SEMINAR FÜR SPRACHWISSENSCHAFT
EBERHARD KARLS UNIVERSITÄT TÜBINGEN

IMPLICATURES
IN UNCOOPERATIVE CONTEXTS

Magisterarbeit in General Linguistics

Anna Pryslopska
Tübingen, 07-Jun-13

Supervisors
Prof. Wolfgang Sternefeld
Dr. rer. nat. Ingo Hertrich

For my parents.

I told you I would write it.

Table of Contents

Table of Figures

Table of Tables

Zusammenfassung

Skalare Implikaturen sind ein angesagtes Forschungsthema in der Linguistik. Drei Denkrichtungen dominieren das Feld mit unterschiedlichen Theorien darüber, wie sie berechnet werden. Neo-Griceaner (Guerts 2009a, Horn 2004, 2006, Geurts und Pouscoulous 2009, Dieussaerta et al. 2011) behaupten, dass semantische und pragmatische Anteile an der Bedeutung in zwei hierarchisch angeordneten Stufen verarbeitet werden. Daher werden Quantoren mit Implikaturen im Vergleich zu rein semantischen Quantoren, wie z.B. *alle*, langsamer verarbeitet. Die so genannten Defaultisten (Storto und Tannenhaus 2004, Grodner et al. 2010, Foppolo 2006, Levinson 2000, Chierchia 2004) hingegen behaupten, dass Hörer pragmatische Bedeutungsanteile automatisch, und daher nicht später als in rein semantischen Fällen herleiten. Relevanztheoretiker sind sich in ihrer Analyse nicht einig; einige (Breheny et al. 2006, 2012, 2013) behaupten, dass Implikaturen zeitnah und mühelos berechnet werden können, während andere (Carston 2009, 2010, 2012, Bott & Noveck 2004, Reboul 2004, De Neys und Schaeken 2007) dies als einen zeit- und ressourcenaufwändigen Prozess ansehen.

Die vorliegende Studie beleuchtet diese Diskussion mithilfe eines *Visual World Paradigm*-Experiments welches unter Anderem die Studie aus Grodner et al. repliziert, jedoch zusätzlich untersucht, ob Implikaturen überhaupt berechnet werden, wenn die Grice'sche Kommunikationsvoraussetzung der Kooperativität nicht gegeben ist. Werden Implikaturen unmittelbar (Defaultisten,) oder verzögert (Relevanztheoretiker) berechnet, oder fehlen sie ganz (Neo-Griceaner?)

In normalen, kooperativen Kontexten wurden Implikaturen automatisch und genauso schnell wie für den rein semantischen Quantor *alle* berechnet. Dies unterschied sich jedoch in nicht-kooperativen Fällen; dort wurden sowohl *einige* als auch *alle* vergleichsweise spät berechnet. Die Ergebnisse bestätigen Grodner et al. und sind für die kooperativen Fälle sowohl mit den Ansichten der Defaultisten, als auch mit jenen der Relevanztheoretiker zu vereinbaren. Jedoch machte keine der oben genannten Theorien die korrekten Vorhersagen für nicht-kooperative Fälle. Eine mögliche Erklärung wäre, dass im unkooperativen Szenario die Berechnung im Allgemeinen aufgeschoben wird, bis der Hörer eine Entscheidung darüber treffen kann, ob der Sprecher vertrauenswürdig ist.

Abstract

Scalar implicatures are a hot topic among linguists. Three dominant schools of thought make different claims about how they are computed. Neo-Griceans (Geurts 2009a, Horn 2004, 2006, Geurts and Pouscoulous 2009, Dieussaerta et al. 2011) argue that semantic and pragmatic meanings are processed in two hierarchically ordered stages, and hence implicatures are delayed in comparison to purely semantic quantifiers like *all*. On the other hand, Defaultists (Storto and Tannenhaus 2004, Grodner et al. 2010, Foppolo 2006, Levinson 2000, Chierchia 2004) claim that listeners arrive at the pragmatic meaning automatically and not later than e.g. for *all*. Relevance theorists are divided in their assessment of pragmatic processing; some (Breheny et al. 2006, 2012, 2013) claim implicatures can be computed immediately and effortlessly, while others (Carston 2009, 2010, 2012, Bott & Noveck 2004, Reboul 2004, De Neys and Schaeken 2007) consider it a time and resource consuming process.

The present study addresses this debate in a visual world experiment which is partially a replication of Grodner et al., but also investigates whether implicatures are computed when a prerequisite for communication in Gricean terms – cooperativity – is lacking. Will implicatures be computed immediately (Defaultists), will they be delayed (Relevance Theorists) or even absent (Neo-Griceans)?

In standard, cooperative contexts implicature computation was automatic and as fast as for *all*. This was different in the uncooperative cases, where the both *some* and *all* were computed later in comparison. The results replicate Grodner et al. They are consistent with both the Defaultist and the Relevance Theoretic view in the cooperative cases. In the uncooperative ones none of the aforementioned theories made the correct predictions. A possible explanation is that the uncooperative scenario the computation is suspended in general until a decision has been made whether to trust the speaker or not.

Acknowledgements

First and foremost I would like to thank Oliver Bott and Fabian Schlotterbeck for their patience, knowledge and encouragement. I am truly indebted and thankful to Michael Franke for the conceptual inception and many fruitful discussions. A great many thanks to my supervisors, Wolfgang Sternefeld and Ingo Hertrich, for their continuous support, from initial advice and contacts to the very last sentence, whilst allowing me to work in my own way. Besides, I am most grateful to Robin Hörnig for his help in analyzing the vast amounts of eye-tracking data. A special thanks to Aleksandar Dimitrov without whom this thesis would never have been finished. Finally, the Sonderforschungsbereich 833 at the University of Tübingen has my deepest gratitude, among other things for funding this research. For any inadequacies or errors that may remain in this work, the responsibility is entirely my own.

Anti-Plagiarism Policy Statement

Hiermit versichere ich diese Arbeit selbständig und nur mit den in der Arbeit angegebenen Hilfsmitteln verfasst zu haben.

Mir ist bekannt, dass ich alle schriftlichen Arbeiten, die ich im Verlauf me ines Studiums als Studien- oder Prüfungsleistung einreiche, selbständig verfassen muss. Zitate sowie der Gebrauch von fremden Quellen und Hilfsmitteln müssen nach den Regeln wissenschaftlicher Dokumentation von mir eindeutig gekennzeichnet werden. Ich darf fremde Texte oder Textpassagen (auch aus dem Internet) nicht als meine eigenen ausgeben.

Ein Verstoß gegen diese Grundregeln wissenschaftlichen Arbeitens gilt als Täuschungs- bzw. Betrugsversuch und zieht entsprechende Konsequenzen nach sich. In jedem Fall wird die Leistung mit „nicht ausreichend“ (5,0) bewertet. In schwerwiegenden Fällen kann der Prüfungsausschuss den Kandidaten/die Kandidatin von der Erbringung weiterer Prüfungsleistungen ausschließen; vgl. hierzu die Prüfungsordnungen für die Bachelor-, Master-, Lehramts- bzw. Magisterstudiengänge.

Datum und Unterschrift

1. Introduction

Cooperation between conversational partners is a prerequisite to the emergence of scalar implicatures. Implicatures are a pragmatic phenomenon arising only if the speaker and hearer work together to reach a common goal through communication. However, this is often not the case in day to day language use. It is easy to imagine scenarios where parties in a dialogue have separate and conflicting goals. Whether it is tricking the customer, telling only half the truth to the parents, misleading the opponent in a game of cards, or arguing with the spouse – everyday communication provides many examples for uncooperative uses of language.

Do implicatures arise in uncooperative situations like the ones mentioned above? Are hearers able to recognize uncooperative expressions and how do they respond to such expressions? How does the computation of implicatures in uncooperative contexts differ from cooperative situations? Can theories of scalar implicatures account for the possible differences? The eye-tracking experiment presented in this paper was designed to answer these questions by investigating implicatures in both cooperative and uncooperative dialog situations.

Firstly, I introduce the phenomenon of implicatures. I present how several influential theories (Grice, Defaultism, Relevance Theory, Contextualism) explain the nature and processing of scalar implicatures. A survey of existing psycholinguistic experiments investigating implicatures follows. Subsequently, I report the eye-tracking study aimed at investigating the processing of scalar implicatures in an uncooperative situation, specifically, a game. The closing chapters contain a discussion of the results and an outlook on future research.

2. Scalar Implicatures

Grice was first to make a distinction between *what is said* and *what is meant* (Grice, 1989, reprint 1991). The words forming a sentence (*what is said)* and their literal meanings are the subject matter of syntax and semantics. Pragmatics on the other hand is concerned with what the utterance *means*.[1] Implicatures are pragmatic phenomena exemplified by the utterance below:

(1) Some cats love catnip.

The speaker of (1) is literally *saying* that to her best knowledge at least one cat and possibly all of them love catnip. The implied *meaning* of the utterance in (1) that the speaker is conveying is, however, that some but not all cats love catnip. She could have used any quantificational expression but chose "some" instead of, for example "many", "few" or "every cat I know". Perhaps the speaker knows a cat that takes no pleasure in catnip. Alternatively, she might be leaving the possibility open, because it is highly unlikely that a single person knows about the preferences of all felines. The nonliteral meaning of (1) in a specific context is *implicated* and is carried by an *implicature*, here specifically a *generalized conversational implicature* (GCI). GCIs are a more universal type of conversational implicature and depend on the utterance itself and not on particular features of the context. In contrast, *particularized conversational implicatures* (PCI) are implicatures that arise in context-specific situations and are not generalizable to other scenarios. A PCI is for instance what speaker B produces in (2):

(2) A: Did you enjoy your vacation in Norway?
B: It was cold and rainy.

Speaker B might have either enjoyed his vacation or have had a horrible time, depending on whether she likes the kind of weather she described. If speaker A is familiar with B's preferences, he can infer the answer to his question. This inference will be based as much on the utterance itself as on the particularities of the situation.

An implicature can be cancelled without causing a contradiction and is consequently distinguishable from a logical entailment. If the speaker further on in the conversation says something that is not compatible with the implicature it is then *explicitly* cancelled. In the example (1) above, if the speaker was to follow up her utterance with "In fact, all of them do love catnip." the meaning of the whole utterance would immediately and without contradiction change to the literal meaning of *some* (*all*). An implicature is *implicitly* cancelled if both speaker and listener share common ground which contains information that is incompatible with the implicature. Consider the following example:

(3) Joanna was looking for a cat last night.

[1] Here I use the word *mean* to describe the interpretation of an utterance after semantic and pragmatic processes have taken place.

The speaker of (3) chose the indefinite article to implicate that the cat Joanna was looking for was not necessarily hers. Maybe she was in an animal shelter looking to adopt one. If this were not the case, the speaker would have said “her cat”. However, it is not difficult to imagine a situation in which both speaker and listener know that Joanna is a crazy cat lady. She has a multitude of cats in her care and the speaker is unsure exactly which one of them went missing that evening, sending the owner into understandable search frenzy. In the former situation, the context allows for the implicature, but the implicature is implicitly cancelled in the latter.

In addition to that, implicatures are also implicitly cancelled in lower-bound contexts (Katsos et al., 2005). The same sentence in two different contexts can trigger an implicature in one (upper-bound) but not in the other (lower-bound; see Chierchia 2004 and the following chapter for discussion).

Grice makes a distinction between *conversational* and *conventional* implicatures. Conversational implicatures are triggered by “certain general features of discourse” (Grice 1991, pp. 26.) Grice formulates these features in the Cooperative Principe (CP) and its adherent maxims. Both the CP and the maxims guide the listener in the computation of conversational implicatures.

> 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.
>
> MAXIMS OF QUANTITY:
>
> a) Make your contribution as informative as is required (for the current purpose of the exchange).
> b) Do not make your contribution more informative than is required.
>
> MAXIMS OF QUALITY: *Try to make your contribution one that is* true.
>
> a) Do not say what you believe to be false.
> b) Do not say that for which you lack evidence.
>
> MAXIM OF RELATION: Be relevant.

MAXIMS OF MANNER: Be perspicuous.

a) Avoid obscure expressions.
b) Avoid ambiguity.
c) Be brief (avoid unnecessary prolixity).
d) Be orderly. (ibid, pp. 26-29.)

Conventional implicatures on the other hand arise when specific words or expressions are used and their meaning has been conventionally agreed upon. Consider the following example:

(4) a. He is rich but kind-hearted.
b. He is rich and kind-hearted.

The difference in meaning between (4a) and (4b) lies in the conjunction: “but” indicates a contrast between being “rich” and being “kind-hearted”, whereas “and” merely connects the two adjectives. They are tied closely to specific words and not the discourse situation (according to Grice, at least). The contrast is not part of the literal utterance but rather is caused by the conventional meaning of *but*, i.e. a conjunction introducing a contrast. The implicature in (4) is not a *conversational implicature* because it cannot be calculated based of the maxims described above. Therefore, it must be a *conventional implicature*.

Scalar implicatures are a kind of GCI[2] and are triggered by words such as “some” and “most”, “sometimes” and “rarely”, “might” and “should”. These and similar expressions can be ordered hierarchically on so-called Horn scales (Horn, 1984) from the logically weakest to the strongest. Examples of Horn scales are for instance the following, ordered left to right from weakest to strongest:

a) <a(n), some, many, most, all>
b) <rarely, occasionally, frequently, always>
c) <may, must>

The logically stronger expressions on a Horn scale asymmetrically entail the weaker ones and the use of a weaker alternative implies the negation of stronger terms. If the speaker chooses a semantically weaker expression she implies that she believes the stronger alternative to be false or that she lacks the information to make a stronger statement.

[2] The terms *scalar implicature* and *generalized conversational implicature* will be used throughout this paper synonymously and they will refer to implicatures triggered by words like “some”.

Scalar implicatures are a very interesting and thoroughly investigated phenomenon, yet still hold many unanswered questions. In what contexts do scalar implicatures arise? Are they truly universal? How are they computed? Is the process of implicature computation a fast or a time consuming one? In the following chapter I will discuss several theories of implicatures that strive to answer these questions and other similar questions.

3. Theories of Scalar Implicatures

The following subsections will give an overview of two schools of thought – the (Neo)Gricean approach and Relevance Theory. I will look at how implicatures are explained in these frameworks. Affiliated with these schools of thought are two theories of implicature computation: Defaultism and Contextualism. They will be outlined adjacent to the Gricean and Relevance Theoretical approaches, respectively.

3.1. (Neo-)Gricean Approach

According to Grice, in order for the hearer to infer the implicature from a given utterance and contextual information both the speaker and hearer are required to behave rationally in the conversational game (cf. Grice (1991), pp. 28-29.) He defined rational behavior in a dialogue in terms of the participants' observance of the CP and its underlying categories, i.e. maxims. Grice first introduces the notion of cooperation in very general terms:

> Our talk exchanges do not normally consist of a succession of disconnected remarks, and would not be rational if they did. They are characteristically, to some degree at least, cooperative efforts; and each participant recognizes in them, to some extent, a common purpose or set of purposes, or at least a mutually accepted direction. (Ibid, pp. 26)

Two independent observations are contained in the above paragraph. The first one relates to the maxim of relation: the agents in a dialog behave rationally and make their contributions relative to what has been said so far. The second assumption is that the speaker and hearer have common goals, which is the gist of the CP.

The maxims are attendant to the CP. Speakers may choose to *violate*, *opt out* or *flout* a maxim. Violation of a maxim results in a lie or an utterance unrelated to the conversation. In the following example the maxim of quality was violated: "There are no cats in Turkey." This is a false statement, as any Turkish Angora breeder will testify. A fantastic source of examples for violation of the maxim of relation is “Monty Python’s Flying Circus”, for instance “My hovercraft is full of eels.” The speaker might opt out by saying “I don’t know.” or “I can’t say for sure.” Finally, flouting a maxim triggers an implicature, like (1) “Some cats like catnip.” (maxims of quantity and quality.)

Implicature computation occurs only in a cooperative setting and provided that conversational maxims are being obeyed. The speaker can produce a conversational implicature by flouting a maxim either to convey an additional meaning or because a conflict between maxims prevents her from fulfilling both.

Scalar items like “some” will usually give rise to the same kind of implicature, although there are instances where no implicatures arise. Griceans claim that generalized conversational implicatures are closely tied to particular words or expressions and that these particular expressions give rise to implicatures across contexts. Semantically, “some” expresses “at least one and possibly all” but pragmatically “at least one but not all”.

The pragmatic processing model based on Grice’s views is called the Literal-First Serial Model (LFS). The LFS requires the recovery of the utterance’s minimal proposition, which includes disambiguation, reference assignment etc., before commencing the pragmatic processing. All pragmatic processes remain frozen until the minimal proposition has been recovered, even in contexts in which the enriched interpretation is clearly favored over the minimal one (cf. Bezuidenhout and Cutting (2002), pp. 443.) In order to recover the *meaning* of an utterance, a hearer must make an additional step. It is possible that in an uncooperative setting the hearer will see no benefit in recovering the utterance *meaning* and may choose not to take this step at all.

Consider example (1), repeated here as (5):

(5) a. Some cats love catnip.
b. $\exists x\ [cat(x) \wedge love\ catnip(x)]$ SOME AND POSSIBLY ALL *cats love catnip*.
c. $\neg\forall x\ [cat(x) \rightarrow love\ catnip(x)]$ NOT ALL *cats love catnip*.

Upon hearing (5)a, a listener will begin analyzing the semantics of the sentence in order to calculate the minimal proposition. She will conclude that the speaker is conveying the information: “There exists at least one cat that loves catnip and it is possible that in fact all cats love catnip.” This proposition is expressed in (5)b.

After having comprehended semantics of (5)a, i.e. what is *said*, the listener can move on to uncovering what is *meant* by it. She will assume that the speaker is not only obeying the CP, but also is capable and willing to contribute to the conversation. Since the speaker decided against using a stronger scale alternative “all” he must know that the stronger term is false, or is lacking confidence that the stronger term can be truthfully applied. In order to not violate the maxim of quality, the speaker is using the weaker scalar alternative and thereby communicating (5)c. The fully developed generalized conversational implicature “Some but not all cats love catnip.” is the combination of both (5)b and (5)c.

Cooperative transactions require the participants to have a common *immediate aim*, different from their respective *ultimate aims*, which may be independent or even conflicting (cf. Grice (1991), pp. 26-29). An example of such ad-hoc cooperation could be a game of Monopoly in which two players temporarily cast aside their differences and join forces to drive another player into bankruptcy; after reaching that goal, they resume former mutual hostility.

The present study is interested in those moments in the game where two or more agents turn against each other. The eye-tracking was designed to test (Neo-)Gricean theories in such contexts. Will implicatures be less likely to arise? Cooperative contexts have been previously investigated and given a plausible explanation but uncooperative ones are largely unknown (see chapter “Experimental Studies” below for more discussion).

The way uncooperative agents interact and communicate cannot be explained by the classical Gricean theory. The lack of cooperation should make communication altogether impossible. Neo-Griceans like Geurts (2009a, 2009b) extend the notion of cooperation to the entire game situation, arguing that the players are cooperative at least in the sense that they are playing a specific game and agree to obey its rules. If the opposing sides had no immediate common aim they would be incapable of playing together at all or would at least get very frustrated with each other, leading to out-of-game hostility.

Grice’s theory speaks little of implicature computation and cannot make predictions for the experiment in that respect. Neo-Griceans argue that the meaning of a sentence is computed in a linear fashion, as illustrated in (5) and the following paragraphs above (cf. Geurts 2009a and 2009b, Geurts and Pouscoulous 2009, Dieussaerta et al. 2011, Horn 2006). Therefore, they predict that the logical form of an utterance will be necessarily computed before any pragmatic processing can take place. Hence, in the example sentence “Some cats love catnip.” the scalar implicature *some but not all* would be more effortful and delayed in time in comparison to “some and possibly all.” In contrast to that, Defaultists – a school of thought associated Griceans – make the opposing predictions.

3.2. Defaultism

Defaultists base their model of implicature computation on the Gricean notion of being *closely tied* to particular words or expressions (cf. Grodner et al. 2010, Foppolo 2006, Freeney et al. 2004). This idea has led some researchers to believe that implicature computation is an automatic and effortless process. According to Defaultists implicatures are a pragmatic phenomenon and are derived cost-free and by default (hence the name) across contexts. The cancelling of an implicature is what requires processing effort.

Two Defaultist views on scalar implicatures will be presented here. The first is one is that of Stephen Levinson (2000) and the second of Gennaro Chierchia (2004). Although both attributed special characteristics of scalar items to their *default* character, Chierchia viewed them as a part of the syntax and Levinson attributed a special status more closely associated with semantics than syntax.

3.2.1. Levinson 2000

The central aim of Levinson's book is – as he puts it – to defend the autonomous and generalized nature of GCIs. Already at the very beginning he allies himself with the Defaultist perspective: "A generalized implicature is, in effect, a default inference, one that captures our intuitions about a preferred or normal interpretation." (Levinson, 2000, pp. 11) Levinson clarifies that his theory is a generative heuristic-driven systemic account and the "normal interpretations" do not rely on routinization:

> The theory of GCI is not of course a theory of conventional idioms, clichés, and formulae, but it is *a generative theory of idiomaticity*–that is, a set of principles guiding the choice of the right expression to suggest a specific interpretation and, as a corollary, a theory accounting for preferred interpretations. (ibid, pp. 24)

Levinson expands the Gricean set of properties of conversational implicatures (cancellability, nondetachability, calculability, nonconventionality, see chapter on scalar implicatures above) by two: reinforcability and universality. The former relates to the fact that, "it is often possible to add explicitly what is anyway implicated with less sense of redundancy than would be the case if one repeated the coded content" (ibid, pp. 15). The latter underlines the autonomous and rational nature of conversational implicatures which *ipso facto* must lead to universality.

Furthermore, he distinguishes three levels of meaning: (a) sentence-meaning; (b) speaker-meaning; (c) utterance(-type) meaning. The first one of these levels is governed by grammar, whereas the second is the pragmatic layer as explained in Gricean terms. The last level is an intermediate one; here the border between syntax and pragmatics is unclear. This level homes presuppositions, speech acts, implicatures, preference organization and so forth. Levinson's theory operates in this "communication" layer of sentence meaning.

The machinery available to the listener allowing her to recover the speaker's intentions and that allows for categorization of GCIs are the three heuristics listed below:

Q-HEURISTIC
What isn't said, isn't.

I-HEURISTIC
What is expressed simply is stereotypically exemplified.

M-HEURISTIC
What's said in an abnormal way isn't normal.
(cf. ibid, Chapter 1, pp. 11-12)

The first heuristic is akin to Grice's Maxim of Quantity (*be as informative as required*, cf. chapter on Scalar Implicatures, ibid pp. 3-4). When the speaker chooses an expression from a contrast set she thereby negates the salient alternatives. The listener may base her inference on the inapplicability of the alternatives, provided of course that the alternatives and the chosen expression differ in informativeness. In our example sentence "Some cats love catnip", we automatically arrive at the "some but not all" inference by the nature of the fact that "some" implies "not all" by default. This inference is driven by the Q-heuristic: since it *isn't said* that "All cats love catnip", it *isn't* the case that all cats love catnip.

The second heuristic is also related to the Maxim of Quantity (*don't be more informative than required*). The application of this heuristic allows the listener to narrow down semantically general expressions to the most stereotypical interpretation. For example in the utterance "Sophie's book on cats is very good." the book can be either one that Sophie wrote, lent to someone, or read, as mandated by the background.

The third heuristic resembles another one of Grice's Maxims, the Maxim of Manner (*be perspicuous*) and is closely related to the I-heuristic. In a nutshell, if the speaker uses an unmarked expression alternative the listener may use the I-heuristic and interpret the utterance in a direct, conventional way. Marked expressions, on the other hand, suggest that stereotypical interpretations are not applicable and should be avoided. What is *not* said creates a comparison to the marked expression, underlining the difference in meaning. Consider the sentence “Leon came in and the cat hissed.” where “Leon” would typically be referenced as “he” if it was indeed him who hissed; hence “the cat” must denote some other individual (cf. ibid, pp. 39).

Levinson recognizes that a single utterance can trigger different or even conflicting subtypes of GCIs. He resolves this problem by postulating that heuristics are thus ordered according to importance or strength: Q heuristic>M heuristic>I heuristic (cf. ibid, pp. 170).

Furthermore, there is what Levinson calls a parasitic relationship between the I and M heuristics: “what is said simply, briefly, in an unmarked way picks up the stereotypical interpretation; if in contrast a marked expression is used, it is suggested that the stereotypical interpretation should be avoided.” (ibid, pp. 170) Thus, the content of a marked expression is complementary to that of a stereotypical expression, had an unmarked form been used.

The processing model suggested by Levinson is one in which two types of both pragmatic and semantic processing occurs. He rejects the traditional view in which semantics processing is a prerequisite for pragmatic processing:

> “[E]ssentially, I argue that the theory of meaning has the components or levels we always thought it has–it is just that we have to reconstrue the kinds of relations that hold between them. (…)The novel suggestion here is that *semantic and pragmatic processes can interleave, in ways that are probably controlled by the constructional types in the semantic representation*.” (ibid, pp. 168) [3]

Let us consider how the example sentence (1), repeated below as (6)a, is computed. First, the syntactic structure of the utterance is analyzed (6)b; the first semantic representation may also be derived. The output of this grammatical analysis is the input to the first pragmatic process (6)c. It is at this stage that the default pragmatic processes, such

[3] Levinson argues that pragmatic processes are both influenced by and exert influence on semantic and even syntactic ones. Specifically, GCIs play a role in such crucial linguistic mechanisms like reference resolution, bridging, scope assignment generality narrowing etc. An example of pragmatic/syntactic interface and reference resolution is quoted from Levinson (2000), pp. 271:

a) The police$_1$ barred the demonstrators$_2$ because they$_2$ advocated violence.
b) The police$_1$ barred the demonstrators$_2$ because they$_1$ feared violence.

Depending on the background knowledge and conjectures about which of the two groups are more likely to be advocating or fearing violence, the reference of *they* can be resolved by crediting either the police or demonstrators with reference.

as GCIs, contribute to utterance meaning and influence the truth-conditional value of the utterance. The output thereof is now the input to model-theoretic semantics (6)d. At this stage contextual information becomes relevant and, for example, a GCI generated earlier may be cancelled. Later still, the output from the second semantic process becomes the input to the second pragmatic stage, where e.g. the process responsible for other inferences such as PCIs takes place. The output of this final process yields the utterance's meaning.

(6) a. Some cats love catnip.
b. $[[ø]_C\ [[[\text{Some}]_D\ [\text{cats}]_{NP}]_{DP}\ [[ø]_I\ [[\text{love}]_V\ [[ø]_D\ [\text{catnip}]_{NP}]_{DP}]_{VP}]_{I'}]_{IP}]_{CP}$
c. Some: some $\wedge \neg$all
d. ⟦ Some cats love catnip.⟧ $\wedge \neg$⟦ All cats love catnip.⟧

3.2.2. Chierchia 2004

Similarly to Levinson, Chierchia argues that scalar items carry a default interpretation. He writes:

> The claim is that there are situations in which (standard) implicatures are by default present and situations where they are by default absent and such situations are determined by structural factors. By default interpretation, I simply mean the one that most people would give in circumstances where the context is unbiased one way or the other. (Chierchia 2004, pp. 12)

His method of deriving the implicature is, however, substantially different from Levinson's. Chierchia sets his focus to the syntax-pragmatics interface and claims that grammatical processes as much as pragmatic ones underlie scalar implicatures computation. He writes:

> [This approach's] guiding idea is that implicatures are introduced locally as soon as possible in the same order in which their trigger (the scalar terms) are introduced in the syntactic tree. (...) Once introduced, implicatures are projected upwards and filtered out or adjusted, as the case may be, much like what happens with presuppositions. (ibid, pp. 9-10)

In other words, scalar implicatures are retrieved phrase by phrase via functional application in a way that resembles recursion for grammatical content. The computation of scalar implicatures occurs in tandem with the semantics of the sentence as it unravels; it thereby contributes to the truth conditions of the phrase they are embedded in.[4]

The overall structure of implicature computation can be summarized in the following steps:

a) Compute (recursively) the plain value of an expression α:
⟦ Some cats love catnip.⟧

b) Compute (recursively) the “scalar” or “strengthened” value of α by exploiting its alternatives and choosing one that is immediately stronger than the target:
⟦ Some cats love catnip.⟧^{S}=All cats love catnip.

c) Introduce direct implicatures locally if licensed:
⟦ Some cats love catnip.⟧ $\wedge$ $\neg$⟦ All cats love catnip.⟧

After the third step, implicatures may be removed encountering a downward entailing item and indirect implicatures may be introduced, if necessary.

This claim stands in contrast to the traditional globalist view. According to globalists, the utterance in its integrity must first be semantically processed before implicatures (as part of pragmatics) can be computed.

Consider the following example:

(7) a. Joanna believes that some cats love catnip.
b. It is not the case that Joanna believes that all cats love catnip.
c. Joanna believes that not all cats love catnip.

From the globalist perspective, the globally computed implicature of (7)a is (7)b, whereas the locally arising implicature is (7)c. It is not the aim of this paper to make further claims about localist and globalist dispute but the reader should be sensitive to the differences in computation and the end product of both Defaultist theories.

Chierchia claims that scalar implicatures influence truth value computation in upward entailing (monotone increasing) contexts but not in downward entailing ones. The former contexts are triggered by phrases such as *some*, *at least*, *most*, whereas the latter by, for example, *less than*, *few*, *no*. Chierchia notes that there are examples of implicatures arising in downward entailing contexts, like:

[4] A scalar implicature is constructed at a local level and may be cancelled or removed at a global level if it no longer fits the semantic environment.

(8) John has not read many books on cats.

The utterance in (8) implies that “John has read some books on cats.” Chierchia calls such cases *indirect* implicatures in contrast to “normal”, *direct* implicatures.

Chierchia’s model goes in tandem with the observation that scalar implicatures arise in some contexts but not in others. For example, in what Chierchia calls the “standard view”, embedded implicatures should not exist. Consider the following utterance:

(9) a) Sophie believes that some cats love catnip.
b) Sophie believes that not every cat loves catnip. *¬ all cats*
c) Sophie believes that every cat loves catnip.
d) It is not the case that Sophie believes that every cat loves catnip.
¬ belief

Given the utterance (9)a, the globalist rendition of the embedded implicature is (9)b. However, if implicatures are rendered locally, (9)c would be the relevant alternative to a) and the implicature would be its negation d). (9)d is much weaker than b) as it only expressed that it is compatible with Sophie’s beliefs that not every cat loves catnip. It does not exclude the possibility expressed in (9)b. (cf. Chierchia 2004, pp. 6) Chierchia’s account provides a good explanation to the problem that scalar implicatures contribute to the truth conditions (semantics) of a phrase, even though they belong to the field of pragmatics.

Overall, Defaultism makes the claim that implicature computation is an effortless, fast and automatic process. If this is true, hearers are expected to compute implicatures, irrespective of whether the context is a cooperative or an uncooperative one. They would subsequently need to decide whether they are content with the interpretation of the speaker’s utterance and, if so, retain the implicature. Alternatively they may consider (contextual or other) information that prompts them to cancel the implicature. There may be instances where it suffices for the hearer to know that “at least some” cats love catnip and whether all of them do is irrelevant (cf. Degen 2007).

One of the reasons why I decided in favor of an eye-tracking experiment, presented later in this paper, is that it provides a marvelous opportunity to measure very exactly the time it takes the participants to fixate the object congruent with the pragmatic meaning of an utterance. Therefore, this method should be able to verify whether implicature computation is indeed swift and undemanding, as the Defaultists would have it. In addition to that, the frequency with which particular objects were fixated allows drawing conclusions about the participants' derived meaning: logical, pragmatic or otherwise. However, the design of the experiment will not permit me to makes claims in support of either globalist or localist theory.

3.3. Relevance Theory

At the core of Relevance Theory is the claim that utterances trigger expectations of relevance. The resemblance to Grice's maxim of relevance is not incidental. Relevance Theory has its roots in the Gricean approach. However, Relevance Theorists refute several other central aspects of the Gricean account, most prominently the following:

> [T]he need for a Co-operative Principle and maxims, the focus on pragmatic processes which contribute to implicatures rather than to explicit, truth-conditional content, the role of deliberate maxim violation in utterance interpretation, and the treatment of figurative utterances as deviations from a maxim or convention of truthfulness. (Wilson and Sperber 2005, pp. 250)

Relevance Theorists claim that expectations of relevance triggered by an utterance are sufficiently predictable and precise to point the hearer to the speaker's intended meaning. The main focus falls on human cognition and its inherent search for relevance. The question arises, "what does it mean to be relevant?" An input is relevant if it may be plausibly connected to the hearer's background knowledge and provides some valuable information. In such a situation, the input is said to yield a *positive cognitive effect,* i.e., *a worthwhile difference to the individual's representation of the world* (ibid, pp. 251). At any given moment there is a multitude of stimuli which are relevant to an individual. The lower the processing costs and the greater the *positive cognitive effect of a* stimulus, the more relevant the stimulus. Wilson and Sperber (2005) define the Relevance Theory's conception of relevance in the following terms:

> RELEVANCE OF AN INPUT TO AN INDIVIDUAL
>
> a) 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.
> b) Other things being equal, the greater the processing effort expended, the lower the relevance of the input to the individual at that time.

They also provide a comprehension procedure which guides the hearer to achieve understanding of the utterance.

COMPREHENSION PROCEDURE

a) Follow a path of least effort in computing cognitive effects: Test interpretive hypotheses (disambiguation, reference resolutions, implicatures, etc.) in order of accessibility.
b) Stop when your expectations of relevance are satisfied.

SUB-TASKS IN THE OVERALL COMPREHENSION PROCESS

a) Constructing an appropriate hypothesis about explicit content (*explicatures*) via decoding, disambiguation, reference resolution, and other pragmatic enrichment processes.
b) Constructing an appropriate hypothesis about the intended contextual assumptions (*implicated premises*).
c) Constructing an appropriate hypothesis about the intended contextual implications (*implicated conclusions*).

The comprehension process is further subdivided into three steps necessary for the computation of implicatures: the retrieval of the (usually incomplete) logical form, the derivation of explicatures (i.e. *what is said*) and the computation of any implicatures (i.e. *what is meant*). These steps are not executed in a sequence but rather unfold in parallel. Furthermore, they are mutually interactive and adjusting in the on-line process of understanding. A *contextual implication* can be deduced only on the basis of both input and context. Implicatures may be *weak* or *strong* depending on whether they are essential in understanding the interpretation or merely relevant and helpful in arriving at an interpretation. Scalar implicatures are defined in this framework much in the same way other implicatures are. The speaker is expected to provide as much information as he considers relevant for the hearer by means of least processing effort in both production and interpretation.

To illustrate how this process unravels, let us consider (10):[5]

(10) Sophie: What should I buy for my kitten?
Joanna: Some cats love catnip.

[5] Cf. Wilson and Sperber (2005). For clarity, the comprehension procedure presented below is structured in a linear fashion, but in reality the process is not a hierarchical one.

Sophie needs to recover the explicature of Joanna's utterance. Firstly, she will recover its semantic structure:

(11) $\exists x\ [cat(x) \wedge love\ catnip(x)]$

In the next step she will disambiguate any potentially ambiguous words (example (12)) and make the link between "cat" and "kitten", which is a young cat.

(12) [cat_1] – feline animal, domesticated or wild
[cat_2] – playboy

(13) [to $love_1$] – adore, like very much
[to $love_2$] – have sexual relations

(14) below is the explicature of Joanna's utterance, i.e. the first enrichment of the logical form of her utterance that fulfills Sophie's expectations of relevance and combines with the implicated premises:

(14) $[\![$Some $cats_1$ $love_1$ catnip$]\!]$

Sophie expects Joanna's utterance in (10) to be optimally relevant to her by answering her question in an informative way. Therefore she accepts Joanna's presumption of relevance. The implicated premise of Joanna's answer is that a number of domesticated felines like catnip very much and (a toy with) catnip might make for a nice present for the kitten.

After having computed the explicature and implicit premises of Joanna's utterance, Sophie moves on to the implicated conclusions. The fully developed meaning of Joanna's answer is that not all cats love catnip but some do; Sophie's kitten might or might not belong to the catnip loving bunch, but Sophie should at least consider purchasing a scented toy.

When the hearer recognizes the informative intention, understanding and communication are achieved. However, if she does not trust the speaker (for example if Joanna was known for being cruel to animals) she is in no way obliged to believe the speaker, and hence fulfill the informative intention. However, it would be very unwise of Sophie to inquire about a cat with someone known for cruelty to animals.

The predictions for implicatures in general made by Relevance Theory are that they will be delayed in time in comparison to utterances *sans* implicatures because the logical form (explicature) of the utterance will always be computed before any pragmatic processes can take place (implicated conclusions.) This trend should be visible in both cooperative and uncooperative contexts, but the latter may influence the processing by making the implicature computation longer. These predictions resemble the Gricean take on implicature derivation but are quite opposite to predications made by Defaultists (see previous chapter).

From a Relevance Theoretical standpoint, it is also possible that uncooperative contexts do not license the computation altogether because it is associated with greater processing effort with possibly little positive cognitive effect. The expectations of relevance may be satisfied with the logical interpretation.

3.4. Contextualism

The relationship between Contextualism and Relevance Theory is similar to that of Griceans and Defaultists: they are frequently associated with one another but differ in several key aspects. Contextualists agree with Relevance Theorists that the comprehension process is divided into several levels and the computation of implicatures from explicatures depends on cognitive effects and processing effort (Carston 1998, 2010, Huang and Snedeker 2009, Reboul 2004.) The sentence comprehension process can be roughly divided into three stages: explicatures, implicated premises and implicated conclusions. The stages are not ordered hierarchically but are interdependent. The first two steps correspond to what Griceans consider *what is said* and the last one (implicated conclusions) to *what is meant*. A hearer may choose to cease computing an utterance if she is content with the logical interpretation (*what is said*) and never get as far as computing the implicated conclusions (*what is meant*) and avoid the effort altogether.

Akin to Relevance Theory, Contextualism identifies the preferred interpretation as the one which achieves the greatest positive cognitive effect and is most relevant in a particular context. With this, Contextualists challenge the nondetachability of implicatures by claiming that neither the logical nor the pragmatic interpretation of a scalar item can be a derived as a "default" one.

The main difference between Contextualists and Relevance Theorists lies in their take on the influence of the context. According to the former, context plays a significant role in the derivation of implicatures because it influences the salience of interpretations. As a result, generalized conversational implicatures are no longer context-independent. They will be computed by the hearer only in situations that provide enough contextual support for the pragmatic interpretation. Even in null-context situations (i.e. without any explicit information about the sentence's context) hearers try to contextualize an utterance, hence context is always a factor influencing the relevance of interpretation alternatives. Furthermore, contextual information comes into play at the stage of *what is said*: "What is said in the Contextualists' enriched sense can contain elements that are determined on the basis of contextual information, and that do not correspond to anything semantically encoded in the sentence uttered." (Bezuidenhout and Cutting 2002) Relevance Theorists, on the other hand, see the role of context mainly in relation to testing interpretative hypothesis (cf. previous chapter and disambiguation example in (12) and its influence on relevance during the entire comprehension process.)

The processing paradigm proposed by Contextualism is based in part on the Relevance Theoretical model and somewhat akin to Chierchia's views. The Local Pragmatic Processing Model (LPP; Garret and Harnish 2007, 2008) assumes that pragmatic processing has a local character. As soon as the first word of a sentence is encountered, the entire pool of pragmatic processes may be called on. Contrary to the Literal-First Serial Model, the LPP attributes a lesser role to the logical form of an utterance, because it considers the logical form insufficient. The lexical concepts underlying the logical form need to undergo pragmatic processing, such as enrichment, bridging etc. The result is a locally processed *ad hoc* concept which is subsequently incorporated in the overall structure of the utterance's proposition, which is *what is said* in an enriched sense. The context shapes the interpretation; *what is said* in an enriched sense may contain elements not present in the semantic representation of an utterance but introduced on the basis of contextual information. Nevertheless, Contextualists agree that an utterance's logical form influenced by the contextual bias is one of the inputs into the recovery of the sentence meaning.

If we return to example (1), in the Contextualist framework the computation of the scalar item would proceed roughly in the same way as Joanna's answer in (10) from the previous chapter, repeated here as (15):

(15)	a) Some cats love catnip.	utterance
	b) $\exists x\ [cat(x) \wedge love\ catnip(x)]$	logical form
	c) ⟦Some $cats_1$ $love_1$ catnip⟧	explicature
	d) Some but not all cats love catnip.	implicature

Contextualists are not in unison as to the timecourse of implicature processing. Some (Breheny et al. 2006, 2012, 2013) claim that implicature computation can be immediate and effortless. Others (Bott & Noveck 2004, Huang and Snedeker 2009, De Neys and

Schaeken 2007) believe that a logical interpretation should be faster than a pragmatic one because computing inferences is always an effortful process. For the eye-tracking experiment, Contextualism makes two diverging predictions: (1) participants will be able to quickly recognize whether the speaker is being cooperative or not and immediately limit themselves to *what is said* in the latter or make the effort of computing *what is meant* in the former; (2) participants will swiftly compute the implicature and based on the speaker type cancel or retain it.

4. Experimental Studies

There is much debate between supporters of the Neo-Gricean, Defaultist and context-driven approaches. In this chapter I will discuss several experimental papers from both groups. However, many scientists stray away from the debate and profess themselves undecided, like Bonnefon et al. (2009), Storto and Tanenhaus (2004) and Bezuidenhout and Cutting (2002). Until recently most experimental findings supported more or less definitely the Contextualist theory, but now more and more scientists point to Defaultism.

This chapter will briefly summarize several studies on scalar implicatures that found evidence for both introduced processing theories. The choice of papers is a highly subjective one and should by no means indicate that the listed studies are the best or most important in the field. My hope is that they give some insight into the current research on implicatures. Sadly, dwelling too long on any of the studies or mention all influential work on scalar implicatures is not within the scope of the present work.

4.1. Neo-Gricean

Presently, there have been few experimental studies that find support for the Neo-Gricean view on implicatures. Geurts and Pouscoulous (2009) investigated embedded implicatures and found relatively few local scalar implicatures in complex sentences. The authors argument extensively against what they call the “conventionalist” view (roughly equivalent to Defaultism) and claim that their results prove that scalar implicatures do not occur in embedded positions. To recapitulate, Defaultists argue that implicatures are always computed as the default interpretations of a scalar trigger. Geurts and Pouscoulous (and Geurts, 2009) propose that embedded implicatures are not really implicatures and that such contexts are special cases.

Dieussaerta et al. (2011) conducted two experiments in which they investigate the effects of cognitive load and repetition of the statements on scalar implicature processing. They found that under high cognitive load and with a limited working memory participants had a harder time reaching a pragmatic interpretation. Furthermore, a higher number of filler statements led to fewer consistent answer patterns. The authors conclude that their findings argue against the automaticity of inferences. Dieussaerta and colleagues argue that default pragmatic processes ought not have been more difficult under high cognitive load than under "normal" cognitive load situations. Furthermore, including a large number of filler sentences prevented the development of a response strategy, and therefore consistency, which – again – should not hold for automatic processes.

4.2. Contextualism

Huang and Snedeker (2009) studied implicatures in visual-world paradigm experiments. Participants heard sentences like "Point to the girl that has some of the socks" and were forced to choose between a girl with two of four socks and one with all soccer balls. The first two experiments revealed a substantial delay in decision-making after the ambiguity has been resolved. The authors attribute this to a lag between semantic and pragmatic processing. The last experiment examined how *some* would be interpreted if the competitors were inconsistent with the semantics (girl with socks vs. girl with no socks). Findings show quick resolution of the target, indicating that pragmatic analysis was the cause of the previous delay. Results provide direct support for the Contextualist view.

De Neys and Schaeken (2007) investigated the nature of implicatures in a dual task experiment. Participants were asked to verify the truthfulness of sentences like "Some oaks are trees." Depending on whether they interpreted the scalar item logically or pragmatically (*some and possibly all* vs. *some but not all*) the sentence would be judged false or true, respectively. Additionally, cognitive resources were strained by the second task: memorization of complex dot patterns during the interpretation process. Findings showed that participants produced more logical than pragmatic decisions when under heavy cognitive load, therefore pointing to the Relevance Theory view.

Breheny et al. (2006) conducted several self-paced reading experiments in which they tested the default nature of scalar implicatures in upper-bound, lower-bound and truly neutral contexts. The first experiment was focused on the scalar alternatives "and" and "or" and revealed that reading times on the implicature trigger *or* in upper-bound contexts were significantly longer than in the lower-bound contexts. The second experiment investigated implicature generation in neutral contexts.[6] The authors manipulated the sentence position of the scalar trigger (sentence-initial or sentence-final) and the explicitness of the implicature by contrasting *some* with *only some*; the latter strongly disambiguated towards "not all". The results show that sentence-final *only some* condition was read the fastest and sentence-final *some* condition was the slowest. Reading times on the implicature segment *some* were significantly lower in the upper-bound contexts.

These findings support the context-driven approach, because implicatures are not generated when there is no contextual information about whether the lower or upper-bound in relevant. They conclude that implicatures are generated only if they are explicitly licensed by context.

In a very well known study, Bott and Noveck (2004) put Contextualist and Defaultist theories to a test. The authors conduct a series of sentence verification experiments in which they focus on less-than-maximally-informative utterances such as "Some elephants are mammals" (Bott and Noveck 2004, pp. 437) These sentences, which the authors call underinformative, are a special case because they are false with the inference and true without it. The first experiment uncovered that subjects are less accurate and take significantly longer to answer correctly when they are prompted for a *some but not all* interpretation rather than a *some and possibly all* one. The second experiment was a modified version of the first one and replicated its results. Longest response times in the third experiment were found when subjects read the underinformative sentences and computed an implicature. The last experiment revealed that the rate of scalar inferences was directly proportional to the length of the response time.

4.3. Defaultism

Grodner et al. (2010) conducted an eye-tracking experiment in order to examine whether the scalar inference is universally delayed. They compared three quantifiers (*summa*, *alla* and *nunna*, phonetically truncated variants of *some of*, *all of* and *none of*) and the point of disambiguation for summa (early or late.) Participants followed spoken instructions directing them to interact with items in a visual world paradigm.

Grodner and colleagues found that subjects fixated on the target compatible with the implicated meaning of *some* prior to a disambiguating noun. The convergence on target

[6] A neutral context, so the authors, is one in which, "there is no information about whether all that is relevant is the lower- bound and no information about whether all that is relevant is the upper-bound" (Breheny, Katsos and Williams 2006, pp. 445.)

was as fast for *some* as for *none* and *all*. They conclude that their findings support the Defaultist view because scalar implicatures were computed immediately and were not delayed in processing relative to the literal interpretation of *some*. Therefore, with sufficient contextual support, processing delays disappear.

One of the main interests of Foppolo (2006) was the cost of implicature computation. In a reaction-time study she manipulated the environment of the scalar item (downward-entailing vs. not downward-entailing) and the type of sentence context (favoring inclusive vs. exclusive interpretation of *or*.) The task was to judge the truthfulness of a sentence in its context. Foppolo found that implicature computation appeared costly only in downward-entailing contexts, hence when their addition leads to a weakening of the overall assertion. She attributes this, however, not to implicature computation costs *per se* but to information loss caused by the presence of the implicature in a downward-entailing context. Contrary to what Contextualism would expect, there was no other cost of implicature derivation in non downward-entailing contexts.

Freeney et al. (2004) studied the implicature derivation in normal vs. underinformative contexts (ex. “Some elephants have trunks”) in children and adults. The authors find that children become sensitive to implicatures in less-than-maximally-informative utterances very early on and 8-year-olds showed much greater sensitivity to the implicature in pragmatically enriched than underinformative contexts. Adult subjects took longer to respond logically to infelicitous *some*-utterances than felicitous *some*-utterances. The results suggest that logical responses are time-consuming but some adults develop the ability to inhibit a pragmatic response in favor of a logical one. Therefore, the authors conclude that their findings are in line with the Defaultist theory, because the pragmatic processing is faster than logical processing in the default cases.

5. Eye-Tracking Experiment

Conversational situations where cooperation is not a given have so far not been systematically investigated. Experimental studies focus strongly on syntactic constraints to implicature derivation. Sadly, communication goals in everyday conversations may rely on half-truths and straight out lies.

The following experiment was aimed at answering the question whether implicatures are at all computed in uncooperative situations. If indeed they are, then how does the computation proceed? My hope is that the experiment will contribute to the debate between

Contextualists and Defaultists. The exact predictions for both theories can be found in section 5.2.3 later in this chapter.

The study was designed as a partial replication of Grodner et al. (2010). However, before plunging into the intricacies of eye-tracking, I had to make sure that the methods I intended to use were not too complicated for the participants and not too obvious. Furthermore, there was some uncertainty as to how well the German sentence materials I intended to use correspond to English *some*. Lastly, I wanted to ensure that the sentence and picture materials were optimal for the experiment task (see section 2 below.)

5.1. Pretest

5.1.1. Aims

The pretest was aimed at confirming that the German quantifier “einige” (*some*) carries the implicature “some but not all”, since I have not come across a study confirming this obvious intuition. I suspected that the German “einige” does not fully equal the English “some”. It is not clear whether the former can be used to mean “one” as its English counterpart can, for example: “Some cat is lying on your front lawn.” In German the same sentence would translate to: “(Irgend)eine Katze liegt auf deinem Rasen.” (*Some/A cat is lying on your front lawn*).

The second rationale behind the pretest was to test the materials for the main experiment which included determining good filler sentences and testing which ratio of target-color to other pebbles works best for the quantifier *some*. The pretest materials are very similar to the cooperative speaker conditions in the main experiment, thereby serving as a good control and comparison for the main experiment.

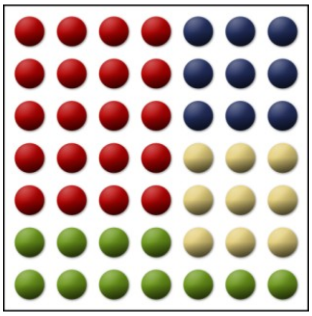
5.1.2. Materials

The sentence material consisted of 16 items in 4 conditions: quantifier *some*+picture *some*, quantifier *some*+picture *all*, quantifier *all*+picture *all*, quantifier *none*+picture *none*. Sample sentence and picture items are in Figure 1 below. As noted above, quantifier *some*+picture *some* items had a varying ratio of target-color to rest of pebbles (10:39, 15:34, 20:29 and 25:24.)

a. *some*+*some* condition

In meinem Sack sind einige Steinchen rot.

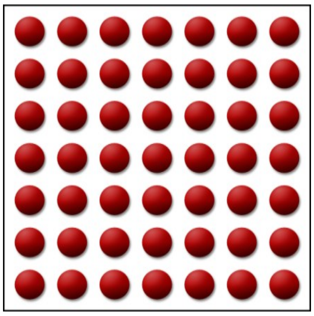
Some pebbles in my bag are red.



b. *some*+*all* condition

In meinem Sack sind einige Steinchen rot.

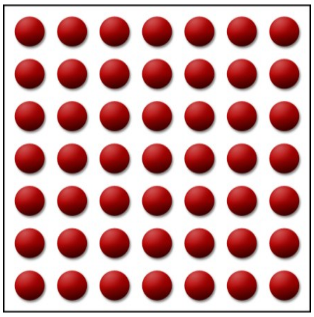
Some pebbles in my bag are red.



c. *all*+*all* condition

In meinem Sack sind alle Steinchen rot.

All pebbles in my bag are red.



d. *no*+*no* condition

In meinem Sack sind keine Steinchen rot.

No pebbles in my bag are red.

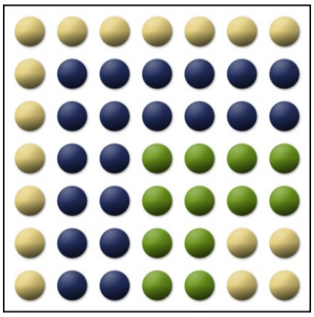


Figure 1. Pretest example item.

A control condition of blatantly false sentence-picture pairs was added with the following quantifier-picture pairs: *no*+*all*, *all*+*some*, *some*+*no* and *exactly one*+*all*. 64 filler sentences with a variety of different structures unrelated to the items ensured that the purpose of the study was obscured. All stimuli were presented in one list.

5.1.3. Procedure

The experiment was an online study done in Webexp2 (Keller et al. 2009). Participants viewed a display screen on which a picture and a sentence were visible and they participants were instructed to make judgments on sentence-picture pairs. If the participant believed that a sentence described the picture truthfully, she was asked to rate the sentence's naturalness on 7-point Likert-scale from 1 (not natural) to 7 (very natural). If, however, she believed that the sentence was a lie altogether she could mark it as such and thereby opt out from giving a rating.

5.1.4. Participants

32 German native speakers (20 female, mean age 25.38 years, range 17-37 years) from Tübingen University participated in the study. None of the volunteers reported colorblindness. They were naïve to the purpose of the study and were paid standard rates for participation.

5.1.5. Results

Response times (RTs) were recorded but could not be well interpreted. The participants were not performing in controlled conditions and could have made unreported breaks during the response phase, thereby skewing the results. RTs ranged from 177 to 174608 ms with a mean of 5965.82 ms and standard deviation 6146 ms. RTs shorter than 1500 ms and longer than 18000 ms were excluded from the analysis as outliers (3.2% of the cases.) The RT distribution was not normal, and hence the three-sigma-rule could not be applied in this case. In order to correct for outliers I chose to exclude RT longer than the mean RT plus two standard deviations. Determining the sensible boundary for extremely short RT was more difficult. The task consisted of reading a ±7 word sentence, comparing it with an image and choosing the answer. This could hardly be completed sensibly in under 1500 ms. Figure 2 presents the distribution of RTs.

For the truth value judgments, sentence-picture pairs of items were accepted as true 96.1% of the time and the false controls were rejected in 92.4%. Table 1 displays truth value and Likert scale judgments of the items (1-4) and controls (5-8) in all four conditions in detail.

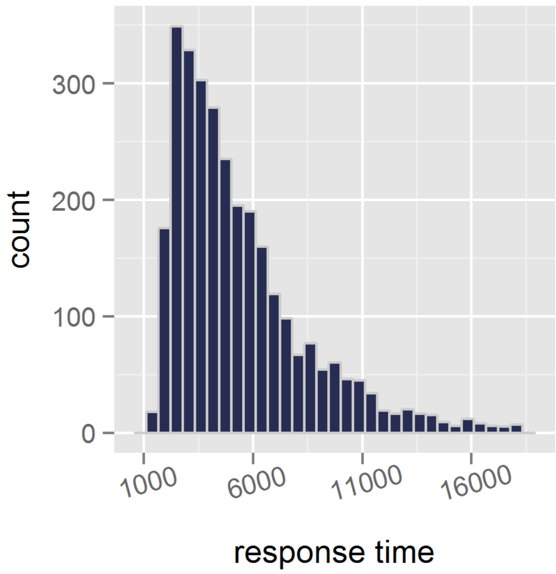


Figure 2. Density plot function of pre-test response times.

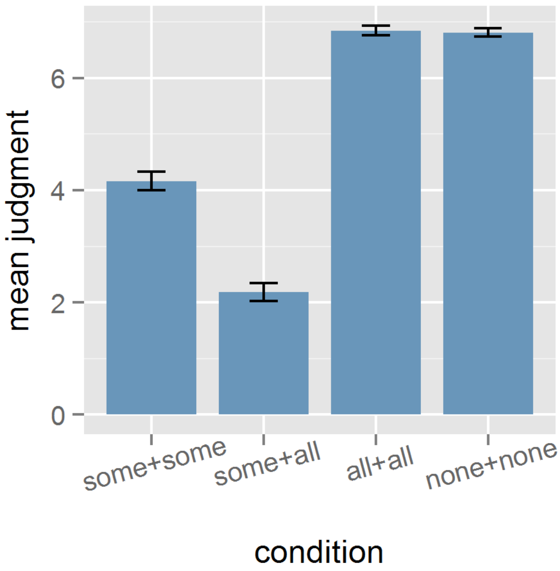


Figure 3. Mean judgments of the critical conditions in the pretest. Error bars are standard errors of the mean.

The participants judged *some*-sentences ("In my bag are *some* red pebbles.") paired with pictures on which a portion of the pebbles was red as true in all of the cases. Sentences with the quantifier *all* and *no* received equally high acceptance rates (100% and 96% respectively). The *some*-sentence paired with an *all*-picture was rejected a bit more frequently, although nowhere near as much as the false controls (11% and 92.4% respectively). The lower acceptance of *some+all* sentence-picture pairs was mainly due to 3 of the participants very consistently rejecting condition 2 items (two 100% and one 75% of the time). Without these three participants, the acceptance rates went up to 97%, equaling the other truthful conditions.

A visual inspection of Figure 3 reveals that there were dramatic differences in judgment between *all+all*, *no+no* and the *some* conditions. Indeed, an ANOVA with subjects as repeated measures reveled that there was a significant difference between the conditions ($F_{subject}$ (3,87)=180.3, $p<0.001$, F_{item}(3,12)=58.6, $p<0.001$.) In addition to high acceptance rates, *all+all* and *no+no* sentence-picture pairs were judged equally and very highly on the Likert scale ($F_{subject}$(1,29)=0.0, $p>0.05$; see Table 1). Both the *some+some* and *some+all* conditions were judged lower in comparison ($F_{subject}$(1,29)=257.9, $p<0.001$).[7]

[7] The subjects excluded from this and the following analysis had missing values in at least one item in one condition.

Condition		Truth value judgments	Judgments on Likert scale
some+some	Mean	1.00	4.16
	N	124	124
	Std. Deviation	.000	1.823
some+all	Mean	.89	2.19
	N	127	113
	Std. Deviation	.314	1.677
all+all	Mean	1.00	6.85
	N	111	111
	Std. Deviation	.000	.926
no+no	Mean	.96	6.81
	N	126	121
	Std. Deviation	.196	.830
no+all	Mean	.06	1.00
	N	126	8
	Std. Deviation	.245	.000
all+some	Mean	.06	1.13
	N	126	8
	Std. Deviation	.245	.354
some+no	Mean	.06	1.75
	N	128	8
	Std. Deviation	.243	2.121
one+all	Mean	.10	1.00
	N	127	13
	Std. Deviation	.304	.000
Total	Mean	.51	4.73
	N	995	506
	Std. Deviation	.500	2.514

Table 1. Truth value judgments and scale judgments for items and false controls in the pretest.

A pairwise comparison between the *some*+*some* and *some*+*all* conditions on z-transformed data revealed that *some* quantifier paired with a *some* pictures is judged significantly higher in acceptability than the same quantifier paired with an *all* picture ($F_{subject}(1,29)=132.4$, $p<0.001$. Not enough of the false controls were given scale judgments to license a comparison of means.

A series of Fisher's exact tests was performed for the truth value judgments. This method was chosen because the data had no variance in two of the conditions (*some*+*some* and *all*+*all*) which rendered both ANOVAs and generalized linear mixed models unreliable. The analysis revealed a significant difference in judgments between *some*+*some* and *some*+*all* conditions for all subjects ($\chi=14.477$, df=1, $p<0.001$). However, this difference was only marginal when the three strongly pragmatic participants were excluded ($\chi=4.001$, df=1, $p<0.063$). There was a marginally significant difference between *some*+*some* and *no*+*no* ($\chi=5021$, df=1, $p<0.06$) as well as between *some*+*all* and *no*+*no* conditions ($\chi=4.533$, df=1, $p<0.054$). The "worst" condition (i.e. *some*+*all*, mean 0.89) was compared with the "best" false control (*one*+*all*, mean 0.1) and the difference was significant beyond doubt.

As mentioned before, the *some*+*some* condition pictures had different numbers of pebbles in each item to test what is the preferred ratio for "some". Pictures with 10 dots or pebbles being in the target color were judged better than all other alternatives with a mean judgment of 5.61 (see Table 2) Indeed, the number of dots of the target color had an influence on the judgments ($F(3,84)=22.6$, $p<0.001$.) The item with the lowest target to rest ratio 10:39 was judged higher than the rest ($F(1,28)=9.8$, $p<0.05$).

Item	Mean Scale Judgment	N	Std. Deviation
Some 10 (10:39)	5.61	31	1.230
Some 15 (15:34)	4.34	32	1.771
Some 20 (20:19)	3.77	30	1.569
Some 25 (25:24)	2.90	31	1.578
Total	4.16	124	1.823

Table 2. Mean judgments on a 7 point Likert scale for condition 1 (some+some).

5.1.6. Discussion

The results show that "einige" (*some*) carries the implicature *not all*. All participants save three found that a sentence with the quantifier *some* fits an *all* picture as well as a *some, no* and *all* quantifier sentences fit *some*, *no* and *all* pictures respectively; blatantly false sentence-picture pairs were rejected much more frequently than *some*+*all* condition. The results provide further incentive for the main experiment, because if "logical" (some+all) and "pragmatic" (some+some) *some* are accepted equally frequently then the quantifier can be safely used in the main experiment with both of its meanings without the logical interpretation being impaired or discarded as a lie.

An important discovery was that three out of the 32 participants were very strongly pragmatic and rejected *some*+*all* sentence-picture pairs. It is a crucial point to make, because strong pragmaticians among test subjects could lead to obscured results. This finding made it clear that personal preference of participants need be looked at during data analysis, licensing perhaps to two separate statistical analyses.

Analysis of scaled judgments for items showed strong effects of condition. Differences in judgments could be attributed to the fact that both *all* and *no* are very clear cut cases whereas *some* is somewhat obscure. The purpose of the pretest was – among other things – to find the optimal picture for *some*. The item analysis found that for the most part the 10:39 ratio of target color pebbles to rest was preferable for *some* and therefore it was adopted for the main experiment. The initial intuition of German native speakers I have consulted before this experiment was, however, that the preferred ratio should be much higher and closer to 30:19. It is possible that the number of target color pebbles for *some* was influenced by filler items which contained quantifiers *most*, *few*, *many*, *a couple*, *half* and numbers thereby pushing *some* into a very particular and small range. However, these finding support Degen and Tannenhaus 2009 results. They found that the quantifier *some* was judged highest on a 7-point Likert scale when paired with pictures of 5 to 7 of 13 gumballs. It is interesting that, contrary to the initial intuition, *some* seems to oscillate around the same low number. The pretest didn't test ratios with less than 10 pebbles for some and perhaps 5:44 ratio would have been an even better fit.

5.2. Game Experiment

The main experiment was in part a replication of Grodner et al. (2010), and in part an expansion towards uncooperative uses of implicatures. The best items and fillers from the pretest formed a template for the materials in the present experiment.

5.2.1. Procedure

The experiment was conducted in a quiet experimental room with an SR Research EyeLink 1000 eye-tracker desktop mount with a 35 mm lens, 9 point calibration and 1k sample rate and pacing interval. The participants sat 62 cm away from the screen and 55 cm away from the camera lens. The items were displayed on a 19 in. Belinea 106055 CRT color monitor with a 1024x768 pixel resolution and 85 Hz refresh rate. Mouse and keyboard were used to navigate in the experiment.

Preceding the experiment the participants were tested for colorblindness using 24 Ishihara plates. Viewing was binocular but only the dominant eye, as determined prior to the experiment by a test similar to the Dolman method, was tracked (left eye for 8 participants.) They were instructed to blink normally and avoid moving the head throughout the eye-tracking part of the experiment but encouraged to take breaks if needed. The eye-tracker was recalibrated after every break, before the exercise and experiment and fixations were manually accepted between trials to enable additional calibration if needed.

Eye movements were recorded only during the last part of the experiment. The auditory stimuli were played via speakers situated to the left and right of the screen.

Participants were randomly assigned to one of two lists. In the first list the female speaker was cooperative and the male uncooperative, whereas in the second one their roles were reversed. Participants were tested in individual sessions of about 1 hour each. The experiment itself was divided into three stages. The first two stages were for training purposes and the third one was the critical testing one. Participants had obligatory breaks between the stages and could take a break in between the trials. They were told that they were going to learn a game and play it for several rounds with two other players, followed by an experiment based on this game.

In the first part participants read a short game manual (a summary of which can be found in Appendix C, pp. iii) and were allowed to ask questions about the game. Afterwards they played nine rounds with a female and male player, one of which was cooperative towards the participant and the other was not. The players' roles changed in accordance with the list the participant was assigned to. The cooperative player shared his pool of points and played on one team with the participant, hence it was in the best interest of these two to be maximally informative with each other. The uncooperative player had a separate pool of points and played against the other two. It was therefore in the best interest of both the other player and the participant to try to deceive him and be minimally informative. All rounds were fixed and the co-players were experimenters and had an answer key. None of the participants reported noticing that the game was predetermined.

The game itself was similar to Skat. The aim was to collect as many pebbles of a certain color as one could. The players were given a card representing the color of pebbles they would be gathering throughout the entire game and one game card each. The game cards were similar to picture items in the experiment. The players asked each other in turns about the color they were collecting, starting with the cooperative and uncooperative players and ending the round with the participant. When it was her turn, a player asked other game participants whether they had her-colored pebbles on their cards. The co-players who were asked had to produce a quantitative utterance; they had to reply truthfully but were allowed to be obscure.[8] Based on their answers, the player who was asking then had to choose from whom she would draw a card, i.e. who she believed had more pebbles in the target color. Whoever she drew a card from drew a new card. At the

[8] For example, when a player had exactly 10 pebbles in the target color, she could say that she had "at least 4", "more than 9", "less than 20", "several", "a couple", "exactly 10", "some", "few" etc. She could not, however, say that she had "more than 10", "exactly 5" or "less than 10".

end of a round each player had two cards, with a total of 6 cards (3 original cards + 1 for every player). At this point players counted the pebbles in their color and the cooperative team summed up and divided theirs by two and compared them to the uncooperative player.

The rounds were designed in such a way that they familiarized the participant with situations she would face in the experiment. The primary aim of this stage was to introduce the two players and explain the differences in their behavior. Secondly, the game ensured that participants were able to produce cooperative and uncooperative utterances at appropriate times and respond to such utterance accordingly. To ensure that the participants had interest in optimizing their game performance, the winner of each round received 0.5 €. The game had pre-set rounds similar to the *no* and *all* condition in which the hostile player was seemingly cooperative in hopes of getting rid of an unfavorable card and acquiring a better one, which happened half the time.

Several participants produced utterances with “some” in this part of the experiment, but the experimenters were instructed not to. The utterances made by participants and experimenters were recorded during this preparatory stage.

The second stage of the experiment was designed to further familiarize the participants with the players, their respective (recorded) voices and types of utterances they make in the eye-tracking experiment. Two sound clips of 16 s. introduced the players. During the presentation of the clips a simplified picture of the person was displayed on a green or red background for the friend and opponent respectively. Next, the participants viewed an image of a set of pebbles and heard a sentence describing the image, as in Figure 4.

The task was to decide whether the voice belonged to the cooperative or uncooperative player. This exercise was repeated ten times per player type and the participants received feedback on their decision. This ensured also that in the main part of the experiment the participants are aware of the speaker context (cooperative or not) from the onset of the recorded utterance. The critical quantifiers were not used in the training phase.

The final part of the experiment was preceded by two dummy trials that were excluded from the analysis. Participants viewed a visual world display with three different pebble sets, as in Figure 5 below. The pictures’ centers were in equal distance from each other. The order of the pictures was varied across items and conditions so that the target could appear in any of the positions. The participants had enough time to familiarize themselves with the display. Additional information about the speaker, i.e. picture and background color, as well as the color of pebbles they were going to make an utterance about were provided at this point. Within this paradigm, subjects' gaze is the dependent measure.

After a period of 6.5 s a blue circle appeared in the center of the screen. Upon clicking on it after a 200 ms delay an utterance was played, for example*: In meinem Sack sind einige Steinchen grün.* ("In my bag are some green pebbles.") The task was to click on the set of pebbles that fitted the player's utterance best. In the experiment stage the friendly player would use *some* to communicate "some but not all", whereas the hostile one would use it to mean either "some but not all" or "some and maybe all". At the end of the experiment the participants filled in an evaluation form.

In meinem Sack ist fast die Hälfte der Steinchen rot.
Almost half of the pebbles in my bag is red.

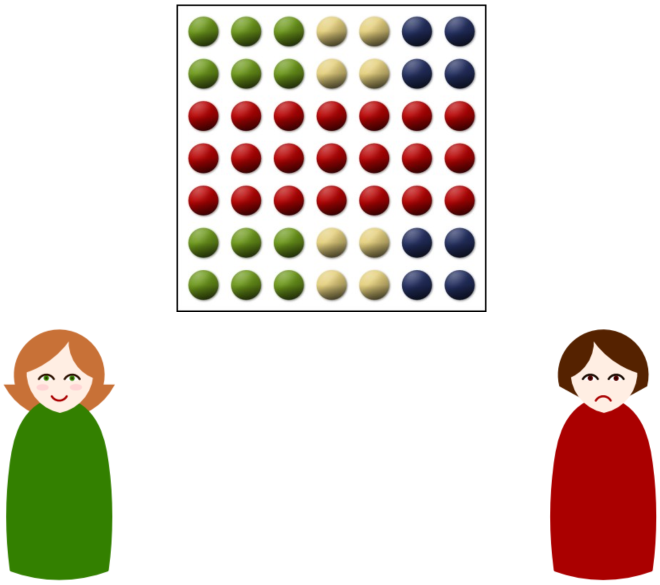


Figure 4. Exercise item in the main experiment.

A: In meinem Sack sind einige Steinchen rot.
In my bag are some red pebbles.

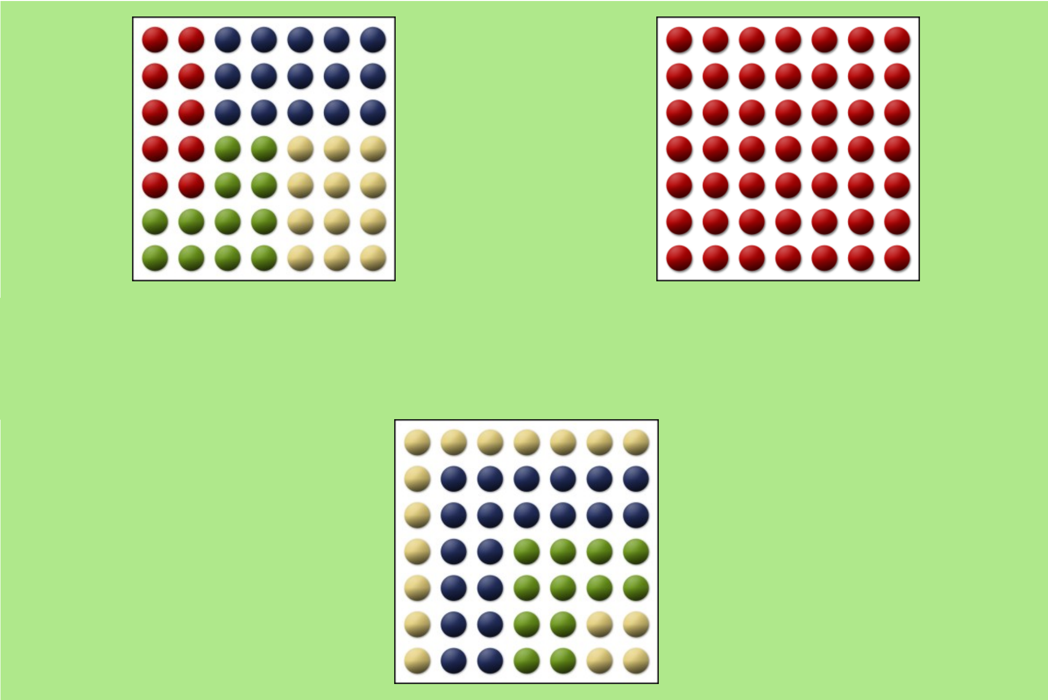


Figure 5. Experiment item in the main experiment.

5.2.2.Materials

Sentence and picture materials were for the most part the same as the ones used in the pretest. The factors were quantifier (*some*, *all*, *none*), picture (*some*, *all*, *none*) and speaker (*cooperative*, *uncooperative*). There were 6 experimental conditions: picture+quantifier *all*, picture+quantifier *none* cooperative and uncooperative, as well as picture+quantifier *some* for the cooperative speaker and picture *some*+quantifier *all* for the uncooperative one.

The pictures for the *some+some* condition were adjusted to the results of the pretest. All 48 items were truthful sentences with respect to the pictures. The items were distributed evenly across two lists in a within subject design. In the first the cooperative player's voice was female and the uncooperative player's male, and in the second they were male and female respectively.

The best fillers from the pretest, i.e. judged true in at least 60% of the cases and having a rating between 2.9 and 6.37, were included in the main experiment. Additional sentence-picture pairs were constructed on the base of those fillers in all colors. The pictures had the same color proportions as the model fillers but a slightly different distribution. This yielded a total of 112 filler sentence-picture pairs. In the fillers there were always one (in the cooperative condition) or two possible targets and a distracter. For the training phase, 20 exercise items similar to but different from the experiment items were created.

The hostile and friendly player utterances were recorded with two German native speakers with previous experience in theatrical or radio performances. One was female and the other male and they were encouraged to speak in a natural pace. To normalize the auditory stimuli and ensure that the recordings were identical across conditions up to the quantifier region, splicing was used for all items. Furthermore, quantifiers were spliced to ensure that the quantifiers started and ended in the same millisecond across conditions. The recordings were played from the hard disk as 44 kHz mono sound clips with a bit rate 700 kBits/s. The sound files for items were judged by three independent listeners as "natural sounding."

5.2.3. Aims, Hypotheses and Predictions

The aim of the present study is twofold. The initial interest was to discover how hearers interpret expressions that carry a GCI in circumstances where cooperation in the classical sense is not presupposed, and even the opposite of it is expected. According to the Gricean view on implicatures, the hearers should not compute them in such a situation, seeing as the CP is not obeyed. From a Relevance Theoretical perspective, the context makes it indeterminate whether an implicature should be retrieved. Defaultists would argue that implicatures will be computed and later perhaps cancelled, whereas Contextualists would see no reason to retrieve a costly implicature.

Another reason behind this experiment was to investigate whether the computation of implicatures is quick and automatic or whether it is a time and resource consuming process. If the present study shows results similar to those of Grodner et al. 2010 and the fixations on the scalar term *some* will be as prompt as on the controls *all* and *none* then the conclusion is that implicature derivation is a very fast and low cost process. The utterances of the cooperative player provide a context that facilitates implicature computation, whereas the utterances made by the uncooperative player provide an ambiguous or indeterminate context that could facilitate both the logical and pragmatic interpretations of *some*.

An example of a situation where it would be profitable for the hostile player to use *some* in the logical, implicature-less sense is the following. The hostile player has in fact very few or even only one pebble in the color she was asked for. By choosing *some* in such a case she is hoping that the listeners do not draw or cancel the implicature, arriving at the "some and maybe all" interpretation. The listener may then be tempted to draw the hostile player's card, which – as mentioned before – is in fact undesirable.

On the other hand, it the hostile player has all pebbles in the desired color, she might hope to trigger the implicature. The listener who computes the implicature will be less likely to draw the card from the hostile player. The listener potentially puts herself in an unfavorable position by choosing a less desirable card. Therefore it was almost always unclear to the participants what "cards" the opposing player was holding.[9]

From a game theoretical point of view the best strategy for a hearer would be to not draw an implicature if the computation of it costs processing effort. On the other hand, if the derivation of implicatures is an automatic process, the hearer will compute an implicature every time. Since the hearer cannot know for certain, and therefore must guess what the speaker is communicating, she can adopt either of the strategies because her chances of being correct are 50/50.

Contextualism would predict that implicature computation of scalar items in the deceptive utterances of the uncooperative speaker be delayed in comparison to the cooperative speaker. This would provide evidence of the influence of the context on the retrieval of implicatures and the costliness of implicature retrieval. However, if implicature computation in the uncooperative cases is not delayed then these findings would support the Defaultist view. This would mean that implicature computation is such an automatic and low cost process that it is preferable for the listener to draw the implicature rather than cancel it. This would mean that listeners are able to perform a computation quickly proving implicature computation to be a very fast process. Taking the context into consideration and calculating an appropriate strategy would have to occur in a matter of milliseconds.

There is also the possibility that implicatures will not be computed at all in the non-cooperative situations. Such findings would be in line with both theories. Contextualists would argue that in this context, arguably unfavorable for the retrieval of implicature, or

[9] An exception was if the hostile player said *none* or *all*. In these cases it was clear how many pebbles of a given color she was holding because they allow only one truthful situation. Sentences with numerals alike *five*, *eight* and *ten* could arguably give rise to implicatures …*and not more* or *exactly*, thereby causing the same kind of difficulty or confusion as the scalar term *some*.

at least less favorable than in a neutral situation, the cost of implicatures would be too high to even consider computing them. Griceans would point out that the uncooperative setting causes the dialogue participants to reject the Cooperative Principle and prevents the implicatures from arising in the first place.

5.2.4. Participants

22 native German speakers (3 male, ages 20-33, mean 24.77 years) from Tübingen University participated in the study. They were volunteers, naïve to the purpose of the study, have not taken part in the pretest and received 8 € compensation. All participants were right handed, had normal or corrected vision, tested negative for color blindness and had no hearing impairment.

5.2.5.Results

Evaluation Form

One person did not fill in any section of the form except for the comments. 61.11% of the other participants could concentrate well throughout the experiment and 29.41% found the experiment too long. They were asked to judge on a 7 point scale (1=very bad and 7=very well) how well the task was explained (mean 6), how well they could distinguish between the friendly and hostile players (mean 6.33) and how well did the friendly and hostile players play (mean 5.1 and 5.04 respectively.) Furthermore, mean difficulty rating of the task on 7 point scale (1=very difficult, 7=very easy) was 5.43. Several participants noted that the experiment was fun.

Response Times and Responses

Response times ranged from 3177 to 18619 ms with a mean of 453.02 ms (see Table 3.) Figure 6 presents the distribution of RTs. The distribution was not normal, similarly to the pretest. Trials in which RTs were larger than the mean and two standard deviations (≈8124 ms) were removed from the analysis. Obviously false responses in *all* and *none* conditions were also excluded from further analysis. Both constraints excluded a total of 4.17% of the cases.

Unlike in the pretest analysis, the shortest RTs were included in the analysis. The participants could answer with a button press only after the complete player utterance was played. There was no novel information in the utterance after the quantifier because the color of pebbles in question was known beforehand. Thus, the participants had enough information at the time of the quantifier to parse the sentence. In consequence, by the end of the recording they must have already computed the entire sentence and either drawn or cancelled an implicature, if required.

Figure 7 displays the mean RTs across the experimental conditions. There were no visible differences in response times between conditions. Moreover, the response times to *all* and *some* quantifiers were not different ($F_{subject}(1,21)=0.42$, $p>0.05$, $F_{item}(1,7)=0.72$, $p>0.05$). Participants responded significantly faster to *none* than *all* ($F_{subject}(1,21)=7.4$, $p<0.05$, $F_{item}(1,7)=8.2$, $p<0.05$, see Figure 8.)

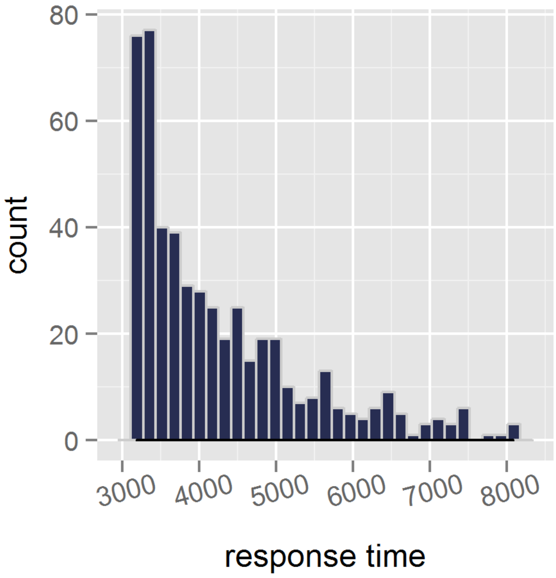


Figure 6. Density plot function of response times in the main experiment.

Condition	Mean	N	Std. Dev.
some cooperative	4299,43	84	1242,957
all cooperative	4265,91	87	1045,472
no cooperative	4115,35	86	960,557
some uncooperative	4472,54	80	1180,761
all uncooperative	4327,89	83	1233,894
no uncooperative	4038,91	86	933,201
Total	4250,14	506	1107,803

N	Valid	528
	Missing	0
Mean		4531,0
Std Error of the Mean		78,190
Standard Deviation		1796,7
Skewness		3,010
Kurtosis		13,090
Minimum		3177
Maximum		18619

Table 3. Response times to conditions in the main experiment.

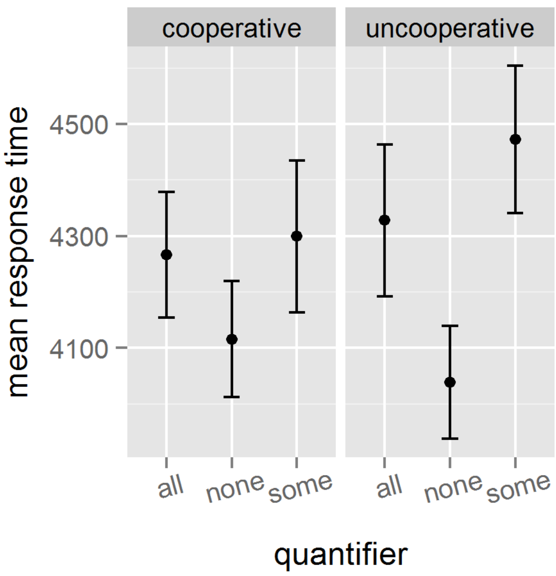


Figure 7. Mean RTs across conditions in the main experiment. Error bars are standard errors of the mean.

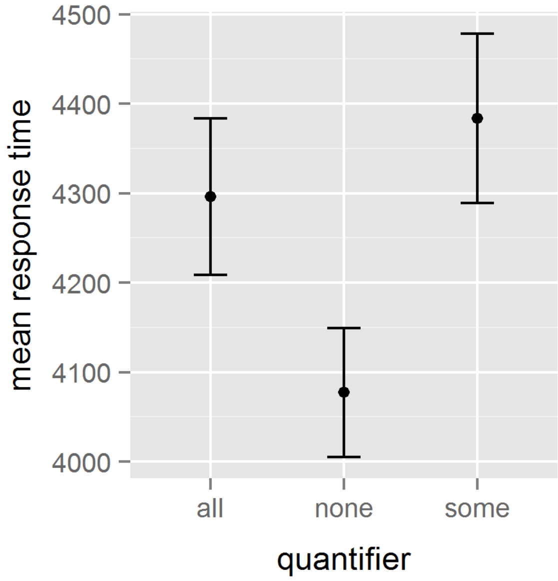


Figure 8. Mean RTs across quantifiers in the main experiment. Error bars are standard errors of the mean

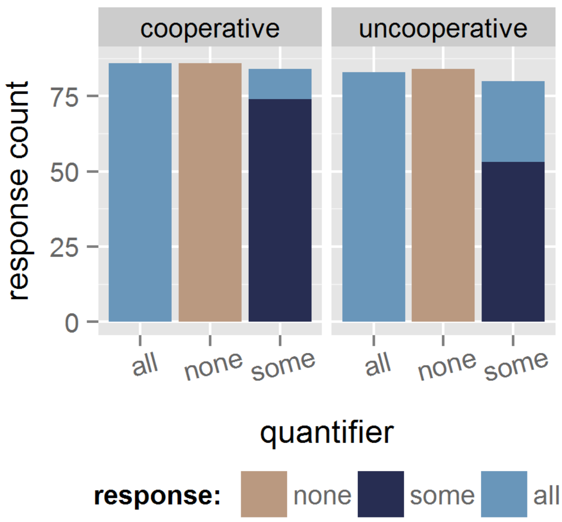


Figure 9. Responses to critical conditions in the main experiment.

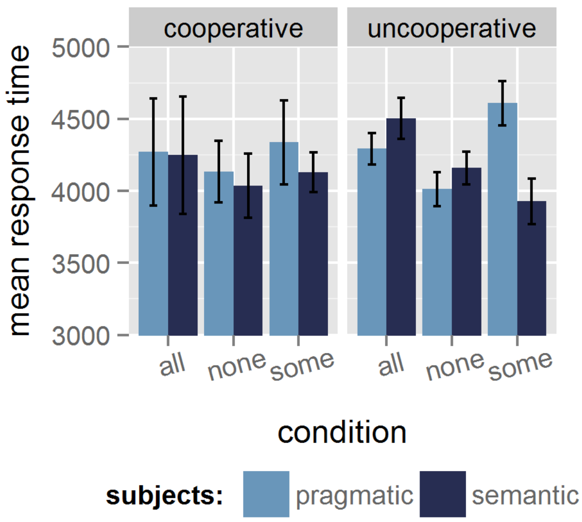


Figure 10. Response times in listener groups in the main experiment. Error bars are standard errors of the mean.

Condition	Responses			
	all	no	some	Total
some cooperative	10	0	74	84
all cooperative	86	0	0	86
no cooperative	0	86	0	86
some uncooperative	27	0	53	80
all uncooperative	83	0	0	83
no uncooperative	0	84	0	84

Table 4. Responses to conditions in the main experiment.

Mouse clicks on pictures in all conditions were recorded. The results depicted in Figure 9 clearly show a difference between the pictures chosen in the cooperative and uncooperative *some* conditions.

In the cooperative *some* condition, participants chose the some picture in 88.1% of the cases and in the uncooperative one only 66.25% of the time (see Table 4 and Figure 9.) Responses to *some* conditions were analyzed with a generalized linear mixed model with random effects *subject id* and *item*, fixed effects cooperation (*yes* or *no*) and dependent variable response (*all* or *some*). When the speaker was uncooperative, participants chose the *all* picture significantly more often than when the speaker was cooperative ($z=-3.6$, $p<0.001$).

The pretest revealed a group of strongly pragmatic participants and for that reason a subject analysis was the next logical step. Indeed, the participants could be divided into two groups based on their consistent responses in the *some* condition. The responses are summarized in Table 5. Four pragmatic participants consistently chose the *some* picture in the cooperative condition and the *all* picture in the uncooperative one. Three participants semi-consistently preferred the *all* picture irrespective of the speaker, accounting for the high error rates. In addition to the possible explanations for such answers given above, the fact that the strongly logical answers are restricted to a handful of participants might indicate that these participants simply did not understand the task or over-generated. The remaining participants were unaffected by the speaker manipulation, although the evaluation form revealed that they could differentiate between the two speakers.

Figure 10 shows a comparative view of response times in the two groups of participants. RTs for all conditions were similar in both groups save for the uncooperative *some* condition, where the non pragmatic participants were slower than pragmatic participants. However, since the number of participants varied greatly in the two groups, a statistical comparison between them was not possible. It remains for further research to look into the differences between strongly pragmatic and ordinary listeners.

Subject	Cooperativity	Response all	Response some	Subject	Cooperativity	Response all	Response some
1	cooperative	0	4	12*	cooperative	1	2
	uncooperative	0	3		uncooperative	4	0
2*	cooperative	0	4	13	cooperative	0	4
	uncooperative	4	0		uncooperative	1	3
3	cooperative	0	4	14+	cooperative	2	2
	uncooperative	0	4		uncooperative	2	2
4*	cooperative	0	4	15	cooperative	0	4
	uncooperative	3	1		uncooperative	0	4
5	cooperative	0	4	16	cooperative	0	4
	uncooperative	0	4		uncooperative	1	0
6	cooperative	0	4	17*	cooperative	1	3
	uncooperative	0	4		uncooperative	3	1
7	cooperative	0	3	18+	cooperative	3	1
	uncooperative	1	3		uncooperative	3	1
8	cooperative	1	2	19	cooperative	0	4
	uncooperative	0	3		uncooperative	0	3
9	cooperative	0	4	20	cooperative	0	4
	uncooperative	0	4		uncooperative	2	2
10+	cooperative	2	1	21	cooperative	0	4
	uncooperative	2	2		uncooperative	1	2
11	cooperative	0	4	22	cooperative	0	4
	uncooperative	0	4		uncooperative	0	3

Table 5. Subject analysis during sentences with quantifier *some* for main experiment.
* Participants who behaved consistently different depending on the speaker.
+ Participants who preferred the *all* picture irrespective of the speaker.

Eye-tracking Data

Fixations shorter than 80 ms were added to the preceding or following fixations if these fixations were within 0.5° of visual angle. The remaining very short fixations were deleted seeing as short fixations are often the result of false saccade planning and not meaningful information processing (Rayner & Pollatsek, 1989). Nearest interest areas (IAs) were used for outlier fixations. The entire sentence duration was split into 50 ms bins.

For the analysis of the visual world data I decided against an analysis of variance due to its disadvantages with this type of data. ANOVAs assume continuous dependent variables and cannot model time. Therefore, they are not well fitted for the kind of data here because the implicature processing theories outlined in the first chapters make important predictions about the temporal aspect of the results. In addition to that there were random effects of *item* and *subject* that I felt needed to be taken into consideration. There is currently much controversy over which analysis method is best suited for visual world paradigm data. It is not the aim of this paper to argue in favor of any of the analysis methods.

Growth curve analysis (Mirman et al. 2009), logistic power peak (Scheepers et al. 2008, Duñabeitia et al. 2009) and many other (Altmann & Kamide 1999, Heller et al. 2008, Ferguson et al. 2010) albeit in many ways superior to ANOVAs, are not very well established. For the analysis of the present study I decided in favor of a generalized linear mixed model (Barr 2008) because it requires fewest data set transformations and is the best established analysis method out of all the models mentioned above.

Baseline

The eye tracking data analysis was split into several parts. First I looked at a time window of 350 ms before quantifier onset to establish a baseline for the analysis of the post-quantifier region. A generalized linear mixed model for *fixations* defined by *cooperation* (cooperative, uncooperative), *picture* (some, all, none) and *quantifier* (some, all, none) with random effects *item* and *subject id* found a highly significant main effect of picture *none* E=-1.13, p<0.001 and *some* E=-2.3, p<0.001 (see Figure 11 and Table 6). The *none* image was fixated much less frequently than both *all* and *some* images. There were several marginally significant interactions which were most likely noise caused by the relatively low number of participants.

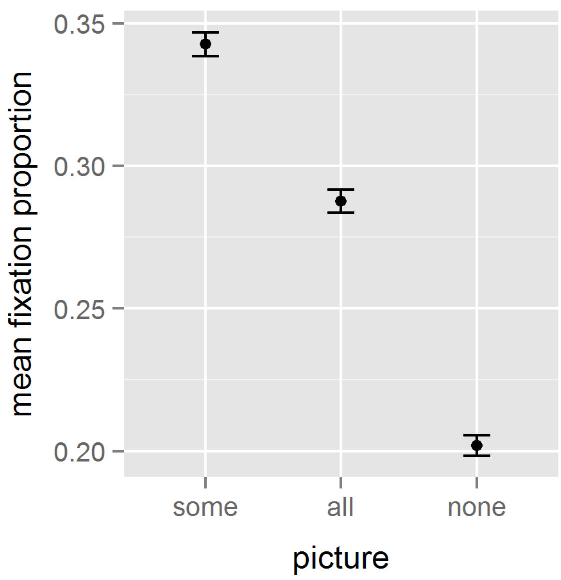


Figure 11. Mean fixation proportions on pictures some, all and none in the baseline.

Fixed effects:	Estimate	Std. Error	z value
(Intercept)	0.223	0.176	1.271
picture some vs. all	-1.126	0.232	-4.855***
picture none vs. all	-2.287	0.287	-7.977***
quantifier some vs. all	-0.045	0.225	-0.202
quantifier none vs. all	-0.015	0.222	0.068
cooperative vs. uncooperative	0.004	0.157	0.028
picture*quantifier some vs. all	0.262	0.328	0.800
picture*:quantifier none vs. all	0.3103158	0.3911711	0.793

Table 6. Generalized linear mixed model fit by the Laplace approximation for the baseline bins. Significance codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1.

Post-quantifier analysis

Next, I focused on the post-quantifier region. I looked at a 1400 ms time interval starting 200 ms before the quantifier onset. The fixations from the baseline were averaged into one bin and used as the reference point in time where no effect of quantifier was found or expected. The total number of bins is therefore 25.

Seeing as the form of the sentences was highly repetitive and that during the interest period the only unknown was the quantifier, it is safe to assume that participants arrived at the semantic and pragmatic form of the utterance before the sentence was completed. This was indeed visible in plotted fixation proportions because about 2000 ms into the sentence (1000 ms after quantifier onset) the proportions of fixations on any of the ROIs dropped dramatically (see Figure 12-Figure 17).

The analysis with factors *picture*, *quantifier, cooperation* and *bin* and random effects of *subject id* and *item* revealed a main effect of image *none* E=-0.8, p<0.01, similarly to the baseline (see Table 6 and Figure 11). There were no effects for any of the images and interactions between image and quantifier in all permutations. Importantly, there was no main effect for cooperation.

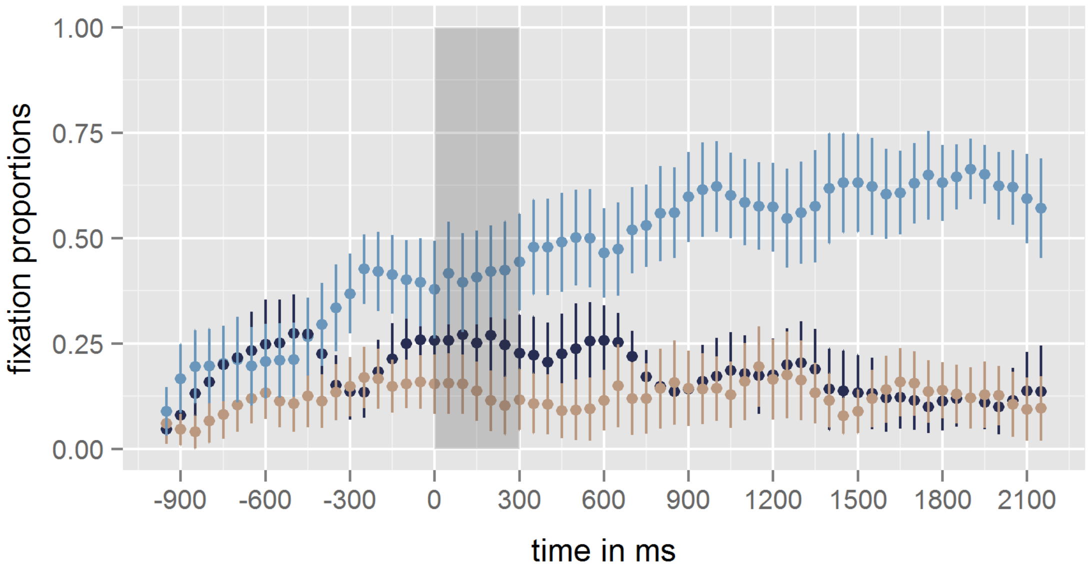


Figure 12. Fixation proportions on images for quantifier *some* and the *cooperative* speaker for the entire utterance time; quantifier duration shaded.

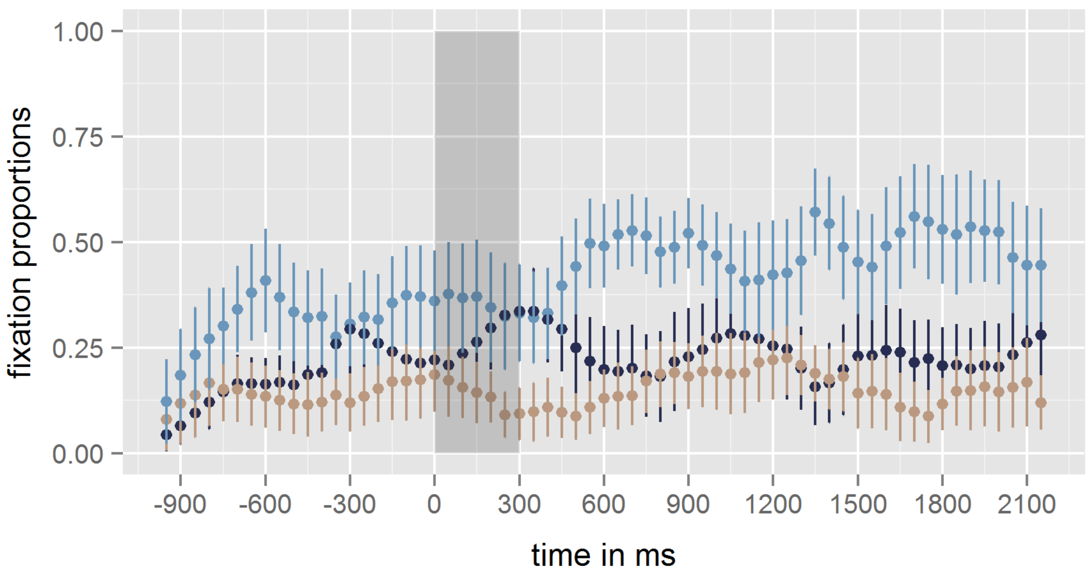


Figure 13. Fixation proportions on images for quantifier *some* and the *uncooperative* speaker for the entire utterance time; quantifier duration shaded.

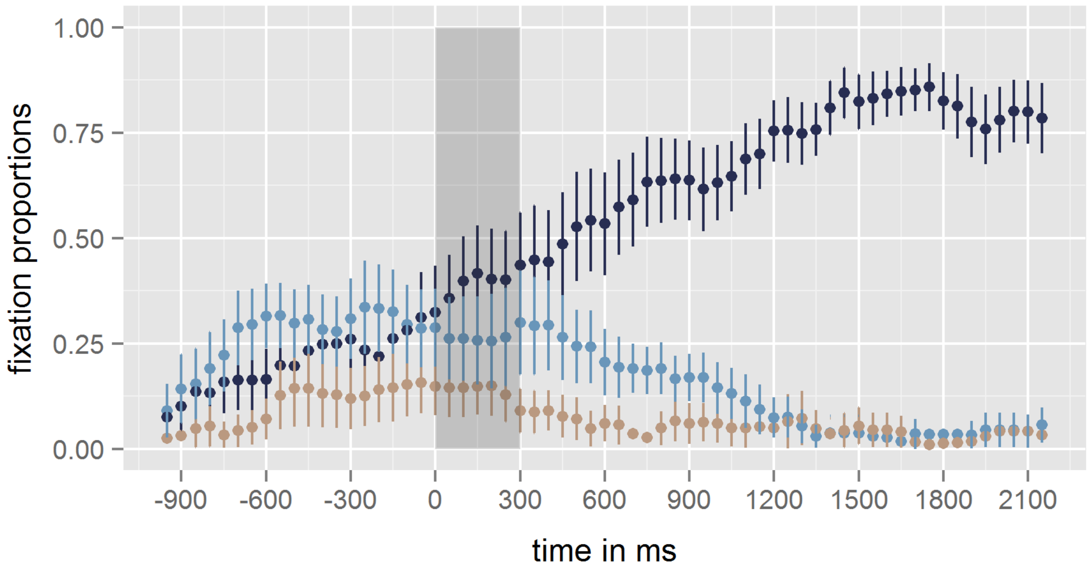


Figure 14. Fixation proportions on images for quantifier *all* and the *cooperative* speaker for the entire utterance time; quantifier duration shaded.

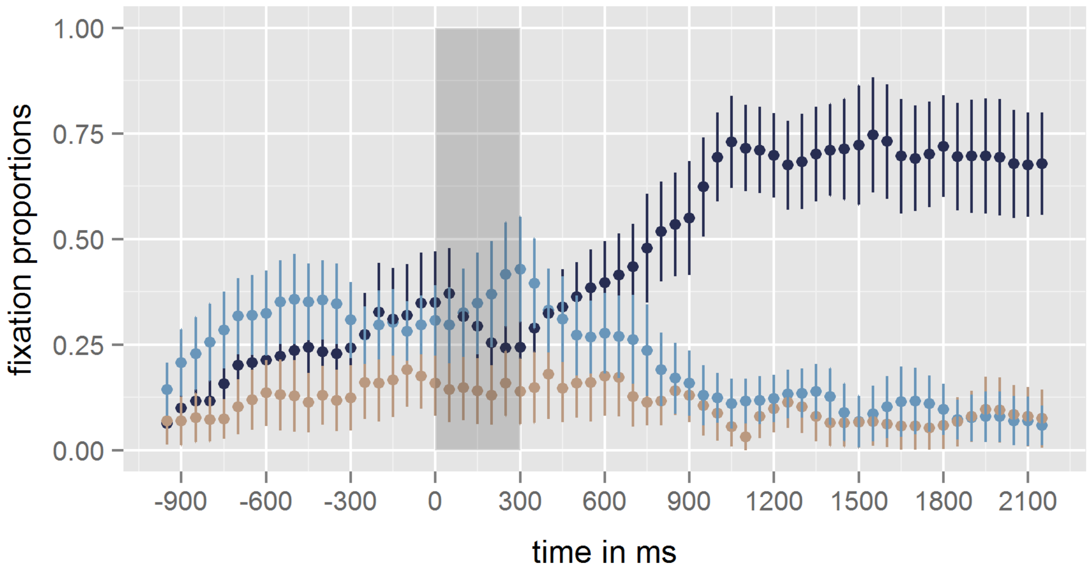


Figure 15. Fixation proportions on images for quantifier *all* and the *uncooperative* speaker for the entire utterance time; quantifier duration shaded.

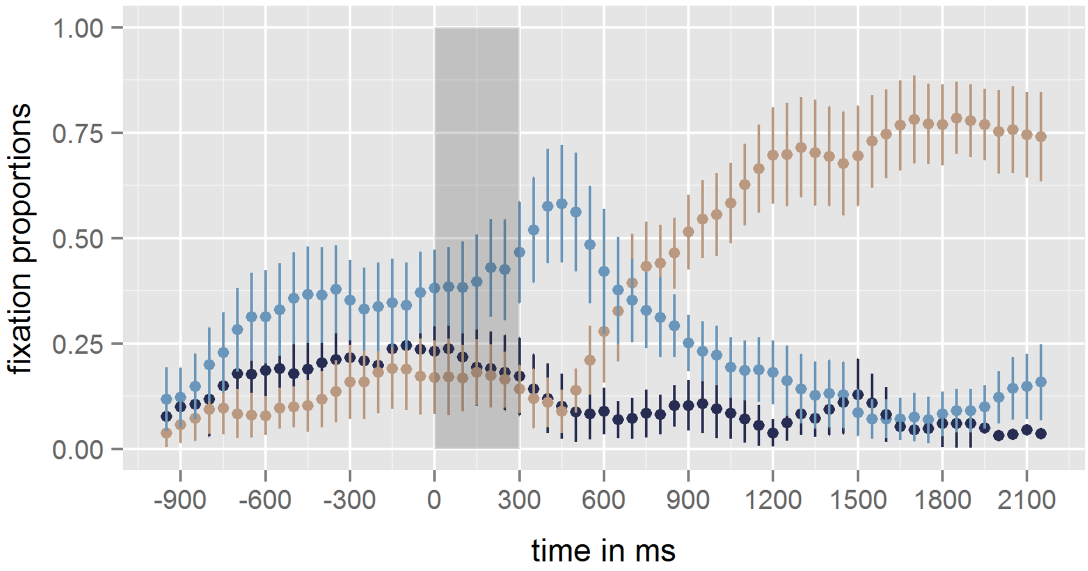


Figure 16. Fixation proportions on images for quantifier *none* and the *cooperative* speaker for the entire utterance time; quantifier duration shaded.

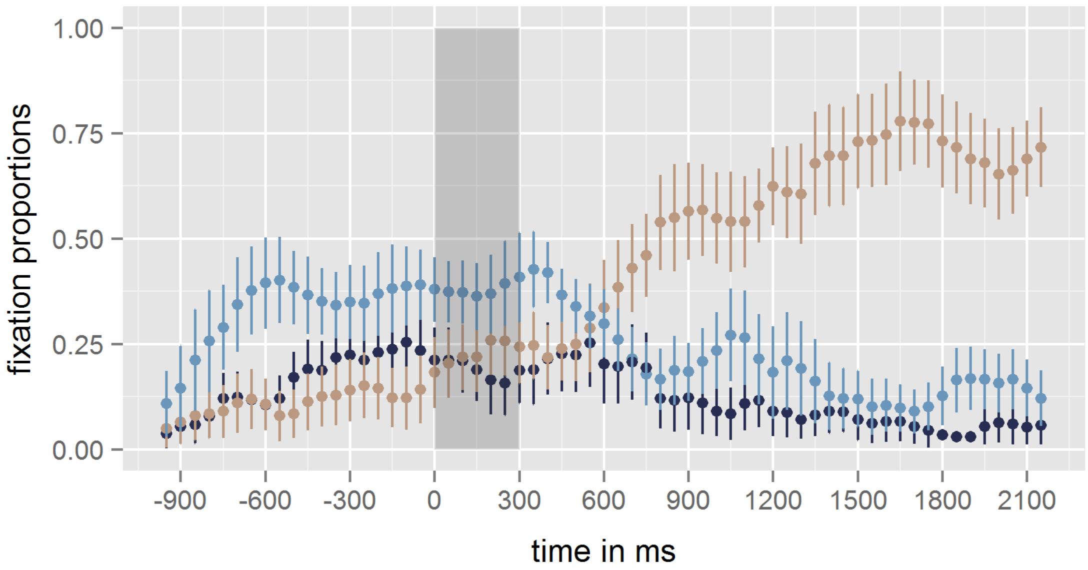


Figure 17. Fixation proportions on images for quantifier *none* and the *uncooperative* speaker for the entire utterance time; quantifier duration shaded.

Unsurprisingly, when listening to *some* sentences, participants fixated the *some* picture significantly more frequently than *all* and *none* pictures; listening to *all* sentences prompted them to fixate the *all* image more than both *some* and *none* (image *some**picture *some*, E=0.8, p<0.05). Interestingly, in the case of quantifier *none* the participants preferred the *some* picture (image *some**quantifier *none*, E=0.7, p<0.05) over *none* and *all* (see Figure 18)

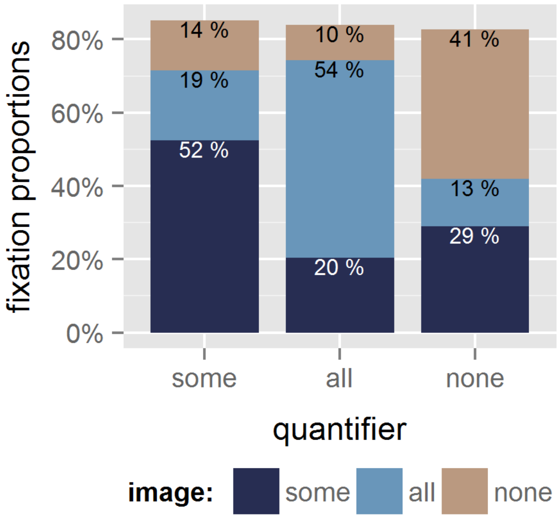


Figure 18. Mean fixation frequencies on the target images for some, all and none, separately for the quantifiers some, all and none, averaged across bins 11-20.

Next, let us consider the time course (factor *bin*) of the interaction between *quantifier* and the target of fixations (*picture*), i.e. the three-way interaction *bin*quantifier*picture,* while ignoring cooperativity for the time being. For the sake of clarity, we further restrict the analysis to the contrasts for the factors *quantifier* and *picture*, where we turn our attention only to the two corresponding contrasts (*some* vs. *all* for both *quantifier* and *picture* and *none* vs. *all* for both *quantifier* and *picture*). The interactions for these contrasts emerge in the analysis, though not equally fast: for *none* vs. *all* the interaction starts already at bin 5 but for *some* vs. *all* it does not appear before bin 11 – 300 ms after the *none* vs. *all* interaction. Both interactions persist up to the end of the interest period (see Table 7.Table **6**) Figure 18 displays a comparative view of the interactions averaged across the joint time window from bin 11 to bin 20.

The comparison between *some* and *all* reveals a very clear pattern of the picture *all* being the main distracter to the target in sentences with the quantifier *some* and *some* being the main distracter for the quantifier *all*. The pattern for *none* vs. *all* looks very similar if averaged across all relevant bins, bin 5 to 20, although fixations on the target images are somewhat less frequent and even below 50% for the quantifier *none* (11 vs. 47% and 32 vs. 15% for picture *some* and picture *all*, respectively, for the quantifiers *none* and *all*). In the case of *none* vs. *all*, the image *some* appears to be the most frequently fixated alternative to the target (20% for quantifier all and 29% for none.)

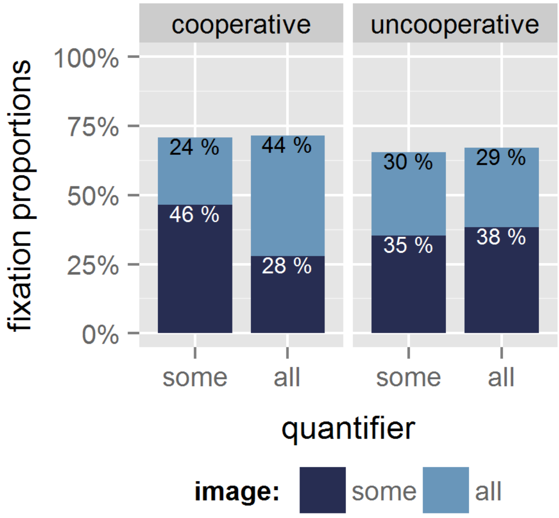


Figure 19. Mean fixation frequencies on the target images for *some* and *all*, separately for cooperative and non-cooperative partners and separately for the quantifiers *some* and *all*, averaged across bins 5-10.

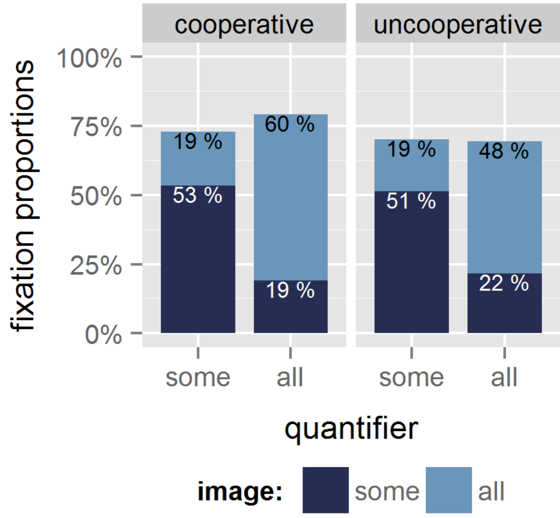


Figure 20. Mean fixation frequencies on the target images for some and all, separately for cooperative and non-cooperative partners and separately for the quantifiers some and all, averaged across bins 11- 20.

Subsequently, we take cooperativity into account and look at the four-way interactions of cooperation, quantifier, image and bin with the restriction in contrasts to those where quantifier and image types correspond to each other (see above.) There is a significant four-way interaction for the contrast some vs. all spread over six bins, 150 to 450 ms after quantifier onset. However, the interaction is only marginally significant in the first and the last of these bins. The fixation frequencies over the bins for this interaction are plotted in Figure 19 below. From Figure 19 one can also read the expected fixation pattern – there are more frequent fixations on the target image (the one corresponding to the quantifier) in cooperative than uncooperative cases. In this early 300 ms period the quantifier appears to play a role for fixations only if the partner was cooperative. Figure 20 shows how this pattern changes afterwards, during bins 11 to 20. Here, irrespective of whether the speaker was cooperative or not, target fixations on the picture corresponding to the quantifier are prevalent, as expected.

The same four-way interaction for the contrast *none* vs. *all* is significant in bins 12 to 16 with the exception of bin 13, where the α-error is around 10%. Unlike in the case of *some* vs. *all*, cooperation has a quantitative rather than a qualitative effect. More fixations on the image corresponding to the quantifier are obtained for cooperative as well as uncooperative partners. Therefore, the quantifier influences fixations in both cases, although to a different degree.

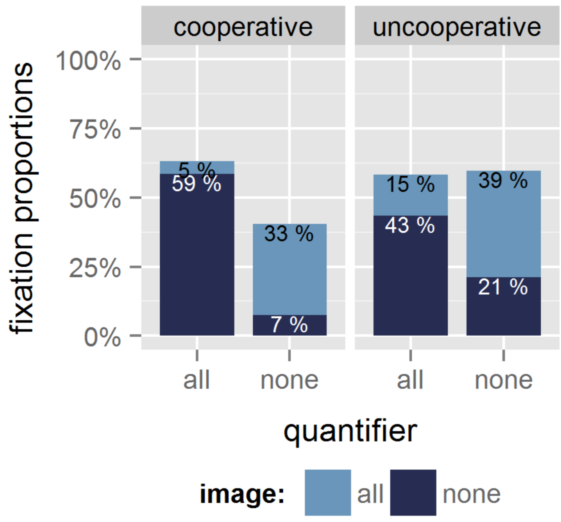


Figure 21. Mean fixation frequencies on the target images for *none* and *all*, separately for cooperative and non-cooperative partners and separately for the quantifiers *none* and *all*, averaged across bins 12-16.

The results from the four-way interaction explain the difference in time between the three-way interactions of *some* vs. *all* and *none* vs. *all*. *Some* vs. *all* was delayed compared to none vs. all because in the uncooperative condition the expected pattern for some vs. all was not visible before bin 11.

In order to find out whether there was a difference in fixation between some and all in two time periods in Figure 19 and Figure 20, the data was subjected to a generalized linear mixed model of fixation proportions with factors picture (some, all), quantifier (some, all), cooperation (yes, no), time period (bins 5-10, bins 11-20) and random effects subject id and item (see Table 7). The four-way interaction picture (some vs. all) × quantifier (some vs. all) × bins (5-10 vs. 11-20) × cooperative (yes vs. no) was highly significant.

Fixed effects	Estimate	Std. Error	z value
(Intercept)	-0.706	0.089	-7.971***
picture (some vs. all)	0.241	0.074	3.271**
quantifier (some vs. all)	0.002	0.074	0.028
cooperative (yes vs. no)	-0.030	0.074	-0.410
bins (5-10 vs. 11-20)	0.0270	0.044	0.612
picture × quantifier (both some vs. all)	-0.396	0.074	-5.379***
picture (some vs. all) × cooperative (yes vs. no)	0.020	0.074	0.277
quantifier (some vs. all) × cooperative (yes vs. no)	-0.026	0.074	-0.356
picture (some vs. all) × bins (5-10 vs. 11-20)	-0.118	0.044	-2.664**
quantifier (some vs. all) × bins (5-10 vs. 11-20)	-0.020	0.044	-0.450
cooperative (yes vs. no) × bins (5-10 vs. 11-20)	-0.015	0.044	-0.336
picture (some vs. all) × quantifier (some vs. all) × cooperative (yes vs. no)	-0.385	0.074	-5.221***
picture (some vs. all) × quantifier (some vs. all) × bins (5-10 vs. 11-20)	0.585	0.044	13.253***
picture (some vs. all) × bins (5-10 vs. 11-20) × cooperative (yes vs. no)	0.028	0.044	0.642
quantifier (some vs. all) × bins (5-10 vs. 11-20) × cooperative (yes vs. no)	0.027	0.044	0.619
picture (some vs. all) × quantifier (some vs. all) × bins (5-10 vs. 11-20) × cooperative (yes vs. no)	0.149	0.044	3.366***

Table 7. Generalized linear mixed model fit by the Laplace approximation for bins 5-10 vs. bins 11-20. Significance codes: 0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1.

6. Discussion

We set off by asking if and how hearers interpret expressions that carry a generalized conversational implicature in circumstances where cooperation cannot be automatically assumed. Are the implicatures be computed swiftly and effortlessly like the neo-Griceans would have it, or are the Relevance Theorists right in claiming that they take time and effort to retrieve?

A first point to note is that the experimental task was not too difficult for the participants and they made a clear distinction between the player types. In the following two sections I will focus on response and gaze data and their implications for the Defaultist-Contextualist debate.

6.1. Response Time and Response

Participants did not need more time to respond to *some* conditions than to *all* conditions. This result points to the Defaultist processing model because implicature processing and the completion of the experiment task take no longer than for an utterance without an implicature. Admittedly though, this is not very strong evidence, so let us turn to response values.

As predicted in section 5.2.3, the *some* quantifier in an uncooperative setting was paired with an *all* picture more frequently than its cooperative *some+all* counterpart. Contextualists would interpret this finding as indication that participants were so strongly influenced by the conversational context that they did not compute the implicature. Alternatively, Defaultists would argue that subjects simply computed and later cancelled the implicature. The lack of difference in response times could be attributed to the fact that the utterances were highly repetitive and computed before the end of the sentence, and thus before the answering period. This uncertainty can be resolved by looking at the subjects' gazes in the visual world directly after hearing the quantifier.

Before we turn to that, there were two other aspects of the response data that are interesting. The relatively high frequency with which the *all* picture was chosen in the cooperative condition was unexpected. It is possible that these responses are simply errors caused by the higher complexity of *some* in comparison to *all* and *none*. A high error rate could also be explained by overly cautious or sensitive participants who did not expect the cooperative player to produce such an inexact utterance. A more precise quantificational expression could have been preferred.

The experiment revealed that there were two main types of participants: (a) those that chose the picture that corresponded to the pragmatic meaning of *some* even in uncooperative situations and (b) those that chose the picture that corresponded to the semantic meaning of *some* in uncooperative context, but otherwise drew the implicature. There was a numerical tendency for the second group to make their response decision to uncooperative *some* faster than the first group, but the data was too sparse to allow for generalization. This difference must be left for future investigation (see Future Work.)

6.2. Fixations

The baseline analysis revealed that the pictures in experimental conditions were not fixated with equal frequency. The *some* image was the most complex one due to the number of colors (four). *None*, on the other hand, was potentially least relevant. The participants were informed what color they were expecting to hear a quantifying utterance about. *None* not only did not have the color in question, but also was paired with a negative quantifier, therefore less expected and potentially more difficult to deal with.

The image *all* distracted the subjects from the pragmatic target in utterances with the quantifier *some*. This finding corresponds to the response data. *Per contra*, in utterances quantified with *all*, the *some* picture was the most distracting one. These findings stress the relationship between the semantics of these two quantifiers. However, generalized over cooperation, the choice of image is made 600 ms after quantifier onset. This is later than predicted by Defaultists and therefore supports Contextualism.

Nevertheless, if we take cooperativity into consideration the picture changes a bit. In a cooperative context, participants fixate the picture corresponding to the pragmatic interpretation of *some* as early as 200 ms after quantifier onset. This effect emerges as quickly as for the *all* and *none* quantifiers. Therefore, the results for cooperative conditions replicate the findings of Grodner et al. (2010) and confirm the predictions made by Defaultists that scalar implicatures are computed automatically and at a very low cost.

One might argue, that there is a limited amount of quantifiers in the German language, and from that set only few start with an “e” and are preceded by *sind* (“are”), meaning that the participants could very accurately predict before the end of the word that an expression of type *einige* or *ein paar* (“a couple”) would ensue. However, all this does not offset the fact that the computations were made at an astonishingly fast pace.

In the uncooperative cases the quantifier has no effect on fixations until 600 ms after quantifier onset. During this time participants are cancelling the implicature and looking for an alternative to the target *some* picture. As predicted by Defaultists, cancellation takes time and effort, albeit still remarkably little. Within slightly more than half a second the subjects cancel the implicature and focus on the semantic interpretation or redraw the implicature.

Nevertheless, there is one limitation on this study that casts some doubt on my interpretation of gazes in the uncooperative cases. The task involved explicit verification, which “might encourage subjects to consider from the start interpretations that would not be considered otherwise." (Storto and Tanenhaus 2004 pp. 5) The fixations on distracters from the target during sentences with quantifier *none* may be caused by this exact need of searching for an alternative interpretation.

To sum up, the study found evidence in support of the Defaultist theory. Scalar implicatures triggered by *some* are computed in cooperative and uncooperative situations alike with astounding speed, although subsequently cancelled in the latter case. The experimental task may have had some influence on participants' behavior and their broadening the range of viable interpretation alternatives to normally inaccessible items. These questions must be left for further research.

7.Future work

Communication in uncooperative contexts is an exciting area for future work. The present study has but scratched the surface of the unknown. Future research could involve a refinement of the methodology and explore atypical language users like children, second language learners or patients with disabilities. At what point do children and language learners acquire the skills that allow them to discern cooperative from uncooperative scenarios? Are lesion patients, who have difficulties with the theory of mind and attributing mental states to themselves and others, able to make the distinction as well?

Another very interesting area for further research is how strongly pragmatic and semantic listener types differ. The second group was underrepresented in the present study so a fair comparison was not possible. However, on the basis of this experiment one could construct a test for determining listener type and study how the listeners differ in behavior in cooperative and uncooperative situations.

Conversations that go beyond the strictly regulated and artificial communication scenarios that are favored by many linguists are a broad area where many of the mechanics are yet to be understood. The present work is an important contribution to the further understanding of the dynamics of language outside the laboratory setting.

Bibliography

Baayen, R. H. (2011). languageR: Data sets and functions with ‘‘Analyzing Linguistic Data: A practical introduction to statistics’’. R package version 1.4.

Baayen, R. H., D. J. Davidson and D. M. Bates (2008). “Mixed-effects modeling with crossed random effects for subjects and items”. *Journal of Memory and Language* 59, pp. 390-412.

Bezuidenhout, A.L. and J. Cutting (2002). “Literal meaning, minimal propositions, and pragmatic processing.” *Journal of Pragmatics* 34(4), pp. 433-56.

Bezuidenhout, A.L. and R.K. Morris (2004). “Implicature, Relevance, and Default Pragmatic Inference.” In Noveck, I. and D. Sperber (eds.), *Experimental pragmatics*, Basingstoke, NY: Palgrave Press, pp. 257-82.

Bonnefon, J.-F., A. Feeney and G. Villejoubert (2009). “When some is actually all: scalar inferences in face-threatening contexts.” *Cognition* 112(2), pp.249-58.

Borg, E. (2009). “On Three Theories of Implicature: Default Theory, Relevance Theory and Minimalism.” In *International Review of Pragmatics* 1, 63-83.

Bott, L. and I. Noveck (2004). “Some utterances are underinformative: the onset and time course of scalar inferences.” *Journal of Memory and Language* 51(3), pp. 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(3), pp. 434-63.

Breheny, R., Ferguson, H. J., & Katsos, N. (2012). “Investigating the timecourse of accessing conversational implicatures during incremental sentence interpretation.” *Language and Cognitive Processes*, (ahead-of-print), 1-25.

— (2013). “Taking the epistemic step: Toward a model of on-line access to conversational implicatures.” *Cognition*, 126(3), 423-440.

Carston, R. (1998). “Informativeness, Relevance and Scalar Implicature.” In: R. Carston and S. Uchida (eds.): *Relevance Theory: Applications and Implications*. Amsterdam: John Benjamins, 179-236.

— (2009). “The explicit/implicit distinction in pragmatics and the limits of explicit communication.” *International Review of Pragmatics*, 1(1), 35-62.

— (2010). “Lexical pragmatics, ad hoc concepts and metaphor: from a relevance theory perspective.” *Italian Journal of Linguistics*, 22(1), 153-180.

— (2012). “Relevance theory.” *Routledge companion to the philosophy of language*, 163-176.

Chierchia, G. (2004). “Scalar Implicatures, Polarity Phenomena, and the Syntax/Pragmatics Interface.” In A. Belletti (ed.), *Structures and Beyond*, 39-103. Oxford: OUP.

Chierchia, G., D. Fox and B. Spector (2008). “The Grammatical View of Scalar Implicatures and the Relationship between Semantics and Pragmatics.” To appear in P. Portner, C. Maienborn and K. von Heusinger (eds.), *Handbook of Semantics*, New York, NY: Mouton de Gruyter.

Degen, J. (2007). “Processing Scalar Implicatures: What Role Does the Question of Default Play for the Debate Between (Neo-)Griceanism and Relevance Theory?” BS in Cognitive Science, Institute of Cognitive Science, University of Osnabrück.

— (2009). *Processing Scalar Implicatures: An Eye-Tracking Study*. MS Thesis in Cognitive Science, University of Rochester. <http://www.bcs.rochester.edu/people/jdegen/docs/Degen2009_MscThesis.pdf> last accessed 31.01.2010

Degen, J. et al. (2009). *Using a novel experimental paradigm to investigate the processing of scalar implicatures*. (Poster presented at the CUNY 2008 Conference on Human Sentence Processing, UC Davis)

Degen, J. and M. Tanenhaus (2009). *Naturalness of lexical alternatives influences interpretation of ‘some’*. (Poster presented at the CUNY 2009 Conference on Human Sentence Processing, UC Davis)

Dieussaerta, K, S. Verkerka, E. Gillarda and W. Schaekena (2011). "Some effort for some: Further evidence that scalar implicatures are effortful." *The Quarterly Journal of Experimental Psychology* 64(12), pp. 2352-2367.

de Jager, T. and R. van Rooij (2007). “Explaining Quantity Implicatures”. In: *Proceedings of the 11th conference on Theoretical Aspects of Rationality and Knowledge*, New York: ACM, pp. 193–202.

De Neys, W. and W. Schaeken (2007). "When People Are More Logical Under Cognitive Load. Dual Task Impact on Scalar Implicature." *Experimental Psychology* 54(2), Hogrefe & Huber Publishers, pp 128-133.

Duñabeitia, J. A. Afonso, A. Avilés, O. C. Scheepers and M. Carreiras (2009). “Qualitative differences in the representation of abstract versus concrete words: Evidence from the visual-world paradigm”. *Cognition* 110 (2), pp 284-292.

Ferguson, H. J., C. Scheepers A. J. and Sanford (2010). “Expectations in counterfactual and theory of mind reasoning.” *Language and Cognitive Processes* 25, pp. 297-346.

Foppolo, F. (2006). “The *cost* of *embedded* Scalar Implicatures.” *SURplus*, Sistema Universitario Ricerca <http://boa.unimib.it/bitstream/10281/9938/3/The_cost_embedded_Scalar_Implicatures.pdf>, last accessed 31.01.2010.

Franke, M. (2008). “What is Game Theoretic Pragmatics?” Draft, <http://www2.sfs.nphil.uni-tuebingen.de/jaeger/lehre/ss09/formalePragmatik/Franke08.pdf>, last accessed 31.01.2010.

Feeney, A., S. Scrafton, A. Duckworth and J. Simon (2004). "The story of some: Everyday pragmatic inference by children and adults." *Canadian Journal of Experimental Psychology/Revue canadienne de psychologie expérimentale*, 58(2), pp. 121-132.

Freenor, M. (2009). “Scalar Implicature in Discourse Representation Theory.” MS Thesis in Logic and Computation, Carnegie Mellon University Pittsburgh, PA. <http://semanticsarchive.net/Archive/Dc5MWY2M/freenor-ms-thesis.pdf> last accessed 31.01.2010.

Garrett, M., & Harnish, R. M. (2007). Experimental pragmatics: Testing for implicitures. Pragmatics & Cognition, 15(1), 65-90.

— (2008). “Skating along the syntactic verge: Experimental pragmatics and understood elements of content.” *Time and Again: Theoretical and Experimental Perspectives on Formal Linguistics*. Amsterdam: Benjamins.

Geurts, B. (2009a). "Scalar Implicature and Local Pragmatics." *Mind & Language* 24(1), pp. 51–79.

— (2009b). *Quantity implicatures*, to appear with CUP.

Geurts, B. and N. Pouscoulous (2009). “Embedded implicatures?!?” *Semantics & Pragmatics* 2(4), pp. 1–34.

Grice, P. (1989, reprint 1991). *Studies in the Way of Words*, Cambridge, MA: HUP.

Grodner, D.J. et al. (2010). ““Some,” and possibly all, scalar inferences are not delayed: Evidence for immediate pragmatic enrichment”. *Cognition* 116(1), pp. 42-65.

Heller, D., D. Grodner and M. K. Tanenhaus (2008). “The role of perspective in identifying domains of reference”. *Cognition* 108 (3), pp. 831-836.

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, Washington: Georgetown University Press, pp. 11-42.

— (2004). "Implicature." In: L. Horn, G. Ward (eds.), *Handbook of pragmatics . Blackwell*, Oxford, 3–28

— (2006). “The Border Wars: A Neo-Gricean Perspective.” In K. von Heusinger and K. Turner (eds.), *Where Semantics Meets Pragmatics*, Oxford: Elsevier, pp. 21-48.

— (2009). “WJ-40: Implicature, truth, and meaning.” *International review of pragmatics*, 1(1), 3-34.

Horton, W.S. and B. Keysar (1996). “When do speakers take into account common ground?” *Cognition* 59/1, pp. 91-117.

Huang, Y.T and J. Snedeker (2009). “Online interpretation of scalar quantifiers: Insight into the semantics–pragmatics interface.” *Cognitive Psychology* 58(3), pp. 376-415.

Jaszczolt, K.M. (2009). “Default “semantics.”” In Heine, B. and H. Narrog (eds.) *The Oxford Handbook of Linguistic Analysis*, Oxford: OUP, pp. 2572-74.

Jäger, G. (2008). “Applications of Game Theory in Linguistics.” *Language and Linguistics Compass* 2(3), pp. 406–21.

Katsos, N. (2007). “The semantics/pragmatics interface from an experimental perspective: the case of scalar implicature.” *Synthese* 165(3), pp. 385-401.

Keller, F., S. Gunasekharan, N. Mayo, and M. Corley (2009). “Timing Accuracy of Web Experiments: A Case Study Using the WebExp Software Package.” *Behavior Research Methods* 41(1), pp. 1-12.

Keysar, B., Barr, D. J., Balin, J. A., & Brauner, J. S. (2000). “Taking perspective in conversation: The role of mutual knowledge in comprehension.” *Psychological Science*, 11(1), 32-38.

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

Mirman, D., J.A. Dixon and J.S. Magnuson (2009). “Statistical and computational models of the visual world paradigm: Growth curves and individual differences.” *Journal of Memory and Language* 59(4), pp. 475-94.

Moll H. and M. Tomasello (2007). “Cooperation and human cognition: the Vygotskian intelligence hypothesis.” *Philosophical transactions of the Royal Society of London* 362(1480), pp. 249-58.

Pickering, M.J. and S. Garrod (2004). “Toward a mechanistic psychology of dialogue.” *The Behavioral and Brain Sciences* 27(2), pp. 169-90.

Rayner, K. and A. Pollatsek (1989). *The Psychology of Reading*, Englewood Cliffs, NJ: Prentice-Hall.

Reboul, A (2004). “Conversational implicatures: nonce or generalized?” In Noveck, I. and D. Sperber (eds.), *Experimental pragmatics*, Basingstoke, NY: Palgrave Press, pp. 322-33.

Russell, B. (2006). “Against Grammatical Computation of Scalar Implicatures”. *Journal of Semantics* 23(4), pp. 361-82.

Scheepers, C., F. Keller and M. Lapata (2008). “Evidence for serial coercion: A time course analysis using the visual-world paradigm*.” Cognitive Psychology* 56, pp. 1-29.

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

Tanenhaus, M.K. and S. Brown-Schmidt (2008). "Language processing in the natural world." *Philosophical transactions of the Royal Society of London* 363(1493), pp. 1105-22.

Wilson, D. and D. Sperber (2005). “Relevance Theory.” In G. Ward and L. Horn (eds.) *Handbook of Pragmatics*. Oxford: Blackwell, pp. 207-319.

Wolter, L., K.S. Gorman, and M.K. Tanenhaus (2011). “Scalar reference, contrast and discourse: Separating effects of linguistic discourse from availability of the referent.” *Journal of Memory and Language*, 65(3), pp. 299-317.

van Rooij, R. (2008). “Games and Quantity implicatures”. *Journal of Economic Methodology* 15(3), pp. 261-74.

Verbrugge, R. and L. Mol (2008). “Learning to Apply Theory of Mind.” *Journal of Logic, Language and Information* 17(4), pp. 489-511.

Appendix A. Sentence Objects

Item Sentences	English Translation
In meinem Sack sind einige Steinchen rot.	*In my bag are some red pebbles.*
In meinem Sack sind einige Steinchen grün.	*…green pebbles.*
In meinem Sack sind einige Steinchen blau.	*…blue pebbles.*
In meinem Sack sind einige Steinchen gelb.	*…yellow pebbles.*
In meinem Sack sind alle Steinchen rot.	*In my bag are all red pebbles.*
In meinem Sack sind alle Steinchen grün.	*…green pebbles.*
In meinem Sack sind alle Steinchen blau.	*…blue pebbles.*
In meinem Sack sind alle Steinchen gelb.	*…yellow pebbles.*
In meinem Sack sind keine Steinchen rot.	*In my bag are no red pebbles.*
In meinem Sack sind keine Steinchen grün.	*…green pebbles.*
In meinem Sack sind keine Steinchen blau.	*…blue pebbles.*
In meinem Sack sind keine Steinchen gelb.	*…yellow pebbles.*

Appendix B. Visual Stimuli

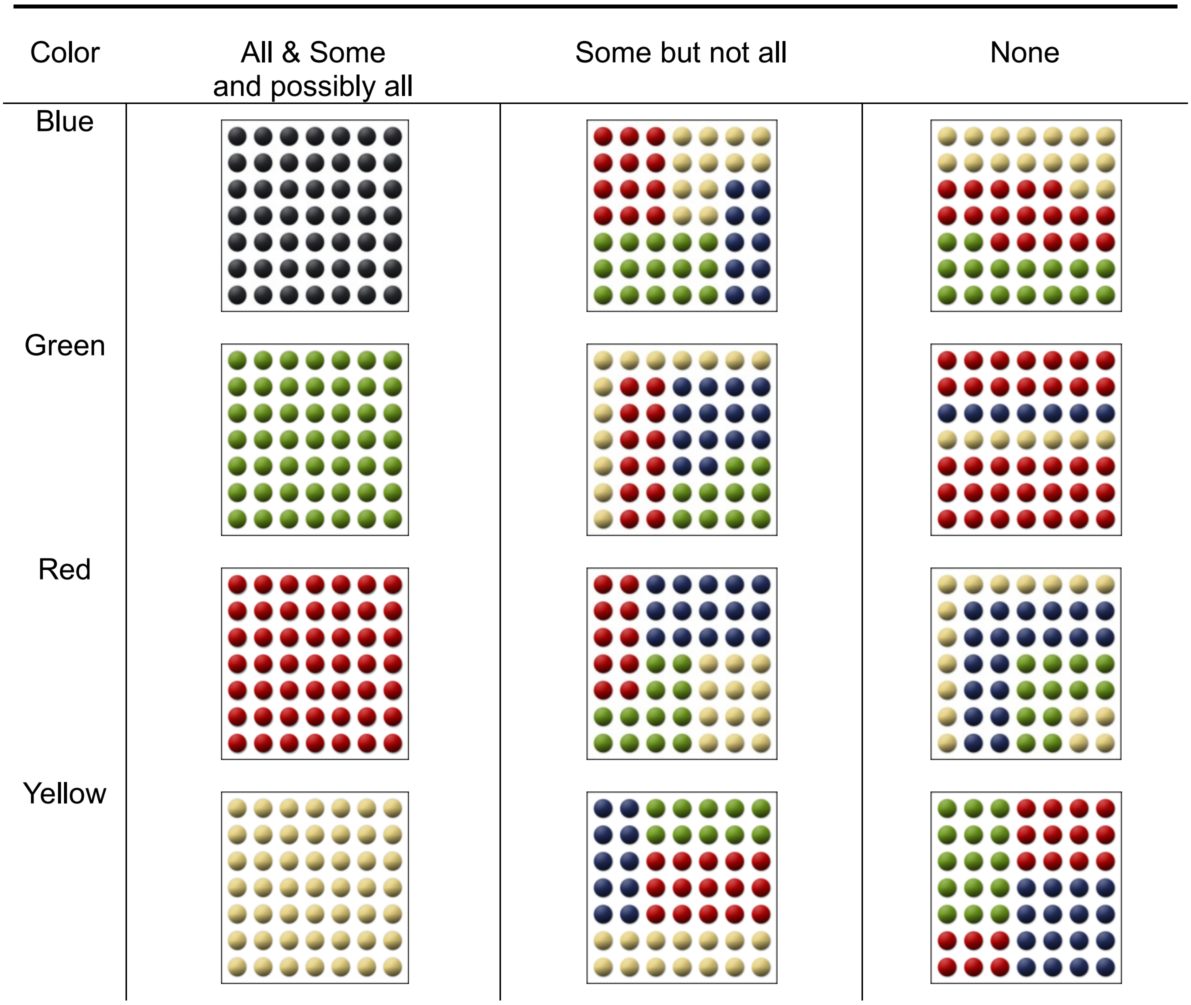


Figure 22. Picture material in the eye-tracking experiment.

Appendix C. Game Instructions

Steinchen

Ziel des Spieles

Ziel des Spieles ist es, in jeder Runde möglichst viele Steinchen in einer Farbe zu sammeln. Die Karten stellen Säcke mit Steinchen dar. Der Gewinner bekommt Punkte für jedes Steinchen in der Farbe, die er sammelt. Der Gewinn wird nach jeder Runde ausgezahlt und die Punkte auf null gestellt.

Spielvorbereitung

Es werden zwei Teams, die gegeneinander spielen, gebildet. Die Spieler eines Teams zählen die Punkte zusammen und teilen sich den Gewinn. Da in diesem Spiel die Teams eine ungleiche Spielerzahl haben, wird die Punktzahl des größeren Teams durch zwei geteilt. Jeder Spieler bekommt eine Karte mit der Farbe, die er im gesamten Spiel sammeln wird. Die anderen Karten liegen in der Mitte auf einem Stapel.

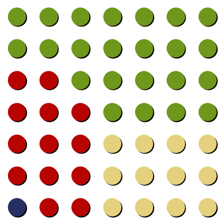
So wird gespielt

Jeder Spieler bekommt eine Karte und schaut sich diese kurz an. Ist ein Spieler an der Reihe, so informiert er seine Mitspieler über die Farbe Steinchen, die er sammelt.

BEISPIEL: „Ich sammle grüne Steinchen."

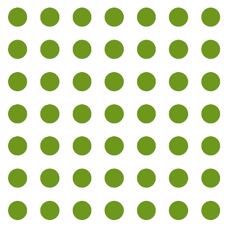
Die anderen Spieler antworten die Reihe nach, indem sie eine Aussage über ihren Steinchensack machen. Die Spieler des gegnerischen Teams antworten zuerst. Die Aussagen müssen wahrheitsgemäß und über die gefragte Farbe sein.

BEISPIEL: „In meinem Sack ist die Hälfte der Steinchen grün."



Wenn ein Spieler des gegnerischen Teams eine Aussage über seinen Steinchensack macht, muss man auch wahrheitsgemäß Antworten, darf aber aus taktischen Gründen täuschen. Nicht zugelassen sind Aussagen wie: "In meinem Sack sind Steinchen grün."

BEISPIEL: „In meinem Sack ist die Hälfte der Steinchen grün."



Nachdem alle Antworten gegeben sind, entscheidet sich der fragende Spieler welchen Steinchensack er ziehen wird. Der Spieler, von dem gezogen wurde, nimmt sich eine neue Karte aus dem Stapel.

Nachdem jeder Spieler einmal gezogen hat, ist die Runde vorbei.

Appendix D. Spectrograms of Sentences

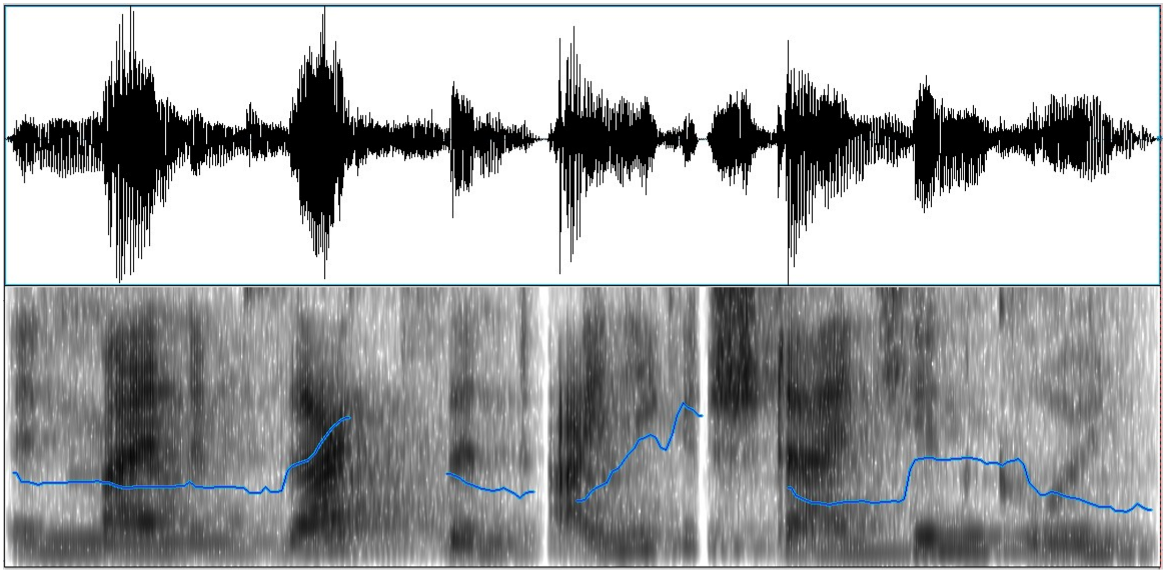


Figure 23. Female speaker *some*-utterance. Pitch marked in blue.

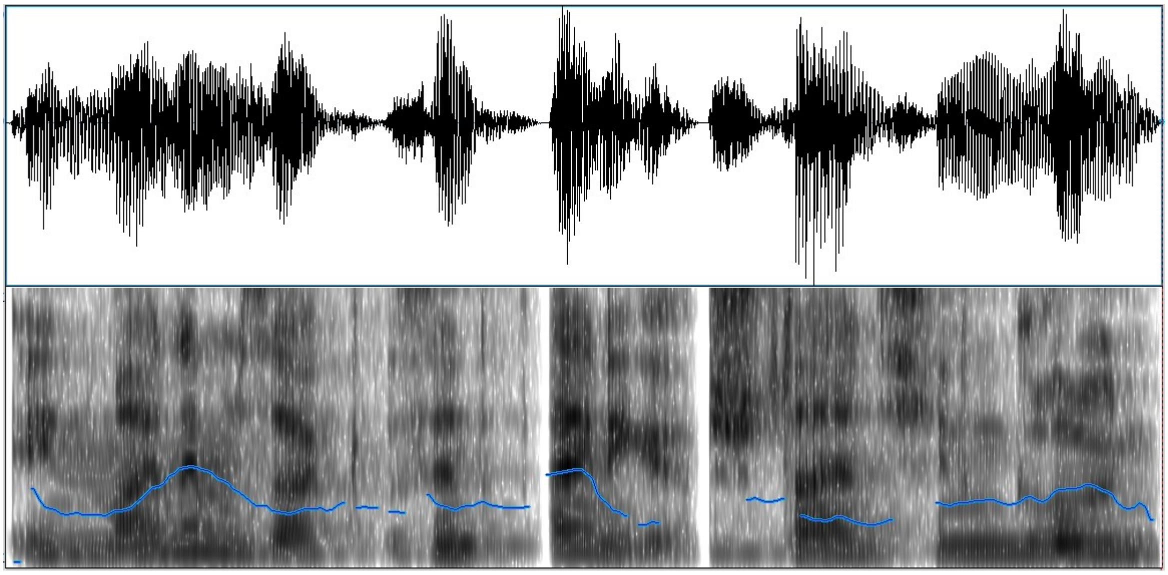


Figure 24. Male speaker *some*-utterance. Pitch marked in blue.

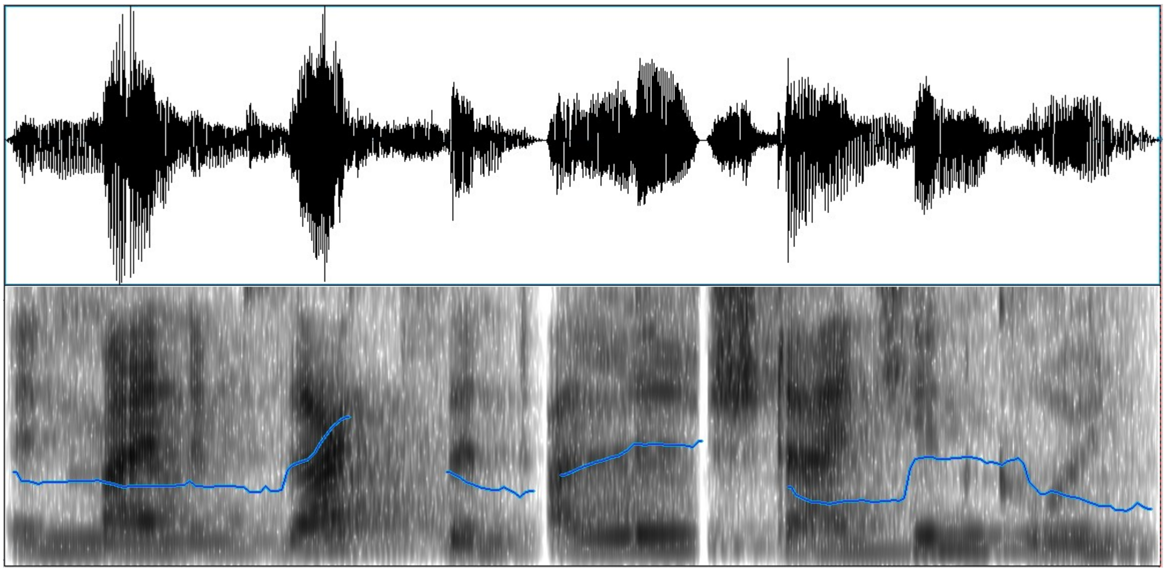


Figure 25. Female speaker *all*-utterance. Pitch marked in blue.

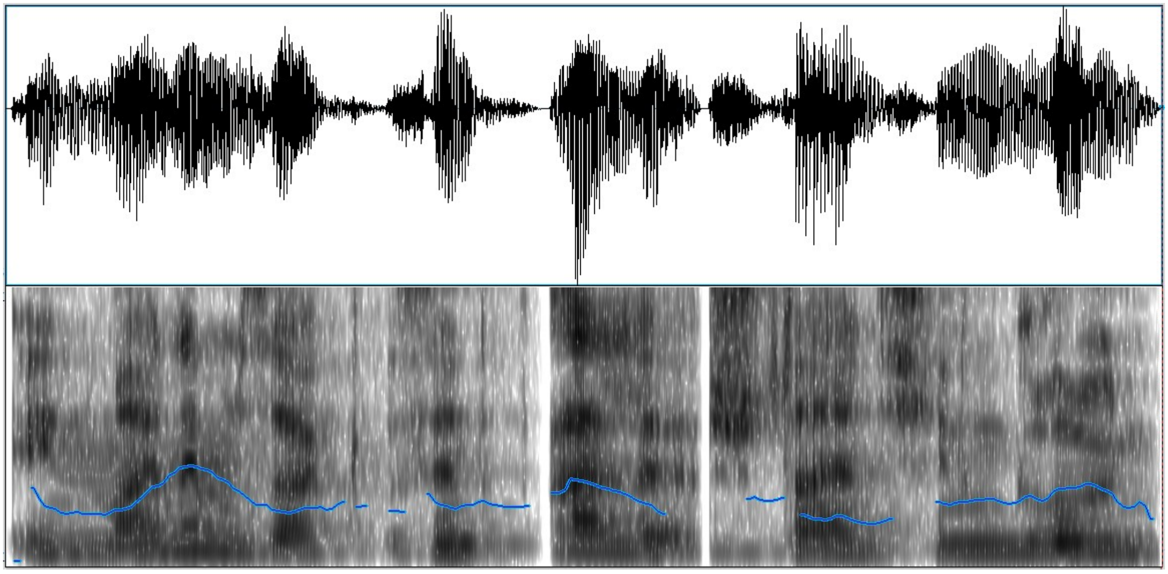


Figure 26. Male speaker *all*-utterance. Pitch marked in blue.

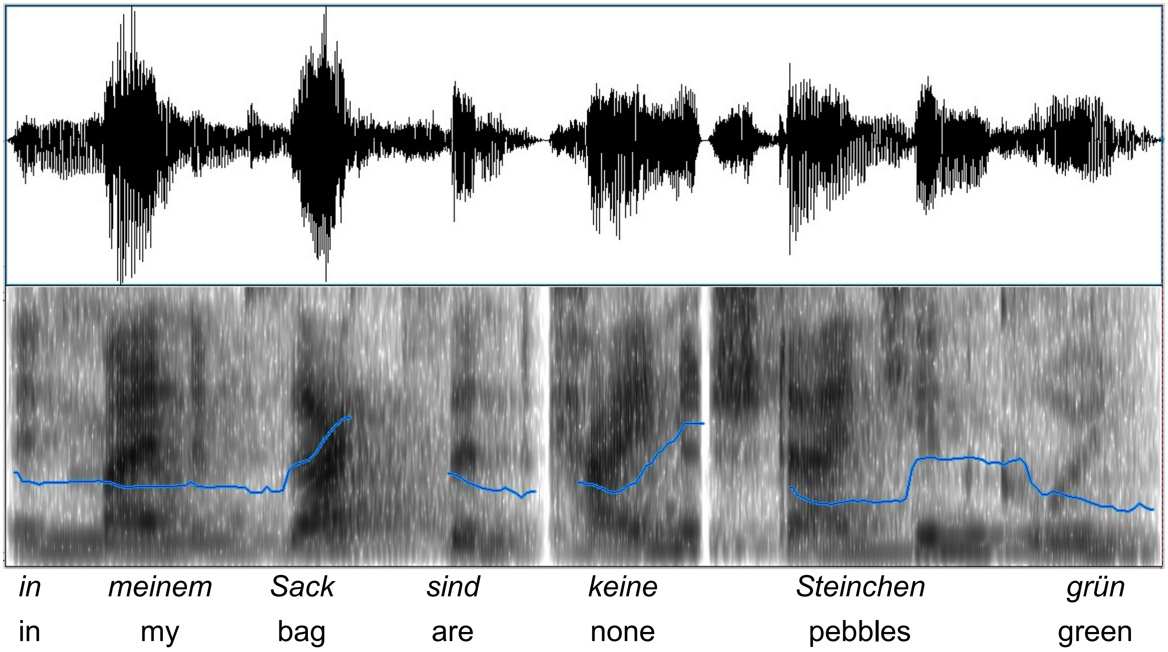


Figure 27. Female speaker *none*-utterance. Pitch marked in blue.

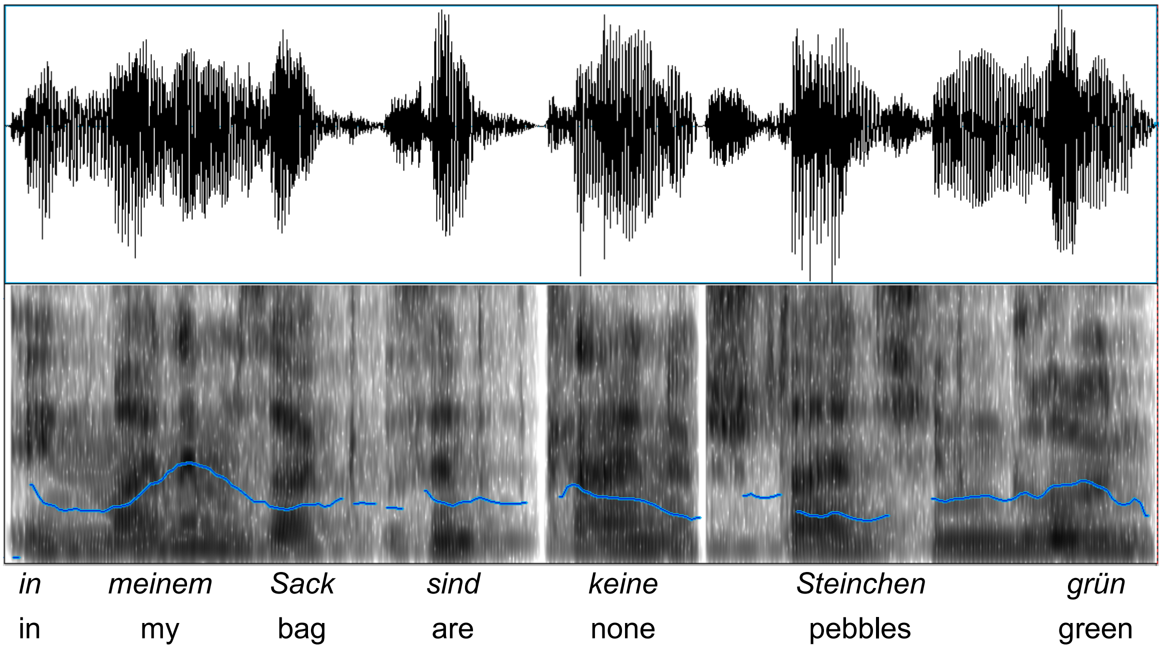


Figure 28. Male speaker *none*-utterance. Pitch marked in blue.