

Embedded implicatures as pragmatic inferences under compositional lexical uncertainty*

Christopher Potts, Daniel Lassiter, Roger Levy, and Michael C. Frank

January 3, 2015

Abstract

How do comprehenders reason about pragmatically ambiguous scalar terms like *some* in complex syntactic contexts? In Gricean theories of conversational implicature, local exhaustification of such terms (‘only some’) is predicted to be impossible if it does not strengthen the literal meaning, whereas grammatical accounts of implicature predict such construals to be available. Recent experimental evidence supports the availability of these local enrichments, but the grammatical theories that this evidence supports do not provide viable mechanisms for weighting such construals against others. We propose a probabilistic model that combines previous work on pragmatic inference under ‘lexical uncertainty’ with a more detailed model of compositional semantics. We show that this model makes accurate predictions about new experimental data on embedded implicatures in both non-monotonic and downward-entailing semantic contexts. In addition, the model’s predictions can be improved by the incorporation of neo-Gricean hypotheses about lexical alternatives. This work thus contributes to a synthesis of grammatical and probabilistic views on pragmatic inference.

1 Conversational implicature: Interacting with grammar

The linguistic forms that discourse participants exchange with each other routinely underrepresent the speaker’s intended message and underdetermine the listener’s inferences. Grice (1975) famously provided a philosophical framework for understanding the driving forces behind such pragmatic enrichment. At the heart of this framework are **conversational implicatures**: social, cognitively complex meanings that discourse participants create jointly in interaction.

Perhaps the best-studied examples of language users going beyond the literal semantics involve weak terms like *some* being strengthened to exclude their communicatively stronger alternatives, giving rise to construals like ‘some and not all’ or ‘only some’. Such inferences are often called **scalar conversational implicatures** (SIs), and they are widely assumed to arise via the same social inferencing mechanisms that are at work in other implicatures. However, this assumption has always been controversial. Even Grice suggested that SIs might be closer to the grammar than other implicatures (p. 56; see also Levinson 2000; Sperber & Wilson 1995; Bach 2006), and recent grammar-driven accounts are framed in direct opposition to an implicature analysis. For

*All the data and code used in this paper are available at <https://github.com/cgpotts/pypragmods>

example, Chierchia et al. (2012: 2316) write, “the facts suggest that SIs are not pragmatic in nature but arise, instead, as a consequence of semantic or syntactic mechanisms”. The ensuing debates have stimulated new insights, pushing researchers to identify and evaluate previously unnoticed consequences of the two broad positions.

Much of the debate between Gricean and grammar-driven accounts has centered around what we informally called **embedded implicatures** — cases where a pragmatically enriched interpretation seems to be incorporated into the compositional semantics. Such readings seem initially to demand implicature-enriched semantic representations. However, many of the relevant examples have received straightforward Gricean accounts in which semantic content and contextual assumptions interact to yield global implicatures that are meaning-equivalent to interpretations that would derive from local pragmatic enrichment (Russell 2006; Geurts 2009). This reduces the power of such examples to decide in favor of one side or the other.

Geurts & Pouscoulous (2009) and Chemla & Spector (2011) study a wide range of embedded implicatures involving scalar terms in the scope of quantified phrases. They show that many of these listener inferences are amenable to Gricean treatments with no need for the pragmatics to intrude on the semantics. However, they identify a class of examples that will not admit such a treatment: scalar terms in the scope of non-monotone quantifiers, as in *Exactly one player hit some of his shots*. In such cases, the interpretation enriched by an embedded implicature (*... some but not all of his shots*) does not entail the literal meaning, whereas the Gricean analysis of scalar terms can only strengthen literal meanings. Geurts & Pouscoulous’s experiments fail to support enrichment in such contexts, whereas Chemla & Spector’s suggest that it is possible.

In this paper, we reproduce the central qualitative result of Chemla & Spector (2011) using more naturalistic experimental stimuli and a more direct method of interpreting participants’ responses. Like Chemla & Spector, we find that scalar terms in non-monotone environments support implicature inferences. We hope that these results bolster Chemla & Spector’s conclusions and help to address the skeptical reactions of Geurts & van Tiel (2013). In our view, this evidence points to a role for compositional semantics in understanding implicatures. But pragmatic factors also play a crucial role: even embedded implicatures arise only with proper contextual support (Russell 2006; Chierchia et al. 2012; Geurts 2011; Chemla & Spector 2011).

To describe the complementary roles of grammar and pragmatics in embedded implicatures, we propose a model that both embraces the compositional insights of Chierchia et al. and characterizes how people arrive at such construals. This model is in the tradition of **rational speech act** models (Frank & Goodman 2012; Goodman & Stuhlmüller 2013) and **iterated best response** models (Franke 2009; Jäger 2012), and is a direct extension of the **compositional lexical uncertainty** model of Bergen et al. (2012) and Bergen et al. (2014). The model accounts for how discourse participants coordinate on the right logical forms (implicature-rich or not), seeking to retain the insights of Gricean accounts while paying close attention to the details of semantic composition.

We show that our model not only captures the qualitative pattern of implicature behaviors that Chemla & Spector found, but also makes quantitative predictions that are highly correlated with people’s actual inferential behavior in context. In addition, we show that these correlations can be improved if the set of refinements is lexically constrained, in keeping with broadly neo-Gricean views of scalar implicature (Horn 1972; Gazdar 1979a,b; Schulz & van Rooij 2006). Our results suggest that the full theory of implicature depends substantively on the fine details of semantic composition *and* broader considerations of rational interaction. This is perhaps a departure from Grice’s (1975) particular conception of pragmatic meaning, but it is well-aligned with his general

theory of meaning and intention (Grice 1989). In view of our experimental results, the chief advantage of our model is that it makes quantitative predictions that are easily and rigorously linked with our human response patterns. In other words, the model makes predictions not only about which pragmatic inferences are possible but also about how likely those inferences are.

Our broader position is that grammar-driven accounts and Gricean accounts are not in opposition, but rather offer complementary insights. When communicating in natural languages, people are relying on linguistic conventions to try to identify and convey each other’s intentions. All sides in the debate acknowledge this mix of grammatical and interactional factors. Grice’s (1975) definition of conversational implicature is interactional, but his maxim of manner embraces a role for language. By introducing additional devices such as Horn scales, Neo-Griceans expand this role into areas Grice addressed with the maxims of quantity, quality, and relevance. Sperber & Wilson (1995) and Bach (1994) characterize many kinds of pragmatic enrichment as inferences about logical forms. And Chierchia et al. (2012) invoke broadly Gricean pressures to explain how speakers and listeners coordinate on whether to posit implicature-rich logical forms or more literal ones. Thus, there is substantially more consensus than the rhetoric often suggests.

2 Implicature, enrichment, and embedding

In this section, we describe embedded implicatures, seeking to identify their relevance for Gricean and grammar-driven frameworks. Under Grice’s (1975) original definition, conversational implicature is an act of social cognition. The original definition is somewhat underspecified, and fleshing it out into a precise formulation is challenging (Hirschberg 1985), but the guiding idea seems clear. The listener assumes that the speaker is cooperative in the Gricean sense of rational interaction. However, the listener is confronted with an utterance U with content p that meets this assumption only if certain additional conditions are met. The listener can resolve this tension by positing that these conditions are in fact met; in many (but not all) cases, this means inferring that the speaker intended for the listener to infer the truth of an additional proposition q , or of a different but related proposition q' . By this reasoning, the listener is able to reconcile the observation that the speaker chose to utter U with the assumption that the speaker is communicating cooperatively.

In the current work, we do not try to make the above description more rigorous. The model that we develop does not depend on an independently formulated definition of implicature, but rather seeks to derive such meanings from more basic considerations about how speakers and listeners reason about each other whenever they interact. Similarly, the model of Chierchia et al. (2012) is noncommittal about the reality of conversational implicatures per se. For them, ‘conversational implicature’ can be seen as an informal label for a certain class of logical forms, rather than a conceptual primitive (see section 3 of this paper). With this in mind, we use the notion of conversational implicature only to articulate the central empirical focus of this paper — embedded scalar terms — and the associated challenges for Gricean accounts.

On the classic Gricean account, scalar implicatures arise when the imperative ‘Be as informative as is required’ (a subclause of the maxim of quantity) is in tension with another pragmatic pressure related to cooperative communication. The opposing force can take many forms, for example, relating to considerations of politeness, discretion, or secrecy, but it is usually attributed to the maxim of quality, which instructs speakers to say only what they have strong positive evidence for. For instance, imagine a sportscaster who has observed the outcome of a single round of a bas-

ketball tournament and is reporting on it as news. If the sportscaster says (1), then she will likely implicate that Player A did not make all of his shots.

- (1) Player A hit some of his shots.

The scalar implicature follows from a straightforward application of the above ideas. We assume the sportscaster is cooperative in the Gricean sense, and knowledgeable and forthcoming about the events. Why, then, did she opt for a weak statement like *Player A hit some of his shots* when a stronger statement like *Player A hit all of his shots* is available and would have been more informative? If knowledge is the only relevant consideration, it must be that she was prevented from using this stronger form because she does not know it to be true. Together with our assumption that she observed the full outcome, she can lack knowledge of this proposition only because it is false, leading to the implicated meaning that Player A did not hit all of his shots. In this way, a listener can enrich the speaker's message.

To make this concrete, suppose that we have two players, A and B, and that we care (for present purposes) only about whether each of them hit none, some but not all, or all of his shots. We can identify these (equivalence classes of) possible worlds with labels like NA, which means that Player A hit none of his shots and Player B hit all of his shots, and SS, which means that both players hit some but not all of their shots. There are $3^2 = 9$ such worlds. The literal semantics of (1) in this context is the proposition given in (2b). Our hypothesized implicature is (2c), the proposition that Player A did not hit all of his shots. The intersection of these two meanings delivers the communicated meaning, (2d).

- | | | | |
|-----|------------------|----------------------------|-----------------|
| (2) | a. Worlds: | NN NS NA SN SS SA AN AS AA | |
| | b. Literal: | SN SS SA AN AS AA | 'at least some' |
| | c. Implicature: | NN NS NA SN SS SA | 'not all' |
| | d. Communicated: | SN SS SA | 'only some' |

There are many proposals for how to formalize this reasoning. The common theme running through all of them is that the implicature is accessible because it is a refinement that strictly entails the original literal content — in this example, because the utterance's literal meaning and the implicature are combined into an enriched meaning by intersection. In Grice's terms, a general claim is further restricted by the interaction of quantity and quality.

The above reasoning extends to examples like (3), in which *some* is in the scope of a universal quantifier, though additional assumptions must be brought in to achieve a comparable implicature.

- (3) Every player hit some of his shots.

In the same general context as before, this can be enriched to convey that every player hit some but not all of his shots. However, making this reasoning precise under a Gricean account requires more assumptions than for (1). If we take the implicature to be the negation of the stronger alternative *every player hit all of his shots*, then the reasoning proceeds as in the first four lines of (4), which takes us to a meaning (4d) that is consistent with one or the other of the players (but not both) having hit all of his shots. To arrive at the target meaning (every player hit some but not all of his shots), we must further assume an auxiliary premise beyond that required for (1). One example of such a premise is that of uniform outcomes (4e); there are many others that will do the job (Spector 2007b).

- | | | | |
|-----|------------------|----------------------------|---------------------------------|
| (4) | a. Worlds: | NN NS NA SN SS SA AN AS AA | |
| | b. Literal: | SS SA AS AA | ‘all hit at least some’ |
| | c. Implicature: | NN NS NA SN SS SA AN AS | ‘not all hit all’ |
| | d. Result: | SS SA AS | ‘all hit some; not all hit all’ |
| | e. Aux. premise: | NN SS AA | ‘uniform outcomes’ |
| | f. Communicated: | SS | ‘all hit only some’ |

Though the need for an auxiliary premise is a noteworthy complication, it seems within the bounds of a Gricean account, and auxiliary premises like these might be independently justified (Russell 2006). As in the previous example, the communicated meaning is an enrichment of the literal content, and Gricean pressures and contextual assumptions deliver the stronger meaning. Geurts & Pouscoulous (2009) and Chemla & Spector (2011) home in on this common theme in scalar implicature calculation and use it to probe the empirical adequacy of the Gricean framework. Examples like (5) drive their discussions. This is a minimal variant of (3) with the subject universal determiner *every* replaced by *exactly one*.

- (5) Exactly one player hit some of his shots.

Many people have the intuition that (5) can be used to describe a situation in which there is exactly one player who scored some but not all of his shots, which is consistent with some players having scored all of their shots. The reading is easy to characterize intuitively: one imagines that *some of his shots* has been locally enriched to *some but not all of his shots*, and that this enriched meaning is the semantic argument to the subject quantifier. What makes this reading notably different from, e.g., (3) is that it does not entail the literal reading, as we see in (6). The literal semantics is the proposition in (6b), whereas the content of the ... *some but not all of his shots* (‘Local’) construal is (6c), which merely overlaps with it.

- | | | | |
|-----|-------------|----------------------------|---------------------------------|
| (6) | a. Worlds: | NN NS NA SN SS SA AN AS AA | |
| | b. Literal: | NS NA SN AN | ‘exactly one hit at least some’ |
| | c. Local: | NS SN SA AS | ‘exactly one hit only some’ |

Any theory in which enriched scalar interpretations are always generated by intersection, as they are in classical Gricean and neo-Gricean accounts, will fail to arrive at (6c). Such theories head inexorably toward a refinement that excludes NA and AN, but they are essentially incapable of ‘introducing’ SA and AS.

The issue is even clearer when a scalar term is in the scope of a downward-monotone operator like *no*, as in *no player hit some of his shots*. In such cases, the embedded enrichment creates a meaning that is strictly entailed by (weaker than) the literal meaning:

- | | | | |
|-----|-------------|----------------------------|----------------------|
| (7) | a. Worlds: | NN NS NA SN SS SA AN AS AA | |
| | b. Literal: | NN | ‘none hit some’ |
| | c. Local: | NN NA AN AA | ‘none hit only some’ |

Gricean theories predict that the ‘Local’ reading is unavailable here, either because of the way pragmatic pressures interact or because *some* is held to be the strongest member of its scale in negative environments, leaving no room for further enrichment. Grammar-driven approaches have tended to agree with the basic empirical assumption, arguing that local enrichment is blocked in environments where it would strictly weaken the literal content (Chierchia 2006).

The empirical evidence is mixed but seems to support the accessibility of these local interpretations. Modifying an earlier design by Geurts & Poussoulous (2009), Chemla & Spector used displays involving geometric patterns to assess whether interpreters could access local-enrichment readings of scalar terms in the scope of non-monotone and downward-monotone operators. Their findings suggest that local enrichment readings are available in both contexts, especially non-monotone ones. Skeptics of local enrichment have found grounds for challenging Chemla & Spector’s findings (see section 5), but we believe that the theoretical challenges posed by embedded implicatures are real. In section 5, we describe a new experiment that reproduces the core findings of Chemla & Spector’s studies.

3 CFS’s grammar-driven model

This section briefly reviews the grammar-driven model of Chierchia et al. (2012) (henceforth CFS). The approach is inspired by those of Chierchia (2004), Sauerland (2001), Spector (2007a), and Fox (2007, 2009). There are two central pieces to the account: a generally available function *ALT* that maps words and phrases to their alternatives, and a covert exhaustification operator *O*.

For *ALT*, the relevant notion of alternative is familiar from theories of questions and focus (Groenendijk & Stokhof 1984; Rooth 1985, 1992): we can assume, as a default, that the alternatives for an expression φ is some subset of the items in the same type-theoretic denotation domain as $\llbracket \varphi \rrbracket$, the meaning of φ . The precise value of the function *ALT* is context-dependent, and discourse participants are presumed to coordinate on it, just as they coordinate on the meanings of deictic or discourse-bound pronouns, elided phrases, and other pragmatically controlled free variables.

The effect of applying the basic exhaustification operator *O* to an expression φ in the context of a given *ALT* is shown in (8) (Spector 2007a; Fox 2007, 2009; Magri 2009; Chierchia et al. 2012).¹

$$(8) \quad O_{ALT}(\varphi) = \llbracket \varphi \rrbracket \sqcap \bigwedge \{ -q : q \in ALT(\varphi) \wedge \llbracket \varphi \rrbracket \not\sqsubseteq q \}$$

The *O* operator maps an expression φ to one that entails $\llbracket \varphi \rrbracket$ and excludes the denotations of expressions in *ALT*(φ) that are strictly stronger than $\llbracket \varphi \rrbracket$. When dealing with truth-functional expressions, we can regard \sqcap as boolean conjunction and \sqsubseteq as a material conditional, but the definition should be thought of as broad enough to include any kind of partial ordering (Hirschberg 1985: §4).

Part of the case for a grammar-driven view is that it uses pieces of semantic theory that are independently needed. In particular, exhaustification is at the heart of Groenendijk & Stokhof’s (1984) theory of questions and their answers. The above operator is a common proposal for the meaning of *only* (for discussion: Rooth 1996; Büring & Hartmann 2001; Beaver & Clark 2008). Schulz & van Rooij (2006) use exhaustification for implicature calculation (see also de Jager & van Rooij 2007). (For critical discussion, see Alonso-Ovalle 2008 and Gajewski 2012.) While CFS are cautious about making direction connections between *O* and these other phenomena (p. 2304), the correspondences are nonetheless noteworthy.

Those are the technical pieces. The proposal can then be summarized easily: *O* operators can optionally appear anywhere in the logical form of a sentence, perhaps subject to additional restrictions and general preferences (see CFS: §4.6). To see the effects that this could have, let’s

¹This is not the operator that CFS ultimately favor, since it requires some implicit restrictions on allowable *ALT* functions in order to get the right inferences. The final version has the same form as (8) but further restricts *ALT*.

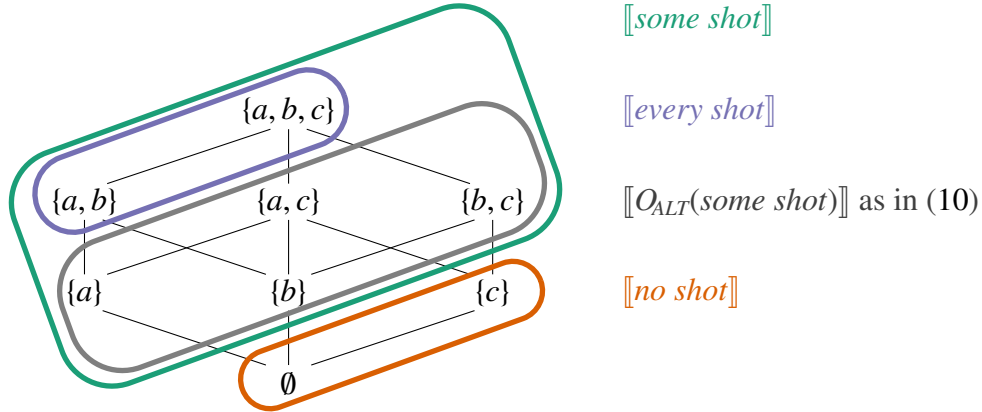


Figure 1: Given a domain $\{a, b, c\}$ with $\llbracket shot \rrbracket = \{a, b\}$, $\llbracket some\ shot \rrbracket$ is equal to the set of sets in the green box, $\llbracket every\ shot \rrbracket$ to the set of sets in the purple box, and $\llbracket no\ shot \rrbracket$ to the set of sets in the orange box. If $O_{ALT}(some\ shot)$ contains $\llbracket every\ shot \rrbracket$, then *some shot* is refined to exclude the purple subset.

return to the examples involving *some* that we reviewed in section 2. Simplifying slightly, let's suppose that *some shot* denotes the set of sets in (9) — the set of all sets Y that have a non-empty intersection with the set of shots.

$$(9) \quad \llbracket some\ shot \rrbracket = \{Y : \llbracket shot \rrbracket \cap Y \neq \emptyset\}$$

Consider a domain of three entities $\{a, b, c\}$, and assume that $\llbracket shot \rrbracket = \{a, b\}$. Then the above is equivalent to the set of sets contained in the green box in figure 1. Now suppose that $ALT(some\ shot)$ is defined as follows:

$$(10) \quad ALT(some\ shot) = \{\llbracket some\ shot \rrbracket, \llbracket every\ shot \rrbracket, \llbracket no\ shot \rrbracket\}$$

- a. $\llbracket some\ shot \rrbracket$ as in (9) (green circle in figure 1)
- b. $\llbracket every\ shot \rrbracket = \{Y : \llbracket shot \rrbracket \subseteq Y\}$ (purple circle in figure 1)
- c. $\llbracket no\ shot \rrbracket = \{Y : \llbracket shot \rrbracket \cap Y = \emptyset\}$ (orange circle in figure 1)

The presence of $\llbracket some\ shot \rrbracket$ has no effect because it is identical to the input. Similarly, all quantifiers that are weaker than the input have no effect if included in the ALT set. The presence of $\llbracket no\ shot \rrbracket$ has no effect because it contradicts the input, so its complement is weaker than the input. The presence of $\llbracket every\ shot \rrbracket$ will, though, be meaningful, as long as we assume that $\llbracket shot \rrbracket \neq \emptyset$. In that case, $O_{ALT}(some\ shot)$ will denote the subset in gray in figure 1. This is equivalent to the intersection of $\llbracket some\ shot \rrbracket$ and the complement of $\llbracket every\ shot \rrbracket$ in the power set of the domain. In other words, it expresses *some and not all*, the intuitively implicature-rich interpretation.

Because O_{ALT} is embeddable, syntactic constituents like $O_{ALT}(some\ shot)$ can appear in the scope of quantifiers. Implicature-rich versions of (1), (3), and (5) are thus available — potentially usable by speakers and inferable by listeners just like any other semantic resolution for an underspecified form in context.

As we noted in the introduction, CFS draw a firm rhetorical distinction between their proposal and the Gricean approach to pragmatics. They state, “the goal of this paper is to challenge the

neo-Gricean approach to SIs” (p. 2303), and, as we said, they later write that “the facts suggest that SIs are not pragmatic in nature but arise, instead, as a consequence of semantic or syntactic mechanisms” (p. 2316). The sense in which their account reflects this position is clear: to characterize implicatures, we need not consider the interactional setting or try to model the speaker and hearer. Rather, we can just describe a specific class of logical forms.

However, this position is tempered by CFS’s pervasive appeals to Gricean reasoning. The authors’ specific examples are generally placed in contexts that support the target implicatures by ensuring that they are relevant, informative, and truthful. They concede that “aspects of the Gricean picture are sound and effective” (p. 2299). And, in summarizing their account, they make explicit the role that Gricean pragmatics must play in helping discourse participants to coordinate on the right logical forms:

one can capture the correlation with various contextual considerations, under the standard assumption (discussed in the very beginning of this paper) that such considerations enter into the choice between competing representations (those that contain the operator and those that do not). (p. 2317)

The coordination problem that Grice sought to solve therefore remains, in the following form. First, in CFS’s theory, the discourse participants must coordinate on the nature of the function *ALT*. Second, because the language permits but does not require silent, embedded *O* operators in many positions, the speaker’s signal frequently underdetermines her intended message; a given surface form *U* might be consistent with logical forms that encode implicatures and those that don’t, depending on where *O* appeared. Crucially, the speaker must rely on the listener to select the right one. Overall, then, implicature calculation now amounts to reasoning about which logical form was intended. How this coordination happens has not been a focus of grammar-driven accounts, but the above quotation suggests that communicative pressures like those Grice identified guide the process.

Summarizing so far, we have evidence from Chemla & Spector’s (2011) experiments that some implicatures require, in some sense, local enrichment of embedded content via enhanced logical forms. Traditional Gricean accounts seem unable to capture such cases, but such accounts excel at characterizing how speakers and listeners coordinate on implicatures in simpler cases. CFS, in contrast, define a model in which local calculation is immediate, but they do not venture an account of how discourse participants coordinate on the right logical forms when more than one is allowed by the grammar. Stepping back, we see that both the Gricean and grammar-driven accounts each clearly have something to contribute. We now turn to the task of developing a synthesis of the two approaches: a model that formally implements pragmatic reasoning over complex, compositionally defined logical forms and that is able to achieve the readings that seem to demand local enrichment. The technical details of the compositional model are different from CFS’s, and the technical details of the pragmatic account are different from Grice, but we hope that it combines the best aspects of both approaches.

4 A compositional lexical uncertainty model

We now present our mixed semantic–pragmatic model, which can be seen as a conceptual fusion of the Montagovian semantic perspective in Lewis (1970), the signaling systems of Lewis (1969),

the probabilistic rational speech acts perspective of Frank & Goodman (2012) and Goodman & Stuhlmüller (2013), and the iterated best response model of Jäger (2007, 2012) and Franke (2009). Our Python implementation of the model is available from the website for this paper.

The model we implement here is a direct extension of the compositional lexical uncertainty model of Bergen et al. (2012) and Bergen et al. (2014) (see also Lassiter & Goodman 2013 for a closely related variant). This model defines production and interpretation as recursive processes in which speakers and listeners reason jointly about the state of world and the precise interpretation of lexical items in context. Our extension simply allows for greater diversity in the semantic lexicon and includes more complex aspects of semantic composition. Thus, in many ways, our central theoretical result is that Bergen et al.’s model predicts embedded implicatures in non-monotone and downward-monotone contexts if it is combined with a full theory of semantic composition.

The model’s crucial feature is **lexical uncertainty**. Discourse participants are not presumed to share a single, fixed lexicon. Rather, they consider many lexica, and their communicative behavior, in both production and interpretation, is guided by their best attempts to synthesize the information from these varied sources. Thus, in the sentences of interest, the discourse participants might entertain multiple senses for an embedded *some*, including not only its ‘at least’ meaning but also the ‘only some’ meaning that corresponds to its enrichment by scalar implicature. This uncertainty carries through the compositional semantics to deliver embedded implicature readings. From this perspective, Chierchia et al.’s model is conceptually very close to lexical uncertainty, in that it requires reasoning about the logical form that a speaker intends to convey; a given token of *some* can take on multiple senses depending on the presence and nature of silent embedded operators in the logical form. Our extension of Bergen et al.’s model shows how this uncertainty guides pragmatic reasoning, and it furthermore shows that the uncertainty need not be fully resolved in order for robust pragmatic inferences to go through.

4.1 Grammar fragment

Table 1 gives the intensional fragment that we use throughout the remainder of this paper, both to explain how our pragmatic model works and to conduct our experimental analyses in section 5. It is our base lexicon, subject to refinement as part of pragmatic inference.

The formal presentation is influenced by that of Muskens (1995): all of the denotations are sets, and the rules of semantic composition (the final four lines) combine them using operations that are formally akin to functional application. Our motivation for this less familiar presentation is that it makes it easy to define a uniform notion of refinement throughout the lexicon.

4.2 Refinement

The grammar in table 1 contains both lexical entries and rules of semantic combination. We assume that the rules are fixed. The lexical entries, on the other hand, are merely a starting point for linguistic communication — a set of somewhat negotiable conventions. You might assume that *couch* and *sofa* are synonymous, but if I say “It’s a couch but not a sofa”, you’ll learn something about my lexical representations and perhaps adjust your own accordingly for the purposes of our interaction. In semantics, we like to imagine that word meanings are fixed across speakers and contexts, but in fact they are sometimes idiosyncratic and usually adaptable (Clark 1997).

Syntax	Denotation of the lefthand side of the syntax rule
$N \rightarrow \textit{person}$	$\{\langle w, x \rangle : x \text{ is a person in } w\}$
$N \rightarrow \textit{shot}$	$\{\langle w, x \rangle : x \text{ is a shot in } w\}$
$V_T \rightarrow \textit{hit}$	$\{\langle w, x, y \rangle : x \text{ hit } y \text{ in } w\}$
$V_I \rightarrow \textit{scored}$	$\{\langle w, x \rangle : \exists y \ x \text{ hit } y \text{ in } w\}$
$D \rightarrow \textit{some}$	$\{\langle w, X, Y \rangle : \{x : \langle w, x \rangle \in X\} \cap \{y : \langle w, y \rangle \in Y\} \neq \emptyset\}$
$D \rightarrow \textit{every}$	$\{\langle w, X, Y \rangle : \{x : \langle w, x \rangle \in X\} \subseteq \{y : \langle w, y \rangle \in Y\}\}$
$D \rightarrow \textit{no}$	$\{\langle w, X, Y \rangle : \{x : \langle w, x \rangle \in X\} \cap \{y : \langle w, y \rangle \in Y\} = \emptyset\}$
$D \rightarrow \textit{exactly one}$	$\{\langle w, X, Y \rangle : \{x : \langle w, x \rangle \in X\} \cap \{y : \langle w, y \rangle \in Y\} = 1\}$
$NP \rightarrow \textit{Player A}$	$\{\langle w, Y \rangle : a \in \{x : \langle w, x \rangle \in Y\}\}$
$NP \rightarrow \textit{Player B}$	$\{\langle w, Y \rangle : b \in \{x : \langle w, x \rangle \in Y\}\}$
$NP \rightarrow \textit{Player C}$	$\{\langle w, Y \rangle : c \in \{x : \langle w, x \rangle \in Y\}\}$
$NP \rightarrow D \ N$	$\{\langle w, Y \rangle : \langle w, \llbracket N \rrbracket, Y \rangle \in \llbracket D \rrbracket\}$
$VP \rightarrow V_T \ NP$	$\{\langle w, x \rangle : \{\langle w, y \rangle : \langle w, x, y \rangle \in \llbracket V_T \rrbracket\} \in \llbracket NP \rrbracket\}$
$VP \rightarrow V_I$	$\llbracket V_I \rrbracket$
$S \rightarrow NP \ VP$	$\{w : \langle w, \llbracket VP \rrbracket \rangle \in \llbracket NP \rrbracket\}$

Table 1: Interpreted grammar fragment. The left column defines a context-free grammar, and the right column gives its recursive interpretation in an intensional model $\langle D, W, \llbracket \cdot \rrbracket \rangle$, where D is a set of entities, W is a set of possible worlds, and $\llbracket \cdot \rrbracket$ is a semantic interpretation function. Notational conventions: $x, y \in D$, $w \in W$, and $X, Y \subseteq (W \times D)$.

The ‘lexical uncertainty’ aspects of our model are designed to capture this variability. The core notion is that of lexical **refinement**, as defined in (11).

- (11) a. Let φ be a set-denoting expression. R is a **refinement** of φ iff $R \neq \emptyset$ and $R \subseteq \llbracket \varphi \rrbracket$.
b. $\mathcal{R}_c(\varphi)$, the set of refinements for φ in context c , is constrained so that $\llbracket \varphi \rrbracket \in \mathcal{R}_c(\varphi)$ and $\mathcal{R}_c(\varphi) \subseteq \wp(\llbracket \varphi \rrbracket) - \emptyset$

The full possible refinement space for a lexical item is the power set of its denotation minus the empty set. In a functional presentation of the interpreted fragment, this could instead be defined in terms of the subfunctions of a given denotation using a cross-categorical notion of entailment. With (11b), we allow that contexts can vary in how much of the full refinement space they utilize. They can be as small as the original denotation (in which case the uncertainty is eliminated), or as large as the full power set (minus the empty set).

The guiding idea is that, in interaction, pragmatic agents reason about possible refinements of their lexical items, with the base lexical meaning serving as a kind of anchor to which each word’s interpretation is loosely tethered. Intuitively, one can imagine that part of what it means to be a responsible interlocutor is to make inferences, based on the speaker’s behavior, not only about the world information she would like to convey, but also about the precise meanings she intends the words she is using to carry in the context of the interaction.

As we noted above, CFS’s model embodies a kind of semantic uncertainty very similar to that considered here. For any given expression that one hears, the speaker might have in mind its literal content $\llbracket \varphi \rrbracket$ or one of the many enrichments available with $O_{ALT}(\varphi)$ for different choices of ALT . Similarly, we admit the trivial refinement $R = \llbracket \varphi \rrbracket$ as well as enrichments (technically, subsets) of it. The major technical difference lies in how these sets of denotations enter into the compositional semantics. For CFS, the alternatives all contribute to a single denotation, whereas our model keeps the alternatives separate during semantic composition, synthesizing them only for pragmatic inference. In terms of figure 1, we saw that CFS’s theory uses O_{ALT} to create a single refined meaning for *some shot*, represented by the set of sets in the gray box (‘some, not all’). Our theory of refinement could create one lexicon for every non-empty subset of the green box. So, in addition to considering ‘some, not all’, we admit lexica that produce $\llbracket \text{some player} \rrbracket = \{\{a, b, c\}\}$ (‘every shot’), lexica that produce $\llbracket \text{some player} \rrbracket = \{\{a, b, c\}, \{a\}\}$ (no obvious paraphrase), and so forth. These are all potential results of $O_{ALT}(\text{some shot})$ for some choice of ALT , and our theory can be regarded as one that reasons in terms of all of these options.

4.3 Pragmatic reasoning

Our pragmatic model combines the logical grammar of section 4.1 with the lexical refinements of section 4.2. The basic ingredients are given in (12). We take as given a context c , an interpreted fragment $\langle \mathcal{G}, D, W, \llbracket \cdot \rrbracket \rangle$ as in table 1, with context free grammar \mathcal{G} , a domain of entities D , a set of worlds W , an interpretation function $\llbracket \cdot \rrbracket$ interpreting expressions of \mathcal{G} in these domains, and a refinement function $\mathcal{R}_c(\varphi)$ that is defined for all lexical items in \mathcal{G} .

- (12) a. M is a subset of the proposition-denoting expressions generated by \mathcal{G} . It is augmented with a null message $\mathbf{0}$ such that $\llbracket \mathbf{0} \rrbracket = W$.
- b. $\mathbf{L} = \{\mathcal{L}' : \mathcal{L}'(\mathbf{0}) = W \text{ and } \mathcal{L}'(m) \in \mathcal{R}_c(m)\}$
- c. $P : W \mapsto [0, 1]$ is a prior probability distribution over worlds.
- d. $C : M \mapsto \mathbb{R}$ is a cost function on messages. For lexical items, costs are specified. For a nonterminal node A with daughters $B_1 \dots B_n$, $C(A) = \sum_{i=1}^n C(B_i)$.
- e. $P_{\mathbf{L}} : \mathbf{L} \mapsto [0, 1]$ is a prior probability distribution over lexica.

In this paper, we do not bias the prior distribution over states P or the prior distribution over lexica $P_{\mathbf{L}}$ in any way, assuming them to be flat. Since we do not have experimental measurements for the priors, this seems like the safest option. (For techniques for measuring and manipulating state priors, see Frank & Goodman 2012 and Stiller et al. 2011.) Similarly, we do not explore different cost functions on messages, assuming them all to be 0 for simplicity. Our cost functions play a role only in disfavoring the ‘null message’ $\mathbf{0}$, which is stipulated to be true in all worlds in all lexica. This null message can be thought of as gathering together all of the messages that are unusable in the context because they don’t make any relevant distinctions. (This message also helps to ensure that our agents always return defined values.)

Our focus is on the space of lexica defined by (12b) given a certain set of relevant messages, as in (12a). Clause (12b) specifies all of the possible lexica given the original interpretation function $\llbracket \cdot \rrbracket$ and \mathcal{R}_c . It is the space opened up by these constructs that allows us to predict where and how embedded implicatures will be perceived as salient. It should be noted in this context that our

decision to refine only lexical items, as in (12b), is made only for simplicity. We could allow arbitrary words and phrases to be refined, as CFS in effect do.

With this background in place, we now define the core lexical uncertainty model. It consists of three inter-related agents:²

$$\begin{aligned}
 (13) \quad & \text{a. } l_0(w \mid m, \mathcal{L}) \propto \frac{\mathbb{I}(w \in \mathcal{L}(m))}{|\mathcal{L}(m)|} P(w) \\
 & \text{b. } s_1(m \mid w, \mathcal{L}) \propto \exp(\log l_0(w \mid m, \mathcal{L}) - C(m)) \\
 & \text{c. } L(w \mid m) \propto P(w) \sum_{\mathcal{L} \in \mathbf{L}} P_{\mathbf{L}}(\mathcal{L}) s_1(m \mid w, \mathcal{L})
 \end{aligned}$$

The first two agents, l_0 and s_1 , are fixed-lexicon agents, and the final listener L reasons over all of the lexica in \mathbf{L} . The most basic agent is l_0 . It defines a conditional distribution over worlds given messages. It does this by simply mapping the truth conditions into a probability distribution and incorporating the state prior. So this is just the semantics turned into a probability distribution for the sake of decision making.

The speaker agent s_1 is already a pragmatic agent, in the sense that it reasons, not about the lexicon directly, but rather about how the listener will reason about the lexicon. The logarithm and exponentiation in this definition allow us to include real-valued costs; where the costs are all 0, it reduces to $l_0(w \mid m)$, by the identity $x = \exp(\log(x))$. Similarly, if the costs are scaled into $[0, 1]$, then the values can be given as a product of the listener probabilities and the costs, analogously to the listener.

Our pragmatic listener is defined in (13c). This agent sums over all of the inferences defined by the lexicon-specific agents s_1 and l_0 . It additionally incorporates the state prior, as l_0 does, and the prior over lexica. This is the agent that we use to characterize listener inferences and define our predictions about our experimental findings.

We have given the compositional lexical uncertainty model in its simplest form, and have gone beyond Bergen et al. only in giving the more complete treatment of semantic composition necessary for the empirical domain under discussion, in allowing uncertainty in the denotations of lexical items of a wider range of semantic types, and in entertaining the possibility of restrictions on the set of possible refinements. But many other elaborations are possible. For instance, one could allow further iteration before the lexical uncertainty listener, by defining generalized versions of these fixed-lexicon agents that can reason about each other. Similarly, one could allow further iteration beyond L , defining speaker and listener agents analogously to their fixed-lexicon counterparts. For exploration of these ideas, we refer to the extended treatment in Bergen et al. 2014. In a similar vein, one can refactor the uncertainty listener so that it makes joint inferences about the lexicon, the world, and perhaps other objects (Smith et al. 2013; Kao et al. 2014). In addition, with recursion allowed on both sides of the lexical-uncertainty listener, one can push the marginalization up and down in the agent hierarchy (Goodman & Lassiter 2015). Finally, one can include a real-valued temperature parameter λ in the speaker agents to control how greedily they try to extract information from the agent they are reasoning about, with higher λ values leading to more aggressive inferential strategies (Sutton & Barto 1998; Bergen et al. 2014).

²Two notational conventions. First, \mathbb{I} is an indicator function, returning 1 if its argument is true, else 0. Second, $P(a \mid b) \propto X$ is read ‘the value $P(a \mid b)$ is proportional to the value X ’. In this context, the exact value of $P(a \mid b)$ can always be obtained by dividing $P(a \mid b)$ by the sum of all the values X' obtained by replacing a by one of its a' .

messages. The prior over states is flat in this example, so this calculation just evenly divides the probability mass over the true states. The pragmatic speaker responds to this agent. Finally, our uncertainty listener sums over these three speakers, achieving a scalar implicature: hearing *Player A scored* leads this listener to assume that the state is *S*.

Lexical uncertainty is not required to achieve this result. If we allow no meanings to be refined, then we deal with the singleton set of lexica containing only the leftmost lexicon. In this small space, the model is effectively equivalent to the rational speech act model of Frank & Goodman (2012) (with potentially small differences relating to how the prior over states is incorporated), and it is also a more thoroughly probabilistic version of the iterated best response model (Franke 2009; Jäger 2007, 2012). Nonetheless, the example illuminates how the lexical uncertainty model works. As the downward arrows indicate, it is useful to start conceptually from *L*. This agent effectively reasons in Gricean terms about three separate lexica; the alternation from speaker to listener and down to the lexicon mirrors the nested belief structure of Grice’s original definition of implicature (sketched at the start of section 2).

Even though we assume an even prior over lexica, useful biases emerge because the space of lexica is structured: there are no lexica in which *aced* is consistent with *S*, but there are two in which *scored* is. This bias carries through the computation to create a strong final bias for the implicature inference. For further discussion of this important point, we refer to Bergen et al. (2014), where it is shown to be essential to generating implicatures based on the principle that marked forms signal marked meanings and unmarked forms signal unmarked meanings (McCawley 1978; Horn 1984; Blutner 1998; Levinson 2000).

Let’s now look at a larger and more complex scenario, one that starts to show the potential of this model to predict embedded implicatures. In this scenario, there are two players. We continue our convention of referring to worlds using sequences like *NN* (‘neither player scored’). The lexical items are *Player A*, *Player B*, *some*, *every*, *no*, *scored*, and *aced*. To start, we assume that, for all lexical items φ , $\mathcal{R}_c(\varphi) = \wp(\llbracket \varphi \rrbracket) - \emptyset$. This creates an enormous space of lexica, and allows the full range of possible interactions between the refinements.

The listener inferences are summarized in table 2. For the most part, they seem aligned with the general view in the literature about how scalar terms interact in contexts like this. For instance, we predict that a proper name *P* will take on the exhaustified sense *only P*, as we would expect given the salience of *every*. In turn, *some* is interpreted as non-specific in virtue of the salience of the two names, and it also leads to a scalar implicature due to the salience of *every*. Perhaps the most striking outcome is that the scalar inference from *scored* to not-aced remains in evidence not just with the proper names but also in the scope of the quantified subjects: the best-guess inference for *every player scored* is *SS*. These effects derive from interacting lexical uncertainty between the subjects and predicates.

Table 2 reveals some drawbacks to unfettered exploration of refinements, however. First, we might expect hearing *some player scored* to lead the listener to assume that the state was either *NS* or *SN*, corresponding to enrichment of both the subject (‘not all players’) and the predicate (‘merely scored’). The current model does not achieve this. In addition, the row for *no player scored* is unintuitive. The best inference is *NN*, which is in line with the literal semantics, but it is striking that the states *NS* and *SN* have some positive probability. This arises because of interacting lexical uncertainty: there are lexica in the space in which *scored* is refined to exclude one of the players. In that case, the negative universal turns out to be true. Only a few lexica support this interaction, ensuring that it cannot become dominant, but it still seems worrisome.

This worry is a touchstone for revisiting an assumption of the model underlying table 2: that the lexical items can be refined in completely arbitrary ways. We take it to be one of the major lessons of neo-Gricean approaches that alternatives are contextually and lexically constrained. CFS’s treatment of O_{ALT} reflects this insight, and our own handling of refinement and lexical uncertainty allows us to incorporate it as well. So it is worth seeing whether we can improve the picture in table 2 by encoding lexical scales in our grammar fragment. We implement this by constraining the refinement sets for several lexical items, as follows:³

- (14) a. $\mathcal{R}_c(\text{Player } A) = \{\llbracket \text{Player } A \rrbracket, \llbracket \text{only Player } A \rrbracket\}$
- b. $\mathcal{R}_c(\text{Player } B) = \{\llbracket \text{Player } B \rrbracket, \llbracket \text{only Player } B \rrbracket\}$
- c. $\mathcal{R}_c(\text{some}) = \{\llbracket \text{some} \rrbracket, \llbracket \text{some and not all} \rrbracket\}$
- d. $\mathcal{R}_c(\text{no}) = \{\llbracket \text{no} \rrbracket\}$
- e. $\mathcal{R}_c(\text{scored}) = \{\llbracket \text{scored} \rrbracket, \llbracket \text{scored and didn't ace} \rrbracket\}$
- f. $\mathcal{R}_c(\text{aced}) = \{\llbracket \text{aced} \rrbracket\}$

The results of working in this more refined refinement space are given in table 3. The picture is mostly unchanged, except we now also achieve the target enrichment for *some player scored*, and the messiness surrounding *no player scored* is fully addressed. The one remaining potential concern about table 3 is that it predicts rather aggressive pragmatic enrichment of the scalar term in the scope of the negative quantifier. As we noted in section 2, it has long been assumed that scalar items in such environments fail to give rise to implicatures. Chemla & Spector (2011) address this question empirically, finding in their experiment low but non-negligible rates of local enrichment in negative environments. We too treat this as an empirical question; in section 5, we present evidence that local enrichments of this sort are indeed salient possibilities for humans.

This completes our overview of the compositional lexical uncertainty model. We now turn to our human-subjects experiments, where we show both that embedded implicatures seem to be real and that the above model achieves excellent fits with human judgment patterns for such sentences.

5 Experiment: Scalars under quantifiers

With our experiment, we sought to obtain a clearer picture of the availability of enriched inferences for scalar terms embedded below quantifiers. We told participants that they were helping to train an automated sportscasting system and asked them to make judgments about sentences in the context of displays like figure 3. This cover story was designed to ensure that implicatures are relevant, that is, worth calculating where available (Chemla & Spector 2011: §3.1; Clifton & Dube 2010). Our goal was to better understand the extent to which certain pragmatic inferences are available, so we sought out a scenario that would be maximally favorable to them.⁴

Our design builds directly on previous studies by Geurts & Pouscoulous (2009) and Chemla & Spector (2011), and there is an obvious correspondence between their geometric shapes and our tournament displays. However, our experiment is not a simple reproduction of theirs. Three major

³We define $\llbracket \text{only Player } A \rrbracket = \{\langle w, Y \rangle : \{a\} = \{x : \langle w, x \rangle \in Y\}\}$, and similarly for $\llbracket \text{only Player } B \rrbracket$, not as a claim about natural language *only*, but rather just for the sake of the simulation.

⁴For studies aimed at understanding the prevalence of implicatures, see Paris 1973; Hendriks et al. 2009.

	NN	NS	NA	SN	SS	SA	AN	AS	AA
<i>Player A scored</i>	0	0	0	.24	.19	.16	.18	.16	.07
<i>Player A aced</i>	0	0	0	0	0	0	.36	.30	.34
<i>Player B scored</i>	0	.24	.18	0	.19	.16	0	.16	.07
<i>Player B aced</i>	0	0	.36	0	0	.30	0	0	.34
<i>some player scored</i>	0	.14	.11	.14	.17	.14	.11	.14	.05
<i>some player aced</i>	0	0	.22	0	0	.19	.22	.19	.18
<i>every player scored</i>	0	0	0	0	.31	.27	0	.27	.14
<i>every player aced</i>	0	0	0	0	0	0	0	0	1
<i>no player scored</i>	.31	.14	.12	.14	.06	.05	.12	.05	.01
<i>no player aced</i>	.18	.19	.08	.19	.14	.06	.08	.06	0
0	.01	.01	.32	.01	.01	.15	.32	.15	0

Table 2: Enrichment in the largest space of refinements supported by this lexicon.

	NN	NS	NA	SN	SS	SA	AN	AS	AA
<i>Player A scored</i>	0	0	0	.45	.11	.22	.15	.05	.02
<i>Player A aced</i>	0	0	0	0	0	0	.42	.36	.22
<i>Player B scored</i>	0	.45	.15	0	.11	.05	0	.22	.02
<i>Player B aced</i>	0	0	.42	0	0	.36	0	0	.22
<i>some player scored</i>	0	.25	.09	.25	.06	.12	.09	.12	.01
<i>some player aced</i>	0	0	.24	0	0	.21	.24	.21	.11
<i>every player scored</i>	0	0	0	0	.61	.16	0	.16	.07
<i>every player aced</i>	0	0	0	0	0	0	0	0	1
<i>no player scored</i>	.61	0	.16	0	0	0	.16	0	.06
<i>no player aced</i>	.19	.17	.10	.17	.13	.07	.10	.07	0
0	.15	.13	.13	.13	.10	.09	.13	.09	.05

Table 3: Enrichment using the lexically-driven (neo-Gricean) refinement sets in (14).

differences are worth highlighting. First, previous experiments in this area have used abstract displays. Together with the inevitable complexity of the sentences involved, this seems likely to be cognitively demanding in ways that could affect the stability and reliability of the responses (for discussion, see Clifton & Dube 2010). Our more naturalistic displays arguably place fewer cognitive demands on participants, allowing them to concentrate on how well the sentences identify the world. Second, Chemla & Spector used wheel-like displays in which lines extend from the vertex to the perimeter. There are potentially many ways this state can be drawn, and some might be more iconic than others. With our displays, it was computationally straightforward to show participants every relevant variant of the condition, which allowed us to factor out this kind of variation in our analyses. Third, Chemla & Spector used a within-subjects design: the individual participants judged every sentence in every context. This means that subjects could draw comparisons across different conditions, which creates opportunities for them to register comparative judgments involving the experimental contexts itself, rather than relying solely on their linguistic intuitions. We hypothesized that a between-subjects design would avoid these potential complications.

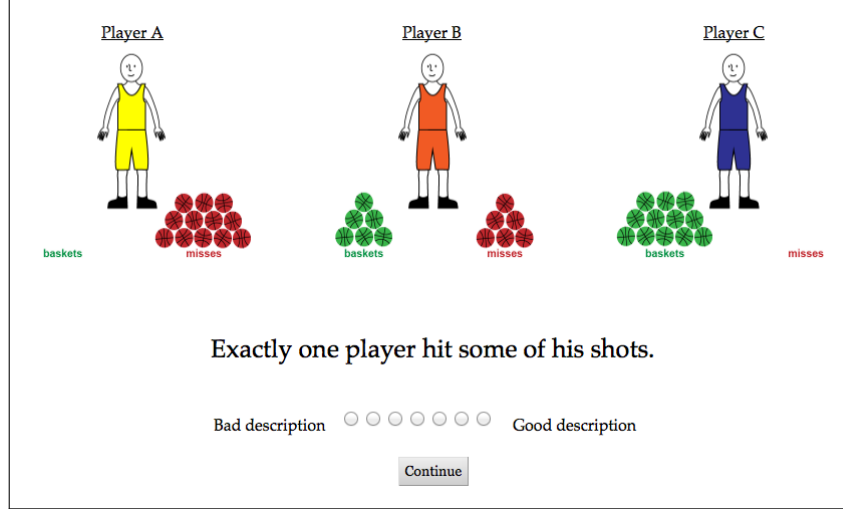


Figure 3: Experiment display.

5.1 Methods

5.1.1 Participants

The experiment had 300 participants, all recruited with Amazon’s Mechanical Turk. No participants or responses were excluded.

5.1.2 Materials

We generated displays like those in figure 3. In each display, each of the three players, A, B, and C, has taken 12 basketball shots (a number small enough for visual display but outside of the subitizing range and thus less likely to introduce competitions from cardinal determiners like *three shots*; Degen & Tanenhaus 2014). The shots were divided into two piles, labeled ‘baskets’ (green) and ‘misses’ (red). The colors of the players clothes were set randomly from a palette of 14 colors.

Target sentences describing the displays were defined as follows:

$$(15) \quad \left\{ \begin{array}{l} \text{Every} \\ \text{Exactly one} \\ \text{No} \end{array} \right\} \text{ player hit } \left\{ \begin{array}{l} \text{all} \\ \text{none} \\ \text{some} \end{array} \right\} \text{ of his shots.}$$

Following previous studies, we put a bound pronoun in the embedded quantifier to try to ensure that the subject took scope over the object. The partitive forms seem likely to further encourage implicature calculation (Reed 1991; Grodner et al. 2010; Degen in press). We chose the verb *hit* over the slightly less marked verb *make* to try to avoid the sense of ‘make’ as in ‘take’ (consistent with missing).

For the target items, there were nine different conditions, corresponding to the worlds in (16), in the notation we’ve been using to identify possible worlds.

$$(16) \quad \{NNN, NNS, NNA, NSS, NSA, NAA, SSS, SSA, SAA, AAA\}$$

This is a subset of the full cross-product of the three outcomes N, S, and A in which player *i* always did at least as well as player *i* + 1, going left to right. Our target sentences were all quantified,

so we don't care about the outcome for any single player, meaning that we don't distinguish, e.g., NNS from NSN, allowing us to work with this smaller set of conditions. In the experiment, the 'order' of each world was randomized, so that, e.g., NSA appeared visually in each of its three orders approximately the same number of times.

5.1.3 Procedure

After reading our consent form, participants were given the following cover story:

We are trying to train an automated sportscasting system to generate color commentary on simple competitions. We'd like you to make judgments about the quality of the comments it generates. We'll use these ratings to train our system further.

After reading the cover story and some instructions, participants were presented with three training items, designed to ensure that participants understood the cover story, displays, and sentences. They then judged 32 sentences, divided into nine target sentences and 23 fillers. The design was between-subjects: no experimental participant judged the same sentence twice.

Each sentence received a total of 300 responses. For the target sentences, each sentence-world pair received between 19 and 44 responses (mean 30); this variation resulted from randomization of the assignment of worlds to sentences.

Target sentences were presented below displays. Participants were asked to evaluate sentences on a seven-point Likert scale ranging from 'Bad description' to 'Good description'. Our rationale for using this scale is that it allows enough space for participants to both register a truth-value assessment and convey information about the quality of the report. (In this sense, our participants acted as listeners who got to observe the speaker's state and assess the extent to which the speaker succeeded in conveying that state with her utterance.)

All the materials and response data for the experiment are available at the website for this paper.

5.2 Results

Figure 4 summarizes the responses by target sentence and the world in which it was evaluated. Overall, participants made judgments that accurately reflected whether sentences were true or false; accuracy was especially clear for the sentences in the first two columns, which do not admit pragmatic enrichment. For these cases, the average ratings are very close to 1 (the lowest value) where the sentence is false, and the ratings average above 6 (the second highest value) where the sentence is true. The pattern of judgments in these conditions thus suggests that our method is appropriate for measuring participants' interpretations.⁵

We now turn to the critical conditions, reporting significance levels for key theoretical comparisons based on the nonparametric Mann-Whitney U test. Responses for '*every...some*' (upper

⁵The only exception to this general pattern is the sentence *No player hit none of his shots* (bottom middle), which received lower than normal ratings in all its true conditions and relatively high ratings for NNN, where it is false on its literal construal. We hypothesize that this pattern reflects a negative concord construal, on which the embedded term is interpreted as equivalent to *any of his shots*, creating a meaning that is true only in NNN. Negative concord of this sort is productive in many dialects of English and understandable in all or nearly all of them. This likely created uncertainty about the intended meaning of the sentence, leading participants to disfavor it in general.

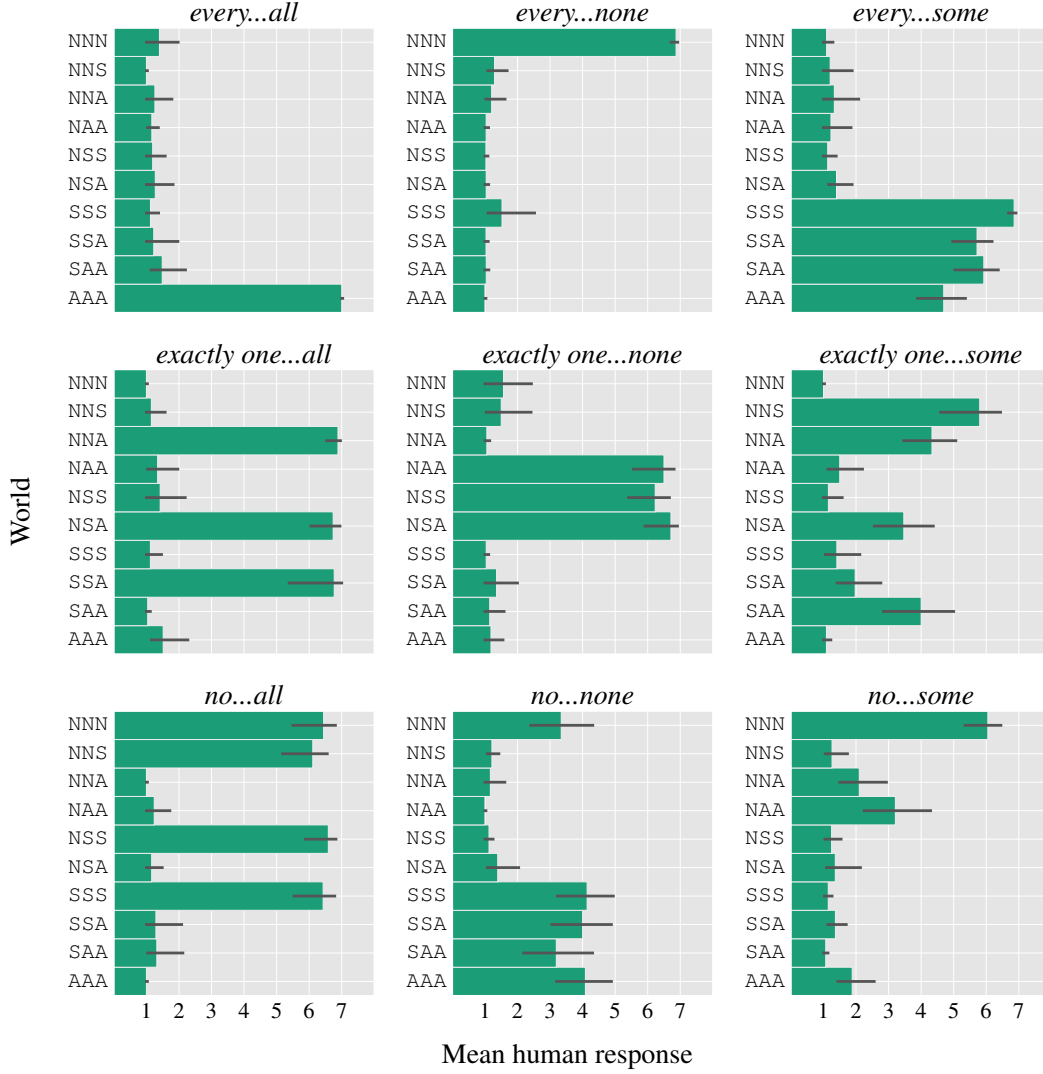


Figure 4: Mean ratings by sentence with bootstrapped 95% confidence intervals.

right) were consistent with the hypothesis that *some* is locally enriched in this condition. In particular, this sentence received its highest ratings in the SSS world. As we reviewed in section 2, example (4), in order to count as a complete report in this world, this sentence requires either local enrichment or a Gricean calculation with auxiliary premises. Worlds SSA and SAA received the next highest ratings (lower than SSS; both at $p < 0.001$), and of all the literally true worlds for this condition world AAA received the lowest rating (lower than SSA and SAA; both at $p < 0.05$). Only a simple Gricean calculation is required to account for the higher rating of SSA and SAA compared with AAA: in the latter world, the salient alternative *every player hit all of his shots* is a more complete description.

Nevertheless, ‘*every...some*’ is not a strong test of the presence of local readings, since the entailment relations between the readings introduce some indeterminacy into the analysis. In particular, since the local enrichment entails the literal reading, we can’t be sure whether the high ratings for SSS derive entirely from the availability of a local enrichment: a literal construal would

suffice to make the sentence true. Furthermore, as discussed in section 2 and in Chemla & Spector, ‘*every...some*’ is of limited utility in distinguishing theoretical proposals anyway. It is the ‘*exactly one...some*’ sentence that allows us to probe most confidently for local readings.

The response pattern for the critical item ‘*exactly one...some*’ is given in the middle right of figure 4. The highest ratings are in the NNS condition, where it is true under its literal and local enrichment construals. However, it also received high ratings in the NSA and SAA worlds (both higher at $p < 0.05$ than in SSA, the world yielding the highest rating among those in which the sentence is false both literally and under all possible enrichments), where it is true only with local enrichment (because two players hit at least some of their baskets in these worlds, ruling out the literal construal). We note also that its more strictly truth-conditional interpretation seems to be salient as well, as it was rated highly on average in the NNA condition.

Finally, the pattern for ‘*no...some*’ also suggests a non-trivial amount of local enrichment: though NNN received the highest average ratings, indicating a preference for a literal construal, the ratings for NNA, NAA, and AAA are fairly high. (The confidence intervals are wide, but a pooled comparison of {NNS, NSA} with {NNA, NAA, AAA} shows the latter set to be significantly higher-rated; $p = 0.006$.) These are the worlds in which no player hit only some of his shots, the local enrichment. This finding seems consistent with the low but non-negligible rates of local enrichment that Chemla & Spector (2011: §4.4.4) report for this quantifier pair. One qualification we should add here is that our sentence is arguably somewhat unnatural in that it places *some*, a positive polarity item (Baker 1970; Israel 1996), in the scope of a negative quantifier. The binding relation between the subject and the pronoun *his* in the embedded phrase should force a surface-scope reading, but we can’t rule out the possibility that participants might have found an inverse-scope construal (‘some shots are such that no player hit them’) that took the scalar term out of the scope of the negation.

We conclude from these responses that local enrichment is possible even in non-monotone environments, and that local enrichment might be available in downward-monotone environments as well. However, our concern is not only whether such readings are possible or impossible, but rather how accurately we can predict their availability on the basis of contextual and world knowledge. We turn now to the task of assessing the ability of the model in section 4 to match both the quantitative and qualitative patterns in our experimental data.

5.3 Model assessment

The pattern of data we observed is sufficiently precise and detailed that extracting its full theoretical value requires more than arbitrary statistical tests of simple null hypotheses — e.g., the null hypothesis that in the ‘*exactly one...some*’ condition, ratings are the same for the worlds admitted by local enrichment as for those excluded under both global and locally-enriched interpretations. This and other such null hypotheses can be rejected with high confidence. Instead, to characterize the patterns of inference that give rise to the observed data, we use a model-comparison approach. In particular, we evaluate four related models that each embody different characterizations of linguistic meaning. By comparing these models, we can gain insights into the aspects of each that contribute to particular patterns of predictions.

For all the models, we take as given the literal semantics described in table 1, as well as the following features of the context:

- (17) a. $D = \{a, b, c\}$
 b. $W = \text{the set in (16)}$
 c. $M = \left\{ Q(\text{player})(\text{hit}(S(\text{shot}))) : \begin{array}{l} Q \in \{\text{exactly one, every, no}\}, \\ S \in \{\text{every, no, some}\} \end{array} \right\}$
 d. $C(\mathbf{0}) = 5; C(m) = 0 \text{ for all } m \in M - \{\mathbf{0}\}$
 e. Flat state prior: $P(w) = P(w')$ for all $w, w' \in W$
 f. Flat lexicon prior: $P_L(\mathcal{L}) = P_L(\mathcal{L}')$ for all $\mathcal{L}, \mathcal{L}' \in \mathbf{L}$

The domain D and worlds W come directly from our human experiment. Similarly, the set of messages M corresponds to (15), with some adjustments to keep the logical grammar simple. We stipulate flat priors and even costs (other than the null message). As noted in section 4, we do not have empirical estimates for these values; though better fits to the human data can be achieved by adding assumptions about them, this risks overfitting to the particular data we have and thus overstating the true accuracy of the models.

The models we consider are defined as follows:

- (18) a. **Literal semantics:** the predicted values are the output of l_0 , as in (13a), run on the messages defined in (17c).
 b. **Fixed-lexicon pragmatics:** the predicted values are the output of the uncertainty listener (13c), but all the lexical items have only themselves as refinements, so that the reasoning is entirely in terms of the base lexicon in table 1.
 c. **Unconstrained refinement:** the inferences of the uncertainty listener (13c) with $\mathcal{R}_c(\text{some}) = \wp(\llbracket \text{some} \rrbracket) - \emptyset$
 d. **Neo-Gricean refinement:** as in ‘Unconstrained refinement’, but with $\mathcal{R}_c(\text{some}) = \{\llbracket \text{some} \rrbracket, \llbracket \text{some and not all} \rrbracket\}$.

These models represent a broad range of approaches to linguistic meaning. The first neglects pragmatics entirely (the model includes a contextual prior over states, but we define it as flat). The second is a version of the rational speech acts (RSA) model of Frank & Goodman (2012) and Goodman & Stuhlmüller (2013). RSA has been shown to capture a broad range of scalar implicatures, but it is known to be limited in its ability to derive manner implicatures and certain classes of embedded implicature (Bergen et al. 2012, 2014). The final two models are full versions of the one we presented in section 4. They represent opposite ends of the spectrum of non-trivial refinements. We saw in connection with table 2 and table 3 that there might be empirical value in greatly constraining the space of refinements.

We employ three methods of comparison: Pearson’s correlation coefficient, which measures the linear correlation between the human responses and the model predictions; Spearman’s rank correlation coefficient, which assesses how closely the human responses and model responses are aligned in terms of the rankings they predict; and the mean-squared error (MSE) of the model predictions as compared with the human responses, which summarizes the distance of the predictions from the human behavior. The use of these three measures allows us to detect models that reproduce quantitative correspondence (Pearson correlation), qualitative correspondence (Spearman correlation), and absolute fit between models and data.

	Pearson	Spearman	MSE
Literal semantics	.94	.76	.0079
Fixed-lexicon pragmatics	.92	.75	.0094
Unconstrained uncertainty	.93	.79	.0052
Neo-Gricean uncertainty	.95	.80	.0046

Table 4: Overall assessment.

Our model predictions are conditional probability distributions over states given messages, and hence constrained to be in the range $[0, 1]$ and to sum to 1. In contrast, our human responses are Likert values in $\{1, 2, 3, 4, 5, 6, 7\}$ with no constraint that they sum to any particular value. To align these values, we rescale the human responses: if x^s is the 10-dimensional vector of mean human responses for target sentence s , then each p^s is the vector of normalized values for that sentence, defined so that $p_i^s = (x_i^s - 1) / \sum_{j=1}^{10} (x_j^s - 1)$. This simply normalizes the Likert responses into a probability distribution; we subtract 1 because the lowest Likert value is 1. The one noteworthy thing about this calculation is that, because it is done on a per-sentence basis, it is not a simple linear rescaling, and so it affects all of our assessment metrics when applied to multiple sentences at once. However, we regard it as the minimal viable linking hypothesis relating our model with our experimental data.

Figure 5 summarizes the models’ predictions alongside the human responses. The predicted values are largely aligned for the examples without *some* in the object position. Where *some* occurs embedded, the models diverge in qualitative terms. For ‘*every...some*’, the patterns are broadly similar, but only ‘Neo-Gricean uncertainty’ is able to mirror the preference ordering of responses seen in the human data. For ‘*exactly one...some*’, only the two uncertainty models are able to predict local enrichment, in that only they assign high probability to the crucial worlds that are false on the literal construal: NSA and SAA. The ‘Literal semantics’ and ‘Fixed-lexicon pragmatics’ models are unable to predict the salience of these construals. Similarly, only the two uncertainty models predict ‘*none...some*’ to have embedded enrichments leading to acceptability for NNA, NAA, and AAA.

Table 4 summarizes our overall quantitative assessment. All of the correlations are extremely high, and the MSE values are extremely low. This is reassuring about the general utility of all of these models for predicting human judgments. Comparing the values is arguably more illuminating. A few patterns stand out. First, ‘Fixed-lexicon pragmatics’ performs the least well overall. Since it has been shown to add substantial value in other areas of language and cognition, we conclude that its particular approach to enrichment is at odds with the patterns for embedded implicatures. The precise causes are hard to pin-point, but the fact that our target implicatures are not always enrichments of the literal content is surely part of the problem. Second, while ‘Unconstrained uncertainty’ is better than literal semantics by the Spearman and MSE metrics, it falls behind it on the Pearson metric. We saw in connection with table 2 that the unintuitive lexica that this model entertains can lead to implausible final patterns. We conjecture that this is the root cause of the model’s poor performance here as well. Third, neo-Gricean uncertainty achieves the best results across all three of our measures. Here again, this is consistent with our expectations based on the large illustrative example from section 4.4, where we saw that this constrained, lexically-driven approach to choosing refinements resulted in the best quantitative and qualitative pattern.



Figure 5: Analysis by target sentence, comparing model predictions with human responses.

	<i>‘every...some’</i>			<i>‘exactly one...some’</i>			<i>‘no...some’</i>		
	P	S	MSE	P	S	MSE	P	S	MSE
Literal	.98	.85	.0008	.77	.70	.0188	.91	.52	.0343
Fixed-lexicon	.93	.87	.0025	.79	.70	.0184	.91	.52	.0343
Unconstrained	.91	.85	.0035	.97	.92	.0010	.74	.49	.0089
Neo-Gricean	.88	.89	.0066	.96	.89	.0029	.89	.88	.0042

Table 5: Assessment of crucial items. ‘P’ = ‘Pearson’; ‘S’ = ‘Spearman’.

The overall analysis given in table 4 understates the value of both uncertainty models when it comes to the distribution of embedded implicatures. Our target sentences provide relatively little space for pragmatic enrichment; in figure 4, the left and middle columns essentially have only literal interpretations, leaving just the right column for our pragmatic models to shine. What’s more, our qualitative review of figure 5 suggests that the right column examples reveal major distinctions. It’s thus worth assessing them quantitatively in isolation. The results of such an assessment are in table 5. The most dramatic pattern is that the two fixed-lexicon models are unable to capture the patterns for embedded implicatures in the non-monotone and downward monotone environments. In contrast, both uncertainty models capture the patterns. These tight fits are evident in figure 5, and it is reassuring to see them reflected in our assessment measures.

It is also striking that the literal model is competitive for *‘every...some’*. This probably traces to the experimentally inconvenient fact that the enriched meaning entails the literal one, making a literal semantics a generally safe bet. It should be noted also that the Spearman coefficient best reflects the ability of a model to capture the preference ordering in the human data, so the ‘Neo-Gricean uncertainty’ model might still be the favored choice for this case.

Finally, it seems that neither uncertainty model is clearly superior to the other for these data. This suggests to us that we have not yet found precisely the right approach to refinements (alternatives). It is tempting to try additional refinement sets to find a single model that wins decisively for all the target examples. We are wary of doing this because, as noted above, it runs the risk of overfitting to our experimental responses; we could easily engineer our own success. However, this is nonetheless a fruitful avenue for future exploration if paired with additional experiments for further validation.

Our model’s performance is sensitive to the space of competitor messages, so it is worth asking how robust these findings are to changes in this area. We have found that the basic pattern is robust to a number of changes to the space of quantifiers. The only noteworthy finding we have to report in this regard is that allowing *exactly one* into object position has a major impact: while SSS remains the best-guess inference for the message *‘every...some’* in this setting, *‘exactly one...some’* and *‘no...some’* effectively lose their embedded implicature readings. This makes intuitive sense given the nature of the model: if the speaker has the option to choose *exactly one of his shots*, and that form is equally costly, then surely her avoidance of that form in favor of *some of his shots* is a signal that she regards the local enrichment as infelicitous.

6 Conclusion

With this paper, we sought a synthesis between Gricean accounts of pragmatic reasoning and grammar-driven ones like that of Chierchia et al. (2012). It seems to us inevitable that both grammar and interaction will play leading roles in the final theory of these phenomena; at some level, all participants in the debate acknowledge this. Our achievement is to unify the crucial components of these approaches in a single formal model that makes quantitative predictions.

The key components of the model we develop are compositional lexical uncertainty and recursive modeling of speaker and listener agents (Bergen et al. 2014). The lexical uncertainty property is in evidence in Chierchia et al.’s account as well, in the form of underspecified logical forms with context dependent meanings. Our model has similar formal mechanisms but also offers an account of how discourse participants reason under this persistent linguistic uncertainty. This leads to an important conceptual point: not all underspecification has to be resolved in order for robust pragmatic enrichment to take place.

The recursive reasoning of our model is characteristic of both Gricean approaches and signaling systems approaches; our model shares formal properties of both but makes quantitative predictions of the sort that can be correlated with human preferences in communication. There are by now many models in the same family as ours (see, e.g., Camerer et al. 2004; Jäger 2012; Smith et al. 2013; Kao et al. 2014; Jäger & Franke 2014), so further exploration is likely to yield an even more nuanced picture.

In addition, we saw that the space of refinements (alternatives) has a significant impact on the final predictions. It would thus be worthwhile to further explore different notions of refinement, seeking better fits with our own experimental patterns and then validating those conclusions in follow-up experiments using our experimental items, or applying the resulting models in new domains. We have made publicly available all the data and code associated with this paper in an effort to encourage these new strands of theory development and quantitative assessment.

References

- Alonso-Ovalle, Luis. 2008. Innocent exclusion in an alternative semantics. *Natural Language Semantics* 16(2). 115–128.
- Bach, Kent. 1994. Conversational implicature. *Mind and Language* 9(2). 124–162.
- Bach, Kent. 2006. The top 10 misconceptions about implicature. In Betty Birner & Gregory Ward (eds.), *Drawing the boundaries of meaning: Neo-Gricean studies in pragmatics and semantics in honor of Laurence R. Horn*, 21–30. Amsterdam: John Benjamins.
- Baker, C. L. 1970. Double negatives. *Linguistic Inquiry* 1(2). 169–186.
- Beaver, David I. & Brady Zack Clark. 2008. *Sense and sensitivity: How focus determines meaning*. Oxford: Wiley-Blackwell.
- Bergen, Leon, Noah D. Goodman & Roger Levy. 2012. That’s what she (could have) said: How alternative utterances affect language use. In Naomi Miyake, David Peebles & Richard P. Cooper (eds.), *Proceedings of the 34th annual meeting of the Cognitive Science Society*, 120–125. Austin, TX: Cognitive Science Society.
- Bergen, Leon, Roger Levy & Noah D. Goodman. 2014. Pragmatic reasoning through semantic inference. Ms., MIT, UCSD, and Stanford.

- Blutner, Reinhard. 1998. Lexical pragmatics. *Journal of Semantics* 15(2). 115–162.
- Büring, Daniel & Katharina Hartmann. 2001. The syntax and semantics of focus-sensitive particles in German. *Natural Language and Linguistic Theory* 19(2). 229–281.
- Camerer, Colin F., Teck-Hua Ho & Juin-Kuan Chong. 2004. A cognitive hierarchy model of games. *The Quarterly Journal of Economics* 119(3). 861–898.
- Chemla, Emmanuel & Benjamin Spector. 2011. Experimental evidence for embedded scalar implicatures. *Journal of Semantics* 28(3). 359–400.
- Chierchia, Gennaro. 2004. Scalar implicatures, polarity phenomena, and the syntax/pragmatics interface. In Adriana Belletti (ed.), *Structures and beyond: The cartography of syntactic structures*, vol. 3, 39–103. New York: Oxford University Press.
- Chierchia, Gennaro. 2006. Broaden your views: Implicatures of domain widening and the “logicality” of language. *Linguistic Inquiry* 37(4). 535–590.
- Chierchia, Gennaro, Danny Fox & Benjamin Spector. 2012. The grammatical view of scalar implicatures and the relationship between semantics and pragmatics. In Maienborn et al. (2012) 2297–2332.
- Clark, Herbert H. 1997. Dogmas of understanding. *Discourse Processes* 23(3). 567–59.
- Clifton, Charles Jr. & Chad Dube. 2010. Embedded implicatures observed: A comment on Geurts and Pouscoulous (2009). *Semantics and Pragmatics* 3(7). 1–13.
- Degen, Judith. in press. Investigating the distribution of *some* (but not *all*) implicatures using corpora and web-based methods. *Semantics and Pragmatics*.
- Degen, Judith & Michael K. Tanenhaus. 2014. Processing scalar implicature: A constraint-based approach. *Cognitive Science*.
- Fox, Danny. 2007. Free choice disjunction and the theory of scalar implicatures. In Sauerland & Stateva (2007) 71–120.
- Fox, Danny. 2009. Too many alternatives: Density, symmetry, and other predicaments. In Tova Friedman & Edward Gibson (eds.), *Proceedings of Semantics and Linguistic Theory* 17, 89–111. Ithaca, NY: Cornell University.
- Frank, Michael C. & Noah D. Goodman. 2012. Predicting pragmatic reasoning in language games. *Science* 336(6084). 998.
- Franke, Michael. 2009. *Signal to act: Game theory in pragmatics* ILLC Dissertation Series. Institute for Logic, Language and Computation, University of Amsterdam.
- Gajewski, Jon. 2012. Innocent exclusion is not contradiction free. Ms., UConn.
- Gazdar, Gerald. 1979a. *Pragmatics: Implicature, presupposition and logical form*. New York: Academic Press.
- Gazdar, Gerald. 1979b. A solution to the projection problem. In Choon-Kyu Oh & David A. Dinneen (eds.), *Syntax and semantics*, vol. 11: Presupposition, 57–89. New York: Academic Press.
- Geurts, Bart. 2009. Scalar implicatures and local pragmatics. *Mind and Language* 24(1). 51–79.
- Geurts, Bart. 2011. *Quantity implicatures*. Cambridge University Press.
- Geurts, Bart & Nausicaa Pouscoulous. 2009. Embedded implicatures?!? *Semantics and Pragmatics* 2(4). 1–34.
- Geurts, Bart & Bob van Tiel. 2013. Embedded scalars. *Semantics and Pragmatics* 6(9). 1–37.
- Goodman, Noah D. & Daniel Lassiter. 2015. Probabilistic semantics and pragmatics: Uncertainty in language and thought. In Shalom Lappin & Chris Fox (eds.), *The handbook of contemporary semantic theory*, Oxford: Wiley-Blackwell 2nd edn.

- Goodman, Noah D. & Andreas Stuhlmüller. 2013. Knowledge and implicature: Modeling language understanding as social cognition. *Topics in Cognitive Science* 5(1). 173–184.
- Grice, H. Paul. 1975. Logic and conversation. In Peter Cole & Jerry Morgan (eds.), *Syntax and semantics*, vol. 3: Speech Acts, 43–58. New York: Academic Press.
- Grice, H. Paul. 1989. *Studies in the way of words*. Cambridge, MA: Harvard University Press.
- Grodner, Daniel J., Natalie M. Klein, Kathleen M. Carbary & Michael K. Tanenhaus. 2010. “Some,” and possibly all, scalar inferences are not delayed: Evidence for immediate pragmatic enrichment. *Cognition* 116(1). 42–55.
- Groenendijk, Jeroen & Martin Stokhof. 1984. *Studies in the semantics of questions and the pragmatics of answers*: University of Amsterdam dissertation.
- Hendriks, Petra, John Hoeks, Helen de Hoop, Irene Krammer, Erik-Jan Smits, Jennifer Spenader & Henriette de Swart. 2009. A large-scale investigation of scalar implicature. In Uli Sauerland & Kazuko Yatsushiro (eds.), *Semantics and pragmatics: From experiment to theory*, 30–50. Houndmills, Basingstoke, Hampshire: Palgrave Macmillan.
- Hirschberg, Julia. 1985. *A theory of scalar implicature*: University of Pennsylvania dissertation.
- Horn, Laurence R. 1972. *On the semantic properties of logical operators in English*. Los Angeles: UCLA dissertation.
- Horn, Laurence R. 1984. Toward a new taxonomy for pragmatic inference: Q-based and R-based implicature. In Deborah Schiffrin (ed.), *Meaning, form, and use in context: Linguistic applications*, 11–42. Washington, D.C.: Georgetown University Press.
- Israel, Michael. 1996. Polarity sensitivity as lexical semantics. *Linguistics and Philosophy* 19(6). 619–666.
- de Jager, Tikitu & Robert van Rooij. 2007. Deriving quantity implicatures. In *Proceedings of theoretical aspects of rationality and knowledge*, New York: ACM Digital Library.
- Jäger, Gerhard. 2007. Game dynamics connects semantics and pragmatics. In Ahti-Veikko Pietari- nen (ed.), *Game theory and linguistic meaning*, 89–102. Amsterdam: Elsevier.
- Jäger, Gerhard. 2012. Game theory in semantics and pragmatics. In Maienborn et al. (2012) 2487–2425.
- Jäger, Gerhard & Michael Franke. 2014. Pragmatic back-and-forth reasoning. In Salvatore Pistoia Reda (ed.), *Pragmatics, semantics and the case of scalar implicature*, 170–200. Houndmills, Basingstoke, Hampshire: Palgrave Macmillan.
- Kao, Justine T., Jean Y. Wu, Leon Bergen & Noah D. Goodman. 2014. Nonliteral understanding of number words. *Proceedings of the National Academy of Sciences* doi:10.1073/pnas.1407479111.
- Lassiter, Daniel & Noah D. Goodman. 2013. Context, scale structure, and statistics in the interpretation of positive-form adjectives. In *Proceedings of salt*, vol. 23, 587–610.
- Levinson, Stephen C. 2000. *Presumptive meanings: The theory of generalized conversational implicature*. Cambridge, MA: MIT Press.
- Lewis, David. 1969. *Convention*. Cambridge, MA: Harvard University Press.
- Lewis, David. 1970. General semantics. *Synthese* 22(1). 18–67.
- Magri, Giorgio. 2009. A theory of individual-level predicates based on blind mandatory scalar implicatures. *Natural Language Semantics* 17(3). 245–297. doi:10.1007/s11050-009-9042-x.
- Maienborn, Claudia, Klaus von Steubner & Paul Portner (eds.). 2012. *Semantics: An international handbook of natural language meaning*, vol. 3. Berlin: Mouton de Gruyter.
- McCawley, James D. 1978. Conversational implicature and the lexicon. In Peter Cole (ed.), *Syntax*

- and semantics*, vol. 7: Pragmatics, 245–259. New York: Academic Press.
- Muskens, Reinhard. 1995. *Meaning and partiality*. Stanford, CA: CSLI/FoLLI.
- Paris, Scott G. 1973. Comprehension of language connectives and propositional logical relationships. *Journal of Experimental Child Psychology* 16(2). 278–291.
- Reed, Ann M. 1991. On interpreting partitives. In Donna Jo Napoli & Judy Anne Kegl (eds.), *Bridges between psychology and linguistics: A swarthmore festschrift for lila gleitman*, 207–223. Hillsdale, NJ: Erlbaum.
- Rooth, Mats. 1985. *Association with focus*: UMass Amherst dissertation.
- Rooth, Mats. 1992. A theory of focus interpretation. *Natural Language Semantics* 1(1). 75–116.
- Rooth, Mats. 1996. Focus. In Shalom Lappin (ed.), *Handbook of contemporary semantic theory*, 271–298. London: Blackwell.
- Russell, Benjamin. 2006. Against grammatical computation of scalar implicatures. *Journal of Semantics* 23(4). 361–382.
- Sauerland, Uli. 2001. On the computation of conversational implicatures. In Rachel Hastings, Brendan Jackson & Zsófia Zvolenszky (eds.), *Proceedings of Semantics and Linguistic Theory 11*, 388–403. Ithaca, NY: Cornell Linguistics Circle.
- Sauerland, Uli & Penka Stateva (eds.). 2007. *Presupposition and implicature in compositional semantics*. Houndmills, Basingstoke, Hampshire: Palgrave Macmillan.
- Schulz, Katrin & Robert van Rooij. 2006. Pragmatic meaning and non-monotonic reasoning: The case of exhaustive interpretation. *Linguistics and Philosophy* 29(2). 205–250.
- Smith, Nathaniel J., Noah D. Goodman & Michael C. Frank. 2013. Learning and using language via recursive pragmatic reasoning about other agents. In *Advances in neural information processing systems* 26, 3039–3047.
- Spector, Benjamin. 2007a. Aspects of the pragmatics of plural morphology. In Sauerland & Stateva (2007) 243–281.
- Spector, Benjamin. 2007b. Scalar implicatures: Exhaustivity and Gricean reasoning. In Maria Aloni, Paul Dekker & Alastair Butler (eds.), *Questions in dynamic semantics*, 225–249. Amsterdam: Elsevier.
- Sperber, Dan & Deirdre Wilson. 1995. *Relevance: Communication and cognition*. Oxford: Blackwell 2nd edn.
- Stiller, Alex, Noah D. Goodman & Michael C. Frank. 2011. Ad-hoc scalar implicature in adults and children. In Laura Carlson, Christoph Hoelscher & Thomas F. Shipley (eds.), *Proceedings of the 33rd annual meeting of the Cognitive Science Society*, 2134–2139. Austin, TX: Cognitive Science Society.
- Sutton, Richard S. & Andrew G. Barto. 1998. *Reinforcement learning: An introduction*. Cambridge, MA: MIT Press.