Pragmatic reasoning about silent only

Michael Franke & Leon Bergen

Abstract

We present several superficially similar, but conceptually distinct probabilistic models of use and interpretation of sentences with nested logical operators. All models are inspired by Gricean ideas of rational communication, but differ in the way and extent to which they take into consideration potential pragmatic enrichments generated by a silent-only operator, following recent grammatical approaches to the computation of scalar implicatures. Based on data from a novel experiment which combines a lexical-choice and an interpretation task, we use Bayesian model comparison to evaluate the predictive adequacy of these models. The data provide strong evidence for the idea that the full range of potential readings made available by grammatical approaches might be needed, and that classical Gricean reasoning may help manage the manifold ambiguity introduced by grammatical approaches to scalar implicatures.

1 Introduction

Listeners are astonishingly efficient in retrieving a speaker’s intended meaning in context, even when an utterance is, in principle, compatible with a plethora of potential conventional meanings. Part of this fast and efficient disambiguation process can be rationalized, following the seminal work of Grice (1989), as a systematic pattern of a listener’s reasoning about what speakers would likely say (in the given context) to convey this meaning or that (e.g. Levinson, 2000; Sperber and Wilson, 1995). Gricean explanations of pragmatic inferences have proven their explanatory value over decades, but were also subject to controversy from the start (e.g. Cohen, 1971). Central to the debate are putative pragmatic inferences, such as scalar implicatures, that occur in the scope of other logical operators. This paper tries to contribute to this foundational discussion.

Scalar implicatures are a special kind of Quantity implicature, so-called because they can be derived from Grice’s Maxim of Quantity, which requires speakers to provide, roughly put, all the relevant information they are able to give (Geurts, 2010). Traditional Gricean accounts rationalize scalar implicatures as the result of reasoning about alternative utterances with stronger logical meanings, so-called scalar alternatives, which the speaker could have but did not supply (Horn, 1972). For example, speakers who utter (1a) may be taken to communicate (1b), because they did not use the stronger alternative in (1c).

(1) a. I own some of Johnny Cash’s albums.
   b. I own some but not all of Johnny Cash’s albums.
   c. I own all of Johnny Cash’s albums.

Gricean explanations of pragmatic enrichments along these lines have met repeated criticism based on the observation that putative scalar implicatures do seem to appear in the scope of other logical operators (Geurts, 2010; Sauerland, 2012). Most recently, an alternative approach, referred to here as
grammaticalism, has been productively applied to a variety of theoretical puzzles and experimental observations (e.g. Chierchia, Fox, and Spector, 2012; Fox, 2007; Fox and Spector, 2018). Grammaticalisms stipulates silent grammatical operators, whose effect is to generate a rich ambiguity of potential pragmatic readings already during the compositional derivation of a sentences’ meaning. For instance, sentence (2a) could receive the reading (2b), where the occurrence of some is enriched to “some but not all” in the scope of the quantifier all by a silent grammatical operator O with a meaning contribution similar to that of only, as schematized in (2c). According to grammaticalism, this is a possible semantic reading of this sentence. Seen in this way, grammaticalism earns its name by trying to incorporate meaning aspects that may appear to be pragmatic in nature into the grammar.

(2) a. All of the aliens drank some of their water.

b. All of the aliens drank some but not all of their water.

c. All of the aliens drank $O(some)$ of their water.

where $O(some) = "some but not all"

There is controversy about which approach, Gricean or grammatical, is superior. Some experimental results suggest that alleged local enrichments as in (2b), which may be hard for traditional Griceanism to explain, do not matter for the explanation of participants’ responses (e.g. Geurts and Pouscoulous, 2009; Geurts and van Tiel, 2013). Other studies suggest that they do (e.g. Chemla and Spector, 2011; Franke, Schlotterbeck, and Augurzky, 2017; Potts et al., 2016). On the other hand, the grammatical approach may be very successful at generating potential readings, but it is, to a large extent, an open issue how, where and when silent-only operators should occur to generate the contextually appropriate reading of an utterance (for some discussion see, e.g., Chierchia, Fox, and Spector, 2012; Fox and Spector, 2018). Seen in this way, it would be desirable to explore in how far Gricean ideas about efficient communication in the light of potential ambiguity and a grammatical approach that generates potential pragmatic readings could be combined to solve each other’s problems.

A step in this direction is taken by a recent extension of the Rational Speech Act model of Frank and Goodman (2012). The Lexical Uncertainty model of Bergen, Levy, and Goodman (2016) includes a listener’s reasoning about the speaker’s lexical entry for potentially underspecified forms. Potts et al. (2016) apply this model to experimental data concerning the interpretation of sentences with nested quantifiers like in (2a). Here, we compare predictions of several probabilistic models, including the Lexical Uncertainty model, based on data from comprehension and production. Concretely, we use Bayesian model comparison (e.g. Jeffreys, 1961; Kass and Raftery, 1995) to evaluate four kinds of models: (i) the vanilla Rational Speech Act model, (ii) the Lexical Uncertainty model, (iii) a variant of (ii) in which speakers are not confined to a single lexical meaning for all occurrences of the same word, (iv) an extension of (iii) which includes all parses generated by a grammatical approach.

A first main contribution of this paper is the introduction of these latter two models. We argue here that the full inclusion of grammatically-supplied pragmatic readings into a Grice-inspired model of pragmatic reasoning requires a slight, but crucial conceptual divergence from the Lexical Uncertainty model. A second main contribution of this paper is to demonstrate the general usefulness of data-driven statistical comparison of models which are not off-the-shelf statistical models (e.g., regression models), but maximally concrete and precise formalizations of extant linguistic theory concerning what could govern lexical choices and interpretations.

Section 2 illustrates the range of potential pragmatic enrichments predicted by grammatical approaches. Section 3 introduces the relevant probabilistic models for reasoning about potential pragmatic enrichments. Section 4 reports on a combined production and comprehension experiment. Sec-
2 Nested Aristotelians and their pragmatic enrichments

The goal of this section is to illustrate differences in the predictions of a Gricean and a grammatical approach to scalar enrichments and to introduce the materials for the experiment reported in Section 4. We consider nested Aristotelians: sentences in which the Aristotelian quantifiers *none*, *some* and *all*—arguably the most basic quantificational operators—appear once in outer (higher scope) position and once in inner (lower scope) position. There are nine nested Aristotelians, exemplified in (3).

\[(3) \{ \text{None} \mid \text{Some} \mid \text{All} \} \text{ of the aliens drank } \{ \text{none} \mid \text{some} \mid \text{all} \} \text{ of their water.} \]

The following will use abbreviations for these sentences, indicating the outer and inner quantifier in order, e.g., as on the right-hand side of Table 1. For example, “AS” is short for *All of the aliens drank some of their water.*

Arguably all nested Aristotelians presuppose the existence of a plurality of relevant aliens. What matters to the truth of nested Aristotelians is whether there are (i) aliens which drank none of their water, (ii) aliens which drank some but not all, and (iii) aliens which drank all. We therefore distinguish seven kinds of world states, each of which Figure 1 provides one example of. Each situation in Figure 1 shows twelve aliens, each with its mug of water. To have a compact representation of relevant world states, we use pictures like \[
\begin{array}{c}
\text{\includegraphics[width=0.1\textwidth]{full_mug.png}}
\end{array}
\]

These pictures indicate which types of aliens exist. If there are aliens which drank none of their water, the pictorial representation for this world state contains a full mug, as in \[
\begin{array}{c}
\text{\includegraphics[width=0.1\textwidth]{full_mug.png}}
\end{array}
\]. otherwise, if there are no aliens which drank none of their water (every alien drank at least some), then the corresponding picture shows no full mug, as in \[
\begin{array}{c}
\text{\includegraphics[width=0.1\textwidth]{empty_mug.png}}
\end{array}
\]. The state \[
\begin{array}{c}
\text{\includegraphics[width=0.1\textwidth]{partial_mug.png}}
\end{array}
\], for example, refers to any possible world in which all of the aliens drank some but not all of their water.

Grammatical approaches to the computation of scalar implicatures assume that candidate pragmatic readings of sentences are generated by a silent operator \(O\) whose meaning contribution is similar to *only* (e.g. Chierchia, Fox, and Spector, 2012). For present purposes, there are three relevant places...
where an $O$-operator might occur in a nested Aristotelian. As shown in (4), it can occur in matrix position thereby applying to the whole sentence and it can apply to the outer or inner quantifier.

(4) $O_M [O_O(Q_O) \text{ of the aliens drank } O_I(Q_I) \text{ of their water}]

There are eight parses we need to consider. A parse determines whether $O$ occurs in matrix ($M$), in outer quantifier ($O$) and in inner quantifier ($I$) position. In the following, we use a notation where, for example, a parse of a sentence with $O$ occurring only in matrix position is denoted as $M__$, a parse with $O$ occurring at all three relevant positions is written as $MOI$ and a parse without any $O$ operator is represented as $___$.

A parse determines the reading of a sentence. If $S$ is a sentence and $p$ its parse, then $[S]^p$ is the reading of $S$ under $p$. A given sentence can have multiple readings based on different parses. Table 1 lists the readings of nested Aristotelians for the different relevant parses, which we consider in this paper. These readings are derived from what is perhaps the simplest instantiation of grammaticalism. The remainder of this section explains in detail how these readings are derived exactly. Readers less interested in these technical details can safely skip forward Section 3, which introduces several conceptually different models of pragmatic reasoning about the use and interpretation of such multiply ambiguous sentences.

**Readings from a vanilla grammatical approach.** When $O$ applies directly to a quantifier $Q$, the resulting meaning $O(Q)$ is that of $Q$ conjoined with the negation of all strictly stronger lexical alternatives. As usual, we assume that some and all are lexical alternatives. We also assume, following standard practice, that none has a lexical alternative in not all, since we analyze none as not some (e.g. Levinson, 2000, p. 80). Consequently, the effect of applying $O$ to $Q$ will be vacuous for none and all, but an application to some gives an *in situ* enrichment to “some but not all.” For example, under a parse $__I$ the sentence “SS” gets the reading $[S]^{__I}$ in (5) that will be true in any state whose pictorial representation includes a half-full mug (see Table 1).

(5) Some of the aliens drank some but not all of their water.

When $O$ applies in matrix position to a sentence $S$ it takes the meaning of $S$ obtained from insertions of $O$ below matrix position, and conjoins that meaning with the negation of suitable sentential alternatives. Let $S$ be a sentence and $p$ its parse. Let $[S]^p_{-?}$ be the meaning of $S$ under the parse $p_{-?}$, which is like $p$ with respect to outer and inner quantifier, but has no $O$ in matrix position. When $O$ applies in matrix position to $S$ whose parse is $p$, the meaning of $O(S, p)$ is obtained by conjoining $[S]^p_{-?}$ of $S$ with the negation of all relevant sentential alternatives $Alt(S, p)$ of $S$, if this operation is non-contradictory:

\[
[O(S, p)] = \begin{cases} [S]^{p_{-?}} \cap_{S' \in Alt(S, p)} [\text{not } S'] & \text{if non-contradictory} \\
[S]^{p_{-?}} & \text{otherwise.}
\end{cases}
\]

We will here assume that the set $Alt(S, p)$ is obtained in two steps, a generation and a filtering step. First, we generate a set $Alt^*(S)$ of potential alternatives by blindly replacing any occurrence of a scalar

---

1The only effect of considering not all a lexical alternative to none is that the sentence “NN” gets an additional reading when $O$ applies in matrix position. But, as the ambitious reader will be able to ascertain later on, this has no noteworthy effect on anything of relevance to our main concerns.
item in \( S \) with any of its lexical alternatives, as defined above. A second step then filters out sentences from \( Alt^*(S) \) which do not entail \( S \).

\[
Alt^*(S) = \{ S' \mid S' \text{ is derived from } S \text{ by replacements of lexical alternatives} \}
\]

\[
Alt(S, p) = \{ S' \in Alt^*(S, p) \mid \llbracket S' \rrbracket \subset \llbracket S \rrbracket \}
\]

For example, consider the sentence “SS” under a parse \(_M_.\). By replacing occurrences of \( \text{some} \) with its lexical alternatives, we obtain \( Alt^*(SS) = \{ SS, AS, SA, AA \} \). Of these “SS” is filtered out. Applying matrix-\( O \) consequently yields the conjunction of the negation of all sentences “AS”, “SA” and “AA” with the literal meaning of “SS” as paragraphed is (6). This is a very strong reading, as the only world state that makes this reading true is

\[
(6) \text{ Some of the aliens drank some of their water, and it is not true that some aliens drank all or that all drank some.}
\]

As a final example, consider again the sentence “SS” but now under a parse \(_M_I_.\). We first consider \( \llbracket SS \rrbracket _I^- \), which gives the set of situations in which there is at least one half-full mug. The set \( Alt^*(SS) \) is as before, but \( Alt(S, p) \) is now empty because none of the sentences in \( Alt^*(SS) \) are, under parse \(_I_.\), strictly stronger than “SS” under parse \(_I_.\). Consequently, additional insertion of \( O \) in matrix position is vacuous in this case. – This is a relative simple instantiation of grammaticalism. Alternatives exist. We will come back to this in Section 6.

3 Models

The instantiation of grammaticalism spelled out in the previous section yields a total of 21 distinct sentence-meaning pairs for the nested Aristotelians, as listed in Table 1. In other words, the potential insertion of silent \( O \)-operators creates massive semantic ambiguity. The main question to be addressed in this section is conceptual, namely how to integrate, in a sound and coherent manner, such multiply and systematically ambiguous sentences into a Grice-inspired model of pragmatic reasoning.

We introduce five models here which differ in the way they use the set of readings in Table 1, or a subset thereof. The starting point is the vanilla Rational Speech Act (RSA) model of Frank and Goodman (2012). It is introduced in Section 3.1 and will serve as a reference point because it associates each sentence with its logical semantics only. We may think of it as only including readings generated from the parse \(_I_.\), i.e., it takes none of the potential ambiguity created by insertion of \( O \)-operators into account. On the other end of the spectrum, there is the so-called Silent Only model (Section 3.4), which considers all the readings listed in Table 1, thereby taking a massive ambiguity of sentences into account during pragmatic reasoning. In between these two extremes are two variants of the RSA model that can be considered to take increasingly large subsets of the readings listed in Table 1 into account (Sections 3.3 & 3.4). While this section focuses on the conceptual question of how these models should plausibly defined so as to make conceptual sense, the resulting empirical question, addressed subsequently in Section 5, is which of these different approaches to reasoning with ambiguous sentences is supported by experimental data.

3.1 The vanilla Rational Speech Act model

The RSA model defines a speaker production rule and a listener comprehension rule, roughly as described by a classical Gricean approach (see Franke and Jäger, 2016; Goodman and Frank, 2016, for
Table 1: Readings produced by a simple instantiation of grammaticalism for different parses applied to the nested Aristotelians.

<table>
<thead>
<tr>
<th>sentence</th>
<th>parses</th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>NN</td>
<td>__I, _O, _OI</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td></td>
<td>M_I, _I, MO, MOI</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>1</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>NS</td>
<td>__I, _O, M_I, MO</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td></td>
<td>__I, _OI</td>
<td>1</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>1</td>
</tr>
<tr>
<td></td>
<td>M_I, MOI</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>1</td>
</tr>
<tr>
<td>NA</td>
<td>__I, _I, _O, _OI</td>
<td>1</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td></td>
<td>M_I, M_I, MO, MOI</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>SN</td>
<td>__I</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td></td>
<td>all others</td>
<td>0</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>SS</td>
<td>__I</td>
<td>0</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td></td>
<td>_I, M_I</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td></td>
<td>_O, MO</td>
<td>0</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td></td>
<td>_OI, MOI</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td></td>
<td>M_2</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>SA</td>
<td>__I</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>1</td>
<td>0</td>
<td>1</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td></td>
<td>all others</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>1</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>AN</td>
<td>all parses</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>AS</td>
<td>__I, _O</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td></td>
<td>__I, _OI, M_I, MOI</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td></td>
<td>M_2, MO</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>1</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>AA</td>
<td>all parses</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>1</td>
</tr>
</tbody>
</table>

(overview). The speaker is assumed to produce true and (preferably/mostly) informative utterances—thereby following the Gricean Maxims of Quality (roughly: be truthful) and Quantity (roughly: be informative) respectively; the listener tries to infer which meaning a speaker most likely had in mind when producing an utterance, and does so on the assumption that the speaker follows the Gricean postulates of truthfulness and informativity.

The speaker production rule $P_S(m | t; \lambda)$ of the vanilla RSA model determines the probability with which the speaker chooses a message (sentence) $m \in M$ when wishing to communicate state (meaning) $t \in T$\(^2\). The model parameter $\lambda$ determines how stringently the speaker selects more informative descriptions over less informative ones. The usual definition utilizes the notion of a literal listener.

\(^2\)The vanilla RSA model and all other models considered here assume that the speaker knows the true world state; an assumption which is arguably warranted by the design of the experiment described in Section 4.
with the following probabilities when we set each message to the Gricean postulates of producing preferably true and informative utterances. Concretely, for in the sense that it models pragmatic inferences derived from the assumption that the speaker adheres that each row is a probability distribution (i.e., summarizes to one).

\[
P_P(m | t; \lambda) \propto \exp(\lambda \cdot \log P(t | m))
\]

The literal listener interprets each message \( m \) based on its semantic meaning, by updating prior beliefs \( P(t) \) with the proposition \( [m] \subseteq T \) that \( m \) is true. However, we may think of the literal listener as a mere technical construct, whose purpose is to anchor the semantic meaning of messages (Franke, 2009). In fact, if the prior probabilities of states are uniform, so that \( P(t) = P(t') \) for all \( t, t' \in T \) (an assumption we will make throughout this paper), the speaker’s choice probabilities, as defined above, can be rewritten in a way that transparently shows the two Gricean constraints of truthfulness and informativity at play:

\[
P_P(m | t; \lambda) \propto [P(t | [m])]^{\lambda} = \left[ \frac{\delta_{\ell \in [m]}}{[m]^{-1}} \right]^\lambda
\]

This reformulation makes clear that the speaker rule will assign probability 0 to any sentence \( m \) which is false in state \( t \) (since then \( \delta_{\ell \in [m]} = 0 \)), as long as there is at least one true sentence available for \( t \).

Between two true messages (for which the \( \delta \)-term evaluates to 1), the message with a stronger semantic meaning, i.e., which is true in fewer world states, will be chosen with a higher probability. The bigger \( \lambda \) the more pronounced this preference for informative utterances.

For example, RSA’s production rule predicts that a speaker would choose utterances to communicate state \( \text{NA} \) with the following probabilities when we set \( \lambda = 5 \):

<table>
<thead>
<tr>
<th>( m )</th>
<th>NN</th>
<th>NS</th>
<th>NA</th>
<th>SN</th>
<th>SS</th>
<th>SA</th>
<th>AN</th>
<th>AS</th>
<th>AA</th>
</tr>
</thead>
<tbody>
<tr>
<td>( P_P(m</td>
<td>t; \lambda = 5) )</td>
<td>0.00</td>
<td>0.00</td>
<td>0.85</td>
<td>0.13</td>
<td>0.01</td>
<td>0.00</td>
<td>0.00</td>
<td>0.00</td>
</tr>
</tbody>
</table>

In words, RSA predicts that speakers would produce only descriptions that are literally true in state \( \text{NA} \), namely “NA”, “SN” or “SS” (see Table 1). Among these, the speaker’s choice probabilities reflect the semantic strength of messages (\( [\text{NA}] = 3 \), \( [\text{SN}] = 4 \) and \( [\text{SS}] = 6 \)). Further example predictions for the RSA speaker rule with \( \lambda = 5 \) are shown in Figure 2. Rows represent states; columns messages. The bars in the cell for state \( t \) and message \( m \) represent the probability \( P_P(m | t; \lambda = 5) \), so that each row is a probability distribution (i.e., summarizes to one).

The comprehension rule of the vanilla RSA model captures a Gricean interpreter. It is Gricean in the sense that it models pragmatic inferences derived from the assumption that the speaker adheres to the Gricean postulates of producing preferably true and informative utterances. Concretely, for each message \( m \) the rule \( P_L(t | m; \lambda) \) assigns a probability to each interpretation \( t \) based on the prior probability \( P(t) \) of a state and the likelihood \( P_P(m | t; \lambda) \) that a (truthful and informative) speaker would use the observed message for this state, following Bayes rule:

\[
P_L(t | m; \lambda) \propto P(t) \cdot P_P(m | t; \lambda)
\]

\(^3\)Here, \( \delta_{\text{boolean}} \) is the delta function which returns 1 if the supplied boolean expression is true and 0 otherwise. The symbol \( \propto \) (for “proportional to”) allows us to define probability distributions more compactly by leaving the normalizing constant implicit: writing \( P(x) \propto F(x) \) for some function \( F(x) \) with domain \( X \) is shorthand for \( P(x) = \frac{F(x)}{\sum_{x \in X} F(x)} \).

\(^4\)The definition given above can be rewritten like so: \( P_P(m | t; \lambda) \propto \exp(\lambda \cdot \log P_L(t | m)) = [P_L(t | m)]^{\lambda} \), which by definition of the literal listener expands to: \( \left[ \frac{P(t)}{\sum_{t'} P(t')} \right]^{\lambda} \). Since \( P(t) = P(t') \) for all \( t, t' \) the prior term cancels out and we retrieve: \( \left[ \frac{\delta_{\ell \in [m]}}{[m]^{-1}} \right]^\lambda \).
Figure 2: Predictions of the production rules of different models for parameter value $\lambda = 5$. Rows represent states, columns messages. An entry in a cell gives the probability assigned to the speaker’s choice of the column-message when trying to communicate the row-state.

Since the interpretation of a message is obtained by normalizing the sender’s likelihoods of sending the message in each state, information about the pragmatic listener’s interpretation can be retrieved from Figure 2 by looking at a column and comparing the relative likelihoods of message choices in different states. For example, “SS” is rather diffuse for $\lambda = 5$; the pragmatic listener assigns some probability to five of the six states where “SS” is true (only state has a semantically very strong alternative that clearly “blocks” this interpretation for “SS”). Concretely, $P_L(\cdot \mid “SS”; \lambda = 5)$ is the following vector of interpretation probabilities:

<table>
<thead>
<tr>
<th></th>
<th>RSA</th>
<th>LU</th>
<th>LI</th>
<th>SO</th>
</tr>
</thead>
<tbody>
<tr>
<td>$P_L(\cdot \mid “SS”; \lambda = 5)$</td>
<td>0</td>
<td>0.002</td>
<td>0.499</td>
<td>0.499</td>
</tr>
</tbody>
</table>

3.2 The Lexical Uncertainty model

The Lexical Uncertainty (LU) model extends the vanilla RSA model by including the listener’s potential uncertainty about the lexical meaning that the speaker assigns to certain expressions (Bergen, Levy, and Goodman, 2016). Like Potts et al. (2016) we are interested here in the case where the listener does not know which lexical entry the speaker has for the word some. The listener may nevertheless try to infer what some most likely means (literally) to the current speaker. This inference, intuitively, proceeds as follows: given that the speaker said $m$ and $m$ can either be interpreted in such
a way that every occurrence of some means “some and maybe all” or in such a way that every occurrence of some means “some but not all”, which pair \((t, l)\) of a state \(t\) and mental lexicon \(l\) is most likely to have caused a Gricean speaker to have produced \(m\)?

Conceptually, the idea that listeners may entertain uncertainty about the lexical meaning a speaker assigns to some is quite distinct from the idea that the grammar produces manifold pragmatic readings for sentences containing words like some by variable insertion of \(O\)-operators at difference scope sites. Nonetheless, we can, for ease of formal parallelism, think of a speaker who considers some to mean “some but not all” as a speaker who always, inflexibly, inserts an \(O\)-operator in front of all occurrences of some when determining the meaning of a sentence; so, this speaker assigns to each nested Aristotelian the meaning it obtains under the parse \(\_0I\) (see Table 1). Similarly, a speaker whose mental lexicon assigns to some its standard logical meaning can be thought of as using the parse ____ for all nested Aristotelians.

Consequently, the LU model can be construed as reasoning about mental lexica \(l\), represented here as parses \(l \in \{____, \_0I\}\). The speaker selects messages \(m\) by the same mechanism as in the RSA model, but based on a semantic interpretation of messages influenced by the speaker’s lexicon \(l\):

\[
P_{S_1}(m | t, l; \lambda) \propto [P(t | [m]])^4]
\]

The production rule of the LU model conditions the speaker’s message choice on that speaker’s fixed lexicon. This will be crucial when comparing the LU model to other models in the following.

The listener does not know the speaker’s mental lexicon but infers which state-lexicon pairs are likely to have caused the speaker to produce the observed utterance, using Bayes rule:

\[
P_{L_1}(t | m; \lambda) \propto P(t) \cdot P(l) \cdot P_{S_1}(m | t, l; \lambda)
\]

This requires the specification of a prior \(P(l)\) for each lexicon \(l\), which is here assumed to be uniform.

If we are interested in \(L_1\)’s interpretation of states only, we can marginalize out uncertainty about the lexicon, like so:

\[
P_{L_1}(t | m; \lambda) = \sum_l P_{L_1}(t, l | m; \lambda)
\]

While Potts et al. (2016) use this latter formula to explain data from a truth-value judgement task, this paper also considers data from an experimental task which is most naturally linked to a production rule. Unfortunately, the production rule \(P_{S_1}\) defined above is not ideally suited for this because it only makes predictions about message-choice probabilities for a given lexicon \(l\). We therefore follow Lassiter and Goodman (2017) and define a speaker \(S_2\) who reasons about pragmatically adequate message choice based on the state-interpretation of \(L_1\). Finally, a pragmatic listener \(L_2\) who reasons about the latter speaker’s choice of messages is defined as before in terms of Bayes rule. Consequently, the final definition of the LU models’ production and comprehension rules are:

\[
P_{S_2}(m | t; \lambda) \propto [P_{L_1}(t | m; \lambda)]^4
\]

\[
P_{L_2}(t | m; \lambda) \propto P(t) \cdot P_{S_2}(m | t; \lambda)
\]

Figure 2 shows the predictions of the LU model’s production rule (for \(\lambda = 5\)). These clearly differ from those of the vanilla RSA model in several places. This also affects interpretation of messages. For example, the sentence “SS” is interpreted as follows:
The most likely interpretation of “SS” is \[
\text{[SS]}.\]
This is because an utterance of “SS” is more likely for a speaker with a lexical entry “some but not all” for some than for a speaker with a standard meaning for some. In turn, this is because “SS” is more informative for the former kind of speaker (\[\|\text{SS}\|_{\text{--}} = 6, \|\text{SS}\|_{\text{OI}} = 3\]) and so more likely to be uttered when true.

### 3.3 The Lexical Intentions model

The LU model treats potential in situ enrichments of some as a consequence of a speaker’s lexicalization of a “some but not all” meaning. This entails that any given speaker assigns to all occurrences of some the same lexical meaning. Consequently, the LU model considers two possible meanings for the sentence “SS”: the standard literal reading (which we can represent using our parse notation as \[\|\text{SS}\|_{\text{--}}\]) and the reading in (7), which corresponds to our \[\|\text{SS}\|_{\text{OI}}\].

(7) Some but not all of the aliens drank some but not all of their water. \((\|\text{SS}\|_{\text{OI}})\)

But it is also conceivable that some speakers, when uttering a sentence with multiple occurrences of some, like “SS”, might mean to convey a reading that results from different lexical meanings for different occurrences, such as in (8a) or (8b), which correspond to readings \[\|\text{SS}\|_{\text{--}}\] and \[\|\text{SS}\|_{\text{O}_-}\] respectively.

(8) a. Some (and maybe all) of the aliens drank some but not all of their water. \((\|\text{SS}\|_{\text{--}})\)

b. Some but not all of the aliens drank some (and maybe all) of their water. \((\|\text{SS}\|_{\text{O}_-})\)

The LU model does not contain speakers of this kind. If we want to model speakers that might associate “SS” with any of the readings in (8), it is not enough to simply include parses \[\text{--I} \] and \[\text{--O}_-\] as additional values that the variable \(l\) can take on, and otherwise leave the LU model unchanged. Why not? – Because the resulting model is conceptually highly implausible. By definition of \(P_{S_1}\), the speaker has a fixed \(l\) and invariably applies it to interpret whatever sentence comes along. That does make sense when \(l\) is instantiated with parses \[\text{--}\] and \[\text{--O}_-\], because we can interpret it as speakers who have lexicalized a particular meaning of some and apply it invariably. But if \(l\) is a parse like \[\text{--O}_-\], we would model a speaker who inflexibly assigns a logical meaning to some in inner position, and a strengthened lexical meaning in outer position, no matter what the sentence and the state to be communicated. This is not a conceptually plausible model of a speaker’s general behavior.

An alternative way is to treat parses as a choice of the speaker, rather than something fixed and immutable. Suppose that all speakers have an ambiguous lexical entry for some; it might mean “some but not all” or “some and maybe all”. Speakers can then choose to mean some in this way or that, depending on whether they deem this beneficial from a communicative point of view. In other words, when carrying an ambiguous lexicon, speakers may utter sentences with different lexical intentions, where the intended meaning for each occurrence of a word might be different. Pragmatic listeners can then try to recover how a speaker may have chosen to mean any single occurrence of some based, as usual, on a model of the speaker’s strategy of producing pairs of sentences and lexical intentions.

The Lexical Intentions (LI) model that formalizes these intuitions looks superficially similar to the LU model, but is conceptually quite different and much simpler. We assume that \(l \in \{\text{--}, \text{--I}, \text{--O}_-, \text{--O}_-\}\)
and define:

\[ P_S(m, l | t ; \lambda) \propto [P(t | [m]^l)]^4 \]
\[ P_L(t, l | m ; \lambda) \propto P(t) \cdot P_S(m, l | t ; \lambda) \]

We then consider the marginals as the predictions of this model.

\[ P_S(m | t ; \lambda) = \sum_l P_S(m, l | t ; \lambda) \]
\[ P_L(t | m ; \lambda) = \sum_l P_L(t, l | m ; \lambda) \]

Technically, the main difference to the LU model is that now the lexical meaning \( l \) is treated not as an argument to be passed into the speaker function, but as an output of it. Conceptually, this means that we do not model speakers who invariably assign a particular lexical meaning to each occurrence of a word, but speakers who choose utterances and their meanings in tandem. They do so, as usual, in such a way as to maximize the informativity of their utterances. Consequently, speakers choose a pair \( \langle m, l \rangle \) as a description of \( t \) only if \( t \in [m]^l \) and they make this choice with a probability proportional to the relative informativity of \([m]^l\), i.e., they prefer pairs \( \langle m, l \rangle \) for which \([m]^l\) is small.

The predictions of LI are subtly different from the two previous models (see Figure 2). For example, the LI model predicts that the listener’s interpretation for the sentence “SS” is:

<table>
<thead>
<tr>
<th>Parsing</th>
<th>Probability</th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
<td>0.00</td>
</tr>
<tr>
<td>1</td>
<td>0.31</td>
</tr>
<tr>
<td>2</td>
<td>0.15</td>
</tr>
<tr>
<td>3</td>
<td>0.38</td>
</tr>
<tr>
<td>4</td>
<td>0.00</td>
</tr>
<tr>
<td>5</td>
<td>0.16</td>
</tr>
<tr>
<td>6</td>
<td>0.00</td>
</tr>
</tbody>
</table>

According to the LI model, “SS” should be associated with interpretation 3 much more readily than predicted by vanilla RSA or the LU model.

### 3.4 The Silent Only model

The LI model assumes that speakers have ambiguous lexical entries for *some*. When pondering the choice of a sentence to communicate the given state, speakers actively choose readings of sentences which makes them true and informative. From here it is only a very small step towards a model that includes all readings supplied by grammaticalism. The Silent *Only* (SO) model is exactly like the LI model, but integrates the all parses \( p \) considered in Table 1:

\[ P_S(m, p | t ; \lambda) \propto [P(t | [m]^p)]^4 \]
\[ P_L(t, p | m ; \lambda) \propto P(t) \cdot P_S(m, p | t ; \lambda) \]

The SO model goes far beyond the LU model of Potts et al. (2016). While the LU model captures certain phenomena, like locally embedded scalar implicatures, in an RSA-style reasoning framework, it does directly engage with or adopt grammaticalism. The SO model, instead, does. Including the full set of grammaticalist readings into the LU model of Potts et al. (2016) makes little conceptual sense for the same reasons laid out in the context of the LI model: it would be quite unnatural speakers who cannot rise above their language’s ambiguity, invariably applying a particular pattern of insertions of O-operators, irrespective of the sentence under consideration and the state to be described. The SO model, instead, models speakers who are masters of the rich ambiguities provided by their language’s
grammar, using these ambiguities to flexibly intent to express this reading or another, depending on what serves communication.

The SO model also makes distinct empirical predictions (Figure 2). For example, the speaker is predicted to use the sentence “SS” with a very high probability in state \[\text{WW}\]. This is because the reading \[\text{[SS]}\] is available for the speaker, which yields a very strong reading that uniquely singles out this world state. Consequently, the listener’s interpretation of “SS” also puts substantial probability on the interpretation \[\text{WW}\]:

\[
P_L(\cdot | \text{“SS”}, \lambda = 5) = \begin{bmatrix} 0.00 & 0.70 & 0.05 & 0.21 & 0.00 & 0.03 & 0.00 \end{bmatrix}
\]

4 Experiment

Design. The models just introduced differ mostly in their quantitative predictions about likelihood of expression and interpretation choices. Introspection is not reliable enough to assess such fine-grained probabilistic predictions, but experimental data may be. As all models make predictions about expression-choice and state-interpretation probabilities, we collected data from tasks that probed rather directly into these model predictions. We will refer to these tasks, perhaps a bit sloppily, as a production and an interpretation task.

Participants. 100 participants with US IP-addresses were recruited via Amazon’s Mechanical Turk, using psiTurk (Gureckis et al., 2016). We excluded data from two participants who did the experiment twice and another three who did not self-identify as native speakers of English.

Materials. The experiment was couched in a cover story about friendly aliens visiting earth. The relevant test sentences are the nested Aristotelians in (3). We used quantification over a mass term (water that the aliens drank) for the inner quantifier so as to stay clear as much as possible of potential typicality effects (e.g. Degen and Tanenhaus, 2015; van Tiel, 2014). It is not easily possible to do this for the outer quantifier as well. The seven relevant states were displayed using pictures like in Figure 1. All pictures contained twelve aliens with full, half-full or empty water jugs. For situations corresponding to a “some but not all” reading of the outer quantifier we used four or six aliens (as appropriate) out of the total twelve, both of which are fairly natural or typical numbers to be denoted by some (Degen and Tanenhaus, 2015; van Tiel, 2014).

Procedure. Participants were first told that the experiment consisted of two parts. Each part started with a background story, which served to introduce the alien scenario, and what was expected of participants, as well as our sentence and picture material. Participants completed seven production trials, one for each world state, in random order. Each trial displayed a picture of the state and participants selected the outer and inner quantifier from a dropdown menu (see Figure 3a). Participants then completed nine interpretation trials, one for each sentence, in random order. Each trial displayed all of the seven states with a short repetition of the task question and a slider bar next to the picture (see Figure 3b). Participants rated how likely they thought the displayed situation is what the speaker had observed.
Results. Figure 4 shows the frequencies with which sentences were selected for each state. Interestingly, sentences with *none* as outer quantifier were used very infrequently. Another interesting point is that sentence “AS” was the most frequently used message in state [□], where its putative local reading in (2b) is true. This seems to conflict with the predictions of the RSA model plotted in Table 1, but consistent with those of the other models. Finally, we see that the most frequently used message in state [■] is “SS,” which goes against the predictions of the RSA and the LU model plotted in Table 1, but is consistent with the SO model.

Each trial in the comprehension task returns nine slider ratings, one for each state. We normalized each of these nine-placed vectors, so that it sums to one. Figure 5 shows averages over these normalized vectors. The interpretation of “AS” sentences reflects the production data in the sense that the (on average) most likely interpretation was state [□], followed by state [■] and finally [■], thereby replicating previous results on interpretation preferences for these sentences (Chemla and Spector, 2011; Franke, Schlotterbeck, and Augurzky, 2017). Finally, the interpretation of the sentence “SS” seems inconsistent with the interpretation predicted by the RSA model (for $\lambda = 5$), but it is hard to assess by visual inspection whether this case might provide strong evidence for or against any of the other models. This is why we turn to formal model comparison next.

5 Model comparison

In order to quantify the relative evidence provided by the experimental data for or against the models introduced in Section 3 we look at Bayes factors (Jeffreys, 1961; Kass and Raftery, 1995). From a Bayesian point of view, a model $M$ consists of a prior $P(\theta \mid M)$ over vectors $\theta$ of values for its parameters and a likelihood function $P(D \mid \theta, M)$, which assigns a likelihood to the observed data $D$ for each vector $\theta$. The marginalized likelihood for model $M$ given data $D$ quantifies how likely $D$ is a
Figure 4: Frequencies of sentence choice for each world state. The black outlines indicate cases where a sentence is true in a given state under at least one of the pragmatic construals from Table 1.

priori for any parameter value:

\[ P(D \mid M) = \int P(\theta \mid M) \cdot P(D \mid \theta, M) \, d\theta. \]

The Bayes factor in favor of model \( M_1 \) over \( M_2 \) is the ratio of marginalized likelihoods: \( \frac{P(D \mid M_1)}{P(D \mid M_2)} \). It quantifies the factor by which our beliefs should shift in favor of \( M_1 \), relative to \( M_2 \), given that we have observed \( D \), since by Bayes rule:

\[
\frac{P(M_1 \mid D)}{P(M_2 \mid D)} = \frac{P(D \mid M_1) \cdot P(M_1)}{P(D \mid M_2) \cdot P(M_2)}. 
\]

Bayes factors are therefore independent of the prior odds of models \( P(M_1) / P(M_2) \), but depend on the priors over parameter values \( P(\theta \mid M_i) \) for each model \( M_i \).

We have production \( D_p \) and comprehension data \( D_c \) and compare models based on how well they explain the conjunction of both, so that \( P(D \mid \theta, M) = P(D_p \mid \theta, M) \cdot P(D_c \mid \theta, M) \). The following explains how probabilistic speaker and listener rules, as defined in Section 3, give rise to a likelihood function for \( D_p \) and \( D_c \) respectively.

The production rule of model \( M \) defines the likelihood \( P_S(m \mid t, \theta, M) \) of a single choice of expression for a state. All of the models from Section 3 use a single parameter \( \lambda \). On top of these, we include two more parameters in each model in order to accommodate for two general observations about \( D_p \). First, our models predict probability zero for messages that are false for a given state. However, we do observe false message choices (see Figure 4). We therefore add a constant error term \( c \varepsilon \) to all predicted message choice probabilities. Second, since no model accounts by itself for the low choice rates of sentences starting with \textit{none}, we include a cost term which is fixed to zero for messages.
starting with some or all, but may be positive for messages starting with none. These costs capture a
general dispreference for particular expressions, and are not to be confused with “processing costs”
from the psycholinguistic literature. Costs are subtracted, following standard practice, so that, e.g.,
for vanilla RSA we obtain:

\[
P_S(m \mid t; \lambda, \epsilon, c) \propto [P_{LL}(t \mid m) - \text{cost}(m)]^\lambda + \epsilon
\]

Finally, if \( D_p \) consists of counts \( n_{ij} \) of the number of times message \( m_j \) was chosen in state \( t_i \), the likelihood \( P(D_p \mid \theta, M) \) is:

\[
P(D_p \mid \theta, M) \propto \prod_{i=1}^{7} \text{Multinomial}(\langle n_{i1}, \ldots, n_{i9} \rangle, \langle P_S(m_1 \mid t_i, \theta, M), \ldots, P_S(m_9 \mid t_i, \theta, M) \rangle).
\]

Comprehension rules give a probability distribution over states for each given message: \( P_L(t \mid m, \theta, M) \). Since these are defined in terms of the speaker choice probabilities, the parameterization of
the listener rules are the same. The comprehension data \( D_c \) consists of probability vectors \( c_{ij} \), where \( c_{ij} \)
is the average of normalized ratings assigned to state \( j \) for message \( i \) (as plotted in Figure 5). We think
of the probability vector \( \langle c_{i1}, \ldots, c_{i7} \rangle \) as a sample from a Dirichlet distribution whose modal value
is the model’s prediction \( \langle P_L(t_1 \mid m_i, \theta, M), \ldots, P_L(t_7 \mid m_i, \theta, M) \rangle \). To allow for more or less deviation
in the realization of observed ratings, we introduce a parameter \( w \), where the higher \( w \) the more we
expect observations that are very close to the model’s predictions:

\[ P(D_c \mid \theta, w, M) \propto \prod_{i=1}^{9} \text{Dirichlet}(c_{i1}, \ldots, c_{i7}) \cdot w \cdot \langle P_L(t_1 \mid m_i, \theta, M), \ldots, P_L(t_7 \mid m_i, \theta, M) \rangle. \]

As for priors over model parameters, we assume that all parameters are independently sampled from flat priors with a sufficiently large support. Priors are the same for all models:

\[ \lambda \sim \text{Uniform}(0, 6) \quad c \sim \text{Uniform}(0, 5) \quad \epsilon = 0.045 \]

Notice that all models fix the error parameter \( \epsilon \) to single value, which is chosen to be close to the average of each model’s maximum likelihood estimate for \( \epsilon \) over all models.

Estimates of marginal likelihoods were obtained by grid approximation on a grid size of 20 for each parameter (e.g., Kruschke, 2015, Chapter 10). Figure 6a shows the resulting Bayes factor approximations in favor of each model when compared to the RSA model. We see that all models are better than the baseline vanilla RSA model. The best model is SO. The Bayes factor in favor of SO when compared against the second best model, LI, is approximately 29; and when compared against the third best model, LU, it is ca. 95. From these results we can also give approximations of modeller’s posterior beliefs after conditioning unbiased priors over models with the observed data:

<table>
<thead>
<tr>
<th>model ( M_i )</th>
<th>RSA</th>
<th>LU</th>
<th>LI</th>
<th>SO</th>
</tr>
</thead>
<tbody>
<tr>
<td>( P(M_i \mid D) )</td>
<td>0.00</td>
<td>0.01</td>
<td>0.033</td>
<td>0.956</td>
</tr>
</tbody>
</table>

In words, if each model is \textit{a priori} equally likely, the posterior probability, after seeing the data, of the SO model is around 0.956. The combined data from production and comprehension provide strong evidence in favor of the silent-only model, suggesting that, within the extensions of RSA-style pragmatic reasoning models, it is best to include the full range of grammatically generated implicature readings.

Since Bayes factors depend on priors over parameter values, Figure 6b additionally shows the results of model comparison using the Bayesian Information Criterion (BIC) (Schwarz, 1978), which relies on the maximum likelihood estimates of the parameters and is therefore independent of priors over parameter values. A model is better the lower its BIC score. Consequently, we retrieve the same ordinal result, as for Bayes factor comparison.

To understand the results better, we can look at production and comprehension data separately. For only production data, the Bayes factor for the SO model is about 30600; the LI model is absolutely no competition. If we look at comprehension data only, the Bayes factor in favor of the SO model is ca. 0.011, i.e., the LI model is roughly 100 times more likely a posteriori, if we start from unbiased prior beliefs. Consequently, the main advantage of the SO model comes from its superior predictions for the production data, which outweigh its weaker predictions for comprehension.

We can zoom in even more and look at each individual condition from production (i.e., the message choices at a single state) and comprehension (i.e., the average slider ratings for a single sentence). Figure 6c plots these results and clearly shows that the main evidence in favor of the SO model comes from its superior predictions for the production data in state \[ \text{[ ]} \]. Figure 7 plots the \textit{prior predictive distributions} for production data under both models. The prior predictive distribution gives the
(a) Bayes factors in favor of a model when compared to RSA, plotted on a log scale.

(b) Model comparison based on BICs (with maximum likelihood estimates for parameters)

<table>
<thead>
<tr>
<th>model</th>
<th>parameter</th>
<th>λ</th>
<th>c</th>
<th>w</th>
<th>BIC</th>
</tr>
</thead>
<tbody>
<tr>
<td>RSA</td>
<td></td>
<td>1.43</td>
<td>1.56</td>
<td>43.85</td>
<td>259.73</td>
</tr>
<tr>
<td>LU</td>
<td></td>
<td>0.94</td>
<td>2.2</td>
<td>39.75</td>
<td>202.06</td>
</tr>
<tr>
<td>LI</td>
<td></td>
<td>1.27</td>
<td>1.52</td>
<td>41.22</td>
<td>200.16</td>
</tr>
<tr>
<td>SO</td>
<td></td>
<td>1.3</td>
<td>1.71</td>
<td>31.87</td>
<td>195.69</td>
</tr>
</tbody>
</table>

(c) Bayes factors in favor of SO over LI per production (left) and comprehension condition (left)

Figure 6: Results from model comparison

marginalized likelihood for each possible data observation. We see that in most conditions the prior predictives of models coincide, and that the main difference in prior predictions is indeed that the LI model underpredicts the choice of sentence “SS” and overpredicts choice of “SN” for condition [1]. The LI model makes better predictions for states [2] and [3] but, as becomes apparent from Figure 7, its predictive advantage over the SO model is less pronounced than SO’s advantage in condition [4]. The reason why the SO model predicts a high frequency of “SS” choices in state [5] is because it makes it possible, so to speak, for a rational speaker say “SS” and intend a global reading. The global reading with M__ is not available to the LI model, but it serves to perfectly single out state [6] from all other states (see Table 1 in main paper) and so it is rational for a speaker to say “SS” and mean [7].

6 Conclusions

We argued in Section 3 that it is conceptually implausible to that the Lexical Uncertainty model of Potts et al. (2016) by just including all readings made available by grammaticalism. We therefore in-
roduced a novel RSA-style reasoning model, the Silent Only model, which is fundamentally different from the LU model. The SO model treats Gricean speakers as capable of choosing one out of several readings supplied by the grammar as the intended meaning of their utterance. Based on empirical data from two tasks that link directly to our models' production and comprehension rules, statistical model comparison in terms of Bayes factors suggests that the SO model might indeed be empirically superior. The main reason for the SO model’s predictive success was identified as the availability of a reading of the sentence “SS” obtained from inserting an $O$-operator at matrix position. This provides suggestive evidence for the idea that the full set of grammatically-induced readings is needed.\footnote{A reviewer asks whether the reading $\{SS\}$ would not be recoverable by a vanilla RSA model, if we allow higher-order reasoning, i.e., a speaker who chooses expressions based on a pragmatic interpreter and so on (more on higher-order reasoning below). This is not the case. Vanilla RSA will not associate “SS” with the interpretation $\text{[SS]}$, even if we allow recursive pragmatic reasoning. The reason is, roughly put, that “SN” is always a better choice in state $\text{[SN]}$ than “SS” for the speaker, so that “SS” has a higher probability of being produced in a different state, so that “SN” effectively blocks “SS” from interpretation $\text{[SN]}$.}

The approach to scalar implicature reasoning formalized in the SO model is not an innocuous synthesis of traditional Gricean and grammatical approaches to scalar implicatures. It is a new breed of its own. Unlike traditional Gricean approaches, the SO model assumes that sentences may have a rich set of potential readings, e.g., as part of the output of a compositional semantic system, including readings that look as if a scalar implicature occurred in the scope of a logical operator. The approach suggested here would be able to accommodate different mechanisms of how these readings are generated, but a grammatical approach to pragmatic readings, involving flexible insertion of $O$-operators clearly suggests itself due to its recent popularity and empirical success. Still, the current approach is not a grammatical approach to scalar implicature either. It is rather a pragmatic account of grammatically-supported meaning enrichments. It puts pragmatic reasoning first again. It embeds grammaticalism and adds predictions concerning the (partial) disambiguation of utterances. It also provides a possible link hypothesis, i.e., a way of deriving precise quantitative predictions from grammatically supplied pragmatic readings.

The account suggested here, even if distinct from grammaticalism, can nonetheless inform grammaticalist theory in multiple ways. On the assumption that a system of rational pragmatic reasoning embeds, harnesses and controls the rich ambiguity supplied by grammar, it becomes possible to ask: which aspects of the underlying grammatical system make the combined system successful as an efficient system of communication from the point of view of language evolution (Brochhagen, Franke, and Rooij, in press), or as an explanation of data from experimental semantics/pragmatics. With regard to the latter, we might, for example, fix the SO model but compare different instantiations of grammaticalism embedded inside of it. As noted in Section 2, there are different instantiations for grammaticalism. There is no unanimity about how grammaticalist readings of complex sentences should be construed. For example, Gotzner and Romoli (2018) argue in favor of a different construction of the set of sentential alternatives $\text{Alt}(S, p)$ than used here. Without going into the details of their construction, the Bayes factor in favor of an SO model based on the simpler construction used here, when compared to Gotzner and Romoli’s more elaborate alternative, is about 1530, suggesting very clearly that, for this data set and the assumption that the SO model is the correct empirical link hypothesis, the simpler construction used here is empirically superior. This is not meant to be a decisive argument against Gotzner and Romoli’s approach, but merely as a gesture towards how the general approach suggested here is able to inform the underlying linguistic theory in future research.

Future work may also scrutinize different models of pragmatic reasoning. For example, the models defined here are all minimal in the sense that they consider the first pair of a production and
comprehension rule, where the latter builds on the former, and which are both directly applicable to
the experimental data at hand. For all other models—except the LU model where we needed to iterate
further for technical reasons—this entailed a traditional Gricean setup: the speaker is assumed to pre-
der information messages where informativity is related to semantic strength, i.e., literal meaning, not
pragmatically enriched meaning of some sort. This means that all of the models considered here as-
sume that speakers do not have a veridical view of the listener’s interpretation; they base their choices
on what could be interpreted to be a simple Gricean heuristic of preferring semantically stronger
messages; they do not engage actively in audience design. Clearly, exploring other models of prag-
matic reasoning in connection with grammatically generated implicature readings, and testing them
in experimental designs which take the possibility of active audience design into account, is clearly a
worthwhile enterprise for future research.

Although our focus here was clearly conceptually, to introduce the SO model and to demonstrate
the use of Bayesian model comparison, more work is clearly necessary on the empirical side. The
scope of our experimental manipulations is admittedly rather small. Recent experimental papers on
scalar implicatures (e.g. Chemla and Spector, 2011; Franke, Schlotterbeck, and Augurzky, 2017) in
complex sentences have focused attention on scalar implicature triggers in non-monotonic environ-
ments like in (9).

(9) Exactly one of the aliens drank some of its water.

Our design was more restricted in order to keep the set of alternative utterances for global Gricean
reasoning confined to what is hopefully a maximally uncontroversial selection. For instance, the
modeling of Potts et al. implicitly assumes that listeners also include (9) as a speaker’s alternative
utterance during the interpretation of any nested Aristotelian. It is debatable whether this is plausible
but it may have non-trivial consequences on model predictions. To sidestep exactly these problems, at
least for the time being, this paper’s experimental set-up is deliberately minimal. Nonetheless, our data
offer enough grip for a rational experimenter to quite substantially update their beliefs about which
model could likely be true.

References

Inference”. In: Semantics & Pragmatics 9.20.

meaning and pragmatic use”. In: Cognitive Science.

Chemla, Emmanuel and Benjamin Spector (2011). “Experimental Evidence for Embedded Scalar Im-
plicatures”. In: Journal of Semantics 28, pp. 359–400.

Chierchia, Gennaro, Danny Fox, and Benjamin Spector (2012). “Scalar Implicature as a Grammatical
Phenomenon”. In: Semantics. An International Handbook of Natural Language Meaning. Ed. by

Cohen, L. Jonathan (1971). “Some Remarks on Grice’s Views About the Logical Particles of Natural
Language”. In: Pragmatics of Natural Language. Ed. by Yehoshua Bar-Hillel. Dordrecht: Reidel,
pp. 50–68.


Franke, Michael and Gerhard Jäger (2016). “Probabilistic pragmatics, or why Bayes’ rule is probably important for pragmatics”. In: *Zeitschrift für Sprachwissenschaft* 35.1, pp. 3–44.


Figure 7: Prior predictive distribution for the production data. Black vertical bars indicate the observed counts for each condition.