Processing Scalar Implicatures:

An Eye-Tracking Study

MSc Thesis in Cognitive Science

Judith Degen

jdegen@bcs.rochester.edu

March 7, 2009

Acknowledgements

I would like to thank my advisor, Peter Bosch, for supporting my every endeavor during my time in Osnabrück. It is from him that I learned most of what I know about pragmatics. Further thanks go to Mike Tanenhaus, Christine Glunlogson, and Florian Jaeger for fruitful discussions about the scalar implicature 'an sich', noticing suspicious bumps in plots, and last-minute help with data analysis challenges. Thank you for bearing with me during the finalization period.

Many thanks also to Mario Negrello and Jackie Griego, for moral support and reading early drafts of this thesis. And of course to Martin Schmidt, my personal python god, for help with pygame.

Finally, I am indebted to Jens, for never failing to challenge and support me throughout a very interesting year indeed.

Abstract

Current empirical research on scalar implicature processing is yielding substantial support for the Context-Driven over the Default account of scalar implicatures (Noveck and Posada (2003), Breheny et al. (2006)). The Context-Driven model is taken to be on par with Relevance Theory, the Default model with Neo-Gricean accounts. It is thus concluded by experimental pragmaticists such as Noveck that Relevance Theory is more likely to provide a correct account of scalar implicaturew, while Neo-Griceanism cannot account for the data.

I review both the theoretical literature (i.e. from the perspective of philosophy of language, semantics, and pragmatics proper) on scalar implicatures and the experimental pragmatics literature dealing with the cognitive basis of scalar implicature processing.

I present an original eye-tracking study on the German scale ⟨alle, einige⟩ (⟨all, some⟩), which aimed at investigating the question of Default. The results provide further support for the Context-Driven model, i.e. the view that processing of pragmatic information is more costly in a cognitive sense than the processing of semantic information.

I finally show, contrary to the established view, that support for the Context-Driven model need not lead to the rejection of (Neo-)Gricean approaches. In particular the assumption that being a Neo-Gricean implies defending a Default notion of scalar implicatures is not warranted.

I conclude with considerations about exciting new questions, beyond the question of Default, that arise within the realm of scalar implicature processing.

Contents

Chapter 1

Introduction

This paper is structured as follows. In section 1.1, I present the phenomenon of scalar implicature and elaborate on its empirical properties. Section 1.2 will review the current theories of scalar implicatures, specifically Neo-Gricean accounts, Relevance Theory, and a structural account. This will lead to the experimental perspective on scalar implicatures, worked out in section 1.3. Here I review the recent experimental findings on different aspects of the processing of scalar implicatures - the questions of Default and Localist processing will be considered, as well as developmental evidence. Section 2 is devoted to presenting an eye-tracking study conducted to investigate whether scalar implicatures arising from the German implicature trigger *einige* (some) are computed by default or rather are part of an effortful, context-driven process. The results show the latter. In section 3 I consider the implications of our study, but also of experimental work on the question of Default more generally, on the theoretical frameworks presented in section 1.2.

1.1 Scalar implicature - the phenomenon

Scalar implicatures arise in pieces of discourse such as the following:

(1) Clara: Do you remember our ski week last year?
Guido: Some rides were perfect powder rides.

Apart from literally conveying that there were perfect powder rides, Guido's answer in (1) is additionally taken to mean both that not all rides were perfect powder rides and that he remembers their ski week. This additional, non-literal meaning was famously explained by Grice (1975) with reference to the notions of *sentence meaning* and *speaker meaning*, which distinguish what is *said* from what is *meant*. Under this view, *sentences* are abstract objects with certain phonological, syntactic, and semantic properties. It is the main business of *syntax* and *semantics* to investigate these properties, which are at the core of linguistic inquiry. On the other hand, *utterances* are realizations of sentences with a concrete extension in time and space. In addition to inheriting all grammatical properties pertaining to the uttered sentence, they have further properties in virtue of having been uttered in a specific situation by a specific speaker addressing a specific audience. That is, these additional properties arise by the utterance's having occurred in a specific *context*. It is these properties that constitute speaker meaning and are the object of study in *pragmatics*. Specifically, for a speaker to mean that P is for her to have the intention that the hearer should realize that in producing the utterance, she intended him to think that P. In the words of Grice (1969):

> "*U* meant something by uttering *x*" is true iff, for some audience *A*,
> *U* uttered *x* intending:
> (i) *A* to produce a particular response *r*
> (ii) *A* to think (recognize) that *U* intends (i).
> (iii) *A* to fulfill (i) on the basis of the fulfillment of (ii).

In (1), Guido for example intends Clara to think that he remembers the ski week, without explicitly saying so. Such additionally conveyed meaning, which arises from the context of utterance, is *implicated*, what carries this meaning is an *implicature*.

The question arises as to how hearers go about inferring implicatures, given an utterance and contextual information. According to Grice, underlying this process is interlocutors' assumptions of mutual rationality. Assuming that discourse is a cooperative activity, speaker and hearer expect each other to follow certain standards. These standards are collapsed in Grice (1975)'s Cooperative Principle:

> Make your conversational contribution such as is required, at the stage at which it occurs, by the accepted purpose or direction of the talk exchange in which you are engaged.

Nine more specific conversational maxims further spell out the Cooperative Principle:

Maxims of Quantity

Quantity-1. Make your contribution as informative as is required (for the current purpose of the exchange).

Quantity-2. Do not make your contribution more informative than is required.

Maxims of Quality

Supermaxim. Try to make your contribution one that is true.

Truthfulness. Do not say what you believe to be false.

Evidencedness. Do not say that for which you lack evidence.

Maxim of Relation

Be relevant.

Maxims of Manner

Supermaxim. Be perspicuous.

Obscurity Avoidance. Avoid obscure expressions.

Ambiguity Avoidance. Avoid ambiguity.

Brevity. Be brief (avoid unnecessary prolixity).

Orderliness. Be orderly.

These maxims guide the process of implicature computation. Consider example (1) again. The implicature that not all rides were perfect powder rides is derived as follows (always assuming that the speaker is following the Cooperative Principle): Quantity-1 requires that the speaker make the most informative statement possible, given the level of informativeness set by the discourse. The more informative statement Guido could have made is *All rides were perfect powder rides*. However, since he did not make that statement, thus violating Quantity-1, it must be because he does not believe it to hold (Truthfulness trumps Quantity-1). Thus, Clara is justified in inferring that Guido indeed intended her to believe that only some, but not all, rides were perfect powder rides.

We have now seen the general mechanism giving rise to the class of implicatures based on a clash between Quantity-1 and Truthfulness. There is a second class of implicatures based on a clash between Quantity-1 and the maxim of Evidencedness. In this case, the implicature that arises is not that the stronger statement does not hold, but rather that the speaker does not know whether or not the stronger statement holds. In our example, such an implicature would arise if, for example, Guido did not go skiing himself, but heard about it from a friend who had told him only about the best rides. In this case, not knowing about the entire ski week, Guido would be equally unjustified as in the originally construed example in using the stronger statement. But Clara, knowing that Guido was not there himself and thus does not know the details about the entire ski day, would not infer from Guido's weaker statement that it was not the case that all the rides were perfect powder rides. Instead, she would infer that Guido does not know whether or not all the rides were, in fact, perfect powder rides. For the purposes of this paper, the interesting class of implicatures is the first, Truthfulness-based one. Henceforth, I do not consider the Evidencedness-based implicatures.

These implicatures, derived as the result of a clash between Quantity-1 and either of the Quality maxims are called *scalar implicatures*. They are scalar because the implicature trigger - *some* in the above example - is the weaker expression in a partially ordered scale consisting of a stronger and a weaker element (Matsumoto (1995)). In our example, the scale involved is ⟨all, some⟩, where *all* is the stronger element, as it logically entails *some*.[1] The ordering relations that may hold between scale elements are manifold: entailment, as in example (1); ranking of entities, states, and attributes; whole/part relationships; type/subtype, instance-of, and generalization/specialization relations (Carston

[1]In first-order predicate logic, $\exists x F(x)$ is implied by $\forall x F(x)$ (though only for non-empty domains, i.e. the fact that *all 13 year old university students are geniuses* does not entail that *there are 13 year old university students who are geniuses*)

(1998)). Generally, making the weaker statement conversationally implicates - but does not semantically imply - that the stronger statement does not hold.

Scalar implicatures exhibit a number of distinctive properties, namely *calculability, nonconventionality, cancelability, nondetachability, reinforceability*, and *universality* (Levinson (2000)). I will briefly review each of these in turn.

1.1.1 Calculability

Scalar implicatures are calculable via the conversational maxims of Quantity-1 and Truthfulness (or Evidencedness). What is required is that the implicature trigger be the weaker element of a partially ordered scale. Above, I provided an example of scalar implicature computation.

1.1.2 Nonconventionality

Grice made a distinction between conventional and conversational implicatures. In contrast to conversational implicatures, conventional implicatures arise as a word's or expression's agreed-upon, conventional meaning. An example is the word *but*, which does more than simply contribute to sentence meaning. Even though for (2a) to be true nothing else is required than that (2b) be true, there is a difference between the two.

(2) a. She is poor *but* honest.

b. She is poor *and* honest.

The difference between (2a) and (2b) is that *but* indicates a contrast between being poor and being honest. This contrast is not part of what is said (i.e. the truth conditions are the same for both sentences[2]), nor is it entailed by what is said. It is not a conversational implicature either, because it does not have to be worked out on the basis of conversational maxims. Rather, it depends solely on the conventional meaning of the word *but* - it is a conventional implicature (Bach (1999)).

[2]But c.f. Potts (2005) for an elaborate account of how conventional implicatures contribute to truth-conditional meaning.

Conversational implicatures on the other hand arise from "certain general features of discourse" alone (Grice (1975): 45). These general features are given by the conversational maxims, on the basis of which conversational implicatures are worked out.

The class of conversational implicatures is further divided into *generalized* and *particularized* conversational implicatures. Generalized conversational implicatures arise in the same systematic manner independently of context. In example (1), the scalar implicature that not all rides were perfect powder rides is a generalized conversational implicature. It would have also arisen if Guido's utterance had been the answer to the question in (3).

(3) Clara: Did Liz like the off-piste?
Guido: Some rides were perfect powder rides.

In addition, Guido's answer in (3) gives rise to the particularized implicature in (4b).[3] However, it not longer implicates that Guido remembers the ski week, which it did when uttered in a context like (1)[4].

(4) a. Not all rides were perfect powder rides.
b. Liz loved the off-piste.
c. # I remember the ski week.

Further examples of generalized conversational implicatures, are the following (5) and (6) taken from Levinson (2000)):

(5) John turned the key and the engine started.
$\rightsquigarrow$ First John turned the key and then the engine started[5].

(6) John caused Bill to die.
$\rightsquigarrow$ John killed Bill indirectly, not in a prototypical manner.

[3]For all non-skiers: for this implicature to arise, it is necessary to understand that *perfect powder rides* denote the ultimate skiing experience.

[4]I use the symbol # throughout to indicate *does not implicate*.

[5]I will use the symbol $\rightsquigarrow$ throughout to indicate *implicates*.

(7) Ralph or Tony went to the party.
$\rightsquigarrow$ Either Ralph or Tony went to the party, but not both.

In example (5), the implicature that John's turning the key occurred before the engine's starting can be attribute to the maxim of Orderliness, which requires that events be told in the sequence in which they occur. Logically, the order of the conjuncts in a statement $p \wedge q$ does not carry any temporal implications.[6] However, when employed in natural language, the first conjunct usually describes an event that precedes the event described by the second conjunct (Partee (1984)).

In example (6), use of the longer *caused ... to die* over the briefer alternative *killed* implicates a non-prototypical process of death causation. Had John killed Bill directly, the maxim of Brevity would have obliged the speaker to use the shorter *killed*. Assuming the speaker is being cooperative, she must have a good reason for violating the maxim of Brevity, e.g. the wish to indicate the non-stereotypicality of John's involvement in Bill's death, complying with the maxim of Relation. Thus, the hearer may infer that John did not kill Bill in a prototypical manner (Levinson (2000)).

Example (7) is an additional example of scalar implicature. The scale involved is ⟨and, or⟩, where *and* is the stronger element, as $p \wedge q$ logically implies $p \vee q$.[7] Analogous to the example of scalar implicature presented above, the maxim of Quantity-1 requires the speaker to employ the stronger scalar alternative. However, the maxim of Truthfulness takes precedence over Quantity-1. Consequently, the speaker, knowing that in fact only Ralph or Tony went to the party, but not both, uses the weaker alternative *or*, which is pragmatically enriched to give rise to the exclusive (*not both*) meaning that the logical *or* operator does not have. Scalar implicature, though only one example of generalized conversational implicature, is the prototypical one.

[6]The logical *and* operator $\wedge$ is commutative, i.e. $p \wedge q$ is semantically equivalent to $q \wedge p$

[7]The logical *or* operator $\vee$ has an inclusive meaning, i.e. p $\vee$ q is not only true if a) p is true and q is false and b) p is false and q is true, but also if c) both p and q are true.

One of the key features of generalized conversational implicatures, according to Grice, is their seeming context-independence, as they arise in the same manner across differing contexts. For example, *p and q*, where p and q denote events, usually implicates that the event denoted by p occurred before the one describe by q. Similarly, *some p are f*, where p denotes individuals and f denotes a property, usually implicates that not all p are f; *p or q* usually implicates that either p is the case or q is the case, but not both.

In contrast to the scalar implicature that not all rides were perfect powder rides in (1), the implicature that Guido remembers the ski week belongs to the class of particularized conversational implicatures, which are clearly context-dependent. That Guido remembers the ski week is certainly not implicated if his utterance that some rides were perfect powder rides is the answer to the question whether Liz liked the off-piste, as in example (3) - no ski week was introduced in the short discourse that anything might be implicated about. In this context though, Guido's utterance gives rise to a different particularized conversational implicature, namely that Liz loved the off-piste.

These two classes of conversational implicatures thus differ significantly in their degree of context-dependency. While generalized conversational implicatures are taken to arise automatically in an almost conventionalized manner, particularized conversational implicatures are subject to the current context (Levinson (2000)).

However, the distinction between generalized and particularized conversational implicature is not accepted by all researchers. In particular, proponents of Relevance Theory, which I will elaborate on in section 1.2.2, believe that all implicatures are equally context-dependent and arise as the result of the same pragmatic processing mechanism. Thus, that generalized conversational implicatures arise automatically and independently of context is taken to be a faulty assumption and the distinction between generalized and particularized conversational implicature is rejected (Carston (1998), Wilson and Sperber (1995)).

I remain fairly neutral with respect to this distinction. While on the one hand acknowledging that it captures a difference between two classes of implicatures descriptively, it may well be that language users employ the same processing mechanisms for both generalized and particularized implicatures.

1.1.3 Cancelability

A key property of implicatures, one that distinguishes them from entailments, is that they are *cancelable*. This is not true of entailments. Sentence (8) entails that there is a cube on the table.

(8) There is a red cube on the table.
→ There is a cube on the table.

It is not possible to make sense of the sentence *There is a red cube on the table, but there is no cube on the table*, i.e. it is not possible to cancel a semantic entailment. The situation is different for implicatures: although the sentence *Some rides were perfect powder rides* conversationally implicates that not all rides were perfect powder rides, it is possible to explicitly cancel this implicature by adding an extra clause, as in *Some rides were perfect powder rides, in fact all of them were.*

Besides being *explicitly* cancelable, scalar implicatures are also *implicitly* cancelable, namely when they are not licensed by the current context. An example of implicit cancellation is example (9) (taken from Levinson (2000): 51).

(9) John: Is there any evidence against them?
Peter: Some of their identity documents are forgeries.
Not all of their identity documents are forgeries.

In this context, John's question is taken to have explicitly fixed the level of expected informativeness. It suffices for him to know that *at least some* of their identity documents are forgeries - whether all of them are is deemed irrelevant. Thus, the *not all* implicature is implicitly cancelled in compliance with Quantity-2, which demands that the speaker should not make her contribution more informative than required.

Contexts in which a scalar implicature is implicitly canceled have been called *lower-bound* contexts (Katsos, Breheny, and Williams (2005)). Conversely, *upper-bound* contexts are those that license the implicature. Given the same sentence in two different contexts, whether this sentence gives rise to a scalar implicature depends on whether it occurs in an upper-bound or lower-bound context. There are two factors that influence a context's boundedness: the implicature trigger's structural context (syntactic) and the extra-sentential conversational context (pragmatic). In example (9) it is the extra-sentential context, i.e. the interlocutors' conversational expectations and compliance with the conversational maxims, that prevents the implicature from arising. Examples of sentences that, independently of being embedded in a conversational context, do not give rise to scalar implicatures due to syntactic constraints are the following (the involved scales are given in angular brackets):

(10) It is not the case that Ralph or Tony went to the party.
Ralph and Tony went to the party.
Scale: ⟨and, or⟩

(11) It is not the case that some of the students failed.
All of the students failed.
Scale: ⟨all, some⟩

(12) Sarah doubts that Mario will try to climb the mountain.
Sarah believes that Mario will succeed in climbing the mountain.
Scale: ⟨succeed, try⟩

(13) Do you believe that he will drop by later?
Do you believe, but not know, that he will drop by later?
Scale: ⟨know, believe⟩

Structural contexts that have been identified as blocking scalar implicatures are - among others - negation, the antecedent of conditionals, embedding under negative propositional attitude verbs, polar questions, 'before'- and 'without'-clauses (Chierchia et al. (2004)). A theory of scalar implicatures must be able to account not only for the presence of implicatures in upper-bound contexts, but also for the absence of implicatures when using the same expression in lower-bound contexts, by providing an account of the cancellation mechanisms at work.

1.1.4 Reinforceability

In addition to being cancelable, scalar implicatures are *reinforceable*. It is often possible to add explicitly what is implicated anyway, without the sense of redundancy that arises when repeating an expression's coded content (Levinson (2000)). Consider again our example of the red cube above, repeated here as (14).

(14) There is a red cube on the table.

(15) There is a red cube on the table and there is a cube on the table.

In (15) we add an extra clause to the original sentence that makes the entailment of there being a cube on the table explicit. Assume that *a red cube* and *a cube* are co-referential. This sentence is strange - upon hearing the first conjunct, we already know that the proposition denoted by the second conjunct is true. Repeating it is thus a violation of the maxim of Relation, which requires that one not make irrelevant contributions.

The same is not true of scalar implicatures. Consider the following examples:

(16) Some rides were perfect powder rides, *but not all of them were*.
Scale: ⟨all, some⟩

(17) Ralph or Tony went to the party, *but they didn't both go*.
Scale: ⟨and, or⟩

(18) Rene tried to climb the mountain, *but he didn't succeed*.
Scale: ⟨succeed, try⟩

(19) Liz believes that he will drop by, *but she doesn't know for sure*.
Scale: ⟨know, believe⟩

Italics indicate the reinforced implicatures. Note that none of these cases give rise to the sense of redundancy that (15) achieves. This is thus a further difference between implicatures and entailments.

1.1.5 Nondetachability

Scalar implicatures are *nondetachable*, which is to say that any expression that carries the same coded content will carry the same scalar implicatures. This means that it is not possible to say the same thing in a different way without also giving rise to the same implicature. This is so because scalar implicatures arise in virtue of what is said, not because of the manner of expression. Grice's example is *try*, which carries some notion of failure, or the potential of failure, as in *Rene tried to climb the mountain* vs. *Rene climbed the mountain*; "this implicature would also be carried if one said *A* [Rene] *attempted to do x, A* [Rene] *endeavored to do x,* or *A* [Rene] *set himself to do x*" (Grice (1978): 185).

That scalar implicatures are nondetachable is directly related to the nature of the maxims involved in their calculation: Quantity-1 and the Quality maxims are so-called *information-selecting* maxims (Matsumoto (1995)). Information-selecting maxims determine the choice between expressions that differ in meaning, e.g. that influence whether a stronger or weaker element on a scale is used. They are distinguished from those maxims that govern linguistic form rather than semantic content, such as the maxim of Brevity. For instance, the utterances *It is possible that the plane will be late* and *It is not impossible that the plane will be late* have the same semantic content, namely that it is possible that the plane will be late, and only differ in their linguistic form. However, the maxim of Brevity requires uttering the former, as it is briefer than the latter. Compare this to the maxim of Quantity-1, clearly an information-selecting maxim, which requires making the most informative statement possible. It governs the choice of the stronger *All men are mortal* over the weaker *Some men are mortal*, which carry different semantic content (the former is a statement about the full set of men, the latter possibly only of a subset). Thus, since information-selecting maxims govern the amount and kind of information conveyed in an utterance, and information-selecting maxims are involved in implicature generation, one would expect implicatures to be nondetachable.

1.1.6 Universality

Finally, scalar implicatures are deemed *universal* - they should arise cross-culturally in virtue of being derived ultimately from fundamental presumptions of rationality expressed by the conversational maxims. Unlike coded meaning they are not arbitrary, they are motivated. However, that these rationality assumptions and the conversational maxims may really be taken to hold cross-culturally is a matter of debate (Keenan (1976) as reported in Gazdar (1979) and Carston (1998))

1.2 Theories of scalar implicature

1.2.1 The Neo-Gricean cluster

The starting point for Neo-Gricean accounts is, as the name says, Grice's theory. What is modified is the number or form of the maxims assumed to be involved in cooperative discourse (e.g. Gazdar (1979), Horn (1984), Levinson (2000)) or an ordering is imposed on Grice's original set of maxims (Matsumoto (1995)). I present two exemplary neo-Gricean approaches to scalar implicatures in some superficial detail, namely Levinson (2000) and Matsumoto (1995).

1.2.1.1 Levinson

Levinson's goal is to give an account of utterance-type maening, an additional kind of meaning which he claims lies between sentence-type meaning and utterance-token meaning. It is more than encoded linguistic meaning but less than the full interpretation of an utterance. Scalar implicatures, and generalized conversational implicatures in general, belong to this level of utterance-type meaning. Although licensed by pragmatic principles, they aren't based on direct computations about the speaker's intentions, but rather on "general expectations about how language is usually used" (Levinson (2000): 22). These general expectations give rise to preferred or default interpretations of certain utterances, and depend on the language's structure. So, for example, the term *some* carries a default rule licensing the *not all* implicature.

Levinson develops three informativeness principles that are based roughly on Grice's Quantity and Manner maxims. Every generalized conversational implicature falls into one of three classes, depending on the principle it is licensed by.

(20) The Q-Principle:

Speaker's maxim: Do not provide a statement that is informationally weaker than your knowledge of the world allows, unless providing an informationally stronger statement would contravene the I-Principle.

Recipient's corollary: Take it that the speaker made the strongest statement consistent with what he knows.

(21) The I-Principle:

Speaker's maxim: Say as little as necessary; that is, produce the minimal linguistic information sufficient to achieve your communicational ends (bearing Q in mind).

Recipient's corollary: Amplify the informational content of the speaker's utterance, by finding the most *specific* interpretation, up to what you judge to be the speaker's M-intended point.

(22) The M-Principle:

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

Recipient's corollary: What is said in an abnormal way indicates an abnormal situation.

These principles are best conceptualized as heuristics that guide interlocutors' behavior in discourse. Recall examples (5), (6), and (7), repeated here as (23), (24), and (25). These are all examples of generalized conversational implicatures.

(23) John turned the key and the engine started.
⇝ First John turned the key and then the engine started[8].

(24) John caused Bill to die.
⇝ John killed Bill indirectly, not in a prototypical manner.

(25) Ralph or Tony went to the party.
⇝ Either Ralph or Tony went to the party, but not both.

The implicature that the engine started *after* John turned the key in example (5) can be attributed to the I-Principle. Due to the hearer's expectation that what is said in a simple or unmarked way represents a stereotypical situation, he automatically enriches the expression *P and Q* to *P and then Q*.

Conversely, the implicature in example (24) is attributed to the M-Principle. Using the marked expression *caused...to die* instead of the simpler *killed* is interpreted as indicating an abnormal, nonstereotypical situation. Thus, the implicature that John killed Bill in an indirect manner (e.g. by continually giving him a higher dose of his prescribed medication) arises.

It is the remaining, the Q-Principle, which licenses scalar implicatures. The applicability of this principle relies on computing a set of clearly established stronger (S) and weaker (W) salient contrasts that form a scale. When an expression is identified as being the weaker member of such a scale, the hearer, assuming that the speaker made the strongest statement consistent with his knowledge, automatically enriches W to¬S. In example (25), the salient scale is ⟨and, or⟩: the statement that Tony *and* Ralph went to the party is informationally stronger than the statement that Tony *or* Ralph went to the party. Thus, *or* being the weaker scale member W, it is automatically enriched by the hearer to *not both* (¬S) upon being encountered in an utterance.

Levinson distinguishes generalized conversational implicatures, which he takes to belong to utterance-type meaning, from particularized conversational

[8] I will use the symbol ⇝ throughout to indicate *implicates*.

implicatures, which he positions on the level of utterance-token meaning. While generalized conversational implicatures are to arise effortlessly, by default, and across contexts, their particularized cousins are relative to the occurrent context. Whether a particularized conversational implicature arises depends on some unspecified maxim of relevance which is responsive to contextual circumstances.

Thus Levinson's is a pragmatic Default view. In a first step, all encountered implicature triggers are automatically pragmatically enriched according to the Q-, M-, and I-Principle (level of utterance-type). This shows that, although a pragmatic phenomenon, scalar implicatures' default nature "causes them to be deeply entangled in grammar and semantics" (Levinson (2000): 169). In a second step, the contextual assumptions are taken into account, potentially canceling the previously generated implicatures if they are not licensed by the context (level of utterance-token). This subsequent cancellation is taken to be an effortful process. Consider the following examples.

(26) Anna: Did all the parents pick up their kids from school?
Ellen: Some did.
$\rightsquigarrow$ Not all parents picked their kids up from school.

(27) Anna: Did any parents pick up their kids from school?
Ben: Some did.
Not all parents picked their kids up from school.

The implicature that not all parents picked up their kids in example (26) is deemed cognitively less effortful to derive than the absence of the same implicature in example (27), because the latter requires an additional cancellation step.

1.2.1.2 Matsumoto

Instead of developing a new system of principles, as Levinson does, Matsumoto (1995) imposes an ordering on Grice's original maxims by introducing the Conversational Condition. The Conversational Condition is intended to predict when the use of a weaker expression (W) instead of a stronger alternative (S) gives rise to a scalar implicature. S and W must constitute an ordered pair of expressions ⟨S, W⟩, which constitutes a *possible* scale. In context, a possible scale may give rise to an implicature, in which case it is a *functional* scale.

(28) The Conversational Condition
The choice of W instead of S must not be attributed to the observance of any information-selecting Maxim of Conversation other than the Quality Maxims and the Quantity-1 Maxim (i.e. , the Maxims of Quantity-2, Relation, and Obscurity Avoidance, etc.).

This is another way of expressing that, if the use of W can be attributed to Quantity-2, Relation, Obscurity Avoidance and possibly other information-selecting maxims, then W will not serve as a trigger for the scalar implicature ¬S. Thus, in every utterance that contains the weaker member W of a possible scale, one must determine if the use of W is in fact attributable to any information-selecting maxim other than Quantity-1 and the Quality maxims. If it is not, the implicature will arise.

Consider the following exchange, a variant of which is first discussed in Grice (1975) and taken up by Matsumoto. Louis and Claudia are planning a trip to Canada, both know that Claudia wants to visit her friend Pierre.

(29) Louis: Where does Pierre live?
Claudia: Somewhere in the South of Canada.
⇝ Claudia does not know which town Pierre lives in.

Using the weaker *somewhere in the South of Canada* as opposed to a more informative town name violates Quantity-1. However, because Claudia does not have evidence for the exact town Pierre lives in, using a town name would violate Quality-1. The Quality maxims, as mentioned in section 1.1, may not be infringed to satisfy Quantity-1. Because the use of *somewhere in the South of Canada* can further not be attributed to any other information-selecting maxims, the implicature arises.

What if the same exchange took place in a slightly different context? Say Claudia knows that Louis is not very proficient in Canadian geography. In this case Claudia's use of *somewhere in the South of Canada* instead of e.g. *Kingston* can be attributed to the maxim of Obscurity Avoidance - had she said *Kingston* she would have violated the maxim of Obscurity Avoidance by employing a term unfamiliar to Louis. Thus, the implicature does not arise.

Alternatively, consider the situation that Louis and Claudia are planning a trip, not to Canada, but around the world (they still want to visit Pierre). Then, if Louis' question is solely a request for information about Pierre's country of residence, Claudia's use of the weaker scale member can be attributed to Quantity-2: in order to avoid being over-informative, she does not name the exact town. Here, again, Louis' question creates a lower-bound context, and the implicature that Claudia does not know Pierre's exact whereabouts does not arise.

As a consequence of his deliberations, Matsumoto reduces the difference between generalized and particularized implicatures to "the question of whether or not the intrinsic semantic relationship between S and W is such that S usually satisfies the Conversational Condition when W is uttered" (Matsumoto (1995): 55). Thus, under this account generalized conversational implicatures can only be viewed as being relatively context-independent, due to the fact that satisfaction of the Conversational Condition is itself always context-dependent.

1.2.1.3 Commonalities and differences

I have presented two substantially different accounts of scalar implicatures, which both carry the label *Neo-Griceanism*. While both proponents agree that some set of conversational principles guides speakers' utterance planning, they disagree as to the nature of those principles. For Matsumoto, it is the original Gricean maxims plus the Conversational Condition. Levinson, on the other hand, develops the Q- and I-principles, similar to Horn (1984)'s Q- and R-principles, to predict the arising of scalar implicatures.

I now turn to a self-proclaimed post-Gricean account of scalar implicatures, which in contrast to the two neo-Gricean accounts presented rejects the involvement of conversational maxims or principles in discourse.

1.2.2 Relevance Theory

Proponents of Relevance Theory share Grice's original intuition that utterances raise expectations of relevance. However, they question whether the Cooperative Principle and any set of maxims are necessary or even appropriate to provide a cognitively realistic account of utterance interpretation (Wilson and Sperber (1995), Carston (1998)). They further reject the distinction between generalized and particularized conversational implicatures - according to Relevance Theory, all pragmatic processes are equally context-dependent. Whether an implicature arises is ultimately determined by the Cognitive Principle of Relevance, which is a meta-principle of cognition meant to replace the Gricean apparatus as a whole.

(30) Cognitive Principle of Relevance:
Human cognition tends to be geared to the maximization of relevance.

A given utterance is processed by integrating the salient context with the assumption that the utterance itself is maximally relevant to the discourse. This process yields full-blown utterance interpretation. Therefore, the utterance will be interpreted as carrying an implicature if the implicature is conducive to relevance.

In Relevance Theory, relevance is defined in terms of cognitive effects and processing effort (Wilson and Sperber (1995)):

(31) Other things being equal, the greater the positive cognitive effects achieved by processing an input, the greater the relevance of the input to the individual at that time.

(32) Other things being equal, the greater the processing effort expended, the lower the relevance of the input to the individual at that time.

A positive cognitive effect is characterized as a worthwhile difference to a person's representation of the world, a true conclusion that matters to that person for example by answering a question, settling a doubt or correcting a mistaken impression. False conclusions are also cognitive effects, but not positive ones.

Processing effort is defined as the "effort of perception, memory, and inference" (Wilson and Sperber (1995): 2) required in the derivation of cognitive effects from an incoming stimulus. The greater the processing effort, it is claimed, the less rewarding the input will be to process, and thus the less deserving of the hearer's attention.

In interpreting an utterance, the hearer assumes a) that the speaker's utterance is relevant enough to be worth processing and b) that the utterance is the most relevant one compatible with the speaker's abilities and preferences. On the basis of these assumptions, the hearer then follows "a path of least effort in computing cognitive effects" (Wilson and Sperber (1995): 8) by testing interpretive hypotheses (implicatures, disambiguations, reference resolutions, etc.) in order of accessibility. She stops this procedure when her expectations of relevance are satisfied. The process of utterance interpretation is thus the simultaneous maximization of cognitive effects and minimization of processing effort.

How does this work when applied to the derivation of scalar implicatures? Consider the following exchange:

(33) Alex: Are all of your friends Buddhist?
Mo: Some are.
$\rightsquigarrow$ Not all of Mo's friends are Buddhist.

That Mo has some friends that are Buddhist is relevant enough to be worth Alex's attention (as indicated by the question). However, it is not sufficient to satisfy Alex's expectations of relevance. Presumably, Mo was able, and not reluctant, to tell Alex whether *all* of his friends are Buddhist[9], and that would have been more relevant to Alex (as indicated by the use of *all* in his question). As Mo did not say that all of them are Buddhist, Alex is entitled to interpret Mo

[9] Assuming that Mo is aware of all of his friends' religious affiliations.

as meaning that only *some, but not all* of his friends are Buddhist.

An almost identical exchange, which however is taken to give rise to a different implicature, is the following (discussed in Carston (1998), but with changed interlocutors' names - A and B in the original).

(34) Alex: Are some of your friends Buddhist?
Mo: Some are.
Not all of Mo's friends are Buddhist.

Here, Carston says, there is no reason to suppose that Mo is implicating that not all of her friends are Buddhist. Rather, Alex has lower expectations of relevance, indicated by explicitly requesting information about only *some* of Mo's friends.

Under the Relevance Theoretic account, the arising of scalar implicatures is reduced to the application of the Principle of Relevance. If a given utterance containing a scalar implicature trigger is more relevant (in terms of positive cognitive effects and processing effort) *with* than *without* the scalar implicature, then it will arise.

Relevance Theory's main problem is vagueness: it is not clear how to quantitatively grasp the notion of cognitive effects and processing effort. Further (and as a result of this), it remains unclear which kind and number of cognitive effects is worth which kind and amount of processing effort. This makes it difficult to arrive at any useful predictions (c.f. Levinson (2000) for an elaborate criticism).

1.2.3 Chierchia's grammatical account

In his 2004 paper "Scalar Implicatures, Polarity Phenomena, and the Syntax/Pragmatics Interface", Chierchia provides a localist, structural account of scalar implicatures. He develops a formalism that allows for computation of scalar alternatives in tandem with the compositional syntax and semantics of a given sentence, thus pushing a part of pragmatics into the realm of grammar. Scalar implicatures are computed locally (subsententially) at scope sites of type *t* instead of globally, which was Grice's original intention. Further, scalar implicatures, under his account, arise only in upward-entailing contexts and are systematically suspended in downward-entailing environments. In the following I explain the notion of upward- and downward-entailing contexts and present Chierchia's formal mechanism.

Upward entailing contexts are generated by upward-entailing (or monotone increasing) items such as *some, every, at least*, which allow for conclusions from the denotation of a verb phrase VP_1 to a verb phrase VP_2, where the denotation of VP_1 is a subset of the denotation of VP_2 (Barwise and Cooper (1981)). For example, if VP_1 is the phrase *entered the race early* and VP_2 the phrase *entered the race*, then sentence (35) entails sentence (36).

(35) Some Republican entered the race early.

(36) Some Republican entered the race.

Downward entailing items such as *no, few, at most* create downward entailing contexts, where the entailment relation is reversed - from superset to subset. Thus, sentence (37) entails sentence (38).

(37) No Republican entered the race.

(38) No Republican entered the race early.

Generally, a function f is downward entailing iff it licenses inferences from a set to its subsets (i.e. iff $f(A)$ entails $f(B)$ whenever $B \subseteq A$).

Chierchia points to downward entailing contexts, which are empirically identified as licensing the downward entailing item *any*, as by default not licensing scalar implicatures. His aim in presenting support for this claim is to make the case for a localist view of implicatures. This position stands in contrast to the traditional globalist view, which assumes a modular view of language processing. In particular, this means that for implicatures to be computed (as part of the pragmatic module), the whole uttered sentence must be processed by the semantic module first (and in consequence, by all lower-level models). A consequence of this is that implicatures cannot arise at the subsentential level. Chierchia's data shows that a) implicatures actually *do* arise subsententially (locally) at scope sites of type t and are projected upwards to larger embedding structures, and b) that implicatures are suspended in downward entailing contexts.

(39) supports claim a).

(39) John believes that some students are waiting for him.
$\rightsquigarrow$ John believes that not all students are waiting for him.

In this example, if the implicature arose globally, it would be that *it is not the case that John believes that all students are waiting for him*. However, this is a weaker implicature compared to the one that is in line with our intuitions, namely the one in (39).[10] Chierchia concludes that implicatures are introduced locally, as soon as possible, in the same order in which their triggers, i.e. the scalar terms, are introduced in the syntactic tree.[11]

(40a) - (46a) are examples of sentences with implicature triggers in downward entailing contexts. None of these give rise to the potential implicatures in (40b) - (46b), supporting claim b).

(40) Negation
   a. It's false that Sue harassed some students.
   b. # Sue harassed all students.

(41) Restriction of downward entailing quantifiers: *no*
   a. No student with an incomplete or failing grade is in good standing.
   b. # While a student with both is.

(42) Scope of downward entailing quantifiers: *no*

[10]The former is compatible with worlds in which there are no students, whereas the latter rules out these worlds.

[11]See Russell (2006) for a rebuttal of this claim. He shows that the derivation of the implicature in (2) from (1) may be achieved pragmatically by introducing the assumption that not only the speaker of the utterance is epistemically adept, but also the subject of the belief-clause.

(1) Mary does not believe that all F are G

(2) Mary believes that not all F are G

a. No student who missed class will take the exam or contact the advisor.

b. # They will do both.

(43) Restriction of *every*

a. Every student who wrote a squib or made a classroom presentation got extra credit.

b. # But not every student who did both got extra credit.

(44) Antecedent of conditionals

a. If Paul or Bill come, Mary will be upset.

b. # But if Paul and Bill both come, Mary won't be.

(45) Negative embedding predicates: dubitatives

a. John doubts that Paul or Bill are in that room.

b. # He doesn't doubt that they both are.

(46) Negative embedding predicates: negative factives

a. John regrets that Paul or Bill are in that room.

b. # He doesn't regret that they both are.

Additional contexts mentioned by Chierchia, Guasti, Gualmini, Meroni, Crain, and Foppolo (2004) are negative factives, negative propositional attitudes, predicates of minimum requirement, polar questions, comparatives, modality of permission, and imperatives. Further, these contexts not only cancel regular scalar implicatures, they additionally introduce new ones on a reversed scale - Chierchia calls this a *recalibration process*. Consider example (47):

(47) In this class, no one read many papers from the reading list.

a. # . . . but someone read all of them.

b. $\rightsquigarrow$ . . . but someone read some of them.

In such contexts, new implicatures are created by the interaction of the embedding negative element and the embedded scalar terms. The relevant scale is the following:

(48) ⟨no student read some paper, no student read many papers, no student read every paper⟩

Entailment holds from left to right. Thus, uttering (47) implicates the negation of the immediately stronger alternative as in (49a), which is equivalent to (2.3).

(49) a. ¬no student read some paper

b. some student read some paper

Implicatures such as in (49a) and (2.3), which arise from an implicature trigger's being embedded in a downward entailing element, Chierchia calls *indirect* implicatures (as opposed to *direct* implicatures that arise in positive contexts).

Chierchia assumes that Logical Forms are interpreted compositionally in the way familiar from the generative tradition. Thus the computation of scalar implicatures proceeds in five steps:

1. Identify the relevant alternatives to a given expression α
2. Compute the strengthened value of α
3. Introduce direct implicatures at scope sites if licensed.
4. Remove implicatures upon encountering a downward entailing operator.
5. Introduce indirect implicatures.

I will discuss each of these steps in turn.

1. Every expression α has a plain value $[\![\alpha]\!]$. In addition, each expression has a "scalar" or "strengthened" value $[\![\alpha]\!]^{S}$. This strengthened value is recursively computed by exploiting the relevant alternatives $[\![\alpha]\!]^{ALT}$. For every α, $[\![\alpha]\!]^{ALT}$ must be of the same type as α and yields only the alternatives introduced by the last scalar element in the tree, with the alternatives to scalar elements in sub-expressions of α having been picked up on the way.

For α lexical:

$$[\![\alpha]\!]^{\text{ALT}} = \begin{cases} \{\alpha_1, \ldots, \alpha_n, \ldots\} & \text{if } \alpha \text{ is part of a scale } \langle \alpha_1, \ldots, \alpha_n, \ldots \rangle \\ \\ \{\alpha\} & \text{otherwise} \end{cases}$$

For $\alpha = [\beta\ \gamma]$, where β is of functional and γ of argumental type:

$$[\![\alpha]\!]^{\text{ALT}} = \begin{cases} \text{Ap}(\{[\![\beta]\!]\}, [\![\gamma]\!]^{\text{ALT}}) & \text{if } \beta^{\text{ALT}} \text{ is a singleton} \\ \\ \text{Ap}([\![\beta]\!]^{\text{ALT}}, \{[\![\gamma]\!]\}) & \text{otherwise} \end{cases}$$

Thus the alternatives for sentence (50a), which contains the two scalar items *some* and *or*, are given in (50b).

(50) a. Some students smoke or drink.

b. Alternatives: [some student$_i$ [t$_i$ smoke or t$_i$ drink]]$^{\text{ALT}}$ =
{ some'(student')(smoke' ∨ drink'), many'(student')(smoke' ∨ drink'), all'(student')(smoke' ∨ drink') }

2. From the set of computed alternatives we want to compute the one that is immediately stronger than the target (which will be negated in the implicature). Thus, given a set of alternatives A and a member of this set β, S_β(A) will be the weakest member of A that asymmetrically entails β. If A does not contain an element that asymmetrically entails β, S_β(A) will be the contradiction $\perp$. For *some students smoke or drink*, S_β(A) is shown in (51).

(51) $\text{S}_{\text{some students smoke or drink}}$(some student smoke or drink$^{\text{ALT}}$) = many students smoke or drink

3. The following definition allows us to compute implicatures from the target and the immediately stronger alternative at scope sites ϕ of type *t*.

(52) $[\![\phi]\!] = [\![\phi]\!]^{\text{S}} \wedge \neg\text{S}(\phi^{\text{ALT}})$

Thus the strengthened value at scope site ϕ is the plain value at ϕ in conjunction with the negation of the immediately stronger alternative to ϕ. A simple example is provided in (53).

(53) John saw some students.

a. $[\![[$ some student$_i$ [John saw t$_i$] $]]\!]^S$ = $[\![[$ some student$_i$ [John saw t$_i$] $]]\!]^S \wedge \neg$ S([some student$_i$ [John saw t$_i$]]$^{\text{ALT}}$)

b. $[\![[$ some student$_i$ [John saw t$_i$] $]]\!]^S$ = some'(student')(λx. saw'(j,x)) $\wedge\neg$ S([some student$_i$ [John saw t$_i$]]$^{\text{ALT}}$)

c. $[\![[$ some student$_i$ [John saw t$_i$] $]]\!]^S$ = some'(student')(λx. saw'(j,x)) $\wedge\neg$ every'(student')(λx. saw'(j,x))

4. Since an expression's strong value should not become weaker than its plain value - what Chierchia calls the *Strength Condition* - we must remove implicatures in downward entailing contexts. This constraint is checked at each step of the derivation and is achieved by comparing an expression's potential strong values with its plain value. To account for this, Chierchia defines functional application as follows:

Strong application (preliminary version):

Suppose $\alpha = [\beta\ \gamma]$, where β is of type <a,b> and γ of type a. Then

$$[\![\beta\ \gamma]\!]^S = \begin{cases} [\![\beta]\!]^S([\![\gamma]\!]^S) & \text{if} [\![\beta]\!] \text{is not DE} \\ \\ [\![\beta]\!]^S([\![\gamma]\!]) & \text{otherwise} \end{cases}$$

Under this definition, if γ contains a scalar term and β is a downward entailing function, the function is applied to γ's plain value in order to comply with the Strength Condition.

5. One step is still missing, namely the introduction of indirect implicatures. This is achieved as follows:

Strong application (final version):

Suppose $\alpha = [\beta\ \gamma]$, where β is of type <a,b> and γ of type a. Then

$$[\![\beta\ \gamma]\!]^S = \begin{cases} [\![\beta]\!]^S([\![\gamma]\!]^S) & \text{if} [\![\beta]\!] \text{is not DE} \\ \\ [\![\beta]\!]^S([\![\gamma]\!]) \wedge \neg \text{S}([\![\beta]\!](\gamma^{\text{ALT}})) & \text{otherwise} \end{cases}$$

Thus, in the case of example (54a), which contains the downward entailing operator *doubt*, and the alternatives to which are given in (54b), we arrive at the meaning in (54c) which is equivalent to (54d).

(54) a. I doubt that John drinks and drives.

b. doubt + {[drink'(j) ∧ drive'(j)], [drink'(j) ∨ drive'(j)]} = {[doubt'([drink'(j) ∨ drive'(j)]), [doubt'([drink'(j) ∧ drive'(j)])]}

c. doubt'([drink'(j) ∧ drive'(j)]) ∧ ¬doubt'([drink'(j) ∨ drive'(j)])

d. doubt'([drink'(j) ∧ drive'(j)]) ∧ believe'([drink'(j) ∨ drive'(j)])

In Chierchia's words: "the theory predicts that sentences have certain default interpretations. Defaults (when incompatible with the context) can be overridden" (Chierchia et al. (2004)). An implicature trigger's meaning is automatically enriched, i.e. the scalar implicature automatically arises. If a downward entailing item is then encountered, the implicature is canceled.

Thus, Chierchia's theory predicts that the implicature arises by default in upper-bound contexts, while in lower-bound contexts it is subsequently canceled. It also arises by default in neutral contexts (contexts in which the implicature is not of interest, and which may be interpreted both with or without the implicature), because in such contexts no structural factors for canceling the default implicature are available. In section 1.3.4 we will see that, if this prediction is to be translated to processing claims, it is a false one. A further problem with this account is that it cannot deal with the absence of scalar implicatures in solely pragmatically motivated lower-bound contexts such as example (9), repeated here as (65).

(55) John: Is there any evidence against them?
Peter: Some of their identity documents are forgeries.
Not all of their identity documents are forgeries.

The sentence uttered by Peter does not contain an implicature blocking downward entailing item, thus Chierchia's purely structural account wrongly predicts the arising of the implicature that not all of their documents are forgeries.

1.2.4 Intermediate summary

Chierchia and Levinson both offer Default accounts, sharing the core idea that scalar implicatures initially arise automatically, independently of context; cancellation occurs in a subsequent step. However, Chierchia's account is a grammatical one, while Levinson's is pragmatic. Under Chierchia's view, scalar implicatures arise as part of compositional semantics, depending on the structural context. According to Levinson, on the other hand, scalar implicatures arise as part of pragmatic processing (after semantic processing) and may be attributed to an interaction between the I- and Q-principle. Matsumoto's account is a pragmatic one also, but in contrast to both Chierchia and Levinson, his account does not make any processing claims. Relevance Theory, a clearly cognitive theory of implicature processing, rejects both the Default claims made by Levinson and Chierchia and the conversational principles that feature in Levinson and Matsumoto's accounts, while retaining Grice's assumption of globalist implicature computation.

The relations between the theories, the specific assumptions they make, and whether or not they claim to be theories about the cognitive processing of implicatures will become relevant in section 3, when I consider the theoretical implications of experimental work on the processing of scalar implicatures, including the study presented in section 2.

1.3 Scalar implicature - processing

1.3.1 From theory to processing claim

Experimental Pragmatics has emerged as a new field of enquiry within the past ten years or so (Noveck (2004), Noveck and Reboul (2008)). Researchers in this field are attempting to find empirically testable predictions and develop new experimental paradigms for resolving issues in pragmatics, theoretical approaches to which reach back into the sixties. Among the main issues under experimental investigation lies the processing of scalar implicatures, first systematically investigated in children and adults by Noveck (2001) and Noveck and Posada (2003). Among the questions to resolve are the following: What are the differences in scalar implicature judgments and processing between children and adults? How do young learners acquire the skill of deriving implicatures? Do implicatures arise locally (sub-sententially) or globally? Is there experimental evidence for the distinction between generalized and particularized implicatures? Can we find experimental evidence for different scale types? How quickly are implicatures computed in comparison to semantic meaning?

All of these questions are deemed relevant to determining which of the two theoretical approaches - Neo-Griceanism or Relevance Theory - best capture the phenomenon of scalar implicature. I will briefly review the work aimed at resolving these issues with a focus on the last - the question of Default - which is the question targeted in our study.

1.3.2 Localism vs. globalism

One of the main questions being investigated is that of Localism: do implicatures arise globally (post-propositionally) after the application of syntax and semantics, as originally stated by Grice, or locally (sub-sententially)? I have presented Chierchia's theoretical arguments for local implicatures in section 1.2.3. However, this question is also experimentally interesting: can we find sub-sentential evidence for implicature generation?

Pouscoulous et al. (2007) addressed the question of Localism in a reasoning task, where participants received booklets with one task such as the following on each page (in French):

(56) a. Wilma said: 'Betty thinks that Fred heard some of the Verdi operas.' Would you infer from this that Betty thinks that Fred didn't hear all the Verdi operas?

b. Wilma said: 'Fred has to hear some of the Verdi operas.' Would you infer from this that Fred isn't allowed to hear all of them?

c. Wilma said: 'All students heard some of the Verdi operas.' Would you infer from this that none of the students heard them all?

Participants answered YES/NO by checking one of two boxes. Pouscoulous et al. (2007) tested embedding of *some* under *think, must*, and *all* and compared these judgments to judgments on unembedded implicatures. They found that implicatures plummeted from over 90% in the unembedded case to 50% for *think*, 27% for *all*, and 3% for *must*. They conclude that scalar implicatures do not arise locally.

This study is problematic in various ways. The first is that the set of embedding elements is a) very limited and b) very diverse. Inferring from reduced implicatures arising under this limited and diverse set of embedding elements that implicatures in general do not arise locally is highly problematic, especially considering that in the case of *think* they in fact did arise in half of all cases. The problem with diversity is that the embedding elements interact differently with the implicature trigger both syntactically and semantically. Thus there is no basis for comparison. Further, the task itself required much reflexion on the sentences by participants. It may well be the case that the task was simply too complex. Online measures may provide more reliable results.

The question of Localism thus remains a fairly open one. I turn now to experimental work on the acquisition of implicatures.

1.3.3 Developmental results

The main result from developmental studies on scalar implicature processing is that children below the age of 9 generally do not interpret scalar items as giving rise to implicatures (Noveck (2001), Papafragou and Musolino (2003), Guasti et al. (2005), Pouscoulous et al. (2007)). In a modal reasoning task, Noveck (2001) investigated children and adults' interpretations of statements like *x might be y*, in contexts in which the stronger statement, ie. *x must by y* were true. A puppet presented these statements and participants were asked to say whether the puppet's claim is right or not. Noveck's finding was that 5-9-year-olds treated *x might be y* logically (i.e. as compatible with the stronger statement) much more often than adults (72%, 80%, and 69% vs. 35%, though the latter is not significantly below chance).

In line with this finding, he reports that 8- and 10-year-old French children are more likely to treat *certains* (some) as compatible with *tous* (all) than adults. He concludes that children tend to be more logical than adults when interpreting scalar terms, while adults are more likely to arrive at pragmatic interpretations.

In a series of follow-up studies, Papafragou and Musolino (2003) investigated the questions of whether all scalar terms are treated in the same way by children and whether children's difficulty in treating scalar items pragmatically reflects a genuine inability to derive scalar implicatures or whether it is due to demands imposed by the experimental task.

They addressed the first question by testing both 5-year-olds and adults on three different scales in Greek: ⟨oli, meriki⟩ (⟨all, some⟩), ⟨tris, dio⟩ (⟨three, two⟩), and ⟨teliono, arxizo⟩ (⟨finish, start⟩). They presented participants with contexts which satisfied the semantic content of the stronger terms on each scale (*all, three, finish*), but described them using the weaker terms (*some, two, start*). Their findings are consistent with Noveck's: adults overwhelmingly reject the underinformative descriptions (92.5% rejections), while children almost never do for two of the three tested scales (12.5% and 10% rejections). However, children seemed to treat the numerical scale differently from the other two scales; they were more likely to reject underinformative descriptions on this scale (65% rejections).

The second question was investigated by training a group of 5-year-olds to detect underinformative statements. Training was achieved by telling them that the puppet describing the contexts, Minnie, sometimes said 'silly things' and that the game was about helping Minnie to 'say things better'. If for a given underinformative description the child failed to correct Minnie, the experimenter eventually would, simultaneously providing a pragmatically felicitous description. Children were then presented with modified versions of the stories in the first experiments (which more readily invited scalar implicatures). This modification led the children to reject underinformative descriptions significantly more often (52.5% rejections for the ⟨all, some⟩ scale, 47.5% rejections for the ⟨finish, start⟩ scale, and 90% rejections for the ⟨three, two⟩ scale).

Papafragou and Musolino (2003) conclude that children do not treat all scalar terms alike, and that whether or not a child can derive a scalar implicature is dependent on their awareness of the goal of the task.

In further follow-up studies on Italian *some*, Guasti et al. (2005) confirm the effect of experimental task demands found by Papafragou and Musolino (2003), but in addition show that children a) do not attain adult-like performance levels on identifying implicatures and b) the effect of training on implicature recognition does not persist (i.e. after re-testing one week later on the same materials but without training, rejection rates of under-informative statements dropped from 52% to 22%).

In some more recent studies, Pouscoulous et al. (2007) showed that both having a visual context against which the utterance is evaluated (compared to relying on world-knowledge) and minimizing the processing costs of generating an implicature greatly increases children's performance on underinformative descriptions.

Together, these findings are taken to show that children, rather than lacking the pragmatic *ability* to derive scalar implicatures, either lack the *processing resources* to do so or have expectations of required informativeness in discourse that differ from those of adults.

That children acquire the skill of scalar implicature computation rather late is interpreted by some as support against Default models of scalar implicature processing, to which Inow turn.

1.3.4 Default vs. context-driven computation

The question of whether scalar implicatures arise by default or as part of an effortful, context-driven process has received much attention over the last few years. There are two opposing views on this question: the Default model, often associated with researchers who hold a Neo-Gricean view, and the Context-Driven model[12], often associated with Relevance Theory (Noveck and Posada (2003)).

The Default view of scalar implicatures arises from a very specific interpre-

[12]The Context-Driven model has also been called the Underspecified model (Bezuidenhout and Morris (2004)) or the Literal-First hypothesis (Grodner et al. (submitted)).

tation of what Grice originally had to say about generalized conversational implicatures. In contrast to particularized conversational implicatures, he claimed, one can sometimes say that “the use of a certain form of words in an utterance would normally (in the absence of special circumstances) carry such-and-such an implicature or type of implicature” (Grice (1975)), i.e. the same expression usually gives rise to the same implicature across contexts. Further, as mentioned in section 1.1, generalized conversational implicatures may be canceled either explicitly, by adding a clause that negates the implicature, or implicitly, if the context is lower-bound. Henceforth, when I speak of *cancellation*, I will (unless indicated otherwise) mean *implicit cancellation*.

From this, Default view proponents conclude that generalized conversational implicatures occur automatically, by default, upon the occurrence of an implicature trigger, and thus intrude effortlessly into the interpretation of an utterance (Noveck (2004)). Under the Default model, scalar implicatures arise independently of context. In a subsequent step they are open to implicit cancellation under consideration of contextual factors. Thus, implicature generation is considered a two-step process, giving rise to generalized conversational implicatures in the first step, while adding particularized conversational implicatures and potentially canceling previously generated implicatures in the second step. Both Chierchia and Levinson’s accounts of scalar implicatures are explicitly Default accounts, and both authors have called for psycholinguistic validation of their claims (Chierchia et al. (2004), Levinson (2000)).

In opposition to the Default model stands the Context-Driven model, which does not consider scalar implicature generation to be an automatic, effortless process. The arising of scalar implicatures is seen as just one of many workings of the interaction of semantics and pragmatics (Carston (1998)). The process of pragmatic enrichment under this account is connected to additional processing effort. Additional processing resources are needed for computing not only an implicature trigger’s simple linguistic meaning, but also its additional, enriched meaning. Further, the Context-Driven model refutes the Default model’s two-step process: if an utterance’s context calls for an implicature-enriched meaning, the implicature will arise - if it does not, it will not. Under the Context-Driven model there is therefore no implicit subsequent cancellation of implicatures, they simply do not arise if they are not licensed by the context. Proponents of Relevance Theory have claimed the Context-Driven model for themselves, since under their account implicatures are calculated as part of the general mechanism of utterance interpretation, which begins with the computation of semantic meaning and pragmatically enriches the interpretation until the hearer's expectations of relevance are satisfied.

The question of Default has been tested with a number of different paradigms: sentence verification and reaction time (Noveck and Posada (2003), Bott and Noveck (2004)), reading time (Breheny et al. (2006)), and eye-tracking (Huang and Snedeker (in press), Grodner et al., Storto and Tanenhaus (2004), Bezuidenhout and Morris (2004)).

For example, Noveck and Posada (2003) conducted a single-sentence judgment task on underinformative sentences such as the following:

(57) Some elephants have trunks.

Such underinformative sentences are interesting in terms of semantic-pragmatic interpretation because they are considered to have two truth values: they are true when taken at face value (lower-bound interpretation), but false if interpreted with the scalar implicature *...but not all elephants have trunks* (upper-bound interpretation). Noveck and Posada (2003) call the lower-bound interpretation of underinformative sentences *logical*, the upper-bound interpretation *pragmatic*. They assessed differences in reation time to such sentences under a logical vs. pragmatic interpretation. Subjects were asked to respond TRUE or FALSE to underinformative sentences by pressing one of two buttons. Reaction times to press the button were measured. In addition, subjects judged control items that were either patently true (example (58)) or patently false (example (59)).

(58) Some houses have bricks.

(59) Some crows have radios.

The predictions from the Default and Context-Driven model are the following: according to the Default view, scalar implicatures are generated by default and are subsequently canceled. Therefore, pragmatic responses (FALSE) should be faster than logical responses (TRUE), because under the logical interpretation the immediately arising implicature has to be canceled in a second effortful step, thus leading to longer reaction times. Under the Context-Driven view, the generation of scalar implicatures is an effortful process, so pragmatic interpretations should take longer to arrive at than logical interpretations of *some*. The results support the Context-Driven model: reaction times of pragmatic responders were almost twice as high as those of logical responders.

Neys and Schaeken (2007) conducted a study using the same paradigm, but additionally had participants memorize dot patterns while rating sentences. Dot patterns were of two difficulties, thus manipulating the cognitive resources available for computing implicatures. The Context-Driven model predicts that participants should be more likely to generate scalar implicatures under low cognitive load, since the implicature generation process requires processing effort. Conversely for the Default model. Neys and Schaeken (2007) found that participants made more logical and fewer pragmatic judgments under higher cognitive load, thus supporting the Context-Driven model.

Similarly, Breheny et al. (2006) conducted two reading time experiments in Greek. The first experiment investigated reading times for upper- vs. loverbound contexts. Twelve pairs of texts like the following (translated) comprised the critical materials.

(60) (*Upper-bound*) 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*. /
(⇝but not both)

(61) (*Lower-bound*) John heard that / the textbook for 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. /
(# but not both)

Every pair contained one upper-bound text and one lower-bound text, consisting of two sentences each. Participants read 12 texts, where half of the materials were taken from the upper-bound, the other half from the lower-bound set. The time taken by participants to read the trigger phrases (italics in the above example) was measured by recording the time taken to press a button and move to the next text segment.

Under the Default view, if cancellation requires processing effort, reading times for the target phrases should be longer for lower-bound than for upper-bound contexts, since in lower-bound contexts the automatically generated implicature must subsequently be canceled, while in the upper-bound contexts cancellation is not necessary. If cancellation is seen simply as the decay of an activated implicature, there should be no difference in processing effort, as measured by reading time. Thus, the Default view predicts that there is either no difference in reading time for upper- vs. lower-bound contexts, or reading times are longer for lower-bound contexts.

The Context-Driven view on the other hand predicts that reading times should be longer for upper-bound contexts, as no implicature is generated in lower-bound contexts in the first place, while implicature generation in upper-bound contexts is an effortful process.

Indeed, the results support the Context-Driven model. Reading times for the target phrase with the implicature trigger in upper-bound contexts were significantly longer than for lower-bound contexts, suggesting that implicatures are not generated effortlessly per default, but rather by an effortful process.

A second experiment tested whether scalar implicatures arise in neutral contexts. In contrast to upper- and lower-bound contexts, neutral contexts are such that it is not of interest whether the utterance is interpreted as carrying the implicature or not. For these cases the Default view predicts that the implicature will arise, as the implicature arises by default and will not be canceled if (the pragmatic and structural) context does not explicitly call for it.

Thus, in comparison with upper-bound contexts, there should be no difference in reading time. The Context-Driven view predicts just the opposite; the implicature does not arise, since implicature generation is an effortful process. Consequently, reading time should be longer for upper-bound than for neutral contexts.

Again the results support the Context-Driven model. Significant differences in reading times between upper-bound and neutral contexts suggest that the implicature does not arise in neutral contexts, contrary to the Default model's prediction.

An off-line metalinguistic task testing whether participants interpret *or* inclusively or exclusively in upward-entailing plus upper-bound contexts, upward-entailing plus lower-bound contexts, and downward entailing contexts is reported in Katsos (2007). Participants were asked to rate the last sentence of short texts for whether the disjunction was exclusive or inclusive on a scale from 1 to 7, where 1 was exclusive, and 7 was inclusive.

(62) a. *Upward-entailing plus upper-bound*
The Head of the Department asked her secretary: Who is giving a lecture on Nuclear Physics for undergraduates? Her secretary replied: Williams or Tyler from the Research Laboratory.

b. *Upward-entailing plus lower-bound*
The Head of the Department asked her secretary: Who is available to give a lecture on Nuclear Physics for undergraduates? Her secretary replied: Williams or Tyler from the Research Laboratory.

c. *Downward-entailing*
The Head of the Department asked her secretary: Who is giving a lecture on Nuclear Physics for undergraduates? Her secretary replied: I expect that if Williams or Tyler from the Research Laboratory are doing so, the undergraduates will get a high quality introduction.

Participants rated disjunctions in the upward-entailing plus upper-bound contexts as rather on the exclusive side of the scale (with a mean of 2.9), while upward-entailing plus lower-bound as well as downward-entailing contexts were rated as inclusive (means of 5.2 and 5.1, respectively). Although ratings were not at the extremes of the scales, they were reliably placed at the predicted end of the scale (assuming that scalar implicatures only arise in upward-entailing plus upper-bound contexts).

These materials were subsequently tested in an on-line self-paced reading study, where Katsos (2007) found that scalar terms were read slower in the upward-entailing plus upper-bound than in the upward-entailing plus lower-bound and downward-entailing contexts (811 vs. 761 and 758 ms, respectively). Since the upward-entailing plus upper-bound context is the only one that licenses the scalar implicature, Katsos (2007) concludes that this is further support for the Context-Driven model.

Huang and Snedeker investigated the question of Default via an eye-tracking study in the visual world paradigm[13]. Participants viewed a display with four quadrants (see Figure 1.1), which typically contained a picture of a boy with two socks and a boy with nothing on the left side, and a girl with two socks (the pragmatic target, i.e. the girl had only a subset of the socks in the display) and a girl with three soccer balls (the logical target, i.e. the girl had all of the soccer balls on the display). Participants were asked to *Point to the girl with some of the socks.*

According to the Context-Driven model, participants should, upon hearing *some*, initially fixate both the logical and pragmatic target equally, since both are consistent with the logical interpretation of *some*. On the other hand, the Default model predicts that the logical target should immediately be rejected. This in turn should result in a rapid increase in fixations on the pragmatic target. The results strongly support the Context-Driven model: participants did not prefer the pragmatic target before the noun's phonetic point of disambiguation (in the example at *ks* of *socks*). In contrast, when participants were asked

[13]Since Tanenhaus et al. (1995), the visual-world paradigm has been widely employed to investigate linguistic questions.

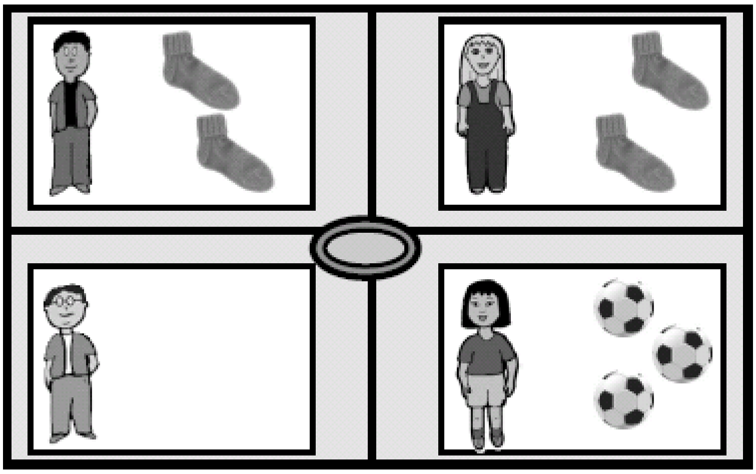


Figure 1.1: Sample display from Huang and Snedeker, for which participants were asked to *Point to the girl with* ***some of the socks/all of the soccer balls***.

to *Point to the girl with all of the soccer balls*, looks to the target increased well before the noun's phonetic point of disambiguation, suggesting that semantic meaning is indeed computed quickly, and implicatures require extra processing effort.

In a reaction to Huang and Snedeker, Grodner et al. conducted a similar experiment aimed at addressing some concerns with the Huang and Snedeker study. They state that simple *some* may not be unambiguously associated with a scalar inference, i.e. *Point to the girl with some socks* does not imply other socks in the discourse. They hypothesize that it is the partitive construction, signaled by *of*, that generates the pragmatic reading and disambiguates the target. In contrast, for *all*, the quantifier itself is sufficient to identify the target. This might lead to delayed looks to the pragmatic target in the *some* case until after the onset of *of*. In addition, participants may have had a bias to initially look at the picture with the most objects.

Grodner et al. thus modified the Huang and Snedeker study in four ways: *some of* was replaced with *summa*, the construction's natural articulation in American English. They assumed the shortened first syllable to provide an earlier phonetic signal for the partitive[14], thus making the timing more comparable

[14]Salverda, Dahan, and McQueen (2003) found evidence that listeners use the duration of the

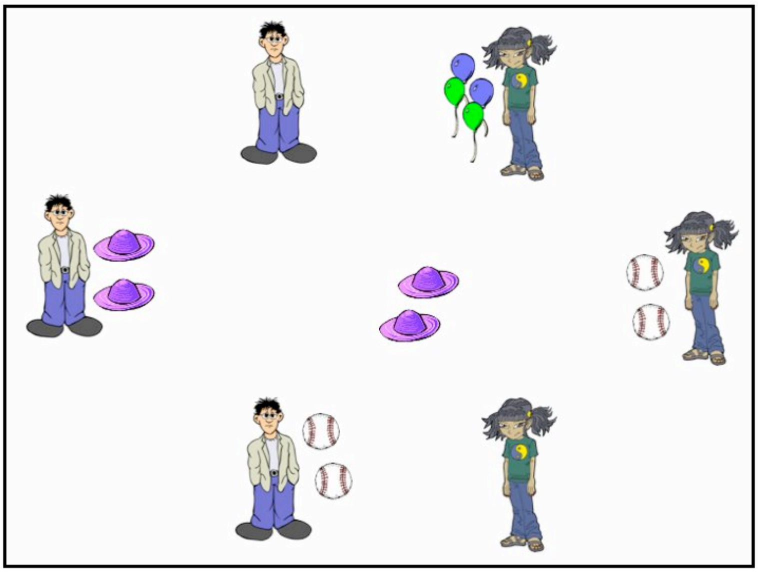


Figure 1.2: Sample display from Grodner et al., for which participants were asked to *Click on the girl with* ***summa the balls/alla the balloons/nunna the items***.

to literal controls. They additionally introduced a competitor that had none of the items (i.e. that was incompatible with both the logical and the pragmatic interpretation of *some*). Further, they included both a *nunna* (for *none of*) and *alla* (for *all of*) condition as literal controls. Finally, they included a baseline condition to compare potential early disambiguation effects of *summa* to, in which there were two individuals with some, but not all, objects of different sets of objects in the display. Figure 1.2 shows a typical display.

In contrast to previous findings, Grodner et al. found no evidence for delayed implicatures. Both in the summa and alla conditions, as well as in the nunna condition, fixations to the target increased 200-300 ms after quantifier onset. In the baseline condition, fixations did not increase until after noun onset. They interpret these results, not as lending support to the Default model, but rather as showing that pragmatic constraints affect the earliest stages of interpretation. Their explanation for the previous results supporting the Context-Driven model is that integrating the resultant interpretation - including the im-

first syllable to distinguish between monosyllabic and polysyllabic words as a vowel unfolds plicature - with the relevant information in the context may require additional processing time, not the computation of the implicature itself (although it remains unclear why this extra processing time was not necessary for participants in their study to integrate the interpretation with the context). In addition, as in the Huang and Snedeker study, they also found an initial bias to look at the individual with the most objects.

The situation is thus an interesting one. Most work to date supports the Context-Driven model, while one study clearly sticks out as not doing so. This situation in itself asks for further elucidation. But there are additional reasons for further pursuing this question: although there has been work on English (Grodner et al., Huang and Snedeker), French (Noveck (2001), Noveck and Posada (2003)), Italian (Guasti et al. (2005)), Greek (Breheny et al. (2006), Papafragou and Musolino (2003)), and Dutch (Geurts and Pouscoulous (2008)), further cross-linguistic investigation is necessary to address the claim that scalar implicatures are a universal phenomenon (see section 1.1.6). In addition, there is one problem running through all the studies on the question of Default: hearer and speaker (if there even is a speaker) are never actively engaged in dialogue. This is strange, considering that Grice's maxims were originally devised as maxims governing cooperative dialogue, not out-of-context, written sentences, which provide the target material in many of the studies mentioned above. In fact, one great improvement of the Huang and Snedeker and Grodner et al. studies over previous studies was introducing an actual speaker (though recorded) that participants listened to, more closely resembling a discourse in a natural setting.

This setting provides the motivation for the eye-tracking study conducted. The Grodner et al. study was modified in the following ways:

1. The issue of universality was addressed by conducting the study in German. The investigated scale was ⟨all, some⟩, which translates to ⟨alle, einige⟩.[15]

[15]But c.f. section 2.4 for considerations about issues related to this translation.

2. In order to a) provide a more interactive, discourse-like setting and b) reduce the baseline fixation bias on the region with the most objects, an interactive task in which participants were themselves involved in setting up the scene that the target sentences were tested on was introduced. The assumption was that setting up the scene on their own (upon hearing spoken recorded instructions) would make the contrast between moving *all* and only *some* of a set of objects salient, thus both facilitating interpretation of pragmatic *some* and reducing baseline fixations to the region with many objects during the target sentence with the scalar term.

The study is reported in the next section.

Chapter 2

The Experiment

2.1 Hypotheses

Considering all I have said in section 1.3.4, the predictions for the processing of scalar implicatures are the following (see Table 2.1 for a summary of hypotheses):

1. If generating the scalar implicature upon hearing *some* is a default process, looks to a target containing some but not all objects of a salient set (in contrast to a competitor containing all objects of a salient set) should

    - increase within 200 - 300 ms of the implicature trigger onset[1] and should
    - increase as quickly as looks to a target containing all objects of a salient set after hearing *all*. That is, there should be no measurable difference in delay of new fixations to the target for *some* as opposed to *all*.

    If, on the other hand, generating scalar implicatures is not a default process, looks to a target containing some but not all objects of a salient set should

[1]Allopenna et al. (1998) found that 200 ms after word onset is the earliest possible point at which to expect effects of any kind of information available during word recognition.

- increase later than after the first 200 - 300 ms following the implicature trigger onset and should
- not increase as quickly as looks to a target containing all objects of a salient set after hearing *all*. That is, there should be a delay for *some*, but not for *all*.

2. Furthermore, fixations should pattern differently for contexts in which the *some but not all*-implicature or the interpretation of *all* disambiguates between target and competitor vs. contexts in which the competitor contains the same number of objects as the target. Specifically, in the disambiguating contexts, new fixations to the target should increase faster for both *some* and *all* than in non-disambiguating contexts.

	some	*all*
Ambiguous	late/late	late/late
Disambiguating	early/later	early/early

Table 2.1: Predictions about the point of disambiguation for the Default/Context-Driven model. *Early* disambiguation is 200-300 ms after quantifier onset, *late* disambiguation is 200-300 ms after noun onset, and *later* disambiguation is later than 200-300 ms after quantifier onset, but earlier than 200-300 ms after noun onset.

2.2 Method

2.2.1 Participants

Participants were 31 Cognitive Science undergraduate and graduate students (17 male and 14 female) from the University of Osnabrück that responded to a call for participants. Their age ranged from 19 to 32 years. They were each paid €5 for their participation. All participants were native speakers of German and had normal or corrected-to-normal vision.

The data of five participants was discarded due to corrupt data files. Thus, all reported data analysis was for 26 participants.

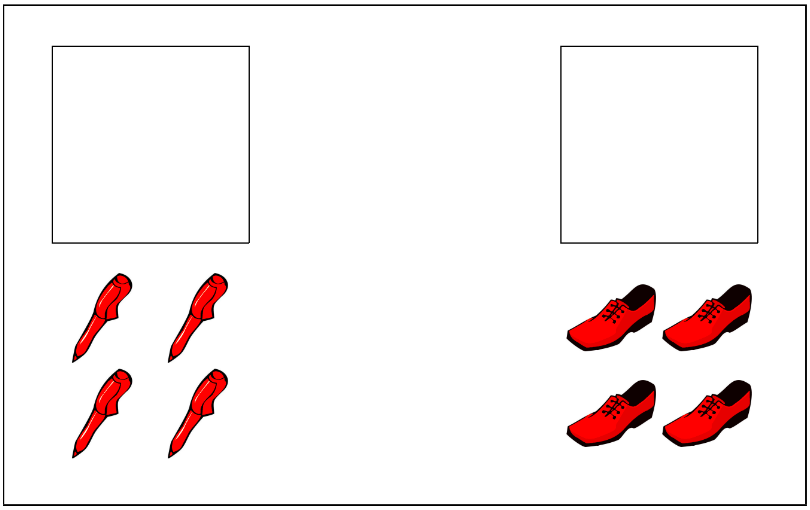


Figure 2.1: Initial object positions in MOVE phase with instructions to move objects to the left or right square.

2.2.2 Procedure

Visual stimuli were presented and participants carried out the interactive task (written in pygame (http://www.pygame.org)) on a 21-inch CRT monitor (Samsung SynchMaster 1100DF, Samsung, Korea) with a refresh rate of 120 Hz and a resolution of 800x1280 pixels. Two loudspeakers ('Z3 Style 2.1', Logitech, Switzerland) were positioned to the left and right of the monitor. The fixations were measured by an Eyelink II eye-tracking system (SR Research, Mississauga, Ontario, Canada), which was controlled by a Pentium 4 PC (Dell Inc., Round Rock, TX, USA).

The experiment took place in a darkened room. Before the experiment began, participants signed a consent form which informed them about the procedure, but not the aim of the experiment. The distance between participants and monitor screen was approximately 80cm. The eye-tracker was set up and calibrated using the 9-dot grid procedure. Only calibration values with a mean error less than $0.5\frac{1}{4}$ during the validation procedure were accepted.

All instructions, both before and within the experiment, were in German. Participants were instructed to move the objects presented to them on the screen according to the auditorily presented instructions. They were told that in certain cases they would not be able to follow the instruction until the end. In these cases they should just wait for the next auditory instruction. Further, they were told that after moving the objects, they would be instructed to click on one of two squares on the screen. Before the experiment, participants were asked whether they had understood the instructions and had the opportunity of asking the experimenter questions about the procedure.

There were 56 trials, where each trial consisted of two phases: the MOVE phase and the CLICK phase. The MOVE phase began with the presentation of two sets of four objects each in two corners of the screen, and two empty squares in the remaining two corners (see Figure 2.1).

Participants then received the auditory instruction to move one set of objects to one of the two squares, e.g. *Schieb die Schuhe ins linke der beiden Vierecke* ('Move the shoes into the left of both squares'). Participants used a mouse to move the objects to their designated target squares via drag and drop. Only one set of objects was moveable after a given instruction. That is, if participants misheard the object name and attempted to move one of the non-moveable objects, the object could not be "picked up" by the mouse. After the first set of objects was moved, participants received the auditory instruction to move the other set of objects to the remaining square. The MOVE phase ended by the mouse cursor jumping to the center of the screen.

Crucially, on certain trials only two of the four initial objects could be moved, that is, the remaining two objects could not be "picked up" by the mouse. When this occurred with the first set of objects, participants heard the instruction to move the other set of objects after only two objects had been moved. When it occurred with the second set of objects, the cursor jumped to the center of the screen after only two objects were moved.

The CLICK phase began with participants fixating the cursor in the center of the screen for two seconds. They then heard the auditory target instruction asking them to click on one of the squares, e.g. *Klick nun auf das Viereck mit einigen der roten Schuhe* ('Click on the square with some of the red shoes'), while viewing the scene they had created as e.g. in Figure 2.2. Once the par-

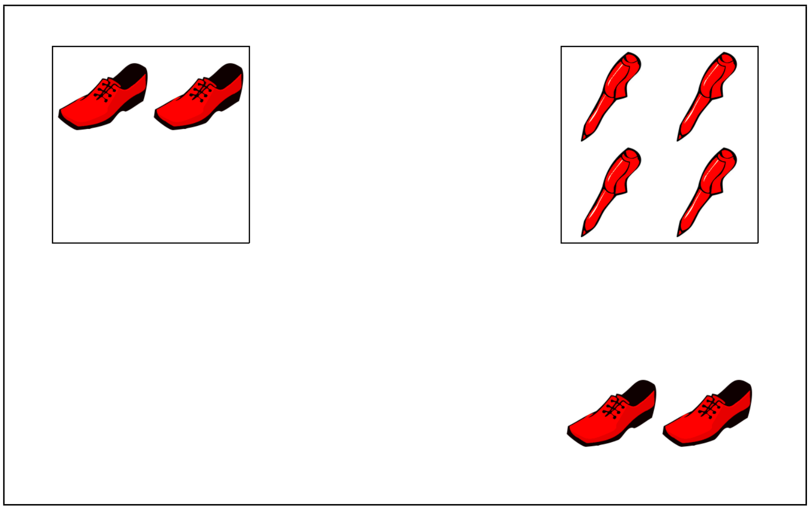


Figure 2.2: Object positions in CLICK phase for *early some* and *early all* condition.

ticipant clicked on one of the squares the display cleared and the next trial began.

After half (28) of the trials, participants could take a short break and remove the eye-tracker if they wished to do so. Of the 30 participants, only 4 took a short break of under two minutes. Of these, only one removed the eye-tracker. For all participants, the eye-tracker was recalibrated according to the above procedure before the beginning of the second half of the experiment.

At the end of the experiment, participants were informed about the goal of the experiment.

2.2.3 Stimuli

2.2.3.1 Visual Stimuli

The moveable object images were 150x150 pixel .png files. The images varied both in color and in displayed object type. For each of the six colors (blue, yellow, gray, green, red, purple) there was a set of eight objects, picked from a total set of ten object types. To ensure that objects' names were unambiguous, a naming study was carried out. Eight participants (who did not participate in the final study) were asked to name the objects as it seemed most natural to

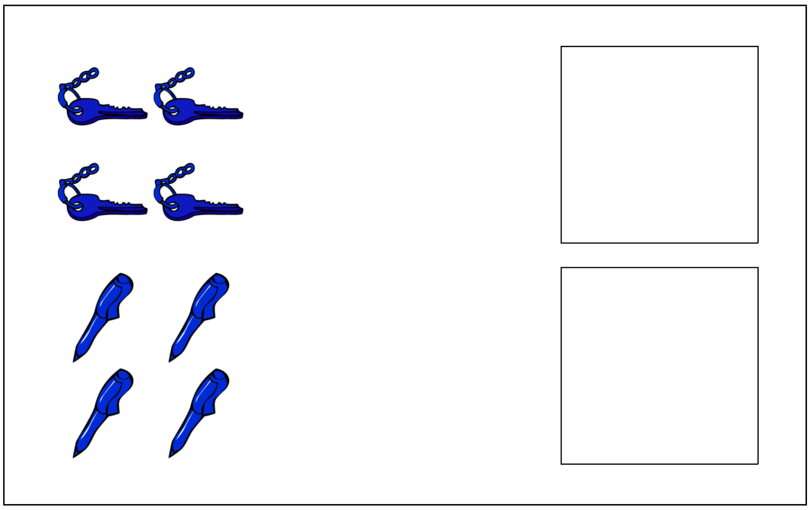


Figure 2.3: Initial object positions in MOVE phase with instructions to move objects to the upper or lower square.

them, with one word each. All objects received unambiguous labels, with the exception of the flower, which was named by one participant as *Tulpe* (‘tulip’) and the pen, which was named by one participant as *Kugelschreiber* (‘ballpoint pen’). See Appendix A for the full list of visual stimuli. All objects were inanimate objects employed in daily life.

2.2.3.2 Auditory Stimuli: MOVE Phase

Auditory instructions for each object type and potential target square position were recorded. That is, there were 40 sentences of the form *Schieb die* OBJECTS *ins* POSITION *der beiden Vierecke* (‘Move the OBJECTS into the POSITION of both squares’), where OBJECTS was an object name and POSITION was one of *linke, rechte, obere, untere* (‘left, right, upper, lower’). The position words were selected as follows: if the target position could be unambiguously referred to by ‘left’ and ‘right’, *linke* or *rechte* was chosen. If this was not possible, as in Figure 2.3, *obere* or *untere* was employed. Participants heard two of these sentences on each trial, one for each set of objects to be moved. The object type was never the same for both sets of objects on any given trial.

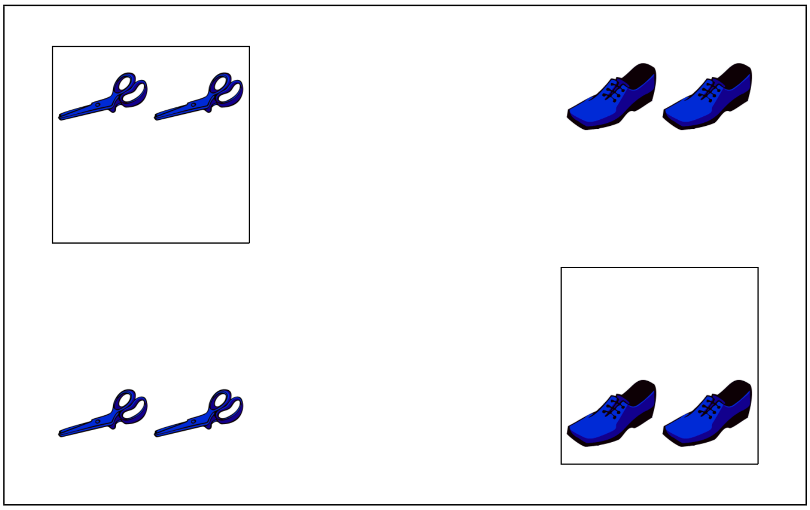


Figure 2.4: Object positions in CLICK phase for EN condition.

2.2.3.3 Auditory Stimuli: CLICK Phase

Auditory clicking instructions were recorded for each object type, color, and quantifier. Since there were two quantifiers - *einigen der, allen* ('some of the', 'all') - there was a total of 96 sentences of the form *Klick nun auf das Viereck mit* QUANTIFIER COLOR OBJECTS. COLOR was always a color adjective: *blauen, gelben, grauen, grünen, roten*, or *lila* ('blue', 'yellow', 'gray', 'green', 'red', 'purple').

Note that the relevant comparison in this study is between *allen* and the partitive construction *einigen der*. It would have been interesting additionally to include the non-partitive construction in the comparison *einigen*: the prediction, if the Default model is correct, being that there should be no difference in fixation patterns after hearing the partitive vs. the non-partitive construction, since in both cases the implicature trigger *einige* should give rise to the implicature, regardless of additional following information. However, this comparison was not included in the present study in order to keep the experiment at a manageable length for the participants. In addition, care was taken to ensure that the quantifiers did not receive contrastive stress.

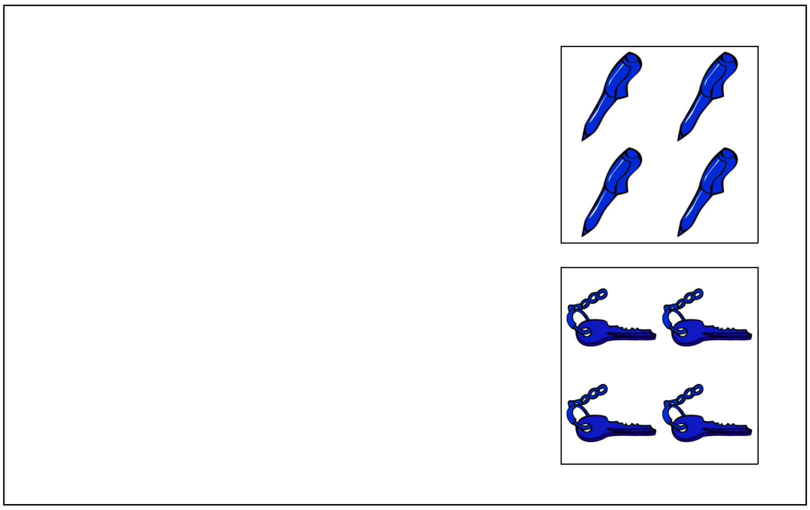


Figure 2.5: Object positions in CLICK phase for the *late all* condition.

2.2.3.4 Conditions

Of the 56 trials, 24 were target trials and 32 were filler trials. The independent variables were context type – disambiguating vs. non-disambiguating – and quantifier – *allen* vs. *einigen der*. Both variables were fully crossed, yielding six trials per condition. In condition 1 (*einigen der* in disambiguating context, henceforth *early some*, since the implicature, if it arises, helps to disambiguate the target early, that is before the onset of the noun), participants moved one set of objects fully and one set of objects partially, resulting in a scene as in figure 2.2. They then heard the target sentence with the quantifier *einigen der*. In condition 2 (*allen* in disambiguating context, henceforth *early all*), participants instead heard the target sentence with the quantifier *allen*. In both disambiguating conditions, whether the first or second set of objects was only partially moved was counterbalanced . In condition 3 (*einigen der* in non-disambiguating context, henceforth *late some*), participants moved both sets of objects only partially, thus leading to a situation where the quantifier would not provide disambiguating information even if interpreted with implicature. Similarly, in condition 4 (*allen* in non-disambiguating context, henceforth *late all*), participants moved both sets of objects fully. See figures 2.4 and 2.5 for examples of scenes in the *late some* and *late all* conditions.

Condition	Object color	Quantifier	Target
early some	same	*einigen der*	partial
early all	same	*allen*	full
late some	same	*einigen der*	partial
late all	same	*allen*	full
F1	same	*zwei*	partial
F2	same	*vier*	full
F3	same	*keinen*	partial
F4	same	*keinen*	full
F5	different	*zwei*	partial
F6	different	*vier*	full
F7	different	*keinen*	partial
F8	different	*keinen*	full

Table 2.2: List of target (*early some* - *late all*) and filler (F1 - F8) conditions.

In all cases, the clicking instruction was directly referential, i.e. there were no instructions that could not be followed because there was no square that did not qualify as a referent.

2.2.3.5 Fillers

Stimuli on the filler trials followed the same general pattern as on the target trials (i.e. initial MOVE phase followed by a CLICK phase). However, there were two differences to target trials: first, on some filler trials object color differed between sets. Second, different quantifiers appeared in the clicking instructions, specifically the quantifier *keinen* ('none') and the number terms *zwei* and *vier* ('two', 'four'). On all filler trials, one set of objects was moved fully while the other was moved only partially. Of the 32 trials, one half were trials with objects of the same color, while on the other half object color differed between sets. For each half, clicking instructions contained the number terms *zwei* and *vier* on 4 trials each. Clicking instructions for the remaining 8 trials contained the quantifier *keinen*, where on four trials the target was the square containing the fully moved set, while on the other four trials the target was the square containing the partially moved set. See table 2.2.3.4 for the full set of conditions.

2.3 Results

2.3.1 Data coding

Trials where participants selected the wrong square (0.9%) and trials with response times three standard deviations slower or faster than the mean (0.28%) were excluded.

In addition, certain pre-processing steps were conducted on the data obtained. Pre-processing was necessary because quantifier and noun on- and offset were not aligned in the auditory stimuli. Quantifier onsets ranged from 1756 to 2352 ms (with a mean of 2033 ms) and had durations of between 289 and 669 ms (with a mean of 480 ms). Because there was a color adjective intervening between the quantifier and the noun, there was greater variability in noun onsets, which ranged from 2601 to 3470 ms (with a mean of 2972 ms). In order to achieve comparability between items, the following formulas were used to adjust on- and offsets of the ith fixation in a way that quantifier onset and offset for the jth item were aligned to the overall quantifier onset and offset.

For fixations beginning or ending within the quantifier:

$$f'(i) = q_{\text{on}} + (f(i) - r_{\text{on}}(j)) \cdot \frac{q_{\text{off}} - q_{\text{on}}}{r_{\text{off}}(j) - r_{\text{on}}(j)} \tag{2.1}$$

The shifted fixation onset or offset is $f'(i)$. q_{on} and q_{off} are the overall quantifier onset and offset times, which were set to 2000 and 2400 ms (i.e. close to the mean quantifier onset and offset times). The actual onset or offset of fixation i is $f(i)$. Finally, $r_{\text{on}}(j)$ and $r_{\text{off}}(j)$ are the real quantifier onset and offset times of item j that fixation i occurred in. By this computation, if a fixation started at the jth item's actual quantifier onset, it was adjusted to start at the overall quantifier onset time (i.e. at 2000 ms).

If a fixation began or ended before quantifier onset, the following formula was employed to either further compress (if quantifier onset of that item was later than q_{on}) or stretch out fixations (if quantifier onset was earlier than q_{on}):

$$f'(i) = f(i) \cdot \frac{q_{\text{on}}}{r_{\text{on}}} \tag{2.2}$$

Similarly, if a fixation began or ended between quantifier offset and noun onset:

$$f'(i) = q_{\text{off}} + (f(i) - r_{\text{off}}(j)) \cdot \frac{n_{\text{on}} - q_{\text{off}}}{rn_{\text{on}}(j) - r_{\text{off}}(j)} \tag{2.3}$$

Here, in which direction a fixation is shifted depends on the real quantifier offset $r_{\text{off}}(j)$ and noun onset $rn_{\text{on}}(j)$, as well as on the overall quantifier offset q_{off} and noun onset n_{on} (which was set to 3000 ms).

Finally, if a fixation onset was between noun onset and noun offset (i.e. end of stimulus), equation 2.3 was modified in the following way:

$$f'(i) = n_{\text{on}} + (f(i) - rn_{\text{on}}(j)) \cdot \frac{n_{\text{off}} - n_{\text{on}}}{rn_{\text{off}}(j) - rn_{\text{on}}(j)} \tag{2.4}$$

Here, n_{off} and $rn_{\text{off}}(j)$ are overall noun offset (at 3500 ms) and real noun offset of item j. Thus, all data was interpreted with quantifier onset at 2000 ms, quantifier offset at 2400 ms, noun onset at 3000 ms, and noun offset at 3500 ms.

2.3.2 Results

A mixed logit model[2] was employed to test the effect of quantifier and context on scalar implicature computation. The analyzed categorical outcome (dependent variable) is the presence of a fixation to the target over its absence. The fixed effect predictors (independent variables) are quantifier (*some* vs. *all*), context (disambiguating (early) vs. non-disambiguating (late)), time (ranging

[2]See Barr (2008) for advantages of using mixed logit models over conventional analyses for analyzing data from visual world eye-tracking experiments.

Predictor	**Coef.** β	SE(β)	z	p
Intercept	-1.323	0.1520	-8.702	< 0.0001
Early	0.106	0.046	2.310	< 0.05
Some	-0.525	0.046	-11.421	< 0.0001
Timeslot	0.006	0.001	9.662	< 0.0001
Early×Some	0.168	0.092	1.827	< 0.07
Early×Time	0.00004	0.00005	1.163	> 0.2
Some×Time	0.002	0.00005	4.424	< 0.0001
Early×Some×Time	-0.0001	0.00007	-1.870	< 0.07

Table 2.3: Result summary: Coefficient estimates β, standard errors $SE(\beta)$, associated Wald's z-score ($= \beta/SE(\beta)$), and significance level p for all predictors in the analysis.

from 0 to 4400 ms in steps of 200 ms), and one parameter for the intercept. Additionally, the analysis includes a random subject intercept and a random slope in time, which can be thought of as the individual adjustment to each subject's rate of converging on the target, and a random item intercept, which can be thought of as the individual adjustment to each item's rate of converging on the target.

Table 2.3 summarizes the parameter estimates β for all fixed effects in the model, as well as the estimate of their standard errors SE(β), the associated Wald's z-score, and the significance level.

Colinearity was < 0.1 for all factors, with the exception of the following: r(Early×Some,Early) = 0.120, r(Early×Time,Early) = -0.280, r(Early×Some×Time,Early) = -0.101, r(Some×Time,Some) = -.287, r(Early×Some×Time,Early×Some) = -0.285.

There are main effects of context, quantifier, and time. When in a disambiguating context, participants were more likely to converge on the target than in a non-disambiguating context, as expected. Further, participants who received a target instruction with *some* were less likely to converge on the target than participants who heard *all*. Finally, the further along in the target sentence, the more likely participants were to converge on the target.

In addition, there is a marginal interaction of context and quantifier. This suggests that hearing *some* in the early condition makes participants more likely to converge on the target than only hearing *some* or only being in a disambiguating context. There is a second, strongly significant interaction between quantifier and time, suggesting that participants in a *some* condition are more likely to converge on the target earlier in time than participants hearing *all*. This may reflect baseline differences in the early conditions between fixations to the square with all moved items vs. the square with only a subset of the items. Generally, people prefer to look at more complex objects.

Finally, there was a marginal three-way interaction between context, quantifier, and time. Although only marginal, this trend suggests that participants were slower at converging on the target when they were in the *early some* condition.

The time course of proportions of fixations for all four conditions is shown in Figures 2.6 to 2.9. Looks to target, competitor, and the two initial object position regions are averaged over subjects and items and proportions are computed over the combined looks to target, competitor, and the two remaining regions.

Comparing the two *early* with the two *late* conditions, we see that looks to the target increase earlier in the former than in the latter conditions, as reported above. We can also detect the marginal three-way interaction: looks to the target converge more slowly in the *early some* than in the *early all* condition. We see the marginal interaction of context and quantifier in that looks in the *early some* condition are more likely to converge on the target than looks in the two *some* conditions overall.

Finally, we can see a baseline preference for the more complex region of interest, i.e. the region with more items. For the *early some* condition, this is the competitor, while for the *early all* condition it is the target. In the late conditions we see no such baseline differences. This makes sense since the regions of interest in the *late some* condition all contain two objects, while in the *late all* condition, the target and competitor regions both contain all objects. Thus, one initial goal was not achieved, namely to reduce baseline preferences for more complex regions by engaging participants in an interactive task. Note

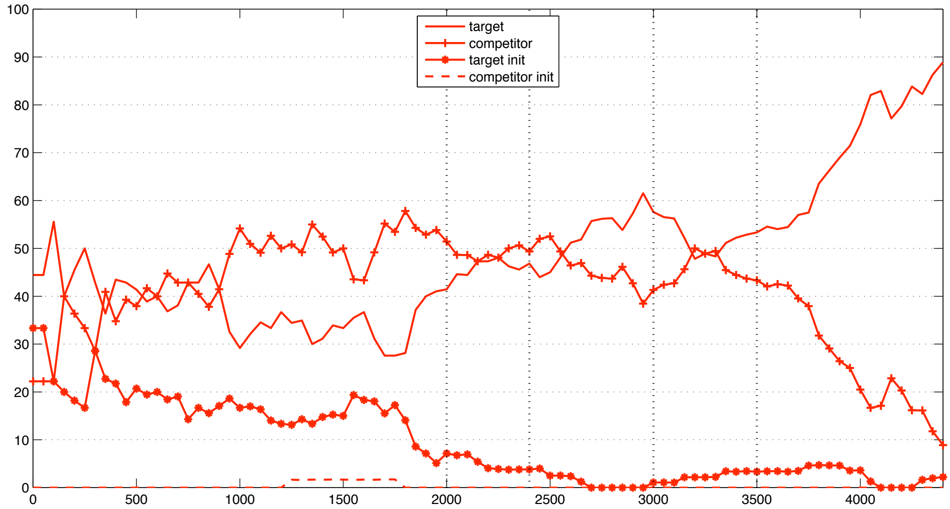


Figure 2.6: Proportions of fixations to the target, competitor, and the two initial object position regions in the *early some* condition, calculated over intervals of 50 ms. The first two and last two vertical lines indicate the quantifier and noun interval.

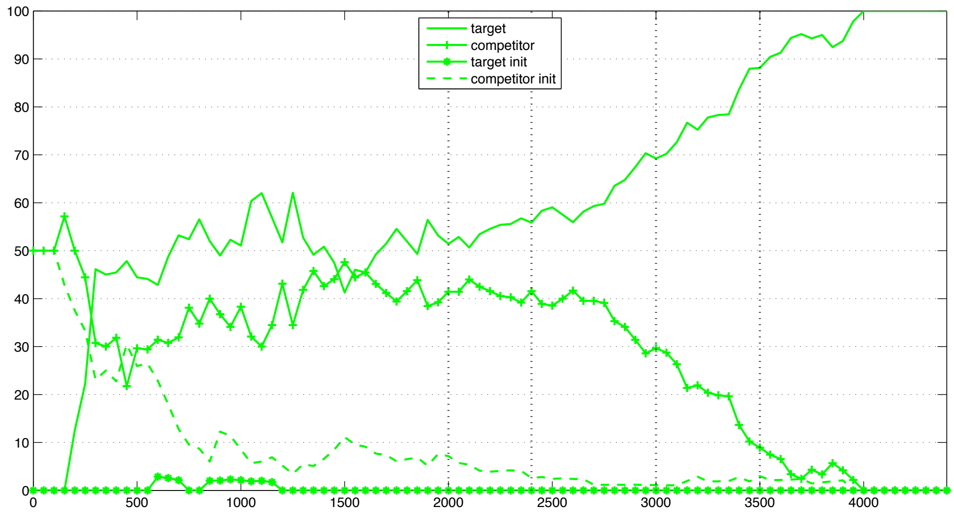


Figure 2.7: Proportions of fixations to the target, competitor, and the two initial object position regions in the *early all* condition, calculated over intervals of 50 ms. The first two and last two vertical lines indicate the quantifier and noun interval.

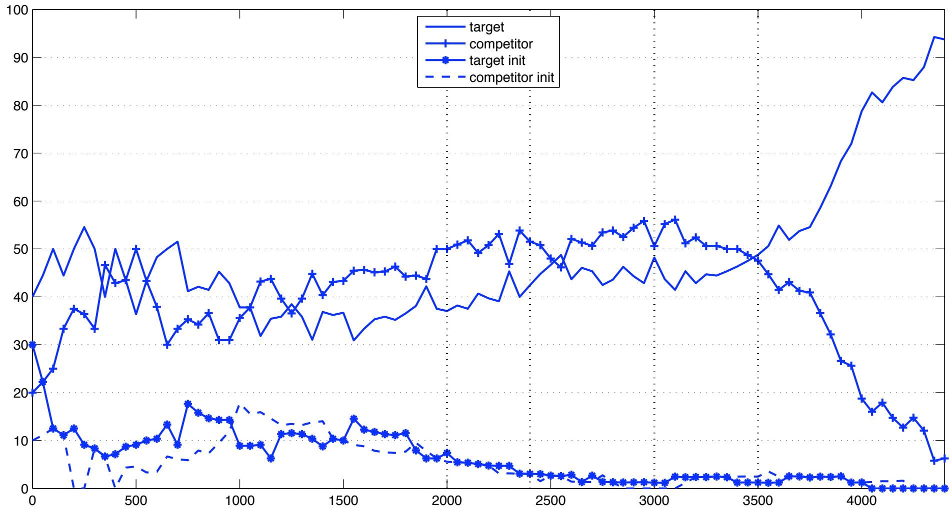


Figure 2.8: Proportions of fixations to the target, competitor, and the two initial object position regions in the *late some* condition, calculated over intervals of 50 ms. The first two and last two vertical lines indicate the quantifier and noun interval.

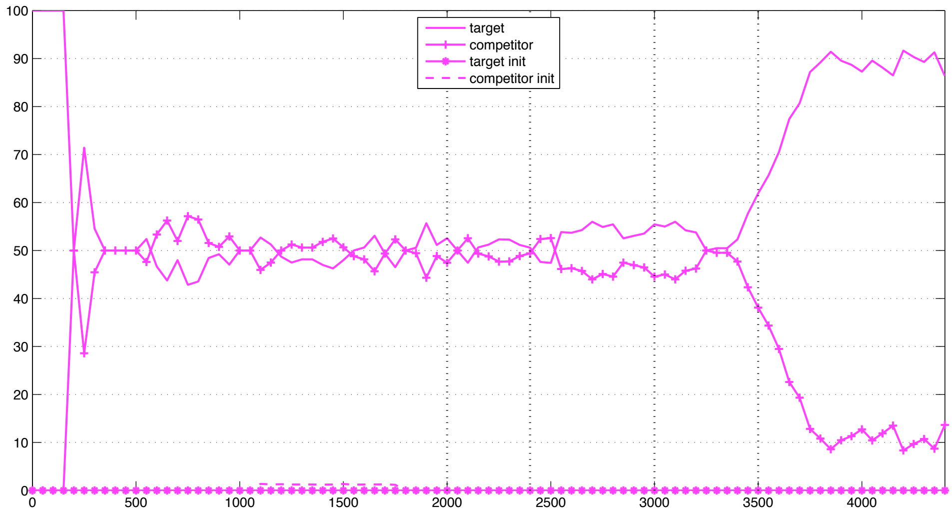


Figure 2.9: Proportions of fixations to the target, competitor, and the two initial object position regions in the *late all* condition, calculated over intervals of 50 ms. The first two and last two vertical lines indicate the quantifier and noun interval.

that there are some remaining looks to the initial object position regions in the *late some* condition, which reflects the fact that there were remaining objects in the initial regions in this condition.

2.4 Discussion

Our results indicate a delay of scalar implicature processing over semantic processing, as evidenced by the delay in increase of fixations on the target in the *early some* condition as compared to the *early all* condition. However, compared to the *late some* and *late all* conditions, where fixations on the target do not increase until 200 ms after the disambiguating noun onset, fixations to the target in the *early some* condition start to increase before noun onset. This shows that, although pragmatic *einige* takes longer to interpret than the literal meaning of *alle*, it does serve a disambiguating function, as evidenced in fixations to the target before noun onset. That is, the implicature arises and is used as disambiguating information. These results further support the Context-Driven model.

I will now dwell somewhat on the problems of this study, the first being one of analysis: since a contingent analysis was not performed, we have no information as to the exact timepoint of disambiguation in all four conditions. Since the three-way interaction was only marginal, it would have been helpful to include this analysis to determine whether the points of disambiguation for the *early some* and *early all* conditions are in fact different.

A further, design-related problem is that participants always moved either two or four objects. That is, the number of moved objects always fell within the subitizing range. In consequence, since subitizing occurs very quickly, a number term may have been more appropriate for labeling the target squares. Participants may have been expecting a number term instead of the more vague *some*, thus leading to a delay in convergence to the target.[3]

A further potential explanation for the delay (other than the standard expla-

[3] Note that this delay is not expected for *all*, since it is a fully referential term.

Table 2.4: Occurrences of *some/all* vs. *einige/alle* and ratio of *all/alle* to *some/einige* in English and German corpora.

	Corpus	Frequency of *some/einige*	Frequency of *all/alle*	Corpus size	Ratio
English	BNC	167114	276755	112181020	1.66
	BROWN	1617	2995	1170811	1.85
	EUROPARL-EN	31662	85115	28447737	2.69
	WSJ	2071	1548	1241041	0.75
German	WSJ	2071	1548	1241041	0.75
	DEWAC	816883	4101638	1627169557	5.02
	GUTENBERG-DE	64638	437722	137336680	6.77
	EUROPARL-DE	28196	80656	2713500	2.86
	SZ	123870	657650	320128929	5.31

nation that scalar implicatures take more time to process than literal meaning) lies in the translation of English *some* to German *einige*. The concern with this translation is that English *some* is a more frequently used word than German *einige*. This is due mostly to the fact that German *some* encodes a mass/count distinction. *Einige* is used for count, *etwas* for mass nouns. Compare examples (63) and (64).

(63) Count determiner

a. Er hat *einige* Bücher von Neal Stephenson gelesen.
he has some books by Neal Stephenson read
'He read some books by Neal Stephenson.'

(64) Mass determiner

a. Er hat *etwas* Wasser aus der Flasche getrunken.
he has some water out the bottle drunk
'He drank some water from the bottle.'

Using the mass determiner with *Bücher* or the count determiner with *Wasser* yields ungrammatical sentences. In addition, German *einige* is less colloquial than English *some*. Often, English *some* is more readily translated to German *manche* or *ein paar*. Table 2.4 shows frequency of occurrence of English *all* and *some* in the British National Corpus (112,181,020 tokens), the Brown corpus (1,170,811 tokens), the English Europarl corpus (28,447,737 tokens), and the

Wall Street Journal corpus (1,241,041 tokens) in comparison to frequency of occurrence of German *alle* and *einige* in the DEWAC corpus (1,627,169,557 tokens), the German Gutenberg corpus (137,336,680 tokens), the German Europarl corpus (27,135,007 tokens), and the Süddeutsche Zeitung corpus (320,128,929 tokens). In addition, the ratio of frequency of *all* to frequency of *some* was calculated for all four English corpora. Conversely for *alle* and *einige* in the German corpora. The mean ratio for the English corpora is 1.74, i.e. *all* occurs almost twice as frequently as *some*. For the German corpora, the mean was at 4.99. These means are significantly different ($F(1,6) = 13.07, p < .05$). This shows that *einige* is indeed much more rarely used in comparison to *alle* than *some* in comparison to *all*.

Since English *some* is found over twice as often in corpora as German *einige*, this difference may give rise to longer lexical access times for *einige* in comparison to *alle*, thus potentially explaining the difference in findings as compared to the Grodner et al. study, which found no delay in scalar implicature processing.

Chapter 3

General Discussion

3.1 Conclusion

Our results - that the pragmatic interpretation of German *einige* is computed more slowly than the interpretation of *alle* - suggests that scalar implicatures require more processing effort than semantic meaning[1], providing further support for the Context-Driven model. In the past, such results have been interpreted as support for Relevance Theory and against Neo-Griceanism (though explicitly it is only Levinson's account that gets mentioned) (Noveck and Posada (2003), Bott and Noveck (2004)). I want to make the point here that this is not a valid inference (c.f. also Blutner (2007)).

To see why support for the Context-Driven model should not be framed as support for Relevance Theory and against Neo-Griceanism, we must have a look at the assumptions underlying the claim that it is. The main assumption is that being a (Neo-)Gricean implies defending a Default view of scalar implicatures. Noveck (2004) claims that "Grice suggests that [...] his so-called 'generalized conversational implicatures', linked in particular to words such as *some, and* or *or* [...] are derived by default" (Noveck (2004): 302).

However, adopting a (Neo-)Gricean view does not imply accepting a Default view on the processing of scalar implicatures. Or simply: Griceanism does not

[1]See however section 2.3.2 for additional factors potentially influencing the delay.

imply Defaultism. Right off it is important to acknowledge that Grice himself did not make any processing claims. As Bach (2005) puts it:

> Grice did not intend his account of how implicatures are recognized as a psychological theory or even as a cognitive model. He intended it as a rational reconstruction. [...] He was not foolishly engaged in psychological speculation about the nature of or even the temporal sequence of the cognitive processes that implements that logic.

Thus, great care is required when translating pragmatic theories that make no psychological claims to models of processing. This care is not taken by researchers who use support for the Context-Driven model to refute the (Neo-)Gricean enterprise as a whole. I will show that, even if one insisted on deriving a processing model from the Gricean theory, this model is better conceived of as Context-Driven, rather than Default (though compatible with both). The question to investigate is thus: why should Grice be thought of as a Defaultist?

According to Grice, what the hearer relies on in calculating a conversational implicature is, among other things, "(2) the Cooperative Principle and its maxims; (3) the context, linguistic or otherwise, of the utterance; (4) other items of background knowledge" (Grice (1975): 49). This means a lot of work for the hearer in deriving scalar implicatures: the flouting of Quantity-1 must be noticed, which requires computing an expression's stronger alternatives against the background of expected informativeness. Further, the flouting will have to be recognized as arising from a clash with one of the Quality maxims, thus leading to the inference - depending on contextual assumptions - that the speaker employed the weaker expression because he either did not believe or did not have sufficient evidence for, the stronger one. This calculation process by no means implies defaultness in the way scalar implicatures are worked out.

Grice's passages on the distinction between generalized and particularized conversational implicatures might shed more light on the matter. He says about generalized conversational implicatures that "the use of a certain form of words in an utterance would normally (in the absence of special circumstances) carry such-and-such an implicature or type of implicature" (Grice (1975): 57). For the case of *some*, this means that *some* usually, ceteris paribus, gives rise to the *but not all* implicature. This notion of scalar implicatures as the preferred way of interpreting implicature triggers may be the main reason for Noveck's view of Grice's theory as a Defaultist one. But this is a misleading intuition: if special circumstances can prevent the implicature from arising, as implied by the above quote, then it is not a context-independent phenomenon. Whatever these ceteris paribus conditions may be, checking whether they are actually given implies expending processing effort. Again, this is not evidence for, but rather against, a Default view of Grice's theory.

There is, however, one passage which is critical and concerns the cancelability of generalized conversational implicatures. As we have seen in section 1.1, a scalar implicature may be canceled in one of two ways:

> It may be explicitly canceled, by the addition of a clause that states or implies that the speaker has opted out, or it may be contextually canceled, if the form of utterance that usually carries it [the implicature] is used in a context that makes it clear that the speaker is opting out (Grice (1975))

Explicit cancellation is not a problem, since Defaultists and Contextualists agree that the implicature must have arisen in such contexts, only to be canceled by the second clause. The problem is rather implicit, or contextual, cancellation. Recall example (9), repeated here as (65). The context is lower-bound, so the implicature is absent.

(65) John: Is there any evidence against them?
Peter: Some of their identity documents are forgeries.
Not all of their identity documents are forgeries.

The question is whether in this case the implicature is generated and then canceled (Default model) or does not arise in the first place (Context-Driven model). Levinson (2000) argues that the fact that Grice speaks of *implicit can-*

cellation in this context implies that he assumes the implicature to arise (for otherwise how could it be canceled?), thus putting Grice in the Defaultist camp.

However, Grice's notion of implicit cancellation, since not a claim about psychological processes, should be viewed as a purely functional one. That is, when he speaks of implicatures as normally carrying an implicature, the psychological process associated with the interpretation of such implicatures may be either a) a default process or b) an effortful process. Similarly, implicit implicature cancellation (functional description) is compatible with both processing claims, namely that a) the implicature is generated and subsequently canceled or b) the implicature does not arise in the first place. Thus, the only sense in which Grice may be portrayed as a Defaultist is in that he claims that in the majority of contexts in which we encounter an implicature trigger such as *some*, we will interpret *some* as meaning *some, but not all*[2]. This notion of Default applies at the output (offline) level, not at the processing (online) level.

I have thus shown that the Gricean account is compatible both with the Default and the Context-Driven model. Further, in section 1.2.1.2 I presented Matsumoto's Neo-Gricean account, which is similarly best construed as a Context-Driven account (though Matsumoto does not make any processing claims, either): since the Conversational Condition must be checked at every implicature trigger before an implicature is generated, the conversational context clearly enters every scalar implicature computation. It is fair to assume (though not necessary), that testing whether the Conversational Condition is satisfied

[2]Levinson (2000) actually makes the argument that it is the fact that implicatures arise in more contexts than not that it makes sense for implicatures to be computed (in the processing sense) by default, assuming the brain is optimal in assigning processing effort. However, Noveck and Sperber (2007) point out that if the number of implicature trigger contexts without implicature exceeds the number of contexts with implicature, then it would be inefficient of brains to generate default implicatures, since this would lead to excessive processing effort devoted to cancellation (assuming that processing costs for generation and cancellation are the same). It is currently an unresolved issue whether there are more contexts in which implicature triggers actually do give rise to an implicature than contexts in which they do not. This problem could be addressed via corpus studies: systematically annotating sentences with implicature triggers for whether an implicature arises or not, though a very big enterprise, may provide insight.

(i.e. that the speaker's utterance is not due to any other information-selecting maxim other than Quantity-1) requires processing effort and is not best conceived of as a default process.

From the considerations that a) Grice's original account is compatible with both the Default and the Context-Driven model and b) there are Neo-Gricean accounts that are better construed as Context-Driven rather than Default accounts, I conclude that the question of Default is not suited to resolving the theoretical debate between Relevance Theory and Neo-Griceanism.[3] It is simply an unfortunate circumstance that the Neo-Gricean proponent calling for psycholinguistic validation of his theory is also a Defaultist, while the predictions of Relevance Theorists, who advocate a psychological theory of language processing, line up with the Context-Driven model's. Empirical evidence for the Context-Driven model should thus not be taken as evidence for Relevance Theory alone, but rather for the union of Relevance Theory and all (Neo-)Gricean approaches compatible with the Context-Driven model.

3.2 Further directions

Should we conclude from this that the question of Default is not a question worth being pursued? The answer to this is a clear *no*. Independently of the theoretical framework, the question of what the differences in processing between semantic and pragmatic information are is worth investigating. There are a number of additional questions related to the question of Default that arise. One factor that should be highly relevant to determining whether or not a scalar implicature arises is the importance of the upper bound. That is, for *some*, is it important whether or not the stronger statement with *all* would actually have been more relevant to the hearer? It may be that in cases where the stronger statement is in fact more relevant than the weaker statement, we may find processing evidence that the implicature arises more quickly than in cases where whether or not the stronger statement is made does not make a

[3] See Degen (2007) for a fuller discussion.

difference to the hearer for the task at hand. Degen and Tanenhaus (in preparation) are addressing this question in reaction-time and eye-tracking settings that manipulate importance of the upper bound.

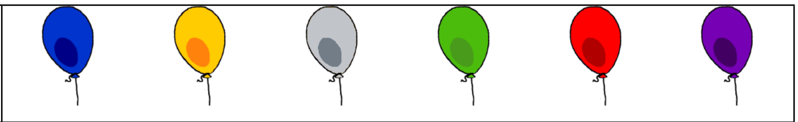
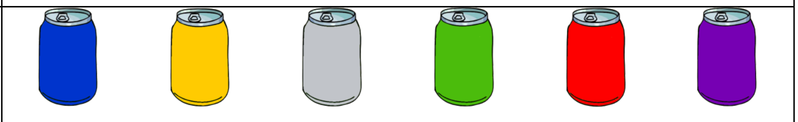
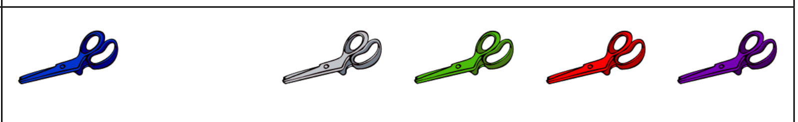
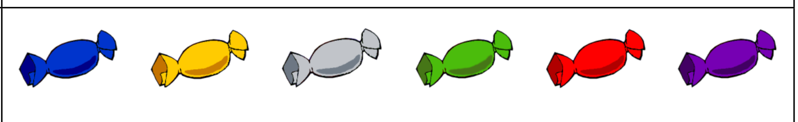
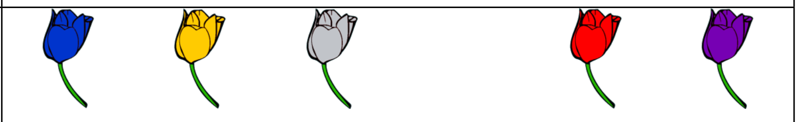
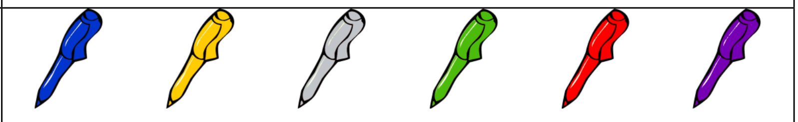
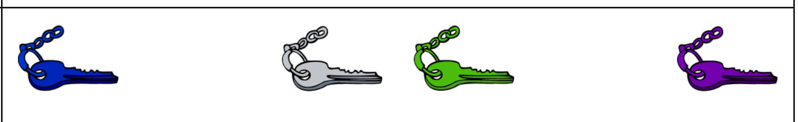
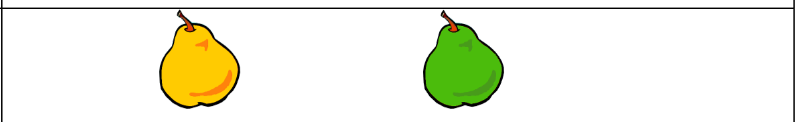
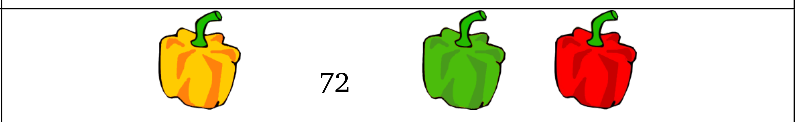
Further, most of the work on the question of Default has focused on the ⟨all, some⟩ (but c.f. Breheny et al. (2006)). This work should be extended to other scales, such as ⟨and, or⟩, but also scales that do not have a logical counterpart, such as ⟨know, believe⟩ or ⟨succeed, try⟩. In fact, Katsos (in press), in ongoing work, has started to address this question by comparing context-independent scales (e.g. ⟨all, some⟩) to context-dependent scales (e.g. ⟨past Santorini, past Syros⟩, where Santorini and Syros are islands on a path to Crete and having passed Santorini implies having passed Syros). He finds no difference in rejections of under-informative utterances containing the weaker element of a context-dependent vs. context-independent scale. However, there is a difference in the nature of these rejections, which tend to be more determined for context-independent (outright "no") than context-dependent ("right, but" + revision) scales.

This may be due to higher frequency of exposure to "standard" scales such as ⟨all, some⟩, ⟨and, or⟩, etc., than to ad-hoc scales such as in the ⟨past Santorini, past Syros⟩ case. Assume that in the large part of cases of encountering a conventional scale, use of the weaker scale member implicates the negation of the stronger scale member. Then, encountering a new instance of a statement employing a weaker scale member that is to be interpreted as not implicating the negation of the statement with the stronger member should lead to high surprisal, and outright rejection of this weaker statement is thus to be expected. For a given ad-hoc scale, on the other hand, which we encounter much less frequently, the implicature will not become the conditioned response, so to say. Thus, since scalar implicatures are not like entailments, it is not surprising that they are unlikely to be rejected outright. It would be a fruitful endeavor to model frequency effects of scale occurrences (derived from corpora) as predictors of rejection (and reaction times of rejection) of underinformative statements.

In sum, then, although experimental work on the question of Default will not resolve the theoretical debate between Neo-Griceans and Relevance Theorists, it brings up a whole new set of exciting questions pertaining to the contextual factors influencing the processing of scalar implicatures.

Appendix A

Stimuli

Ballons	
Dosen	
Schuhe	
Scheren	
Bonbons	
Blumen	
Stifte	
Schlüssel	
Birnen	
Paprikas	

Bibliography

Allopenna, P. D., J. S. Magnuson, and M. K. Tanenhaus (1998). Tracking the time course of spoken word recognition using eye movements: Evidence for continuous mapping models. *Journal of Memory and Language 38*, 419 – 439.

Bach, K. (1999). The myth of conventional implicature. *Linguistics and Philosophy 22*, 327–366.

Bach, K. (2005). The top ten misconceptions about implicature. In B. Birner and G. Ward (Eds.), *Festschrift for Larry Horn*. Amsterdam: John Benjamins.

Barr, D. J. (2008). Analyzing 'visual world' eyetracking data using multilevel logistic regression. *Journal of Memory and Language 59*, 457 – 474.

Barwise, J. and R. Cooper (1981). Generalized quantifiers and natural language. *Linguistics and Philosophy 4*, 159–219.

Bezuidenhout, A. L. and R. K. Morris (2004). Implicature, relevance and default pragmatic inference. In I. A. Noveck and D. Sperber (Eds.), *Pragmatics*, Chapter 12, pp. 257–282. Palgrave Macmillan Ltd.

Blutner, R. (2007). Some experimental aspects of optimality-theoretic pragmatics. Talk given at Experimental Pragmatics 2007, Berlin.

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

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

Carston, R. (1998). Informativeness, relevance and scalar implicature. In R. Carston and S. Uchida (Eds.), *Relevance Theory: Applications and Implications*, pp. 179–236. Amsterdam: John Benjamins.

Chierchia, G., M. T. Guasti, A. Gualmini, L. Meroni, S. Crain, and F. Foppolo (2004). Semantic and pragmatic competence in children's and adults' comprehension of or. In I. A. Noveck and D. Sperber (Eds.), *Experimental Pragmatics*, Chapter 13, pp. 283–300. Palgrave Macmillan Ltd.

Degen, J. (2007). Processing scalar implicatures: What role does the question of Default play for the debate between (Neo-)Griceanism and Relevance Theory? *Publications of the Institute of Cognitive Science, University of Osnabrück 3-2007*.

Degen, J. and M. K. Tanenhaus. Using a novel experimental paradigm to investigate the processing of scalar implicatures. In preparation.

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

Geurts, B. and N. Pouscoulous (2008). No scalar inferences under embedding. In F. in: Paul Egré and G. Magri (Eds.), *Presuppositions and implicatures*. MIT Working Papers in Linguistics.

Grice, P. (1969). Utterer's meaning and intentions. *Philosophical Review 78*(2), 147–177. Reprinted in Grice 1989, pp. 86-116.

Grice, P. (1975). Logic and conversation. *Syntax and Semantics 3*, 41–58. Reprinted in Grice 1989, pp. 22-40.

Grice, P. (1978). Further notes on logic and conversation. *Syntax and Semantics 9*, 183–197. Reprinted in Grice 1989, pp. 41-57.

Grodner, D., N. M. Klein, K. M. Carbary, and M. K. Tanenhaus (submitted). Rapid interpretation of pragmatic "some". submitted.

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), 667–696.

Horn, L. (1984). Towards a new taxonomy of pragmatic inference: Q-based and r-based implicature. In D. Schiffrin (Ed.), *Meaning, Form, and Use in Context*, Linguistic Applications, GURT84, pp. 11–42. Washington: Georgetown University Press.

Huang, Y. and J. Snedeker (submitted). From meaning to inference: Evidence for the distinction between lexical semantics and scalar implicature in online processing and development. submitted.

Katsos, N. (2007). The semantics/pragmatics interface from an experimental perspective: the case of scalar implicatures. *Synthese*, DOI: 10.1007/s11229–007–9187–4.

Katsos, N. (in press). Evaluating under-informative utterances with context-dependent and context-independent scales: experimental and theoretical implications. In *Experimental Semantics and Pragmatics*. Palgrave Studies in Pragmatics, Language and Cognition.

Katsos, N., R. Breheny, and J. Williams (2005). The interaction of structural and contextual constraints during the on-line generation of scalar inferences. In B. Bara, L. Barsalou, and M. Bucciarelli (Eds.), *Proceedings of the 27th Annual Conference of the Cognitive Science Society*, pp. 1108–1113. Mahwah, NJ: Erlbaum.

Keenan, E. (1976). The universality of conversational postulates. *Language and Society 5*, 67–80.

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

Matsumoto, Y. (1995). The conversational condition on Horn scales. *Linguistics and Philosophy 18*, 21–60.

Neys, W. D. and W. Schaeken (2007). When people are more logical under cognitive load - dual task impact on scalar implicature. *Experimental*

Psychology 54(2), 128–133.

Noveck, I. and A. Posada (2003). Characterizing the time course of an implicature: an evoked potentials study. *Brain and Language 85*, 203–210.

Noveck, I. A. (2001). When children are more logical than adults. *Cognition 78*, 165–188.

Noveck, I. A. (2004). Pragmatic inferences related to logical terms. In I. A. Noveck and D. Sperber (Eds.), *Experimental Pragmatics*, Chapter 14, pp. 301–321. Palgrave Macmillan Ltd.

Noveck, I. A. and A. Reboul (2008, November). Experimental pragmatics: a gricean turn in the study of language. *Trends in Cognitive Sciences 12*(11), 425–431.

Noveck, I. A. and D. Sperber (2007). The why and how of experimental pragmatics: The case of 'scalar inferences'. In N. Burton-Roberts (Ed.), *Advances in Pragmatics*. Basingstoke: Palgrave.

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

Partee, B. (1984). Nominal and temporal anaphora. *Linguistics and Philosophy 7*(3), 243–286.

Potts, C. (2005). *The logic of conventional implicatures*. Oxford University Press.

Pouscoulous, N., I. A. Noveck, G. Politzer, and A. Bastide (2007). A developmental investigation of processing costs in implicature production. *Language Acquisition 14*(4), 347–375.

Russell, B. (2006). Against grammatical computation of scalar implicatures. *Journal of Semantics 23*, 361–382.

Salverda, A., D. Dahan, and J. McQueen (2003). The role of prosodic boundaries in the resolution of lexical embedding in speech comprehension. *Cognition 90*, 51–89.

Storto, G. and M. K. Tanenhaus (2004). Are scalar implicatures computed online? In *Proceedings of WECOL*.

Tanenhaus, M. K., M. J. Spivey-Knowlton, K. M. Eberhard, and J. Sedivy (1995). Integration of visual and linguistic information in spoken language comprehension. *Science* (5217), 1632–1634.

Wilson, D. and D. Sperber (1995). Relevance theory. In G. Ward and L. Horn (Eds.), *Handbook of Pragmatics*, pp. 607–632. Oxford: Blackwell.