

# **Probabilistic reasoning and the computation of scalar implicatures**

by

Benjamin Russell

B.A., San Francisco State University, 2002

A dissertation submitted in partial fulfillment of the  
requirements for the degree of Doctor of Philosophy  
in the Department of Cognitive, Linguistic and Psychological Sciences at Brown  
University

PROVIDENCE, RHODE ISLAND

May 2012

© Copyright 2012 by Benjamin Russell

This thesis by Benjamin Russell is accepted in its present form  
by the Department of Cognitive, Linguistic and Psychological Sciences as satisfying the  
thesis requirement for the degree of Doctor of Philosophy.

Date\_\_\_\_\_

\_\_\_\_\_

Pauline Jacobson, Advisor

Recommended to the Graduate Council

Date\_\_\_\_\_

\_\_\_\_\_

Kai von Fintel, Reader

Date\_\_\_\_\_

\_\_\_\_\_

Laurence R. Horn, Reader

Date\_\_\_\_\_

\_\_\_\_\_

Laura Kertz, Reader

Approved by the Graduate Council

Date\_\_\_\_\_

\_\_\_\_\_

Sheila Bonde, Dean of the Graduate School

## Acknowledgements

Many people (more than I will remember to name here) contributed to the development of this thesis in all sorts of ways. Most important, there is no way I could have reached this point without the help, advice, insight, and patience of my advisor and mentor, Polly Jacobson. Thanks also go to my committee members. Larry Horn's bibliographic suggestions and keen observations were indispensable. Kai von Fintel, in addition to providing fruitful discussions throughout the process, helped me focus on a number of outstanding issues in a near-final draft. Laura Kertz generously agreed to read and comment on the dissertation without ever having met me, and provided comments that greatly improved it. Julie Sedivy, a former member of the committee, was a great teacher and helped guide the work in a much earlier incarnation.

Dan Grodner provided countless hours on the phone and Google chat, helping me work through technical problems and forcing me to consider the predictions that my theory made. Tahir Butt contributed friendship and the iBook G4 that most of the manuscript was typed on. Chris Barker gave me generous encouragement, and, indeed, convinced me to scrap the research I was doing on conditional coordination in favor of a dissertation on scalar implicatures. Conversations with Chris Potts, Bart Geurts, Nausicaa Pouscoulous, Robert van Rooij, Uli Sauerland, Yael Sharvit, Benjamin Spector, Daniel Rothschild, Svetlana Godjevac, Anna Szabolcsi, and Brian Weatherson all influenced my thinking in ways that are reflected in this thesis. Mark Johnson and Tom Griffiths both kindly tried to help with a technical problem.

Finally, this thesis never would have been finished without my family's support, especially my parents Paul and Joanne Russell and parents-in-law Patty Nunn and John and Shelley Hébert.

Deepest thanks go to my wife, Michele. This thesis is dedicated to Louis Bernard Russell.

# Contents

<b>Acknowledgments</b>	<b>iv</b>
<b>1 Introduction</b>	<b>1</b>
1.1 The origins of implicature . . . . .	2
1.2 Horn/Gazdar Generation . . . . .	9
1.2.1 Horn’s schema . . . . .	9
1.2.2 Gazdar’s generative formalism . . . . .	12
1.3 Empirical challenges for Gazdar/Horn-style generation . . . . .	21
1.3.1 An informal Gricean reply to Chierchia . . . . .	25
1.4 Limitations of formal theories . . . . .	34
<b>2 Probabilistic Reasoning for Conversation</b>	<b>37</b>
2.1 Bayesian and Gricean reasoning . . . . .	37
2.2 Probabilistic preliminaries . . . . .	39
2.3 <i>The mathematics of rationality?</i> . . . . .	44
2.4 The new Bayesian cognitive psychology . . . . .	46
2.5 Conversational reasoning: terminology and notation . . . . .	50
2.6 Defining implicature probabilistically . . . . .	54
2.7 Calculating scalar implicatures using Bayes’s Rule . . . . .	58
<b>3 Relevance</b>	<b>62</b>
3.1 Introduction . . . . .	62
3.2 Positive relevance and assertion . . . . .	62
3.2.1 A constraint on assertions . . . . .	62
3.2.2 Carnap’s definition of relevance . . . . .	65
3.2.3 Properties of Carnap relevance . . . . .	67
3.2.4 Relevance is relative to the addressee’s beliefs . . . . .	72
3.2.5 Sperber and Wilson’s Relevance . . . . .	73
3.3 Applications for Carnap relevance . . . . .	79
3.3.1 An analysis of <i>but</i> . . . . .	80

3.3.2	Merin’s program applied to scalar implicature . . . . .	85
3.4	Pragmatics and the PRAG . . . . .	88
3.4.1	Numeral expressions . . . . .	88
3.4.2	Interaction with context . . . . .	89
3.4.3	Other expressives . . . . .	91
3.4.4	<i>so</i> . . . . .	91
3.5	Relevance and informativeness . . . . .	92
3.6	Relevance-based implicature computation . . . . .	99
3.7	Relevance implicatures . . . . .	102
3.8	Consequences of the PRAG . . . . .	105
3.8.1	Grammatical computation and the PRAG . . . . .	105
3.8.2	Implicature vs. inference . . . . .	106
3.9	Relevance is Gricean . . . . .	107
<b>4</b>	<b>A probabilistic theory</b>	<b>109</b>
4.1	Utterance expectation . . . . .	109
4.2	The implicatures of disjunction and the role of prior probabilities . . . . .	122
4.3	Disjunction, complexity, and the maxim of Manner . . . . .	124
4.4	The symmetry problem . . . . .	129
4.5	Presuppositional scalar implicature . . . . .	133
4.6	Bayesian vs. game-theoretic pragmatics . . . . .	138
4.7	Conclusion . . . . .	148
<b>5</b>	<b>Against grammatical computation</b>	<b>150</b>
5.1	Introduction . . . . .	150
5.2	Sharvit and Gajewski’s response to Russell’s response to Chierchia . . . . .	151
5.2.1	Gricean “embedded” implicature in more detail . . . . .	153
5.2.2	Probabilities make a more nuanced prediction . . . . .	154
5.2.3	Neg-raising and the probabilistic approach to implicature . . . . .	158
5.2.4	Formal Derivation . . . . .	161
5.2.5	Implications for Sharvit and Gajewski’s challenge . . . . .	163
5.3	Cancellation with <i>or</i> . . . . .	166
5.3.1	Hurford’s Constraint and local implicatures . . . . .	168
5.3.2	The softness of Hurford’s Constraint . . . . .	169
5.3.3	Truth-conditional effects of local implicatures? . . . . .	173
5.3.4	The interaction of grammatical implicatures with Gricean clausal implicatures . . . . .	177
5.3.5	Gricean clausal implicatures for entailment-related disjuncts . . . . .	181
5.3.6	More on the hybrid grammatical + pragmatic system . . . . .	186

5.4	Conclusion . . . . .	188
<b>6</b>	<b>Concluding remarks</b>	<b>190</b>

# List of Figures

4.1	A partition induced by $ALT = \{\neg\omega, \neg\sigma, \omega, \sigma\}$ and three $ALT$ -maximal belief states. . . . .	112
4.2	A partition induced by $ALT = \{\neg\omega, \neg\sigma, \omega, \sigma\}$ and three $ALT$ -maximal belief states. . . . .	114
4.3	Computation of scalar implicature in simple case. Left column lists elements of $ALT$ ; top row lists $ALT$ -maximal beliefs under consideration; and $r_H(x)$ is the assumed relevance value of each alternative. . . . .	120
4.4	Computation of scalar implicature with the inclusion of symmetric alternative $o\omega$ . . . . .	121
4.5	The implicatures of disjunction. Note $AB$ abbreviates $AB$ . . . . .	122
4.6	The inclusion of the alternative “ $A$ or $B$ and not both” pushes the interpretation towards epistemic weakness (as appended to Figure 4.5). . . . .	124
4.7	An utterance of “ $b$ or $sp$ ” implicates $[b, osp]$ over $[osp]$ with odds of 2 : 1. . . . .	125
4.8	$N$ -values with simplicity ( $s(x)$ ) incorporated. . . . .	128
4.9	Scalar reasoning including a “some and not all” alternative with high relevance . . . . .	130
4.10	Presuppositional scalar reasoning favors common grounds $[_{CG}\neg p]$ and $[_{CG}p, \neg p]$ over $[_{CG}p]$ for an utterance of $Jbp$ . . . . .	137
4.11	Presuppositional scalar reasoning for strong/weak $\times$ present/absent presupposition. . . . .	138
4.12	Values for $S_i(\ulcorner x^\neg \urcorner   \mathcal{B})$ with flat priors. . . . .	146
4.13	Values for $H_i(\mathcal{B}   \ulcorner x^\neg \urcorner)$ with flat priors. . . . .	146
4.14	Values for $S_i(\ulcorner x^\neg \urcorner   \mathcal{B})$ with skewed priors. . . . .	147
4.15	Values for $H_i(\mathcal{B}   \ulcorner x^\neg \urcorner)$ with skewed priors. . . . .	147
5.1	Scalar reasoning yields rough equivalence of “epistemic strong” and “ignorance” scalar implicatures. . . . .	162
5.2	Equivalence of $[o\omega]$ and $[o\omega, \sigma]$ . . . . .	182



# Chapter 1

## Introduction

Most of the time, *some* implies *not all*, *possible* implies *not certain*, and *warm* implies *not hot*. These implications are reversible: *not all* implies *some*, and so on. Until recently, the status of these implications has not, aside from the objections of occasional critics, been very controversial: they are conversational implicatures. The idea of implicatures was established by Grice in his 1967 William James lectures and elaborated by Horn in his 1972 dissertation. But the task of formalizing a system that accurately and adequately generates and explains scalar implications has proven difficult. This is perhaps due to the fact that conversational implicature is, following Grice, a product of nothing more than the rationality and cooperativity of humans engaged in conversation. Therefore, a tenable formal theory of rationality is an apparent prerequisite for any theory of conversational implicatures: aspects of this theory should account for inferences about speakers' intentions when they use words like *some* and *possible*. Applying precise, formal, cognitive theories of reasoning and rationality to generate such inferences will be the principal project of this dissertation.

In this first chapter, I set out the foundations of the theory of implicature, beginning with Grice's early formulation and explicating two early formal implementations. I then outline a number of empirical challenges to this Gricean approach approach, and argue

that a Gricean theory, to meet these challenges, must be formalized. Finally, I briefly consider the various formal systems that have been devised for generating scalar implicatures, and argue that, although these proposals have made many significant advances in the understanding of scalar implicature, they mostly miss the point of Gricean pragmatics: that implicature is a product of rationality.

## 1.1 The origins of implicature

Grice's theory of conversational implicature can be seen as a direct response to and extension of Austin's (1956) and Searle's (1966) observations about assertability: sentences are only assertable in contexts where there is some doubt as to their truth. It's weird to say, "I sat in my chair intentionally," to an addressee who does not entertain the possibility that the sitting was accidental. Grice takes the idea that there are conditions on assertability in two directions: he asks where the conditions come from, and how they can be characterized more completely. As to where they come from, Grice reasons that the utterance of an assertion is licit if and only if it is rational—more generally, that language use conforms to principles of rational behavior: "Our talk exchanges do not normally consist of a succession of disconnected remarks, and would not be rational if they did. They are characteristically, to some degree at least, cooperative efforts; and each participant recognizes in them, to some extent, a common purpose or set of purposes, or at least a mutually accepted direction . . . at each stage, *some* possible conversational moves would be excluded as conversationally unsuitable." (p. 18) Given this foundation, Grice provides explanations for the anomaly of a number of utterance/context pairs considered by Searle and Austin, plus some new ones.

The flip side of explaining anomaly of utterances in contexts is explaining the existence of certain implications associated with felicitous utterances. The logic of this is simple: if a particular proposition has been asserted, the context must be one that makes

the proposition assertable. A hearer, uncertain about what the context is (say, about the beliefs of the speaker), can deduce that it is a context that provides all the necessary conditions for its assertion (say, that the speaker believes some other proposition is false). The study of assertability, then, provides, in principle, much more than just insight into certain anomalous utterances; it is the foundation of the study of rational inference about language use, or *conversational implicature*.

For Grice, assertability boils down to rationality: if it doesn't make sense to say something, you shouldn't say it. Grice's approach to the problem of how to characterize what is and is not rational to assert is to provide a set of four maxims which rational conversants ought to follow: *Quantity* (make your contribution (1) as informative and (2) no more informative than is required); *Quality* (do not say what you believe to be false or lack evidence for); *Relation* (be relevant); and *Manner* (be brief and orderly, avoiding obscurity of expression and ambiguity). At face value, it is difficult to find anything obviously irrational in these maxims. But there is no reason to think (and Grice does not imply) that these four maxims constitute a comprehensive theory of rational use of language. To start, the *reasoning* part of rationality is left completely out of this formulation—the maxims characterize (again, perhaps incompletely or incorrectly) cooperation, but how do we reason about this cooperation? Deductively? How do we adjudicate between the maxims when they conflict? Grice has little to say about such issues. But Grice's key insight, and it is practically a truism, is that in reasoning about an utterance, hearers consider the consequences of a rational speaker making that utterance cooperatively. These consequences are the utterance's conversational implicatures.

In addition to this conceptual foundation of implicature, Grice sets down diagnostics for implicature; each of them supported by pretty minimal argumentation. The first, and most widely cited, is cancellability. Grice argues that cancellability follows from the dependence of implicature calculation on the assumption of the Cooperative Principle: since speakers may always opt out of cooperation, they always have the option of preventing

their interlocutors from drawing unwanted inferences. But it is not empirically clear that speakers are indeed able to cancel implicatures by opting out of cooperation. Cancellation is certainly often achieved without opting out of the cooperative principle, but rather by a speaker's cooperatively instructing an addressee that an inference they are likely to draw is not part of her intended meaning. Indeed, it is rather hard to cancel an implicature in a non-cooperative way—after all, if a speaker is not being cooperative, why should she bother to inform her interlocutor? E.g., consider the implicatures associated with a range of responses to (1.1):

(1.1) Tom: Does Mary speak Portuguese?

Elis: Her husband does. . .

- a. . . .and I think she might, too.
- b. . . .but why should I tell you anything?
- c. . . .I cannot say more/my lips are sealed. (Grice, p. 30)

Here, we see a neat dissociation between cancellation and non-cooperativity. In the first continuation, Elis cancels any implicature to the effect that she thinks Mary does not speak Portuguese. Presumably she recognizes that she may have implied this, and makes it clear that her omission of potentially relevant information is due only to relative ignorance on her part. This is clearly cooperative behavior: she explicitly guides her interlocutor towards explicit knowledge of her epistemic state. In the second continuation, despite the fact that Elis makes it as clear as possible that she has opted out of the cooperative principle (indeed, she is belligerent), the implicature that Mary does not speak Portuguese is still clearly felt. So, not only can implicatures be cancelled without opting out of the cooperative principle; but a speaker can opt out of cooperation without canceling an implicature. This undermines the cooperative principle as the Gricean foundation of implicature.<sup>1</sup>

---

<sup>1</sup>It may also point to a theory where it is impossible to opt out of the assumption of cooperation: as long as a speaker is bothering to say something to an addressee, even if that utterance is hostile or even intended

Nonetheless, Grice's characterization of implicature does make cancellation a trivially necessary property of implicatures. Implicatures are inferences rationally derived from an utterance. Inferences associated with entailments are also, of course, rationally-derived, but these inferences are necessitated by an utterance's semantic content. When Grice talks about implicatures, he is talking about aspects of meaning that are *not* linguistically encoded; i.e., that do not correspond to entailments. Since implicatures are not entailed, it is trivially true that they may be cancelled. The demonstration that an utterance may be true without the inference following is also a demonstration that the inference may be canceled. Another way to see this is to note that implicatures are inferences about what the context must be in order for a speaker to felicitously make an assertion. Such inferences will not, in general, be deterministic because not all facts of the context are known by the interlocutor. Because of this, reversing assumptions about one or more facts may reverse or remove an implicature.<sup>2</sup>

Nondetachability, the idea that implicatures shouldn't be tied to particular lexical items, but should rather remain when synonymous items are substituted, is an oft-repeated diagnostic for implicature. But nondetachability is a gradable property, even according to Grice, who defines it thus: "insofar as the calculation that a particular conversational implicature is present requires ... only a knowledge of what has been said ... and insofar as the manner of expression plays no role in the calculation, it will not be possible to find another way of saying the same thing, which simply lacks the implicature in question ..." (p. 39). The property of nondetachability separates conversational from conventional implicatures: you can remove the "contrasting" meaning of a sentence with *but*, but retain its core meaning, by replacing it with *and*:

(1.2) a. She is poor but she is honest.

---

to deceive, the interlocutors are engaged in a cooperative enterprise of information transfer.

<sup>2</sup>This discussion is not meant to imply that all non-entailed aspects of meaning are implicatures. Rather, cancellability is a necessary property of implicatures, due to their status as non-entailments.

- b. She is poor and she is honest.

But this characterization of nondetachability makes it a poor diagnostic, since one can only determine the degree to which a particular inference is nondetachable if one knows to what degree knowledge other than of what is said is required in its inference, and to what degree the inference depends on the manner of expression. In a general sense, then, nondetachability is a non-diagnostic: for if a pair of ostensibly synonymous expressions exist for which the first has an implication that the second does not, perhaps it is an implicature that depends somehow on the manner of expression. For the present purposes, then, nondetachability will not be considered a useful diagnostic.

This leaves us with just cancellability,<sup>3</sup> which is a shame, since presupposition seems also to be a cancellable aspect of meaning, as does *implicature*: inferences like the one from “I haven’t had breakfast” to the weaker “I haven’t had breakfast today”. Nonetheless, we will proceed under the operating assumption that conversational implicatures are an “aspect of what is meant in a speaker’s utterance without being part of what is said” (Horn, 2004).

Not only does Grice leave us with very little in the way of useful diagnostics, he also leaves us without anything like a formal theory of implicature computation. Implicatures are simply those non-entailed inferences that are derived from knowledge of what was said, how it was said, and what context it was said in. Grice was able to devise rational explanations of how an utterance of “Jones is having dinner with a woman” implies, but doesn’t entail, that he’s not having dinner with his wife. But, as observed by Kroch (1972), there is something essentially post-hoc about this style of linguistic analysis. Generative linguists, from the 1950’s on, have expected theories to make predictions: syntactic theories should generate well-formed sentences and fail to generate ungrammatical ones;

---

<sup>3</sup>I omitted in the preceding discussion a third diagnostic for implicature from Grice: calculability, the property of an inference that can be worked out. But a theory of implicature calculation is a formal theory of implicature, so using calculability as a diagnostic is putting the cart before the horse.

semantic theories should generate model-theoretic meanings that match intuitions. For a pragmatic theory of implicature to generate inferences, it has to be formalized.<sup>4</sup> This dissertation is an attempt to formalize an aspect of Gricean pragmatics.

In saying that the present project is a formalization of a Gricean theory of scalar implicature, I do not mean that it will slavishly follow Grice, assuming the maxims he identifies are the right ones, trying to prove that Grice's analyses of minute points are correct, and latching onto every bit of terminology that Grice introduces. Indeed, I've already questioned the necessity and usefulness of Grice's diagnostics, and the validity of the Cooperative Principle as the foundation of implicature computation. Further, much of Grice's terminology will be of little use here. One such terminological innovation is the distinction Grice makes between generalized and particularized implicatures. Later authors, most notably Levinson (1983, 2000), place considerable stock in this distinction, leading to separate theories for generalized and particularized implicature, ones where generalized implicatures are the product of conventionalized rules. But arguments for such a theoretical distinction are weak. Grice, who devotes just three pages to generalized conversational implicature in "Logic and Conversation", does not argue directly that the status of such implicatures is fundamentally distinct from that of particularized implicatures. A generalized conversational implicature arises, Grice says, when "the use of a certain form of words in an utterance would normally (in the absence of special circumstances) carry such-and such an implicature or type of implicature" (p. 37). So, even in Grice's terms, an implicature is generalized only to the degree that it is *normally* observed. Only if evidence could be found for a distinct bifurcation between implicatures normally found and those not can the generalized/particularized distinction be maintained as a concept with theoretical significance. In the absence of such evidence, the distinction is only useful descriptively—a generalized implicature is a (type of) implicature that tends to be found

---

<sup>4</sup>In writing that a pragmatic theory ought to generate inferences, I am glossing over one side of the pragmatic coin. A full-fledged pragmatic theory, of course, ought to explain both a speaker's intended implications and a hearer's derivation of inferences.

across a broader variety of contexts than particularized ones. The present work will take the view that within the class of scalar implicatures, as argued by Hirschberg (1991), there is the full range, from particularized to generalized. But before that can be shown, what is meant by *scalar* implicatures must be made precise.

The famous example of generalized implicature Grice discusses in “Logic and Conversation” is distinct from but closely related to the now canonical examples of scalar implicatures. In saying “Jones is having dinner with a woman,” a speaker implies that it is not Jones’s wife that he is having dinner with. In the details of this example Grice gives the outline of what is more or less the standard explication of scalar implicatures: “The implicature is present because the speaker has failed to be specific in a way in which he might have been expected to be specific, with the consequence that it is likely to be assumed that he is not in a position to be specific” (p. 38). In general, *specific* is replaced with *informative* in the standard literature on scalar implicature; aside from that distinction, this is a more or less modern characterization of scalar implicature.<sup>5</sup> This characterization

---

<sup>5</sup>A source of considerable confusion in subsequent literature (in explication, if not in actual theorizing), is Grice’s statement that such behaviors are “classifiable as a failure . . . to fulfill the first maxim of Quantity.” Grice’s wording of Quantity 1 is “Make your contribution as informative as is required” for the current purposes of the exchange. Can a *failure* to follow this maxim really result in this kind of implicature? Suppose the analysis is apt, and a speaker who makes a scalar implicature has not made their contribution as informative as is required. Then, in the case of indefinites (“a woman”) that Grice discusses, what is required must be the exact relation of the woman to Jones. Nothing follows from this unless a ordering of maxim-following is assumed: when a speaker fails to adhere to a particular maxim, it is because this adherence would lead to the violation of some higher-ranked maxim. In this case, the higher-ranked maxim would have to be Quality, a violation of which entails saying something which is false or for which you lack evidence. So a hearer can infer that, in order to provide the required information (the relation of the woman to Jones), a speaker would have had to violate Quality, a graver offense. So, for every relation *R*, the speaker must know it is false that Jones is having dinner with his *R*, or at least lack evidence for this proposition.

This is a rather complex view of the mechanisms of implicature generation, and it raises a lot of questions: how are the other maxims ordered? How do we know that the violation of Quantity wasn’t necessary to avoid violating, say, Manner? Grice himself seems not to be convinced that such a mechanism is the right one. By way of explanation, he gives an analogy that provides a clue to his real thoughts about generalized, and particularly scalar, implicatures: “Information, like money, is often given without the giver’s knowing to just what use the recipient will want to put it” (p. 38). I take this to mean that a speaker may well not even know precisely what “the purpose of the exchange” is; but simply that she is required to provide as much information about the topic of inquiry (in this case, the identity of Jones’s dining companion) as she can. And this requirement suggests a simpler view of the reasoning process for implicature: a speaker who (felicitously) uses an indefinite has not *failed* to provide as much information as is required; rather, she has provided exactly as much as is required, which is exactly as much as she can (at least, for some conversations. See Green (1995) for a strong argument that what is required varies depending on the conversation at hand). Assuming “a woman” is as much information as can be provided, a hearer can deduce that the



employs two concepts that must be fleshed out if it is to serve as a viable theory of scalar implicature: what does it mean for a speaker to be *specific* or *informative*, and how can we tell what level of specificity/informativity addressees *expect*? These are questions that should be answerable in any given theory of scalar implicatures that claims to be Gricean, and these answers should be plausible: ideally, they should be motivated by general theories of *specificity* or *informativity*, and, perhaps more importantly, of *expectation*.

## 1.2 Horn/Gazdar Generation

### 1.2.1 Horn's schema

The need for a predictive theory of implicature led to early attempts to formalize Gricean scalar implicature generation. In Horn's (1972) thesis, he gives the following schematic description of scalar inference: (p. 112)

- (1.3) Given a quantitative scale of  $n$  elements  $p_1, p_2, \dots, p_n$  and a speaker uttering a statement  $S$  which contains an element  $p_i$  on this scale, then
- a. the listener *can* infer  $\neg S_{p_j}^{p_i}$  for all  $j < i$  ( $j \neq n$ )
  - b. the listener *must* infer  $\neg S_{p_n}^{p_i}$ .
- (where  $S_b^a$  denotes the result of substituting  $b$  for all occurrences of  $a$  in  $S$ )

Horn's formalism is apparently intended not as a theory of an implicature generation, but as a schematic description of the inferences made by rational conversants. The general idea is that, when a speaker asserts a relatively weak proposition, listeners infer the negation of stronger propositions.

---

speaker doesn't know more, thus deriving an implicature based on the assumption that maxims are inviolable, without resorting to maxim-ordering. So, in this dissertation, I will not talk about failure to fulfill the maxims as a way of generating ordinary implicatures—rather, the assumption that speakers will not fail to fulfill the maxims is crucial for the generation of conversational implicatures.

But not every stronger proposition is negated, of course: Horn limits his definition to propositions associated with sentences that differ minimally from what the speaker uttered—by substituting for a single word with another element of that word’s *quantitative scale*. If Horn had not limited competition somehow, his mechanism would generate scalar implicatures in great excess, since every sentence is “relatively weak” compared to some sentence or other (except, maybe, a contradiction, on the view that contradictions entail, and are therefore stronger than, everything else). In particular, because every sentence  $S$  is entailed by an utterance of  $S$  and *the sky is blue*, every sentence that does not entail *that the sky is blue* would implicate that the speaker does not believe the sky is blue. This is clearly not the case. This problem is avoided by supposing that  $S$  and *the sky is blue* is not an alternative (or competitor) to the speaker’s utterance  $S$ : neither hearers or speakers consider it as something the speaker could just as easily have said (except, perhaps, in certain unusual contexts). Horn limits application of his schema using quantitative scales, ordered lists of strength-related words. Horn orders them so that, for a given scale  $p_1, \dots, p_n$ , the strength relation  $p_i \subset p_j$  holds if  $j < i$ .

In many ways, Horn’s formulation is conceptually attractive and empirically fruitful. In particular, it generates garden-variety scalar implicatures, on the assumption that *some*, *all* and *possible*, *certain* are quantitative scales:

- (1.4)    a.    *If a speaker utters* Some of the protesters fled,  
                     *the listener must infer*  $\neg$  All of the protesters fled.
- b.    *If a speaker utters* It is possible that Harry has been arrested,  
                     *the listener must infer*  $\neg$  It is certain that Harry has been arrested.

However, three caveats are in order. First, the distinction made between possible and necessary inferences in (1.3a) and (1.3b) is somewhat suspect. It is meant to capture, for example, the strength of the implicature from (1.5a) to (1.5c) over that from (1.5a) to

(1.5b).

- (1.5)    a.    Some of my friends are Zoroastrians.  
           b.    It is not the case the most of my friends are Zoroastrians.  
           c.    It is not the case that all of my friends are Zoroastrians.

But this distinction comes basically for free from entailment facts: (1.5b) entails (1.5c). And, in general, a weaker inference is drawn in more circumstances than a stronger (entailing) one, since every time the strong inference is drawn, the weaker one is entailed, but not vice versa. So, since every world where (1.5b) is true must also be one where (1.5c) is true, but not vice versa, the second inference will be less probable, or drawn less often, than the former, capturing the distinction between *may* and *must* in Horn's formulation. So it seems (1.3a) and (1.3b) can be collapsed.

Second, as pointed out in Gazdar (1979), the definition makes no allowance for downward-entailing operators; it predicts listeners can infer (1.6b) from an utterance of (1.6a):

- (1.6)    a.    It's not possible that Arthur will arrive before dinner.  
           b.    It's certain that Arthur will arrive before dinner.

(It is not the case that it's not certain that Arthur will arrive before dinner.)

In point of fact, such inferences are not licensed. Moreover, inference patterns are actually the reverse of this: "It's not certain that Arthur will arrive before dinner" implies *It's possible that Arthur will arrive before dinner*. This is a key fact for theories of scalar implicatures, and I will refer back to it as the DE reversal property.

In addition, the epistemic status of the implicature is not considered (though this is carefully considered in Horn (1989)): does the listener infer that the speaker knows  $S_{p_j}^{p_i}$  is false, or that she does not know  $S_{p_j}^{p_i}$  is true, or that she does not believe  $S_{p_j}^{p_i}$  is true, or

what?

### 1.2.2 Gazdar's generative formalism

Gazdar (1979) brings additional formal rigor to the theory of scalar implicature, but he is noncommittal about the relation of his formalism to Gricean theory: he says it “may be seen as a special case of Grice’s quantity maxim, or as an alternative to it, or as merely a conventional rule for assigning one class of conversational meanings to one class of utterance” (p. 49). Whichever it is, one thing is clear: Gazdar intends his theory to be a generative one, meant to explain implicature generation. Moreover, he believes this theory is at least *compatible* with the idea that “implicatures derive from general conversational principles”, that his rules are plausible consequences of principles of rational conversation, in contrast to a system that “cannot be plausibly explained in terms of its conversational function.”

Gazdar’s model is two-tiered: one set of rules generates potential implicatures (as well as potential presuppositions) from utterances; another set of rules then incorporates these potential implicatures into the conversational context. The first aspect of Gazdar’s formal system is the computation of *potential* implicatures, which he calls im-plicatures (with a hyphen). Gazdar’s formalism is more subtle and complex, with more nuanced predictions, than it is often given credit for (in, e.g., Chierchia 2004, Sauerland 2004, or Russell 2006), so I will take the time to spell out the formal details. He begins, following Horn (1972), with a definition of *simple scalar expression alternatives*:<sup>6</sup>

- (1.7) A pair of sentences  $\varphi_\alpha$  and  $\varphi_\beta$  are *simple scalar expression alternatives* with respect to  $\alpha$  and  $\beta$  iff  $\varphi_\alpha$  is identical to  $\varphi_\beta$  except that in *one* place where  $\varphi_\alpha$  has  $\alpha$ ,  $\varphi_\beta$  has  $\beta$ , and there are no logical functors having wider scope than  $\alpha$  in  $\varphi_\alpha$ ,

---

<sup>6</sup>I’ve combined two definitions into one here, since one is not useful without the other.

and  $\alpha$  and  $\beta$  are quantitative scalemates.<sup>7</sup>

Crucially, quantitative scales are taken to be “given to us”; Gazdar does not know what it is that determines whether a pair of strength-related expressions will enter into scalar implicature computation, instead implicitly depending on the diagnostic ubiquitous in the literature: if  $\varphi_\alpha$  and  $\varphi_\beta$  are expression alternatives with respect to  $\alpha$  and  $\beta$ , and a scalar implicature-type relationship can be observed between the pair, then  $\alpha$  and  $\beta$  are quantitative scalemates.

The key definition is of potential (scalar) implicatures, or scalar im-plicatures:<sup>8</sup>

(1.8)  $\psi$  (scalar) im-plicates  $x$  iff

- a.  $x = K \neg \llbracket \varphi \rrbracket$  for some sentence  $\varphi$ ,
- b.  $\psi = X \cap \varphi' \cap Y$ , for some expressions  $X$  and  $Y$ , possibly null,
- c.  $\llbracket \psi \rrbracket \subseteq \llbracket \varphi' \rrbracket$ ,
- d.  $\varphi$  and  $\varphi'$  are simple scalar expression alternatives, and
- e.  $\llbracket \varphi \rrbracket \subset \llbracket \varphi' \rrbracket$ .

Given this relation, we can provide a function that computes sets of implicatures for a given utterance.<sup>9</sup>

---

<sup>7</sup> Quantitative scales are defined in Gazdar as lists ordered by decreasing informativeness, in terms of generalized semantic strength; i.e., if  $\alpha_i$  and  $\alpha_j$  are scalemates, the one with the lower index entails the other. So  $\langle all, most, some \rangle$  is a quantitative scale, but  $\langle most, all, some \rangle$  is not. I’ve chosen to eliminate this convention, which simplifies certain aspects of the formalism without changing the substance of the theory.

<sup>8</sup> Gazdar defines everything in set-theoretic terms, which may be the cause of certain formal difficulties in Gazdar’s work. I have recast these definitions in relational terms, thus making them more intuitive and avoiding any formal difficulties. Obviously, anything defined as a function from objects to sets of objects, as Gazdar tries to do, may be equivalently defined as a relation between objects and elements of those sets, which is what I’ve done here. Intuitively: if we’ve generated the set of  $S$ ’s implicatures, we can define the relation  $S$  implicates  $p$  in terms of this set. Formally: Let  $f$  have domain  $D$  and range  $R$ , where  $R$  is a set of sets. Then define the set of pairs  $\{(x, y) : x \in D \wedge \exists A \in R[f(x) = A \wedge y \in A]\}$ . It is even more straightforward to define the function in terms of the relation, as I’ve done in (1.9).

<sup>9</sup> This is equivalent to the function  $f_s$  in Gazdar’s original work.

$$(1.9) \quad f_s(\psi) = \{x : \psi \text{ scalar im-plicates } x\}$$

This function maps sentences to their garden-variety scalar implicatures as a special case.

For example:

$$(1.10) \quad \begin{aligned} &\text{Some of the musicians left. } (= U) \\ &\text{im-plicates } K \neg \text{All of the musicians left. } (= I) \end{aligned}$$

Since:

- a.  $I = K \neg \text{All of the musicians left.}$
- b.  $U = \text{null} \frown \text{Some of the musicians left} \frown \text{null.}$
- c.  $\llbracket \text{Some of the musicians left} \rrbracket \subseteq \llbracket \text{null} \frown \text{Some of the musicians left} \frown \text{null} \rrbracket.$
- d. *Some of the musicians left* and *All of the musicians left* are simple scalar expression alternatives.
- e.  $\llbracket \text{All of the musicians left} \rrbracket \subset \llbracket \text{Some of the musicians left} \rrbracket.$

But the mechanism is designed to generate more than garden-variety scalar implicatures. In addition to predicting that sentences im-ply the negation their stronger scalar alternatives, sentences im-ply the negation of the stronger scalar alternatives of their proper substrings, so long as they entail those substrings—this is what (1.8b) and (1.8c) do.<sup>10</sup> To see how this works, consider the following:

$$(1.11) \quad \begin{aligned} &\text{a. George knows that Dick tried to stage a coup.} \\ &\text{im-plicates} \end{aligned}$$

---

<sup>10</sup>There are odd consequences of doing this in terms of substrings, since the grammar would, apparently, have to “come back” and attempt to provide a syntactic and semantic analysis for substrings before the implicature generation procedure could proceed—and such substrings may well have no associated syntactic parse. The system could perhaps be fixed up by replacing (1.8b) with “ $\psi = X \frown \varphi' \frown Y$ , for some constituent  $\varphi'$  and expressions  $X$  and  $Y$ , possibly null,” but this requires a grammar where syntactic structure “sticks around” after a derivation is complete, and moreover further highlights the unnatural quality of reasoning about substrings, rather than alternative utterances.

- b.  $K \neg$  Dick succeeded in staging a coup.
- (1.12) a. Dick broke the law, but Scooter has taken most of the blame.  
*im-plicates*
- b.  $K \neg$  Scooter has taken all of the blame.

The sentence in (1.11a) contains (as a substring) and entails *Dick tried to stage a coup*, satisfying (1.8b) and (1.8c). This embedded sentence has a simple scalar expression alternative in *Dick succeeded in staging a coup*, so its negation is im-plicated. Gazdar’s system’s account of such implicatures has gone largely unnoticed in the literature (Gazdar himself does not devote any discussion to this consequence of his system), and (at the risk of getting ahead of myself—see Section 1.3 for further discussion) it is worth noting that Gazdar had a global (operating on sentences, or strings) mechanism for the computation of such implicatures, something that Chierchia (2004) challenged the field to achieve.<sup>11</sup>

Gazdar also generates another class of implicatures, which he calls *clausal quantity im-plicatures*. These are the implicatures of epistemic uncertainty about certain embedded sentences:<sup>12</sup>

- (1.13) If John sees me then he will tell Margaret.  
 $\sim\sim$ I don’t know that John will see me.

- (1.14) My sister is either in the bathroom or in the kitchen.  
 $\sim\sim$ I don’t know that my sister is in the bathroom.

<sup>11</sup>To be clear, Gazdar handles only a small class of the “embedded” implicature examples that have arisen in recent literature—those where a scalar term is embedded below a factive propositional attitude—so I do not mean to imply that Gazdar solved the “embedded” implicature problem in 1979. Rather, I merely point out that Gazdar demonstrated that a global implicature computation mechanism is capable of deriving “embedded” implicatures.

<sup>12</sup>I leave out the formal details of Gazdar’s clausal implicature generation mechanism. It is worthwhile to note that clausal implicatures and scalar implicatures are not unified in the way they are in many more recent theories, like Sauerland (2004) and Russell (2006) (and in the system developed in 4).

Gazdar understands clausal implicatures in the following Gricean terms: “*if* one utters a compound or complex sentence having a constituent which is not itself entailed or presupposed by the matrix sentence and whose negation is likewise neither entailed nor presupposed, *then* one would be in breach of the maxim of quantity if one knew that sentence to be true or false, since one could have been more informative by producing a complex sentence having the constituent concerned, or its negation, as an entailment or a presupposition.” Of course, Gazdar recognizes that, in principle, it may well not be the case that a speaker could have produced such a complex sentence, but argues that “natural languages provide their users with pairs of sentences of *roughly equivalent brevity* which differ only in that in one, one or more constituent clauses are not entailed.”

This is a bit of a leap of faith, backed up by a few examples (*if/since*, *think/know*, *or/and*). Moreover, the *or/and* scale is not really amenable to this kind of reasoning—after all, suppose a speaker knows *one* of the coordinated sentences is true, but is unsure of the other. In this case, she could not have (felicitously) produced a complex sentence having the constituent concerned as an entailment or a presupposition, since the disjunct that the speaker is unsure of would be entailed if the speaker chose conjunction instead.<sup>13</sup>

The apparatus given so far generates potential implicatures (im-plicatures); these become actual implicatures via a relatively straightforward mechanism. Contexts are understood as (consistent) sets of propositions; these are updated by what Gazdar calls *satisfiable incrementation*, notated  $\cup!$  and defined as follows.

$$(1.15) \quad A \cup! B := A \cup \{X : X \in B \wedge \forall Y [Y \in A \rightarrow X \cap Y \neq \emptyset]\}$$

This takes the union of one set with those elements of a second set that are (pairwise) consistent with the elements of the first. Given an utterance  $e$  with a set of potential implicatures  $f_s(e)$  in an utterance context  $c$ , context is updated according to the following

---

<sup>13</sup>Of course, the speaker could have used a little extra effort to produce something like *A and maybe B*.



update function  $u$ :<sup>14</sup>

$$(1.16) \quad u(c, e) = (c \cup \{K[e]\}) \cup f_s(e)$$

The new context is satisfiably incremented first by the content of the utterance; then by its implicatures.

Given  $u$ , Gazdar is able to provide a simple definition of scalar implicature.<sup>15</sup>

$$(1.17) \quad \text{A proposition } x \text{ is a } \textit{scalar implicature} \text{ of an utterance } e \text{ in a context } c \text{ iff } x \in f_s(e) \cap u(c, e).$$

That is, scalar implicatures are those propositions generated by  $f_s$  that may be satisfiably incremented in the context of utterance.

Gazdar's formalism provides a precise model of implicature generation. It makes clear predictions about the observed implicatures of utterances in context. (It is, to my knowledge, still the only formal model of the conversational effect of implicatures—a lot of theories generate them, but then don't tell us what to do with them.) The first result Gazdar obtains is that implicatures are intrasententially canceled in sentences like (1.18).

$$(1.18) \quad \text{Some of the boys, in fact all of them, were there.}$$

This sentence has an im-plicature that the speaker knows it is not the case that all of the boys were there.<sup>16</sup> However, the assertion of the sentence adds the proposition  $K(\textit{some}$

<sup>14</sup>Gazdar also includes the addition of presuppositions, computed by another function,  $f_p$ , to  $u$ . The present study is not focused on presuppositions, so I leave  $f_p$  out for simplicity.

<sup>15</sup>Gazdar actually defines quantity implicatures as the union of scalar and clausal im-plicatures that are added to a context; however, since clausal implicatures are not the topic of the present study, I omit these from Gazdar's definitions.

<sup>16</sup>It is actually not completely straightforward to generate this in Gazdar's system. The ostensible way to do it is to consider the competitor *All of the boys, in fact all of them, were there*—this presumably entails (1.18), so its negation is im-plicated. The resulting strange sentence entails *Not all of the boys were there*, the desired im-plicature. It is worth noting that computation along these lines constitutes a significant departure from Gricean reasoning: nobody would ever utter the strange alternative, so its role in the computation can't

*were there AND all were there*) (equivalent to  $K(\text{all were there})$ ) to the context before the implicature can be added. And since the implicature contradicts the assertion, the implicature is not added in.

It should be noted that this analysis depends on the assumption that (1.18) means the same thing as the sentences in (1.19).

- (1.19)    a. #Some of the boys were there and all of the boys were there.  
             b. #Some of the boys were there and all of them were.

Gazdar's theory therefore predicts, indirectly, that such sentences will be straightforward examples of implicature cancellation, just like (1.18). In this way, Gazdar misses the intuition that implicature cancellation *does* something—it is not just the avoidance of inconsistency, but is, itself, an operation on the conversational state, and there are (pragmatic) conditions on its application. Moreover, as Levinson (2000) points out, cancellation may occur at any time after an implicature is made, and there is no way in Gazdar's ordered-incrementation approach for this to happen: once  $u$  is applied to an utterance in context, the implicatures thereby generated are in the context for good.

Finally, productive and explicit as it is, it is not clear that Gazdar's definition makes the right predictions about empirical data, specifically the implicatures of complex sentences. This is easier to see by looking at examples than it is by considering the definitions in the abstract. Let us return to the example in (1.11a).

---

have any real Gricean basis. Of course, the reference to substrings in Gazdar's implicature computation mechanism does not have a clear Gricean foundation, either. It would, therefore, not be a particularly pernicious conceptual departure to allow competitors to eliminate parentheticals, by changing the condition in (1.8b) to

- (1.i)     $\psi = X \frown \varphi' \frown Y$ , for some expressions  $X$  and  $Y$ , possibly null, or  $\psi = W \frown \pi \frown Z$ , where  $\pi$  is a parenthetical and  $W \frown Z = \varphi'$ .

The first clause, from Gazdar's original definition, allows for the "deletion" of material preceding or following a substring; this new second clause allows for the deletion of parentheticals.

(1.11a) George knows that Dick tried to stage a coup.

In uttering the sentence, a speaker entails (via presupposition) the sentence *Dick tried to stage a coup*. This sentence is less informative than the alternative *Dick succeeded in staging a coup*, and the speaker could have replaced the embedded sentence with this stronger alternative, therefore making a stronger entailment. From here, we must make the following leap: the speaker must have failed to embed the stronger alternative because she does not believe it is true. But there is, of course, another reason the speaker might have failed to utter the stronger sentence: because she does not believe the complex alternative *George knows that Dick succeeded in staging a coup* is true.<sup>17</sup> In other words, an utterance of (1.11a) does two things: it tells the hearer (through presupposition) that Dick tried to stage a coup, and it tells the hearer about George's belief about this fact. The stronger scalar alternative, "George knows that Dick succeeded in staging a coup", may well have a presupposition that is valid in context, while at the same time making a claim about George's beliefs that is invalid.

To illustrate a possible context, suppose the speaker and addressee share knowledge of the following chain of events: George is at his ranch in Texas when he gets a phone call from a trusted advisor saying that Dick is in the process of attempting to suspend the constitution and assume dictatorial powers. George is somewhat alarmed by this, but not so much so that he goes back to Washington. This is all that's mutually known about the coup situation when the speaker utters (1.11a) in response to a request for "an update on the coup situation." Now, upon hearing (1.11a), the addressee is in a quandary: there are two good reasons the speaker might have had for not uttering *George knows that Dick succeeded in staging a coup*. One might be that she does not believe that the coup has succeeded (leading to Gazdar's predicted implicature); but she also may believe the coup

---

<sup>17</sup>Of course, there is a deep difference between the status of the entailments of the two alternatives: one is a presupposition, whereas the other is asserted meaning. See Section 1.3.1 below for discussion of the significance of this distinction.

has succeeded but that George's knowledge is limited to the attempt. A Gricean system should be able to generate either (or both) of these implicatures, depending on whether the interlocutors mutually believe that the coup succeeded, but Gazdar's only generates the first. The second implicature, that George's knowledge is limited to the coup attempt, is not generated in Gazdar's system for the simple reason that complex sentences do not compete with each other.

Gazdar's decision to exclude complex sentences from the set of alternatives in his system is deliberate. He is worried about cases where scalar elements appear below negation, among other "logical functors", citing the lack of observed implicature from (1.20a) to (1.20b), which Horn's (1972) schematic formalism predicts to go through, as noted in (1.6a) above.

- (1.20)    a.    It is not the case that Paul ate some of the eggs.  
              b.    Paul ate all of the eggs. (reducing the double negation of *It is not the case that it is not the case that Paul ate all of the eggs.*)

Gazdar keeps his system from generating such undesirable implicatures by limiting competition to simple expression alternatives, though these competing simple sentences may be embedded in larger utterances.

In a system where complex sentences do not compete with complex sentences, no complex implicatures are expected: all scalar implicatures are of the form  $K \neg \varphi$  for some simple (with respect to the pertinent scalar term) sentence  $\varphi$ . The predicted lack of complex implicatures has a number of apparent empirical shortcomings. One is that scalar elements located below downward-entailing operators are predicted not, in general, to enter into implicature generation. But this fails to account for the DE reversal property, recognized in Horn (1972) and Fauconnier (1975), that the pattern of scalar implication is reversed by downward-entailing operators.<sup>18</sup>

---

<sup>18</sup>Chomsky (1972) noticed these data too, but argued that "Not many arrows hit the target" bears a non-

- (1.21) a. Scooter won't necessarily get off scot free.  
           *implies* Scooter will possibly get off scot free.
- b. Not all the articles of the constitution were violated.  
           *implies* Some of the articles of the constitution were violated.

Gazdar, because he excludes sentences with logical operators from competing, cannot generate such inferences.

### 1.3 Empirical challenges for Gazdar/Horn-style generation

The general approach taken by Horn and Gazdar, where scalar implicatures are a matter of a weak assertion implicating the negation of a strong alternative, was refined in many ways in the decades following its development (Soames, 1982; Hirschberg, 1991; Horn, 1989; Green, 1995). Beginning around 2000, however, studies by Landman, Levinson, and Chierchia began to challenge the empirical viability of generating scalar implicatures globally, as an operation on sentences or utterances.

I focus here on the challenges raised in Chierchia (2004)<sup>19</sup> The most compelling of these are ostensibly “embedded” implicatures of the kind in (1.22a).

- (1.22) George believes some of his advisors are crooks.
- a.  $\neg\neg$ George believes not all of his advisors are crooks.
- b.  $\neg\neg$ It is not the case that George believes all of his advisors are crooks.

---

cancellable presupposition that some did.

<sup>19</sup>The Landman data and arguments are largely coextensive with Chierchia's. The Levinson data, on the other hand, mainly involve *intrusive* implicatures, which receive special intonational marking and may be subject to a *metalinguistic* treatment, as proposed by Horn (1985) and subsequent works (though see McCawley (1991), Carston (1996), and Geurts (1998) for arguments against a metalinguistic account).

Chierchia argues that global Horn-Gazdar generation cannot generate the observed implicature in (1.22a): Gricean agents can only make conclusions about the negation of competing utterances, not embedded clauses within those utterances, and so they should only make global inferences like the one in (1.22b). Horn’s schema predicts only that a listener draws the inference in (1.22b), whereas Gazdar’s predicts no scalar implicature at all for (1.22), since the scalar term appears in an embedded sentence that is not entailed by its matrix clause.<sup>20</sup>

Chierchia argues that the interpretation in (1.22a), which looks like an embedded scalar implicature, is just that: i.e., there is a mechanism in the grammar that generates “scalar implicatures” locally and projects them through the semantic composition. As a result of the system he develops, a given linguistic expression  $\nu$  has both its regular meaning,  $\llbracket \nu \rrbracket$ , and also a strong meaning,  $\llbracket \nu \rrbracket^s$ , which is enriched with grammatically-generated “scalar implicatures.”

In Chierchia’s system, (1.22) has the strong meaning in (1.22a), derived by the semantically strengthened embedded clause in (1.23) composing up to yield the meaning in (1.23b).

- (1.23)    a.     $\llbracket \text{some of his advisors are crooks} \rrbracket^s$   
                   $\approx \llbracket \text{some and not all of his advisors are crooks} \rrbracket$
- b.     $\llbracket \text{believes some of his advisors are crooks} \rrbracket^s$   
                   $= \llbracket \text{believes} \rrbracket^s(\llbracket \text{some of his advisors are crooks} \rrbracket^s)$   
                   $\approx \llbracket \text{believes some and not all of his advisors are crooks} \rrbracket$

Thus, the grammatical mechanism Chierchia introduces provides a way to account for ob-

---

<sup>20</sup>Gazdar actually does predict that (1.22) has a *clausal* implicature, to the effect that it is possibly true and possibly false that some of George’s advisors are crooks. On top of the unfortunate facts that such implications are generally not intuitively felt to be present for sentences like (1.22), and that the data that the clausal implicature generation mechanism is designed for are readily handled with simplified scalar apparatus (see Section 1.3.1 below), clausal implicatures don’t get us any closer to the inference in (1.22a).

served scalar implicatures in complex sentences that are beyond the reach of Horn/Gazdar generation.<sup>21</sup>

Moreover, Chierchia's system provides a way to generate conversational implicatures for a sentence with a weak scalar term embedded below a factive verb, like (1.24a):

- (1.24)    a.    George knows some of his advisors are crooks.  
              b.    George knows all of his advisors are crooks.  
              c.    All of George's advisors are crooks.

Chierchia observes that an utterance of (1.24a) implies not just the negation of (1.24b), as predicted by Horn's schema, nor does it just imply the negation of (1.24a)'s presupposition in (1.24c), as Gazdar predicts; it often implies both: that not all of George's advisors are crooks and he knows it.

Chierchia's system generates both of these inferences. Since implicatures are added to an expression's meaning in the compositional semantics at each type *t* meaning, the embedded sentence *some of his advisors are crooks* (strong) means roughly *some of his advisors are crooks and not all of his advisors are crooks*. When it combines with *know*, this implicature-enriched meaning makes two contributions: it enters into the *belief* relation with George, and it becomes presupposed content, to be projected through the semantic composition by whatever means ordinary presuppositions project.

- (1.25)    a.    Asserted: George believes some of his advisors are crooks and not all are.  
              b.    Presupposed: Some of George's advisors are crooks and not all are.

At the end of the day, then, a grammatical system is able to generate an ostensibly correct interpretation for (1.24a).

---

<sup>21</sup>I have omitted the formal details of Chierchia's theory here, in the interest of perspicuity. For a relatively straightforward formal implementation of a local, grammatical implicature generation mechanism, see Fox (2007).

In addition to problematic “embedded implicature” phenomena, Chierchia points out a case of apparently gross over-generation by Horn/Gazdar: when *some* appears in one of the disjuncts of *or*, a bizarre implicature about the other disjunct is predicted. For example, (1.26) is predicted to implicate (1.26b), not (1.26a).

(1.26) George ate some of the fries or the apple pie.

- a. It is not the case that George ate all of the fries.
- b. It is not the case that George ate all of the fries or the apple pie.

This implicature entails that George did not eat the apple pie, so an utterance of (1.26) would imply that George did not eat the apple pie, an obviously undesirable implicature.<sup>22</sup>

A local theory, in contrast, is able to make the right prediction:

(1.27)  $\llbracket \text{George ate some of the fries or the apple pie} \rrbracket^s$   
 $= \llbracket \text{George ate some of the fries or George ate the apple pie} \rrbracket^s$   
 $\approx \llbracket \text{George ate some and not all of the fries or George ate the apple pie} \rrbracket$

Again, Chierchia’s grammatical system accounts for the data by computing implicatures locally, severing scalar implicature computation from Gricean reasoning altogether.

Taking stock, Chierchia provided three key cases that pose a challenge for Gazdar-style scalar implicature generation: scalar items in embedded positions, scalar items under factive propositional attitudes, and scalar terms in disjunctions. By demonstrating the failure of Gazdar generation to generate readily observed scalar implicatures for only moderately complex sentences, Chierchia called into question the status of scalar implicatures as pragmatic inferences, generated along Gricean lines.

<sup>22</sup>In fact, Gazdar predicts no scalar implicatures of the type in (1.26b), since the element containing *some* is not “entailed” by the matrix clause (assuming Chierchia’s conjunction-reduction style syntactic analysis of the sentence). But Horn-style generation does indeed make the bizarre prediction. Subsequent grammatical theories like the ones found in Chierchia et al. (to appear) do not depend on the conjunction-reduction analysis.



### 1.3.1 An informal Gricean reply to Chierchia

Chierchia's proposal to grammaticize scalar implicature has deep consequences for theories of language and conversation. Moving scalar implicatures into the grammar means a wholesale departure from the Gricean program: what has been known as *scalar implicature*, the inference from *some* to *not all*, or from *possibly* to *not necessarily*, has nothing to do with reasoning or rationality, but is simply delivered by grammar on a par with the literal semantics delivered by sentences with *only*.

Russell (2006), employing informal arguments, suggests that such a departure is not clearly warranted by the empirical data, and that an account of Chierchia's problematic examples may indeed be available in a non-grammatical, purely pragmatic theory of implicature. These arguments are built around the idea that Horn's and Gazdar's generation mechanisms are not high fidelity implementations of Gricean reasoning, but that they are roughly equivalent to the application of Gricean principles in a range of special cases, most of them very simple. Through a more careful consideration of the pragmatic conversational reasoning that rational interlocutors should engage in, solutions to Chierchia's puzzles may be derived.

In the case of embedded scalar terms, a series of inferences, each with Gricean justification, is capable of bringing the hearer to empirically attested interpretations. First, note that Chierchia understates the complexity of the data in examples like (1.22).

(1.22) George believes some of his advisors are crooks.

- a.  $\sim\sim$ George believes not all of his advisors are crooks.
- b.  $\sim\sim$ It is not the case that George believes all of his advisors are crooks.

Chierchia is right to claim that there are contexts where an utterance of (1.22) implies George believes not all of his advisors are crooks, and pragmatic theory should be able to account for this implication. But the implication is certainly not always present when the

weak, global implicature is; some contexts license just the inference in (1.22b). Consider the following context, where the assumption that George is opinionated about each of his advisors is explicitly denied.

- (1.28) George has not yet formed an opinion about all of his advisors, but, at this point, he believes some of them are crooks.

Here, there is no implicature that George believes not all of his advisors are crooks, but there is a clear implicature that it is not the case that George believes they're all crooks.

The implication in (1.22a) entails the one in (1.22b); in other words, it is a stronger implication. A very general way to account for both implications, then, is to suppose that (1.22b) is more basic than (1.22a), and that the latter is derived from the former by some extra inferences. Indeed, (1.22a) follows from (1.22b) in every context where George has *some* belief about whether all of his advisors are crooks: if it is not the case that George believes that all of his advisors are crooks, and George has some belief about whether all of his advisors are crooks, then George must believe that not all of his advisors are crooks.

Still, this kind of informal Gricean derivation for scalar implicatures begins to break down for examples that are more complex: recall Chierchia's example with a scalar term below a factive propositional attitude:

- (1.24a) George knows that some of his advisors are crooks.

First, notice that the data are more complex than Chierchia recognizes. There are contexts where the "implicature of the presupposition" is defeated and yet the ordinary scalar implicature still arises, posing a serious problem for his analysis. Consider (1.24a) in the following slightly richer context:

- (1.29) The public has long been aware that every last one of George's advisors is a

crook. And now (even) George knows that some of his advisors are crooks.

Here, the right theory will predict that the implicature that George believes not all of his advisors are crooks can still go through while the inference that not all of the advisors are crooks does not. In Chierchia's theory, clauses with scalar terms become essentially ambiguous: they may be strengthened or not. In other words, either the "embedded" implicature is canceled (and is found neither in the presupposition nor the assertion), or not (and is found in both).<sup>23</sup> There is no room for the intermediate interpretation found in (1.29) where the presupposition's "implicature" is canceled but the assertion's is not. A rule could be formulated to selectively remove implicatures within the presuppositional system, but this would require additional stipulation, making Chierchia's system still more complex.

A Gricean analysis of such inferences must begin with a careful consideration of their status. The first thing to notice is that they are relatively weak compared to garden-variety scalar implicatures. Indeed, many speakers I consulted do not have the intuition that (1.24a) implies *that not all the advisors are crooks* at all, but only that it is compatible with such a situation. Such sentences are apparently equally felicitous whether or not all of the advisors are crooks—an expression like *in fact* is not needed to cancel that supposition like it is to cancel an ordinary scalar implicature.

- (1.30)    a.    George knows that some of his advisors are crooks, and (in fact/*I* know) they all are.
- b.    Some of George's advisors are crooks, and #(in fact)/(#*I* know) they all are.

In Chapter 4 below, I provide a formal analysis of these inferences and their properties with

---

<sup>23</sup>Cancellation, in Chierchia's system, is a reanalysis of a sentence whereby Strong Application does not apply.

respect to context sensitivity. For now, a brief characterization of the Gricean mechanisms underlying this inference can be given. There are two possibilities. First (and probably wrong): the source of the inference is the assumption that George is well-informed about the integrity of each of his advisors. This assumption, plus the implicature that it is not the case that George believes all his advisors are crooks, licenses the inference that not all of George's advisors are crooks.

Against this approach, such inferences seem to persist even when the factive verb is found below negation (I've used examples with *possible/certain* because of potential positive-polarity effects getting in the way with *some*):

- (1.31) George doesn't know that failure in Iraq is possible.<sup>24</sup>  
*implies* Failure in Iraq is not certain.

Certainly *know* in such an environment does not presuppose or imply that its subject is well-informed about anything, let alone that George is well-informed about the odds of failure in Iraq. This suggests a possible second type of explanation for these phenomena: Gricean scalar reasoning about presupposed content. It is natural to assume that Gricean reasoning incorporates such information: if speakers are aware of competing utterances, they should be aware of the presuppositional content of those utterances as well as their assertoric content (see Geurts 2011 for a discussion of Gricean reasoning about presupposed content). In a case like this, a speaker, considering the alternative *George doesn't know that failure in Iraq is certain*, might choose to use a sentence with a weaker presupposition because she does not want a stronger one to be accommodated. Spelling this out, we have:

- (1.32) Utterance assertion:  $\neg$  George believes failure is possible.

---

<sup>24</sup>This implicature also has peculiar cancellation facts; it doesn't require *in fact* or any other standard expression: consider the continuation, "... and even Dick doesn't realize it's certain."

Utterance presupposition: Failure is possible.

Alternative's assertion:  $\neg$  George believes failure is certain.

Alternative's presupposition: Failure is certain.

Notice that whereas the alternative's assertion is weaker than the utterance's, its presupposition is stronger. The speaker, then, could have uttered a sentence with a weaker assertion and a stronger presupposition, but chose not to. The reasons for making such a conversational move will depend on context in complicated ways. This complexity can not be readily captured using the informal arguments employed in Russell (2006). But it is worth noting that this kind of nuance ought to be expected on a reasoning-based approach, but is quite mysterious, given a mechanical, grammatical approach to implicature computation.<sup>25</sup>

Finally, recall Chierchia's disjunction example:

(1.26) George ate some of the fries or the apple pie.

In reply to this challenge, Sauerland (2004) provides a global system for implicature computation, adapted from Gazdar's, with some simplifications and some added complexity. First, Sauerland simplifies Gazdar's notion of scalar expression alternatives: he no longer requires that they be *simple* with respect to the scalar elements they contain:

(1.33) A pair of sentences  $\varphi_\alpha$  and  $\varphi_\beta$  are *scalar expression alternatives* with respect to  $\alpha$  and  $\beta$  iff  $\varphi_\alpha$  is identical to  $\varphi_\beta$  except that in *one* place where  $\varphi_\alpha$  has  $\alpha$ ,  $\varphi_\beta$  has  $\beta$ , and  $\alpha$  and  $\beta$  are quantitative scalemates.

Second, Sauerland simplifies the implicature-generation mechanism, discarding the sub-string apparatus of Gazdar, and weakens it, in line with views from Horn (1989) and

---

<sup>25</sup>Though it is captured, to some degree, by the formal system in Chapter 4.

Soames (1982) that the basic form of scalar implicature is  $\neg K$ , not  $K\neg$ . The formulation he gives simply says that in saying  $\varphi$ , a speaker implicates that she doesn't know a stronger scalar alternative to  $\varphi$  is true.

(1.34)  $\varphi \rightsquigarrow \neg K\psi$  iff:

- a.  $\psi \subset \varphi$ , and
- b.  $\psi$  and  $\varphi$  are scalar expression alternatives.

This seems a lot closer to Gricean reasoning than Gazdar's formalism: a speaker who utters a weak sentence and knows a stronger sentence is true is not being as informative as possible, and so is in violation of the maxim of Quantity. Moreover, there is no bothering with substrings, an aspect of Gazdar's definition that is also hard to justify on Gricean grounds.

Thus simplifying Gazdar's apparatus, Sauerland's key innovation is the introduction of a pair of unrealized lexical items  $L$  and  $R$ , treated as quantitative scalemates to *and* and *or*. The semantics of  $L$  and  $R$  is unlike other coordinating conjunctions: rather than merely taking Boolean meet or join of the elements it coordinates,  $L$  takes two phrases and returns the semantic value of the one on the left (i.e., the second one it combines with), so that, for example, *Charles sold that bottle L gave it away* means the same thing as *Charles sold that bottle*.  $R$  is analogous, returning the semantics of the right conjunct.

Recall that Horn/Gazdar Generation depends crucially on the notion of quantitative scales: scalar implicature generation only applies to pairs of sentences that are scalar expression alternatives, as in Gazdar's definition (1.7). The  $L$  and  $R$  operators make each disjunct in a sentence an alternative to the matrix disjunction, so that, for example, *George ate the apple pie* is an alternative to *George ate some of the fries or the apple pie*.<sup>26</sup>

---

<sup>26</sup>Notice that if the  $L$  and  $R$  operators were realized, we'd have a (potentially unsolvable) version of the "symmetry" problem (term due to von Stechow and Heim (1998)):

(1.i) John left  $R$  the sky is blue.

Finally, an operation to generate secondary implicatures is proposed, whereby an implicature of the form  $\neg K\varphi$  is strengthened to  $K\neg\varphi$ , whenever this does not conflict with the (primary) scalar implicatures of an utterance; so, for example, *Some of the students left* has a primary implicature of  $\neg K$  *All of the students left* and a secondary implicature of  $K\neg$  *All of the students left*, since this is compatible with the utterance's assertion and primary implicature.

These two extra bits of apparatus solve Chierchia's disjunction problem. An utterance of (1.26) has the following primary implicatures:

- (1.35)    a.     $\neg K$  George ate some of the fries.  
              b.     $\neg K$  George ate the apple pie.  
              c.     $\neg K$  George ate all of the fries or the apple pie.

Notice (1.35c), no longer an epistemic strong implicature, does not entail that George did not eat the apple pie. Moreover, none of these primary implicatures can be strengthened without contradicting some other primary implicature or the assertion itself:  $K\neg$  *George ate some of the fries*, plus the assertion that George ate some of the fries and the apple pie, entails  $K$  *George ate the apple pie*. But this contradicts the primary implicature in (1.35b). Likewise, (1.35b) cannot be strengthened, *mutatis mutandis*. Strengthening (1.35c) entails  $K\neg$  *George ate the apple pie*, which, combined with the assertion, yields  $K$  *George ate some of the fries*, contradicting (1.35a).

Though Sauerland's solution is indeed global, it is worth pointing out that Gricean considerations suggest a solution without the need of stipulated  $L$  and  $R$  operators and

---

*is weaker than its scalar alternative* John left and the sky is blue.  
*and so implicates*  $\neg K$  John left and the sky is blue.

That is, an utterance of "John left" would implicate  $\neg K$  *the sky is blue*. See section 4.4 for more discussion of the symmetry problem.

Sauerland argues that the unrealized nature of  $L$  and  $R$  is a plausible consequence of Grice's maxim of Manner. It is worth pointing out that, given the unrealized nature of  $L$  and  $R$ , Gricean scalar reasoning makes no sense:  $ALB$  is not something the speaker would have said if he believed  $A$  were true.

primary and secondary implicatures. First, note that the key to Sauerland's analysis is a closer consideration of the epistemic status of scalar implicatures: again, scalar reasoning only licenses epistemic weak inferences, *à la* Horn (1989) and Soames (1982). So scalar implicatures are inferences of the form  $\varphi \rightsquigarrow \neg K\psi$ , rather than the stronger  $\varphi \rightsquigarrow K\neg\psi$ . Given these weaker inferences, corresponding strong inferences of  $K\neg\psi$  may be generated when the hearer assumes that the speaker knows *whether*  $\psi$  is true ( $K\psi \vee K\neg\psi$ )—i.e., when the speaker is *competent* with respect to the truth of  $\psi$ . Treating scalar implicatures as this kind of two-stage inference is not only more faithful to Grice's theory; it also provides a solution to Chierchia's puzzle, for (1.26) now has the implicature in (1.36), which does not have the bad entailments of (1.26b).

(1.36)  $\neg K$  George ate all of the fries or the apple pie.

And (1.36) entails the following two epistemic facts about the speaker.

- (1.37)    a.  $\neg K$  George ate all of the fries.  
             b.  $\neg K$  George ate the apple pie.

In contexts where the hearer can assume the speaker is competent—i.e., knows whether George ate all the fries—(1.37a) can be strengthened, yielding the desired implicature in (1.26a) (in all other contexts, (1.37a) is predicted to be the strongest implicature drawn, which is very much in keeping with my intuitions).

The implicatures above are an accurate representation of the observed interpretation of (1.26). But the reader may wonder why (1.37b) is not strengthened, along the lines of (1.37a), to the obviously undesirable inference (1.38)?

(1.38)  $K\neg$  George ate the apple pie.



This can be explained in intuitive terms: a sentence's scalar implicature cannot be strengthened if this leads to contradiction with another of its basic implicatures. First, notice that  $p$  and  $q$  are stronger than (entail)  $p \vee q$ , and  $p$  and  $q$  are each certainly easier to say than  $p$  or  $q$ , so if a speaker has chosen to use  $p$  or  $q$ , she must not know that either  $p$  or  $q$  is true. So the weak (clausal) implicatures in (1.39) are generated along with the weak implicatures in (1.36).<sup>27</sup>

- (1.39)    a.     $\neg K$  George ate some of the fries.  
              b.     $\neg K$  George ate the apple pie.

These implicatures, combined with the uncontroversial assumption that speakers believe what they say (Grice's maxim of quality), are incompatible with the strengthening of (1.37b) to (1.38). The derivation is as follows:

- (1.40)    a.     $K$  (George ate some of the fries  $\vee$  George ate the apple pie) (from (1.26))  
              b.     $K$  ( $\neg$  George ate the apple pie  $\rightarrow$  George ate some of the fries) (equivalent to (1.40a))  
              c.     $K\neg$  George ate the apple pie  $\rightarrow K$  George ate some of the fries (from (1.40b), assuming a distribution axiom for knowledge modality)  
              d.     $\neg K$  George ate some of the fries  $\rightarrow \neg K\neg$  George ate the apple pie (contrapositive of (1.40c))  
              e.     $\neg K$  George ate some of the fries (from (1.39a))  
              f.     $\neg K\neg$  George ate the apple pie (modus ponens)

The conclusion in (1.40f) contradicts, and therefore "blocks," (1.38). This shows the observed implicature (1.26a) can apparently be generated by a global Gricean theory without

---

<sup>27</sup>These are the implicatures labeled *clausal* in Gazdar (1979); like Sauerland, they are treated as scalar here. See Chapter 5 for further discussion of clausal implicatures.

any undesirable consequences.<sup>28</sup>

## 1.4 Limitations of formal theories

Gazdar motivates his formalism with the logic of an informal Gricean account of implicature derivation via the maxim of quantity (p. 51):

Anyone uttering a [weak] sentence who was in a position to utter a [stronger scalar alternate] sentence would be being less informative than he could be since the [stronger] sentence makes a stronger claim about the world than the [uttered] sentence. Thus, if the speaker is being cooperative and observing the maxim of quantity, it follows that in uttering the [weaker] sentence he is implicating the negation of the [stronger] sentence.

But this reasoning is not instantiated by the rules that Gazdar formulates. That is, although Gricean scalar reasoning and Gazdar's formal system have similar output for a certain class of sentences, one is simply an instantiation of rationality, whereas the other is a collection of definitions and rules. Only on the assumption that rationality is an amalgam of rules and, moreover, that among these rules is (1.8), can Gazdar's theory claim the conceptual foundation of Grice's.

In the 30 years since Gazdar put forth his system, other global, formal models of scalar implicature generation have been proposed. In general, these theories also lack the conceptual foundation of Grice. Sauerland's (2004) system, outlined above, depends on extra *L* and *R* operators and a stipulated distinction between primary and secondary

---

<sup>28</sup> See Chapter 4 for a formal derivation of this case. The metaphor used here, that a clausal implicature "blocks" the strengthening of an epistemic weak implicature due to an undesirable epistemic strong implicature, is somewhat misleading and, moreover, does not have an obvious rational basis: why should one implicature (the epistemic weak, clausal implicature) take precedence over another? If this were so, it would mean that implicature generation follows a prescribed set of steps, with certain classes of implicatures generated before others, and this ordering would have to be stipulated. The theory presented in Chapter 4 provides a Gricean solution without any stipulation of ordering effects.

implicatures,<sup>29</sup> and generation of scalar implicatures is based on a formal rule, not on Gricean reasoning. At best, the formal rule is inspired by informal considerations. In this light, Sauerland's system begins to seem more like Chierchia's—it is simply a global grammatical mechanism, rather than a local one.

Over the past decade, additional global systems for generating scalar implicatures have been developed. van Rooij and Schulz (2004) define a series of operators that work on propositions to give back model-theoretic objects, based on a set of formal definitions. These operators are inspired by both the exhaustive interpretation of questions developed by Groenendijk and Stokhof (1984) and a theory called Predicate Circumscription from Artificial Intelligence (McCarthy, 1980), intended to model the way humans “jump to conclusions,” given partial information. But neither the operators nor the definitions seem to be a product of rationality; rather, they are formal objects, analogous to grammatical objects, that approximate the desired inferences. Moreover, it is propositions, not utterances, that people reason about in van Rooij and Schulz's theory. Because of this, scalar implicatures are not derived from any consideration of what the speaker could have said but didn't.

At the same time, theories like those of Horn, Gazdar, and Sauerland that do depend on alternative utterances all rely on quantitative scales for implicature generation.<sup>30</sup> And quantitative scales are decidedly un-Gricean—why should we limit our application of the maxim of Quantity to certain formally-defined alternatives? The answer generally given in the literature is that the formally-defined alternatives are a substitute for a principled application of the maxim of Manner: unless two sentences are scalar alternatives, we can't guarantee that there is no Manner-based reason for failing to utter the stronger one, like, say, that it would take too long to say. But this need for guarantees cuts us off from the

---

<sup>29</sup>This distinction is, of course, inspired by the Gricean/Soamesian epistemic step (see discussion surrounding (2.16)), but it is stipulated that primary implicatures take precedence over secondary implicatures, preventing strengthening in the clausal case.

<sup>30</sup>Though see Hirschberg (1991) and van Kuppevelt (1996) for theories where the scales involved in scalar reasoning are drawn from contextually-supplied partial orders or from discourse structure.

richness of scalar implicature phenomena. We presumably don't limit our reasoning to alternatives that are exactly equally marked—it would be rational, it seems, to consider slightly more complex or wordy alternatives, if their relevance or informativity was sufficiently high.

In weighing the Gricean merits of the new formal global theories, present tools seem to fail us. After all, how can we even make arguments at all without explicit, detailed agreement about what it means to be rational? The state of the field can be understood, I think, as a result of pragmatic theory's development by linguists. That is, in the majority of these theories, scalar implicature is generated by rules that are devised for the specific purpose of generating scalar implicatures and related inferences. Broadly speaking, they are not based upon more general theories of reasoning. This fact is understandable: the assumption that implicature is a product of rational inference means that a theory of implicature is dependent on a formal theory of reasoning. Rather than start at such a fundamental level, implicature theorists have treated implicature like any other linguistic phenomenon: something to be explained by a rule-based system. For this reason, formal theories of scalar implicature generation tend to be grammatical, whether they are global or not.<sup>31</sup>

---

<sup>31</sup>An important exception to this claim is found in Franke (2009), which generates scalar implicatures and models an array of other pragmatic phenomena by treating conversation as a signalling game, based on the idea that games provide a way to model rational multi-agent interaction. See section 4.6 for discussion.

## Chapter 2

# Probabilistic Reasoning for Conversation

### 2.1 Bayesian and Gricean reasoning

The cornerstone of Gricean pragmatics is the idea that human rationality drives conversational inference: the grammar encodes the meaning that an utterance comes with, and rational conversational partners use the fact that the speaker chose an utterance with that meaning to draw non-encoded inferences. In the Gricean program, implicature is more closely related to figuring out why someone handed you a screwdriver, rather than a wrench, than it is to getting a type  $t$  meaning out of an  $\langle et, t \rangle$  and an  $\langle et \rangle$ . And yet it is rare for formal theories of linguistic pragmatics to seriously consider theories of rationality.<sup>1</sup> The exceptions to this can be sorted into two of different schools of thought: logical approaches, which include both deductive and non-monotonic logics (Tokarz, 1994; Spector, 2007b; van Rooij and Schulz, 2004), both developed as normative theories of inference; and, more recently, game-theoretic approaches (Ross, 2006; Benz and Van Rooij, 2007;

---

<sup>1</sup>That is not to say that the theories are not motivated and inspired by rational considerations.

Franke, 2009), which are presented as representations of rational human interaction.<sup>2</sup>

A new approach is explored here. Recent work in cognitive psychology has suggested that probability theory, combined with structured cognitive representations, provides the proper tools to model diverse cognitive phenomena, including reasoning (Tenenbaum et al., 2007; Carey, 1985; Murphy and Medin, 1985; Gopnik and Meltzoff, 1997). The idea that this should be so is not particularly new; the earliest work in probability theory was intended as a way to model proper human reasoning under uncertainty. However, for most of the history of cognitive science, probability theory has been shunned. The first of two main reasons for this is Chomsky's 1957 early argument, contrary to the programs of structuralists and behaviorists, that probabilistic models cannot begin to predict the generative power of human language: a "statistical model for grammaticalness," Chomsky argued, would necessarily treat the grammatical sentence "Colorless green ideas sleep furiously" the same as the ungrammatical "Furiously sleep ideas green colorless," since they are both observed with a frequency of 0. The second is experimental work, starting with studies in the 1980s by Kahneman and Tversky, apparently demonstrating that humans violate the laws of probability theory in certain predictable ways when making judgments of likelihood (see, e.g., Kahneman and Tversky 1984). The present work does not aim to challenge Chomsky's arguments; indeed, it is an underlying assumption of the present work that a (basically non-probabilistic) generative grammar delivers model-theoretic meanings to the reasoning system.<sup>3</sup> Kahneman and Tversky's claims have much more direct relevance to the present thesis: if humans do not actually reason according to the laws of probability, how can Grice's program be advanced by a probabilistic theory of

---

<sup>2</sup>Of course, there is a logical underpinning to all theories—indeed, even theories of syntax—and logic, as I note, can be seen as a normative model of inference. But it should be clear that syntax is not understood as a product of rationality.

<sup>3</sup>This assumption is not critical; indeed, the nature of the system that builds sentences and their meanings is basically irrelevant to the present endeavor: the system can be probabilistic or deterministic, as long as it associates a model-theoretic meaning with sentences. Also note that Pereira (2000) contends that Chomsky's argument "relies on the unstated assumption that any probabilistic model necessarily assigns zero probability to unseen events," and that models that avoid this flaw using smoothing methods had been published prior to Chomsky's critique.

implicature computation? I argue below that this is not the right question. Under certain assumptions, probability theory is *the* theory of rationality under uncertain information, so *any* theory of implicatures with a truly Gricean foundation *must* be a probabilistic one.

In order to pursue this program, in this chapter, I provide arguments that probability theory is a reasonable way to model rational inference and lay out the formal groundwork necessary to address the problem of scalar implicature within a probabilistic setting. First, the mathematical foundation of probability theory is briefly presented, followed by early twentieth century arguments (Ramsey, 1931; De Finetti, 1964) that probability theory is *the* theory of reasoning under uncertainty. Then recent work that uses probabilistic tools to model human cognition, particularly human reasoning, is discussed, and a parallel to Gricean implicature computation is drawn. Having thus motivated the project of using probabilistic tools for implicature computation, I lay out the formal specifics of the problem: what exactly is it that a probabilistic system for implicature computation ought to generate? The definition of implicature within a probabilistic theory of reasoning has a simple and immediate consequence: implicature generation depends entirely on conversants' theories of conversation, in particular theories of how speaker beliefs influence a hearer's expectations about the speaker's choice of utterance. Given this way of looking at the problem, a criterion for presence of scalar implicature is derived: A hearer infers that a speaker who utters a weak sentence does not believe a stronger alternative is true whenever assuming the speaker believes the stronger alternative reduces the hearer's expectation that the speaker will utter the weak alternative.<sup>4</sup>

## 2.2 Probabilistic preliminaries

The origin of probability theory lies in the observation that beliefs are not usually black-and-white: I'm not sure whether Albert Pujols will hit a home run tomorrow, or whether

---

<sup>4</sup>See discussion in section 2.5 and Section 3.8.2 of the distinction between inference and implicature; the assumption in the present work is that inference is the basic phenomenon.

he hit one yesterday; I'm not even sure whether it's still baseball season. If someone asks me any of these questions, I will admit that I don't know the answer. But my beliefs about these propositions are not simply undefined: I think I heard someone talking about Albert Pujols hitting a home run yesterday, so it is more likely, in my mind, that he hit one yesterday than that he will hit one tomorrow. The proposition that he'll hit one on both days is also certainly less likely to be true than the proposition that he hit one on either, and the proposition that he'll either hit one or he won't is likeliest of all; it can't turn out to be false. And suppose I find out it's not baseball season any more: I know I should then ratchet down my degree of belief in the proposition that Pujols will hit a home run tomorrow. These considerations lead to the assumption that the strength of a belief is a question of degree, with degrees linearly ordered with respect to each other. There are indefinitely many possible proposition-ordering functions; of these, it has been argued that only those that satisfy the axioms of probability theory can represent degrees of belief for a rational agent. Probability theory viewed as a system for graded belief of a rational agent is generally known as *Bayesian* probability theory.

Before giving the traditional arguments for a Bayesian approach to rationality, I give the formal details of probability theory. Probability theory can be understood as a function that assigns degrees of belief to propositions, or sets of worlds. Given the set of all possible worlds  $\Omega$ , we can say that a function  $f$  whose domain is the power set of  $\Omega$  (i.e., the set of propositions) and whose codomain is the real numbers (so that  $f : \mathcal{P}(\Omega) \rightarrow \mathbb{R}$ ) satisfies the axioms of probability theory (Kolmogorov's axioms) iff:<sup>5</sup>

1.  $0 \leq f(E)$  for all  $E \subseteq \Omega$ .
2.  $f(\Omega) = 1$ .

---

<sup>5</sup>These axioms are usually given in full generality over  $\sigma$ -algebras, but we have no application here for probability functions defined over anything smaller than the full power set of the set of worlds.



3. For any sets  $E_1, E_2$ , each contained in  $\Omega$ , where  $E_1 \cap E_2 = \emptyset$ ,

$$f(E_1 \cup E_2) = f(E_1) + f(E_2).^6 \quad (2.1)$$

If we think of the degree of belief in a proposition as its weight, a probability measure is one that assigns each proposition a non-negative weight; the weight of the entire space of worlds is 1; and the weight of the union of disjoint propositions is equal to the sum of their weights. Each of these axioms, even the somewhat more complex last axiom, has a strong intuitive justification: (1) there is a degree of belief lower than all others, equal to the degree of belief in a contradiction; (2) it is a certainty that the entire space contains the real world somewhere in it; and (3) the degree of belief in the union of disjoint propositions (the proposition that at least one of the propositions is true) is simply the sum of the degrees of belief in each proposition individually (note that this depends on the assumption that the propositions are *disjoint*; otherwise we would “count” the weight of the intersection twice).

Given these axioms, *conditional probability* may be defined as follows:

$$P(A|B) = \frac{P(A \cap B)}{P(B)}. \quad (2.2)$$

The conditional probability  $P(A|B)$  corresponds to the degree of belief in  $A$ , given the assumption that  $B$  is true. The intuition behind conditional probabilities is that, by restricting the universe to a given proposition (conditionalizing on that proposition), it is possible to consider the probability of a second proposition in this narrowed universe. For example, we can conditionalize the proposition  $A = \text{that Harry will order pepperoni}$  on the proposition  $B = \text{that pepperoni is 50\% off}$ . This simply amounts to assuming  $B$  is true; then considering the probability of  $A$ . At the risk of over-explaining this point: suppose

---

<sup>6</sup>In order to prove certain theorems, a generalization from *finite* to *countable* additivity is needed: “For any countable sequence of sets  $E_1, E_2, \dots$ , each contained in  $\Omega$ , where  $E_m \cap E_n = \emptyset$  for all  $m \neq n$ ,  $f(\bigcup_i E_i) = \sum_i f(E_i)$ .” For our purposes, finite additivity is sufficient.

Harry prefers most other toppings to pepperoni, so his probability of ordering pepperoni is quite low, say  $P(A) = 0.05$ , a one-in-twenty chance. Now, suppose further that pepperoni is 50% off one day a week, so  $P(B) = \frac{1}{7} = 0.14$ . Now, suppose the probability of both propositions being true, pepperoni being 50% off and Harry ordering it, is not much lower than the probability of Harry ordering pepperoni (i.e., most of the worlds where Harry orders pepperoni are ones where it is 50% off), so  $P(A \cap B) = 0.04$ . This allows us to calculate the conditional probability  $P(A|B)$ , or the probability that Harry will order pepperoni, given that pepperoni is 50% off:

$$P(A|B) = \frac{P(A \cap B)}{P(B)} \quad (2.3)$$

$$= \frac{0.04}{0.14} \quad (2.4)$$

$$= 0.28 \quad (2.5)$$

This should demonstrate the close connection between atomic propositions, their intersections, and conditional probability.

Moreover, conditional probabilities may be inverted via a simple formula: Bayes' Rule, the centerpiece of Bayesian probability theory, given as follows:

$$P(H|E) = \frac{P(E|H)P(H)}{P(E)} \quad (2.6)$$

The propositional variables  $H$  and  $E$  can be understood to stand for *hypothesis* and *evidence*, respectively<sup>7</sup>. In applications, Bayes's Rule is used to figure out, given some evidence  $E$ , how likely a hypothesis  $H$  is to be its explanation; the rule says this can be calculated in terms of how likely the hypothesis and evidence were beforehand (their *prior probabilities*), as well as the *likelihood* that the evidence would be observed if, in fact, the hypothetical explanation were in effect.

---

<sup>7</sup>The proof of Bayes's Rule is a fairly direct consequence of the definition of conditional probability in equation 2.2 above:  $P(H|E) = \frac{P(H \cap E)}{P(E)} = \frac{P(H \cap E)}{P(E)} \frac{P(H)}{P(H)} = \frac{P(H \cap E)}{P(H)} \frac{P(H)}{P(E)} = P(E|H) \frac{P(H)}{P(E)}$ .

So, going back to the pepperoni example, suppose you observe Harry ordering pepperoni (the evidence), and you want to know why. Bayes's Rule provides a straightforward way to calculate how likely a hypothesized explanation (that pepperoni is 50% off) is, given this observed evidence. We know  $P(A) = 0.05$ ,  $P(B) = 0.14$ , and  $P(A|B) = 0.28$ . Plugging these values into the Bayes's Rule formula, we obtain:

$$P(B|A) = \frac{P(A|B)P(B)}{P(A)} \quad (2.7)$$

$$= \frac{0.28 \cdot 0.14}{.05} \quad (2.8)$$

$$= 0.8; \quad (2.9)$$

i.e., the probability that pepperoni is 50% off, given the evidence that Harry ordered pepperoni, is 0.8, or 80%. This reflects the intuition that the most likely explanation for Harry's ordering pepperoni is that pepperoni was discounted.

A somewhat more sophisticated example is provided in Pearl (2000). Suppose, at a casino, a person at the next gambling table calls out "twelve", and we want to know what the odds are that the person is playing roulette rather than dice. It is straightforward to directly calculate the probabilities  $P(\text{twelve}|\text{dice})$  and  $P(\text{twelve}|\text{roulette})$  (each game is a random event with a certain number of equally likely outcomes, only one of which is "twelve", so the probability is simply obtained by dividing 1 by the number of possible outcomes); it is not as easy to directly calculate the corresponding probabilities  $P(\text{dice}|\text{twelve})$  and  $P(\text{roulette}|\text{twelve})$ .<sup>8</sup> Bayes's rule makes the calculation trivial: the latter can be calculated directly in terms of the former.

---

<sup>8</sup>None of these probabilities, it seems to me, are ones ordinary people have intuitions about—even with special training (in the mathematics of probability), considerable reflection, if not pen-and-paper calculation, is necessary. The sorts of probabilities that people do have (qualitative) intuitions about are, hopefully, the ones needed for implicature computation. Numerical examples like this are given for purposes of illustration because it is easy to establish uncontroversial distributions for them.

$$\frac{P(dice|twelve)}{P(roulette|twelve)} = \frac{P(twelve|dice)P(dice)}{P(twelve|roulette)P(roulette)} = \frac{\frac{1}{36}P(dice)}{\frac{1}{38}P(roulette)} \quad (2.10)$$

If we assume the prior probabilities of dice and roulette are equal (there are the same number of dice tables as roulette tables), we get odds of 38 : 36 in favor of dice.

## 2.3 *The mathematics of rationality?*

Early in the 20th century, work by Ramsey (1931) and De Finetti (1964) argued that probability theory is the mathematics of reasoning under uncertainty, that no other theory correctly (normatively) models the inferences of a rational agent. Given certain assumptions, they showed that any measure of degrees of belief that did not adhere to the axioms of probability theory was necessarily an irrational measure: a gambler with such beliefs could have a “Dutch Book”, a system of bets that would ensure his loss, written for him. In these arguments, a normative gambler is one who is willing to place bets for any amount up to and including  $n$  dollars for a return of one dollar on a proposition that she believes to degree  $n$ ; as well as accept bets, with a promise to return a dollar if the proposition turns out to be true, for any amount greater than or equal to  $n$ . So, for example, a rational gambler will bet up to fifty cents for a dollar return on a fair coin toss; he will also take your bet for any amount greater than or equal to fifty cents, in return for his promise of paying out a dollar. The Dutch Book argument shows that such a gambler can only save himself from losing if his beliefs obey the laws of probability theory.

First, suppose the agent violates the first axiom by believing some proposition  $p$  to degree  $n < 0$ : then he should accept bets for  $n$  dollars, with a return of 1 dollar. That means he’s paying out of his pocket (since  $n < 0$ ) for the privilege of paying out an additional dollar if  $p$  turns out to be true. He always loses at least  $n$  dollars on such a bet,

and he loses  $n + 1$  if  $p$  turns out to be true. Suppose he violates the second axiom: he believes that a sure bet (the whole set of outcomes) ought to cost  $m$  dollars, where  $m$  is greater (or less) than 1. Then he must place (or take, if  $m < 1$ ) a bet for  $m$  dollars, in either case he is guaranteed to lose  $|1 - m|$ . Finally, suppose he violates the third axiom: that there are two mutually exclusive propositions  $p$  and  $q$  that he believes to degrees  $m$  and  $n$ , respectively, but that he believes  $p \cup q$  to degree  $k \neq m + n$ . That means either  $k > m + n$  or  $k < m + n$ . If the former holds, the agent has to place a bet for  $k$  on  $p \cup q$ , and take bets for  $m$  and  $n$  for  $p$  and  $q$ , respectively. He has, before the values of  $p$  or  $q$  are known, a balance of  $m + n - k$ , which is negative. After the values of  $p$  and  $q$  are revealed, no net exchange of money can take place (either  $p$  and  $q$  are both false, in which case there is no payoff on any bet, or just one is true, in which case the agent collects 1 and pays out 1, for a net of 0; they can't both be true, since they are mutually exclusive); so the agent's negative balance remains—he is guaranteed a loss. Analogously, if  $k < m + n$ , the agent can be made to place bets on  $p$  and  $q$ , and to take one for  $p \cup q$ , so that he has a negative balance—again, no change can take place when the values of  $p$  and  $q$  are revealed.

Dutch Book arguments have been widely criticized, especially in the philosophy literature. These critiques generally attack the assumptions needed for the argument to go through: that rational agents have linear ordered degrees of belief, or that rational agents must place and accept bets in the normative way prescribed by the arguments (Kyburg, 1978; Seidenfeld and Schervish, 1983). I cannot reply to these issues here except to say that none of the arguments demonstrates any obvious irrationality (i.e., a behavior that, after careful examination, an agent would have to agree was a mistake) follows from any of the assumptions. In the absence of a such an argument, probability theory can be taken as a normative theory of belief, if not *the* normative theory.<sup>9</sup>

Grice's theory is that we draw conversational inferences because it is the rational thing to do. Given the Dutch Book argument, the system of belief behind rational inferences

---

<sup>9</sup>An argument that a measure called *possibility* is normative may be found in Snow (2001).

is a probabilistic one. This does not entail that the actual cognitive/neural system behind pragmatic inference is fundamentally probabilistic, nor that human behavior generally conforms to the laws of probability theory. After all, people often do behave in demonstrably irrational ways—we make mistakes that we recognize as such, upon reflection. The source and consequences of this irrationality will not be explored in the present work. What will be explored are the consequences of a theory that assumes strict rationality in this sense. That is, I aim to develop a theory of pragmatics that assumes agents are rational (in a probabilistic way), and develop the consequences of this theory. If Grice was right, the rationality assumption, implemented formally in probabilistic terms, should be sufficient to generate a broad class of implicatures—irrationality will only lead to a failure to calculate implicatures correctly.

## **2.4 The new Bayesian cognitive psychology**

In addition to the classical Dutch Book arguments for probability, a large body of interrelated work by cognitive psychologists has begun to implicate a Bayesian basis for an array of cognitive functions, including memory, categorization, word learning, perception, planning, and, most relevant to the present endeavor, reasoning under uncertainty. These studies generally supplement traditional structured or symbolic representations with probabilistic models. Tenenbaum et al. (2007) describe the approach as “rational domain-general statistical inference guided by appropriately structured intuitive domain theories”, and credit the approach to Carey (1985), Murphy and Medin (1985), and Gopnik and Meltzoff (1997). Intuitive domain theories are simply mental representations of various domains: they may provide taxonomic relations between, say, a group of species with respect to genetics, directed graphs representing causal relations, syntactic parse trees, and others. Probabilistic methods inform three aspects of these domain theories: they explicate their actual content (knowledge is probabilistic); they allow the knowledge to be used

(inference is probabilistic); and they provide a means by which the knowledge may be acquired (learning is probabilistic) (see Tenenbaum and Griffiths 2001). (For the purpose of the present study, I ignore the third aspect: I assume the hearer can acquire the correct intuitive theory of conversation, and aim simply to identify that theory and show how it may be used to generate conversational inferences.)

Tenenbaum and Griffiths (2001) provide a simple example of a domain theory and an inference derived with it. A doctor wants to know what levels of a particular hormone are healthy. She examines a healthy patient whose hormone level is 60, on a scale of 0 to 100. The doctor's hypothesis space, then, is narrowed to hypotheses where 60 is a healthy level. Beyond that, however, how does the doctor use this partial information to reason about the likelihood of various other levels—say 40—being healthy? How does the value of 50 compare to 40 (it should be more likely)? How does the difference between 20 and 30 differ from that between 40 and 50? The simplified theory that the doctor brings to this domain is that there is a *range* of levels that are healthy: every number between an unknown minimum and an unknown maximum represents a health level. The probability that a given level is healthy decreases as that level gets more distant from 60. This is because every hypothesis entertained is one where 60 is healthy; in relatively fewer hypotheses is 50 healthy, and still fewer make 40 healthy. Given the doctor's domain theory, very partial information (that 60 is healthy) gives the rational doctor quite a lot of information about what other levels are likely to be healthy (and those that are not).

There is a parallel here to implicature generation: the doctor takes partial information that results from a hidden property and uses it to draw an inference about the content or character of that hidden property—hearers take partial information (a speaker's utterance), and use the information it conveys to draw conclusions about the hidden knowledge state of the speaker. But there is a deep difference in *how* the evidence emerges: in the case of the hormone level, the exemplar is assumed to be randomly selected from the healthy patients; in the case of implicature, a rational agent chooses to assert a specific proposition

from their set of beliefs, with a particular purpose in mind. In other ways, the domain of implicature generation is considerably more complex than the one the doctor faces: hypotheses are not limited in a straightforward way to “ranges” of values; rather, hearers must consider a field of propositions (or worlds), and draw an inference about how likely each proposition (or world) is to be in a speaker’s knowledge state. Nonetheless, the tools used for the simple type of inferential process discussed in Tenenbaum and Griffiths can be applied straightforwardly in a theory of implicature generation—all that differs is the domain theory.

It is instructive to examine the simple example in Tenenbaum and Griffiths in more detail before discussing the more complex domain theory needed for scalar implicature computation. The doctor’s intuitive theory of the domain is that there is an interval, a range of real numbers, that constitutes the proper generalization about healthy hormone levels. Each hypothesis the doctor considers, then, is an interval of numbers  $[m, n]$  (including hypotheses where there is a single healthy value, or intervals where  $m = n$ ).<sup>10</sup> For simplicity, we assume that the doctor’s instruments only measure to integer precision (following Tenenbaum and Griffiths, who later generalize the model to continuous intervals and probability densities), and that the hypotheses she considers are ranges  $[m, n]$  of integers. Formally, the doctor’s task is to evaluate the probability of each hypothesis  $h$  against the background theory  $T$  (that hormone levels have healthy *ranges*), and data  $e$  (that 60 is healthy):  $P(h|e, T)$ .

By Bayes’s Rule, the probability for any hypothesis  $h$  is:<sup>11</sup>

$$P(h|e, T) = \frac{P(e|h, T)P(h|T)}{\sum_{h' \in H_T} P(e|h', T)P(h'|T)}. \quad (2.11)$$

<sup>10</sup>The intuitive theory of the domain might have been different: it could have been that there are two disconnected healthy ranges, or that individuals in one population have a different healthy range than individuals in another population. These theories could be formalized, and this would lead to different inferences.

<sup>11</sup>Notice the background theory  $T$  is present here, though it was absent in Bayes’s Rule as presented in equation (2.6). Think of  $T$  as the universe being considered right now: one where there is an integer interval that constitutes the healthy range of hormones. Every probability is conditioned upon  $T$  (in addition to any other conditions), since every probability is being considered with respect to this universe.



Focusing on just the numerator (as the denominator is the same for every hypothesis), we can state this in terms of proportionality (the symbol  $\propto$  means *is proportional to*):

$$P(h|e, T) \propto P(e|h, T)P(h|T). \quad (2.12)$$

So, for a given hypothesis  $h = [m, n]$ , we have:

$$P(h = [m, n]|e, T) \propto P(e|h = [m, n], T)P(h = [m, n]|T). \quad (2.13)$$

The prior distribution over the hypotheses can be instantiated in a number of ways. The simplest is to give a uniform prior, where each hypothesis receives the same probability. Then, we have, simply,

$$P(h = [m, n]|e, T) \propto P(e|h = [m, n], T) \quad (2.14)$$

$$= \frac{1}{(n - m) + 1} \quad (2.15)$$

So the probability of a given hypothesis after the observation of a single healthy value is inversely proportional to its size (the number of integer values in the interval, or  $(n - m) + 1$ ). This means that the hypothesis where the interval contains just one value receives a higher probability than any other hypothesis. But there is just one such compatible hypothesis after an observation of a healthy value, whereas there are many compatible hypotheses that involve wider intervals.

A number of additional inferences may be made given this simple apparatus. One is that the probability that a given value is healthy is concave decreasing as that value's distance from the observed value  $e$  increases. More interesting is that the slope of this decrease becomes steeper as more healthy values are observed: ten healthy patients with values contained in  $[m, n]$  make healthy values outside  $[m, n]$  much less likely. The example demonstrates that Bayesian techniques applied to a very simple theory of a domain

(again, here that theory is that the values of healthy hormone levels form an interval) provide a way to reason about the hidden character of an entity in the environment. By analogy, similar methods could work to model the reasoning of hearers about the hidden knowledge states of speakers.

## 2.5 Conversational reasoning: terminology and notation

So what, exactly, is it that a hearer thinks is likely when he derives a scalar inference consequent to a speaker's utterance? The implicature relation is between utterances and propositions. Utterances are tied to contexts that specify speaker, hearer, time, place, and so on, so one can talk equally well about the speaker of an utterance implicating a certain proposition. Intermediate between propositions and utterances are sentences: every utterance has an associated sentence (or phrase or fragment); every (declarative) sentence has an associated proposition.<sup>12</sup> For this reason, it is sentences that will be Greek letter-bearers. The utterance of a sentence  $\nu$  in a context  $c$  will be notated with square quotes:  $\ulcorner \nu, c \urcorner$ . Usually, the context will be understood (though not left unspecified in the discussion); in this case, I will write  $\ulcorner \nu \urcorner$ . Note that  $\ulcorner \nu, c \urcorner$  is a proposition: the set of worlds where the speaker utters  $\nu$  in context  $c$ . As a shortcut, functions and operators that are ordinarily understood to take propositional arguments will take sentences directly: for example,  $B_s \nu$  means the belief operator  $B$  for the agent  $s$  applies to the proposition  $\llbracket \nu \rrbracket$  (making the simplifying decision to ignore syntactic ambiguity). When subscripts on beliefs are omitted, unless otherwise indicated, the speaker  $s$  is assumed to be the belief holder. Hopefully, this bracket-saving shortcut will not cause any confusion.

There is one terminological complaint a reader is likely to have that, it seems to me, is well-justified. I use *implicature* indiscriminately between its correct usage, roughly, a

---

<sup>12</sup>I will try to be careful not to talk about sentences implicating or being implicated, but I will do this occasionally, for reasons of brevity. In these instances, I mean the obvious thing: that the utterance of the sentence implicates, or that the proposition the sentence denotes is implicated.

proposition a speaker means without saying; and its other common usage, the *inference* that a hearer makes, given a speaker's utterance. The present study takes a hearer-centric approach to implicature; in this approach, the only way to even talk about a speaker implicating a proposition is to assume they can work out the inference a hearer is likely to draw from their utterance, and that they make the utterance with the intention that the hearer will draw that inference. In this sense, conversational implicature is more complex than conversational inference. To simplify matters, I focus here on the development of a theory of conversational inference, leaving aside the questions of whether the speaker *intended* for the hearer to draw an inference.<sup>13</sup> But because the phenomenon that is discussed in the literature is conversational *implicature*, I will often use that term instead.

The calculation of implicatures in a probabilistic framework requires a theory that predicts, in a given context, that the hearer has a sufficiently high degree of belief in a proposition about the speaker's beliefs, conditional upon the speaker's utterance. Scalar implicatures are fundamentally epistemic weak propositions: a speaker who has failed to assert a strong proposition is thought not to be in a position to assert it. In other words, the strong proposition is not entailed by the speaker's belief state—for all the speaker knows, the stronger proposition could be false. It is often said that a sentence with *some* implicates the negation of the corresponding sentence with *all*, but I assume here that this inference is only arrived at through a strengthening of the epistemic weak proposition: to the extent that the speaker is opinionated,<sup>14</sup> the implicature can be strengthened to epistemic strong, and if he is well-informed, it can be enriched to include simple negation.

The first careful identification of the epistemic weak nature of scalar implicature is in a response by Soames (1982) to Gazdar's (1979) claim that scalar inference consists in gen-

<sup>13</sup>See Gauker (2001) for an argument that inference is the basic phenomenon, and that many hearer-generated inferences are mis-labelled as speaker-generated implicatures. See also discussion in Section 3.8.2 below.

<sup>14</sup>The opinionatedness assumption—that the hearer believes  $B_s x \vee B_s \neg x$ —has also been called an *ex-perthood* assumption (Sauerland, 2005), an *authority* assumption (Zimmermann, 2000; Geurts, 2005), and a *competence* assumption (van Rooij and Schulz, 2004; Franke, 2009). See also Soames (1982) and Harnish (1976).

eration of potential implicatures of the form  $K_s \neg \chi$  (the speaker knows  $\chi$  is false), which may then be discarded or kept, in which case they become actual implicatures. Soames points out that epistemic weak scalar implicatures are common, and that epistemic strong ones are only derived when there is reason to assume that the speaker knows whether the stronger proposition is true. The example he gives is an utterance of

(2.16)     Some of the birds we tagged last summer have migrated to California.

Soames notes that “it is easy to imagine contexts in which someone who uttered [(2.16)] would not be presumed to know the truth value of” the stronger alternative. Subsequent works, including Horn (1989), Sauerland (2004), and Russell (2006) provide further arguments for the importance of deriving epistemic weak implicatures prior to the derivation of epistemic strong implicatures (see discussion in section 1.3.1). So the basic task of a formal theory of scalar implicature generation will be to derive inferences of the form  $\neg B\sigma$ , where  $\sigma$  is a stronger alternative to the speaker’s utterance. Before we can talk about how to do that in a probabilistic system, I will lay out formal preliminaries for the requisite epistemic operators.

It is straightforward to represent epistemic weak/strong distinctions in a probabilistic theory of implicature computation, since probability distributions are defined over the set  $\Omega$  of possible worlds, which means that the events we reason about are the same objects as the propositions built by a compositional model-theoretic semantics. The proposition that an individual  $a$  believes proposition  $q$  is simply the set of worlds where every world compatible with  $a$ ’s belief is an element of  $q$ . This can be formalized (along the same lines as formal semantic treatments of belief) as follows: an accessibility relation maps an individual  $a$  and a world  $w$  to a set of worlds: the smallest set of worlds that are compatible with  $a$ ’s beliefs in world  $w$ . Such a relation, notated  $R_a[w]$ , will be taken as primitive.

The belief operator  $B$  can then be defined:

$$B_a q := \{w : R_a[w] \subseteq q\} \quad (2.17)$$

Such propositions, being sets of worlds, receive probabilities in individuals' distributions. This means that interlocutors have beliefs about each others' beliefs: in a given conversation, the hearer can think it's probably true that the speaker thinks that George left.<sup>15</sup>

I remain agnostic for now about what, if anything, assertions (and other speech acts) “do”: that is, what sorts of conversational representations do we have? There is fine evidence (Stalnaker, 1978, 2002) for the existence of *common ground*, but this will not be crucial to the development of the theory here.<sup>16</sup> Instead, we simply start with a probabilistic system of belief wherein agents can also represent the (categorical) beliefs of other agents.

Putting all the notation together, a theory of utterance expectation that allows hearers to calculate conversational implicatures should provide values for conditional probabilities

---

<sup>15</sup>It may seem strange to focus on categorical speaker beliefs, given the focus on graded hearer beliefs being pursued here. There are two reasons for this. First, the introduction of probability distributions over probability distributions would add unneeded complexity to the theory (see Shirazi and Amir 2007 for the development of a probabilistic modal logic and demonstration that the system is intractable). The second reason is the nature of the reasoning task: the hearer is reasoning about whether the speaker's belief state allows him to assert some proposition  $p$ . The relevant (idealized) question about those belief states is: does the speaker believe  $p$  in that belief state? So the approach taken here is that hearers have probabilistic beliefs about speakers' categorical beliefs.

<sup>16</sup>The use of probabilistic belief states suggests the possibility of probabilistic common ground, something Merin (1999) has called *common prior*. Both common ground and common prior are potentially compatible with the approach taken here. Here are three possibilities: speaker and hearer each represent a basically Stalnakerian common ground, which acts as the domain for assertions, which are functions that return common ground intersected with the assertion's associated proposition. From here, hearers have to figure out why speakers put a particular proposition in the common ground—inferences to this effect taking place in individuals' belief states (though perhaps some of these inferences also affect the common ground, so that it is common knowledge after you say “ $x$ ” that you want me to believe  $x$ ). A second possibility is that common ground is replaced by an analogous probabilistic representation, or common prior. In such an approach, the effect of an assertion is to conditionalize the common prior on that assertion; it should be noted that common prior may raise some deep formal problems (Aumann, 1976). A final possibility is to do away with common representations altogether, and say that agents have knowledge about conversation, so that when a speaker asserts something, a hearer infers that the speaker intends for him to believe that proposition, and that the speaker believes the proposition, and so on.

of the form

$$P(\ulcorner \chi \urcorner | B\nu),$$

where  $\chi$  and  $\nu$  are arbitrary sentences. In other words, the intuitive theory of utterance expectation should, given inputs from the context, provide hearers with estimates of how likely it is that the speaker will say “ $\chi$ ”, given the assumption that the speaker believes  $\nu$ .

## 2.6 Defining implicature probabilistically

In the previous section, I posited that a probabilistic theory of implicature will deliver a sufficiently high probability (in the hearer’s beliefs) for some speaker belief  $\mathcal{B}$ , conditional upon the assumption that the speaker uttered some sentence  $\nu$ . In the following discussion, three ways of measuring *sufficiently high* will be examined; only one seems adequate. For now, I want to talk generally about conversational implicatures, applying the general characterization of implicatures to the scalar case later (I will use  $\ulcorner \nu \urcorner$  and  $\mathcal{B}$  discussion of conversational implicatures generally, and  $\ulcorner \omega \urcorner$  and  $\neg B\sigma$  for the scalar case).

A first possibility for the characterization of “sufficiently high” is that a proposition is implicated by the utterance of a sentence just in case the likelihood of the truth of the proposition, given the utterance, is high (again, because hearer perspective is taken,  $P(\mathcal{B})$  refers to the hearer’s degree of belief that the proposition  $\mathcal{B}$  about the speaker’s beliefs is true). Formally, this would be:

(2.18) Definition of implicature (first attempt):

The utterance of a sentence  $\nu$  *conversationally implicates*  $\mathcal{B}$  iff  $P(\mathcal{B} | \ulcorner \nu \urcorner) > n$ ,  
for some contextually determined standard  $n$ .

But the implicature relation intuitively has to do with an utterance *making* a particular proposition more likely, not that the proposition is simply highly probable after the speaker

makes the utterance. The definition in (2.18) will, for example, make every utterance implicate everything that is already presumed true, like that the moon is not made of green cheese. An utterance might even make such a proposition *less* likely, but given the definition in (2.18), if it is still sufficiently likely, it will qualify as an implicature. For example, in a discussion between conspiracy theorists, a contribution of, “The moon isn’t really composed of the minerals they say it is,” would, presumably, lead hearers to belief states where green cheese is still very unlikely to be the moon stuff. But it may be more likely than it was when the hearer believed the moon was made up of a known set of minerals. It seems incorrect, in such a scenario, to suppose that the speaker has implicated that the moon is not made of green cheese.<sup>17</sup>

This suggests that the theory must account for the intuition that, when an utterance of  $\nu$  implicates  $\mathcal{B}$ ,  $\mathcal{B}$  has to be *more likely* to be true given the fact that the speaker said  $\nu$  than it was before the speaker said  $\nu$ .

---

<sup>17</sup> A statistical approach underscores this point for the domain of scalar implicatures: under minimal simplifying assumptions, propositions with *all* are statistically dramatically less likely than the corresponding propositions with *some*. Consider situations where each individual in a domain’s inclusion in a given predicate is independent of every other individual’s inclusion, and each individual has the same probability  $q$  of inclusion. Then the probability that all individuals in the domain are included in the predicate is

$$P(all) = q^i, \quad (2.19)$$

where  $i$  is the number of individuals in the domain. The probability that none of the individuals are included in the predicate is

$$P(none) = (1 - q)^i, \quad (2.20)$$

so the probability that at least some individuals are included is

$$P(some) = 1 - P(none) = 1 - (1 - q)^i, \quad (2.21)$$

and the probability that some but not all are included is

$$P(some \wedge \neg all) = P(some) - P(all) = (1 - (1 - q)^i) - q^i. \quad (2.22)$$

Suppose the predicate *show up at the party* has the above properties (independence and equal probability among individuals), that each individual has a probability of 0.5 of showing up, and that Pete has six friends. Using the formula above, the probability that at least some of Pete’s friends will show up is 0.984; the probability they’ll all show up is 0.0156. The probability that some but not all will show up is 0.969, and the odds of *some but not all* to *all* showing up are sixty-two to one. Here is a table showing the odds for worlds where Pete has two to sixteen friends.

## (2.23) Definition of implicature (second attempt):

The utterance of a sentence  $\nu$  *conversationally implicates*  $\mathcal{B}$  iff  $P(\mathcal{B}|\ulcorner \nu \urcorner) > P(\mathcal{B})$ .

This definition does away with  $n$ , so a proposition does not simply have to exceed some constant value to be implicated. Instead, implicatures are propositions that become *more* likely following a speaker's utterance:  $\ulcorner \nu \urcorner$  implicates  $\mathcal{B}$  iff the probability of the speaker having belief  $\mathcal{B}$ , given his utterance of  $\nu$ , is greater than the plain probability of the speaker having belief  $\mathcal{B}$ . This move seems to be a hyper-correction: now, the conspiracy theorist in the green cheese example above implicates that the moon *is* made of green cheese. But this is simply a consequence of the literal interpretation of the speaker's utterance: by eliminating worlds where the moon is made of the usual minerals, the speaker has also eliminated worlds where the moon is not made of green cheese, making it more likely that the moon is made of green cheese. Implicatures are more than just the consequences of literal interpretation.

(2.i)	elements	$P(all)$	$P(some)$	$P(some \wedge \neg all)$	$some \wedge \neg all : all$
	2	0.25	0.75	0.5	2:1
	3	0.125	0.875	0.75	6:1
	4	0.0625	0.9375	0.875	14:1
	5	0.0312	0.9687	0.9375	30:1
	6	0.0156	0.9843	0.9687	62:1
	7	0.0078	0.9921	0.9843	126:1
	8	0.0039	0.9960	0.9921	254:1
	9	0.0019	0.9980	0.9960	510:1
	10	0.0009	0.9990	0.9980	1022:1
	11	0.0004	0.9995	0.9990	2046:1
	12	0.0002	0.9997	0.9995	4094:1
	13	0.0001	0.9998	0.9997	8190:1
	14	0.0000	0.9999	0.9998	16382:1
	15	0.0000	0.9999	0.9999	32766:1
	16	0.0000	0.9999	0.9999	65534:1

In the last column are the odds ratios in favor of *some but not all*: even in a domain with only 3 elements, *some but not all* is six times as likely as *all*, and the odds only improve as the number of elements increases (*some* isn't really used for domains with fewer than three or four elements). So, under reasonable assumptions, speakers are much less likely to believe *all* than they are to believe *some but not all*. So the definition in (2.18) would predict scalar implicature arises in most cases, simply as a consequence of brute statistical fact and the literal meanings of *some* and *all*.



Indeed, the notion of literal interpretation is central to the concept of implicature. An implicated proposition is not entailed by the literal meaning of an utterance; rather, an implicature magnifies a subset of the literal meaning. A rough way of looking at the literal interpretation of an utterance  $\ulcorner \nu \urcorner$  is simply conditionalizing on the proposition  $B\nu$ , that the speaker believes what he said. Within  $B\nu$  is included the possibility that the speaker further has belief  $\mathcal{B}$  and the possibility that the speaker has belief  $\neg\mathcal{B}$  (assuming that both propositions are compatible with  $B\nu$ ). Prior to the speaker's utterance of  $\nu$ , the hearer's belief distribution assigns a particular ratio to this bipartition of  $B\nu$ ; this ratio is  $\frac{P(B\nu \wedge \mathcal{B})}{P(B\nu \wedge \neg\mathcal{B})}$ . If the hearer simply finds out the speaker believes  $\nu$ , there is no shift in this ratio:<sup>18</sup>

$$\frac{P(B\nu \wedge \mathcal{B}|B\nu)}{P(B\nu \wedge \neg\mathcal{B}|B\nu)} = \frac{P(B\nu \wedge \mathcal{B})}{P(B\nu)} \frac{P(B\nu)}{P(B\nu \wedge \neg\mathcal{B})} \quad (2.24)$$

$$= \frac{P(B\nu \wedge \mathcal{B})}{P(B\nu \wedge \neg\mathcal{B})} \quad (2.25)$$

But implicature does involve a shift in this ratio: not only does the speaker's utterance indicate  $B\nu$ ; it also reduces the proportion of the partition occupied by  $\mathcal{B}$ . This gives the following definition of implicature:

(2.26) Definition of Implicature

Given a set  $\mathcal{B}$ , an utterance of  $\nu$  implicates  $\mathcal{B}$  just in case

$$\frac{P(B\nu \wedge \mathcal{B}|\ulcorner \nu \urcorner)}{P(B\nu \wedge \neg\mathcal{B}|\ulcorner \nu \urcorner)} > \frac{P(B\nu \wedge \mathcal{B})}{P(B\nu \wedge \neg\mathcal{B})}$$

This says that an utterance of  $\nu$  implicates  $\mathcal{B}$  if the fact of its utterance, over and above the simple assumption that the speaker believes  $\nu$ , increases the probability of  $\mathcal{B}$ .<sup>19</sup> Applying

<sup>18</sup>Note for readers with limited exposure to probability: conditional probabilities always have a single |, and everything to the left of the bar must be “parsed” together:  $P((B\nu \wedge \mathcal{B})|B\nu)$  is the only parse for  $P(B\nu \wedge \mathcal{B}|B\nu)$ .

<sup>19</sup>Notice here that the hearer is reasoning about whether an utterance of  $\ulcorner \nu \urcorner$  implicates  $\mathcal{B}$ . We only consider non-trivial cases where  $B\nu$  is possible and has a non-null intersection with  $\mathcal{B}$  and  $\neg\mathcal{B}$ ; this avoids any divide-by-zero problems.

the definition in (2.26) to the green cheese case, no implicature would be expected. The speaker's assertion tells the hearer nothing more about his belief that the moon is made of green cheese than simply conditionalizing on the speaker's belief in non-standard moon stuff.

A feature of this definition is that implicatures may be stronger or weaker, depending on the circumstance. A potential drawback is that, if the difference between the two ratios is minuscule, the implicature may be so weak that it is not intuitively detectable, even though (2.26) indicates an implicature is present. The flip side of this is that the system will make predictions about the relative strength of implicatures: if the system predicts  $P(\mathcal{B}|\ulcorner \nu, c \urcorner) > P(\mathcal{B}|\ulcorner \nu', c \urcorner)$  then an utterance of  $\nu$  ought to implicate  $\mathcal{B}$  more strongly in context  $c$  than an utterance of  $\nu'$ .

A conceptual peculiarity in the definition is that a lot of facts that are related to an utterance end up being implicatures. So, for example, the proposition *that the speaker believes he said  $\nu$*  will count as an implicature: certainly it is more likely, conditional upon the assumption that the speaker said  $\nu$ , than the proposition *that the speaker believes he did not say  $\nu$* , given that the speaker said  $\nu$ , whereas the prior probability (given no assumptions about what the speaker said) of the speaker's belief that he said  $\nu$  is presumably lower than the probability that the speaker does not believe he said  $\nu$ . I am inclined to think that such peculiarities are basically innocuous.

## 2.7 Calculating scalar implicatures using Bayes's Rule

In the realm of conversation, speakers undoubtedly impose different intuitive theories depending on the task at hand. I will argue here that only one theory, that of utterance expectation, is necessary for the computation of conversational implicatures. Utterance expectation is at the heart of informal discussion of Gricean scalar reasoning: the speaker has said something less informative *than he would have been expected to* on the assumption that

he was in a position to assert something more informative. Therefore, the assumption that the speaker had knowledge of the more informative proposition must be false—i.e., the speaker’s utterance of a weak sentence implicates the speaker does not believe a stronger proposition.

A crucial point in this reasoning process is the calculation of what the speaker would have been expected to say on the assumption that he had a particular belief. Traditional neo-Gricean approaches hedge their bets on this point: the speaker is expected to say the most informative thing possible, as long as that doesn’t make them say something too irrelevant, or too prolix, or untrue. In most cases it is hard to predict which of Grice’s maxims a speaker will prioritize, and discussion is, instead, limited to (often illuminating) post-hoc analyses of which maxims were at play in leading a speaker to utter what he uttered. Probability theory gives us a way to make clear predictions about such matters: a formal theory of what speakers with certain beliefs are expected to say in certain conversational settings combines with general inference mechanisms to yield implicatures.

In section 2.6, I submitted that a theory of implicature generation should predict that

$$\frac{P(B\nu \wedge \mathcal{B}|\ulcorner \nu \urcorner)}{P(B\nu \wedge \neg \mathcal{B}|\ulcorner \nu \urcorner)} > \frac{P(B\nu \wedge \mathcal{B})}{P(B\nu \wedge \neg \mathcal{B})} \quad (2.27)$$

when an utterance of  $\nu$  produces an observed implicature of  $\mathcal{B}$ . In scalar terms, we are interested in the case where a weak sentence  $\omega$  has a stronger alternative  $\sigma$  and an utterance of  $\omega$  produces an implicature of  $\neg B\sigma$ . So, applying the formula above, we need a theory that predicts:

$$\frac{P(B\omega \wedge \neg B\sigma|\ulcorner \omega \urcorner)}{P(B\sigma|\ulcorner \omega \urcorner)} > \frac{P(B\omega \wedge \neg B\sigma)}{P(B\sigma)} \quad (2.28)$$

In light of the earlier discussion of Tenenbaum and Griffiths, this may be recast as a problem of reasoning about the posterior probability of a latent cause (speaker’s lack of belief in  $\sigma$ ), given the observation that the speaker uttered  $\omega$ . To compute this, a speaker can

apply Bayes's Rule, to obtain:

$$P(B\omega \wedge \neg B\sigma | \ulcorner \omega \urcorner) = \frac{P(\ulcorner \omega \urcorner | B\omega \wedge \neg B\sigma)P(B\omega \wedge \neg B\sigma)}{P(\ulcorner \omega \urcorner)}, \quad (2.29)$$

for the numerator, and

$$P(B\sigma | \ulcorner \omega \urcorner) = \frac{P(\ulcorner \omega \urcorner | B\sigma)P(B\sigma)}{P(\ulcorner \omega \urcorner)} \quad (2.30)$$

for the denominator. Putting them together, the term  $P(\ulcorner \omega \urcorner)$  cancels, and we are left with the following condition on scalar implicature:

$$\frac{P(\ulcorner \omega \urcorner | B\omega \wedge \neg B\sigma)P(B\omega \wedge \neg B\sigma)}{P(\ulcorner \omega \urcorner | B\sigma)P(B\sigma)} > \frac{P(B\omega \wedge \neg B\sigma)}{P(B\sigma)}. \quad (2.31)$$

This further simplifies to:

$$\frac{P(\ulcorner \omega \urcorner | B\omega \wedge \neg B\sigma)}{P(\ulcorner \omega \urcorner | B\sigma)} > 1, \quad (2.32)$$

which is equivalent to:

$$P(\ulcorner \omega \urcorner | B\omega \wedge \neg B\sigma) > P(\ulcorner \omega \urcorner | B\sigma). \quad (2.33)$$

What this means is that scalar implicature, from a probabilistic standpoint, is present when the following criterion is satisfied:

(2.34) Scalar implicature criterion (simplified):

Let  $\sigma \subset \omega$  be sentences. Then an utterance of  $\omega$  implicates  $\neg B\sigma$  if and only if:

$$P(\ulcorner \omega \urcorner | B\omega \wedge \neg B\sigma) > P(\ulcorner \omega \urcorner | B\sigma). \quad (2.35)$$

This, in turn, means that whenever an utterance of a weak sentence is less likely, given that the speaker believes a stronger alternative is true, than it is given that the speaker believes

the weak sentence but not the strong sentence, the theory predicts that a scalar implicature is present. Moreover, when this condition does not hold (or to the extent that the difference in probabilities is small), no scalar implicature is predicted (or the implicature is predicted to be felt weakly).

Stepping back, the simple move of applying Bayesian reasoning to the problem of implicature generation (in the most straightforward way possible), yields a theory that makes implicature computation a simple matter of comparing two conditional probabilities. One is the probability that a speaker will make a weak utterance, given a weak belief; the other is the probability that the speaker will make the weak utterance, given a relatively strong belief.

# Chapter 3

## Relevance in pragmatic theory

### 3.1 Introduction

In the previous chapter, I laid out the groundwork for the probabilistic generation of scalar implicatures: hearers derive scalar implicatures as a result of their theories of conversation, which lead them to expect certain utterances over others, depending on the context. It is not controversial to say that hearers expect speakers to make a contribution that is relevant in the context. In this chapter, I explore the empirical and formal properties of relevance, establishing a new generalization about constraints placed upon assertions by relevance. This sets the stage for the incorporation of relevance into a probabilistic theory of implicature in Chapter 4.

### 3.2 Positive relevance and assertion

#### 3.2.1 A constraint on assertions

Much of the literature on scalar implicature implicitly or explicitly assumes that the message communicated by a weak sentence  $\omega$  is equivalent to the scalar-strengthened sentence

$\omega \wedge \neg\sigma$ . So, for example, a speaker communicates the same thing by uttering (3.1) as he would by uttering (3.2a) or (3.2b):

(3.1) Some of the students passed the exam.

(3.2) a. Some but not all of the students passed the exam.

b. Some of the students passed the exam, but not all of them did.

But the set of contexts where (3.1) may be uttered is different from the contexts for (3.2a) or (3.2b). In particular, the latter two may be uttered when the negation of the stronger sentence is *the point* of the speaker's utterance, but (3.1) apparently may not. Consider a context where you and I are high school teachers, and a good performance by our students on a standardized exam helps us to qualify for a particular grant. I've just come back from grading the exams, and I say to you:

(3.3) a. Crap! Not all of our students passed!

b. Crap! Only some of our students passed!

c. Crap! Some but not all of our students passed!

d. Crap! Some of our students passed, but not all of them did!

e. #Crap! Some of our students passed!

In (3.3a)–(3.3d), it is part of the assertion that not all of the students passed, a proposition with undesirable consequences.<sup>1</sup> But in an ordinary utterance of (3.3e), this is part of implicated meaning. The badness of (3.3e) shows that *some* and *some but not all* are not interchangeable: the contexts that support the use of *some* are not the same as those that support *some but not all*. In particular, it seems the latter, but not the former, may be used

---

<sup>1</sup>This is related to the observation in Horn (2002) that *almost* can be used for good news when *barely* is used for bad news. In Horn's terms, the assertoric inertia of the sentence uttered has to be aligned with the interjection.

in this context to deliver bad news.<sup>2</sup>

Conversely, note also that *some* may be used in contexts where *not all* is unacceptable: in the context above, my breaking of good news shows the complementary pattern:

- (3.4)    a. #Whew! Not all our students passed!  
           b. Whew! Not all but some of our students passed!<sup>3</sup>  
           c. Whew! Some of our students passed.

When it is felicitous, the utterance of (3.4a) usually implicates that some of the students passed. And, indeed, the utterance of (3.4c) implicates that not all of the students passed. But it is only (3.4c), and not (3.4a), that may be used in this context. Again, we see that it is the asserted content, not implicated content, that must be a relief for the speaker.

These facts, upon cursory inspection, may be summarized as follows: when a speaker prefaces an assertion with an interjection like *Crap*, it is the content of the assertion, not the implicature, that must be aligned with the polarity of the interjection. The proposal I will make in this chapter is that the facts in (3.3) are a consequence of the organization of conversation according to the point being made by the speaker, specifically the fact that speakers seek to maximize the relevance of the assertions they make. To flesh out this proposal about conversation, an excursion into the proper characterization of relevance is necessary. With a suitable formal definition of relevance in hand, the empirical generalization may be stated simply; from there, the role of relevance in the analysis of other phenomena, including conventional implicature, is demonstrated.

<sup>2</sup>Some speakers find (3.3e) marginally acceptable with a pitch accent on *some*. This fact is potentially related to cases of “metalinguistic negation” discussed by Horn (1985).

<sup>3</sup>For discussion of the reversed order of *not all* and *some* here, see the discussion of *but* in Section 3.3.1 below. Kai von Stechow (p.c.) notes that the following sequence is acceptable: “Crap! Some of our students failed, but at least not all of them did.” Here it seems the point signaled by the interjection is only in effect for the first conjunct; one can imagine a parenthetical *whew!* after *but*.



### 3.2.2 Carnap's definition of relevance

The earliest formal definition of relevance is due to Carnap (1950), who defines it as a relation between propositions (more precisely, as a function from pairs of propositions to degrees, or real numbers): a proposition  $E$  is relevant to another proposition  $H$  to the degree that learning its truth would affect the probability assigned to  $H$ . Formally, this is:

(3.5) Definition of (*Carnap*) *relevance*:

$$r_H(E) = P(H|E) - P(H)$$

The value returned by  $r$  is a measure of how dramatically belief in  $H$  would be changed by learning  $E$ : how different  $P(H|E)$  is from  $P(H)$ . On this definition, one proposition is *relevant* to another iff it receives a non-zero  $r$ -value with respect to that proposition. Consider the proposition *that it will rain tomorrow*. Suppose you think its odds of being true are even, that its probability is 0.5. Now, if an acquaintance tells you, “It’s supposed to rain tomorrow,” your degree of belief in the proposition *that it will rain tomorrow* (call this  $H$ ) will increase—you’ve received positively relevant information about  $H$ . On the other hand, an utterance of “They say it’ll be nice for the next few days” should cause your degree of belief in  $H$  to decrease—this is what it means for information to be negatively relevant. Negative  $r$  values mean negative relevance; positive values positive. Moreover, the numerical  $r$  values of Carnap relevance provide a relevance ordering of propositions with respect to a given proposition  $H$ : given two propositions, the one whose assumption increases  $P(H)$  more is more relevant to  $H$ . If they both have the same effect on  $H$ , they are equally relevant to  $H$ .

In the example (3.3) above, the proposition *that not all of our students passed* is positively relevant to the proposition *that we will not receive the grant*. In terms of Carnap relevance, the probability that we will not receive the grant given that not all of our stu-

dents passed is higher than the probability that *we will not receive the grant* received prior to the information that not all of the students passed, so the  $r$ -value is greater than 0. On the other hand, the proposition *that some of our students passed* is negatively relevant to the proposition *that we will not receive the grant* (i.e., it is positively relevant to the proposition *that we will receive the grant*, following the identity proved in footnote 7 below).

The role of a given point  $H$  to which an assertion  $X$  is expected to be Carnap relevant can be understood within a broader theory of discourse structure. According to the theory developed in Roberts (1996), discourse is hierarchically organized into superquestions and subquestions, with assertions contributing partial answers to questions under discussion (see also Ginzburg; van Kuppevelt 1996; 1996).<sup>4</sup> Roberts notes (attributing the observation to Lewis 1969) that “a question, if accepted, dictates that the interlocutors choose among the alternatives which it proffers.” In terms of Carnap relevance, this amounts to a choice of an answer, or *point*,  $H$  to which the speaker’s assertion  $X$  must be relevant.

In (3.3), the speaker has used an interjection to mark his assertion as undesirable. That means that, given the question under discussion (or *issue*, since it is, in this case,

---

<sup>4</sup> Roberts provides her own definition of relevance. In her theory, relevance is defined with respect to a question under discussion  $q$ , and depends on a definition of answerhood.

- (3.i) A *partial answer* to a question  $q$  is a proposition which contextually entails the evaluation to true or false of at least one element of  $\llbracket q \rrbracket$ .

Then relevance for assertions is defined as follows:

- (3.ii) An assertion  $a$  is *relevant* to the question under discussion  $q$  iff  $a$  introduces a partial answer to  $q$ .

The definition is not suitable for the present enterprise, for two reasons. First, it is categorical relevance, so it will not allow us to compare the relative relevance of two assertions. Second, as Buring (2003, fn. 7) notes, Roberts’ definition of partial answerhood will not treat answers like “Presumably” as even partial answers, since there is no contextual entailment. Buring adapts Roberts’ definition of answerhood to incorporate probability; contextual entailment is no longer required.

- (3.iii)  $A$  is an answer to  $Q$  if  $A$  shifts the probabilistic weights among the propositions denoted by  $Q$ .

Buring’s definition captures some of what Carnap relevance does: if  $E$  is Buring relevant to a polar question  $\{H, \neg H\}$ , then  $P(H|E) \neq P(H)$ , so  $E$  is Carnap relevant to  $H$  (likewise, if  $E$  is Carnap relevant to  $H$ ,  $E$  is Buring relevant to  $\{H, \neg H\}$ ). However, Buring relevance is still categorical, and it is not clear how to measure how much a proposition shifts the probabilities of a question’s potential answers around.

a polar question)  $\{H = \text{that we will get the grant}, \neg H = \text{that we will not get the grant}\}$  only those assertions that are positively relevant to the undesirable  $\neg H = \text{that we will not receive the grant}$  are felicitous.<sup>5</sup>

This leads us to the following hypothesized generalization:

(3.6) *Positively Relevant Assertion Generalization (PRAG)*

An assertion's propositional content must be positively relevant to the point the speaker is making. That is, given a sentence  $\chi$  with propositional content  $\llbracket \chi \rrbracket$ , and a proposition  $H$  that is the point being made (i.e., a partition in the denotation of a question under discussion), a speaker may not utter  $\chi$  unless  $r_H(\llbracket \chi \rrbracket) > 0$ .

This generalization suggests an important distinction between asserted and implicated meaning: it is asserted meaning, not implicated meaning, that must be relevant to the point the speaker is making. Before returning to test the accuracy of the PRAG's predictions, I digress briefly into the properties of Carnap relevance and the question of whether it is an appropriate tool for the analysis of conversational phenomena.

### 3.2.3 Properties of Carnap relevance

First, a brief examination of the formal properties of Carnap relevance and their general consequences reveals that the definition conforms well to a number of intuitive characteristics of conversational relevance. Relevance is a symmetric relation for all  $E, H$  where

---

<sup>5</sup>I am not sure if the interjection signals that the propositional content of an assertion is undesirable, or that the point being made by the contribution is undesirable. The facts are rather difficult to tease out. But consider the following example. While perusing the results, one of the teachers says,

(3.i)      Crap! That student I hate got a perfect score.

In this case, the teacher's assertion is positively relevant to a desirable  $H$  (*that we will receive the grant*). But it seems the teacher can't be speaking to the question of grant approval, but rather to the question of whether the student he hates will succeed. This provides some evidence that the interjection signals the speaker's attitude towards the conversational point he is making, not his attitude towards the content of his assertion.

$E \not\subset H$  and  $H \not\subset E$ : for non-zero-probability  $E$  and  $H$ , if  $r_H(E) > 0$ , then  $r_E(H) > 0$ .

Proof:

$$r_H(E) > 0 \tag{3.7}$$

$$\text{iff } P(H|E) - P(H) > 0 \tag{3.8}$$

$$\text{iff } P(H|E) > P(H) \tag{3.9}$$

$$\text{iff } \frac{P(H \cap E)}{P(E)} > P(H) \tag{3.10}$$

$$\text{iff } P(H \cap E) > P(H)P(E) \tag{3.11}$$

$$\text{iff } \frac{P(E \cap H)}{P(H)} > P(E) \tag{3.12}$$

$$\text{iff } P(E|H) > P(E) \tag{3.13}$$

$$\text{iff } P(E|H) - P(E) > 0 \tag{3.14}$$

$$\text{iff } r_E(H) > 0 \tag{3.15}$$

This is basically in line with intuitions: if  $E$  is relevant to some point  $H$ , making the point  $H$  itself would be relevant to  $E$ . For example: *I'm hungry* can be used with the point that *you should cut the pie now*, and the point of saying *you should cut the pie now* could be that *I'm hungry*.<sup>6</sup> The fact that relevance is symmetric should not be confused with the wrong idea that the numerical relevance of  $E$  to  $H$  is equal to that of  $H$  to  $E$ ; except in special cases,  $r_H(E) \neq r_E(H)$ .

Carnap relevance conforms to other intuitions about relevance—particularly irrelevance. That is, unless the probability of  $H$  given  $E$  is different from the probability of  $H$ ,  $E$  is irrelevant to  $H$ ; Carnap relevance assigns  $E$  an  $r$ -value of 0 with respect to  $H$ . For example, knowing that it's 16 degrees in Madrid doesn't make it any more or less likely that Clinton will be the next president, so the first proposition's  $r$ -value with respect to the

---

<sup>6</sup>This way of looking at relevance may remind the reader of Grice's examples of "relevance implicatures": it is possible to devise contexts where *you should cut the pie now* implicates *I'm hungry*, and vice versa. See section 3.7 below for a cursory analysis of such implicatures.

second is 0; Carnap relevance mirrors the intuition that the first proposition is irrelevant to the second.

Further, if knowing  $E$  makes  $H$  more likely,  $E$  is positively relevant to  $H$ : if the speaker is arguing that Clinton will be the next president, the proposition *that Clinton will be the Democratic nominee* is relevant. But notice that what is natural is to say that the proposition is relevant, not that it is *positively* relevant. This is also the case for negatively relevant propositions: *that farmers love Obama* is negatively relevant to Clinton's nomination (it has a negative  $r$  value), but we would still say that this proposition is relevant if speaker and hearer are interested in whether Clinton will be the nominee. Crucially, if an assertion is positively relevant to  $H$ , its negation is negatively relevant, as a result of the symmetry property proved beginning in (3.7):

$$r_H(X) > 0 \Rightarrow r_X(H) > 0 \quad (3.16)$$

$$\Rightarrow P(X|H) - P(X) > 0 \quad (3.17)$$

$$\Rightarrow P(X|H) > P(X) \quad (3.18)$$

$$\Rightarrow -P(X|H) < -P(X) \quad (3.19)$$

$$\Rightarrow 1 - P(X|H) < 1 - P(X) \quad (3.20)$$

$$\Rightarrow P(\neg X|H) < P(\neg X) \quad (3.21)$$

$$\Rightarrow P(\neg X|H) - P(\neg X) < 0 \quad (3.22)$$

$$\Rightarrow r_{\neg X}(H) < 0 \quad (3.23)$$

$$\Rightarrow r_H(\neg X) < 0 \quad (3.24)$$

Ostensibly, then, speaker and hearer are sensitive to the polarity (either positive or negative) of an assertion's relevance (as the examples supporting the PRAG show), but the word *relevant* in English simply applies to propositions with non-zero  $r$ -values.

A related question is how to compare the relevance of positively and negatively rele-

vant propositions. As mentioned above, relevance is a comparative notion: one proposition may be more or less relevant than another with respect to a third. For example, it would be more relevant to  $H$  (*that Clinton will be the Democratic nominee*) to know that Clinton carried the state of Iowa than it would be to know that she carried Jefferson County, Iowa. But consider a situation where  $W$  has a large negative  $r$ -value with respect to  $H$ , and  $R$  has a very small positive  $r$ -value; e.g., let  $W$  be *that she is unpopular in Iowa*, and  $R$  be *that she is popular in Rhode Island*. Intuitively  $W$  seems more relevant to  $H$  than  $R$ , although  $R$ 's  $r$ -value is numerically greater (since a relatively small positive number is, of course, greater than a “large” negative number).

But large negative  $r_H$  values translate directly into large positive  $r_{\neg H}$  values. So, in this case, the fact that  $W$  is highly negatively relevant to  $H$  means that  $W$  is highly positively relevant to  $\neg H$  (indeed,  $r_H(E) = -r_{\neg H}(E)$  in general<sup>7</sup>). So  $W$  is more relevant than  $R$ , on the condition that the speaker is making the point that  $\neg H$ . And a speaker who chooses to utter  $W$  and speak to the point  $\neg H$  has made a contribution whose positive relevance is greater than that of an utterance of  $R$ , speaking to the point that  $H$ .

It is more common in the linguistic pragmatics literature, and often more intuitively natural, to talk about relevance to a question than relevance to a proposition. In the example above, the proposition  $W = \text{that Clinton is unpopular in Iowa}$  gets us closer to an answer to the question *whether Clinton will win the Democratic nomination*. Indeed, we can talk about how relevant an answer is to a question, in a sense that makes  $W$  intuitively more *relevant* to the question  $\{H, \neg H\}$ . In Schulz and van Rooij (2006), relevance is defined as the highest Carnap relevance value of the assertion to the propositions in the

---

<sup>7</sup>Proof:  $r_H(E) = P(H|E) - P(H) = (1 - P(\neg H|E)) - P(H) = (1 - P(H)) - P(\neg H|E) = P(\neg H) - P(\neg H|E) = -(P(\neg H|E) - P(\neg H)) = -r_{\neg H}(E)$ .

question:<sup>8</sup>

$$qr_Q(E) = \max\{r_H(E) | H \in Q\}. \quad (3.25)$$

That is, a proposition  $E$ 's relevance to a question  $Q$  is the maximum of the propositional relevance of  $E$  to each of the component propositions of  $Q$ . Then, with respect to the polar question  $\{H, \neg H\}$ ,  $W$  gets a higher relevance value than  $R$  because its propositional relevance to  $\neg H$  is higher than  $R$ 's propositional relevance to  $H$ .<sup>9</sup>

Finally, extra information seems, intuitively, to make a relevant proposition less relevant, contrary to what is provided by Carnap relevance. That is, with respect to the question of whether Clinton will get the nomination, consider the propositions *that she will win the popular vote* ( $V$ ), and *that she will win the popular vote and it is sixteen degrees in Madrid* ( $V \cap M$ ). Both propositions make the probability of Clinton winning the nomination high—winning the nomination is strongly correlated with winning the popular vote in the primaries. So both propositions have equal relevance values:

$$r_H(V) = P(H|V) - P(H) = P(H|V \cap M) - P(H) = r_H(V \cap M). \quad (3.26)$$

Intuitively, there is something wrong with this—the extra information about the weather in Madrid makes “Clinton will win the popular vote, and it’s sixteen degrees in Madrid” a less relevant contribution.<sup>10</sup>

---

<sup>8</sup>An apparent technical problem with the definition: the epistemic framework for Schulz and van Rooij (2006) is not probabilistic; they instead make the move of defining  $P(q|p)$  as  $\text{card}(q)/\text{card}(p)$ , which, even in a finite setting, is not “a simplified measure of the probability of  $q$  given that  $p$  is true”, but is more like the odds ratio of  $q$  to  $p$ . In a probabilistic setting, where  $P(q|p)$  is simply conditional probability, the definition of utility value they provide is perfectly sound; I simply assume here that such a probabilistic setting obtains.

<sup>9</sup>Merin (1999) suggests a definition of question relevance: “a proposition  $E$  is *relevant to a question*  $Q$  iff it is relevant to at least one element” of the partition that  $Q$  denotes (p. 18, fn. 30). This definition provides a reasonable test of whether a given proposition is a relevant answer to a question, but does not provide a way to compare the relevance of two propositions, both of which have non-zero relevance values.

<sup>10</sup>In van Rooij (2003) and Potts (2006), within slightly different frameworks for defining relevance, this problem is solved by modifying the definition of relevance so that, among equally relevant alternatives, those that are less informative are more relevant. But, in the present study, relevance is a number between  $-1$  and  $1$ . The problematic cases are those where the  $r$ -values of two propositions are equal. In a van Rooij/Potts-style system, there is no way to measure how much the extra information decreases the  $r$ -value. Any attempt to do so would likely lose some of the elegant formal properties of simple Carnap relevance.

But it is not quite right to say that extra information makes an utterance less relevant. It seems more natural to me to stipulate that the propositions  $V$  and  $V \cap M$  above are indeed equally relevant to  $H$ , but to allow that  $V \cap M$  contains some irrelevant information. That is, the proposition that it is sixteen degrees in Madrid is irrelevant—it does not, presumably, affect Clinton’s chances of winning the nomination. Nobody could argue with the claim that  $V \cap M$  contains irrelevant information; on the other hand, intuitions about whether  $V \cap M$  as a whole is less relevant than  $V$  are less sharp. So, because the two utterances are associated with propositions of equal  $r$ -value, the fact that one is preferred over the other will require another explanation. This explanation will be that it takes extra words to convey the extra information in  $V \cap M$ . Additional details of this will come in section 4.3, where a hypothesis about which forms are preferred to others, *ceteris paribus*, is developed. As a result, an utterance with denotation  $V \cap M$  would be dispreferred due to the fact that it requires a significantly more complex form, but does not deliver any additional relevant information.

### 3.2.4 Relevance is relative to the addressee’s beliefs

There is one aspect of relevance that has so far been left vague: we’ve said  $E$  is relevant to  $H$  if assuming  $E$  changes the probability of  $H$ . This, of course, means that relevance is defined relative to a probability distribution. But, in light of the decision in Chapter 2 to take a subjective approach to probability, we need to ask whose subjective probability distribution relevance is computed with respect to, for the purposes of conversation. It is obvious that relevance cannot be defined relative to the speaker’s belief state: the speaker’s job is not to assert a proposition that gives the speaker himself new information about the topic of conversation. The alternative is that relevance is defined relative to the hearer’s beliefs: the speaker’s contribution, then, is relevant if it changes the hearer’s beliefs about  $H$ , which is pretty much in line with intuitions.

Note that this creates a problem for the speaker: it is his job to utter something relevant



(in conformance with the PRAG), but, because he does not know what the hearer's actual belief state is, he doesn't know exactly what the effects of his utterance will be. That means a speaker must use his beliefs about the hearer's beliefs to select a contribution that is likely to be relevant. This makes the speaker's job hard, and may well make speakers error-prone (failing to utter something relevant). Difficulty notwithstanding, the claim is that speaker has uttered something relevant only if the proposition he asserts actually increases the probability of  $H$  in the hearer's distribution.

All desirable inferences are generated; no undesirable inferences are.

### 3.2.5 Sperber and Wilson's Relevance

In this chapter, I have argued that Carnap relevance is an appropriate, precise, and intuitive measure of conversational relevance. I would be remiss if I continued without at least briefly addressing another notion of relevance that is prominent in pragmatics research: the one proposed in the Relevance Theory (RT) of Sperber and Wilson (1986). The idea behind RT is that conversational implicature (and communication, generally) is a consequence of a very general cognitive principle: the principle of Relevance. Relevance is characterized both as a property of *assumptions* relative to *contexts*, and as a property of assumptions relative to *individuals*. Putting off the characterization of the technical term *context*, relevance is defined as follows:

(3.27) *Relevance* (to contexts)

- a. "an assumption is relevant in a context to the extent that its contextual effects in this context are large"
- b. "an assumption is relevant in a context to the extent that the effort required to process it in this context is small" (p. 125).

(3.28) *Relevance* (to individuals)

- a. “an assumption is relevant to an individual to the extent that the contextual effects achieved when it is optimally processed [presumably by the individual in question -BR] are large”
- b. “an assumption is relevant to an individual to the extent that the effort required to process it optimally is small” (p. 145).

Many of the terms in Sperber and Wilson’s theory are given special, technical senses, and they are characterized in unusual ways. So it is necessary to delve a bit into these. The first is the characterization of gradable properties like Relevance, by using sets (always pairs, in their case) of *extent conditions*—these can be thought of as the dimensions of relevance. The idea that gradable properties may be graded along more than one dimension is not unique to Sperber and Wilson’s Relevance: in formal semantics, multidimensional adjectives are discussed in Kamp (1975), Bierwisch and Lang (1989), and elsewhere. An example given by Kamp is the intelligence scale: both the ability to manipulate people and the ability to manipulate numbers affect intelligence. It can therefore be hard to decide who is more intelligent: Fred, who manipulates people well, but cannot manipulate numbers; or Mary, who cannot manipulate people, but is very good with numbers. In terms of *extent conditions*, Sperber and Wilson would say something like “An individual is intelligent to the extent that she is good at manipulating people, and an individual is intelligent to the extent that she is good at manipulating numbers.” Relevance, characterized as in (3.27) and (3.28), is a function of two dimensions: contextual effect and effort (or ease).

But the function itself is left unspecified. Presumably, Sperber and Wilson’s extent conditions are intended to be *ceteris paribus*: contextual effects of two assumptions being equal, the more relevant one is the one that is easier to process. (In later work, Wilson and Carston (2007) makes this *ceteris paribus* aspect of extent conditions explicit.) These two conditions are at the crux of the theory, and yet neither their individual evaluation nor their

interaction is spelled out. Levinson (1989) has suggested that Relevance of an assumption  $a$  might be the quotient:

$$(3.29) \quad r(a) = \frac{e(a)}{c(a)},$$

where  $e(a)$  is a measure of the contextual effects of  $a$  and  $c(a)$  is the cost, or effort, of processing  $a$ . This captures the presumably *ceteris paribus* nature of the condition, but, of course, still leaves the measures  $c$  and  $e$  unspecified.

Sperber and Wilson measure contextual effects in terms of cardinality of finite sets of non-trivial inferences (these sets are always finite because Sperber and Wilson set up a special logic that does not allow the derivation of infinitely many non-trivial inferences). This measure is criticized in Gazdar and Good (1982), who claim that this measure divorces SW Relevance from relevance as it is ordinarily understood. Gazdar and Good write: “a moment’s reflection reveals that the cardinality of a set of inferences is a number which has no bearing on relevance” (p. 90); in particular, a sentence with a lexical ambiguity like “George Best walked to the ball” might yield 93 non-trivial contextual implications for the round-object interpretation, and 87 for the formal-dance-party interpretation. They claim that “these numbers obviously have nothing to do with the relative relevance of the two interpretations.” The idea is that relevance just *can’t* depend on how much follows—their argument is an appeal to core intuitions. This is not obvious to me. That is, one can at least *imagine* a formal system that yields more contextual implications for a given interpretation in a context just in case that interpretation is more (commonsense) relevant in that context: more background assumptions about the proposition allowing more deductive steps, or something like that.

Gazdar and Good make a stronger argument: SW Relevance misses the important relationship between (commonsense) relevance and the interests of the participants in a conversation. They observe that an utterance of, “George is a Mason,” is clearly more

relevant when addressed to someone who is interested in and asking about the Masons, but has next to no knowledge about them, than when addressed to someone who is very knowledgeable but not at all interested, despite the fact that the latter addressee will be able to make many more inferences—George has been through such-and-such rites, and has access to such-and-such resources, etc.—yielding a higher SW Relevance rating for the uninterested addressee. Gazdar and Good’s argument then, is that SW Relevance misses the point: relevance has to do with the interests of conversational participants—what they want to know—not how suggestive (in terms of raw number of inferences) an interpretation is.

Merin (1999) provides another version of this argument. Merin notes that SW Relevance might say something about the ratio of informativity to effort, but relevance is not reducible to these two notions. This, Merin argues, is because relevance is fundamentally relational: it makes no sense to talk about the relevance of a proposition unless it is with respect to some other idea, concept, proposition, question, or goal. Neither informativity nor effort has this relational character: informativity has to do with the extent to which a belief state is changed by conditioning on a proposition (in Sperber and Wilson’s terms, how many non-trivial implications follow from this proposition), and effort apparently has to do with various factors—memory demands, complexity—that have nothing to do with the idea to which a proposition may or may not be relevant. Merin writes: “What SW-relevance... misses out on altogether is relevance.”

Although Sperber and Wilson are less than totally transparent on these issues, their characterization of relevance is, in fact, fundamentally relational (contra Merin), and it seems their characterization of context is intended to be responsive to the present interests of the interlocutors (contra Gazdar and Good). This depends on an idiosyncratic notion of context (SW context), whereby only carefully selected propositions are present in a given SW context. Because they are not precise about what makes up SW contexts in general, the discussion here is framed in terms of a SW context they devise.

- (3.30)
- a. People who are getting married should consult a doctor about possible hereditary risks to their children.
  - b. Two people both of whom have thalassemia should be warned against having children.
  - c. Susan has thalassemia.

These three propositions (and nothing else) are assumed to make up the entire SW context. Relative to this SW context, SW Relevance ranks (3.31a) higher than (3.31b) and (3.31c):

- (3.31)
- a. Bill, who has thalassemia, is getting married to Susan.
  - b. Susan, who has thalassemia, is getting married to Bill.
  - c. Bill, who has thalassemia, is getting married to Susan, and 1967 was a great year for French wines.

The reason (3.31a) is more SW Relevant than (3.31b) is that (3.31a), combined with the SW context, yields both implications *that Bill and Susan should be warned against having children* and *that Bill and Susan should consult a doctor about possible hereditary risks to their children*, whereas (3.31c) only yields the implication *that they should consult a doctor*. (3.31a) is more SW Relevant than (3.31c) because the latter requires greater processing effort without any corresponding advantage in contextual effects—with respect to the assumptions in (3.30), (3.31a) and (3.31c) have exactly the same contextual effects.

This ranking would not hold generally in SW contexts other than (3.30). If propositions existed in the SW context for which *that 1967 was a great year for French wines* yielded contextual effects, the extra processing effort required for (3.31c) might be worth it, and so that sentence might be more SW Relevant than (3.31a). What Sperber and Wilson seem to do with this example is restrain the SW context to propositions which are Carnap relevant to a proposition like *that Bill and Susan should take precautions before deciding to have children*. Given this fact, SW relevance is indeed a relative notion: it is

relative to the set of propositions in a SW context. Moreover, this fact makes SW contextual effects sort of a proxy for Carnap relevance by building Carnap relevance into the definition of SW contexts.

These considerations provide a potential answer to Gazdar and Good's and to Merin's arguments that SW relevance does not correspond to intuitive relevance. Gazdar and Good assert, with respect to their example, that contextual effects are much greater for an uninterested Masons expert than a rapt Masons novice, contrary to the intuitive relevance of the sentence "George is a Mason" when uttered to each individual. But Sperber and Wilson tightly constrain contexts with respect to the conversations at hand; one factor that affects context will be the interests of the conversational participants. Therefore, propositions that the Masons expert *knows*, like *that Masons must endure prolonged exposure to freezing temperatures*, are not part of the context, and do not yield contextual effects. So despite the fact that the Masons expert is able to draw more inferences from the utterance, these inferences are not *contextual effects*, since the expert is not interested in Masons, and therefore the context does not contain any propositions having to do with the Masons.

Nonetheless, SW Relevance will be of little use for the present endeavor. Most important, SW Relevance is stubbornly resistant to formalization, and, as emerged from the discussion of SW contexts above, depends on a seemingly arbitrary stipulation of the elements that make up a context. Moreover, Sperber and Wilson's ideas about how contexts are determined seem to be circular. Specifically, they rule out all conceptions whereby a context for an utterance is determined prior to, or independent of, the utterance, instead defining contexts relative to utterances. That is, a context is a function from utterances to sets of propositions, so that if a speaker asserts a proposition, the context is the set of propositions which, combined with their utterance, yield contextual effects. But this means that the determination of relevance of an utterance depends on how many contextual effects may be derived; this number, in turn, depends on which propositions make up the context; the determination of the elements in the context depends on the speaker's

choice of utterance; and the speaker's choice of utterance depends on the relevance of the utterance, closing the loop. It seems that such a process could not be formalized in a way that made it terminate.

SW Relevance, then, conforms at best marginally well to intuitive notions of relevance and depends on a specialized notion of context that is itself dependent on relevance. Sperber and Wilson are clear that their intent is not to try to match their relevance strictly to intuitive understandings of relevance. Instead, they provide a new characterization of relevance that, they say, provides a theory of all communication (and much of the rest of cognition). In contrast, Carnap relevance hews quite closely to the common sense notion of relevance and has a simple and precise mathematical formalization. Moreover, rather than treating the measure as the whole theory of communication, I intend to consider it merely as a guiding force, much in line with Grice's original theory. Indeed, Grice himself was skeptical about the directionless nature of SW Relevance: "Wilson and Sperber... seem to be disposed to sever the notion of relevance from the specification of some particular direction of relevance."

### **3.3 Applications for Carnap relevance**

In this Chapter, I have identified a generalization about relevance (the PRAG), proposed Carnap relevance as a way to measure relevance, and examined the intuitive properties of Carnap relevance that make it an appropriate tool to measure conversational relevance. On the assumption that speakers are expected to be relevant, the introduction of this tool will be essential for the development of a formal theory of Utterance Expectation in Chapter 4 (recall that, in Chapter 2, I argued that a theory of implicature, given Bayesian reasoning, can be reduced to a theory of Utterance Expectation.) Before moving on to the formal theory, I think it is worthwhile to examine the role of relevance in key pragmatic applications, with the overarching goal of further motivating Carnap relevance as a guiding force

in conversation.

### 3.3.1 An analysis of *but*

Merin (1999) builds a far-reaching pragmatic theory with Carnap relevance at its center,<sup>11</sup> inspired by the argument-based pragmatics of Ducrot (1973).<sup>12</sup> Merin employs Carnap relevance to analyze a number of empirical phenomena, including conditions on the use of *but*, *even*, and *also*; scalar adjectives like *warm*; particularized scalar implicature; and generalized scalar implicature.

The most straightforward application of relevance that Merin proposes is a simple condition governing the use of *but*. The felicitous use of *but* is generally thought to require some kind of contrast between conjuncts. Whereas the usual story is that the content of the conjuncts contrasts (e.g., that one makes the other less likely), Merin proposes that the relevance of the conjuncts must contrast: *A but B* is felicitous iff

$$r_H(A) < 0, r_H(B) > 0, \text{ and } r_H(A \wedge B) > 0.^{13} \quad (3.32)$$

As Merin notes, in a context where the issue is *whether Wilkins can stop the rot at Batbridge*, an utterance of “Wilkins is a man of principle, but he is now Chancellor of Crowford” does not imply that Crowford tends to have crooks as Chancellor—rather, his current employment makes him an unlikely candidate for the salvation of Batbridge (here, *H* is

<sup>11</sup>Merin actually uses a modification of Carnap relevance called the “log-likelihood-ratio for simple *H*”:  $\log \frac{P(E|H)}{P(E|\neg H)}$ . As Merin states, these functions are “continuous monotone increasing” functions of each other; for all present intents and purposes, that makes them equivalent. Because Carnap’s definition is easier to intuitively grasp, that is what I have used in the present project.

<sup>12</sup>Ducrot’s work has not, to my knowledge, been translated into English.

<sup>13</sup>Merin actually reverses the  $<$ ’s and  $>$ ’s in this definition—but this is equivalent, modulo negation of the point being made. I believe it is more in line with intuitions to think that the first conjunct is a concession (i.e., an argument against the point being made by the speaker), with the second point being the speaker’s argument for their point. Indeed, the concessive (König, 1990) *although* shows the same asymmetry: the speaker who argues that Wilkins is not a viable choice to stop the rot at Batbridge may argue, “Although Wilkins is a man of principle, he is now Chancellor of Crowford.” Moreover, this way of stating the condition on *but* is also in line with the PRAG, if a sentence *A but B* is assumed to assert the proposition  $A \wedge B$ , where Merin’s definition would make such sentences an exception to the PRAG.



*that Wilkins can not stop the rot at Batbridge*).<sup>14</sup> That is, in general, there is no requirement that there be a direct contrast between the conjuncts of *but*.

A similar observation appears as early as Dummett (1973), who provides an example in which a club committee (presumably in the United Kingdom) is discussing which speakers to invite. Someone utters, “Robinson always draws large audiences, but he is in America for a year.” Dummett writes that the utterer “is not suggesting that a popular speaker is unlikely to go to America, but that, while Robinson’s popularity as a speaker is a reason for inviting him, his being in America is a strong reason against doing so.” Notably, Dummett does not extend this observation to an analysis of *but*; instead, he concludes simply that *but* is used to “hint that there is some contrast, relevant to the context, between the two halves of the sentence: no more can be said, in general, about what kind of contrast is hinted at.” Other authors have similarly asserted that the understood contrast associated with *but* is “ineffable” (Potts, 2005; Horn, 2008).

But Merin’s analysis provides a straightforward prediction of these facts. The speaker’s point *H* here is plausibly *that we should not invite Robinson*, and the two conjuncts have opposing relevance with respect to *H*, satisfying Merin’s condition on *but*. Far from leaving the contrast required by *but* as ineffable, Merin’s analysis provides a straightforward characterization of the conditions governing the use of *but*. The analysis also sheds some light on the question of why it has been incorrectly assumed that a direct contrast between conjuncts is necessary, since often when *A* and *B* contrast relative to *H* ( $r_H(A) < 0$  and  $r_H(B) > 0$ ), they also contrast directly with each other ( $r_A(B) < 0$ ). In other words, the fact that *A* and *B* must “contrast” with respect to their relevance to *H* leads to the impression that *A* and *B* contrast simpliciter.

Merin limits discussion to simple root clause conjoining instances of *but*, and it is not

---

<sup>14</sup>Discussions about stopping the rot at Batbridge may be outside the idiom of some readers. For those readers: in a conversation between Sacramento Zoo administrators about who to hire to handle the big cats, an utterance of “Wilkins is great with lions, but he is under contract to handle the big cats at the San Diego Zoo” does not imply that there is a contrast between being great with lions and being under contract to handle the big cats at the San Diego Zoo.

immediately clear how to extend the analysis to cases of conjunction of non-sentential phrases, or in embedded contexts:

- (3.33)    a.    Every graduate student but few undergrads understood the point of the talk.  
              b.    If you finish your chicken but don't eat your vegetables, you might not get dessert.  
              c.    Marv said that Shaq is huge but that he is agile. (Bach, 1999, p. 348)

I will not extend Merin's basic relevance-based analysis to all cases of embedded and non-clause-conjoining *but*, but I will show that it has relatively broad application.

First, Bach (1999) provides putative evidence that, in an indirect quotation of a sentence containing *but*, the quotation must also use *but*:

- (3.34)    Marv: Shaq is huge but he is agile.  
              a.    Marv said that Shaq is huge but that he is agile.  
              b.    #Marv said that Shaq is huge and that he is agile.

Bach argues that this datum shows that *but* is part of "what is said", in Grice's terminology, as part of a larger program to label conventional implicature a myth.<sup>15</sup> But a little context, combined with Merin's analysis of *but*, allows us to examine this case in somewhat more detail, both showing that Bach's claim is false and illuminating the reason he made the mistake he did.

Let's return to Dummett's example: the club committee is deciding whom to invite to speak for the upcoming year's series of talks. *A* is absent from the meeting, but he sends *B* as his designee. When telling his preferences to *B* prior to the meeting, *A*, however, was under the impression that the talks would all take place in America. *A* tells

---

<sup>15</sup>Bach assumes that the content of an indirect quotation that must be attributed to the quoted speaker must be part of "what is said" by that speaker.

*B*: “It’s a pity there’s no money budgeted for travel. Otherwise, I’d like them to consider Robinson. Robinson always draws large audiences, but he is going to be in London all year.” At the meeting, *B* learns that the series of talks will be in London, after all. The committee members ask *B* for *A*’s opinion. *B* says: “*A* said that Robinson always draws large audiences, and that he’s going to be in London all year.” I think it is clear, in this case, that *B* has done the right thing by replacing the *but* with *and*: in the context of the meeting, both conjuncts are positively relevant, and it would therefore be infelicitous to use *but* here.

Why did Bach think that the data in (3.34) reflected a requirement to use *but* in an indirect quotation of an assertion with *but*? I think it is because it is easy to imagine a shared *H* for the two utterance contexts, to think that both Marv’s original utterance and the speaker quoting Marv are providing evidence for the same thing. What the data presented above show is that, by manipulating the contexts of the original assertion and its indirect quotation, what counts as relevant also shifts. The data also undermine Bach’s contention that the conventional implicature associated with *but* is part of what is said.

Let us now turn to instances where *but* conjoins non-proposition-denoting elements. Because Merin’s analysis of *but* depends on contrasting relevance, and relevance is a relation between propositions, the theory does not automatically extend to such cases. I will not develop a full-fledged extension here, but I will present data that suggest that relevance is operative in the coordination of generalized quantifiers with *but*, undermining previous claims that quantifiers of opposing monotonicity must be conjoined with *but*, not *and*.

Barwise and Cooper (1981) report that conjunction of generalized quantifiers with opposing monotonicity requires the use of *but*. The datum is:

(3.35) No student \*and/but every professor attended the talk.

The intuition is similar, if somewhat attenuated, when it is sentences, rather than general-

ized quantifiers, that are conjoined:

(3.36) No student attended the talk, \*?and/but every professor did.

But these sentences are presented out of context. Merin's *but* analysis predicts that it is not the monotonicity of the quantifiers but their relevance to the point being made that determines the appropriateness of *but* coordination.

To test this prediction, consider a context where both propositions in (3.36) are positively relevant to the point being made: suppose the department's professors wanted to discuss some highly confidential issue, like whether a certain departmental merger should go through. A professor and her spouse are having a conversation at dinner:

(3.37) You remember these endless murmurs about a possible merger with department X? Well, there was a talk in our department today.

- a. ...No student attended, and/#?but every professor did...
- b. ...No student and/#?but every professor attended...

So we were able to actually vote on the merger afterward.

Clearly the pattern does not reverse: *but* is still pretty good in this context. But because both *no student attended the talk* and *every professor attended the talk* are positively relevant to the proposition *that we were able to vote on the merger*, *and* is more natural. Moreover, the pattern is maintained in (3.37b), where *but* conjoins generalized quantifiers.

A detailed compositional semantics for *but* that allows the relevance condition to apply with respect to expanded propositions, rather than the denotata of the phrases *but* conjoins, will not be pursued here. And the story about the data in (3.37b) may be more complicated than Merin's simple relevance condition (*but* is, after all, pretty good here, even though

though both conjuncts have the same relevance polarity). But the example should illustrate that Merin's analysis of *but*, which requires the introduction of relevance into the theory of conversation, makes illuminating new predictions about the effect of context on data in the literature.<sup>16</sup>

### 3.3.2 Merin's program applied to scalar implicature

Despite Merin's successful analysis of *but*, Merin's program has a number of problems, both in its application to empirical phenomena and in its conceptual foundation. Most critically, Merin's stated intent is to "treat natural language semantics as a branch of pragmatics," apparently at odds with the vast majority of current work in linguistic semantics (this dissertation included), which assumes that natural language semantics is something delivered by grammar, and that pragmatics is fundamentally separate from grammar (which is not to say that the two do not interact).<sup>17</sup> Merin's program, following Ducrot (1973), requires assumptions that the "paradigmatic discourse situation is not one of simple 'information transfer', but of explicit or tacit *debate*" and that speakers "do not engage in idle gossip, do not proselytize for the sake of it, but speak to a *point*." As an empirical claim, these assumptions seem false: interlocutors are rarely adversaries (more often, they cooperate), and they obviously do engage in idle gossip, etc.

In Merin's adversarial model, conversation is a bargaining situation, where the speaker's offer (assertion) is limited by what she thinks the hearer is willing to accept; analogously, if I offer you ten dollars for a painting, I don't think you'd accept five. Likewise, the hearer's concession (belief) is limited by what she thinks the speaker is willing

---

<sup>16</sup>I exclude here what have been called "contrastive" uses of *but*, of the type "Not three but four students left." This seems to have a metalinguistic character (evidenced by the contrast between "Not some but all" and "#All but not some"), and apparently the conditions on its use are not related to relevance.

<sup>17</sup>This mainstream view applies even to the work advocating a shift of scalar-implicature-type meanings to the grammar Chierchia (2004); Sharvit and Gajewski (2008); Gajewski and Sharvit (to appear); Fox (2007); Chierchia et al. (to appear). These authors do not propose elimination of the distinction between semantics and pragmatics; they simply argue for the shift of an aspect of speaker meaning (scalar implicature) from pragmatics to semantics.

to offer: if I concede to sell you the painting for ten dollars, I don't think you'd pay fifteen.

But the bargaining analogy is not really intuitively apt for the conversational setting. If you accept my offer of ten dollars, I then know that you'd accept any amount greater than ten dollars as well; when I make the offer, you know I'd be willing to pay any lesser amount. But in conversation, when you accept my claim, say, that Bush has mismanaged the war in Iraq (towards  $H = \text{that Bush is a lousy president}$ ), I don't think you're willing to accept everything less relevant than a claim of mismanagement; nor am I willing to accept everything more relevant. My claim, in any intuitive understanding of the term, does not include such propositions as *that Bush is an undercover Russian agent*, even though this would make him a worse president than the accusation of mismanagement. I am not, in ordinary circumstances, willing to accept this proposition, even when I am arguing that Bush is a bad president.

The key application of Merin's adversarial model is the calculation of particularized scalar implicatures as a hearer's refusal to accept any proposition that is more relevant than the speaker's assertion. Applied to a case from Hirschberg (1991), this produces reasonable results. At a job interview, the candidate ( $a$ ) is asked by the interviewer ( $q$ ),

- (3.38)     $q$ : Do you speak Portuguese?  
               $a$ : My husband does. ( $= F$ )  
               $\sim$ I don't speak Portuguese. ( $= \neg E$ )

Assume the point  $H$  is *that  $q$  be hired*. Since this would be made more likely by the interviewee speaking Portuguese than it would by her husband's possession of this skill ( $r_H(E) > r_H(F) > 0$ ), the speaker has made an assertion whose relevance is relatively low. The rational hearer, in Merin's system, accepts no proposition more relevant than the speaker's demand, so  $E$  is rejected by the hearer. Therefore the speaker means, in uttering  $F$ , that  $E$  is false, generating the implicature.

In general, whenever  $r_H(X) > r_H(F)$ , the rational hearer will exclude  $X$  after hearing an utterance of  $F$ , so the system is capable of generating a great mass of particularized scalar implicatures. Merin limits the number of implicatures by restricting such  $X$  to the elements of a set  $\mathcal{S}$  of propositions that are “selected by lexical and local relevance structure.” But Merin does not spell this out in any greater detail. Note, briefly, that if  $\mathcal{S}$  is not constrained, the approach certainly generates too many implicatures, leading to a variation of the symmetry problem,<sup>18</sup> wherein, given relatively innocuous assumptions, utterances generally lead to a contradictory implicature. Suppose, for example, that  $\mathcal{S} = \mathcal{P}(\Omega)$ , that is,  $\mathcal{S}$  contains every proposition. Suppose further that the individual worlds in  $\Omega$  each receive an equal probability (a fairly standard assumption). Now suppose that a speaker utters  $X$ , positively relevant to  $H$ , where  $X$  contains at least one world, call it  $w$ , not in  $H$ . Let  $v$  be an arbitrary world; I show that  $r_H(H \cup \{v\}) \geq r_H(X)$ :

$$r_H(X) = P(H|X) - P(H) \tag{3.39}$$

$$\leq P(H|H \cup \{w\}) - P(H) \tag{3.40}$$

$$\leq P(H|H \cup \{v\}) - P(H) \quad (\text{strictly less when } v \in H) \tag{3.41}$$

$$= r_H(H \cup \{v\}) \tag{3.42}$$

Since  $r_H(H \cup \{v\}) \geq r_H(X)$ , we can conclude that whenever  $X \not\subseteq H$ , every world  $v$  is contained in some proposition that is more relevant than the speaker’s utterance  $X$ . The consequence of this is that every utterance is predicted to implicate that nothing is true.

Because Merin does not spell out the details of how  $\mathcal{S}$  is determined, we are left with a problem at least as great as the problem of symmetry and Horn scales (again, see section 4.4 for discussion): the system is consistent and generates observed implicatures only if the set  $\mathcal{S}$  is properly constrained. Classic Horn-Gazdar generation is built on the stipulation of Horn scales; Merin generation depends on the stipulation of  $\mathcal{S}$  sets.

---

<sup>18</sup>Term due to von Fintel and Heim (1998); see section 4.4 for discussion and a partial solution.

### 3.4 Pragmatics and the PRAG

Stepping back, this chapter opened with a hypothesized generalization (the PRAG) that accounted for the non-interchangeability of *some* and *some but not all* following interjections. Because the PRAG is stated in terms of Carnap's formalization of relevance, we delved into an application of Carnap relevance for the analysis of *but*, thus providing independent motivation for the presence of Carnap relevance in conversation. Now we return to the PRAG, exploring the consequences for pragmatic theory of the hypothesis that speakers' assertions must be positively Carnap relevant.

#### 3.4.1 Numeral expressions

With respect to the data that motivated the PRAG, (3.3e) was argued to be bad because it makes an assertion that is not positively relevant to the point being made. The flip side of this is that, if the PRAG is correct, it can be used to diagnose implicature: because asserted content, but not necessarily implicated content, must be positively relevant to the speaker's main point, implicated content can be identified as that content that may be negatively relevant to the speaker's main point. In particular, this diagnostic may be brought to bear on numeral expressions, the semantics of which have been debated over the last few decades: do they mean *at least n* (Horn, 1972; Levinson, 2000; Chierchia, 2004; Fox and Hackl, 2006), or *exactly n* (Horn, 1992; Bultinck, 2004; Geurts, 2006)? Horn (1992) uses question-answering to illustrate a clear distinction between canonical scalar terms like *many* or *warm* and numerals like *two* and *three*:

(3.43)    *Q* : Do you have two children?

*A*<sub>1</sub> : No, three.

*A*<sub>2</sub> : ?# Yes, (in fact) three.

(3.44)    *Q* : Did many of the guests leave?



$A_1$  : #No, all of them.

$A_2$  : Yes, (in fact) all of them.

Here, a polar question about a numerical quantity is answered negatively when the actual quantity is greater, but an analogous question using a (relatively) weak quantifier is answered positively when a stronger quantifier applies. This is consistent with the view that *all* is semantically stronger than *many*, but *three* is not in a strength relation with *two*.

The PRAG makes a corroborating prediction for cardinals. Suppose Samson arrives late to a date, and Delilah, a total neurotic, exclaims, “Thank god! I thought you’d been killed in a traffic accident!” Samson can reply, with irritation:

(3.45) (Of course I wasn’t killed!) I was (#at least) fifteen minutes late!/#I was pretty late!

The probability that Samson is killed increases as his lateness increases: so the probability that he’s been killed given that four hours have passed and he still has not arrived is much greater (though still probably low) than the probability that he’s been killed given that fifteen minutes have passed.<sup>19</sup> But if *fifteen* means the same as *at least fifteen*, it cannot be positively relevant that Samson was fifteen minutes late—this proposition could only increase the probability that Samson was killed, since it simply excludes worlds where Samson’s lateness is minimal, and therefore unlikely to be an indication of his death.

### 3.4.2 Interaction with context

The probabilistic approach to relevance taken here makes it highly context-sensitive: changes in common knowledge affect the calculation of relevance. Consider the follow-

---

<sup>19</sup>For Samson to consider his contribution relevant here, he has to think Delilah imagines he is later than he really is. Indeed, Samson’s response doesn’t work if Delilah says, “You’re ten minutes late! I thought you’d been killed!”

ing scenario: A coach, on his way to the field, receives a cell phone call from his assistant coach.

(3.46) Mutually known by coach and assistant:

- a. Exactly two team members can score goals: Lucy and Kate.
- b. Exactly two team members can play goalie: Christine and Kate.
- c. Lucy is on vacation in Italy with her family.

The assistant informs the coach that Christine has called to say she has the flu. Goalie is an absolutely essential position, so it is mutually known that Kate will be playing goalie for the duration. The coach whines, “How are we going to score any goals?!?” His assistant, who knows which players have arrived, can’t appease him by saying, “Don’t worry, we’ve got one capable goal-scorer”, even though it is true, since it is not relevant in this context: the coach cares about the proposition *that we will score goals* or *that we will win the game*, and in this context, having one goal-scorer doesn’t help.

Moreover, the utterance can be used in this situation by the coach to undermine his assistant’s reassurances:

(3.47) Assistant: Don’t worry; I’m sure we’ll do fine.

Coach: We have one goal-scorer (and she has to play goalie!)

Here, contextual factors affect the probability distributions against which relevance is calculated, leading an utterance that would be irrelevant or negatively relevant in one context (and therefore infelicitous), positively relevant in another. Moreover, this reflects the kind of context sensitivity that a Gricean theory of implicature needs.

### 3.4.3 Other expressives

The inventory of English expressives is not, of course, limited to interjections like *crap* and *whew*. If the PRAG is correct, a whole host of expressives are predicted to interact with the relevance of corresponding assertions.

(3.48) Searle #(only) washed some of the fucking dirty dishes!

The expressive signals relevance to a point that the speaker has negative feelings about (e.g., that the house is still a mess, that Searle is a jerk); yet the content of the speaker's assertion is negatively relevant with respect to those points, in violation of the PRAG. Speaker-oriented adverbs (Jackendoff, 1972; Potts, 2005) provide another bit of evidence. Consider the following, from a review of a published conference proceedings:

(3.49) Unfortunately, some papers which were presented at the conference are not reported.

This sentence cannot be used when what is unfortunate is that not all of the papers are unreported (i.e., that some papers are reported). It must be the omissions that are being lamented, not the inclusions. *Unfortunately*, like *crap*, indicates that the content of the following assertion is relevant to a proposition viewed as undesirable by the speaker.

### 3.4.4 *so*

So far, conversational points have been either implicit or explicitly specified in the form of a preceding question or a context where some particular *H* is assumed to be the operative point. But speakers, it seems, sometimes tell their interlocutors what point they're speaking to, by using connectives like *so*.<sup>20</sup> Here, we see a sharp distinction between positive

<sup>20</sup>There are two possible analyses available here, and I will not try to figure out which one is right. One is to say that *A, so B* is felicitous only if *B* is *H* (that is, *B* is the point being made by the speaker). The other

and negative relevance:

- (3.50)    a.    Some of the students brought whiskey, so the chaperones were furious.  
              b.    ?Not all, but some, of the students brought whiskey, so the chaperones were  
                               furious.  
              c.    #Not all of the students brought whiskey, so the chaperones were furious.

Since  $H$  is the proposition *that the chaperones were furious*, only the positively relevant *some of the students brought whiskey* is felicitous as a first conjunct. It is not surprising (indeed it is expected) to find that relevance provides a useful tool for the analysis of *so*, which conveys (but does not assert) a conventional meaning closely related to that of *but*. And, indeed, a similar item, *therefore*, is at the heart of Grice's original argument for conventional implicature, with the example, "He is an Englishman; he is, therefore, brave."<sup>21</sup>

### 3.5 Relevance and informativeness

This dissertation is about scalar implicature, not about relevance. So far, I have presented evidence that sentences with scalar implicatures have two parts: the bit that is classically understood to be asserted (*at least some...*), and the bit that is implied (*not all*), and that the former, and not the latter, are sensitive to conditions on relevance. But the centrality of relevance in the organization of conversation has greater significance for theories of scalar implicature. Notice that, in the context surrounding (3.3), the strong alternative with *all* is more relevant than the weak alternative with *some*. We are even more likely to get the grant if all of the students pass the exam than we are if some (and maybe all)

---

is to simply say  $A, so B$  is felicitous only if  $r_B(A) > 0$ . In the latter case, the whole conjoined sentence may be relevant to a third point. For simplicity, I assume the former case, where  $B$  is the actual point being made by the speaker, is the operative one.

<sup>21</sup>See also discussion in König 1990 on the duality between the adverbials *therefore* and *although*, which extends in a straightforward way to the coordinators *so* and *but*.

of them pass: asserting *all* in this case would be making an even more relevant contribution. Indeed, Merin’s particularized scalar implicature generation mechanism supplants informativeness with relevance entirely: scalar reasoning is reasoning about what more relevant propositions could be asserted.

This shift in the theory of scalar implicature, however, risks divorcing scalar implicature from the scalarity of words like *some*: it is, after all, possible for strictly stronger expressions to be less relevant to some point  $H$  than their weaker counterparts. An obvious case where this is so is one where  $H = \omega$  and not  $\sigma$ : here  $\omega$  is positively relevant, but  $\sigma$  is, of course, highly negatively relevant. To put it more generally:

$$(3.51) \quad \textit{Fact: For } \sigma \subset \omega, \text{ there exist } H \text{ such that } r_H(\sigma) \leq r_H(\omega).^{22}$$

This means that, for for any  $\sigma$  that is more informative than  $\omega$ , there is a context that makes  $\omega$  more relevant than  $\sigma$ .

This poses a problem for theories of scalar implicature that shift the foundation from informativeness to relevance: why is garden-variety scalar implicature so ubiquitous if it depends on contextually-determined relevance relations? For easy reference in the sequel, I give the problem a label below:

$$(3.56) \quad \textbf{Problem:} \text{ Relevance relations between weak and strong scalar alternatives are flexible; yet scalar implicature seems to be “generalized” across contexts.}$$

First, note that informativeness and relevance are not completely independent: they do

---

<sup>22</sup>Proof: Let  $H = \omega \wedge \neg\sigma$ . Then

$$r_H(\sigma) = P(\omega \wedge \neg\sigma | \sigma) - P(\omega \wedge \neg\sigma) \tag{3.52}$$

$$= 0 - P(\omega \wedge \neg\sigma) \tag{3.53}$$

$$\leq P(\omega \wedge \neg\sigma | \omega) - P(\omega \wedge \neg\sigma) \tag{3.54}$$

$$= r_H(\omega). \tag{3.55}$$

interact. For one thing, entailment-related alternatives are always positively relevant to each other. That is, when  $\sigma \subset \omega$ ,  $P(\sigma|\omega)$  cannot be less than  $P(\sigma)$ ; so,  $r_\sigma(\omega) \geq 0$  (strictly greater when  $P(\omega) \neq 1$ ).<sup>23</sup> So weaker propositions (in terms of entailment), are never negatively relevant to their stronger counterparts. (Obviously, stronger propositions are also positively relevant to their weaker counterparts—they entail them.)

(3.57) *Fact:* If  $\sigma \subset \omega$ , then  $r_\sigma(\omega) \geq 0$  and  $r_\omega(\sigma) \geq 0$ .

Moreover, whenever  $\omega$  is more relevant to the point being made  $H$  than  $\sigma$ ,  $\omega \wedge \neg\sigma$  is, in turn, more relevant than  $\omega$ . Here is the proof:

(3.58) **Claim:** Let  $\omega$ ,  $\sigma$ , and  $H$  be propositions, with  $\omega \subset \sigma$ . Then  $r_H(\omega) > r_H(\sigma)$  if and only if  $r_H(\omega \wedge \neg\sigma) > r_H(\omega)$ .

---

<sup>23</sup>Proof:  $P(\sigma|\omega) = P(\sigma \cap \omega)/P(\omega) = P(\sigma)/P(\omega) \geq P(\sigma)$ , since  $P(\omega) \leq 1$ . If  $P(\omega) \neq 1$ , the inequality is  $>$ , or strictly greater than, so  $r_\sigma(\omega) > 0$ .

*Proof:*

$$r_H(\omega \wedge \neg\sigma) > r_H(\omega) \quad (3.59)$$

$$\iff P(H|\omega \wedge \neg\sigma) - P(H) > P(H|\omega) - P(H) \quad (3.60)$$

$$\iff P(H|\omega \wedge \neg\sigma) > P(H|\omega) \quad (3.61)$$

$$\iff \frac{P(\omega \wedge \neg\sigma|H)P(H)}{P(\omega \wedge \neg\sigma)} > \frac{P(\omega|H)P(H)}{P(\omega)} \quad (3.62)$$

$$\iff \frac{P(\omega \wedge \neg\sigma|H)}{P(\omega \wedge \neg\sigma)} > \frac{P(\omega|H)}{P(\omega)} \quad (3.63)$$

$$\iff \frac{P(\omega|H) - P(\sigma|H)}{P(\omega) - P(\sigma)} > \frac{P(\omega|H)}{P(\omega)} \quad (3.64)$$

$$\iff P(\omega)(P(\omega|H) - P(\sigma|H)) > P(\omega|H)(P(\omega) - P(\sigma)) \quad (3.65)$$

$$\iff P(\omega)P(\omega|H) - P(\omega)P(\sigma|H) > P(\omega)P(\omega|H) - P(\sigma)P(\omega|H) \quad (3.66)$$

$$\iff -P(\omega)P(\sigma|H) > -P(\sigma)P(\omega|H) \quad (3.67)$$

$$\iff P(\sigma)P(\omega|H) > P(\omega)P(\sigma|H) \quad (3.68)$$

$$\iff P(\sigma)P(\omega|H)P(H) > P(\omega)P(\sigma|H)P(H) \quad (3.69)$$

$$\iff \frac{P(\omega|H)P(H)}{P(\omega)} > \frac{P(\sigma|H)P(H)}{P(\sigma)} \quad (3.70)$$

$$\iff P(H|\omega) > P(H|\sigma) \quad (3.71)$$

$$\iff P(H|\omega) - P(H) > P(H|\sigma) - P(H) \quad (3.72)$$

$$\iff r_H(\omega) > r_H(\sigma) \quad (3.73)$$

This fact means that relevance will apply to classic scalar implicature data in a principled way: whenever the weaker alternative is more relevant than the stronger alternative, the scalar strengthened alternative is more relevant still. What this means is that the relevance of  $\omega$ , the weak alternative, is always between the relevance of  $\sigma$  and  $\omega \wedge \neg\sigma$ . To put it yet another way, a weak sentence always falls between the strong alternative and the upper-bounded alternative on the relevance scale.

The fact that a weak sentence  $\omega$  falls between  $\sigma$  and  $\omega \wedge \neg\sigma$  on the relevance scale

means that, whenever  $\omega$  is more relevant than  $\sigma$ ,  $\omega$  has a more relevant alternative:  $\omega \wedge \neg\sigma$ . On the assumption that scalar implicature computation is relevance-based (at the risk of getting a bit ahead of ourselves; see Chapter 4 for details), this result will tend to lead to a scalar-type implicature of  $\neg B(\omega \wedge \neg\sigma)$ . This (combined with an assumption of  $B\omega$ ) is equivalent to an implicature of  $\neg B\neg\sigma$ , or, for all the speaker knows,  $\sigma$  is possible.

What does this mean? The essence of the result is that, when an utterance does not generate a scalar implicature (i.e., when an inference to  $\neg B\sigma$  is not licit), an anti-scalar implicature is generated: (i.e., an inference of  $\neg B\neg\sigma$ ). In probabilistic terms, this says that when the theory assigns a low probability to  $\neg B\sigma$ , it also assigns a high probability to  $\neg B\neg\sigma$ . As intuitive as this seems, it doesn't have to be the case: you could have a tendency to exclude the "middle space" of  $\neg B\sigma \wedge \neg B\neg\sigma$ , meaning that  $B\neg\sigma$  increases as  $B\sigma$  increases. The result in (3.73) tells us that the opposite tends to be the case: if relevance relations provide that  $B\sigma$  is high, they also provide that  $B\neg\sigma$  is low: relevance-based implicature computation pushes speaker belief as a "contiguous mass," not into two discrete directions.

At this point, it seems most prudent to return to the solid ground of empirical evidence to explore the observed relationship between relevance relations and scalar implicature. The result above states that, whenever a weak sentence is more relevant than a semantically stronger sentence, it is less relevant still than a "weak and not strong" alternative. To start, consider the following triplet:

- (3.74)    a.    John has a dog. (weak)  
              b.    John has a chihuahua. (strong)  
              c.    John has a non-chihuahua dog. (weak and not strong)

For context:  $A$  and  $B$  are trying to decide whose house to burglarize, and they have been considering the merits of various enemies' houses.  $A$  lists various reasons that John's



house might be a good one to rob: he rarely locks his doors and windows; his neighbors are never home; he keeps his valuables out in the open. Suddenly, something occurs to *B*: “Wait!” he says. Now, if *B* utters (3.74a), he has said something positively relevant to *H* = *that we should not burglarize John’s house*; a dog could cause trouble and even serious injuries for burglars. The stronger sentence (3.74b) is less relevant, since a chihuahua is only likely to pose minor logistical challenges to a burglary. And, in line with the result in (3.58), it would be even more relevant to know that John has a non-chihuahua dog.

Focus here on the proposition *that John does not have a chihuahua*. In contexts where the hearer doesn’t know whether John has a dog, it is negatively relevant to *H*: it provides a weak argument that John’s house is a good candidate for burglary. But if we narrow, by intersection, the worlds to those where John has a dog, the relevance of *not a chihuahua* is reversed: it now provides a (weak) argument that John’s house is a bad burglary candidate. The result above in (3.58) tells us that this arrangement will always be the case: a semantically weak proposition is more relevant than a stronger counterpart when restriction of the context to the weak proposition’s worlds reverses the relevance of the strong term.

Notice that the speaker who utters “dog” seems to imply that he does not think the dog is a chihuahua. This flies in the face of the relevance-based implicature computation hypothesis:  $\omega \wedge \neg\sigma$  (i.e., “a non-chihuahua dog”) is *more* relevant than  $\omega$ , and so a hearer should be able to conclude that, for all the speaker knows, the dog is a chihuahua. Interestingly enough, we find the opposite result with “Doberman” or any other especially ferocious breed of dog: it is more relevant to say “John has a Doberman” than “John has a dog”, yet it seems that if the speaker utters “dog”, he also seems to imply he does not have knowledge that the dog is a Doberman. But the implicature does not persist when we consider breeds of dog that are “ordinary” in terms of their ferociousness: a Labrador, a Saint Bernard, a bloodhound, and so on. For these breeds, “John has a dog” has roughly equivalent relevance to “John has a Labrador/Saint Bernard/bloodhound,” and an utterance of the former does not seem to implicate the speaker does not believe the latter.

What is going on here? We seem to have an implicature from  $\omega \rightsquigarrow \neg B\sigma$  only if  $r_H(\sigma) \neq r_H(\omega)$ .<sup>24</sup> What does this look like in the case of *some* and *all*? That is, can we come up with contexts where *all* is less relevant than *some*?

It is possible, if somewhat artificial, to construct contexts where *some* is more relevant than *all*. Imagine a family reunion, where *A* and his wife *B* commonly know that *A*'s cousin *C* is a total jerk and just sucks the air out of every room he is in, but the other nine cousins are fun:

(3.75)     *A*: I'm bored.

*B*: Some of your cousins are hanging out in the basement.

*B*'s assertion is relevant to *that A should hang out in the basement*, assuming *A* is likely to have fun in the basement if it is true. The more informative (stronger) assertion "All of your cousins are hanging out in the basement" is negatively relevant, since it entails *C* would be present and eliminate any possibility of fun.

The key to understanding this example is to note that the speaker indicates she does not believe  $\sigma$ , but not because she would say  $\sigma$  if she believed it— $\sigma$  is irrelevant (or negatively relevant). The question for a rational hearer, then, is: What would the speaker do if she believed  $\sigma$ ? The answer, it seems, is that she would utter something else entirely, like "It's beautiful outside" or "The old people are playing cards upstairs." That is, the speaker would select a different *H*, since her beliefs would not support making the point *that A should hang out in the basement*.<sup>25</sup>

<sup>24</sup>Notice non-equality is not a sufficient condition (they seem to have to be significantly different).

<sup>25</sup>See also section 4.4 below for a scenario that is, perhaps, clearer, and a more precise derivation of a speaker's expected behavior in such a scenario.



- b. The presence of foreign newspapers at the station ( $FS$ ) make the presence of Italian newspapers at the station more likely:  $P(IS|FS) > P(IS)$  (in terms of relevance,  $r_{IS}(FS) > 0$ ).
- c. The presence of foreign newspapers at the Palace ( $FP$ ) make the presence of Italian newspapers at the Palace more likely:  $P(IP|FP) > P(IP)$  (in terms of relevance,  $r_{IP}(FP) > 0$ ).
- d. The propositions  $BP$  and  $BS$ , that the palace and station, respectively, are stocked with British newspapers, are irrelevant to the question of whether there are Italian papers:  $r_{IS}(BS) = 0$  and  $r_{IP}(BP) = 0$ .

Benz points out that, given this background, a speaker who knows  $BP$  certainly also knows  $FP$ : knowing there are British papers at the palace entails knowing there are foreign papers at the palace. And, although  $BP$  is irrelevant,  $FP$  is relevant, if the speaker is making the point that there are Italian papers. But Benz points out that a speaker who knows there are British papers at the palace, but does not know anything else about foreign papers, should not utter  $FP$ , even though he knows it is true and even though  $FP$  is positively relevant to  $IP$ . In other words, despite the fact that the speaker knows  $FP$ , and despite the fact that it is positively relevant, it is not a good assertion to make in this scenario.

In the scenario, a speaker who utters  $FP$  must be making the point  $IP$ , since, out of the possible answers to the question *where can I get an Italian Newspaper*,  $FP$  is positively relevant only to  $IP$ . But a speaker whose knowledge of  $FP$  comes only from the fact that they know that there are British papers at the palace ought not to be making the point  $IP$ . Speaking imprecisely, there is something about the nature of the speaker's knowledge of  $FP$ , which is based entirely on knowledge of  $BP$ , that does not provide a good foundation to make the point  $IP$ . To drive the point further home, imagine a scenario where the speaker's knowledge of  $FP$  is based on having seen a truck marked "Foreign

Newspaper Delivery Service” at the Palace loading dock; in this scenario, nothing has changed about the relevance relations, and yet the speaker would be able to felicitously assert *FP*. Benz, then, has shown that relevance alone is not enough to predict goodness of assertions.

Benz argues that this result shows that a conversational reasoning must be built on a game theoretic foundation: the complexity of reasoning about how a rational speaker would reason about a rational hearer’s reasoning about a rational speaker (and so on) requires more than simple relevance-based optimization. Instead, Benz advocates a game theoretic approach based on coordinating actions to yield the highest payoff for the hearer (e.g., that the hearer obtains an Italian newspaper). Game theory can be very generally applied to situations where multiple agents are acting (and, indeed, a game strategy that embodies classical Horn/Gazdar generation is conceivable; see Section 4.6 for applications of game theory to the problem of implicature generation), so Benz’s argument is not against the centrality of relevance in scalar implicature computation. Rather, Benz’s argument is that simply replacing informativeness with relevance for (basically) classical Horn/Gazdar generation does not yield good predictions. If an utterance is good to the degree that it is true and relevant, there must be some other balancing force.

What it would take to salvage a relevance scale approach is a theory that predicts that a rational speaker would not choose *IP* as his point if all he knew was *BP*. That is, the speaker doesn’t have a good reason to believe *IP*, and, therefore, he shouldn’t pretend that *FP* is a good reason for the hearer to believe it. Grice’s original Quality maxim directs speakers, “Do not say that for which you lack adequate evidence.” What Benz has shown is that, for certain contexts, speakers should not speak towards *H* for which they lack adequate evidence.

To cast the point in a somewhat different light, suppose the speaker in the above example does decide to speak to *H=that you go to the Palace*. In that case, he has done nothing wrong by uttering *FP*: he has uttered a true sentence that makes *H* more likely. Where

the speaker went wrong was in selecting  $H$ ; after all, he doesn't think  $H$  will result in a good outcome for his hearer. Indeed, one can imagine a somewhat scatterbrained speaker saying, "Well, there are foreign papers at the Palace, but those may just be British papers; on the other hand, the station is often well stocked with papers from all countries. I wish I could be more helpful." Here, the speaker doesn't have a perfect  $H$  to choose, so he speaks to both *that you go to the Palace* and *that you go to the Station*. There seems to be nothing in this that provides a firm footing for an argument against relevance scale approaches in general; instead, it simply points to the need for such approaches to eventually develop a theory of  $H$  selection.<sup>27</sup>

### 3.7 Relevance implicatures

The argument in Chapter 2 is that a theory of scalar implicature is a matter of comparing expectations about the speaker's choice of utterance depending on the speaker's belief state. I have proposed in this chapter that speakers are expected to make assertions that are positively Carnap relevant to the point  $H$  they are making. Putting these pieces together, then, a Bayesian hearer will assume a given  $H$  is the point of the speaker's utterance, calculate the probability that the speaker would have chosen the utterance he chose given that assumption plus the assumption that the speaker has a stronger belief, and compare that to the corresponding probability given the assumption that the speaker does not have the stronger belief (see Chapter 4 for details). But the flip side of this way of thinking about implicatures is that, given some assumptions about the speaker's utterance and beliefs, Grice's classic cases of *Relevance* implicatures may be computed: this is just a matter of making inferences about the selection of the point  $H$  a speaker is making. A brief exposition of how this might occur follows.

One of Grice's classic examples is:

---

<sup>27</sup>See section 3.7 below for discussion, and section 4.4 for a rudimentary attempt to develop such a theory.

(3.77) A: Smith doesn't seem to have a girlfriend these days.

B: He has been paying a lot of visits to New York lately.

Grice posits that B, in making his assertion about Smith's visits to New York, has implicated that Smith may have a girlfriend in New York. B, Grice argues, would be infringing on the maxim "Be relevant" unless he thinks Smith's visits to New York have something to do with his having a girlfriend there. To simplify the case, suppose A had simply asked the question, "Does Smith have a girlfriend these days?" thus defining the issue at stake as  $\{G, \neg G\}$ , where  $G = \text{that Smith has a girlfriend}$ . Moreover, let us assume we're in a context where the content of B's assertion ( $= NY$ ) is positively relevant to  $G$ , (i.e.,  $r_G(NY) > 0$ ); that is, it is understood that assuming Smith has been visiting New York frequently increases the probability that Smith has a girlfriend. The hearer A, then, simply must calculate the following two probabilities about the identity of the point ( $pt$ ) the speaker is addressing:  $P(pt = G | \ulcorner NY \urcorner)$  and  $P(pt = \neg G | \ulcorner NY \urcorner)$ . Using Bayes Rule to calculate the second probability, the hearer may obtain:

$$P(pt = \neg G | \ulcorner NY \urcorner) = \frac{P(\ulcorner NY \urcorner | pt = \neg G)P(pt = \neg G)}{P(\ulcorner NY \urcorner)} \quad (3.78)$$

But, because  $r_G(NY) > 0$ , it follows that  $NY$  is negatively relevant to  $\neg G$  (i.e.,  $r_{\neg G}(NY) < 0$ ; see Footnote 7 for a proof). Taking the PRAG for granted, B's assertion must be positively relevant to the point they are making. Therefore,

$$P(\ulcorner NY \urcorner | pt = \neg G) = 0, \quad (3.79)$$

so

$$P(pt = \neg G | \ulcorner NY \urcorner) = 0, \quad (3.80)$$

which means

$$P(pt = G | \ulcorner NY \urcorner) = 1, \quad (3.81)$$

since

$$P(pt = G | \ulcorner NY \urcorner) = 1 - P(pt \neq G | \ulcorner NY \urcorner) \quad (3.82)$$

$$= 1 - P(point = \neg G | \ulcorner NY \urcorner) \quad (3.83)$$

$$= 1 - 0 \quad (3.84)$$

$$= 1. \quad (3.85)$$

This calculation is essentially independent of the probabilistic framework (the 1 and 0 probabilities are the first clue that this is true). The point here is that, given the PRAG, the hearer knows that, no matter what point  $H$  the speaker is making,  $r_H(NY) > 0$ . Because  $r_G(NY) > 0$  and  $r_{\neg G}(NY) < 0$ , the hearer can conclude that  $H = G$ , rather than  $H = \neg G$ . To sum up, assuming the speaker is speaking to the question *whether Smith has a girlfriend* and that the answer  $NY$  is positively relevant to an affirmative answer, the speaker must be speaking to the affirmative answer.

But this inference falls somewhat short of Grice's claim: that the speaker has implicated that it is possible that Smith has a girlfriend (in New York). So far, I have said nothing about what beliefs a speaker must have in order to speak to a point. It is possible to conceive of a theory of conversation where it is permitted to speak to points you know are false, as well as one where you must only speak to points you believe are true. However, the preliminary evidence seems to indicate that speakers are required to speak to points they do not believe are impossible (unless they indicate otherwise),<sup>28</sup> capturing Grice's (accurate, it seems to me) intuition that B's utterance implicates that, for all he knows, it is possible that Smith has a girlfriend.

---

<sup>28</sup>We have already seen in Section 3.3.1 that speakers often make assertions towards points they believe are false. For example, the speaker who utters "Robinson always draws a large crowd, but he is in America for the year", makes one assertion towards the point  $H = \text{that John should be invited}$ , another assertion towards the point  $\neg H$ , and a compounded assertion that, on balance, argues for  $\neg H$ . There is a wealth of expressions in English that seem to signal a point has been made counter to the beliefs of the speaker. Among these are a large class of concessives, including *nonetheless*, *nevertheless*, *that said*, *etc.*. Nonetheless, as an isolated answer to a polar question, it is reasonable to think that a speaker must not utter something they believe is relevant to an answer they think is impossible.



## 3.8 Consequences of the PRAG

### 3.8.1 Grammatical computation and the PRAG

As discussed in Chapter 1, grammatical systems for implicature computation (Chierchia, 2004; Fox, 2007) make implicated material part of the compositionally-derived semantics of a sentence. So, in particular, “John believes some of his students left” means  $believe(j)(some(left) \wedge \neg all(left))$ . This move is made to account for the observation that usually, when *some* appears outside a downward-entailing environment, it seems to be interpreted as *some and not all*. Likewise, *not all* tends to be interpreted as *some and not all*. But the data in (3.3) supporting the PRAG show that the two sentence types are not interchangeable in contexts where their relevance is in opposition.<sup>29</sup>

This fact apparently means that theories that incorporate scalar implicature into the meaning of utterances via grammatical rules will have to add some extra apparatus that identifies the “implicated” part of meaning as distinct from the plain value, and will then have to stipulate a pragmatic rule that says it is the former, and not the latter, that must be relevant to the point of the utterance, as well as being operative in other phenomena like conditions on the use of *but*. By contrast, in a purely pragmatic account of implicature, the PRAG emerges as a natural property of any theory that recognizes that it is rational to choose a partition (or partitions) of the question under discussion and assert something that increases the hearer’s belief in that partition.

At the same time, these considerations impact many non-grammatical theories of implicature that identify the implicature computation task with deciding what meaning the speaker intended: upper-bounded (implicature-enriched) or literal. These theories encompass many of the more promising formal theories, including the Bidirectional Optimality

<sup>29</sup>An interesting point: *some and not all* is not particularly good in the examples in (3.3), either. The use of *and* rather than *but* forces the proposition’s relevance to be calculated as a whole—so sentences like this will be preferred when both *some* and *not all* are positively relevant to the same point. This distinction warrants some study; in particular, of the difference in distribution between *some* and *some and not all*.

Theory of Blutner and the game-theoretic approach in Parikh (1991) and Ross (2006). In these theories, if a hearer decides to pair a speaker's utterance with an implicature-enriched meaning, that seems to be compatible with reasoning that the speaker wanted to make an upper-bounded point. Because these theories predict the pairing of an utterance of either  $\ulcorner \omega \urcorner$  or  $\ulcorner \neg \sigma \urcorner$  with an intended speaker meaning of  $\omega \wedge \neg \sigma$ , they will be hard pressed to explain the difference in felicity for the two utterances: both (can) convey the same message, so they should be licensed in the same contexts.

To reiterate, the evidence presented in this chapter for the PRAG tells us otherwise: a speaker's point is the asserted, literal meaning; implicatures are computed on top of that, crucially depending on the fact that the speaker's point is the asserted meaning. This is broadly compatible with a theory of conversation that makes the assertion of relevant material a primary goal.

### 3.8.2 Implicature vs. inference

Although implicature and inference are often treated as two sides of the same coin, one need not be present for the other to occur. It is obviously the case that a hearer can make conversational inferences without the speaker intending them: consider *that Professor Plum is the killer*, inferred by the sleuth when Professor Plum asserts, "I was in the Conservatory all night." Likewise, an implication can fall on deaf ears: indirect speech acts like "It sure is hot in here" don't yield a request to open the window if the speaker knows that the window can be opened but the hearer does not; yet it is reasonable to say that the speaker in this case has implied that he wants the hearer to open the window.

The PRAG is broadly compatible with a hearer-centered approach. A speaker whose assertions conform to the PRAG does not have to compute his utterance's scalar implicatures during utterance planning. The selection of a PRAG-compliant assertion depends on the assertion's semantics and not, apparently, on scalar-implicated propositions. This is

a small vindication for the hearer-centered approach over treatments where implicatures are aspects of meaning necessarily intentionally conveyed by cooperative conversants. On the speaker-centered view, utterance planning is more complex: a speaker must compute not only the semantics of a candidate assertion, but all the implicatures of this assertion as well. Inserting implicature computation into the utterance planning process adds considerable complexity to the process. This complexity might eventually be necessary, but it is not necessary in order for a speaker to comply with the PRAG, which simply posits that the asserted content of a speaker's utterance must be positively relevant.

On this view, the aspect of meaning known as scalar implicature is not part of speaker meaning; it corresponds, instead, to the inferences of hearers. Speakers can, and undoubtedly often do, go through the steps they imagine their interlocutor will, figuring out which inferences they are likely to draw. But for at least one criterion related to the assertability of utterances (the PRAG), it is the plain assertion, irrespective of implicatures, that a speaker must consider.

### 3.9 Relevance is Gricean

In this chapter, I examined the role of relevance in conversation: what are the obligations of participants in a conversation in terms of relevance; what does it mean for a participant to make a relevant contribution; and how can the relevance of one assertion be compared to another? Carnap's definition of relevance provides a foundation for answering these questions: participants are obligated to provide a contribution that is positively Carnap relevant to the point they are making; i.e., to an answer to the question under discussion.

The introduction of formally-defined relevance also makes it possible to refine the logic of Gricean scalar reasoning. Horn/Gazdar-based theories of scalar implicature generation are based on informativity (*if the speaker knew something more informative were true, she would have said that instead*), essentially assuming that speakers are expected

to say the most informative thing they know. But this is widely recognized to be an oversimplification: there are always lots of (infinitely many) more informative things a speaker could say, many of which they know are true (like “ $\alpha$  and the sky is blue”). Indeed, Grice’s original formulation of the conversational maxims addresses this by including maxims of Quantity<sub>2</sub>, Relation, and Manner: implicature reasoning only goes through to the extent that these maxims do not interfere. In more careful studies of scalar implicature (Horn, 1989; Geurts, 2009; Katzir, 2007), scalar implicature counterfactual reasoning includes qualifications: if a speaker knew a *relevant*, more informative proposition were true, she would have said that instead. In this chapter, I have gone a step further than that: a cooperative speaker is not expected to make a contribution that is relevant and maximally informative; instead, speakers are simply expected to make a contribution that is maximally informative *about the point they are making*—in other words, maximally relevant.<sup>30</sup>

Grice himself seems to have come to recognize the centrality of relevance for the determination of quantity. In his 1987 “Retrospective Epilogue,” he writes, “To judge whether I have been undersupplied or oversupplied with information seems to require that I should be aware of the identity of the topic to which the information in question is supposed to relate; only after the identification of a focus of relevance can such an assessment be made.” Grice here recognizes that a speaker’s contributions are expected to be informative *about something*: a question under discussion, or a topic of conversation. The main argument in this chapter is that the operative notion of informativity for conversation—that is, information about the topic at hand—is relevance to a point, or answer to such a question, measured as defined as in Carnap (1950).

---

<sup>30</sup>This view is probably compatible with arguments in Green (1995) that speakers are expected to provide exactly as much information as the current exchange requires. Green’s examples show that speakers are only expected to maximize informativeness to the extent that the context requires that. Converting informativeness into relevance provides a way to determine how much information (and about what) a context requires.

# Chapter 4

## A probabilistic theory of scalar implicature

### 4.1 Utterance expectation

In Chapter 2, I argued that the problem of scalar implicature generation can be solved with a theory of hearers' expectations about conversation. In Chapter 3, in anticipation of a role for relevance in these expectations, I presented a formal theory of relevance. In this chapter, I seek to lay out a theory of what hearers expect from speakers' utterances. As put forth in Chapter 2, a theory of utterance expectation is simply a way to assign a conditional probability: the probability that the speaker will utter a given sentence  $\chi$ , given the assumption that he has some belief  $\mathcal{B}$ . In this chapter, I will put the tools developed so far together to provide a theory of the probability hearers assign to speakers' assertions.

The assignment of such probabilities will depend on a ranking of utterances with respect to beliefs: if it is better for the speaker to utter  $\chi$  than  $\eta$  when she has belief  $\mathcal{B}$ , then  $P(\ulcorner \chi \urcorner | \mathcal{B}) > P(\ulcorner \eta \urcorner | \mathcal{B})$ .<sup>1</sup> But, even assuming that there is a well-defined function that

---

<sup>1</sup>A notational reminder:  $\ulcorner \chi \urcorner$  is an utterance of the sentence  $\chi$ , belief states are notated with script like  $\mathcal{B}$ , and so  $P(\ulcorner \chi \urcorner | \mathcal{B})$  is the probability that the speaker will utter  $\chi$ , given the assumption that he has belief state  $\mathcal{B}$ . Also,  $B_\chi$  is the proposition that the speaker believes the proposition denoted by  $\chi$ .

tells us, given a context and a belief  $\mathcal{B}$ , how good an utterance of  $\chi$  is, it is not trivial to translate that function into a probability distribution. The approach I take in this chapter uses *proportionality* to translate the function into a probability distribution; in other words, the probability of a speaker choosing an utterance in a given context is proportional to how good the utterance is in that context. So if  $\chi$  is twice as good as  $\eta$  when a speaker has belief  $\mathcal{B}$ , the speaker is twice as likely to utter  $\chi$  as he is to utter  $\eta$ .

In such a system, we must define Utterance Expectation relative to those beliefs that entail the belief or lack of belief of each possible alternative. This is because the question of whether a speaker should utter  $\omega$  if he believes  $\omega$  depends on whether he believes  $\sigma$  or not. In particular, then, we can't consider  $B\omega$  as a hypothetical belief for utterance expectation when there are alternatives that are neither supported nor contraindicated by  $B\omega$ .

To spell this out, the theory of Utterance Expectation depends on a new definition of *ALT*-maximal belief, which is defined relative to a set of alternative sentences *ALT*:

- (4.1) **Maximal Belief:** A proposition  $\mathcal{B}$  is a *maximal belief* relative to a set of sentences *ALT* (or an *ALT*-maximal belief) iff, for every  $\chi \in ALT$ ,  $\mathcal{B} \subseteq B\chi$  or  $\mathcal{B} \subseteq \neg B\chi$ .

This says *ALT*-maximal beliefs are ones that spell out in detail whether each alternative is believed or not. The number of *ALT*-maximal beliefs depends both on the cardinality of *ALT* and on the nature of the alternatives. With respect to a situation where  $ALT = \{\omega, \sigma\}$ , three *ALT*-maximal beliefs are:  $B\sigma$ ,  $B\omega \wedge \neg B\sigma$ , and  $\neg B\omega$ . In the first case, either the speaker believes  $\sigma$  is true (and can therefore utter  $\ulcorner \sigma \urcorner$  or  $\ulcorner \omega \urcorner$ ); in the second, the speaker does not believe  $\sigma$  is true (and therefore can utter  $\ulcorner \omega \urcorner$  but not  $\ulcorner \sigma \urcorner$ ); and in the third, the speaker does not believe  $\omega$  is true, and can therefore not utter either  $\omega$  or  $\sigma$ . (Note that  $B\omega$  is not *ALT* maximal, since it leaves unspecified whether the speaker

believes, and can therefore utter,  $\sigma$ .)

Given this definition, it is possible to specify a proportional theory of utterance expectation relative to *ALT*-maximal beliefs.<sup>2</sup> To begin, consider a simplified model that depends only on Quality—whether or not the speaker believes  $\nu$ —ignoring relevance and manner for the moment.

(4.2) **Utterance Expectation (*ALT*-maximal; quality based):** For any sentence  $\nu$ , set of alternatives *ALT*, and *ALT*-maximal belief  $\mathcal{B}$ :

$$P(\ulcorner \nu \urcorner | \mathcal{B}) = \frac{P(B\nu | \mathcal{B})}{\sum_{\chi \in \text{ALT}} P(B\chi | \mathcal{B})}. \quad (4.3)$$

What the definition says is that for an *ALT*-maximal belief  $\mathcal{B}$ , probabilities are distributed equally among those alternative utterances that  $\mathcal{B}$  supports.

Before exploring the consequences of the definition in (4.2), it will be useful to introduce some notation.<sup>3</sup> First, the definition of *ALT*-maximal beliefs suggests partitioning the set of worlds into regions that fully specify the truth or falsity of each alternative. Expanding the example somewhat to handle an *ALT* set of  $\{\neg\omega, \neg\sigma, \omega, \sigma\}$ , the partition simply includes the regions  $\omega \wedge \neg\sigma$  (where  $\omega$  is true but  $\sigma$  is false, and vice versa for the negated sentences),  $\sigma$  ( $\omega$  and  $\sigma$  are both true), and  $\neg\omega$  (neither  $\omega$  nor  $\sigma$  is true). The three

---

<sup>2</sup>There is a potential downside with limitation of conversational reasoning to *ALT*-maximal beliefs: as the set of alternatives grows, *ALT*-maximal beliefs become less and less interesting. For example, if *ALT* contains  $\omega = \text{Some of the students got a score higher than 90}$  in a context where 70 is a passing grade,  $[\sigma]$  is not a belief the speaker can reason about, since the belief that all students passed does not entail a belief about whether any students got a score higher than 90. Instead, *ALT*-maximal beliefs are things like  $[o_1 o_2 \sigma]$ , where  $o_1$  indicates a negation of the proposition that some of the students scored higher than 90 and  $o_2$  indicates a negation of the proposition that all of the students passed. Perhaps this is a limitation in the theoretical apparatus that could be overcome with some formal gymnastics. On the other hand, perhaps it tells us something deeper: Gricean reasoning fails over large alternative sets, and it therefore makes sense to limit the set of alternatives being considered in principled ways.

<sup>3</sup>This notation is based on the notation used in Franke (2009), where bracketed lists appear as subscripts, in which information states are partitioned relative to the alternatives available along lines similar to the *ALT*-maximality approach here.

regions (for notational simplicity, I use  $o\omega$  for  $\omega \wedge \neg\sigma$ ) are illustrated in Figure 4.1a.

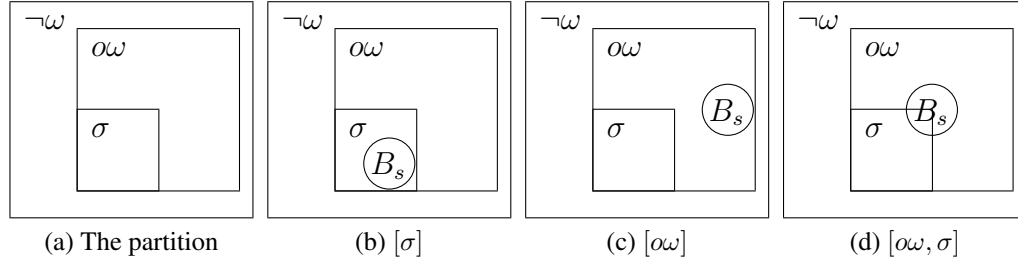


Figure 4.1: A partition induced by  $ALT = \{\neg\omega, \neg\sigma, \omega, \sigma\}$  and three  $ALT$ -maximal belief states.

An  $ALT$ -maximal belief can be specified as a comprehensive list of each region of the partition that contains worlds that the speaker believes are possible. With respect to this model,  $[\sigma]$  is shorthand for the belief  $B\sigma$ , which is  $ALT$ -maximal (it supports an utterance of both  $\ulcorner\sigma\urcorner$  and  $\ulcorner\omega\urcorner$ , but not  $\ulcorner\neg\sigma\urcorner$  or  $\ulcorner\neg\omega\urcorner$ ). To spell out the other beliefs,  $[o\omega]$  is the proposition  $B(\omega \wedge \neg\sigma)$  and  $[o\omega, \sigma]$  is the proposition  $\neg B\neg(\omega \wedge \neg\sigma) \wedge \neg B\neg\sigma \wedge \neg B(\omega \wedge \neg\sigma) \wedge \neg B\sigma$ . For now, beliefs designated by the bracket notation are speaker beliefs, but, when other agents' beliefs are represented in the same notation, I will include a subscript on the opening bracket:  $[_s\sigma]$  is  $B_s\sigma$ , other subscripts will denote other agents.

Without making extra assumptions, this theory of utterance expectation is not enough to produce scalar implicatures in a simplified model where the set of alternative utterances is  $ALT = \{\sigma, \omega, \neg\sigma, \neg\omega\}$ . Recall the simplified scalar implicature criterion in (2.34):

(2.34) Scalar implicature criterion (simplified):

Let  $\sigma \subset \omega$  be sentences. Then an utterance of  $\omega$  implicates  $B\omega \wedge \neg B\sigma$  if and only if:

$$P(\ulcorner\omega\urcorner | B\omega \wedge \neg B\sigma) > P(\ulcorner\omega\urcorner | B\sigma). \quad (4.4)$$

Notice  $B\omega \wedge \neg B\sigma$  is not  $ALT$ -maximal, since there is an alternative  $\neg\sigma$  for which it entails neither  $B\neg\sigma$  nor  $\neg B\neg\sigma$ . Instead, we must consider the two beliefs that, together,



make up  $B\omega \wedge \neg B\sigma$ :  $[o\omega]$ , which supports an utterance of  $\ulcorner \neg \sigma \urcorner$ , and  $[o\omega, \sigma]$ , which does not. (Notice that we have simply split the speaker belief associated with epistemic weak implicatures,  $B\omega \wedge \neg B\sigma$ , into the speaker belief associated with epistemic strong implicature,  $[o\omega]$ , and the speaker belief associated with an “ignorance” implicature,  $[o\omega, \sigma]$ .) Focusing for now on the epistemic strong implicature, this is simply:

$$P(\ulcorner \omega \urcorner | [o\omega]) > P(\ulcorner \omega \urcorner | [\sigma]). \quad (4.5)$$

Computing these two values using the Quality-based theory of utterance expectation in (4.2), we get:

$$P(\ulcorner \omega \urcorner | [o\omega]) = \frac{P(B\omega | [o\omega])}{\sum_{\eta \in ALT} P(B\eta | [o\omega])} \quad (4.6)$$

$$= \frac{1}{P(B\sigma | [o\omega]) + P(B\omega | [o\omega]) + P(B\neg\sigma | [o\omega]) + P(B\neg\omega | [o\omega])} \quad (4.7)$$

$$= \frac{1}{0 + 1 + 1 + 0} \quad (4.8)$$

$$= \frac{1}{2}, \quad (4.9)$$

and:

$$P(\ulcorner \omega \urcorner | [\sigma]) = \frac{P(B\omega | [\sigma])}{\sum_{\eta \in ALT} P(B\eta | [\sigma])} \quad (4.10)$$

$$= \frac{1}{P(B\sigma | [\sigma]) + P(B\omega | [\sigma]) + P(B\neg\sigma | [\sigma]) + P(B\neg\omega | [\sigma])} \quad (4.11)$$

$$= \frac{1}{1 + 1 + 0 + 0} \quad (4.12)$$

$$= \frac{1}{2}. \quad (4.13)$$

In this case, then,  $P(\ulcorner \omega \urcorner | [o\omega]) = P(\ulcorner \omega \urcorner | [\sigma])$ . The intuition behind this is that, if the speaker holds the first belief ( $[o\omega]$ ), there are two compatible (and therefore equally likely) utterances:  $\omega$  and  $\neg\sigma$  (“some” and “not all”). If the speaker holds the second belief ( $[\sigma]$ ),

there are also two compatible utterances:  $\omega$  and  $\sigma$  (“some” and “all”).

In addition, there is a third *ALT*-maximal knowledge state in which a speaker might utter  $\omega$ :  $[o\omega, \sigma]$ , the “ignorance” belief state where the speaker knows one of the two alternative states is true, but is uncertain which one it is. With respect to this state, we have:

$$P(\ulcorner \omega \urcorner | [o\omega, \sigma]) = \frac{P(B\omega | [o\omega, \sigma])}{\sum_{\eta \in ALT} P(B\eta | [o\omega, \sigma])} \quad (4.14)$$

$$= \frac{1}{1 + 0 + 0 + 0} \quad (4.15)$$

$$= 1. \quad (4.16)$$

The intuition here is simple: if the speaker knows “some” is true but is unsure whether “all” is, there’s only one thing he can say: “some”. So the probability of this utterance in this belief state is 1.

To help visualize what is going on here, consider again the three epistemic states from above and the available utterances in those states. In Figure 4.2, the circle represents the speaker’s belief state; the utterances inside the circle are those compatible with that belief state.

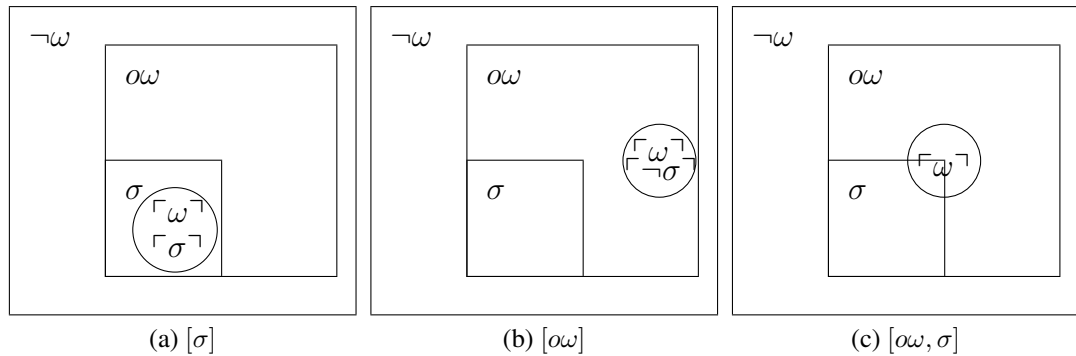


Figure 4.2: A partition induced by  $ALT = \{\neg\omega, \neg\sigma, \omega, \sigma\}$  and three *ALT*-maximal belief states.

Because the model is set up to only attend to the maxim of Quality, probabilities are

split evenly among alternatives compatible with a given belief state. This gives us the following probabilities:

$$P(\ulcorner\omega\urcorner|\ulcorner\sigma\urcorner) = P(\ulcorner\sigma\urcorner|\ulcorner\sigma\urcorner) = \frac{1}{2} \quad (4.17)$$

$$P(\ulcorner\omega\urcorner|\ulcorner o\omega\urcorner) = P(\ulcorner\neg\sigma\urcorner|\ulcorner o\omega\urcorner) = \frac{1}{2} \quad (4.18)$$

$$P(\ulcorner\omega\urcorner|\ulcorner o\omega, \sigma\urcorner) = 1 \quad (4.19)$$

What these results say is that, when the speaker utters  $\ulcorner\omega\urcorner$ , scalar reasoning favors the “ignorance” knowledge state  $\ulcorner o\omega, \sigma\urcorner$  over the alternatives where the speaker knows more. This, in turn, means that  $P(\ulcorner\omega\urcorner|B\omega \wedge \neg B\sigma) > P(\ulcorner\omega\urcorner|B\sigma)$ , satisfying the scalar implicature criterion from (2.34), since, assuming  $P(\ulcorner o\omega\urcorner)$  and  $P(\ulcorner o\omega, \sigma\urcorner)$  are both non-zero,

$$P(\ulcorner\omega\urcorner|B\omega \wedge \neg B\sigma) = P(\ulcorner\omega\urcorner|\ulcorner o\omega\urcorner \cup \ulcorner o\omega, \sigma\urcorner) \quad (4.20)$$

$$= \frac{P(\ulcorner\omega\urcorner|\ulcorner o\omega\urcorner)P(\ulcorner o\omega\urcorner) + P(\ulcorner\omega\urcorner|\ulcorner o\omega, \sigma\urcorner)P(\ulcorner o\omega, \sigma\urcorner)}{P(\ulcorner o\omega\urcorner) + P(\ulcorner o\omega, \sigma\urcorner)} \quad (4.21)$$

$$= \frac{\frac{1}{2}P(\ulcorner o\omega\urcorner) + 1P(\ulcorner o\omega, \sigma\urcorner)}{P(\ulcorner o\omega\urcorner) + P(\ulcorner o\omega, \sigma\urcorner)} \quad (4.22)$$

$$> \frac{\frac{1}{2}P(\ulcorner o\omega\urcorner)}{P(\ulcorner o\omega\urcorner)} \quad (4.23)$$

$$= \frac{1}{2}, \quad (4.24)$$

whereas

$$P(\ulcorner\omega\urcorner|\ulcorner\sigma\urcorner) = \frac{1}{2}. \quad (4.25)$$

So the scalar implicature criterion is satisfied for cases with non-trivial priors. It is satisfied, however, because an ignorance implicature is generated; in general, we want to be able to generate an epistemic strong implicature.

What working through the consequences of the quality-only theory with respect to this model suggests is that the theory of utterance expectation can not be based on the maxim of Quality alone. This result is a desirable one: if speakers simply selected a true utterance at random, we wouldn't expect hearers to be able to calculate scalar implicatures.<sup>4</sup> What Chapter 3 suggested is that speakers don't just select randomly from among those utterances they believe; instead, they favor utterances that are more relevant. With respect to the model at hand, this will distribute the probability weight unevenly among those utterances compatible with a belief: if  $\sigma$  is four times as relevant as  $\omega$ , the speaker will be four times more likely to utter it if she believes  $[\sigma]$ .

The most straightforward way to incorporate relevance into a theory of utterance expectation is to multiply relevance by probability of belief in the definition in (4.2):

(4.26) **Utterance Expectation (Quality and relevance based):** For any sentence  $\nu$ , conversational point  $H$ , set of alternatives  $ALT$ , and  $ALT$ -maximal belief  $\mathcal{B}$ :

$$P(\ulcorner \nu \urcorner | \mathcal{B}) = \frac{P(B\nu | \mathcal{B})r_H(\nu)}{\sum_{\chi \in ALT} P(B\chi | \mathcal{B})r_H(\chi)}.$$

A caveat: ordinary Carnap relevance will not work here as an instantiation of the  $r$  relation, since Carnap relevance values can be negative. Moreover, because of the PRAG (see Chapter 3), which says that speakers must make positively relevant assertions, the proper notion of relevance for present purposes is *positive relevance*; that is, a proposition's  $r$  value is the same as Carnap relevance if positive, and otherwise 0:

$$pr_H(\chi) := \max(\{0, r_H(\chi)\}) \quad (4.27)$$

---

<sup>4</sup>Though see discussion in 4.6 below of Franke (2009), where iterated reasoning produces scalar implicatures in an essentially Quality-only theory.

So, since  $pr_H(\nu)$  is non-negative for every  $\nu$ ,  $P(\ulcorner \nu \urcorner | \mathcal{B})$  is non-negative. To avoid confusion ( $pr$  is often used to mean *probability*), I use  $r$  for positive relevance from here on.<sup>5</sup>

There is one more technical obstacle to address before exploring the applications of the relevance-based theory of utterance expectation. Notice that, because Utterance Expectation applies only to *ALT*-maximal beliefs, there is no straightforward way to compute a value like  $P(\mathcal{B} | \ulcorner \nu \urcorner)$ , since that involves the Bayesian transformation

$$\frac{P(\ulcorner \nu \urcorner | \mathcal{B})P(\mathcal{B})}{P(\ulcorner \nu \urcorner)}, \quad (4.28)$$

and  $P(\ulcorner \nu \urcorner)$  can't be derived straightforwardly using Utterance Expectation, since  $\Omega$ , the set of all worlds, is not *ALT*-maximal.<sup>6</sup> Moreover, there are non-*ALT* maximal beliefs featured in the general definition of implicature from (2.26), repeated below.

#### (2.26) Definition of Implicature

<sup>5</sup>Note that absolute value of relevance is not what we want: assuming a speaker is speaking towards the point  $H$ , a highly negatively relevant utterance would receive a high probability, in violation of the PRAG. See discussion surrounding example (3.6) for elaboration.

<sup>6</sup>But note that the definition in (4.26) can be expanded to provide utterance expectation for arbitrary unions of mutually exclusive *ALT*-maximal beliefs:

(4.i) For any  $\nu \in ALT$  and speaker belief state  $\mathcal{B}$  that is partitioned into *ALT*-maximal beliefs  $\mathcal{B}_1, \dots, \mathcal{B}_n$ ,

$$\begin{aligned} P(\ulcorner \nu \urcorner | \mathcal{B}) &= \sum_{i=1}^n P(\ulcorner \nu \urcorner | \mathcal{B}_i)P(\mathcal{B}_i | \mathcal{B}) \\ &= \sum_{i=1}^n \frac{P(B\nu | \mathcal{B}_i)r_H(\nu)}{\sum_{\chi \in ALT} P(B\chi | \mathcal{B}_i)r_H(\chi)} P(\mathcal{B}_i | \mathcal{B}). \end{aligned}$$

This formulation is a simple consequence of the axioms of probability theory: it simply takes the conditional probabilities for each *ALT*-maximal belief and adds them up, weighted according to the likelihood of the belief. For technical reasons, this more general definition will not be used; instead, I will use the version based on odds of *ALT*-maximal beliefs given in (4.29).

Given a set  $\mathcal{B}$ , an utterance of  $\nu$  implicates  $\mathcal{B}$  just in case

$$\frac{P(B\nu \wedge \mathcal{B} | \ulcorner \nu \urcorner)}{P(B\nu \wedge \neg \mathcal{B} | \ulcorner \nu \urcorner)} > \frac{P(\ulcorner \nu \urcorner \wedge \mathcal{B})}{P(B\nu \wedge \neg \mathcal{B})}$$

Here, even if we limit  $\mathcal{B}$  to *ALT*-maximal beliefs, therefore ensuring that  $B\nu \wedge \mathcal{B}$  is also *ALT*-maximal, its negation  $\neg \mathcal{B}$  will not, in general, be *ALT*-maximal.

This obstacle can be surmounted by considering odds ratios of *ALT*-maximal beliefs relative to a given utterance. In this framework we talk about an utterance favoring one speaker belief over another, rather than implicating a belief simpliciter:

(4.29) Definition of implicature (odds version): Given a sentence  $\nu$ , a set of alternatives *ALT*, and *ALT*-maximal beliefs  $\mathcal{B}_1$  and  $\mathcal{B}_2$  such that  $\mathcal{B}_1 \subseteq B\nu$  and  $\mathcal{B}_2 \subseteq B\nu$ , an utterance of  $\ulcorner \nu \urcorner$  *implicates*  $\mathcal{B}_1$  over  $\mathcal{B}_2$  iff:

$$\frac{P(\mathcal{B}_1 | \ulcorner \nu \urcorner)}{P(\mathcal{B}_2 | \ulcorner \nu \urcorner)} > \frac{P(\mathcal{B}_1 | B\nu)}{P(\mathcal{B}_2 | B\nu)}$$

This way of treating implicatures as an odds ratio, though ostensibly somewhat clunky, has the advantage of allowing the comparative evaluation of inferences: it permits us to ask how likely is it that George thinks not all of his advisors are crooks, vs. George thinking it is possible that only some are crooks and possible that all are crooks (i.e., “embedded” implicature interpretations). Attractive empirical consequences of this will be outlined in Chapter 5.

The odds-based criterion also lends itself to a simple and straightforward way of calculating implicatures. First, notice that the odds ratio in (4.29) may be reduced substantially:

we can simplify the left hand side as follows:

$$\frac{P(\mathcal{B}_1|\ulcorner\nu\urcorner)}{P(\mathcal{B}_2|\ulcorner\nu\urcorner)} = \frac{\frac{P(\ulcorner\nu\urcorner|\mathcal{B}_1)P(\mathcal{B}_1)}{P(\ulcorner\nu\urcorner)}}{\frac{P(\ulcorner\nu\urcorner|\mathcal{B}_2)P(\mathcal{B}_2)}{P(\ulcorner\nu\urcorner)}} \quad (4.30)$$

$$= \frac{P(\ulcorner\nu\urcorner|\mathcal{B}_1)P(\mathcal{B}_1)}{P(\ulcorner\nu\urcorner|\mathcal{B}_2)P(\mathcal{B}_2)} \quad (4.31)$$

$$= \frac{\frac{r_{H\nu}P(B\nu|\mathcal{B}_1)}{N(\mathcal{B}_1)} P(\mathcal{B}_1)}{\frac{r_{H\nu}P(B\nu|\mathcal{B}_2)}{N(\mathcal{B}_2)} P(\mathcal{B}_2)} \quad (4.32)$$

$$= \frac{N(\mathcal{B}_2) P(\mathcal{B}_1)}{N(\mathcal{B}_1) P(\mathcal{B}_2)}. \quad (4.33)$$

I use the notation  $N(\mathcal{B})$  to abbreviate the following sum, relative to a belief  $\mathcal{B}$ :

$$N(\mathcal{B}) := \sum_{\chi \in ALT} r_{H\chi} P(B\chi|\mathcal{B}) \quad (4.34)$$

Plugging this into (4.29), we obtain a new implicature criterion: an utterance of  $\ulcorner\nu\urcorner$  implicates  $\mathcal{B}_1$  over  $\mathcal{B}_2$  iff:

$$\frac{N(\mathcal{B}_2) P(\mathcal{B}_1)}{N(\mathcal{B}_1) P(\mathcal{B}_2)} > \frac{P(\mathcal{B}_1|B\nu)}{P(\mathcal{B}_2|B\nu)}. \quad (4.35)$$

Rearranging terms and simplifying, we obtain:

$$\frac{N(\mathcal{B}_2)}{N(\mathcal{B}_1)} > \frac{P(\mathcal{B}_1|B\nu)}{P(\mathcal{B}_2|B\nu)} \frac{P(\mathcal{B}_2)}{P(\mathcal{B}_1)} \quad (4.36)$$

$$= \frac{P(B\nu|\mathcal{B}_1)P(\mathcal{B}_1)P(B\nu)}{P(B\nu|\mathcal{B}_2)P(\mathcal{B}_2)P(B\nu)} \frac{P(\mathcal{B}_2)}{P(\mathcal{B}_1)} \quad (4.37)$$

$$= \frac{P(B\nu|\mathcal{B}_1)}{P(B\nu|\mathcal{B}_2)} \quad (4.38)$$

$$= 1, \quad (4.39)$$

where the last step is justified by the fact that  $\mathcal{B}_1$  and  $\mathcal{B}_2$  both entail  $B\nu$ .

This leads to the simple implicature criterion below:

(4.40) Implicature Criterion (odds version): Given a sentence  $\nu$  and  $ALT$ -maximal beliefs  $\mathcal{B}_1$  and  $\mathcal{B}_2$  such that  $\mathcal{B}_1 \subseteq B\nu$  and  $\mathcal{B}_2 \subseteq B\nu$ , an utterance  $\ulcorner \nu \urcorner$  *implicates*  $\mathcal{B}_1$  over  $\mathcal{B}_2$  iff:

$$N(\mathcal{B}_2) > N(\mathcal{B}_1).$$

This makes implicature computation little more than an exercise in tabulation. Let us return to the example above, where  $ALT = \{\omega, \sigma, \neg\omega, \neg\sigma\}$ . Assume a stereotypical relevance arrangement, where  $r_H(\sigma) > r_H(\omega) > 0$ . As outlined in Chapter 3 (specifically, proven starting in (3.16)), since the first two alternatives are positively relevant to the point being made, the latter two are negatively relevant. What this means is that those alternatives will drop out of consideration: their positive relevance value is 0. This leads to a nearly trivial computation of scalar implicature for this example, illustrated in Figure 4.3. What the example shows is that  $[\sigma]$  is considered an unlikely belief for an utterer

	$r_H(x)$	$[o\omega]$	$[\sigma]$	$[o\omega, \sigma]$
$\ulcorner \omega \urcorner$	0.1	1	1	1
$\ulcorner \sigma \urcorner$	0.4	0	1	0
$\ulcorner \neg\sigma \urcorner$	0	1	0	0
$\ulcorner \neg\omega \urcorner$	0	0	0	0
$N(\mathcal{B})$		0.1	0.5	0.1

Figure 4.3: Computation of scalar implicature in simple case. Left column lists elements of  $ALT$ ; top row lists  $ALT$ -maximal beliefs under consideration; and  $r_H(x)$  is the assumed relevance value of each alternative.

of  $\ulcorner \omega \urcorner$ , relative to  $[o\omega]$  and  $[o\omega, \sigma]$ , because there is an alternative available in that belief state that is not available in the other two—namely,  $\sigma$ —and, moreover, that alternative is more relevant than the alternative that is available in the other two states. Specifically, the table shows scalar reasoning for an utterance of  $\ulcorner \omega \urcorner$ , in this model, favoring  $[o\omega]$  to  $[\sigma]$  by a ratio of 5 : 1, and likewise for  $[o\omega, \sigma]$ .

In this example, we are able to see what is special about scalar cases, where one utterance alternative asymmetrically entails another. The column marked  $[\sigma]$  is compatible



with two utterances— $\lceil \omega \rceil$  and  $\lceil \sigma \rceil$ , whereas the columns  $[o\omega]$  and  $[o\omega, \sigma]$  are only compatible with one:  $\lceil \omega \rceil$ . The extra alternative available in  $[\sigma]$  “steals” some of the probability weight from  $\lceil \omega \rceil$ , and thus makes it less likely that a speaker who utters  $\lceil \omega \rceil$  believes  $[\sigma]$ . In even more intuitive terms: a speaker who utters  $\lceil \omega \rceil$  is unlikely to believe  $[\sigma]$  because if the speaker had believed  $[\sigma]$  he would have been more likely to utter  $\lceil \sigma \rceil$ . (Notice there is no scalar effect between belief states  $[o\omega]$  and  $[o\omega, \sigma]$ ; further discussion of this fact is saved for section 4.2.)

What happens when we “open up” the *ALT* set to allow the symmetric alternative, “some but not all,” abbreviated  $o\omega$ , into the picture? First, note that the result in (3.58) requires  $r_H(o\omega) < r_H(\omega)$ , since we assumed  $r_H(\sigma) > r_H(\omega)$ . For concreteness, suppose  $r_H(o\omega) = 0.05$ . As the table in Figure 4.4 shows, the scalar implicature interpretation is preserved. Here the strength of the implicature to  $[o\omega]$  is weaker, at 10 : 3 in favor of  $[\omega]$

	$r_H(x)$	$[o\omega]$	$[\sigma]$	$[o\omega, \sigma]$
$\lceil \omega \rceil$	0.1	1	1	1
$\lceil \sigma \rceil$	0.4	0	1	0
$\lceil o\omega \rceil$	0.05	1	0	0
$N()$		0.15	0.5	0.1

Figure 4.4: Computation of scalar implicature with the inclusion of symmetric alternative  $o\omega$ .

over  $[\sigma]$ . This seems intuitive: the consideration of another alternative that the speaker might utter in  $[o\omega]$  makes it somewhat less likely that the speaker holds this belief when he utters  $\lceil \omega \rceil$ .<sup>7</sup>

The tabular format for calculating implicatures allows us to immediately see the effect of including more alternatives, like “many” and “most”. Because *Bmany* and *Bmost* are entailed by  $[\sigma]$  and contradicted by  $[o\omega]$ , their inclusion increases  $N([\sigma])$  but not  $N([o\omega])$ .

<sup>7</sup> It should not escape the reader’s notice that, with respect to this 3-element *ALT* set,  $\lceil \omega \rceil$  is predicted to implicate  $[\sigma]$  over  $[o\omega]$  when  $r_H(o\omega) > r_H(\omega)$  (or, equivalently, by the result in (3.58), when  $r_H(\omega) > r_H(\sigma)$ ). See discussion in section 4.3 for a principled basis on which this effect is attenuated, due to “Manner”-based complexity considerations, and section 4.4, which addresses the symmetry problem in full.

It is therefore straightforward to see that their inclusion strengthens the implicature of  $[o\omega]$  over  $[\sigma]$ , regardless of their relevance values (which will tend to be relatively high, strengthening the implicature further). Still more alternatives exist that strengthen the implicature: “some and maybe even all”, for example, is supported by  $[\sigma]$  and  $[o\omega, \sigma]$ , but not not by  $[o\omega]$ . What would alternatives look like that would weaken that implicature (or even reverse it)? Those would be alternatives that are entailed by  $[o\omega]$  and contradicted by  $[\sigma]$ . Such alternatives are relatively difficult to find. For example, alternatives like “most but not all” are *not* supported by  $[o\omega]$ .

## 4.2 The implicatures of disjunction and the role of prior probabilities

We have seen how the probabilistic system generates implicatures for sentences with simple scalar terms like *some*. More complex is *or*, since it has, in addition to the epistemic strong *not all* implicature, epistemic weak clausal implicatures: “*A* or *B*” implies the speaker neither believes nor disbelieves either disjunct is true. To see the calculation of these implicatures in the present system, consider the table in Figure 4.5. The epistemic

	$r_H(x)$	$[A]$	$[B]$	$[AB]$	$[A, AB]$	$[B, AB]$	$[A, B]$	$[A, B, AB]$
“ <i>A</i> ”	0.3	1	0	1	1	0	0	0
“ <i>B</i> ”	0.3	0	1	1	0	1	0	0
“ <i>A</i> and <i>B</i> ”	0.6	0	0	1	0	0	0	0
“ <i>A</i> or <i>B</i> ”	0.1	1	1	1	1	1	1	1
$N()$		0.4	0.4	1.3	0.4	0.4	0.1	0.1

Figure 4.5: The implicatures of disjunction. Note  $AB$  abbreviates  $AB$ .

weak implicature is clear: an utterance of “*A* or *B*” implicates  $[A, B]$  (that is, the speaker believes exactly one of *A* and *B* is true) over  $[AB]$  (that is, the speaker believes both are true) with 13 : 1 odds. Interestingly, the table gives even odds to  $[A, B]$  and  $[A, B, AB]$ :

the epistemic strong and epistemic weak interpretations. This is desirable since the epistemic strong inference depends on beliefs about how well-informed (or opinionated) the speaker is. In this model, speaker opinionatedness is a question of prior beliefs:  $P([A, B])$  vs.  $P([A, B, AB])$ . So, if we assume (prior to scalar reasoning) the speaker is likely to be opinionated about  $AB$  (e.g.,  $P([A, B]) = 0.3$  and  $P([A, B, AB]) = 0.03$ ), we obtain:

$$\frac{P([A, B]|\lceil A \text{ or } B \rceil)}{P([A, B, AB]|\lceil A \text{ or } B \rceil)} = \frac{N([A, B, AB])}{N([A, B])} \frac{P([A, B])}{P([A, B, AB])} \quad (4.41)$$

$$= \frac{0.1}{0.1} \frac{0.3}{0.03} \quad (4.42)$$

$$= \frac{10}{1}. \quad (4.43)$$

That is, the posterior odds of the epistemic strong to the epistemic weak interpretation are equal to the prior odds (in this case, highly skewed in favor of the epistemic strong interpretation). This should be a clear demonstration of the formalism's close adherence to intuitions about scalar reasoning: the epistemic weak interpretation is created by scalar reasoning (which takes into account the available alternatives), and the epistemic strong interpretation is a consequence of prior beliefs about the speaker's opinionatedness.

The disjunction example provides a few more important predictions. One is that  $[A]$  is indistinguishable from  $[A, AB]$  (and similarly for  $[B]$ ). These pairs, again, will be distinguished by their prior probabilities; the utterance of “ $A$  or  $B$ ” implicates that the speaker probably does not hold the belief  $A$  or the belief  $B$ , but, if the hearer did in fact hold one or the other—say  $A$ —scalar reasoning does not help the hearer decide whether the speaker thinks  $B$  is possible.

Again, the results for disjunction do not depend on limiting the alternatives to those listed in Figure 4.5. Consider the effect of the addition of the alternative “ $A$  or  $B$  and not both”, as shown in Figure 4.6. The most notable difference between Figure 4.5 and Figure 4.6 is that the  $N$  values for  $[A, B]$  and  $[A, B, AB]$  are now different, with  $N([A, B]) > N([A, B, AB])$ . This means that an utterance of “ $A$  or  $B$ ” implicates

	$r_H(x)$	$[A]$	$[B]$	$[AB]$	$[A, AB]$	$[B, AB]$	$[A, B]$	$[A, B, AB]$
$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$
“ $A$ or $B$ and not both”	0.05	1	1	0	0	0	1	0
$N()$		0.45	0.45	1.3	0.4	0.4	0.15	0.1

Figure 4.6: The inclusion of the alternative “ $A$  or  $B$  and not both” pushes the interpretation towards epistemic weakness (as appended to Figure 4.5).

$[A, B, AB]$  over  $[A, B]$  by a factor of 3 : 2; in other words, scalar reasoning favors the ignorance implicature. Recall, from above, that the presence or absence of epistemic strong implicature is entirely a consequence of prior beliefs about speaker opinionatedness, so, if there is a strong prior belief that the speaker is opinionated, an epistemic strong interpretation is still preferred. Moreover, the fact that the inclusion of “ $A$  or  $B$  and not both” as a potential alternative utterance weakens the epistemic strong implicature is intuitive: a speaker who has the belief  $[A, B]$  could have unambiguously conveyed that fact by uttering this alternative utterance, slightly undermining the epistemic strong implicature. This point should further underscore the (neo)-Gricean point that epistemic strong implicatures are not the result of scalar reasoning: indeed, scalar reasoning actually disfavors epistemic strong interpretations.

### 4.3 Disjunction, complexity, and the maxim of Manner

Consider the following problematic case of disjunction from Chierchia (2004):

(4.44) Kai ate the broccoli or some of the peas.

Chierchia pointed out that Horn-style Generation would negate the alternative “Kai ate the broccoli or all of the peas” and lead the hearer to conclude that Kai did not eat the broccoli. Sauerland (2004) and Russell (2006) argue that independent considerations about the

epistemic status of implicatures provide an argument that higher-fidelity Gricean systems do not make such a prediction. The current system accords well with these arguments. To see why, let  $b$  be *that Kai ate the broccoli*,  $sp$  be *that Kai ate some of the peas*,  $osp$  be *that Kai ate some but not all of the peas*, and  $ap$  be *that Kai ate all of the peas*. Note that  $B\neg(b \vee ap)$  is not *ALT*-maximal with respect to a realistic set of alternatives: in particular, it does not entail  $Bsp$  or its negation. However, note that the speaker’s belief in the truth of (4.44), plus the inference the speaker believes *that Kai did not eat the broccoli or all of the peas*, yields the conclusion that the speaker believes *that Kai ate some and not all of the peas and did not eat the broccoli*; i.e.,  $[osp]$ . This belief is *ALT*-maximal, so the system allows the hearer to reason about how likely this belief is relative to, say,  $[b, osp]$  (the speaker believes exactly one of  $b, osp$ ). As illustrated in Figure 4.7, scalar reasoning

	$r_H(x)$	$[osp]$	$[b, osp]$
“ $sp$ ”	0.02	1	0
“ $ap$ ”	0.04	0	0
“ $b$ ”	0.04	0	0
“ $b$ and $sp$ ”	0.1	0	0
“ $b$ and $ap$ ”	0.1	0	0
“ $b$ or $sp$ ”	0.02	1	1
“ $b$ or $ap$ ”	0.04	0	0
$N()$		0.04	0.02

Figure 4.7: An utterance of “ $b$  or  $sp$ ” implicates  $[b, osp]$  over  $[osp]$  with odds of 2 : 1.

here leads the hearer to prefer  $[b, osp]$  to  $[osp]$  with 2 : 1 odds; the magnitude of this ratio can be shifted by changing relevance scores of the alternatives, but the direction cannot. What this means is that the pernicious inference that Chierchia identified is not generated in this system; instead, the system disfavors the inference that Kai did not eat the broccoli.

However, it is worth pointing out that the odds are intuitively better for  $[osp]$  than we would expect: an utterance of (4.44) is rarely compatible with worlds like  $[osp]$  where the speaker believes Kai did not eat the broccoli. Within the current theory, there are three possible sources of this problem: we assumed the relevance of  $sp$  was too low relative

to that of *b or sp*; we neglected to consider relevant alternatives that are compatible with *[osp]* but not with *[b, osp]*; or the system's failure to take into account the complexity of utterances makes "*b or sp*" unrealistically likely when the speaker believes *[osp]*. The first two considerations seem relatively weak: the set of listed alternatives is a plausible one, and the relative relevance of the two alternatives seems plausible as well (a situation where *sp* is much more relevant than *b or sp* is unusual: eating the broccoli is probably at least as good as eating some of the peas, say for purposes of obtaining dessert). But the last option is one I have deferred until now, and, intuitively, one reason a speaker doesn't say "Kai ate the broccoli or some of the peas" when the speaker knows very well that Kai ate some of the peas and not the broccoli is that he went to a bunch of extra effort to include the broccoli in his utterance.<sup>8</sup> What we need, then, is a way to incorporate Grice's maxim of Manner.

Katzir (2007) developed a simple formal theory to model Manner, but the theory delivered not a numeric value for the complexity of utterances but a binary distinction between those alternatives that could be considered for scalar implicature computation (Katzir remains agnostic in that work about whether these alternative are operative in a grammatical or non-grammatical scalar implicature computation process) and those that are, for manner-based reasons, ruled out. What we need for the problem at hand is not a mechanism to rule out alternatives, but rather one that makes "*sp*" a particularly good alternative, relative to "*b or sp*", when the speaker believes "*sp*" is true. (At any rate, the Katzir system treats "*sp*" as a legitimate alternative to "*b or sp*".)

Instead, we need a numeric value that can be assigned to an utterance in context. From a profoundly naive perspective, the simplicity (or form-based goodness) of a speaker's utterance can be estimated as the prior probability of that form, given a field of possible forms (i.e., the sentences of the language). A very rough approximation of such a value can be derived from bigram data, so that, given some string of words  $s = w_1, \dots, w_n$ , adding

---

<sup>8</sup>The other intuitive reason for avoiding the longer sentence in *[osp]* is that there is a clausal implicature that is not compatible with this belief state. But that is what we're trying to explain here.

a word  $w_{n+1}$  reduces the simplicity by the probability of that word's following  $w_n$ .<sup>9</sup> We can estimate that probability with corpus data. For example, the word *the* appears 28.2 million times in the Corpus of Contemporary American English (COCA); the phrase *the man* appears 50,756. This frequency data gives us an estimate of the probability of *man* following *the*: by dividing, we get a frequency of 0.0021; which means *the* is 469 times simpler than *the man*. By this measure, a sentence with *some and not all* is 77.3 million times more complex than a corresponding sentence with *some*.<sup>10</sup> Spelling it out, we have:

(4.45) Definition: Let  $f(w)$  measure the probability of a word, with  $f(w|x)$  measuring the probability of a word  $w$ , given that the preceding word is  $x$ . Then the *simplicity* of a word  $w$  is  $s(w) := f(w)$ . The *simplicity* of a string of words  $w_1, \dots, w_n$ , where  $n > 1$ , is:

$$s(w_1, \dots, w_n) := s(w_1, \dots, w_{n-1})f(w_n|w_{n-1}).$$

Does this definition seem realistic? Is “Some of the students left” really 77 million times simpler than “Some and not all of the students left”? I remain agnostic about this question. Nonetheless, it is fairly clear that a working definition of complexity will provide orders of magnitude greater complexity for sentences for each additional word they contain. For purposes of illustration, I adopt the simplifying assumption that  $f(w|x) = 0.10$  for all  $w, x$ , and  $f(w) = 0.10$  for all  $w$  (this is highly unrealistic, but it makes the calculations simpler while showing the general effect of geometrically increasing complexity.)

Each utterance's probability, then, should be determined by not just its relevance and whether or not the speaker believes it, but also by its simplicity. This can be incorporated

<sup>9</sup>This measure of complexity is likely to be very remote from the right one; for the time being, we simply require a measure that significantly increases complexity with each word added.

<sup>10</sup>It is also worth mentioning that *some but not all* appears 54 times whereas *some and not all* appears just once in COCA. According to the analysis of *but* in section 3.3.1, the former cannot compete with *some*: its sum relevance must be in favor of *not all*, which must be opposed to the relevance of *some*. This provides support for the suggestion that speakers are more likely to say *some but not all* than they are to say *some and not all*, which indicates that, in conversational settings, *all* is expected to be more relevant than *some*.

in the Utterance Expectation formula as follows:

- (4.46) **Utterance Expectation (Quality, relevance, and manner based):** For any sentence  $\nu$ , conversational point  $H$ , set of alternatives  $ALT$ , and  $ALT$ -maximal belief  $\mathcal{B}$ :

$$P(\ulcorner \nu \urcorner | \mathcal{B}) = \frac{P(B\nu | \mathcal{B})r_H(\nu)s(\nu)}{\sum_{\chi \in ALT} P(B\chi | \mathcal{B})r_H(\chi)s(\nu)}.$$

This new Utterance Expectation formula yields the table in 4.8 for an utterance of “Kai ate the broccoli or some of the peas” ( $b$  or  $sp$ ). The odds in favor of  $[b, osp]$  vs  $[osp]$ ,

	$r_H(x)$	$s(x)$	$[osp]$	$[b, osp]$
“ $sp$ ”	0.02	$10^{-6}$	1	0
“ $ap$ ”	0.04	$10^{-6}$	0	0
“ $b$ ”	0.04	$10^{-4}$	0	0
“ $b$ and $sp$ ”	0.1	$10^{-9}$	0	0
“ $b$ and $ap$ ”	0.1	$10^{-9}$	0	0
“ $b$ or $sp$ ”	0.02	$10^{-9}$	1	1
“ $b$ or $ap$ ”	0.04	$10^{-9}$	0	0
$N()$			$2.002 \times 10^{-8}$	$2 \times 10^{-11}$

Figure 4.8:  $N$ -values with simplicity ( $s(x)$ ) incorporated.

taking simplicity into consideration, are 1001 : 1, which seems closer to intuitions than the previous prediction of 2 : 1.

This complexity measure raises a number of questions: does adding an extra word to an utterance really increase its complexity by an order of magnitude (and therefore cut its utility by a corresponding amount)? My intuition is that speakers do not seem excessively burdened by considerations of brevity: speakers are happy to talk and talk without adding a great deal to their message or pushing a lot closer to their point. Is it really so much worse to say “The sky was devoid of clouds” than “The sky was clear”? It seems, however, that there is little risk of pernicious implicature generation in such cases.



The system, though it assigns a much lower prior probability to the former utterance, will not generate implicatures because there is no knowledge state that supports an utterance of “The sky was devoid of clouds” that fails to support “The sky was clear”, and vice versa. So the extra effort associated with the longer utterance is inert for purposes of scalar reasoning: even if the latter is considered much more likely to be uttered than the former, relative to a given belief state, there is never a contrasting belief state, never a belief state that makes the former more likely to be true than the latter. Inversely, then, there is no expected implicature associated with the increased complexity. This example points to a key advantage of the Bayesian approach: depending on the circumstances, a difference in utterance expectation can lead to the generation of an implicature, as in the disjunction case, or not, as in the “devoid of clouds” case.

## 4.4 The symmetry problem

In contexts where a weak scalar assertion  $\omega$  is positively relevant, a stronger alternative  $\sigma$  is typically more positively relevant still, and the so-called “symmetric” alternative  $\omega \wedge \neg\sigma$  has low positive relevance or is negatively relevant. In such cases, the importance of the symmetric alternative in scalar reasoning is diminished or nullified, due both to its relevance score and, following considerations outlined in section 4.3, to its higher complexity (or low simplicity score). But there is nothing about the formal definition of relevance being used here, nor is there anything about intuitive relevance, that prevents a reversal of this typical arrangement, where  $r_H(\sigma) < r_H(\omega) < r_H(\omega \wedge \neg\sigma)$  (see section 3.5 for elaboration). In this section, I consider the predictions of the system for such cases.

Consider the following context.<sup>11</sup> An academic department is planning series of talks, with the intention of provoking controversy within the department, which has gotten a little boring. Because the department’s goal is to trigger passionate intellectual debate, it

---

<sup>11</sup>Thanks to Kai von Fintel for providing this example.

prefers speakers who have both strong supporters and strong detractors. In this context, a contribution of “Some of the faculty love Chomsky” ( $\omega$ ) is (plausibly) positively relevant to  $H = \text{that we should invite Chomsky}$ . Note also that “Some and not all of the faculty love Chomsky” ( $\omega \wedge \neg\sigma$ ) is still more relevant, as more debate is likely to ensue if there are non-Chomsky-lovers in the audience. And, by extension, “All of the faculty love Chomsky” is negatively relevant: it makes Chomsky a worse candidate.

Consistent with this context, assume we have the relevance scores in figure 4.9. Given

	$r_H(x)$	$s(x)$	$[o\omega]$	$[o\omega, \sigma]$	$[\sigma]$
$\lceil \omega \wedge \neg\sigma \rceil$	0.3	$10^{-9}$	1	0	0
$\lceil \omega \rceil$	0.1	$10^{-6}$	1	1	1
$\lceil \sigma \rceil$	0	$10^{-6}$	0	0	0
$N(\mathcal{B})$			$1.003 \times 10^{-7}$	$10^{-7}$	$10^{-7}$

Figure 4.9: Scalar reasoning including a “some and not all” alternative with high relevance

the tabulation, we see that there is no scalar effect predicted for an utterance of  $\lceil \omega \rceil$ : all three belief states have (nearly) identical sums, and the small disadvantage generated for  $[o\omega]$  is likely to be overshadowed by prior probabilities.

We have thus avoided deriving an implicature of  $[\sigma]$ , partially addressing the symmetry arguments. But intuitively, an utterance of  $\omega$  does still implicate that the speaker is not in belief state  $[\sigma]$ . That is, even under non-canonical relevance relations (where  $\omega \wedge \neg\sigma$  is more relevant than  $\sigma$ ) it seems an upper-bounding inference that the speaker does not believe  $\sigma$  goes through.

The system is capable of generating such inferences, but it does so in largely different manner than ordinary scalar inferences, and the analysis requires an assumption about how speakers choose the point  $H$  they will speak to. Specifically, given this assumption, the system predicts that a speaker in belief state  $[\sigma]$  will be expected to choose  $\neg H$  (*that we should not invite Chomsky*), and implicature computation will proceed from there. Recall that the Utterance Expectation formula in (4.46) assumes the hearer knows what point  $H$

the speaker is speaking to. This is an idealization meant to simplify the task of calculating implicatures. But in this case, we need to actually calculate  $P(pt = H)$ .

The following definition provides a first formal approximation of the (hearer's subjective) probability that the speaker will choose a given point  $H$  (in notation,  $P(pt = H)$ ) from among the propositions in the question under discussion.

(4.47) The probability that the speaker will choose, from among the partitions in the  $QUD$ , the point  $H \in QUD$ , given the speaker's  $ALT$ -maximal belief state  $\mathcal{B}$ , is defined as follows:

$$P(pt = H|\mathcal{B}) = \frac{\max_{\chi \in ALT}(r_H(\chi)P(B\chi|\mathcal{B})s(\chi))}{\sum_{H' \in QUD} \max_{\chi \in ALT}(r_{H'}(\chi)P(B\chi|\mathcal{B})s(\chi))}$$

This says that a speaker is more likely to choose a point for which there are relevant, simple alternatives.

Note that  $r_{\neg H}(\omega)$  and  $r_{\neg H}(\omega \wedge \neg\sigma)$  are negative, and suppose  $r_{\neg H}(\sigma)$  is, in line with intuitions, quite high (Chomsky won't inspire heated debate if everybody loves him), say 0.5. Then, using the formula in (4.47), we have:

$$P(pt = H|[\sigma]) = \frac{0.1 \times 1 \times 10^{-6}}{(0.1 \times 1 \times 10^{-6}) + (0.5 \times 1 \times 10^{-6})} \quad (4.48)$$

$$= \frac{1}{6}, \quad (4.49)$$

and

$$P(pt = H|[o\omega]) = \frac{0.3 \times 1 \times 10^{-6}}{(0.3 \times 1 \times 10^{-6}) + 0} \quad (4.50)$$

$$= 1. \quad (4.51)$$

What this says is that the probability of the speaker speaking to  $H$  (= *that we should invite*

*Chomsky*) in belief state  $[o\omega]$  is 1, whereas the probability of the speaker speaking to  $H$  in belief state  $[\sigma]$  is  $\frac{1}{6}$ .

Given these numbers, plus the fact that an only an insignificant scalar effect was found in figure 4.9 (which justifies the “approximate” cancellation of terms in equation (4.53), we can use Bayes’ Rule to compute:

$$\frac{P([\sigma]|\ulcorner\omega\urcorner \wedge pt = H)}{P([o\omega]|\ulcorner\omega\urcorner \wedge pt = H)} = \frac{P(pt = H|[\sigma])}{P(pt = H|[o\omega])} \frac{P(\ulcorner\omega\urcorner|[\sigma] \wedge pt = H)}{P(\ulcorner\omega\urcorner|[o\omega] \wedge pt = H)} \frac{P([\sigma])}{P([o\omega])} \quad (4.52)$$

$$\approx \frac{P(pt = H|[\sigma])}{P(pt = H|[o\omega])} \frac{P([\sigma])}{P([o\omega])} \quad (4.53)$$

$$= \frac{1}{6} \frac{P([\sigma])}{P([o\omega])} \quad (4.54)$$

This says that, in this context, an utterance of  $\ulcorner\omega\urcorner$  favors  $[o\omega]$  over  $[\sigma]$ , with an odds ratio of 6 : 1.

Let’s step back to put together the elements of this partial solution to the symmetry problem. First, the context entails that  $\omega \wedge \neg\sigma$  is more relevant than  $\omega$ , and  $\sigma$  is negatively relevant. The fact that  $\omega \wedge \neg\sigma$  is significantly more complex than  $\omega$  neuters its scalar influence, so there is no scalar effect away from  $[o\omega]$ , even though  $\omega \wedge \neg\sigma$  is more relevant than *some*. At the same time, the fact that  $\sigma$  is positively relevant to  $\neg H$  suggests that the speaker would be more likely to speak to  $\neg H$  (rather than  $H$ ) if the speaker were in belief state  $[\sigma]$ . That fact leads to a Bayesian inference that the speaker is not in belief state  $[\sigma]$ .

This story does not handle all possible scenarios. Cases where all three alternatives,  $\omega \wedge \neg\sigma$ ,  $\omega$ , and  $\sigma$  are positively relevant, but  $r_H(\omega \wedge \neg\sigma) > r_H(\omega)$ , are still a puzzle for the system.<sup>12</sup> With respect to a case like this, the system predicts no scalar implicature, contrary to intuitions. I leave it as an open question how/whether an upper-bounding implicature is derived for such cases, but suggest that the alternative utterance involving “not all” may play a role (to the extent that  $r_{\neg H}(\neg\sigma)$  is high).<sup>13</sup>

<sup>12</sup>For what it’s worth, these conditions seem quite rare in conversation; it is very difficult, if not impossible, to devise a plausible context where those relevance relations hold.

<sup>13</sup>Matsumoto (1995) argues that, due to the implicatures of sentences like “It was warm yesterday, and it

## 4.5 Presuppositional scalar implicature

Arguments for grammatical computation of scalar implicature often point out that, when scalar terms are embedded below factive verbs, the presupposition also seems to be upper-bounded:

(4.55) George knows that some of his employees are embezzling.

In section 1.3.1, I argued that the grammatical theory is actually too inflexible to handle the observed implicatures associated with factives: the implicature of the presupposition can be defeated while the implicature of the assertion remains. In this section, I examine the ways a Bayesian implicature computation mechanism can be applied to the domain of presupposition strength.

The standard view of presupposition, put forth by Strawson (1950), and originally due to Frege, is that a sentence's presuppositions must be true for the sentence's truth value to be defined: *The King of France is bald* is true if there is a unique salient King of France and he is bald, false if there is a unique salient King of France and he is not bald, and receives no truth value in a world where there is no King of France, or when there are multiple salient Kings of France.

To account for the anomaly of sentences like "A father of the victim arrived at the scene," Heim (1991, as reported and translated in Sauerland 2008) proposes a new pragmatic principle, which she ties to Grice's maxim of Quantity, called *Maximize Presupposition*:

---

is a little bit more than warm today," simplicity alone cannot prevent the derivation of scalar implicatures. I cannot reply to Matsumoto's argument here, except to say that (1) on the present account, it is an interaction of simplicity and relevance that prevents derivation of symmetric scalar implicatures, and (2) the nature of simplicity scores has not been explored: it is possible that the presence of *a little bit more than warm* in a parallel construction makes its *s* score closer to that of *warm*. For a related implementation that makes *a little bit more than warm* equally complex to *warm*, see Katzir (2007). Also see Block (2008) and Lassiter (2010) for two arguments that Matsumoto's example is not a real problem.

(4.56) Make your contribution presuppose as much as possible!

On its face, the principle is plausible enough: if a speaker faces a choice between an utterance that presupposes a bit of information and one that does not presuppose it, the speaker must choose the utterance that presupposes the information if it is, indeed, taken for granted.<sup>14</sup> Sauerland has used this principle as the foundation of an analysis of phenomena that he calls implicated presuppositions (also called anti-presuppositions; cf. Chemla 2008, Schlenker 2006, Percus 2006, Singh 2011, and Spector 2007a). A speaker who fails to utter a sentence that makes a particular presupposition indicates he does not believe that presupposition may be taken for granted. So, for example, if I utter “John believes Mary is cheating” instead of “John knows Mary is cheating,” I indicate that it is not taken for granted that Mary is cheating. Sauerland’s logic may be spelled out as follows: A speaker has uttered a sentence that does not make a particular presupposition. There is a comparable alternative that does make that presupposition. Because the speaker failed to make that presupposition, he implies that he does not take the presupposition for granted.

It is intuitive to extend this logic beyond the pairs of present-or-absent presuppositions that Sauerland considers to include pairs that make a relatively strong or weak presupposition. For example, (4.55), which presupposes that some of George’s employees are embezzling, has an alternative that makes the stronger presupposition that all of George’s employees are embezzling:

(4.57) George knows that all of his employees are embezzling.

---

<sup>14</sup>For canonical presupposition cases—those where speaker and hearer both take a presupposition for granted, like *that I have two hands* in “Both my hands are dirty”—there is no clear Gricean motivation for the principle, since the proposition is totally uninformative. Gricean motivation for *Maximize Presupposition* can be found, however, in cases of accommodation (Stalnaker, 1978; Lewis, 1979; Thomason, 1990), where presupposition is used to add (non-controversial) information to the discourse. See Geurts (2011) for discussion of scalar reasoning about accommodated presuppositions. Also see the discussion from Gazdar (1979), reproduced surrounding example (1.14) above, which lumps reasoning about entailed and presupposed content together.

A speaker who maximizes presupposition would be obligated to utter the sentence in (4.57) whenever it is taken for granted that all of George's employees are embezzling. Therefore, an utterance of (4.55) should imply that it is not taken for granted that all of the employees are embezzling; I will refer to this inference as a *presuppositional implicature*.

One key caveat for this story is that a speaker might be unable to maximize presupposition without violating other pragmatic principles. For example, maximizing presupposition may require the speaker to utter something for which he lacks evidence, violating the maxim of Quality: the speaker may believe that George's belief is that three of his five advisors are embezzling, but it is taken for granted that all five are embezzling. In this case, the speaker can't utter the sentence that makes the maximal presupposition, (4.57), because the speaker doesn't believe the sentence is true.<sup>15</sup>

It is not difficult to incorporate reasoning about presuppositions into the Bayesian system. I assume a basic, standard view of presupposition, whereby there is some common ground (CG) that contains those propositions mutually taken for granted by speaker and hearer. The key fact about the CG for presuppositional implicature is that it is an idealization: neither speaker nor hearer actually know for sure exactly which set of propositions are mutually taken for granted. In a probabilistic system, then, a hearer assigns each potential common ground a prior probability; this probability, of course, can be affected by the utterance a speaker chooses. To see the framework of the problem, consider a simple example with two alternatives:<sup>16</sup>

- (4.58)    a.    John believes Mary is pregnant. (= *Jbp*)  
               b.    John knows Mary is pregnant. (= *Jkp*)

<sup>15</sup>This is a key difference between implicated presuppositions and presuppositional implicatures. In implicated presuppositions, the utterance and its alternative have the same assertoric content, and reasoning is simply about whether a presupposition is present or absent. In presuppositional implicature, alternatives have stronger or weaker asserted content *and* stronger or weaker presupposed content.

<sup>16</sup>I gloss over the well-known fact that knowledge is not simply true belief, or belief plus a presupposition of truth (knowledge also requires justification, plus something else, as demonstrated by Russell (1948) and later by Gettier (1963)).

There are three types of common grounds of interest here: those that entail Mary is pregnant, those that entail she is not pregnant, and those where neither proposition is taken for granted. We can use bracket notation to represent these three types of common grounds as follows:  $[_{CG}p]$ ,  $[_{CG}\neg p]$ ,  $[_{CG}p, \neg p]$ . For the sake of simplicity, suppose the hearer's prior belief is that all three CGs are equally likely. Now, to incorporate presupposition into the theory of utterance expectation, we need only to add a term that says, assuming that the CG is  $\mathcal{C}$ , a speaker can only utter a sentence with presupposition  $p$  if the  $\mathcal{C}$  entails  $p$ . If we let  $\phi(\chi) = \{p : p \text{ is a presupposition of } \chi\}$ , and let  $\mathcal{C}()$  be a function that takes a proposition and returns a truth value for that proposition in  $\mathcal{C}$ , then this relevant term for utterance expectation will simply be the product:

$$\Phi_{\mathcal{C}}(\chi) = \prod_{p \in \phi(\chi)} \mathcal{C}(p) \quad (4.59)$$

This value is simply 1 if all the presuppositions of  $\chi$  are met in common ground  $\mathcal{C}$ , and 0 otherwise. This value can be incorporated into Utterance Expectation as follows:

(4.60) **Utterance Expectation (incorporating presupposition):** For any sentence  $\nu$ , conversational point  $H$ , set of alternatives  $ALT$ ,  $ALT$ -maximal belief  $\mathcal{B}$ , and common ground  $\mathcal{C}$ :

$$P(\ulcorner \nu \urcorner | \mathcal{B}, \mathcal{C}) = \frac{P(B\nu | \mathcal{B})r_H(\nu)s(\nu)\Phi_{\mathcal{C}}(\nu)}{\sum_{\chi \in ALT} P(B\chi | \mathcal{B})r_H(\chi)s(\chi)\Phi_{\mathcal{C}}(\chi)}.$$

Now, since the alternatives  $Jbp$  and  $Jkp$  are equally complex by the measure being used here (since they have the same number of words), the simplicity term  $s(x)$  will drop out of the calculation. Likewise, I assume the two sentences are equally relevant, since presupposition is essentially a non-at-issue aspect of meaning. Given these assumptions, we



can calculate implicated presuppositions as illustrated in Figure 4.10. Here, we see that

	$[_{CG}p]$	$[_{CG}\neg p]$	$[_{CG}p, \neg p]$
“ $Jbp$ ”	1	1	1
“ $Jkp$ ”	1	0	0
$N()$	2	1	1

Figure 4.10: Presuppositional scalar reasoning favors common grounds  $[_{CG}\neg p]$  and  $[_{CG}p, \neg p]$  over  $[_{CG}p]$  for an utterance of  $Jbp$ .

the utterance of  $Jbp$  disfavors the common ground  $[_{CG}p]$ , since that common ground also supports the utterance  $Jkp$ , which makes a presupposition.

Note that the model here supports the generation of implicated presuppositions without assuming any additional principle of usage like *Maximize Presupposition*. The result depends only on the incompatibility of sentences that make presuppositions with common grounds that do not contain those presuppositions.

Similar reasoning may be brought to bear on presuppositional implicature examples. Consider the sentence (4.57) and its three alternatives:

- (4.61)    a.    George believes that some of his employees are embezzling. (=  $Gb\omega$ )  
              b.    George believes that all of his employees are embezzling. (=  $Gb\sigma$ )  
              c.    George knows that some of his employees are embezzling. (=  $Gk\omega$ )  
              d.    George knows that all of his employees are embezzling. (=  $Gk\sigma$ )

There are seven relevant possibilities for the common ground in this example, ranging from very weak, making no entailments about whether or how many employees are stealing (using bracket notation, this is  $[_{CG}\neg\omega, \omega, \sigma]$ ); to very strong, entailing that all employees are in fact stealing (in notation,  $[_{CG}\sigma]$ ). For simplicity, I consider only the “strong” common ground possibilities:  $[_{CG}\neg\omega]$ ,  $[_{CG}\omega]$ ,  $[_{CG}\sigma]$ . George’s possible beliefs, meanwhile, include  $[_{G}\omega]$ ,  $[_{G}\omega, \sigma]$ , and  $[_{G}\sigma]$ . (There are, in addition, the speaker’s beliefs to contend with, but I make the simplifying assumption that the speaker is opinionated,

so that  $[_{G\omega}, \sigma]$  is an abbreviation for  $[_s[_{G\omega}, \sigma]]$ .) A hearer can then use Bayesian reasoning to differentiate between the likelihoods for these common grounds and beliefs, as illustrated in Figure 4.11 (again, I omit simplicity scores, since they are constant among alternatives, but assume the *all* alternatives are significantly more relevant than the *some* alternatives). The table in 4.11 contains a great deal of information, but the key thing to

		CG $\mathcal{B}_G$	$[\neg\omega]$			$[_{o\omega}]$			$[_{\sigma}]$		
			$[_{o\omega}]$	$[_{\sigma}]$	$[_{o\omega}, \sigma]$	$[_{o\omega}]$	$[_{\sigma}]$	$[_{o\omega}, \sigma]$	$[_{o\omega}]$	$[_{\sigma}]$	$[_{o\omega}, \sigma]$
“ $Gb\omega$ ”	0.1		1	1	1	1	1	1	1	1	1
“ $Gb\sigma$ ”	0.5		0	1	0	0	1	0	0	1	0
“ $Gk\omega$ ”	0.1		0	0	0	1	1	1	1	1	1
“ $Gk\sigma$ ”	0.5		0	0	0	0	0	0	0	1	0
$N()$			0.1	0.6	0.1	0.2	0.7	0.2	0.2	1.2	0.2

Figure 4.11: Presuppositional scalar reasoning for strong/weak  $\times$  present/absent presupposition.

recognize is that the utterance of  $\lceil Gk\omega \rceil$  strongly favors a common ground of  $[_{CG\omega}]$  over a common ground of  $[_{CG\sigma}]$  when George’s belief is  $[_{G\sigma}]$ . In contrast, when George’s belief is  $[_{G\omega}]$ , scalar reasoning does not distinguish between the common grounds  $[_{CG\omega}]$  and  $[_{CG\sigma}]$ ; only prior probabilities distinguish those.

This property of the table in Figure 4.11 accords well with my intuitions about presuppositional scalar implicature: we don’t always conclude from an utterance of  $Gk\omega$  that it is not taken for granted that  $\sigma$ . Only some of the possible speaker beliefs license this implicature. On average, however, scalar reasoning does favor the common ground  $[_{CG\omega}]$  following an utterance of  $Gk\omega$  over the common ground  $[_{CG\sigma}]$ , and so the Bayesian system has at least achieved a first-pass presuppositional implicature.

## 4.6 Bayesian vs. game-theoretic pragmatics

In the past decade, a line of research has developed that attempts to account for pragmatic phenomena generally, and scalar implicatures in particular, using formal tools from the

mathematics of game theory. Game theory seeks to model social interaction by treating each participant as a player in a game, with a set of available moves and associated pay-offs, generally conditional upon the state of the world and the moves of the other players. Proponents of game theoretic pragmatics emphasize that games, in and of themselves, are not theories of pragmatic phenomena; rather, the theory is specified by the model of players' strategies relative to a game. Game theoretic pragmatics posits that game theory, because it provides the tools to construct of a formal model of rational, interactive agent behavior, is well-suited to the analysis of conversation (a rational, interactive two-agent activity).

The game theoretic approach provides for richer models of conversational behavior than the Bayesian framework put forth in this dissertation. Whereas the Bayesian approach models the conversational inference task as an essentially one-agent process, with hearers using their knowledge of the conversational setting and speaker beliefs to draw inferences, game theoretic approaches model inference as a fundamentally two-participant process. Speakers and hearers coordinate their actions so that speakers rationally choose utterances and hearers rationally choose interpretations depending on a given state of the world. The parameters around which speakers and hearers coordinate vary from theorist to theorist: I focus here on the theory put forth in Franke (2009) (see also Franke 2011), which is similar to work by Jäger (2007), but many other works build game theoretic models in various other ways (Parikh, 1991; Ross, 2006; Benz, 2007; Benz and Van Rooij, 2007). This heterogeneity of theories is a conceptual advantage of the game theoretic approach: while they theories basically agree that it is rational for speakers and hearers to coordinate their actions to maximize their utility, each theory makes different claims about what constitutes an action's utility, and how speakers and hearers reason about each other, each with its own set of empirical predictions.

A game theoretic pragmatic model has two main parts: the game model itself, and the solution concept, or the strategy of the players. Franke puts forth a basic model,

formally a *signalling game* where the speaker selects a message and the hearer selects an interpretation, and payoff is high for speaker and hearer when the hearer correctly selects, as the interpretation of the speaker's utterance, the real world. Roughly speaking, speakers base their moves on the information they have about the world, and hearers base their moves on the information they receive from a speaker's utterance. With respect to the scalar implicature problem, a conversational game awards high points to speaker and hearer when their moves result in a hearer interpretation of  $o\omega$  when that proposition is true, and an interpretation of  $\sigma$  when that is true. There are various solution concepts that could be employed by the two players to converge on this pairing, many of which are ostensibly not an accurate model of human behavior. For example, the speaker can adopt a strategy of always uttering "All" when he knows  $o\omega$  is true, and always uttering "Some" when he knows  $\sigma$  is true. If the hearer, in turn, always interprets "All" as  $o\omega$  and interprets "Some" as  $\sigma$ , such a speaker has settled on an optimal solution concept. But, intuitively and empirically, this is not what speakers and hearers do.

The solution concept Franke puts forth is called *Iterated Best Response* (IBR), with speakers and hearers reasoning about the best response to an interlocutor who behaves in a certain limited rational manner. The iterations of the game can be understood as "she thinks I think"-type higher order beliefs, with a Level- $n$  player basing his moves on those of a Level- $n - 1$  opponent. Without spelling out the technical details of the theory, basic scalar implicature does not require many iterations:<sup>17</sup>

Level 0: Speaker utters "Some" when  $o\omega$  is true and is indifferent between "All" and "Some" when  $\sigma$  is true.

Level 1: Hearer interprets "All" as  $\sigma$  and interprets "Some" as  $o\omega$  (provided the prior probability of  $\sigma$  is not too high).

Level 2: Speaker utters "Some" when  $o\omega$  is true and utters "All" when  $\sigma$  is true.

---

<sup>17</sup>This is what Franke calls the  $S_0$  sequence, assuming flat priors. There is a corresponding  $H_0$  sequence that begins with a literal-interpreter hearer.

Level 3: Etc.

In Level 0, speakers are “literal”: “All” is false when  $\omega$  is true, so the speaker has to say “Some”, but both candidates are compatible with  $\sigma$ , which allows the speaker to choose freely between them. By Level 1, however, a hearer has commenced scalar implicature type behavior. This is because “Some” is produced all the time when  $\omega$  is true, but only half the time when  $\sigma$  is true. This means  $\omega$  has twice the expected utility for a Level-1 hearer when the speaker utters “Some.” The Level-2 speaker, then, has a straightforward best strategy: in a given state of the world, choose the utterance that the hearer will map to that state. It should be clear that higher-level speakers and hearers will not change behavior once they settle on such an optimal mapping.

The game theoretic approach clearly has much in common with the Bayesian approach: both depend crucially on a formalization of the hearer’s expectations about the speaker’s behavior. The game theoretic approach takes the further step, however, of making the speaker’s behavior depend on the hearer’s expectations. The IBR model can be understood as a way of starting with probabilistic expectations and, eventually, settling on deterministic expectations (so that  $P([\sigma]|\ulcorner\omega\urcorner) \in \{0, 1\}$ ). A Level-0 speaker simply distributes probabilities equally over each compatible alternative utterance, given a belief, and a Level-1 hearer reasons probabilistically about this behavior to assign each possible utterance an optimal belief (again, distributing probabilities equally over beliefs that are equally optimal, relative to a Level-0 speaker).

With respect to the present Bayesian theory, which is purely hearer-driven (a hearer uses expectations about speaker behavior to reason about speaker beliefs) iterated reasoning is analogous to computation of higher order inferences. For example, in the Bayesian theory, a hearer’s expectations could lead to an inference that speaker who utters “Some of the students left” is 3 times as likely to believe *that some and not all left* than *that all left*. It is natural to imagine that a rational speaker would anticipate the hearer’s undertaking this reasoning process and, in turn, adjust her choice of utterance. For example, the

speaker who is tempted to utter “Some of the students left” when she believes *that all of the students left*, would naturally reason: the hearer is going to think I believe *that some and not all of the students left* if I utter “Some of the students left”, so I’ll utter “All of the students left” instead. The rational hearer, in turn, could conclude that the speaker has essentially no chance of uttering “Some of the students left” when she believes *that all of the students left*. This natural back-and-forth reasoning process closely mirrors the IBR model: iterated reasoning about an interlocutor tends to strengthen an inference.

This points to a key difference between the models. The IBR system always dials down an interpretation to one of the partitions: it does not tell the hearer how likely one partition is versus another, but instead provides a black and white pairing of utterance with partition.<sup>18</sup> For example, the IBR theory might predict that the utterance of “Some of the students passed” is paired with the speaker belief  $[o\omega]$  while an utterance of “All of the students passed” is paired with the belief  $[\sigma]$ ; the Bayesian approach, on the other hand, might predict that an utterance of “Some of the students passed”, in a given context, implicates  $[o\omega]$  over  $[\sigma]$  by a factor of 4 : 1. In some ways, categorical implicature generation seems desirable: implicature has certainly traditionally been treated as a categorical phenomenon, with implicatures either present or absent, not a question of degree. On the other hand, the probabilistic nature of inferences in the Bayesian theory has a number of benefits. It provides a straightforward mechanism for implicature reversal: no “backing up” to undo an implicature computation mechanism is necessary (see section 4.7 for discussion of implicature cancellation/reversal). In the Bayesian system, discovering that the speaker actually believes  $\sigma$  simply leads the hearer to narrow his belief distribution down to the twenty per cent of worlds where the speaker utters “Some” and believes  $\sigma$ .<sup>19</sup> Finally,

<sup>18</sup>The general game-theoretic approach, of course, does not require hearers to settle on a black-and-white interpretation: hearer moves can, of course, be assignments of probabilities to various possible hearer beliefs. If such moves are iterated (i.e., higher-order speakers then make their moves according to the probabilities the hearer will assign), it seems such theories tend to converge on the same results as Franke’s deterministic game theoretic model. To my knowledge, the game theoretic approaches that have been advanced in the literature so far have not derived probabilistic implicatures. See equations 4.65 through 4.67 below for a rudimentary game-theoretic strategy that produces probabilistic implicatures.

<sup>19</sup>Some recent research, including Magri (2009) and Singh (2010), has suggested that, contrary to the

there are empirical facts that seem better addressed by a theory that generates probabilistic scalar inferences; see, in particular, arguments below in Section 5.2 surrounding so-called “embedded implicature” phenomena.

Another difference between Franke’s system and the Bayesian system proposed here is the status of the symmetry problem. In Franke’s system, the symmetric alternative “some and not all” can’t help but beat out “some” in  $[o\omega]$ , so the game will inevitably map the former utterance to  $[o\omega]$  while it maps the latter to  $[o\omega, \sigma]$ . Franke’s solution is essentially to limit the alternatives in his game to Horn-scalar alternatives, opting to provide a theory of reasoning about alternatives, not a theory of which alternatives are available for reasoning. The present theory must also limit the set of alternatives entertained for scalar reasoning, but it need not exclude symmetric alternatives (see section 4.4).

Finally, the two approaches diverge with respect to their treatment of the various epistemic implicatures. To handle epistemic nuance, Franke *lifts* the signalling game to make it a task of pairing utterances with belief states of the hearer, rather than simply states of the world. The three speaker epistemic states under consideration for the scalar case, as in the Bayesian theory, are  $[o\omega]$ ,  $[\sigma]$ , and  $[o\omega, \sigma]$ . A Level-0 speaker maps each state to any compatible utterance from the alternative set  $\{\omega, \sigma\}$  (recall that the symmetric alternative must be excluded) as follows:

$$[o\omega] \mapsto \ulcorner \omega \urcorner \quad (4.62)$$

$$[\sigma] \mapsto \ulcorner \omega \urcorner, \ulcorner \sigma \urcorner \quad (4.63)$$

$$[o\omega, \sigma] \mapsto \ulcorner \omega \urcorner \quad (4.64)$$

In response, the Level-1 hearer maps an utterance of  $\sigma$  to  $[\sigma]$ : this is the only belief state that supports the utterance. But where should the hearer map  $\ulcorner \omega \urcorner$ , which is compatible

---

tradition of cancellability, scalar implicatures are mandatory; this view, supported by a small but relatively compelling body of evidence, is apparently compatible with a game-theoretic approach, but not the Bayesian one.

with all three belief states? As in the Bayesian approach, the answer to this question depends entirely on the relative prior probabilities of  $[o\omega]$  and  $[o\omega, \sigma]$ . In the IBR approach, the Level-1 hearer maps  $\ulcorner \omega \urcorner$  to whichever belief state has a larger prior probability.<sup>20</sup>

Franke argues that this reflects the opinionatedness-based explanation of epistemic strengthening argued for in many global accounts, including Sauerland (2004), van Rooij and Schulz (2004), Russell (2006), and Spector (2007b). In the IBR model, Franke identifies the assumption of speaker opinionatedness with with an assumption that narrow beliefs (in this case, those with just one partition) have higher prior probabilities than broader beliefs (with two partitions). Because the resolution of scalar implicatures in cases like the one above depends on such inequalities, it seems the IBR model, like the Bayesian theory, reflects Gricean intuitions about epistemic strong scalar implicatures.

But the two theories actually make deeply divergent predictions about epistemic scalar implicatures. Under the IBR model, if priors specify  $P([o\omega]) = P([o\omega, \sigma])$ , an utterance of “Some” is mapped without preference to the two knowledge states  $[o\omega]$  and  $[o\omega, \sigma]$ ; i.e.,  $P([o\omega]|\ulcorner \omega \urcorner) = \frac{1}{2}$  and  $P([o\omega, \sigma]|\ulcorner \omega \urcorner) = \frac{1}{2}$ . If, however,  $P([o\omega]) > P([o\omega, \sigma])$ , even if it is only very slightly greater (e.g.,  $P([o\omega]) = 0.30000001$  and  $P([o\omega, \sigma]) = 0.3$ ), the model maps “Some” to  $[o\omega]$  deterministically; i.e.,  $P([o\omega]|\ulcorner \omega \urcorner) = 1$  and  $P([o\omega, \sigma]|\ulcorner \omega \urcorner) = 0$ . In contrast, the Bayesian model favors the epistemic strong over the epistemic weak interpretation to the degree that the prior probabilities favor the strong interpretation.<sup>21</sup>

This comparison between the approaches is not meant to be an argument against game theory as a framework for pursuing Gricean pragmatics, but rather to point out key differences between a particular game theoretic approach (the one that is most closely related to the approach taken in this thesis) and the Bayesian approach that is being developed in this thesis. Indeed, a number of potential hybrid approaches are conceivable.

<sup>20</sup>I am glossing over a detail here, which is that, if the hearer assigns a sufficiently large prior probability to  $[\sigma]$ , the hearer is predicted to map  $\ulcorner \omega \urcorner$  to  $[\sigma]$ .

<sup>21</sup>Another way of looking at this is that continuously shifting priors lead to jumps in posteriors for the IBR model, whereas continuously shifting priors lead to continuously shifting posteriors in the Bayesian.



Here is one hybrid model. First, compare the initial (Level 1) step of the hearer’s iteration in Franke’s system to the Quality-only (relevance-free and simplicity-free) theory of utterance expectation presented in (4.2). In Franke’s system, the Level 1 hearer assumes the speaker has chosen his utterance randomly from those alternatives that are supported by his beliefs and selects, through Bayesian reasoning, the belief state that is most likely to have produced that utterance. In the quality-only Bayesian system, the hearer uses the same information not to select a “best” belief state, but to assign each belief state a posterior probability. Thus, rather than settling on a single interpretation for a message, a hearer can, at each iteration, assign a probability to each possible belief.

To make Franke’s system generate probabilistic inferences, we simply take a Level- $n$  hearer’s posterior probabilities at face value, so that a Level- $n + 1$  speaker allocates his utterances probabilistically based on the probability that a Level- $n$  hearer will assign them to a given belief state. In formulas, this is:

$$S_0(\ulcorner x \urcorner | \mathcal{B}) := \frac{P(Bx | \mathcal{B})}{\sum_{y \in ALT} P(By | \mathcal{B})} \quad (4.65)$$

$$H_n(\mathcal{B} | \ulcorner x \urcorner) := \frac{S_{n-1}(\ulcorner x \urcorner | \mathcal{B}) P(\mathcal{B})}{\sum_{\mathcal{B}' \in \mathcal{B}_{ALT}} S_{n-1}(\ulcorner x \urcorner | \mathcal{B}') P(\mathcal{B}')} \quad (4.66)$$

$$S_n(\ulcorner x \urcorner | \mathcal{B}) := \frac{H_{n-1}(\mathcal{B} | \ulcorner x \urcorner)}{\sum_{y \in ALT} H_{n-1}(\mathcal{B} | \ulcorner y \urcorner)} \quad (4.67)$$

A little explanation is in order: here,  $S_n$  and  $H_n$  are Level- $n$  speaker and hearer probability distributions, and  $\mathcal{B}_{ALT}$  is an *ALT*-maximal partition of the belief space. There is some question here about how to interpret the prior probability function  $P$ : if it is understood as the hearer’s prior probability distribution, we must say that the (idealized) speaker has access to the hearer’s beliefs. Moreover, we have to say that the idealized speaker is expected to behave according to the hearer’s prior probabilities about the speaker’s beliefs. This makes some sense if we think of the entire game as a hearer-conducted exercise: “given my idealization of the speaker as someone who knows what I believe, and as some-

one who has the beliefs I think he has, he is likely to behave in a certain way. Depending on how he actually behaves, it is rational for me to adjust my posterior beliefs in a corresponding way. Once I do that, the speaker would be more rational to behave in (another) certain way. Etc.”<sup>22</sup>

The adaptation of the IBR model to allow probabilistic reasoning retains many of the properties of Franke’s system, but makes different predictions about the epistemic results. I can not prove it here, but the sequences  $H_{2n+1}$  and  $S_{2n}$  seem to converge, at least for simple examples. To get a sense of the output of the model, consider the following tables for  $ALT = \{\omega, \sigma\}$  and  $\mathcal{B}_{ALT} = \{[o\omega], [\sigma], [o\omega, \sigma]\}$ , with uniform priors of  $1/3$  over  $\mathcal{B}_{ALT}$ . The tables in Figures 4.12 and 4.13 show identical results to Franke’s system: speakers

$i$	$x$	$[o\omega]$	$[\sigma]$	$[o\omega, \sigma]$
0	$\omega$	1	0.5	1
	$\sigma$	0	0.5	0
2	$\omega$	1	0.17	1
	$\sigma$	0	0.83	0
4	$\omega$	1	0.07	1
	$\sigma$	0	0.93	0
$\vdots$		$\vdots$		
*	$\omega$	1	0	1
	$\sigma$	0	1	0

$i$	$x$	$[o\omega]$	$[\sigma]$	$[o\omega, \sigma]$
1	$\omega$	0.4	0.2	0.4
	$\sigma$	0	1	0
3	$\omega$	0.46	0.08	0.46
	$\sigma$	0	1	0
5	$\omega$	0.48	0.03	0.48
	$\sigma$	0	0.93	0
$\vdots$		$\vdots$		
*	$\omega$	0.5	0	0.5
	$\sigma$	0	1	0

Figure 4.12: Values for  $S_i(\ulcorner x \urcorner | \mathcal{B})$  with flat priors. Figure 4.13: Values for  $H_i(\mathcal{B} | \ulcorner x \urcorner)$  with flat priors.

are driven away from uttering  $\omega$  when they believe  $[\sigma]$ , so hearers conclude  $[\sigma]$  is not the speaker’s belief when he utters  $\omega$ . Moreover, with a uniform prior, since speakers must utter  $\omega$  in equally likely states  $[o\omega]$  and  $[o\omega, \sigma]$ , the hearer must assign each of these beliefs equal probability for an utterance of  $\omega$ .

The systems, however, make distinct predictions when priors are skewed. Recall that a prior that favored  $[o\omega]$  over  $[o\omega, \sigma]$  was predicted to generate an epistemic strong implica-

<sup>22</sup>Note that I omit  $H_0$ , assuming that the sequence starts with a literal speaker. This means that  $H_n$  is only defined for odd  $n$ , and  $S_n$  for even  $n$ . It is straightforward to define the  $H_0$  sequence to fill in these gaps.

ture; conversely, if  $[o\omega, \sigma]$  received a higher prior, an ignorance implicature was predicted. This had the potentially undesirable consequence of leading to implicature “flipping” behavior: slight shifts in priors led to big flips in implicature. The probabilistic IBR system does not lead to this effect: suppose, for example,  $[o\omega]$ ’s prior probability is twice that of  $[o\omega, \sigma]$ ; specifically, that  $P([o\omega]) = \frac{2}{5}$ ;  $P([\sigma]) = \frac{2}{5}$ ;  $P([o\omega, \sigma]) = \frac{1}{5}$ . The resulting tables

$i$	$x$	$[o\omega]$	$[\sigma]$	$[o\omega, \sigma]$
0	$\omega$	1	0.5	1
	$\sigma$	0	0.5	0
2	$\omega$	1	0.2	1
	$\sigma$	0	0.8	0
4	$\omega$	1	0.11	1
	$\sigma$	0	0.89	0
$\vdots$			$\vdots$	
*	$\omega$	1	0	1
	$\sigma$	0	1	0

Figure 4.14: Values for  $S_i(\ulcorner x \urcorner | \mathcal{B})$  with skewed priors.

$i$	$x$	$[o\omega]$	$[\sigma]$	$[o\omega, \sigma]$
1	$\omega$	0.5	0.25	0.25
	$\sigma$	0	1	0
3	$\omega$	0.59	0.12	0.29
	$\sigma$	0	1	0
5	$\omega$	0.62	0.07	0.31
	$\sigma$	0	0.93	0
$\vdots$			$\vdots$	
*	$\omega$	$\frac{2}{3}$	0	$\frac{1}{3}$
	$\sigma$	0	1	0

Figure 4.15: Values for  $H_i(\mathcal{B} | \ulcorner x \urcorner)$  with skewed priors.

in Figures 4.14 and 4.15 reveal that the basic scalar implicature is unchanged: the hearer can conclude that the speaker has no chance of believing  $[\sigma]$  if he says  $\ulcorner \omega \urcorner$ . But the epistemic status is different: probabilities are distributed proportionally to  $[o\omega]$  and  $[o\omega, \sigma]$ .<sup>23</sup> In other words, the system predicts that the posterior odds for these two belief states is equal to their prior odds, as we found in section 4.2 above.

In broad strokes, the difference between a game theoretic approach and the Bayesian approach presented in this chapter is the way of conceptualizing the hearer’s task. In the Bayesian Utterance Expectation approach, the hearer employs a simple intuitive theory of speaker behavior—essentially, that speakers prefer utterances that are simpler and more relevant—and uses it to draw probabilistic inferences about a speaker’s beliefs. In the game theoretic approach put forth by Franke (2009), the hearer’s task is to coordinate

<sup>23</sup>An interesting fact, left to be explored, is that when  $P([\sigma])$  receives a sufficiently high prior (specifically, when  $P([\sigma]) \geq 0.5$ ), a hearer no longer gets  $H_*([\sigma] | \ulcorner \omega \urcorner) = 0$ ; instead, this probability begins to gradually increase with  $P([\sigma])$ .

with the speaker to pair an utterance with an interpretation. The theories therefore make qualitatively different predictions: one provides graded predictions, whereas the other provides deterministic predictions. Because of this, the Bayesian approach is able predict that one implicature may be felt more strongly than another; see Chapter 5, especially Section 5.2.2, for applications. The Bayesian approach also has the advantage of simplicity: implicature computation is simply a matter of reasoning about alternative utterances the speaker could have used, while the game theoretic approach requires hearers and speakers to engage in “back-and-forth” reasoning about what they think each other thinks. Iterations in the game-theoretic approach enrich the theory with the ability to model higher order reasoning, and, as outlined by Franke, this has numerous applications, including the generation of free choice inferences and the pragmatics of conditionals. Nonetheless, the two approaches are, broadly speaking, compatible. In this section, specifically equations (4.65)–(4.67), I suggested a possible way to reconcile the approaches by introducing probabilistic reasoning about the speaker’s belief state to the game theoretic approach.

## 4.7 Conclusion

In this chapter, a simple intuitive theory of Utterance Expectation, based on the idea that speakers are expected to produce relevant, simple utterances that they believe are true, was developed. Based on this intuitive theory, it was shown that hearers are able to calculate a range of scalar implicatures. This calculation of scalar implicatures does not depend on the exclusion of “symmetric,” or “some and not all”-type, alternatives. A measure of complexity was developed that allows the theory to predict more complex utterances are much less probable than simpler utterances, but does not force speakers to make simple utterances. Finally, the theory was extended to provide a preliminary account of implicated presuppositions and presuppositional implicature, and was compared with a game-theoretic approach to similar phenomena.

I conclude this chapter with a brief discussion of the status of scalar implicatures generated in the Bayesian model. In particular, I wish to provide a response to those who may feel the scalar implicature is more than just a matter of increasing the probability of speaker knowledge states. For example, it may be somewhat unsatisfying that the theory does not predict that an utterance of “John ate some of the food” does not lead the hearer to conclude that John did not eat all of the food, but simply (i) lowers the hearer’s subjective probability that the speaker believes John ate all of the food ( $[\sigma]$ ), and (ii) raises the probabilities that the speaker believes John ate some (but not all) of the food ( $[o\omega]$ ) and that the speaker believes John ate some (and possibly, but not necessarily, all) of the food ( $[o\omega, \sigma]$ ).

But the hallmark of conversational implicature, of course, is its non-encoded nature: a speaker who implicates  $p$  has some wiggle room, since he didn’t actually say it. The Bayesian approach to implicature computation not only reflects the non-deterministic nature of implicature; it also provides a straightforward mechanism for implicature cancellation. The Bayesian theory predicts  $P([\sigma]|\ulcorner\omega\urcorner)$  is relatively low, but, in general, it will not predict this probability is zero. By providing a small probability that the implicated material is not true, the theory allows further conditioning on the negation of the implicated material: “In fact, I’m sure he ate all of it.” Simple conditional updating on assertions provides for the reversal of the implicature, so that the effect of the scalar implicature calculation is neutralized.

## **Chapter 5**

# **Against grammatical computation of scalar implicatures**

### **5.1 Introduction**

In recent years, a number of sophisticated arguments for the existence of “implicatures” in embedded positions have been put forth. In this chapter, I examine these arguments in light of the probabilistic theory developed in the preceding chapters. The first line of argument, raised in Sharvit and Gajewski (2008) (and extended in Gajewski and Sharvit (to appear)) is that a Gricean theory depends on the neg-raising (cf. Horn 1978) properties of propositional attitudes to derive the strengthening of implicatures necessary to produce “embedded” implicature interpretations. A second line of arguments, made in Chierchia et al. (to appear), suggest that acceptability patterns of entailment-related disjuncts provide evidence for grammatical generation of scalar implicatures.

The broad theoretical claim made by this dissertation is that the degree to which an implicature is “felt” is simply on the probability, calculated by the hearer, of the speaker’s belief in a stronger proposition. The calculation of this probability depends, in turn, on the relevance, defined probabilistically, of the speaker’s utterance compared to alternative

utterances, and on comparative simplicity. These theoretical claims have a number of empirical consequences related to the objections of Sharvit and Gajewski and of Chierchia et al. In particular, treating implicature as probabilistic inference leads to nuanced predictions about the relationship between neg-raising and “embedded implicature” phenomena that are not made by a grammatical approach as advocated by Sharvit and Gajewski. In addition, the probabilistic, relevance-based approach predicts variations in the data surrounding entailment-related disjuncts that are apparently outside the reach of the Chierchia et al. grammatical mechanism.

## 5.2 Sharvit and Gajewski’s response to Russell’s response to Chierchia

Recall from Chapter 1 that Chierchia (2004) argued that sentences with scalar terms embedded below *believe* provide a challenge to Gricean theories of scalar implicature. Chierchia’s example was:

(5.1) John believes that some of his students are waiting for him.

Chierchia points out that Gricean scalar reasoning only provides for the negation of the stronger alternative “John believes that all of his students are waiting for him”, but that the sentence is generally understood as implying something stronger: that John believes that not all of his students are waiting for him.

As outlined above in section 1.3.1, Russell (2006) provides an informal Gricean derivation of such ostensibly embedded implicatures. The derivation depends on the opinionatedness of John about the strong alternative: if John is opinionated (he either believes all of his students are waiting, or he believes not all of them are waiting), this fact combines with the global scalar implicature to yield an “embedded” implicature (note the

similarity between this analysis and the analysis of the strengthening of epistemic weak to epistemic strong implicature from Sauerland (2004), van Rooij and Schulz (2004), and Russell (2006)). In that work, I suggested that “embedded” implicatures’ dependence on an assumption of opinionatedness would make such inferences highly context-sensitive. Indeed, in contexts where such opinionatedness is explicitly ruled out, the “embedded” implicature fails to arise (see discussion surrounding (1.28) in Chapter 1).

Sharvit and Gajewski (2008) provide a challenge to this analysis. They point out that the apparently default opinionatedness of the subject of *believe* is connected to the fact that *believe* is a neg-raising verb (that is, that a sentence of the form *not believe p* is usually understood to mean *believe not p*). For example, the following sentence is typically interpreted to mean that John believes it is not raining.

(5.2) John doesn’t believe it’s raining.

Sharvit and Gajewski note that neg-raising is sufficient to predict “embedding” implicature behavior for neg-raising verbs: if surface wide-scope negation is automatically interpreted as narrow-scope negation, the scalar implicature  $\neg$  *John believes all of the students left* is automatically strengthened to *John believes  $\neg$  all of the students left*.<sup>1</sup> In effect, Sharvit and Gajewski argue the Russell (2006) analysis depends on the neg-raising property of *believe*, as far as it goes in explaining the presence of “embedded” scalar implicatures below *believe*. The idea that “embedded” implicatures are derived by the neg-raising properties of the embedding predicate can be termed the neg-raising analysis of “embedded” implicature.

The neg-raising analysis of “embedded” implicature makes a decisive prediction: predicates that do not have the neg-raising property should not have “embedded” implicatures. Sharvit and Gajewski then go on to provide evidence that contradicts this prediction:

---

<sup>1</sup>This really depends on the analysis of neg-raising: to the extent that neg-raising depends on the speaker actually saying “believe”, strengthening will not be expected to apply to implicated material.



they claim that examples with the predicate *be certain*, which is not a neg-raiser (since, for example, “I am not certain Harry left” does not imply that I am certain Harry didn’t leave), give rise to “embedded” implicature interpretations, patterning just like neg-raisers like *believe*. For example, Sharvit and Gajewski posit that an utterance of the sentence in (5.3) usually carries the “embedded” implicature that John’s certainty is that only one of them disappeared.

(5.3) John is certain that the boss or her assistant has disappeared.

This fact, Sharvit and Gajewski argue, provides evidence for a grammatical view of scalar implicatures: if *believe* and *certain* have different neg-raising properties but pattern identically with respect to “embedded” implicature phenomena, “embedded” scalar implicature cannot simply be a consequence of neg-raising.

### 5.2.1 Gricean “embedded” implicature in more detail

Whereas Sharvit and Gajewski argue that the analysis only works for neg-raisers, the suggestion in Russell (2006) is not simply that “embedded” implicatures are due to neg-raising. Rather, I argued that an “embedded” implicature interpretation arises for *believe* only when certain background assumptions are present in a given context. But a question remains about the connection between neg-raising and such background assumptions. To answer this question in detail, it is useful to look at an example of embedded implicature in detail. Suppose the belief of John is  $B_j$ ,  $\omega$  is *Rudi speaks French or German*, and  $\sigma$  is *Rudi speaks French and German*. For an utterance of “John believes Rudi speaks French or German,” Gricean scalar reasoning generates the inference  $\neg B_j \sigma$ . From here, the assumption that John is opinionated about the stronger alternative (i.e.,  $B_j \sigma \vee B_j \neg \sigma$ )

yields the “embedded implicature” interpretation. Schematically:

$$\lceil B_j \omega \rceil \quad \text{“John believes Rudi speaks French or German”} \quad (5.4)$$

$$\downarrow \quad \text{Gricean scalar reasoning + well-informed speaker} \quad (5.5)$$

$$\neg B_j \sigma \quad \text{NOT J believes R speaks F and G (epistemic strong, global)} \quad (5.6)$$

$$\downarrow \quad \text{opinionated John } ((B_j \sigma \vee B_j \neg \sigma)) \quad (5.7)$$

$$B_a \neg \sigma \quad \text{J believes NOT R speaks F and G (“embedded”)} \quad (5.8)$$

The last step in the derivation requires an assumption that John is opinionated.<sup>2</sup> This means that, regardless of the verb or predicate, the step will depend on whether the context provides that John is opinionated with respect to that verb or predicate.

There is a problem with this analysis: because the derivation beginning in (5.4) depends on the truth of  $(B_j \sigma \vee B_j \neg \sigma)$ , the analysis predicts that the “embedded” implicature only arises when the context specifically lays out the truth of John’s opinionatedness. But, in general, contexts that strictly entail the opinionatedness of agents like John are probably rare. In most contexts, it is reasonable to think opinionatedness is not taken for granted, but is rather considered more or less likely, depending on various contextual factors, including the propositional attitudes involved, the proposition, and related facts about the world, including what is known about the related experiences and beliefs of agents like John.

### 5.2.2 Probabilities make a more nuanced prediction

The problem with the Russell analysis is that it depends on the truth of  $B_j \sigma \vee B_j \neg \sigma$  for the generation of “embedded” implicature. The analysis was put forth in a basically non-

<sup>2</sup> To the extent that neg-raising verbs like *believe* carry an opinionatedness assumption as a semantic presupposition (as Sharvit and Gajewski claim, following Bartsch (1973)), that step may be automatic for neg-raisers.

probabilistic setting, and it therefore is unable to handle assumptions about how likely John’s opinionatedness is. But the probabilistic approach, because it uses prior probabilities to account for epistemic strengthening (see section 4.2), extends straightforwardly to “embedded” implicature phenomena, providing a degree of belief in the “embedded” implicature interpretation relative to the “wide scope” interpretation. In the derivation, strengthening is achieved probabilistically: the “embedded” interpretation will depend on how likely it is that John is opinionated. (I use “opinionated” generally, with respect to a propositional attitude or other proposition-taking predicate  $B^i$  and a proposition  $\chi$ , to mean  $B^i\chi \vee B^i\neg\chi$ .) The degree of belief in the strengthening of a global implicature to an “embedded” implicature, under this theory, should simply depend on the prior probability of the “embedded” interpretation compared to the probability of the wide scope interpretation, or on the conditional probability  $P(B_j\neg\sigma|\neg B_j\sigma)$ .

This approach makes immediate predictions for the effect of varying propositional attitudes on “embedded” scalar implicatures. Specifically, the theory makes a prediction about the relationship between the semantic strength of a propositional attitude (e.g.: “pretty sure” versus “absolutely sure”) and the robustness of “embedded” implicature interpretations. Simply put, the prediction is that a semantically stronger propositional attitude will display a less robust “embedded” implicature than a weaker propositional attitude.<sup>3</sup> In broad, informal terms, this is because *absolutely sure that X* has a strictly smaller extension than *pretty sure that X*, so *absolutely sure that X*  $\cup$  *absolutely sure that  $\neg X$*  also has a smaller extension than *pretty sure that X*  $\cup$  *pretty sure that  $\neg X$* . That means that opinionatedness is less likely relative to *absolutely sure* than it is relative to *pretty sure*, and so the “embedded” implicature is less likely to be true with the former than with the latter.

In other words, given the epistemic strong implicature for the two attitudes, strengthening to the “embedded” implicature interpretation is more likely with *pretty sure* than

---

<sup>3</sup>The predicted data are examined below in Section 5.2.5.

with *absolutely sure*. A simple calculation confirms this. Let  $B_1$  be *absolutely sure* and  $B_2$  be *pretty sure*. Note that, with respect to the two attitudes, entailment relations are given as follows:

$$\begin{array}{ccc} B_{1j}\neg\sigma & \subset & \neg B_{1j}\sigma \\ \cap & & \cup \\ B_{2j}\neg\sigma & \subset & \neg B_{2j}\sigma \end{array} \quad (5.9)$$

These entailments, in turn, give rise to the following inequalities between conditional probabilities:

$$P(B_{1j}\neg\sigma|\neg B_{1j}\sigma) < P(B_{2j}\neg\sigma|\neg B_{2j}\sigma) \quad (5.10)$$

Proof:<sup>4</sup>

$$P(B_{1j}\neg\sigma|\neg B_{1j}\sigma) = \frac{P(B_{1j}\neg\sigma)}{P(\neg B_{1j}\sigma)} \quad (5.11)$$

$$< \frac{P(B_{2j}\neg\sigma)}{P(\neg B_{1j}\sigma)} \quad (5.12)$$

$$< \frac{P(B_{2j}\neg\sigma)}{P(\neg B_{2j}\sigma)} \quad (5.13)$$

$$= P(B_{2j}\neg\sigma|\neg B_{2j}\sigma). \quad (5.14)$$

The inequalities in conditional probabilities show that, with respect to the weaker propositional attitude  $B_2$ , the global epistemic strong implicature is closer to the “embedded” implicature.

Another way to look at this is to recognize that, as the propositional attitude weakens, the number of worlds that are in neither the positive nor negative extension shrinks.<sup>5</sup> Con-

---

<sup>4</sup>Another way to see that this is true is to observe the chain  $B_{1j}\neg\sigma \subset B_{2j}\neg\sigma \subset \neg B_{2j}\sigma \subset \neg B_{1j}\sigma$ , and to notice that the terminal elements of the chain must be further apart (have more worlds that are in one but not the other) than the middle elements.

<sup>5</sup>Note that, for some attitudes, this number will be zero. For example, every world is contained in the union of *possible* with *possible not* (see discussion around (5.18a) for elaboration). For these attitudes, “neg-raising” is essentially impossible, since the narrow scope negation is weaker than wide scope. Therefore, the generalization about weakening attitudes leading to increased probability of “opinionatedness” applies only to attitudes for which there is a gap between positive and negative extensions, or those which Löbner (1987) calls *intolerant*.

sider, for example, “more likely than not” or, equivalently, “the odds are in favor of”: if it is not the case that John thinks the odds are in favor of  $p$ , John must think the odds are in favor of  $\neg p$  (except for the special case when John thinks  $p$  and  $\neg p$  are equally likely). For this propositional attitude, then, the “embedded” implicature is expected to emerge almost automatically: the “embedded” implicature and the global, epistemic strong implicature are nearly synonymous. Indeed, this prediction is borne out by the facts:

(5.15) John thinks the odds are in favor of most of the students passing.

Here, extension of the global implicature and the “embedded” implicature are nearly equivalent.

- (5.16) a. Global: NOT(John thinks the odds are in favor of all students passing)  
 $= P_j(all-pass) \leq P_j(\neg all-pass)$
- b. Local: John thinks the odds are in favor of NOT(all students passing)  
 $= P_j(all-pass) < P_j(\neg all-pass)$

In these implicatures,  $P_j$  represents John’s belief distribution: the only difference between the globally-generated implicature and the “embedded” implicature is the exclusion of the state where John thinks *all-pass* and  $\neg all-pass$  are equally likely. If John’s beliefs don’t assign exactly equal probabilities to *all-pass* and  $\neg all-pass$ , the global and local implicatures are identical. That means that a theory that generates an implicature *that John doesn’t think the odds are in favor of all students passing* automatically generates the local implicature, *that John thinks the odds are in favor of not all passing* (unless John thinks the odds are exactly evenly split between all passing and not all passing). For such predicates, then, local and global theories make nearly the same predictions, since the local and global implicatures are nearly identical.

### 5.2.3 Neg-raising and the probabilistic approach to implicature

I have argued that it is wrong to conflate neg-raising and the Gricean view of “embedded implicature” argued for in Russell (2006). But the opinionatedness-based strengthening approach to “embedded” implicature is, indeed, closely related to Horn’s (1978) view of neg-raising, where he argues that neg-raising predicates all exhibit a relatively slim difference between the meanings with wide and narrow scope negation: “It is the closeness of the external (contradictory) readings of *not likely*, *not believe*, *not advisable* to *likely not*, *believe not*, *advisable not*, respectively, which renders the negated predicates potential neg-raisers, and the relative distance of *not possible*, *not realize*, *not obligatory* from *possible not*, *realize not*, *obligatory not* which removes these from contention.”

For each of these predicates (with the exception of *possible*—see discussion below), there is a partition created by  $B\chi$ ,  $\neg B\chi \wedge \neg B\neg\chi$ ,  $B\neg\chi$ . The distance between “the external (contradictory) reading” of  $\neg B$  and  $B\neg$ , then, can be understood as the size of the middle partition: those worlds where  $\neg B$  is true and  $B\neg$  is false, or where the agent is unopinionated. And that, in turn, is determined by the semantic strength of  $B$ , which increases as the end partitions shrink and the middle partition grows. In sum, then, the size of the middle partition, or the gap between  $B\chi$  and  $B\neg\chi$ , is a direct consequence of the semantic strength of the predicate  $B$ . So, whereas Horn’s generalization is spelled out in terms of the slimness of the difference between the wide and narrow scope readings, the prediction made above about “embedded” implicatures is given in terms of the semantic strength of the propositional attitude. These are two sides of the same coin.

Opinionatedness-based “embedded” implicature computation and neg-raising should not, however, be conflated. In particular, Horn points out that the slimness of gap property is not a sufficient predictor of neg-raising behavior, since near-synonyms *guess* and *suppose* pattern differently in some dialects of English, as do near-synonyms *espérer* and *souhaiter* in French. In light of this fact, he argues that neg-raising involves “short-

circuited implicature,” whereby a pragmatic meaning may or may not become conventionalized (when neg-raising becomes conventionalized, it can provide for the licensure of polarity items, a grammatical phenomenon), with seemingly arbitrary lexical exceptions.

For “embedded” implicature, there do not seem to be the same arbitrary exceptions:

(5.17) Q: “Why do you think you got heckled when you gave your lecture about how to succeed in the music industry?”

A: “I guess/suppose most of these kids have never heard of Nirvana.”

Does the speaker here guess/suppose that some of these kids have heard of Nirvana? I think so: my intuition is that the implicature is felt equally strongly in both cases. That is, because *guess* and *suppose* are roughly equivalent in terms of semantic strength, opinionatedness-based strengthening predicts they will have similar “embedded” implicature patterns (contrary to their differing neg-raising patterns); and, indeed, this prediction is borne out. It is precisely the non-conventionalized nature of the Gricean theory of “embedded implicature” that accounts for its distinctness from neg-raising.<sup>6</sup>

Among the predicates listed by Horn—*likely*, *believe*, *advisable*, *possible*, *realize*, and *obligatory*—there is one that does not fit with the others: *possible*. What is different about *possible* is that *not possible* entails *possible not*: “It is not possible that it will rain tomorrow” entails “It is possible that it will not rain tomorrow.” Because of this, *possible* and *possible not* overlap, giving rise to the following contrast:

---

<sup>6</sup>Larry Horn (p.c.) notes that there are weak intolerant deontic modals like *better*, *ought to*, *be supposed to*, or *should*, all of which are well-behaved neg-raisers, but don’t seem to receive a local implicature interpretation:

- (5.i) a. I don’t think you {better/ought to} hang around here anymore.  
       = I think you {better not/ought not to} hang around here anymore.  
       b. You {better/ought to} apologize to your mother or your father.  
       ≠ You {better not/ought not to} apologize to your mother and your father.

I leave the deontics for future research. See Ippolito (2010) for discussion of other predicates.

- (5.18) a. It's possible that John left, and it's possible that he didn't.  
 b. #It's likely that John left, and it's likely that he didn't.  
 c. #It is obligatory that John leave, and it is obligatory that he not leave.

In Horn (1989), this fact is recognized and tied to Löbner's (1987) definition of *tolerant* quantifiers: those for which  $Q(P)$  and  $Q(\neg P)$  may both be true at the same time. For tolerant attitudes like *possible*, the “slimness of gap” property does not apply; there is no gap at all.<sup>7</sup> So, whereas a narrow-scope negation reading is stronger than wide-scope for *likely*, *believe*, *advisable*, *realize* (modulo factive presupposition), and *obligatory*, the wide-scope negation reading is stronger for *possible*. That means that it does not make sense to think of “embedded” implicatures under *possible* as a strengthened version of the global implicature; they are simply entailments of the global implicature.

Consider the following: in an article about typefaces, the following claim is made:

- (5.19) It's possible that most of the 'Helvetica' seen in the '70s was actually not Helvetica.

Here, the author is (robustly) understood to believe it is possible that not all of the purported '70s 'Helvetica' was phony. In a Gricean theory, however, this does not depend on a strengthening of the global implicature as in the derivation starting with (5.4). Rather, the global implicature actually entails the “embedded” implicature: If the speaker believes it is not possible that all of the purported Helvetica was phony, the speaker must also believe it is possible that not all of the purported Helvetica was phony.

The flip side of this property of *possible* is that the grammatical view of “embedded” implicature predicts that, by default, scalar terms embedded under *possible* would have the weak embedded reading.<sup>8</sup> The example above suggests this is not the case: the speaker

<sup>7</sup>See Horn (1989, pp. 325–329) for useful discussion.

<sup>8</sup>This actually only applies to those grammatical theories that generate local implicatures by default, rather than generating the strongest implicature possible.



who says that it is possible that most of the purported Helvetica was phony strongly implies that he does not believe it is possible that all of the purported Helvetica was phony.

#### 5.2.4 Formal Derivation

Embedded implicature, in terms of the theory developed in Chapter 4, is strongly tied to epistemic strong implicatures. In Chapter 4, I argued that epistemic strong implicatures are simply a matter of relative prior probabilities: how likely is it for a speaker to be well-informed? Likewise, I have argued in this chapter that the strength of an “embedded” implicature is a question of how likely an agent is to hold a strong, rather than weak, propositional attitude with respect to the implicature-containing proposition.

The derivation of a simple example is straightforward. Suppose the speaker is telling the hearer about a friend named Jones, a math teacher who has taken over leadership of a dismally under-performing Mathletes team at an inner-city school, and who, the speaker argues, is doing a stellar job.

(5.20) Jones believes that some of his students have improved enough to make the national team. ( $= Jb\omega$ )

Here, the strong scalar alternative,  $Jb\sigma$ , would surely be more relevant: Jones thinks he has made an across-the-board dramatic improvement in his Mathletes’ performance. Modestly, then, let’s suppose  $r_H(Jb\sigma) = 0.5$ ,  $r_H(Jb\omega) = 0.3$ , and  $r_H(Jb\omega\omega) = 0.2$ .

Before computing implicatures, note that the introduction of propositional attitudes expands the relevant set of possible speaker belief states considerably. The bracket notation here makes the belief-holder explicit:  $[_S]$  is a speaker belief state and  $[_J]$  is John’s belief state. This leads to beliefs ranging from  $[_S[_J\sigma]]$  to  $[_S[_J\omega\omega]]$ ,  $[_J\sigma]$ ,  $[_J\omega\omega, \sigma]$ . Limiting ourselves to those speaker beliefs that are not uncertain about John’s beliefs (i.e., limiting ourselves to opinionated speakers), the question of “embedded” implicature becomes the

following ratio:

$$\frac{P([S[Jo\omega]]|\ulcorner Jb\sigma\urcorner)}{P([S[Jo\omega, \sigma]]|\ulcorner Jb\sigma\urcorner)}. \quad (5.21)$$

Recall from the calculations beginning in (4.30) that

$$\frac{P([S[Jo\omega]]|\ulcorner Jb\sigma\urcorner)}{P([S[Jo\omega, \sigma]]|\ulcorner Jb\sigma\urcorner)} = \frac{N([S[Jo\omega, \sigma]])}{N([S[Jo\omega]])} \frac{P([S[Jo\omega]])}{P([S[Jo\omega, \sigma]])} \quad (5.22)$$

As in Chapter 4, this ratio of  $N$  terms may be easily computed through a simple tabulation, as illustrated in Figure 5.1.<sup>9</sup> The table reveals that

	$r_H(x)$	$s(x)$	$[S[Jo\omega]]$	$[S[Jo\omega, \sigma]]$
“ $Jb\omega$ ”	0.5	1	1	1
“ $Jb\sigma$ ”	0.7	1	0	0
“ $Jbo\omega$ ”	0.2	$10^{-3}$	1	0
$N()$			0.5002	0.5

Figure 5.1: Scalar reasoning yields rough equivalence of “epistemic strong” and “ignorance” scalar implicatures.

$$\frac{N([S[Jo\omega, \sigma]])}{N([S[Jo\omega]])} = \frac{0.5}{0.5002} = 0.9996. \quad (5.23)$$

This means that, for all intents and purposes, the ratio of  $N$  values is 1. So the degree to which the “embedded” implicature is felt depends nearly entirely on the ratio of the prior probabilities. As I argued above, the ratio of the priors will favor  $[S[Jo\omega]]$  if the attitude in question is *belief*, rather than certainty. So the formal theory developed Chapter 4 makes two desirable predictions about “embedded” implicature: 1) like epistemic strong implicature, it depends on prior probabilities; and 2) substituting a stronger propositional attitude under which the scalar term is embedded weakens the “embedded implicature” interpretation.

<sup>9</sup>As in 4.3, I simplify the simplicity values  $s(x)$ , simply assuming the simplest alternatives receive a value of 1, and that *and not all* decreases simplicity by a factor of a thousand.

### 5.2.5 Implications for Sharvit and Gajewski's challenge

I have outlined a prediction made by the probabilistic approach: semantically weaker propositional attitudes produce stronger (more robust) “embedded” implicatures.<sup>10</sup> Sharvit and Gajewski substitute the non-neg-raiser *certain* for *believes* and seem to take it for granted that both verbs behave the same with respect to “embedded” implicatures. However, as outlined above, the prediction of a probabilistic Gricean theory is that *certain*, with its greater semantic strength than *believe* (I can believe Obama will win the election without being certain of it, but I can't be certain of it without believing it), will have a weaker “embedded” implicature interpretation than *believe*.

Consider the following context: Obama has just unveiled his proposal for healthcare reform, which includes an estimate of \$500 billion in savings to the Medicare program. A reporter asks Obama: “So you're going to pay for health care reform by cutting health care for seniors?” Consider the following pair of replies, followed by inferences drawn by the reporter:

- (5.24)    a.    Obama: “My budget director believes that most of the changes we have planned for Medicare will actually improve outcomes for seniors.”  
               b.    Reporter infers: The budget director believes that some of the changes will not improve outcomes for seniors.
- (5.25)    a.    Obama: “My budget director is certain that most of the changes we have planned for Medicare will actually improve outcomes for seniors.”  
               b.    Reporter infers: The budget director is certain that some of the changes will not improve outcomes for seniors.

---

<sup>10</sup>It is arguably also a prediction made by any rationality-based, Gricean approach. I have tried to argue in Chapter 2 that a rationality-based, Gricean approach is, by necessity, also a probabilistic approach. At any rate, there is no obvious way to talk about the strength with which a particular implicature is felt without identifying implicatures with probabilities.

In which case has the reporter drawn a more solid inference? My intuition is that the inference in (5.24) is more solid than the one in (5.25), and this accords well with the intuitions of numerous informants I've consulted. As outlined in 5.2.4, this is exactly what the probabilistic system predicts.

Further examples with stronger and weaker propositional attitudes highlight this result:

- (5.26)    a.    Obama: "My budget director is absolutely certain that most of the changes we have planned for Medicare will actually improve outcomes for seniors."
- b.    Reporter infers: The budget director is absolutely certain that some of the changes will not improve outcomes for seniors.
  
- (5.27)    a.    Obama: "My budget director thinks it is fairly likely that most of the changes we have planned for Medicare will actually improve outcomes for seniors."
- b.    Reporter infers: The budget director thinks it is fairly likely that some of the changes will not improve outcomes for seniors.

According to the intuitions of informants, the contrast here is rather stark: in (5.27), the reporter's inference feels relatively solid: the epistemic strong, wide scope implicature (that it is not the case that the director thinks it is fairly likely that all of the changes will improve outcomes) is largely coextensive with the "embedded" implicature (that the budget director thinks it is fairly likely that not all of the changes will improve outcomes). On the other hand, in (5.26), the reporter is on shakier ground—the budget director might be absolutely certain that some of the changes won't improve outcomes, but that does not seem to have been implied: maybe the budget director is uncertain about the consequences of the rest of the changes; maybe he thinks it is fairly likely that they will also improve outcomes, without being absolutely certain. The utterance itself does not seem to make it more likely that the director is absolutely certain that there are changes planned that will

not improve outcomes.<sup>11</sup>

In addition to these deliberately crafted examples, limitless data are available from Google. A UC Davis professor, making an argument that the fate of the remains of Kennewick Man (a skeleton found in the shallows of the Columbia River near Kennewick, Washington) should be determined by the Native Nations of the region, writes:

(5.28) It is absolutely certain that most of Kennewick Man's descendants are Native Americans.

<http://nas.ucdavis.edu/Forbes/kennwick.html>

Would such an utterance indicate that the speaker thinks it is absolutely certain that not all of Kennewick Man's descendants are Native Americans? It is pretty clear to me that the utterance is more likely to be uttered when the speaker is agnostic about whether all descendants are Native Americans.<sup>12</sup> For comparison, suppose the writer had written: "It is pretty likely that most of Kennewick Man's descendants are Native Americans." With a weaker predicate, the "embedded" implicature interpretation becomes more likely, as expected: the speaker does not have strong evidence that most of Kennewick Man's descendants are Native Americans, so he is likely to think it is pretty likely that not all of the descendants are Native Americans.

For one further example, consider the following warning posted on a travel web site:

(5.29) You can be absolutely certain that most of the hotels in or near Haadrin Beach

<sup>11</sup>In addition to informal intuitions, I conducted a pilot study with 15 subjects and found a significant interaction between semantic strength of propositional attitude *absolutely certain*/*pretty likely* and "embedded"/global interpretation ( $F(1, 14) = 12.277; p = 0.004$ ). Whereas subjects rated the likelihood of a global implicature high for both attitudes, the mean likelihood judgment for the "embedded" implicature interpretation was lower for *absolutely certain* than for *pretty likely* (means 4.6 and 8.2, respectively, on a scale of 0 to 10). See Chemla and Spector (2011) for related results; though they measure appropriateness of an utterance in a given situation.

<sup>12</sup>Some of my informants have an even stronger intuition: that the speaker thinks it is fairly certain that all of Kennewick Man's descendants are Native Americans, but, for whatever reason, does not want to "go there."

will have a minimum stay of 3 nights - most actually up to 5 or even 7 nights.

<http://www.phanganvideos.com/>

Can you be absolutely certain that there are hotels without a 3 night minimum? I'd say not. But what if the message instead read: "You can be somewhat confident that most of the hotels in or near Haadrin Beach will have a minimum stay of 3 nights"? Could you be somewhat confident that there are hotels without a 3 night minimum? My intuition is that you probably could.

The prediction that the strength of an "embedded" implicature is inversely related to the semantic strength of the propositional attitude that the scalar term is embedded below (modulo other contextual factors) is a straightforward consequence of a non-grammatical theory implemented with a probabilistic formalism, as laid out in Section 5.2.4. A grammatical theory, on the other hand, would be at pains to explain this difference in strength. Grammatical theories generally predict that the "embedded" implicature is generated by default, or that the strongest implicature is generated by default. To predict a distinction in the strength of these implicatures, a grammatical system would apparently have to appeal to a Gricean mechanism for generating graded pragmatic inferences when default implicatures are canceled. But if grammatical theories need Gricean principles to handle facts about "embedded" implicatures, a major argument in favor of grammatical theories dissolves.

### 5.3 Cancellation with *or*

A second line of argumentation in favor of a grammatical theory of scalar implicature is built around *or*-cancellation, or cases where a scalar implicature is cancelled through the use of disjunction. For example:

(5.30) Louis wants you to read Winnie the Pooh or Paddington, or both.

Recently, Sharvit and Gajewski (2008) and Chierchia et al. (to appear) have argued that the behavior of such sentences—the fact that any exclusive *or* implicature is cancelled—provides evidence that a scalar implicature is computed locally, on the *or* that coordinates *Winnie the Pooh* and *Paddington*. To examine this argument, I begin by taking a step back to talk about cancellation more generally.

The term *cancellation* is generally applied to cases where a scalar term appears, yet a scalar implicature is ultimately not present. These cases include sentences like:

(5.31) John solved some of the problems correctly; in fact, he solved all of them correctly.

Such cases are generally valid when there is an explicit question of whether John solved some of the problems correctly, and the answer to the question is affirmative. But, because this affirmative answer tends to carry an implicature that John did not solve all of the problems correctly, the speaker who believes he did solve all the problems retracts this implication.<sup>13</sup>

To spell this out, a somewhat more realistic context for (5.31) is:

(5.32) A: Is it true that John solved some of the problems correctly?

B: Yes! In fact, he solved all of them correctly.

Often, the canceling phrase appears in the same utterance as the scalar term, as a parenthetical that immediately follows the scalar.<sup>14</sup>

<sup>13</sup>See van Kuppevelt (1996) for elaboration and qualifications.

<sup>14</sup>Cases like the one in (5.33a) are typically referred to as *suspension* of the implicature; all that is asserted is that the negation of the implicature is possible (see Horn 1989 for discussion).

- (5.33)    a.    You will find solutions to some, and perhaps most, of the assigned homework problems in the back of the textbook.
- b.    Many infants will spit up after some, or even all, feedings or during burping because their digestive tracts are immature.

And sometimes, a strong scalar is coordinated with a weak scalar using *or*, without any apparent change in intonation or modulation of prosody. For example, the following sentence, in a discussion of Medicare supplemental health insurance policies, sounds natural with the following punctuation:

- (5.34)    These policies may pay for some or all of the Medicare coinsurance amounts; some or all deductibles; and certain services not covered by the Original Medicare Plan at all.

In such a case, something very different seems to be going on. Ostensibly, the sentence does not carry a scalar implicature at all: by virtue of the disjunction with the strong scalar *all*, the implicature usually associated with *some* is apparently defeated; it never arises.

### 5.3.1 Hurford's Constraint and local implicatures

Hurford (1974) observed that, in many cases, it is bad to utter a sentence of the form *A or B* when either disjunct entails (more generally, is semantically stronger than) the other: a person who utters *John lives in California or Sacramento* seems to think Sacramento is not in California. However, disjunctions of the form *L or R or both* are much better: *George will order fries or a shake, or both*. Hurford argued that such *or both* examples provide evidence for an English *or* that is ambiguous between exclusive and inclusive interpretations. The argument is simple: the lower disjunction (*fries or a shake*) must be exclusive (i.e., must mean *fries or a shake, but not both*); otherwise, it would be illicit



in a disjunction with *both*, (which abbreviates *fries and a shake*), since *both* entails an inclusive *fries or a shake*, just as Sacramento entails California.

Chierchia et al. (to appear) do not subscribe to Hurford’s ambiguous *or* analysis. They note (following Gazdar (1979)) that exceptions to Hurford’s generalization are not limited to *or both* cases; they also include other scalar terms, like *Most or all of the officials will be willing to share their fries*. Instead, Chierchia et al. argue that two conclusions may be drawn from Hurford’s observations. The first is that Hurford’s observation about entailment-related disjuncts is the result of a constraint (presumably in the grammar) that says *L or R* is illicit when *L* entails *R* or vice versa. The second is that apparent violations of Hurford’s Constraint involving scalar terms are licit because the weak scalar disjunct is enriched with locally-generated scalar implicatures. Chierchia et al. posit a grammatical computation mechanism that, for the sake of simplicity, involves the free application of a covert *only*, which they call *exh* or *O* (I will use the latter here, for brevity), in the syntactic composition. Such an operator, of course, affects a sentence’s truth conditions:  $\llbracket p \rrbracket \neq \llbracket \text{only } p \rrbracket = \llbracket O(p) \rrbracket$ .<sup>15</sup> The application of *O* in the grammar narrows the interpretation of the weak disjunct so that it is no longer in an entailment relation with the other disjunct. In this way, the presence of a grammatical scalar implicature-computation operator would account for the insensitivity of scalar terms to Hurford’s Constraint.

### 5.3.2 The softness of Hurford’s Constraint

Chierchia et al.’s analysis depends on a sharp distinction between scalar terms, for which Hurford’s Constraint may be foiled by local scalar implicature computation, and non-scalar, entailment-related pairs, for which Hurford’s Constraint violations “crash”, due to the impossibility of local scalar implicature computation. But this distinction is blurry: numerous examples of non-scalar Hurford’s Constraint violations that do not “crash” may

---

<sup>15</sup>This is an approximation that glosses over the distinction between presupposition and assertion: *O(p)* asserts both *p* and the negation of a stronger alternative to *p*, whereas *only(p)*, of course, presupposes *p* while asserting the negation of stronger alternatives.

be found in the “wild”.<sup>16</sup>

- (5.35) We also rent only the most modern limos to our customers, because we believe that when you look for a limo service in **Northern California or San Francisco**, you want the best limousine service possible.

<http://www.sfolimousine.com/>

- (5.36) Over the last seven years the Summit has featured over 300 groups from around the United States as well as Europe, Australia, and Japan including exclusive first time debut performances by artists who have never played in **California or San Francisco**.

<http://www.outsound.org/summit/summit.html>

- (5.37) The patient zero story—in its various versions, that Gaetan Dugas, a Canadian flight assistant, “brought AIDS to” **California or San Francisco or the United States**—is based on a study by David Auerbach and Darrow published in 1984.

<http://www.nybooks.com/articles/4227>

- (5.38) Christmas in **France or Paris** is memorable and magical.

<http://gofrance.about.com/>

- (5.39) French property and in particular Paris property have consistently been favoured investments for Irish and UK investors. Leaseback investments in France and Paris are popular as well. But a growing number of investors who are buying property in **France or Paris** are opting for the freedom of freehold French properties.

<http://www.parism2.com/>

---

<sup>16</sup>Not all of these sentences sound natural to all speakers, especially out of context. But a significant portion of them are judged to be natural by all of my informants, and that is sufficient for the point made here.

- (5.40) With plastic **deck furniture or chairs** there is really no need to use any expensive product or chemical, **Dawn dish soap or dish soap in general** with some warm water and a scrubby is really all you need for the elements that they are going to come across outside and that would be how you clean plastic deck chairs and furniture.”  
[http://www.ehow.com/video\\_4951794\\_cleaning-plastic-deck-chairs.html](http://www.ehow.com/video_4951794_cleaning-plastic-deck-chairs.html)
- (5.41) Every now and again, people tend to change their surroundings. We update wall colors, change the drapes. Have new flooring installed. Sometimes we purchase new **furniture or chairs**.  
[http://www.ehow.com/how\\_4693151\\_paint-old-vinyl-chair.html](http://www.ehow.com/how_4693151_paint-old-vinyl-chair.html)
- (5.42) In order to standardize the quality, look and type of furniture across the campus, the University has recently signed contracts with two vendors—one for furniture and another supplier for chairs. If you are planning on purchasing new **furniture or chairs** for your lab or office, please see the Storeroom staff.  
<http://www.biology.ualberta.ca/facilities/storeroom/?Page=6666>
- (5.43) Funds cannot be used for non-technology purchases such as **furniture or chairs**.  
<http://blog.lyrasis.org/gatesgrant2008/?tag=acceptable-purchase>
- (5.44) “I’ll get a call usually that a family has been evicted and that I would have to show up either because there have been dogs or cats that have been left behind for either a few hours, a few days,” said Gonzales. “Sometimes it’s been weeks.

By the time I've gone in I've had to pull out **an animal or a cat** that's on the verge of dying."

<http://www.cpbm.org/article/economy-leaves-horses-and-pets-hungry-abandoned>

- (5.45) "It's so interesting to me that when **an animal or a dog** is out of its usual context, like at a funeral, a lot of people react like they've never seen a dog before," Patty DeJohn said. "At a funeral, it's awkward. People don't know what to say to each other. Animals are something people mutually like, so it brings them together."

<http://blog.cleveland.com/sunmessenger/2009/09/south-euclid-funeral-home-uses.html>

These data have significant consequences for the Chierchia et al. theory. The assumed presence of Hurford's Constraint means one disjunct never entails the other, so examples like (5.35) through (5.45) must be analyzed as containing the exhaustivity operator *O*, applied to the weak disjunct. In (5.45), this means the exhaustivity operator applies to *animal*, thereby leading to an interpretation where *animal* means *animal and not dog*. This means that the theory must permit substantially free application of *O*: *California or San Francisco*, *France or Paris*, *furniture or chairs*, and *an animal or a cat* are all attested, so *O* must apply to *California*, *France*, *furniture*, and *animal*, in addition to scalar terms like *some*, *most*, *or*, and *possible*. The fact that *O* rescues these sentences from Hurford's constraint violations, then, raises the question of why *O* does not rescue every Hurford's Constraint violation. In other words: if *O* can apply freely to non-scalar, as well as scalar, terms, why do Hurford's Constraint violations ever "crash"?<sup>17</sup>

<sup>17</sup>See also Rohdenburg (1985) for numerous examples of entailment related disjuncts, including:

- (5.i) This recipe calls for a pound of fat or suet.

- (5.ii) Home can be a dangerous place for an infant or child with insatiable curiosity.

### 5.3.3 Truth-conditional effects of local implicatures?

Additional data presented by Chierchia et al. in support of the claim that Hurford's Constraint interacts with *O* raises further questions. Earlier arguments for grammatical computation have not included predictions of hard, truth-conditional effects of grammatically-computed implicatures. This is because scalar implicatures are generally cancellable, so it is stipulated that the presence of grammatically computed implicatures is optional (though cf. Magri 2009 and Singh 2010, who argue for mandatory application). This means every sentence is (many-ways) ambiguous, and a sentence like *some of the students left* may be judged true when all of the students left, provided that it is disambiguated along non-*O*-containing lines.

Chierchia et al. argue that disambiguation without *O* is not always possible. They reason that, assuming Hurford's Constraint is operative, potential violations would force the insertion of *O*, creating an obligatory change in the truth conditions of a sentence. In ordinary *some or all*-type sentences, no change in overall truth conditions is predicted: the union of  $\llbracket \text{some and not all} \rrbracket$  with  $\llbracket \text{all} \rrbracket$  is equivalent to  $\llbracket \text{some} \rrbracket$ , just like the union of the *O*-free  $\llbracket \text{some} \rrbracket$  with  $\llbracket \text{all} \rrbracket$ . But in sentences where the application of *O* strengthens the first disjunct beyond simple negation of the second disjunct, there is a "middle space" that is excluded, changing the truth conditions of the sentence as a whole. They provide the following datum:

(5.46) Peter either solved both the first and the second problem or all of the problems.

This sentence, Chierchia et al. claim, is false when Peter solved just the first three problems (but not the fourth, fifth, and sixth). This is predicted by the theory because the second disjunct, *all of the problems*, is semantically stronger than the first disjunct, *both the first and the second problem*. Therefore, Hurford's Constraint requires the insertion of an *O*,

---

These examples were presented within an argument against Kempson's (1980) polysemy proposal for words like *dog* with strength-related terms like *bitch*.

changing the interpretation of the first disjunct to *only the first and the second problem*.<sup>18</sup>

Chierchia et al. make a similar claim for the following sentence:

- (5.47) It is either the case that every student solved some of the problems, or that Jack solved all of them and all the other students solved only some of them.

This sentence, they claim, has only the reading where the first disjunct is interpreted to mean every student solved some but not all of the problems. The empirical claim, then, is that the sentence is false if any student other than Jack solved all the problems. The second disjunct here entails the first, when interpreted with a wide-scope implicature (i.e., not every student solved all of the problems), so the insertion of *O* below *Every student* is “forced” by Hurford’s constraint. These predictions of truth-conditional effects are impossible for Gricean theories, which predict instead that implicatures never affect truth conditions.

To the extent that the truth conditions of these sentences are indeed those claimed by Chierchia et al., these data may provide important evidence for the grammatical generation of scalar implicatures. But it seems to me that these two examples push the limits of pragmatic competence—it is not clear to me that either of them is felicitous at all, so it is especially hard to judge their truth conditions. The evidence that truth conditions are indeed affected by a covert *O* seems to be based only on the authors’ intuitions about the sentences in isolation. Perhaps intuitions may be more reliably generated in conversational context: it seems that it is hard to accuse the speaker of such a sentence of falsehood when the local implicature is absent.<sup>19</sup>

<sup>18</sup> It is notable that the authors have here used a non-scalar (*the first and the second problem*) to “force” an “embedded” implicature. Again, the authors do not address the question of why Hurford’s Constraint violations ever occur, given the ability of *O* to apply to non-scalar terms.

<sup>19</sup> Even given these contexts, I, and many of my informants, have hazy intuitions about these examples. Perhaps this is because it is hard to come up with a context where a sentence like (5.46) is appropriately relevant.

(5.48) A: Did Peter do well on the exam?

B: ?He solved the first and the second problem or all of the problems.

C: #? False! He solved the first three problems!

Intuitions are similarly shaky, but leaning away from truth-conditional effects, with respect to questioning:

(5.49) A: Is it true that Peter solved the first and second problem or all of the problems?

a. B: ?I guess...? He solved the first three.

b. B: #? Certainly not. He solved the first three.

In addition, consider the following real-life example, from an article abstract:

(5.50) However, simultaneous inactivation of the first three or all five ubiquitination sites in gp18LP led to a massive increase in subviral particles released by these mutant glycoproteins that were readily detectable by electron microscopy analysis upon expression of the ubiquitination-deficient glycoprotein by itself or in a proviral context.

Is the sentence false if inactivation of the first four ubiquitination sites led to an increase in particles? It seems to me that it is not false, but simply misleading: why hasn't the author told us about the massive increase that followed simultaneous activation of the first four sites? Maybe the activation-of-the-fourth-site experimental data was compromised, so the speaker does not know whether subviral particles were released. The idea that the sentence is misleading in certain contexts broadly accords with the idea that the inference is pragmatic, not grammatical. In contrast, the Chierchia et al. position is that the logical form is forced by Hurford's Constraint to be *O(first three) ∨ all*, so the author of (5.50) is not just misleading if simultaneous activation of the first four sites led to an increase; he

is mistaken or lying.

It is worth pointing out that it would be no surprise, given a Gricean, cancellation-based theory, that (5.46) is weirder than “Peter solved some or all of the problems.” That prediction occurs simply because, in ordinary contexts, an utterance of *Peter solved both the first and the second problems* does not just imply that Peter did not solve all of the problems. Rather, the utterance implies that Peter solved the first and the second problems and no others. It is an odd conversational move, then, to cancel an implicature other than the one that is most likely to be generated. This prediction is consistent with the fact that most Google hits for sentences of the type *the first  $n$  or all* quantify over sets that have  $n + 1$  elements. For example, in an article about Shakespeare:

- (5.51) His first plays to be produced in London are thought to be either the first two or all three parts of Henry VI around 1590–92.

Here, the writer is plausibly cancelling the implicature that no more than the first two parts of Henry VI are thought to be Shakespeare’s first plays to be produced in London.<sup>20</sup>

<sup>20</sup>As further evidence, Google returns estimated hits as follows:

# of hits	phrase
33,900,000	“the first two or all three”
18,700,000	“the first two or all four”
4	“the first two or all five”
5	“the first two or all six”
5,010,000	“the first three or all four”
7	“the first three or all five”
5	“the first three or all six”
840,000	“the first four or all five”
348,000	“the first four or all six”
1	“the first four or all seven”

I do not want to make too much of these data, as the number of hits provided by Google are often unreliable estimates. Nonetheless, we see a strong preference against phrases like “the first two or all five”, plausibly explained by a preference for a second disjunct that cancels the implicature of the first disjunct. Note that “the first two or all four” and “the first four or all six” are OK, which would be expected as there are many domains where only the even numbers are possible.



### 5.3.4 The interaction of grammatical implicatures with Gricean clausal implicatures

The most compelling arguments in Chierchia et al. have to do with the interaction of clausal implicatures with entailment-related disjuncts. The authors argue that a grammatical theory, combined with Gricean scalar reasoning, provides a tidy explanation of *or both/or all*-style cancellation facts, while such predictions are out of reach for a purely Gricean theory.<sup>21</sup> Consider the following sentence:

(5.52) Peter solved some or all of the problems.

In this sentence, applying *O* to *some* has no direct truth conditional effect:

(5.53)  $\llbracket \text{some and not all} \rrbracket \cup \llbracket \text{all} \rrbracket = \llbracket \text{some} \rrbracket \cup \llbracket \text{all} \rrbracket = \llbracket \text{some} \rrbracket$

However, Chierchia et al. point out that the application of *O* does have an effect on Gricean reasoning about the sentence. Without the application of *O*, the sentence is synonymous with its first disjunct:  $\llbracket \text{some or all} \rrbracket = \llbracket \text{some} \rrbracket$ . But, when *O* is applied to the first disjunct, we have  $\llbracket O(\text{some}) \rrbracket \subset \llbracket O(\text{some}) \text{ or all} \rrbracket$ ; i.e., the first disjunct asymmetrically entails the utterance. Because of that, standard Gricean scalar reasoning can generate the usual four so-called clausal implicatures:

(5.54) Clausal implicatures with exhaustified disjuncts

- a.  $\neg BO(\text{some})$
- b.  $\neg Ball$
- c.  $\neg B\neg O(\text{some})$
- d.  $\neg B\neg all$

---

<sup>21</sup>Where Chierchia et al. use *A or B or both*-type examples, I replace them here with *some or all*-type examples, to avoid the confusion of piling up *ors*.

The last clausal implicature, listed in (5.54d), says that, for all the speaker knows, Peter solved all of the problems, effectively “cancelling” the epistemic strong scalar implicature that would have been associated with an utterance of “Peter solved some of the problems.” In effect, Chierchia et al. have combined a grammatical theory of scalar implicatures with a pragmatic theory of clausal implicatures to provide an account of *or*-style cancellation.

Let us pause for a moment to explicitly characterize the theory that Chierchia et al. have proposed for *or*-cancellation. The system is a hybrid: the grammar delivers upper-bounded meanings that have long been considered a product of rational inference, but that system of rational inference is still in place, and it delivers clausal implicatures. The hybrid nature of the system is necessitated by the fact that clausal implicatures are necessarily epistemic weak (they are inferences about which beliefs the speaker does not have). There is, therefore, no straightforward way to generate them in the grammar. For clarity, here is the two-stage pragmatic account of clausal implicatures.

(5.55) Clausal implicatures of a disjunction “*L* or *R*” (two-stage pragmatic)

- a.  $B(L \vee R)$  (speaker believes what he says; maxim of Quality)
- b.  $\neg BL$  (scalar reasoning, since  $L \subset L \vee R$ )
- c.  $\neg BR$  (scalar reasoning:  $R \subset L \vee R$ )
- d.  $\neg B\neg L$  (consequence of (5.55a) and (5.55c))
- e.  $\neg B\neg R$  (consequence of (5.55a) and (5.55b))

On this analysis, following Sauerland (2004) and Russell (2006), clausal implicatures (5.55b) and (5.55c) are just scalar implicatures: they represent utterances the speaker could have made that entail the speaker’s actual utterance and that are, in fact, easier to say (i.e., they are shorter).<sup>22</sup> Clausal implicatures (5.55d) and (5.55e) are derived immediately from these first two: if the speaker does not believe *L* is true, he must believe *R* is possible

---

<sup>22</sup>This is in contrast to *L and R* as an alternative to *L*, say, or to *R*: in this case, the speaker can’t utter the stronger *L and R* without extra effort.

(otherwise, since he believes  $L \vee R$  is true, he would have to believe  $L$  is true), deriving clausal implicature (5.55e) (clausal implicature (5.55d) is produced similarly).

To generate clausal implicatures for entailment-related disjuncts, Chierchia et al. simply adapt this reasoning to run on an utterance whose disjuncts contain grammatically-computed scalar implicatures. This grammatical + pragmatic account is spelled out as follows:<sup>23</sup>

- (5.56) Clausal implicatures of a disjunction “ $L$  or  $R$ ” (grammatical + pragmatic)
- a.  $B(O(L) \vee O(R))$  (what speaker has “said” includes upper-bounded disjuncts; maxim of Quality)
  - b.  $\neg BO(L)$  (scalar reasoning, since  $O(L) \subset (O(L) \vee O(R))$ )
  - c.  $\neg BO(R)$  (scalar reasoning, since  $O(R) \subset (O(L) \vee O(R))$ )
  - d.  $\neg B\neg O(L)$  (consequence of (5.56a) and (5.56c))
  - e.  $\neg B\neg O(R)$  (consequence of (5.56a) and (5.56b))

The fact that Chierchia et al. must invoke Gricean reasoning to account for clausal implicatures sheds some light on an unacknowledged feature of their grammatical system. In particular, it further highlights the necessity of free application of  $O$  to disjuncts, whether they are scalar items or not. In a disjunctive sentence  $L$  or  $R$ , consider the following two possibilities for the generation of the exclusive disjunction interpretation (the location of  $O$ ’s is given in the first line, with a gloss of the interpretation below):

- (5.57) a.  $O(L \vee R)$   
 b.  $(L \vee R) \wedge \neg(L \wedge R)$

- (5.58) a.  $O(L) \vee O(R)$

---

<sup>23</sup>This division of labor between grammar and pragmatics underscores the importance of ceasing to use the term “implicature” for grammatically-generated upper bounded meanings. Fox and Spector (2008), recognizing this fact, introduce the term “embedded exhaustification.”

$$\text{b. } (L \wedge \neg R) \vee (R \wedge \neg L)$$

These end up being truth-conditionally equivalent: both exclude worlds in the intersection of  $L$  and  $R$ . Intuitively, the former logical form matches more closely the motivation for grammatical implicature computation: it is the only one that depends on the *and/or* scale. But the system cannot generate clausal implicatures from this form, since  $L$  does not entail  $O(L \vee R)$ , except for the special case where  $L$  and  $R$  are disjoint sets (in which case the exclusive disjunction interpretation is trivial). In other words, to generate both exclusive *or* (grammatically) and clausal implicatures (non-grammatically), the theory depends on each disjunct being an alternative to the other, and on the insertion of  $O$  below *or*.

This is a setback for the Hurford's Constraint plus local exhaustification theory of the Hurford's generalization data. Because the simple application of the implicature-computing operator  $O$  to the disjunction creates an exclusive disjunction that is not, in general, entailed by its disjuncts, exclusive disjunction must be achieved by the application of  $O$  to each disjunct individually in order to generate clausal implicatures. Since clausal implicatures seem to be present in almost every ordinary example of disjunction, this application of  $O$  to individual disjuncts must be particularly free. Again, that underscores the puzzle, for grammatical theorists, of why Hurford's Constraint violations happen at all.

Moreover, the theory that  $O$  applies optionally and freely has problems of its own. As pointed out by Magri (2009), the following sequence is infelicitous.

(5.59) #John gave the same grade to all of his students. He gave some of them an A.

If  $O$  is a grammatical operator that applies freely and optionally, the sequence in (5.59) ought to be perfectly felicitous, given a parse where the second sentence is  $O$ -free. The fact that the sequence is not at all felicitous leads Magri to propose a system where  $O$  applies obligatorily, which leads to further considerations that I will not explore here.

In contrast, there is a straightforward Gricean story to tell about the sequence. First, it is only felicitous if both sentences are true. That means that it is only felicitous if John gave all of his students an A. In that case, the speaker could have uttered the stronger "...He gave all of them an A." A Gricean hearer would then search for a reason for the speaker to have chosen the weaker utterance, but such a reason is not readily obvious. The hearer is then left to wonder about the speaker's intent.

The Gricean story, then, predicts that if some reason for uttering a weak sentence is made available, the oddness of (5.59) should be ameliorated. One potential reason would be that the speaker is carrying out a logical deduction aloud. Suppose the interlocutors are trying to find out what grade John gave to Harry. On a table before them is the transcript of a tape recording in which John is heard to say, "As it happened, each of my students received exactly the same grade." Next to the transcript are three students' papers the investigators have obtained, and each of them is marked with an A. Gesturing to the transcript, the speaker says, "John gave the same grade to all of his students." Then, gesturing to the three graded papers, he says, "He gave some of them an A." He continues, "So, even though we do not have Harry's paper in evidence, we may conclude that Harry received an A." As a Gricean would predict, no contradictory scalar implicature arises: the speaker's reason for making a weak assertion is not that it is the strongest assertion he has evidence for, but that it is the strongest assertion supported by a particular piece of evidence: the stack of three graded papers. Combining those two bits of evidence, the speaker is able to conclude that Harry received an A.

### 5.3.5 Gricean clausal implicatures for entailment-related disjuncts

The reason the hybrid account of *or*-type cancellation is so compelling, despite its flaws, is that a purely pragmatic account of the cancellation data runs up a much more significant obstacle. The most natural pragmatic explanation of the *or both*-style cancellation facts, like the hybrid grammatical explanation, depends on clausal implicatures; in particular, it

depends on fourth clausal implicature in (5.55) above, instantiated in the *or all*-type case as  $\neg B\neg all$ . But this implicature cannot be derived through the intuitive two-stage pragmatic schema outlined in (5.55). That would require the derivation of the clausal implicature  $\neg Bsome$  (along the lines of (5.55b)), which contradicts the speaker's belief in the literal meaning of his utterance (Quality). The only non-trivial clausal implicature that can be derived for entailment-related disjuncts like *some or all* is  $\neg Ball$ : that the speaker does not believe *all* is true. ( $\neg B\neg some$  is trivial, since it is entailed by the Quality inference  $Bsome \vee all$ , which is equivalent to  $Bsome$ .)

This difficulty in deriving clausal implicature (5.55e) for entailment-related disjunctions is a real problem for the Gricean account. In the formalism developed in this dissertation, the problem is that, with respect to the alternatives under consideration, the belief states  $[o\omega]$  and  $[o\omega, \sigma]$  look exactly the same. The table in Figure 5.2 predicts a scalar

	$r_H(x)$	$s(x)$	$[o\omega]$	$[o\omega, \sigma]$	$[\sigma]$
" $\omega$ "	0.01	1	1	1	1
" $\sigma$ "	0.04	1	0	0	1
" $\omega \vee \sigma$ "	0.01	$10^{-4}$	1	1	1
$N()$			0.010001	0.010001	0.050001

Figure 5.2: Equivalence of  $[o\omega]$  and  $[o\omega, \sigma]$ .

effect of roughly 1 : 1 : 5 for  $[o\omega]$ ,  $[o\omega, \sigma]$ , and  $[\sigma]$ , respectively. What this means is that, although a  $\neg B\sigma$  implicature is predicted, scalar reasoning does not differentiate between  $[o\omega]$  and  $[o\omega, \sigma]$ . The desired behavior, of course, would be for scalar reasoning to favor  $[o\omega, \sigma]$ : an ignorance inference.

Before considering possible solutions, note that this is not, strictly speaking, a case of cancelling a scalar implicature, or preventing one from arising. I argued in Chapter 4 that the *scalar* implicature is just the epistemic weak inference that  $[o\omega]$  or  $[o\omega, \sigma]$  is true and  $[\sigma]$  is false. What needs explaining in this case is the fact that this weak inference is not strengthened to  $[o\omega]$ . I argued in Chapter 4 that such strengthening happens due to a prior

belief that the speaker is opinionated. To explain the *or both*-type cancellation facts, then, the theory must find a way for the utterance of *some or all* to negate the prior belief that the speaker is opinionated.

In a probabilistic setting, the difference between epistemic weak and strong is one of degree. But probability alone does not seem a sufficient tool to solve the *or* cancellation conundrum. This is because *some* and *some or all* are semantically equivalent. Notice that  $r(\omega) = r(\omega \vee \sigma)$  and  $P(B\omega|\mathcal{B}) = P(B(\omega \vee \sigma)|\mathcal{B})$  for all  $\mathcal{B}$ , since  $\omega$  and  $\omega \vee \sigma$  are the same propositions. Moreover, since the two sentences have the same denotation, their relevance is also indistinguishable. Therefore, since neither Quality- or Relevance-based considerations can differentiate between the alternatives  $\omega$  and  $\omega \vee \sigma$ , the only thing left is simplicity:  $\omega$  is significantly simpler than  $\omega \vee \sigma$ .

This means that “some or all” can only be an overall worse choice than “some” (since it is semantically equivalent yet longer). So the explanation for the speaker’s failure to use  $\omega$  or  $\sigma$  cannot be due to the fact that neither is good enough in a given belief state; after all, the speaker chose something that is worse, *a priori*, than one of those possibilities. This suggests a potential solution to the problem: the interpretation of  $\ulcorner \omega \vee \sigma \urcorner$  is the result of a hearer trying to figure out why a speaker uttered something *a priori* worse than a clearly available alternative. Maybe the speaker is flouting conversational expectations precisely because he does not want to make the implicatures he would make by uttering “some”.

The theory of *M*-implicatures (term due to Levinson 2000, with the phenomenon discussed much earlier in McCawley 1978 and Horn 1984, and referred to under the term “division of pragmatic labor” in Horn 1984) proposed in Franke (2009) may suggest a game-theoretic way of implementing the idea.<sup>24</sup> The broad generalization behind *M*-implicatures is that more marked forms tend to be associated with less stereotypical interpretations: McCawley’s example was that an utterance of “Black Bart caused the sheriff

<sup>24</sup>As Franke acknowledges, this analysis of *M*-implicatures is closely related to the bidirectional optimality theory analysis first developed in Blutner (1998).

to die” implicates that Black Bart did not kill the sheriff in a stereotypical way (otherwise, the speaker would have said “Black Bart killed the sheriff”). For the cancellation with *or* case, this reasoning would say that the more marked form “some or all” is associated with the less stereotypical meaning  $[o\omega, \sigma]$ .

To see how this plays out, assume, as above, that the set of alternatives is  $ALT = \{\omega, \omega \vee \sigma, \sigma\}$ , and speakers, *ceteris paribus*, prefer simpler utterances (in Franke’s terms, speakers are sensitive to message cost). Then a Level-0 speaker should behave as follows:

$$[o\omega] \mapsto \ulcorner \omega \urcorner \quad (5.60)$$

$$[o\omega, \sigma] \mapsto \ulcorner \omega \urcorner \quad (5.61)$$

$$[\sigma] \mapsto \ulcorner \sigma \urcorner, \ulcorner \omega \urcorner \quad (5.62)$$

A Level-1 hearer, in response, would interpret  $\ulcorner \omega \urcorner$  according to the relative priors for  $[o\omega]$  and  $[o\omega, \sigma]$ , mapping it to whichever is higher, and would interpret  $\ulcorner \sigma \urcorner$  as  $[\sigma]$ . But the Level-1 hearer would not know what to do with  $\ulcorner \omega \vee \sigma \urcorner$ , since a Level-0 speaker doesn’t produce it in any of the three states. In Franke’s terms,  $\ulcorner \omega \vee \sigma \urcorner$  is a *surprise* message.

I will not go into the details developed by Franke for the interpretation of surprise messages here. In broad terms, Franke’s system assigns the simplest utterance to the belief state with the highest prior probability and assigns more complex utterances, one by one, to progressively lower probability states.

That means the system maps everything out correctly when  $P([o\omega]) > P([o\omega, \sigma])$ . But when  $P([o\omega, \sigma])$  starts out with a higher probability, the system reverses: it predicts “some or all” will be mapped to  $[o\omega]$ . This doesn’t happen.

(5.63) I am no expert on these machines, so bear with me. When I descaled the machine it made the seals in the pistons feel squeaky. I opened the back of the machine so that I could get a pretty good, but not great, view of the pistons.



- a. The descaling had removed the lubricant from some of the pistons.
- b. The descaling had removed the lubricant from some or all of the pistons.

In the example, the context sets up the prior probabilities so that the speaker is not likely to be fully informed about the state of all of the pistons—he only got a pretty good view of them—so  $P([o\omega]) < P([o\omega, \sigma])$ . That means that Franke’s surprise message story predicts (5.63a) is mapped to  $[o\omega, \sigma]$  and (5.63b) is mapped to  $[o\omega]$ . The first prediction is somewhat plausible: the context made it unlikely for the speaker to have a strong belief, so the hearer infers that the speaker’s belief is weak. But the second prediction is implausible: a rational hearer would never interpret this to mean that the speaker is sure that some and not all of the pistons had been stripped.

Moreover, there is an apparent problem with any story that depends on the relative proximity of “some or all” to elevate  $[o\omega, \sigma]$  over  $[o\omega]$ . Such a story depends on the speaker failing to utter “some”, and choosing something longer but semantically equivalent instead. But it does not in any way rely on the fact that the speaker chose to lengthen his utterance by adding “or all”. There are other innumerable possible utterances that have this same property of being longer than *some* but adding no extra information, like “some or some”, “some or most”, “some and not none”, etc. But none of these utterances may be used to indicate that the speaker believes *all* is possible. In particular, the example *some or most* demonstrates that a perfectly felicitous utterance may have the property of being longer than *some*, be synonymous with *some*, and still fail to imply that, for all the speaker believes, *all* is possible.

These considerations leave the pragmatic view with an unsolved puzzle. Between Quality, Relevance, and Manner (specifically, complexity), the only available story to differentiate “some” from “some or all” is to say “some or all” implicates  $[o\omega, \sigma]$  because it is synonymous with, but more complex than, *some*. The existence of numerous other expressions in English that share this property but do not implicate  $[o\omega, \sigma]$  weakens this

story considerably. There are conceivable other stories to pursue within the pragmatic approach, including conventional implicature and dynamic semantics possibilities, but those will have to be left for future research.

### 5.3.6 More on the hybrid grammatical + pragmatic system

It is worthwhile to step back to consider nature of the system that Chierchia et al. have devised to model implicature cancellation with *or*. The system uses a local, grammatical mechanism to narrow the meaning of a weak disjunct, then generates global, epistemic weak implicatures based on the narrowed semantics of the disjuncts. The resulting system has interesting redundancies. The propositional content of speaker's utterance "Some of them left" is ambiguous between *some and not all left* and *some left*; in (at least) the latter case, the speaker has the stronger alternative *all left*, so the speaker, using pragmatic, scalar reasoning, can get at least halfway to the propositional content of the first, grammatically-generated interpretation. In a context that assumes speaker opinionatedness (and Chierchia et al. must acknowledge that such contexts do, in fact, exist), that implicature will be strengthened, so that the speaker has asserted *some left*, but has conveyed that he believes that *some and not all left*. Given the ambiguity inherent in the grammatical system, such contexts would favor disambiguation of the speaker's utterance as *some and not all left*. But that means that the computation of grammatical implicatures (or disambiguation between the presence or absence of *O*) depends on Gricean computation of just the same implicatures.

Let us consider a more complex and subtle case where the grammatical system has two separate mechanisms for generating implicatures that would, pre-theoretically, seem to be unified phenomena. The authors point out that, in a sentence where *some* is embedded below a modal like *required*, like (5.64), adding *or all* makes the utterance incompatible with a situation where we are not allowed to read both.

(5.64) We are required to read some or all of the books on the list.

The grammatical theory has a straightforward, purely grammatical analysis of this fact: the sentence's logical form is  $O(\Box(O(\text{some}) \vee (\text{all})))$ . The matrix-level  $O$  negates alternatives  $\Box(O(\text{some}))$  and  $\Box(\text{all})$ —in other words, we are not required to read just some, and we are not required to read all. This logically entails that we are allowed to read all (if we weren't, the sentence couldn't be true unless we were required to read just one or the other).

Notice how closely this echoes the reasoning in the matrix clausal implicature case.

(5.65) John read some or all of the books on the list.

In this case, an utterance of “some or all” leads to an implicature of  $\neg B\neg all$ , as outlined in (5.54). In the modal case, utterance of “We are required to read some or all” produces an implicature of  $\neg\Box\neg all$ . Moreover, the logical steps taken to get to both interpretations are essentially the same. But, according to the Chierchia et al. theory, the second inference is dependent upon Gricean reasoning, and the first is not. To spell out the similarity in logic, but difference in source, of the two types of inference, consider the two derivations within the grammatical system, represented schematically:

(5.66) Derivation of epistemic possibility clausal implicatures in a grammatical system:

- a.  $O(\text{some}) \vee all$  (logical form of speaker's utterance)
- b.  $B(O(\text{some} \vee all))$  (Gricean Quality)
- c.  $\neg BO(\text{some}) \wedge \neg Ball$  (scalar reasoning for primary clausal implicatures)
- d.  $\neg B\neg O(\text{some}) \wedge \neg B\neg all$  (logical consequence of (5.66b) and (5.66c))

(5.67) Derivation of disjunction-under-modal permission implicatures in a grammati-

cal system:

- a.  $O\Box(O(L) \vee R)$  (logical form of speaker's utterance)
- b.  $\neg\Box O(L) \wedge \neg\Box R$  (application of widest scope  $O$ )
- c.  $\neg\Box\neg O(L) \wedge \neg\Box\neg R$  (logical deduction)

To summarize: grammatical theories cannot straightforwardly generate clausal implicatures in the grammar, due to the epistemic component of such implicatures. Grammatical theorists, therefore, resort to traditional Gricean scalar reasoning to generate clausal implicatures (see section 5.3.4 for discussion). But, because modals and speaker beliefs both involve quantification over possible worlds, the negation of a modal by application of  $O$  and negation of a belief through Gricean reasoning lead to largely parallel implicatures. If the grammatical theory is right, then, it has the surprising consequence of making two ostensibly similar inferences the result of altogether different processes: one pragmatic, the other purely semantic.

## 5.4 Conclusion

In this chapter, I have examined two recent arguments in favor of grammatical computation of scalar implicatures. The first argument, extended by Sharvit and Gajewski (2008), is that propositional attitude verbs do not have the neg-raising properties needed to generate “embedded” scalar implicature phenomena. This argument, I suggested, has a deep empirical shortcoming: “embedded” implicature phenomena are, in fact, affected by the semantic strength of a propositional attitude (which is, in turn, tied to the neg-raising property). In particular, I argued that “embedded” implicatures are derived more weakly for strong propositional attitudes like *certain* than for weaker propositional attitudes like *believe*. This, in turn, provides evidence in favor of Gricean computation (especially Gricean computation that is sensitive to degree of speaker belief, like the Bayesian system devel-

oped in Chapter 4).

The second argument addressed in this chapter is the argument made by Chierchia et al. (to appear) that *some or all*-type cancellation has a straightforward explanation in a grammatical system while it remains a mystery to a Gricean system. This argument carries more weight: due to the identical propositional content of *some* and *some or all*, a Gricean system cannot generate a clausal implicature of *the speaker believes it is possible that all*. I suggested that higher-order reasoning (along game-theoretic lines) could provide an explanation similar to the analysis of *M*-implicatures, or “division of pragmatic labor” (Horn, 1984) phenomena (though I noted that such an approach would have some obstacles related to prior probabilities).

But I also noted that the grammatical approach to *or* cancellation has a number of obstacles and peculiarities: most importantly, it does not provide an explanation, or even an account, of Hurford-type *France or Paris* violations. Under the embedded exhaustification analysis, disjuncts have to be able to be exhaustified regularly, and there is no reason to exclude the possibility of exhaustification on disjunctions like *France or Paris*. In addition, the grammatical theory requires Gricean generation of clausal implicatures (which are a variety of scalar implicature, under most modern Gricean theories), and this creates an unexpected and, it seems to me, unparsimonious distinction between matrix-level clausal implicatures and “embedded” implicatures below exhaustified modals.

## Chapter 6

### Concluding remarks

Grice, in 1967, laid out the groundwork for an approach to pragmatic analysis: if an aspect of an utterance's meaning follows from conditions on its assertion, that aspect of meaning can be calculated as a conversational implicature. Grice's theory has allowed semanticists to keep the grammar relatively simple for the past 40 years: in particular, meanings for weak scalar words like *some* and *warm* and *possible* and *attractive* do not exclude *all*, *hot*, *certain*, or *beautiful*. Instead, these upper-bounded meanings are derived through conversational reasoning: if a speaker believed the corresponding assertion with a stronger term applied, the speaker would have made that assertion.

But Grice's theory has little use for linguists if it does not make clear predictions about the interpretation of utterances containing scalar terms. That, presumably, is why linguists immediately set about the task of formalizing Grice's theory. These formalizations of the theory used various tools: simple grammar-like rule systems that operate on sentences; logical systems for rational inference about alternatives; optimization procedures that pair weak scalar forms with upper-bounded meanings; and game theoretic models that incorporate back-and-forth speaker and hearer reasoning to pair weak scalar forms with upper bounded speaker belief states.

In this thesis, I advocated the idea that Gricean scalar reasoning—that the speaker

might be expected to make a stronger assertion if he had a belief that supported it—is essentially a probabilistic inference task for the hearer: given the observation that the speaker made a weak assertion, what is the probability that the speaker has a strong belief? A theory of scalar implicature, I argued, should predict that this probability is low. A key advantage of probability theory is that calculating this probability (and thereby deriving a scalar implicature) can be carried out in terms of a related, more intuitively simple probability: given a speaker’s belief state, how likely is he to choose a particular assertion? A hearer’s estimation of this probability is his subjective theory of Utterance Expectation.

The calculation of scalar implicature, then, is a matter of setting out a theory of Utterance Expectation—that is, a theory of hearers’ expectations about speakers’ utterances. In this thesis, I argued that those expectations are the result of three factors: belief, relevance, and simplicity (closely tied to Grice’s maxims of Quality, Quantity/Relation, and Manner). With simple measures of these three factors, Utterance Expectation assigns a higher probability to utterances that the speaker believes, that are more relevant, and that are simpler. Applying Bayes’ Rule to Utterance Expectation values makes it possible to derive likelihood for various speaker beliefs, given the observation of an utterance. The derivation of simple scalar implicatures is immediate: the availability of a stronger alternative in a stronger belief state makes it less likely that the speaker will choose that alternative.

But it is also a fairly immediate result that scalar reasoning does not distinguish between ignorance (for all the speaker knows, the stronger alternative is true) and epistemic strong (the speaker believes the stronger alternative is false) implicatures. What the theory predicts, instead, is that prior probabilities of ignorance and epistemic strong belief states adjudicate between these states “after” the scalar implicature is derived. This prediction is in line with informal arguments in the literature that scalar reasoning only permits the derivation of an epistemic weak implicature, which is only strengthened when a background assumption exists that the speaker is opinionated.

This analysis of the epistemic status of implicatures is mirrored in problematic “em-

bedded” implicature cases. For these utterances, the probabilistic system again predicts no scalar effect to distinguish the “embedded” implicature from the global scalar implicature. Instead, these interpretations are purely a consequence of the ratio of their prior probabilities. This makes a key prediction: the robustness of an “embedded” scalar implicature is, *ceteris paribus*, inversely related to the semantic strength of the predicate that the scalar term is embedded below. Empirical evidence supports this prediction: “Harry is absolutely certain that some of the students will be late” has a weaker “embedded” implicature than “Harry thinks it is pretty likely that some of the students will be late.”

The probabilistic framework developed in this thesis is especially well-suited to make these types of predictions about the magnitude of an implicature. In contrast, previous theories generate categorical, rather than probabilistic, scalar implicatures, and they are therefore not equipped to make predictions about the strength of implicatures. Further investigation should involve experimental measurement of implicature magnitude. This should shed light on both the nature of Utterance Expectation (e.g., how important is simplicity?), and on the nature of scalar inferences themselves (through, e.g., a comparison with judgments about ambiguous sentences).



# Bibliography

- Aumann, R. J., 1976. Agreeing to disagree. *Annals of Statistics* 4 (6), 1236–1239.
- Austin, J. L., 1956. Ifs and cans. *Proceedings of the British Academy* 42, 109–132.
- Bach, K., 1999. The myth of conventional implicature. *Linguistics and Philosophy* 22 (4), 327–366.
- Bartsch, R., 1973. Negative transportation gibt es nicht. *Linguistische Berichte* 27 (7).
- Barwise, J., Cooper, R., 1981. Generalized quantifiers and natural language. *Linguistics and Philosophy* 4 (2), 159–219.
- Benz, A., 2007. On relevance scale approaches. In: Puig-Waldmüller, E. (Ed.), *Proceedings of Sinn und Bedeutung* 11. Universitat Pompeu Fabra, Barcelona, pp. 91–105.
- Benz, A., Van Rooij, R., 2007. Optimal assertions, and what they implicate. a uniform game theoretic approach. *Topoi* 26 (1), 63–78.
- Bierwisch, M., Lang, E., 1989. *Dimensional Adjectives: Grammatical Structure and Conceptual Interpretation*. Springer-Verlag, Berlin.
- Block, E., 2008. Is the symmetry problem really a problem?, unpublished manuscript, NYU.
- Blutner, R., 1998. Lexical pragmatics. *Journal of Semantics* 15 (2), 115–162.

- Bultinck, B., 2004. *Numerous Meanings: the meaning of English cardinals and the legacy of Paul Grice*. Elsevier Science, Amsterdam.
- Büring, D., 2003. On d-trees, beans, and b-accent. *Linguistics and Philosophy* 26 (5), 511–545.
- Carey, S., 1985. *Conceptual change in childhood*. MIT Press, Cambridge.
- Carnap, R., 1950. *Logical Foundations of Probability*. University of Chicago Press, Chicago.
- Carston, R., 1996. Metalinguistic negation and echoic use. *Journal of pragmatics* 25 (3), 309–330.
- Chemla, E., 2008. An epistemic step for anti-presuppositions. *Journal of Semantics* 25 (2), 141–173.
- Chemla, E., Spector, B., 2011. Experimental evidence for embedded scalar implicatures. *Journal of semantics* 28 (3), 359–400.
- Chierchia, G., 2004. Scalar implicatures, polarity phenomena, and the syntax/pragmatics interface. In: Belletti, A. (Ed.), *Structures and beyond*. Vol. 3 of *The Cartography of Syntactic Structures*. Oxford University Press, Oxford, pp. 39–103.
- Chierchia, G., Fox, D., Spector, B., to appear. The grammatical view of scalar implicatures and the relationship between semantics and pragmatics. In: Maienborn, C., von Stechow, K., Portner, P. (Eds.), *Semantics: An International Handbook of Natural Language Meaning*. Mouton de Gruyter.
- Chomsky, N., 1957. *Syntactic Structures*. Mouton, The Hague.
- Chomsky, N., 1972. Some empirical issues in the theory of transformational grammar. In: Peters, P. S. (Ed.), *Goals of Linguistic Theory*. Prentice-Hall, Englewood Cliffs, NJ.

- De Finetti, B., 1964. Foresight: Its logical laws, its subjective sources. In: Kyburg, Jr., H. E., Smokler, H. E. (Eds.), *Studies in Subjective Probability*. John Wiley and Sons, New York, pp. 93–158, originally published in 1937 under the title ‘*La Prevision: Ses Lois Logiques, Ses Sources Subjectives*’.
- Ducrot, O., 1973. *La preuve et le dire*. Mame, Paris.
- Dummett, M. A., 1973. *Frege: Philosophy of Language*. Duckworth, London.
- Fauconnier, G., 1975. Pragmatic scales and logical structure. *Linguistic Inquiry* 6 (3), 353–375.
- Fox, D., 2007. Free choice disjunction and the theory of scalar implicature. In: Sauerland, U., Stateva, P. (Eds.), *Presupposition and Implicature in Compositional Semantics*. Palgrave Macmillan, New York, pp. 71–120.
- Fox, D., Hackl, M., 2006. The universal density of measurement. *Linguistics and Philosophy* 29 (5), 537–586.
- Fox, D., Spector, B., 2008. Economy and embedded exhaustification, handout from a talk given at MIT.
- Franke, M., 2009. Signal to act: Game theory in pragmatics. Ph.D. thesis, Universiteit van Amsterdam.
- Franke, M., 2011. Quantity implicatures, exhaustive interpretation, and rational conversation. *Semantics and Pragmatics* 4 (1), 1–82.
- Gajewski, J., Sharvit, Y., to appear. In defense of the grammatical approach to local implicatures. *Natural Language Semantics*.
- Gauker, C., 2001. Situated inference versus conversational implicature. *Noûs* 35 (2), 163–189.

- Gazdar, G., 1979. *Pragmatics: Implicature, Presupposition, and Logical Form*. Academic Press, New York.
- Gazdar, G., Good, D., 1982. On a notion of relevance. In: Smith, N. (Ed.), *Mutual Knowledge*. Academic Press, London, pp. 88–100.
- Gettier, E., 1963. Is justified true belief knowledge? *Analysis* 23 (6), 121–123.
- Geurts, B., 1998. The mechanisms of denial. *Language* 74 (2), 274–307.
- Geurts, B., 2005. Entertaining alternatives: disjunctions as modals. *Natural Language Semantics* 13 (4), 383–410.
- Geurts, B., 2006. Take “five”: the meaning and use of a number word. In: Vogeeler, S., Tasmowski, L. (Eds.), *Non-definiteness and plurality*. Benjamins, Philadelphia, pp. 311–329.
- Geurts, B., 2009. Scalar implicature and local pragmatics. *Mind and Language* 24, 51–79.
- Geurts, B., 2011. *Quantity implicatures*. Cambridge University Press.
- Ginzburg, J., 1996. Dynamics and the semantics of dialogue. In: Seligman, J., Westerstahl, D. (Eds.), *Logic, language and computation*. Vol. 1. Center for the Study of Language and Information, Stanford, CA.
- Gopnik, A., Meltzoff, A., 1997. *Words, thoughts, and theories*. MIT Press.
- Green, M., 1995. Quantity, volubility, and some varieties of discourse. *Linguistics and Philosophy* 18 (1), 83–112.
- Groenendijk, J., Stokhof, M., 1984. On the semantics of questions and the pragmatics of answers. In: Landman, F., Veltman, F. (Eds.), *Varieties of Formal Semantics*. Foris Publications, Dordrecht, pp. 143–170.

- Harnish, R., 1976. Logical form and implicature. In: Bever, T., Katz, J., Langendoen, D. (Eds.), *An integrated theory of linguistic ability*. Crowell, pp. 313–391.
- Heim, I., 1991. Artikel und definitheit. In: von Stechow, A., Wunderlich, D. (Eds.), *Semantik: ein internationales Handbuch der Zeitgenössischen forschung*. de Gruyter, Berlin, pp. 487–535.
- Hirschberg, J., 1991. *A Theory of Scalar Implicature*. Garland Publishing Company, New York.
- Horn, L. R., 1972. On the semantic properties of the logical operators in English. Ph.D. dissertation, University of California at Los Angeles, Los Angeles, California.
- Horn, L. R., 1978. Remarks on neg-raising. In: Cole, P. (Ed.), *Pragmatics*. Vol. 9 of *Syntax and Semantics*. Academic Press, New York, pp. 129–220.
- Horn, L. R., 1984. Toward a new taxonomy for pragmatic inference: Q-based and R-based implicature. In: Schiffrin, D. (Ed.), *Meaning, form, and use in context: Linguistic Applications*. Georgetown University Press, Washington, pp. 11–42.
- Horn, L. R., 1985. Metalinguistic negation and pragmatic ambiguity. *Language* 61 (1), 121–174.
- Horn, L. R., 1989. *A Natural History of Negation*. Chicago University Press, Chicago.
- Horn, L. R., 1992. The said and the unsaid. In: *SALT II: proceedings from the second conference on Semantics and Linguistic Theory*. Ohio State University, Columbus.
- Horn, L. R., 2002. Assertoric inertia and NPI licensing. In: Adronis, M., Debenport, E., Pycha, A., Yoshimura, K. (Eds.), *CLS 38-2: The Panels*. Chicago Linguistic Society, pp. 55–82.
- Horn, L. R., 2004. Implicature. In: Horn, L. R., Ward, G. (Eds.), *The handbook of pragmatics*. Blackwell, pp. 3–28.

- Horn, L. R., 2008. On f-implicature: Myth-analysis and rehabilitation, paper presented at the University of Michigan Fall 2008 Workshop in Philosophy and Linguistics.
- Hurford, J., 1974. Exclusive or inclusive disjunction. *Foundations of Language* 11, 409–411.
- Ippolito, M., March 2010. Embedded implicatures? remarks on the debate between globalist and localist theories. *Semantics and Pragmatics* 3 (5), 1–15.
- Jackendoff, R., 1972. *Semantic Interpretation in Generative Grammar*. The MIT Press, Cambridge, Massachusetts.
- Jäger, G., 2007. Game dynamics connects semantics and pragmatics. In: Pietarinen, A.-V. (Ed.), *Game Theory and Linguistic Meaning*. Elsevier, Oxford, pp. 103–118.
- Kahneman, D., Tversky, A., 1984. Choices, values, and frames. *American psychologist* 39 (4), 341.
- Kamp, H., 1975. Two theories about adjectives. In: Keenan, E. L. (Ed.), *Formal Semantics of Natural Language*. Cambridge University Press, Cambridge, England, pp. 123–155.
- Katzir, R., 2007. Structurally-defined alternatives. *Linguistics and Philosophy* 30, 669–690.
- Kempson, R., 1980. Ambiguity and word meaning. In: Greenbaum, S., Leech, G., Svartvik, J. (Eds.), *Studies in English Linguistics*. Longman, London/New York, pp. 7–16.
- König, E., 1990. Concessive connectives and concessive sentences: cross-linguistic regularities and pragmatic principles. In: Hawkins, J. A. (Ed.), *Explaining Language Universals*. Oxford: Blackwell. Blackwell, pp. 145–166.
- Kroch, A., 1972. Lexical and inferred meaning for some time adverbs. *Quarterly Progress Report of the Research Laboratory of Electronics* 104.

- Kyburg, H., 1978. Subjective probability: criticisms, reflections, and problems. *Journal of Philosophical Logic* 7 (1), 157–180.
- Lassiter, D., 2010. Why symmetry is not a problem for a Gricean theory of scalar implicature, handout from a talk given at Utterance Interpretation and Cognitive Models III in Brussels.
- Levinson, S. C., 1983. *Pragmatics*. Cambridge University Press, Cambridge, England.
- Levinson, S. C., 1989. Review of *Relevance*, by Dan Sperber and Deirdre Wilson. *Journal of Linguistics* 25, 455–472.
- Levinson, S. C., 2000. *Presumptive Meanings*. The MIT Press, Cambridge, Massachusetts.
- Lewis, D., 1969. *Convention: A philosophical study*. Harvard University Press.
- Lewis, D., 1979. Scorekeeping in a language game. *Journal of philosophical logic* 8 (1), 339–359.
- Löbner, S., 1987. Quantification as a major module of natural language semantics. In: Groenendijk, J., de Jongh, D., Stokhof, M. (Eds.), *Studies in discourse representation theory and the theory of generalized quantifiers*. Foris, pp. 53–85.
- Magri, G., 2009. A theory of individual-level predicates based on blind mandatory scalar implicatures. *Natural Language Semantics* 17 (3), 245–297.
- Matsumoto, Y., 1995. The conversational condition on Horn scales. *Linguistics and Philosophy* 18 (1), 21–60.
- McCarthy, J., 1980. Circumscription—a form of non-monotonic reasoning. *Artificial intelligence* 13 (1-2), 27–39.
- McCawley, J., 1978. Conversational implicature and the lexicon. In: Cole, P. (Ed.), *Pragmatics*. Vol. 9 of *Syntax and Semantics*. Academic Press, pp. 245–59.

- McCawley, J., 1991. Contrastive negation and metalinguistic negation. In: Dobrin, L., Nichols, L., Rodríguez, R. M. (Eds.), *Chicago Linguistic Society*. Vol. 27. pp. 189–206.
- Merin, A., 1999. Information, relevance, and social decisionmaking: Some principles and results of decision-theoretic semantics. In: Moss, L. S., Ginzburg, J., de Rijke, M. (Eds.), *Logic, Language and Computation*. Vol. 2. CSLI Publications, Stanford, California, pp. 179–221.
- Murphy, G., Medin, D., 1985. The role of theories in conceptual coherence. *Psychological review* 92 (3), 289.
- Parikh, P., 1991. Communication and strategic inference. *Linguistics and Philosophy* 14, 473–513.
- Pearl, J., 2000. *Causality: Models, Reasoning, and Inference*. Cambridge University Press, Cambridge, England.
- Percus, O., 2006. Antipresuppositions. In: Ueyama, A. (Ed.), *Theoretical and empirical studies of reference and anaphora: Toward the establishment of generative grammar as an empirical science*. Japan Society for the Promotion of Science, pp. 52–73, report of the Grant-in-Aid for Scientific Research (B), Project No. 15320052.
- Pereira, F., 2000. Formal grammar and information theory: together again? *Philosophical Transactions of the Royal Society of London. Series A: Mathematical, Physical and Engineering Sciences* 358 (1769), 1239–1253.
- Potts, C., 2005. *The Logic of Conventional Implicatures*. Oxford Studies in Theoretical Linguistics. Oxford University Press, Oxford.
- Potts, C., 2006. Conversational implicatures via general pragmatic pressures. In: McCready, E. (Ed.), *Preproceedings of Logic Engineering and Natural Language Semantics*. Japanese Society for Artificial Intelligence, Tokyo, pp. 65–79.



- Ramsey, F. P., 1931. Truth and probability. In: Braithwaite, R. B. (Ed.), *The Foundations of Mathematics: Collected Papers of Frank P. Ramsey*. Routledge and Kegan Paul, London, pp. 156–198.
- Roberts, C., 1996. Information structure in discourse: Towards an integrated formal theory of pragmatics. In: Yoon, J.-H., Kathol, A. (Eds.), *OSU Working Papers in Linguistics*. Ohio State University, pp. 91–136.
- Rohdenburg, G., 1985. Dogs, bitches and other creatures. *Journal of Semantics* 4 (2), 117–135.
- Ross, I., 2006. Games interlocutors play: New adventures in compositionality and conversational implicature. Ph.D. thesis, University of Pennsylvania.
- Russell, B., 1948. *Human knowledge: Its scope and its limits*. New York: Simon & Schuster.
- Russell, B., 2006. Against grammatical computation of scalar implicatures. *Journal of Semantics* 23 (4), 361–382.
- Sauerland, U., 2004. Scalar implicatures in complex sentences. *Linguistics and Philosophy* 27 (3), 363–391.
- Sauerland, U., 2005. The epistemic step, paper presented at Experimental Pragmatics, Cambridge University, Cambridge, UK.
- Sauerland, U., 2008. Implicated presuppositions. In: Steube, A. (Ed.), *Sentence and Context*. Mouton de Gruyter, Berlin.
- Schlenker, P., 2006. Maximize presupposition and Gricean reasoning. Manuscript, UCLA and Institute Jean-Nicod.
- Schulz, K., van Rooij, R., 2006. Pragmatic meaning and non-monotonic reasoning: The case of exhaustive interpretation. *Linguistics and Philosophy* 29, 205–250.

- Searle, J., 1966. Assertions and aberrations. In: Williams, B. O., Montefiore, A. (Eds.), *British Analytical Philosophy*. Routledge and Kegan Paul, London, pp. 44–54.
- Seidenfeld, T., Schervish, M., 1983. A conflict between finite additivity and avoiding dutch book. *Philosophy of Science* 50 (3), 398–412.
- Sharvit, Y., Gajewski, J., 2008. On the calculation of local implicatures. In: Chang, C., Haynie, H. (Eds.), *Proceedings of the 26th West Coast Conference on Formal Linguistics*. Cascadia Press, pp. 411–419.
- Shirazi, A., Amir, E., 2007. Probabilistic modal logic. In: Holte, R., Howe, A. (Eds.), *Proceedings of the 22nd AAAI Conference on Artificial Intelligence*. Vol. 22. pp. 489–495.
- Singh, R., 2010. Oddness and ignorance inferences, handout of talk presented at MOSAIC 2, McGill University, Montreal.
- Singh, R., 2011. Maximize presupposition! and local contexts. *Natural Language Semantics* 19 (2), 149–168.
- Snow, P., 2001. The reasonableness of possibility from the perspective of Cox. *Computational Intelligence* 17 (1), 178–192.
- Soames, S., 1982. How presuppositions are inherited: A solution to the projection problem. *Linguistic Inquiry* 13, 483–545.
- Spector, B., 2007a. Aspects of the pragmatics of plural morphology: On higher-order implicatures. In: Sauerland, U., Stateva, P. (Eds.), *Presupposition and Implicature in Compositional Semantics*. Palgrave-Macmillan.
- Spector, B., 2007b. Scalar implicatures: Exhaustivity and Gricean reasoning. In: Aloni, M., Butler, A., Dekker, P. (Eds.), *Questions in Dynamic Semantics*. Elsevier, Oxford, pp. 225–250.

- Sperber, D., Wilson, D., 1986. *Relevance*. Harvard University Press, Cambridge, Massachusetts.
- Stalnaker, R., 1978. Assertion. In: Cole, P. (Ed.), *Pragmatics*. Vol. 9 of *Syntax and Semantics*. Academic Press.
- Stalnaker, R., 2002. Common ground. *Linguistics and Philosophy* 25 (5), 701–721.
- Strawson, P., 1950. On referring. *Mind* 59 (235), 320–344.
- Tenenbaum, J., Griffiths, T., 2001. Generalization, similarity, and Bayesian inference. *Behavioral and Brain Sciences* 24 (04), 629–640.
- Tenenbaum, J., Kemp, C., Shafto, P., 2007. Theory-based Bayesian models of inductive reasoning. In: Feeney, A., Heit, E. (Eds.), *Inductive reasoning: Experimental, developmental, and computational approaches*. Cambridge University Press, pp. 167–204.
- Thomason, R., 1990. Accommodation, meaning, and implicature: Interdisciplinary foundations for pragmatics. In: Cohen, P. R., Morgan, J., Pollack, M. (Eds.), *Intentions in communication*. MIT Press.
- Tokarz, M., 1994. Non-axiomatizability of Grice's implicature. *Studia Logica* 53 (2), 343–349.
- van Kuppevelt, J., 1996. Inferring from topics. *Linguistics and Philosophy* 19 (4), 393–443.
- van Rooij, R., Schulz, K., 2004. Exhaustive interpretation of complex sentences. *Journal of Logic, Language and Information* 13, 491–519.
- van Rooij, R., 2003. Questioning to resolve decision problems. *Linguistics and Philosophy* 26, 727–763.

- von Fintel, K., Heim, I., 1998. Notes on implicature. Lecture Notes from a course on Pragmatics in Linguistic Theory at MIT.
- Wilson, D., Carston, R., 2007. A unitary approach to lexical pragmatics: Relevance, inference and ad hoc concepts. In: Burton-Roberts, N. (Ed.), *Pragmatics*. Palgrave Macmillan, London, pp. 230–259.
- Zimmermann, T. E., 2000. Free choice disjunction and epistemic possibility. *Natural Language Semantics* 8 (4), 255–290.