucal case and higher cardinals assign genitive case, as illustrated in (i), were four assigns paucal and thousand genitive case:

\begin{verbatim}
i) četyre tysjači šagov four thousand-Pauc steps-Gen
\end{verbatim}

14 Anecdotal evidence shows that children learn numbers through the generative schemas proposed. It is frequent to hear Italian-speaking children using a conjunction schema to generate numbers above 29, where 30 (*trenta*) becomes *venti-dieci* (lit. *twenty-ten*), 31 *venti-undici* (twenty-eleven) and so on.

15 While all languages that have complex numbers appear to exploit the same machinery, there are languages that seem to lack altogether complex numbers and only have a few lexically entries for numerals (see Dehane, 1997).
words, thus of their syntax and semantics, are supported by the availability of a number line with an inherent scalar character independent of language. On the other, two generative mechanisms are available that are responsible for the generation of a linguistic number line that matches the pre-linguistic number line. There is nothing analogue to this for the other scales. For example, there doesn’t seem to be a generative mechanism for deriving the scale associated to quantifiers, nor there is a pre-linguistic scale associated to quantifiers. Given the considerations just sketched above, no wonder that the lexical scalar character of numbers falls more readily into place.

10. Conclusive remarks on SI computation in children

In this chapter, I presented a series of experiments that investigated children’s ability to generate one kind of pragmatic inferences, i.e., scalar implicatures. As a start, I explored the emergence of this ability by means of a developmental study in which different age groups were tested using the same method and material. Then, I investigated which piece of knowledge that is considered necessary for the derivation of SIs children actually master. Finally, I examined the role played by a change in the experimental design in priming the activation of the scale associated to different scalar items, assuming that a lexical component is involved in SI computation.

Based on data from Italian, I showed that there is a developmental effect in the generation of the SI associated to some, with 6- and 7-year-olds behaving as adults and generating SIs much of the time and 4- and 5-year-olds being more equivocal, in spite of the fact that they have quite sophisticated linguistic knowledge of the items involved, know how to meet relevance expectations and know that utterances differ in terms of quantity of information. If it is true overall that 5-year-old children as a group are less prone than adults to generate SIs, examining the performance of individual subjects I observed a bimodal distribution, with one group always deriving SIs and one never doing so. I conjecture that this result, which characterizes my first experiment as well as other
experiments discussed in the literature, is in fact an artefact of the experimental design adopted, in which children are presented with a set of four or five statements including the same scalar item. I conjectured that previous studies actually underestimated children’s ability to derive SIs. In the mean time, I predicted that they would improve their performance if encouraged to do so. Specifically, I predicted that the use of a within-subject design (in which the same subject was shown different types of scalar items) would have primed subjects to derive more implicatures than in previous studies. In a sense, I expected the specific design to contribute to activate the scalar component associated to the critical terms. This conjecture was developed on the basis of two considerations:

(16)

(a) children’s bimodal distribution in previous experiments reflected a subjective choice (in Chierchia’s term, this reflects the choice in the selection of the feature associated to the scalar term) and was not to be interpreted as a failure due to limitations in children’s cognitive system (cf. discussion in 9.5);

(b) the ability of deriving SIs depends on the fact that the scalar term is correctly and fully lexicalised, i.e. both the basic meaning and the scalar component are represented in the lexicon (as attested by the Enriched Lexicon Hp, cf. 9.5).

The results obtained seem to conform to my predictions. Specifically:

(17)

(a) children overall improved, showing that they are in fact able to derive implicatures (contrary to previous findings);

(b) previous distribution is to be interpreted as a subjective choice in feature selection;
children’s performance depends on the type of scalar item, ultimately on the fact that the particular scalar item is fully lexicalised (variation is predicted - and found - among different items).

To conclude, we can say that some developmental factor is involved in the generation of SIs. But this is not, as previously thought, anything linked to children’s cognitive system (such as memory limitations). More simply, what it is developing seems to be the lexical (scalar) component associated to each scalar term. In this respect, knowing for example when to use *some* and *all* means knowing their basic meaning. Deriving the SI associated to *some*, however, requires at least two further steps. In the first place, the scalar component (i.e., the set of scalar alternative) must be lexicalized. In the second place, the access to it must get automatized. And of course it takes time to carry out these steps. How much also depends on the nature of the scalar term.

Under this view, children may fail to derive SIs because they are still unsure about the eventuality that a given element belongs to a scale or not. This may happen when the memory trace of the scalar component associated to a particular item is not yet in place or it is not stable enough to permit an automatic retrieval of the scale.
Appendix A

EXPERIMENT 1 (TVJT) – Chapter 2

List of critical statements containing *or*

Ogni bambino che ha il panino o il biscotto si è seduto  
*Every boy that has a sandwich or a cookie sat down*

Ogni signora ha messo il cappello o il cappotto  
*Every lady put on a hat or a coat*

Nessuna scimmia che ha la banana o il biscotto fa il bagno  
*No monkey that has the banana or the cookie takes a bath*

Nessun pagliaccio porta la torta o la Coca-Cola  
*No clown brings a cake or a Coke*

Una bambina che ha preso il gelato o la fragola e’ andata in barca  
*A girl that took the ice-cream or the strawberry got on a boat*

Una bambina si è portata a casa il cane o il gatto  
*A girl took a dog or a cat home*
EXPERIMENT 2 (E-PRIME) – Chapter 2

List of critical statements containing or

Se un Blimpo ha un curpo o un dorf, ha anche una matita
*If a Glimp has a curp or a dorf, he also has a pencil*  
DE

Se un Plosco ha un samp o un fert, ha anche una chiave
*If a Plosc has a samp or a fert, he also has a key*  
DE

Se un Flonto ha un vilpo o un murl, ha anche una pianta
*If a Flont has a vilp or a murl, he also has a plant*  
DE

Se un Blimpo ha una macchinina, ha anche un perpo o un dorf
*If a Glimp has a toy-car, he also has a perp or a dorf*  
NON-DE

Se un Plosco ha una conchiglia, ha anche un fert o un gurf
*If a Plosc has a shell, he also has a fert or a gurf*  
NON-DE
Appendix C

EXPERIMENT 3A (E-PRIME) – Chapter 4

List of critical donkey sentences

Ogni Blimpo che ha un curpo lo conserva nel barattolo
Every Glimp that has a curp preserves it in a jar

Ogni Plosco che ha un sampo lo protegge con una foglia
Every Plosc that has a samp protects it with a leaf

Ogni Flonto che ha un vilpo lo tiene dentro un secchio
Every Flont that has a vilp keeps it in a bin

Nessun Blimpo che ha un perpo lo protegge con una foglia
No Glimp that has a perp protects it with a leaf

Nessun Plosco che ha un gurfo lo tiene dentro il secchio
No Plosc that has a gurf keeps it in a bin

Nessun Flonto che ha un murllo lo conserva nel barattolo
No Flont that has a murl preserves it in a jar

Qualche Blimpo che ha un dorfo lo tiene dentro un secchio
Some Glimp that has a dorf keeps it in a bin

Qualche Plosco che ha un ferto lo conserva nel barattolo
Some Plosc that has a fert preserves it in a jar

Qualche Flonto che ha un tarplo lo protegge con una foglia
Some Flont that has a tarp protects it with a leaf
EXPERIMENT 3B (E-PRIME) – Chapter 4

List of biasing context (only English version) + critical donkey sentences

In this planet, many Plosc own samps, objects extremely sensitive to the sunlight. For this reason, they always have to be protected under a leaf when the sun shines. Thus, in sunny days...

Ogni Plosco che ha un sampo lo protegge con una foglia
Every Plosc that has a samp protects it with a leaf

On planet Flont there’s the plague of the termites. To face it, Flonts use special traps, the vilps. Fighting termites causes a lot of stress to the Flonts. To improve the situation, they consulted a famous magician who suggests them a remedy: they have to sacrifice a vilp putting it into a bin full of water; a vilp in the water doesn’t function as a trap anymore, but releases an anti-stress substance. Let’s see if Flonts are less stressed given that...

Ogni Flonto che ha un vilpo lo tiene dentro un secchio
Every Flont that has a vilp keeps it in a bin

On planet Plosc it’s easy to get lost and gurfs are fundamental instruments for navigation. Usually, gurfs are kept into bins full of water to recharge. But no one ever goes out leaving all his gurfs at home into bins. Let’s see if, when they go out...

Nessun Plosco che ha un gurfo lo tiene dentro il secchio
No Plosc that has a gurf keeps it in a bin

The radio news at 11.00 urged the population to put all their gurfs into bins made of lead because if they’re kept outside they are radioactive. Let’s see if, after the bulletin...

Qualche Blimpo che ha un dorfo lo tiene dentro un secchio
Some Glimp that has a dorf keeps it in a bin

On planet Flont, someone reports that if you keep tarps under a leaf for seven days, they become aphrodisiac. It seems that some Flont still believes in this rumour...

Qualche Flonto che ha un tarpo lo protegge con una foglia
Some Flont that has a tarp protects it with a leaf
Appendix E

EXPERIMENT 4 (Questionnaire) – Chapter 4

List of critical items with every + donkey anaphora

Ogni Blimpo che ha un curpo lo conserva nel barattolo
(a) Ogni Blimpo che possiede dei curpi conserva tutti i curpi che ha nel barattolo
(b) Ogni Blimpo che possiede dei curpi ne conserva uno di quelli che ha nel barattolo

Every Glimp that has some curps preserves all the curps he has in a jar
(a) Every Glimp that has some curps preserves all the curps he has in a jar
(b) Every Glimp that has some curps preserves one of the curps he has in a jar

Ogni Flonto che ha un vilpo lo tiene dentro un secchio
(a) Ogni Flonto che possiede dei vilpi tiene tutti i vilpi che ha dentro un secchio
(b) Ogni Flonto che possiede dei vilpi ne tiene uno di quelli che ha dentro un secchio

Every Flont that has some vilps keeps all the vilps he has in a bin
(a) Every Flont that has some vilps keeps all the vilps he has in a bin
(b) Every Flont that has some vilps keeps one of the vilps he has in a bin

Ogni Plosco che ha un sampo lo protegge con una foglia
(a) Ogni Plosco che possiede dei sampi protegge tutti i sampi che ha con una foglia
(b) Ogni Plosco che possiede dei sampi ne protegge uno di quelli che ha con una foglia

Every Plosc that has some samps protects all the samps he has with a leaf
(a) Every Plosc that has some samps protects all the samps he has with a leaf
(b) Every Plosc that has some samps protects one of the samps he has with a leaf
Appendix F

EXPERIMENT 5 (TVJT) – Chapter 5

List of critical statements containing *qualche*

Qualche puffo va in barca  
*Some Smurf is going on a boat*

Qualche scimmia mangia il biscotto  
*Some monkey is eating a biscuit*

Qualche bambina guarda la televisione  
*Some girl is watching TV*

Qualche pagliaccio pesca  
*Some clown is fishing*

Qualche soldato va a cavallo  
*Some soldier is riding a horse*
Appendix G

EXPERIMENT 6 (TVJT) – Chapter 5

List of critical statements containing *alcuni dei*

*Alcuni dei puffi vanno in barca*
*Some of the Smurfs are going on a boat*

*Alcune delle scimmie mangiano il biscotto*
*Some of the monkeys are eating a biscuit*

*Alcune delle bambine guardano la televisione*
*Some of the girls are watching TV*

*Alcuni dei pagliacci pescano*
*Some of the clowns are fishing*

*Alcuni dei soldati vanno a cavallo*
*Some soldiers are riding the horse*
Appendix H

EXPERIMENT 7 (TVJT) – Chapter 5

List of statements containing all

Tutti gli scoiattoli mangiano il gelato
All chipmunks are eating ice-cream

Tutti i puffi vanno in barca
All the Smurfs are going on a boat

Tutti i cani dormono
All dogs are sleeping

Il pagliaccio nasconde tutti i biscotti
The clown is hiding all the biscuits

Paperino lava tutte le macchine
Duffy Duck is washing all the cars

Tutti i bambini dormono
All the boys are sleeping

La mamma lava tutti i bambini
Mum is washing all the kids

Sandokan cattura tutti i puffi
Sandokan is catching all the Smurfs

List of statements containing some

Ninja lava alcune delle pecore
Ninja is washing some of the sheep

La bambina lava alcuni dei piatti
The girl is washing some of the dishes

Alcuni dei bambini suonano la chitarra
Some kids are playing guitar

Alcuni degli indiani comprano un panino
Some Indian guys are buying a sandwiches
Appendix I

EXPERIMENT 8 (FJT) – Chapter 5

<table>
<thead>
<tr>
<th>Puppet 1 (RACOON ALBERTO)</th>
<th>Puppet 2 (SEAL SILVANA)</th>
</tr>
</thead>
</table>
| Tutti gli scoiattolini fanno la doccia  
All chipmunks are taking a shower | Qualche scoiattolino si fa la doccia  
Some chipmunk is taking a shower |
| Tutti gli scoiattolini si svegliano  
All chipmunks are waking up | Qualche scoiattolino si sveglia  
Some chipmunk is waking up |
| Tutti gli scoiattolini vanno sull’altalena  
All chipmunks are playing on the swing | Qualche scoiattolino va sull’altalena  
Some chipmunk is playing on the swing |
| Qualche scoiattolino va in autobus con Pippo  
Some chipmunk is going on the bus with Goofy | Tutti gli scoiattolini vanno in autobus con Pippo  
All chipmunks are going on the bus with Goofy |
| Qualche scoiattolino suona la chitarra  
Some chipmunk is playing the guitar | Tutti gli scoiattolini suonano la chitarra  
All chipmunks are playing the guitar |
Appendix L

EXPERIMENT 9 (TVJT) – Chapter 5

List of critical statements containing tutti and qualche

Tutti i nanetti mangiano
All the dwarfs are eating

Qualche puffo va in barca
Some Smurf is going on a boat

Tutte le pecore bevono
All the sheeps are drinking

Qualche scimmia mangia il biscotto
Some monkey is eating a cookie

Tutti i puffi giocano con le macchinine
All the Smurfs are playing with toy-cars

Qualche bambina guarda la televisione
Some girl is watching TV

Tutti i signori vanno in piscina
All the sirs are swimming

Qualche pagliaccio pesca
Some clown is fishing

Tutti i bambini vanno in treno
All the kids are going by train

Qualche soldato va a cavallo
Some soldier is riding a horse
Appendix M

EXPERIMENT 10 (TVJT) – Chapter 5

<table>
<thead>
<tr>
<th>LIST 1</th>
<th>LIST 2</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Warm-ups</strong></td>
<td><strong>Warm-ups</strong></td>
</tr>
<tr>
<td>Tutti i soldati sono andati a cavallo</td>
<td>Tutte le scimmie hanno mangiato il biscotto</td>
</tr>
<tr>
<td><em>All the soldiers rode a horse</em></td>
<td><em>All the monkeys ate a cookie</em></td>
</tr>
<tr>
<td>Batman ha comprato due pere</td>
<td>Il gufo ha acceso due candele</td>
</tr>
<tr>
<td><em>Batman bought two pears</em></td>
<td><em>The owl lighted two candles</em></td>
</tr>
<tr>
<td><strong>Critical (underinformative) statements</strong></td>
<td><strong>Critical (underinformative) statements</strong></td>
</tr>
<tr>
<td>Cenerentola ha decorato un pezzo dell’albero</td>
<td>Pinocchio ha fatto un pezzo del puzzle</td>
</tr>
<tr>
<td><em>Cinderella decorated a piece of the tree</em></td>
<td><em>Pinocchio made a piece of the puzzle</em></td>
</tr>
<tr>
<td>Alcuni dei puffi sono andati in barca</td>
<td>Alcuni dei pagliacci sono andati a pescare</td>
</tr>
<tr>
<td><em>Some of the Smurfs went by boat</em></td>
<td><em>Some of the clowns went fishing</em></td>
</tr>
<tr>
<td>Due scimmie hanno mangiato il biscotto</td>
<td>Due puffi sono andati in barca</td>
</tr>
<tr>
<td><em>Two monkeys ate a cookie</em></td>
<td><em>Two Smurfs went by boat</em></td>
</tr>
<tr>
<td>Lo gnomo ha raccolto alcune delle carote</td>
<td>L’indiano ha lavato alcune delle pecore</td>
</tr>
<tr>
<td><em>The dwarf picked up some of the carrots</em></td>
<td><em>The Indian washed some of the sheep</em></td>
</tr>
<tr>
<td><strong>Control false items</strong></td>
<td><strong>Control false items</strong></td>
</tr>
<tr>
<td>Baloo ha chiuso tutti i bidoni</td>
<td>Lo gnomo ha raccolto tutte le carote</td>
</tr>
<tr>
<td><em>Baloo closed all the bins</em></td>
<td><em>The dwarf picked up all the carrots</em></td>
</tr>
<tr>
<td>Gargamella a costruito tutto il recinto</td>
<td>Mogwli ha colorato tutta la stella</td>
</tr>
<tr>
<td><em>Gargamel build the whole paddock</em></td>
<td><em>Mowgli colored the whole star</em></td>
</tr>
<tr>
<td><strong>Control true items</strong></td>
<td><strong>Control true items</strong></td>
</tr>
<tr>
<td>Pinocchio ha fatto un pezzo del puzzle</td>
<td>Gargamella ha costruito un pezzo del recinto</td>
</tr>
<tr>
<td><em>Pinocchio made a piece of the puzzle</em></td>
<td><em>Gargamel build a piece of the paddock</em></td>
</tr>
<tr>
<td>Mowgli ha colorato tutta la stella</td>
<td>Cenerentola ha decorato tutto l’albero</td>
</tr>
<tr>
<td><em>Mowgli colored the whole star</em></td>
<td><em>Cinderella decorated the whole tree</em></td>
</tr>
<tr>
<td>Il gufo ha acceso due delle candele</td>
<td>Baloo ha chiuso alcuni dei bidoni</td>
</tr>
<tr>
<td><em>The owl lighted up all the candles</em></td>
<td><em>Baloo closed some of the bins</em></td>
</tr>
<tr>
<td>L’indiano ha lavato due pecore</td>
<td>Batman ha comprato due pere</td>
</tr>
<tr>
<td><em>The Indian washed two sheep</em></td>
<td><em>Batman bought two pears</em></td>
</tr>
</tbody>
</table>
### Warm-ups

<table>
<thead>
<tr>
<th>List 3</th>
<th>List 4</th>
</tr>
</thead>
<tbody>
<tr>
<td>Tutti i puffi sono andati in barca</td>
<td>Tutti i pagliacci sono andati a pescare</td>
</tr>
<tr>
<td><em>All the Smurfs went by boat</em></td>
<td><em>All the clowns went fishing</em></td>
</tr>
<tr>
<td>Baloo ha chiuso due bidoni</td>
<td>Lo gnomi ha raccolto due carote</td>
</tr>
<tr>
<td><em>Baloo closed two bins</em></td>
<td><em>The dwarf picked up two carrots</em></td>
</tr>
</tbody>
</table>

### Critical (underinformative) statements

<table>
<thead>
<tr>
<th>List 3</th>
<th>List 4</th>
</tr>
</thead>
<tbody>
<tr>
<td>Gargamela ha costruito un pezzo del recinto</td>
<td>Mowgli ha colorato un pezzo della stella</td>
</tr>
<tr>
<td><em>Gargamel build a piece of the paddock</em></td>
<td><em>Mowgli colored a piece of the star</em></td>
</tr>
<tr>
<td>Alcuni dei soldati sono andati a cavallo</td>
<td>Alcune delle scimmie hanno mangiato il</td>
</tr>
<tr>
<td><em>Some of the soldiers rode a horse</em></td>
<td><em>biscotto</em></td>
</tr>
<tr>
<td>Due pagliacci sono andati a pescare</td>
<td>Due soldati sono andati a cavallo</td>
</tr>
<tr>
<td><em>Two clowns went fishing</em></td>
<td><em>Two soldiers rode a horse</em></td>
</tr>
<tr>
<td>Batman ha comprato alcune delle pere</td>
<td>Il gufo ha acceso alcune delle candele</td>
</tr>
<tr>
<td><em>Batman bought some of the pears</em></td>
<td><em>The dwarf lighted up some of the candles</em></td>
</tr>
</tbody>
</table>

### Control false items

<table>
<thead>
<tr>
<th>List 3</th>
<th>List 4</th>
</tr>
</thead>
<tbody>
<tr>
<td>L’indiano ha lavato tutte le pecore</td>
<td>Batman ha comprato tutte le pere</td>
</tr>
<tr>
<td><em>The Indian washed all the sheep</em></td>
<td><em>Batman bought all the pears</em></td>
</tr>
<tr>
<td>Cenerentola ha decorato tutto l’albero</td>
<td>Pinocchio ha fatto tutto il puzzle</td>
</tr>
<tr>
<td><em>Cinderella decorated the whole tree</em></td>
<td><em>Pinocchio made the whole puzzle</em></td>
</tr>
</tbody>
</table>

### Control true items

<table>
<thead>
<tr>
<th>List 3</th>
<th>List 4</th>
</tr>
</thead>
<tbody>
<tr>
<td>Mogwli ha colorato un pezzo della stella</td>
<td>Cenerentola ha decorato un pezzo dell’albero</td>
</tr>
<tr>
<td><em>Mowgli colored a piece of the star</em></td>
<td><em>Cinderella decorated a piece of the tree</em></td>
</tr>
<tr>
<td>Pinocchio ha fatto tutto il puzzle</td>
<td>Gargamela ha costruito tutto il recinto</td>
</tr>
<tr>
<td><em>Pinocchio made the whole puzzle</em></td>
<td><em>Gargamel built the whole paddock</em></td>
</tr>
<tr>
<td>Lo gnomi ha raccolto alcune delle carote</td>
<td>L’indiano ha lavato alcune delle pecore</td>
</tr>
<tr>
<td><em>The dwarf picked up some of the carrots</em></td>
<td><em>The Indian washed some of the sheep</em></td>
</tr>
<tr>
<td>Il gufo ha acceso due candele</td>
<td>Baloo ha chiuso due bidoni</td>
</tr>
<tr>
<td><em>The owl lighted up two candles</em></td>
<td><em>Baloo closed two bins</em></td>
</tr>
</tbody>
</table>
References


References


Fox, D. and D. Pesetsky (2004). *Cyclic linearization and typology of movement.* Handout, MIT.


Ladusaw, W. (1979). Negative Polarity Items as Inherent Scope Relations, University of Texas at Austin.


