The Question Under Discussion Focus Condition for Scalar Implicatures

Arjen Zondervan
0039071

MA Thesis
Linguistics: The Study of the Language Faculty
UiL-OTS, Universiteit Utrecht
Supervisors: dr. I.C.M.C. Mulders, dr. K. Szendrői

August 2006

TABLE OF CONTENTS

CHAPTER 0: INTRODUCTION

One of the most fascinating properties of the human language system or perhaps of human cognition in general, is that language users are able to deduce extra meaning from utterances, that cannot be traced back to the meaning of the words that made up the utterance. Without this ability, human communication like we engage in every day would be very difficult, if not impossible. This extra meaning is deduced in many different ways, as participants in a conversation rely on principles of cooperativity, relevance, mind-reading and many more. This thesis considers one particular kind of the inferences that make up this extra meaning: scalar implicatures. This kind of inference is of special interest to linguistics as it is one of the most clean types of extra meaning: In contrast to other inferences that rely on an interplay of a number of different factors that are hard to formalize, scalar implicatures are based on principles of informational strength, which can be formalized by the logical property of entailment. Because of this, they form an interesting window into the magical world of extra meaning, that is called *pragmatics* in linguistics.

Scalar implicatures are based on the idea that certain words are in a scale with each other. If a lower term of such a scale is used, it is often felt to imply that the higher term does not hold. This inference is called a *scalar implicature*. Well-known examples of these scalar implicatures are the inference from *some* to *not all* and from *or* to *not and*. I will focus on the last one in this work. In Chapter 1 I will explain the reasoning behind scalar implicatures, and place it the context from which it originated: the work of the philosopher Paul Grice.

In the last couple of years, the number of experimental studies on scalar implicatures has increased drastically. Especially researchers working on first language acquisition have been investigating this topic. Their interest is triggered by the well-known observation that young children seem to have difficulties with drawing scalar implicatures. Besides the acquisition question, there are two main points of debate in the present literature on scalar implicatures: The first is the question whether scalar implicatures are calculated after the semantics has assigned a meaning to the whole sentence (so *globally*), or whether scalar implicature calculation occurs during compositional semantics itself, while the meaning of the sentence is being built up from its parts (so *locally*). Researchers that follow Grice's classical approach claim the first is the case, but recently Gennaro Chierchia argued for the second option, convincing many researchers of the need for a local view. The second main point of debate is whether scalar implicatures are default inferences that arise irrespective of the context they appear in, or whether they only arise in certain contexts. Views that belong to either side are respectively called *defaultist* theories and *contextualist* theories. Sections 2.1 and 2.2 of this work will be dedicated to respectively the global vs. local debate and the defaultist vs. contextualist debate, the two main current debates in the theory of scalar implicatures.

The main contribution of this work is the introduction of a condition that up till now has not been taken into account in experiments on scalar implicatures. I claim this condition is decisive for scalar implicature calculation. The predictions of this condition are shared by three contextualist approaches that are discussed in Section 2.2: Relevance Theory of Sperber & Wilson (2002), the exhaustivity-approach of Van Rooy (2002) and the topic/comment account of Van Kuppevelt (1996). All three theories have their own way of explaining these predictions, but I will reduce these three approaches to one condition that relies on the Question Under Discussion (QUD) of the context the scalar term appears in. This condition, labeled the *QUD Focus Condition*, states that only if the scalar term appears in the part of the

answer that was questioned by the QUD of the context, a scalar implicature will arise. This part of the answer is signaled by focus. This condition, the main hypothesis of this work, and its predictions for the scalar implicature of *or*, are introduced in Section 2.3.

As I will show, the QUD Focus Condition can account for data from previous experiments on scalar implicatures that up till now seemed contradictory. In Section 3.1, a number of these previous experiments, most of which use the Truth Value Judgment Task (TVJT), are reconsidered. This work also presents an experiment that tested the predictions of the QUD Focus Condition. To be able to compare the data to previous studies, a TVJT was used to assess scalar implicatures. However, I varied the QUD (and corresponding focus structure) that was triggered by the context in such a way that in one condition the scalar appeared in the part of the target sentence that was questioned by the QUD, while in the other condition it did not. This way it is possible to isolate the workings of the QUD Focus Condition. The experiment is presented in Section 3.3. The thesis concludes with a critical discussion about the suitability of a TVJT to assess scalar implicatures, and suggestions for further research.

CHAPTER 1: BACKGROUND

1.1 What are implicatures?

The term *implicature* dates back to Paul Grice's "William James Lectures" delivered at Harvard in 1967. An implicature is a part of the meaning of a sentence that is not given by compositional semantics: it is not built up from the meaning of its parts. Consider for example the exchange in (1), taken from Levinson (1982):

(1) A: Can you tell me the time?
B: Well, the milkman has come.

B's answer in (1) actually means something like (2):

(2) B: *No I don't know the exact time of the present moment, but I can provide some information from which you may be able to deduce the approximate time, namely* the milkman has come.

However, there is nothing in the meaning of the parts of B's utterance: *Well, the, milkman, has* and *come*, from which the italicized part of (2) can be derived. Something similar is going on in (3), also from Levinson (1982):

(3) The flag is white.

Sentence (3) seems to mean: *The flag is entirely white*. However, there is nothing in the meaning of *The, flag, is* and *white*, from which it can be derived that the flag is *entirely* white. This is shown by a variant of (3):

(4) The flag is white, red and blue.

In (4) the same words are used (along with three more), but the meaning that the flag is entirely white does not arise. So if something in the compositional semantics of (3) yielded the *entirely white* meaning, then why does this meaning not arise in (4), and lead to a contradiction of some sort because of what follows? These examples show that there is a part of the meaning of a sentence that cannot be ascribed to compositional semantics, but comes about in another way. One of the ways in which it can come about is by *implicature*[1].

At the basis of all implicatures lies Grice's Cooperative Principle:

(5) The Cooperative Principle: "Make your contribution such as required, at the stage at which it occurs, by the accepted purpose or direction of the talk exchange in which you are engaged."

According to Grice, in normal conversation every speaker adheres to this principle. Grice made the idea of the Cooperative Principle explicit by listing four *maxims of conversation*, guidelines every speaker follows in normal conversation, listed here in (6)-(9):

[1] This term should not be confused with *implication*, which is a more general term for all meaning that is not given by compositional semantics. This however also includes other pragmatic inferences like presupposition.

(6) The maxim of Quality:
Try to make your contribution one that is true, specifically:
(i) Do not say what you believe to be false.
(ii) Do not say that for which you lack adequate evidence.

(7) The 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.

(8) The maxim of Relevance:
Make your contribution relevant.

(9) The maxim of Manner:
(i) Avoid obscurity.
(ii) Avoid ambiguity.
(iii) Be brief.
(iv) Be orderly.

Implicatures come about because a hearer knows a speaker is following the maxims. To see how this works, reconsider the exchange in (1). Because A knows that B is following the maxim of Relevance, A knows the meaning of B's utterance cannot just be about the milkman having come, but that this is somehow relevant to the question. This is how the part: *I can provide some information from which you may be able to deduce the approximate time* of (2) comes about. Also, because A knows B is following the maxim of Quantity (i) (which can be paraphrased as: *Be as informative as possible for the current purposes*), A knows B does not know the exact time, otherwise he would have told him. So the other part of (2), *No I don't know the exact time of the present moment*, is deduced by the hearer is this way. The extra meaning of (3) can be deduced from the maxims in a similar way. The hearer knows the speaker is following the maxim of Quantity (i), so he will assume the speaker is being as informative as possible. Upon hearing (3), the hearer will therefore conclude that the speaker knows the flag does not have any other colors, otherwise he would have said so. If the flag is in fact white, red and blue and the speaker knows this, then according to the maxim of Quantity (i) he would utter (4) and not (3), as in that case (4) is more informative than (3). So the extra meaning of (1) and (3) are in fact implicatures, that can be calculated with Grice's maxims of conversation.

1.2 What are scales?

Some linguistic expressions (often words) together form entailment scales. These are always expressions of the same category (verb, adverb etc.). They form an entailment scale because a sentence that contains one item, asymmetrically entails the same sentence with the other item.[2] Consider for example (10) and (11):

(10) Bruce Springsteen's concert was excellent.
(11) Bruce Springsteen's concert was good.

Sentence (10), with the adjective *excellent*, entails (11), which is the same sentence but with *excellent* replaced by *good*: if the concert was excellent, it also has to have been good. This entailment does not work the other way around though, as the concert might have been good,

[2] Sentence A entails sentence B iff in every situation in which A is true, B is also true. Sentence A *asymmetrically* entails sentence B iff sentence A entails sentence B and sentence B does not entail sentence A.

but not good enough to be called *excellent*. So *excellent* and *good* form an entailment scale. The expression that entails the other is considered to be informationally stronger, as it is compatible with less situations. Consider for example the possible situations in (12):

(12) Situation 1: The concert was excellent (and therefore also good).
Situation 2: The concert was good (and not excellent).
Situation 3: The concert was neither excellent nor good.

Sentence (10), containing *excellent*, is only true in Situation 1. Sentence (11) however, containing *good*, is compatible with Situation 1 and 2. So the expression *excellent* rules out more possible situations, and is therefore considered to be more informative. To catch this intuition, one might think of the difference in informativity between a weather forecast that says: *It might rain and it might be sunny* and one that says: *It is definitely going to rain*. The former is compatible with a lot more situations than the latter and therefore more likely to be right, but because the latter rules out more situations, it is way more informative.

Entailment scales are, probably because they are an ordered set, often represented in between '<' and '>'. Whether the stronger item is put on the left or the right of the weaker item is trivial and it differs in the literature, I will put the stronger item on the right. Entailment scales can consist of more that two members (and often do), as a sentence can asymmetrically entail another sentence that in its turn can entail another sentence and so on.[3] Entailment scales can consist of many types of expressions (adjectives, verbs, quantifiers, connectives, adverbs and more), but always one type per scale. Some other examples of entailment scales are given in (13) (taken from Horn, 1972):

(13) <warm, hot>
<sometimes, often, always>
<want to *V*, try to *V*, succeed in *V*-ing>
<possibly *p*, *p*, necessarily *p*>
<possible that *p*, probable that *p*, certain that *p*>
<may, should, must>
<cool, cold>
<like, love>

1.3 What are scalar implicatures?

Scalar implicatures (SIs) are based on the maxim of Quantity (i) (*Be as informative as possible for the current purposes*). We just saw that the items in an entailment scale differ in informativity: stronger items are more informative than their weaker counterparts. So based on the maxim of Quantity, you expect speakers to always use the strongest item on the entailment scale that is compatible with his or her knowledge. This means however, that if a speaker uses a weaker item from an entailment scale, a hearer can conclude that the same sentence with a stronger item is not compatible with the speaker's knowledge. In a situation in which it is obvious that the speaker has complete knowledge of the actual situation, the hearer can even conclude that the sentence with the stronger item is untrue. Reconsider (10) and (11):

[3] And because entailment is transitive, sentences containing items that are more than one place higher on the scale (than the weaker item) all entail the sentence containing the weaker item.

(10) Bruce Springsteen's concert was excellent.
(11) Bruce Springsteen's concert was good.

Let us assume the speaker actually went to the concert, and let us ignore the subjectivity of the judgment about the concert. Then based on the reasoning above, if this speaker utters (11), the hearer will conclude that the stronger (10) is untrue. So (11) gets the extra meaning: *it is not the case that Bruce Springsteen's concert was excellent*. This is a scalar implicature of sentence (11). The complete meaning of the sentence now becomes: *Bruce Springsteen's concert was good and it is not the case that Bruce Springsteen's concert was excellent.* So by using an item from an entailment scale, a speaker automatically labels the same sentence with a stronger item from the same scale as being untrue. Chierchia (2004) nicely summarizes the reasoning behind scalar implicatures in the following way:

i. The speaker said (11) and not (10), which would have also been relevant.
ii. (10) entails (11) [*good* and *excellent* are part of a scale].
iii. If the speaker had the info that (10), she/he would have said so [Quantity].
iv. The speaker has no evidence that (10) holds.
v. The speaker is well informed.
Therefore
vi. It is inlikely/not the case that (10) holds.

A good example of how much this mechanism is applied in everyday language is the entailment scale <some, all>. In everyday language, *some* is often felt to be ambiguous between *some and possibly all* and *some and not all*. This can be shown by the fact that people have difficulties in determining if for instance (14) is true or false:

(14) Some elephants have trunks.

Often it is felt that strictly speaking this sentence is true, but it is odd. This can be explained if we consider the scalar implicature of a sentence like (14). But in order to do so we first need to establish the basic meaning of *some*. In the linguistic literature, the basic meaning of *some* is considered to be *some and possibly all*. One of the reasons to assume this is that there are contexts in which the other possible meaning, *some and not all*, is definitely ruled out. Consider for example (15):

(15) It is not the case that some students went to the party.

If we interpret *some* in (15) as *some and not all*, we get (16):

(16) It is not the case that some and not all students went to the party.

Sentence (16) would be true in a situation in which no students went to the party, but also in a situation in which all students went to the party. But that is not what (15) intuitively means: it only means that no students went to the party. The sentence should be false if all students went to the party. This meaning we can only get if we interpret *some* as *some and possibly all*, yielding (17):

(17) It is not the case that some and possibly all students went to the party.

Now that we have established the basic meaning of *some* as *some and possibly all*, consider (18) and (19):

(18) All students went to the party.
(19) Some (and possibly all) students went to the party.

Obviously, (18) asymmetrically entails (19). This means *some (and possibly all)* and *all* form an entailment scale, just like *excellent* and *good* above, with *all* being the stronger item and *some* the weaker. So if a speaker uses *some*, the hearer can conclude, on the basis that the speaker is being as informative as possible, that the same sentence with *all* is untrue. The scalar implicature of e.g. *Some men are liars* therefore is: *It is not the case that all men are liars*. This is exactly how the other meaning of *some*, namely *some and not all*, comes about. It is a combination of the basic meaning of *some*: *some and possibly all* and the scalar implicature it usually gets: *it is not the case that all*. Now let us reconsider the oddity of (14). The scalar implicature of (14) is: *it is not the case that all elephants have trunks*. This is obviously false and that is why people have difficulties with sentences like these.

There is more evidence that the *and not all* part is not part of the semantics of *some*, but comes about by an implicature. One of the characteristic properties of implicatures is that they are defeasable (or cancelable), by adding extra material to a sentence or a discourse. Consider (20):

(20) Bruce Springsteen's concert was good. In fact, it was excellent.
(21) # The color of the book was black. In fact, it was red.

Above we saw that the implicature of the first sentence of (20) is: *it is not the case that Bruce Springsteen's concert was excellent*. However, by adding the extra sentence, this implicature is canceled. But notice that the two sentences are still perfectly compatible with each other, they do not yield a contradiction. In (21) however, the two sentences do form a contradiction and therefore an unfelicitous sequence (signalled by the symbol '#'). The difference is that in (21), the meaning of the sentence given by compositional semantics is contradicted, while in (20) only the extra meaning of the implicature is contradicted. So an important difference between semantic meaning and implicatures is that only the latter can be canceled by adding extra material. Now let us return to *some*. Consider (22) and (23):

(22) Some men are liars. In fact, all men are.
(23) # Some men are liars. In fact, no men are.

As (22) shows, *not all men are liars* has to be an implicature of the sentence and not part of the semantics of *some*, as it can be canceled without problems. This is contrasted with (23), where it is shown that the meaning of *some* that can be paraphrased as *at least some*, is part of the semantics of *some*, as it cannot be canceled. This is clear evidence in favor of *some and possibly all* as the basic meaning of *some*, where *and not all* is a scalar implicature.

1.4 The scalar implicature associated with sentences containing *or*

Another expression whose apparent ambiguity is caused by a scalar implicature is the connective *or*. In normal conversation, *or* often seems to mean *either-or*:

(24) That day he went to the store and bought a newspaper or a magazine.

Sentence (24) seems to mean that he bought one of the two, but not both. This is often referred to as the *exclusive-or* interpretation. However, there are contexts in which *or* obviously has to be interpreted inclusively, so in which it means: *one of the two or both*:

(25) If the King or the Queen comes to my party, I will be very pleased.
(26) Could everyone who has a question or a comment wait till after the talk?

Obviously, if after uttering (25) both the King and the Queen show up at my party, I am still going to be very pleased. Similarly, if I am attending this talk in (26) and I have both a question and a comment, I will not blurt them out right away, but I will wait till the end of the talk, just like everyone who has either a question or a comment. So in these contexts, *or* obviously should be interpreted as *one of the two or both*, which is also sometimes referred to as *and/or*. Contexts like these have led linguists to follow logicians in their basic meaning of the connective *or* as inclusive-*or*. But then why do normal situations like (24) not reflect this meaning? Just like in the case of *some* and *all*, there is a scalar implicature at work. If we take inclusive-*or* as the meaning of *or* we can see how this works. Consider (27) and (28) and the possible situations in (29):

(27) The dog ate the hamster and the goldfish.
(28) The dog ate the hamster or the goldfish.

(29) Situation 1: The dog ate both the hamster and the goldfish.
Situation 2: The dog ate only the hamster.
Situation 3: The dog ate only the goldfish.
Situation 4: The dog ate neither the hamster nor the goldfish.

Sentence (27), with *and*, is only compatible with Situation 1, in all other situations it is untrue. Sentence (28), with inclusive-*or*, is compatible with Situation 1, 2 and 3 and is only untrue in Situation 4. So in each situation in which (27) is true (only Situation 1), (28) is also true. So *and* and *or* form an entailment scale, with *and* as the stronger item and *or* the weaker. Now we can apply the Gricean reasoning we have seen a couple of times already above. If a speaker utters a sentence with *or* (like (28)), the hearer can conclude the sentence with *and* is untrue, because otherwise the speaker would have chosen this more informative sentence. So the scalar implicature of *A or B* is: *it is not the case that A and B*. This gives us the exclusive-*or* interpretation we get so often in everyday language. It is built up from the basic inclusive meaning of *or* and the *not and* implicature.

Just like we did for *some*, we can check the components of the meaning of *or*, by checking for cancelability.

(30) The dog ate the hamster or the goldfish. In fact, it ate both.
(31) # The dog ate the hamster or the goldfish. In fact, it ate neither.

Sentence (30) shows the *and not both*-component of the meaning of *or* can be canceled and therefore it cannot be a part of the semantics, it is an implicature. By contrast, the *at least one of the two*-component cannot be canceled, as (31) shows, so that is part of the semantics of *or*.

We have seen the exclusive reading of *or* is actually a result of a scalar implicature. There is however one important assumption at the basis of all scalar implicatures: the assumption that the speaker has complete knowledge of the situation. Obviously, *or* is also often used when a speaker does not have complete knowledge. Consider for instance the following situation: After a party, there are cigarette buts all over the living room floor, although there were ashtrays everywhere. The host of the party knows that only two people, Bjorn and Benny, have been smoking at the party. When asked who was so rude to put out their cigarettes on the floor instead of using the ashtrays, the host might say (32):

(32) Bjorn or Benny must have put out his/their cigarettes on the floor.

As the choice of pronoun already indicates, the interpretation of *or* is definitely inclusive here. This is because the Gricean reasoning that the speaker used *or* because he knows *and* is untrue, cannot take place here. Obviously the speaker has no complete knowledge, which is also signaled by the modal construction with *must have*. This is called an *epistemic limitation*, which blocks scalar implicatures. In this work, these situations will not be considered. Some authors also claim that in calculating a scalar implicature, a hearer cannot conclude that the speaker knows the stronger item is untrue, but only that the speaker is uncertain whether the stronger item is true. This would lead to much weaker implicatures, of the form: *it is uncertain if* instead of: *it is not the case that*. For now I will only consider situations in which the speaker has complete knowledge of the situation, and this is known to the hearer. In these situations we can draw the strong implicatures described above.

CHAPTER 2: MAIN POINTS OF DISCUSSION ON SCALAR IMPLICATURES

In this chapter, I will consider the two main points of discussion on SIs of the last couple of years. The first debate is about at which point in sentence processing the SI-calculation takes place. Because a lot of recent work on SIs is based on this question, I will discuss it here. This work will not make a contribution to this particular discussion though. However, some of the experimental work that I will discuss later is driven by this discussion. The second main point of discussion on SIs is the question whether SIs depend on the context. This question will be the starting point of the main contribution of this work.

2.1 Global vs. local

The theory on SIs as outlined in Chapter 1 is often referred to as the "Gricean" view, and its followers are often called "neo-Griceans". See e.g. Horn (1972), Gazdar (1979) and Atlas & Levinson (1981). The neo-Gricean view has been relatively unchallenged for at least the last two decades of the 20th century. However, in 2000 Gennaro Chierchia wrote an article (published in 2004) in which he proposes a radically different view on the calculation of SIs[4]. What triggered the dispute between Chierchia and the neo-Griceans is the question what happens if scalar terms appear in embedded positions. Based on this kind of sentences Chierchia proposes a theory of how SIs are computed *locally*, instead of the global neo-Gricean mechanism. This chapter illustrates two embedding contexts that are at the heart of the discussion: Embedding of a scalar under a logical operator (or another scalar) and embedding of a scalar in a downward entailing (DE) context.

2.1.1 Embedding under a logical operator or another scalar

On the neo-Gricean view, SIs are calculated after the compositional semantics of the whole sentence is done. So the semantics provides the input value for SIs and other pragmatic inferences. The neo-Gricean mechanism for calculating SIs is to negate the sentence with a stronger scalar term and to add this to the meaning of the sentence. Consider for example how this works for sentence (1). I will use the symbol '~>' for the SI of the sentence.

(1) a. Bruce Springsteen's concert was good.
b. ~> It is not the case that Bruce Springsteen's concert was excellent.
c. Bruce Springsteen's concert was good, but it is not the case that it was excellent.

The SI of (1a) is (1b). The resulting complete meaning of (1a) therefore is (1c). Now the only possible scope the negation (*it is not the case that*) of the SI can take is over the whole sentence (widest scope), because the semantics of the whole sentence is already closed off once the SI is calculated. To put it in other words: the SI cannot look inside the semantics of the sentence, so it can only scope over its result. It can only be calculated *globally*: over the whole sentence. For examples like (1), where the scalar term is not embedded under a logical operator or another scalar, this mechanism works fine. However, it runs into problems for sentences like (2), taken from Chierchia (2000):

[4] Although it was first published in 2004, I will refer to it as Chierchia (2000) to keep the timeline correct, as many works before 2004 take his article into account.

(2) Mary is either working at her paper or seeing some of her students.

Here we can ignore the possible SI of *or*, as the exclusive-*or* meaning is explicitly stated by using *either... or*. What is interesting is the potential SI of *some*. Intuitively, the complete meaning of (2) is (3):

(3) Mary is either working at her paper or seeing some, though not all, of her students.

But let us see what happens when we apply the neo-Gricean mechanism to (2). Replacing *some* with the stronger *all* and negating the whole sentence gives us the SI in (4):

(4) ~> It is not the case that [Mary is either working on her paper or seeing all of her students].

But this SI not only negates the second disjunct, but also the first. This is wrong as (5) is obviously not an implicature of (2):

(5) Mary is not working at her paper.

It seems that as a result of the assumption that the SI can only take widest scope, the negation ends up at the wrong place: instead of taking scope over the whole sentence, it should have been embedded in the second disjunct, as in (3). So here the neo-Gricean global mechanism results in an SI that is too strong. But there are also examples where the global mechanism predicts SIs that are too weak. Consider for example (6), taken from Sauerland (2004):

(6) Kai had the broccoli or some of the peas last night.

Here, the scalar term *some* is embedded under another scalar term: *or*.[5] Intuitively, this sentence has two SIs, according to the two scalar terms, given in (7):

(7) a. ~> Kai did not have the broccoli and some of the peas last night.
b. ~> Kai did not have all of the peas last night.

It is unclear however how to arrive at the SI in (7b) on the neo-Gricean global view. According to Gazdar (1979), no SIs arise if a scalar term is embedded under another operator. He would therefore only predict the SI in (8), which is equivalent to (7a):

(8) It is not the case that [Kai had the broccoli and some of the peas last night].

However, if we add this implicature to the meaning of (6), it is still true in the situation in which Kai had no broccoli but all of the peas. As Sauerland points out, if you were to utter (6) and I know for a fact that Kai had all of the peas, my answer would be (9a), and not (9b):

(9a) No, he had all of the peas last night.
(9b) # Yes, he had all of the peas last night.

This indicates that the complete meaning of (6) does indeed negate that he had all of the peas, so the SI in (7b) is calculated. On Gazdar's mechanism however, there is no way to get it.

[5] Notice the difference with (2), where the possible SI of *or* was not relevant, as *either or* was used.

Alternatively, we might try to replace both scalars by their stronger counterparts and then negate the whole sentence. This would yield (10):

(10) It is not the case that [Kai had the broccoli and all of the peas last night].

But if we add (10) to the meaning of (6), it is still true in the situation in which Kai had no broccoli but all of the peas. So it seems that here, the global mechanism calculates SIs that are too weak.

Chierchia (2000) provides four more embedding environments in which the global mechanism seems to yield the wrong predictions: *believe*-type verbs, factive verbs, quantifiers and nonmonotone quantifiers. According to Chierchia, these data indicate that the global mechanism should be abandoned and replaced by a local mechanism for SI calculation, which he calls the *Semantic Core Model*.

2.1.2 The Semantic Core Model

Chierchia's Semantic Core Model states that 'implicatures are introduced locally as soon as possible in the same order in which their triggers (the scalar terms) are introduced in the syntactic tree' (Chierchia (2000)). SIs are introduced bottom-up right after the scalar term enters the compositional semantics[6]. This means that SIs are introduced not after, but during compositional semantics. This is an important claim as it rejects the traditional modular view that pragmatic inferences cannot interfere with compositional semantics. According to Chierchia, SIs are part of the semantic core, hence the name of the model. So how does this model work exactly? Each scalar term is compared locally (at each scope site of type t) to its scalar alternatives. Based on this comparison, an SI is calculated locally, negating the weakest alternative that asymmetrically entails the plain meaning. Chierchia calls this a *direct* implicature. Now, the stronger meaning of the constituent, i.e. the meaning including the SI, is combined with constituents higher in the tree.

Let us see how this works for the examples above. According to the Semantic Core Model, in (2) the SI of *some* is introduced in the second disjunct, before the derivation arrives at *or*. So the SI is already introduced at: t_{Mary} *is seeing some of her students*, resulting in: t_{Mary} *is seeing some, though not all, of her students*, which is then combined with the rest of the sentence, yielding (3) as the correct complete meaning of (2). We no longer get the wrong entailment (5). The same goes for (6), where the SI that Kai did not have all of the peas is already introduced in the second disjunct: $t_{Kai\ had}$ *some, though not all, of the peas*, before the other scalar *or* is encountered.[7] So the SI (7b) we were unable to arrive at on the global view, is calculated directly in the second disjunct on the Semantic Core Model.

[6] Chierchia is not very clear as to where exactly the SI is introduced. He says: "Let's assume that this is done at any scope site (of type t)..."

[7] So Chierchia predicts that *or* combines with the stronger meaning of the second disjunct. The meaning before the SI of *or* is taken into account will thus be: (i) *Kai had the broccoli or some, though not all, of the peas last night*. The SI of *or* will be: (ii) *it is not the case that [Kai had the broccoli and some, though not all, of the peas]*. Although this SI is compatible with a situation in which Kai had both the broccoli and all of the peas, this situation is already ruled out by (i). The complete meaning is exactly what we want: (iii) *Kai had the broccoli or some, though not all, of the peas and it is not the case that Kai had the broccoli and some of the peas.*

Sauerland (2004) proposes a neo-Gricean solution for scalars embedded under another scalar: Take all possible combinations of the scalars and their one-step stronger alternatives. This procedure is similar to taking the Cartesian product of two sets. The negation of each of these possible combinations is an SI of the sentence, as long as they asymmetrically entail the plain meaning. Consider for example (11), and its three SIs based on the <most, all> and <good, excellent> scales. [8]

(11) Most players were good.
i. ~> It is not the case that [all players were good.]
ii. ~> It is not the case that [most players were excellent.]
iii. ~> It is not the case that [all players were excellent.]

This mechanism gives the right results for this sentence. Notice the difference with our attempt in (10), where both scalars were replaced by their stronger counterparts and the whole sentence was negated. That would result in only the SI (iii) here, which would still be compatible with a situation in which all players were good or a situation in which most players were excellent. So Sauerland's mechanism is definitely an improvement over that. However, as Sauerland himself notes, this mechanism still does not give the right predictions for sentences like (6). To solve this problem, he introduces a new scale for disjunction, which I will not go into here.

2.1.3 Embedding in a downward entailing context

Another question that has been claimed to differentiate between the neo-Gricean view and the Semantic Core Model is how scalar terms behave embedded under negation or other downward entailing (DE) contexts. Consider (12), taken from Gualmini (2001), and the possible situations in (13):

(12) John did not write a paper or make a presentation.

(13) Situation 1: John wrote both a paper and made a presentation.
Situation 2: John only wrote a paper.
Situation 3: John only made a presentation.
Situation 4: John neither wrote a paper nor made a presentation.

In (12), three of the four possible situations *are negated* (Situation 1-3). So also the situation in which John did both (Situation 1). This shows that under negation, *or* is interpreted as inclusive-*or*. The *not and*-SI in (14), which is equivalent to (15), does not arise:

(14) It is not the case that [John did not write a paper and make a presentation].
(15) John wrote a paper and made a presentation.

This blocking of SIs is not limited to negation, but is observed in all DE-contexts. Consider for example (16a-c), also based on Gualmini (2001):

(16) a. None of the students wrote a paper or made a presentation.
b. Everyone who wrote a paper or made a presentation received a good grade.

[8] Notice that the SIs in (i) and (ii) entail the SI in (iii). Sauerland puts SIs that are entailed by others between parentheses.

c. John fulfilled the requirements of the Syntax seminar without writing a paper or making a presentation.

In (16a-c), *or* is interpreted inclusively. That these are all DE-contexts can easily be shown by their ability to license the NPI *any*:[9]

(17) a. None of the students wrote any paper.
b. Everyone who wrote any paper received a good grade.
c. John fulfilled the requirements of the Syntax seminar without writing any paper.

That it is not just negation that makes the difference can further be illustrated by minimal pairs such as (18a-b):

(18) a. John graduated before he wrote a paper or made a presentation.
b. John graduated after he wrote a paper or made a presentation.

Only (18b) seems to get the *not and* SI, as *before* in (18a) creates a DE-context. But DE-contexts do not only block implicatures, they trigger new ones. Consider (19) and (20):

(19) Michael Jackson didn't insult all of the judges.
(20) Nobody went to the barbeque and the party.

Sentence (19) seems to have as an implicature that Jackson did insult *some* of the judges. And (20) seems to communicate that some people did go to the barbeque *or* the party. Some authors, e.g. Gualmini (2001) claim that the neo-Gricean view cannot account for these new implicatures. Obviously, the account of Gazdar (1979) cannot account for it, as he claims scalars that are embedded under other operators (like negation) do not get any implicatures. However, other neo-Griceans like Atlas & Levinson (1981) and Horn (1989) noticed that negation simply reverses scales. Consider how this works for the <or, and> scale and the <some, all> scale: While *A and B* entails *A or B*, *not (A or B)* entails *not (A and B)*. Similarly, while *all A VP* entails *some A VP*, *not (some A VP)* entails *not (all A VP)*. So under negation or other DE-operators, the relevant scales are flipped and become: <and, or> and <all, some>. Knowing this, the neo-Gricean reasoning still holds: SIs are based on the maxim of Quantity, which is an informativity constraint. So in contexts in which *or* is more informative than *and* and *some* is more informative than *all* (DE-contexts), using a weaker term indicates the negation of the stronger term, just like in non-DE contexts. So the neo-Gricean view can perfectly explain the rise of new SIs in DE-contexts: the fact that the entailments are reversed will immediately cause the SIs to go the other way. The blocking of SIs we saw in (18) is also a direct result of the scales being reversed: in a DE-context, *or* is the strongest item on the scale and that item typically does not get an SI, as there is no stronger item to be negated.

The Semantic Core Model has a different way of dealing with DE-contexts, although the rationale behind it, the informativity constraint, is the same. Let us recap how implicatures were calculated in the first place in this model: Above I said the following:

> Each scalar term is compared locally (at each scope site of type t) to its scalar alternatives. Based on this comparison, an SI is calculated locally, negating the weakest alternative that

[9] Chierchia (2000) notes that (although) 'since the seminal work of Ladusaw (1979), DE-ness has been identified as a key property in the licensing of NPIs ..., it is controversial whether DE-ness suffices in characterizing NPIs..., it is thus useful to keep *any* licensing and DE-ness separate.' For simplicity, I will not do this.

asymmetrically entails the plain meaning. Chierchia calls this a *direct* implicature. Now, the stronger meaning of the constituent, so the meaning including the SI, is taken and combines with constituents higher in the tree.

But what happens if a higher constituent contains a DE-operator? Chierchia states constituents are combined by functional application (as most Montague-style compositional semantics theories assume). Every time this happens, the *Strength Condition* has to be satisfied: The strong value (meaning including the locally introduced SIs), cannot become weaker than the plain value (meaning without SIs). This is simply a version of the informativity constraint or the maxim of Quantity we saw above on the neo-Gricean view. So when a DE-operator is encountered, the strong meaning is no longer more informative than the plain meaning, so the strong meaning is abandoned and the plain meaning is taken. However, in order to also account for the new implicatures that arise in DE-contexts, Chierchia's system contains a comparison of the result of functional application with the plain meaning to the result of functional application with its embedded scalar alternatives. If functional application with an embedded scalar alternative is more informative, this meaning is negated and added to the plain meaning as an SI. Chierchia calls this an *indirect* SI. The mechanism can be nicely illustrated using a scale that has (at least) three items, like <sometimes, often, always>. Consider (21):

(21) It is not the case that John has often been to class.

Working bottom-up, we first encounter the scalar *often*. At the first scope site of type t, the embedded IP *John has often been to class*, the direct SI is introduced, resulting in the complete meaning of the IP: *John has often, but not always, been to class*. However, the constituent *it is not the case (that)* obviously is DE, so when applying functional application of the strong meaning of the IP and this constituent, the Strength Condition is violated: *It is not the case that John has often and possibly always been to class* (the plain meaning) rules out more possible situations than *It is not the case that John often and not always has been to class* (the strong meaning). So the plain meaning is taken. Additionally, the result of functional application with the scalar alternatives of the embedded *often* are considered. Now it turns out that functional application with another scalar alternative of *often*, *sometimes*, leads to a more informative statement than with *often* itself: *It is not the case that John has sometimes been to class* rules out more situations than *It is not the case that John has often been to class*. So we add the negation of this stronger alternative as an SI, given in (22), which is equivalent to (23):

(22) ~> It is not the case that [it is not the case that John has sometimes been to class].
(23) John has sometimes been to class.

This way the Semantic Core Model lands on the SI in (23), which is what we want. However, we saw above that the neo-Gricean mechanism also had no problems in predicting these SIs. The informativity constraint (from the maxim of Quantity) and the Strength Condition seem to add up to the same thing. The only difference seems to be that while the neo-Griceans rely on entailment to flip the scales, the Semantic Core Model has a built-in mechanism that itself compares plain meaning to all scalar alternatives. So while in the above example the neo-Griceans would rely on the flipped scale to predict the SI *not(not sometimes)*, the Semantic Core Model compares the plain meaning to all alternatives (so also the meaning with for instance *always*), to arrive at the same SI.

Another difference between Chierchia and at least one neo-Gricean, Horn (2005), is the status of the new SIs that arise in DE-contexts. As Chierchia already indicates by using the label 'indirect' for these implicatures, he claims they are weaker than the direct ones in non-DE contexts: 'Although there is a systematicity in how implicatures get associated with negative sentences, such implicatures appear to be somewhat weaker and flimsier than their positive counterparts' (Chierchia (2000)). He hypothesizes this might be the result of the fact that the granularity of the scale seems to rely on contextual factors. Members of an entailment scale sometimes do not seem to be negated by an SI, simply because they are not relevant in the context. Consider for example the <some, all> scale I referred to above. Based on entailment, *most* also belongs on this scale, yet sentences like (24a) do not seem to have the SI (24b), but (24c), skipping *most*:

(24) a. It won't happen that every student will complain.
b. ~> Most students will complain.
c. ~> Some students will complain.

Now the stronger scalar items in DE-contexts are always on the lower end of an absolute scale. Consider for instance the numerals, which are often claimed to be SI-triggers too. Their stronger scalar alternatives in non-DE contexts are higher on the absolute scale of say, the natural numbers. However, in a DE-context their stronger scalar alternatives are lower on this absolute scale. Chierchia claims that because their stronger alternatives are lower on this absolute scale, their contextual importance might be lower. He gives money as an example: having very little money is in many contexts the same as having no money, so we might feel the SIs do not arise because they are not relevant anyway.[10] Horn (2005) disagrees and claims both are just as strong. His arguments are mainly anecdotal, as he mentions 'the very first invocation of the term 'scalar implicature', that according to the Oxford English Dictionary was given in Horn (1972), to account for the fact that *not all* implicates *not none*, i.e. *some*.' He also mentions that Chomsky (1972) states the inference from *not many* to *some* is stronger than a 'garden-variety conversational implicature'. I think the discussion on the status of SIs in DE-contexts is not decided, and as for this work it is of no great importance, I will leave this issue open here.

We can conclude that the main point of discussion between the neo-Gricean view and the Semantic Core Model is whether SIs are computed globally, after compositional semantics, or locally, during semantics. This difference primarily leads to different predictions when a scalar is embedded under a logical operator or another scalar. Both views seem to have a mechanism that can account for the SI blocking and the rise of new SIs in DE-contexts. Both views rely on an informativity constraint, albeit that both have their own way of implementing this constraint.

So although both views have more in common than is often acknowledged, can we find arguments for which view is better? The above data of complex sentences show that the neo-Gricean global mechanism is a bit a too inflexible. Extensions of it such as the proposal by Sauerland (2004) improve the flexibility but rely on a lot of SI-calculation per sentence, which seems uneconomical. Chierchia's local view is a lot more flexible, but it relies on a comparison at each functional application, which also seems uneconomical. Especially the fact that (for instance after a DE-operator is encountered) the initial SI is canceled, and

[10] The fact that Chierchia refers to context to explain that indirect SIs are weaker is interesting. As will be discussed in the next section, Chierchia's theory is usually assumed to be a *defaultist* theory, a theory that claims SIs are triggered by default (irrespective of context), and only possibly canceled later on.

subsequently *all* scalar alternatives of the embedded scalar (including the one that was just canceled) are again compared in the new situation, seems uneconomical. The neo-Griceans simply rely on an external fact to predict that the SI will go the other way: the reversal of entailment scales in DE-environments. However, it is of course this large number of comparisons of Chierchia's system that makes it so flexible that it can handle the more complex sentences that the neo-Griceans cannot. However, as the difference between the two views is also procedural (*How and at which point did the SI come about?*), it is hard to tease the two apart by only relying on intuitions about the meaning of complex sentences, especially as our intuitions are less clear cut for these sentences. Perhaps experiments that conduct measurements during the interpretation process are a better way to find evidence for either of the two views. I will discuss some of these in Section 3.2.

Although the discussion between the neo-Gricean view and the Semantic Core Model is not directly relevant in this work, the question how SIs actually come about will prove very important later on. In the discussion of the experiment I will present, the question arises whether SIs are calculated completely unconsciously or whether subjects have some kind of conscious control over them. And if they do, is this conscious control only a possibility of canceling the SI after it is unconsciously calculated, or can certain considerations of the hearer (e.g. about the intentions of the speaker) block their calculation in the first place? Chierchia's approach seems to assume initial SI-calculation without conscious control, while on the neo-Gricean view, it is not that obvious. It is for instance unclear what happens on the neo-Gricean view when the speaker can no longer be assumed to be cooperative: does this block initial SI-calculation by the hearer or does the SI arise anyway, to be canceled by the hearer who realizes it cannot be grounded in cooperativity? I will return to this matter later. First, in the next section I will discuss the other current main point of discussion on SIs, which will be the starting point for the rest of this work.

2.2 Defaultists vs. contextualists

Besides the discussion about scalars in embedded contexts we considered in the previous section, there is another important area of debate concerning SIs. This is the question whether SIs are default inferences or whether they only arise in certain contexts. The authors that support the first option can be labeled as "defaultists". In the last section we saw that supporters of the Semantic Core Model belong to this camp: The local introduction of SIs always takes place, irrespective of context, they are at best canceled by constituents higher in the tree. The neo-Griceans usually (e.g. Levinson (1983)) take SIs to be *generalized* conversational implicatures (see below) which have as a hallmark that they arise independent of context. So based on that, the neo-Gricean view can also be considered a defaultist theory.[11] Followers of the other hypothesis are often called "contextualists". In this section I

[11] However, the position of the neo-Gricean view in the defaultist vs. contextualist debate is not so clear. In the same work in which he claims SIs are generalized CIs, Levinson (1983) says: ...implicatures can just disappear when it is clear from the context of utterance that such an inference could not have been intended as part of the utterance's full communicative import' and backs this claim up by the typical example in which no SI seems to arise when a certain number is a prerequisite:

Q: Has John really got the requisite number of cows?
A: Oh sure, he's got three cows all right.

Levinson claims the *and not more* SI for the cardinal number *three* does not arise here. However, if SIs are generalized CIs and therefore arise irrespective of context, why does the SI not arise here? In later work,

will contrast the defaultist view with three contextualist approaches: Relevance Theory (RT) of Sperber & Wilson (2002), Exhaustivity of Van Rooij (2002), and a contextualist theory based on topic/comment by Van Kuppevelt (1996).

2.2.1 The defaultist view

Grice divided conversational implicatures (CIs) into two kinds: *generalized* conversational implicatures and *particularized* conversational implicatures. Generalized CIs do not need a special context or a special scenario to arise, while particularized CIs do. Consider for example (25), taken from Levinson (1983):

(25) I walked into a house.
(26) The house was not my house.

Based on the first submaxim of Quantity (*Be as informative as possible for the curent purposes*), the use of the indefinite article *a* seems to trigger the implicature in (26): that the house was not my house (because if it was my house, according to Quantity I would have used the more informative *my*). This is an example of a generalized CI: no special context or scenario is needed to get this implicature. This is contrasted by (27), which does not seem to have (28) as an implicature. Only the special context of (29) does it get this implicature:

(27) The dog is looking very happy.
(28) Perhaps the dog has eaten the roast beef.
(29) A: What on earth has happened to the roast beef?
B: The dog is looking very happy.

So in order for (27) to get the implicature in (28), we need a special context such as the one created by A's question in (29). Therefore (28) is an example of a particularized CI.

Generalized CIs have two important properties: They are *lexical* and *automatic*. By 'lexical' is meant that they are triggered by certain lexical items, and not by bigger entities such as the sentence as a whole, the LF of the sentence, the propositional content of the sentence or the context in which the sentence appears. In (25) and (26) we saw that use of the indefinite expression *a X* triggers the implicature that *X* is not closely related to the speaker, independent of the sentential context.[12] That generalized CIs are automatic is a result of what was stressed above: they do not need any special context. So they arise automatically and will be part of the complete meaning of the sentence, unless something cancels them (if we were for instance to add to (25): *It was my house*). Defaultists claim SIs are generalized CIs too. They are lexical, as they are triggered by lexical items, namely the scalar terms, and they are calculated automatically, irrespective of context.

Levinson (2000) claims the SIs do arise by default, but at a level that is intermediate between the semantics of a sentence and the complete meaning of the sentence. He calls this level *utterance-type meaning*, which consists of default (or 'presumptive') interpretations 'which are carried by the structure of utterances, given the structure of language, and not by virtue of the particular context of utterances.' We can definitely label this as a defaultist theory.

[12] Although there is a famous counterexample: *I broke a finger*. However, here the second submaxim of Quantity is claimed to be at work: "Do not say more than necessary". A hearer will easily conclude that it was your own finger, so being overly specific by using *my* would violate this maxim. The competition between the first and second submaxim of Quantity has been criticized by e.g. Carston (1995), but I will not go into that in this work.

2.2.2 Contextualist approach 1: Relevance Theory

Relevance Theory (RT) has been developed by Sperber & Wilson (S&W) in a series of papers from 1986 onwards[13]. They reject Grice's Cooperative Principle and the four maxims, and replace them by one *Principle of Relevance*. This principle states that human cognition is aimed at optimizing relevance. While Grice's maxims especially apply to conversation, S&W claim the Principle of Relevance is a general property of cognition, created by evolution. But what does it mean to optimize relevance? According to S&W, the goal of information processing is to yield a *positive cognitive effect*. S&W describe this in the following way: 'a worthwhile difference to the individual's representation of the world – a true conclusion for example' (Wilson & Sperber (2002)). The most important type of positive cognitive effect is a *contextual effect*: a conclusion deducible from the (linguistic) input and the context together, by means of for instance implications or strengthenings (of meaning). So a contextual effect comes about when a hearer can draw a conclusion (for instance about the world around him) based on both the linguistic input (the encoded message) and the context (can be anything: world knowledge, the situation in which speaker and hearer are, etc.), and the hearer could not have drawn this conclusion from either input or context alone. However, RT states this contextual effect should come about for as little effort as possible. This balance is captured in what S&W simply call *Optimal Relevance*, given here in (30):

(30) Optimal Relevance
An ostensive stimulus is optimally relevant to an audience iff:
a. It is relevant enough to be worth the audience's processing effort.
b. It is the most relevant one compatible with communicator's abilities and preferences.

Condition (30a) is based on the fact that we are willing to try and interpret utterances because every utterance comes with the presumption that it is worth while.[14] Condition (30b) actually captures why the balance between the maximalization of contextual effect and the minimalization of effort causes a match between speaker- and hearer meaning: It tells the hearer that he will not have to work too hard to get the intended effect, he can just land on the easiest deducible meaning that satisfies his expectations of relevance and stop after that. As long as his expectations of relevance are satisfied, he does not have to go and look for other meanings. This is because (30b) instructs the speaker to formulate his utterance in such a way that it is the most easily accessible meaning that satisfies the expectations of relevance of the hearer. So the hearer needs to search no further, as the meaning that is easiest to access is the one the speaker intended to convey.

So how does RT deal with SIs? Consider for example the *not all* SI associated with the scalar term *some*. According to RT, this implicature will only be calculated by the hearer if it is needed to satisfy his expectations of relevance. If this expectation is already satisfied without the SI, the least effort property of (30) predicts it will not be calculated. Accordingly, a speaker will only expect a hearer to calculate it if it is relevant. So whether the SI arises depends on whether it is relevant in the context. Contexts in which SIs arise are, as Carston (1995) puts it, '[...] ones in which a contrast between *only some* and *all* is relevant, where a crucially different range of contextual effects will arise depending on whether the whole

[13] This paragraph is based on Wilson & Sperber (2002).
[14] S&W call this the *Communicative Principle of Relevance*: Every ostensive stimulus conveys a presumption of its own optimal relevance.

domain or merely a proper subset of the domain is involved.' She gives the following example:

(31) Mary failed some of her exams.

Generally, in the context of failing exams, it is important whether someone failed all or only some of her exams. So on the *some and possibly all* meaning of *some* (without the SI), (31) will not satisfy the expectations of relevance of the hearer and will therefore not fulfill (30a), the guarantee of an adequate contextual effect. So in contexts like these, the SI arises to satisfy relevance. However, consider the exchange in (32), where according to Carston no SI is calculated:

(32) A: If any of the students have failed I'll be in trouble.
B: I'm afraid some of them have.

This is what Carston calls a 'threshold case': Only the lower bound (*at least one*) is relevant. The only thing that is relevant to A, the hearer of B's utterance, is whether or not at least one student failed, because then he is in trouble. So B's utterance already fulfills the expectations of relevance by confirming that at least one student failed (by using *some*), without limiting the upper bound of *some* by a *not all* SI. As Carston puts it: 'There is no such inference because it is not required for the hearer to satisfy himself that the presumption of optimal relevance has been fulfilled.' This goes straight against the defaultist view, who would predict an SI for (32), based on the presence of a scalar term, independent of context.

Another major point of discussion between the neo-Gricean view and Relevance Theory is whether or not implicatures are part of *what is said*: the truth-conditional content of a sentence. As we saw above, the neo-Gricean view does not count implicatures to belong to truth conditional content, as they are cancelable. However, RT claims some implicatures (including SIs) do in fact affect truth conditions, and are therefore part of what is said, so should be called *explicatures* instead of implicatures. RT claims many LFs are truth-conditionally incomplete: they have open slots, which have to be filled by inferences. These open slots are for example (taken from Horn (2005)): the meaning of genitives (*Salman's book* – the one he reads, wrote, owns?), unspecified comparison sets (*Chris is tall* – for an adult (fe)male?), or other expressions with free variable slots (*Lee is ready* (for what?)). RT claims these slots have to be enriched, consisent with the principle of relevance, and that these enrichments are explicatures, part of what is said. Horn (2005) however shows that some of these enrichments seem to be neither part of what is said, nor part of what is implicated. Consider for example the common enrichment of *A and B* to *A and B in that order* in (33):

(33) John and Mary had a baby and they got married.
(34) John and Mary had a child and got married, but not necessarily in that order.
(35) If John and Mary got married and had a child, their parents will be pleased, but if they had a child and got married, their parents will not be pleased.

Sentence (33) will normally be interpreted as: *John and Mary had a baby and they got married, in that order*. However, as (34) shows, this inference can be canceled, so it does not seem to be part of the truth conditional content of the sentence. But (35) shows that it definitely has effect on truth conditions: if the antecedent of the two conditionals had the same truth conditions, (35) would be a contradiction, which it obviously is not. Because its truth conditional effect, Horn claims these inferences are indeed not implicatures. However, they

are not explicatures either, as they are not explicit and they are cancelable. Eventually, he follows Bach (1994, 2001) and decides to keep calling these inferences *implicatures*, until a better term comes along. For the present purposes, the labeling of SIs as implicatures or explicatures is not important, as we are focusing here on whether they are default inferences or contextually dependent.[15] In the next section, I will consider another approach that claims they are dependent on context.

2.2.3 Contextualist approach 2: Exhaustivity

Van Rooij (2002) proposes an account of SIs in terms of *exhaustivity*. He does not reject Grice's cooperative principle and the maxims, like RT does, but another step in the neo-Gricean reasoning: the conclusion that can be drawn about the knowledge of the speaker. Gazdar (1979) put forward the view that if a speaker uses a weaker scalar term, it can be concluded by the hearer that the speaker knows the stronger alternative does not hold.[16] If we combine this view with the assumption that entailment alone is enough for two items to form a scale, we would predict (36) to have the SI in (37), as *John is sick* entails (36):

(36) Someone is sick.
(37) ~> It is not the case that John is sick.

Of course the assumption that entailment alone is enough for two items to form a scale is questionable: We already saw in Chapter 1 that items in a scale are usually the same type of expression, contrary to the proper noun *John* and the quantifier *someone* in this example. Another criterion, stemming from Grice's submaxim of Brevity, might be that both items have to be lexicalized to the same degree (blocking for instance *iff* as a scalar alternative to *if*). However, if we ignore these problems and assume <someone, John> is a scale, Gazdar's account makes too strong a prediction, as intuitively (36) does not have the SI in (37). Based on these considerations, Soames (1982) claimed that instead of concluding that the speaker knows the stronger item is untrue, we can only conclude that the speaker does not know whether the stronger sentence holds[17]. But this conclusion seems to be too weak: How can we explain the SI for sentences like *A or B* (that *A and B* is untrue) if we conclude that all the

[15] Although in the experiment, we will rely on the power of SIs to affect truth conditions. This is merely a result of the experimental method adopted (for reasons of comparison to other experiments), than a statement about the status of SIs as part of what is said.

[16] I also adopted this 'idealization' of an all-knowing speaker in Chapter 1. For the experimental part of this work, this idealization is harmless, as an epistemic limitation is ruled out by the experimental task. In the present section however, we want to illustrate the workings of exhaustification versus the traditional neo-Gricean mechanism, and for that we do need to question this idealization.

[17] In Van Rooij & Schulz (2004), five types of conclusions about the speaker's knowledge are distinguished:

(i) the *factive strong reading* of Horn (1972): the stronger item is untrue (¬s')
(ii) the *epistemic strong reading* of Gazdar (1979): the speaker knows or believes the stronger item is untrue (□¬s')
(iii) the *factive weak reading* of scalar items under negation (also Gazdar (1979)): no implicature at all
(iv) the *epistemic weak reading* of Soames (1982): the speaker thinks it is possible that the stronger item is untrue (◊¬s')
(v) the *ignorance* reading (Gazdar (1979) for clausal implicatures): the speaker does not know whether the stronger reading is true or not (◊¬s' ∧ ◊s').

Above, we already collapsed the first two types, which will be our idealized implicature in the rest of this work. For simplicity we will we also collapse the two last types in the discussion in this section. As was already shown in the section on embedded implicatures, the third type seems to be non-existent.

speaker knows is that *A and/or B* is true, and he does not know whether they both are? Van Rooij claims that all that can be concluded about the knowledge of the speaker, is that he knows the exhaustified reading is true, and nothing more. So what exactly is the exhaustified reading?

Van Rooij follows Hirschberg (1985) and states that SIs seem to depend on the context, and specifically on the question that sentences containing scalar terms are supposed to answer. He gives (38) as an example:

(38) There are cookies or chocolates in the box.
(39) ~> There are not both cookies and chocolates in the box.

Sentence (38) normally has (39) as an SI. However, if (38) is an answer to for instance the *yes/no*-question (40), this SI seems not to arise:

(40) Are there cookies or chocolates in the box?

Van Rooij therefore states that assertions that contain scalar terms should always be analyzed as answers to questions, and that exhaustivity is the key factor in determining whether the SI arises or not. He says: '...almost all typical quantity implicatures can be alternatively analyzed on the assumption that assertions are exhaustified answers to questions.' (Van Rooij (2002)). But what is an exhaustified answer? Consider the question in (41) and three possible answers in (42a-c):

(41) Who is sick?
(42) a. John.
b. John and Bill.
c. A man.

Groenendijk & Stokhof (1984) (from here on G&S) noticed that answer (42a) conveys that *only* John is sick, and no one else. Similarly, (42b) says only John and Bill are sick, and (42c) that only a man is sick. This is because the answers in (42a-c) are interpreted as *exhaustive*: they are considered to give a complete list of the people that are sick (in the relevant domain). G&S introduce an exhaustivity operator (*exh*), that picks out this meaning from for instance the answers in (42a-c). Two equivalent definitions of *exh* are given in (43):

(43) *exh* $= \lambda \boldsymbol{P} \lambda P\, [\boldsymbol{P}(P) \wedge \neg \exists P' : \boldsymbol{P}(P') \wedge P' \neq P \wedge \forall x[P'(x) \rightarrow P(x)]\,]$
$= \lambda \boldsymbol{P} \lambda P\, [\boldsymbol{P}(P) \wedge \neg \exists P' : \boldsymbol{P}(P') \wedge P' \subset P]$

In (43), $\boldsymbol{P}$ is the denotation of the term answer (of type <<e,t>,t>), so for instance λPP(j) for (42a) (*John*), and P is the property underlying the wh-question (type <e,t>), so for instance λxSICK(x) for (41) (*being sick*). What *exh* actually does is pick out the minimal elements of a set of sets. Let's see how this works for the answers in (42a-c). If we assume a domain of three people, John, Bill and Mary, λPP(j) (*John* in (42a)) corresponds to the following set of sets: {{j},{j,b},{j,m},{j,b,m}}. Now if we apply *exh* to λPP(j), we get (after some λ-conversions): $\lambda P \forall x[P(x) \leftrightarrow x = j]$. This corresponds to the quantifier {{j}}. So *exh* picked out the minimal element of λPP(j), {j}, so this is the set of individuals that is sick. This corresponds to the reading that only John is sick, which is what (42a) intuitively means. Similarly for (42b), $\lambda P[P(j) \wedge P(b)]$ (*John and Bill*) is the following generalized quantifier: {{j,b},{j,b,m}}. If *exh* is applied to the λ-term, we get $\lambda P \forall x[P(x) \leftrightarrow x = j \vee x = b]$, which

correponds to {{j,b}}, so *only* John and Bill being sick, which is what we want. Finally, the answer *a man* in (42c), can in our domain be translated as λP[P(j)∨P(b)] (*John or Bill*, since John and Bill are the only men in our domain), which corresponds to: {{j},{j,b},{j,m},{j,b,m},{b},{b,m}}. Applying *exh* gives λP∀x[[P(x) ↔ x = j] ∨ [P(x) ↔ x = b]], so the set of sets {{j},{b}}. This means only John or only Bill is sick, which is intuitively correct.

Van Rooij adopts G&S's exhaustivity operator and explains the SI-data with it. Consider for instance the answer *John or Bill* to (41). Since we translated *a man* as λP[P(j)∨P(b)] (*John or Bill*), we already saw what *exh* gives us: the set of sets {{j},{b}}. Crucially, the set {j,b} is not picked out by *exh*, so if we take *John or Bill* to be an exhaustive answer (which we will, considering (41)), we can conclude that the speaker knows that it is not the case that John and Bill are both sick. So the *not and* SI of *or* is arrived at directly by exhaustivity. This contrasts with Soames' account, on which we can not explain this SI. Also, exhaustivity does not make too strong predictions like the one we saw of Gazdar(1979) in (37). The answer *a man* in (42c) would, on Gazdar's account have the SI: *it is not the case that John is sick*, as *John is sick* entails *A man is sick*. However, the exhaustive answer {{j},{b}} does not predict this incorrect SI, it is compatible with John being sick.

Another interesting property of the exhaustivity approach that Van Rooij points out is that it can account for some of the embedded environments we saw in Section 2.1. Consider for example (44), from Chierchia (2000):

(44) John believes that some men are sick.
(45) ~> It is not the case that [John believes all men are sick].
(46) John believes that some men are sick and it is not the case that John believes all men are sick.
(47) John believes that some, though not all, men are sick.

The neo-Gricean mechanism would predict (45) to be the implicature of (44), and (46) the complete meaning of (44). This is too weak however, because it does not follow from (46) that John actually disbelieves that all men are sick. The complete meaning (46) is compatible with a situation in which John does believe that some men are sick and he simply has no beliefs about whether all men are sick or not. The complete meaning that Chierchia's local mechanism predicts, given in (47), does rule out this situation. Van Rooij claims exhaustivity can also explain this SI, if for instance (44) is an answer to *Who does John believe is sick?* The exhaustive interpretation of *some men* does not pick out the set of sets that contain all men, so it yields the *some but not all* reading.

Exhaustivity can also account for the absence of SIs in some contexts, like the one we saw in (38)-(40) above. Crucially, the question determines what is an exhaustive answer and as we stated above, the speaker can only be assumed to know that this exhaustive answer is true, but nothing more. Consider for example (48):

(48) a. Q: Who has 2 children?
b. A: John has 2 children.
c. John doesn't have more than 2 children.

Cardinal numbers such as *2*, are often considered to be scalar terms too (on a scale with the other numbers), and to trigger the SI *not more than n*, resulting in an *exactly n* (here: *exactly*

2) reading[18]. Without the question in (48a), (48b) would normally get the SI in (48c). However, as an answer to (48a), the SI seems to be absent. Some authors (e.g. Levinson (1983), see fn. 11) have claimed that in a case like this the SI is canceled, because it is not relevant. However, on an exhaustivity account, the SI does not arise in the first place: As an answer to (48a), (48b) is equivalent to the term answer *John*. The exhaustivity operator will therefore apply to *John*, making sure that the meaning of (48b) is that *only* John has 2 children in our current domain. However, as this is an exhaustive answer to the question, this is the only thing we can conclude the speaker to know: It is John and only John who has 2 children. According to the exhaustivity view, the speaker cannot be assumed to know that a stronger statement that is not part of the exhaustive answer, for instance that John has 3 children, is untrue. This is why the SI (48c) will not arise on the exhaustivity view. An important property of the exhaustivity view is that the question determines which part of the sentential answer is interpreted exhaustively, and it is only that part in which SIs can arise. In the next section I will consider another contextualist approach, that makes roughly the same predictions but is based on different notions.

2.2.4 Contextualist approach 3: Topic/comment

Van Kuppevelt (1996) also acknowledges that the generation of scalar implicatures seems to be dependent on the question the sentence is supposed to answer. He says: '...on each discourse level the generation of a scalar inference is determined by the explicit or implicit (sub)topic-forming question...' His analysis relies on the topic/comment structure of sentences: he claims SIs are only generated if the scalar term is in a constituent that has comment status. Let's start by considering what exactly is meant by topic and comment. Van Kuppevelt gives the following definition:

> By definition, a topic T_p is *that which is being questioned* by means of a contextually induced explicit or implicit question Q_p. The corresponding comment C_p is provided by answer A_p. C_p is *that which is asked for* by Q_p. [his italics]

Here, topic T_p is the intension of the topic term of the question, e.g. the intension of *the one(s) who is/are laughing* for (49):

(49) Q: Who is laughing?
A: Alan is laughing.

So T_p of (49), (T_{49}), is the intension of *the one(s) who is/are laughing*. Assuming a domain of two people, Alan and Brian, T_{49} is the following:

(50) $T_{49} = \{<w_1,\{a\}>,<w_2,\{b\}>,<w_3,\{a,b\}>\}$

So the intension of the topic term is the set of possible extensions of this term, the set (51):

(51) $\{\{a\},\{b\},\{a,b\}\}$

[18] However, whether the *exactly n* meaning of cardinals comes about by SI is a matter of debate. See e.g. Horn (2005), who says: '...the RT approach convincingly overturns the original Gricean line on the meaning of cardinal operators (lower-bounded by meaning, upper-bounded by implicature).'

This set corresponds to the possible answers to (49Q): *Alan, Brian, Alan and Brian*. The comment C_p is the actual extension of the topic term, the one that is picked out by the answer. For answer (49A) this is {a}. So in procedural terms, what happens is the following: Topic-forming questions are the result of textually given *indeterminacies*: a non-uniquely referring term which is questioned and therefore becomes a topic. In (49) this is the topic term *the one(s) who is/are laughing*. At the moment of questioning, the extension of the topic term is undetermined: the topic range does not contain a unique value, see e.g. (51). The answer determines the unique value of the topic term in the actual world/situation (the comment C_p), (e.g. {a}). With this, a necessary condition for topic-hood (undeterminateness of the topic term) is no longer met, resulting in 'topic closure'.

Now what does this mean for SIs? Consider (52):

(52) Q: Who has fourteen children?
A: (i) Nigel$_{comment}$ has fourteen children. (ii) In fact he has twenty.

Defaultists predict a *not more than* SI in (52A), which is then canceled by the second sentence (52A(ii)). However, Van Kuppevelt notes the following:

> It seems that in cases where the assumed implicature inducing expression is not part of the comment, [defaultist, AZ] implicature theory assumes, without evidence, two extra processes, namely a process of generation and one of cancellation [...] both processes are redundant given the semantic interpretation 'at least fourteen' of the lexical item of which it is assumed that it gave rise to the implicature.'

He goes on to show that (52A(i)) still has an indeterminacy, a non-uniquely referring term that can become a topic. As noted above, questions are the result of these indeterminacies, and (52A(i)) still gives rise to the (possibly implicit) question how many children Nigel exactly has, as (52'Q2) is completely natural:

(52') Q1: Who has fourteen children?
A1: Nigel$_{comment}$ has fourteen children.
Q2: <How many children does he have?>
A2: He has twenty$_{comment}$.

On the defaultist account, (52'A1) would trigger the *and not more* implicature, resulting in the *exactly fourteen* reading, leaving no indeterminacy and ruling out a following question like (52'Q2) about the exact number. Van Kuppevelt concludes that the possibility of another question (and therefore the presence of an indeterminacy), shows that *fourteen* in (52'A1) does not get the *not more than* SI. According to Van Kuppevelt, the decisive factor for SIs is whether or not the scalar term has comment status: If it does, the SI is calculated, otherwise not. Consider for instance the following minimal pair:

(53) Q: Who$_{topic}$ bought four books?
A: Harry$_{comment}$ bought four books. (In fact he bought seven.)

(54) [Harry did a lot of shopping this afternoon.]
Q: How many$_{topic}$ books did he buy?
A: He bought four$_{comment}$ books. (#In fact he bought seven.)

In (54), because *four* has comment status (it is *that which is asked for* by Q), it gets the *exactly four* reading, and as a result a continuation with a higher cardinal (*seven*) is infelicitous.[19] [20] In (53) though, the implicature never arises because *four* does not have comment status. Therefore, a continuation with a higher cardinal is felicitous.[21]

Summarizing, Van Kuppevelt's approach is contextual in the sense that, just like Van Rooij's exhaustivity approach, it relies on an implicit or explicit question in the context. SIs are not calculated automatically, but only if they have comment status. If they do not have comment status, they are not calculated and possibly canceled afterwards, but they are not calculated at all – contrary to what the defaultists claim. This is signaled by the fact that scalars with no comment status still allow questions about their exact value. In the following section, I will consider what the three contextualist accounts above have in common and how their predictions can be tested against those of the defaultists.

2.3 Common properties and predictions of the contextualist approaches

In the previous section we considered the defaultist vs. contextualist debate. Defaultists claim SIs are triggered by certain lexical items and are calculated automatically, independent of the sentence or context the scalar appears in. Contextualists claim that the context is decisive for the calculation of SIs: SIs are not calculated automatically with every scalar term, they only appear if the context satisfies certain conditions. For RT, this condition is the expectation of relevance of the hearer, for exhaustivity it is being part of an exhaustive answer and for the topic/comment approach the condition is having comment status relative to a contextually induced question. In this section I will try to reduce these three conditions to one, more general condition: the condition that a scalar should be in the part of the sentence that answers the Question Under Discussion. After that, I will show that focus can function as a signal of this answerhood. This will allow us to test the predictions of this new contextualist condition against the predictions of the defaultists.

2.3.1 What the contextualist approaches have in common: the Question Under Discussion

As was already noted above, a similarity between the exhaustivity account and the topic/comment account is that both theories rely on the relation between the sentence and the possible question it is supposed to answer, to predict the presence or absence of SIs in this sentence. Both Van Rooij and Van Kuppevelt also agree that this question can be explicitly stated, as in natural conversation, or it can be present in the context, explicit or implicit. In RT, it is the expectation of relevance of the hearer that determines whether or not SIs are calculated: the interpretation that satisfies this expectation with least effort, wins. However, the question-based approach and the relevance-based approach are actually more closely

[19] Van Kuppevelt, like S&W, claims that SIs are entailments rather than pragmatic inferences. He claims in (54), the SI cannot be canceled as it is blocked by contradiction, and contradiction can only arise if the *and not more* SI is entailed by the answer, so Van Kuppevelt concludes the SI has to be an entailment. As I pointed out in section 2.2.2, this discussion is not relevant for the present purposes (but see fn. 15).

[20] Other theories could claim (54A) is simply not a good answer or even an untrue answer to (54Q) if Harry bought 7 books. For instance Van Rooij's exhaustivity would reject (54A) on the basis that it is not an exhaustive answer. I will ignore this for the sake of Van Kuppevelt's argument.

[21] A theory of implicature cancellation would be able to account for (53), by claiming the implicature is present but canceled. However, it cannot account for why in (54) this cancellation is problematic.

related than they seem at first. Consider first the case of an explicit question, that is actually stated by someone in the conversation. This person now becomes the hearer, and his expectations of relevance of the speaker's utterance will be based on to what extent the speaker's utterance is an answer to his question. So the interpretation that will satisfy his expectations of relevance with least effort is the interpretation that best answers his question. So here for RT too, if we consider the sentence relative to the question, we can predict what the interpretation will be, and we can predict whether or not the SI will be calculated. Now how about the situation in which there is no explicit question? According to exhaustivity and topic/comment, we should now consider the implicit question, raised by the context, to be the question relative to which the answer should be interpreted. That means the main question the context gives rise to will fulfill this task. But this is exactly the issue on which the expectations of relevance of the hearer will be based: it is the most relevant thing in the context. So here again RT and the question-based approaches seem to give the same instruction: consider the sentence relative to the main question in the context, which is the most relevant thing under discussion right now. So intuitively, the main question of a context and the most relevant issue are one and the same thing. From now on, I will refer to this as the Question Under Discussion (QUD).

Consider for example a variant of (48) above, in which there is no explicit question *Who has two children?*, but this question is the QUD of the context. This might be the case if for instance some friends are showing each other an ad in the newspaper that says that each person that brings two children to today's ball game, gets in for free. They are wondering if they have to alert any of their (other) friends about this. One of them might utter (55), even if John has in fact ten children:

(55) John has two children.

According to exhaustivity, the exhaustive answer to the QUD *Who has two children?*, here is *John*. The numeral *two* does not get a *not more* SI as it is not part of the exhaustive answer. Van Kuppevelt would claim that only the constituent *John* has comment status relative to the QUD, so *two* gets no SI. Furthermore, RT makes exactly the same prediction: the expectations of relevance of the hearer are set in the context of the wondering about whether they have any other friends that might profit from this ad. So the very same QUD will be considered by them: *Who (of our other friends) has two children?* The expectations of relevance of the hearer obviously do not include how many kids John actually has. So, as the scalar *two* is not the relevant part of the answer, it gets no SI.

What all three contextualist approaches actually do, is split a sentence up into a part that actually answers the QUD (e.g. *John* in (55)), and a part that does not (*has two children* in (55)). This other part is already present in the QUD, and is kept constant in the answer. This part is sometimes referred to as the *presupposed part*.[22] Hamblin (1973)'s theory of the semantics of questions also reflects this split: He states that the denotation of a question is the set of all possible answers, irrespective of the actual answer. So, the denotation of *Who has two children?* looks like this: {John has two children, Jack has two children, Bob has two children, ...}, with all possible answers (relative to a domain) as set members. Groenendijk & Stokhof (1994) adopt Hamblin's view and describe the possible answers as *partitions of*

[22] For my purposes it does not matter whether this is really presupposition or not. What is important is that it can not be changed in the answer without giving infelicity: *I borrowed shoes yesterday* is not a felicitous answer to *What did you buy two days ago?*. I leave it up to presupposition experts to decide if this infelicity is caused by presupposition failure.

logical space: The logical space is completely filled because all possibilities are represented as a partition of it. So the denotation of *Who has two children?* can be visualized as follows:

(56)

Nobody has two children
John is the one that has two children
Jack is the one that has two children
Bob is the one that has two children
John and Jack are the ones that have two children
...
Everybody has two children

This theory of the semantics of questions clearly reflects the split that the contextualist approaches make use of: the split between the questioned part and the presupposed part of a question, and the corresponding split between the answering part and the presupposed part of an answer. I therefore propose the following generalization of the three contextualist approaches, which I will for now label the "QUD Condition":[23]

(57) **QUD Condition**: An SI will arise in a sentence iff the scalar term (with which the SI is associated) is in a constituent that answers the QUD of the context that the sentence is part of.

So, although the three contextualist accounts above differ in the underlying reasons of why the scalar should be in the QUD-answering part of the sentence, all three of them make this prediction. In the next section, I will go on to show how focus can play a role in signaling which part of the sentence is the QUD-answering part. First however, I will illustrate with some examples the importance of the QUD if one takes the QUD Condition seriously.

In the experimental literature, the most commonly used technique to assess SIs is by a Truth Value Judgment Task (TVJT). Participants are presented with a situation (picture or story) and have to judge a sentence, containing a scalar term, on truth or falsity relative to this situation. Usually, the sentence is such that the SI is decisive for truth or falsity. This way, the answers of the participants reflect whether or not they calculated the SI. A famous example is the experiment of Noveck (2001). He presented participants with sentences like: *Some elephants have trunks*. Obviously, the only reason why participants would judge this sentence untrue, is based on the SI: *and not all elephants have trunks*. However, Noveck presented these sentences with no context at all. Therefore, it was impossible for subjects to distill from context what the QUD is, what is at issue here. It is often claimed that in absence of context, people still search for a QUD and possibly accommodate one. Participants could have accommodated the QUD: *Do some or all elephants have trunks?* Obviously, *some* would be in the part that answers this QUD, so the QUD Condition would predict an SI and an 'untrue' answer. However, an array of other QUD's is possible, for instance: *Do some elephants (that are for some reason salient) have trunks?* or *What do some elephants have for grabbing stuff?*

[23] Notice that the sentence itself is also of importance for determining what the QUD is. This is of great importance in a Truth Value Judgment Task, where a sentence is presented after a story: Story and target sentence together determine the QUD.

With these QUD's, *some* is not in the answering part of the sentence, so the QUD Condition predicts no SI. As it turns out, adult participants had great difficulty with these sentences out of context, answering 'untrue' in only 59% of the cases. Percentages around 50% are often taken as a signal of guessing behavior. If the participants were following the QUD Condition, this behavior makes sense: As the QUD was unclear, they had no criterion to determine whether the SI should be calculated or not. All they could do is guess what the QUD is, accommodate this QUD and decide based on that.

A context in which the QUD is obvious, is for instance Chierchia et al. (2001)'s description of a test item of their Experiment 2:

> In one trial, children [and adults, AZ] were presented with a story about four boys at the summer camp who are choosing which toys they want to play with. There are a lot of toys they can choose: some skateboards, some bikes, a boat and a truck. After considering the possible choices, the four boys [each, AZ[24]] take both a skateboard and a bike. At this point, a puppet produced the following target sentence:
>
> "Every boy chose a skateboard or a bike."

Chierchia et al. do not provide more information on their items than the description above. However, in the description above, there is only one thing salient in the story: the choice the boys make. So the QUD the sentence answers has to be: *What did the boys choose (from the set of toys)?* The choice of the boys is central in the whole story, as can be read off parts of the story like: *after considering the possible choices...* Also, other possible QUD's, like: *Which of the boys chose a skateboard or a bike?* are made unlikely because the members of the set of boys are never distinguished, they are always referred to as *(the) four boys*, so that question never arises. The members of the set of toys though, are distinguished: *There are a lot of toys: ... some skateboards, some bikes, a boat and a truck...* so it is obvious that the QUD is about the toys. As *or* is in the part of the target sentence that answers the question (*a skateboard or a bike*), according to the QUD Condition people will calculate the implicature *not a skateboard and a bike* and judge the sentence untrue. As I said, it is not clear whether in the items Chierchia et al. really did not distinguish the boys etc, but adults did indeed judge this kind of sentence untrue in 100% of the cases.

This example can be contrasted with an experiment by Janse et al. (2005). In one of their stories, three wizards were introduced (seperately), that came to a meeting in the forest. Two of them had a moustache and brought a torch and an oil lamp. The third wizard did not have a moustache and brought nothing. The target sentence was: *Every wizard that has a moustache, brought a torch or an oil lamp*. Here, contrary to Chierchia's story, the members of the subject set are introduced separately and distinguished (by the property of having a moustache or not). This distinction returns in the target sentence, that is only about the wizards with a moustache. So, this combination of story and sentence gives rise to at least two QUD's, one questioning the subject (*Which wizards brought a torch or an oil lamp?*) and the other questioning the object (*What did every wizard that has a moustache bring?*).[25] Only on the latter QUD is *or* in the part of the sentence that answers the QUD and will an SI arise according to the QUD Condition. Although the experimental situation was not completely

[24] I hope Chierchia et al. did not have this ambiguity in the actual stories.
[25] I will ignore the pair-list question: *What did which wizard bring?* here, but it is a possible QUD.

comparable to Chierchia's, Janse et al. observed only 17% 'untrue' answers, signaling that many participants went for the subject-oriented QUD, that was blocked in Chierchia's trial.[26]

The very different results of the three experiments above that were conducted to test for SIs show that the QUD can be a decisive factor that is often overlooked in past experiments. In the next chapter of this work I will reconsider the results of some more past experiments in terms of the QUD Condition. Now I will move on to the question how we can make the predictions of this condition operational. As it turns out, focus is a good tool to do just this.

2.3.2 Focus as a signal of QUD-answer

One of the leading theories on the semantic effects of intonational focus is the one by Rooth (1985). His idea is that next to its 'ordinary' semantic value, each sentence has a 'focus semantic value', that interacts with semantic and pragmatic processes. This focus semantic value is 'the set of propositions obtainable from the ordinary semantic value by making a substitution in the position corresponding to the focused phrase.' (Rooth (1992)). Intuitively, this is the set of alternatives from which the ordinary semantic value is picked. So for sentence (58), in which *Mary* is focused, the focus semantic value is (59), which can be formalized as in (60):

(58) $[_S$ [Mary]$_F$ likes Sue].
(59) {Mary likes Sue, Debbie likes Sue, Wendy likes Sue, ...}
(60) $[[[_S$ [Mary]$_F$ likes Sue] $]]^f$ = {like (x,s)| $x \in$ D} (where D is the domain of individuals).

This immediately reminds us of Hamblin's semantics of questions. We saw for instance in (56) that the denotation of questions is claimed to be the set of all its possible answers. Now (59) looks very much like the set of possible answers to a question like: *Who likes Sue?* Rooth (1992) also notices this similarity. He claims that answer formation is one of the processes with which the focus semantic value interacts: It seems that for a sentence to be a possible answer to a question, it has to be in the focus semantic value of the actual answer. Rooth (1992) says: 'We might say that the function of focus in an answer is to signal other propositions which are potential answers in the context of a question.' This is very useful for our purposes: If focus signals which part of the answer can be substituted to form an alternative answer to the same question, it signals exactly the part we are interested in: the part that actually answers the question. This corresponds to the well-known idea that focus signals *new information* in a sentence. The rest of the sentence is old information: it was already present in the QUD. Only the part that was not already given by the question, gets focus.

So if we consider an answer to a QUD and we take the QUD Condition into account, an SI can only arise in the part of the sentence that has focus. We can now extend the QUD Condition to (57'), and change its name to the *QUD Focus Condition*.

(57') **The QUD Focus Condition for Scalar Implicatures**: An SI will arise in a sentence iff the scalar term (with which the SI is associated) is in a constituent that answers the QUD of the context that the sentence is part of, and therefore has focus.

[26] Note that an important difference between the two experimental situations was that Janse et al. presented the target sentence separately from the story. As a result of this setup, participants had to keep the story in mind and were not able to look back at the situation, which participants in Chierchia's study were.

Let us see what predictions this makes for the SI we will be concerned with in the rest of this work: the *not and* SI of the scalar term *or*. Consider the following two minimal pairs (61-62) and (63-64):

(61) Q: Who/which man had A or B?
A: John$_F$ had A or B.

(62) Q: What/which weapon does John have?
A: John has A or B$_F$.

(63) Q: What/which weapon did A or B have?
A: A or B had a gun$_F$.

(64) Q: Who/which man had a gun?
A: A or B$_F$ had a gun.

In (61A), *or* is not in a constituent that has focus, so we do not expect the *not and* SI to arise. That means that in a TVJT experiment, there is no reason to judge (61A) untrue in a situation in which John has both A and B (as long as John is the only one who has A and/or B). In (62A) however, *or* is part of the focused constituent *A or B*, so the QUD Focus Condition predicts an SI here. That means we expect participants of a TVJT to judge this sentence false in a situation where John has both A and B. The same reasoning holds for the minimal pair (63-64), where *or* is part of the subject. In (63) it has no focus, so no SI is predicted, while in (64) it has focus, and therefore we expect an SI. In the experiment in the next chapter, I will focus on the difference between sentence-types like (61) and (62). According to the contextualist QUD Focus Condition they form a minimal pair, while the defaultist view predicts no difference between them: in both sentence-types an SI should be calculated.

CHAPTER 3: EXPERIMENTS ON SCALAR IMPLICATURES

3.1. Earlier experimental studies reconsidered

In section 2.3.1 I already illustrated how taking into account the QUD can shed new light on contradictory results from earlier experimental studies on SIs. The last couple of years have seen an increasing number of experiments in this area, especially from the perspective of L1-acquisition, as it is well known that young children seem to have difficulties in calculating SIs. As we have seen, the adult data that is gathered in these experiments, mostly as control data, provide contradictory results. In this section I will reconsider some more of these studies on SIs, by considering what QUDs were at work.

Although most experimental studies on SIs that can be found in the literature were done in the last couple of years, Noveck (2004) refers to some classic reasoning studies that addressed the same issue. For example Evans & Newstead (1980) conducted experiments about disjunction, one with adults and one with children and adults. They already found the guessing-like behavior, reporting that 43% of the adult group ruled out a sentence of the form *A or B* when both A and B were the case.[27] Paris (1973) already found the discrepancy between adults' and children's interpretation of *or*, which was a matter of interest for developmental studies already back then, as studies by Sternberg (1979), Smith (1980) and Braine & Rumain (1981) illustrate.

Noveck (2001)

I will take the study by Noveck (2001), that was already referred to in section 2.3.1, as the starting point of the explosion of experiments on SIs of the last couple of years, although there probably were many predecessors. A reason to start here is that this experiment seems to have formed a baseline for much work that was done afterwards. Noveck (2001) first presents two experiments on the scalar modals *might* vs. *must*, which I will not go into here. The most interesting experiment is his Experiment 3, in which 6 types of sentences (differing in truth and felicity) containing *some* or *all* were presented to participants. The sentences were presented by the experimenter without any context and participants were asked to say if they agreed or not. Noveck tested two groups of children and one adult control group (N=15), I will focus on the last. The two most interesting types of sentences were sentences with *some*: one type that is true both with and without SI, e.g. *Some birds live in cages*, and the other type that is true without SI and false with SI, e.g. *Some elephants have trunks*. On the former type, 99% of the sentences was agreed on by the adult participants, while only 41% of the latter

[27] Remarkably, they actually presented sentences (given as rules) with *either ... or*, and still the majority of the participants judged the sentence 'true' when both disjuncts were true. How can this be? Noveck (2004) cites a rule of Evans and Newstead (1980): *Either there is a P or a 4*, which was given with some letter-number pairings. If these pairings included a pair consisting of P and 4, 57% of the participants said the rule is satisfied ('true'). This might have been the result of the fact that the sentences were presented as rules. Rules often provide a *sufficient condition*, which is typically embedded in a DE-environment such as the antecedent of a conditional (*If you photograph these objects or touch them, you will be fined*), the restrictive clause of a universal like *everybody* or *every time* (*Everybody who photographs these objects or touches them, will be fined* or *Every time you photograph these objects or touch them, you will be fined*), or under negation (*Do not photograph these objects or touch them*). As we have seen above, DE-environments block the *not and* implicature of *or*. This might be the reason why even though *either...or* was explicitly stated, it still got interpreted inclusively in 57% of the cases. Also, if the disjuncts were both interpreted as a sufficient condition, it is common sense that in the examples above, someone who both photographs and touches the objects has obviously broken the rule. So if the disjuncts in the rules of Evans & Newstead were interpreted as sufficient conditions, the 57% 'true' judgments make sense.

type was agreed on, signaling that in 59% of the cases participants calculated the *not all* SI. Above I already indicated that this result is expected on the QUD Focus Condition: As no context is given, it is impossible to distill the QUD. Participants could only guess a QUD, accommodate this QUD and answer based on that. QUDs like *Do some or all elephants have trunks?* are possible, on the basis of which (according to the QUD Focus Condition) an SI should be calculated, as *some* is in the answering part of the sentence. However, QUDs like *Do some elephants (that are for some reason salient) have trunks?* or *What do some elephants have for grabbing stuff?* are also possible, and as here *some* is not in the answering part, no SI is predicted by the QUD Focus Condition. We can therefore claim that participants in this experiment were guessing about what the QUD is, resulting in percentages around 50%.

Chierchia et al. (2001)
Another study I already referred to above is the one by Chierchia et al. (2001). Their Experiment 1 consisted of a story that was acted out, followed by a target sentence that contained *or* in the restrictor of a universal. Chierchia et al. describe a typical trial the following way :

> In a typical trial, children were told a story about Snow White and four dwarves at a picnic. When the dwarves are ready to choose some food, Snow White invites them to choose healthy food, and she promises a jewel to all the dwarves who will choose healthy food. She reminds them that bananas and strawberries are healthy food. Three of the dwarves want to receive a jewel, so they choose fruit. But they are very hungry, so they choose both a banana and a strawberry. One of the dwarves says he doesn't care about jewels, and he chooses potato chips. Snow White only gives a jewel to the dwarves who have chosen a banana and a strawberry. At the end of this story, a puppet produced the following target sentence:
>
> "Every dwarf who chose a banana or a strawberry received a jewel."

As the story is quite complex, it is not straightforward to decide what the QUD is. The most salient thing seems to be the rewarding of the dwarves that chose healthy food. A possible QUD along these lines is the yes/no question: *Did every dwarf that chose a banana or a strawberry indeed receive a jewel?* Of course, as an answer to this QUD, *or* is in the given part, not in the answering part (which would actually be the implicit *yes*), so no SI is predicted. We could try to come up with a QUD that would put *or* in the answering part, for instance: *Which dwarves received a jewel?* However, in the target sentence, which would be an answer to this QUD, *or* is in the restrictor of *every*, which is a DE-environment. As we saw in section 2.1, DE-environments block the *not and* implicature of *or*, based on flipping of the scale (neo-Griceans) or application of the Strength Condition (Chierchia). So whatever the QUD is here, no SI is predicted. Not surprisingly, Chierchia et al. found 95.5% 'true' answers among their adult control group (N=11).[28]

In section 2.3.1 I already showed that the items in Chierchia et al.'s Experiment 2 only gave rise to one possible QUD, which gave rise to the SI. Their Experiment 3 seems to suffer from the same problem. Instead of using a TVJT, Chierchia et al. conduct what they themselves call a *Felicity Judgment Task*. Like in Experiment 1, a story was acted out. After the story, two puppets both uttered a different sentence that was meant as a description of the story. Participants were asked to reward the puppet that 'said it better' with a coin. Here's Chierchia et al.'s description of a trial :

[28] This is exactly what Chierchia et al. expected, as one of their aims of this experiment was to show the absence of this SI in a DE-environment.

> In one trial, children were told a story about some farmers who were cleaning their animals. After looking at all of the animals, each farmer decided to clean a horse and a rabbit. At this point, the two puppets provided an alternative description of the story:
>
> (1) Every farmer cleaned a horse or a rabbit.
> (2) Every farmer cleaned a horse and a rabbit.

As no judgment had to be made but only a choice between two utterances that only differ in felicity, the name *Felicity Judgment Task* is plainly wrong, so I will call this task a *Felicity Choice Task*. A problem with this task is that immediately the undesirable QUD *Did every farmer clean a horse or a rabbit or did every farmer clean a horse and a rabbit?* arises.[29] That QUD only makes sense under an exclusive interpretation of *or*, leaving (2) as the only true statement and therefore the better answer. Even if we ignore this QUD, the trial has the same problem I pointed to above for Experiment 2. Next to the undesirable meta-QUD above, the description of the trial only leaves open an object-oriented QUD like: *Which of his animals did every farmer clean?* The whole story is pointed at the farmers' choice: there is a contrast set of animals from which some are chosen and some are not, explicitly introduced in *After looking at all of the animals*, but there is no contrast set of farmers. The farmers are never distinguished, so a subject-QUD like *Who/Which farmer cleaned a horse or a rabbit?* will not arise. Relative to the only possible QUD of the trial, the object-oriented QUD, the connective (*or* in (1) and *and* in (2)) is in the part that answers the QUD, so the part where SIs are calculated. Obviously, participants will prefer *and* over *or* here: (1) with SI is simply untrue in the given situation. Chierchia et al. did not have an adult control group in this experiment, but even the tested children (N=15) did well, rewarding the puppet that uttered the sentence with *and* in 93.3% of the trials.

Gualmini (2001)

Another interesting set of experiments is described by Gualmini (2001). He wanted to demonstrate that in a DE-environment, such as the one triggered by the quantifier *none (of the)*, SIs do not arise.[30] In his Experiment 1 and 2, he presented participants with stories like the following summarized example: *Three monkeys are looking for a banana and a frisbee. Two of them found neither. One found both.* The target sentence was: *None of the monkeys found the banana or the frisbee*. Gualmini reasons that if *or* gets the *not and* SI, like it does in non-DE contexts, the complete meaning of this sentence would be: *None of the monkeys found either the banana or the frisbee*. In the given story, this sentence is true: none of the monkeys found only one of the two. However, the adult group (N=11) judged the sentence 'untrue' in 100% of the cases, signaling an inclusive interpretation of *or*. So the *not and* SI is indeed absent in this context, as was expected in a DE-environment. On the QUD-based approach, there is another reason why we would never expect an SI here: a sentence starting with the quantifier *none (of the)* can only be an answer to a subject-oriented QUD like: *Which of the monkeys found the banana or the frisbee?* or *How many of the monkeys found the*

[29] Notice that in such a question, *or* and *and* automatically get stress (to contrast them with each other). Usually, when *or* itself is stressed, this is a signal that it should be interpreted exclusively. This type of stress should not be confused with (intonational) focus, which I use as an indication of QUD-answerhood: there, the whole constituent has focus, not just the scalar term.

[30] For some reason, Gualmini thinks this is a "puzzle for the neo-Gricean view", and therefore support for the Semantic Core Model. However, we saw in section 2.1 that neo-Griceans like Atlas & Levinson (1981) and Horn (1989) can perfectly explain this as the entailment scale is reversed in a DE-environment.

banana or the frisbee?.[31] The target sentence is not a good answer to a object-oriented QUD like: *What did the monkeys find?* Relative to a subject-oriented QUD, *or* is not in the part of the sentence that answers it, so the QUD Focus Condition predicts no SI anyway.

In Experiment 3 and 4, Gualmini tries to find evidence for the flipped <and, or> scale in a DE-environment. Experiment 3 had a different task than Experiment 1 and 2: subjects were asked to reject not just sentences that were false, but also sentences that were infelicitous, I will call this a *Truth/Felicity Judgment Task*.[32] Gualmini describes a trial in which three polar bears get to choose from a set of three skateboards, a basketball and a frisbee. Eventually, each polar bear chooses a skateboard and nothing else. The target sentence is: *Every polar bear chose a skateboard, but none of the polar bears chose the frisbee and the basketball*. As we saw in section 2.1, under a DE-operator like *none*, the <and, or> scale is flipped: *or* is the more informative connective: while *and* only rules out the situation in which both conjuncts are the case, *or* rules out the situation in which both disjuncts are the case and the situations in which one of the disjuncts is the case. Gualmini expected participants to reject the target sentence based on that: the same sentence with *or* instead of *and* is also true for this story and it is more informative and therefore more felicitous. However, adults (N=15) judged the sentence 'okay' in 90% of the trials.

In Experiment 4, Gualmini conducted a Felicity Choice Task with the same type of story and again a target sentence with *none* (so comparable to the one above). Participants were asked to choose between the sentence with *and* and the sentence with *or*. Again, adult participants (N=16) did not behave as expected: they only had a preference for *or* in 41% of the cases, and had no preference at all in 50% of the cases. Gualmini is puzzled by these results, he says: 'Such results show that adults refrain from applying scalar implicatures in particular experimental contexts' and: 'This leaves us with no experimental evidence that the information strength associated with alternative representations of the target sentence yields a reversed scale in downward entailing contexts.' There is however one important thing that Gualmini overlooked. As for instance Szabolcsi & Haddican (2004) discuss, *A and B* embedded under negation, can also mean *neither*. They even claim that for unstressed *and*, this interpretation is preferred over the *not both* reading that Gualmini assumed. Szabolcsi & Haddican give the following examples:

(3) Mary didn't take hockey AND algebra.
'not both'

(4) Mary didn't take hockey and algebra.
(i) 'neither'
?? or * (ii) 'not both'

So for stressed AND in (3), *not both* is the only possible interpretation. However for unstressed *and* in (4), *neither* is the preferred interpretation, and *not both* is even judged marginal or unacceptable. This sheds new light on the findings of Gualmini. As *not...and* in his target sentence was interpreted as *neither*, this is the same meaning as the sentence with *or*. No wonder participants accepted it in Experiment 3, and no wonder that participants had

[31] The yes/no question *Did any of the monkeys find the frisbee or the banana?* is also a possible QUD. Here, *or* is also not in the part of the sentence that answers the QUD.
[32] It is questionable whether this task is something that can be asked of ignorant participants, who might not fully grasp the concept of felicity. Should they for instance also reject sentences that are ugly or strange, or sentences they themselves would not use?

no preference in 50% of the cases in Experiment 4: what preference is there to have between two equivalent sentences? The 41% preference for *or* might be caused by participants taking *and* to be stressed AND, which is not unlikely as it automatically get stress if it is contrasted with the same sentence with *or* (see fn. 29). However, Gualmini does not mention how the sentences were presented so whether this is possible.

Papafragou & Musolino (2003)
Experiment 1 of Papafragou & Musolino (2003) resembled the one by Noveck (2001) in so far that the target sentence contained the scalar item *some*, where the stronger *all* was also true. But contrary to Noveck's experiment, here a context story was provided, and acted out before the subjects. So the subjects saw for instance four toy horses, which all jumped over a fence. After that, a puppet said: *Some horses jumped over the fence* and the subjects were asked to tell whether she "said it well". So this experiment is really a *Felicity Judgment Task* (contrary to Chierchia et al.'s Experiment 3, which was actually a *Felicity Choice Task*). Contrary to Noveck's study, here the adults (N=30) rejected the sentence in 92.5% of the times, signaling that the SI was calculated. Of course, this experiment might yield a different result than Noveck's because here not truth but felicity was judged. But the main difference is that here a context, and therefore a QUD, was provided. Notice that neither the horses nor their actions were distinguished, and neither have a contrast set in the story, so QUDs like *Which horses jumped over the fence?* or *What did some horses do/jump over?* are impossible. It seems there is only one thing in this context that is relevant, and that is the number of horses that jumped over the fence. The QUD is: *How many (of the) horses jumped over the fence?* Relative to this QUD, *some* is in the QUD-answering part so an SI is expected.[33]

Guasti et al. (2005)
Guasti et al. (2005) first tried to replicate the findings of Noveck (2001) in their Experiment 1. Their adult control group (N=19) calculated an SI in 50% of the cases, which is comparable to the 59% of Noveck, and is typical of the guessing behavior participants exhibit in the absence of context (and therefore absence of QUD). Guasti et al.'s Experiment 4 resembled Papafragou & Musolino's (P&M) experiment I just discussed. After watching a video in which a story was acted out, participants were asked to judge whether a sentence uttered by a puppet was 'right' or 'wrong' (Felicity Judgment Task). Just like in P&M's experiment, sentences contained the scalar *some*, while the story also allowed the stronger *all*. The percentage of SIs of the adult control group (N=12) was 83%, a little lower than P&M's 92.5%. Just like in P&M's study, the only possible QUD in the trials is one that questions the number of Ns that VP: *How many (of the) N VP?* (or: *Did all N VP or did some N VP?*). So again, relative to every possible QUD, *some* is in the part of the sentence that answers the QUD, so the SI is no surprise on the QUD Focus Condition.

Papafragou & Tantalou (2005)
Papafragou & Tantalou (2005) criticize earlier TVJT experiments on SIs. They claim the task brings along three problems: Firstly, the experimental setup makes it impossible to determine whether the potential SI should be considered as part of what the speaker intended to communicate. In many of the above TVJT experiments, the puppet whose utterances have to be judged, is introduced as an incompetent speaker. Therefore it is impossible to tell whether the SI was intended by the speaker or not. Secondly, as the puppet is usually only asked to "describe what happened", there is no specific expectation of informativeness. This runs

[33] The obvious contrast between *all horses* that the subjects saw jump over a fence and the target sentence that says *some horses* might also trigger the QUD: *Did some or all horses jump over the fence?* Of course here, *some* is also in the part of the sentence that answers this QUD, so the QUD Focus Condition also expects an SI.

parallel to my QUD approach: in many of the TVJT experiments, it is unclear what the main issue is the target sentence is supposed to answer. Thirdly, as the target sentences are removed from natural conversation, participants have to estimate the experimenter's goals to decide whether or not they should let the SI be decisive in the truth value judgment. I think the first and the third point of criticism add up to the same thing: a participant in a TVJT can deliberately choose to include or exclude SIs in his truth value judgment. He will do so based on an estimation of the intentions of the speaker (or the experimenter). I will return to this important point below in Section 3.3.4, when I discuss the shortcomings of a TVJT as a diagnostic for SIs.

Papafragou & Tantalou designed a new experiment that they claim does not have these shortcomings. Their experiment goes as follows: A story is acted out in which it is made explicit that someone (e.g. a toy bear) is expected to complete a task (e.g. color all the stars on a drawing). The bear takes off and comes back after a while. The experimenter asks: *Did you color all the stars?* The bear says: *I colored some.* The participants in this experiment were instructed to reward the bear if he met the expectations. As the experiment is set up with a natural conversation and there is a clear expectation, a better estimate of the speaker's (in this case the bear's) intentions can be made by the participants. Also, the speaker has a natural motivation to use a weaker form: as he was unable or unwilling to complete the task, he just reported his partial progress. There are no adult data on this experiment, but 4- to 6-year old children (N=10) withheld the reward, signaling calculation of an SI, 77.5% of the time. So let us see if the QUD-approach also makes this prediction. The QUD in this case is explicitly stated: *Did you (the bear) color all the stars?* This is a yes/no question. Normally, a yes/no question can only be answered by 'yes' or 'no', or an affirmative echoic declarative such as: *I colored all the stars*, which is an implicit 'yes'. However, what is happening here seems to be a special mechanism. A part of the presupposed part of the sentence is changed in the answer: *all* is replaced by *some*, giving rise to an implicit 'no'-answer. This can only be done by giving stress to *some*. This stress indicates that this quantifier (*some*) is contrasted with the one in the question (*all*). Now of course, contrasting it with *all* already gives us the *some but not all* interpretation of *some*, which would otherwise arise by an SI. Therefore I will consider these cases in which the scalar term itself is stressed as the result of a separate mechanism. I will ignore them and focus on the cases where the scalar itself does not get separate stress (see also fn. 29).

Feeney et al. (2004) Experiment 1 and 2

Feeney et al. (2004) start with a replication of Noveck (2001). They found 35% SIs for their adult control group (N=32). Surprisingly, they did not find a much lower percentage of SIs for 7-8 year old children. While Noveck found only 11% SIs for this group, Feeney et al. found 34%, a percentage not significantly different from the adult control group. Feeney et al.'s Experiment 2 shows that with adding context, the percentage SIs for the children even goes up to 79%. They used contexts in which the SI is highly relevant. For instance a context in which a little girl has eaten all three of the sweats that were on a plate. Her mother asks: *What have you been doing with the sweets?* The girl answers: *I've eaten some of the sweets*. However, just like in the experiment above by Papafragou & Tantalou (2005), something in the presupposed part of the QUD is changed, from *the sweets* to *some of the sweets*: Obviously, as the mother is standing at the empty plate, the QUD is about *all* of the sweets, so the use of *some* in the answer is marked. This indicates that the changed item should be contrasted with something, the most relevant contrasting thing being *all of the sweets* that the QUD was about. That the same mechanism is at work here (as in Papafragou & Tantalou (2005)) is also reflected by the fact that in order for the sentence to be a good answer to the

mother's question, the scalar *some* has to get stress: *I've eaten SOME of the sweets*.[34] Like Papafragou & Tantalou's items, I think the occurrences of SIs that are the result of stress on the scalar term itself should be considered as a separate mechanism.

Janse et al. (2005)

Another experiment I already discussed in the previous chapter was the experiment we conducted about a year ago (Janse et al. (2005)). This study was originally an eye-tracking study, but the percentage SIs turned out surprisingly low at 17%. As I already pointed out, this might have been an effect of the specific setup of the experiment: Participants were not able to look back at the story they just read, so they had to keep it in mind. Insecurity about the story might have caused them to accept more sentences that were untrue if the SI was calculated. Also, although there was no explicit time-pressure built into the eye-tracking procedure, participants might have felt less free to take as much time as they would have in the experiments I discussed above. As results by e.g. Bott & Noveck (2004) suggest, participants are less likely to calculate SIs if they are under time pressure. This might have been another factor that made the percentage of SIs drop. But the most interesting property of the Janse et al. study is that here, contrary to many of the experiments above, (at least) two QUDs were possible that differed in which part of the sentence was questioned. Consider for example one of the trials:

> (5) Story: *Three wizards come to a meeting in the forest at night. It is dark so they brought torches and oil lamps.*
> *The first wizard has a moustache. He brought a torch and an oil lamp.*
> *The second wizard also has a moustache and he also brought an torch and an oil lamp.*
> *The third wizard has no moustache, and he brought nothing.*
>
> target sentence:
> Elke tovenaar die een snor heeft, heeft een fakkel of een olielamp bij zich.
> *Every wizard that a moustache has, brought a torch or an oil lamp (with him)*
> "Every wizard that has a moustache, brought a torch or an oil lamp."

In for instance Chierchia et al. (2001)'s Experiment 2 and 3, we saw that the subject set was never distinguished and the only relevant thing was the object set, resulting in only object-oriented QUDs. Here, contrary to Chierchia et al.'s story, the members of the subject set are introduced separately and distinguished (by the property of having a moustache or not). Also, the target sentence picks out two wizards from the subject set, namely the wizards with a moustache. So this combination of story and sentence does gives rise to the subject-oriented QUD: *Which wizards brought a torch or an oil lamp?* This QUD can be paraphrased by using a superset for *torch* and *oil lamp*: *Which wizards brought lights?* This question clearly indicates that it is irrelevant what exactly the different wizards brought, the only thing that is important is which wizards were the ones that brought any lights at all. But the object-oriented QUD *What did every wizard that has a moustache bring?* is still an option. Only on this QUD is *or* in the part of the sentence that answers the QUD and will an SI arise according to the QUD Focus Condition.[35] So the fact that (at least) two QUDs are possible here explains

[34] It is possible to utter this answer without extra stress on the scalar. However, then another constituent, such as *eaten,* has to get extra stress: *I've EATEN some of the sweets*. But then, *eaten* is contrasted with something else, e.g. *thrown away*. This is obviously not the right interpretation of the sentence.

[35] For simplicity, I will ignore the pair list question *Which wizards brought what?* The target sentence cannot really be interpreted as a pair list, unless the part *and the wizard without a moustache brought nothing* is added. It might be claimed that this part is implicated, so for that reason I will exclude this QUD in my experiment by using *only* to trigger focus on one constituent.

why we did not find the close-to-100% SIs that most of the experiments above (that had a context story) found.[36] A cause of the QUD-ambiguity of the trials of Janse et al. is the fact that the written sentence allows several focus constructions. As I pointed out earlier, focus can act as a signal of QUD-answerhood. In the experiment I will present in section 3.3, I will use the focus operator *only* to trigger a certain focus construction and with that only leave one possible QUD.

A remark on possible cross-linguistic differences
Above I ignored the fact that the experiments I just discussed, were conducted in different languages: French (Noveck (2001)), English (Chierchia et al. (2001), Gualmini (2001)), Greek (Papafragou & Musolino (2003), Papafragou & Tantalou (2005)), Italian (Guasti et al. (2005)), and Dutch (Janse et al. (2005)). There is of course a possibility that cross-linguistic differences make a clean comparison impossible. Although it is an interesting question whether SIs differ cross-linguistically, I will not develop this argument further. I will assume that SIs work the same in (at least these) different languages.

3.2 Time span studies

All of the above studies take the answers of the participants on some version of the TVJT to assess in which environments SIs arise and in which environments they do not. Another experimental path that has been followed in the last couple of years has as its aim to settle the debate between the defaultist and the contextualist view by considering the time span in which subjects read or judge sentences with scalar terms. The starting point of this approach is that the defaultist view and the contextualist view differ in their explanation of what happens in contexts in which a scalar term is present but no SI is calculated. Defaultists claim that the SI initially is calculated, but it is for some reason canceled afterwards (for instance because it leads to contradiction with following material). Contextualists claim the SI is not calculated in the first place (for instance because it is not relevant the context). The assumption of this line of research is that the calculation and cancellation of SIs both have a processing cost that should yield an observable delay in processing of the sentence. On this assumption, defaultists would predict that sentences in which an SI is calculated and canceled should take longer to process than sentences in which the SI is calculated and not canceled, which should take longer to process than comparable sentences without a scalar term. Schematically, the defaultist prediction looks like this ('<' meaning 'faster than')

(6) sent. without scalar < sent. with scalar and SI < sent. with scalar and no SI (canceled)

Contextualists predict that only sentences in which the SI actually arises are slower. There should be no difference in processing time between a sentence without a scalar and a sentence with a scalar that (for contextual reasons) gets no SI, as in the latter the SI is not calculated in the first place. Schematically ('=' meaning 'no difference in processing time'):

(7) sent. without scalar = sent. with scalar and no SI < sent. with scalar term and SI.

[36] That the percentage of SIs is even lower than around 50% might be the result of participants preferring the subject-oriented QUD to the object-oriented QUD. A possible explanation for this is the following: The subject of the target sentence actually picks out two of the three wizards (and excludes the one without a moustache), the object of the target sentence does not exclude an object, both objects (torch and oil lamp) are included in the sentence (although under disjunction). In my own experiment I will make sure this picking out (and excluding other members of the reference set) only happens in the constituent that is questioned by the QUD.

In this section I will discuss some of the experimental studies in which the time span of scalars is measured with this difference in predictions between the defaultist and the contextualist view in mind. I will discuss them as this line of research has proven quite popular the last couple of years.

Bezuidenhout & Morris (2004)
Bezuidenhout & Morris (2004) (B&M) conducted an eye-tracking study on sequences like (8) and (9):

(8) Some books had color pictures. In fact all of them did, which is why the teachers liked them.
(9) The books had color pictures. In fact all of them did, which is why the teachers liked them.

Sentence (8) is a classical cancellation context, where on the defaultist view the second sentence cancels the *not all* SI of *some*. In (9), *some* is replaced by *the*, which obviously does not trigger a *not all* SI, so no cancellation happens in the *in fact* sentence. B&M claim that according to the defaultist view, the cancellation of the SI should lead to a delay in the processing of the second sentence of (8), which will not be present in processing (9). They state this delay is expected on the *them did* region, as not before that point it is clear the SI is canceled. B&M adopt a contextualist view that is based on underspecification (they call this model *UM*): readers will keep open both meanings of a scalar (with and without SI), and decide based on later contextual evidence which meaning is taken.[37] So in (9), when readers encounter *all*, this is potential contextual evidence about which meaning of *some* was meant. B&M claim that UM therefore predicts a delay in reading time on the word *all*, that will not be present in processing (9). Their results do show a delay on *all*, but not on *them did*. B&M claim this to be evidence for UM. A weak point of this study however is the assumptions that were made about the regions in which the delays are expected. One could easily argue that the cancellation process of the SI is already triggered by *all*, or even by *in fact*. In that case, the delay on *all* can be taken as evidence for the defaultist view.

Bott & Noveck (2004)
Bott & Noveck (2004) (B&N) measured reaction times of judgments on the sentences of the type of Noveck (2001) like *Some elephants have trunks*. B&N call these sentences *underinformative*, as the more informative *all* could have been used. Just like in Noveck (2001), no context is provided. In their Experiment 1, B&N explicitly instructed participants to in one session interpret *some* as *some and possibly all* (the 'logical' meaning (without SI), yielding a 'true'-answer) and in the other session as *some and not all* (the 'pragmatic' meaning (with SI), yielding a 'false' answer).[38] The defaultist and the contextualist view differ on what happens in the logical condition: According to the defaultist view, the SI is calculated and has to be canceled, while according to the contextualist view, the SI is not calculated in the first place.[39] So the defaultist view expects longer reaction times on the logical condition (SI calculation and SI cancellation) than the pragmatic condition (only SI

[37] I will not discuss this view further here. This section only has as its aim to show how time span studies can be used to tease apart defaultist and contextualist predictions.
[38] Of course, the order of sessions was counterbalanced over subjects: half of them got the logical session first and then the pragmatic session, and the other half got the sessions in reversed order.
[39] It is questionable whether participants calculate the SI if they are beforehand told not to, but I will ignore this for the sake of B&N's argument.

calculation), while the contextualist view expects longer reaction times on the pragmatic condition (SI calculation) than on the logical condition (no SI calculation). These predictions were already given above in (6) and (7), only in slightly different terms. The words of the sentence were presented one by one on a computer screen, reaction time was measured from the presentation of the last word. B&N found longer reaction times for the pragmatic condition, supporting the contextualist view.

However, as B&N themselves note, this difference in reaction times might be caused by the fact that the pragmatic condition gave rise to a 'false' answer, while the logical condition got a 'true' answer, and it has often been claimed that 'true' answers are faster than 'false' answers. B&N try to correct this in their Experiment 2, where they precede the sentences with: *Mary says the following sentence is true/false*, in such a way that both conditions give rise to a 'agree' response. So the pragmatic condition always got *Mary says the following sentence is false*, while the logical condition always got *Mary says the following sentence is true*. However, it can still be claimed that it takes more time to agree on a statement that includes negation than one that does not. In Experiment 3 B&N leave the instructions behind, taking the spontaneous answers of the participants as indication of logical or pragmatic interpretation. Again they found that answering pragmatically, so with SI, takes longer than answering logically. This supports the contextualist view, although the response bias (comparing 'true' to 'false' answers) is still present. Below I will discuss how Feeney et al. (2004) try to avoid this bias.

In Experiment 4, B&N varied the time available to the participants to judge a sentence. Half of the participants was given a relatively long time per item to respond, and the other half got a shorter time (the mean reaction time of the logical condition of Exp. 1). After every item, the participant was told whether he answered in time or not. Only responses that were given in time were included in the analysis. It turned out that SIs (so pragmatic responses) went down from 44% in the long condition to 28% in the short condition. B&N take this as evidence for the contextualist view, as this view claims SIs are costly, non-automatic inferences.

Feeney et al. (2004) Experiment 3

In Section 3.1 I discussed two TVJT studies of Feeney et al. Following those, they also performed an interesting time span experiment. They repeated Bott & Noveck's Experiment 3, but instead of focusing on the reaction time difference between logical and pragmatic responses to the underinformative *Some elephants have trunks*-sentences, they compared the reaction times of logical responses to these sentences to logical responses to the felicitous *Some birds live in cages*-sentences. As the logical response to both sentences is 'true', the response bias is no longer present. Feeney et al. found that the logical answers (so: 'true') to the underinformative sentences take longer than a 'true' answer to the felicitous sentences. These data can easily be explained by the defaultist view: In both sentence-types an SI is calculated because a scalar is present. On the felicitous sentence the SI can stay as it is true under a *some but not all* interpretation. However, as the logical answer indicates, in the underinformative sentences it is canceled, which causes the longer reaction time. The contextualist view cannot explain this: According to this view, both sentence-types do not get an SI: for the felicitous sentences the SI is irrelevant: the sentence is true both with and without SI, so it will not get one, and the 'true' answer to the underinformative sentence signals that it did not get an SI. So we could interpret Feeney et al.'s findings as evidence for the defaultist view. Feeney et al. themselves however, interpret these data in a completely different way: they claim that the delay is due to the fact that the subjects that answer logically to the underinformative sentences deliberately resist the SIs. They refer to work on

decontextualization: separating the logical form of what one is reasoning about from its content (e.g. Stanowich (1999)). This ability is associated with general intellectual ability. Therefore, Feeney et al. also conducted a counting span task on their subjects and found a correlation between memory span and number of logical responses to underinformative *some* statements. Feeney et al.'s idea of deliberately resisting of SIs is similar to my explanation in terms of QUD-guessing: Participants might deliberately resist SIs because they are guessing about the intentions of the experimenter. As the task does not provide them with a clue about what is at issue here, they might guess that a logical answer (so without SI) is preferred, just as they might guess a QUD that would trigger a logical answer is intended. It is therefore important to make sure that the QUD is clear in the items. I will return to the issue of deliberately ignoring SIs in section 3.3.4, when I discuss the downsides of the TVJT as a tool to assess SIs.

Breheny et al. (2005)
Breheny et al (2005) conducted a self-paced reading experiment, which is a redo of an experiment by Bezuidenhout & Cutting (2002). They compared reading times on segments that included the scalar or in so called 'upper-bound' contexts to 'lower-bound' contexts. See e.g. (10) and (11) (slashes indicate self-paced reading regions):

(10) Upper-bound context:
John was taking a university course / and working at the same time. / For the exams / he had to study / from short and comprehensive sources. / Depending on the course, / he decided to read / the class notes or the summary. /

(11) Lower-bound context:
John heard that / the textbook of Geophysics / was very advanced. / Nobody understood it properly. / He heard that / if he wanted to pass the course / he should read / the class notes or the summary. /

In (10) it is relevant whether John decided to read both the class notes and the summary, or just one of the two. One could for instance think of a question like: *Did John read both the class notes and the summary for each course?* However in (11) it is irrelevant whether he would read both: it only states what he should *at least* do. According to Breheny et al., lower-bound contexts like (11) exclude SIs. On the defaultist view, the SI in (11) would therefore be calculated but canceled, resulting in longer reading time for (11) than (10), in which the SI is not canceled. This prediction was already given in (6) above. Contrastively, the contextualist view claims the SI in (11) never arises, so (10), which does give rise to an SI, would yield longer reading times (see (7) above)). Breheny et al. did indeed find longer reading times for (10) than for (11), supporting the contextualist view. However, as they themselves point out, this result is the opposite of the result of Bezuidenhout & Cutting (2002). Breheny et al. blame this on the scalar items of Bezuidenhout & Cutting that included cardinal numbers, the status of which as scalar items is a topic of discussion in the literature.

In this section some studies were addressed that try to settle the defaultist-contextualist debate by measuring time spans of sentences with scalars. Although some of the results look promising, these studies have their downsides: they rely on indirect data: differences in reaction times or reading times can be caused by a huge number of factors, which are hard to control for. Especially when reaction times on a TVJT are measured, the whole decision making process might involve many steps that can cause a delay on one condition and not on the other (think of the response-bias of Noveck & Bott's study). There is no guarantee that

this delay can be attributed to the SI calculation or cancellation only. Next to that, the assumptions that SI-calculation and cancellation both bring along a processing cost that causes an observable delay is not an innocent one: perhaps the processor has an overcapacity that can easily handle this extra processing cost, or perhaps certain shortcuts are at work that decrease the extra processing cost so that it does not create a delay. The aim of this section was to provide some insight in the work that is done on SIs using time span data. In the next section, I will present an experiment that will test the QUD-based approach to SIs.

3.3 An experiment testing the QUD-approach

3.3.1. Introduction

In this section I will present an experiment designed to test the main hypothesis of this work, the QUD Focus Condition. In section 2.3.1, I put together three contextualist approaches to SIs, claiming they all make the following prediction.

(12) **The QUD Focus Condition for Scalar Implicatures**:
An SI will arise in a sentence iff the scalar term (with which the SI is associated) is in a constituent that answers the QUD of the context that the sentence is part of, and therefore has focus.

The last part of this condition was added in section 2.3.2, where I discussed the fact that the part that answers the QUD is signaled by focus. In the same section, I presented the predictions of the QUD Focus Condition for the *not and* SI of *or*: The QUD Focus Condition predicts that in (13) no SI will arise, while in (14) it will:

(13) Q: Who/which man had A or B?
A: $\underline{\text{John}}_F$ had A or B.

(14) Q: What/which weapon does John have?
A: John has $\underline{\text{A or B}}_F$

So how can we test these predictions? Let us first recap how the QUD Focus Condition explained the results of the experiments we saw in section 3.1. There were some experiments in which no context at all was provided, like the one of Noveck (2001). These typically resulted in around 50% SIs, possibly because participants were guessing the QUD. Then there were a couple of experiments that did provide a context story, but the story and target sentence only allowed one possible QUD. If the scalar in the target sentence was in the answering part (e.g. in Chierchia et al. (2001)'s Experiment 2 and 3, Papafragou & Musolino (2003)'s Experiment 1, Guasti et al. (2005)'s Experiment 4), this resulted in SI-scores for adults close to 100%. If the scalar was not in the part of the sentence that answered the only possible QUD, SI scores went down and approached 0% (e.g. Gualmini (2001)'s Experiment 1 and 2). [40] Some experiments found that while *or* seemed not to be in the part that answered the QUD, SIs still were calculated (e.g. Papafragou & Tantalou (2005), Feeney et al. (2004)'s Experiment 2). However, these experiments relied on a special mechanism in which the scalar itself is stressed to mark its contrast with the QUD. Finally, in a study by Janse et al. (2005), each item allowed for two possible QUDs, for one of which *or* was in the answering part and

[40] Note that in Gualmini (2001), the 0% SIs was also predicted by the DE-context.

for the other it was not. Because the target sentences were presented in written form, they allowed both focus constructions, corresponding to the two possible QUDs. So focus could not disambiguate the QUD in this experiment. The low percentage of SIs (17%) suggested that for some reason, participants preferred the subject-oriented QUD over the object-oriented QUD.

If we take the above results into consideration, it seems what we need to test the QUD Focus Condition is an experiment in which, like in (13) and (14), the QUD is varied while the target sentence is the same. Relative to one QUD *or* should be in the QUD-answering (thus focused) part and relative to the other QUD it should not. If we then find SIs for the sentence with the first QUD and we do not find SIs for the same sentence with the second QUD, this is evidence for the QUD Focus Condition. However, if we want to compare (and contrast) our results to the results of the experiments discussed in section 3.1, we have to provide context stories (and not just explicit questions like in (13) and (14)) that only trigger the SI we want it to trigger. Thereby, we have to make sure that for each item, the only possible focus construction of the target sentence is the one that is compatible with our intended QUD. In the next section I will show how I tried to achieve these two goals for the experimental items.

3.3.2 The experiment

Items

How do we get a context story to trigger a certain QUD and no other? We saw that one of the reasons that e.g. Chierchia et al. (2001)'s context stories only gave rise to one possible QUD was that only the members of one set (in their case the object set) were introduced separately and that it was only from this set that the target sentence picked out members and excluded others. The description of one of their items is repeated here:

(15) In one trial, children [and adults, AZ] were presented with a story about four boys at the summer camp who are choosing which toys they want to play with. There are a lot of toys they can choose: some skateboards, some bikes, a boat and a truck. After considering the possible choices, the four boys [each, AZ] take both a skateboard and a bike. At this point, a puppet produced the following target sentence:

"Every boy chose a skateboard or a bike."

As is immediately clear, the part of the target sentence where the action is, is the object: While the undifferentiated subject set is completely included in the target sentence (every boy), two members of the object set are picked out (a skateboard and a bike), while two others are not picked (a boat and a truck). The only possible QUD therefore is object-oriented: *What did the boys pick?* For our experiment, we want to have two different context stories for each target sentence: combined with the one context story the target sentence should give rise to a subject-oriented QUD (as in (13)) and combined with the other story it should give rise to an object-oriented QUD (as in (14)). So we have to make sure that in the first case, only members of the subject set are introduced separately in the story and picked from by the target sentence, and in the second case the same should apply only to members of the object set.

Our second goal, to get the focus construction right for the QUD we are aiming for, would be simple for an experiment in which the items are presented orally. We can use the main stress of the sentence: According to e.g. Reinhart (1995), the possible foci of a sentence are the

constituents that contain the main stress.[41] So if we simply put the main stress on the subject of the target sentence in the condition in which we want to trigger a subject-oriented QUD, and on the object in the object-oriented QUD condition, we should be home free. However, for practical reasons, the experiment was carried out in written form. Therefore, I decided to use the focus-sensitive operator *only*. The standard assumption is that the focus of the sentence is in the c-command domain of *only* (see e.g. Reinhart (2004)). So if we put *only* in the subject NP or the object NP, this constituent automatically gets focus.[42] To see how this works, consider the versions of (13) and (14) with *only*:

(13') Q: Who/which man had A or B?
A: Only John had A or B.

(14') Q: What/which weapon does John have?
A: John has only A or B.

Sentence (13'A), in which the subject contains *only*, can only be an answer to the subject-oriented (13'Q), not to the object-oriented (14'Q), so the subject has focus. Similarly, sentence (14'A), in which the object contains *only*, can only be an answer to the object-oriented (14'Q), not to the subject-oriented (13'Q), so the object has focus. So the only thing we have to do to get the focus right for the QUD we are aiming for is to put *only* in the right constituent.

So one can trigger a particular QUD in a context story by only distinguishing and picking out members of one set and get the focus construction right by using *only*. Exploiting these techniques, I came up with the following skeletons for the two conditions (16a-b), below each skeleton the situation of the context story is schematized:

(16a) **Condition 1 (Subject focus):**
A and B agreed to (both) verb set.
A verbed setmember1. A also verbed setmember2.
B usually verbs too, but (for reason x) this time B did not verb any of the set.
target: Only A verbed setmember1 or setmember2.

A	1+2
B	∅

(16b) **Condition 2 (Object focus):**
A promised to verb set.
A verbed setmember1. A also verbed setmember2.
A intended to verb setmember3, but (for reason x) did not.
target: A verbed only setmember1 or setmember 2.

A	1+2
∅	3

[41] I will not go into the projection properties of focus, which is a matter of debate (see Reinhart (2004)).
[42] As *only* is inside the NP, its c-command domain is just the rest of the NP.

An example of a story pair according to these skeletons is (17a) and (17b). The experiment was conducted in Dutch[43]:

(17a) Harry Potter en Hermione hebben afgesproken vandaag allebei wat verlichting mee te nemen naar de bijeenkomst in het bos. Harry heeft een fakkel meegenomen. Hij heeft ook een olielamp meegenomen. Hermione is meestal zeer betrouwbaar, maar dit keer is ze vergeten ook maar iets van verlichting mee te nemen.
target: Alleen Harry heeft een fakkel of een olielamp meegenomen.

Harry Potter and Hermione agreed to both bring some lights to the gathering in the woods today. Harry brought a torch. He also brought an oil lamp. Hermione is usually a very reliable person, but this time she forgot to bring any lights at all.
target: Only Harry brought a torch or an oil lamp.

(17b) Harry Potter heeft beloofd dat hij vandaag wat verlichting mee zal nemen naar de bijeenkomst in het bos. Hij heeft een fakkel meegenomen. Hij heeft ook een olielamp meegenomen. Hij heeft er nog even aan gedacht om ook een zaklantaarn mee te nemen, maar dat werd teveel om te dragen.
target: Harry heeft alleen een fakkel of een olielamp meegenomen.

Harry Potter promised to bring some lights to the gathering in the woods today. He brought a torch. He also brought an oil lamp. He thought of bringing a flashlight too, but that was too much to carry.
target: Harry brought only a torch or an oil lamp.

In (17a), the target sentence only picks out a member of the subject set {Harry, Hermione}, and not of the object set {oil lamp, torch}. Also, the contrast between Harry, who did bring lights, and Hermione, who did not, is given extra attention in the *usually... this time* sentence. This together triggers the subject-oriented QUD: *Who brought a torch or an oil lamp?* or *Who brought lights?*, where the question what lights they brought is unimportant. This is confirmed by the focus on *Harry*, triggered by *only* in the subject. The scalar *or* is not in the focused part in this condition, so the QUD Focus Condition expects no SI and therefore a 'true' answer.

In contrast, in (17b), it is the object set of three {torch, oil lamp, flashlight} from which two members are picked by the target sentence. The subject set is no longer important, as it has only one member that is not contrasted with any other person. Here, the contrast between the two objects that were brought along and the one that was not gets extra attention by the sentence that expresses that the third thing was almost picked, but eventually was not (*he thought of bringing... but did not*). These two things trigger the object-oriented QUD: *What/Which lights did Harry bring?* This QUD is also signaled by the focus on *a torch or an oil lamp*, triggered by *only* in the object. Here, the scalar *or* is in the focused constituent, so according to the QUD Focus Condition we do expect the *not both* SI here and therefore a 'false' answer.

[43] Note that in both stories, explicit use of *and* was avoided, and instead two separate sentences were used, connected by *also*. So the sentence: *Harry brought a torch and an oil lamp* was avoided. This was a precaution to make sure participants would actually interpret the target sentence, and not just judge the sentence 'untrue' based on a superficial discrepancy, without interpreting it.

I created 20 story-pairs like (17a-b) according to the skeleton in (16a-b) (so altogether 40 stories). I also created 25 filler stories with target sentences with *only*, which were slightly more difficult than the test items. This was achieved by using bigger contrast sets and using negation in the story or in the target sentence. To make sure participants were not confused by fillers that did not have the clear subject- or object-orientation, each filler was either unambiguously subject-oriented (having only a contrast set for the subject), or unambiguously object-oriented (only contrast set for object). Each target sentence of a filler also had *only* in the corresponding constituent. The number of subject-oriented fillers and object-oriented fillers was balanced (13x subject-oriented, 12x object-oriented), just like the number of true and false fillers (13x true, 12x false). As the test items all contain *or*, 20 of the 25 fillers contained *and* (and none contained *or*) balancing the total of occurrences of *or* and *and* in the experiment (both 20).[44] A number of the fillers also included the *usually ... this time* contrast (which was used in Condition 1) or the *wanted to ... but did not* contrast (of Condition 2), to avoid that participants were able to deduce what the test items are.

Method and Procedure

The experiment was conducted in Dutch. Participants were presented with a written TVJT, containing 45 stories with target sentences that contained *alleen (only)*. Participants were led to believe that the experiment was about the interpretation of *alleen*. They were instructed to judge the sentences as 'true' or 'false'. It was explicitly pointed out to the participants (both by the experimenter and in the written instructions) that they should judge the sentences on truth, and ignore strangeness or uglyness of sentences. The instruction further stated that the experiment would take about 15-20 minutes. There was no time limit, all participants completed the experiment in 15 minutes or less.

Design

Each questionnaire consisted of 20 test items (10x Condition 1 and 10x Condition 2) and 25 fillers. The stories were presented in a pseudo-random order. Four versions of the questionnaire were used. Version 1 is given in the appendix. Version 1 and Version 2 had the same order of items, but for every test item the other condition (so the other story of the story-pair) was presented. So if Version 1 had for instance story (17a) above, Version 2 had (17b) in the same place. Version 3 and 4 were respectively Version 1 and 2, but in reversed order. The first test item of Version 1 was a Condition 1 item, and the last a Condition 2 item (and therefore the first test item of Version 3 was a Condition 2 item, and the last was a Condition 1 item). The first test item of Version 2 was a Condition 2 item, and the last was a Condition 1 item (and therefore the first test item of Version 4 was a Condition 1 item and the last was a Condition 2 item), so this was balanced over versions. The versions were evenly distributed over the participants. The questionnaire never contained three or more test items directly after each other (irrespective of condition). In each questionnaire it occurred five times that a test item was presented directly after another test item (3 times the condition was different, 1 time two items of Condition 1, and 1 time two items of Condition 2).

Participants

31 undergraduates of the Faculty of Arts of Universiteit Utrecht participated in the experiment (23 female, 8 male). All of them filled out the questionnaire at the same time. All participants were native speakers of Dutch and had no knowledge of formal logic. No participants seemed

[44] However, none of the fillers contained *and* in object position. This is again (see fn. 43) a precaution to stop participants from rejecting items based on a superficial comparison to target sentences with *and*, so to make sure participants are actually interpreting the target sentences.

to have any difficulties in completing the experiment. All participants answered almost all fillers correctly, so no subjects were excluded from analysis. [45]

Results

The mean percentage SIs for Condition 1 was 54.2% and for Condition 2 was 84.2%. If we break down these means by the histograms of both conditions, we immediately see how the difference came about:

Figure 1a

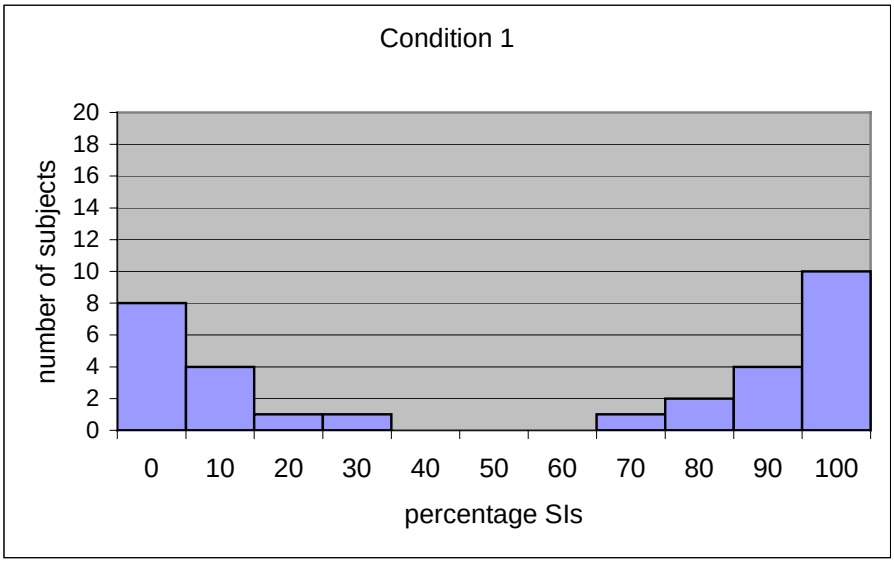
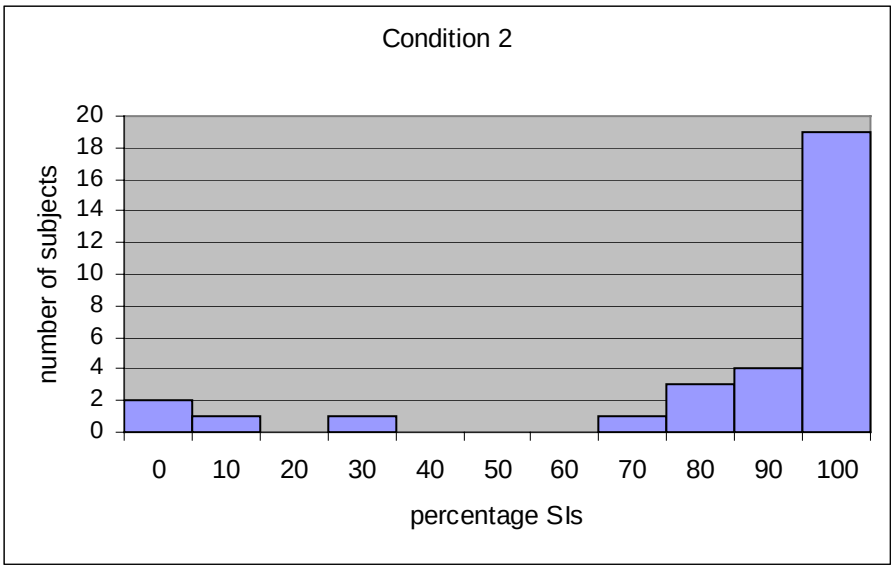


Figure 1b



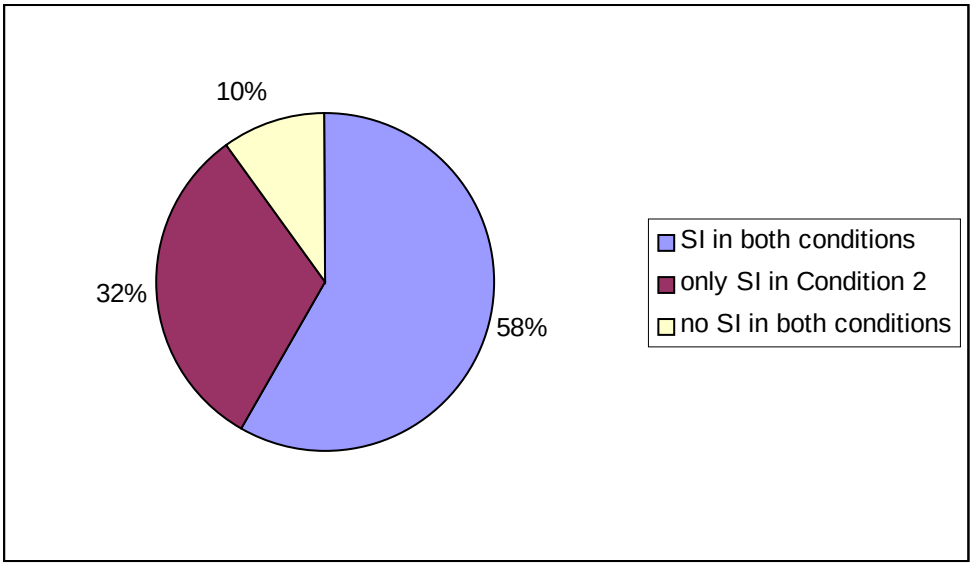
As the histogram of Condition 1 clearly shows, the participants are split up into two groups: one group that did calculate an SI in Condition 1 and a group that did not. The histogram of

[45] 30 out of 31 participants got 22 or more of the 25 fillers correct. One participant got 18 out of 25 fillers correct, so it is questionable whether this participant should be included in the analysis. However, as this participant did not behave according to QUD Focus Condition, I decided it would not bias the data in favor of the hypothesis so I did include the participant in the analysis.

Condition 2 clearly does not show this split: most participants calculate an SI in this condition.

Subject-analysis
Out of the 31 participants, 10 participants (32%) answered mainly 'true' to Condition 1 and mainly 'false' to Condition 2.[46] There were no participants that answered mainly 'false' to Condition 1 and mainly 'true' to Condition 2. 18 participants (58%) answered mainly 'false' to both conditions. 3 participants (10%) answered mainly 'true' to both conditions. As a 'false' answer indicates an SI, the SI-calculating behavior of the subject group looks like this:

Figure 2



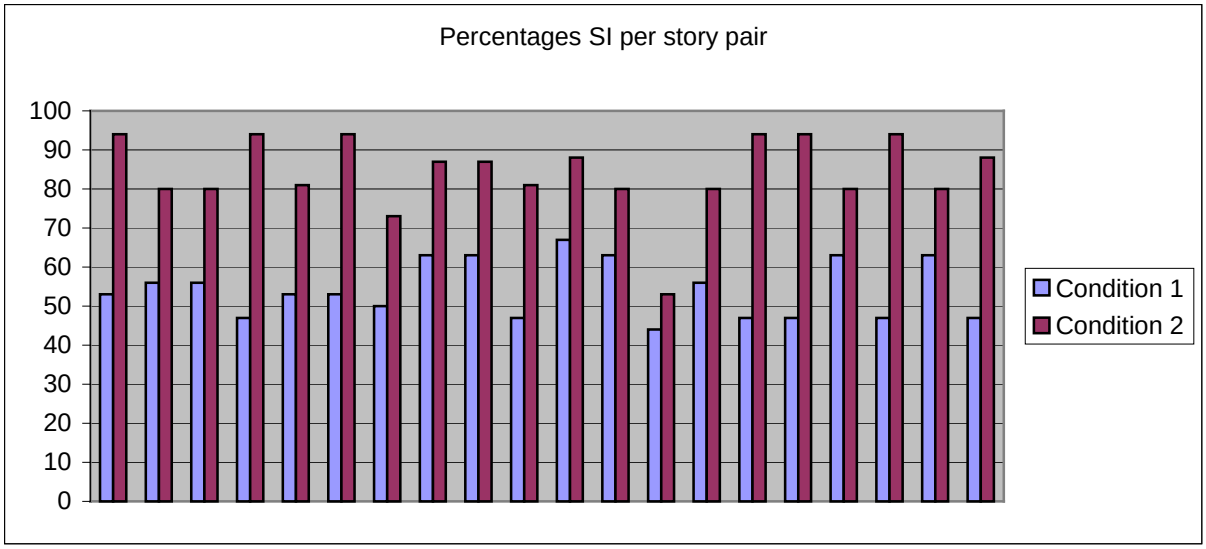
So 32% of the participants answered according to the QUD Focus Condition, calculating SIs only in Condition 2. Importantly, no participants showed the opposite behavior to the QUD Focus Condition: only calculating SIs in Condition 1. However, 58% of the participants made no difference between the two conditions and calculated the SI on both conditions, while 10% calculated no SI in either condition.

An analysis of variance revealed that over subjects there was a significant effect of condition ($F_1(1,30) = 15.67$, $p<.0005$). The scores per subject are given in the appendix. So although only 32% of the participants answered according to the QUD Focus Condition, calculated over all subjects there was a significant effect of condition on SI-calculation.

Items-analysis
The results over items are very consistent (see the appendix for the percentages per story-pair). This reflects the consistent answering of the participants per condition, and shows that all of the story-pairs worked well in triggering the different conditions. An analysis of variance revealed that over items there was a significant effect of condition ($F_2(1,19) = 121.07$, $p<.0005$). Figure 3 below, in which the scores of both conditions on all 20 story pairs are put next to each other, shows the consistent scores over the stories.

[46] 'Mainly' means at least 7 out of 10 items (70%). Participants answered very consistently per condition. 83% of the 62 percentages SIs per condition (31 participants times two conditions) were 10% or less SIs or 90% or more SIs (SIs were signaled by a 'false' answer). No scores of 40-60% SIs were observed.

Figure 3



Conclusions

What is clear from the data of our experiment is that there is a significant effect of condition (and therefore of QUD+focus), both on the over subject analysis as on the over items analysis. This indicates that the SI-calculating behavior of subjects is significantly different in the two conditions: Less SIs are calculated in Condition 1. As the two conditions only vary in QUD and the corresponding focus structure, this has to be the cause of this particular difference. This can be taken as evidence for the effect of the QUD and focus on SI-calculation. The effect is nicely shown in Figure 1a and 1b above, which show that in Condition 1, where *or* was not in the focused part of the sentence, a significant part of the participants did not calculate the *not and* SI, while they did so in Condition 2, where *or* was in the focused part of the sentence. The effect is supported by the fact that none of the participants behaved in the opposite way (only calculating the SI when *or* was not focused).

However, as is clearly shown in Figure 2, there is also a large group of subjects (58%) that still calculate SIs on both conditions. This indicates that other factors than focus might have also triggered SI-calculation. This suggests that QUD-answerhood (and thus focus) might better be considered as a *sufficient condition* for SI-calculation: *If* the scalar term is in the part of the sentence that answers the QUD, the SI will arise (instead of *iff*). In any case, it is clear from this experiment that whether or not the scalar term has focus affects SI-calculation. This cannot be explained on a defaultist view of SIs and therefore it supports the contextualist view. The effect of focus should be taken into account in both theoretical and experimental future research on SIs.

3.3.3 Discussion

One of the main questions that arises when we consider the outcome of the experiment is why, contrary to the QUD Focus Condition, the majority of the subjects calculate SIs in both conditions. Above we already suggested that the QUD Focus Condition might better be considered to be a sufficient condition for SI-calculation, and that other factors also affected SI-calculation. However, in this section I will consider two factors that were the result of the

specific setup of the experiment and that might have caused the SIs on Condition 1, that were unexpected on our initial QUD Focus Condition. The first possible factor is that there was a carry-over effect from the items of Condition 2 to the items of Condition 1. The second is the possibility that in Condition 1, a pair-list QUD was triggered instead of the subject-oriented one we aimed at.

Possible carry-over effects

As was shown in (16a) and (16b), the skeletons of the two conditions had many similarities: I tried to only vary the things that triggered the specific QUD, so that if a difference was found it would have to be due to the role of the QUD. However, a downside of this is that the items of both conditions looked very much alike. Most importantly, items of both conditions had a target sentence that contained *or* where *and* was also possible. As participants usually try to answer consistently, this might have caused judgments on one condition to have spilled over to the other condition. The most plausible direction of this carry-over effect is from items of Condition 2 to items of Condition 1: If a subject first encounters an item of Condition 2, he will calculate the SI, yielding the or_{excl} meaning and judge the sentence 'untrue' based on the fact that both disjuncts are the case. So the next time he encounters *or* in a context where both disjuncts are the case, irrespective if this time it is a Condition 1 item, he might reject this sentence by analogy to the Condition 2 item, instead of really interpreting the Condition 1 item. A carry-over effect the other way around is also possible, but only if the subject was aware of the fact that *or* was used where *and* was also possible: Answering 'true' to a Condition 1 item means that the SI was not calculated, so we can not tell whether subjects noticed that *and* was also a possibility.[47] If they did, it is possible they kept this in mind and also answered 'true' to Condition 2 items by analogy. However, there were only 3 subjects that answered 'true' to both conditions, so this scenario seems implausible.

There were however 18 subjects that answered 'untrue' to both conditions, so they might have carried over Condition 2 answers to Condition 1 items. Consider the following scenario:

Scenario 1: The first test item a subject encounters is a Condition 2 item. This is judged 'untrue'. After that, all Condition 1 items are also judged 'untrue'.

So the first test item a subject encounters is important. The versions of the questionnaire differed in this respect: The first test item of Version 1 and 4 was a Condition 1 item, while in Version 2 and 3 it was a Condition 2 item. So Scenario 1 only applies to Version 2 and 3. Based on this, we would expect significantly more subjects to answer 'untrue' to both conditions here than in Version 1 and 4, which started with a Condition 1 item. However, this is not the case: 9 subjects that filled out Version 2 or 3 answered 'untrue' to both conditions, while 8 subjects that filled out Version 1 and 4 did. However, the carry-over effect for Version 1 and 4 might have started later, after the first Condition 2 item was encountered, according to the following two scenarios:

Scenario 2a: The first test item is a Condition 1 item. This one is judged 'true'. The second test item is a Condition 2 item, that is judged 'untrue'. After that, all Condition 1 items are judged 'untrue'.

Scenario 2b: Same as Scenario 2a, but now the first Condition 1 item is corrected from 'true' to 'untrue'.

[47] I will return to this in the next section.

Scenario 3: Subject initially answers 'true' to Condition 1 items. However, after a couple of Condition 2 items, he starts answering 'untrue' to Condition 1 items too.

Let us consider Scenario 2a: Only two times was the switch from 'true' to 'untrue' after the first item observed for items of Condition 1, so this also provides no compelling evidence of carry-over.[48] How about Scenario 2b? In 6 questionnaires one or more answers to test items were corrected, only two of which suggested possible carry-over from Condition 2 to Condition 1. Finally, also Scenario 3 is not supported by the data: Only once a clean switch on items of Condition 1 from 'true' to 'false' in a later part of the questionnaire was observed, and that was after having answered 'true' to only two Condition 1 items. We can conclude that although carry-over effects from Condition 2 to Condition 1 seem plausible, there is no concrete evidence for them in the data.

Pair-list QUD in Condition 1

In Section 3.1 I explained the results of previous studies on SIs in terms of the QUD Focus Condition. There, I claimed that SI-percentages around 50% are typical of QUD-guessing: Subjects answered according to a QUD for which the scalar is in focus half of the times and according to a QUD for which it is not in focus the other half of the times. This was caused by the fact that the context triggered these two possible QUDs, or no QUD at all (resulting in accommodation of the two types of QUDs). However, in Condition 1 of our experiment, we also found a mean of 54.2% SIs. As Figure 1a showed, this was caused by two groups of subjects, one group calculating SIs, and one group not calculating them. So a possible explanation of the behavior of the group that still calculated SIs in Condition 1, is that they still considered the target sentence to be an answer to an object-oriented QUD, despite our efforts to trigger a subject-oriented QUD in that condition. Or perhaps even a subject- and object-oriented QUD was possible. For instance, as the stories of Condition 1 were about different people (e.g. Harry and Hermione) who interacted with different objects (e.g. a torch and an oil lamp), this might have triggered the QUD: *Who interacted with what?* Notice that in Conditon 2 this QUD is less plausible, as there was only one person in the story. Now this alternative QUD also questions the object, so *a torch or an oil lamp* would also get focus and therefore an SI would be calculated. We would however have to assume that subjects would allow two foci (subject and object) in one sentence. Also, as this alternative QUD needs a pair-list answer, we would have to assume that the subjects that got this QUD, interpreted the target sentence (*Only Harry brought a torch or an oil lamp*) as an implicit pair-list answer, like: *Harry brought a torch or an oil lamp, and Hermione brought nothing*. So the possibility that Condition 1 also triggered a QUD like *Who brough what?* would explain why a group of subjects calculated the SI there, but it relies on some assumptions that cannot be backed up by our experiment.

3.3.4 General discussion: the suitability of TVJT to assess SIs

The task that is most often used in experiments on SIs is the TVJT, which is why I used it in the experiment too. However, it is by no means obvious that this task is suitable to show the presence or absence of SIs. Here, I discuss two serious problems for this particular task: the question how to interpret the truth value judgments and the possibility that conscious control over SIs influences the judgments. This section is about the shortcomings of the TVJT in general, so it will not reconsider the results of our experiment.

[48] Although these cases were in Version 1 or 4.

How to interpret the truth value judgments

A well-known criticism of experiments that use the TVJT to assess SIs is that pragmatic inferences are hard to judge in terms of truth and falsity because is it merely *felicity* that is at issue. The researchers that conducted the TVJT experiments I discussed above (including ourselves) all adopt some non-trivial assumptions that allow them to draw conclusions about the environments in which SIs are drawn. While they usually take these assumptions for granted, I will discuss them here and consider whether they might bias the data. I will start with the following two:

(21) assumption: Hearers have no conscious control over SI-calculation, just like they have no conscious control over other components of meaning.
(22) assumption: SIs effect truth conditions of the sentences that gave rise to them.

Let us consider these one by one. Assumption (21) is based on the idea that hearers have no conscious control over the meaning the processor gives to sentences. Of course, after this initial meaning is established, hearers can manipulate this meaning. They can for instance throw out an irrelevant meaning of an ambiguous word or sentence, or hypothesize over what the speaker might have meant if the meaning makes no sense. However, they have no control over which meanings are triggered in the first place. Assumption (21) claims this is also the case for SI-calculation: Hearers have no conscious control over the initial SI-calculation.[49] In the next subsection I will discuss the possibility of hearers consciously canceling an SI after they initially did have it.

The second assumption, (22), is of vital importance to all experiments that rely on the TVJT to assess SIs. It says that the extra meaning that is added to a sentence by an SI, has influence on the truth conditions of this sentence: If a situation is not compatible with the SI of a sentence (if it has been calculated), the sentence becomes untrue, not just infelicitous, in this situation. This assumption is implicitly accepted by all researchers that hope to draw conclusions about SIs based on TVJT experiments: if the SIs do not influence truth conditions in this direct way, it is impossible to derive their presence from truth value judgments.

However the following assumption is often taken to be a straightforward extension of assumption (22):

(23) assumption: An 'untrue'-response to a target sentence that is true without an SI, is evidence that the SI is calculated.

Now (23) is a classical case of the fallacy that is so often made with logical implication: $p \rightarrow q$ does not entail $q \rightarrow p$. An SI can change the truth conditions of a sentence from 'true' to 'false' (assumption (22)), but this does not mean the 'untrue'-response is necessarily the result of this SI. After all, there might have been another cause of the 'untrue' answer. Let us for the sake of the argument, pursue such an option.

In Chapter 1 we saw that SIs arise in situations in which a speaker uses a weaker scalar item and the hearer concludes the stronger item is not the case. So the hearer does not know what the actual situation is, otherwise he would not have to draw conclusions based on what the speaker said. Now in the TVJT experiments, the hearer (in this case the participant) does

[49] To make this more clear, in what follows I will sometimes use terms like *have the SI* or *got the SI* instead of *calculated the SI* which sounds too much like a controlled process.

know the actual situation. The implicit claim (23) in the TVJT experiments is that if he marks the sentence with a weaker scalar term as 'untrue', this is evidence of the fact that he calculated an SI for the sentence. However, there are (at least) two ways in which this 'untrue' answer might have come about, which are not necessarily the same thing: Subjects might have read the sentence, got an SI and therefore (e.g.) the exclusive meaning of *or*, which made the sentence untrue. Alternatively, subjects might just have noticed that *and* would be a better option in the target sentence, and as ignorant participants cannot be expected to clearly distinguish infelicity from falseness (they simply search for discrepancies between the story and the sentence), they might have judged the sentence 'untrue' based on this infelicity alone. The question is however, whether these two things are the same thing: Is it possible that subjects who recognize the stronger alternative, and therefore know the scale, do not calculate the SI that negates the stronger alternative? Strictly speaking, recognizing the possibility of a stronger alternative (and knowing the scale) is only a *necessary condition* for SI-calculation, not a *sufficient* one. So is it possible for subjects to have the inclusive reading of *or* (without SI) but judge the sentence 'untrue' based on *and* being possible? We would then have to explain why someone who recognizes the possibility of a stronger item, would not calculate the SI, while others would. However, if it is possible, this might have led to an overestimation of the percentage of SIs in the TVJT experiments. The two possible causes of an 'untrue' answer then are: [50]

Possibility 1: The subject did get the SI and judged the sentence 'untrue' based on the exclusive interpretation of *or*.
Possibility 2: The subject did not get the SI, but did notice the possibility of *and* instead of *or*, and judged the sentence 'untrue' based on that.

However, for the 'true'-answers we do not have this problem. There are (at least) three possibilities how participants might have landed on such an answer:

Possibility 3a: The subject did not get the SI and did not notice the possibility of *and* instead of *or*.
Possibility 3b: The subject did not get the SI but did notice the possibility of *and* instead of *or*. However, he decided this does not make the sentence untrue.
Possibility 4: The subject did get the SI, but deliberately ignored / canceled it.

I will consider Possibility 4 in the next subsection. Possibility 3a and 3b show that for the 'true'-answers, the possibility of a stronger alternative does not interfere with the number of SIs. So for 'true'-answers this does not bias our conclusions about SIs. However, here Possibility 4 is an interfering factor that makes it hard to draw conclusions about SIs.

As the speculations above show, the possibility that there are several different factors that can cause subjects to answer 'true' or 'untrue' would make it hard to asses SIs with a TVJT. Below I will consider another possible interfering factor: the deliberate cancellation of SIs.

Deliberately ignoring SIs based on the intentions of the speaker
In section 3.2 I discussed Experiment 3 of Feeney et al. (2004). They compared reaction times on underinformative sentences like *Some elephants have trunks* to sentences that were not

[50] Notice the difference in conscious control between the two possibilities. In Possibility 1, the exclusive interpretation comes about unconsciously (by SI) and that immediately makes the sentence untrue. In Possibility 2 however, there is a conscious step: The recognition of the possibility of a stronger term is unconscious, but the decision whether this makes the sentence untrue is a conscious one.

underinformative like *Some birds live in cages*. They only compared reaction times of 'true'-answers to both conditions, so answers that were not based on the SI. It turned out that it took significantly longer to answer 'true' to the underinformative sentences than it took to answer 'true' on the other sentence-type. This obviously shows something is going on in the underinformative condition. Feeney et al. claim the difference is the result of a deliberate cancellation of the SI: Although participants initially get an SI, they choose to ignore it. This would of course be devastating for TVJT experiments: if subjects that get an SI in the first place have conscious control over whether they answer according to the SI or ignore it, it is impossible to determine by their answers whether they had the SI.

Feeney et al. claim the reason for deliberately canceling SIs is that the intention of the speaker (in this case the experimenter) is unclear. This makes sense as the task provided no context story, so this is similar to QUD-guessing: guessing what is at issue here. However, this brings us to a more general problem: One of the conditions of SI-calculation, is that the intentions of the speaker are clear. SIs are based on Grice's Cooperativity Principle. So only when you know a speaker is being cooperative, you can draw this kind of inferences. The cooperativity of the speaker is however not obvious in a TVJT. After all, the task is to check the utterances of the speaker to see if they are true relative to a story. But this inherently triggers doubts about the speaker's cooperativeness: the fact that he might not be telling the truth leaves open the possibility that he is being uncooperative. Now the question is whether this would block the initial unconscious SI-calculation or whether the initial SI is consciously canceled later on, as Feeney et al. suggest.

To see how the intentions of the speaker can block / cancel SIs, consider for instance a sentence like (24), taken from Feeney et al.'s Experiment 2:

(24) I've eaten some of the sweets.

If a hearer does not know what the actual situation is, say a mother comes home and her child utters (24), it will be less likely that the mother will interpret (24) according to the *not all* SI here: after all, the child may be trying to conceal that it ate all the sweets. In Feeney et al's Experiment 2, participants knew the child ate all of the sweets, and they judged (24) 'untrue' based on the fact that they recognized the uncooperative use of the weaker form. So the estimations of the hearer of the intentions of the speaker are decisive for whether or not subjects answer according to SIs. In an experiment by Mol et al. (2005), participants had to guess each other's secret color code (consisting of four colors). Each turn, they had to guess a code and they received feedback from the other player. They also had to indicate to the experimenters how they interpreted the other player's feedback. The goal of each player was to guess the other player's code before he did. So instead of conveying as much information as possible, players tried to convey as little as possible, creating a completely non-cooperative situation. Results of this study show that this kind of game-theoretical reasoning about the other player's intentions can shift both language use and interpretation from more pragmatic to more logical, so decreasing the environments in which SIs are used. So the intention of the speaker is an important factor for SIs, that the TVJT experiments seem to have little grip on.

Many TVJT experiments try to control for this by introducing an 'incompetent speaker' (e.g. a puppet that says silly things). The problem with this is that subjects might think this speaker is not able to draw certain pragmatic inferences such as SIs, or that he is not aware of certain entailment scales. If for instance an incompetent speaker utters *Some horses jumped over the fence* while all of them did, a participant might initially have an SI, but still judge this

sentence ‘true’ based on the idea that the incompetent speaker did not deliberately use a weaker term, that he did not mean to deny the stronger term by using the weaker. So only if a hearer knows the speaker is cooperative and competent, will he actually answer according to an SI. These two things are however very questionable in a TVJT.

3.3.5 Suggestions for further research

In 3.3.3 I discussed two possible biasing factors of the present experiment: carry-over effects and a possible pair-list QUD in Condition 1. Carry-over effects from one condition to another can be ruled out by splitting the subject group in two, testing one condition per subject. For this a large set of subjects is needed though. The second factor was the possibility of a pair-list QUD for Condition 1. However, this can be avoided by using pre-recorded target sentences, using the main stress of the sentence to indicate which constituent has focus, instead of using *only*. Putting the main stress on the subject will rule out a pair-list QUD. These two options are relatively small adjustments to the current experiment, which might make an even stronger case for the QUD Focus Condition.

However, the last section discussed the shortcomings of a TVJT to assess SIs. Some authors (e.g. Papafragou & Musolino (2003), Guasti et al. (2005)) have therefore switched to a Felicity Judgment Task (FJT), in which subjects are asked to judge if the speaker “said it well”. This is a step forward: On a TVJT we do not know how subjects will judge a sentence with a weaker item, it might be the case that they ignore it and still judge the sentence ‘true’, or they might judge the sentence ‘untrue’ based on it. On a FJT we definitely expect subjects to answer ‘no’ when a weaker item is used, so we can conclude that if they answer ‘yes’, they did not notice the infelicity of the weak item. So if a FJT results in more ‘no’ answers than a TVJT, we can conclude that in the TVJT at least some subjects found the possibility of a stronger alternative does not make the sentence untrue. However, a ‘no’-answer on a FJT can still mean two things: the subject got the SI that makes the target sentence untrue, and judges the sentence infelicitous because it is untrue, or the subject did not get the (e.g. *not both*) SI, but just rejected the infelicity of the weaker term. So we still need a task that clearly distinguishes SIs from infelicity caused by a weak item.

A possible task would be an act-out variant of our experiment, which would look like this: A subject is not presented with a story but just with a QUD (and perhaps to reinforce this a reason why this QUD is very important). The subject therefore does not know what the actual situation is. It is made clear that the speaker knows the actual situation, so the answer to the QUD, and that he is cooperative and competent. In Condition 1 the QUD is (like in our experiment): *Who brought a torch or an oil lamp?* The subject is given two puppets (Harry and Hermione) and two toy-objects (a torch and an oil lamp). The sentence that is uttered by the speaker is (main stress is indicated by capitals):

(25) HARRY brought a torch or an oil lamp.

The subject is now asked to show all possible combinations of puppets and objects that follow from (25). The QUD Focus Condition predicts that subjects would come up with three possible situations: one in which Harry has only a torch, one in which he has only an oil lamp and one in which he has both (and in all three situations Hermione has nothing). Contrastively, in Condition 2 the QUD would be: *What did Harry bring?* The subject gets one puppet and three toy-objects (torch, oil lamp and flashlight). The sentence is:

(26) Harry brought A TORCH OR AN OIL LAMP.

Now the QUD Focus Condition expects the subjects to act out only two possibilities: One in which Harry brought only a torch an one in which he brought only an oil lamp. If this is the case, this can only mean he calculated the *not both* SI. Now we have direct evidence of how the subjects interpreted *or*: with or without SI. We no longer have to rely on truth value judgments that are hard to interpret as they could have been caused by a number of other factors.

Finally, it would be interesting to further investigate the game-theoretical reasoning about the other player's intentions and the corresponding SI-behavior. A way to do this is to conduct experiments like the one I just described, but vary the cooperativity of the speaker that can be assumed by the hearer. We could for instance tell the participant that the speaker's goal is to make him find out what the real situation is like with a couple of clues. We could then for instance ask the subject to act out all the possibilities that are left after that clue. If we now present sentences like (26) as a 'first clue' it would be interesting to see if the participant will still rule out the third possibility, or keep it in because the speaker is not completely cooperative.

CHAPTER 4: CONCLUSION

In this work I introduced a new condition that predicts in which environments SIs will arise, based on the Question Under Discussion. I showed that this condition can explain adult data on SIs that up till now seemed contradictory. The different percentages of SIs that were observed were actually predicted on the QUD Focus Condition: close to 0% SIs if the scalar was not in the part of the sentence that was questioned by the QUD and close to 100% if it was. Percentages around 50% typically indicated that two types of QUD were possible: One type for which the scalar was in the part that was questioned by the QUD and one type for which it was not. This was caused by lack of context, like in the Noveck (2001) study, or by a context that triggered more than one possible QUD, like in Janse et al. (2005). The fact that percentages around 50% were observed was a typical result of the guessing of a QUD. While none of the previous studies that yielded these seemingly contradictory results took the QUD into account, the above data suggest that it should be taken into account.

The results of the TVJT experiment we conducted provide support for the QUD Focus Condition. We found a significant effect of condition on both the subjects and the items analysis, and as the conditions only differed in QUD and corresponding focus-structure, this effect has to be caused by the QUD Focus Condition. The fact that 32% of the subjects made a difference between the two conditions in the way we expected on the QUD Focus Condition and none of the participants made the opposite difference, supports the QUD Focus Condition. Finally, on the condition where the scalar was in the focused part we observed that participants answered in two groups: one group did not calculate the SI, while the other group did. This split was not observed in the other condition: A large majority of subjects calculated an implicature there. This shows that the QUD Focus Condition caused a group of participants not to calculate an SI when *or* was not in the questioned part, while they did calculate an SI when the scalar was in the questioned part.

The results of the experiment are not equivocal though: Still 58% of the participants answered ‘untrue’, so according to the SI, to both conditions. We explored the possibility of carry-over effects causing this, but although it seems plausible, no concrete evidence was found. The other possible explanation for the 58% that had SIs on both conditions, is the option that a pair-list question was triggered in the condition in which *or* was not in the focused part. This explanation relied on some non-trivial assumptions however. We suggested two changes to the experiment that would rule out these two possibilities. For now, we can conclude that the experiment showed QUD and focus have an effect on SI-calculation, that has to be taken into account in future research. However, perhaps the QUD Focus Condition should be considered as a *sufficient condition* for SI-calculation.

The present work tried to get us another step closer to the workings of the remarkable human ability to deduce extra meaning from utterances. We argued that the Question Under Discussion is decisive for SI-calculation: We will deduce from an utterance with a scalar term the extra meaning that the stronger item does not hold, if the scalar term that is used appears in a part of the utterance that is focused, as it answers the question that is at issue in the present conversation: the Question Under Discussion.

REFERENCES

Atlas, J. (1993). The importance of being "only": testing the neo-Gricean versus neo-entailment paradigms. *Journal of Semantics* 10: 301-18.

Atlas, J. & S.C. Levinson (1981). It-Clefts, Informativeness, and Logical Form: Radical Pragmatics. In P. Cole, ed., *Radical Pragmatics*. New York: Academic Press.

Bach, K. (1994) Conversational implicature. *Mind & Language* 9, 124-162.

Bach, K. (2001). You don't say? *Synthese* 127, 11-31.

Bezuidenhout, A.L. & J.C. Cutting (2002). Literal meaning, minimal propositions, and pragmatic processing. *Journal of Pragmatics*, 34, 433–456.

Bezuidenhout, A.L. & R.K. Morris (2004). Implicature, Relevance and Default Pragmatic Inference. In: I. Noveck and D. Sperber (eds.) *Experimental Pragmatics*. Palgrave Macmillan.

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

Braine, M.D. & B. Rumain (1981) development of comprehension of 'or': Evidence for a sequence of competencies. *Journal of Experimental Child Psychology* 31: 46-70.

Breheny, R, N. Katsos and J. Williams (2005). Are generalised conversational implicatures generated on-line by Default? *Cognition* xx: 1-30.

Carston, R. (1995) Quantity maxims and generalised implicatures. *Lingua* 96. 213-24.

Chierchia, G. (2000). see Chierchia (2004).

Chierchia, G. (2004). Scalar implicatures, polarity phenomena and the syntax/pragmatic interface. In A. Beletti (Ed.), *Structures and Beyond*. Oxford University Press.

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

Chomsky, N. (1972) Some empirical issues in the theory of transformational grammar. In S. Peters (ed.), *Goals of Linguistic Theory*, 63-130. Prentice-Hall, Englewood Cliffs, NJ.

Evans, J.S.B. & S.E. Newstead (1980). A study of disjunctive reasoning. *Psychological Research* 41, 373-88.

Feeney, A., S. Scrafton, A. Duckworth and S.J. Handley (2004). The story of some: Everyday pragmatic inference by children and adults. *Canadian Journal of Experimental Psychology*, 58:90-1.

Gazdar G. (1979). *Pragmatics: Implicature, Presupposition, and Logical Form*. New York: Academic Press.

Grice H.P. (1967). see Grice (1975).

Grice, H. P. (1975). Logic and conversation (from the William James lectures, Harvard University, 1967). In P. Cole & J. Morgan (Eds.), *Syntax and semantics 3: Speech acts* (pp. 41-58). New York: Academic Press.

Groenendijk, J. & M. Stokhof (1984). *Studies on the Semantics of Questions and the Pragmatics of Answers*, dissertation, University of Amsterdam.

Groenendijk, J. & M. Stokhof (1994). Questions. In J. van Benthem & A. ter Meulen (eds.) *Handbook of Logic and Language*. Amsterdam: Elsevier.

Gualmini, A. (2001). *The Unbearable Lightness of Scalar Implicatures.* Doctoral research paper, University of Maryland at College Park.

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

Hamblin, C.L. (1973). Questions in Montague English. *Foundations of Language* 10, 41-53.

Hirschberg, J. (1985) *A theory of Scalar Implicature*, Ph.D. thesis, UPenn.

Horn, L. (1969). A presuppositional analysis of *only* and *even*. *Proceedings of the Annual Meeting of the Chicago Linguistics Society*. Volume 38. University of Chicago.

Horn, L. (1972). *On the Semantic Properties of Logical Operators in English*. Ph.D. diss., UCLA; distributed by IULC.

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

Horn, L. (1996) Exclusive company: *only* and the dynamics of vertical inference. *Journal of Semantics* 13.1: 10-40.
Horn, L. (2005) The Border Wars: a neo-Gricean persepctive. In K. Turner & K. von Heusinger (eds.), *Where Semantics Meets Pragmatics*. Elsevier.
Ippolito, M. (2005). An implicature analysis of 'Only'. Ms. Boston University.
Janse, I., F. Landa and A.J. Zondervan (2005) Paper Experimental Psycholinguistics, unpublished ms. Universiteit Utrecht.
van Kuppevelt, J. (1996). Inferring from topics. Scalar implicatures as topic-dependent inferences. *Linguistics and Philosophy* 19, 393-443.
Ladusaw, W. (1979). *Polarity Sensitivity as Inherent Scope Relation*. Ph.D. diss. University of Texas, Austin; distributed by IULC.
Levinson, S.C. (1982) *Pragmatics*, Cambridge: Cambridge University Press.
Levinson, S.C. (2000). *Presumptive Meanings : the theory of generalized conversational implicature*. MIT Press. Cambridge, MA.
MacCawley, J.D. (1981). *Everything that Linguists Have Always Wanted to Know about Logic But Were Ashamed to Ask*. University of Chicago Press.
Mol, L., R. Verbrugge, & P. Hendriks (2005). Learning to reason about other people's minds. In: *Proceedings of the Joint Symposium on Virtual Social Agents*, SSAISB 2005 Convention, AISB (The Society for the Study of Artificial Intelligence and the Simulation of Behavior), University of Hertfordshire, Hatfield.
Noveck, I. (2001). When children are more logical than adults: Experimental investigations of scalar implicature. *Cognition* 78: 165-188.
Noveck, I. (2004) Pragmatic Inferences Related to Logical Terms. In: I. Noveck and D. Sperber (eds.) *Experimental Pragmatics*. Palgrave Macmillan.
Papafragou A. & J. Musolino (2003). Scalar implicatures at the semantics-pragmatics interface. *Cognition* 80, 253-282.
Papafragou, A., & N. Tantalou (2004). Children's computation of implicatures. *Language Acquisition* 12: 71-82.
Paris, S.G. (1973). Comprehension of language connectives and propositional logic relationships. *Journal of Experimental Child Psychology* 16: 278-91.
Reinhart, T. (1995) Interface Strategies. *OTS Working Papers in Linguistics*, University of Utrecht, Utrecht, the Netherlands.
Reinhart, T. (2004). The Processing Cost of Reference Set Computation: Acquisition of Stress Shift and Focus. *Language Acquisition* 12, 109-155.
Roberts, C. (2006). *Only*, presupposition and implicature. submitted to *Journal of Semantics*.
van Rooij, R. (2002). Relevance Implicatures. Ms., ILLC, Amsterdam, available at http://semanticsarchive.net/Archive/WIyOWUyO/Implicfinal.pdf
van Rooij, R. & K. Schulz (2004). Exhaustive interpretation of complex sentences. *Journal of Logic, Laguage and Information* 13(4): 491-519.
Rooth, M. (1985). *Association with Focus*. PhD thesis, University of Massachusetts, GSLA, Amherst, MA.
Rooth, M. (1992). A theory of Focus Interpretation. *Natural Language Semantics* 1, 75-116.
Sauerland, U. (2004). Scalar implicatures in complex sentences. *Linguistics and Philosophy* 27. 367-91.
Smith, C.L. (1980). Quantifiers and question answering in young children. *Journal of Experimental Child Psychology* 30: 192-205.
Soames, S. (1982) How presuppositions are inherited: A solution to the projection problem. *Linguistic Inquiry* 13, 483-545.
Sperber, D. & D. Wilson (1986). *Relevance: Communication and Cognition*, Harvard University Press, Cambridge, MA.
Stanowich, K.E. (1999). *Who is rational? Studies of individual differences in reasoning*. Hillsdale, NJ: Lawrence Erlbaum Associates, Inc.
Sternberg, R.J.(1979). Developmental patterns in the encoding and combination of logical connectives. *Journal of Experimental Child Psychology* 28: 469-98.

Szabolcsi, A. & B. Haddican (2004). When Conjunction Meets Negation: A Study in Cross-linguistic Variation. *Journal of Semantics* 21(3):219-249.
Wagner, M. (2005) NPI-licensing and Focus Movement. *Proceedings of SALT XV*. Cornell University.
Wilson, D. & D. Sperber (2002) Relevance Theory. In Horn, L.R. & Ward, G. (eds.) (2004) *The Handbook of Pragmatics*. Oxford: Blackwell, 607-632.

APPENDIX PART I: PERCENTAGES PER SUBJECT AND PER ITEM

Figure 1: percentage SIs per subject (roman numbers indicate version)

Subject	Cond. 1	Cond. 2
I.1	10%	30%
I.2	90%	100%
I.3	10%	10%
I.4	90%	90%
I.5	0%	90%
I.6	0%	0%
I.7	20%	80%
II.1	0%	80%
II.2	90%	100%
II.3	0%	100%
II.4	100%	100%
II.5	100%	100%
II.6	100%	100%
II.7	10%	100%
II.8	0%	0%
III.1	100%	100%
III.2	100%	100%
III.3	30%	100%
III.4	100%	90%
III.5	100%	100%
III.6	10%	100%
III.7	100%	80%
III.8	0%	100%
IV.1	80%	100%
IV.2	100%	100%
IV.3	80%	70%
IV.4	70%	90%
IV.5	100%	100%
IV.6	0%	100%
IV.7	0%	100%
IV.8	90%	100%

Figure 2: percentage SIs per story-pair

Story-pair	Cond. 1	Cond. 2
1	53%	94%
2	56%	80%
3	56%	80%
4	47%	94%
5	53%	81%
6	53%	94%
7	50%	73%
8	63%	87%
9	63%	87%
10	47%	81%
11	67%	88%
12	63%	80%
13	44%	53%
14	56%	80%
15	47%	94%
16	47%	94%
17	63%	80%
18	47%	94%
19	63%	80%
20	47%	88%

APPENDIX PART II: VERSION I OF THE QUESTIONNAIRE

Here, for every story the number of the story-pair and the condition or the number of the filler is given between parentheses. This is only given here for convenience of the reader. It was of course not included in the actual questionnaire. For each test item and the instruction, the English translation is given below (of course also not included in the actual questionnaire).

Instructie:
Hieronder staan 45 korte verhaaltjes. Elk verhaaltje wordt gevolgd door een stelling met *alleen*. Geef bij elke stelling aan of deze waar is of onwaar, door rechts 'waar' of 'onwaar' te omcirkelen. De verhaaltjes zijn kort dus als het goed is doe je er tussen de 15 en 20 minuten over.
N.B: Het gaat erom of de zin *waar* is of niet. Probeer alleen dit te beoordelen en niet of een zin vreemd of lelijk is.

Instruction:
45 short stories are given below. Each story is followed by a statement with only. *Mark each statement as true or untrue, by encircling 'true' or 'untrue' on the right. The stories are short so normally it would take you between 15 and 20 minutes.*
N.B. What matters is if the sentence is true *or not. Try to judge only this and not if a sentence is strange or ugly.*

1. (Filler 4) André, Koos, Lee en Johnny moeten vanavond samen optreden maar ze zijn allemaal ziek. Ze spreken af allemaal zelf een vervanger te regelen. Johnny is echter te ziek om de telefoon op te pakken dus hij regelt geen vervanging. Lee heeft genoeg vrienden die voor hem willen invallen, maar vandaag lukt het hem niet iemand te strikken. De anderen hebben wel een vervanger geregeld.

Alleen André en Koos hebben vervanging geregeld.

2. (Filler 18) Superman, Spiderman en Batman hebben afgespoken vandaag zoveel mogelijk bejaarden te redden. Spiderman redt een een bejaarde die met zijn rollator in de tramrails zit. Superman redt een andere bejaarde die met zijn baard in zijn broekpers vastzit. Batman ziet nergens een bejaarde in nood.

Alleen Batman heeft geen bejaarde gered.

3. (Story 6 Cond 1) Wouter en Martijn hebben afgesproken vandaag allebei vogels te fotograferen in het bos. Wouter fotografeert een lijster. Even later fotografeert hij ook een ekster. Martijn, die gewoonlijk wel weet waar hij de vogels kan vinden, heeft vandaag geen geluk en het lukt hem niet ook maar één vogel te fotograferen.

Alleen Wouter heeft een lijster of een ekster gefotografeerd.

Wouter and Martijn agreed to both go and photograph birds in the forest. Wouter photographs a thrush. A little later he also photographs a magpie. Martijn, who usually knows exactly where to find the birds, has no luck today and he doesn't manage to photograph even one bird at all.

Only Wouter photographed a thrush or a magpie.

4. (Filler 9) Peppie en Kokkie gaan het huis schilderen. Ze beginnen met de kozijnen. Als ze daarmee klaar zijn besluiten ze de deuren niet te doen om genoeg verf over te houden voor de muren. Maar nog voordat ze daarmee beginnen gooien ze per ongeluk het verfblik om en kunnen ze dus niet verder.

Peppie en Kokkie hebben alleen de kozijnen geschilderd.

5. (Story 9 Cond 2) Gerrit heeft beloofd vandaag meubels te maken voor de winkel. Hij maakt een stoel. Even later maakt hij ook een tafel. Hij had de opdracht gekregen ook een bank te maken, maar hij is niet snel genoeg geweest om die af te maken.

Gerrit heeft alleen een stoel of een tafel gemaakt.

Gerrit promised to make furniture for the store today. He makes a chair. A little later he also makes a table. He was also given the task to make a couch, but he didn't get around to doing that.

Gerrit made only a chair or a table.

6. (Filler 10) Silvie en Rafaël gaan worst, bier en schnitzels te halen in de supermarkt. Ze pakken een blik worst. Daarna willen ze het bier gaan pakken maar Rafaëls favoriete merk is uitverkocht dus besluiten ze geen bier te kopen. Even later zien ze dat de schnitzels die ze wilden kopen wat aan de dure kant zijn, dus die laten ze maar liggen.

Silvie en Rafaël hebben alleen worst gekocht.

7. (Filler 15) Starsky en Hutch gaan hun schitterende auto pimpen. Ze pimpen eerst de velgen. Daarna wordt de achterbumper flink verlaagd. Ze willen ook het interieur pimpen, maar daar moeten ze nog even voor sparen.

Starsky en Hutch hebben alleen de achterbumper gepimpt.

8. (Filler 6). Björn, Benny, Agnetha en Frida gaan naar een verkleedfeest. Ze hebben afgesproken allemaal verkleedkleren aan te trekken. Agnetha is echter bang dat er verder niemand verkleed gaat dus besluit ze toch geen verkleedkleren aan te trekken. De anderen hebben wel allemaal verkleedkleren aangetrokken.

Alleen Björn en Benny hebben verkleedkleren aangetrokken.

9. (Story 1 Cond 1) Harry Potter en Hermione hebben afgesproken vandaag allebei wat verlichting mee te nemen naar de bijeenkomst in het bos. Harry heeft een fakkel meegenomen. Hij heeft ook een olielamp meegenomen. Hermione is meestal zeer betrouwbaar, maar dit keer is ze vergeten ook maar iets van verlichting mee te nemen.

Alleen Harry heeft een fakkel of een olielamp meegenomen.

Harry Potter and Hermione agreed to both bring some lights to the gathering in the woods today. Harry brought a torch. He also brought an oil lamp. Hermione is usually a very reliable person, but this time she forgot to bring any lights at all.

Only Harry brought a torch or an oil lamp.

10. (Story 4 Cond 1) Joep en Paulus hebben afgesproken vandaag allebei wat versiering te kopen voor het feest. Joep koopt slingers. Hij koopt ook vlaggetjes. Paulus is normaal geen lui persoon, maar vandaag heeft hij het niet kunnen opbrengen om naar de feestwinkel te gaan om ook maar iets aan versiering te kopen.

Alleen Joep heeft slingers of vlaggetjes gekocht.

Joep and Paulus agreed to both buy some decorations for the party today. Joep buys some paper chain. He also buys flags. Paulus is usually pretty motivated to help set up te party, but today he couldn't put himself to go to the store to buy any decoration.

Only Joep bought paper chain or flags.

11. (Filler 11) Gert en Hermien gaan de dieren in de dierentuin voederen. Ze voederen eerst de olifanten. Daarna willen ze de apen voederen maar ze kunnen niet dichtbij genoeg komen. Als ze daarna bij de tijgers aankomen worden ze door een verzorger weggestuurd en geven ze het verder op.

Gert en Hermien hebben alleen de olifanten gevoederd.

12. (Story 17 Cond 2) Tom heeft zijn publiek beloofd een paar mooie trucs op zijn skateboard te doen. Hij maakt eerst een kickflip. Daarna maakt hij een backflip. Hij wilde ook nog een ollie doen, maar dat lukt helaas vandaag niet.

Tom heeft alleen een kickflip of een backflip gemaakt.

Tom promised his audience to do some cool tricks on his skateboard. First he does a kickflip. After that he does a backflip. He wanted to do an ollie too, but that didn't work out today.

Tom did only a kickflip and a backflip.

13. (Filler 23) Hielke en Sietse zitten aan het ontbijt. Hun moeder heeft ze roggebrood, suikerbrood en krentenbrood voorgeschoteld. Ze willen graag goed eten om snel grote sterke mannen te worden, maar omdat ze erop uit moeten met de Kameleon laten ze het allebei bij een plak krentenbrood.

Hielke en Sietse hebben alleen geen roggebrood gegeten.

14. (Filler 1) Lance, Bernard, Eddy en Miguel gaan morgen wielrennen en hebben afgesproken allemaal een helm te dragen. Het is de dag erna echter zo warm dat Eddy besluit geen helm op te zetten. Miguel is meestal zeer voorzichtig, maar vandaag heeft hij besloten toch geen helm te dragen. De anderen hebben wel keurig een helm op.

Alleen Lance en Bernard hebben een helm opgezet.

15. (Story 8 Cond 2) Lanceloet heeft beloofd vandaag wild te schieten voor het banket van de koning. Hij schiet een konijn. Even later schiet hij ook een fazant. Hij had eigenlijk ook nog een hert willen schieten, maar dat lukt hem niet.

Lanceloet heeft alleen een konijn of een fazant geschoten.

Lanceloet promised to shoot some game (animals living in the wild) for the King's banquet today. He shoots a rabbit. A little later he also shoots a pheasant. He wanted to shoot a deer too, but he didn't succeed.

Lanceloet shot only a rabbit or a pheasant.

16. (Story 15 Cond 1) Van Basten en Rijkaard hebben afgesproken zoveel mogelijk mooie acties te maken op het voetbalveld. Rijkaard maakt een mooie volley. Later maakt hij ook een schitterend hakje. Van Basten, die normaal gesproken de ene mooie actie na de andere maakt, ziet geen kans om ook maar iets van een mooie actie te maken.

Alleen Rijkaard heeft een volley of een hakje gemaakt.

Van Basten and Rijkaard agreed to make as much beautiful actions as possible on the field. Rijkaard makes a beautiful volley. Later he makes a brilliant heel-ball. Van Basten, who normally makes one beautiful action after another, does not succeed in making any beautiful action at all.

Only Rijkaard made a volley or a heel-ball.

17. (Filler 20) Van Gaal, Advocaat, Cruijff en Hiddink zijn een potje aan het basketballen. Advocaat scoort aan de lopende band. Cruijff en Hiddink doen verwoede pogingen, maar uiteindelijk lukt het alleen Hiddink daadwerkelijk te scoren. Van Gaal vindt de anderen maar dom en probeert niet eens te scoren.

Alleen Van Gaal heeft geen bal in de basket gekregen.

18. (Story 20 Cond 1) Katja en Bridget hebben afgesproken vandaag huisdieren te fotograferen. Bridget fotografeert een hond. Even later fotografeert ze een kat. Katja is normaal gesproken erg oplettend, maar vandaag ziet ze niet één huisdier en kan ze er dus ook geen fotograferen.

Alleen Bridget heeft een hond of een kat gefotografeerd.

Katja and Bridget agreed to photograph as pets today. Bridget photographs a dog. A little later she also photographs a cat. Katja is normally very alert, but today she doesn't see any pets so she can't photograph any either.

Only Bridget photographed a dog or a cat.

19. (Filler 8) John, Paul, George en Ringo gaan naar een feest. Ze hebben afgesproken allemaal wat drank van het feest mee naar buiten te smokkelen. Paul vindt het uiteindelijk toch te onaardig en besluit geen drank te stelen. De anderen houden zich wel aan de afspraak en stelen drank van het feest.

Alleen John en Paul hebben drank gestolen.

20. (Story 16 Cond 1) Arnon en Leon hebben zich voorgenomen deze zomer lekker veel boeken te lezen. Leon leest een roman. Hij leest ook een thriller. Arnon, die meestal wel de tijd neemt om een boek te lezen, komt er deze zomer helemaal niet aan toe om ook maar iets te lezen.

Alleen Leon heeft een roman of een thriller gelezen.

Arnon and Leon planned to read a lot of books this summer. Leon reads a novel. He also reads a thriller. Arnon, who usually takes the time to read a book, doesn't get round to reading anything at all this summer.

Only Leon read a novel or a thriller.

21. (Story 19 Cond 2) Jennifer wordt in de trein aangesproken door een sociaal werker die fruit verzameld voor de daklozen. Ze geeft hem een appel. Ze geeft hem ook een sinaasappel. Ze wilde hem ook een peer geven, maar die bleek ze vandaag niet bij zich te hebben.

Jennifer heeft alleen een appel of een sinaasappel gegeven.

Jennifer is on the train and she is being asked for fruit by a social worker who's collecting fruit for the homeless. She gives him an apple. She also gives him an orange. She wanted to give him a pear too. but it turned out she didn't bring one today.

Jennifer gave only an apple or an orange.

22. (Filler 21) Er is een feest in de Sesamstraat en Pino en Ieniemienie hebben een fles wodka, een fles Passoa en een fles whisky meegenomen. De fles wodka hebben ze er zo doorheen gejast, maar ze doen langer over de fles Passoa. Als die op is zijn ze dan ook zodanig dronken, dat ze het verder voor gezien houden.

Pino en Ieniemienie hebben alleen de whisky niet meer opgedronken.

23. (Story 5 Cond 1) Sofie en Julie hebben afgesproken vandaag allebei langs de deuren te gaan om oude kleding in te zamelen voor Roemeense straatkindertjes. Sofie krijgt een trui mee. Later krijgt ze ook een broek mee. Julie, die normaal heel goed is in het inzamelen, heeft die dag minder geluk. Het lukt haar niet ook maar iets aan kleding mee te krijgen.

Alleen Sofie heeft een trui of een broek meegekregen.

Sofie and Julie agreed to both go door to door today to collect old clothing for Rumanian street children. Sofie gets a sweater. Later she also gets a pair of pants. Julie, who usually is very good at collecting, is less lucky that day. She gets no clothing at all.

Only Sofie got a sweater or pants.

24. (Filler 12) Gerard en Gordon moeten een badkamer schoonmaken. Ze maken eerst de WC schoon. Daarna moeten ze de douche schoonmaken maar die slaan ze stiekem over. Ze willen aan de wasbak beginnen maar besluiten dat ze er geen kracht meer voor hebben.

Gerard en Gordon hebben alleen de WC schoongemaakt.

25. (Filler 22) Sugar Lee Hooper wil eens wat aan de lijn gaan doen. Ze besluit haar dagelijkse dieet van taart, chocola en friet eens wat te herzien. Na lang wikken en wegen besluit dat ze niet zonder taart en friet kan, maar dat ze de rest van haar dagelijkse dieet resoluut afzweert.

Suger Lee Hooper heeft besloten alleen geen chocola meer te eten.

26. (Filler 2) Geri, Victoria, Melanie en Michelle gaan morgen picknicken en ze hebben afgesproken allemaal een taart te bakken. Geri heeft het echter erg druk en komt er niet aan toe een taart te bakken. Melanie heeft vandaag ook geen tijd en bakt dus ook geen taart. De anderen hebben wel keurig een taart gebakken.

Alleen Victoria en Michelle hebben een taart gebakken.

27. (Story 2 Cond 2) Jan heeft beloofd vandaag wat versnaperingen te kopen voor het klassenfeest. Hij heeft chips gekocht. Hij heeft ook koekjes gekocht. Hij wilde ook snoep kopen, maar daar is hij niet meer aan toe gekomen.

Jan heeft alleen chips of koekjes gekocht.

Jan promised to bring snacks to the class party. He brought potato chips. He also brought cookies. He wanted to bring some candy too, but he didn't get round to buying that.

Jan brought only chips or cookies.

28. (Filler 7) Maurits, Bernard, Constantijn en Friso hebben vanavond een familiebijeenkomst op het paleis. Omdat de chauffeurs staken moeten ze zelf vervoer regelen. Constantijn gokt er echter op dat de staking op tijd is afgelopen en regelt geen vervoer. De anderen hebben wel vervoer geregeld.

Alleen Maurits en Bernard hebben vervoer geregeld.

29. (Story 14 Cond 2) Sukke gaat naar een feestje en heeft eraan gedacht wat presentjes mee te nemen. Hij heeft een bloemetje meegenomen. Hij heeft ook een wijntje meegenomen. Hij wilde eerst ook een doosje bonbons meenemen, maar dat is hij vergeten.

Sukke heeft alleen een bloemetje of een wijntje meegenomen.

Sukke is going to a party and thought of bringing some small gifts. He brought some flowers. He also brought a bottle of wine. He wanted to bring a box of chocolates too, but he forgot.

Sukke only brought some flowers or a bottle of wine.

30. (Story 3 Cond 2) Marco heeft beloofd vandaag vis te vangen voor het avondeten. Hij vangt een zalm. Even later vangt hij ook een schol. Hij wilde eigenlijk ook nog een makreel vangen, maar dat lukt hem vandaag niet.

Marco heeft alleen een zalm of een schol gevangen.

Marco promised to catch some fish for dinner today. He catches a salmon. A little later he also catches a plaice. He actually wanted to catch a mackerel too, but he doesn't manage to do so.

Marco caught only a salmon or a plaice.

31. (Filler 16) Maverick, Goose, Iceman en Jester gaan weer eens ouderwets op Migs jagen. Jester moet al snel onverrichter zake terugkeren met motorproblemen. Iceman en Goose schieten allebei een Mig neer. Het lijkt Maverick niet te gaan lukken, maar vlak voordat zijn brandstof opraakt weet hij toch nog een Mig neer te halen.

Alleen Jester heeft geen Mig neergehaald.

32. (Filler 25) Asterix en Obelix zijn er door Regelnix op uitgestuurd om eten te zoeken voor het dorpsfeest. Ze moeten druiven, kippen, aardappels en een everzwijn meenemen. De druiven stelen ze meteen van de Romeinen. Na enig zoeken vinden ze ook een aantal kippen. Normaal gesproken vinden ze alles makkelijk, maar dit keer vinden ze verder niks meer.

Asterix en Obelix hebben alleen geen everzwijn gevonden.

33. (Story 18 Cond 1) Edwin en Daphne hebben afgesproken indruk te maken op de mensen in het zwembad door acrobatische sprongen van de duikplank te maken. Daphne maakt een salto. Bij de volgende sprong maakt ze een schroef. Edwin, die normaal gesproken een goede schoonspringer is, krijgt het niet voor elkaar om ook maar één acrobatische sprong te maken.

Alleen Daphne heeft een salto of een schroef gemaakt.

Edwin and Daphne agreed to impress the people in the swimming pool by making acrobatic jumps off the diving board. Daphne does a summersault. At her next jump she also does a twist. Edwin, who normally is a good diving board diver, doesn't manage to make even one acrobatic jump.

Only Daphne did a summersault or a twist.

34. (Filler 13) Ray en Anita hebben beloofd de stallen uit te mesten. Ze mesten eerst de paardenstal uit. Daarna mesten ze de koeienstal uit. Ze gaan door naar de varkensstal, maar daar stinkt het zo dat ze besluiten ermee op te houden.

Ray en Anita hebben alleen de koeienstal uitgemest.

35. (Filler 14) Suske en Wiske hebben beloofd een gedeelte van de afwas te doen. Ze wassen eerst de schalen af. Daarna wassen ze de pannen af. Vervolgens willen ze doorgaan met de borden. Maar zodra ze zien hoe die eruit zien hebben ze geen zin meer en besluiten ze te stoppen met afwassen.

Suske en Wiske hebben alleen de schalen afgewassen.

36. (Story 11 Cond 1) Svetlana en Natasja hebben afgesproken allebei van huis broodbeleg mee te nemen voor de picknick. Natasja neemt wat kaas mee. Ze neemt ook wat ham mee. Svetlana heeft normaal gesproken veel broodbeleg in huis, maar momenteel heeft ze niks aan broodbeleg in huis en kan ze dus ook niks meenemen.

Alleen Natasja heeft kaas of ham meegenomen.

Svetlana and Natasja agreed to both bring some toppings from home to the picknick. Natasja brings some cheese. She also brings some ham. Svetlana usually has a lot of toppings at home, but at the moment she doesn't have any toppings at home so she can't bring any either.

Only Natasja brought cheese or ham.

37. (Filler 19) Monet, Van Gogh, Gauguin en Rembrandt zitten samen sudoku's te maken. Van Gogh heeft er al snel één opgelost. De anderen hebben het er moeilijker mee, en uiteindelijk geven ze het tot groot genoegen van Van Gogh zelfs op.

Alleen Gauguin en Rembrandt hebben geen sudoku opgelost.

38. (Story 12 Cond 2) Irene heeft beloofd vandaag groente mee te nemen voor haar kookprogramma. Ze neemt wat kool mee. Ze neemt ook wat lof mee. Ze wilde ook wat bonen meenemen, maar die is ze straal vergeten.

Irene heeft alleen kool of lof meegenomen.

Irene promised to bring vegetables to his cooking show today. He brings some cabbage. He also brings some chicory. He wanted to bring some beans too, but he completely forgot.

Irene brought only cabbage or chicory.

39. (Filler 17) Jan-Peter, Rita, Wouter en Femke hebben allemaal beloofd een feesthoedje te dragen bij het volgende debat. Femke komt keurig met een feesthoedje op aanlopen. Jan-Peter heeft ook braaf een feesthoedje opgezet. De anderen hebben het toch niet aangedurfd daadwerkelijk een feesthoedje op te zetten.

Alleen Wouter en Rita hebben geen feesthoedje op.

40. (Filler 24) Roy heeft beloofd regenkleding mee te nemen voor zijn zeilmaatje. Hij neemt een regenbroek mee. Hij neemt ook een regenjas mee. Hij wilde ook een zuidwester en een regenlaarzen meenemen, maar die is hij vergeten.

Roy heeft alleen geen zuidwester meegenomen.

41. (Story 13 Cond 2) Rob zit in zijn bak naar gereedschap te zoeken voor een klus. Hij vindt wat schroefjes. Even later vindt hij ook wat spijkers. Hij dacht dat er ook nog wat boutjes inzaten, maar er blijkt verder niks in te zitten.

Rob heeft alleen schroefjes of spijkers gevonden.

Rob is searching his container for tools for a job. He finds some screws. A little later he finds some nails. He thought there were some bolts in there too, but it turns out there's nothing in there anymore.

Rob found only screws or nails.

42. (Filler 5) Helmut, Heinz, Franz en Dieter gaan naar het strand. Ze hebben afgesproken allemaal een kuil te graven. Dieter merkt echter al voordat hij begint met graven dat hij last van zijn rug heeft en besluit toch maar geen kuil te graven. De anderen graven wel allemaal een mooie kuil.

Alleen Helmut en Heinz hebben een kuil gegraven.

43. (Story 10 Cond 1) Paul en Michel hebben afgesproken allebei wat drinken te halen voor de gasten op hun feestje. Paul haalt wat bier. Even later haalt hij ook wat wijn. Michel houdt zich normaal gesproken keurig aan zijn afspraken, maar nu is hij straal vergeten ook maar iets aan drinken te halen.

Alleen Paul heeft bier of wijn gehaald.

Paul and Michel agreed both to go and get some drinks for the guests at their party. Paul gets some beers. A little later he also gets some wine. Michel usually sticks to his promises, but this time he totally forgot to get any drinks at all.

Only Paul got beer or wine.

44. (Story 7 Cond 2) Bill heeft beloofd om vandaag flink wat apparaten te repareren voor een elektronicahandel. Hij repareert een tv. Later repareert hij ook een radio. Hij wilde ook nog een cd-speler repareren, maar daar heeft hij helaas geen tijd meer voor.

Bill heeft alleen een tv of een radio gerepareerd.

Bill promised to fix a bunch of appliances for an electronics store today. He fixes a tv. Later he also fixes a radio. He actually also wanted to fix a cd-player, but he didn't get round to doing that.

Bill only fixed a tv or a radio.

45. (Filler 3) Nick, Kevin, Howie en Brian hebben allemaal een lekke band. Omdat ze morgen willen gaan fietsen hebben ze afgesproken allemaal hun band te plakken. Kevin komt erachter dat hij helemaal niet weet hoe dat moet dus hij plakt zijn band niet. Howie houdt zich normaal gesproken aan zijn afspraken maar plakt nu ook zijn band niet. De anderen hebben wel hun band geplakt.

Alleen Nick en Brian hebben hun band geplakt.