RESEARCH REPORT

Theory-driven statistical modeling for semantics and pragmatics:
A case study on grammatically generated implicature readings

MICHAEL FRANKE
University of Osnabrück

LEON BERGEN
University of California, San Diego

Computational probabilistic modeling is increasingly popular in linguistics, but its relationship with linguistic theory is ambivalent. We argue here for the potential benefit of theory-driven statistical modeling, based on a case study situated at the semantics-pragmatics interface. Using data from a novel experiment, we employ Bayesian model comparison to evaluate the predictive adequacy of four models that differ in the extent to and manner in which grammatically generated candidate readings are taken into account in four probabilistic pragmatic models of utterance and interpretation choice. The data provide strong evidence for the idea that the full range of potential readings made available by recently popular grammatical approaches to scalar-implicature computation might be needed, and that classical Gricean reasoning may help manage the manifold ambiguity introduced by grammatical approaches to these. The case study thereby shows a way of bridging linguistic theory and empirical data with the help of probabilistic pragmatic modeling as a linking function.

Keywords: pragmatics, computational modeling, scalar implicature, grammaticalism, Bayesian model selection

1. INTRODUCTION. The last two decades have brought an empirical turn to several areas of theoretical linguistics, including syntax (Sprouse 2007), semantics (Bott et al. 2011), and pragmatics (Noveck & Sperber 2004). It is exciting when a discipline with a rich theoretical inventory becomes more experimental. But there are also challenges. An important issue concerns the precise linking of established theoretical ideas to potential experimental observations (Chemla & Singh 2014). Based on a case study on scalar implicature, this research report argues for a benefit of casting semantic theory explicitly into a probabilistic model of pragmatic decisions, by means of which empirical data can be analyzed in a theory-driven way.

The most common template for the empirical testing of a (formal or verbal) theory $T$ is to derive specific predictions from $T$, in the form of a hypothesis $H$, ideally such that competitor theories (or, in the absence of such: common sense) would not make $H$ appear very likely. An experiment is then designed to assess $H$ in light of empirical data $D$. The most frequent method to see whether $H$ is corroborated by $D$ is to use a statistical model $M_{ST}$, which is chosen from a restricted repertoire of standardized tools appropriate for the kind of experiment and data at hand (e.g. a generalized regression model). Interestingly, the original theory $T$ does not usually inform the statistical model $M_{ST}$ with which the data are analyzed. The model $M_{ST}$ could be used in geology, medicine, or economics. This is particularly troublesome for disciplines with a strong theoretical foundation that could be used to formulate more specific, theory-driven models $M_T$ which directly link theory $T$ to data $D$. Indeed, researchers in several areas of science likewise argue for the integration of more theory-driven statistical modeling, for example, in ecology (McElreath 2016), psychology (Lee & Wagenmakers 2015), and also linguistics (Brasoveanu & Dotlacić 2020).

To demonstrate the possible value of theory-driven modeling in an empirically oriented formal linguistics, this paper considers a case study on the familiar notion of scalar implicature (Geurts 2010, Grice 1975, Horn 2004). In response to certain shortcomings of a traditional Gricean account of implicature, a recently popular and
empirically successful alternative approach postulates that potential implicature-like meaning enrichments are generated in the grammar, thereby creating a massive semantic ambiguity (e.g. Chierchia, Fox, & Spector 2012, Fox 2007, Fox & Spector 2018). We here address how these multiple semantic readings affect a speaker’s choice of utterance and a listener’s choice of interpretation, which is something not directly predicted by semantic theory alone, but necessary for linking its predictions to experimental data. We introduce four different models of utterance and interpretation choices, as variants of the recently popular ‘rational speech act’ models (Frank & Goodman 2012, Franke & Jäger 2016, Goodman & Frank 2016). Models differ in the way that semantic ambiguity affects choices of utterance and interpretation (see Fig. 2 below). The key contribution here is a novel model, the so-called “global intentions model”, in which speakers choose both an overtly observable utterance and its intended meaning (the latter stemming from the set of grammatically generated implicature meanings). This theory-driven model $M_T$ can be linked directly to experimental data from a novel small-scale experiment introduced here. We then use Bayesian model comparison (e.g. Jeffreys 1961, Kass & Raftery 1995), thereby demonstrating how theory-driven statistical modeling could make for an additional, helpful tool for testing linguistic theories more directly based on empirical data.

The paper is structured as follows. We first introduce the grammatical approach to scalar implicature (§2), and then develop the relevant probabilistic models for utterance and interpretation choices in the light of multiple semantic ambiguity (§3). We report on a small-scale combined production and comprehension experiment in §4, and §5 uses this data to compare models.1

2. Grammatically generated scalar-implicature readings. Scalar implicatures are a special kind of quantity implicature, so-called in reference to Grice’s maxim of quantity, which requires speakers to be as informative as possible, given the current goals of the conversation (Grice 1975). For example, speakers who utter 1a may be taken to communicate 1b, because they did not utter the logically stronger, and therefore more informative, alternative sentence 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.
   d. I own some of Johnny Cash’s albums, and it’s not true that I own all.

Traditional Gricean accounts rationalize scalar implicatures as inferences a listener can draw based on reasoning about the speaker’s actions. The action to be rationalized in the example above, so the mainstream account goes, is that the speaker uttered the complete sentence in 1a. An alternative action of the speaker is an utterance of 1c. The implicature reading in 1b therefore arises by conjoining the literal meaning of the sentence in 1a with the negation of the complete sentential alternative in 1c, resulting in 1d, which is equivalent to 1b.

This logic has difficulty explaining that an utterance of 2a intuitively conveys the meaning in 2b. This inference does not follow from 2d, which is the result of conjoining the negation of the meaning of the alternative sentence 2c with the literal meaning of 2a (but see Geurts 2010 for an alternative solution).

1 Materials, data, and modeling code can be retrieved from https://osf.io/fuary/.
(2) a. A soldier showed some signs of malaria.
   b. A soldier showed some but not all signs of malaria.
   c. A soldier showed all signs of malaria.
   d. A soldier showed some signs of malaria, and it’s not true that a soldier showed all signs.

In response to problems like this, an alternative approach has gained prominence recently (e.g. Chierchia, Fox, & Spector 2012, Fox 2007, Fox & Spector 2018). Grammaticalism assumes that the sentence in 2a is semantically ambiguous. It has a literal reading, where *some* just means ‘some and maybe all’. But it also has a reading where *some* is enriched to mean ‘some but not all’ in the scope of *a soldier*. Grammaticalism consequently earns its name by trying to incorporate meaning aspects that may appear to be pragmatic in nature into the grammar.

To illustrate the range of readings made available by grammaticalism, consider the set of nested aristotelians, which is the focus of this paper. Nested Aristotelians are 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) {None | Some | All} of the aliens drank {none | some | all} of their water.

In what follows, we use abbreviations for these sentences, indicating the outer and inner quantifier in order, as on the left-hand side of Table 1 below. 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 that drank none of their water, (ii) aliens that drank some but not all, and (iii) aliens that drank all. We therefore distinguish seven kinds of world states, each of which Figure 1 provides an example of. Each situation in Fig. 1 shows twelve aliens, each with its glass of water. To have a compact representation of relevant world states, we use pictures like [- -] or [ - -]. These pictures indicate which types of aliens exist. If there are aliens that drank none of their water, the pictorial representation for this world state contains a full glass, as in [ - -]; otherwise, if there are no aliens that drank none of their water (every alien drank at least some), then the corresponding picture shows no full glass, as in [- -].
The state \(-\)\(-\), for example, refers to the set of worlds in which all of the aliens drank some but not all of their water.

Grammaticalism assumes that pragmatic readings of sentences are generated by a silent exhaustification operator \(\text{Exh}\) whose meaning contribution is similar to that of the particle \(\text{only}\) (e.g. Chierchia, Fox, & Spector 2012). For present purposes, there are three relevant places where an \(\text{Exh}\)-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.

\[
\text{(4) } \text{Exh}_M [\text{Exh}_O (Q_O) \text{ of the aliens drank } \text{Exh}_I (Q_I) \text{ of their water}]
\]

A parse determines the reading of a sentence. It determines whether \(\text{Exh}\) occurs in matrix (\(M\)), in outer quantifier (\(O\)), or in inner quantifier (\(I\)) position. We use a notation where, for example, a parse of a sentence with \(\text{Exh}\) occurring only in matrix position is denoted as ‘\(M\)’, a parse with \(\text{Exh}\) occurring at all three relevant positions is written as ‘\(\text{MOI}\)’, and a parse without any \(\text{Exh}\) operator, thus a literal reading, is represented as ‘lit’. If \(S\) is a sentence and \(p\) its parse, then \[[S]^p\] is the reading of \(S\) under \(p\). Sentences can have multiple readings based on different parses. Table 1 lists the readings of nested Aristotelians. These readings are derived from what is perhaps the simplest instantiation of grammaticalism (see the appendix for technical details).

<table>
<thead>
<tr>
<th>SENTENCE</th>
<th>PARSES</th>
<th>WORLD STATE</th>
</tr>
</thead>
<tbody>
<tr>
<td>NN</td>
<td>lit, I, O, OI</td>
<td>0 0 0 0 0 1 1 1</td>
</tr>
<tr>
<td>M, MI, MO, MOI</td>
<td>0 0 0 0 0 1 1 0</td>
<td></td>
</tr>
<tr>
<td>NS</td>
<td>lit, O, M, MO</td>
<td>1 0 0 0 0 0 0 0</td>
</tr>
<tr>
<td>I, MI, MOI</td>
<td>0 0 0 1 1 0 0 1</td>
<td></td>
</tr>
<tr>
<td>NA</td>
<td>lit, I, O, OI</td>
<td>1 1 0 0 1 0 0 1</td>
</tr>
<tr>
<td>M, MI, MO, MOI</td>
<td>0 1 0 0 1 0 0 0</td>
<td></td>
</tr>
<tr>
<td>SN</td>
<td>lit, I</td>
<td>1 1 1 1 0 0 0 0</td>
</tr>
<tr>
<td>all others</td>
<td>0 1 1 1 1 0 0 0</td>
<td></td>
</tr>
<tr>
<td>SS</td>
<td>lit</td>
<td>0 1 1 1 1 1 1 1</td>
</tr>
<tr>
<td>I, I</td>
<td>0 1 1 1 0 0 0 0</td>
<td></td>
</tr>
<tr>
<td>O, MO</td>
<td>0 1 1 1 0 0 0 0</td>
<td></td>
</tr>
<tr>
<td>OI, MOI</td>
<td>0 1 0 0 1 1 0 0</td>
<td></td>
</tr>
<tr>
<td>M</td>
<td>0 1 0 0 0 0 0 0</td>
<td></td>
</tr>
<tr>
<td>SA</td>
<td>lit, I</td>
<td>0 0 1 1 0 0 1 1</td>
</tr>
<tr>
<td>all others</td>
<td>0 0 1 1 0 0 1 0</td>
<td></td>
</tr>
<tr>
<td>AN</td>
<td>all parses</td>
<td>1 0 0 0 0 0 0 0</td>
</tr>
<tr>
<td>AS</td>
<td>lit, O</td>
<td>0 0 0 0 0 0 1 1</td>
</tr>
<tr>
<td>I, I, MI, MOI</td>
<td>0 0 0 0 1 1 0 0</td>
<td></td>
</tr>
<tr>
<td>M, MO</td>
<td>0 0 0 0 1 1 0 0</td>
<td></td>
</tr>
<tr>
<td>AA</td>
<td>all parses</td>
<td>0 0 0 0 0 0 1 0</td>
</tr>
</tbody>
</table>

Table 1. Readings produced by the instantiation of grammaticalism described in the appendix. For each nested Aristotelian, the table shows whether this sentence is true (1) or false (0) in each world state under a given parse. For example, the first line of the table gives the truth conditions of sentence ‘NN’ for parses without \(\text{Exh}\) in matrix position, all of which happen to be equivalent.

In sum, grammaticalism stipulates silent grammatical operators, whose effect it is to generate a rich but conventionally predictable semantic ambiguity. It thereby predicts the availability of intuitively attractive readings for cases, like in 2, that are difficult to explain for traditional Gricean accounts. However, it remains controversial whether all of these readings are important for explaining relevant empirical data (e.g. Chemla & Spector 2011, Franke, Schlotterbeck, & Augurzky 2017, Geurts & Pouscoulous 2009, Geurts & van Tiel 2013, Potts et al. 2016). Moreover, it is, to a large extent, an open issue how the massive ambiguity generated in the grammar is resolved in context, by
some, perhaps pragmatic, mechanism (for some discussion see e.g. Chierchia, Fox, & Spector 2012, Fox & Spector 2018). Seen in this way, it would be desirable to explore how far a combination of, on the one hand, Gricean ideas about efficient communication in the light of potential ambiguity and, on the other hand, a grammatical approach that generates potential pragmatic readings could solve each approach’s problems. This is the goal of the models introduced next.

3. Models. We consider four models, all of which rely on a general, Grice-inspired picture of a speaker trying to maximize the amount of information conveyed by an utterance and a listener consequently interpreting the speaker’s utterance based on such information-maximizing behavior. Each model predicts a probability with which a speaker chooses each nested Aristotelian to describe any given world state (from Fig. 1) and the degree of belief a listener assigns to each world state after hearing any given nested Aristotelian. The models differ, however, in the way they integrate grammatically supplied ambiguity.

Figure 2 gives a condensed overview of the main conceptual differences. First, the vanilla rational speech act (RSA) model of Frank and Goodman (2012) associates each sentence with its logical semantics only (§3.1). On the other end of the spectrum, the global intentions model (§3.4) considers all of the readings listed in Table 1. In between these two extremes are two models that take increasingly large subsets of the readings listed in Table 1 into account (§3.2, §3.3). Importantly, however, models also differ in the manner in which semantic ambiguity affects speakers’ and listeners’ choices. The lexical uncertainty model of Potts et al. (2016) features a speaker-production part in which each speaker is assumed to have a fixed, inflexible lexicon with exactly one meaning attached to each lexical item (§3.2). In addition, we also introduce two conceptually different, novel models in which all speakers are aware of the full semantic ambiguity generated by either their lexicon (lexical intentions model; §3.3) or grammar (global intentions model; §3.4), and in which they flexibly use the potential this ambiguity creates for the purpose of maximizing information flow in communication.

Figure 2. Schematic representation of main conceptual differences in the speaker-production part of the four models compared here.
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 & Jäger 2016, Goodman & Frank 2016 for an overview). The speaker is assumed to follow the Gricean maxims of quality (roughly: be truthful) and quantity (roughly: be informative). 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 $P_S(u \mid t; \alpha)$ of the vanilla RSA model determines the probability with which the speaker chooses an utterance $u \in U$ when the true state is $t \in T$. The model parameter $\alpha$ determines how stringently the speaker selects more informative descriptions over less informative ones (see below). The usual definition utilizes the notion of a literal listener ($\text{LL}$) and looks like this.\(^3\)

\begin{align*}
(5) & \quad P_{\text{LL}}(t \mid u) = P(t \mid [u]) \propto P(t) \delta_{[u] \subseteq [t]} \\
(6) & \quad P_S(u \mid t; \alpha) \propto \exp(\alpha \cdot \log P_{\text{LL}}(t \mid u))
\end{align*}

The literal listener, defined in equation 5, interprets each utterance $u$ based on its semantic meaning, by simple Bayesian updating $P(t \mid [u])$ of prior beliefs $P(t)$ with the proposition $[u] \subseteq T$ that $u$ is true.\(^4\)

It helps to think of the literal listener as a mere technical construct, whose purpose is to anchor the semantic meaning of utterances (Franke 2009). If the prior probabilities of states are uniform, so that $P(t) = P(t')$ for all $t, t' \in T$ (an assumption we make throughout), equation 6 can be rewritten to show the two Gricean constraints of truthfulness and informativity at play:\(^5\)

\begin{equation}
(7) \quad P_S(u \mid t; \alpha) \propto [P(t \mid [u])]^\alpha = \left[ \frac{\delta_{[t] \subseteq [u]}}{[u]^{-1}} \right]^\alpha
\end{equation}

This reformulation makes clear that, as long as there is any true utterance in $t$, any false utterance has zero probability of being selected in $t$ (since $\delta_{[t] \subseteq [u]} = 0$). Between two true utterances (for which the $\delta$-term evaluates to 1), the utterance with a stronger semantic meaning—that is, the one that is true in fewer world states—will be chosen with a higher probability. The bigger $\alpha$, the more pronounced this preference for informative utterances is.

For example, RSA’s production rule predicts that a speaker would choose utterances to communicate state $[u]$ with the following probabilities when we set $\alpha = 5$ (an arbit-

\(^2\) The vanilla RSA model and all other models considered here assume that the speaker knows the true world state, an assumption that is arguably warranted by the design of the experiment described in §4. See Goodman & Stuhlmüller 2013 and Herbstritt & Franke 2019 for extensions to cases with (higher-order) uncertain speakers.

\(^3\) Here, $\delta_{\text{boolean}}$ is the delta function that 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 speaker’s choice of expression, defined in equation 6, is derived from two assumptions (see e.g. Franke & Jäger 2016 for details). First, the speaker seeks to minimize the distance (in terms of Kullback-Leibler divergence) between their probabilistic belief and that of the literal listener. Second, the speaker tends toward optimal choices probabilistically (modeled with a so-called softmax function with rationality parameter $\alpha$).

\(^5\) Equation 6 can be rewritten like so: $P_S(u \mid t; \alpha) \propto \exp(\alpha \cdot \log P_{\text{LL}}(t \mid u)) = [P_{\text{LL}}(t \mid u)]^\alpha$, which by definition of the literal listener expands to: $\frac{P(t) \delta_{[t] \subseteq [u]}}{\sum_{t'} P(t') \delta_{[t'] \subseteq [u]}}$. Since $P(t) = P(t')$ for all $t, t'$ the prior term cancels out and we retrieve: $\left[ \frac{\delta_{[t] \subseteq [u]}}{[u]^{-1}} \right]^\alpha$.\(^3\)
trary value chosen here only for illustration; §4’s model comparison does not hinge on specific values for \( \alpha \).

\[
\text{Table 1: Predictions of the production rules of different models for parameter value } \alpha = 5.
\]

\[
\begin{array}{cccccccccc}
\text{model} & \text{NN} & \text{NS} & \text{NA} & \text{SN} & \text{SS} & \text{SA} & \text{AN} & \text{AS} & \text{AA} \\
\text{RSA} & 0.00 & 0.00 & 0.85 & 0.13 & 0.01 & 0.00 & 0.00 & 0.00 & 0.00 \\
\text{LU} & 0.00 & 0.00 & 0.85 & 0.13 & 0.01 & 0.00 & 0.00 & 0.00 & 0.00 \\
\text{LI} & 0.00 & 0.00 & 0.85 & 0.13 & 0.01 & 0.00 & 0.00 & 0.00 & 0.00 \\
\text{GI} & 0.00 & 0.00 & 0.85 & 0.13 & 0.01 & 0.00 & 0.00 & 0.00 & 0.00 \\
\end{array}
\]

In words, RSA predicts that speakers would produce only descriptions that are literally true in state \(\mathbb{1}^{\text{true}}\), namely ‘NA’, ‘SN’, or ‘SS’ (see Table 1). Among these, the speaker’s choice probabilities reflect the semantic strength of the utterances (\(\|\text{NA}\| = 3\), \(\|\text{SN}\| = 4\), and \(\|\text{SS}\| = 6\)). Further example predictions for the RSA speaker rule with \(\alpha = 5\) are shown in Figure 3 below.

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 truthfulness and informativity. Concretely, for each utterance \(u\), the rule \(P_L(t \mid u; \alpha)\) assigns a probability to each interpretation \(t\) based on the prior probability \(P(t)\) of a state and the likelihood \(P_S(u \mid t; \alpha)\) that a (truthful and informative) speaker would use the observed utterance for this state, following Bayes’s rule.

\[
P_L(t \mid u; \alpha) \propto P(t) \cdot P_S(u \mid t; \alpha)
\]

For example, consider the pragmatic listener’s belief after hearing ‘SS’, when fixing \(\alpha = 5\) for illustration.

\[
\text{Table 2: Predictions of the comprehension rules of different models for parameter value } \alpha = 5.
\]

\[
\begin{array}{cccccccccc}
\text{model} & \text{NN} & \text{NS} & \text{NA} & \text{SN} & \text{SS} & \text{SA} & \text{AN} & \text{AS} & \text{AA} \\
\text{RSA} & 0.00 & 0.00 & 0.85 & 0.13 & 0.01 & 0.00 & 0.00 & 0.00 & 0.00 \\
\text{LU} & 0.00 & 0.00 & 0.85 & 0.13 & 0.01 & 0.00 & 0.00 & 0.00 & 0.00 \\
\text{LI} & 0.00 & 0.00 & 0.85 & 0.13 & 0.01 & 0.00 & 0.00 & 0.00 & 0.00 \\
\text{GI} & 0.00 & 0.00 & 0.85 & 0.13 & 0.01 & 0.00 & 0.00 & 0.00 & 0.00 \\
\end{array}
\]

The sentence ‘SS’ is literally false in state \(\mathbb{1}^{\text{true}}\), so this state receives zero probability in the pragmatic listener’s interpretation. But the other states, except for \(\mathbb{1}^{\text{false}}\) and \(\mathbb{1}^{\text{true}}\), also receive (almost) zero probability, because the pragmatic listener would expect the speaker to choose a different expression in these states with much higher likelihood (see Fig. 3).

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, & 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 *u* and *u* can be interpreted in such a way that either every occurrence of *some* means ‘some and maybe all’ or every occurrence means ‘some but not all’, which pair (*t*, *l*) of a state *t* and mental lexicon *l* is most likely to have caused this speaker to have produced *u*?

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 *Exh*-operators. With slight abuse of notation, however, since *some* is the only Aristotelian quantifier susceptible to local meaning change by an *Exh*-operator, we can think of a speaker who considers *some* to mean ‘some but not all’ as a speaker who always, inflexibly, uses parse OI (see Table 1). Similarly, a speaker with a literal meaning for ‘some’ can be thought of as using the parse lit everywhere.

The LU model has listeners reason about the speaker’s mental lexicon *l*, represented here as parses *l* ∈ {lit, OI}. Crucially, any given speaker is assumed to have one of the two mental lexica, but not both (see Fig. 2). The speaker selects utterances *u* by the same mechanism as in the RSA model, but based on a semantic interpretation of utterances influenced by the speaker’s lexicon *l*.

\[
\text{(11)} \quad P_{S_1}(m \mid t, l; \alpha) \propto [P(t \mid \llbracket m \rrbracket)]^a
\]

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

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’s rule.

\[
\text{(12)} \quad P_{L_1}(t, l \mid m; \alpha) \propto P(t) \cdot P(l) \cdot P_{S_1}(m \mid t, l; \alpha)
\]

The prior *P(l)* over lexica *l* is here assumed to be uniform.

Equation 13 defines the listener’s beliefs for each pair of state *t* and mental lexicon *l*. From this, we can derive the listener’s beliefs about only the state *t*, by taking the weighted sum over the mental lexica (so-called ‘marginalization’).

\[
\text{(13)} \quad P_{L_1}(t \mid m; \alpha) = \sum_l P_{L_1}(t, l \mid m; \alpha)
\]

While Potts et al. (2016) use this latter formula to explain data from a truth-value judgment task, this paper also considers data from an experimental task that 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 makes predictions only for a given lexicon *l*. We therefore follow Lassiter and Goodman (2017) and define a speaker *S_2* who reasons about pragmatically adequate utterance choice based on the state-interpretation of *L_1*. Finally, a pragmatic listener *L_2* who reasons about the latter speaker’s choice of utterances is defined as before in terms of Bayes’s rule. Consequently, the final definitions of the LU models’ production and comprehension rules are as follows.

\[
\text{(14) a.} \quad P_{S_2}(m \mid t; \alpha) \propto [P_{L_1}(t \mid m; \alpha)]^a
\]

\[
\text{b.} \quad P_{L_2}(t \mid m; \alpha) \propto P(t) \cdot P_{S_2}(m \mid t; \alpha)
\]

Figure 3 above shows the predictions of the LU model’s production rule (for *α* = 5). These clearly differ from those of the vanilla RSA model in several places. This also affects interpretation of utterances. For example, the sentence ‘SS’ is interpreted as follows.
The most likely interpretation of ‘SS’ is \( \text{[\text{lit}]} \), which is a maximally informative utterance \( \text{[\text{lit}]} \) for a speaker with lexicon OI.

### 3.3. The Lexical Intentions Model

The LU model introduced in the previous section 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 (see Fig. 2). Consequently, the LU model considers two possible meanings for the sentence ‘SS’: the standard literal reading \( \text{[\text{lit}]} \) and the reading \( \text{[\text{OI}]} \) in 16.

(16) Some but not all of the aliens drank some but not all of their water.  \( \text{[\text{SS}]} \)

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 17a or 17b, which correspond to readings \( \text{[\text{I}]} \) and \( \text{[\text{O}]} \), respectively.

(17) a. Some (and maybe all) of the aliens drank some but not all of their water.  \( \text{[\text{SS}]} \)

b. Some but not all of the aliens drank some (and maybe all) of their water.  \( \text{[\text{SS}]} \)

The LU model does not contain speakers of this kind. To model speakers that associate ‘SS’ with any of the readings in 17, it is not enough to just include parses I and O as additional values for the variable \( l \) in the LU model. The result would be conceptually highly implausible. By equation 11, the speaker has a fixed lexicon \( l \) and invariably applies it to interpret whatever sentence comes along. That does make sense when \( l \) is instantiated with parses lit and OI, 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 O, equation 11 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.

Alternatively, we can treat parses as a choice of the speaker. 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 choosing utterances and lexical intentions (see Fig. 2).

The lexical intentions (LI) model formalizes these intuitions. It resembles the LU model superficially, but is conceptually different and also simpler. We assume that \( l \in \{\text{lit, I, O, OI}\} \) and define production and comprehension rules.

(18) a. \( P_S(m, l | t; \alpha) \propto [P(t | m)]^l \)  

b. \( P_L(t, l | m; \alpha) \propto P(t) \cdot P_S(m, l | t; \alpha) \)

As before, we then take weighted sums to retrieve predictions of utterance and interpretation choice probabilities.

(19) a. \( P_S(m | t; \alpha) = \sum_l P_S(m, l | t; \alpha) \)

b. \( P_L(t | m; \alpha) = \sum_l P_L(t, l | m; \alpha) \)
The main difference between equations 11 and 18 is the position of the lexicon parameter \( l \). In the LI model the lexical meaning \( l \) is treated not as an argument to be passed into the speaker function, like in the LU model, but as an output of it. Conceptually, this means that the LI model does 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 \( \langle u, l \rangle \) as a description of \( t \) only if \( t \in \llbracket u \rrbracket \), and they make this choice with a probability proportional to the relative informativity of \( \llbracket u \rrbracket \); that is, they prefer pairs \( \langle u, l \rangle \) for which \( \llbracket u \rrbracket \) is small.\(^6\)

The predictions of LI are subtly different from those of the two previous models (see Fig. 3). For example, the LI model predicts that the listener’s interpretation for the sentence ‘SS’ is as in 20.

\[
\begin{align*}
P_L(\cdot | 'SS'; \alpha = 5) &= 0.00 \quad 0.31 \quad 0.15 \quad 0.38 \quad 0.00 \quad 0.16 \quad 0.00
\end{align*}
\]

According to the LI model, ‘SS’ should be associated with interpretation \( \llbracket u \rrbracket \) much more readily than predicted by vanilla RSA or the LU model.

### 3.4. The Global Intentions 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 that make them true and informative. From here it is only a very small step toward a model that includes all readings supplied by grammaticalism. The global intentions (GI) model is exactly like the LI model, but integrates all parses \( p \) considered in Table 1 (see Fig. 2).

\[
\begin{align*}
(21) \ a. \ P_S(m, p | t; \alpha) &\propto [P(t | \llbracket m \rrbracket)]^a \\
(21) \ b. \ P_L(t, p | m; \alpha) &\propto P(t) \cdot P_S(m, p | t; \alpha)
\end{align*}
\]

The GI 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 not directly engage with or adopt grammaticalism. The GI model, by contrast, does. Including the full set of grammaticalist readings into the LU model makes little conceptual sense for the same reasons laid out in the context of the LI model: it would be quite unnatural to imagine speakers who cannot rise above their language’s ambiguity, invariably applying a particular pattern of insertions of *Exh*-operators, irrespective of the sentence under consideration and the state to be described. The GI model, instead, models speakers who are masters of the rich ambiguities provided by their language’s grammar, using these ambiguities to flexibly intend to express this reading or another, depending on what serves communication.

The GI model also makes distinct empirical predictions (Fig. 3). For example, the speaker is predicted to use the sentence ‘SS’ with a very high probability in state \( \llbracket u \rrbracket \). This is because the reading \( \llbracket SS \rrbracket \) 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 \( \llbracket u \rrbracket \).

\[
\begin{align*}
P_L(\cdot | 'SS'; \alpha = 5) &= 0.00 \quad 0.70 \quad 0.05 \quad 0.21 \quad 0.00 \quad 0.03 \quad 0.00
\end{align*}
\]

\(^6\) Notice that the LI model, and the following GI model, assumes that the literal listener ‘miraculously’ recognizes the speaker’s intended meaning. Since the literal listener is a technical construct, the implication is that speakers are not assumed to take into account a realistic model of their interlocutor. Speakers are modeled as using a (myopic, conventional) heuristic manner of choosing utterances and intended meanings, which, however, can then be recognized by pragmatic listeners to retrieve the speaker’s intended meaning.
4. Experiment.

4.1. Design. The models just introduced differ mostly in their quantitative predictions about the 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 probe these model predictions directly. We refer to these tasks as a production and an interpretation task.

4.2. Participants. One hundred 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.

4.3. 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 avoid typicality effects (e.g. Degen & Tanenhaus 2015, van Tiel 2014). The seven relevant states were displayed using pictures as in Fig. 1. All pictures contained twelve aliens with full, half-full, or empty glasses. 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 & Tanenhaus 2015, van Tiel 2014).

4.4. 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 4a). 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 4b). Participants rated how likely they thought it was that the situation displayed is what the speaker had observed.

Figure 4. Illustrations of experimental trials. (NB: in interpretation trials the pictures were actually shown on the left of the text and slider bar, but are here shown vertically stacked for ease of presentation.)
4.5. Results. Figure 5 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 utterance in state -NN-, which seems to conflict with the predictions of the RSA model plotted in Table 1. Finally, we see that the most frequently used utterance in state -NN- is ‘SS’, which goes against the predictions of the RSA and the LU model plotted in Table 1, but is consistent with the GI model.

Each trial in the comprehension task returns nine slider ratings, one for each state. We normalized each of these nine-placed vectors. Figure 6 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 -NN-, followed by state -NS-, and finally -NA-, thereby replicating previous results on interpretation preferences for these sentences (Chemla & Spector 2011, Franke, Schlotterbeck, & Augurzky 2017). Finally, the interpretation of the sentence ‘SS’ seems inconsistent with the interpretation predicted by the RSA model (for α= 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 §3, we look at Bayes factors (Jeffreys 1961, Kass & Raftery 1995). From a Bayesian point of view, a model $M$ consists of a prior $P(\theta | M)$ over vectors $\theta$ of values for its parameters and a likelihood function $P(D | \theta, M)$, 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 priori for any parameter value.

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

The Bayes factor in favor of model $M_1$ over $M_2$ is the ratio of marginalized likelihoods: $P(D | M_1) / P(D | M_2)$. 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’s rule:

\[
\frac{P(M_1 | D)}{P(M_2 | D)} = \frac{P(D | M_1) P(M_1)}{P(D | M_2) P(M_2)}
\]
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 their conjunction: $P(D \mid \theta, M) = P(D_p \mid \theta, M) \cdot P(D_c \mid \theta, M)$. The following explains how the models defined in §3 give rise to a likelihood function for $D_p$ and $D_c$, respectively.

The production rule of model $M$ defines the likelihood $P_S(u \mid t, \theta, M)$ of a single choice of expression for a state. All of the models from §3 use a single parameter $\alpha$. 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 utterances that are false for a given state. However, we do observe false utterance choices (see Fig. 5). We therefore add a constant error term $\epsilon$ to all predicted utterance choice probabilities. Second, since no model accounts by itself for the low choice rates of sentences starting with none, we include a cost term $c(u)$, which is fixed to zero for utterances starting with some or all, but may be positive for utterances 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, for example, for vanilla RSA we obtain the following.

\[
(25) \quad P_S(u \mid t; \alpha, \epsilon, c) \propto [P_{LL}(t \mid u) - c(u)]^{\alpha} + \epsilon
\]

Finally, if $D_p$ consists of counts $n_{ij}$ of the number of times utterance $u_j$ was chosen in state $t_i$, the likelihood $P(D_p \mid 0, M)$ is as follows.

\[
(26) \quad P(D_p \mid 0, M) \propto \prod_{i=1}^{7} \text{Multinomial} \left( \langle n_{i1}, \ldots, n_{i9} \rangle, \langle P_S(m_1 \mid t_i; 0, M), \ldots, P_S(m_9 \mid t_i; 0, M) \rangle \right).
\]

Comprehension rules give a probability distribution over states for each given utterance: $P_L(t \mid u, 0, M)$. 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 utterance \( i \) (as plotted in Fig. 6). We think of the probability vector \( \langle c_1, \ldots, c_7 \rangle \) as a sample from a Dirichlet distribution whose modal value is the model’s prediction \( \langle P_L(t_1 \mid u_i, 0, M), \ldots, P_L(t_7 \mid u_i, 0, M) \rangle \). 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.

\[
(27) \quad P(D_c \mid \theta, w, M) \propto \prod_{i=1}^{9} \text{Dirichlet}(\langle c_{i1}, \ldots, c_{i7} \rangle, w \cdot \langle P_L(t_1 \mid u_i, 0, M), \ldots, P_L(t_7 \mid u_i, 0, M) \rangle)
\]

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.

\[
(28) \quad a \sim \text{Uniform}(0, 6) \quad c \sim \text{Uniform}(0, 5) \quad w \sim \text{Uniform}(10, 80) \quad \epsilon = 0.045
\]

Notice that all models fix the error parameter \( \epsilon \) to a 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 twenty for each parameter (e.g. Kruschke 2015:Ch. 10). Figure 7a 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 GI. The Bayes factor in favor of GI 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 modeler’s posterior beliefs after conditioning unbiased priors over models with the observed data, as in Table 2.

<table>
<thead>
<tr>
<th>MODEL ( M_i )</th>
<th>( P(M_i \mid D) )</th>
</tr>
</thead>
<tbody>
<tr>
<td>( M_{\text{RSA}} )</td>
<td>0.01</td>
</tr>
<tr>
<td>( M_{\text{LU}} )</td>
<td>0.033</td>
</tr>
<tr>
<td>( M_{\text{LI}} )</td>
<td>0.956</td>
</tr>
</tbody>
</table>

Table 2. Approximations of posterior beliefs.

In words, if each model is a priori equally likely, the posterior probability, after seeing the data, of the GI model is around 0.956. The combined data from production and comprehension provide strong evidence in favor of the GI 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 7b 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 GI model is about 30,600; the LI model is absolutely no competition. If we look at comprehension data only, the Bayes factor in favor of the GI model is ca. 0.011; that is, the LI model is roughly 100 times more likely a posteriori, if we start from unbiased prior beliefs. Consequently, the main advantage of the GI model comes from its superior predictions for the production data, which outweigh its weaker predictions for comprehension.
Zooming in to look at each individual condition from production (i.e. the utterance choices at a single state) and comprehension (i.e. the average slider ratings for a single sentence), Figure 7c shows that the main evidence in favor of the GI model comes from its superior predictions for the production data in state . Figure 8 plots the prior predictive distributions for production data under both models. The prior predictive distribution gives the 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 the choice of ‘SN’ for condition . The LI model makes better predictions for states and , but, as becomes apparent from Fig. 8, its predictive advantage over the GI model is less pronounced than GI’s advantage in condition . The reason why the GI model predicts a high frequency of ‘SS’ choices in state is that it makes it possible that a rational speaker utters ‘SS’ while intending a global reading. The global reading with M is not available to the LI model, but it serves to perfectly single out state from all other states (see Table 1), so it is rational for a speaker to say ‘SS’ and mean .

6. Conclusions. We addressed the general question of how multiple ambiguity affects a speaker’s pragmatic choice of utterance and a listener’s interpretation. Focusing on implicature readings of complex sentences, we showcased theory-driven statistical modeling as a method of directly linking linguistic theory to empirical data. We formulated the ‘global intentions’ model, which assigns a conceptually plausible role to the full richness of meanings generated by grammatical approaches to scalar implicature inside of a Gricean model of probabilistic pragmatic reasoning (see Fig. 2). The GI model is not an innocuous synthesis of traditional Gricean and grammatical approaches to scalar implicatures. It is a new breed of its own. It puts general pragmatic reasoning first and
embeds grammaticalism as one potential mechanism of generating conventionally associated readings for sentences. It thereby also provides a possible link hypothesis, that is, a way of deriving precise quantitative predictions for empirical data for a wide range of cases. Statistical model comparison in terms of Bayes factors indeed suggests that the GI model might be empirically superior to the investigated alternatives.

Theory-driven statistical modeling of the kind executed here can inform linguistic theory in multiple ways. We might, for example, fix the GI model but compare different instantiations of grammaticalism embedded inside of it. Consider Gotzner and Romoli’s (2018) proposal of a different construction of the set of sentential alternatives $\text{Alt}(S, p)$ than used here. Without going into the details of their account, the Bayes factor in favor of a GI model based on the simpler construction used here (see the appendix), when com-

**Figure 8.** Prior predictive distribution for the production data. Black vertical bars indicate the observed counts for each condition.
pared to Gotzner and Romoli’s more elaborate alternative, is about 1,530, suggesting that
the simpler construction is empirically superior. This is not meant to be a decisive argu-
ment against Gotzner and Romoli’s approach, but an example of how theory-driven mod-
eling opens up new ways to quantitatively compare elaborate and fine-grained theoretical
proposals in the light of experimental data.

The main reason for the GI model’s predictive success was identified as the avail-
ability of a reading \([SS]^M = \{\text{reading}\}\) of the sentence ‘SS’ obtained from inserting an Exh-
operator at matrix position. This provides suggestive evidence for the idea that the full
set of grammatically induced readings is needed. Indeed, the vanilla RSA model does
not associate ‘SS’ with \(\text{reading}\), because ‘SN’ is always a better choice in state \(\text{reading}\) than
‘SS’ for the speaker. This highlights an important implication of the GI model: it uses
two different types of sentential alternatives. On the one hand, there are what we could
call grammatical sentential alternatives, which are used to derive the grammat-
ically derived readings of candidate utterances, as detailed in the appendix. For exam-
ple, by common construction ‘SN’ is not a grammatical alternative to ‘SS’, so that the
reading \([SS]^M = \{\text{reading}\}\) can be derived. On the other hand, there is the set of utter-
ance alternatives, which we here equated, naturally, with the set of utterances of-
fered in the experimental design (exactly like Potts et al. 2016). The set of utterance
alternatives to ‘SS’ does include ‘SN’. It is not necessary to equate these two sets of al-
ternatives, because it is conceivable, for example, that the grammatical module that
generates readings for sentences is (fairly) encapsulated from context. Utterance alter-
natives might be much more flexibly adapted in the context, as they provide the set of
actual utterances the speaker is currently aware of in their choice. Future work should
explore the potential generated by this distinction of grammatical vs. utterance alterna-
tives further.

Future work should also scrutinize different models of pragmatic reasoning. For ex-
ample, all of the models considered here assume that speakers do not have a veridical
view of the listener’s interpretation; they base their choices on a simple Gricean heuris-
tic of preferring semantically stronger utterances; they do not engage actively in audi-
ence design. Clearly, exploring other models of pragmatic reasoning in connection with
grammatically generated implicature readings, and testing them in experimental de-
signs that take the possibility of active audience design into account, is a worthwhile
enterprise for future research.

Although our focus here was conceptual and methodological, more work is clearly
necessary on the empirical side as well. The scope of our experimental manipulations
is admittedly rather small. Recent experimental papers on scalar implicatures (e.g.
Chemla & Spector 2011, Franke, Schlotterbeck, & Augurzky 2017) in complex sen-
tences have focused attention on scalar-implicature triggers in nonmonotonic environ-
ments like in 29.

(29) 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 se-
lection. For instance, the modeling of Potts et al. (2016) implicitly assumes that listen-
ers also include 29 as a speaker’s alternative utterance during the interpretation of any
nested Aristotelian. To sidestep such controversial assumptions, 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.
When \( \text{Exh} \) applies directly to a quantifier \( Q \), the resulting meaning \( \text{Exh}(Q) \) is that of \( Q \) conjoined with the negation of all lexical alternatives that are strictly stronger, that is, which entail \( Q \) but are not entailed by it. As usual, we assume that some and all are lexical alternatives of each other. We also assume, following standard practice, that none and not all are, since we analyze none as not some (e.g. Levinson 2000:80). Consequently, the effect of applying \( \text{Exh} \) to \( Q \) will be vacuous for none and all, but \( \text{Exh}(\text{some}) \) gives an in-situ enrichment to ‘some but not all’. For example, under a parse I the sentence ‘SS’ gets the reading \([SS]\) in A1 that will be true in any state whose pictorial representation includes a half-full glass (see Table 1).

(A1) Some of the aliens drank some but not all of their water.

When \( \text{Exh} \) applies in matrix position to a sentence \( S \), it takes the meaning of \( S \) obtained from insertions of \( \text{Exh} \) 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 \( \text{Exh} \) in matrix position. When \( \text{Exh} \) applies in matrix position to \( S \) whose parse is \( p \), the meaning of \( \text{Exh}(S, p) \) is obtained by conjoining \( [S]_{p|M} \) of \( S \) with the negation of all relevant sentential alternatives \( \text{Alt}(S, p) \) of \( S \), if this operation is noncontradictory (but see e.g. Gotzner & Romoli 2018 for discussion of more elaborate definitions).

(A2) \( [\text{Exh}(S, p)] = \left[ [S]_{p|M} \cap \text{Exh}(S, p) \right] [\text{not } S]_p \) if noncontradictory

We here assume that the set \( \text{Alt}(S, p) \) is obtained in two steps, a generation and a filtering step. First, we generate a set \( \text{Alt}^r(S) \) of potential alternatives by blindly replacing any occurrence of a scalar item in \( S \) with any of its lexical alternatives, as defined above. A second step then filters out sentences from \( \text{Alt}^r(S) \) that do not entail \( S \).

(A3) a. \( \text{Alt}^r(S) = \{ S' \mid S' \text{ is derived from } S \text{ by replacements of lexical alternatives} \} \)

b. \( \text{Alt}(S, p) = \{ S' \in \text{Alt}^r(S, p) \mid [S']_p \subset [S]_{p|M} \} \)

For example, consider the sentence ‘SS’ under a parse I. By replacing occurrences of some with its lexical alternatives, we obtain \( \text{Alt}^r(\text{SS}) = \{ \text{SS, AS, SA, AA} \} \). Of these, ‘SS’ is filtered out. Applying matrix-\( \text{Exh} \) consequently yields the conjunction of the negation of all sentences ‘AS’, ‘SA’, and ‘AA’ with the literal meaning of ‘SS’, as in A4. This is a very strong reading, as the only world state that makes this reading true is \( \square \).

(A4) 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 MI. We first consider \([SS]\), which gives the set of situations in which there is at least one half-full glass. The set \( \text{Alt}^r(\text{SS}) \) is as before, but \( \text{Alt}(\text{SS}, p) \) is now empty because none of the sentences in \( \text{Alt}^r(\text{SS}) \) are, under parse lit, strictly stronger than ‘SS’ under parse I. Consequently, additional insertion of \( \text{Exh} \) in matrix position is vacuous in this case.

REFERENCES


7 The only effect of considering not all a lexical alternative to none is that the sentence ‘NN’ gets an additional reading when \( \text{Exh} \) applies in matrix position. But this has no noteworthy effect on anything of relevance to this paper’s main concerns.


[mchfranke@gmail.com] [lbergen@ucsd.edu] [Received 12 March 2018; revision invited 3 June 2018; revision received 1 March 2019; revision invited 1 June 2019; revision received 9 October 2019; accepted 16 October 2019]