Experimental Investigations into how Children and Adults Process the Implicature Associated with the Scalar Term Some

SCRAFTON, SUSAN

Experimental Investigations into how Children and Adults Process the Implicature Associated with the Scalar Term *Some*

Susan Scrafton

One Volume

Submitted for the degree Doctor of Philosophy

Durham University, Department of Psychology, 2009

Table of Contents

List of Tables

List of Figures

This thesis is dedicated to

Frank Robinson

&

Grant F. Milne

two of nature's gentlemen

Declaration

None of the data or material contained in this thesis has been submitted for previous or simultaneous consideration for a degree in this or any other country.

Part of the work in this thesis has been published as:

Scrafton, S. & Feeney, A. (2006). Dual processes, development and scalar implicatures. *Proceedings of the 28th Annual Conference of the Cognitive Science Society*, 774-779.

This research was funded by a Durham University Wolfson Research Institute Studentship and a Durham University Postgraduate Teaching Fellowship.

Acknowledgements

This thesis represents not only my work as a PhD researcher but the end of my journey as a student, a journey that began many years ago when I returned to education as a mature student. To all those friends and family who have encouraged me every step of the way I would like to say a big thank you, your support was very much appreciated. There were many twists and turns, not to mention the occasional dead end, but I got there in the end, and yes I now have a proper job!

Particular thanks in relation to this thesis go to:

Aidan Feeney, my supervisor, without whom there would be no thesis. Thank you for your belief and encouragement, and giving me the opportunity to achieve something I would never have thought possible. Your patience over the years was very much appreciated. The parsnips finally got buttered!

All my participants, without whom there would be no data and therefore no thesis. Special thanks to: Mr Welsh, the staff and children of Nunthorpe Primary School; Mrs Porteous, the staff and children of Roseworth Primary School; Mr Howse, the staff and children of Caldicotes Primary School; Mr McFarlane, the staff and children of Badger Hill Primary School; Mr Walke, Ms Robinson and the children of Thornhill Park School; Mr Rees and the children of Westlands School; Mr Rochester, Miss McDonnell and the children of King James 1st Community Arts College; Mr Wallace and the children of Framwellgate School; Mrs Orridge, Mrs Riddel and the children of High Clarence Primary School; and St. Mary’s Church Youth Group. You were all stars.

Charlotte and Linda, and Claire and Catherine, for giving up a weekend to model for the storyboards. May your fame spread far and wide.

Sue Leekam, for advice on carrying out research on children with Autism Spectrum Disorders, and Greig Coull and Lucy Carey for sharing data and insights.

Keira, Adam, Dan, Ali, Bronia, Ian and Pim, who now know more about scalar implicature than any sane person should. Thank you for being sounding boards, giving stats advice, having sociable lunchtimes and keeping me going. A big thank you also to Dave for all the IT help.

Steve and Catherine and my new colleagues in the Foundation Centre for being flexible and giving me time to finish my thesis.

Holly and Beth, my daughters, for being such great kids (and now adults). Follow your dreams.

Paul, my wonderful husband, without whose support the journey would never have begun in the first place. Thank you for everything. Love you lots!

Abstract

A scalar implicature is the use of a weak term from a scale to implicate that a stronger term in the scale is not the intended meaning. For example, *some* is often interpreted as meaning *some but not all*, whereas logically its meaning is at least one. It is only in recent years that scalar implicature has progressed from its role as an explanation for poor reasoning performance in adults to its current status as the subject of experimental investigation. As a result, relatively little is known about scalar implicature and the literature contains seemingly contradictory findings and untested assumptions. The primary aim of this thesis was to investigate the quantifier *some* in order to clarify and extend our existing knowledge of the scalar implicature associated with the term. In Chapter 1 the literature is reviewed and three research questions are identified in relation to the implicature: What is the developmental trajectory of sensitivity to the implicature? What contexts facilitate sensitivity to the implicature? And which contemporary theory best captures the processing of the scalar term? The experiments in Chapter 2 primarily examined how sensitivity to scalar implicature develops. The results revealed that contrary to assumptions in the literature, sensitivity does not develop linearly but in a U-shaped fashion. Consequently children can be more pragmatic than adults. In addition, sensitivity was seen in 3-year-old children, which is earlier than has previously been shown in the literature. Chapter 3 explored the role of context in facilitating sensitivity to scalar implicatures in children. It focused on Feeney, Scrafton, Duckworth and Handley's (2004) claim that deception contexts help children to detect implicatures. The findings revealed that deception contexts can aid sensitivity but an important factor is the motivation behind the deception attempt. Thus, the highest rates of sensitivity to the implicature were observed in conditions where there was an

obvious benefit to the speaker if her attempted deception was successful. Chapter 4 compared how *some* is processed by children and adults, and the experiments revealed that adults may be subject to processing difficulties that the children appear not to be. Sensitivity in adults was affected by a secondary task, their logical response times on *infelicitous some* trials were longer than on control trials and they appeared to experience difficulties in resolving response conflicts. Contrary to assumptions in the literature, a logical response in adults is not necessarily indicative of failure to detect an implicature but could represent either cancellation of the implicature or the detection of conflict between two possible interpretations. In all the experimental chapters, the findings are discussed in relation to theoretical accounts of scalar implicature, with the conclusion that no current theory in its present form fully captures the nature of the phenomenon. Overall, it is concluded that there are assumptions in the literature which appear to be incorrect and therefore future research must be mindful of using untested assumptions to interpret new results. In addition it is argued that theoretical explanations of scalar implicature need to be revised and should be used as a way of generating hypotheses rather than the tool by which results are interpreted. *Some* is interesting not only for what it can tell us about the interface between semantics and pragmatics but also for what it reveals about the relationship between inference and reasoning.

Chapter 1

Literature Review

1.1. Introduction

Human communication is a sophisticated process and is often seen as that which sets us apart from other animals. As Sperber (1995) has pointed out, our exchanges about hopes, fears and knowledge are far more complicated than the rudimentary codes of bees and monkeys. For us, verbal communication is a lot more than simple sound to meaning pairings, it typically involves inference. Inference involves going from the grammatical content of an utterance to its pragmatic content; in other words from what is literally said to what is actually meant. Thus, understanding an utterance involves understanding what the speaker intended to communicate. There are many occasions in conversation where inferences have to be made in order to understand the speaker's intended meaning, such as the interpretation of irony or metaphor. However this thesis focuses on one particular form of conversational inference, that of scalar inference, and more specifically the scalar implicature associated with that inference.

A scalar inference involves a listener evaluating a speaker's utterance against a set of informationally ordered alternatives, for example a scale, in order to draw a conclusion that was not specifically stated. A scalar implicature, on the other hand, is the speaker's use of a weak term from that set to convey the information that a stronger term in the scale is not the intended meaning (Bach, 2006; Horn, 1972, 2004). Implicatures are associated with a variety of scales such as connectives, where the set is <*and, or*>, modals <*must, might be*>, numerals <…*three, two one*>, and quantifiers <*all, some*>, the latter of which is the primary focus of this thesis.

The implicature associated with *some*, and also *or*, was recognised as far back as 1867 by Mill (as cited in Horn, 2004) and much has been written about it, especially within the field of linguistics where people have sought to identify the nature of the implicature (e.g. Geurts, 2009; Hirschberg, 1991; Recanti, 2003; Russell, 2006; Saurland, 2004; Spector, 2007). The presence of scalar implicatures has also been acknowledged within the reasoning field with regards to the ability of people to reason according to the laws of logic (Begg & Harris, 1982; Polizter, 1986; Wilkins, 1928). However, the field of experimental pragmatics, in which scalar implicatures are the direct focus of study, has only really emerged as a discipline this century (Noveck & Sperber, 2004). Consequently experimental investigations of scalar implicatures are in their infancy, and much of our understanding of the nature of scalar implicatures comes from a somewhat limited empirical data base. In addition, comparisons between studies have been difficult as they have been conducted in different languages, required different responses, used different tasks, and tested different ages, or a number of variables have been manipulated within the same study. This has led to two problems; there are seemingly contradictory findings for a number of research questions, and there is a reliance on untested assumptions to interpret new results, assumptions which may prove to be incorrect.

The primary concern of this thesis is the scalar implicature associated with the quantifier *some*, and the thesis's aim is to examine experimentally three areas where there are contradictory findings or untested assumptions in order to clarify our existing knowledge about the scalar implicature, and thus to extend that knowledge. Before the specific aims of each experiment and the layout of the rest of the thesis are presented, this chapter begins with an appraisal of the phenomena of scalar implicature. This appraisal starts with an overview of scalar implicature, where the definition of the term is clarified with examples. The overview continues with the charting of the origins of the term scalar implicature from Grice's (1989) seminal work on conversational inference, to the work of Horn (1972), who coined the term scalar implicature to reflect the notion that these implicatures rely on a quantificational scale for their derivation. Following this overview, two opposing contemporary accounts of scalar implicature processing are described; the default account and the relevance account. The appraisal ends with the presentation and evaluation of the current empirical findings concerned with the scalar implicature associated with *some*.

1.2. Overview of Scalar Implicature

As we have already seen, a scalar implicature is the use of a weak term from a scale to implicate that a stronger term in the scale is not the intended meaning. For example consider the following utterances:

(1) Some of Paul's children are girls

(2) Bring a savoury dish or a pudding

Logically speaking the scalar terms used in the utterances, *some* and *or*, are compatible with the stronger terms in their respective scales; *some* is defined as meaning more than one so it does not preclude *all*, whilst *or* is considered to be true if at least one of the disjuncts is true, and so does not preclude both disjuncts being true (Copi & Cohen, 1994). Therefore fellow interlocutors would be entitled to interpret the utterances as:

(3) Some, maybe all, of Paul's children are girls

(4) Bring a savoury dish or a pudding, maybe both

However uttering (3) or (4) when you know that the stronger term in each set is the case, whilst not logically fallacious, could be considered to be pragmatically infelicitous. So the utterances are often interpreted as meaning:

(5) Some, but not all, of Paul's children are girls

(6) Bring a savoury dish or a pudding, but not both

In each case the listener has inferred that the speakers' use of a weak term implicates that the stronger term is not the case. If *all* and *and* were the intended meanings the speakers would have used those terms. As *all* and *and* have not been used, *not all* and *not both* must be the case.

1.2.1. Grice

Paul Grice (1989) was the first person to systematically explain how inferences in conversation are derived, and coined the term conversational implicature to describe this process. For Grice, communication is made up of two elements; what is said and what is implicated. Although he did not deal with scalar implicatures particularly, one of his aims was to reconcile the differences in meaning between formal logical expressions, such as ∃ and V, and their counterparts in natural language, such as *some* and *or*.

According to Grice, conversation is rarely a series of disconnected remarks, there is generally a sense of cooperation on the part of each participant, with each recognising a common purpose or mutually accepted direction. He developed a Cooperative Principle, which required participants to:

> 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 in which you are engaged. (Grice, 1989, p. 26)

Grice (1989) also proposed that discourse is conducted according to a number of maxims, or rules, which govern expectations of quantity, quality, relation and manner. The purpose of spelling out these maxims was to show that a speaker can convey information indirectly by saying something which, on the face of it, may appear irrelevant, but given the nature of the discourse, is highly meaningful. In order to preserve the cooperativeness of the statement the listener adds additional information to interpret the sentence according to what the speaker actually meant rather than what the speaker literally said.

The implicatures in (5) and (6) arise out of the maxim of quantity.

Maxim of Quantity

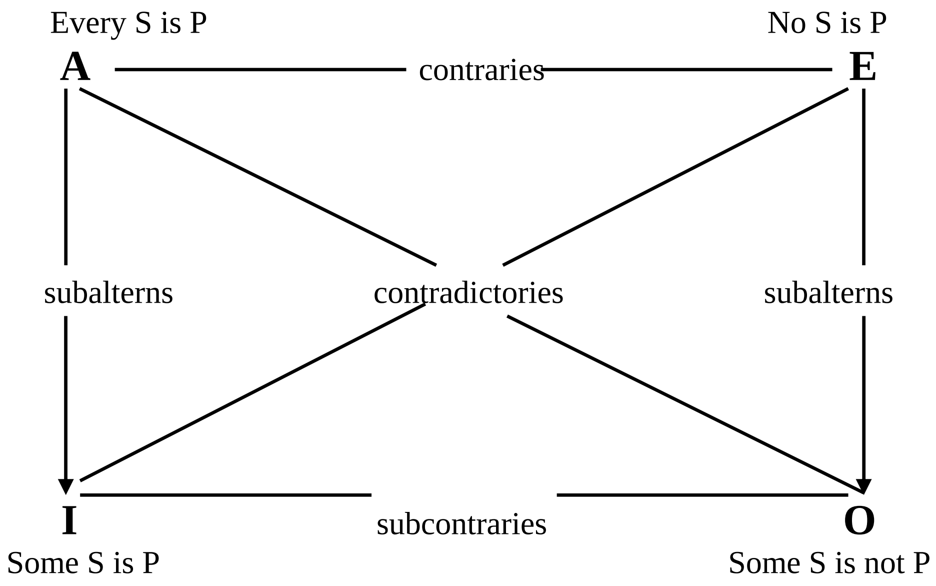
> (i) Make your contribution as informative as is required (for the current purposes of the exchange).
>
> (ii) Do not make your contribution more informative than is required. (Grice, 1989, p. 26)

On hearing (1) s*ome of Paul's children are girls*, or (2) *bring a savoury dish or a pudding*, the listener assumes that the speaker is adhering to the Cooperative Principle and observing the maxim of quantity. Therefore they infer that as the speaker chose to use the quantifier *some* instead of *all*, and the connective *or* instead of *and*, the more informative terms must not be the case. So the sentences are interpreted as (4) *some, but not all, of Paul's children are girls*, and (5) *bring a savoury or a pudding, but not both*. Thus the literal meaning of the sentence is computed, found wanting, and then added to; it is this additional information, the conversational implicature associated with the original utterance and the inference drawn from it, that preserves the cooperative nature of the discourse.

There is one further point relating to Grice's theory which is worth noting, as it links to the contemporary pragmatic theories which are discussed in section 1.3 of this chapter. According to Grice (1989) there is a distinction to be made between particularised and generalised implicatures. Particularised implicatures are those implicatures that are made only on a particular occasion by virtue of the context in which the utterance occurs, whereas generalised implicatures are those associated with certain words regardless of the context in which the utterance occurs. Unfortunately Grice offers only one example to make this distinction; the indefinite article (p.37). The indefinite article *a* normally carries the assumption that the speaker is not in a position to be more specific, so the utterance *X is meeting a woman this evening* carries the generalised implicature that the woman is not his wife/girlfriend/daughter. As we will see in section 1.3, the debate over whether the quantifier *some* normally carries the implicature *not all* regardless of context is still going on today.

1.2.2. Horn

The implicatures in (5) and (6) came to be known as a scalar implicatures through the work of Larry Horn (1972). He suggested that terms such as *all* and *some* and *and* and *or* lie on a scale of informativeness where the initial term in the scale is the more informative term. This informativeness arises from one-way entailment relations; strong terms entail weaker terms but not vice-versa. Thus strong terms are more informative because if you are in a situation where you know a strong term to be true you also know that the weak term is true. Implicatures arise from these entailment relationships and the logical relationships in the traditional square of opposition (see Figure 1.1; also Horn 1972, p.207 and Parsons, 2008).



The four corners of the square represent four logical forms:

A is a universal affirmative of the form every S is P

E is a universal negative of the form no S is P

I is a particular affirmative of the form some S is P

O is a particular negative of the form some S is not P

Corresponding A and E statements are contrary: they can both be false at the same time but they cannot both be true.

Corresponding A and O, and E and I statements are contradictory: members of each pair cannot be true or false simultaneously.

An I statement is unilaterally entailed by an A statement; the same applies to O of E.

Corresponding I and O statements are subcontrary: they can be simultaneously true but they cannot be simultaneously false.

Figure 1.1. Traditional Square of Opposition

Horn (1972, 2004) points out that the assertion of either of the two subcontraries, I or O, quantity implicates the other. Therefore, whilst *some* is entailed by *all* and so is logically compatible with it, it implicates *not all*. Therefore although the I statement, *some of Paul's children are girls*, is logically distinct from the O statement, *some of Paul's children are not girls*, what is communicated by both statements is essentially the same, *some of Paul's children are girls and some aren't.* So, scalar terms are lower-bounded by their semantic meaning and upper-bounded by a quantity implicature. This applies not just to quantifiers and connectives but generalises to all operators in entailment scales, such as modals *<necessarily, possibly>*, *<must, may>*, adverbs *<always, often, sometimes>*, adjectives *<hot, warm>* and numerals *<three, two, one>*.[1]

1.3. Contemporary Theories

A number of contemporary theories of pragmatics have been applied to the phenomena of scalar implicatures in order to explain their derivation. In experimental pragmatics these theories tend to be divided into two main camps; default (Chiercha, 2004: Gazdar, 1979; Horn, 1972, 2004, 2009; Levinson, 1983, 2000) and relevance based accounts (Carston, 2002; Sperber & Wilson 1985/1995).

1.3.1. Default theories

One of the main proponents of the notion of scalar implicatures as default is Levinson (1983, 2000). Levinson's account of implicatures develops from, and extends, the Gricean framework of communication. He proposes that instead of thinking about Grice's maxims as behavioural norms or rules, they should instead be

[1] Although in later work, Horn (2009) acknowledges that numerals may be a special case.

viewed as “inferential heuristics” (2000, p.35). He identified three heuristics; I, M and Q.

The I heuristic can be thought of as standing for informativeness and relates to Grice’s second maxim of quantity, “do not make your contribution more informative than necessary” (Grice, 1989, p. 26). The relevant heuristic is “what is expressed simply is stereotypically exemplified” (Levinson, 2000, p. 37). In other words, simple expressions are interpreted according to what is most typical. For example, the conditional *if* is interpreted as the biconditional *iff*, as in the statement *if, and only if, you mow the lawn I will give you 5 pounds.*

The M heuristic relates to Grice’s maxim of manner, “be perspicuous” (Grice, 1989, p. 27). The relevant heuristic is “what’s said in an abnormal way isn’t normal” (Levinson, 2000, p. 38). This heuristic is complimentary and opposite to the I heuristic, in that if a statement is marked in some way then the stereotypical interpretation should be avoided. Using Levinson’s example, the simple expression *Bill stopped the car* carries the I implicature that he did so in the typical fashion, by using the foot brake, whereas *Bill caused the car to stop* carries the M implicature that the car was stopped in an unusual way, such as by the application of the handbrake.

The Q heuristic relates to Grice’s first maxim of quantity, “make your contribution as informative as is required” (1989, p. 26). This heuristic, considered to be the most important and the one responsible for scalar implicatures, is “what isn’t said, isn’t” (Levinson, 2000, p.35). In other words, reference is made to what might have been said but it is taken that it is not the case. For example, the statement *some boys came to the party* Q-implicates that *not all boys came* through the heuristic that as *all* was not said it cannot be so.

Although accepting Grice's two levels of communication, the "said" and the "implicated" (Grice, 1989, p. 118), Levinson (1983, 2000) proposes that there is a third layer, what he terms "utterance- type-meaning" (2000, p. 22). This additional layer represents Generalised Conversational Implicatures (GCI's), and comes before, and is distinguished from, the other class of conversational implicatures, the Particularised Conversational Implicatures (PCI's). The meaning of a GCI is not based on the semantics of the utterance or, as is the case for PCI's, on the context-driven speaker's intentions, but on a general expectation of how language is normally used. Levinson illustrates the difference between GCI's and PCI's with the following example (2000, p. 16-17):

> Consider the sentence *some of the guests are already leaving,* and consider that it might be uttered in two rather different contexts;
>
> *Context 1*
>
> A: "What time is it?"
>
> B: "Some of the guests are already leaving"
>
> *PCI*: 'It must be late.'
>
> *GCI*: 'Not all the guests are already leaving.'
>
> *Context 2*
>
> A: "Where's John?"
>
> B: "Some of the guests are already leaving."
>
> *PCI*: 'Perhaps John has already left.'
>
> *GCI*: 'Not all the guests are already leaving'

As can be seen from the example, the different contexts produce two different PCI's, but there is also a GCI which is common to both. This GCI arises from the general expectations of how language is normally used, given the structure of the language, and it is these expectations that give rise to default inferences.

So, according to Levinson (1983, 2000) scalar implicatures are part of a whole group of GCI's, which are so readily inferred that they are the preferred or default interpretation. These implicatures automatically go through unless they are cancelled. Thus the mere presence of a weak scalar term will result in a scalar inference regardless of context. Therefore any statement of the form *some a's are b's* will, all other things being equal, have the default interpretation *not all a's are b's*. According to Levinson, the ubiquitous nature of scalar implicatures in everyday conversation can only be accounted for by this default interpretation.

Although the experimental findings of this thesis are compared with predictions made by Levinson's default account, there is another default account that has received attention within the literature, Chierchia's (2004) Sematic Core Model (see also, Chierchia, Crain, Guasti, Gualmini & Meroni, 2001; Chierchia et al. 2004; Gualmini, Crain, Meroni, Chierchia, & Guasti, 2001; Guasti, Chierchia, Crain, Foppolo, Gualmini, & Meroni, 2005). Given its relevance to language processing an overview of the model will now be given. Although Chierchia is in agreement with Levinson regarding the general interpretation of the term *some,* his mechanism underpinning the default status is very different. He proposes that the semantic and pragmatic processing of a sentence occur alongside one another. That is, the literal meaning and pragmatic meaning are assigned phrase by phrase with implicatures being factored in recursively as soon as they occur. This local processing contrasts sharply with the standard neo-Gricean view where scalar implicatures are processed globally, that is pragmatic enrichment takes place after semantic processing has assigned a literal meaning to the whole sentence. According to Chierchia, a different model is needed from that of the standard neo-Gricean account to explain the presence of embedded implicatures and the absence of scalar implicatures in certain linguistic contexts. It should be noted that there is evidence that negation blocks scalar implicatures (see Horn, 1989; also Gualmini, 2004; Musolino, Crain & Thornton, 2000; Musolino & Lidz, 2006) but this line of enquiry is beyond the scope of this thesis.

Chierchia (2004) suggests that embedded implicatures are a problem for the traditional neo-Gricean account. He argues that as scalar implicatures place statements against a background of alternates, and the choice of statement from those alternates implicates the negation of all informationally stronger alternates, embedded implicatures should not exist because implicatures come about from the negation of the whole sentence. However, if a whole sentence is negated with an embedded scalar term then contradiction can occur. For example, consider the following statement taken from Chierchia et al. (2004, p. 286):

(7) Mary is either reading a paper or seeing some students.

The negated stronger alternate is

(8) It is not the case that (Mary is either reading a paper or seeing every student)

which entails

(9) Mary is not reading her paper

which contradicts the original statement in (7). The only way for contradiction not to occur, Chierchia suggests, is if implicatures are dealt with phrase by phrase. In that way the negation applies only to the second disjunct, the phrase in which the scalar term occurs.

Chierchia (2004) proposes that the absence of scalar implicatures in downward entailing (DE) contexts, that is contexts that license inferences from super sets to their subsets, also necessitates a model which has interacting semantic and pragmatic modules. According to Chierchia, there is a relationship between contexts where scalar implicatures fail to arise and those linguistic contexts which license the negative polarity item *any*; contexts that are DE. Both the addition of scalar implicatures and the licensing of *any* are driven by a need to be more informative than the relevant alternates allow. But informativeness is relevant to the polarity of the context. When a scalar term is encountered in a DE context, its implicature is not more informative than the literal meaning and so it is automatically removed. Consider the following examples taken from Chierchia et al. (2004, p. 288):

(10) Paul invited John and Bill

Situation 1 = Paul invited John and Bill

(11) Paul invited John or Bill

Situation 1 = Paul invited John and Bill

Situation 2 = Paul invited John

Situation 3 = Paul invited Bill

Examples (10) and (11) are non-DE contexts. The implicature arises in (11) because this statement is less informative than (10); (10) is true for only one situation whereas (11) is true for three situations. However in DE contexts, (12) and (13), the situation is reversed and the statement containing *or* is more informative than the one involving *and*, thus there is no reason to compute the implicature associated with *or*.

(12) Paul didn’t invite John or Bill

Situation 1 = Paul invited neither John nor Bill

(13) Paul didn’t invite John and Bill

Situation 1 = Paul invited neither John nor Bill

Situation 2 = Paul invited John

Situation 3 = Paul invited Bill

Therefore a mechanism is needed which allows embedded scalar implicatures to be computed but removes them in DE contexts. DE contexts are structurally determined, whilst implicatures are pragmatically determined, therefore a model is needed which is sensitive to both modules.

As can be seen from the above accounts, Chierchia (2004) and Levinson (1983, 2000) agree on the default status of the term *some*, but disagree on the mechanism by which the implicature is processed. Although this difference is interesting, the experiments to be described in this thesis were not designed to arbitrate between the accounts. As Levinson's default account has received most attention in the literature, it will be the one focused on in this thesis.

1.3.2. Relevance theories

Standing in opposition to default theories are relevance based accounts, the best known of which, Sperber and Wilson's Relevance Theory (RT) (Sperber & Wilson, 1985/1995; see also Noveck & Sperber, 2007; Sperber, 1994; Sperber & Wilson, 2002; Wilson & Sperber, 2002a, 2004), will be described here.

RT shares some assumptions with Gricean and Neo-Gricean theories, namely that the whole point of human communication is the expression and recognition of intentions, and that listeners are guided towards speakers' meanings by the expectations that the speakers' utterances create. However RT substantially departs from these theories on at least three points; points that have implications for the processing of scalar implicatures. The first departure concerns Grice's (1989) cooperative principle or maxims; according to Sperber and Wilson (1985/1995) there is no need for these as communication can be fully explained in terms of relevance. Secondly, comprehension does not involve global processing; it is an online process whereby the explicit and implicit meanings of an utterance are developed in parallel. Following on from this, the last point concerns literal interpretation. According to RT, unlike Grice, arriving at the most relevant interpretation of an utterance does not mean that the literal interpretation has to be accessed. These points will be expanded on and illustrated below.

According to RT, human cognition, in its desire for increasing efficiency, has evolved to detect stimuli that are most relevant to us. That is stimuli whose processing makes a worthwhile difference to our view of the world. Linguistic utterances are particularly relevant because they carry an assumption of their own relevance. This is because in conversing with someone we are laying claim to their attention, and in doing so are overtly signalling our intention to provide relevant information. How relevant an input is for a listener is a function of the effect gained from its processing and the effort required to carry out that processing. All other things being equal, the greater the positive cognitive effects the greater the relevance, but the greater the processing effort expended the lower the relevance. It is these expectations of relevance, not Grice's (1989) maxims, which guide a listener to a speaker's meaning.

With regards to the processing of utterances, RT proposes that an utterance can only be fully understood in conjunction with the context in which it occurs. So unlike default theory, scalar terms do not have an automatic interpretation. RT proposes that the goal of a listener is to construct a hypothesis about the speaker's meaning which satisfies the expectation of relevance raised by the utterance. The listener achieves this by constructing appropriate hypotheses about the explicit content, the intended contextual assumptions, and the intended contextual implications. However these steps are not carried out sequentially, but in parallel, against a backdrop of assumptions and expectations of how relevant the utterance will be for the listener. Additionally, revision can occur as the utterance unfolds. So, during the comprehension process the listener takes the linguistically encoded sentence meaning and follows the path of least effort to compute cognitive effects and stops when conditions of relevance are met. However, people can differ in terms of their interpretation strategy. Sperber (1994) proposes that the level of understanding a listener has about a speaker's intention towards them regarding benevolence and competence determines which interpretation the listener settles on. Sperber identifies three different levels, which have in common the path of least effort; a naïve optimist assumes benevolence and competence and stops at the first relevant interpretation, a cautious optimist assumes only benevolence and stops at the first relevant interpretation that he considers the speaker would have thought relevant to him, whereas a sophisticated communicator, who assumes neither benevolence or competence, stops at the first relevant interpretation that he considers that the speaker would have thought *seemed* relevant to him.

With regards to literal interpretations versus pragmatic interpretations, RT proposes that the literal interpretation of a scalar term need not be accessed if it is not relevant. Consider the utterance "John is a soldier" in the following example taken from Sperber and Wilson (2002, p. 20).

Peter: Can we trust John to do as we tell him and defend the interests of the Linguistic Department in the University Council?

Mary: John is a soldier!

Whilst Peter might have a mental representation of the concept of a soldier, not all the qualities and characteristics attributed to that concept will be equally activated. Interpretations are arrived at in order of accessibility and the degree to which they satisfy his expectations of relevance. So Peter might access representations relating to duty and following orders before settling on the interpretation *John identifies with the goals of the team*. Once this interpretation is reached other interpretations are not even considered, especially the literal interpretation, *John is a member of the military*. Whilst the military interpretation may be literally true it is not relevant to Peter's understanding of Mary's utterance and therefore it is not computed. If Peter's question had been about John's profession then the literal interpretation of Mary's reply would have computed because in that case it would have been relevant. In fact Wilson and Sperber (2002b) go so far as to argue that language use is not governed by truthfulness. Thus a term may be broadened out or narrowed down from its literal interpretation to suit the purposes of the interaction.

So according to RT scalar terms are always interpreted in context with consideration being given to the speaker's intentions. Therefore the scalar term *some* can be interpreted as *may be all*, *not all* or even *just one*, depending on what is relevant to the hearer (Noveck & Sperber, 2007). This interpretation in turn is dependent on the size of the cognitive effect achieved in relation to the amount of effort required. Even though it may appear that *some* has more than one meaning, only one will be available at one time because forcing the listener to choose between interpretations would cause unnecessary effort and would go against the principle of optimal relevance. Therefore it is not a case of processing the literal meaning of a term, finding it wanting and therefore enriching it to achieve the intended meaning, listeners access the most relevant interpretation and settle on it. If the context does not prime the enriched interpretation of a scalar term, then RT predicts that a literal interpretation should take less time and effort to arrive at than an enriched one, because an enriched interpretation always involves the processing of meaning construction (Noveck & Sperber, 2007).

So it can be seen from the description of the theoretical accounts above, that default theories and relevance based accounts make very different predictions regarding the processing of the scalar term *some*. These differences are illustrated in Table 1.1, taken from Noveck and Sperber (2007).

Table 1.1

Contrasting predictions of Generalised Conversational Implicature Theory and Relevance Theory regarding the speed of processing of the scalar term some according to Noveck and Sperber (2007)

Interpretation of *Some*	Generalised Conversational Implicature Theory	Relevance Theory
Literal	Default enrichment + context sensitive cancellation	No enrichment
	hence slower	**hence faster**
Enriched	Default enrichment	Context-sensitive enrichment
	hence faster	**hence slower**

In section 1.4.3. of this chapter we will take a closer look at how the predictions regarding processing times have been tested.

1.4. Empirical Studies of the Scalar Implicature Associated with *Some*

A "cottage industry" (Noveck, 2004, p. 305) has sprung up in recent years intent on furthering our knowledge of the scalar implicature associated with the quantifier *some*, and it has undoubtedly had a fruitful start; new methodologies have been devised, existing ones normally associated with other disciplines have been utilised, and theories have been experimentally tested. But whilst our understanding of the phenomena has increased, key questions still remain unanswered. This literature review will be organised around three such questions. It will consider why the questions remain unanswered and explain why it is important that they are answered. In order to do this, the literature review will be primarily concerned with the implicature associated with *some*, although as scalar implicatures are regarded as universal in nature, experimental evidence will also be considered from other scalar terms. In addition, whilst most of the work surveyed is drawn from the area of experimental pragmatics, consideration will also be given to other areas of study where the presence of scalar implicatures has also been documented.

Experimental investigations involving the scalar implicature associated with *some* have tended to pursue three lines of enquiry; establishing when sensitivity to the implicature emerges, establishing in what contexts sensitivity is seen, and adjudicating between default and relevance based accounts of implicature processing. These three questions form the structure of the review.

1.4.1. What is the developmental trajectory of sensitivity to the scalar implicature associated with some?

Within the field of experimental pragmatics there has been a debate as to whether children are insensitive to scalar implicatures following Noveck's (2001) claim that weak scalar terms "tend to be treated logically by young competent participants and more pragmatically by older ones" (p. 165). Adults' ability to detect scalar implicatures are generally never questioned (Guasti et al, 2005; Hurewitz, Papafragou, Gleitman, & Gelman, 2006; Noveck, 2001; Papafragou & Musolino, 2003), and the implicit assumption within the literature has been that as age increases logical interpretations of scalar terms decrease, so children's logical responding should decrease linearly until it reaches adult levels of pragmatic responding. Consequently even if children show sensitivity to scalar implicatures it should never be at a rate greater than that of adults (Guasti et al., 2005; Noveck, 2004; Noveck & Reboul, 2008; Noveck & Sperber; 2007; Papafragou & Musolino, 2003). However given that no study has investigated a wide range of children's ages this assumption may be ill founded.

Knowing the developmental trajectory of sensitivity to the scalar implicature associated with *some*, at what age it appears, how it develops and whether the development is linear, is necessary for a number of reasons. Firstly, if meaningful comparisons are to be made between age groups, the assumption of linearity is not enough, it needs to be established. Secondly, knowing who shows sensitivity will also help with understanding how scalar implicatures are processed and what cognitive demands they place on an individual. Finally, knowing the trajectory will also help resolve the issue which is often discussed alongside this topic, namely what factors facilitate the detection of scalar implicatures. Before considering the evidence from studies of scalar implicature let us first turn to the reasoning literature, which has long acknowledged the presence of scalar implicatures.

The idea that adults reliably interpret *some* as meaning *not all* has been documented in the adult reasoning literature since the turn of the last century (Wilkins, 1928) and is often the reason advanced as to why they do so badly on certain reasoning tasks (Begg & Harris, 1982; Ceraso & Provitera, 1971; Johnson-Laird, 1975; Neimark & Chapman 1975; Newstead, 1989; Politzer, 1986). According to Evans, Newstead and Byrne (1993) “there is ample evidence that people interpret quantifiers according to conversational implicatures rather than logic” (p. 221).

Begg and Harris (1982) suggest that adults use the rules of conversation to interpret *some* rather than the rules of formal logic. In the case of syllogistic reasoning the authors propose that adults treat the syllogism as an obscure attempt at conversation rather than an isolated logical proposition, and because they are more adept at communicating than they are at reasoning, they choose to reason according to conversational principles. Begg and Harris asked adults to interpret categorical propositions of the type used in syllogisms, by choosing appropriate set-theoretical relations. The authors found that *some* and *some not* were given similar interpretations, both being interpreted to mean *some but not all*, even though participants had been instructed that *some* did not preclude *all*.

Begg (1987), following on from his previous work (Begg & Harris 1982), investigated why people interpret *some* as meaning *not all*. He hypothesised that *some* serves a communicative function of restricting attention to the minority rather than the majority. He gave adult participants information about the occupations of fifty men and fifty women. Regardless of the occupation, such as writer or artist, there were equal numbers of both male and female representatives; however the occupations were not equally represented, there were more writers than artists. The results revealed that in relation to the phrase *some men are ...*: the phrase was considered to be more misleading, as determined by assigned numerical grades, when the occupation that followed was writers, as opposed to artists; the term artists was chosen above writers on forced choice statements; and artist was chosen over writer for tasks that required completion of the sentence. Begg concluded that there was a strong preference for those statements that directed attention to the smaller rather than larger subset of those named individuals (for other work that has looked at the focusing properties of quantifiers see the large corpus of work by Moxey & Standford [e.g. Moxey & Sanford, 1993a, 1993b, 2000; Paterson, Sanford, Moxey & Dawydiak, 1998; Sanford & Moxey, 2003, 2004: Sanford, Moxey & Paterson, 1996]).

Bara, Bucciarelli, and Johnson-Laird (1995) also found that adults favoured a *not all* interpretation of *some* when asked to choose a set that represented the quantifier. In addition, when asked to say what conclusion followed from a syllogism, the adults' answers revealed that the majority did not consider that *some* could mean *possibly all*. For example, when given a syllogism of the form *some A are B, some B are C* and asked what conclusion followed, the majority of adults, instead of replying that there was no valid conclusion, answered *some A are C*, suggesting that they had not considered the logical interpretation of *some* which would allow the conclusion *all A are C.*

However there is evidence that this issue may not be as clear cut as first thought. In a follow up on previous work (Newstead, 1989), Newstead (1995; Roberts, Newstead & Griggs, 2001) challenges the idea that interpretational errors manifest themselves in syllogistic reasoning. He acknowledges that people make interpretational errors when faced with quantifiers on simple interpretational tasks but feels that "there is very little empirical evidence to support the claim that Gricean errors of this kind are a major determinant of syllogistic reasoning performance" (Newstead, 1995, p. 645). He substantiates this claim by reviewing previously published data in relation to whether Gricean interpretations of quantifiers can account for performance in syllogistic reasoning. He extended his review to carry out a series of experiments to determine when Gricean errors occur. The tasks he set for participants involved the interpretation of quantifiers, and both the construction and evaluation of syllogisms. He found that in the simple interpretational tasks Gricean errors were common, however they became less common as the logical demands of the task increased, and with regards to syllogisms they appeared to be virtually non-existent. Newstead interpreted these results as indicating that an increase in task complexity leads participants to consider alternatives. So it would appear that whilst there is ample evidence that adults in reasoning tasks can and do interpret *some* to mean *not all* it is by no means certain that they always do so. It appears that for adults the cognitive demands of the reasoning task can affect the interpretation of *some* that is given.

The results in the reasoning literature regarding children's interpretation of *some* are also not clear cut, and there is evidence for both pragmatic and logical interpretations. Badzinski, Cantor and Hoffner (1989) examined 4- to 11-year-old children's comprehension of relational quantifiers. The participants heard sentences relating to the activities of a group of children, such as *one day after school, some of the children went to the park.* Seven cardboard figures of boys or girls and a board that was divided in two with tape were placed in front of the participants. The participants' task was to place on one side of the board the number of figures they thought had participated in the activity, and on the other side the number of figures who had not. The mean number of figures chosen as having participated in the activity was 3.7 across the age groups, and by three of the four age groups. In addition, no child from any of the age groups assigned all seven figures when the quantifier used in the sentence was *some*.

Similarly, Neimark and Chapman (1975), investigating the interpretation that students from 7th grade through to college (age 13 years and upwards) favoured for the quantifiers *all* and *some*, found that *all* and *no* were interpreted logically whereas *some* and *some not* were interpreted pragmatically. Using a mimeographed task, participants had to consider set relations that were described in classical propositional form, such as *some A are B*, and then encircle all the appropriate alternatives, which were depicted in Venn diagram form. The results revealed that as age increased so did the frequency with which *some* was interpreted according to logical convention.

However in a contradictory finding, Bara et al. (1995) found that whilst adults did not consider the *possibly all* meaning of *some* (as was mentioned earlier) children appeared to do so. Children aged 9 and 10 years when asked what conclusion followed from *some A are B*, *some B are C* overwhelmingly replied *all A are C*, which is compatible with interpreting *some* to mean *possibly all*.

With regards to the question of what the developmental trajectory of sensitivity to the scalar implicature associated with *some* is, the assumption that adults routinely make pragmatic interpretations of *some* seems to be quite entrenched in the reasoning field, even though, as we have seen, the empirical evidence for certain reasoning tasks does not always support this view. As for children's interpretations the evidence is mixed. There is support for Noveck's view that *some* is first treated logically before pragmatic interpretations emerge (Bara et al., 1995) but there is also evidence of children being pragmatic in their interpretations (Badzinski et al., 1989) and perhaps more importantly of children being more pragmatic than adults (Neimark & Chapman, 1975). However, it must be remembered that the focus of these studies, and the other reasoning studies reviewed, was not the scalar implicature associated with *some* (the possible exceptions being Newstead [1995] and Roberts et al. [2001]), and so it is not necessarily the case that these results will generalise to situations were the scalar term is manipulated. To examine the evidence from such cases we turn to the experimental pragmatics literature.

As was mentioned previously, it was Noveck's (2001) claim that logical interpretations are supplanted by pragmatic interpretations that sparked the debate over whether children are insensitive to scalar implicatures. In support of this claim Noveck, using a sentence evaluation task, found that adults rejected infelicitous statements, such as *some elephants have trunks*, 57% of the time whereas 8- and 10-year-old children did so only 11% and 15% of the time respectively. The task required participants to say whether they agreed or disagreed with a set of statements. The statements were based on three types of information; factually universal information that is best described by *all*, such as cats have ears, factually existential information that is best described by *some*, such as dogs have spots, and absurd information where neither quantifier would be suitable, such as garages sing. Each statement was preceded by *some* or *all*, which resulted in six subgroups; *true all, false all, absurd all, felicitous some, infelicitous some* and *absurd some*. The statements that were not infelicitous acted as controls to ensure that the participants had a good understanding of the quantifiers. The children's ability to deal with *some* statements that were not pragmatically infelicitous, such as *some children are blonde* and *some fish are made of leaves*, ruled out the possibility that the children were randomly choosing one interpretation over another. Instead, Noveck suggested that although the children were linguistically competent, pragmatic interpretations of scalar terms were generally not yet available to them at these ages.

Noveck's (2001) claim was strengthened with the additional finding that 7- and 9-year-old children treated the modal term *might* logically. The children and a group of adults were shown three boxes; one was open and contained a toy parrot and a toy bear, one was open and contained a parrot, and one was closed. The participants were told that the closed box had the same contents as one of the open boxes. When faced with the statement *there might be a parrot in the box*, the children overwhelmingly chose to accept the statement 80% and 69% of the time respectively, as opposed to the adults who chose to accept it only 35% of the time.

Although Noveck (2001) pointed out that he wished to avoid making the claim "that the general capacity to infer implicatures is linked to a particular chronological age" (p. 184), as his was the initial study, the ages of the child participants and the method used have acted as benchmarks against which others have tested.

The statement evaluation task has also been used with 7-year-old children in both English (Feeney, Scrafton, Duckworth & Handley, 2004) and Italian (Guasti et al., 2005). The children in these studies also favoured a logical interpretation of *some*, with 66% and 87% logical response rates respectively, in tasks using statements based on those used by Noveck (2001). The children's near ceiling responses on the control statements suggested that they generally had a good understanding of the quantifiers *all* and *some*, therefore their logical responses to the *infelicitous some* statements implied that they had failed to detect the implicature. However the adults' logical response rates in these studies were also quite high, 65% and 50%, and it is difficult to imagine that the adults were unable to detect the implicature. This would suggest that even if adults have the ability to detect the implicature it can at times go undetected. So by the same token, maybe children have the pragmatic ability to detect the implicature, it is just not being demonstrated with this methodology.

On the other hand, studies involving a change in methodology have also found that children appear to be insensitive to the scalar implicature associated with *some*. Pouscoulous, Noveck, Politzer, and Bastide (2007) found that 9-year-old French children agreed with a puppet's statement *some turtles are in the boxes*, 91% of the time, when all the turtles were in boxes, compared with a 53% response rate in adults. Similarly, in a study involving toy horses jumping over a log, Papafragou and Musolino (2003) found that 5-year-old Greek children thought that a puppet had answered well when she said *some of the horses jumped over the log* when all the horses had. In addition these children also accepted the use of the term *start* in a task when *finish* would have been more appropriate, such as *the little girl started making the puzzle*, when she had been seen to finish it. It has been suggested that the failure to elicit a pragmatic interpretation of *some* may be due to the meta-linguistic judgement that is required in many of the studies (Pouscoulous, et al., 2007), however Huang and Snedeker (2005) found that 5-year-old children failed to calculate scalar implicatures in a study using an eye tracking paradigm to examine the on-line comprehension of scalar terms. In a visual world task children were asked to point to a character with *some* of a set of objects, where there were two possible referent points; a girl with a half share of one set of objects (e.g. socks) and a girl with a whole set of different objects (e.g. soccer balls). The reference points were semantically ambiguous for the onset of the noun (e.g. soc...) and the children only looked at the target after it had been lexically disambiguated, suggesting that their interpretation of *some* was compatible with *all*. And in a task involving the use of stickers to signify the interpretation of scalar terms, Hurewitz et al. (2006) report that 3- and 4-year-old children were no better than chance at choosing a picture to represent the statement *the alligator took some of the cookies.* Children were just as likely to choose a picture showing the alligator with 4/4 cookies as they were to choose the picture were he had 2/4 cookies. Interestingly though, the same children were well above chance in choosing an exact interpretation of the number term *two*.

Given these findings the conclusion might be drawn that in general children between 3- and 10-years-of-age do not reliably detect the implicature associated with *some*. Furthermore there is evidence of *some* being interpreted logically from work looking at how children determine which noun phrase is quantified. Smith (1980), whose work formed the basis for Noveck's methodology, found that children aged between 4 and 7 years answered affirmatively when faced with an infelicitous question, such as *do some birds have wings?* And with regards to other scalar terms there is evidence that children appear insensitive to the implicature associated with *or* (Braine & Rumain, 1981; Paris, 1973), and additional evidence for the scalar term *might* (Noveck, Ho & Sera, 1996).

However what is not always commented on is the pattern of responding by individual participants. Although not reported in the paper, the children in the Feeney et al. (2004) study were not equivocal in their responses. Thus they tended to be consistently pragmatic or consistently logical. Therefore although the pragmatic interpretations could have involved some noise, in that the children who favoured a logical response occasionally gave a pragmatic interpretation, 7 of the 24 children consistently gave a pragmatic interpretation, accounting for 28% of the data. Hence over a quarter of the children were in fact showing sensitivity to the scalar implicature. Papafragou and Musolino (2003) and Pouscoulous et al. (2007) do not report the distribution of the participants' scores, but 10% of the children in the Guasti et al. (2005) study and 13% of the children in Noveck (2001) study also consistently gave a pragmatic response. More surprisingly, 9 out of the 24 children (3- and 4-years-of-age) in the Hurewitz et al. (2006) study chose the partial set two out of three times.

Similarly, if the nature of the distributions of the adult responses is examined then it reveals that adult responses to scalar terms are bi-modally distributed. Whilst Noveck (2001) stated that his adults were equivocal, with logical response rates of 41%, the distribution reveals that participants tended to be consistent in their interpretations; two thirds of his participants tended to favour a pragmatic interpretation whilst the other third always gave a logical response. Guasti et al. (2005) also found a bi-modal distribution, with the same number of adults consistently accepting the infelicitous statements as rejecting them. Therefore there is evidence of adults not giving pragmatic interpretations and of children being able to detect scalar implicatures.

There are also studies whose methodology has made them more successful in eliciting sensitivity to scalar implicatures in children. Guasti et al. (2005) gave 7-year-old Italian children a Truth Value Judgement Task (TVJT) where they had to decide whether a puppet's statement was a good or bad description of a scenario that had been acted out in front of them. They found that on 75% of occasions the children rejected the puppet's infelicitous statement, *some soldiers are riding a horse*, when all the soldiers were seen riding horses. In addition, 11 of the 15 participants consistently rejected the statement suggesting that they had detected the scalar implicature. It is also worth noting that the children's response rate was not significantly different from the adults' 83% logical response rate. High sensitivity rates to scalar implicatures were also seen in children in a study carried out by Pouscoulous et al. (2007). These authors used action-based tasks involving boxes and tokens with 4-, 5- and 7-year-olds to investigate scalar implicature detection. They found that when the researcher made the request *I would like some boxes to contain a token*, when *all* the boxes already contained a token, the youngest participants removed a token from at least one of the boxes 68% of the time, whilst the oldest children did so 83% of the time. This high rate of pragmatic responding was very close to the 86% rate for the adults.

As can be seen from the studies described above, the age of acquisition of sensitivity to scalar implicatures seems to vary with 4-year-old children being sensitive in one study (Pouscoulous et al., 2007) and 10-year-old children being insensitive in another (Noveck, 2001). One study that sheds some light on these seemingly contradictory findings is that of Feeney et al. (2004). They gave 7-year-old children a statement evaluation task based on that used by Noveck (2001) and a storyboard task. The storyboard task depicted, in photographs, a girl interacting with a set of objects. The children's task was to decide whether the girl gave a truthful response when she was asked by her mother what she had done to the objects. When the children heard the girl say she had interacted with *some* of the objects when she had been seen to interact with *all*, for example *I've eaten some of the biscuits* when she had eaten *all*, they rejected her response 79% of the time. However in the statement evaluation task the children rejected the *infelicitous some* statements only 43% of the time. This would suggest that sensitivity to scalar implicatures may not be a case of all or nothing; once sensitivity begins to emerge, it may not always be seen.

So the present picture regarding the developmental trajectory of sensitivity to scalar implicatures in children appears to be linear, in that there is no evidence from the experimental pragmatics literature of children being more pragmatic than adults. However it must be remembered that the age groups investigated in the studies have tended to have a large gap between the oldest child participants and the adults. Also, no study has included adolescents and so it remains to be seen how sensitive this age group is to scalar implicatures. It is difficult to say at what age sensitivity develops. It does appear that once sensitivity is acquired it is not always seen, as there are also adults who give logical interpretations. Whether logical interpretations in adults represent a failure to detect the scalar implicature or the cancellation of the implicature is still debateable. It also appears that sensitivity to scalar implicatures, at least for children, is context dependent (Feeney et al., 2004). The factors that facilitate this detection are the subject of the next part of the review.

1.4.2. What contexts facilitate sensitivity to the implicature associated with some?

The role of context in the detection of scalar implicatures has received very little research attention, and the small number of studies that have been carried out have tended to have a different focus depending on whether the participants are adults or children. Within the adult age group the focus has tended to be on identifying contexts in which the scalar implicature is endorsed less often. Two contexts where this appears to be the case are ones which threaten the face of the listener, in other words the self-image a person has (Bonnefon, Feeney & Villejoubert, 2009), and those whereby it is enough to know that *some* is the case and knowing *all* adds no additional information, known as lower-bound contexts (Breheny, Katsos & Williams, 2006). With regards to children the focus has been on identifying contexts that facilitate their detection of scalar implicatures, and as shown in section 1.4.1. factors which are often implicated include shared conversational background (Guasti et al., 2005), naturalistic context (Guasti et al., 2005) and deception (Feeney et al., 2004). However, unlike the previously mentioned studies concerning adults detection of implicatures (Bonnefon et al. 2009; Breheny et al., 2006), it is still not fully understood which contexts aid facilitation in children and how they do so. The experimental infancy of the topic and the tendency for researchers to manipulate a number of variables in a single study, have not only made it difficult to assess the relative contribution of each contextual factor, they have led to claims, for example, that a shared conversational background facilitates sensitivity (Guasti et al., 2005) when sensitivity is not always seen in studies that have provided such a background (Hurewitz et al., 2006; Papafragou & Musolino, 2003; Pouscoulous et al., 2007).

The finding that children have difficulty detecting scalar implicatures with the statement evaluation task appears robust, with children in a number of studies giving a *maybe all* interpretation for *some* (Noveck, 2001; Feeney et al. 2004; Guasti et al., 2005). In addition, Feeney et al.'s (2004) finding that there was a difference in children's response rates to *infelicitous some* between the statement evaluation and storyboard tasks suggests that it is the statement task itself that hinders the detection of scalar implicatures in children, rather than the non-availability of scalar inferences per se.

Guasti et al. (2005) suggest that the source of difficulty, in relation to the detection of scalar implicatures in the statement evaluation task, may lie in the conversational background; the statement task relies on encyclopaedic knowledge that may not be common to both parties. According to these authors the crucial factors for enhancing the detection of scalar implicatures are "the availability of the relevant evidence and naturalness of the situation" (p.692). Guasti and colleagues found that 7-year-old children had high rates of pragmatic responding when it involved a TVJT. This methodology typically involves two experimenters, one acting out short stories using props and the other playing the role of a puppet sitting watching alongside the participants. At the end of the story the puppet recounts what they think has taken place within the story (Crain & Thornton, 1998). Guasti et al. modified the task so that the participants watched a video where the experimenter acted out stories using props whilst holding a puppet. At the end of each story the puppet recounted what had happened in the story, and the child's task was to say whether it was a good or bad description of what had happened. For example, one story featured five soldiers who had to go and collect a treasure. The soldiers had a discussion about the merits of going by motorbike or on horseback, motorbikes were considered to be fast but expensive, and the decision was made that they should all go on horseback. Thus the scenario made the evidence needed to evaluate the under-informative statement immediately available to the participants. This claim for the facilitatory role of a naturalistic context is supported by the work of Papafragou and Tantalou (2004), who propose that conversational contexts allow children to formulate felicity expectations and to derive scalar implicatures when the speaker's contribution results in infelicity. However, with regards to a shared conversational background, insensitivity has also persisted in studies that have utilised this factor. As was seen in the previous section, the children in the studies of Papafragou and Musolino (2003) and Pouscoulous et al. (2007) had low rates of pragmatic responding. According to Hurewitz et al. (2006) it is not the visual scene per se but the nature of the discourse that accompanies the scene which determines whether or not implicatures are detected. Therefore whilst a shared conversational background may be necessary for the detection of scalar implicatures in children, it is not sufficient.

Children's difficulties with scalar implicature detection can also be framed in terms of effort. According to Pouscoulous et al. (2007) children have fewer cognitive resources than adults, thus children fail to recognise infelicity in the statement evaluation task and tasks involving naturalistic contexts because the cognitive demands of the task are so great that they interfere with the processing of the implicature. They suggest that sensitivity can be facilitated if the cognitive demands of the task are decreased. They propose that one demanding aspect of the task is the required mode of response; the meta-linguistic judgement that a child usually has to make. Whilst it is the case that modifying the task, so that the child performed an action rather than giving a verbal response, resulted in an increase in pragmatic responding, the children in Hurewitz et al.'s (2006) study were at chance in their detection of scalar implicatures with a task that required an action for a response. In addition other studies have shown spontaneous implicature detection with meta-linguistic judgements (Feeney et al., 2004; Guasti et al., 2005). Therefore whilst this type of response may be cognitively demanding it does not necessarily block scalar implicature detection. In addition, Pouscoulous et al. also made another two changes; distractor items were removed and the scalar term was changed from *certains* to *quelques*. These additional changes make it difficult to assess the relative contribution of the mode of response.

An alternative way to look at the issue of context is to frame the difficulties in terms of effect; in other words what is there to be gained from drawing the implicature. According to Feeney et al. (2004) children do not detect the implicature in the statement evaluation task because there are no cognitive effects to be gained from doing so, therefore they do not expend the effort required to process it. However in the storyboard task employed in the study, the girl's infelicitous reply to her mother could be construed as evidence of an intention to deceive. For example, the girl's claim that she had eaten *some* of the biscuits when it was shown that she had in fact eaten *all* of them could be seen as an attempt to avoid telling her mother that she had eaten *all* the biscuits, without actually lying. The authors propose that it is beneficial for the children to detect deception, and thus the children expend the required effort to process the implicature. This claim that the cognitive effects to be gained from the storyboards arise from the detection of deception seems plausible, as the detection of deception would seem to be beneficial; indeed there are claims that we have an evolved ability to detect cheaters (Cosmides & Tooby, 2005). However, the relative contribution of deception to the detection of scalar implicatures is difficult to assess as it is confounded with another factor; the shared conversational background.

So the role of context in terms of the facilitation of sensitivity to scalar implicatures in children is unclear. The claims of the various authors are difficult to assess given the number of manipulations that are carried out within the same study (Pouscoulous et al., 2007), other potential confounds (Feeney et al., 2004) and contradictory findings (Guasti et al., 2005). Clearly until systematic controlled studies are carried out in relation to context, the question in the title of this subsection will remain unanswered. However, finding out what role context plays in the facilitation of sensitivity to scalar implicatures is important for two reasons. Firstly, it is necessary to establish that the claims made by the individual studies are valid; that is sensitivity, or insensitivity, is due to the claimed factor. This is not easy to establish when a number of variables have been manipulated. Secondly knowing how and why context facilitates sensitivity will aid understanding of how scalar implicatures are processed, and this in turn will have implications for theoretical accounts of scalar implicature, which are the focus of the next part of this review.

1.4.3. Which contemporary theory best captures the processing of the scalar term some?

Within the field of experimental pragmatics there has been a debate over which of the contemporary theories of pragmatics, default or relevance based accounts, best captures scalar implicature processing. It appears that different aspects of the empirical data support both accounts. Part of this contradiction may be explained by the different assumptions that researchers have made regarding the nature of the process responsible for the computation of scalar implicatures. Some researchers see this processing as involving at most two steps (Bott & Noveck, 2004; Noveck & Posda, 2003; De Neys & Schaeken, 2007), whilst others see it as involving up to three steps (Feeney et al., 2004). More research is needed not only to increase our understanding of the nature of the processing of scalar implicatures but to see whether either of the contemporary pragmatic theories can fully account for the empirical data.

The debate over which theory of scalar implicature best captures how scalar terms are processed centres on the issue of speed, with Levinson's (1983, 2000) default theory being contrasted against that of Sperber and Wilson's (1985/1995) Relevance Theory. As was discussed earlier in this chapter, Levinson's default theory proposes that scalar implicatures go through automatically, and consequently a logical interpretation should be slower than a pragmatic interpretation as it would represent cancellation of the implicature. For RT the opposite is true, for contexts where the pragmatic interpretation is not primed. For this account, a logical response represents insensitivity to the implicature and therefore this would be quicker than a pragmatic interpretation, which involves not only attributing to the speaker a meaning provided by the linguistic decoding but also attributing meaning construction based on the context (Noveck & Sperber, 2007). Experimental studies have tested the theoretical claims of Levinson, and Sperber and Wilson by analysing the time taken by adults to make logical and pragmatic interpretations to *infelicitous some* (Bott & Noveck, 2004; De Neys & Schaeken, 2007; Feeney et al, 2004; Noveck & Posda, 2003).

Bott and Noveck (2004) examined the time course of scalar implicatures in a series of experiments where they presented categorical statements to adults which required a true/false judgement. They compared the time taken for participants to make a logical response to *infelicitous some* statements, such as *some elephants are mammals,* with the time taken to make a pragmatic response and found that pragmatic interpretations took significantly longer. They also found that the number of pragmatic responses to the *infelicitous some* statements decreased when the time available to make a response decreased. They concluded that the scalar inference is made after the logical meaning of *some* has been processed. Similar findings came from Noveck and Posda (2003) who presented participants with three different types of *some* statements; patently true, patently false and under-informative statements. In a between-participants analysis they found that participants who responded logically to the under-informative statements were faster in their response times than those who responded pragmatically. In addition, the logical responders were also significantly faster in responding correctly to the other conditions; the patently true and patently false items. The authors concluded that implicatures are part of a late arriving, effort-demanding decision process.

Similar conclusions come from the only study to date that has directly manipulated the effect of cognitive load on the processing of scalar implicatures. De Neys and Schaeken (2007) gave participants a sentence verification task involving *infelicitous some* statements, whilst at the same time burdening their executive cognitive resources with a memory task. They found fewer pragmatic interpretations were made when a complex dot pattern had to be memorised compared to when the dot pattern was simple. According to the authors the decrease in pragmatic responses under load indicates that detecting scalar implicatures requires the involvement of effortful cognitive processes. They suggest that if effort was involved in the logical interpretation, through the cancellation of the implicature as suggested by default theory, then more pragmatic interpretations would have been expected under load.

With regards to how scalar implicatures are processed by children there are no reaction time or direct load manipulation studies. However, a corollary of the studies investigating sensitivity to scalar implicatures in children has been that if children have difficulties detecting the scalar implicature when adults have none, then scalar implicatures must require more effort than the children have at their disposal. According to Pouscoulous et al. (2007) if the cognitive demands of a task are decreased, thus minimising the amount of effort required by the children to process the implicature, then pragmatic response rates increase. The authors made a number of changes to a task where young children had a low pragmatic response rate. In the original task children had to make judgements about the quantity of animals in a box, for example *some turtles are in the boxes*. In the modified version the authors removed distracter items, changed the response from a meta-linguistic judgement to one that was based on an action, and changed the quantifier from *certains* to *quelques*. This modified version resulted in an increase in pragmatic responding amongst younger children. However, even though the manipulations would intuitively seem to decrease the demands placed on cognitive resources, there was no independent measure of the effort required for each manipulation to assess its load factor.

Although the studies described above lend support to relevance based accounts their findings appear to rely on the assumption that a logical interpretation of a scalar term indicates that the scalar implicature has not been detected. This assumption is made explicit by Guasti et al. (2005) who comment that “adult subjects in the Noveck study failed to compute the implicature associated with *some* 41% of the time” (p. 671), and by Katsos, Breheny and Williams (2005) who state that in terms of processing resources the scalar implicature is “costly and cancellation never occurs” (p. 1109).

However, Feeney et al. (2004) proposed that this at-most-two step process was too simple. They suggested that for certain adults there was a third possible step; the cancellation of a scalar implicature after its computation. Like the previously described reaction time studies, Feeney et al. gave participants a statement evaluation task. However rather than comparing reaction times between logical and pragmatic interpretations to *infelicitous some* statements, which they felt was problematic as two different responses were involved, they compared the time taken to say yes to an *infelicitous some* statement with the time taken to give the same response to a statement containing *felicitous some*. They found that the response times differed, with participants taking significantly longer to give a logical response to the infelicitous statement. The authors suggested that this increase in response time was due to the increased processing time involved in the inhibition of the pragmatic implicature. They supported their argument with the finding that logical responding to the *infelicitous some* statements was positively correlated with a measure of cognitive capacity; they suggested that resisting the pragmatic inference required effort and this was achieved by people high in cognitive resources. Thus, contrary to the assumptions made in other studies (Bott & Noveck, 2004; De Neys & Schaeken, 2007; Katsos et al., 2005; Noveck & Posda, 2003) a logical response is not necessarily indicative of a failure to detect the implicature, and both the processing and cancellation of the implicature appear to require effort.

Earlier attempts to discriminate between the theories of Levinson (1983, 2000) and Sperber and Wilson (1985/1995) rest on the assumption that one of the accounts must be right. However both theories appear to have some support. There is evidence that the processing of the implicature involves effort, in line with relevance based accounts. There is also evidence that the cancellation of a scalar implicature involves effort which is in line with default theory. However, this could also fit in with relevance based accounts as, contrary to Katsos et al.'s (2005) statement, RT does allow the cancellation of implicatures even after they have been processed, if expectations of relevance change during the course of comprehension (Garrett & Harnish, 2009; Sperber & Wilson, 2002). Further experimental investigation is needed to test these theories to see whether either approach can adequately accommodate experimental findings.

1.5. Summary and Overview of Experimental Work

There is experimental evidence that in certain contexts adults and children differ in terms of their sensitivity to the scalar implicature associated with *some*, with adults being more sensitive than children. This has led to two assumptions; sensitivity to scalar implicatures develops linearly and the processing of scalar implicatures is effortful. A corollary of this is that adults' and children's interpretations reflect the same processes; a logical interpretation is indicative of a failure to detect the scalar implicature whereas a pragmatic interpretation indicates sensitivity. However, our understanding of adults and children's interpretations of *some* is far from complete and a number of issues remain unresolved. Firstly, the linearity of the developmental trajectory of sensitivity to scalar terms has not been established, merely assumed. Secondly, there is evidence to suggest that Noveck's (2001) assumption that logical interpretations are supplanted by pragmatic interpretations is too simplistic; an additional step may be available. Also, finding the same type of responding in children and adults appears not to be necessarily indicative of the same processing mechanisms; a logical response may reflect either the failure to detect a scalar implicature or the cancellation of the implicature. Following on from this point, although experimental evidence appears to support a relevance based account of scalar implicature processing, there is evidence that effort is implicated in both the processing of the scalar implicature and the cancellation of the implicature. Finally, the role of context in the facilitation of sensitivity to scalar implicatures remains unclear.

The broad aim of the experiments in this thesis is to address the outstanding issues outlined above. Chapter 2 describes four experiments which explore the developmental trajectory of sensitivity to the scalar implicature associated with *some* with three different methodologies, and in a number of age groups ranging from preschool children through to adults. In addition the first experiment in the chapter examines the universal nature of scalar terms by including the connective *or*. Chapter 3 contains two experiments that investigate the role of context in the facilitation of sensitivity to scalar implicatures, by examining the contribution of factors, namely shared conversational background and deception, that have been implicated in that facilitation. The last experimental chapter, Chapter 4, investigates the processing of scalar terms in children and adults to see if the cost of processing a scalar implicature is the same for both age groups, and examines the time course of interpretations in relation to current pragmatic theories of scalar implicature. The final chapter of the thesis, Chapter 5, synthesizes the findings of the experimental chapters, and offers an interpretation based around the three research questions. It also considers the implications and limitations of the research, with suggestions for how future research should proceed.

Chapter 2

The Development of Sensitivity to Scalar Implicatures

2.1. Aim

Although they do not show universal pragmatic responding, adults' experimental detection of scalar implicatures is well accepted (Noveck, 2001). However children's spontaneous detection of scalar implicatures has been the subject of some debate (Feeney et al., 2004; Guasti, et al., 2005; Hurewitz et al., 2006; Noveck, 2001; Papafragou & Musolino, 2003; Papafragou & Tantalou, 2004; Pouscoulous et al., 2007). Also, implicit within the scalar implicature literature is the assumption that the developmental trajectory of sensitivity to scalar terms is linear, even though no study has investigated a wide range of age groups to establish that this is the case. The aim of the four experiments in this chapter was to investigate the development of spontaneous sensitivity to scalar implicatures.

2.2. Introduction

As seen in Chapter 1, scalar implicatures have become the focus of experimental investigation (eg. Bott & Noveck, 2004; Feeney et al., 2004; Noveck, 2001; Noveck & Posda, 2003; Papafragou & Musolino, 2003; Papafragou & Tantalou, 2004), rather than being merely a phenomenon invoked to explain adults' poor performance on reasoning tasks. However, this work is in its experimental infancy, consequently empirical evidence to support claims that have been made is limited. Thus questions still remain regarding the development of sensitivity to scalar implicatures.

Although the literature that is relevant to the aims of this chapter was reviewed in Chapter 1, a brief summary will now be given.

The literature on scalar implicature appears to be divided between studies which suggest that children have difficulties detecting the scalar implicature associated with *some* and those suggesting that it is possible for even young children to detect it.

Studies that have found high rates of logical responding in children, and thus are suggestive of children having difficulty with detecting scalar implicatures, have tended to be those that have used a statement evaluation task (Feeney et al., 2004; Guasti et al., 2005; Noveck, 2001). Nevertheless, high rates of logical responding have also been seen in studies that could be described as scenario evaluation tasks (tasks where children are asked to evaluate a statement based on a scene that they can see) (Papafragou & Musolino, 2003; Pouscoulous et al., 2007). However attention must be drawn to the rates of logical responding of adults in these studies. Whilst the assumption is made that adults' are capable of detecting scalar implicatures their rates of logical responding vary greatly, from very low rates of logical responding, 8% in Papafragou and Musolino's (2003) study, to a high rate of 65% in Feeney et al.'s (2004) study (a rate which incidentally was not significantly different from that of the 7-year-old participants). Moreover, other studies have revealed equivocal rates of logical responding in adults (Noveck, 2001; Guasti et al., 2005; Pouscoulous et al., 2007). This opens up the suggestion that the insensitivity seen in the children may not be a reflection of their pragmatic abilities per se but more a reflection of the impoverished pragmatic nature of the task.

A number of studies have shown that children are not always insensitive to the scalar implicature associated with *some*. Pragmatic response rates have ranged from 77% in 4-year-olds (Papafragou & Tantalou, 2004), to 52% (Papafragou & Musolino, 2003) and 75% (Pouscoulous et al., 2007) in 5-year-old children.

However these findings have a somewhat limited application to establishing whether or not children can spontaneously detect scalar implicatures; training was used in the Papafragou and Musolino experiment, conditions of under-informativeness were not set up in the Papafragou and Tantalou study, and Pouscoulous et al. investigated sensitivity to scalar implicatures through the production of an action rather than a linguistic response.

Spontaneous sensitivity requiring a linguistic response, that is to say observed in the absence of training, has been seen in children in two experiments; the fourth experiment of Guasti et al. (2005) and the second experiment of Feeney et al. (2004). Both studies found rates of pragmatic responding in 7-year-olds equal to or greater than 75%. Unfortunately both of these studies are limited in what they reveal about the developmental trajectory of sensitivity to scalar implicatures in children. The Guasti et al. study showed that 7-year-old children can be adult-like in their detection of scalar implicatures, as their 25% logical response rate did not significantly differ from the 17% logical response rate of the adults. However, as the study investigates only two age groups, this finding represents a snapshot of sensitivity rather than showing when sensitivity develops and how its development proceeds. The Feeney et al. study is important in that it appears to be the only within-participants experiment investigating sensitivity to the scalar implicature associated with *some*[2]. It revealed that 7-year-old children appeared to be insensitive to scalar implicatures on one task but not on another; a 57% logical response rate in a statement evaluation task compared to a 21% logical response rate in a storyboard task. This suggests that sensitivity is not an all or nothing case; once sensitivity emerges in children it is not always seen, which incidentally also appears to be the

[2] Hurewitz et al. (2006) use a within-participants design to investigate sensitivity to numbers compared to quantifiers.

case for adults. Unfortunately only one age group of children was investigated in this study, so the criticisms that were made of the Guasti et al. study regarding developmental findings also apply to this study. In addition no adult data were collected for the Feeney et al. study so it is not known how the adults would respond to the storyboards. In a later experiment in the same paper, Feeney et al. found very high rates of logical responding in adults on a timed statement evaluation task. The authors' analysis of the adults' responses to *infelicitous* and *felicitous some* led them to conclude that there was a group of adults who seemed to be detecting the scalar implicature and then cancelling it. If this was the case and the adults were to do the same in the storyboard task then it is possible that a greater rate of pragmatic responding may be seen in children compared to the adults.

Although the Feeney et al. (2004) study was limited in its findings, its methodology offers the potential to uncover the developmental trajectory of sensitivity to scalar implicatures. The use of two tasks in a within-participants design helps to establish whether sensitivity is task dependent or if it has not yet developed. Likewise the trajectories can be compared to see if development is the same in both tasks. In addition, the storyboard task sets up conditions of under-informativeness, which was a criticism of the Papafragou and Tantalou (2004) study. Furthermore, the control storyboards establish that the children have a good understanding of the terms *all* and *some*, and the high rates of logical responding achieved with this task suggest that it will be suitable for younger children. The storyboard also requires a meta-linguistic judgement, which allows direct comparison with the statement evaluation task. Yet for all these merits, the storyboard methodology is a relatively untested technique[3] and therefore replication of the Feeney et al. finding is required.

Therefore the aim of the experiments in this chapter was to investigate the development of spontaneous sensitivity to scalar implicatures using the statement evaluation task and the storyboard methodology; to establish at what age sensitivity is seen and whether sensitivity develops linearly. To this end the first experiment investigated sensitivity in two groups of children, Year 1 children (5-6-years-of age) and Year 4 children (8-9-years-of-age), and a group of adults. The second experiment expanded the previous experiment's age range to include two adolescent age groups; Year 7 children (11-12-years-of-age) and Year 10 children (14-15-years-of-age). The third experiment focused on preschool children aged 3-4-years-old, and the fourth experiment followed up the findings of Experiment 3 with a novel methodology.

2.3. Experiment 1: Sensitivity to scalar implicatures in children and adults

2.3.1. Introduction

This first experiment followed up Feeney et al.'s (2004) results in two ways; it examined whether the facilitatory effect of context seen in the 7-year-old children extended to other age groups and it also examined whether the quantifier findings extended to another type of scalar term, the connective *or*.

It would be expected that the findings from one scalar term would generalise to another as scalar terms are often considered to be homogenous in the derivation of their meaning, in that they have a lower-bounded literal interpretation, whilst their upper-bounded interpretation is arrived at by means of an implicature (Grice, 1989;

[3] The storyboard methodology has been used to look at scalar implicatures in second language acquisition (Slabakova, in press)

Horn, 1972). Therefore for comparable tasks it might be expected that context will affect interpretations of *some* and *or* in a similar manner. Indeed it has been suggested that findings from one scalar term do generalise to another (Noveck, 2001). However, there is work from cardinals (Huang & Snedeker, 2005; Hurewitz et al., 2006; Papafragou & Musolino, 2003) that suggests that children do not treat all scalar terms in the same way. Thus finding out whether the dissociation seen with *some* is seen with *or* would help resolve this debate. Before describing the aims of the study and the experiment itself, a brief explanation of the different interpretations of *or* will be given, followed by a summary of the findings on *or* from the literatures on experimental pragmatics and reasoning.

Logical and pragmatic interpretations of the connective or

Logically, *or* is defined as a truth functional connective that links two disjuncts. *Or* is taken to be logically true if at least one of the disjuncts is true. Thus it is considered true if both disjuncts are true, and so, is compatible with *and*. For example, it is commonly understood that the statement, *the applicant should have a degree or five years experience*, does not rule out someone applying who satisfies both requirements. However in natural language *or* tends to be seen as offering an alternative between two disjuncts. For example the statement, *we will go to the cinema or the zoo*, suggests that one or other of the attractions will be visited, but not both; if both attractions are to be visited then it would be expected that the stronger term *and* would have been used. The difference in meaning for *or* can be made distinct by preceding the connective with the terms *inclusive*, meaning maybe both, or *exclusive*, meaning not both. The difference between the two definitions can be seen in Table 2.1.

Table 2.1

Truth table for the connective or

Proposition		Inclusive Or	Exclusive Or
P	Q		
True	True	True	False
True	False	True	True
False	True	True	True
False	False	False	False

As can be seen from Table 2.1 the only distinction between the two forms of *or* is the truth value for the proposition where both disjuncts are true; this proposition is true for *inclusive or*, but false for *exclusive or* (for a review of the debate over whether the two interpretations of *or* constitute two semantically different terms see Jennings, 1994; and Simons, 2000).

Children's and adults' sensitivity to the implicature a*ssociated with or*

With regards to the development of sensitivity to the implicature associated with *or*, the findings from both children and adults are mixed.

In the literature on experimental pragmatics high inclusive/logical response rates have tended to suggest that both children and adults are insensitive to the implicature associated with *or* (Chierchia et al., 2001; Chierchia, Crain, Guasti & Thornton, 1998; Chierchia et al., 2004; Gualmini, Crain & Meroni, 2000; Gualmini, Meroni & Crain, 2001). For example, Chierchia et al. (2004) found that 5-year-old children overwhelmingly allowed both disjuncts to be true when it involved making a bet on the outcome of a story. When faced with a situation where a puppet said *I bet that Batman will take a cake or an apple*, the children honoured the bet 95% of the time when Batman was seen to take both. The adults also honoured the bet the majority of the time, with a compliance rate of 60%. Similarly Chierchia et al. (2001) found that the statement *every dwarf who chose a banana or a strawberry received a jewel* was accepted by children 92% of the time, when it was the case that dwarves who received the jewel had chosen both. The adults displayed an even higher acceptance rate of 95%. However, there is evidence from other studies for universal pragmatic responding in adults. When Chierchia et al. (2001) changed the scenario and statement to *every boy chose a skateboard or a bike*, the adults always considered this to be false when both items were chosen. The children's logical response rate in this study also dropped to 50%.

One problem with interpreting the results of this literature, is that *or* is quite often in the scope of other logical operators, such as the quantifier *every* (Chierchia et al., 2001), the conditional *if* (Chierchia et al., 2004; Gualmini et al., 2000) and the modal *can* (Gualmini et al., 2001). It is unclear what effect these other logical operators have on the detection of a scalar implicature as some researchers propose that linguistic structure alone is responsible for the interpretation of the connective (Cherchia et al., 2001; Gualmini et al., 2001), whilst others have investigated extralinguistic factors, such as the effect of predicting an outcome that involves a wager (Chierchia et al., 1998; Chierchia et al., 2004).

Findings in the reasoning literature have led researchers to suggest that adults do not necessarily favour a pragmatic interpretation of *or*, unlike the work on *some*, with high rates of inclusive responding being found (Braine & Rumain, 1981; Evans & Newstead, 1980; Paris, 1973; Winer, 1990). For example, Evans and Newstead found that the majority of adults considered the proposition *either the letter is an A or the number is a 3* to be true when it was followed by the letter/number pair A3.

Similarly, adults in the Paris study who were asked to decide whether a spoken description, such as *the bird is in the nest or the shoe is on the foot,* matched an accompanying slide presentation, considered the statement to be true 75% of the time when both items were present. In addition the children in this study also appeared to favour an inclusive interpretation, with the majority considering the statement to be true if both items were present. Logical response rates ranged from 97% in 7-year-olds to 75% in 16-year-olds.

The Paris (1973) study also highlights an important point when considering inclusive/exclusive interpretations of *or*, that of considering the whole truth table. Regardless of which interpretation is given to *or*, the connective should be accepted as true when one of the disjuncts is true. A closer look at the truth table in the study reveals that errors are made on those propositions that involve mixed forms (true/false and false /true) and vary from 70% for 7-year-old children to 24% for college students. This acceptance of *or* as true when both disjuncts are true, but false when one disjunct is false, suggests the participants are interpreting *or* as conjunction.

The tendency to interpret *or* as conjunction has also been found in research that has used a different methodology to that used by Paris (1973). For example, Suppes and Feldman (1971) asked children between the ages of 4 and 6 years to *give me the things that are green or square,* and found that children offered green squares 47% of the time, which was the majority response rate. Interestingly this response rate was far greater than that of where either the green items were offered (25%), or where the square items were offered (6%); both being indicative of an exclusive interpretation. The inclusive interpretation of *or,* offering all the green things and all the square things as well as the green squares, had only a 3% response rate. In another example Neimark and colleagues (Nitta & Nagano, 1966, cited in Neimark & Slotnick, 1970; Neimark & Slotnick, 1970; Neimark, 1970) investigated set relations, using mimeographed tasks, in 7- to 19-year-olds, and found that when asked to circle the sets that corresponded to the statement *black or birds*, the participants tended to circle the set that contained only black birds, rather than including the sets that contained black things and birds. Even though errors declined sharply with age, the older participants made the conjunction error a quarter of the time.

However it is not just children who make conjunction errors in their interpretation of *or*. Gualmini et al. (2004) found that adults always rejected the statement *John or Paul can lift the piano* when it was shown that both were able to. Moreover when Evans et al. (1993) reviewed adult studies on disjunctive reasoning they found that when the whole truth table was considered, rather than just individual rates of inclusive/exclusive responses, more adults gave exclusive rather than inclusive interpretations.

Experimental aims and predictions

The aim of this experiment was three-fold; firstly to examine developmental trends in people's sensitivity to scalar terms, secondly to investigate whether interpretations of scalar terms change in relation to context, and finally to examine the homogeneity of scalar terms. Given the findings of Feeney et al. (2004), this experiment employed the same methodology to establish that their effect was reliable, to investigate whether a dissociation in context was seen in other age groups, namely adults and Year 1 children, and to see whether the findings from the scalar term *some* extended to the scalar term *or*. To accommodate this final aim a statement evaluation task and storyboards comparable with those used to investigate *some* and *all* were devised for the connectives *or* and *and*, and were incorporated with the quantifier statements and storyboards.

Four predictions were made: First, based on the dissociation in context seen in the Feeney et al. (2004) study it was predicted that the Year 4 children would show limited sensitivity to scalar implicatures in the statement evaluation task but significant sensitivity in the storyboard task. Following on from this, it was anticipated that the Year 1 children would be insensitive to scalar implicatures in the statement task, although it was expected that some of them would be able to detect the implicature in the storyboard task. Thus the second prediction was that a dissociation in context would be seen in the Year 1 children. With regards to the adults it was unclear how they would respond between contexts, given that it is generally accepted that they can detect scalar implicatures but don't always do so. However, taking into account the assumptions about the developmental trajectory of sensitivity to scalar terms, the third prediction was that the adults would show the same or increased rates of sensitivity in both tasks compared to the children. The final prediction was that as scalar terms are considered to be homogenous in their derivation, the findings from *some* would generalise to *or*.

2.3.2. Method

Participants

Seventy one people participated in the experiment. These were split into three age groups: twenty three Year 1 children, age range 5 years 7 months to 6 years 6 months; twenty four Year 4 children, age range 8 years 7 months to 9 years 6 months; and twenty four adult participants, age range 18 years to 60 years. The children attended a school in the North East of England and the adults were recruited from users of a church hall in the same area. The head teacher gave consent for the children whilst the adults gave their own consent.

Materials

The participants heard a list of 48 randomly ordered statements and 32 randomly ordered storyboards. Half of the statements and storyboards contained the quantifiers *some* or *all*, and half contained the connectives *or* or *and*.

Quantifier statement materials. These materials were based on the statement materials of Noveck (2001) and concerned three types of information; factually universal, factually existential and absurd. A factually universal statement is one where the quantifier *all* best describes how many of a group possess a particular feature, such as *all elephants have trunks*. A factually existential statement is one where the quantifier *some* best describes how many of a group possess a particular feature, such as *some animals have stripes*, whilst an absurd statement is one where neither quantifier correctly describes it, such as *garages sing*. Each statement described a relationship between a quantifier and a noun and was presented in one of the following subgroups: *True all* statements, for example *all* giraffes have long necks; *false all* statements, for example *all* dogs have spots; *felicitous some* statements, for example *some* houses have a garage; *infelicitous some* statements, for example *some* snow is cold; *absurd all* statements, for example *all* policeman are made of jelly; *absurd some* statements, for example *some* peas wear shorts. These subgroups and examples can be seen in Table 2.2.

Table 2.2

The quantifier subgroups, with examples, for each task in Experiment 1

Sub Group	Statements	Storyboards
True All	All elephants have trunks	I ate all the biscuits (ate 3/3 biscuits)
False All	All dogs have spots	I dirtied all the towels (dirtied 2/3 towels)
Absurd All	All garages sing	
Felicitous Some	Some animals have stripes	I broke some of the tiles (broke 2/3 tiles)
Infelicitous Some	Some giraffes have long necks	I picked some of the flowers (picked 3/3 flowers)
Absurd Some	Some policemen are made of jelly	

Each statement was presented with either *all* or *some* preceding it, although each participant received only one form, which resulted in two sets of materials (see appendix AI for statements used).

Quantifier storyboard materials. The storyboards were taken directly from Feeney et al. (2004). An example of a storyboard can be seen in Figure 2.1.

Charlotte finds 3 sweets

She eats the first sweet

She eats the second sweet

She eats the third sweet

Charlotte’s Mum says,

“Charlotte what have you done with the sweets?”

Figure 2.1. An example of a quantifier storyboard used in Experiment 1.

Each of the storyboards contained four or five coloured photographs, depending on the subgroup, and depicted a theme. There were 16 themes (see appendix AII for themes used) each of which centred on the activities of a girl in relation to a set of objects she had found, such as eating sweets or picking flowers. In each storyboard a girl called Charlotte finds three objects and interacts with either two or three of them. In each case her mother asks what she had done to the objects. Charlotte replies by stating that she has interacted with *all/some* of them, such as, *I'‘ve eaten all/some of the biscuits*. Participants saw each theme in one of four forms corresponding to four of the subgroups of the statement materials; four *true all*, four *false all*, four *felicitous some* and four *infelicitous some* storyboards. There were no absurd forms of the storyboards. The subgroups with examples can be seen in Table 2.2. Each storyboard was presented with either *all* or *some* appearing in Charlotte's reply, although each participant received only one form, which resulted in two sets of materials.

Connective statement materials. The connective statements, containing *and* or *or*, were designed to match four of the subgroups of the quantifier statements. Thus there were *true and*, *false and*, *felicitous or* and *infelicitous or* statements, but no absurd connective statements. The quantifier and connective statements also matched in that they contained the same noun, for example an elephant was mentioned in both sets of statement materials. An example statement for each subgroup can be seen in Table 2.3.

Table 2.3

The connective subgroups, with examples, for each task for Experiment 1

Sub Group	Statements	Storyboards
True And	An elephant has a trunk *and* ears	I ate the Twix and the Kit Kat (ate both)
False And	A dog has a tail *and* an umbrella	I pulled Katy's hair and Beth's hair (pulled one girl's hair)
Felicitous Or	A book has pages *or* a nose	I cut the leg off the dog or the leg off the teddy (cut one)
Infelicitous Or	A giraffe has a long neck *or* a tongue	I broke the white tile or the green tile (broke both)

Two sets of connective statements were prepared, which can be seen in Appendix AI.

Connective storyboard materials. As was the case with the connective statement materials, the connective storyboards were designed to match the subgroups of the quantifier storyboards. Thus there were *true and*, *false and*, *felicitous or and infelicitous or* storyboards. The themes chosen for the storyboards also matched the themes used in the quantifier storyboards (see appendix AII for themes used). An example connective storyboard can be seen in Figure 2.2.

Claire sees two tiles

She breaks the white tile

She breaks the green tile

Claire's mum says, "Claire what have you done to the tiles?"

Figure 2.2. An example of a connective storyboard used in Experiment 1.

In the connective storyboards a girl called Claire finds two objects and interacts with either one or both of them. In each case her mother asks her what she has done to the objects, and Clare replies that she has interacted with one *and/or* the other, for example *I've eaten the Twix and/or the Kit Kat*. Each theme had four subgroups, corresponding to the subgroups of the quantifier storyboards, and an example of each can be seen in Table 2.3. Participants saw each theme in one of the four subgroups. Two sets of materials were prepared.

Each set of connective statements was paired with a set of quantifier statements, which were then randomly ordered. The same was done for the connective and quantifier storyboards. This resulted in two sets of materials which were ordered and counterbalanced to produce four groups.

Procedure

Participants received instructions on how to complete each task immediately before they attempted that task. For the statement task participants were told they would hear a series of statements and they had to decide whether each statement was true or false. For the storyboard task participants were told they would see a number of photographs and hear accompanying text. They had to decide whether the response made by the child was true or false. For both tasks participants had an answer sheet on which to record their responses; a tick if they considered the statement/response to be true, a cross if they considered it to be false.

All of the children were tested in small groups and received their materials in three parts to prevent fatigue effects. The statements were presented in one sitting whilst the storyboards were split into two sittings with sixteen storyboards in each. The adults were tested either individually or in small groups and completed the tasks in one sitting.

2.3.3. Results

Alpha levels were set at $p \leq .05$ for all the inferential tests carried out in this and subsequent experiments. No adjustments were made for multiple comparisons[4].

The participants' responses were coded with respect to their logical correctness. The correct response for the absurd statements is always false. The correct logical response for the true/felicitous *all* and *and* materials is true, whilst the correct logical response for the false/infelicitous *all* and *and* materials is false. However for *some* and *or* the correct logical response is always true regardless of truth/felicity.

The data from one Year 1 child was excluded from the analysis due to poor responding on the *true all* and *true and* storyboards.

Quantifier data

The mean scores and standard deviations for the remaining 70 participants according to truth, term and task can be seen in Table 2.4. Each participant scored 7 or 8 on the absurd statements, and this result was not included in further analysis.

[4] There are a number of procedures available to make alpha-adjustments when multiple statistical tests are carried out in a study; the Bonferroni correction being one of the most common. The rational for the use of such methods is to control for type I errors. However the use of these methods is hotly debated (see Matsunaga, 2007; O'Keefe, 2003; Perneger, 1998; Rothman, 1990; Tutzauer, 2003). Taking this debate into account, and in view of the infancy of the phenomena being studied and the possibility of type II errors, the decision was made to simply report test statistics, *p* values and effect sizes and discuss possible interpretations of each result.

Table 2.4

Mean number of logical responses (and standard deviations) for the quantifier terms in Experiment 1

Truth/felicity	Term	Task	Age		
			Year 1 (n = 22)	Year 4 (n = 24)	Adult (n= 24)
True/ Felicitous	All	Statements	3.86 (0.35)	3.88 (0.34)	3.88 (0.34)
		Storyboards	3.77 (0.53)	3.96 (0.20)	3.92 (0.28)
	Some	Statements	3.50 (0.67)	3.96 (0.20)	3.96 (0.20)
		Storyboards	2.91 (1.34)	3.12 (1.36)	3.88 (0.45)
False/ Infelicitous	All	Statements	3.41 (0.59)	3.83 (0.48)	3.96 (0.20)
		Storyboards	3.68 (0.72)	3.67 (0.64)	3.88 (0.34)
	Some	Statements	2.68 (1.81)	2.63 (1.79)	2.17 (1.55)
		Storyboards	0.86 (1.39)	0.37 (0.71)	1.67 (1.81)

Note: The maximum number of correct logical responses for each statement type was 4.

Statement data. Collapsing results across the age groups, there were significantly fewer logical responses given to the *some* trials (mean 6.30) than to the *all* trials (mean 7.61), Wilcoxon matched pairs signed ranks test, $z = 4.79$, $p < .001$, $r = .40$, and to the false/infelicitous trials (mean 6.23) compared with the true/felicitous trials (mean 7.69), Wilcoxon matched pairs signed ranks test, $z = 5.47$, $p < .001$, $r = .46$. These results also held for the individual age groups, all *p's* < .05, all *r's* ≥ .31 (see appendix BI, Tables 1 & 2, for statistics for each age group).

The number of logical responses given to each type of statement by each age group was analysed using Friedman tests. These analyses revealed that the effect of statement type for the Year 1 children was approaching significance, χ^2 (3) = 7.64, $p = .053$, and was significant for the other two age groups, all *p's* ≤ .001 (see appendix BI, Table 3, for the statistics for the individual age groups) . Follow up Wilcoxon signed ranks tests revealed that the Year 4 and adult age groups gave significantly fewer logical responses to the *infelicitous some* statements compared to any other statement type, all *p's* ≤ .009, all *r's* ≥ .38, whilst for the Year 1 age group the *infelicitous some* statements significantly differed only from the *true all* statements, $z = 2.53$, $p = .011$, $r = .38$ (see appendix BI, Table 4, for statistics for the comparison of each statement type for each age group).

Storyboard data. Collapsing results across the age groups, revealed that significantly fewer logical responses were given to the *some* trials (mean 4.29) than to the *all* trials (mean 7.63), Wilcoxon matched pairs signed ranks test, $z = 6.86$, $p < .001$, $r = .58$, and to the false/infelicitous trials (mean 4.71) compared with the true/felicitous trials (mean 7.20), Wilcoxon matched pairs signed ranks test, $z = 6.62$, $p < .001$, $r = .56$. These results also held for the individual age groups, all *p's* < .001, all *r's* ≥ .52 (see appendix BI, Tables 5 & 6, for statistics for each age group).

The number of logical responses given to each type of storyboard by each age group was analysed using Friedman tests and these were significant for all age groups, all *p's* < .001 (see appendix BI, Table 7). This analysis was followed up with Wilcoxon signed ranks tests. This analysis revealed that all age groups gave significantly fewer logical responses to the *infelicitous some* storyboards compared to any other storyboard type, all *p's* ≤ .001, all *r's* ≥ .50 (see appendix BI, Table 8, for statistics for the comparison of each storyboard type for each age group).

Infelicitous some data. The nature of the distribution of the *infelicitous some* responses for both tasks can be seen in Figure 2.3.

In the statement task, the distributions of the Year 1 and Year 4 children were unimodal, with the majority of children always giving a logical response, although in both age groups there were a number of children who always gave a pragmatic response. The adults, however, were much more equivocal, and their distribution was more evenly spread. The distributions in this task contrast sharply with the distributions for the *infelicitous some* data in the storyboard task. The distributions of the Year 1 and Year 4 children were once again unimodal, but the majority of children always gave a pragmatic response, and there were very few children who always gave a logical response. The adult data is also unimodal, with the majority of adults always giving a pragmatic response, however there were also a number of adults who always gave a logical response.

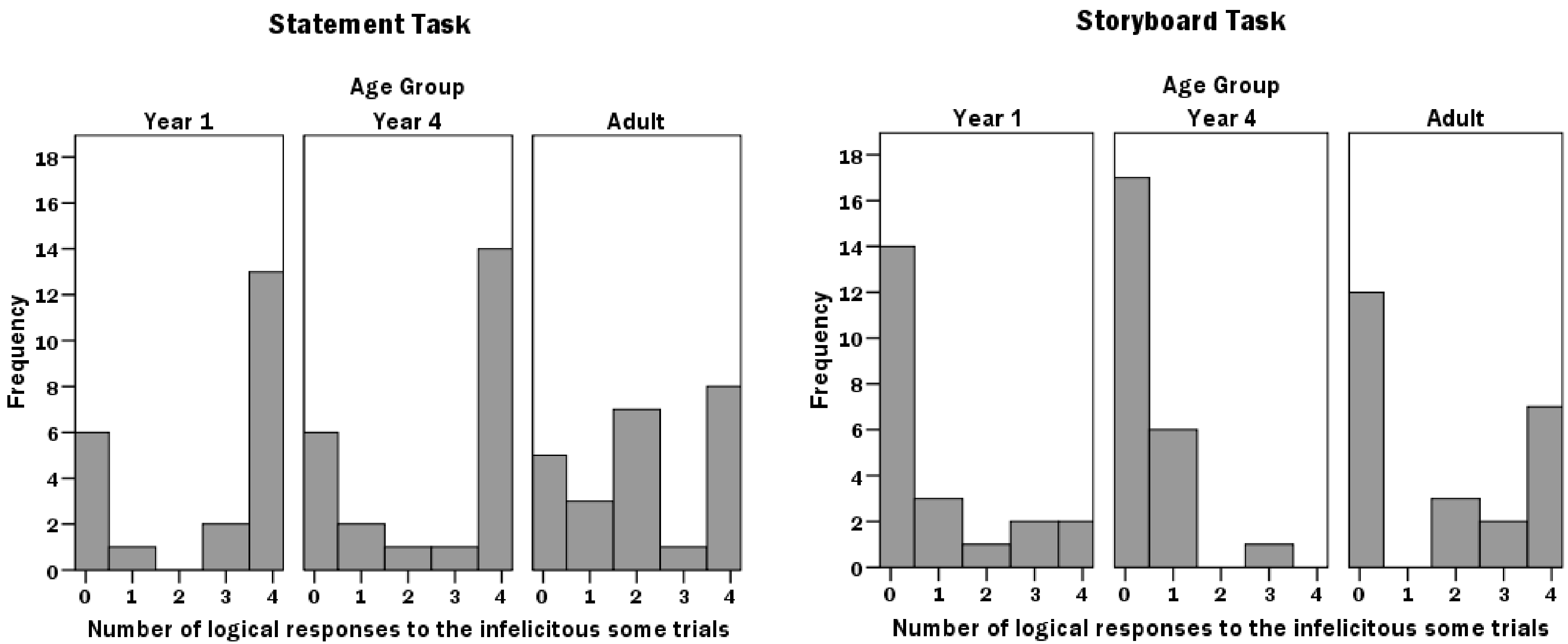


Figure 2.3. Distribution of participants according to the number of logical responses given to the infelicitous some trials, according to age and task, in Experiment 1

Collapsing across age groups, significantly fewer logical responses were given to the *infelicitous some* trials in the storyboard task (mean 0.97) than in the statement task (mean 2.49), Wilcoxon matched pairs signed ranks test, $z = 5.06$, $p < .001$, $r = .43$. The effect of task for each age group can be seen in Figure 2.4.

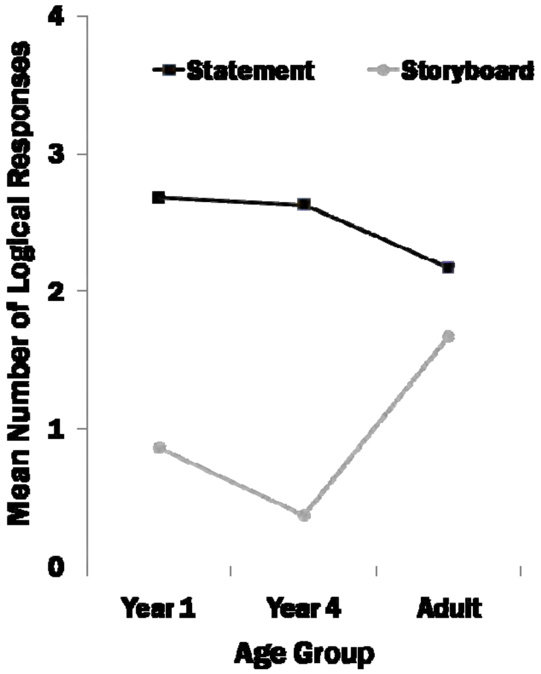


Figure 2.4. Mean number of logical responses to the infelicitous some trials, according to age and task, in Experiment 1.

Significantly fewer logical responses were given in the storyboard task by the Year 1 children, Wilcoxon matched pairs signed ranks test, $z = 3.17$, $p \leq .002$, $r = .48$, and by the Year 4 children, $z = 3.70$, $p < .001$, $r = .53$, but there was no significant effect of task for the adults, Wilcoxon matched pairs signed ranks test, $z = 1.21$, $p = .226$, $r = .17$.

With regards to the number of logical responses given to the *infelicitous some* trials between the age groups, there was no significant difference in the statement task, Kruskal-Wallis test, $H(2) = 1.63$, $p = .443$, but the difference in the storyboard task was approaching significance, Kruskal-Wallis test, $H(2) = 5.53$, $p = .063$. Additional analysis on the storyboard finding revealed that the there was no difference in the number of logical responses given by the Year 1 and Year 4 children, Mann Whitney test, $U = 231$, $p = .381$, $r = .13$, but the Year 4 children gave significantly fewer logical responses than the adults, Mann Whitney test, $U = 190$, $p = .022$, $r = .33$.

Connective data

The mean scores and standard deviations for the participants according to truth, term and task can be seen in Table 2.5.

Table 2.5

Mean number of logical responses (and standard deviations) for the connective terms in Experiment 1

Truth/felicity	Term	Task	Age		
			Year 1 (n = 22)	Year 4 (n = 24)	Adult (n= 24)
True/ Felicitous	And	Statements	3.91 (0.29)	3.96 (0.20)	3.96 (0.20)
		Storyboards	3.91 (0.29)	3.75 (0.68)	3.96 (0.20)
	Or	Statements	0.00 (0.00)	1.58 (1.59)	1.13 (1.66)
		Storyboards	1.05 (1.25)	1.88 (1.75)	2.00 (1.77)
False/ Infelicitous	And	Statements	3.95 (0.21)	3.71 (0.62)	4.00 (0.00)
		Storyboards	3.77 (0.53)	4.00 (0.00)	3.92 (0.28)
	Or	Statements	3.41 (0.73)	3.12 (1.57)	0.92 (1.25)
		Storyboards	1.41 (1.47)	2.17 (1.83)	0.67 (1.27)

Note: The maximum number of correct logical responses for each statement type was 4.

Statement data. Collapsing results across the age groups, significantly fewer logical responses were given to the *or* statements (mean 3.39) than to the *and* statements (mean 7.83), Wilcoxon matched pairs signed ranks test, $z = 7.02$, $p < .001$, $r = .59$, and to the true/felicitous statements (mean 4.87) compared to the false/infelicitous statements (mean 6.34), Wilcoxon matched pairs signed ranks test, $z = 4.36$, $p < .001$, $r = .37$. These results also held for the Year 1 and Year 4 children, all p's $\leq .006$, all r's $\geq .40$. The difference in logical response rates between *and* and *or* also held for the adults, Wilcoxon signed ranks test, $z = 4.22$, $p < .001$, $r = .61$, however the difference in logical responding to the true/felicitous and false/infelicitous statements was not significant for this age group, $z = 0.29$, $p = .774$, $r = .04$ (see appendix BI, Tables 9 & 10, for statistics for each age group).

The number of logical responses given to each type of statement by each age group was analysed using Friedman tests, and these were significant for each age group, all p's $< .001$ (see appendix BI, Table 11). Follow up tests for the Year 1 children revealed that whilst the *infelicitous or* statements (mean 3.41) differed from all other statement types, Wilcoxon matched pairs signed ranks test, all p's $\leq .008$, all r's $\geq .40$, *felicitous or* statements (mean 0.00) also differed from all other statement types, all p's $<.001$, all r's $\geq .64$. For the Year 4 children *infelicitous or* (mean 3.12) differed from *true and* (mean 3.96) and *felicitous or* (mean 1.58), whilst *felicitous or* differed from all other statement types, all p's $\leq .001$, all r's $\geq .47$. In the adult participants *infelicitous or* and *felicitous or* did not differ from each other, $z = 0.36$, $p = .73$, $r = .05$, but each differed from the other statement types, all p's $< .001$, all r's $\geq .58$ (see appendix BI, Table 12, for statistics for the comparison of each statement type for each age group).

Storyboard data. Collapsing results across the age groups, revealed that significantly fewer logical responses were given to the *or* storyboards (mean 3.07) than to the *and* storyboards (mean 7.77), Wilcoxon matched pairs signed ranks test, $z = 6.90$, $p < .001$, $r = .58$, but there was no difference between the false/infelicitous storyboards (mean 5.31) compared with the true/felicitous storyboards (5.53), Wilcoxon matched pairs signed ranks test, $z = 0.76$, $p = .446$, $r = .06$. The difference between *or* and *and* also held for the individual age groups, all *p's* < .001, all *r's* ≥ .53 (see appendix BI, Table 13, for statistics for each age group). There was no difference between the true/felicitous and false/infelicitous storyboards for the Year 1 children, Wilcoxon matched pairs signed ranks test, $z = 0.93$, $p = .367$, $r = .14$, but for the Year 4 children and the adults fewer logical responses were given to the false/infelicitous storyboards than the true/felicitous storyboards, all *p's* ≤ .030, all *r's* ≥ .31 (see appendix BI, Table 14, for statistics for each age group).

The number of logical responses given by each age group to each type of storyboard was analysed using Friedman tests and these were significant for all age groups, all *p's* < .001 (see appendix BI, Table 15, for statistics for each age group). Follow up tests for the Year 1 and Year 4 children revealed that the *infelicitous or* and *felicitous or* storyboards did not differ from each other, Year 1, z = 0.99, $p = .321$, $r = .15$, Year 4, z = 1.46, $p = .144$, $r = .21$, but *infelicitous or* and *felicitous or* each differed from the other storyboard types, all *p's* ≤ .002, all *r's* ≥ .45. In the adult participants *infelicitous or* had significantly fewer logical responses than all the other storyboard types, all *p's* ≤ .003, all *r's* ≥ .43, however *felicitous or* also had significantly fewer logical responses than the other storyboard types, all *p's* ≤ .003, all *r's* ≥ .43 (see appendix BI, Table 16, for statistics for the comparison of each storyboard subgroup for each age group).

Infelicitous or data

The nature of the distributions for the *infelicitous or* responses in both tasks can be seen in Figure 2.5. In the statement task the distributions of the Year 1 and Year 4 children were unimodal, with the majority of children always giving a logical response. The distribution of the adults was also unimodal but their majority response was to always be pragmatic. In the storyboard task the majority of the Year 1 children always gave a logical response, although the remainder of the children were evenly distributed across the scores. The Year 4 children's data was bimodal with the majority of children either always giving a logical response or never giving one. The adults' distribution was similar to that of the statements task with the majority never giving a logical response.

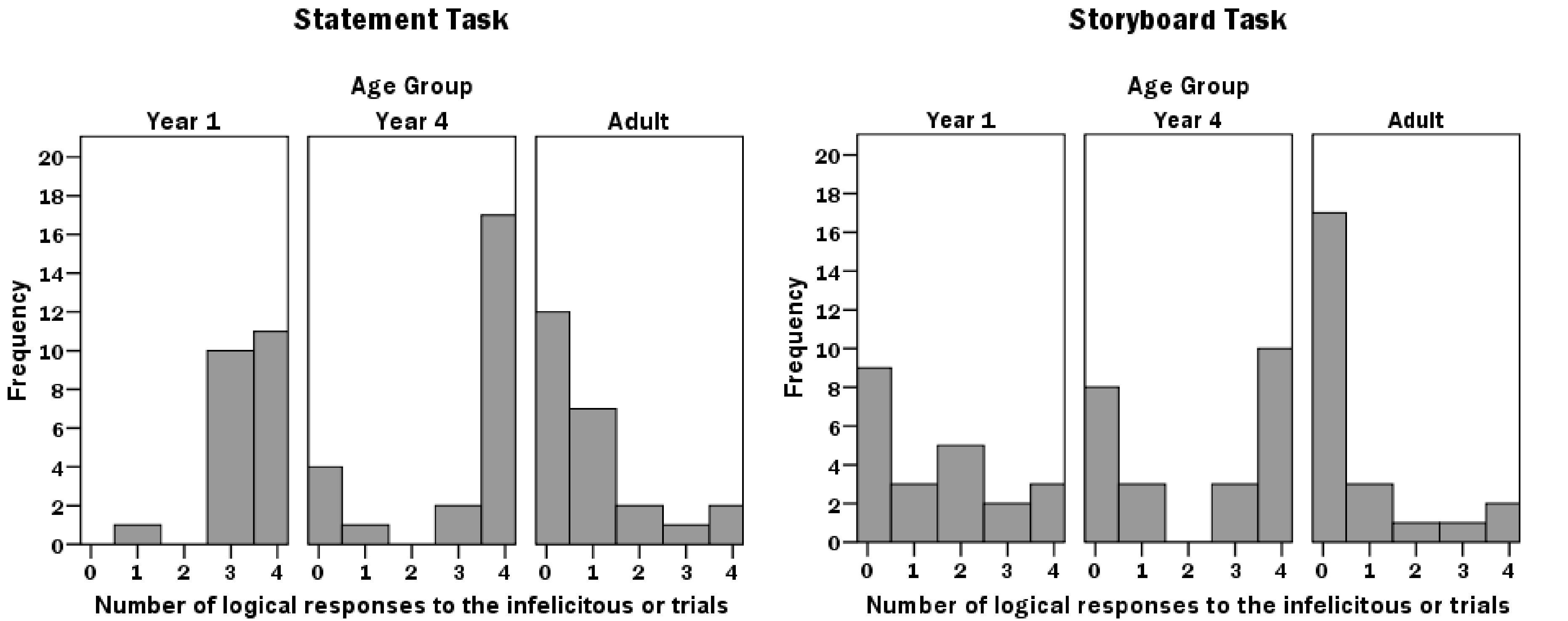


Figure 2.5. Distribution of participants according to the number of logical responses given to the infelicitous or trials, according to age and task, in Experiment 1.

Collapsing across age groups, significantly fewer logical responses were given to the *infelicitous or* trials in the storyboard task (mean 1.41) than in the statement task (mean 2.46), Wilcoxon matched pairs signed ranks test, $z = 4.29$, $p < .001$, $r = .36$. The effect of task for each age group can be seen in Figure 2.6.

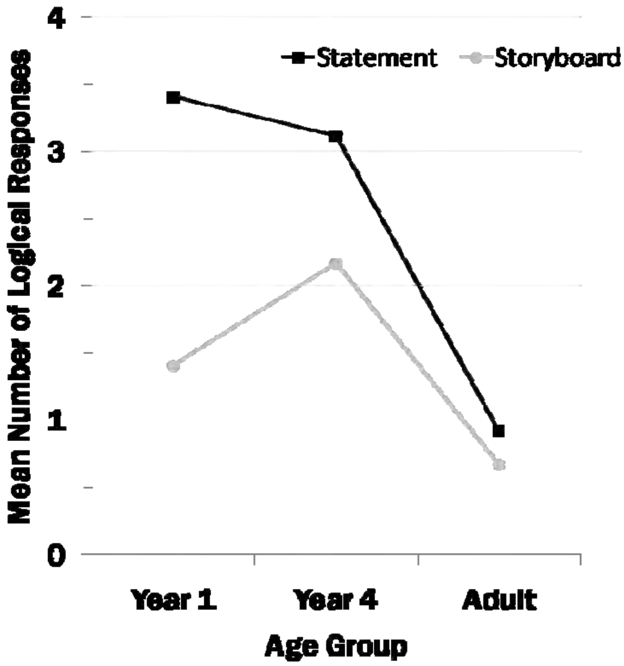


Figure 2.6. Mean number of logical responses to the infelicitous or trials, according to age group and task, for Experiment 1.

Significantly fewer logical responses were given in the storyboard task by the Year 1 children, Wilcoxon matched pairs signed ranks test, $z = 3.88$, $p < .001$, $r = .58$, and by the Year 4 children, Wilcoxon matched pairs signed ranks test, $z = 2.26$, $p = .024$, $r = .33$, but there was no significant effect of task for the adults, Wilcoxon matched pairs signed ranks test, $z = 0.90$, $p = .369$, $r = .13$.

With regards to the number of logical responses given to the *infelicitous or* trials between the age groups, there was a significant difference in the statement task, Kruskal-Wallis test, $H(2) = 27.87$, $p < .001$, and in the storyboard task, Kruskal-Wallis test, $H(2) = 9.53$, $p = .009$. Following up these results with Mann-Whitney tests revealed there was no significant difference in the number of logical responses given by the Year 1 compared to the Year 4 children in either the statement task, $U = 235.00$, $p = .460$, $r = .11$, or the storyboard task, $U = 201.50$, $p = .152$, $r = .21$, but the Year 4 children gave significantly more logical responses than adults in both tasks; statement task, $U = 102.50$, $p < .001$, $r = .58$, storyboard task, $U = 158.00$, $p = .003$, $r = .42$.

2.3.4. Discussion

Two of the predictions made at the start of the experiment were confirmed: The Year 4 children showed greatly increased sensitivity to the scalar implicature associated with *some* in the storyboard task and a dissociation in context was seen in the Year 1 children. However, the third prediction that the development of sensitivity to scalar implicatures would be linear was not supported as the adults had a higher logical response rate to the *infelicitous some* trials in the storyboard task compared to the Year 4 children. Finally, given the problems that were encountered with *or*, it was difficult to ascertain whether the findings from *some* generalised to *or*, and thus the fourth prediction remained untested.

Given the difficulties participants had with the connective control statements (this will be expanded on shortly), the connective and quantifier results will be discussed separately.

With regards to the number of logical responses given to the *infelicitous some* trials the Year 4 results are in keeping with the findings of Feeney et al. (2004) in that the 8-year-old children show sensitivity to the scalar implicature. This sensitivity is somewhat limited in the sentence evaluation task, with a logical response rate of 66%, but nonetheless this response rate is significantly lower than those of the control statements. The logical response rate is also equal to that found by Feeney et al. and lower than those of Noveck (2001) and Guasti et al. (2005). In addition very high rates of pragmatic responding were seen in the storyboard task, with a pragmatic response rate of 91%. This is even higher than the 79% pragmatic response rate found by Feeney et al., suggesting that their conclusion, that children are capable of detecting scalar implicatures, is valid. These results indicate that these children are more likely to reject an infelicitous statement, such as *some elephants have trunks*, than they are a felicitous one, such as *some houses have a garage*. However, they are more likely to detect infelicity when it is part of a naturalistic context, such as in the storyboards, rather than in a contrived situation, such as the statement evaluation task. This finding also supports studies that have found high rates of pragmatic responding to *infelicitous some* using naturalistic contexts (Guasti et al., 2005; Papafragou & Tantalou, 2004; Pouscoulous et al., 2007).

The results from the Year 1 children show that when given pragmatically enriched materials children as young as 5-years-of-age can spontaneously detect the implicature associated with *some*, even when it requires a meta-linguistic judgement. Their pragmatic response rate of 77% in the storyboard task was far greater than the 52% seen in Papafragou and Musolino's (2003) between-participants experiment, and it was achieved without training. Although sensitivity to the scalar implicature was not significant in the statement task, the 67% logical response rate observed in this experiment was less than the 85% of logical responses given by the 10-year-olds in Noveck's study. In addition, in this study there were six children who always rejected the *infelicitous some* statements, suggesting that for some young children the implicature is readily available even with pragmatically impoverished materials.

The adults, as expected, were equivocal in their response to the *infelicitous some* statements, with a logical response rate of 54%, which falls between the 66% rate reported in Feeney et al. (2004) and the 41% rate reported by Noveck (2001). Surprisingly the adults did not show significantly more sensitivity in the storyboard task. Even more unexpected was the finding that their 42% logical response rate in the storyboard task was significantly higher than that of the Year 4 children. This result goes against the assumption that children should never be more pragmatic than adults (Noveck, 2001; Papafragou, 2003). This finding is also difficult to explain with the viewpoint that a logical response is only indicative of insensitivity to the implicature (Noveck, 2001; Bott & Noveck, 2005; Guasti, 2005). If this is the case it is hard to explain why adults, who are considered to have more cognitive resources, should be less sensitive to the implicature than the children. The finding is best accounted for by Feeney et al.'s suggestion that a logical response can also represent the cancellation of the implicature, and it is the cancellation that requires effort. This point is taken up and focused on in Chapter 4.

The results from the statement and storyboard tasks for the *infelicitous or* trials show that children are capable of detecting the implicature associated with the connective. Logical response rates of 35% for the Year 1 children and 54% for the Year 4 children in the storyboard task were significantly lower than the logical response rates of 79% and 78% respectively in the statement task. Thus children were far less likely to accept a statement such as *I've spilt the Coco-Cola or the Irn Bru* when they had seen that the speaker had spilled both, than they were to accept a statement such as *the snow is cold or white.*

However consideration of the control statements reveals that the picture for *or* is not as straightforward as it was for *some*. As was pointed out in the introduction to this experiment, when considering the interpretation people give to *infelicitous or*, the interpretation must be considered alongside the responses to the control statements. A high logical response rate to the *infelicitous or* statements combined with a low logical response rate to the *felicitous or* statements suggests that *or* is being interpreted as conjunction rather than disjunction. Regardless of whether one is willing to accept two true disjuncts being present, the presence of only one true disjunct should be accepted, thus logical response rates to *felicitous or* should be near ceiling. Unfortunately in the statement task both the Year 1 and Year 4 children had low rates of logical responding on the *felicitous or* trials, 0% and 28% respectively, and high rates of logical responding on the *infelicitous or* trials, 85% and 78% respectively, suggesting that conjunction was their preferred interpretation. However there is evidence that the errors do not reflect a simple case of interpreting *or* as conjunction. In the storyboard task both the Year 1 and Year 4 children had low rates of logical responding on the *felicitous or* and the *infelicitous or* trials, and this pattern was also seen in the adults, moreover there were significant differences between the logical response rates for the *infelicitous or* and the *true and* trials in both tasks for all the age groups.

The floor effect, seen in the Year 1 children in the statement task, whereby no logical responses were given to the *felicitous or* trials, is particularly worrying. The statements were designed with the youngest participants in mind, and particular consideration was given to the attributes of the noun that rendered the disjunct false. It was important that attributes were chosen which the children would know were not a defining feature of the noun, so to be obviously false. For example, it is well known that elephants have feet and not hooves, but as hooves are the feet of other animals, choosing this attribute may have confused the young children. However this uncharacteristic attribute may also have been responsible for unforeseen problems. There were several reports from the adult participants that it was surprising to be asked whether *an elephant has ears or socks*. Instead of considering the nature of the disjunct the participants said they focused on the unexpected attribute and dismissed the sentence accordingly; it could be the case that the children also did this.

It is also possible that the problems the participants had interpreting *or* could have influenced their responses to *some*, as both groups of children had lower than expected logical response rates to the *felicitous some* trials, 73% for the Year 1 children and 78% for the Year 4 children. The rate for the Year 4 children seems low when compared to the 90% rate found by Feeney et al. (2004). Because of concerns that the statement methodology might not be ideal for studying the implicature associated with *or*, the remaining experiments concentrate solely on investigating the scalar implicature associated with *some*.

2.4. Experiment 2: Sensitivity to scalar implicatures in Year 1, Year 4, Year 7, and Year 10 children and adults

2.4.1. Introduction

The finding from Experiment 1, that the Year 4 children were more pragmatic than the adults, was unexpected given the current assumptions in the scalar implicature literature. The main aim of this experiment was to attempt to replicate this finding. A secondary aim was to investigate sensitivity in two groups of adolescents. These two age groups, Year 7 and Year 10 children warrant attention for three reasons: First, investigations into scalar implicature have tended to focus on children under 11, therefore the developmental trajectory between this age and adulthood is unknown. Second, studies comparing one group of children to a group of adults cannot be considered to be truly developmental. They do provide a snap shot of what is happening at a particular age, but unless a range of age groups are tested on the same task it is difficult to say how development proceeds. Finally, if it does turn out that children are more pragmatic than adults, then the response patterns of the adolescents may help to formulate an hypothesis as to why that might be the case, as studies involving adolescent participants have been informative for theories of reasoning (Jacobs & Klaczynski, 2002; Klaczynski, 2001, 2004; Klaczynski & Cottrell, 2004).

2.4.2. Method

Participants

One hundred and nine people participated in the experiment. These were split into five age groups: twenty four Year 1 children, age range 5 years 9 months to 6 years 9 months; twenty three Year 4 children, age range 8 years 9 months to 9 years 9 months; twenty two Year 7 children, age range 11 years 10 months to 12 years 10 months; twenty one Year 10 children, age range 14 years 10 months to 15 years 10 months; and nineteen adult participants, age range 20 to 55 years. The Year 1, Year 4, and Year 7 groups were recruited from two schools in the North East of England. The Year 10 group was recruited from a local school and a youth group. The adult participants were recruited at Durham University. The head teachers gave consent for the children’s participation. The adults gave their own consent.

Materials

Participants heard a list of 24 statements and saw a series of 16 storyboards. The materials were identical to those used for the quantifiers *all* and *some* in Experiment 1 (see Appendix AI). As in Experiment 1, two sets of statement materials were prepared and the statements within each set were randomly ordered. The same was done with the storyboard materials. There were two pairings of statements and storyboard materials and the order of presentation was counterbalanced.

Procedure

Participants received instructions on how to complete each task immediately before they attempted that task. For the statement task participants were told that they would hear a series of statements and they had to decide whether each was true or false. For the storyboard task participants were told that they would see a number of photographs and hear accompanying text, and they had to decide whether the response made by the child was true or false. For both tasks participants had an answer sheet on which to record their responses. All the participants were tested in small groups and the tasks were completed in one sitting.

2.4.3. Results

As in the previous experiment the participants' responses were coded with respect to their logical correctness. Before the data were analysed the scores of two Year 1 children, one Year 4 child, four Year 10 children and two adults who all scored at or below chance levels in one of the *true all* conditions were removed. The mean scores and standard deviations for the remaining 100 participants according to term, truth, task and age can be seen in Table 2.6. Each participant scored 7 or 8 on the absurd statements and this result was not used in any further analysis.

Table 2.6.

Mean number of logical responses (and standard deviations) according to term, truth, task and age for Experiment 2.

All						
Truth/ Felicity	Task	Age				
		Yr 1	Yr 4	Yr 7	Yr 10	Adult
		n = 22	*n* = 22	*n* = 22	*n* = 17	*n* = 17
True/ Felicitous	Statements	3.91	4.00	3.86	3.76	3.88
		(0.29)	(0.00)	(0.35)	(0.44)	(0.33)
	Storyboards	3.95	3.86	4.00	3.88	3.82
		(0.21)	(0.35)	(0.00)	(0.33)	(0.39)
False/ Infelicitous	Statements	3.91	3.91	4.00	4.00	3.94
		(0.29)	(0.29)	(0.00)	(0.00)	(0.24)
	Storyboards	3.73	3.95	3.91	3.94	3.59
		(0.63)	(0.21)	(0.29)	(0.24)	(1.00)
Some						
Truth/ Felicity	Task	Age				
		Yr 1	Yr 4	Yr 7	Yr 10	Adult
		n = 22	*n* = 22	*n* = 22	*n* = 17	*n* = 17
True/ Felicitous	Statements	3.73	3.95	3.95	4.00	4.00
		(0.46)	(0.21)	(0.21)	(0.00)	(0.00)
	Storyboards	3.59	3.68	4.00	3.94	3.94
		(0.50)	(0.48)	(0.00)	(0.24)	(0.24)
False/ Infelicitous	Statements	3.95	1.95	0.18	1.35	2.18
		(0.21)	(1.84)	(0.50)	(1.66)	(1.63)
	Storyboards	0.95	0.27	0.14	1.35	1.35
		(1.36)	(0.55)	(0.35)	(1.84)	(1.62)

Note: The maximum number of correct logical responses for each statement type was 4.

Statement data

Collapsing results across the age groups, significantly fewer logical responses were given to the *some* statements (mean 5.86) than to the *all* statements (mean 7.84), Wilcoxon matched pairs signed ranks test, $z = 7.17$, $p < .001$, $r = .51$ and to the false/infelicitous statements (mean 5.89) compared to the true statements (mean 7.81), Wilcoxon matched pairs signed ranks test, $z = 7.05$, $p < .001$, $r = .50$. These results also held for all the individual age groups, *p's* < .01, *r's* ≥ .49, except the Year 1 children, $p > .05$ for both cases (see appendix BII, Tables 1 & 2, for statistics for each age group).

The number of logical responses given to each type of statement by each age group was analysed using Friedman tests, and where significant these were followed up with Wilcoxon signed ranks tests. This analysis revealed that except for the Year 1 children, χ^2 (3) = 5.71, $p = .127$, all other age groups gave significantly fewer logical responses to the *infelicitous some* statements compared to any other statement type, all *p's* ≤ .003, all *r's* ≥ .49 (see appendix BII, Tables 3 & 4, for statistics for the comparison of each statement type for each age group).

Storyboard data

Collapsing across age groups, significantly fewer logical responses were given to the *some* storyboards (mean 4.58) than to the *all* storyboards (mean 7.74), Wilcoxon matched pairs signed ranks test $z = 8.32$, $p < .001$, $r = .59$, and to the false/infelicitous storyboards (4.59) compared to the true/felicitous storyboards (mean 7.73), Wilcoxon matched pairs signed ranks test, $z = 8.51$, $p < .001$, $r = .60$. These results also held for all the individual age groups, $p < .01$, $r \geq .48$ (see appendix BII, Tables 5 and 6, for statistics for each age group).

The number of logical responses given to each type of storyboard was analysed using Friedman tests and these were found to be significant for all age groups, all *p's* $< .001$, all *r's* $\geq .49$ (see appendix BII, Table 7, for statistics for each age group). This analysis was followed up with Wilcoxon signed ranks tests. This analysis revealed that all age groups gave significantly fewer logical responses to the *infelicitous some* storyboards compared to any other storyboard type, all *p's* $\leq .005$, all *r's* $\geq .48$ (see appendix BII, Table 8, for statistics for the comparison of each storyboard type for each age group).

Infelicitous some data

The nature of the distribution of the *infelicitous some* data can be seen in Figure 2.7. The distributions were not normal in either task for any age group. In the statement task two of the age groups, Year 1 and Year 7, had unimodal distributions, although they were polarized in their modes, whilst the other three age groups were bimodal. In the storyboard task all the distributions for the age groups were unimodal on zero.

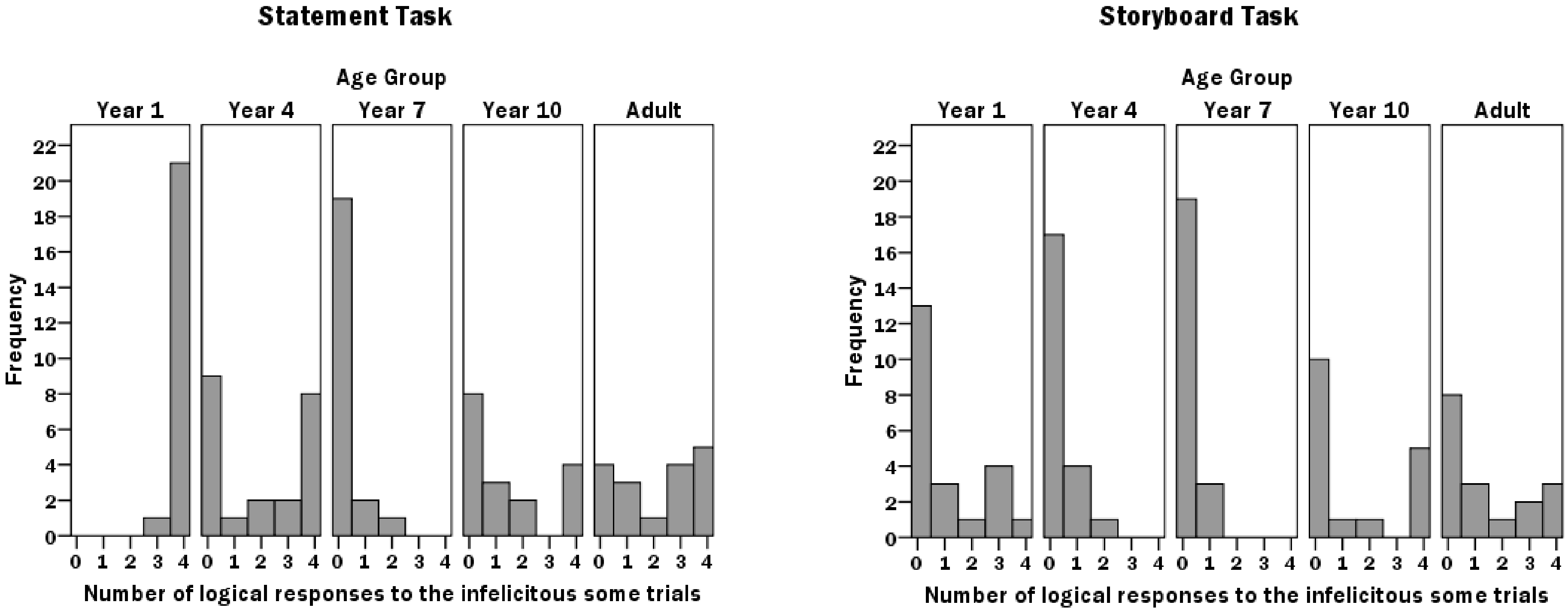


Figure 2.7. Distribution of participants according to the number of logical responses given to the infelicitous some trials, according to age and task, for Experiment 2.

With regards to the *infelicitous some* statements, the means for which can be seen in Figure 2.8, a Kruskal-Wallis test revealed that the number of logical responses significantly differed according to age, $H(4) = 48.20$, $p < .001$.

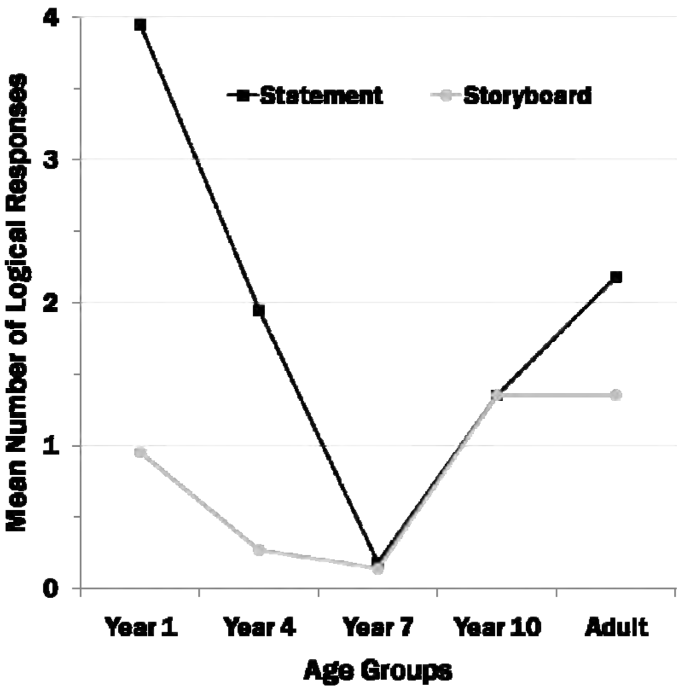


Figure 2.8. Mean number of logical responses to the *infelicitous some* trials for each age group according to task in Experiment 2.

Comparing each age group with its neighbour, Mann-Whitney tests revealed that the Year 1 children gave significantly more logical responses than the Year 4 children, $U = 93$, $p < .001$, $r = .63$, these Year 4 children gave significantly more logical responses than the Year 7 children, $U = 115.5$, $p = .001$, $r = .52$, but these Year 7 children gave significantly fewer logical responses than the Year 10 children, $U = 107$, $p = .005$, $r = .45$, who did not differ significantly from the adults, $U = 105.5$, $p = .164$, $r = .24$. In addition, a Mann-Whitney test revealed that the number of logical responses given by the Year 7 children was significantly fewer than the number given by the adults, $U = 56.5$, $p < .001$, $r = .67$.

With regards to the *infelicitous some* storyboards, the means for which can also be seen in Figure 2.8, a Kruskal-Wallis test revealed that the number of logical responses differed significantly according to age, $H(4) = 11.80$, $p = .019$. Comparing each age group with its neighbour, Mann-Whitney tests revealed that the Year 7 and Year 10 groups differed from each other, $U = 126.5$, $p = .025$, $r = .36$ 9 (see appendix BII, Table 9, for other age group comparison statistics). Further analysis also revealed that both the Year 4 children, Mann-Whitney tests, $U = 117.5$, $p = .021$, $r = .37$, and the Year 7 children, $U = 104.5$, $p = .004$, $r = .46$, gave significantly fewer logical responses than the adults.

In relation to differences in responding between tasks and collapsing across age groups, a Wilcoxon matched pairs signed ranks test revealed that significantly fewer logical responses were given to the *infelicitous some* trials in the storyboard task compared with the statement task, $z = 5.61$, $p < .001$, $r = .40$. The effect of task for each age group can be seen in Figure 2.8. Significantly fewer logical responses were given to the *infelicitous some* trials in the storyboard task by the Year 1 children, Wilcoxon matched pairs signed ranks test $z = 4.05$, $p < .001$, $r = .61$, and by the Year 4 children, $z = 3.23$, $p \leq .001$, $r = .49$, but there was no significant effect of task for the other age groups, all *p's* > .05 (see appendix BII, Table 10, for statistics for each age group).

The nature of the pattern of responding between the tasks is summarized in Table 2.7.

Table 2.7

Number of participants for each pattern of responding according to age group in Experiment 2

	Statement > Storyboard	Statement = Storyboard	Statement < Storyboard
Year 1	22	2	0
Year 4	13	8	1
Year 7	2	17	3
Year 10	3	11	3
Adult	8	8	1

Note: The column headings show the relationship between the number of logical responses given to the infelicitous some trials in both of the experimental conditions.

As can be seen from Table 2.7 the pattern of responding was different for each age group; although one shared feature was that for all age groups only a minority of participants, if any, responded more logically to the storyboards as compared with the statements. Although not apparent from the table, there were different response patterns between the age groups for those participants who gave the same number of responses in both tasks. The Year 1 same responders were all logical, the Year 4 and Year 7 same responders were all pragmatic, whereas the Year 10 and the adult same responders were a mixture of logical and pragmatic responders.

2.4.4. Discussion

The results of the experiment attest to the reliability of the finding that children can be more pragmatic than adults. Not only were the Year 4 children more pragmatic than the adults in the storyboard task, as seen in Experiment 1, but the Year 7 children were more pragmatic than the adults in both tasks.

In the statement task logical responding dominated the Year 1 age group after which it declined to the point of near extinction in the Year 7 children, before re-emerging in late adolescence and adulthood. An overall increase in pragmatic responding in the storyboard task obscured the differences between the age groups but as can be seen in Figure 2.8 the developmental trajectory in both tasks was U-shaped. The pattern of responding in the Year 1, Year 4 and Year 7 children followed the current assumption that pragmatic responding increases with age (Guasti et al., 2005; Noveck, 2001; Noveck & Reboul, 2008; Noveck & Sperber, 2007; Papafragou & Musolino, 2003). However the pattern of responding that followed from the Year 7 children through to the adults does not fit with this assumption and is consequently problematic. An explanation for this pattern of responding is given in the Chapter Discussion.

It is also worth noting, given the ease with which the Year 7 children appeared to detect the scalar implicature, that the interpretations given by the adults did not always appear to be effortless. A number of the adults were heard to voice out loud which interpretation they should chose based on what they perceived to be the demands of the task. For instance one participant was heard to say that he couldn’t decide whether he should chose the interpretation that was actually true, i.e. that if Charlotte ate all the sweets then she had actually eaten some of them, or whether he should chose the interpretation that was expected in that type of dialogue, i.e. if someone asks what you have done you should be as informative as possible.

The findings of this experiment are problematic for current pragmatic theories. Default theories (Chierchia, 2004; Gazdar, 1979; Horn, 1972, 2004; Levinson, 1983, 2000) suggest that for adults the detection of a scalar implicature is automatic whilst its cancellation is effortful. Applying the same rationale to children is problematic as to do so appears to necessitate the claim that Year 1 children were able to cancel the implicature at a greater rate than the adults. Relevance theories (Carston, 2002; Sperber & Wilson, 1985/1995) on the other hand propose that effort is involved in processing the scalar implicature. If this is the case then it is hard to explain why the Year 7 children have a higher pragmatic response rate compared to the adults. RT would also find it difficult to account for the adult participants being faced with a choice over which interpretation to make, as this deliberation goes against the principle of optimal relevance. Whether either of the two theories can give an adequate account of scalar implicatures will be addressed in the General Discussion.

2.5. Experiment 3: Sensitivity to scalar implicatures in preschool children

2.5.1. Introduction

This experiment extended the age range of the previous experiments to investigate sensitivity in preschool children, to see whether 3- and 4-year-olds were capable of spontaneously detecting the implicature associated with *some* in tasks that required meta-linguistic judgements. Although researchers have found that training has been required to facilitate the detection of scalar implicatures (Papafragou & Musolino, 2003) and others have suggested that meta-linguistic judgements may be a source of difficulty for young children (Pouscoulous et al., 2007), the high rates of pragmatic responding shown by the 6-year-old children in the storyboard condition of the previous experiments suggests that younger children may also show sensitivity in this task. Preschool children not only possess a wide repertoire of inference making skills before they start school (Buckley, 2003), they are also capable of, and able to detect, deception (Bussey, 1992; Lewis, Stanger & Sullivan, 1989; Sodian, Taylor, Harris & Perner, 1991). As Feeney et al. (2004) claimed that it was the potentially deceptive nature of the protagonist’s utterances in the storyboards task that facilitated detection of the scalar implicature in children, it is possible that the preschool children will also be able to expend the effort necessary to process the implicature, in order to gain the positive cognitive effects associated with the detection of an attempted act of deception.

Thus the aim of this experiment was to investigate whether spontaneous sensitivity to scalar implicatures in tasks requiring a meta-linguistic judgement could be found in children younger than had previously been tested .To this end a group of 3- and 4-year-olds were tested with modified versions of the quantifier materials from Experiment 1. Given the high levels of pragmatic responding seen in the 6-year-olds in the previous experiments of this chapter, the prediction was made that the preschool children would show spontaneous sensitivity in the pictureboard task, a task equivalent to the storyboard task of the previous experiments, but not in the statement task.

2.5.2. Method

Participants

Twenty three children, aged between 3 years and 4 years 6 months (mean age 3 years 10 months) participated in the study. They all attended a school nursery in the North East of England. Consent for their participation was given by the head teacher.

Materials

The children heard a list of 24 statements and saw a series of 16 pictureboards with an accompanying soft toy.

Statements. The 24 statements took the same form as those used in the previous experiments (see appendix AIII for the statements used). Each statement was presented with both *all* and *some* preceding it, although each participant received only one form, which resulted in two sets of materials. In each set the *all* and *some* statements were blocked, this was to minimise possible confusion in the children as to what quantifier they were judging. The order of the quantifier blocks was counterbalanced which resulted in four conditions. The order of the statements within each block was randomized.

Pictureboards. The 16 pictureboards were a simplified version of the storyboards used in the previous experiments. It was thought that the number of photographs and amount of text in the storyboards might be too demanding for the young children in terms of the amount of cognitive effort required to maintain concentration. To control for this the number of photographs was reduced to two. An additional aid to concentration involved replacing the model in the story with a soft toy. The soft toy was present in the testing room and the child was able to hold it whilst they evaluated the pictureboard. An example pictureboard is shown in Figure 2.9.

Figure 2.9. Example of pictureboard used in Experiment 3.

As can be seen from Figure 2.9 each pictureboard contained two photographs which were placed back-to-back; the initial photograph showing a soft toy with three objects and the reverse showing the same soft toy and the result of the toy's actions. There were four themes, drinking juice, spilling juice, eating sweets and eating cake. Each theme had four forms consistent with the *true all*, *false all*, *felicitous some* and *infelicitous some* trials of the storyboards in the previous experiments.

A different toy was used in each of the pictureboards and the children saw each of the four forms of the pictureboards. The pictureboards were blocked according to the *some/all* response of the toy and the order of presentation of the blocks was counterbalanced amongst the children.

Each set of pictureboards was paired with a set of statements, and the order of presentation was counterbalanced amongst the children.

Procedure

Immediately prior to the task commencing each child was assessed to see if they had an understanding of what *true* meant. This assessment consisted of asking the children a number of questions, such as, *if I said you had an elephant sitting on your head, would that be true*? All of the children, who were assessed individually, appeared to understand the term and gave answers of yes for a true state of affairs and no for untrue.

For the statement task each child was told that they would hear a sentence and if they thought that the sentence was true they should say yes but if they thought that it was not they should say no. The researcher recorded the children's answers.

For the pictureboard task the pictureboards were arranged in a semicircle with the accompanying soft toy placed behind the pictureboard. The researcher introduced the child to the soft toy and told the child about a job that had been given to the toy, which the child would see in the photographs. The researcher then asked the toy about the task it had been given and invited the child to turn the photograph over. The researcher then informed the child of the toy's response and the child responded with a yes if they thought that the toy's response was true and no if it was not; this response was recorded by the researcher. For example:

Researcher (pointing at photograph)	Bat I gave you three bottles to mind, what did you do with them?
Researcher (to child)	Shall we look to see what Bat did with them?
Child turns the pictureboard over	
Researcher	Bat says he drank some of the bottles.

The children did not always answer spontaneously and sometimes the final sentence had to be repeated.

All the children completed the tasks on an individual basis and in one sitting.

2.5.3. Results

As in the previous experiments the children's responses were coded with respect to their logical correctness. The data from four children was removed at this point; two for low correct responding on the absurd statements[5] and two for low correct responding on the *false all* pictureboards[6].

Examination of the remaining data revealed that the majority of responses in the tasks were yes/no but there was a significant difference in the type of response between the tasks with the children giving more yes/no responses in the statement

[5] Low correct responding on the absurd statements could be indicative of a positive response bias.
[6] Low correct responding on the *false all* statements could be due to lack of experience, but on the pictureboards it could be indicative of failure to consider the quantifier.

task, Wilcoxon matched pairs signed ranks test, $z = 3.19$, $p = .001$, $r = .52$. The mean percentages for the two tasks can be seen in Table 2.8.

Table 2.8

Type of response, expressed as a percentage, given in Experiment 3

	Task	
Type of Response	Statements	Pictureboards
Yes/No	99.0	85.5
Other	1.0	14.5

The other response category took two forms, those which could be easily interpreted as meaning no, such as *he didn't*, and those where no interpretation was possible, such as where the child merely stated a numeral[7]. Only those responses which could be interpreted to mean yes/no were included in the analysis, which resulted in two further children being excluded as they had no score for the *felicitous some* trials in the pictureboard condition. Consequently analysis was carried out on the proportions of answers given for the remaining 17 children. The mean proportion of logical responses according to truth, term and task are shown in Table 2.9. With regards to the absurd statements all of the children correctly judged all of the statements to be false and this finding was not included in any further analysis.

[7] The vast majority of this type of response occurred in the *felicitous some* pictureboard condition.

Table 2.9

Mean proportion (standard deviations) of logical responses, according to truth and term and task, for Experiment 3

(n = 17)	Statements	Pictureboards
True All	.92 (0.19)	.97 (0.12)
False All	.40 (0.36)	.96 (0.09)
True Some	.83 (0.26)	.84 (0.34)
Infelicitous Some	.86 (0.22)	.46 (0.46)

Statement data

The children gave significantly fewer logical responses to the *all* statements (mean proportion .68) than to the *some* statements (mean proportion .88), Wilcoxon matched pairs signed ranks test, z = 2.35, p = .019, r = .40 and to the false/infelicitous statements (mean proportion .66) than to the true/felicitous statements (mean proportion .90), Wilcoxon matched pairs signed ranks test, z = 3.37, p = .001, r = .58.

The number of logical responses given to each type of statement was analysed using a Friedman test, and this was significant, χ^2 (3) = 20.90, p > .001. Follow up Wilcoxon signed ranks tests revealed that the children gave significantly fewer logical responses to the *false all* statements compared to any other statement type, all *p's* $\leq$ .004, all *r's* $\geq$ .50 (see appendix BIII, Table 1, for statistics for the comparison of each statement type).

Pictureboard data

The children gave significantly fewer logical responses to the *some* pictureboards (mean proportion .65) than to the *all* pictureboards (mean proportion .96), Wilcoxon matched pairs signed ranks test, z = 3.24, p = .001, r = .56. The difference in the number of logical responses to the false/infelicitous pictureboards (mean proportion .73) compared to the true/felicitous pictureboards (mean proportion .89) was approaching significance, Wilcoxon matched pairs signed ranks test, z = 1.81, p = .070, r = .31.

The number of logical responses given to each type of pictureboard was analysed using a Friedman test, and this was significant, χ^2 (3) = 14.22, p = .003. Follow up Wilcoxon signed ranks tests revealed that the children gave significantly fewer logical responses to the *infelicitous some* pictureboards when compared with the *true all* and *false all* pictureboards, all *p's* $\leq$ .004, all *r's* $\geq$ .50 and the comparison with the *felicitous some* pictureboards was approaching significance, z = 1.82, p = .069, r = .31 (see appendix BIII, Table 2, for statistics for the comparison of each pictureboard type).

Infelicitous some data

The nature of the distribution of the *infelicitous some* data can be seen in Figure 2.10.

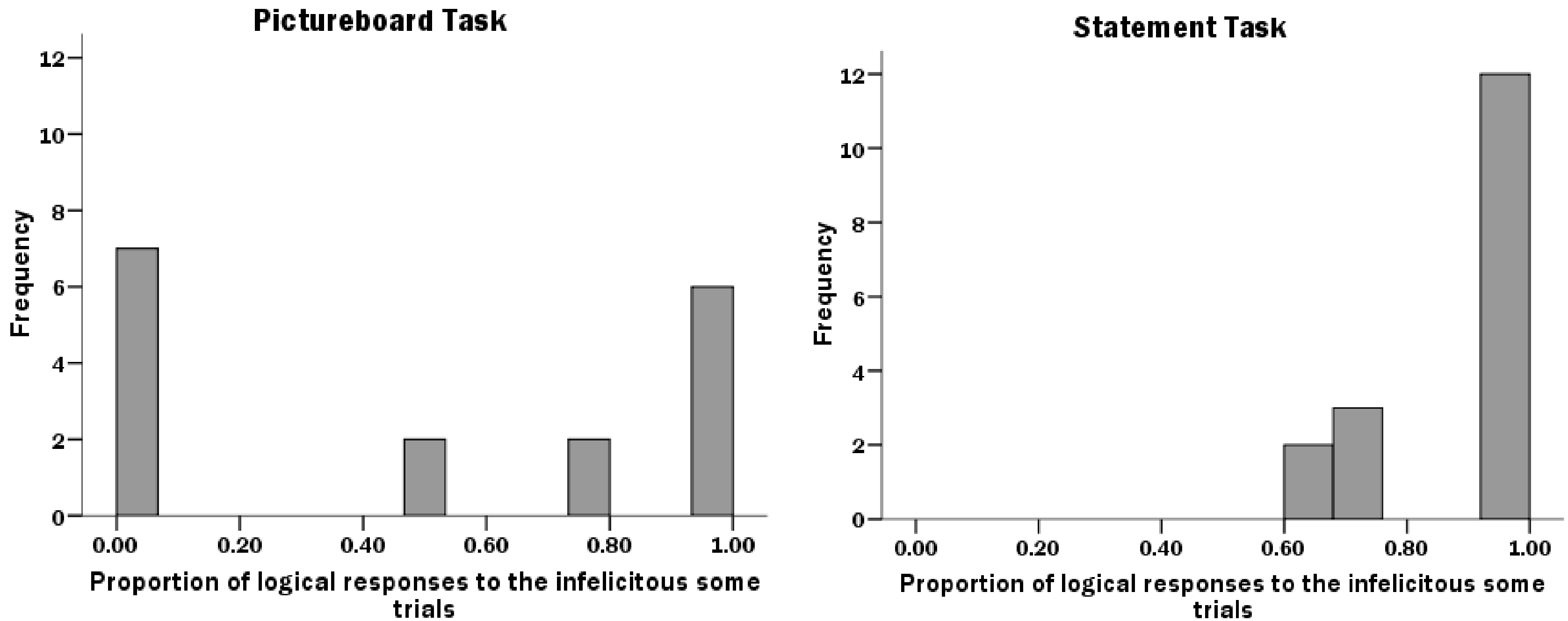


Figure 2.10. Distribution of participants according to the proportion of logical responses given to the infelicitous some trials in Experiment 3.

In the statement task the distribution of children was unimodal; twelve children always gave a logical response and five children gave a logical response the majority of the time. However in the pictureboard task the distribution was bimodal; seven children never gave a logical response, six children always gave one and four children were equivocal.

With regards to the number of logical responses given to the *infelicitous some* trials between tasks, significantly fewer logical responses were given in the pictureboard task than in the statement task, Wilcoxon matched pairs signed ranks test, z = 2.49, p = .013, r = .43. Performance on neither task depended on whether it was attempted first or second, Mann Whitney tests, statement task, U = 35, p = 1.000, r = 0.00, pictureboard task, U = 28, p = .460, r = .13.

Examination of the nature of the pattern of responding between tasks revealed that ten children gave more logical responses in the statement task, five gave the same number in both tasks, whilst only two gave more logical responses in the pictureboard task.

2.5.4. Discussion

The prediction made at the start of the experiment, that nursery children would show spontaneous sensitivity in the pictureboard task but not in the statement task was confirmed. This experiment suggests that at least some preschool children can spontaneously detect the implicature associated with *some* even when they have to make a meta-linguistic judgement. Although the children in this experiment overwhelmingly accepted the *infelicitous some* statements as true, the logical response rate in the pictureboard task dropped to under 50%. The finding also supports the suggestion made in the previous experiments that sensitivity is not an all or nothing affair; whilst sensitivity is developing it is not necessarily seen in all contexts. Not one of the children showed sensitivity in the statement task, implying that detection is context dependent and not readily available if the context is not supportive.

However, two problems were encountered with the materials:

First, although the statement condition produced near ceiling yes/no responses, the *false all* control statements were not answered as well as anticipated. One reason for this could be a lack of world knowledge. As the statements task involves retrieval of encyclopaedic knowledge, judging a statement such as *all dogs have spots* to be false requires the retrieval of counterexamples, and finding counterexamples can be a difficult process for children (Beck, Robinson, Carroll & Apperly, 2006). The nouns used in the statements were chosen because it was thought that the children would be familiar with them, however it could be that the children possessed a limited knowledge of the nouns and this meant that a counterexample was difficult to find.

Within the picture board condition the problem was one of coding; the children gave answers which could not always be coded easily. Problems with the data arose in the *felicitous some* condition, when the children answered *two* instead of giving a yes/no response. Although this response is correct in that it describes how many objects were interacted with, the fact that they did not give the answer in the required format meant that this data could not be used, as there was no of knowing whether they regarded the sentence as true or false. With regards to the difference in logical responses rates between the target and control pictureboards, this loss of data could be the reason that the difference between the *infelicitous* and *felicitous some* trials did not reach significance, even though there was a moderate effect size. Whilst this lack of significance could indicate that the children's understanding of the term *some* is tenuous, interestingly these children also tended to respond pragmatically to the *infelicitous some* trials, which would suggest that whilst the children may not know what *some* means they know it does not mean *all*. This in turn suggests that the children's rate of sensitivity was underestimated. One possible reason for the reduction of yes/no responses in this condition could be that the pictureboards are more cognitively demanding than the statements, in that they require the children to concentrate for a longer period of time and remember information. Although the condition was designed to facilitate detection it could be that the children lose concentration and are unsure of giving a definitive answer.

To overcome these problems a second experiment was designed which featured a condition which involved the children more directly.

2.6. Experiment 4: Further investigation into sensitivity to scalar implicatures in preschool children

2.6.1. Introduction

Although the results of the previous experiment suggested that preschool children were able to detect scalar implicatures, the results were not as conclusive as one would hope for. It was suggested that one of the problems in the last experiment could be due to the children failing to maintain engagement with the materials. Given that children a few years older than the preschoolers had no problem giving yes/no responses in a condition requiring more concentration (Experiment 2 of this thesis), it was thought that a task that directly involved the children, given their egocentric nature (Piaget, 2001; Piaget & Inhelder, 2000), may overcome this problem. Bearing this in mind, one of the characteristics that the pictureboards and storyboards from the previous experiments in this thesis have in common is that the main character could be said to be intending to deceive. Relevance Theory (Sperber & Wilson, 1985/1995) would predict that the cognitive gains from detecting deception would be worth the effort of the extra processing. If this is the case, then a task that involved the direct deception of the child participants may increase the cognitive effects to be gained from the detection of the deception and thus increase sensitivity to scalar implicatures in those children. To this end a new set of stimulus materials was introduced, consisting of a set of paper bags and their contents. It was predicted that the rate of logical responding to the *infelicitous some* trials in this new condition would be significantly different from the control trials.

2.6.2. Method

Participants

Twenty one children, aged between 3 years and 4 years 6 months (mean age 3 years 10 months) participated in the study. They all attended a nursery in the North East of England. The head teacher gave consent for the children's participation.

Materials

The children were tested with sixteen pictureboards and sixteen bags.

Pictureboards. The pictureboards were the same as those used in Experiment 3. As in that experiment they were blocked according to the *some/all* response of the toy and the order of presentation of the blocks was counterbalanced amongst the children.

Bags. Each of the bags was made of brown paper and contained three objects of the same type, such as balls or chocolates. The bags took the same form as the pictureboards, regarding quantifier and felicity, through the manipulation of how many objects the researcher took and how many she claimed to have taken:

true all - researcher takes 3 objects and claims to have taken *all* *false all* - researcher takes 2 objects and claims to have taken *all*

felicitous some - researcher takes 2 objects and claims to have taken *some*

infelicitous some - researcher takes 3 objects and claims to have taken s*ome*

The children saw four of each form of the bags, thus there were four *true all*, four *false all*, four *felicitous some* and four *infelicitous some* trials. A different set of toys or sweets was used for each trial (see Appendix AIV for objects used).The trials were blocked according to the *some/all* claim of the researcher and block order was counterbalanced amongst the children.

Each set of pictureboards was paired with a set of bags, and condition order was counterbalanced.

Procedure

As in Experiment 3 each child was assessed prior to the first task taking place to see if they had an understanding of what *true* meant. This assessment consisted of asking the children the same questions as in that experiment, such as *if I said you had an elephant sitting on your head, would that be true*? All of the children, who were assessed individually, appeared to understand the term and gave answers of yes for a true state of affairs and no for untrue.

The procedure for the pictureboard task was the same as Experiment 3. Thus the pictureboards were arranged in a semicircle with the accompanying soft toy placed behind the pictureboard. The researcher introduced the child to the soft toy and told them about a job that had been given to the toy, which the child would see in the photographs. The researcher then asked the toy about the task it had been given and invited the child to turn the photograph over. The researcher then informed the child of the toy’s response and the child responded with a yes if they thought that the toy's response was true and no if it was not; this response was recorded by the researcher.

For the bags task, each child was shown the 16 bags and told that they would have the opportunity to either talk about or play with the contents of each bag. After this interaction had taken place the researcher ensured that the child was holding all the items from the bag. The researcher then took either two or three items from the child and made a statement about the number of items taken. The child's task was to decide whether the researcher's statement was true. For example:

Researcher Shall we look and see what's inside this bag?

(Bag is handed to child)

Researcher Do you know what they are?

(to child)

Child Bouncy balls.

Researcher That's right. Shall we bounce them?

(The balls are bounced and the researcher engages the child in conversation about the colour of the balls, which one bounced the highest, etc.)

Researcher Would you hold the balls for me?

(Researcher ensures that child is holding all the bouncy balls)

Researcher I'm going to take this bouncy ball from you.

(Researcher takes bouncy ball from child. This step is repeated either once or twice)

Researcher I've taken some of the bouncy balls. Is this true?

In order to avoid fatigue the children were tested individually in two sittings, which were one week apart. They saw a set of sixteen pictureboards in one sitting and a set of sixteen bags in the other.

2.6.3. Results

As in the previous experiments the children's responses were coded with respect to their logical correctness. The data from one child was removed at this point for low correct responding on the *felicitous some* trials in both tasks.

Examination of the remaining data revealed that the majority of responses in the tasks were yes/no but the children sometimes answered with a different response, as in Experiment 3. The difference in the type of response between the tasks was approaching significance, Wilcoxon matched pairs signed ranks test, z = 1.79, p = .074, r = .28, with the children giving more yes/no responses in the bag task. The mean number of responses for the two tasks can be seen in Table 2.10.

Table 2.10

Type of response given in Experiment 4, expressed as a percentage

	Task	
Type of Response	Pictureboards	Bags
Yes/No	86.25	96.56
Other	13.75	3.44

The same criteria were applied to accepting those responses that could be interpreted to mean yes/no, as in Experiment 3, which resulted in the loss of data from four children as they had no score for the *felicitous some* trials in one of the tasks. Consequently analysis was carried out on the proportions of answers given for the remaining 16 children. The mean proportions of logical responses according to truth, term and task are shown in Table 2.11.

Table 2.11

Mean proportion (and standard deviations) of logical responses according to truth, term and task for Experiment 4

(n = 16)		Pictureboards	Bags
True/ Felicitous	All	.98 (0.06)	.98 (0.06)
	Some	.81 (0.30)	.92 (.16)
False/ Infelicitous	All	.77 (0.34)	.91 (0.26)
	Some	.55 (0.43)	.45 (0.45)

Pictureboard data

In the pictureboard task the children gave significantly fewer logical responses to the *some* trials (mean .68) than to the *all* trials (mean .88), Wilcoxon matched pairs signed ranks test, z = 2.02, p = .043, r = .36, and to the false/infelicitous trials (mean .66) compared to the true/felicitous trials (mean .90), Wilcoxon matched pairs signed ranks test, z = 2.58, p = .010, r = .46. The number of logical responses given to each trial was analysed using a Friedman test, and this was significant, χ^2 (3) = 11.92, p = .008. Follow up Wilcoxon signed ranks tests revealed that the children gave significantly fewer logical responses to the *infelicitous some* trials when compared with the *true all* trials, z = 2.85, p = .004, r = .50 but not to the *false all* or the *felicitous some* trials, all p's $\geq$ .114, all r's $\leq$ .24 (see appendix BVI,

Table 1 for statistics for the comparison of the target and control trials for the pictureboard task).

Bag data

For the bag task the children gave significantly fewer logical responses to the *some* trials (mean .68) than to the *all* trials (mean .95), Wilcoxon matched pairs signed ranks test, $z = 2.74$, $p = .006$, $r = .48$ and to the *false/infelicitous* trials (mean .68) compared to the *true/felicitous* trials (mean .95), Wilcoxon matched pairs signed ranks test, $z = 2.99$, $p = .003$, $r = .53$. The number of logical responses given to each trial was analysed using a Friedman test, and this was significant, χ^2 (3) = 22.03, $p > .001$. Follow up Wilcoxon signed ranks tests revealed that the children gave significantly fewer logical responses to the *infelicitous some* trials than to any other trial, all *p's* ≤ .019, all *r's* ≥ .42 (see appendix BIV, Table 2 for statistics for the comparison of the target and control trials for the bag task).

Infelicitous some data

The nature of the distribution of the *infelicitous some* data for both tasks can be seen in Figure 2.11. In the pictureboard task responses were evenly distributed at the ends and middle of the scale; five children never gave a logical response, seven children always, or nearly always, gave one, and four children were equivocal. Whilst in the bag task the distribution was bimodal with nine children never, or rarely, giving a logical response and seven children always, or nearly always, giving a logical response. The nature of the pattern of responding between tasks was that eight children gave more logical responses in the pictureboard task, two gave the same number in both tasks, whilst six gave more logical responses in the bag task.

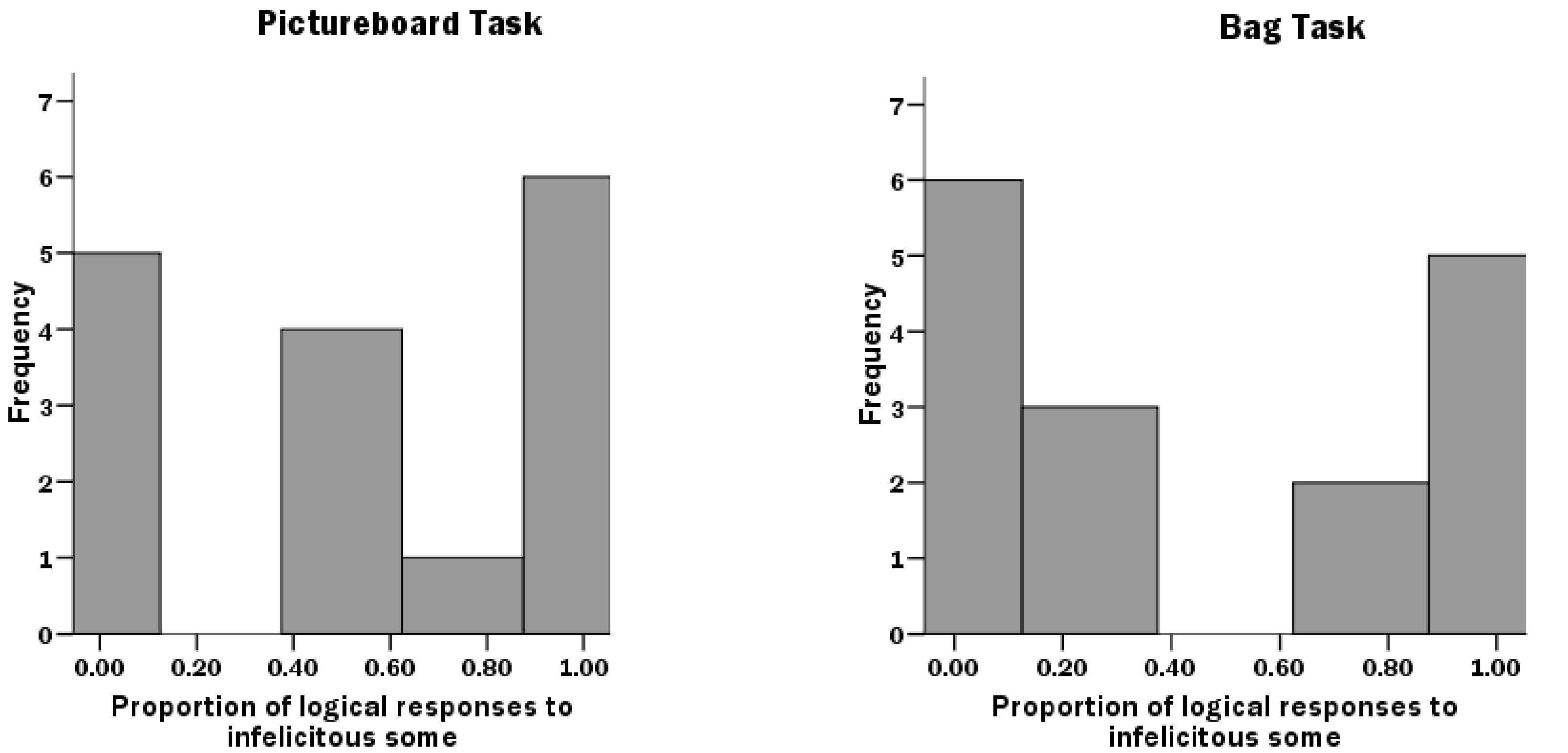


Figure 2.11. Distribution of participants according to the proportion of logical responses given to the infelicitous some trials for the pictureboard and bag tasks in Experiment 4.

Analysis revealed that there was no significant difference in the proportion of logical responses given to the *infelicitous some* trials between tasks, Wilcoxon matched pairs signed ranks test, $z = 0.57$, $p = .569$, $r = .10$, nor was there any significant difference in the proportion of logical responses to the *infelicitous some* trials based on the order in which the tasks were given, $U = 28$, $p = .826$, $r = .04$. However there was a significant practice effect, with the children giving fewer logical responses to the *infelicitous some* trials in the task that was presented second (mean .31), compared to the task that was presented first (mean .69), Wilcoxon matched pairs signed ranks test, $z = 2.28$, $p = .023$, $r = .40$. Additional analysis of this practice effect revealed that the children gave significantly fewer logical responses to the *infelicitous some* trials than to any other trial for the task they received second, Wilcoxon signed ranks tests, all p's $\leq .004$, all r's $\geq .51$, however in the task they received first the *infelicitous some* trials only differed significantly from the *true all* trials, Wilcoxon signed ranks test, $z = 2.39$, $p = .017$, $r = .60$ (see appendix BIV, Table 3 for statistics showing the comparison of target trials to control trials for each task).

2.6.4. Discussion

The prediction made at the start of the experiment, that children would give fewer logical responses to the target trials compared to the control trials in the bag task was confirmed. This finding helps develop the current understanding of sensitivity to scalar implicatures in very young children in a number of ways. It shows that sensitivity can occur spontaneously, without the need for training, and that sensitivity can be seen in tasks that require a meta-linguistic judgement. Furthermore, this result was achieved in an experiment that set up conditions of under-informativeness, where the children displayed a good understanding of *all* and *some*.

One unexpected finding of this experiment was the effect of practice; regardless of which task came first the majority of children showed increased sensitivity on the task that they completed second. Explicit training has been shown to have a beneficial (Papafragou & Musolino, 2003), if somewhat limited (Guasti et al., 2005), effect on implicature detection. However, the children in this experiment received no information about the goals of the task, nor any feedback on their responses, and there was a delay of one week between the administration of the tasks. Repeated testing on theory of mind tasks has been shown to improve experimental scores over control groups scores (Flynn, 2006) and one explanation put forward is that it is children's own perception of conflicting evidence rather than training from adults that causes them to change their beliefs, and that time is essential in allowing children time to ruminate on this change (Slaughter & Gopnik, 1996). Another possible explanation may lie with the control items. It has been proposed that the control items may add to the complexity of the task for children (Pouscoulous et al., 2007) and that a smaller difference in the number of control to target items may facilitate detection in adults (De Neys & Schaeken, 2007), but it may be the case that in these experiments the control items helped to make the difference between the *felicitous some* and *infelicitous some* trials more salient and the children are able to exploit this difference the second time around.

The findings of this experiment have implications for theoretical accounts of scalar implicature. With regards to the practice effects, both default and relevance based accounts can offer an explanation: For relevance theories the trials in the first task may have highlighted the difference between the *some* and *all* trials, and between the *true* and *false* trials. The children may have been able to make use of this knowledge on the second task, and thus less effort may have been required to see infelicity in the second task. A corollary of this is that if less effort is required, it can result in more effect, thus the children perform more pragmatically in task two. For default theories, again the structured nature of the tasks may have facilitated the learning of the pragmatic meaning of the term, and thus enabled the children to perform more pragmatically in the second task. With regards to the lack of sensitivity that is seen in some children within each task, then again both accounts can offer an explanation; the default account through lack of acquisition, and relevance theory in terms of effort or effects. However both accounts are less convincing for those children who show consistent sensitivity in one task but not the other. If *some* generates a default interpretation it is not clear why it is activated in one context and not the other, when the pragmatic demands of the two tasks are ostensibly the same. Similarly, as sensitivity is seen in both tasks it suggests that the effect is worth the effort, and the effort is not overly demanding, so it is unclear why the implicature is relevant in one task but not the other.

2.7. Chapter Discussion

The general aim of the four experiments in this chapter was to investigate the development of spontaneous sensitivity to scalar implicatures; that is, to establish at what age sensitivity is seen and whether it develops linearly.

The experimental findings show that children as young as 3-years-old can detect scalar implicatures and that the developmental trajectory of sensitivity to scalar implicatures is not linear, it appears to be U-shaped. These findings indicate that some of the current assumptions in the experimental pragmatics literature regarding the development of sensitivity to scalar implicatures are incorrect; young children can detect scalar implicatures even when required to make a meta-linguistic response, and children can be more pragmatic than adults.

Looking at the combined results of Experiments 2 and 3 in Table 2.12 the following picture for the individual age groups emerges.

Table 2.12

Types of responder, expressed as a percentage and broken down by age group, from the combined results of Experiments 2 and 3

Year group	Response Categories				
	Logical/ Logical	Logical/ Pragmatic	Pragmatic/ Pragmatic	Pragmatic/ Logical	Other
Preschool	47	41			12
Year 1	23	73			4
Year 4		41	45		14
Year 7			95		5
Year 10	24		59		18
Adult	24	24	35	5	12

Note: In the response category the pairings show the statement context followed by the storyboard context. Logical ≥ 3 logical responses, Pragmatic ≤ 1 logical response, Other represents responses involving at least one score of 2.

Preschool children are able to detect the scalar implicature but only if the context is supportive, thus no child is pragmatic in both tasks. The same is also true of the Year 1 children, although more children are able to detect the scalar implicature in the storyboard task compared with the previous age group. The Year 4 children are effectively split between those who are reliant on the context to make the implicature and those who can make it regardless of the task. In addition no child in this age group is wholly logical in the storyboard task. The Year 7 children are universally sensitive to the implicature regardless of context. Pragmatic responding regardless of context is also seen in the majority of the Year 10 children, although nearly a quarter of this age group respond logically across both tasks. Adult responses are mixed and no one type of response dominates. This age group also has the only case of a participant responding pragmatically in the statement task and logically in the storyboard task.

Although the U-shaped trajectory may seem surprising, U-shaped development is seen in a number of other fields, such as drawing (Gardner & Winner, 1982), face perception (Carey, 1982), maths (McNeil, 2007; Stavy, Strauss, Orpaz & Carmi, 1982) and language development (Ervin & Miller, 1963; Bowerman, 1982). In fact, such is the prominence of U-shaped development in the learning of the English past tense, that models of human learning (Plunkett & Marchman, 1993; Rumelhart & McClelland, 1986; Westerman, 1998) are judged against their ability to model this phenomenon. In addition, other fields can easily explain the finding that children are more pragmatic than adults, even though it goes against current assumptions regarding scalar implicatures. In the cognitive developmental literature Piaget (2001; Piaget & Inhelder, 2000) proposes that logical thought does not emerge until adolescence, with the development of formal operations. These operations allow a child to move away from reasoning about concrete situations involving step-by-step processes, to reasoning systematically about all logical relations involving abstract concepts. So whilst even young children may appear to be capable of logical thought on tasks that involve situations that they are familiar with (Dias & Harris, 1988; Hawkins, Pea, Glick & Scribner, 1984;

Harris & Nunez, 1996), they fail on tasks which are decontextualised and require abstract thought (Wildman & Fletcher, 1977, Piaget & Inhelder, 2000). Similarly in the reasoning literature dual process theories (Evans & Over, 2004; Sloman, 1996; 2002; Stanovich, 1999; Stanovich & West, 2000; for recent reviews see Evans, 2003; Osman, 2004) suggest that there are two separate but interacting processes for reasoning: Heuristic processes are fast, parallel, automatic, unconscious and driven by context. Analytic processes are slow, conscious, controlled, constrained by working memory capacity and responsible for decontextualised thinking (Stanovich, 1999), which underlies our ability to reason or decide independently of context. As logical thinking is cognitively demanding it is therefore more likely to be carried out by those who are high in cognitive resources.

So how might sensitivity to scalar implicatures develop? Any explanation must be constrained by the pattern of responding that is seen within the individual age groups, as well as comparisons between the age groups to consider when particular types of response emerge and disappear. Bearing these constraints in mind, as well as the U-shaped trajectory, and drawing on research from literatures additional to experimental pragmatics, one possible explanation could be as follows: The detection of the scalar implicature in the two youngest age groups relies on context, thus pragmatic responding is possible but only if the context is supportive. The complete lack of pragmatic/pragmatic responders suggests that the detection of the implicature is not automatic for these age groups. This finding, and the finding that the numbers of logical/logical responders declines with age, suggests that the initial understanding of the scalar term *some* is consistent with a logical reading (Noveck, 2001). It is unlikely that the logical responses of these two age groups are due to the same processes of the logical/logical responders in the Year 10 and adult age groups as there is a complete absence of this type of responder in Years 4 and 7. It is more likely that the preschool and Year 1 children have failed to detect the implicature, rather than that they have detected the scalar implicature and then cancelled it.

As children get older scalar implicatures are more easily detected, evidenced by the increase in pragmatic responding from preschool to Year 4, although some of the children in the latter age group are still reliant on context. The logical responses in the Year 4 group only occur in the statement task suggesting that this response represents a failure to detect the implicature rather than the cancellation of it , for if the children had used logical thought to influence their interpretation then logical responding in the storyboard task would have been expected. However no child consistently gave logical responses in the storyboard task.

By Year 7 there is no evidence of logical responding in either of the tasks. This universal pragmatic responding suggests that the scalar implicature is easily detected, so much so that it could be automatic. Logical responding re-emerges in Year 10 and the children are largely split into logical/logical and pragmatic/pragmatic responders. This re-emergence of logical responding coupled with the lack of effect of context within participants suggests that logical responders in this age group have detected the implicature and cancelled it (Feeney et al., 2004). The relatively small numbers of Year 10 and adult logical/logical responders suggests that this kind of logical thinking is not available to all (Piaget, 2001; Piaget & Inhelder, 2000).

Responses in adulthood are much more varied and it is more difficult to say what a logical response represents in the logical/pragmatic responders. It may be that the adults, or at least some of the adults, have both types of response available and chose an interpretation based on the task, thus a logical response could be indicative of cancellation of the scalar implicature. Or it could be that the task is so easy for the adults that minimal cognitive resources are allocated to the statement task and the implicature goes undetected, whereby a logical response would represent insensitivity to the implicature.

To summarise, it is hypothesised that the initial interpretation of *some* is consistent with a logical reading but as age increases so does sensitivity to its scalar implicature. Initially context is heavily relied upon to detect the scalar implicature but this declines with age to the point where the implicature is so readily detected that pragmatic responding is universal. After this point, logical responding re-emerges; this could be indicative of cancellation of the implicature in some participants, or insensitivity in others.

Chapter Three

Deception Contexts and Scalar Implicatures

3.1. Aim

The effect of context on children's performance in cognitive tasks is well documented (e.g. Cummins, 1995; Donaldson, 1978; McGriggle & Donaldson, 1974; Girotto, Gilly, Blaye & Light, 1989; Schwartz & Goldman, 1996) but the role of context in the facilitation of the detection of scalar implicatures in children is unclear. The aim of the two experiments in this chapter was to investigate claims that have been made regarding how context facilitates sensitivity in children, and in particular Feeney et al.'s (2004) claim that deception contexts facilitate the detection of scalar implicatures.

3.2. Introduction

Whilst initial experimental investigations into the detection of scalar implicatures suggested that children were insensitive to them (Noveck, 2001; Papafragou & Musolino, 2003), it is now known that this is not necessarily the case. Children have demonstrated sensitivity to scalar implicatures in a number of studies (Feeney et al., 2004; Guasti et al., 2005; Pouscoulous, et al., 2007), and in Chapter 2 of this thesis we saw sensitivity in children as young as 3-years-of-age. Three factors that have been implicated in facilitating this increase in sensitivity are shared conversational backgrounds (Guasti et al, 2005), strong discourse support (Hurewitz et al., 2006) and deception contexts (Feeney et al., 2004). Evidence and support for claims that have been made about the importance of these factors is somewhat limited, as was seen in Chapter 1, consequently how contexts facilitate the detection of scalar implicatures remains unclear. As a means of establishing the rationale for the experiments in this chapter the evidence presented in Chapter 1 will be briefly reviewed.

According to Guasti et al. (2005) high rates of logical responding in the statement evaluation task may be due to the fact that the background against which the scalar implicature is evaluated is not common to the researcher and participant. For example, in relation to the statement *some bikes have wheels,* a researcher may intend that the set to be considered is the set of fully functioning bikes. However, a participant may extend the set to include bikes that are broken, and if the statement is evaluated against this extended set then it is felicitous rather than infelicitous. Guasti et al. found that when they used a naturalistic context that made the relevant evidence immediately available, 7-year-old children appeared adult-like in their sensitivity rates to scalar implicatures. The authors claim that this shared background coupled with a naturalistic conversational context sets up informativeness expectations in line with Gricean principles. However, it would appear that whilst a shared conversational background may facilitate sensitivity to scalar implicatures it does not always do so, as other studies have revealed high rates of logical responding even in the presence of this factor (Huang & Snedeker, 2005; Hurewitz et al., 2006; Noveck, 2001; Papafragou & Musolino, 2003; Pouscoulous et al., 2007).

According to Hurewitz et al. (2006), who considered work on scalar implicatures and studies on developmental sentence-processing (eg. Hurewitz, Brown-Schmidt, Thorpe, Gleitman & Trueswell, 2000; Trueswell, Sekerina, Hill & Logrip, 1999), it is not just a shared background that children need, as children fail to derive inferences even when the visual scene supports such interpretations; the conversational discourse needs to set up informativeness expectations (eg. Papafragou, 2003; Papafragou & Tantalou, 2004). Thus, they propose that a shared conversational background is insufficient; children need strong discourse support to derive pragmatic inferences.

Feeney et al. (2004) suggested that children's insensitivity to scalar implicatures in the statement evaluation task may be due to lack of cognitive effects; there is nothing to be gained from drawing the implicature in the statement task, therefore children do not expend the effort required to do so. The authors claimed that the dissociation due to context that they found in children is due to the cognitive effects gained from the use of a deception context in the storyboard task. As the detection of an intention to deceive is beneficial, the children expend the necessary effort required to process the scalar implicature. This claim for the facilitatory role of deception contexts seems plausible given that we supposedly have an evolved ability to detect cheaters (Cosmides & Tooby, 2005) and this type of context has produced high levels of pragmatic responding in children as young as three-years-of age despite requiring a meta-linguistic judgement (see Pouscoulous et al., 2007, for comments on type of responding). However the authors provide no evidence for their claim, and their manipulation of context was confounded with another factor, namely that the storyboards also provide a shared conversational background.

The main aim of the experiments in this chapter was to investigate Feeney et al.'s (2004) claim that deception contexts facilitate the detection of scalar implicatures. As was pointed out above, the authors provided no evidence to support their claim, nor do they provide a definition of deception. Deception is a multi-faceted term meaning different things to different researchers (Hyman, 1989; Masip, Garrido & Herrero, 2004; Mitchell & Thompson, 1986; Zuckerman, DePaulo & Rosenthal, 1981), thus the use of this term does not pinpoint what it is about the storyboard context that results in greater sensitivity to scalar implicatures.

Deception is often seen as synonymous with lying (Ekman, 1989; Lewis, 1993; Masip et al. 2004). Whilst implicatures are not lies in the sense that they are logically true[8] they are considered to be deceptive in the sense that they are under-informative (Hopper & Bell, 1984; Vincent & Castlefranchi, 1981). Under-informativeness though is a feature of any task that involves scalar implicatures, therefore it cannot be this feature that facilitates sensitivity in the storyboard task compared to the statement evaluation task. Feeney et al. (2004) suggest that it is the agent's attempts to mislead an audience, that is Charlotte's attempt to mislead her mother, which results in a greater pragmatic response rate. This suggestion seems plausible given that one feature of the bags task used in Experiment 4, which resulted in sensitivity being seen in 3-year-old children, was the deceptive interaction between the child and the researcher, and this feature is absent in the statement evaluation task.

If sensitivity to scalar implicatures is facilitated by the attempts of an agent to mislead an audience, then the manipulation of this factor should result in differing levels of sensitivity. To this end the first experiment involved manipulating the presence or absence of an agent, whilst the second experiment extended the methodology of the first to include materials designed to manipulate the motivation to deceive.

[8] However see Meibauer, 2005, for a definition of lying that includes not only false assertions but true assertions that carry false implicatures.

3.3. Experiment 5: The effect of an agent in deception contexts on sensitivity to scalar implicatures in Year 1 and Year 4 children

3.3.1. Introduction

According to Feeney et al. (2004) the dissociation seen in children, between the statement and storyboard tasks, could be due to the cognitive effects associated with the detection of an attempt by an agent to deceive an audience. However, the tasks not only differed in terms of cognitive effects, but also in terms of shared conversational background. The storyboard task, unlike the statement task, makes the relevant evidence immediately available, and it is claimed that this feature also facilitates sensitivity to the implicature (Guasti et al., 2005). To investigate whether the attempt of an agent to deceive an audience makes an additional contribution to the detection of scalar implicatures, over and above that made by the shared conversational background, the visual component of the storyboard task was held constant whilst the dialogue was manipulated so as to make reference to an agent in one condition but not in the other. Thus, although none of the photographs included an agent, in one condition the dialogue made specific reference to an agent. If the increased sensitivity to the implicature that is seen in the storyboard task is due to an agent's attempt to deceive, then such increased sensitivity (i.e. decreased logical responding), would be expected in the agent condition rather than the non-agent condition.

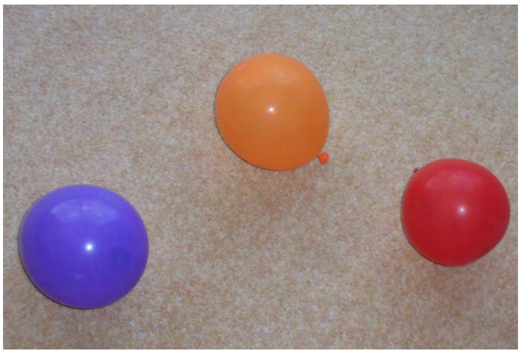
3.3.2. Method

Participants

Fifty three children participated in the experiment. There were twenty five Year 1 children (age range: 5 years 6 months – 6 years 6 months; mean age: 6 years 1 month) and twenty eight Year 4 children (age range: 8 years 6 months – 9 years 6 months; mean age: 9 years), all from a school in the North East of England. Consent for their participation was given by the head teacher.

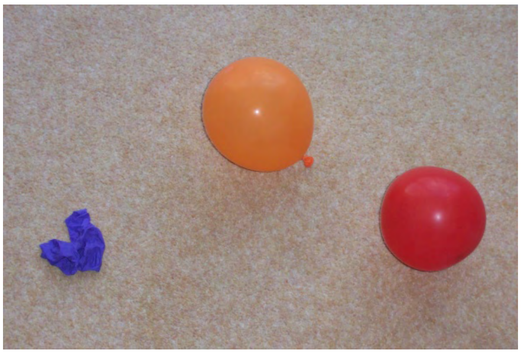
Materials

The children saw 32 storyboards, with 16 storyboards in each condition. The storyboards were based on those used in previous experiments and contained the same control trials, *true all*, *false all* and *felicitous some*, and target trials, *infelicitous some*. The visual appearance of the storyboards was held constant, although each storyboard depicted a different set of objects, whilst the dialogue accompanying the storyboards was manipulated to make reference to an agent in one condition but not in the other (for the themes used for the storyboards see Appendix AV). An example storyboard for each condition can be seen in Figure 3.1.



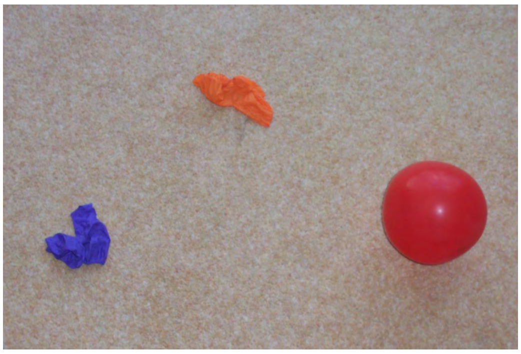
(A) Charlotte finds three balloons

(NA) Here are three balloons



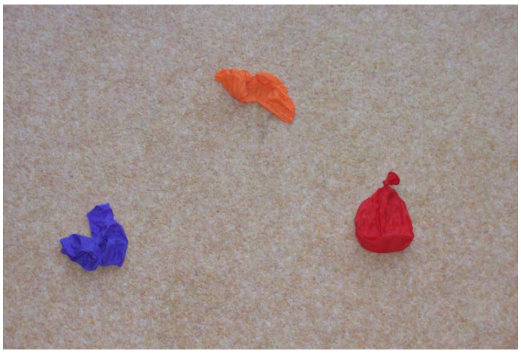
(A) She pops the first balloon

(NA) The first balloon has been popped



(A) She pops the second balloon

(NA) The second balloon has been popped



(A) She pops the third balloon

(NA) The third balloon has been popped

(A) Charlotte's mum says to Charlotte, Charlotte what have you done to the balloons? Charlotte says, I popped some of the balloons.

(NA) Some of the balloons have been popped.

(A) = Agent Condition (NA) = Non-agent Condition

Figure 3.1. Example of a storyboard used for each condition in Experiment 5.

Visual appearance. As can be seen in Figure 3.1 each storyboard contained four photographs. The first photograph introduced three objects, and the following photographs depicted an interaction with either two or three of the objects. If an interaction happened with only two of the objects the fourth photograph was a repeat of the preceding one.

Dialogue. As can be seen in Figure 3.1, in the agent condition the dialogue took the form of a story about a girl called Charlotte. In the story Charlotte finds three objects, she interacts with either two or three of the objects, she is then asked by her mum what she has done to the objects and she gives a reply. In the non-agent task no reference is made to any one interacting with the objects. The objects and what has happened to them are simply referred to and there is a final statement about the end state of affairs.

One potential problem with manipulating the dialogue is that the concluding statement in the non-agent condition is in the passive voice, and an understanding of the passive voice generally occurs later than that of the active voice (Buckley, 2003), which is the grammatical form of the agent condition. However, research suggests that whilst passive sentences may be syntactically complex (Harris & Coltheart, 1996) the children in this experiment should be able to comprehend the statements given their age and the construction of the passive. This is because the passive sentences used were both truncated (no agent present) and non-reversible (the object could not carry out the verb). Both of these factors have been found to facilitate comprehension of the passive voice in children (Bever, 1970; Buckley, 2003; Harris, 1976; Maratsos, Kuczaj, Fox & Chalkley, 1979; Tager-Flusberg, 1993; Van der Lely, 1996). In addition the use of the passive voice moves the focus of attention from the agent to the object (Tomasello & Brooks, 1999); this coupled with the truncated form whereby no agent is mentioned could serve to enhance the experimental manipulation of agent and non-agent conditions.

Each storyboard appeared with the quantifiers *all* and *some*, although each child received only one form. This resulted in two sets of storyboards for each condition. The order of presentation of the conditions, agent/non-agent, was counterbalanced resulting in four groups for each age group.

Procedure

The experiment was completed in two sittings with a gap of a week between each condition. The children took part in small groups and received instructions on how to complete the task immediately prior to attempting each task. For the agent condition they were told that they would hear stories about a girl called Charlotte who finds different objects. Their job was to decide whether the statement made by Charlotte, in response to her mum asking what she has done to the objects, was true or false. In the non-agent condition the children were told that they would see a series of storyboards and that in each story board there would be a collection of objects. Their job was to decide whether the statement made about the final state of affairs was true or false. The children had an answer sheet on which to record their answers.

3.3.3 Results

As in the previous experiments the children's responses were coded with respect to their logical correctness. At this point the data of five Year 1 children were removed, four because of poor responding in the *true all* trials of the agent condition and one for the same trials in the non-agent condition, and from one Year 4 child because of below or at chance responding on the control trials in both conditions.

The mean scores and standard deviations for the remaining 47 children according to truth, term, condition and age can be seen in Table 3.1.

Table 3.1

Mean number of logical responses (and standard deviations) according to truth, term, condition and age for Experiment 5

		Year 1 (n = 20)		Year 4 (n = 27)	
		Agent	Non-agent	Agent	Non-agent
True/ Felicitous	All	3.95 (0.22)	3.80 (0.52)	4.00 (0.00)	4.00 (0.00)
	Some	3.15 (1.14)	3.35 (0.93)	4.00 (0.00)	3.96 (0.19)
False/ Infelicitous	All	3.80 (0.52)	3.30 (1.17)	3.96 (0.19)	3.96 (0.19)
	Some	0.75 (1.07)	1.35 (1.60)	0.15 (0.36)	0.04 (0.19)

Note. The maximum score for each category is 4.

To be assured that the passive sentence construction of the non-agent condition had not caused comprehension difficulties, analysis was carried out on the number of logical responses given to the control statements between conditions. Wilcoxon signed ranks tests revealed that there was no difference in responses rates for either age group, Year 1, z = 1.23, p = .207, r = .20, Year 4, z = 1.00, p = .317, r = .14.

Agent data

In the agent condition the Year 1 children gave significantly fewer logical responses to the *some* trials (mean 3.90) than to the *all* trials (mean 7.76), Wilcoxon matched pairs signed ranks test, $z = 4.05$, $p < .001$, $r = .64$, and to the *false/infelicitous* trials (mean 4.56) compared to the *true/felicitous* trials (mean 7.10), Wilcoxon matched pairs signed ranks test, $z = 3.24$, $p = .001$, $r = .51$. The Year 4 children also gave significantly fewer logical responses to the *some* trials (mean 4.14) than to the *all* trials (mean 7.96), Wilcoxon matched pairs signed ranks test, $z = 4.87$, $p < .001$, $r = .66$, and to the *false/infelicitous* trials (mean 4.12) compared to the *true/felicitous* trials (mean 8.00), Wilcoxon matched pairs signed ranks test, $z = 4.86$, $p < .001$, $r = .66$.

Friedman tests were used to analyse the number of logical responses given to the control and target trials in the agent condition by each age group. The results were significant for both age groups, Year 1, χ^2 (3) = 44.42, $p < .001$, and Year 4, χ^2 (3) = 79.58, $p < .001$. Follow up Wilcoxon signed ranks tests revealed that both groups of children gave significantly fewer logical responses to the *infelicitous some* trials than to any of the control trials, all p's $\leq .001$, all r's $\geq .52$ (see Appendix BIV, Tables 1 and 2 for statistics showing the comparison of each type of trial in the agent condition).

Non-agent data

In the non-agent condition the Year 1 children gave significantly fewer logical responses to the *some* trials (mean 4.70) than to the *all* trials (mean 7.10), Wilcoxon matched pairs signed ranks tests, z = 3.01, $p = .003$, $r = .48$, and to the *false/infelicitous* trials (mean 4.66) compared to the *true/felicitous* trials (mean 7.16), Wilcoxon matched pairs signed ranks test, $z = 3.43$, $p = .001$, $r = .54$. The Year 4 children also gave significantly fewer logical responses to the *some* trials (mean 4.00) than to the *all* trials (mean 7.96), Wilcoxon matched pairs signed ranks test, $z = 5.11$, $p < .001$, $r = .70$ and to the *false/infelicitous* trials (mean 4.00) compared to the *true/felicitous* trials (mean 7.96), Wilcoxon matched pairs signed ranks test, $z = 5.04$, $p < .001$, $r = .69$.

Friedman tests were used to analyse the number of logical responses given to the control and target trials in the agent condition by each age group. The results were significant for both age groups, Year 1, χ^2 (3) = 21.98, $p < .001$, and Year 4, χ^2 (3) = 78.16, $p < .001$. Follow up Wilcoxon signed ranks tests revealed that both groups of children gave significantly fewer logical responses to the *infelicitous some* trials than to any of the control trials, all *p's* $\leq$ .005, all *r's* $\geq$ .44 (see Appendix BV, Tables 3 and 4 for statistics showing the comparison of each type of trial in the non-agent condition).

Infelicitous some data

The natures of the distributions of the *infelicitous some* data for both conditions can be seen in Figure 3.2. In the agent condition the majority of children in both age groups were always pragmatic; only two Year 1 children were logical and these were logical responders in the non-agent condition. In the non-agent condition the Year 4's distribution was almost identical to that of the agent condition, whilst for the Year 1 children there is a shift towards an increase in logical and equivocal responders.

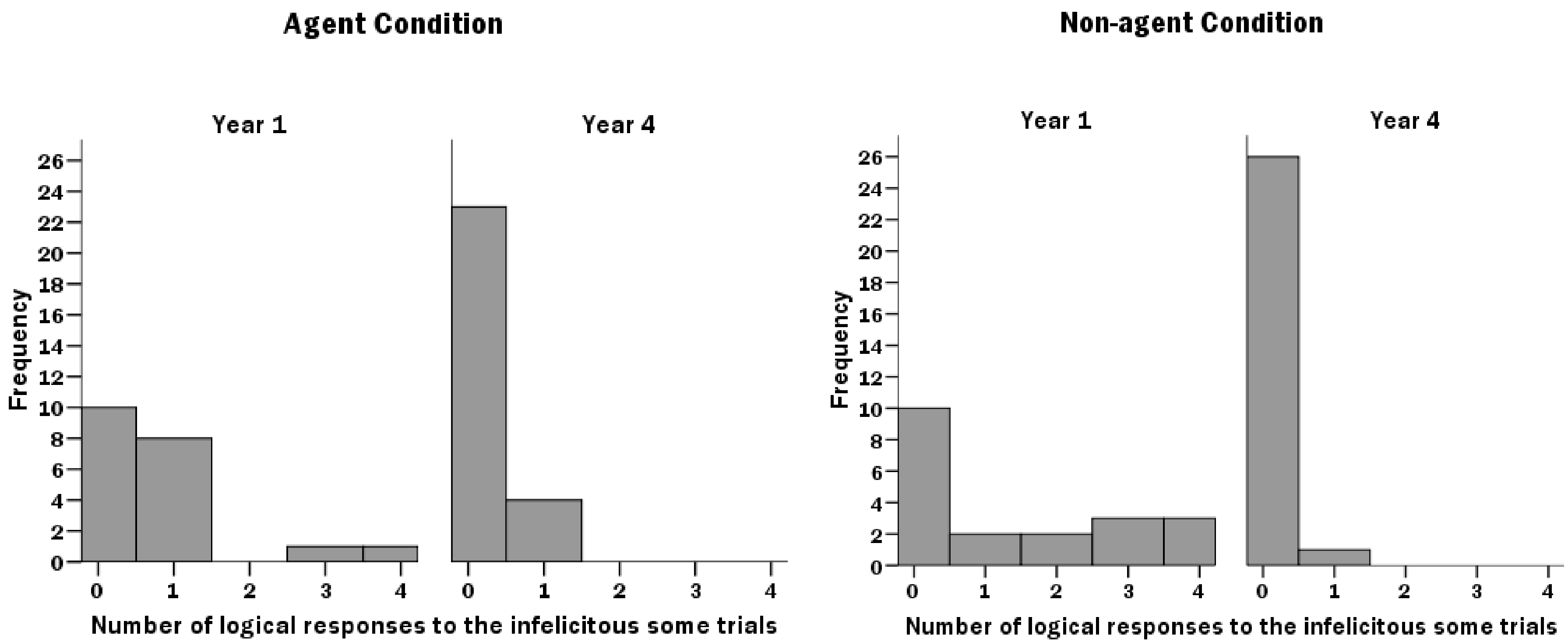


Figure 3.2. Distribution of participants according to the number of logical responses given to the infelicitous some trials for the agent and non-agent conditions in Experiment 5.

Analysis of the *infelicitous some* data, the means of which can be seen in Figure 3.3, revealed that the Year 1 children gave significantly more logical responses in both conditions when compared with the Year 4 children, Mann-Whitney tests, non-agent condition, $U = 171$, $p = .008$, $r = .39$, agent condition, $U = 141$, $p < .001$, $r = .55$. With regards to responding between conditions, the Year 1 children gave significantly more logical responses in the non-agent condition than in the agent condition, Wilcoxon matched pairs signed ranks test, $z = 2.07$, $p < .039$, $r = .33$, but there was no significant difference for the Year 4 children, Wilcoxon matched pairs signed ranks test, $z = 1.34$, $p = .18$, $r = .18$.

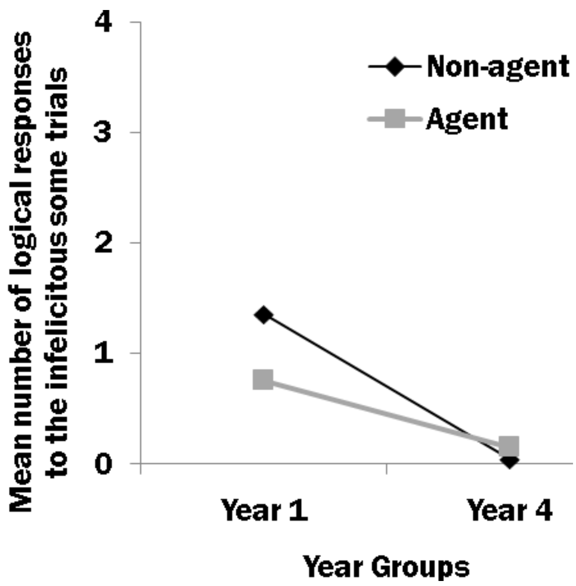


Figure 3.3. Mean number of logical responses to the *infelicitous some* trials, for the agent and non-agent conditions, for each age group in Experiment 5.

The order in which the conditions were presented was also examined to see if it affected the rate of logical responding to the *infelicitous some* trials; the means (and standard deviations) for the Year 1 children can be seen in Table 3.2 and for the Year 4 children in Table 3.3.

Table 3.2

Year 1 children's mean number of logical responses (and standard deviations) for the infelicitous some trials, according to condition and order of presentation, for Experiment 5

Year 1	Order	Condition	
(*n* = 20)		Agent	Non-agent
	Agent First	0.50	0.25
		(0.54)	(0.46)
	Non-agent First	0.92	2.08
		(1.31)	(1.68)

Collapsing across conditions, the Year 1 children who received the non-agent condition first (mean 1.50) gave more logical responses to the *infelicitous some* trials than those children who received the agent condition first (mean 0.37) and this was approaching significance, Mann-Whitney test, $U = 26$, $p = .078$, $r = .39$. The medium effect size suggests that the lack of significance may be due to the small number of participants, rather than there being no real effect. Within conditions, there was no effect of order in the agent condition, Mann-Whitney test, $U = 44$, $p = .732$, $r = .05$, but there was an effect of order in the non-agent condition, with those children who received this condition first giving more logical responses compared to the children who received it second, Mann-Whitney test, $U = 20$, $p = .020$, $r = .52$.

Table 3.3

Year 4 children's mean number of logical responses (and standard deviations) for the infelicitous some trials, according to condition and order of presentation, for Experiment 5

Year 4 (*n* = 27)	Order	Condition	
		Agent	Non-agent
	Agent First	0.00	0.00
		(0.00)	(0.00)
	Non-agent First	0.31	0.08
		(0.48)	(0.28)

Order effects were also seen in the Year 4 children. Collapsing across conditions, the children who received the non-agent condition first (mean 0.39) gave significantly more logical responses to the *infelicitous some* trials than the children who received the agent condition first (mean 0.00), Mann Whitney test, $U = 56$, $p = .012$, $r = .34$. Within conditions, there was no effect of order in the non-agent condition, Mann Whitney test, $U = 84$, $p = .299$, $r = .14$, but there was in the agent condition with those children who received this task second giving more logical responses than the children who received it first Mann Whitney test, $U = 63$, $p = .027$, $r = .42$.

The data was also analysed for practice effects. These were seen in the Year 1 children, with significantly more logical responding in the task that was presented first (mean 1.45) compared to the task that was presented second (mean 0.65), Wilcoxon matched pairs signed ranks test, $z = 2.58$, $p = .010$, $r = .41$. Practice effects were not observed in the Year 4 children, Wilcoxon matched pairs signed ranks test, $z = 1.34$, $p = .180$, $r = .18$.

3.3.4. Discussion

The prediction made at the start of the experiment, that children would show increased sensitivity to the scalar implicatures in the agent condition was partially confirmed; increased sensitivity in the agent condition was found in the Year 1 children, but not in the Year 4 children, who had near universal pragmatic responding in both conditions.

The expected difference in response rates between the conditions was not seen in the Year 4 children; their rate of logical responding was close to floor level. As the visual scene was held constant in both conditions it could be proposed that the shared conversational background was responsible for the high level of sensitivity to the scalar implicatures, rather than the deception context. However, comments made by the children during testing lead to an alternative explanation. The children often commented, *I wonder who has eaten the sweets,* or, *I wonder who has torn the magazines.* These comments suggest that although no-one was present in the non-agent condition, as all the pictures depicted the results of actions, the children inferred the involvement of an actor. Indeed other research has shown that very young children can infer the involvement of a hidden causal agent, without direct perceptual evidence of the agent's existence, if the context involves objects being acted upon (Saxe, Tenenbaum & Carey, 2005; Saxe, Tzelnic & Carey, 2007). If the Year 4 children did infer the involvement of an agent then this could provide an alternative account for the lack of difference in the response rates between the two conditions.

The greater pragmatic response rate of the Year 1 children in the agent condition suggests that a deception context which involves an agent's attempts to mislead an audience facilitates sensitivity to scalar implicatures over and above the benefits gained from a shared conversational background, thus supporting Feeney et al.'s (2004) claim. However, it should be noted that this was not the only facilitatory factor for this age group; practice effects were also observed. This finding though does raise the question of whether an important aspect of facilitation is the motivation behind the agent's attempt to deceive. It could be said that in minimising her involvement with the objects, Charlotte could benefit by lessening the consequences of her actions. In other words, if Charlotte's mother would be angry to learn that all the paint pots have been knocked over, Charlotte, by implying that she hasn't knocked over all of them, may escape some punishment. Indeed deception to avoid punishment is the most common form of deception in early childhood (Lewis, 1993). In the non-agent condition there was nobody to benefit from the deception. Therefore the possibility remains that sensitivity to scalar implicatures is facilitated by situations where there is a benefit to the agent of being under-informative. Because of the floor effects found with the Year 4 children in this experiment, the next experiment focused on Year 1 children.

3.4. Experiment 6: Further investigation into deception contexts and sensitivity to scalar implicatures in Year 1 children

3.4.1. Introduction

The previous experiment provided evidence for a facilitating effect of an agent in deception contexts, particularly amongst the youngest participants. An important question about this result is whether the motivation of the agent is a key aspect of this facilitation effect. In order to investigate this question, an additional set of storyboards was introduced. This descriptive set differed from the interactive set used in the previous experiment in terms of the nature of the involvement with the objects. In the storyboards used in the previous experiment the objects were interacted with, either explicitly in the agent condition or implicitly in the non-agent condition. In the new set of storyboards the objects are merely described and no responsibility can be attached to the speaker for their current state. For example, one descriptive storyboard features three bags of sweets and the bags are described with reference to their contents. Thus Charlotte has the motivation to deceive in the interactive set but not in the descriptive set. It was predicted that if sensitivity is facilitated by the potential benefits to an agent, then there should be more pragmatic responding in the agent condition of the interactive set compared to the descriptive set. In order to replicate the effect of agency in the previous experiment, the agent and non-agent conditions were included for both sets of storyboards. Once again the rationale for this manipulation is that if sensitivity relies on the presence of an agent who is under-informative then more pragmatic responding might be expected in the agent conditions regardless of the nature of the agent's involvement with the objects.

3.4.2. Method

Participants

Fifty five children participated in the experiment (age range 5 years 8 months to 6 years 9 months; mean age 6 years 3 months). They were the Year 1 age groups from three schools in the North East of England. Consent for their participation was given either by the head teacher or the parents.

Materials

There were two sets of storyboards; an interactive set and a descriptive set. The interactive storyboards were visually identical to the storyboards in the previous experiment. Modification was made to the dialogue to make the storyboards compatible with the new descriptive set. Reference to Charlotte's mother was removed, so Charlotte made a statement about what she had done to the objects rather than responding to her mother's query. The new descriptive set of storyboards described characteristics of objects. The descriptive storyboards had the same layout as the interactive ones, and for each set of storyboards there were two conditions, as in the previous experiment, agent and non-agent. The themes used for the descriptive storyboards can be seen in Appendix AV, and an example of a descriptive storyboard with accompanying dialogue for each condition can be seen in Figure 3.4.

(A) Charlotte finds three bags of sweets

(NA) Here are three bags of sweets

(A) The first bag contains liquorice all sorts

(NA) The first bag contains liquorice all sorts

(A) The second bag contains liquorice all sorts

(NA) The second bag contains liquorice all sorts

(A) The third bag contains liquorice all sorts

(NA) The third bag contains liquorice all sorts

(A) Charlotte says, "Some of the bags contain liquorice all sorts"

(NA) Some of the bags contain liquorice all sorts

(A) = Agent Condition (NA) = Non-agent Condition

Figure 3.4. Example of a descriptive storyboard for each condition in Experiment 6.

As can be seen from Figure 3.4, in the agent condition Charlotte is still the focus of the dialogue, although a change has been made to the dialogue from the previous experiment. In this experiment she makes a statement about what she has done (interactive set) or what she has found (descriptive set) rather than giving a response to her mother's query. In the non-agent condition the interaction or characteristics of the objects are described and a final statement is made about the objects.

Each child saw one set of storyboards, either the interactive or descriptive set, with the agent and non-agent conditions. There were sixteen storyboards in each condition of which twelve were control trials, *true all, false all* and *felicitous some* and four were the *infelicitous some* target trials. The order of presentation of the agent and non-agent conditions was counterbalanced, this resulted in eight groups.

Procedure

The children took part in small groups and received instructions on how to complete each task immediately prior to the task taking place. For the agent condition they were told that they would hear stories about a little girl called Charlotte who finds different objects. Their task was to decide whether the statement made by Charlotte at the end of the story about what she had done, was true or false. In the non-agent condition the children were told that they would see a series of storyboards and that in each story board there would be a collection of objects. Their job was to decide whether the statement made about the final state of affairs was true or false. The children had an answer sheet on which to record their answers. The tasks were completed in two sittings with a gap of two weeks between each.

3.4.3. Results

As in the previous experiments the responses of the children were coded according to their logical correctness. At this point the data from seven children were removed because of their poor performance on the control trials in both conditions. The mean scores and standard deviations for the remaining 48 participants according to truth, term, set and condition can be seen in Table 3.4.

Table 3.4

Mean number of logical responses (and standard deviations) according to truth, term, set and condition for Experiment 6

		Interactive (n = 22)		Descriptive (n = 26)	
		Agent	Non-agent	Agent	Non-agent
True/ Felicitous	All	3.86 (0.35)	3.82 (0.50)	3.77 (0.43)	3.65 (0.56)
	Some	3.59 (0.91)	3.09 (1.19)	3.12 (1.42)	3.12 (1.11)
False/ Infelicitous	All	3.73 (0.55)	3.73 (0.46)	3.73 (0.53)	3.58 (0.90)
	Some	0.64 (1.00)	1.05 (1.21)	2.35 (1.60)	2.58 (1.60)

Note. The maximum score for each category is 4.

Interactive set data

In the agent condition of the interactive set of storyboards fewer logical responses were given to the *some* trials (mean 4.23) compared to the *all* trials (mean

7.59), Wilcoxon matched pairs signed ranks test, $z = 4.11$, $p < .001$, $r = .62$, and to *false/infelicitous* trials (mean 4.36) compared to the *true/felicitous* trials (mean 7.45), Wilcoxon matched pairs signed ranks test, $z = 4.04$, $p < .001$, $r = .61$. In the non-agent condition fewer logical responses were also given to the *some* trials (mean 4.14) compared to the *all* trials (mean 7.55), Wilcoxon matched pairs signed ranks test, $z = 4.20$, $p < .001$, $r = .63$, and to the *false/infelicitous* trials (mean 4.77) compared to the *true/felicitous* trials (mean 6.91), Wilcoxon matched pairs signed ranks test $z = 3.28$, $p = .001$, $r = .50$. Friedman tests were used to analyse the number of logical responses given to the control and target trials for both conditions and these were significant, agent χ^2 (3) = 46.68, $p < .001$, and non-agent χ^2 (3) = 43.10, $p < .001$. Follow up Wilcoxon signed ranks tests revealed that significantly fewer logical responses were given to the *infelicitous some* trials than any of the control trials in both conditions, all *p's* $\leq .001$, all *r's* $\geq .50$ (see Appendix BVI, Table 1 for statistics showing the comparison of target trials with control trials for both conditions)

Descriptive set data

In the agent condition of the descriptive set of storyboards fewer logical responses were given to the *some* trials (mean 5.46) compared to the *all* trials (mean 7.50), Wilcoxon matched pairs signed ranks test, $z = 3.84$, $p < .001$, $r = .53$, but not to the *false/infelicitous* trials (mean 6.08) compared to the *true/felicitous* trials (mean 6.88), Wilcoxon matched pairs signed ranks test, $z = 1.44$, $p = .150$, $r = .20$. In the non-agent condition fewer logical responses were also given to the *some* trials (mean 5.69) than to the *all* trials (mean 7.23), Wilcoxon matched pairs signed ranks test, $z = 3.13$, $p =.002$, $r = .43$, but not to the *false/infelicitous* trials (mean 6.15) compared to the *true/felicitous* trials (mean 6.77), Wilcoxon matched pairs signed ranks test, z = 1.28, p =.201, r = .18. Friedman tests were used to analyse the number of logical responses given to the control and target trials for both conditions and these were significant, agent χ^2 (3) = 15.30, p = .002, and non-agent χ^2 (3) = 7.89, p = .048. Follow up Wilcoxon signed ranks tests revealed that significantly fewer logical responses were given in both conditions to the *infelicitous some* trials when compared to the *true all* and *false all* trials, all p's $\leq$ .017, all r's $\geq$ .33 but not to the *felicitous some* trials, all p's $\geq$ .132, all r's $\leq$ .21 (see Appendix BV, Table 2 for statistics showing the comparison of target trials with control trials for both conditions).

Infelicitous some data

The distribution of the *infelicitous some* data in the interactive set for both conditions, which can be seen in Figure 3.5, is unimodal; the majority response in both conditions is to always be pragmatic. With regards to the pattern of responding between conditions, ten of the children were consistent in their responses, which tended to be pragmatic, eight were more logical in the non-agent condition compared to the agent condition, whilst four were more logical in the agent condition compared to the non-agent condition.

A different distribution and pattern of responding can be seen in Figure 3.5 for the descriptive set of storyboards. Although the distributions are again unimodal, the majority response is to always be logical. As to the pattern of responding between conditions; ten children were more logical in the non-agent condition compared to the agent condition, four were more logical in the agent condition compared to the non-agent condition, whilst twelve children were consistent between conditions. Of these twelve consistent responders only three were pragmatic responders.

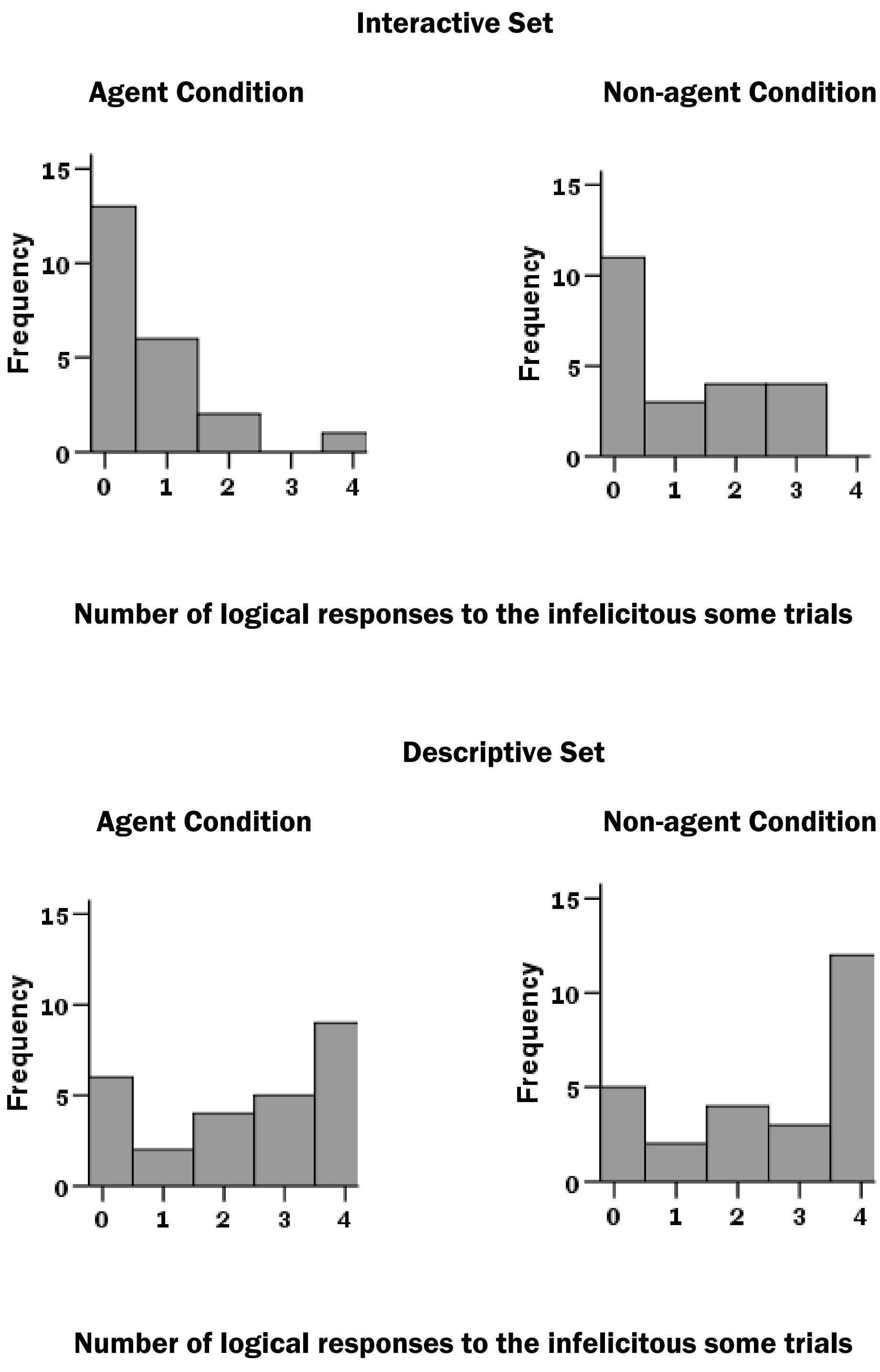


Figure 3.5. Distribution of participants according to the number of logical responses given to the infelicitous some trials in both conditions and for both sets of storyboards in Experiment 6.

With regards to responses to the *infelicitous some* trials between sets, significantly more logical responses were given to the descriptive set for both the agent, Mann Whitney test, $U = 122.5$, $p < .001$, $r = .50$, and non-agent conditions, Mann Whitney test, $U = 131.5$, $p = .001$, $r = .48$. However within each set there was no significant difference between the number of logical responses given to the agent and non-agent conditions, Wilcoxon matched pairs signed ranks tests, interactive set, $z = 1.45$, $p = .147$, $r = .22$, descriptive set, $z = 0.89$, $p = .373$, $r = .12$.

The order in which the conditions were presented was also examined to see if it affected the rate of logical responding to the *infelicitous some* trials; the means (and standard deviations) for the interactive set of storyboards can be seen in Table 3.5 and for the descriptive set in Table 3.6.

Table 3.5

Mean number of logical responses (and standard deviations) for the infelicitous some statements of the interactive set of storyboards, according to condition and order of presentation, for Experiment 6

Interactive	Order	Condition	
(n = 22)		Agent	Non-agent
	Agent First	0.62	0.46
		(1.19)	(0.78)
	Non-agent First	0.67	1.89
		(0.71)	(1.30)

Collapsing across conditions in the interactive set of storyboards revealed that significantly more logical responses were given to the *infelicitous some* trials by the children who received the non-agent condition first (mean 2.56) than those who received the agent condition first (mean 1.08), Mann-Whitney test, $U = 29.00$, $p = .042$, $r = .43$. Within conditions, there was no effect of order in the agent condition, Mann-Whitney test, $U = 47.50$, $p = .404$, $r = .18$, but there was an effect in the non-agent condition, with those children who received this condition first giving more logical responses compared to the children who received it second, Mann-Whitney test, $U = 22$, $p = .009$, $r = .56$.

Table 3.6

Mean number of logical responses (and standard deviations) for the infelicitous some statements of the descriptive set of storyboards, according to condition and order of presentation, for Experiment 6

Descriptive	Order	Condition	
($n = 26$)		Agent	Non-agent
	Agent First	2.38	2.08
		(1.66)	(1.85)
	Non-agent First	2.31	3.08
		(1.60)	(1.19)

No order effects were seen in the descriptive set of storyboards either collapsing across conditions, Mann-Whitney test, $U = 71.00$, $p = .482$, $r = .14$, or within conditions, Mann-Whitney tests, agent, $U = 81.00$, $p = .853$, $r = .04$, non-agent, $U = 58.50$, $p = .158$, $r = .28$.

The data was also analysed for practice effects; the means (and standard deviations for which can be seen in Table 3.7.

Table 3.7

Mean number of logical responses (and standard deviations) for the infelicitous some trials, for each set of storyboards, according to time of presentation of the conditions, for Experiment 6

	1st Condition	2nd Condition
Interactive set	1.14	0.55
(*n* = 22)	(1.36)	(0.74)
Descriptive set	2.73	2.19
(*n* = 26)	(1.46)	(1.70)

Practice effects were seen in the interactive set of storyboards, with children giving significantly more logical responses to the *infelicitous some* trials in the condition they received first compared to the condition they received second (mean 0.55), Wilcoxon signed ranks test, $z = 2.09$, $p = .036$, $r = .32$. Practice effects were not observed in the descriptive set of storyboards, Wilcoxon signed ranks test, $z = 1.53$, $p = .127$, $r = .21$.

3.4.4. Discussion

To summarise the results of this experiment; the descriptive set of storyboards produced significantly less pragmatic responding in both the agent and non-agent conditions compared to the interactive set of storyboards, but there was no significant difference in pragmatic response rates within each set of storyboards between the agent and non-agent conditions.

The difference in logical response rates to the *infelicitous some* trials that was seen between sets, but not within sets, suggests that the motivation behind an attempt to deceive plays a facilitatory role in the detection of scalar implicatures. The statement made by Charlotte (the agent condition) for the *infelicitous some* trials of both sets of storyboards could be seen as untruthful. However, in the interactive set of storyboards Charlotte obviously gains from the infelicitous statement if her interlocutor fails to detect the implicature. It is difficult to imagine what the cognitive gains are from uttering an infelicitous statement about the colour of paint charts.

Unlike the previous experiment, the number of logical responses to the *infelicitous some* trials between the conditions of the interactive set of storyboards was not significantly different. However in this experiment there was an overall decrease in logical responding to the *infelicitous some* trials compared to Experiment 5 (Exp. 5, mean = 2.10 ; Exp. 6, mean = 1.69) especially in the non-agent condition of this experiment (Exp. 5 non-agent, mean = 1.35; Exp. 6 non-agent, mean 1.05), which could possibly account for this finding. There was a weak to medium effect of agency. $r = .22$, and an order effect was still observed in the non-agent condition, which would suggest that agency is still important for some children. However, practice effects in the interactive storyboard set are again seen, suggesting that context is not the only facilitating factor.

With regards to pragmatic theories, the results of this experiment are best accounted for by relevance theories. These theories can account for the difference in response rates to the *infelicitous some* trials between the descriptive and interactive sets through effort and/or effect; there are possibly greater cognitive effects in detecting an untruth that benefits the speaker of the utterance. Additionally the practice effects that are found in the interactive set can be explained in terms of set structure; the trials in the first task may have highlighted the difference between the *some* and *all* trials, and between the *true* and *false* trials. This may have enabled the children to make comparisons between the terms which helped them to see the under-informative nature of the *infelicitous some* trials, and thus enabled them to perform more pragmatically in task two. The lack of practice effects in the descriptive set can also be explained; although the effort has decreased, the effect to be gained does not warrant the extra processing of the implicature.

Levinson’s (1983, 2000) default theory could account for the practice effects; the structured set of examples may have helped the children to see the under-informative nature of the *infelicitous some* trials, and thus enabled them to perform more pragmatically in task 2. However the fact that the practice effects are not seen in the descriptive set casts doubt on this explanation. The descriptive set has essentially the same structure as the interactive set and would therefore allow the same comparisons, thus it is difficult to see why learning is not facilitated in the descriptive set without recourse to the context. The difference in response rates to the *infelicitous some* trials between the interactive and descriptive sets is also difficult for default theories. As the children were randomly assigned to the different sets, a similar level of linguistic knowledge would be expected between the groups. However whilst the majority response pattern in the interactive set was to be consistently pragmatic between the conditions, this was the minority response in the descriptive set.

3.5. General Discussion

The experiments presented in this chapter show that context plays an important role in the detection of scalar implicatures. However, a supportive context alone does not necessarily result in the detection of scalar implicatures and effects of age, order of presentation and practice are also seen.

The two experiments examined the role of deception in the facilitation of sensitivity to scalar implicatures, and in doing so also considered the role of shared conversational background.

The high levels of pragmatic responding seen in the experiments suggest that deception does facilitate the detection of scalar implicatures in children, thus supporting the claim of Feeney et al. (2004). However, the contribution of deception is multifaceted. Deception has many definitions or features (Chisholm & Feehan, 1977; Vincent & Castlefranchi, 1981) including being under-informative (Adler, 1997), involving an agent (Hyman, 1989) and misleading with the truth to benefit the speaker (Sutter, 2009), all of which would apply to the experimental manipulations carried out in the chapter. However, the difference in the level of logical responding to the *infelicitous some* trials between the interactive and descriptive sets of storyboards in Experiment 6 suggests that the aspect of deception which contributes to the facilitation of sensitivity to scalar implicatures in the storyboard task is the motivation behind the act of deception, in other words sensitivity is more likely to be seen when there is a reason for being under-informative.

The finding that sensitivity to scalar implicatures is more likely to be seen in contexts when there are gains to made from the deceptive act, raises the interesting possibility that detection of scalar implicatures could be facilitated by contexts which carry consequences for the interpretation of scalar terms. For example Guasti et al. (2005) found high rates of pragmatic responding in a study that would fit this explanation. In their study we are told that part of the scenario involved a discussion between the soldiers regarding the benefits and costs of each mode of transport;

motorbikes were considered fast but expensive and therefore horses were chosen. This scenario seems to convey the message that the choice of transport is important and that cost is the reason for the soldiers' choice. Maybe the children are aware of this and seek to be as informative as possible as a result. A lack of consequences could explain children's, and adults, insensitivity in the statement evaluation task, and also children's insensitivity in Papafragou and Musolino's (2003) study, where it doesn't seem to matter whether some or all of the elephants have trunks or whether some or all of the horses jumped over a fence. This could be a fruitful line of future enquiry.

With regards to claims made about shared conversational backgrounds (Guasti et al. 2005), the findings from the descriptive set of storyboards once again show that even when the difference between what is known and what is claimed is immediately available, sensitivity is not necessarily seen in young children. However, the descriptive storyboards produced a higher level of pragmatic responding in the Year 1 children compared with the statement condition in Experiment 2 for the same age group. This suggests that whilst neither deception nor a shared conversational background are sufficient for scalar implicature detection, each is beneficial to young children.

One interesting finding of this chapter was the universal pragmatic responding of the Year 4 children in Experiment 5. A possibility for future research would be to extend the age range in that experiment to see if sensitivity to the implicature follows the same U-shaped trend that was seen in Experiment 1, thus confirming the proposal that the developmental trajectory of sensitivity to scalar implicatures is not linear. In addition, finding out when consistent logical responding re-emerges will have implications for theories of children's cognitive development. This point is expanded on in the General Discussion in Chapter 5.

With regards to theoretical accounts of scalar implicature, although both relevance and default theories can explain many of the results of this chapter, the effects of practice are probably best accounted for by relevance theories. The lack of such effects in the descriptive set of Experiment 6 suggests that the structure of the task and a second presentation does not necessarily facilitate a pragmatic interpretation of the term *some*. However, although context is a central feature of relevance theories, and it has been shown that this is an important factor in the detection of scalar implicatures, it appears that effort and effect determine whether an implicature is detected; thus a supportive context does not necessarily result in sensitivity to a scalar implicature. The role of effort in scalar implicature detection is the subject of the next chapter.

Chapter Four

The Processing of Scalar terms in Children and Adults

4.1. Aim

How scalar terms are processed by adults and children remains unclear. Researchers stand divided over the issue in adults, with the debate centring on whether implicatures are processed automatically or derived according to context, and there is no processing account for children, merely the assumption that scalar implicatures are more effortful for this age group as compared to adults. The U-shaped result of Experiment 2, along with the anecdotal reports of the adult participants, suggests that adult processing may be more complex than contemporary theories propose, whilst children's processing may not be as cognitively demanding as has previously been assumed. The aim of the two experiments in this chapter was to investigate and compare scalar implicature processing in children and adults.

4.2. Introduction

The default account of Levinson (1983, 2000) and Sperber and Wilson's (1985/1995) RT give rise to very different views on the automaticity of scalar implicatures. Levinson proposes that scalar implicatures are so readily inferred that they are the default interpretation and go through automatically; any effort involved is in their subsequent cancellation. On the other hand Sperber and Wilson suggest that scalar implicatures are context dependent, and in the absence of contextual priming a pragmatic interpretation takes longer than a logical one.

Experimental studies have tested these theoretical claims in one of two ways; by examining the effect of cognitive load on interpretations of *infelicitous some* (De Neys & Schaeken, 2007) or analysing the time taken to make logical and pragmatic interpretations (Bott & Noveck, 2004; De Neys & Schaeken, 2007; Feeney et al., 2004; Noveck & Posda, 2003). These studies, all conducted on adult participants, have tended to show that pragmatic interpretations take longer, and involve more effort, compared to logical interpretations, thus supporting relevance based accounts[9]. For example, De Neys and Schaeken (2007) gave their participants a sentence verification task involving under-informative sentences, whilst at the same time burdening their executive cognitive resources with a memory task. They found that participants made fewer pragmatic interpretations under load. Bott and Noveck (2004), also using a sentence verification task, compared the time taken by participants to give a logical interpretation to *infelicitous some* trials, to the time taken by the same participants for a pragmatic interpretation, and found that pragmatic interpretations took longer.

However, one point to note with the above mentioned studies (Bott & Noveck, 2004; De Neys & Schaeken, 2007; Noveck & Posda, 2003) is that the experimental designs are based on interpreting scalar implicature processing as having at most two steps and interpreting the results in line with this assumption. But, Feeney et al. (2004) suggested that this two-step process was too simple to account for the way all adults processed scalar terms. They proposed that there were at least three possible courses of action available to adults with regards to the processing of scalar terms; firstly the inference may go undetected, secondly the implicature could be made, and thirdly the adult participant may be aware of the pragmatic interpretation of the scalar term but prefers to give a logical response and so cancels the implicature. Rather than comparing reaction times between logical and pragmatic interpretations, which they felt was problematic as two different

[9] The one exception being the study of Feeney et al. (2004), which found that for some adults a logical interpretation appeared to be a costly process.

responses were involved, Feeney et al. compared the time taken to say yes to *infelicitous some* trials with the time to give the same response to *felicitous some* trials. They found that the response times differed, with participants taking significantly longer to give a logical response to the *infelicitous some* trials. They suggested that this increase in response time was due to the increased processing time involved in the inhibition of the pragmatic implicature. Thus a logical response was not necessarily indicative of a failure to detect the implicature; it could represent a preferred interpretation. The authors supported their argument with the finding that logical responding to the *infelicitous some* trials was positively correlated with a measure of cognitive capacity, and suggested that because resisting the pragmatic inference required effort, this was achieved primarily by people high in cognitive resources.

It is interesting to consider the results of Experiment 2, the U-shaped developmental trajectory with the re-emergence of logical responding in adolescence, in the context of Feeney et al.'s (2004) proposal that the processing of scalar terms may be more complex than has previously been supposed. It seems unlikely that all of the adults who answered logically had failed to detect the implicature, given the universal pragmatic responding of the Year 7 children. It seems more likely that at least some of the adults chose not to give a pragmatic response. And, coupled with the anecdotal comments of some of the adult participants, these results appear consistent with the suggestion that both logical and pragmatic interpretations are available to at least some adults. They raise the further possibility that choosing between these interpretations may be an additional source of processing difficulty in adults.

With regards to the processing of scalar terms in children there is a paucity of research, with no studies directly manipulating cognitive load or measuring reaction times. Therefore questions still remain as to whether children process scalar terms in line with adults, if less efficiently, or whether the processing is qualitatively different.

The assumption in the literature appears to be that if the pragmatic interpretation of a scalar term takes longer than a logical interpretation in adults, the same will also be true for children, and that the processing cost of the implicature for children will be greater than for adults. This assumption is supported by work which found that sensitivity to implicatures increases when cognitive demands are lessened. Pouscoulous et al. (2007) made a number of changes to a task in which they found low pragmatic responding in young children. In the original task children had to make judgements about the amount of animals that were in a box, for example *some turtles are in the boxes*. In the modified version the authors removed distracter items, changed the response from a meta-linguistic judgement to one that was based on an action, and changed the quantifier from *certains* to *quelques*. This modified version resulted in an increase in pragmatic responding amongst younger children. However, even though the manipulations would intuitively seem to decrease the demands placed on the children's cognitive resources, there was no independent measure of the effort associated with each manipulation.

Yet there is reason to suspect that the effort involved in interpreting scalar terms may not be the same for children and adults. In Experiment 2 the re-emergence of logical interpretations suggested that the older adolescents and adults may have two responses available, whereas the universal pragmatic responding of the younger adolescents suggests that this age group only had one interpretation available. It may be that for the adults, choosing between the alternate interpretations is an additional source of effort. If this is the case, then contrary to the assumption that processing scalar terms is never less costly for children than it is for adults, the reverse may be true. That is, adults may incur processing costs that are not incurred by children. The aim of the two experiments described in this chapter was to test this possibility.

4.3. Experiment 7: The effect of load on the detection of scalar implicatures

4.3.1. Introduction

There is a view in the experimental pragmatics literature that the processing of scalar implicatures is effortful. However much of the evidence for this view comes from indirect measures of effort, such as the association often drawn between increased reaction times for pragmatic interpretations of scalar terms and effortful processing, and the linking of the ability to detect scalar implicatures with increased access to cognitive resources. One study that sought a more direct measure of the role of effort in the processing of scalar terms is that of De Neys and Schaeken (2007). They used dual task methodology to investigate the effect of cognitive load on the detection of scalar implicatures. They burdened the executive cognitive resources of adults by having them memorise dot patterns, either complex (load condition) or simple (control condition), whilst evaluating statements containing scalar terms. They found that the number of pragmatic interpretations decreased in the load condition and claimed that this is because the detection of scalar implicatures is effortful, thus providing support for a relevance theory view. However, this conclusion is based on a two-step model of implicature processing, and as mentioned in the introduction to this chapter this model may be too simple. There may be two sources of difficulty in the statement evaluation task; detecting the implicature, and choosing between a logical and pragmatic interpretation. Therefore one aim of this experiment was to test the validity and reliability of De Neys and Schaeken's study and to establish whether there appears to be any evidence of choice making between pragmatic and logical interpretations in adults.

An additional aim of this experiment was to examine the role of effort in children's processing of scalar terms. As was mentioned in the introduction to this chapter, there is very little research that has directly focused on whether the processing of scalar terms is effortful for children, the possible exception being Pouscoulous et al. (2007)[10]. The assumption in the literature appears to be that if scalar implicatures are effortful for adults then they will be even more so for children. However, there are two experiments in this thesis where universal pragmatic responding has been seen in children; the Year 7 age group in Experiment 2, and the Year 4 age group in Experiment 5. The general interpretation of *some* as meaning *not all* by these participants suggests that they had little difficulty processing the implicature. One possibility is that because the implicature is so frequent in older children, the pragmatic interpretation has become the default interpretation, and thus automatic and effortless. Another possibility is that detecting the implicature requires effort in children, just as in adults, but because they only have one interpretation available to them, the children do not have to expend extra effort in deciding whether to respond logically or pragmatically.

In order to decide between these two possibilities, this experiment examined scalar implicature processing in a group of Year 6 children and a group of adults through the use of a dual task methodology, similar to that employed by De Neys and Schaeken (2007). One potential problem with their design was that each of their

[10] But see comments made in the previous section of this chapter.

participants received the same amount of cognitive load. It is therefore not clear that the load affected all participants equally. For the current experiment it is imperative that load affects all participants equally, both within and between age groups. In response to this potential confound a number of modifications were made to the design of the De Neys and Schaeken study. Prior to the presentation of the tasks each participant had their load capacity determined through the use of a visual pattern test (Gathercole, Pickering, Ambridge & Wearing, 2004; Holmes, Adams & Hamilton, 2008). This was to ensure that they could be given a load relative to this measure, thereby ensuring that all participants were equally affected by the load.

A number of predictions are possible for this experiment. Given the assumption that implicatures are more demanding for children than adults, then the effect of load should be as great, or greater, in children than in adults. Alternatively, if choosing between interpretations is an additional source of difficulty for some adults, then the effect of load should be greater in the adults compared to the children.

4.3.2. Method

Participants

Fifty three adults and thirty children participated in the study. The adults were recruited through a campaign at Durham University and an advertisement in a local magazine. Their ages ranged from 18 years to 65 years (mean age 30 years) and all gave their own consent. The children were a Year 6 class, aged from 10 years 3 months to 11 years 2 months (mean age 10 years 7 months) from a primary school in the North East of England. Consent for their participation was given by the head teacher.

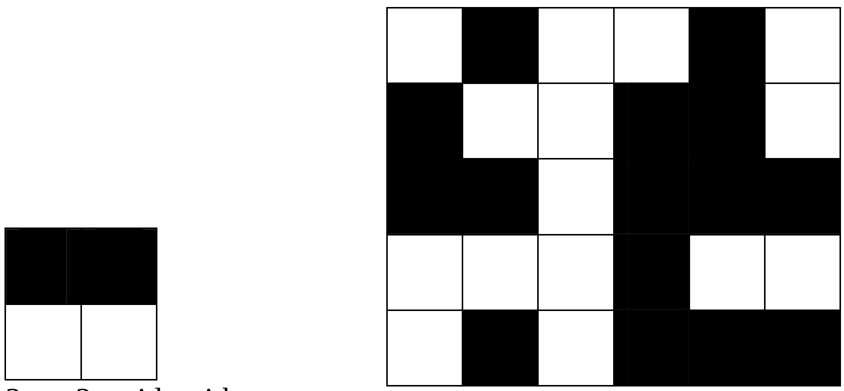
Materials

Each participant heard two sets of 28 statements, under conditions of cognitive load or no load, and a post test consisting of 16 statements. Their load capacity was determined prior to them hearing the statements, using the Thames Valley Test Company's Visual Patterns Test (VPT) (Della Sala, Gray, Baddeley & Wilson, 1997).

Statements. The statements were based on those used in previous experiments in this thesis, with control and target statements (see Appendix AVII for statements used). De Neys and Schaeken (2007) gave their participants 10 *infelicitous some* target statements and 10 filler statements to evaluate. However they raised concerns that the ratio of target to filler statements might result in participants guessing the rationale of the experiment. To combat this, a smaller ratio was used; there were 8 target statements and 20 filler statements. The filler statements comprised 8 *felicitous some*, 4 *absurd all*, 4 *absurd some*, 2 *true all* and 2 *false all* statements. The purpose of these was to disguise the aim of the experiment, to determine that participants had a good grasp of the term *some*, and to establish that the load condition did not interfere with their general linguistic competence. Two sets of materials were prepared. The sets were counterbalanced between participants with regards to order of presentation and load condition, resulting in four groups.

A potential problem was identified with the statement evaluation task in Chapter 2, namely it was not certain that the statements which were considered to be universal by the researcher were considered to be such by the participants. If a participant considers the universal sentences to be existential then they will not consider them to be under-informative when the statements are preceded by the term *some*. In other words they will consider the *some* sentences to be felicitous rather than infelicitous. In order to ascertain that each participant was considering the whole set, and not a partial set in relation to the *infelicitous some* statements, a post test was carried out. The post test consisted of the sixteen *infelicitous some* statements which had previously been presented to the participant, only this time the quantifier *some* was replaced by *all*.

Cognitive Load. Prior to the presentation of the experimental conditions, each participant undertook the VPT. This test was used to assess the participant's visual pattern capacity, in order to determine the amount of load they received in the load condition. The VPT is a measure of short-term visual memory and requires participants to reproduce a visual pattern. The visual pattern is created through filling in half the squares in a grid. The grids vary in size; the smallest measuring 2 x 2 and the largest 5 x 6 (see Figure 4.1 for example visual patterns).



2 x 2 grid with two filled squares

5 x 6 grid with fifteen filled squares

Figure 4.1. Examples of test patterns in the Visual Patterns Test[11]

The number of filled squares in a grid determines the level of difficulty; this ranges from 2 to 15 filled squares. For each level of difficulty there are three patterns, each of which employs the same sized and shaped grid. The patterns are presented on a series of stimulus cards, with each participant seeing the card for three seconds, after which it is covered up, and they are required to replicate it on a response sheet, which consists of blank grids the same size and shape as the target pattern. Testing continues until the participant cannot correctly reproduce any of the three patterns that make up a particular level of difficulty. For the purposes of this experiment, participants were given a level score based on the difficulty level of the last set of patterns where all three patterns were correctly replicated.

The cognitive load condition consisted of 24 patterns, one for each statement, where the difficulty level of the pattern matched that of the level score given in the assessment test. The grids and patterns were based on those of the VPT, and an example can be seen in Figure 4.2.

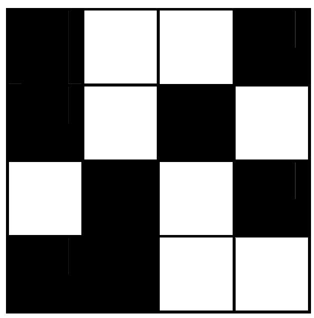


Figure 4.2. Example of a level 8 difficulty pattern used in Experiment 7

Procedure

The participants were run individually and received instructions on how to complete the tasks immediately prior to each task taking place. Testing began with the visual pattern assessment, followed by the two sets of statements, with and without load, and ended with the post test.

For the initial assessment of their visual pattern capacity, participants were told that their task was to replicate the patterns on the grids provided. They were told that the patterns would get progressively harder and not to worry if they got them wrong. They received feedback on their performance after each pattern. The assessment stopped when they got all the patterns wrong in a group of three, or if they reached the end.

For the no load condition the participants were told to answer true or false to a series of statements read out by the researcher and if they were unsure of an answer they should go with the one that they thought was the best fit. The answers were recorded by the researcher.

In the load condition the participants were presented with a pattern. Before replicating the pattern on a grid, the participant had to answer true or false to a stimulus statement, which was recorded by the researcher. The researcher monitored the participant’s performance on the replication of the pattern to ensure that they had sufficiently engaged with the task, and the participants received feedback on whether the pattern, but not the statement, was correct.

In the post test the participants were told to answer true or false to a series of statements read out by the researcher. At the end of the test if the participant had answered false to any of the statements they were asked to give a reason for this response.

4.3.3. Results

The data of one adult participant was removed before analysis because of failure to engage sufficiently with the pattern replication task in the load condition.

The responses of the remaining 52 adults and 30 children were coded with respect to their logical correctness as in the previous experiments. Analysis of the post test revealed that 65% of the adults (34/52) and 93% of the children (28/30) always responded true to the *true all* statements, thus these participants considered these statements to be under-informative when the statements were preceded by the term *some* in the experimental conditions. The analysis will focus on these participants[12].

Control Statements

Analysis of the control statements revealed that both sets of participants answered these well. The means (and standard deviations) can be seen in Table 4.1.

Table 4.1

Mean number of logical responses (and standard deviations) for the control statements, according to age and condition, for Experiment 7

	Condition	
	No Load	Load
Adult	19.82	19.91
(*n* = 34)	(0.46)	(0.29)
Child	19.79	19.75
(*n* = 28)	(0.42)	(0.44)

Note. The maximum score in each cell is 20

There was no significant difference in the mean number of logical responses given to the control statements between the adults and children for either condition, Mann Whitney *U* tests, load $U = 399.00$, $p = .087$, $r = .22$, no load $U = 447$, $p = .536$,

[12] If the data from those participants who did not consider all of the post test statements to be true is included in the analysis , the findings for this experiment still hold. See Appendix BVII for this analysis.

r = .08. Nor was there any difference in the number of logical responses given to the control statements between the load and no load condition within either age group, Wilcoxon matched pairs signed ranks tests, adults z = .91, p = .366, r = .11, children, z = .33, p = .739, r = .04.

Infelicitous some statements

The mean number of logical responses (and standard deviations) for the *infelicitous some* statements for each age group can be seen in Table 4.2.

Table 4.2

Mean number of logical responses (and standard deviations) for the infelicitous some statements, according to age and condition, for Experiment 7

	Condition	
	No Load	Load
Adult	5.56	5.82
(n = 34)	(3.30)	(3.14)
Child	4.64	4.39
(n = 28)	(3.67)	(3.75)

Note. The maximum score in each cell is 8

Adult responses. The nature of the distribution of the *infelicitous some* data for both conditions can be seen in Figure 4.3. As can be seen from the figure the distribution of the adults is unimodal in both conditions. The majority response, within each condition, is consistent logical responding. There is little pragmatic responding over-all and very little consistent pragmatic responding within each condition.

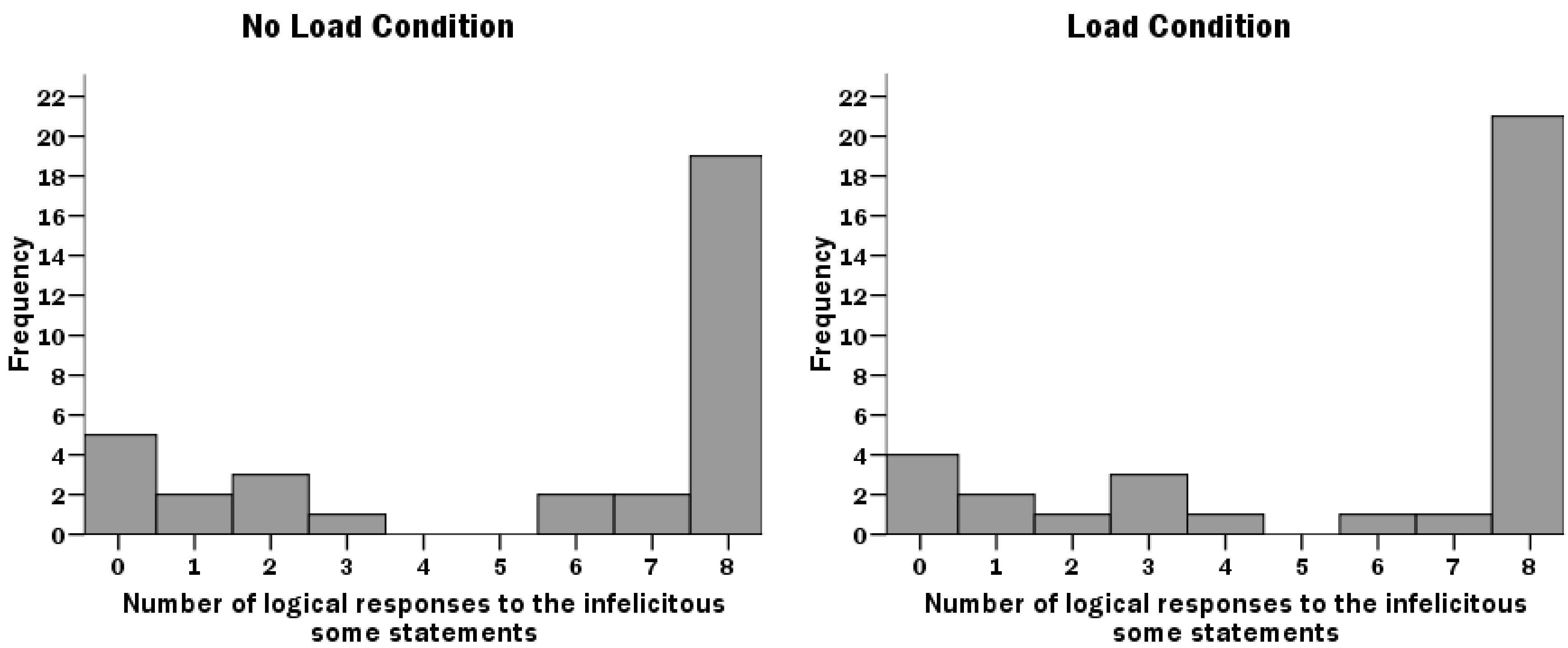


Figure 4.3. Distribution of adults according to the number of logical responses given to the infelicitous some statements, for both conditions, in Experiment 7.

With regards to the number of logical responses given to the *infelicitous some* statements between conditions, shown in Table 4.2, there was no significant difference, Wilcoxon matched pairs signed ranks test, z = 0.74, p = .457, r = .09.

The order in which the conditions were presented was also examined to see if it affected the rate of logical responding to the *infelicitous some* statements; the means (and standard deviations) for which can be seen in Table 4.3.

Table 4.3

Adults mean number of logical responses (and standard deviations) for the infelicitous some statements, according to condition and order of presentation, for Experiment 7

Adult	Order	Condition	
(n = 34)		No load	Load
	No load first	4.29	4.59
		(3.55)	(3.52)
	Load first	6.82	7.06
		(2.53)	(2.16)

Note. The maximum score in each cell is 8

This analysis revealed that compared to those adults who received the no load condition first, the adults who received the load condition first gave significantly more logical responses to the *infelicitous some* statements in both the load condition, Mann-Whitney test, U = 92.50, p = .040, r = .35, and the subsequent no load condition, Mann-Whitney test, U = 79.50, p = .014, r = .42.

Analysing the rates of logical responding for the condition that was presented first revealed that adults who attempted the load condition first gave significantly more logical responses to the *infelicitous some* trials (mean 7.06) compared to those adults who attempted the no load condition first (mean 4.29), Mann-Whitney test, $U = 75.5$, $p = .009$, $r = .45$.

Children's responses. The nature of the distribution of the *infelicitous some* data for both conditions can be seen in Figure 4.4. As can be seen from the figure the distribution of the children' responses is unimodal in both conditions, although not to the same extent as the adults. A quarter of the children are consistently pragmatic in the no load condition and there is a similar number of wholly pragmatic responders in the load condition.

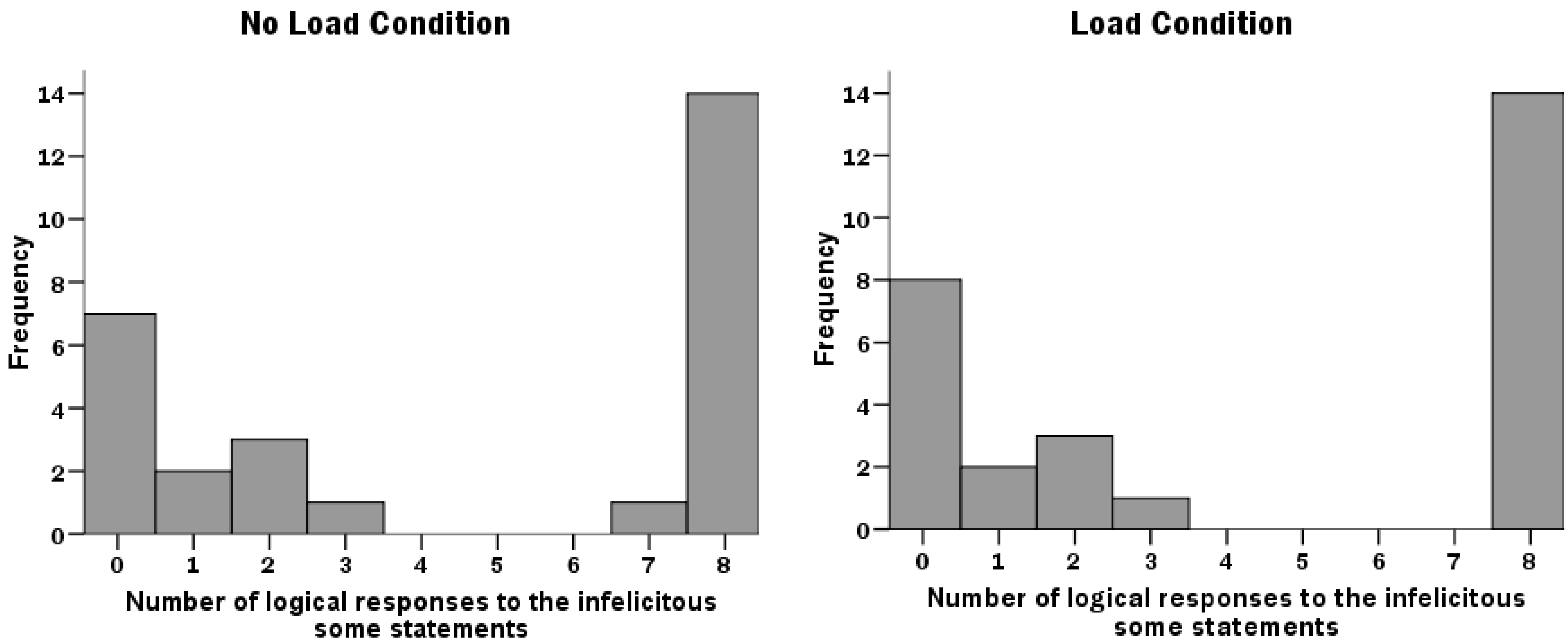


Figure 4.4. Distribution of children according to the number of logical responses given to the infelicitous some statements, for both conditions, in Experiment 7.

With regards to the number of logical responses given to the *infelicitous some* statements between conditions, the means of which can be seen in Table 4.2, like the adults there was no significant difference, Wilcoxon matched pairs signed ranks test, $z = 0.74$, $p = .457$, $r = .09$. However, unlike the adults, there was no effect of order of presentation, the means of which can be seen in Table 4.4, Mann Whitney tests, no load condition, $U = 74.00$, $p = .243$, $r = .16$, load condition, $U = 79.00$, $p = .356$, $r = .12$.

Table 4.4

Children's mean *number of logical responses (and standard deviations) for the infelicitous some statements, according to condition and order of presentation, for Experiment 7*

Children	Order	Condition	
($n = 28$)		No load	Load
	No load first	5.46	5.15
		(3.26)	(3.78)
	Load first	3.93	3.73
		(3.97)	(3.71)

Note. The maximum score in each cell is 8

Analysing the rates of logical responding for the condition that was presented first revealed that although the children who attempted the no load condition first gave more logical responses to the *infelicitous some* trials (mean 5.46) compared to those children who attempted the load condition first (mean 3.73), this difference was not significant, Mann-Whitney test, $U = 71.5$, $p = .204$, $r = .24$.

Comparison of adult and children's results

Although the adults gave more logical responses than the children to the *infelicitous some* statements in both conditions, see Table 4.2, these differences were not significant, Mann Whitney tests, no load, U = 422.50, p = .409, r = .11, load, U = 383.00, p = .144, r = .19. This is not the case though if load order is taken into account, see Tables 4.3 and 4.4. The adults and children who received the no load condition first did not differ significantly in the number of logical responses given to the *infelicitous some* statements in this condition, Mann Whitney test U = 83.50, p = .236, r = .22, or the subsequent load condition, Mann Whitney test U = 101.50, p = .680, r = .08, but there was a significant difference in the mean number of logical responses given to the *infelicitous some* statements between the adults and children who received the load condition first. The adults gave significantly more logical responses in both the load condition, Mann Whiney test, U = 68.50, p = .021, r = .44, and the no load condition, Mann Whitney test, U = 79.00, p = .034, r =.37.

4.3.4. Discussion

To summarise the results of this experiment: there was no effect of cognitive load on children's or adult's processing of scalar terms; there was an effect of load order in the adult participants, with those adults who received the load condition first giving more logical interpretations in both conditions compared to those adults who received the no load condition first; and when load order was taken into account, the children gave more pragmatic interpretations to the *infelicitous some* trials than the adults.

According to De Neys and Schaeken (2007), the addition of cognitive load interferes with adults' ability to draw scalar implicatures. If this is the case, then one would expect to see significantly more logical responding amongst adults in the load condition compared to the no load condition. This was not the finding of this experiment, although a between-participants comparison based on the first condition attempted by the participants revealed that there was significantly more logical responding in the load condition compared to the no load condition.

It could be argued that the changes made in this experiment to the design used by De Neys and Schaeken (2007) meant that the weak effect of load within participants was not detected. This experiment had fewer participants, the ratio of target to filler statements was decreased, a no load condition was used rather than a low load condition, and the cognitive load was adjusted in line with each participants visual cognitive capacity. These changes could also account for the difference in rates of pragmatic responding between the studies, 76 % in the De Neys and Schaeken study and 29% in this experiment. Indeed, it could be claimed that there was insufficient pragmatic responding in this experiment to uncover the small effect of load. However, the between-participants effect of load, and the effect of order in the adult participants both suggest that cognitive load does affect adult participants' processing of scalar terms. The latter finding also suggests that the effect of load is more complex than was claimed by De Neys and Schaeken.

The participants who received the load condition first gave significantly more logical responses in both conditions compared to those participants who received the no load condition first, but within participants there was no difference between conditions. This suggests that the high level of logical responding in the no load condition, when it followed the load condition, was the result of these adults developing a strategy in the load condition and sticking with it throughout the experiment.

The need for a strategy can be explained in terms of effort. If there are two sources of effort for the adult, the processing of the implicature and deciding whether to give a pragmatic or logical response, then having already experienced the demands of the secondary task in the visual assessment, the adults may decide to conserve effort by deciding to always answer logically when faced with the term *some,* and as it serves them well in the load condition they continue to use this in the no load condition. If the adults had not adopted a strategy then the same level of logical responding would be expected in both of their conditions.

The suggestion of an effortful process of deciding between response alternatives in adult participants could also explain De Neys and Schaeken's (2007) findings. As their sample consisted of university undergraduates it seems unlikely that none of their participants had the analytical ability to compute the implicature whilst being aware that it could be cancelled. The small increase (6%) in logical responding observed under load in their study could reflect a strategic switch to logical responding in a relatively small number of participants. Unfortunately, as De Neys and Schaeken's results do not contain any analysis of order effects, it is not possible to comment on whether there was evidence of strategic responding in their sample.

With regards to the processing of scalar terms in children, the results of this experiment suggest that when 10-year-old children draw the implicature they do so with little difficulty, as their response rates are not affected by load, and there is no evidence of strategy formation. This absence of difficulty in the children suggests that their processing may be different to that of the adult participants. This proposal is further supported by the finding that when the logical response rates of those participants who received the load condition first are considered, the children give more pragmatic responses than the adults in both conditions. This appears to indicate that the processing of scalar implicatures can be more effortful for adults than it is for children, which goes against current assumptions in the literature. This idea is consistent with the claim that the children, unlike the adults, have only one response available; that is, either they do not detect the implicature and so respond logically, or detect it and respond pragmatically, because once the implicature has been detected they are unable to cancel it.

The findings of this experiment raise the possibility that adults' processing of scalar terms is different to children's. However what remains unanswered is whether detection of the implicature in adults is more effortful than in children. That is, because children have only one response available to them, the pragmatic interpretation may effectively be a default for them. For adults, on the other hand, because more than one response is available, the inference does not go through automatically. Another possibility is that the developmental difference that was observed may be due to the effort some adults have to exert in deciding between the pragmatic and logical response. Children who have only one response available to them do not experience this cognitive cost. The next experiment was designed to address these possibilities.

4.4. Experiment 8: The time course of scalar implicatures

4.4.1. Introduction

The findings of the previous experiment suggested that children and adults may process scalar terms differently; both age groups are able to detect the implicature but only the adults are able to cancel the implicature once it has been made. Previous research (Bott & Noveck, 2004; De Neys & Schaeken, 2007;

Noveck & Posda, 2003) has indicated that adults' pragmatic interpretations of *infelicitous some* take longer than logical interpretations, and if the assumption in the literature that children's processing costs will be equal to or greater than that of the adults holds, then one would expect to see the same trend in children. One of the main aims of this experiment was to test this prediction on the basis that if, as the previous experiment suggested, children have little difficulty processing scalar implicatures, then the cost to the children may be less than to the adults, and the opposite result may be seen.

Another main aim of this experiment was to investigate the proposal that because some adults are aware that either of two responses is possible, they incur a response selection cost that is not incurred by children. To this end the time courses of logical responses to the *infelicitous* and *felicitous some* statements were examined. If the proposal holds, one would expect to see logical responses to the *infelicitous some* statements taking longer than the same response to the *felicitous some* statements in adults, but not in children.

An additional aim was to examine the relationship between the type of interpretation given to the *infelicitous some* statements and participants' working memory capacity. Given that Feeney et al. (2004) found that memory span was positively correlated with logical response rates to *infelicitous some*, and there have been similar claims in other work on individual differences in reasoning (Newstead, Handley, Harley, Wright & Farrelly, 2004; Stanovich & West, 1998; Torrens, Thompson & Cramer, 1999) the same relationship was predicted for the adults in this study. However, with regards to the children the opposite relationship was predicted. Taking into account Noveck's (2001) suggestion that logical interpretations give way to subsequent pragmatic interpretations, the developmental pattern that was seen for the sensitivity to scalar implicatures in the experiments of Chapter 2, and the finding of Morsanyi and Handley (2008) that heuristic responding in children is based on cognitive capacity, it was predicted that logical responding in the children would be negatively associated with counting span.

4.4.2. Method

Participants

Thirty adults and thirty seven children participated in the experiment. The adults were recruited through a campaign at Durham University and an advertisement in a local magazine. Their ages ranged from 18 years to 55 years (mean age 29 years) and all gave their own consent. The children were aged between 10 years 10 months and 12 years 9 months (mean age 11 years 9 months) and were recruited from Year 6 and Year 7 classes from two schools in the North East of England. Consent for their participation was given by either their parents or the head teacher.

Materials and Procedure

Each participant completed two tasks; a statement verification task consisting of an initial set of 48 statements and an 8 statement post test, and a working memory counting span task. The set of 48 statements and the counting span task were presented on a computer screen, whilst the post test was read out loud by the researcher.

Statement task. The 48 statements used were based on the ones used in the previous experiments. Each participant received 8 target *infelicitous some* statements and 40 control statements, which were comprised of equal numbers of *true all*, *false all*, *absurd all*, *felicitous some* and *absurd some* statements (see Appendix AVIII for statements used).

Each statement was presented with either *all* or *some* preceding it, although each participant received only one form, which resulted in two sets of materials. The order of the statements within each set was randomly assigned by the computer. The participant responded true or false to each statement using a button box, this response and the time taken to make it were recorded by the computer. In an attempt to equate reading times between statements, each statement was four syllables long.

As in the previous experiment, in order to ascertain that each participant was considering the whole set and not a partial set in relation to the *infelicitous some* statements, a post test was carried out. The post test consisted of the eight *infelicitous some* statements which had previously been presented to the participant only this time the quantifier *some* was replaced by *all,* and the sentences were presented by the researcher. The participant again had to respond true or false to the statements. If they responded false they were asked for an explanation.

Counting span task. Each participant had their working memory capacity measured through a counting span task. Counting span tasks are reliable, valid, and widely-used measures of Working Memory (WM) capacity and suitable for a number of populations (Conway et al., 2005). The task is designed to “force WM storage in the face of processing in order to engage executive attention processes” (Conway et al., p. 773) and has been used in previous studies of individual differences and reasoning abilities (e.g. Barrouillet & Lecas 1999; Handley, Capon, Beveridge, Denis, & Evans, 2004). The counting span task used here was a computerised version of that used by Case, Kurland and Goldberg (1982). Each participant was presented with screens containing blue and green dots, see Figure 4.5. For each screen, their task was to count out loud the number of blue dots, whilst pointing at them, and to remember the total for each screen. The number of blue dots ranged from 2 to 7. The number of screens presented in a series ranged from two to five screens. Once the participant had seen all the screens in a series they had to recall out loud the totals from each screen; these totals were recorded by the researcher. Participants saw three sequences of two, three, four and five screens.

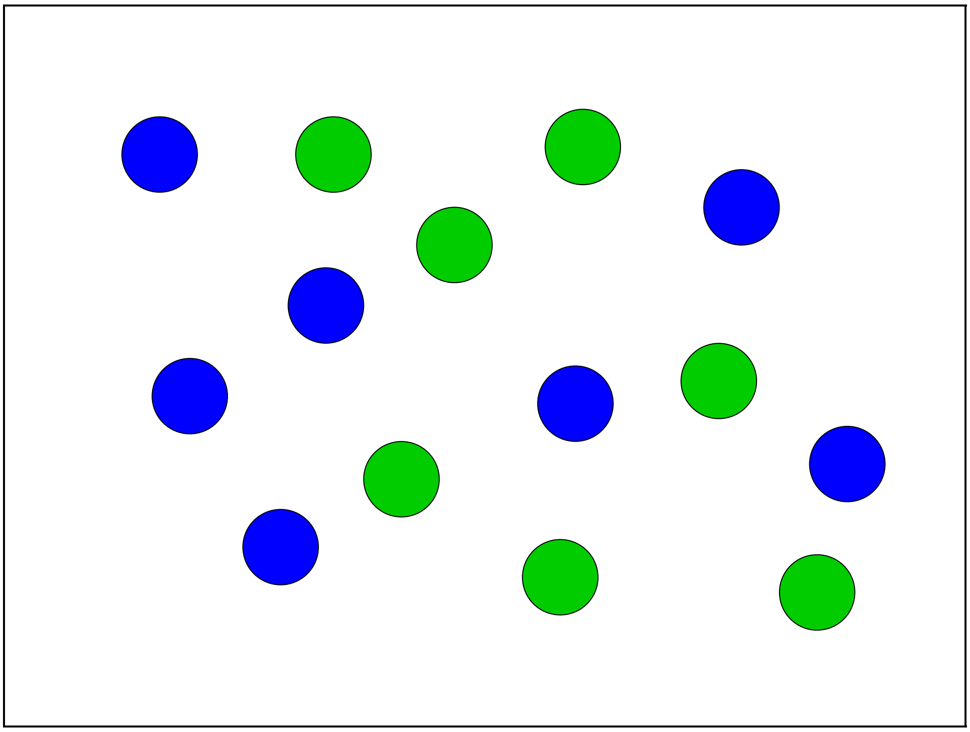


Figure 4.5. An example of a screen used in the counting span task of Experiment 8.

4.4.3. Results

Data treatment

Statement verification task. As in previous experiments the participants' responses were coded with respect to their logical correctness. In order to ensure that participants had not traded speed for accuracy, participants were excluded if they got at least three absurd statements wrong, or at least four other control statements wrong in one sub-group. This resulted in the loss of the data from one adult participant and two children. In addition, only those control statements that were correct were included in the reaction time analysis. This resulted in the loss of 5.4% of the data from that analysis. Also, those responses to control and target statements that were considered to be outliers, that is, 3 standard deviations from the age group mean reaction time for that type of statement were also removed from both the behavioural and reaction time data. This resulted in a loss of 2.3 % of the control data and 3.9% of the responses to *infelicitous some* data. Finally, with regards to the *infelicitous some* data, only those statements that the participant considered to be true of *all,* as recorded in the post test, were included in the analysis, this resulted in the loss of 4.7 % of the data.

Participants' scores were converted into proportions, in order to take into account the results of the post test, thus for the *infelicitous some* trials the figure reported is the number of logical responses as a proportion of the number of statements the participant considered to be true of *all*.

Counting span task. A participant's counting span was measured according to the formula 1 + (n / 3) where n is the total number of correctly recalled sets (Barrouillet & Lecas, 1999). Sets were only considered up to the point where the participant failed to recall all three sets at a particular level (Case, Kurland & Goldberg, 1982).

Data analysis

Behavioural rates of responding. The mean proportions of the logical responses according to truth, term and age group are shown in Table 4.5.

Table 4.5

Mean proportions (and standard deviations) of logical responses for each statement type, according to age group, in Experiment 8

	Statement			
	True All	False All	Felicitous Some	Infelicitous Some
Adult	.92	.93	.96	.63
(n = 29)	(.09)	(.10)	(.06)	(.36)
Children	.88	.87	.92	.81
(n = 35)	(.11)	(.13)	(.11)	(.29)

In the adults significantly fewer logical responses were given to the *some* statements (mean proportion .80) compared with the *all* statements (mean proportion .93), Wilcoxon matched pairs signed ranks test, z = 2.58, p = .010, r = .34, and to the *false/infelicitous* statements (mean proportion .78) compared with the *true/felicitous* statements (mean proportion .95), Wilcoxon matched pairs signed ranks test, z = 3.89, p < .001, r = .51. However this was not the case for the children. There was no significant difference between the mean proportions for the *some* (mean proportion .87) and *all* (mean proportion .88) statements, Wilcoxon matched pairs signed ranks test, z = 0.58, p = .560, r = .07, or the *true/felicitous* (mean proportion .90) and *false/infelicitous* (mean proportion .84) statements, Wilcoxon matched pairs signed ranks test, z = 1.85, p = .065, r = .22.

The mean proportion of logical responses given to each type of statement was analysed using Friedman tests and was found to be significant for the adults, χ^2 (3) = 19.94, p < .001, but not for the children, χ^2 (3) = 4.16, p = .245. A follow up

Wilcoxon matched pairs signed ranks test for the adults revealed that they gave fewer logical responses to the *infelicitous some* statements than any other statement type, all *p's* $\leq$ .001, all *r's* $\geq$ .42 (see Appendix BVIII, Table 1, for individual statistics).

The nature of the distribution of the *infelicitous some* data for both age groups can be seen in Figure 4.6. In both age groups the majority response was to be consistently logical, with this being more pronounced in the children. There was evidence of responding right across the scale for both age groups, although this was more pronounced in the adult group.

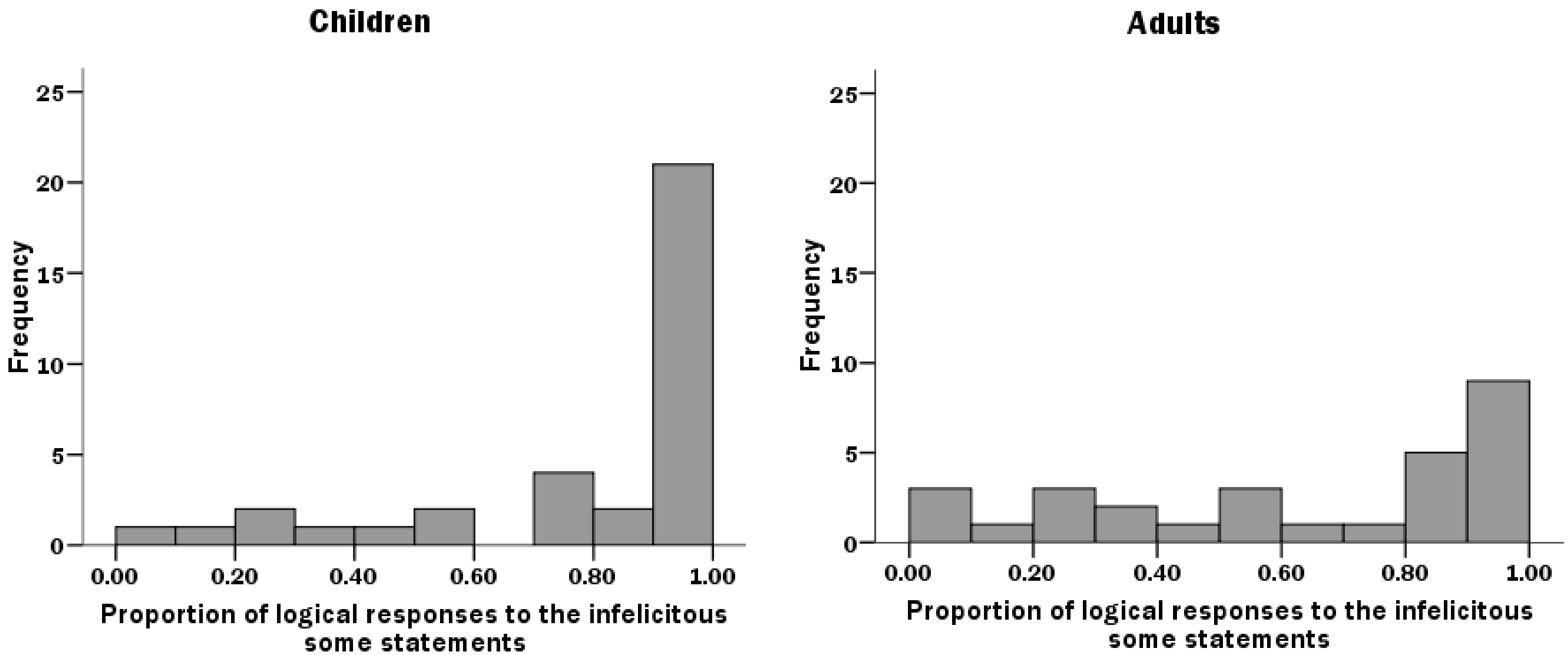


Figure 4.6. Distribution of participants according to the proportion of logical responses given to the infelicitous some statements by both age groups in Experiment 8

With regards to the mean proportion of logical responses given to the *infelicitous some* statements between the age groups, the adults gave significantly fewer than the children, Mann-Whitney test, $U = 344$, $p = .020$, $r = .29$.

Reaction time data. The mean reaction times for logical responses to the control statements were analysed using unrelated *t* tests, which revealed that for each type of control statement the children were significantly slower than the adults, all *p's* < .001, all *r's* $\geq$ 0.58 (see Appendix BVIII, Table 2, for the analysis of each type of control statement). To take this difference into account, reaction times were considered to be outliers if they were 3 standard deviations from the age group mean reaction time for each type of trial.

Participants were divided into three groups depending on the type of responses they gave to the *infelicitous some* statements; all logical responses, all pragmatic responses, or a mixture of both responses. The number of each type of responder for each age group can be seen in Table 4.6.

Table 4.6

The number of participants for each type of responder, for each age group, in Experiment 8

Age Group	Type of Responder		
	Logical	Pragmatic	Mixed
Adult	9	3	17
Children	21	0	14

One of the experiment's aims was to investigate whether pragmatic responses to *infelicitous some* statements take longer than logical responses. As there were so few pragmatic responders to compare with logical responders, the reaction times of the logical and pragmatic responses of the mixed responders were compared. The mean reaction times of the mixed responders for each type of statement can be seen in Table 4.7.

Table 4.7

Mean reaction time in milliseconds (and standard deviations) for the control and target statements of the adult and child mixed responders for Experiment 8

	Statement				
	True All	False All	Felicitous Some	Infelicitous Some (logical)	Infelicitous Some (pragmatic)
Adult	1480	1438	1479	1707	2053
(*n* = 17)	(329)	(300)	(292)	(566)	(758)
Children	2446	2224	2436	2542	3160
(*n* = 14)	(422)	(294)	(515)	(468)	(738)

Analysis of the reaction times of the *infelicitous some* statements for the mixed responders revealed that pragmatic responses took significantly longer than logical responses in both age groups, related *t* tests, adults, t (16) = 2.09, p = .053, r = .50, children, t (13) = 3.61, p = .003, r = .70. The proportional increase (pragmatic reaction time / logical reaction time) in pragmatic response times compared to the logical response times was compared between the age groups. An independent *t* test revealed that there was no significant difference between the groups, t (29) = 0.146, p = .885, r = .03.

A second aim of the experiment was to examine whether logical responses to both types of *some* statements differ. Amongst the mixed responders related *t* tests revealed that there was no significant difference between the mean reaction times for the logical responses to the *infelicitous* and *felicitous some* statements for either age group, adults t (16) = 1.81, p = .089, r = .41 children t (13) = 0.73, p = .476, r = .20. The effect size in the adult age group coupled with the p value suggests that there is a moderate effect which the small sample size may be obscuring.

The reaction times of the logical responders were also analysed. The mean reaction times for each type of statement for this group of responders can be seen in Table 4.8.

Table 4.8

Mean reaction time in milliseconds (and standard deviations) for the control and target statements of the adult and child logical responders for Experiment 8

	Statement			
	True All	False All	Felicitous Some	Infelicitous Some (logical)
Adult	1316	1340	1209	1393
(n = 9)	(360)	(316)	(416)	(321)
Children	2299	2256	2227	2370
(n = 21)	(447)	(413)	(434)	(535)

The mean reaction times for the logical responses to the *infelicitous* and *felicitous some* statements were compared. Related t tests revealed that there was no difference in the reaction times of the children, t (20) = 1.61, p = .124, r = .33, but there was a difference for the adults, t (9) = 2.95, p = .019, r = .70, with the response times to the *infelicitous some* statements being significantly slower than the response times to the *felicitous some* statements.

With regards to the pragmatic responders, the extremely small sample size, $n = 3$, negates the possibility of meaningful statistical analysis. However as can be seen from Table 4.9, the mean reaction times for their pragmatic responses to the *infelicitous some* statements were longer than for any of the control statements.

Table 4.9

Mean reaction time in milliseconds (standard deviations) for the control and target statements of the adult pragmatic responders for Experiment 8

	Statement			
	True All	False All	Felicitous Some	Infelicitous Some
Adult	1869	1587	1561	2263
(n = 3)	(400)	(199)	(79)	(442)

Individual differences. The mean counting span scores achieved by the two age groups was 3.80 (sd = .66) for the adults and 3.09 (sd = .66) for the children, and the difference between these two scores was statistically significant, Mann-Whitney test, $U = 229$, $p \leq .001$, $r = .48$.

The correlations between counting span and logical responses to the statements can be seen in Table 4.10.

Table 4.10

The correlations between counting span and logical responses to the statements, according to age, for Experiment 8

		True all	False all	Felicitous some	Infelicitous some
Counting span	Adult	-.076	.240	-.181	-.211
		(.693)	(.210)	(.348)	(.273)
	Child	-.219	-.131	.166	.017
		(.207)	(.453)	(.340)	(.921)

Note: Probabilities in parentheses

As can be seen from Table 4.10 counting span does not significantly correlate with the number of logical responses given to any of the statement trials for either age group.

The counting span scores of the mixed responders were compared to those of the logical responders and were found not to be significantly different for either age group, all *p's* > .296 (see Appendix BVIII, Table 3, for individual age group statistics).

The counting span scores were also correlated with the mean reaction times for a pragmatic response to the *infelicitous some* statements in both age groups and were found not to be significant for either group, adults, Pearson's r = .061, p = .797, children, Pearson's r = -.268, p = .353.

The correlation between the counting span scores of the adult logical responders and the mean reaction times for a logical response to the *infelicitous some* statements was also computed but there was no significant relationship, Pearson's r = -.110, p = .778.

4.4.4. Discussion

To summarise the results of this experiment: pragmatic responses to the *infelicitous some* statements took longer than logical responses in the mixed responders of both age groups; logical responses to the *infelicitous some* statements took longer compared to the *felicitous some* statements, but only in the adult participants, not in the children; and no relationship was found between WM capacity and the type of response given to the *infelicitous some* statements.

According to previous research (Bott & Noveck, 2004; De Neys & Schaeken, 2007; Noveck & Posda, 2003) adults' pragmatic responses to *infelicitous some* statements take longer than logical responses and the results of the adult mixed responders of this experiment support this idea. Interestingly, despite children's apparent lack of difficulty with scalar implicatures in the previous experiment, the children's pragmatic responses to the *infelicitous some* statements took longer than their logical responses. Furthermore the processing cost to the children was proportionally the same as that for the adults. These results provide evidence against Levinson's (1983, 2000) default view that pragmatic interpretations are automatic.

However the results of the adult logical responders in this experiment suggest that processing is more complex than has previously been claimed. The responses of this group of adults to the *infelicitous some* statements took longer than the same response to the *felicitous some* statements. This finding suggests a number of possibilities: It could be seen as supporting Feeney et al.'s (2004) claim that a logical response can be the result of the cancellation of a scalar implicature and not necessarily an indication that the implicature has gone undetected. The difference could also be reflecting an awareness that there is more than one type of response available. Alternatively the slow-down could represent a choice in interpretation with implicit conflict detection of the two responses taking place (De Neys & Glumicic, 2008)

The finding that there was no difference between the response times to the *felicitous* and *infelicitous some* statements for the children who were classed as logical responders, suggests that these children are neither detecting the implicature and cancelling it, nor have more than one response available.

So the picture that appears to emerge from this experiment is that in children the implicature either goes undetected, in which case they answer logically, or the implicature is detected and the children answer pragmatically. Adults also follow this processing course; however for some adults there is an additional source of difficulty, which is related to the availability of conflicting logical and pragmatic responses.

The lack of a significant relationship between counting span and the number of logical responses given to the *infelicitous some* statements fails to support Feeney et al.'s (2004) proposal that extra cognitive processing, of the kind required to cancel the implicature as opposed to that required to detect the implicature, is more likely to be found in individuals of higher cognitive capacity. However, there is a problem with this type of analysis. High logical response rates are also associated with failure to detect the implicature and as seen in the experiments in this thesis involving very young children, you do not have to be high in cognitive capacity to fail to detect a scalar implicature. There is no way of determining how a sample of logical responders is divided between those who are insensitive to the implicature and those who have cancelled it. In Feeney et al.'s study the sample consisted entirely of undergraduates and half of these were logical responders, just the type of person who might be expected to be high in cognitive capacity and capable of detecting the scalar implicature and cancelling it. Whereas in this experiment the sample contained participants of varying educational levels, only a third of whom were logical responders. Therefore it might be expected that more of the participants in this study were insensitive to the implicature which would have an adverse effect on a correlational analysis. In addition the sample in this experiment was very small for such analysis.

One surprising aspect of this experiment was the high number of logical responses given to the *infelicitous some* statements by the children. Given that universal pragmatic responding was seen in children aged not much older than these in Experiment 2, more pragmatic responding was expected. However the high logical response rate may be due to the methodology. The logical response rates for the control sentences were lower than would normally be seen; proportions in Experiment 2 were close to, if not, 1, however in this experiment they fell between .87 and .92. It could be that the children were trading speed for accuracy, and there is evidence that under conditions of speeded response people are less likely to detect scalar implicatures (Bott & Noveck, 2004).

4.5. Chapter Discussion

The two experiments presented in this chapter suggest that there are differences in adults and children's processing of scalar terms and that for some adults effort is required for more than the detection of the implicature. However the nature of these effort-demanding processes is still somewhat unclear.

One of the main findings of the experiments is the further evidence to support the view that a logical response in adults does not necessarily mean that the implicature has gone undetected. The use of a response strategy by adults in the dual task paradigm of Experiment 7, and the longer mean reaction times to the *infelicitous some* statements as compared to the *felicitous some* statements in the adult logical responders in Experiment 8, suggest that a logical response can be effortful. If the implicature had not been detected by the adults it would not require processing, therefore one plausible explanation to account for the effort expended is that it is required to cancel the implicature. This proposal goes against the conventional view that a logical response is evidence that the implicature has not been detected (De Neys & Schaeken, 2007; Guasti et al., 2005; Noveck, 2001) and has implications for how responses are interpreted. Against this interpretation, however, is the finding that amongst mixed responders, logical responses to the *infelicitous some* statements took less time than pragmatic responses. The opposite finding would be predicted if participants were effortfully detecting and then effortfully cancelling the implicature. An alternative possibility is that some participants, aware that two responses are possible, make a choice, on a trial by trial basis about whether to respond pragmatically or logically. However, the order effects in Experiment 7 are problematic for this explanation as they suggest that participants adopt a response strategy early on. A final possibility is that some participants respond strategically but the presence of conflict between pragmatic and logical responses leads to a slowdown in reaction times. This account is consistent with the data in this experiment and consistent with recent findings of De Neys and colleagues (De Neys & Glumicic, 2008; Franssens & De Neys, 2009).

Another important finding concerns the processing of scalar implicatures by children. The finding that children are unaffected by load even though the pragmatic interpretation takes longer to process than a logical interpretation suggests that children have only one response once the implicature has been detected. Thus, although the experiments support the standard view that the detection of scalar implicatures requires effort (Bott & Noveck, 2005; Noveck, 2001, Noveck & Posda, 2003; Pouscoulous et al., 2007), how much effort is difficult to say.

Taken together, the findings of the experiments go against the conventional view that the processing of implicatures is always more effortful for children than adults (Noveck, 2001; Papafragou & Musolino, 2003). Whilst it may be true that initial detection of implicatures is cognitively demanding for young children, it does appear that older children experience less difficulty in processing scalar implicatures than some adults do.

The results of these experiments have implications for default and relevance based accounts of scalar implicature processing. The reaction time study indicates that the scalar implicature associated with *some* is not an automatic interpretation; a pragmatic interpretation takes longer to produce than a logical interpretation that represents failure to detect the implicature. This goes against Levinson's (1983, 2000) default account. However both studies show that logical responding can cause adults additional processing problems in terms of time and effort and the best explanation of this is that the logical response represents a cancellation of the implicature. This would be in keeping with the default view that the undoing of the implicature requires time and effort. However this type of processing is not necessarily ruled out by a relevance theory view. RT allows for expectations of relevance to be readjusted during the comprehension process (Sperber & Wilson, 2002), which appears to not discount the idea that once an implicature has been computed it can be revaluated and cancelled. However what is ruled out by RT is the idea that conflict could cause a slow down in reaction times. For RT two interpretations should never be activated at once as the effort of having to choose between them goes against the principle of optimal relevance.

One problem with the reaction time methodology employed in Experiment 8 is that it may lead to a trade off between speed and accuracy. Clearly, for the reaction time data to reflect underlying processes, participants need to answer as soon as they have processed the statement. However if the participants put themselves under a time constraint, in that they think that they should answer as quickly as possible, then there is the possibility that this will result in an increased number of scalar implicatures going undetected (Bott & Noveck, 2005). The children’s poorer than expected correct response rates to the control statements and the high rates of logical responding to the *infelicitous some* statements suggests that they had traded speed for accuracy.

In summary, the processing of scalar implicatures is multifaceted. Logical responses in adults can indicate either the failure to detect the implicature or the cancellation of it. Logical responses in children however, appear to indicate that the scalar implicature has gone undetected. When children do detect scalar implicatures, then a pragmatic response takes longer to compute, but how effortful this processing is remains unclear given the finding that detection of scalar implicatures is ubiquitous for some age groups in previous experiments in this thesis. These findings have implications for theoretical accounts of scalar implicature processing as well as for proposals about the developmental trajectory of sensitivity to scalar implicatures.

Chapter 5

General Discussion

5.1 Introduction

Although the implicature associated with the quantifier *some* has long been acknowledged (Mill, 1867, as cited in Horn, 2004; Wilkins, 1928), there exists only a limited empirical data base. An additional problem has been the difficulty in comparing existing studies because of differences in methodology, languages, and ages studied, or the manipulation of a number of variables within the same study. This has led to two problems, seemingly contradictory findings for a number of research questions and a reliance on untested assumptions to interpret new results.

The primary aim of this thesis was to clarify, and thus extend, our existing knowledge of the scalar implicature associated with *some*. To do this three research questions were asked relating to the developmental trajectory of sensitivity to scalar implicatures, the role of context in facilitating sensitivity and contemporary pragmatic theories that seek to explain the phenomenon.

Eight experiments were conducted with a range of age groups, using a number of methodologies. A summary of the experimental results is presented in section 5.2., and they are interpreted in relation to the three research questions in subsections of 5.3. The implications of the findings of this thesis in relation to current pragmatic theories of scalar implicature processing and the field of experimental pragmatics are covered in section 5.4., and the limitations of the experiments are discussed in section 5.5. Future lines of enquiry are mentioned throughout the General Discussion but three particular areas that warrant further investigation are discussed in section 5.6. Finally, concluding remarks are presented in section 5.7.

5.2. Summary of Results

Although the results of the experiments were presented in the relevant chapters, for ease of exposition a brief summary of the results for each experiment follows.

Experiment 1 examined interpretations given to the scalar terms *some* and *or* by children from Years 1 and 4, and a group of adults, using statement evaluation and storyboard methodologies. With regards to the quantifier results, the number of logical responses given to the target *infelicitous some* trials differed from the number given to the control trials for the Year 4 children and the adults, but not for the Year 1 children. However in the storyboard task the logical response rate of the *infelicitous some* trials differed from that of the control trials for all age groups. Both groups of children gave more pragmatic responses in the storyboard task, but the adults showed no dissociation in context. Surprisingly, the Year 4 children had a higher pragmatic response rate for the *infelicitous some* trials in the storyboard task, than the adults. With regards to the results for *or*, although it appeared that the children were showing sensitivity to the scalar implicature in the storyboard task, poor logical response rates for the control trials meant that it was not possible to say whether the findings from *some* generalised to *or*.

Experiment 2 followed up the surprising finding from the initial experiment that children appeared to be more pragmatic than adults in the storyboard task. The experiment examined interpretations given to *some* in children from Years 1, 4, 7 and 10, and a group of adults using the statement and storyboard tasks. Logical response rates in the statement task for the target *infelicitous some* trials differed from those of the control trials for all age groups except the Year 1 children. However in the storyboard task the number of logical responses given to the *infelicitous some* trials was lower than the control trials for all the age groups. Once again the Year 4 children gave more pragmatic interpretations to the *infelicitous some* trials in the storyboard task compared to the adults. In addition, in the same task the Year 7 children also had a higher pragmatic response rate than the adults. Moreover, this group of children had a higher pragmatic response rate compared to the adults in the statement evaluation task. As in the previous experiment the Year 1 and Year 4 children gave more pragmatic interpretations to the *infelicitous some* trials in the storyboard task compared to the statement evaluation task.

Experiment 3 expanded the age range of the previous experiments to look at rates of sensitivity in preschool children using the statement evaluation task and a novel pictureboard task. None of the children showed sensitivity to scalar implicatures in the statement task, whereas the results for the pictureboard task indicated that some of the children were sensitive to the implicature.

Experiment 4 followed up the findings of the previous experiment and, in addition to the pictureboard task, introduced a new bag methodology to the preschool children. Whilst the results of the pictureboard task remained inconclusive, the new methodology resulted in lower rates of logical responding to the *infelicitous some* trials compared with the control trials. An unexpected finding was the effect of a one week gap in the presentation of the second task. Children gave more pragmatic interpretations to the *infelicitous some* trials for the task they attempted second regardless of what that task was.

Experiment 5 took a different approach to investigating sensitivity by manipulating conditions within the storyboard methodology to examine the effect of context on Year1 and Year 4 children. The experimental manipulation was the explicit mention of an agent in one condition but not the other. Rates of logical responding to the *infelicitous some* trials were lower than the control trials for both age groups in both conditions. The Year 1 children gave more pragmatic interpretations when the agent was explicitly mentioned, but no dissociation was seen in the Year 4 children, who had universal pragmatic responding. Effects of order were seen in both age groups and practice effects were observed in the Year 1 children.

Experiment 6 followed up the results of the Year 1 children in the previous experiment to examine the effect of motivation to deceive on sensitivity to scalar implicatures. A new set of storyboards were introduced, which involved description of, rather than interaction with, a set of items. Both sets of storyboards had agent and non-agent conditions. Rates of logical responding to the *infelicitous some* trials were lower compared to the control trials in both conditions of the interactive set of storyboards, but not for the descriptive set. The children also gave fewer logical responses to the interactive set regardless of condition. Order and practice effects were observed in the interactive set but these did not occur in the descriptive set.

Experiment 7 followed up a finding from De Neys and Schaeken (2007) to look at the effect of cognitive load on interpretations given to *some* in a statement evaluation task by Year 6 children and a group of adults. The dual task methodology revealed that whilst there was no effect of cognitive load overall on adult's interpretations, there was an effect of load order, with more logical responding being seen in both conditions for those adults who received the load condition first. Interestingly no effects of load or load order were seen in the children. In addition, when the effect of load order was taken into account, the adults who received the load condition first gave more logical responses in both conditions compared to the children.

The final study, Experiment 8, measured reaction times in a statement evaluation task and examined the relationship between behavioural responses and working memory capacity in 10-12-year-olds and adults. With regards to the behavioural responses fewer logical responses were given to the *infelicitous some* trials compared to the control trials by the adults but not the children, and no relationship was found between counting span and the type of interpretation given to *infelicitous some*. Analysis of the reaction times revealed that for those participants who gave both logical and pragmatic responses to *infelicitous some*, pragmatic interpretations took longer regardless of age. For those participants who gave only logical responses to the *infelicitous some* trials, reaction times to these trials were longer than to the *felicitous some* trials, but only for the adult participants.

5.3. Interpretation of Findings

This thesis sought to answer three research questions. The findings from the experiments are interpreted and discussed in terms of those questions.

5.3.1. What is the developmental trajectory of sensitivity to scalar implicatures?

The findings from a number of experiments in this thesis have shown that the answer to the above question is more complex than had previously been assumed. It appears from these experiments that sensitivity is not simply a case of logical interpretations being supplanted by pragmatic ones (Noveck 2001).

Sensitivity to scalar implicatures appears to emerge around 3-years-of-age (Experiments 3 & 4), which is younger than has previously been documented in the literature for tasks that set up conditions of under-informativeness and where interpretations of the scalar term can be evaluated against control trials. However sensitivity at this age is very much context and task dependent, and whilst children are usually consistent in their interpretations within tasks this is not the case across tasks. Thus if a scalar term is set in a pragmatically enriched context which directly involves, and has implications for, the child participant (Experiment 4) then sensitivity is much more likely to be seen than if the context is pragmatically impoverished, such as the statement evaluation task, where the cognitive effects to be gained are minimal (Experiments 1, 2 & 3). Indeed in such impoverished contexts, sensitivity to scalar implicatures is not always seen in older children (Experiment 8) and some adults (Experiments 1, 2 & 8), which is consistent with previous research (Feeney et al., 2004; Guasti et al., 2005; Noveck, 2001, Pouscoulous et al., 2007)

Children up to the age of 11 often show a dissociation in context (Experiments 1, 2, 3 & 5) indicating that whilst sensitivity to scalar implicatures is developing it is not an all or nothing matter, and shows the reliance of children on other pragmatic cues to detect the implicature, rather than just the mere presence of the scalar term. However, the same could also be said of some adults, whose ability to detect scalar implicatures would be considered to be developed. Whilst a dissociation in context was not observed for the adults as a group (Experiment 2), individual patterns of responding show that there are some adults whose interpretation of the scalar term changes according to the context in which it appears.

Sensitivity to scalar terms appears to develop in a non-linear manner, which is contrary to the assumptions in the literature (Noveck, 2001; Papafragou & Musolino, 2003). One of the most important findings of this thesis has been the discovery that the development of sensitivity to scalar implicatures appears to be U-shaped, and in opposition to the popular belief that children should never be more pragmatic than adults (Noveck, 2001; Papafragou & Musolino, 2003), lower rates of logical responding in children, as compared with adults, were seen in three experiments in this thesis (Experiments 1, 2 and 7). This finding raises a number of questions in relation to what a logical response actually represents and has implications for theoretical accounts of scalar implicatures, which are discussed in sections 5.3.3 and 5.4, and the interpretation of previously published findings.

5.3.2. What contexts facilitate sensitivity to the implicature associated with some?

The initial focus on context in this thesis was specifically on the facilitating effect of deception contexts in implicature detection. However in considering those contexts a number of factors that had previously been implicated in aiding sensitivity were also examined. Although Feeney et al.’s (2004) claim for the facilitatory nature of deception in relation to the detection of scalar implicatures gained support from a number of experiments (Experiments 1, 2, 4 & 5), it also became apparent that deception alone is not enough to elicit sensitivity. The findings from Experiment 6 indicated that a scalar implicature will not necessarily be detected if an utterance is under-informative or if someone is attempting to deceive their interlocutor. Another important factor appears to be that there are benefits to be gained from the deception. This finding, considered alongside Feeney et al.’s proposal, led to the suggestion of another possible facilitating factor; the consequences of a particular interpretation. In other words, maybe sensitivity is facilitated by any scenario where the participant has reason to give a pragmatic interpretation. Although some of the storyboard contexts studied in this thesis contain very good examples of such scenarios, there may be other examples in the literature. For example in the Guasti et al. (2005) study participants see a scenario where the mode of transport is discussed in relation to cost. Therefore the expectation could be set up in participants that they should be as informative as possible because how many soldiers chose which particular mode of transport is important. It is interesting that 75% of the 7-year-old children in this experiment detect the scalar implicature compared to 13% of 7-year-olds in a statement evaluation task in another experiment in this study.

The claim of Guasti et al. (2005) that shared background knowledge facilitates sensitivity to scalar implicatures appears to be supported by higher pragmatic response rates in the younger children in the storyboard/pictureboard task (Experiments 1, 2 & 3), when compared to their rates of pragmatic responding in the statement evaluation task. However, this feature alone is not enough for sensitivity to be seen in some children (Experiment 6), supporting Hurewitz et al.’s (2006) claim that children do not always make use of the visual scene.

Hurewitz et al.’s (2006) claim for the benefits of strong discourse support, as opposed to visual scene or situational support, also receives some support from the findings in this thesis. This feature is present in the storyboards where Charlotte is in conversation with her mother (Experiments 1 & 2) and where the researcher is in conversation with the participants (Experiment 4). In both experiments clear informativeness expectations are set up. However it could be argued that the agent conditions of the interactive and descriptive storyboards (Experiment 6) contain the same discourse support, and yet there is a significant difference in the rates of pragmatic responding between the two conditions. Therefore it may be the case that whilst discourse support helps sensitivity, the dialogue needs to contain something in addition to the flouting of conversational principles.

One thing that is apparent from the findings of the thesis is that no single contextual feature that has been proposed in the literature is guaranteed to result in sensitivity to scalar implicatures. It seems likely that contexts that are based on naturalistic conversations, where there are clear informativeness expectations, with readily available background knowledge, and where the interpretation of the scalar term carries consequences, will result in more pragmatic interpretations than contrived pragmatically impoverished contexts, where the background is not established and there is nothing to motivate a particular interpretation. However, it remains to be seen whether a context can ever be supportive enough to produce sensitivity in children younger than 3-years-of-age.

5.3.3. Which contemporary theory, default or relevance, best captures the processing of the scalar term some?

Experiments 7 and 8 tested predictions made by two current pragmatic theories of scalar implicature; the default account of Levinson (1983, 2000), and Sperber and Wilson's (1985/1995) Relevance Theory. The findings from these experiments, as well as others in the thesis, lead to the conclusion that neither theory in their present form can adequately account for the empirical data.

The theories of Levinson (1983, 2000) and Sperber and Wilson (1985/1995) both give an account of adult processing of scalar terms therefore the evidence from adults and children will be considered separately.

Levinson's (1983, 2000) proposal that the default interpretation of *some* is *not all* and that this interpretation is triggered automatically whenever the scalar term appears regardless of context is not borne out by the reaction times of the adult mixed responders in Experiment 8. For this group of participants a pragmatic response took longer than a logical response. There is however support from the same experiment for his proposal that the cancellation of the implicature requires time and effort. Logical responses to *infelicitous some* took longer than the same response to *felicitous some* in the logical responders, suggesting that for some adults a logical response can be a late arriving process.

The findings from the children also present mixed support for Levinson's 1983, 2000) default account. It appears from the findings of the younger children that the initial interpretation of *some* is logical, and children will give this interpretation for a number of years in pragmatically impoverished contexts. This finding alone is not problematic for Levinson's theory as it could be argued that the inference has not yet become automatic, however the finding of a dissociation in context is. For Levinson (2000, p. 22), the inference is triggered by the structure of the utterance rather than the context of the utterance. It is difficult to explain why children can detect the implicature in the storyboard/bag task and not the statement evaluation task (Experiments 1, 2, 5 & 6) without reference to the context of the utterance. In addition the developmental trajectory would suggest that the children are using context to make pragmatic inferences rather than to cancel them.

The practice effects seen in the children could be accounted for by Levinson (1983, 2000). The repeated presentation and structured nature of the stimulus materials may have made the children more aware of the structure of the language and expectations of how the language is used, allowing them to make more pragmatic inferences. However the lack of practice effects in the descriptive set of storyboards (Experiment 6) weakens this proposal as the structure is essentially the same as in other experiments.

The findings from some of the experiments in this thesis (Experiments 2, 5 & 7) suggested that there may be a point in development where *some* is the automatic interpretation. Universal pragmatic responding and pragmatic responding seemingly unaffected by cognitive load, suggested that the detection of scalar implicatures in these children did not appear to be an effortful process. However the reaction times of children (Experiment 8) revealed that a pragmatic interpretation took longer to arrive at than a logical interpretation. Horn (2009) points out though, that not all default accounts posit effortless processing for a pragmatic interpretation. He draws attention to the fact that Grice's view of conversational inferences always involved additional processing to go from what is said to what is meant. Horn, using the example of his regular mode of transport, identifies that his default status, Monday to Friday, is to drive to work. This does not mean though that the procedure is automatic, he still has to carry out the actions of driving; thus, "an implicature may arise in a default context without thereby constituting a default or automatic inference" (p. 23).

Relevance Theory's account of adults' processing of scalar terms receives mixed support. The reaction times (Experiment 8) support the predictions of Noveck and Sperber (2007) that enriching the literal meaning of *some* takes time. However, this thesis's proposal that adults can experience a conflict in choosing an interpretation goes against RT's proposal that there can never be more than one interpretation available otherwise the principle of optimal relevance would be violated.

Anecdotal findings from the adults also suggest problems for RT's proposed three levels of interpretation strategy. According to Sperber (1994), a sophisticated interpretation, where the hearer assumes neither competence nor benevolence on the part of the speaker, involves the hearer stopping not at the first relevant interpretation that comes to mind, but at the first interpretation they think the speaker would have thought relevant. These complex meta-representations do not have to be conscious or involve difficulty. Yet some adults in Experiment 2 consciously deliberated over interpretations, systematically thinking through what interpretation they should give. In this case it would appear that Noveck and Sperber's criticism of

Levinson (1983, 2000) that the benefit of deriving the inference by default would be offset by the frequency of cancellation could be used against RT. That is, if interpretations are often revised by adults what is the benefit of optimal relevance? The principle of optimal relevance states that every ostensive stimulus (a stimulus worth the hearer's attention) carries the presumption that its processing is not only relevant enough to be worth the effort but that it is more relevant than any other available stimulus. The point of optimal relevance is that effort is not required to choose between stimuli. But if hearers can, and do, revise their interpretations then the benefits of optimal relevance appear to be offset by the effortful revisions.

In terms of children's processing then RT can handle most of the findings. It can explain the dissociation in context and practice effects in terms of the pay off between effect and effort. It can also explain the lack of practice effects in the descriptive set of storyboards (Experiment 6) with reference to the same principal; less effort is required to process the second presentation of the materials but the cognitive effect to be gained does not warrant the extra processing.

What neither theory seems to capture is the complex dynamic nature of scalar implicature processing: Scalar implicatures can be detected by 3-year-olds and missed by 10-year-olds, and they appear to be easier to process for 10-year-olds than for some adults. Some age groups will be homogenous in their interpretations of scalar terms on certain tasks, other age groups will be homogenous regardless of task, whilst for other age groups interpretation seems to be a matter of individual preference. It seems that the interpretation of scalar terms can be unconscious or require deliberation. Inference pervades language understanding and is the "core of the understanding process" (Shank, 1976, p.168), and because of this, it must for the most part be quick and relatively effortless so that successful communication can take place. Scalar inferences on the other hand appear to go beyond this. Moshman (1998, pp. 952-953) identifies three qualitatively different cognitive processes; inference (typically automatic and unconscious), thinking (the deliberate co-ordination of inferences to serve a purpose) and reasoning (constraining one's thinking on the basis of epistemic constraints). These would seem to apply to the processing of scalar implicatures. What we need is a developmental theory that can account for the dynamic changes not only at the semantic/pragmatics interface but at the inference/reasoning border.

5.4. Implications

The findings of this thesis have major implications for the current pragmatic theories that are used to explain the processing of scalar implicatures, but there are also more general implications for the area of experimental pragmatics.

The ability of Levinson's (1983/2000) default account and Sperber and Wilson's (1985/1995) Relevance Theory to adequately account for the empirical data was discussed in detail in section 5.3.3. and therefore the main arguments will not be reiterated here. However the main conclusion of this thesis in relation to the theories is that the empirical data presents problems for both camps. One of the main problems for both theories is the failure to give an adequate account of the interpretational strategies that appear to be available to adults. Each theory also neglects to give any account of the development of processing. Findings from children and adults cannot be adequately accounted for by recourse to suppositions about how they might behave. The complex developmental picture of sensitivity that is evident in this thesis suggests that both theories need serious revision to cope with the dynamic changes that occur.

The general implications for experimental pragmatics relate mainly to the interpretation of findings based on untested assumptions, without empirically testing those assumptions to be confident that they hold. Within any new field there is a need to establish an empirical data-base in order to further our knowledge of the topic under investigation. Whilst there have been undeniable methodological advances made in the study of scalar implicature it is unclear that assumptions made in interpreting recent experimental results are justified. For example, De Neys and Shaeken (2007) interpreted their findings based on the assumption of a two-step model, i.e. that a logical response only represents failure to detect the implicature rather than detection and subsequent cancellation of the implicature. The results described in Chapter 4 suggest that this assumption may not be justified. Thus, before further progress can be made, researchers may wish to question the assumptions against which they interpreted their experimental findings, and consider alternative explanations of their data.

The developmental findings also have methodological implications for the study of scalar implicature. The experiments indicate that many factors affect sensitivity to scalar terms and also show how susceptible interpretations can be to change when sensitivity is emerging. One point to bear in mind for future research is that there may be periods in development that are more unstable than others and therefore experimental manipulations may have a greater effect on one age group than another. It is also difficult to make comparisons between studies at the present time because of the manipulation of a number of experimental variables or the failure to control for different variables. This thesis has shown that it does not seem safe to assume that findings will generalise across languages or even similar contexts, and given that individual scalar terms can result in unique interpretation problems, i.e. the interpretation of disjunction as conjunction, there does not appear to be homogeneity in the derivation of scalar terms. In order to establish reliability and validity, systematic investigation of individual variables needs to be carried out to ascertain what contribution each variable makes to the detection of scalar implicatures.

5.5. Limitations

Problems within individual experiments were raised and discussed in the relevant chapters, therefore this section focuses on more general methodological issues.

One limitation of the research carried out in this thesis is the sole use of the experimental method. Given that this thesis investigates an aspect of language, and language is primarily used for communication between people, then the controlled artificial nature of experimentation may seem an inappropriate method. However, as Noveck and Sperber (2004) point out, pragmatics has a long history of philosophical and linguistic theories developed from individual intuitions of how an utterance would be uttered in hypothetical situations, and of comparing these intuitions against corpora, but what linguistics has lacked is the rigour of scientific methods to test these intuitions. Therefore the experimental method has much to offer the study of scalar implicatures as long as an interpretation of the experimental findings is tempered with an appreciation that this might not be how people would respond outside the laboratory.

The reliance on behavioural responses to ascertain participants' interpretations is also a limitation. This has been the method of choice for existing experimental investigations into scalar implicatures. However this thesis has shown that, particularly in relation to logical responses, the same response from different participants may not mean the same thing or reflect the same processing. There has been a move to examine on-line processing with the use of eye-tracking, evoked potentials, and reaction time studies, however the field may benefit from a multi-method approach. Studies which combine behavioural responses, with the on-line methods described above, and are complimented by data from verbal protocols or people's reflections on why they chose a particular response would help us understand how scalar implicatures are processed and how particular methods affect participants interpretations (for example the use of the storyboard methodology with adults in Experiment 2 and the button box with children in Experiment 8).

Another limitation is the use of a dichotomous forced choice response. In all experiments participants were required to answer true or false. Given what this thesis is proposing in relation to adults' processing, it is difficult to know whether a response reflects a definite answer, or a participant's preferred interpretation for a particular context. In other words does the response reflect unconscious inferencing or deliberative reasoning? It could be said if a participant is using a logical reasoning strategy then there will be only one interpretation for *infelicitous some*, a logical interpretation. However, reasoning does not always take place according to logical norms, and reasoning according to pragmatic principles could be seen as attempt to come up with the best interpretation for a particular situation. The use of open ended responding may help clarify this issue and also help discover whether there are points in development where children experience any conflict in interpretation.

The use of cross-sectional studies also provides a limited picture of changes in development. Whilst this method can be used to highlight age related differences and developmental trends, it does not provide us with a picture of how change unfolds. One way to examine the dynamics of change would be to employ the microgentic method (for an introduction to this method see Flynn, Pine & Lewis, 2007) with the appropriate analysis (Cheshire, Muldoon, Francis, Lewis & Ball, 2007). The microgenetic method has examined developmental change in a number of different domains, such as theory of mind and inhibitory control (Flynn, 2006, 2007), and language development (van Dijk & van Geert, 2007). Its ability to investigate spontaneous change through the repeated presentation of tasks over the period of change, and to induce change by providing different forms of intervention, could help to uncover the dynamics of the developmental change we see in the U-shaped trajectory.

5.6. Future Directions

Although follow up experiments and ways to increase the reliability and validity of the results are mentioned in the discussion sections of the individual experiments as well as in other sections of this General Discussion, this section identifies three lines of enquiry where future research could make valuable contributions to our understanding of *some* and where findings from *some* could aid understanding in other fields.

5.6.1. Investigation of the U-shaped trajectory

The first line of enquiry is to follow up the U-shaped developmental trajectory. U-shaped trajectories are interesting not only because they are unusual, but also because they stimulate process level explanations (Siegler, 2004). To this end two further experiments are suggested; a replication of Experiment 2, to establish that the U-shaped finding is reliable, and a follow up of the result of Experiment 5. Experiment 5 employed the storyboard methodology, where the experimental manipulation was the presence/absence of an agent, and universal pragmatic responding was seen in the Year 4 children. It would be interesting to extend the age range upwards in this experiment for two reasons; firstly to see if the developmental trajectory of sensitivity to the scalar implicature is U-shaped, and secondly, if the trajectory is U-shaped, to see when logical responding re-emerges. If the re-emergence of logical responding seen in Experiment 2 is linked to the development of a formal reasoning system then it would be expected that logical responding would re-emerge around early adolescence (Piaget, 2001; Piaget & Inhelder, 2000). However if it is not linked to the development of reasoning processes but is related to the development of pragmatic ability or meta-pragmatic knowledge then it may be observed much earlier in childhood (Anderson-Wood & Smith, 1997; Bernicot, Laval & Chaminaud, 2007). Further investigation of the trajectory could help determine whether the U-shape distribution represents a change from language processing to reasoning.

5.7.2. Investigation of the relationship between scalar inferences and other pragmatic inferences

The last point of the previous paragraph brings us to the second line of enquiry, namely investigating the similarities and differences between scalar inferences and other pragmatic inferences. Most of the research into pragmatic skills has tended to investigate them in isolation (Adams, 2002), though there have been attempts to investigate a number of phenomena together (Bucciarelli, Colle & Bara, 2003; Loukusa, Leinonen & Ryder, 2007; Ryder & Leinonen, 2003) and to see how pragmatic skills relate to pragmatic knowledge (Bernicot et al., 2007). In addition there is no single theory or protocol that covers the development of pragmatic capacity and identifies ages at which children are expected to produce and comprehend different speech acts (Adams, 2002; Bara, Bosco & Buccarelli, 1999).

Separate pieces of research on individual pragmatic skills which are component characteristics of scalar implicatures, indicate that they occur at different ages. For example, the ability to infer information from stories develops between 4 and 6 years of age (Paris & Upton, 1976), as does the transition from literal to non-literal interpretations of speech (Eson & Shapiro, 1982). Children can provide enough information for an interlocutor to decide what the point of reference is by the age of seven, but the ability to detect ambiguity is not complete for another two years (Lloyd, Camaioni & Ercolani, 1995). Furthermore, meta-pragmatic knowledge can be present from the age of 6 (Andersen-Wood & Smith, 1997), but not fully developed until adolescence (Gombert, 1992) (for a summary of results on the development of pragmatic abilities see Adams, 2002). In addition, as is the case for scalar implicatures, pragmatic skills and meta-pragmatic knowledge are affected by context (Laval, 2003; Ninio & Snow, 1996; Tomasello, 2000). This line of enquiry would help to establish whether pragmatic responses to scalar implicatures develop in-line with expectations from other pragmatic research and also whether a U-shaped development trajectory is seen with other inferences.

5.6.3. Investigating sensitivity to scalar implicatures in people with Autism Spectrum Disorders

The third line of enquiry is the investigation of atypical populations to understand scalar implicatures. The use of atypical populations to understand development is a standard psychological paradigm and has been used to investigate other pragmatic abilities (Bara et al., 1999; Bishop, 2000; Tager-Flusberg, 1999). People whose disorders fall within the autism spectrum would be ideal candidates as pragmatics is seen as a core deficit of this disorder (Tager-Flusber, 1981). It might be expected that people with autism would be insensitive to scalar implicatures given their tendency to interpret utterances literally (Attwood, 1998; Bogdashina, 2005; Happé, 1993), the difficulties they have with non-literal speech such as irony, jokes and metaphors (Baron-Cohen, 1997; Dennis, Lazenby & Lockyear, 2001; Happé, 1994), and their disregard for Grice's maxims (Surian, Baron-Cohen & Van der Lely, 1996). However, in a review of pragmatic abilities in people with Asperger's syndrome or high functioning autism, Loukusa and Moilanen (in press) concluded that whilst the pragmatic abilities of such people may be less developed than those of the typically developing population, as they were still able to answer many different kinds of pragmatic questions their pragmatic abilities should be seen as inefficient rather than deficient (Loukusa & Moilanen, in press; Loukusa et al., 2007).

To date there appears to be only one published study that has investigated sensitivity to scalar implicatures in an atypical population[13]. Pijnacker, Hagoort, Buitelaar, Teunisse & Geurts (2009) looked at sensitivity to scalar implicatures in adults with autism spectrum disorders (ASD), using a statement evaluation task. The authors found that there was no difference in logical responses rates to the *infelicitous some* sentences between the adults with ASD and matched controls. The adults with ASD had a logical response rate of 24% whilst the matched controls' response rate was 29%, indicating high levels of sensitivity in both groups.

There appear to be no published studies investigating sensitivity to scalar implicatures in children with autism. However, data ere collected for an investigation for this thesis, but for reasons beyond the control of the author could not be

[13] Noveck, Guelminger, Georgieff and Labruyere (2007) investigated interpretations given to ambiguous *every…not* sentences, such as *every* horse did *not* jump over the fence, in adults with autism. However the authors themselves point out that whilst there may be a strong overlap with the work on scalar implicatures, there are important differences regarding the ambiguity and strength of the terms.

completed. Thirteen children, age range 9-16-years-of-age, with autism as part of their diagnosis were presented with statements and storyboards similar to those used in the rest of the thesis, and measures of IQ (Wechsler Abbreviated Scale of Intelligence [Psychological Corporation, 1999]) and language (British Picture Vocabulary Scale-II [Dunn, Dunn, Whetton & Burley, 1997]) were also taken (see Appendix CI for the methodology and results of this study). The children had low logical response rates to the *infelicitous some* trials in both tasks, 36.5% in the statement task and 9.5% in the storyboard task. The individual measures taken suggested that language ability was marginally associated with logical response rates to *infelicitous some*, whilst IQ was negatively associated with it. These findings, though involving a small number of participants, indicate that children with autism can detect scalar implicatures, despite associated research suggesting that they may have difficulties.

Studying sensitivity to scalar implicatures in atypical populations would not only be beneficial for our understanding of the nature of scalar implicatures, but there has also been a call for studies to enhance our understanding of developmental traits in pragmatic inferencing (Loukusa & Moilanen, in press). In addition, the statement and storyboard methodologies are suitable for studying the effect of context on interpretations of words, an area that is much debated within the autism literature (Frith & Snowling, 1983; Jolliffe & Baron-Cohen, 1999; López & Leekam, 2003). Furthermore there has been a call for more research on individual types of inference in order to help clinical assessments of pragmatic abilities determine the source of inference failure in particular individuals (Adams, 2002).

5.7. Concluding Remarks

The eight experiments contained in this thesis present a number of novel findings which contribute to our understanding of the scalar implicature associated with the quantifier *some*. Using existing methodologies, as well as developing new ones, the experiments revealed that some of the previous assumptions in the literature regarding the phenomena appear to be incorrect; sensitivity to scalar implicatures is not linear, children can give more pragmatic interpretations than adults on the same task, in some cases scalar implicatures can be more effortful for adults than for children, and a logical response is not necessarily indicative of failure to detect a scalar implicature. The findings also indicated that there are many factors that contribute to scalar implicature detection whilst sensitivity is developing; age, context, method of eliciting response, shared conversational background, order of presentation and practice. Interpretations of the findings have implications for current pragmatic theories that seek to explain the processing of scalar implicatures, in that they suggest that these theories need to be revised in order to accommodate the empirical data and raise the possibility that reasoning processes are also involved in the interpretation process. The way in which the field of experimental pragmatics carries out investigations of scalar implicature may also need to be revised to provide reliable valid findings that are able to be replicated. Overall this thesis indicates that the phenomenon of scalar implicature is more complex and multifaceted than had previously been assumed.

References

Adams, C. (2002). Practitioner Review: The assessment of language pragmatics. *Journal of Child Psychology and Psychiatry*, 43, 8, 973-987.

Adler, J. (1997). Lying deceiving or falsely implicating. *Journal of Philosophy*, 94, 9, 435-452.

Andersen-Wood, L. & Smith, B.R. (1997). *Working with Pragmatics*. London: Winslow Press.

Attwood, T. (1998). *Asperger's syndrome: A guide for parents and professionals*. London: Jessica Kingsley Publishers.

Bach, K. (2006). The top 10 misconceptions about implicature. In B.J. Birner & G. Ward (Eds.). *Drawing the boundaries of meaning: Neo-Gricean studies in pragmatics and semantics in honour of Laurence R. Horn.* Amsterdam: Benjamins.

Badzinsk, D.M. Cantor, J. & Hoffner, C. (1989). Children's Understanding of Quantifiers. *Child Study Journal*, 19, 4, 241-258.

Bara, B.G., Bucciarelli, M. & Johnson-Laird, P.N. (1995). Development of Syllogistic Reasoning. *The American Journal of Psychology*, 108, 2, 157-193.

Baron-Cohen, S. (1997). Hey! It was just a joke! Understanding propositions and propositional attitudes by normally developing children, and children with autism. *Israel Journal of Psychiatry,* 34, 174-178.

Barruoillet, P. & Lecas, J.F. (1999). Mental models in conditional reasoning and working memory. *Thinking and Reasoning*, 5, 4, 289-302.

Beck. S.R., Robinson, E.J., Carroll, D.J, & Apperly, I.A. (2006). Children's thinking about counterfactuals and future hypotheticals as possibilities. *Child Development*, 77, 2, 413 – 426.

Begg, I. & Harris, G. (1982). On the Interpretation of Syllogisms. *Journal of Verbal Learning and Verbal Behavior*, 21, 595-620.

Begg, I. (1987). Some. *Canadian Journal of Psychology*, 41, 62-67.

Bernicot, J., Laval, V. & Chaminaud, S. (2007). Non-literal language forms in children: In what order are they acquired in pragmatics and meta-pragmatics? *Journal of Pragmatics*, 39, 12, 2115-2132.

Bever, T.G. (1970). The cognitive basis for linguistic structures. In J.R. Hayes (Ed.), *Cognition and the Development of Language*. New York: Wiley.

Bishop, D.V.M. (2000). Pragmatic language impairment: A correlate of SLI, a distinct subgroup, or part of the autistic continuum? In D.V.M. Bishop & L.B. Leonard (Eds.), *Speech and Language Impairments in Children: Causes, characteristics, intervention and outcome* (pp. 99-114). Hove: Psychology Press.

Bogdashina, O. (2005). *Communication Issues in Autism and Asperger Syndrome - Do we speak the same language?* London: Jessica Kingsley Publishers.

Bonnefon, J.F., Feeney, A. & Villejoubert, G. (2009). When some is actually all: Scalar inferences in face-threatening contexts. *Cognition*, 112, 2, 249-258.

Bott, L. & Noveck, I.A. (2004). Some utterances are underinformative: The onset and time course of scalar inferences. *Journal of Memory and Language*, 51, 3, 437-457.

Bowerman, M. (1982). Starting to talk worse: Clues to language acquisition from children's late speech errors. In S. Strauss & R. Stavy (Eds.), *U-shaped Behavioral Growth* (pp.101-146). New York: Academic Press.

Braine, M. & Rumain, B. (1981). Development of Comprehension of “or”: Evidence for a sequence of competencies. *Journal of Experimental Child Psychology*, 31, 1, 46-70.

Breheny, R., Katsos, N. & Williams, J. (2006). Are generalised scalar implicatures generated by default? An on-line investigation into the role of context in generating pragmatic inferences. *Cognition*, 100, 3, 434-463.

Bucciarelli, M., Colle, L. and Bara, B.G. (2003). How children comprehend speech acts and communicative gestures. *Journal of Pragmatics*, 35, 2, 207-241.

Buckley, B. (2003). *Children’s Communication Skills: From Birth to Five Years.* London: Routledge.

Bussey, K. (1992). Lying and Truthfulness: Children's Definitions, Standards, and Evaluative Reactions. *Child Development*, 63, 1, 129-137.

Carston, R. (2002). *Thoughts and Utterances: The Pragmatics of Explicit Communication*. Oxford: Blackwell

Case, R., Kurland, D. M., & Goldberg, J. (1982). Operational Efficiency and the Growth of Short-Term Memory Span. *Journal of Experimental Child Psychology*, 33, 386-404.

Ceraso, J. & Provitera, A. (1971). Sources of Error in Syllogistic Reasoning. *Cognitive Psychology*, 2, 400-410.

Cheshire, A., Muldoon, K.P., Francis, B., Lewis, C.N. & Ball, L.J. (2007). Modelling change: New opportunities in the analysis of microgenetic data. *Infant and Child Development*, 16, 1, 119-134.

Chierchia, G. (2004). Scalar Implicatures, Polarity Phenomena and the Syntax/Pragmatics Interface. In A. Belletti (Ed.), *Structures and Beyond*. Oxford: Oxford University Press.

Chierchia, G., Crain, S., Guasti, M.T. & Thornton, R. (1998). Some and Or: A study on the emergence of logical form. In *Proceedings of the 22th Boston University Conference on Language Development*, 97-108. Sommerville, MA: Cascadilla Press.

Chierchia, G., Crain, S., Guasti, M.T., Gualmini, A. & Meroni, L. (2001). The acquisition of disjunction: Evidence for a grammatical view of scalar implicatures. In *Proceedings of the 25th Boston University Conference on Language Development*, 157-168. Sommerville, MA: Cascadilla Press.

Chierchia, G., Guasti, M.T., Gualmini, A., Meroni, L., Crain, S. & Foppolo, F. (2004). Semantic and Pragmatic competence in Children's and Adults' Comprehension of *Or*. In I. A. Noveck and D. Sperber (Eds.), *Experimental Pragmatics*. Basingstoke: Palgrave Macmillian.

Chisholm, R.M. & Feehan, T.D. (1977). The intent to deceive. *The Journal of Philosophy*, 74, 3, 143-159.

Conway, A. R. A., Kane, M. J., Bunting, M. F., Hambrick, D. Z., Wilhelm, O. & Engle, R. W. (2005). Working memory span tasks: A methodological review and user's guide. *Psychonomic Bulletin and Review*, 12, 769-786.

Copi, I.M. & Cohen, C. (1994). *Introduction to logic*. New Jersey: Macmillan.

Cosmides, L. & Tooby, J. (2005). Neurocognitive Adaptations Designed for Social Exchange. In D.M. Buss (Ed.). *Evolutionary Psychology Handbook*. New York: Wiley.

Craine, S. & Thornton, R. (1998). *Investigations in Universal Grammar: A guide to experiments on the acquisition of syntax and semantics*. Cambridge, MA: MIT Press.

Cummins, D.D. (1995). Naive theories and causal deduction. *Memory and Cognition*, 23, 5, 646-658.

De Neys, W. & Glumicic, T. (2008). Conflict monitoring in dual process theories of reasoning. *Cognition*, 106, 3, 1248-1299.

De Neys, W. & Schaeken, W. (2007) When People are More Logical under Cognitive Load: Dual Task Impact on Scalar Implicature. *Experimental Psychology,* 54, 2, 128-133.

Della Sala, S., Gray, C., Baddeley, A. & Wilson, L. (1997). Visual Patterns Test. Suffolk: Thames Valley Test Company.

Dennis, M., Lazenby, A. L. & Lockyear, L. (2001). Inferential language in high-function children with Autism. *Journal of Autism and Developmental Disorders*, 31, 1, 47-54.

Dias, M.G. & Harris, P.L. (1988). The effect of make-believe play on deductive reasoning. *British Journal of Developmental Psychology*, 6, 3, 207-221.

Donaldson, M. (1978). *Children's Minds*. London: Fontana Press.

Dunn, L. M., Dunn, L.M., Whetton, C. & Burley, J. (1997). *The British Picture Vocabulary Scale (2nd Ed.)*. London: nfer-Nelson.

Ekman, P. (1989). *Why kids lie: how parents can encourage truthfulness*. New York: Penguin.

Ervin, S.M. & Miller, W.R. (1963). Language development. In H.W. Stevenson (Ed.) *Child Psychology: The sixty second yearbook of the National Society for the study of Education (Pt. 1).* Chicago: Chicago Press.

Esson, M.E. & Shapiro, A.S. (1982). When 'Don't' means 'Do'; Pragmatic and cognitive development in understanding an indirect imperative. *First Language*, 3, 8, 83-91.

Evans, J.St. B.T. & Newstead, S.E. (1980). A study of disjunctive reasoning. *Psychological Research*, 41, 373-388.

Evans, J.St.B.T. (2003). In two minds: dual-process accounts of reasoning. *TRENDS in Cognitive Sciences,* 7, 10, 454-459.

Evans, J.St.B.T. & Over, D.E. (1996). *Rationality and Reasoning*. Hove: Psychology Press.

Evans, J.St.B.T., Newstead, S.E. and Byrne, R.M.J. (1993). *Human Reasoning: The psychology of deduction*. Hove: Earlbaum.

Feeney, A., Scrafton, S., Duckworth, A. & Handley. (2004). The Story of Some: Everyday pragmatic inferences by children and adults. *Canadian Journal of Experimental Psychology*, 58, 2, 121-133.

Flynn, E. (2006). A microgenetic investigation of stability and continuity in theory of mind development. *British Journal of Developmental Psychology,* 24, 3, 631-654.

Flynn, E. (2007). The role of inhibitory control in false belief understanding. *Infant and Child Development*, 16, 1, 53-69.

Flynn, E., Pine, K. & Lewis, C. (2007).Using the microgenetic method to investigate cognitive development: An introduction. *Infant and Child Development*, 16, 1, 1-6.

Franssens, S. & De Neys, W. (2009). The effortless nature of conflict detection during thinking. *Thinking and Reasoning*, 12, 2, 105-128.

Frith, U. & Snowling, M. (1983). Reading for meaning and reading for sound in autistic and dyslexic children. *Journal of Developmental Psychology*, 1, 329-342.

Gardener, H. & Winner, E. (1982). First imitations of artistry. In S. Strauss & R. Stavy (Eds.), *U-shaped Behavioral Growth* (pp.147-168). New York: Academic Press.

Garrett, M. & Harnish, R.M. (2009). Q-Phenomena, I-Phenomena and Implicature: Some experimental pragmatics. *International Review of Pragmatics*, 1, 84-117.

Gathercole, S.E., Pickering, S.J., Ambridge, B. & Wearing, H. (2004). The Structure of Working Memory from 4 to 15 Years of Age. *Developmental Psychology*, 40, 2, 177–190.

Gazdar, G. (1979). *Pragmatics: Implicature presupposition and logical form*. New York: Academic Press.

Geurts, B. (2009). Scalar Implicatures and Local Pragmatics. *Mind and Language*, 24, 1, 51-79.

Girotto, V., Gilly, M., Blaye, A. & Light, P. (1989). Children's Performance in the Selection Task: Plausibility and familiarity. *British Journal of Psychology*, 80, 79-95.

Gombert, J.E. (1992). *Metalinguistic Development*. Chicago: The University of Chicago Press.

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

Gualmini, A. (2004). Some Knowledge Children Don't Lack. *Linguistics*, 42, 5, 957-982.

Gualmini, A., Crain, S. & Meroni, L. (2000). The inclusion of disjunction in conditional sentences. In *Proceedings of the 24th Boston University Conference on Language Development*, 367-378. Sommerville, MA: Cascadilla Press.

Gualmini, A., Crain, S., Meroni, L., Chierchia, G. & Guasti, M.T. (2001). At the Semantics/Pragmatics Interface in Child Language. In *Proceedings of Semantic and Linguistic Theory XI*, 231-247. Ithaca, NY: CLC Publications.

Gualmini, A., Meroni, L. & Crain, S. (2001). The inclusion of disjunction in child grammar: Evidence from modal verbs. Proceedings of *NELS 30*: 247-257. Amherst, MA: GLSA, University of Massachusetts.

Guasti, M.T., Chierchia, G., Crain, S., Foppolo, F., Gualmini, A. & Meroni, L. (2005). Why Children and Adults Sometimes (but not always) Compute Implicatures. *Language and Cognitive Processes,* 20, 5, 667-696.

Handley, S., Capon, A., Beveridge, M., Dennis, I. & Evans, J.St.B.T. (2004). Working memory, inhibitory control and the development of children's reasoning. *Thinking & Reasoning*, 10, 2, 175-195.

Happé, F.G.E. (1993). Communicative competence and theory of mind in autism: A test of relevance theory. *Cognition*, 48, 2, 101-119.

Happé, F.G.E. (1994). An advanced test of theory of mind: Understanding of story characters' thoughts and feelings by able autistic, mentally handicapped, and normal children and adults *Journal of Autism and Developmental Disorders,* 24, 2, 129-154.

Harris, M. (1976). The influence of reversibility and truncation on the interpretation of the passive voice by young children. *British Journal of Psychology*, 67, 3, 419-428.

Harris, M. & Coltheart, M. (1986). *Language processing in children and adults: An introduction.* London: Routledge.

Harris, P.L. & Nunez, M. (1996). Understanding of Permission Rules by Preschool Children. *Child Development*, 67, 4, 1572-1591.

Hawkins, J., Pea, R.D., Glick, J. & Scribner, S. (1984). “Merds that laugh don’t like mushrooms”: Evidence for deductive reasoning by preschoolers. *Developmental Psychology*, 20, 4, 584-594.

Hirschberg, J. (1991). *A Theory of Scalar implicature*. New York: Garland Publishing.

Holmes, J., Adams, J.W. & Hamilton, C.J. (2008). The relationship between visuospatial sketchpad capacity and children's mathematical skills. *European Journal of Cognitive Psychology*, 20, 2, 272-289.

Hopper, R. & Bell, R.A. (1984). Broadening the deception construct. *Quarterly Journal of Speech*, 70, 3, 288-302.

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

Horn, L. (1989). *A Natural History of Negation*. Chicago: University of Chicago Press.

Horn, L.R. (2004). Implicature. In L.R. Horn & G. Ward (Eds.), *Handbook of Pragmatics.* Oxford: Blackwell.

Horn, L.R. (2009). WJ-40: Implicature, truth, and meaning. *International Review of Pragmatics*, 1, 3-34.

Huang, Y.T. & Snedeker, J. (2005, April). *What exactly do numbers mean?* Paper presented at the conference on Experimental Pragmatics: Exploring the Cognitive Basis of Conversation, University of Cambridge, United Kingdom.

Hurewitz, F., Brown-Schmidt, S., Thorpe, K., Gleitman, L.R. & Trueswell, J.C. (2000). One frog, two frog, red frog, blue frog: Factors affecting children's syntactic choices in production and comprehension. *Journal of Psycholinguistic Research*, 29, 6, 597-626.

Hurewitz, F., Papafragou, A., Gleitman, L. & Gelman, R. (2006). Asymmetries in the Acquisition of Numbers and Quantifiers. *Language, Learning and Development*, 2, 2, 77-96.

Hyman, R. (1989). The Psychology of Deception. *Annual Review of Psychology*, 40, 135-154.

Jacobs, J.E., & Klaczynski, P.A. (2002). The development of decision making during childhood and adolescence. *Current Directions in Psychological Science*, 4, 145-149.

Jennings, R.E. (1994). *The Genealogy of Disjunction.* New York: Oxford University Press.

Johnson-Laird, P.N. (1975). Models of Deduction. In R.J. Falmagne (Ed.), *Reasoning: Representation and Process.* New York: Wiley.

Jolliffe, T., & Baron-Cohen, S. (1999). A test of central coherence theory; linguistic processing in high-functioning adults with autism or Asperger's syndrome: Is local coherence impaired? *Cognition*, 71, 2, 149-185.

Katsos, N., Breheny, R. & Williams, J. (2005). *Proceedings of the Annual Conference of the Cognitive Science Society*, 27, 1108-1113.

Klaczynski, P.A. (2004). A dual-process model of adolescent development: Implications for decision making, reasoning, and identity. In R.V. Kail (Ed.), *Advances in Child Development and Behavior, vol. 31* (pp. 73-123). San Diego, CA: Academic Press.

Klaczynski, P.A. (2001). Framing effects on adolescent task representations, analytic and heuristic processing, and decision making: Implications for the normative-descriptive gap. *Journal of Applied Developmental Psychology,* 22, 289-309.

Klaczynski, P.A., & Cottrell, J.E. (2004). A dual-process approach to cognitive development: The case of children's understanding of sunk cost decisions. *Thinking & Reasoning*, 10, 2, 147-174.

Laval, V. (2003). Idiom comprehension and metapragmatic knowledge in French children. *Journal of Pragmatics*, 35, 5, 723-739.

Leinonen, E., Letts, C.A. & Smith, B.R. (2000). *Children's Pragmatic Communication Difficulties.* London: Whurr.

Levinson, S.C. (1983). *Pragmatics*. Cambridge: Cambridge University Press.

Levinson, S.C. (2000). *Presumptive Meanings: The Theory of Generalized Conversational Implicature*. Cambridge: MIT Press.

Lewis, M. (1993). The development of deception. In M. Lewis & C. Saarni (Eds.) *Lying and Deception in Everyday Life.* New York: Guilford Press.

Lewis, M., Stanger, C. & Sullivan, M.W. (1989). Deception in 3 year olds. *Developmental Psychology,* 25, 3, 439-443.

Lloyd, P., Camaioni, L. & Ercolani, M. (1995). Assessing referential communication skills in the primary school years: A comparative study. *British Journal of Developmental Psychology,* 13, 1, 13-29.

López, B. & Leekam, S. (2003). Do children with autism fail to process information in context? *Journal of Child Psychology and Psychiatry,* 44, 2, 285-300.

Loukusa, S., Leinonen, E., Kuusikko, S., Jussila, K., Mattila, M., Ryder, N., et al. (2007). Use of context in pragmatic language comprehension by children with Asperger syndrome or high-functioning Autism. *Journal of Autism and Developmental Disorders,* 37, 6, 1049-1059.

Loukusa, S., Leinonen, E. & Ryder, N. (2007). Development of pragmatic language comprehension in Finnish-speaking children. *First Language*, 27, 3, 82, 279-296.

Loukusa, S. & Moilanen, I. (In Press). Pragmatic inference abilities in individuals with Asperger syndrome or high-functioning autism. A review. *Research in Autism Spectrum Disorders*, doi:10.1016/j.rasd.2009.05.002.

Maratsos, M., Kuczaj, S.A., Fox, D.E. & Chalkley, M.A. (1979). Some empirical studies in the acquisition of transformational relations: Passives, negatives, and the past tense. In W. A. Collins (Ed.), *Children's language and communication*. New Jersey: Lawrence Erlbaum.

Marcus, G.F., Pinker, S., Ullman, M., Hollander, M. Rosen, T.J., Xu, F., et al. (1992). Overregularization in Language Acquisition. *Monographs of the Society for Research in Child Development,* 57, 4, i-178.

Masip, J., Garrido, E. & Herrero, C. (2004). Defining deception. *Anales di Psicologia,* 20, 1, 147-171.

Matsunaga, M. (2007). Familywise Error in Multiple Comparisons: Disentangling a Knot through a Critique of O’Keefe’s Arguments against Alpha Adjustment. *Communication Methods and Measures,* 1, 4, 243-265.

McGarrigle, J. & Donaldson, M. (1974). Conservation Accidents. *Cognition*, 3, 4, 341-350.

McNeil, N.M. (2007). U-shaped development in Math: 7-year-olds out perform 9-year-olds on equivalence problems. *Developmental Psychology,* 43, 3, 687-695.

Meibauer, J. (2005). Lying and falsely implicating. *Journal of Pragmatics*, 37, 1373-1399.

Mitchell, R.W. & Thompson, N.S. (Eds.) (1986). Deception: Perspectives on human and non-human deceit. New York: Staten University of New York Press.

Morsanyi, K. & Handley, S.J. (2008). How smart do you need to be to get it wrong? The role of cognitive capacity in the development of heuristic-based judgement. *Journal of Experimental Child Psychology*, 99, 18-36.

Moxey, L.M. and Sanford, A.J. (1993a). *Communicating Quantities: A psychological perspective.* Hove: LEA

Moxey, L.M. & Sanford, A.J. (1993b). Prior Expectation and the Interpretation of Natural Language Quantifiers. *European Journal of Cognitive Psychology*, 5, 1, 73-91

Moxey, L.M. & Sanford, A.J. (2000). Communicating Quantities: A review of psycholinguistic evidence of how expressions determine perspectives. *Applied Cognitive Psychology*. 14, 3, 237-255.

Musolino, J. & Lidz, J. (2006). Why Children Aren't Universally Successful with Quantification. *Linguistics,* 44, 4, 817-852.

Musolino, J., Crain, S. & Thornton, R. (2000). Navigating negative quantificational space. *Linguistics*, 38, 1, 1-32.

Neimark, E.D. (1970). Development of comprehension of logical connectives; Understanding of 'Or'. *Psychonomic Science*, 21, 217-219.

Neimark, E.D. & Chapman, R.H. (1975). Development of the Comprehension of Logical Quantifiers. In R.J. Falmagne (Ed.), *Reasoning: Representation and Process.* New York: Wiley.

Neimark, E.D. & Slotnick, N.S. (1970). Development of the understanding of logical connectives. *Journal of Educational Psychology*, 61, 451-460.

Newstead, S.E. (1989). Interpretational Errors in Syllogistic Reasoning. *Journal of Memory and Language*, 28, 78-91.

Newstead, S.E. (1995). Gricean Implicatures and Syllogistic Reasoning. *Journal of Memory and Language*, 34, 644-664.

Newstead, S.E., Handley, S.J., Harley, C., Wright, H. & Farrelly, D. (2004). Individual differences in deductive reasoning. *Quarterly Journal of Experimental Psychology,* 57A, 33-60.

Ninio, A. & Snow, C.E. (1996). *Pragmatic Development*. Colorado: Westview Press.

Noveck, I.A. (2001). When children are more logical than adults: Experimental investigations of scalar implicature. *Cognition*, 78, 165-168.

Noveck, I.A. (2004). Pragmatic inferences related to logical terms. In I. A. Noveck and D. Sperber (Eds.), *Experimental Pragmatics*. Basingstoke: Palgrave Macmillian.

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

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

Noveck, I.A. & Reboul, A. (2008). Experimental Pragmatics: a Gricean turn in the study of language. *Trends in Cognitive Science*, 12, 11, 425-431.

Noveck, I.A. & Sperber, D. (2004). Introduction. In I. A. Noveck & D. Sperber (Eds.), *Experimental Pragmatics*. Basingstoke: Palgrave Macmillian.

Noveck, I.A. & Sperber, D. (2007). The Why and How of Experimental Pragmatics: the case of 'Scalar Inferences'. In N. Burton-Roberts (Ed.) *Pragmatics*. Basingstoke: Palgrave Macmillan.

Noveck, I.A., Ho, S. & Sera, M. (1996). Children's understanding of epistemic modals. *Journal of Child Language*, 23, 3, 621-643.

O'Keefe, D.J. (2003). Colloquy: Should Familywise Alpha be Adjusted? Against Familywise Alpha Adjustment. *Human Communication Research*, 29, 3, 431-447.

Osman, M. (2004). An Evaluation of Dual-Process Theories of Reasoning. *Psychonomic Bulletin and Review*, 11, 6, 988-1010.

Papafragou. A. (2003). Aspectuality and scalar structure. In *Proceedings of the 27th Boston University Conference on Language Development*, 626-637. Sommerville, MA: Cascadilla Press.

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

Papafragou, A. & Tantalou, N. (2004). Children's Computation of Implicatures. *Language Acquisition: A Journal of Developmental Linguistics*, 12, 1, 71-82.

Paris, S.G. (1973). Comprehension of language connectives and prepositional logical relationships. *Journal of Experimental Child Psychology*, 16, 278-291.

Paris, S.G. & Upton, L.R. (1976). Children's memory for inferential relationships in prose. *Child Development*, 47, 3, 660-668.

Parsons. T. (2008). The Traditional Square of Opposition. *The Stanford Encyclopedia of Philosophy (Fall 2008 Edition).* Edward N. Zalta (ed.), URL = <http://plato.stanford.edu/archives/fall2008/entries/square/>.

Paterson, K.B., Sanford, A.J., Moxey, L.M. & Dawydiak, E.J. (1998). Quantifier polarity and referential focus during reading. *Journal of Memory and Language,* 39, 2, 290-306.

Perneger, T.V. (1998). What's wrong with Bonferroni adjustments? *BMJ,* 316, 7139, 1236-1238.

Piaget, J. (2001). *The Psychology of Intelligence* (M.Piercy & D.E. Berlyne, Trans.). London: Routledge. (Original work published in 1947).

Piget, J. & Inhelder, B. (2000). *The Psychology of the Child* (H. Weaver, Trans.). New York: Basic Books. (Original work published in 1966).

Pijnacker, J., Hagoort, P., Buitelaar, J., Teunisse, J.P. & Geurts, B. (2009). Pragmatic inferences in high-functioning adults with Autism and Asperger Syndrome. *Journal of Autism and Developmental Disorders,* 39, 4, 607-618.

Plunkett, K. & Marchman, V. (1993). From rote learning to system building: Acquiring verb morphology in children and connectionist nets. *Cognition*, 48, 1, 21-69.

Politzer, G. (1986). Laws of language use and formal logic. *Journal of Psycholinguistic Research,* 15, 1, 47-92.

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

Recanati, F. (2003). Embedded Implicatures. *Philosophical Perspectives*, 17, 1, 299-332.

Roberts, M.J., Newstead, S.E. & Griggs, R.A. (2001). Quantifier Interpretation and Syllogistic Reasoning. *Thinking and Reasoning*, 7, 2, 173-204.

Rothman, K.J. (1990). No Adjustments are Needed for Multiple Comparisons. *Epidemiology*, 1,1, 43-46.

Rumelhart, D.E. & McClelland, J.L. (1986). On Learning Past Tenses of English Verbs. In D.E. Rumelhart & J.L.McClelland (Eds.), *Parallel Distributed Processing, Vol. 2* (pp. 216–271). Cambridge, MA: MIT Press.

Russell, B. (2006). Against Grammatical Computation of Scalar Implicatures. *Journal of Semantics*, 23, 4, 361-382.

Ryder, N. & Leinonen, E. (2003). Use of Context in Questioning Answering by 3-, 4- and 5-year-old children. *Journal of Psycholinguistic Research*, 32, 4, 397-415.

Sanford, A. J. & Moxey, L. M. (2003). New Perspectives on the Expression of Quantity. *Current Directions in Psychological Science*, 12, 240-243.

Sanford, A.J., Moxey, L.M. & Paterson, K.B. (1996). Attentional focusing with quantifiers in production and comprehension. *Memory and Cognition*, 24, 2, 144-155.

Sanford, A.J. & Moxey, L.M. (2004). Exploring Quantifiers: Pragmatics meets the psychology of comprehension. In I. A. Noveck and D. Sperber (Eds.) *Experimental Pragmatics* . Basingstoke: Palgrave Macmillian.

Sauland, U. (2004). Scalar Implicatures in Complex Sentences. *Linguistics and Philosophy*, 27, 3, 367-391.

Saxe, R., Tenenbaum, J.B. & Carey, S. (2005). Secret Agents: Inferences about hidden causes by 10- and 12-month-old infants. *Psychological Science*, 16, 12, 995-1001.

Saxe, R., Tzelnic, T. & Carey, S. (2007). Knowing who dunnit: Infants identify the causal agent in an unseen causal interaction. *Developmental Psychology*, 43, 1, 149-158.

Schank, R.C. (1976). The role of memory in language processing. In C Cofer (Ed.), *The Nature of Human Memory*. San Francisco: W.H. Freeman.

Schwartz, D.L. & Goldman, S.R. (1996). Why people are not like marbles in an urn: An effect of context on statistical reasoning. *Applied Cognitive Psychology*, 10, 99-112.

Siegler, R.S. (2004). U-shaped interest in U-shaped development- and what it means. *Journal of Cognition and Development*, 5, 1, 1-10.

Simons, M. (2000). *Issues in the semantics and pragmatics of disjunction.* New York: Garland.

Slabakova, R. (in press). Scalar implicatures in second language acquisition. *Lingua* doi:10.1016/j.lingua.2009.06.005.

Slaughter, V. & Gopnik, A. (1996). Conceptual Coherence in the Child's Theory of Mind: Training Children to Understand Belief. *Child Development*, 67, 6, 2967-2988.

Sloman, S.A. (1996). The Empirical Case for Two Systems of Reasoning. *Psychological Bulletin*, 119, 1, 3-22.

Sloman, S.A. (2002). Two Systems of Reasoning. In T. Gilovich, D. Griffin and D. Kahneman (Eds.), *Heuristics and Biases: The Psychology of Intuitive Judgment*. Cambridge: Cambridge University Press.

Smith, C.L. (1980). Quantifiers and Question Answering in Young Children. *Journal of Experimental Child Psychology*, 30, 191-205.

Sodian, B., Taylor, C., Harris, P. L. & Perner, J. (1991). Early deception and the child's theory of mind: False trails and genuine markers *Child Development,* 62, 3, 468-483.

Spector, B. (2007). Scalar Implicatures: Exhaustivity and Gricean Reasoning. In M. Aloni, A. Butler & P Dekker (Eds.), *Questions in Dynamic semantics* (pp. 225-250). Oxford: Elsevier.

Sperber, D. (1994). Understanding verbal understanding. In J. Khalfa (Ed.), *What is Intelligence?* Cambridge: Cambridge University Press.

Sperber, D. (1995). How do we communicate? In J. Brockman & K. Matson (Eds.), *How things are: A science toolkit for the mind*. New York: Morrow.

Sperber, D. & Wilson, D. (1985/1995). *Relevance: Communication and cognition.* Oxford: Blackwell.

Sperber, D. & Wilson, D. (2002). Pragmatics, Modularity and Mind-reading. *Mind and Language,* 17, 1-2, 3-23.

Stanovich, K.E. (1999). *Who is Rational? Studies of individual differences in reasoning.* Hillsdale, NJ: Lawrence Erlbaum Associates, Inc.

Stanovich, K.E. & West, R.F. (1998). Individual differences in rational thought. *Journal of Experimental Psychology: General,* 127, 161-188.

Stanovich, K.E. & West, R.F. (2000). Individual Differences in Reasoning: Implications for the rationality debate? *Behavioral and Brain Sciences,* 23, 645-726.

Stavy, R., Strauss, S., Orpaz, N. & Carmi, G. (1982). In S. Strauss & R. Stavy (Eds.), *U-shaped Behavioral Growth* (pp.11-36). New York: Academic Press.

Suppes, P. & Feldman, S. (1971). Young children’s comprehension of logical connectives. *Journal of Experimental Child Psychology,* 12, 304-317.

Sutter, M. (2009). Deception through telling the truth?! Experimental evidence from individuals and teams. *The Economic Journal*, 119, 534, 47-60.

Surian, L., Baron-Cohen, S. & Van der Lely, H. (1996). Are children with autism deaf to Gricean maxims? *Cognitive Neuropsychiatry,* 1, 1, 55-71.

Tager-Flusberg, H. (1981). On the nature of linguistic functioning in early infantile autism. *Journal of Autism and Developmental Disorders,* 11, 1, 45-56.

Tager-Flusberg, H. (1993). Putting words together: Morphology and syntax in the preschool years. In J.B. Gleason (Ed.), *The Development of Language* (pp. 151-194). New York: Macmillian.

Tager-Flusberg, H. (1999). Language development in atypical children. In M. Barrett (Ed.), *The Development of Language* (pp.311-348). Hove: Psychology Press.

The Psychological Corporation. (1999). *Wechsler Abbreviated Scale of Intelligence*. San Antonio: The Psychological Corporation.

Tomasello, M. (2000). The social-pragmatic theory of word learning. *Pragmatics*, 10, 4, 401-414.

Tomasello, M. & Brooks, P.J. (1999). Early syntactic development: A construction grammar approach. In M. Barrett (Ed.), *The Development of Language* (pp. 161-190). Hove: Psychology Press.

Torrens, D., Thompson, V.A. & Cramer, K. (1999). Individual differences and the belief-bias effect: Mental models, logical necessity, and abstract reasoning. *Thinking and Reasoning,* 5, 1-28.

Trueswell, J.C., Sekerina, I., Hill, N.M. & Logrip, M.L. (1999). The kindergarten-path effect: Studying on-line sentence processing in young children. *Cognition*, 73, 2, 89-134.

Tutzauer, F. (2003). On the Sensible Application of Familywise Alpha Adjustment. *Human Communication Research*, 29, 3, 455-463.

Van der Lely, H.K.J. (1996). Specifically language impaired and normally developing children: Verbal passive vs. adjectival passive sentence interpretation. *Lingua*, 98, 4, 243-272.

van Dijk, M. & van Geert, P. (2007). 'Wobbles, humps and sudden jumps': A case study of continuity, discontinuity and variability in early language development. *Infant and Child Development*, 16, 1, 7-33.

Vincent, J.M. & Castelfranchi, C. (1981). On the art of deception: How to lie while saying the truth. In H. Parret, M. Sbisà & J. Verschueren (Eds.), *Proceedings of the Conference on Pragmatics, Possibilities and Limitations of Pragmatics* (pp. 749-777). Amsterdam: John Benjamins.

Westermann, G. (1998). Emergent modularity and U-shaped learning in a constructivist neural network learning the English past tense. *Proceedings of the 28th Annual Conference of the Cognitive Science Society*, 1130-1135.

Wildman, T.M. & Fletcher, H.J. (1977). Developmental increases and decreases in solutions of conditional syllogism problems. *Developmental Psychology*, 13, 6, 630-636.

Wilkins, M.C. (1928). The effect of changed material on ability to do formal syllogistic reasoning. In R.S. Woodworth (Ed.), *Archives of psychology*. Columbia University: New York.

Wilson, D. & Sperber, D. (2002a). Relevance theory: A tutorial. In Y. Otsu (Ed.), *Proceedings of the Third Tokyo Conference on Psycholinguistics.* Tokyo: Hituzi Shobo.

Wilson, D. & Sperber, D. (2002b). Truthfulness and Relevance. *Mind*, 111, 443, 583-632.

Wilson, D. & Sperber, D. (2004) Relevance Theory. In L.R. Horn and G Ward (Eds). *The Handbook of Pragmatics*. Blackwell Publishing, 2005. Blackwell Reference Online. 21 November 2008. <http://www.blackwellreference.com/subscriber/book?id=g9780631225485_9780631225485>.

Winer, G.A. (1990). Variations in the use of the term or. *Journal of Psycholinguistic Research*, 19, 1, 1-20.

Zuckerman, M., DePaulo, B.M. & Rosenthal, R. (1981). Verbal and non-verbal communication of deception. In L. Berkowitz (Ed.), *Advances in Experimental Social Psychology* (Vol. 14, pp. 2-59). New York: Academic Press Inc.

Appendix AI

Statement Materials Used in Experiments 1 and 2

Trial	Quantifier Statements	Trial	Connective Statements		Trial
TA/ IS	All/some snow is cold	TA/IO	Snow is cold and/or white	Snow is cold and/or green	FA/FO
TA/ IS	All/some elephants have trunks	TA/IO	An elephant has a trunk and/or ears	An elephant has a trunk and/or socks	FA/FO
TA/ IS	All/ some books have pages	TA/IO	A book has pages and/or a title	A book has pages and/or a nose	FA/FO
TA/ IS	All/some giraffes have long necks	TA/IO	A giraffe has a long neck and/or a tongue	A giraffe has a long neck and/or a jumper	FA/FO
TA/ IS	All/some cats have ears	TA/IO	A cat has fur and/or eyes	A cat has fur and/or wings	FA/FO
TA/ IS	All/some pennies are round	TA/IO	A penny is round and/or brown	A penny is round and/or blue	FA/FO
TA/ IS	All/some schools have teachers	TA/IO	A school has classrooms and/or teachers	A school has classrooms and/or spaceships	FA/FO
TA/ IS	All/some airplanes have wings	TA/IO	An aeroplane has wings and/or an engine	An aeroplane has wings and/or slippers	FA/FO
FA/FS	All/ some flowers are yellow	TA/IO	A flower has petals and/or a stem	A flower has petals and/or a piano	FA/FO
FA/FS	All/some birds live in cages	TA/IO	A bird has feathers and/or a beak	A bird has feathers and/or a computer	FA/FO

FA/FS	All/some dogs have spots	TA/IO	A dog has a tail and/or a nose	A dog has a tail and/or an umbrella	FA/FO
FA/FS	All/some children are blonde	TA/IO	A child has toes and/or fingers	A child has toes and/or a beak	FA/FO
FA/FS	All/some animals have stripes	TA/IO	An animal breathes and/or eats	An animal breathes and/or sings	FA/FO
FA/FS	All/ some bicycles are pink	TA/IO	A bicycle has handle bars and/or pedals	A bicycle has handle bars and/or gloves	FA/FO
FA/FS	All/some houses have a garage	TA/IO	A house has doors and/or windows	A house has doors and/or wheels	FA/FO
FA/FS	All/ some cars are red	TA/IO	A car has wheels and/or doors	A car has wheels and/or feathers	FA/FO
Absurd	All/ some mice eat trains				
Absurd	All/some garages sing				
Absurd	All/some peas wear shorts				
Absurd	All/some bubbles breath fire				
Absurd	All/ some policemen are made of jelly				
Absurd	All/some crayons have noses				
Absurd	All/ some teapots are fluffy				
Absurd	All/ computers are made of chocolate				

Appendix AII

Storyboard themes for Experiments 1 and 2

Quantifier	Connective
Eating biscuits	Eating biscuits
Eating chocolate	Eating chocolate
Breaking tiles	Breaking tiles
Picking flowers	Picking petals off flowers
Putting salt in tea	Put vinegar in tea
Pulling doll’s heads off	Cut legs off toys
Pulling children’s hair	Pulling children’s hair
Tying shoelaces together	Putting Cheerioes in the shoes
Letting tyres down	Puncturing tyres
Eating sweets	Eating sweets
Putting pepper in sandwiches	Pepper in sandwich
Dirtying towels	Dirtying towels
Eating cake	Eating cakes
Putting worms in lunch box	Spider in lunch box
Drawing on books	Tearing books
Spilling pop	Spilling bottles

Appendix AIII

Statement Materials Used in Experiments 3

Universal	Existential	Absurd
Airplanes have wings	Bicycles are blue	Chairs laugh
Books have pages	Cars are red	Clocks cry
Rabbits have ears	Clothes have zips	Mice eat trains
Elephants have trunks	Flowers are pink	Spoons wear hats
Giraffes have long necks	Paint is white	Doors dance
Schools have teachers	Animals have stripes	Plates run
Snow is cold	Birds live in cages	Carrots wear shorts
Fish swim	Drinks are made of chocolate	Policemen are made of jelly

Appendix AIV

Toys used in Bags task in Experiment 4

Bouncy balls

Yo-yos

Musical pipes

Kazoos

Purses

Whistles

Bats and balls

Cars

Chocolate buttons

Chomps

Lollies

Caramel Freddos

Milky bars

Fudges

Chocolate Freddos

Dairy milks

Appendix AV

Storyboard themes for Experiments 5 and 6

<u>The objects and method of interaction</u>

Eating lollies	Eating crisps
Pulling dolls heads off	Breaking Lego houses
Spilling glasses of juice	Drinking pop in bottles
Cereal in slippers	Drawing on books
Eating chocolate eggs	Eating chocolate biscuits
Crushing cans	Breaking eggs
Knocking paint pots over	Drinking juice in glasses
Putting food colouring in water	Tying shoelaces together
Eating cakes	Eating chocolate bars
Popping balloons	Tearing magazines
Knocking vases over	Drinking juice in cartons
Dirtying towels	Drawing on posters
Eating boiled sweets	Eating cheeses
Breaking tiles	Cutting fruit
Spilling mugs of tea	Drinking milk in glasses
Putting worms in containers	Putting grass in sandwiches

Appendix AVI

Descriptive storyboard themes for Experiment 6

Objects that were described:

Birthday cards	Buns
Teddies	Jigsaws
Photoframes	Trophies
Playing cards	Pick ’n’ mix
Plates	Videos
Tarts	Receipts
Easter eggs	Uno cards
Paint cards	Postcards
Tiles	Biscuits
Beans	Wooden spoons
Socks	Kitchen utensils
Medals	Chocolate mousses
Oranges	Digestives
Fruit	Newspapers
Travel brochures	Ornaments
Mugs	Toys

Appendix AVII

Sentences used in Experiment 7

Set 1

Infelicitous some

Some books have pages

Some schools have teachers

Some elephants have trunks

Some snow is cold

Some rabbits have ears

Some airplanes have wings

Some giraffes have long necks

Some fish swim

Felicitous some

Some paint is white

Some animals have stripes

Some birds live in cages

Some houses have a garage

Some children have blonde hair

Some bicycles are pink

Some dogs have spots

Some cars are red

Absurd

All chairs laugh

All clocks cry

All policemen are made of jelly

All spoons wear hats

Some doors dance

Some plates run

Some carrots wear shorts

Some mice eat trains

True all

All fire is hot

All rain is wet

False all

All dresses have pockets

All drinks are fizzy

Set 2

Infelicitous some

Some feathers are light

Some babies breathe

Some people have faces

Some zebras have stripes

Some hedgehogs have spikes

Some cats can purr

Some authors write

Some dogs can bark

Felicitous some

Some shoes are black

Some women wear glasses

Some books have colour pictures

Some toys are dolls

Some dogs are brown

Some chalk is white

Some bags are pink

Some men have beards

Absurd

All teapots are fluffy

All spoons tell jokes

All forks giggle

All fruit have legs

Some windows dream

Some peas wear jumpers

Some garages sing

Some cups shout

True all

All cows can moo

All winds blow

False all

All pets are cats

All food is chocolate

Appendix AVIII

Sentences used in Experiment 8

Universal	Existential	Absurd
snails have shells	men have hats	books have legs
ducks have beaks	shoes are black	pies have ears
owls have wings	bags are pink	chairs have eyes
days have hours	cars are red	doors have feet
grapes are fruit	tools are saws	doors are dogs
cod are fish	pets are mice	hats are fruit
swans are birds	sweets are mints	plums are fish
shirts are clothes	toys are dolls	cats are trees
snow is cold	bikes are blue	books have legs
drinks are wet	clothes have zips	pies have ears
fire is hot	skirts are grey	doors have feet
months have days	leaves are brown	chairs have eyes
beef is meat	food is bread	doors are dogs
drills are tools	shapes are cubes	hats are fruit
twigs are wood	meat is pork	plums are fish
squares are shapes	fish are pets	cats are trees

Appendix BI

Additional statistical analysis for Experiment 1

Table 1

Wilcoxon signed ranks analysis of the number of logical responses given to the all statements compared to the some statements for each age group

	Year 1	Year 4	Adult
z	2.03	2.47	3.56
p	.042	.013	.000
r	.31	.36	.51

Table 2

Wilcoxon signed ranks analysis of the number of logical responses given to the true quantifier statements compared to the false quantifier statements for each age group

	Year 1	Year 4	Adult
z	2.62	3.20	3.56
p	.009	.001	.000
r	.40	.42	.51

Table 3

Friedman analysis of the number of logical responses given to each quantifier statement type for each age group

	Year 1	Year 4	Adult
χ^2	7.67	16.41	35.22
df	3	3	3
p	.053	.001	.000

Table 4

Wilcoxon signed ranks analysis of the number of logical responses given to the target trials compared to the control trials in the quantifier statement task, for each age group

		Year 1	Year 4	Adult
True all/ Infelicitous some	*z*	2.53	2.83	3.56
	p	.011	.005	.000
	r	.38	.41	.51
False all/ Infelicitous some	*z*	1.56	2.62	3.59
	p	.119	.009	.000
	r	.24	.38	.52
Felicitous some/ Infelicitous some	*z*	1.79	2.85	3.59
	p	.074	.004	.000
	r	.27	.41	.52

Table 5

Wilcoxon signed ranks analysis of the number of logical responses given to the all storyboards compared to the some storyboards for each age group

	Year 1	Year 4	Adult
z	3.99	4.29	3.63
p	.000	.000	.000
r	.60	.62	.52

Table 6

Wilcoxon signed ranks analysis of the number of logical responses given to the true quantifier storyboards compared to the false quantifier storyboards for each age group

	Year 1	Year 4	Adult
z	3.54	4.16	3.76
p	.000	.000	.000
r	.53	.60	.54

Table 7

Friedman analysis of the number of logical responses given to each quantifier storyboard subgroup for each age group

	Year 1	Year 4	Adult
χ^2	33.07	54.43	37.84
df	3	3	3
p	.000	.000	.000

Table 8

Wilcoxon signed ranks analysis of the number of logical responses given to the target trials compared to the control trials in the quantifier storyboard task

		Year 1	Year 4	Adult
True all/	*z*	3.95	4.46	3.75
Infelicitous some	*p*	.000	.000	.000
	r	.60	.64	.54
False all/	*z*	3.96	4.36	3.71
Infelicitous some	*p*	.000	.000	.000
	r	.60	.63	.54
Felicitous some/	*z*	3.26	4.16	3.71
Infelicitous some	*p*	.001	.000	.000
	r	.50	.60	.54

Table 9

Wilcoxon signed ranks analysis of the number of logical responses given to the and statements compared to the or statements for each age group

	Year 1	Year 4	Adult
z	4.21	3.65	4.22
p	.000	.000	.000
r	.63	.53	.61

Table 10

Wilcoxon signed ranks analysis of the number of logical responses given to the true connective statements compared to the false connective statements for each age group

	Year 1	Year 4	Adult
z	4.19	2.75	.29
p	.000	.006	.774
r	.63	.40	.04

Table 11

Friedman analysis of the number of logical responses given to each connective statement type for each age group

	Year 1	Year 4	Adult
χ^2	56.73	32.73	54.54
df	3	3	3
p	.000	.000	.000

Table 12

Wilcoxon signed ranks analysis of the comparison of the number of logical responses given to the target trials compared to the control trials in the connective statement task for each age group

		Year 1	Year 4	Adult
True and/ Infelicitous or	*z*	2.67	2.39	4.20
	p	.008	.017	.000
	r	.40	.34	.61
False and/ Infelicitous or	*z*	3.05	1.44	4.20
	p	.002	.150	.000
	r	.46	.21	.61
Felicitous or/ Infelicitous or	*z*	4.22	3.28	0.35
	p	.000	.001	.73
	r	.64	.47	.05
True and/ Felicitous or	*z*	4.52	3.87	4.01
	p	.000	.000	.000
	r	.68	.56	.58
False and/ Felicitous or	*z*	4.60	3.56	4.06
	p	.000	.000	.000
	r	.69	.51	.59

Table 13

Wilcoxon signed ranks analysis of the number of logical responses given to the and storyboards compared to the or storyboards for each age group

	Year 1	Year 4	Adult
z	4.12	3.64	4.23
p	.000	.000	.000
r	.62	.53	.61

Table 14

Wilcoxon signed ranks analysis of the number of logical responses given to the true connective storyboards compared to the false connective storyboards for each age group

	Year 1	Year 4	Adult
z	0.93	2.17	3.01
p	.367	.030	.003
r	.14	.31	.43

Table 15

Friedman analysis of the number of logical responses given to each connective storyboard type for each age group

	Year 1	Year 4	Adult
χ^2	51.21	37.74	53.81
df	3	3	3
p	.000	.000	.000

Table 16

Wilcoxon signed ranks analysis of the number of logical responses given to the target trials compared to the control trials in the connective storyboard task for each age group

		Year 1	Year 4	Adult
True and/	*z*	3.86	3.15	4.29
Infelicitous or	*p*	.000	.002	.000
	r	.58	.45	.62
False and/	*z*	3.79	3.37	4.27
Infelicitous or	*p*	.000	.001	.000
	r	.58	.49	.62
Felicitous or/	*z*	0.99	1.46	2.99
Infelicitous or	*p*	.321	.144	.003
	r	.15	.21	.43
True and/	*z*	4.16	3.45	3.74
Felicitous or	*p*	.000	.001	.000
	r	.63	.50	.54
False and/	*z*	4.01	3.69	3.71
Felicitous or	*p*	.000	.000	.000
	r	.60	.53	.54

Appendix BII

Additional statistical analysis for Experiment 2

Table 1

Wilcoxon signed ranks analysis of the number of logical responses given to the all statements compared to the some statements for each age group

	Year 1	Year 4	Year 7	Year 10	Adult
z	1	3.38	4.35	3.26	2.83
p	3.17	.001	.000	.001	.005
r	.15	.51	.66	.56	.49

Table 2

Wilcoxon signed ranks analysis of the number of logical responses given to the true statements compared to the false statements for each age group

	Year 1	Year 4	Year 7	Year 10	Adult
z	1.39	3.48	4.31	3.26	2.92
p	.166	.000	.000	.001	.003
r	0.21	.52	.65	.56	.51

Table 3

Friedman analysis of the number of logical responses given to each statement type for each age group

	Year 1	Year 4	Year 7	Year 10	Adult
χ^2	5.71	31.53	62.23	28.46	29.734
df	3	3	3	3	3
p	.127	.000	.000	.000	.000

Table 4

Wilcoxon signed ranks analysis of the number of logical responses given to the target trials compared to the control trials in the statement task for each age group

		Year 4	Year 7	Year 10	Adult
True all/ Infel some	*z*	3.40	4.31	3.26	2.96
	p	.001	.000	.001	.003
	r	.51	.65	.56	.51
False all/ Infel some	*z*	3.30	4.46	3.27	2.96
	p	.001	.000	.001	.003
	r	.49	.67	.56	.51
True some/ Infel some	*z*	3.38	4.40	3.27	3.09
	p	.001	.000	.001	.002
	r	.51	.66	.56	.53

Table 5

Wilcoxon signed ranks analysis of the number of logical responses given to the all storyboards compared to the some storyboards for each age group

	Year 1	Year 4	Year 7	Year 10	Adult
z	4.04	4.18	4.40	3.13	2.77
p	.000	.000	.000	.000	.006
r	.61	.63	.66	.54	.48

Table 6

Wilcoxon signed ranks analysis of the number of logical responses given to the true storyboards compared to the false storyboards for each age group

	Year 1	Year 4	Year 7	Year 10	Adult
z	3.92	4.12	4.46	3.23	3.45
p	.000	.000	.000	.001	.001
r	.59	.62	.67	.55	.59

Table 7

Friedman analysis of the number of logical responses given to each storyboard type for each age group

	Year 1	Year 4	Year 7	Year 10	Adult
χ^2	48.89	56.38	63.39	29.36	30.00
df	3	3	3	3	3
p	.000	.000	.000	.000	.000

Table 8

Wilcoxon signed ranks analysis of the number of logical responses given to the target trials compared to the control trials in the storyboard task for each age group

		Year 1	Year 4	Year 7	Year 10	Adult
True all/ Infel some	*z*	4.13	4.28	4.46	3.24	3.34
	p	.000	.000	.000	.001	.001
	r	.62	.65	.67	.56	.57
False all/ Infel some	*z*	4.01	4.35	4.40	3.27	2.81
	p	.000	.000	.000	.001	.005
	r	.60	.66	.66	.56	.48
True some/ Infel some	*z*	3.89	4.22	4.46	3.21	3.36
	p	.000	.000	.000	.001	.001
	r	.59	.64	.67	.55	.58

Table 9

Mann-Whitney analysis of the number of logical responses to the infelicitous some storyboards for adjacent age groups

	Year 1 and Year 4	Year 4 and Year 7	Year 7 and Year 10	Year 10 and Adult
U	185	218.5	126.5	138.5
z	1.62	0.82	2.24	0.23
p	.104	.410	.025	.821
r	.24	.12	.36	.03

Table 10

Wilcoxon signed ranks analysis of the number of logical responses given to the infelicitous some trials between tasks for each age group

	Year 7	Year 10	Adult
z	.45	.00	1.85
p	.655	1.00	.064
r	.07	0	.32

Appendix BIII

Additional statistical analysis for Experiment 3

Table 1

Wilcoxon signed ranks analysis of the number of logical responses given to the target trial and control trials in the statement task

	True all/ False all	True all/ Fel some	True all/ Infel some	False all/ Fel some	False all/ Infel some	Fel some/ Infel some
z	3.37	1.61	1.10	2.91	3.17	1.63
p	.001	.107	.271	.004	.002	.102
r	.58	.28	.19	.50	.54	.28

Table 2

Wilcoxon signed ranks analysis of the number of logical responses given to the target and control trials in the pictureboard task

	True all/ False all	True all/ Fel some	False all/ Fel some	True all/ Infel some	False all/ Infel some	Fel some/ Infel some
z	.38	1.77	1.50	2.91	2.87	1.82
p	.705	.077	.136	.004	.004	.069
r	.07	.30	.26	.50	.50	.31

Appendix BIV

Additional statistical analysis for Experiment 4

Table 1

Wilcoxon signed ranks analysis of the number of logical responses given to the target trials compared to the control trials in the pictureboard task

	True all/ Infel some	False all/ Infel some	True some/ Infel some
z	2.85	1.38	1.58
p	.004	.169	.114
r	.50	.24	.28

Table 2

Wilcoxon signed ranks analysis of the number of logical responses given to the target trials compared to the control trials in the bag task

	True all/ Infel some	False all/ Infel some	True some/ Infel some
z	2.88	2.35	2.96
p	.004	.019	.003
r	.51	.42	.52

Table 3

Wilcoxon signed ranks analysis of the number of logical responses to the target trials compared to the control trials, by order of presentation

	True all/ Infel some	False all/ Infel some	True some/ Infel some
First task			
z	2.39	0.81	0.81
p	.017	.419	.421
r	.60	.20	.20
Second Task			
z	3.31	2.91	3.22
p	.001	.004	.001
r	.59	.51	.57

Appendix BV

Additional statistical analysis for Experiment 5

Table 1

Wilcoxon signed ranks analysis of the number of logical responses to the target trials compared to the control trials in the agent condition for the Year 1 children

	True all/ Infel some	False all/ Infel some	True some/ Infel some
z	3.91	3.89	3.29
p	.000	.000	.001
r	.62	.62	.52

Table 2

Wilcoxon signed ranks analysis of the number of logical responses to the target trials compared to the control trials in the agent condition for the Year 4 children

	True all/ Infel some	False all/ Infel some	True some/ Infel some
z	4.92	4.87	4.92
p	.000	.000	.000
r	.67	.66	.67

Table 3

Wilcoxon signed ranks analysis of the number of logical responses to the target trials compared to the control trials in the non-agent condition for the Year 1 children

	True all/ Infel some	False all/ Infel some	True some/ Infel some
z	3.53	2.78	3.16
p	.000	.005	.002
r	.56	.44	.50

Table 4

Wilcoxon signed ranks analysis of the number of logical responses to the target trials compared to the control trials in the non-agent condition for the Year 4 children

	True all/ Infel some	False all/ Infel some	True some/ Infel some
z	5.11	5.04	5.11
p	.000	.000	.000
r	.70	.69	.70

Appendix BVI

Additional statistical analysis for Experiment 6

Table 1

Wilcoxon signed ranks analysis of the number of logical responses to the target trials compared to the control trials for both conditions of the interactive set of storyboards

		True all/ Infel some	False all/ Infel some	True some/ Infel some
Agent condition	*z*	4.17	4.10	3.92
	p	.000	.000	.000
	r	.63	.62	.59
Non-agent condition	*z*	4.09	3.98	3.30
	p	.000	.000	.001
	r	.62	.60	.50

Table 2

Wilcoxon signed ranks analysis of the number of logical responses to the target trials compared to the control trials for both conditions of the descriptive set of storyboards

		True all/ Infel some	False all/ Infel some	True some/ Infel some
Agent condition	*z*	3.49	3.19	1.51
	p	.000	.001	.132
	r	.48	.44	.21
Non-agent condition	*z*	2.77	2.39	1.22
	p	.006	.017	.222
	r	.38	.33	.17

Appendix BVII

Additional analysis for Experiment 7

Additional analysis was carried out on all the participants, taking the post-test data into account. One participant was excluded as he responded false to all the statements in the post-test.

Control Statements

Analysis of the control statements revealed that both sets of participants answered these well. The means (and standard deviations) can be seen in Table 1.

Table 1

The mean number of logical responses (and standard deviations) for the control statements, according to age and condition, for Experiment 8

Age	Condition	
	No Load	Load
Adult	19.71	19.78
	(0.64)	(0.46)
Child	19.80	19.77
	(0.64)	(0.43)

Note. The maximum score in each cell is 20

There was no significant difference in the mean number of logical responses given to the control statements between the adults and children for either condition, Mann Whitney U tests, no load, U = 744.00, p = .772, r = .03, load, U = 740.00, p = .729, r = .03. Nor was there any difference in the number of logical responses given to the control statements between the load and no load condition within either age group, Wilcoxon matched pairs signed ranks tests, adults, $z = .83$, $p = .405$, $r = .04$, children, $z = .33$, $p = .739$, $r = .04$.

Infelicitous some statements

In order to take into account the results of the post test, that is those *infelicitous some* statements that the participants considered to be *felicitous*, the *infelicitous some* scores were converted into a proportion of the number of statements the participants considered to be true of *all*. The mean proportion of logical responses (and standard deviations) for the *infelicitous some* statements for each age group can be seen in Table 2.

Table 2

Mean proportion of logical responses (and standard deviations) to the infelicitous some statements, according to age and condition, for Experiment 8

Age	Condition	
	No Load	Load
Adult	.66	.70
	(.43)	(.40)
Child	.58	.55
	(.45)	(.47)

Adult responses. The distribution of the adult data was unimodal in both conditions. The majority response, within each condition, was consistent logical responding. There was little pragmatic responding over all and very little consistent pragmatic responding within each condition. With regards to the number of logical responses given to the *infelicitous some* statements between conditions there was no significant difference, Wilcoxon matched pairs signed ranks test, $z = 0.83$, $p = .405$, $r = .04$. Analysis of the effects of order of presentation of the experimental conditions revealed that the adults who received the load condition first gave significantly more logical responses to the *infelicitous some* statements in both the load condition, Mann Whitney test, $U = 221.50$, $p = .030$, $r = .30$, and the subsequent no load condition, Mann Whitney test, $U = 191.5$, $p = .006$, $r = .39$ compared to those adults who received the no load condition first. Analysing the rates of logical responding for the condition that was presented first revealed that adults who attempted the load condition first gave significantly more logical responses to the *infelicitous some* trials compared to those adults who attempted the no load condition first, Mann-Whitney test, $U = 194.5$, $p = .007$, $r = .38$.

Children's responses. The distribution of the *infelicitous some* data for the children was unimodal in both conditions, with the majority of children giving consistent logical responses, although a quarter of the children were consistently pragmatic in the no load condition and there was a similar number in the load condition. With regards to the number of logical responses given to the *infelicitous some* statements between conditions there was no significant difference, Wilcoxon matched pairs signed ranks test, $z = 0.33$, $p = .739$, $r = .04$. However, unlike the adults, there was no effect of order of presentation in either the no load condition, Mann Whitney test $U = 88.50$, $p = .290$, $r = .19$, or the load condition, Mann Whitney test $U = 94.50$, $p = .417$, $r = .15$. Analysing the rates of logical responding to the infelicitous some trials for the condition that was presented first revealed that there was no significant difference, between conditions, Mann-Whitney test, $U = 85.5$, $p = .239$, $r = .21$.

Comparison of adult and children's results

Although the adults gave more logical responses than the children to the *infelicitous some* statements in both conditions, see Table 2, these differences were not significant, Mann Whitney tests, no load, U = 684.50, p = .394, r = .09, load, U = 635.50, p = .162, r = .16. This is not the case though if load order is taken into account. The adults and children who received the no load condition first did not differ significantly in the number of logical responses given to the *infelicitous some* trials in this condition, Mann Whitney test U = 160.5, p = .252, r = .18, or the subsequent load condition, Mann Whitney test U = 192.50, p = .778, r = .04. However, there was a significant difference in the mean number of logical responses given to the *infelicitous some* sentences between the adults and children who received the load condition first. The adults gave significantly more logical responses in both the load condition, Mann Whiney test, U = 108.50, p = .018, r = .38, and the no load condition, Mann Whitney test, U = 115.50, p = .029, r =.35.

Appendix BVIII

Additional analysis for Experiment 8

Table 1

Wilcoxon signed ranks analysis of the adults mean proportions of logical responses for the infelicitous some trials compared to the control trials

	True all/ Infel some	False all/ Infel some	True some/ Infel some
z	3.42	3.18	3.79
p	.001	.001	.000
r	.45	.42	.50

Table 2

Unrelated t test analysis of children's mean reaction times (m/s) to the control trials compared to the adults

	True All	False All	Absurd All	Felicitous Some	Absurd Some
Adult reaction time	1514	1471	1292	1485	1423
Child reaction time	2497	2380	2236	2392	2202
df	58.51	62	56.04	62	62
t	7.58	7.92	6.95	7.48	5.61
p	.001	.001	.001	.001	.001
r	.70	.71	.68	.69	.58

Table 3

Unrelated t test analysis of mixed responders counting span scores compared to logical responders for both age groups

Age group	Mean counting span score		*t*	*df*	*p*	*r*
Adult	mixed	3.81	0.25	24	.807	.05
	logical	3.74				
Child	mixed	3.23	1.06	33	.296	.18
	logical	2.99				

Appendix CI

Sensitivity to scalar implicatures in children with autism

Method

Participants

Thirteen children participated in the experiment. They were recruited from two specialist schools in the North East of England. The children were aged between 9 years 8 months and 16 years 3 months (mean age 14 years 1 month). All the children had autism as a part of their diagnosis. Consent for the experiment was obtained from one of the children's parents or the head teacher of the school.

Materials

The children heard a list of 24 statements and saw a series of 16 picture books.

Statements. The statements were the same as those used in Experiment 3 (see Appendix CII Table 1 for statements used). Each statement was presented with both *all* and *some* preceding it, although each participant received only one form, which resulted in two sets of materials. The statements within each set were randomly ordered.

Picture books. The picture books were based on the storyboards used in previous experiments in this thesis and featured Charlotte and the three objects that she interacts with. An example of a picture book can be seen in the Figure C1.

Page 1

Here is a story about a girl called Charlotte

Page 2

Charlotte finds three sandwiches

Page 3

She puts grass in the first sandwich

Page 4

She puts grass in the second sandwich

Page 5

She puts grass in the third sandwich.

Charlotte’s Mum says to Charlotte, “Charlotte what have you done to the sandwiches?

Figure C1. An example of a picture book, with accompanying dialogue.

As can be seen in Figure C1 each picture book contained five pages; each page was a single laminated photograph. The first photograph in each picture book was the same, showing Charlotte and her mother. The second photograph introduced the three objects, whilst the third and fourth photographs showed the result of Charlotte's interaction with the first and second objects respectively. The last photograph depicted either the result of Charlotte's interaction with the third object or was a repeat of the previous photograph, depending on which trial it was representing. There was a different set of objects in each of the picture books. Each photograph was accompanied by dialogue which followed a similar format to the storyboards used in the previous experiments of this thesis. There were 16 themes for the picture books (see Appendix CII Table 2 for themes used) and each theme appeared with the quantifiers *all* and *some*, although each child received only one form. This resulted in two sets of storyboards. There were two pairings of statements and picture book materials and order of presentation was counterbalanced amongst the children resulting in four groups.

In addition to the receiving the experimental materials the children were tested on measures of language and IQ:

The language measure used was the British Picture Vocabulary Scale-II (BPVS-II; Dunn, Dunn, Whetton & Burley, 1997). This individually administered, multi-choice test measures a participant's receptive vocabulary. This particular test was chosen for a number of reasons; ease of administration, and the relatively short testing time, both of which are important considerations when testing children with autism. An additional reason was the BPVS-II's links with the Peabody Picture Vocabulary Test (PPVT; Dunn & Dunn, 1981, 1997). The PPVT is the most widely used measure for matching children with autism to control groups of typically developing children (Tager-Flusberg, 1999) and the second edition was adapted to provide the original BPVS (Dunn, Dunn & Whetton, 1981). For the test each participant was presented with a spoken word stimulus, they then had to pick from a group of four black and white illustrations, the one that best represented the meaning of the stimulus item. The items in the BPVS are arranged into 14 sets, with 12 items in each set, and each set gets progressively harder. For each child a basal set was established, this is the set in which they made only one error, and testing continued until the ceiling set was reached, a set where the participant made eight or more errors. A raw score was obtained by subtracting the number of errors made by the participant in the sets from the basal through to the ceiling, from the last item number of the ceiling set. This raw score was then converted to a standardised score, and an age equivalent score was then obtained.

The IQ measure used was the Wechsler Abbreviated Scale of Intelligence (WASI; Psychological Corporation, 1999). This individually administered test is designed to tap into various facets of intelligence. It was chosen as it provides a short reliable measure of intelligence and is linked to the Wechsler Intelligence Scale for Children- III (WISC-III; Wechsler, 1991) and the Wechsler Adult Intelligence Scale-III (WAIS-III; Wechsler, 1997), the most commonly used intelligence tests (Long, 2000; Mackintosh, 1998). The WASI yields three traditional IQ scores; verbal, performance, and full scale, and consists of four subtests; vocabulary, block design, similarities and matrix reasoning. Each child was tested on all four subtests. A raw score was obtained for each of the subtests, which was converted into a *T* score, and these were converted into WASI IQ age equivalent scores.

Results

As in the previous experiments of this thesis the children's responses were coded according to their logical correctness. Each child scored 8 on the absurd statements and this data was not included in any further analysis. The mean scores and the standard deviations for the children, according to truth, term and task can be seen in Table C1.

Table C1

Mean number of logical responses (and standard deviations) according to truth, term and task

		Statements	Picture books
True/Felicitous	All	3.92	4.00
		(0.28)	(0.00)
	Some	3.54	3.69
		(0.78)	(0.86)
False/Infelicitous	All	3.85	3.92
		(0.38)	(0.28)
	Some	1.46	0.38
		(1.77)	(0.87)

Note: The maximum number of correct logical responses for each statement type was 4.

Statement task

In the statement task significantly fewer logical responses were given to the *some* trials than to the *all* trials, Wilcoxon matched pairs signed ranks test, $z = 2.97$, $p = .003$, $r = .58$, and to the false/infelicitous trials compared to the true/felicitous trials, Wilcoxon matched pairs signed ranks tes,t $z = 2.56$, $p = .009$, $r = .50$.

A Friedman test was used to analyse the number of logical responses given to the control and target trials, and this was significant, χ^2 (3) = 19.94, p > .001. Follow up Wilcoxon signed ranks tests revealed that the children gave significantly fewer logical responses to the *infelicitous some* trials compared to any of the control trials, all p's ≤ .008, all r's ≥ .52 (see Appendix CII, Table 3, for statistics showing the comparison of each control trial to the target trials).

Picture book task

The picture book task followed the same pattern of results. The children gave significantly fewer logical responses to the *some* trials compared to the *all* trials, Wilcoxon matched pairs signed ranks test z = 3.27, p = .001, r = .64, and to the false/infelicitous trials compared to the true/felicitous trials, Wilcoxon matched pairs signed ranks test, z = 3.17, T = 0, p = .002, r = .62.

The i*nfelicitous some* trials also received significantly fewer logical responses compared to any of the control trials, Friedman test, χ^2 (3) = 34.01, p > .001, follow up Wilcoxon signed ranks tests, all p's = .001, all r's ≥ .63 (see Appendix CII, Table 4, for statistics showing the comparison of each control trial to the target trials).

Infelicitous some trials

The nature of the distribution of the *infelicitous some* data for both tasks can be seen in Figure C2.

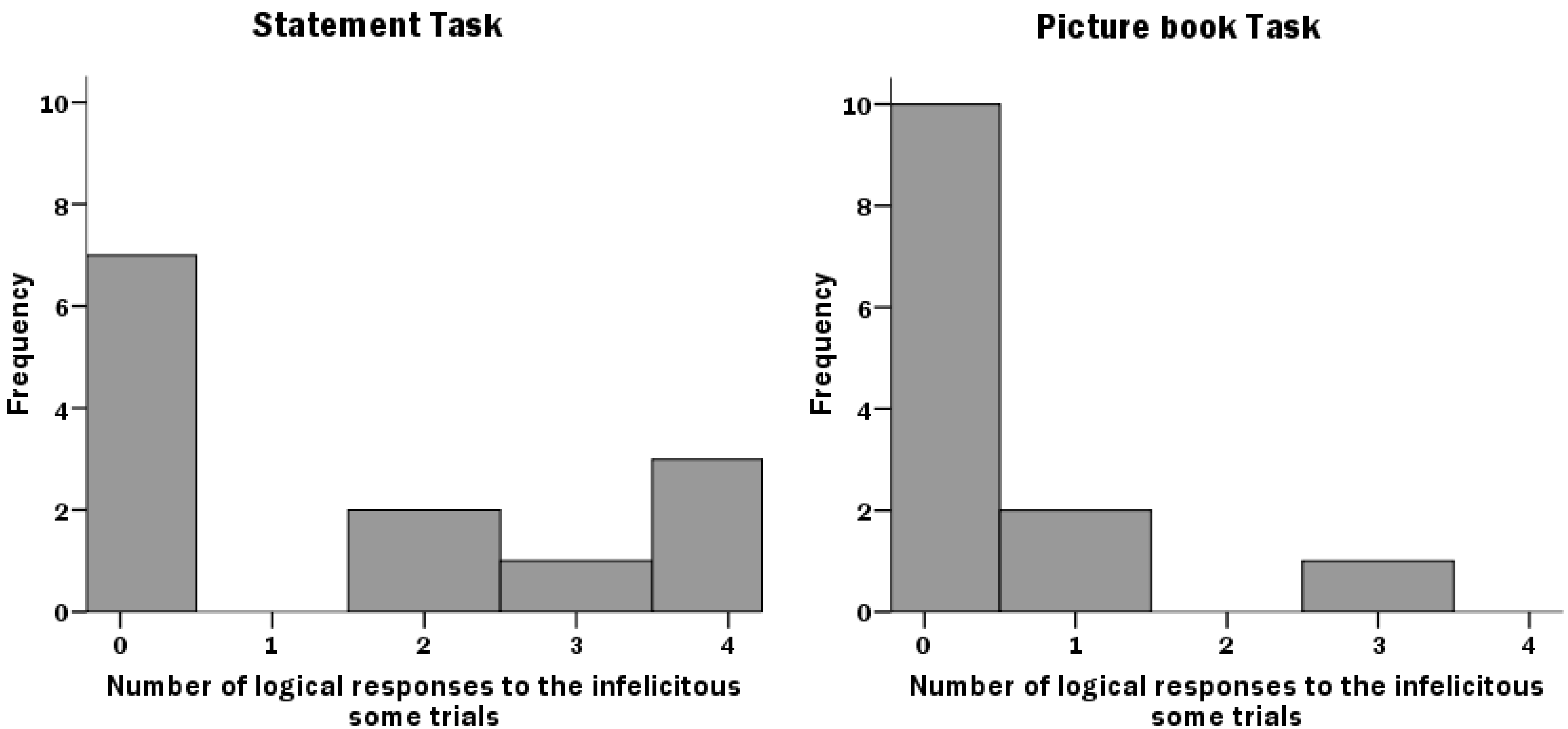


Figure C2. Distribution of participants according to the number of logical responses given to the infelicitous some trials for both tasks.

In the statement task, seven children never gave a logical response, three children always gave one, one child gave one most of the time, whilst two were equivocal. In the picture book task, ten children never gave a logical response, two children almost never did and one child gave a logical response the majority of the time. The nature of the pattern of responding between tasks was that six children gave more logical responses in the statement task, six gave the same number in both tasks, which was zero, whilst only one child gave more logical responses in the picture book task.

With regards to the number of logical responses given to the *infelicitous some* trials between tasks, although fewer logical responses were given in the picture book task this difference was not significant, Wilcoxon matched pairs signed ranks test, $z = 1.64$, $p = .101$, $r = .32$. The medium effect size suggests that the lack of significance may be due to the small number of participants, rather than there being no real effect. There was no significant difference in the total number of logical responses based on the order the children received the tasks, Mann Whitney test, $U = 20.00$, $p = .880$, $r = .04$, nor was there a practice effect, Wilcoxon matched pairs signed ranks test, $z = 0.17$, $p = .863$, $r = .03$.

The data from the *infelicitous some* trials was also analysed with regards to the language and IQ measures. Summary statistics for the measures can be seen in Table C2 (see Appendix CII, Table 5, for each participant's individual measures).

Table C2

Descriptive statistics for the language and IQ measures

	Mean (standard deviation)	Range
Chronological Age	161 (20.90)	119 - 195
Age Equivalent WASI	143.08 (21.66)	109 - 183
Age Equivalent BPVS	135.85 (29.75)	97 - 204

Note: Each measure is given in months

The scores for the *infelicitous some* trials were correlated with the participant's chronological age, their WASI equivalent age, and their BPVS equivalent age. Within each task there were no significant correlations, all p's ≥ .068, however there were a number of weak to moderate relationships (see Appendix CII, Table 6, for correlation statistics). Collapsing across the tasks revealed that there was a significant relationship between the total number of logical responses and chronological age, Kendall's $tau_{b,}$ τ = -.25, p = .269, there was a weak/moderate relationship between the number of logical responses and the BPVS equivalent age, Kendall's $tau_{b,}$ τ = -.44, p = .051, and a moderate relationship between the number of logical responses and WASI equivalent age, Kendall's $tau_{b,}$ τ = -.53, p = .019.

References

Dunn, L. M. & Dunn, L. M. (1981). *Peabody Picture Vocabulary Test—Revised.* Circle Pines, MN: American Guidance Service.

Dunn, L. M. & Dunn, L. M. with Williams, K. (1997). *Peabody Picture Vocabulary Test – Third Edition.* Circle Pines, MN: American Guidance Service.

Dunn, L. M., Dunn, L,M. & Whetton, C. (1981). *The British Picture Vocabulary Scale.* London: nfer-Nelson.

Dunn, L. M., Dunn, L,M., Whetton, C. & Burley, J. (1997). *The British Picture Vocabulary Scale (2nd Ed.).* London: nfer-Nelson.

Long, M. (2000). *The Psychology of Education*. London: RoutledgeFalmer.

Mackintosh, N.J. (1998). *IQ and Human Intelligence*. Oxford: Oxford University Press.

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

Tager-Flusberg, H. (1981). On the nature of linguistic functioning in early infantile autism. *Journal of Autism and Developmental Disorders,* 11, 1, 45-56.

Tager-Flusberg, H. (1999). The challenge of studying language development in autism. In L. Menn and N. Bernstein Ratner (Eds.), *Methods for Studying Language Production.* Mahwah, NJ: Lawrence Erlbaum Associates.

The Psychological Corporation. (1999). *Wechsler Abbreviated Scale of Intelligence*. San Antonio: The Psychological Corporation.

Wechsler, D. (1991). *Wechsler Intelligence Scale for Children – Third Edition*. San Antonio: The Psychological Corporation.

Wechsler, D. (1997). *Wechsler Adult Intelligence Scale – Third Edition*. San Antonio: The Psychological Corporation.

Appendix CII

Additional material for Appendix CI

Table 1

Statements used in autism study

Universal	Existential	Absurd
airplanes have wings	bicycles are blue	chairs laugh
books have pages	cars are red	clocks cry
rabbits have ears	clothes have zips	mice eat trains
elephants have trunks	birds live in cages	spoons wear hats
fish swim	flowers are pink	doors dance
snow is cold	paint is white	plates run
schools have teachers	animals have stripes	carrots wear shorts
giraffes have long necks	drinks are made of chocolate	policemen are made of jelly

Table 2

Themes used for picture books

Eating lollies	Spilling juice	Drinking pop	Popping balloons
Ripping magazines	Breaking tiles	Dirtying towels	Worms in containers
Putting cereal in slippers	Eating chocolate eggs	Breaking Lego houses	Eating chocolate bars
Pulling toys heads off	Knocking paint pots over	Spilling mugs of tea	Putting grass in sandwiches

Table 3

Wilcoxon signed ranks analysis of the number of logical responses given to the target trials compared to the control trials in the statement task

	True all/ Infel some	False all/ Infel some	True some/ Infel some
z	2.87	2.84	2.63
p	.004	.004	.008
r	.56	.56	.52

Table 4

Wilcoxon signed ranks analysis of the number of logical responses given to the target trials compared to the control trials in the picture book task

	True all/ Infel some	False all/ Infel some	True some/ Infel some
z	3.35	3.31	3.21
p	.001	.001	.001
r	.66	.65	.63

Table 5

Participant's individual age, IQ and language measures

Participant number	1	2	3	4	5	6	7	8	9	10	11	12	13
Age (in months)	119	138	151	155	156	156	157	158	161	176	186	188	195
WASI Age equivalent	138	109	127	152	138	120	160	145	116	143	162	167	183
BPVS Age equivalent	149	135	102	173	142	121	117	204	97	109	150	123	144

Table 6

Kendall's tau correlations for the number of logical responses given to the infelicitous some trials for each task and the combined total of each task, and the age, IQ and language measures

	Chronological age	WASI equivalent age	BPVS equivalent age
Statement task	-.094	-.423	-.327
	.685	.068	.157
Picture book task	-.363	-.242	-.360
	.131	.314	.132
Both Tasks	-.250	-.530*	-.439
	.269	.019	.051