Probabilistic pragmatics, or why Bayes' rule is probably important for pragmatics

Michael Franke & Gerhard Jäger*

University of Tübingen, Seminar für Sprachwissenschaft

1 Introduction

Pragmatics is about language use in context. This involves theorizing about speakers' choices of words and listeners' ways of interpreting. More often than not, this also involves a certain amount of noise and uncertainty: speakers and listeners may not know exactly what the relevant contextual parameters are, they may make mistakes, believe that their interlocutor is uncertain and possibly prone to err, etc. We believe that taking this picture seriously can, despite its apparent messiness, inspire a stringent formal approach to pragmatics that lends itself to precise empirical testing. We call it *probabilistic pragmatics* here, to emphasize the role that probabilities play in it. But it contains much more. In the following, we try to sketch its main characterizing features in relation to other approaches and give some example applications. We argue that probability models are the natural and most practicable tool for modeling the richness of pragmatic phenomena which are affected by many unknown contextual factors.

Sections 2 and 3 characterize probabilistic pragmatics. Section 2 discusses different levels of analysis in pragmatic theory, so as to contrast probabilistic pragmatics with alternative approaches. Section 3 discusses key properties of probabilistic pragmatics. Sections 4, 5 and 6 sketch examples of applications. Section 4 introduces a baseline model for reasoning about referential expressions to demonstrate how the probabilistic modeling, inspired by classical pragmatic theory, can be fit to experimental data. Section 5 exemplifies further ways in which probabilistic pragmatics can shed light on gradient patterns in empirical data. The leading example for illustration is that of scalar implicature. Section 6 argues that considering (multiple levels of) gradient subjective contextual uncertainty, as captured by a probability distribution, is essential to understanding indirect speech acts. This section demonstrates how explicit representations, inspired from game theory, of interlocutors' preferences and likely dialogue moves help tackle indirectness of speech in non-cooperative contexts.

2 Levels of analysis within pragmatic theory

Paul Grice's work on conversational implicatures (Grice, 1975) has greatly inspired the shaping of theoretical pragmatics (e.g. Gazdar, 1979; Horn, 1972; Atlas and Levinson, 1981; Levinson, 1983; Horn, 1989; Levinson, 2000). It thereby also shaped experimental approaches to pragmatic phenomena, of which recent years have seen more and more (e.g. Noveck and Sperber, 2004; Meibauer and Steinbach, 2011). We take Grice's ideas as a starting point here.

*Author names appear in alphabetical order

Grice showed that paying attention to regularities of language use helps reconcile a semantic analysis of natural language in terms of standard logics with meaning intuitions that seem to contradict such analyses. Crucial in Grice's approach was the formulation of Maxims of Conversation, which are speaker-oriented rules of conduct, such as *Be relevant!*, his Maxim of Manner. The Maxim of Quantity requires, for example, that speakers provide all the relevant information they are capable of providing. Listeners, in turn, can derive pragmatic inferences based on the assumption that speakers adhere to these rules. Whether these rules are normative or merely matter-of-fact may be inessential for the purpose of deriving pragmatic inferences, but it is important for our purposes here to note that Grice thought of these regularities not as arbitrary, but derivable from general considerations of rationality:

> "[O]ne of my avowed aims is to see talking as a special case or variety of purposive, indeed rational, behaviour." (Grice, 1975, p. 47)

> "I would like to be able to think of the standard type of conversational practice not merely as something that all or most do *in fact* follow but as something that it is *reasonable* for us to follow, that we *should not* abandon." (Grice, 1975, p. 48)

Probabilistic pragmatics follows Grice in assigning an important role to goal-oriented, optimal behavior. But probabilistic pragmatics is not particularly interested in maxims; it targets the more foundational level of explaining pragmatic phenomena by appeal to reasons and purposes.

There are many levels of analysis at which pragmatic theory can operate. Figure 1 gives five such levels. There may be more, but these suffice for our present purposes. With the exception of the level of processes (which we will discuss in Section 3), there is a linear order. From top to bottom, we go from descriptive to explanatory, from specific to general, from detail to big picture. Different levels of analysis are motivated, at least in part, by a different weighing of research questions. In the following, we focus on contrasting the levels of constraints and principles, on the one hand, with that of reasons, on the other, because this most clearly demarcates probabilistic pragmatics, which operates at the level of reasons, from a big chunk of contemporary work in formal pragmatics, whose focus is on constraints and principles.

On the level of constraints, we are interested in formulating generalizations of pragmatic interest with a rather specific scope. A prominent example from the recent literature is Hurford's constraint (Hurford, 1974; Chierchia, Fox, and Spector, 2012), which is the generalization that a disjunction of the form "*A* or *B*" sounds pragmatically infelicitous if *A* entails *B* or vice versa, such as in *John is in Paris or in France.* It hardly needs an argument why we should be interested in generalizations of this kind; they are the building blocks for an empirical basis of pragmatic theory.

Related to the level of contraints, but slightly more general in scope is the level of principles. Like constraints, principles aim to capture relevant regularities. Unlike constraints, they may apply to a larger set of phenomena. (The distinction is vague and flimsy; it is only drawn for illustration.) A prominent example is the strongest meaning principle (Dalrymple et al., 1998), according to which the logically strongest reading of ambiguous sentences is the preferred interpretation. This has been suggested for disambiguating reciprocals (Dalrymple et al., 1998), plural predication (Winter, 2001), complex implicature cases (Chierchia, Fox, and Spector, 2012) and vague predicates (Cobreros et al., 2012). In this sense, its scope is more general than that of, say, Hurford's constraint, while still being a generalizing description of relevant observations (e.g., meaning intuitions).

The levels of contraints and principles are mainly concerned with a tight description of the observable facts and so chiefly answer *what?*-questions. While there is no denying that this is important for an empirically oriented theoretical linguistics, there are other important criteria for scientific theory

level	example	question
constraints	Hurford's constraint: In a disjunction "A or B," A may not entail B.	what?
$\updownarrow$		
principles	Strongest meaning principle: Prefer strongest reading of an ambiguous sentence.	what?
$\updownarrow$		
maxims	Maxim of Quantity: Maximize flow of relevant information.	what/why?
$\updownarrow$		
reasons	Optimal language use: Be rational (or, at least, try to be)!	why?
processes	Naive serial modularity: Compute truth-conditions, then apply pragmatics.	how?

Figure 1: Levels of analysis for pragmatic theory. Constraints are descriptions of data observations with a narrow focus. Principles are more general descriptions. Maxims are general descriptions of (speaker) behavior that aim to explain various data observations. Reasons are explanations driven by more general considerations for maxims, principles, constraints, or, directly, for data observations. Processes describe how behavior is produced: all previous levels could or should relate to processes, but do not have to do so, logically speaking.

formation that emphasize the need to address *why?*-questions too. These are concerned with parsimony of explanation, reducibility, coherence, plausibility and general intelligibility. This is where we see the added benefit of pragmatic theory formation at the levels of maxims and reasons, which try to derive the data-driven generalizations of the former levels from more general ideas. While maxims are still partly generalizing descriptions of (speaker) behavior, the level of reasons tries to explain maxims, principles, constraints or, directly, pragmatic behavior in a uniform picture. Ideally, we would like to be able to follow the lead of the natural sciences and reduce multiple constraints, principles and data-observations to a smaller set of common assumptions that not only describe what is happening in a uniform manner, but that also embeds pragmatic theory in a larger context of human cognition.

A lot of formally-oriented research in pragmatics takes place at the level of constraints and principles. The method is to formulate, using mathematical notation or structures (such as a logic, or an algebraic model structure), a set of assumptions from which particular observations can be derived (given possibly implicit background assumptions accepted by the community). This method is flexible and so enables good fits to observable data. But sometimes assumptions made to explain data observations can be and should be explained or motivated by appeal to more fundamental ideas. This is where the level of reasons comes in, giving reasons for pragmatic facts, not just generalizing descriptions.

Consider two examples from the recent literature. First, let us look at gradable adjectives like *tall/short* or *bent/straight*. A well-motivated (though not undisputed) formal semantics for these expressions uses degrees (e.g. Cresswell, 1977; von Stechow, 1984; Rotstein and Winter, 2004; Kennedy and McNally, 2005). The degree-based approach assumes that a gradable adjective A denotes a function $[\![A]\!]_{\langle e,d\rangle} = \lambda x_e \,.\, \mathbf{A}(x)$, mapping an individual x of type e onto the degree $d = \mathbf{A}(x)$ to which x has property A. Truth-conditions for a sentence like (1) are derived by comparing John's and Bill's degrees of tallness (being antonyms, *tall* and *short* "live" on the same scale of degrees) with contextually supplied thresholds θ_{tall} and θ_{short}. A gloss is provided in (2).

(1) John is tall and Bill is short.

(2) John's degree of tallness is above θ_{tall} and Bill's degree of tallness is below θ_{short}.

What remains to be explained is how thresholds are fixed in a given context to form truth-conditions as in (2). We focus here on one observation, just for concreteness of an example. Some antonym pairs like *tall/short* are non-overlapping and even non-complementary, i.e., there is a middle ground where neither *tall* nor *short* applies so that θ_{tall} should be strictly bigger than θ_{short} across the board. This does not follow from the semantics sketched so far.

Solt (2011) discusses a possible solution, which she attributes to von Stechow (2006). Non-overlap and non-complementarity of antonyms can be explained if we assume that positive form adjectives are not compared to point-valued thresholds, but to a non-trivial interval. If the relevant point of comparison for *tall*, say, is the upper bound of that interval, and that of *short* is the lower bound, non-overlap and non-complementarity can be derived from this structural assumption.

Many explanations in formal semantics/pragmatics are similar to this. An assumption about abstract structural properties or operations entails (in a system of accepted background principles) the datum to be explained. But in our view, this particular case of a structure-driven explanation is not very convincing. It does not feel like we learn *why* antonym pairs like *tall/short* should be non-complementary. This feeling is corroborated by the existence of pragmatic explanations that derive non-complementarity from assumptions about language use (e.g. Franke, 2012; Lassiter and Goodman, 2014; Qing and Franke, 2014). For instance, if the use of adjectives is shaped by the desire to facilitate referential communication in statistically variable contexts, then it follows that *tall* and *short* will be used only for those individuals that are remarkably taller or shorter than average (Franke, 2012).

Not everything can be easily bent to an optimality-driven functional explanation. Here is a nice (borderline) example. Spector (2014) argues that the French complex disjunction *soit ... soit* is a positive polarity item whose distribution can be explained by the assumption that it must occur in the scope of a (hidden) exhaustifity operator. Exhaustification is a formal operation with a long history in formalizing pragmatics (e.g. Groenendijk and Stokhof, 1984; von Stechow and Zimmermann, 1984; Schulz and van Rooij, 2006; Fox, 2007). With this, Spector's assumption of obligatory exhaustification entails the relevant pattern of observed distribution of *soit ... soit*. As the distributional pattern in question is far from trivial, this is an impressive achievement. The assumption of obligatory exhaustification explains, in a structural and descriptive sense, the observed empirical data. Still, the question does arise whether we must accept obligatory exhaustification of a lexical item like *soit ... soit* as a primitive, or whether it can be explained by "deeper principles." Spector sees little chance of that for a traditional Gricean approach (see his Section 4.2), but Lauer (2014) argues that obligatory implicatures are as such consistent with and even predicted by Gricean theory in certain cases. We remain uncommitted here and stress that we would not want to claim (insanely) that everything of pragmatic interest needs to be reduced and explained as optimal language use; structural assumptions about language play a pivotal role. But where it is or seems possible to reduce structural assumptions to general principles of language use, we believe that it is fruitful and insightful to try.

Probabilistic pragmatics aims at the level of reasons. It aspires to explain language use by considerations of rationality or optimality. (We will enlarge on this in the next section.) This is not the only conceivable strategy for a pragmatic theory that aims at the level of reasons, but, we believe, a plausible one that is in line with Gricean ideas (for whatever that is worth) and also one that has already demonstrated its abilities in the past. Probabilistic pragmatics is meant to complement, but not necessarily to compete with more representation-driven, structural and descriptive approaches, unless these claim that they are all that is ever needed. We believe that scientific progress comes from integrating multiple perspectives: one task, many tools.

3 Probabilistic pragmatics

Probabilistic pragmatics is a research program with a diverse and lively base of proponents that has grown impressively over the last couple of years. As we conceive it here, probabilistic pragmatics subsumes game theoretic approaches (e.g. Parikh, 2001; Benz, 2012; Jäger, 2012; Clark, 2012; Franke, 2013; Mühlenbernd, 2013; de Jaegher and van Rooij, 2014; Rothschild, 2013; Pavan, 2014; Deo, 2015), as well as "Bayesian approaches" (e.g. Frank and Goodman, 2012, 2014; Kao et al., 2014; Zeevat, 2014; Bergen, Levy, and Goodman, 2014; Lassiter and Goodman, 2014; Potts et al., 2015). With some due neglect of detail and variance between different contributions, five properties characterize probabilistic pragmatics: it is (i) probabilistic (*duh!*), (ii) interactive, (iii) rationalistic or optimality-based, (iv) computational and (v) data-oriented. Other approaches within pragmatics share some of these properties, but no other shares all. Some of these properties are intrinsically connected. Some entail further properties of interest: e.g., (i), (ii) and (iii) carry us into a Bayesian approach (in a sense to be explained below).

Probabilistic. Pragmatics is a fuzzy and gooey affair. Figuring out what a speaker meant at some occasion in a given context can be tricky. Even when it feels rather clear, there can hardly be perfect certainty about what that speaker thought the point of conversation was, which alternative utterances she may have been aware of (e.g., the extent of her active lexicon and preferences in her production grammar) and the like. Speakers and listeners are also not infallible and may make mistakes. If so, speakers and listeners may anticipate that listeners and speakers make mistakes and act accordingly. None of this needs to happen consciously (see below), but happen it does. Psycholinguists acknowledge this without shame or ado (e.g. Degen and Tanenhaus, 2015).

A defining feature of probabilistic pragmatics is that it takes various sources of uncertainty about the context into account and that it models this uncertainty with probability distributions. Some approaches may try to marginalize the role of probabilities to obtain an almost qualitative system of reasoning (e.g. Franke, 2011); others may want to make good use of the fuzziness of non-trivial probability distributions. Here are two reasons why the latter strategy makes sense. For one, probability can be needed for explanatory purposes, such as, e.g., in modeling vague language use (Frazee and Beaver, 2010; Franke, 2012; Lassiter and Goodman, 2014; Qing and Franke, 2014). We will see examples in Sections 5 and 6. For another, models that make probabilistic predictions about speakers' and listeners' choices lend themselves to straightforward empirical testing; they come, if designed properly, with a testable likelihood function ready-made for plugging into your statistical analyses. We will see examples in Section 4. It is possible, perhaps plausible, that ways of representing uncertainty other than probability distributions can do similar, perhaps better, work (cf. Halpern, 2003, for many alternatives), but probability theory is simply the most established and well-known, and beats its competitors in terms of practical applicability by a margin, especially when it comes to statistical testing of a model's predictions.[1]

[1]Cohen (2009), Zeevat (2014) and Goodman and Lassiter (2014) provide further arguments and perspectives on the use of probabilistic approaches within semantics and pragmatics.

Moreover, probabilistic pragmatics, as we conceive it, is **Bayesian in a the weak sense**, because it uses probabilities as representations of the language users' *subjective* uncertainty. We argue that this is important to explain certain subtle pragmatic phenomena, like indirect speech acts (see Section 6). Probabilistic pragmatics is also Bayesian in a stronger sense. This follows from other characteristic properties, as we will explain below.

Interactive. Pathological cases aside, pragmatics is business between speakers and hearers. Whenever two of these meet, they do so in a context. Sure, for reasons of theoretical elegance, say, if that is our notion of elegance, we can dispense with the pragmatic two-mind problem and strip context down to a single algebraic representation. Good work comes from such abstraction: witness exhaustification-based approaches to pragmatic inference (e.g. van Rooij and Schulz, 2004; Schulz and van Rooij, 2006; Fox, 2007; Chierchia, Fox, and Spector, 2012) or approaches like inquisitive semantics (Ciardelli, Groenendijk, and Roelofsen, 2013). But other approaches see added value in explicitly handling speaker, listener and context and the interaction between these. Approaches that do are the intentions-first approach of Geurts (2010), the dynamic pragmatics of Lauer (2013), relevance theory (Sperber and Wilson, 1995, 2004) and many approaches in psycholinguistics (e.g., the holistic constraint-based approach of Degen and Tanenhaus, 2015). Borderline cases of interactive approaches are Neo-Gricean work (e.g. Horn, 1984; Levinson, 2000) and bidirectional optimality theory (Blutner, 1998, 2000; Blutner and Zeevat, 2009).

Probabilistic pragmatics therefore considers explicitly the role of production and comprehension. It does not conflate the two. Neither does it assume that speakers and listeners must have the same perspective on the relevant contextual parameters (see Franke, 2014a, for an extreme case of modeling divergences). When it comes to fitting a model to empirical data (see Section 4), this allows a much more straightforward map of a model's prediction to response patterns from experiments that relate clearly to either production or comprehension (cf. Degen and Goodman, 2014). Approaches that do away with an explicit distinction between speaker and hearer are in much muddier waters and must often rely on linking hypotheses, which are implicit and hence not properly evaluated, about how a given theoretical approach can even make predictions about (behavioral) data from an experiment (cf. Chemla and Singh, 2014, for related discussion).

Rationalistic. A key assumption of probabilistic pragmatics, as we conceive it here, is that pragmatic behavior is (approximately) rational, or better put: optimally adapted to solve a particular purpose. This is an empirical hypothesis, one that must be assessed indirectly by assessing the overal success of models that instantiate it. It is not a necessary assumption to make for a pragmatic theory that uses probabilities. But it is also not something that we picked from a lucid dream. In fact, it brings pragmatic theory into the confines of *rational analysis* as formulated by John R. Anderson:

> "A *rational analysis* is an explanation of an aspect of human behavior based on the assumption that it is optimized somehow to the structure of the environment. . . . [T]he term does not imply any actual logical deduction in choosing optimal behavior, only that the behavior will be optimized." (471 Anderson, 1991)

Rational analysis has been applied to many aspects of cognition, such as memory and categorization (e.g. Anderson, 1990, 1991), reasoning (e.g. Oaksford and Chater, 1994; Hahn and Oaksford, 2007) or inductive learning (Tenenbaum, Griffiths, and Kemp, 2006; Tenenbaum et al., 2011).

Rationality or optimality is an endstate that actual language users may not reach. Probabilistic pragmatics therefore happily considers noisy approximations to optimal choice. Whether these noisy realizations are themselves rational (e.g., a tradeoff between exploration and exploitation) is another matter. The crucial idea is that the assumption of optimality structures theory formation and explains *why* we see particular patterns of behavior.

For a rationalistic pragmatics to bite, we must specify what the goal or purpose is that pragmatic behavior is hypothesized to be optimal for. Again, this is an empirical issue. There can be different models within this approach that postulate different goals. Mostly, linguists assume that language use is shaped by the desire to communicate effectively. Some see the function of structuring thought as a reason for the evolution of grammar (Chomsky, 2011). This may be reasonable. But it is still a far step from there to see soliloquy as the motor for the evolution of (shared!) conventional meaning and conversational practices. Beyond blind cooperation, some see a role for egocentric motives, argumentation and non-cooperative linguistic behavior as well (e.g. Anscombre and Ducrot, 1983; Merin, 1999; Rubinstein, 2000; de Jaegher, 2003; de Jaegher and van Rooij, 2014; Blume and Board, 2014). Be that as it may be, we believe that the purpose of use that explains different pragmatic phenomena can be different. Each particular model should be evaluated by the particular assumptions that it makes.

A rationalistic analysis of why a particular pragmatic behavior or phenomenon is rational or optimal is often formulated in terms of the beliefs and preferences of language users. The predictions of the model are then derived by looking at what would be rational or optimal choices given the assumed beliefs and preferences. Still, probabilistic pragmatics is not necessarily committed to the idea that these beliefs, preferences and choice mechanisms are actively entertained and executed each time a pragmatic decision is made. Rather probabilistic pragmatics can be thought of as a computational-level analysis in the sense of Marr (1982). This is why Figure 1 also contrasts the level of reasons with the level of processes. Probabilistic pragmatics need not be totally unrelated to predictions about processes. There could (and some say: should) be some effort to relate computational-level rationalistic explanations *why* we see certain behavior to specifications of mechanisms *how* this behavior can be implemented, especially in the light of issues of computational complexity (e.g. van Rooij et al., 2014).

It is important to stress that the relation of probabilistic pragmatics to processing accounts is basically the same as that of other positions in theoretical pragmatics. These, too, need auxiliary assumption to spell out how a given abstract account makes predictions about processing-related observations such as reading or reaction times, or eye- or mouse movement (cf. Chemla and Singh, 2014). Nonetheless, different abstract theories will constrain the set of plausible processing theories in some way or other. Probabilistic pragmatics, for example, is domain-general, holistic and yet uncommitted with respect to the issue of modularity. Let us briefly elaborate.

Probabilistic pragmatics is domain-general in the sense that it is constrained by the same considerations of plausibility as rationalistic explanations in other domains: when we want to make particular assumptions in a particular rationalistic model, these are subject to domain-general criteria of plausibility. If a purported model needs to assume, for proper fit to the data in question, that, say, an agent responds rationally to only a subset of the speaker's utterances, but not to others, then this would clearly seem strange in the light of common-sense assumptions about rationality. At least in this sense, probabilistic pragmatics is domain-general. In contrast, many structural, mechanistic approaches within theoretical pragmatics are not constrained by common-sense in this way, and make good use of this freedom.

Being domain-general in this sense does not commit probabilistic pragmatics to being non-modular. Despite appearances, it is perfectly consistent with this approach to maintain that there is a specialized "pragmatics module" that carries out the computations in question. That the same general constraints on theory formation apply in other domains, does not mean that the same abstract cognitive system (or, even more ridiculously, the same brain area) must carry out these computations. Probabilistic pragmatics, as we see it, is open to the idea that pragmatic reasoning is a piece of special-purpose cognition, finely attuned to the specific affordances of this domain that may or may not be found in other areas (to this extent) as well, such as the processing of many layers of contextual clues, the execution of highly recursive planning strategies, or the representation of (higher-order) mental states. In this respect, probabilistic pragmatics differs from the modular version of a traditional Gricean approach to conversational implicature, as sketched by Chemla and Singh (2014).

Similarly, probabilistic pragmatics differs in its processing-related predictions from the version of a grammatical approach to implicature calculation given by Chemla and Singh (2014). A grammatical approach to implicatures (Chierchia, Fox, and Spector, 2012) strongly suggests a serial processing architecture: aspects relating to the epistemic or doxastic state of the speaker are computed *after* the basic implicatures of a sentence are computed. In contrast, probabilistic pragmatics is much happier with a holistic theory of pragmatic processing. Contextual information about the likely epistemic state of the speaker, or any other speaker-related parameter, can be taken into account immediately (Goodman and Stuhlmüller, 2013). Contextual and pragmatic considerations can affect phonological decoding, parsing and semantic analysis early on. There is no commitment here to a serial architecture; on the contrary, the probabilistic approach is particularly happy to marry a deeply holistic, multi-source approach to linguistic processing.

Bayesian (in a strong sense). From the idea that a pragmatic theory should be probabilistic, interactive and rationalistic, it is only a small step to assuming that it is Bayesian in a strong sense as well. Not only do we represent language users' subjective uncertainty in terms of probability distributions, we also use Bayes' rule to describe how, in particular, listeners' interpretations are a form of abductive reasoning, inferring the most likely (epistemic or intentional) state that would have triggered the speaker (under a reasonable model of utterance production) to say what he actually said (and not something else). We will see in Section 4 how this kind of reasoning is captured by Bayes' rule.

Computational. Probabilistic pragmatics is computational in the sense that it would like to offer mathematical models: it is a formal approach within theoretical pragmatics. The reason for this are clear: implications of hypotheses can be assessed and ideas and operations can be communicated with greater fidelity. Moreover, probabilistic pragmatic models are often implementable, and in fact implemented, in computer simulations. This is because, being interactive and taking noisy contextual parameters into account, it can become quite tedious to calculate predictions and implications, especially for parameterized models. Probabilistic pragmatics does not want to resort to hand-waviness; it would like to make precise predictions about empirical data (see below). In particular, it sets out to tackle more and more of the complexity that a psycholinguistic picture of pragmatic phenomena suggests. This is why computational models are handy and useful within this approach.

Data-oriented. Obviously, probabilistic pragmatics would like to explain empirical data. Otherwise it should not aspire to play in the garden of linguistic theory. Like much other work in formal pragmatics, part of the empirical data to be explained is based on introspective meaning intuitions and generalizations over these (as accepted by the practice of the community). But, as other theoretical work in pragmatics does too, the focus is increasingly on explaining empirical data from laboratory experiments or, occasionally, corpora or other sources. The main difference that probabilistic pragmatics brings along in this respect is that it can, by its very nature, go a step further: it often comes ready-made to predict, not only particular categorical features of the data, but the full quantitative pat-

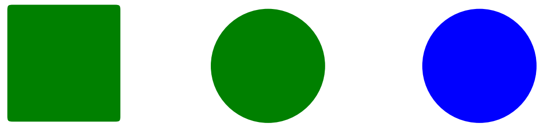


Figure 2: Example context for a reference game trial, after (Frank and Goodman, 2012).

tern found in a data set. This will become clear when we look at some of the experimental approaches outlined in Sections 4 and 5.

4 Reference games

Reference to objects, abstract or concrete, is basic to communication. Reference games are heavily simplified laboratory tasks designed to investigate production and comprehension of referential expressions in a confined, controlled environment. Reference games with different kinds of stimuli, different kinds of experimental measures, and slightly different empirical goals have been studied in the recent literature (Stiller, Goodman, and Frank, 2011; Degen and Franke, 2012; Frank and Goodman, 2012; Degen, Franke, and Jäger, 2013; Baumann, Clark, and Kaufmann, 2014; Carstensen, Kon, and Regier, 2014; Franke and Degen, 2015; Qing and Franke, 2015).

Consider the example in Figure 2. There are three possible referents. From left to right: a green square, a green circle and a blue circle. Speaker and hearer both know that these (and only these) are the referents at stake. The speaker's task is to identify a given referent. In many experimental set-ups, the speaker's choice is constrained to, say, a single property that is true of the referent (see Gatt et al., 2013, for criticism). The listener's task is to guess which referent the speaker had in mind for a given description.

For example, if the speaker wants to refer to the green square and his options are signaling "green" and "square," what should he choose? A Gricean speaker should choose "square" because that is a more informative description than "green." What should the listener choose if he hears "green?" By the same Gricean logic, the listener should choose the green circle, because a Gricean speaker who would have wanted to refer to the green square would have said "square." This is what fully rational agents would be expected to do, if they want to cooperatively play the communication game, and this is the starting point of rational analysis.

But even before looking at any data, the rational analyst may wish to add assumptions about computational limitations, the environment or other factors that might prevent agents from instantiating perfectly optimal behavior (Anderson, 1990, 1991). In the present case, it may well be that speakers have preferences for using shape properties rather than color properties or the other way around. It may be that listeners' choice of referents is influenced by contextual salience, not only by reasoning about informativity of utterances. Also, agents could make mistakes in calculating what the optimal choice is. A probabilistic pragmatic approach would integrate such factors. Here is a sketch.

A Gricean speaker, who adheres to the Maxim of Quantity, prefers more over less informative utterances. Most often, it is implicitly assumed that informativity is measured with respect to literal meaning, not pragmatically refined meaning. We do that here as well. To capture literal meaning in a probabilistic setting, define a dummy literal interpreter by a conditional probability distribution that maps each property (that is true of at least one referent) and each referent to a probability:

$$P_{\text{literal}}(\text{choose referent } r \mid \text{receive property } p) = \begin{cases} \frac{1}{\text{\# of referents with property } p} & \text{if } p \text{ is true of } r \\ 0 & \text{otherwise.} \end{cases}$$

If the speaker wants to refer to r and believes in literal interpretation, the above conditional probabilities are the speaker's expectations about the chance of referential success when choosing p. Suppose that furthermore the speaker has a preference for properties (e.g., shape over color, or nouns over adjectives). We capture this by adding a bonus or malus $f(p)$ to the expected communicative success of signaling p, so that we obtain an integrated measure of the speaker's *expected utilities*:

$$\text{EU}_{\text{speaker}}(\text{wish to refer to } r, \text{choose } p\ ;\ \text{parameter } f) = P_{\text{literal}}(r \mid p) + f(p)\,.$$

The notion of expected utility in the context of a reference game is explained further in Info Box 1.

Expected utilities are numerical measures of the (subjective) desirability of choice options, given the choice maker's uncertainty. In reference games, we assume that speakers want listeners to infer the intended referent. If we neglect speaker's preferences for certain utterances/properties over others, we can formally define the **utility** of a concrete round of play in which the intended referent is r and the listener infers referent r' as $\text{U}(r, r') = 1$ if $r = r'$ and 0 otherwise. (Any other pair of numbers, one bigger than the other, would produce the same results, up to rescaling.) **Expected utility** is then the probability-weighted average of utilities that the speaker expects to obtain from a given property choice p when wishing to refer to r: $\text{EU}(r, p) = \sum_{r'} P_{\text{literal}}(r' \mid p)\ \text{U}(r, r')$. Given the binary zero/one utilities, this simplifies to $\text{EU}(r, p) = P_{\text{literal}}(r \mid p)$, which is what we used in the main text. For illustration, the table above gives the expected utilities of property choices (columns) for intended referents (rows) for the reference game in Figure 2 (based only on informativity, i.e., neglecting speaker preferences for particular property choices).

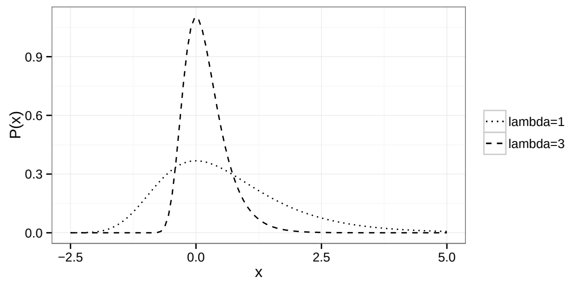
	property			
referent	square	green	circle	blue
	1	.5	0	0
	0	.5	.5	0
	0	0	.5	1

Infobox 1: Expected utilities

The optimal choice for a speaker who wants to refer to r is any property p that maximizes expected utility. But if speakers make mistakes in calculating these expected utilities, they might sometimes choose other properties too. This would be infrequent if large mistakes are infrequent. Also, we would expect that if errors occur, it is more likely that a sub-optimal property p' is picked than one that is even worse than p'. In other words, we would expect choice probabilities to be a monotonic function of expected utilities. Many probabilistic choice functions implement this (based on different ideas about what the underlying noise or error source is). A convenient and well-motivated choice is the soft-max function (e.g. Luce, 1959; McFadden, 1976; Sutton and Barto, 1998; Goeree, Holt, and Palfrey, 2008), which has a free parameter λ that captures the inverse error rate in calculating expected utilities. (See Info Box 2 for further explanation.) The production probability of choosing p given r is:

$$P_{\text{prod}}(\text{choose } p \mid \text{wish to refer to } r\ ;\ \text{parameters } \lambda,\, f) = \frac{\exp\left(\lambda \cdot \text{EU}_{\text{speaker}}(r, p; f)\right)}{\sum_{p'} \exp\left(\lambda \cdot \text{EU}_{\text{speaker}}(r, p'; f)\right)}\,.$$

The soft-max function can be derived from the assumption that a decision maker chooses the best option given a noise-perturbed assessment of expected utilities. Concretely, if the expected utilities of options $o_1, \dots, o_n$ are $u_1, \dots, u_n$, the decision maker assesses perturbed utilities $v_1, \dots, v_n$, where $v_j = u_j + x_j$. The probability that the agent chooses option o_i is then $P(o_i) = P(\forall j \neq i : \; u_i + x_i > u_j + x_j)$. This probability depends on the distribution of the error terms x_j. To derive the soft-max choice rule, we assume that all x_j are independently sampled from a so-called **Gumbel distribution** with location parameter $\mu = 0$ and scale parameter $\beta = {}^1\!/\!_\lambda$ every time a choice is made. The Gumbel distribution is a special case of an extreme-value distribution and gives the probability that the maximum of a set of samples has a particular value x. In our context, we can imagine that the decision maker considers spuriously a number of quixotic features of option o_j despite u_j, the biggest value of which is added to u_j. The plot above gives examples of the probability density of error x_j for different values of λ. Under this stochastic distribution of error terms, the above probability $P(o_i)$ reduces to the soft-max choice rule $P(o_i) = {}^{\exp(\lambda u_i)}\!/\!_{\sum_j \exp(\lambda u_j)}$ (e.g., Train, 2009, Chapter 3). Consequently, the λ parameter of the soft-max function can be seen as the inverse scale parameter of the noise-perturbation of expected utilities. It determines the variance of the error distribution: $\text{Var}(x_j) = {}^{\pi^2}\!/\!_{6\lambda^2}$. When used in the soft-max function, values for λ do not have an absolute meaning, but must be interpreted relative to differences between the values of expected utilities that are being perturbed.



Infobox 2: The soft-max probabilistic choice rule

A Gricean listener assumes that the speaker abides by the Maxim of Quantity. If the speaker may make mistakes, listeners likely anticipate that. To infer which referent a speaker had in mind, a rational listener would apply Bayes' rule:

$$P_{\text{comp}}(\text{choose } r \mid \text{receive } p \text{ ; parameters } \lambda, f) = \frac{P(r) \cdot P_{\text{prod}}(p \mid r; \lambda, f)}{\sum_{r'} P(r') \cdot P_{\text{prod}}(p \mid r'; \lambda, f)} .$$

Here, $P(r)$ is the prior probability that the speaker wants to refer to referent r. Frank and Goodman (2012) suggest to measure this empirically. Their prior elicitation condition asked subjects which referent they thought a speaker had in mind who gave a referential description in an unknown language. Data from this prior elicitation condition was then fed into the model as an empirical estimate of *a priori* salience of referents.

To illustrate how a probabilistic approach lends itself to modeling empirical data, let's consider a small data set (taken from Qing and Franke, 2015). The data comes from three experimental conditions: (i) prior elicitation ($N = 240$), (ii) comprehension ($N = 360$) and (iii) production ($N = 432$). Every subject, recruited via Amazon's Mechanical Turk, saw only one condition, and only one trial of that condition (a one-shot experiment). All conditions used a forced-choice paradigm. So, in the comprehension condition, subjects were presented with the context in Figure 2 and had to select a ref-

	property				
referent	square	green	circle	blue	N
🟩	.94	.06	0	0	144
🟢	0	.44	.56	0	144
🔵	0	0	.17	.83	144

(a) production

	referent			
property	🟩	🟢	🔵	N
none	.3	.12	.58	240
green	.36	.64	0	180
circle	0	.35	.65	180

(b) prior elicitation & comprehension

Table 1: A small data set from a reference game experiment. Each table shows the proportion of choices. Columns are choice options, rows are choice situations (experimental trials). The first row of the table on the right is the data from the prior elicitation condition (see main text).

erent for either "green" or "circle." In the production condition, subjects saw the context in Figure 2 and had to select a property for a designated referent.

Data from these experiments is shown in Table 1. The data from the prior elicitation condition is shown as the first row in the table on the right, together with the comprehension data. Inspection of the data suggests that speakers did conform to the Gricean postulate of informativity, at least in tendency: the majority of speakers selected property "square" to describe the green square and property "blue" to describe the blue circle. But there also seems to be a tendency to prefer shape-properties: the number of subjects that chose "circle" to refer to the green circle is higher than the number of subjects that chose "green" in this case, and the number of subjects who chose "square" to describe the green square is higher than the number of subjects that chose "blue" to describe the blue circle. Next, data from the prior elicitation condition suggests that the uniquely colored object may be most salient, followed by the unique shape. The interpretation of descriptions "green" or "circle" is hard to assess with the naked eye. It could well be a merge of prior salience and reasoning about the speaker's production probabilities, as the model suggests, but to assess this properly, we need a stringent model fit. In any case, it seems that there are subtle quantitative patterns in the data that a probabilistic model would like to catch.

One of the nice properties of probabilistic pragmatics, mentioned in Section 3, is that models often deliver a directly applicable likelihood function for data observations from a suitable experiment. In general terms, we get:

$$P(\text{possible data point } d \mid \text{model}, \text{concrete parameter values}),$$

directly from the theoretical model. For example, the production probabilities $P_{\text{prod}}(p \mid r\,;\,\lambda, f)$ defined above give us a parameterized likelihood function for each possible observation in the production experiment. If, for example, $\lambda = 3$ and $f(p) = 0$ for all p, then we predict that a choice of "square" in the production comprehension has a probability of ca. 0.82. That we observed 135 out of 144 subjects choosing "square," then has a likelihood of $\text{Binomial}(135\,;\,n = 144, p = 0.82) \approx 2.61e^{-5}$.[2] If we increase λ to 4, this increases to ca. 0.88. If we assume $\lambda = 4$ and that $f(p) = -0.1$ for color properties, the predicted probability of choice "square" is 0.92. This is just to demonstrate how probabilistic pragmatics is able, in principle, to make precise predictions about expected choice frequencies.

Looking at only one condition is not enough, of course. If we take the whole data set into account, we can ask whether there are plausible values for λ and specifications of f that make the model match

[2] Subjects could choose only properties that were true of the specified referent in the experiment, so this is a binary choice. The function $\text{Binomial}(k\,;\,n, p)$ gives the probability that k out of n trials are hits if the probability of a hit on each trial is p.

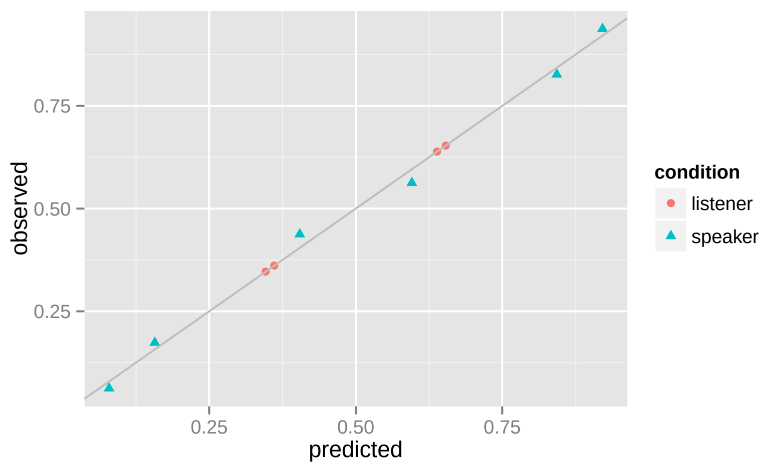


Figure 3: Prediction-observation plot. The model's predictions for the best-fitting parameters (see main text) are plotted on the horizontal axis, the empirically observed frequencies on the vertical axis. The perfect model would show all data points on the gray unit line.

the observed choice frequencies in a satisfactory way. Since we have a likelihood function, we can use it to find parameter values that make the observed data most likely (Myung, 2003). Assuming a fixed $f(p) = x$ for both color terms, we calculate that the best fitting parameters for our little toy data are $\hat{\lambda} \approx 4.13$ and $\hat{x} \approx -0.1$ for the production data and $\hat{\lambda} \approx 3.46$ and $\hat{x} \approx -0.23$ for the comprehension data.[3] Under best-fitting parameters the model's predictions are almost perfectly aligned with the observed choice frequencies. This is visualized in Figure 3.

Actually, the presented parameter fit based on our small data set is not enough to warrant substantial conclusions about the absolute quality of the presented model. The point here is mere illustration, nothing more. But it should suffice to see that probabilistic pragmatic modeling does provide a handle on subtle gradient aspects of empirical data. The real work, however, would begin basically where we must now leave it. In concrete applications, we would like to learn about model parameters from data-driven inferences. Moreover, we would like to compare probabilistic models that differ in theoretically relevant ways, based on their ability to predict the data. For example, Qing and Franke (2015) show by statistical model comparison that variants of the approach sketched here do worse overal, if the comprehension rule does not consider empirically measured salience. Franke and Degen (2015) show by the same method that, if we take individual-level data into account, other production and comprehension rules appear credible as well.

5 Gradience in pragmatic data

Experimental approaches to pragmatic phenomena are increasingly popular. Frequently, experimental data show gradient patterns that are unexpected under established models from theoretical pragmatics, which are often built from set theory and variations on standard logics. Probabilistic pragmatics does not enter the ring to knock out the establish contesters. Rather, it is one way of reconciling black-and-

[3] We could, of course, look for a single parmeter pair that best fits both production and comprehension data. Here, we simply assume that the listener need not have faithful estimates of the production parameters. It is zooming in on issues like this, that make for the daily (nightly?) work of the probabilistic pragmatics enthusiasts.

white theoretical ideas with empirically attested shades of gray.

Take the case of scalar implicatures. Under appropriate contextual conditions, an utterance of the logically weaker sentence in (3-a) suggests (3-b) because of the close association of logically stronger *all* with its "scale mate" *some*.

(3) a. Kiki borrowed some of Bubu's records.
b. Kiki borrowed some but not all of Bubu's records.

There are many accounts of scalar implicature, some of which disagree fundamentally (e.g. Geurts, 2010; Sauerland, 2012; Chemla and Singh, 2014, for overview). Nonetheless, we suspect that most would agree that whether an enriched implicature meaning is assessed in a given context depends on many factors (relevance, availability of alternatives, the question under discussion, ...). Still, most prominent formal accounts —all highly successful given the standards of theoretical linguistics— treat scalar implicature as if it was a binary phenomenon. Exceptions exist, but are rare (e.g. Russell, 2012; Goodman and Stuhlmüller, 2013). Yet there are empirical observations that are hard to reconcile with a categorical formal picture. We will look at some presently. The solution could be to say that formalization can only carry this far; or to deny that there is anything of theoretical interest to the attested gradience; or to start building models (one step at a time) around the existing theoretical ideas that gradually work towards integrating gradient contextual clues into formal accounts that are capable of predicting how different factors affect the strength of a scalar inference in context. Unsurprisingly, we would like to take the latter option.

There are many ways in which scalar implicatures appear fuzzy and gooey. Firstly, the readiness with which interpreters draw the scalar inference from *some* to *some but not all* depends on many contextual cues and additional factors and may not be as high as formal accounts may implicitly suggest (e.g. Degen and Tanenhaus, 2015; Degen, 2015). Secondly, beyond the *some/all* case, there is substantial variability in the strength of scalar inferences, depending on the lexical items at stake (e.g. Doran et al., 2009; van Tiel et al., 2014). Pairs like *some/all* or *sometimes/always* invite scalar implicature answers in suitable experimental settings much more strongly than pairs like *big/enormous* or *attractive/stunning*. Finally, when we look more closely at the preferred pragmatic interpretation of *some* in terms of actual quantities described, we see systematic patterns that seem to call out for a quantitative account. Let's consider this last point in slightly more detail.

Suppose there are 10 circles. We tell you that some of the circles are white. How many of the 10 circles do you think are white? Likely, you would guess 4 or 5. And you would probably also consider 2 less likely than 6. But it would be unlikely that this is simply because you bring prior expectations to bear on the situation from general world knowledge. After all, what should a sane person expect *a priori* about the likely coloring of 10 circles mentioned in a linguistics paper? It seems more plausible that what we consider likely quantities is mediated by our linguistic interpretation of the sentence "Some of the circles are white."

Relevant data on sentences of the form "Some of the *A*s are *B*s" have been collected by presenting subjects with pictures that varied the cardinality $|A \cap B|$ (the target set size) and asking them to rate how well the sentence described the picture on a Likert scale (Degen and Tanenhaus, 2011; van Tiel and Geurts, 2013; Degen and Tanenhaus, 2015; van Tiel, 2014). Mean ratings for different target set sizes from two of these experiments are shown in Figure 4.[4] The plot also shows what a simple-minded application of standard theoretical approaches would give us: semantic meaning of *some* would exclude the case where no *A* is *B*; a scalar inference would exclude the case where all *A*s are *B*s; and we would otherwise expect the description to be just fine. It is hard to imagine that anybody

[4] The plot is taken from Franke (2014b). See there and the original papers for details about the experiments.

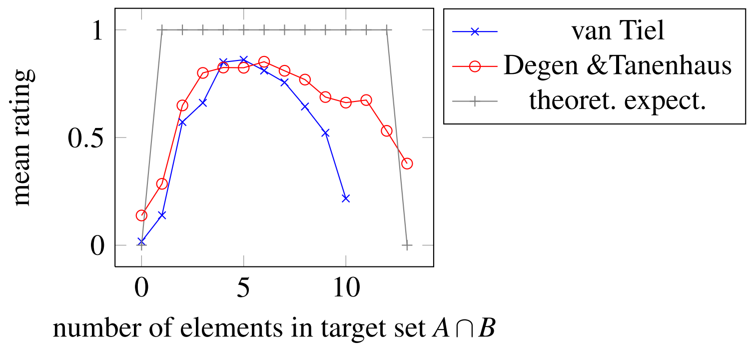


Figure 4: Mean Likert-scale ratings of sentences of the form "Some of the *A*s are *B*s" for different target set sizes in different experiments (see main text). The gray line gives the expected applicability under standard linguistic theory (for $|A| = 14$).

working in theoretical pragmatics would commit to the empirical predictions sketched by the gray line in Figure 4. Some element of noise surely must perturb the picture. But what, how and why exactly? To answer these questions in a stringent and empirically assessable way, a quantitative approach seems necessary.

One such approach hinges on *prototypes* and *typicality*. This idea has been shaped into a precise quantitative model by van Tiel (2014). In simplified terms, the model aims to predict mean ratings as a function of the distance between target set size $|A \cap B|$ and the prototypical interpretation of "Some of the *A*s are *B*s." The prototypical interpretation is taken to be the target set size with the highest mean rating. With some further assumptions in place, van Tiel's model matches his observed data very well. This is not the place to discuss strengths and weaknesses of van Tiel's general approach (see Cummins, 2014, for discussion). We mention this work, because it provides a nice contrasting example of a successful quantitative approach that does not share some of the fundamental properties of probabilistic pragmatics as characterized in Section 3. While it is computational and data-oriented, it is not probabilistic, interactive and rationalistic. In particular, it does not explain the data as the result of goal-oriented language use.

As an alternative to a typicality-based explanation, Franke (2014b) gives an extension of the basic probabilistic model of Section 4. The extension tries to explain the relevant data (e.g., Figure 4) by assuming that acceptability ratings reflect the production probability $P_{\text{prod}}(\text{use quantifier } q \mid \text{target set size } sn)$ with which a speaker would like to use quantifier $q = \textit{some}$ to describe a cardinality n. For this to work, two theoretically interesting changes have to be made to the simple baseline model from Section 4. Firstly, for referential language use, communicative success is likely to be binary: either the right referent is inferred (success) or not (failure). But when communicating quantities, it may be better to guess 5 when the real quantity is 4 than to guess 8. How bad a certain difference between interpretation and actual quantity is is a free parameter. In other words, we include a parameter for the extent of allowable pragmatic slack. Secondly, the model assumes that there could be different levels of salience of different alternative expressions. The model includes alternatives *none*, *all*, *many*, *most* (with some dummy semantic meaning) and also numeral expressions *one*, *two*, *three* but treats the salience of each of as a set of free parameters that are estimated from the data. So, instead of assuming that the notion of alternativeness is categorical and fixed for good by grammar, we have a gradient competition model between differently salient alternatives. With these assumptions in place, the probabilistic model explains the observed ratings very well and also yields empirical estimates of gradient alternativeness (see Franke, 2014b, for details). Whether the typicality-based model of van

Tiel, the probabilistic model sketched here, or yet some other alternative comes out first in stringent model comparison, is a matter open for future investigation.

Another way in which the interpretation of *some* shows interesting gradient behavior is in its interaction with prior world knowledge. If instead of talking about the colors of circles, we hear (4), then it does seem to influence our estimate of the quantity of flunkers whether the students in question (normally) perform very well or whether that particular test was known to be hard.

(4) Some of the students failed the test.

The probabilistic model sketched in Section 4 would have little trouble integrating prior expectations and merging them with quantity reasoning. Most obviously, the comprehension probabilities P_{comp} would simply need to factor prior expectations into Bayes' rule, much like the model of Section 4 did with the estimated salience of referents.

But in certain cases, the interaction between quantity reasoning and world knowledge can be puzzling. Consider the case below, taken from Geurts (2010).

(5) Cleo threw all her marbles in the swimming pool. Some of them sank to the bottom.

General world knowledge would have us expect that all of the marbles sank. But an utterance of (5) seems to suggest rather pointedly that not all of the marbles sank. This is not predicted from the simple and obvious extension of the probabilistic model mentioned in the previous paragraph. What's going on?

There are many solutions to this problem in related frameworks (e.g. Franke, 2009; Rothschild, 2013; Franke, 2014a). Let's focus here on a recent proposal that nicely ties in with other very promising work. Degen, Tessler, and Goodman (2015) present a simple extension of the basic model of Section 4 in which the listener does not only infer the likely world state (e.g., the number of sunken marbles or the number of flunkers). The listener also tries to infer what the relevant prior expectations should be (according to the speaker), in particular whether anything abnormal or unexpected is going on. This is a *joint inference*: an inference about several (possibly related) parameters of interest. Concretely, in the case of (5), the listener reasons about two things: (i) how many marbles sank, according to the speaker (the world state), and (ii) how likely every number of sinking marbles is according to the speaker (the speaker's prior beliefs about world states). Given an utterance of (5), this joint-inference system predicts that listeners will come to believe that most likely (according to the speaker) not all of the marbles sank and that there is something fishy about the world in the sense that the speaker's prior beliefs about world states are not that all marbles are likely to sink (maybe the pool is filled with fish oil or the marbles are hollow). This is because this assumption best explains why the speaker uttered (5). Hence, the joint-inference model explains why, despite prior expectations of the listener to the contrary, we get a scalar implicature reading of (5). Degen, Tessler, and Goodman show that this model makes astute predictions about empirically observed listeners' interpretations of sentences like (5) for many items that differ with respect to the associated prior expectations.

Modeling listeners' joint inferences about several parameters at once seems very promising and has already proven its worth in other applications. Consider the general idea. All of these are, or can be, highly interdependent: conclusions about (pragmatic) inferences licensed by an utterance, the contextual resolution of semantic variables (pronouns, deixis, temporal reference, . . .), inferences about the question under discussion that the speaker meant to address, the level of pragmatic slack assumed feasible by the speaker etc. The idea of a joint inference is that listeners would, on occasion and perhaps frequently, infer many or all of these in one swoop. Probabilistic pragmatics has little trouble modeling such holistic inferences. Lassiter and Goodman (2014) model the interpretation of vague gradable adjectives as a joint inference about the contextually relevant threshold of use for a word like *tall* and the most likely interpretation (say, someone's body height). A joint inference that extends to uncertainty about the lexical meanings that a speaker entertains has been applied to otherwise perplexing manner implicatures (Bergen, Levy, and Goodman, 2012, 2014) and alleged embedded implicature readings (Potts et al., 2015). Non-literal interpretations can be captured by joint inference models that allow for uncertainty about the question under discussion (Kao, Bergen, and Goodman, 2014; Kao et al., 2014). In sum, there seems to be a lot of potential in modeling holistic inferences about multiple interdependent unknowns in a probabilistic modeling approach.

6 Indirect speech acts

The previous two case studies dealt with empirical domains where the data to be explained are of a gradient, continuous nature. This section is more in line with traditional pragmatics in the sense that it concerns introspective judgments about pragmatic interpretations, i.e., data that are *prima facie* categorical. We will argue that assumptions about quantitative relations between subjective probabilities are a helpful analytical tool in such a setting as well.

The problem Indirect speech acts seem to pose a challenge to a rationalistic model of pragmatics. They seem to constitute blatant violations of Gricean Maxims (especially Quantity and Manner, but sometimes also Quality and Relation). If the Gricean Maxims express principles of rational communication, indirect speech acts appear to be examples of irrational behavior. Following the overall tradition of Brown and Levinson (1987), we will make a case here, however, that indirect speech can in fact be conceived as rational behavior if the assumptions and goals of the interlocutors are properly taken into account, and sketch a mathematical implementation of this idea which is heavily influenced by game theory (especially by the *Iterated Best Response* model of game theoretic pragmatics; (cf. Franke, 2011; Jäger, 2012; Franke and Jäger, 2014)).

In the chapter *Games People Play* of his popular book *The Stuff of Thought: Language as a Window into Human Nature* (Pinker, 2008), as well as in a series of journal publications (Pinker, 2007; Pinker, Nowak, and Lee, 2008; Lee and Pinker, 2010), Steven Pinker discusses a variety of examples and offers an informal solution to the apparent paradox that people choose complicated and error-prone ways of communicating things that could be expressed in a perfectly clear and perspicuous way. According to Pinker, three factors are at play here: successful indirect speech acts (1) maintain plausible deniability, (2) establish shared knowledge but not common knowledge of the intended content, and (3) avoid mixing of relationship types.

Plausible deniability is perhaps best illustrated with Pinker's example of a veiled bribe:

> "The veiled bribe is another recognizable plot device, as when the kidnapper in *Fargo* shows a police officer his drivers's license in a wallet with a fifty-dollar bill protruding from it and suggests, 'So maybe the best thing would be to take care of that here in Brainerd.'" (Pinker, 2008, p. 374)

If the police officer is corrupt, he will let the speaker off the hook, but if he is honest, the speaker still avoids being charged for bribing an officer.

Stalnaker (2005) gives another example which fits into this context (even though he does not offer a specific analysis):

> "In May, 2003, the US Treasury Secretary, John Snow, in response to a question, made some remarks that caused the dollar to drop precipitously in value. The *Wall Street Journal* sharply criticized him for 'playing with fire,' and characterized his remarks as 'dumping on his own currency,' 'bashing the dollar,' and 'talking the dollar down.' What he in fact said was this: 'When the dollar is at a lower level it helps exports, and I think exports are getting stronger as a result.' This was an uncontroversial factual claim that everyone, whatever his or her views about what US government currency policy is or should be, would agree with. Why did it have such an impact?" (Stalnaker, 2005, p. 82)

If we suppose that John Snow knew what he was doing, he might have chosen to avoid a more direct statement because his indirect statement left the option open to deny such intentions later on (what, according to Stalnaker, he in fact did).

Pinker illustrates the second motivation—creating shared knowledge but avoiding to establish common/mutual knowledge—with a sexual proposition that is couched in indirect terms.

> "Say a woman has just declined a man's invitation to see his etchings. She knows—or at least is highly confident—that she has turned down an invitation for sex. And he knows that she has turned down the invitation. But does he know that she knows that he knows? And does she know that he knows that she knows? A small uncertainty within one's own mind can translate into a much bigger uncertainty when someone else is trying to read it." (Pinker, 2008, p. 418)

Regarding the third motivation—avoid the mixing of relationship types—, Pinker offers the following example:

> "How about this: You want to go to the hottest restaurant in town. You have no reservation. Why not offer fifty dollars to the maitre d' if he will seat you immediately? This was the assignment given to the writer Bruce Feiler by *Gourmet* magazine in 2000.[5] The results are eye-opening.
>
> The first result is predictable to most people who imagine themselves in Feiler's shoes: the assignment is terrifying. Though no one, to my knowledge has ever been arrested for bribing a maitre d', Feiler felt like the kidnapper in *Fargo* [...]
>
> The second result is that when Feiler did screw up the courage to bribe a maitre d', he thought up an indirect speech act on the spot. He showed up at Balthazar, a popular restaurant in Manhattan, and with sweaty skin and a racing heart he looked the maitre d' in the eye, handed him a folded twenty-dollar bill, and mumbled, 'I hope you can fit us in.' Two minutes later they were seated, to the astonishment of his girlfriend. On subsequent assignments he implicated the bribes with similar indirectness:
>
>> I was wondering if you might have a cancellation.
>>
>> Is there any way you could speed up my wait?
>>
>> We were wondering if you had a table for two.
>>
>> This is a really important night for me." (Pinker, 2008, p. 399)

[5] B. Feiler, "Pocketful of Dough,", *Gourmet*, October 2000. Cited after Pinker (2008), 457.

According to Pinker, a direct speech act such as "I will pay you 20 dollar if you let me jump the line." would not have worked because a maitre d' is in a position of dominance/authority. Doing a business transaction such as providing a free table without wait in exchange for 20 dollar is incompatible with a relation of authority; it is typical of a relationship of reciprocity/exchange/fairness. These relationship types are incompatible, so by openly accepting the bribe, the maitre d' would have forgone his position of authority. The indirect speech act offered him a way to take the money while *saving face*; he could pretend to maintain the authority relationship type while acting according to the reciprocity type.

The three factors Pinker mentions undoubtedly do play a role in the pragmatics of indirect speech acts, but they perhaps do not cover the whole story. Consider a situation where a mobster wants to coerce a restaurant owner into paying protection money:

(6) Your little daughter is very sweet. She goes to the school in Willow Road, I believe.

This is clearly a veiled threat. If there are witnesses and the mobster is tried for extortion in a court of law, there will be no plausible deniability though. Every judge or juror will recognize the threat as such. Also, there is no ambiguity about the type of social relationship speaker and hearer are in here. It might be argued that (6), as opposed to a direct speech act such as (7), does not create common knowledge of a threat. After all, (6) could, in principle, be an innocent remark, while no such misunderstanding is possible with (7).

(7) If you do not pay your protection money, we will kidnap your daughter.

So while (6) does not establish common knowledge, it does establish common knowledge that it is very likely that the speaker wants to convey a threat. But why should this subtle difference motivate the speaker to prefer (6) over (7)?

In general, there are actually two problems to be addressed: (a) why do indirect speech acts work in the first place, and (b) when and why is it rational to prefer indirect speech acts over direct ones?

Why do indirect speech acts work? Let us return to Stalnaker's example, repeated here as (8-a):

(8) a. When the dollar is at a lower level it helps exports, and I think exports are getting stronger as a result.
 b. The US Treasure will take measures to lower the dollar's exchange rate.

How does (8-a) convey the information that the speaker intends to take action leading to a weaker dollar, i.e., the literal content of (8-b)? Here is a sketch of a rationalistic explanation couched in decision and game theory.

Suppose the speaker S is in one of two states — or, in the language of game theory, has one of two possible *types*:

- t_1: S will take actions to reduce the dollar's value.
- t_2: S will take no actions to reduce the dollar's value.

The listener L has some prior assumptions about the relative likelihood of these two types, which can be represented as a *prior probability distribution* P: $0 < P(t_1) < 1$ is the listener's level of credence that S is of type t_1; $P(t_2) = 1 - P(t_1)$.

How likely is it that S would utter (8-a) in t_1, and in t_2? For either type, the statement expresses an economic truism, but for t_1 it would be a useful argument to justify his intentions. It is conceivable that t_2 utters this sentence, just to say something meaningless during a public hearing, but as there myriads of meaningless statements to choose from, this likelihood is small. Let us use the following notation for the production probability that the speaker emits (8-a) (where s, the signal, symbolizes (8-a)) when she is of either type:

- $P(s|t_1)$: Likelihood that S utters s if he is in t_1.
- $P(s|t_2)$: Likelihood that S utters s if he is in t_2.

Given the considerations above, it seems fair to assume that $P(s|t_1) > P(s|t_2)$.

As in the previous sections, we assume that the listener L will use Bayes' rule to compute the *posterior probability distribution* over S's types, given the signal observed:

$$P(t|s) \quad = \quad \frac{P(s|t)P(t)}{\sum_{t'} P(s|t')P(t')}$$

Since $P(s|t_1) > P(s|t_2)$, it follows that $P(t_1|s) > P(t_1)$, i.e., the posterior probability of t_1 is higher than its prior probability.[6] Note that $P(t_1|s)$ will still be smaller than 1 as long as $P(s|t_2) > 0$.

If we assume that S would never lie, statement (8-b) does have 0 likelihood in t_2, so upon observing it, the posterior probability of t_1 would be 1. So, choosing the indirect formulation (8-a) serves to make the content of (8-b) more likely for L without making it certain. As the actions of currency traders heavily depend on how they assess the probability of future economic events, the reactions to John Snow's remark in 2003 seem entirely plausible.

Why be indirect? The previous considerations illustrate the reasoning of a rational listener. Let us now consider a rational speaker S who wants to use her signal to influence some decision of some equally rational listener L. Here is another example of indirect speech. Suppose S visits a bazaar and sees a beautiful carpet that she desperately wants to buy. The price has to be negotiated. She could initiate her interaction with the carpet dealer with one of the three statements in (9).

(9) a. This rug has somewhat faded colors, but the pattern is kind of nice. $(= s_1)$
b. This is a beautiful carpet. $(= s_2)$
c. I have decided to buy this carpet. $(= s_3)$

[6] Here is the derivation. Please note that $P(t_1|s) + P(t_2|s) = 1$. If $P(s|t_2) = 0$, $P(s|t_1) = 1$. As we assumed above that $P(t_1) < 1$, in this case trivially $P(t_1|s) > P(t_1)$. Now let us assume $P(s|t_2) > 0$.

$$\begin{aligned}
P(s|t_1) &> P(s|t_2) \\
\frac{P(s|t_1)}{P(s|t_2)} &> 1 \\
\frac{P(s|t_1)P(t_1)}{P(s|t_2)P(t_2)} &> \frac{P(t_1)}{P(t_2)} \\
\frac{P(s|t_1)P(t_1)/P(s)}{P(s|t_2)P(t_2)/P(s)} &> \frac{P(t_1)}{P(t_2)} \\
\frac{P(t_1|s)}{P(t_2|s)} &> \frac{P(t_1)}{P(t_2)} \\
\frac{P(t_1|s)}{1-P(t_1|s)} &> \frac{P(t_1)}{1-P(t_1)} \\
P(t_1|s) - P(t_1)P(t_1|s) &> P(t_1) - P(t_1)P(t_1|s) \\
P(t_1|s) &> P(t_1)
\end{aligned}$$

Before we can analyze the potential impact of these statements, let us set up a model of the (possible) subsequent business interaction. For simplicity's sake, we assume that the potential seller, L, will offer a certain price once, or say nothing. The potential buyer, S, has the choice to accept or reject the offer if one is made.

To simplify things further, we take it that L will offer either a high price (35\$) or a low price (15\$). The carpet has an inherent value v_s for S and v_l for L. If both interlocutors are rational, a transaction can only take place if $v_l \leq v_s$. If the price paid is strictly between v_l and v_s, both parties will benefit from the transaction.

We consider three possible types of S (where 1\$ is the unit of values):

- t_1^s is not really interested in owning the carpet: $v_s = 0$.
- t_2^s has a moderate interest in the carpet: $v_s = 20$.
- t_3^s has a strong interest in the carpet: $v_s = 40$.

Then there are two possible types of L:

- t_1^l is moderately interested in selling the carpet: $v_l = 30$.
- t_2^l is strongly interested in selling the carpet: $v_l = 10$.

L assumes *a priori* that S is not terribly interested in the carpet: $P(t_1^s) = 0.9, P(t_2^s) = 0.09$ and $P(t_3^S) = 0.01$, while S prior assumption is that L is probably only moderately interested in selling: $P(t_1^l) = 0.9$ and $P(t_2^l) = 0.1$. Finally, we assume that both interlocutors would prefer not to embark upon a conversation that does not result in a sale. For concreteness' sake, we say that both value the time lost this way with 1\$.

The structure of this strategic situation can be represented as the *extensive game* which is depicted in Figure 5.

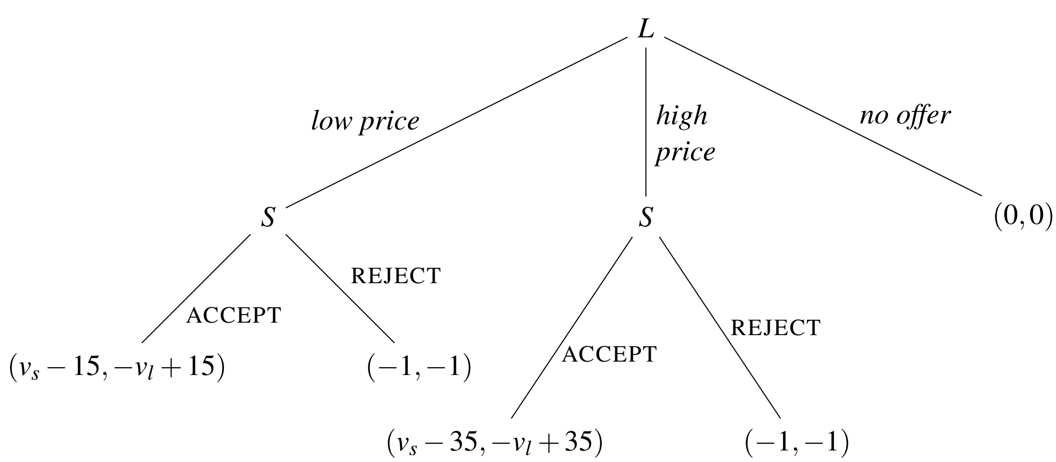


Figure 5: The carpet sale game

The game is played top-down starting at the root. Branching node labels indicate which interlocutor's turn it is. The game is over if a leaf is reached; the numbers at the leaves indicate S's and L's payoff.

If the values of v_s and v_l are known, this kind of game can be solved via *backward induction.* In our setting, these values are not known to the other conversation partner, but they can estimate them using their prior probabilities. The expected value of v_s is $E(v_s) = P(t_1^s)\cdot 0 + P(t_2^s)\cdot 20 + P(t_3^s)\cdot 40 = 2.2$. The expected value of v_l is $E(v_l) = P(t_1^l)\cdot 30 + P(t_2^l)\cdot 10 = 28$. Filling in these values leads to the game in Figure 6.

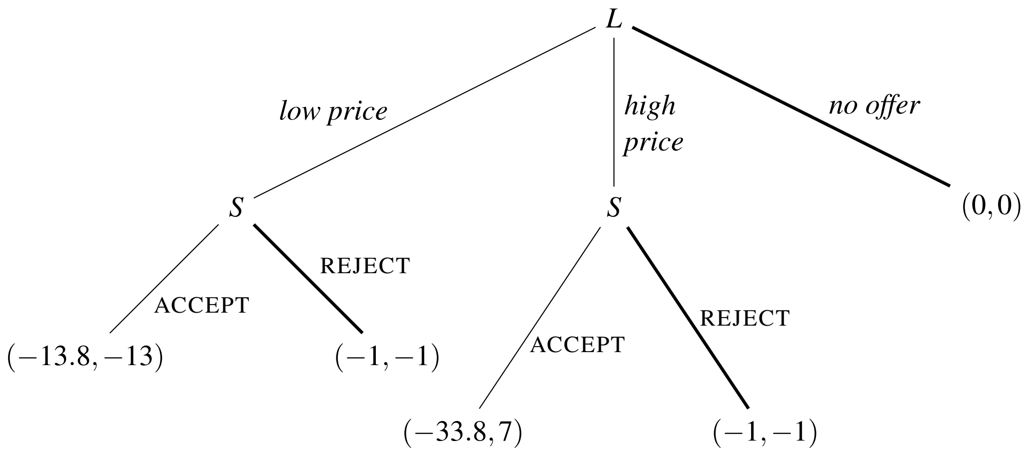


Figure 6: The carpet sale game: expected utilities

In both S-nodes, REJECT is the rational choice. So L has to expect that both *low price* and *high price* leads to a payoff of -1, which makes *no offer* the rational choice. (The rational choices are indicated by bold lines in Figure 6.) This is not surprising, as $E(v_l) > E(v_s)$, i.e., there is no possible price that would lead to a positive expected payoff for both interlocutors.

If S is type t_1^s, she is satisfied with this outcome. If she is in another type, she has an interest though in changing L's prior assumptions about her type, as this possibly induces him to make an offer that might be profitable for her. *Pre-play communication* offers her such an option. If she sends a signal to L prior to his first move, she has the chance to manipulate his beliefs about her type, and thus his choice of action.

As before, we assume that L holds a belief about S's production probabilities that are conditional on S's type. An example is displayed in Table 2. The values are somewhat arbitrarily chosen, with the intention to be both plausible and to yield intuitively plausible results in connection with the carpet sale game. The parameter x is to be interpreted as a very small constant, since there is a large

	s_1	s_2	s_3
t_1^s	$9x$	x	0
t_2^s	$4x$	$30x$	$20x$
t_3^s	x	$150x$	$500x$

Table 2: Production probabilities (conditional on type)

number of other possible utterances, so the probabilities of s_1, s_2, and s_3, in absolute terms, are very small. The numbers chosen are compatible with the intuition that a disinterested S will likely not give any indication of interest at all, and if she does, she will only show mild interest. A moderately interested S is more likely to express her interest, and she will likely choose somewhat stronger terms. A strongly interested S is most likely to say so, probably in clear and direct terms. — Note that we are considering a non-strategic speaker at this point, i.e., a person who expresses her views without considering possible social consequences. Likewise, we are considering a listener L who is rational but socially inapt; he is capable of applying Bayes' rule, but it does not cross his mind that S might try to manipulate him.

With L's prior beliefs and beliefs about the speaker's production behavior, L's posterior distribution over S's types upon observing a signal is as given in Table 3 (from Bayes' rule; rounded to two decimal digits). The strategic situation L is facing upon receiving a signal is shown in Figure 7. Upon

	t_1^s	t_2^s	t_3^s
s_1	0.96	0.04	0.00
s_2	0.18	0.53	0.29
s_3	0.00	0.26	0.74

Table 3: L's posterior distribution

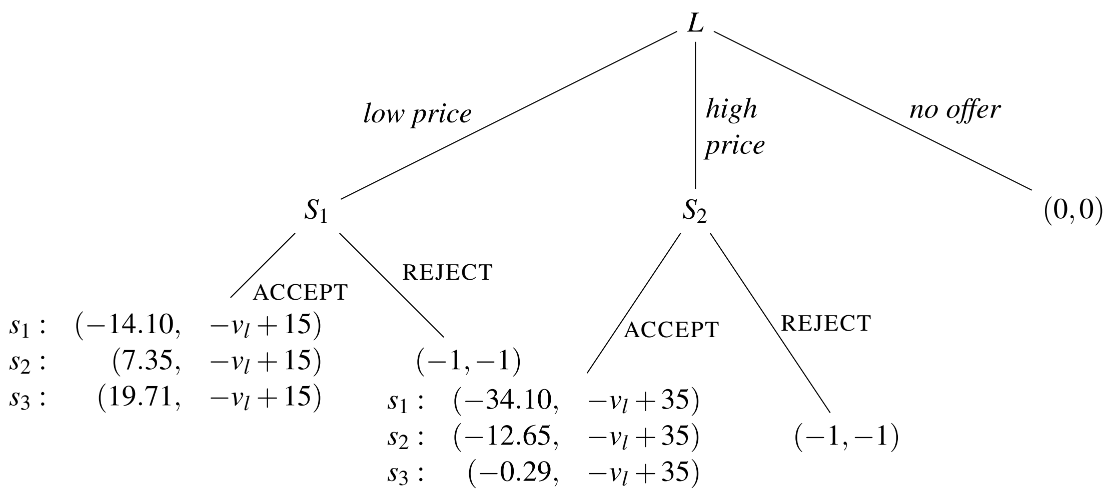


Figure 7: The carpet sale game: L's posterior expected utilities

observing s_1, L will conclude that S would reject either offer. This leaves *no offer* as his best choice. If he observes s_2, he expects *low price* to be accepted and *high price* to be rejected. Observing s_3 leads him to expect that both offers would be accepted.

If L is of type t_1^l, only the high price would secure him a profit. For t_2^l, both the high and the low price would be profitable. So L's best responses to the different signals are as in Figure 8.

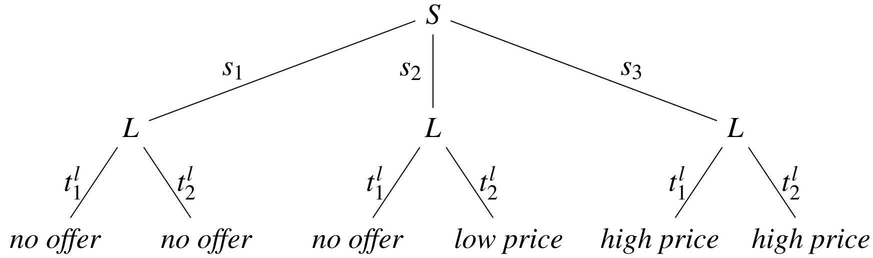


Figure 8: Best response strategy of a rational but naive L

When S plans her utterance (if she thinks strategically, that is), she can go through these calculations and determine her expected utilities depending on her type and the different possible messages. They are given in Figure 9.

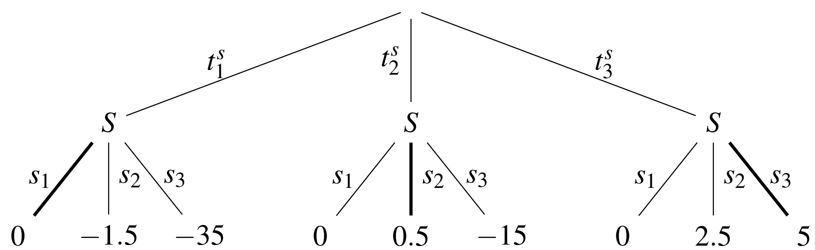


Figure 9: The carpet sale game: S expected utilities, given a rational but naive L

S's utility-maximizing choices (indicated by bold lines) are sending s_1 if she is not really interested, s_2 if she is moderately interested, and s_3 if she is strongly interested to buy the carpet. So the rational thing to do for S is to send a direct signal if her stakes are high and a moderately indirect signal if her stakes are low.

We might also consider a pragmatically sophisticated L who is able to anticipate S's calculations. For such an L, message s_i acquires the pragmatic meaning "I am of type t_i^s." His best response to this would still be as in Figure 8 though. The strategy pair: *S sends message s_i iff she is of type t_i^s and L responds as in Figure 8* are in equilibrium, i.e., they are rational responses to each other.

There is another twist to this story. Suppose S is more optimistic that L strongly wants to sell the carpet. Let us say that S's prior probabilities are $P(t_1^l) = P(t_2^l) = 0.5$. We call this the *eagerness assumption*. Then S's expected utilities shift to those in Figure 10. In this scenario, even a strongly interested buyer will only indicate her interest by using the indirect message s_2.

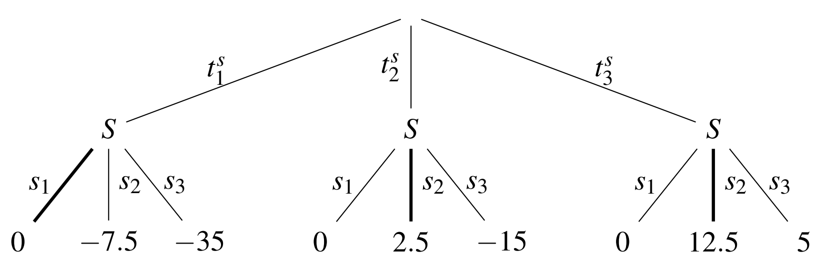


Figure 10: The carpet sale game: S expected utilities, given a rational but naive L, under the eagerness assumption

We refrain from a formal analysis of a setting where L is unsure about S's prior assumptions about him, as this would go beyond the scope of this article. However, the lesson to be drawn from these considerations can be formulated as such: A rational and sophisticated speaker in a negotiation situation will use an indirect message *if her stakes are low or if she believes that her opponent's stakes are high*.

The bigger picture So far, we tacitly assumed that the interlocutor's prior assumptions about each other were common knowledge. If this is not the case, indirect signals carry secondary information about these prior assumptions. Being very indirect may, in the appropriate setting, indicate exactly that: *Your stakes in this are higher than mine!* Conversely, a direct signal then carries the secondary message: *My stakes are higher than yours!*

If these prior assumptions are not common knowledge, this kind of secondary information will inform further levels of recursive strategic reasoning. This provides further motivation for using indirect speech acts. When the mobster in our example (6) above is indirect, he perhaps tries to communicate: "Your stakes are high, as I will hurt your daughter if you don't pay. My stakes, on the other hand, are low since I bribed the police and can pretty much do what I want in this neighborhood." Likewise, an indirect sexual innuendo is apt to carry the secondary message: "Your stakes are high since I am very, very attractive. My stakes are low because, well, I am attractive and there are many other potential partners if you should reject me." Conversely, a direct sexual come-on communicates: "Your stakes are low since I consider myself to be rather unattractive. My stakes are high because I haven't had sex in quite some time and if you reject me, it will stay that way." Seen this way, it is unsurprising that the indirect approach is likely to be more effective.[7]

In this example, the interlocutors had an immediate interest to communicate implicatures about their degree of interest in buying or selling the carpet, since both have a chance to save/earn more money by manipulating the other person's assumptions about each other. A similar case can be made to some other instances of indirect speech. Consider Pinker's example about the kidnapper trying to bribe a police officer. Here it makes sense for the speaker to convey the impression that his stakes are low —that he just wants to save the paperwork of an official ticket—, lest the police officers suspects the speaker has graver reasons to avoid contact with the police. In other cases, the interlocutors' higher-order expectations about each other do not have an obvious impact on what a rational course of action will be. For the maitre d' in the other example of an attempted bribe discussed by Pinker, it arguably shouldn't make a difference how important it is for the patron to get a table quickly. Even trickier are instances of indirect speech that are chosen for politeness reasons. Why would it be rational to downplay one's interest in another glass of wine by saying "I wouldn't mind another glass of wine"? The more direct "Pour me another glass of wine!" should be more apt if the intention behind the speech act is to get another glass of wine.

Intuitively, there are at least two conceivable motivations at play here: The speaker does not want to come across as over-eager to get more wine, and she does not want to convey the impression that she is in a position to give the listener an order. Both considerations are irrelevant for the interaction at hand, but they affect the speaker's *reputation*, i.e., the assumptions that other people, both the addressee and inadvertent listeners, form about the speaker. These assumptions may in turn affect

[7] Communicating that the speaker's stakes are low may backfire in certain situations. Compare the two attempts to a marriage proposals from the TV series *Dexter* (S3E4, October 19, 2008, Showtime):

- **First attempt:**
  Dexter: My insurance would cover you. Rita, if we got married, we'd have joint assets. You wouldn't have to worry. Let's not forget about marital deductions. With Astor and Cody as dependants...
  Rita vomits
- **Second attempt:**
  Dexter: My life has always felt like an unanswered question... a string of days and nights waiting for something to happen, but I didn't know what. Rita, we're connected. Wherever I am, I feel you, and the kids... with me. You're what makes me real. I want us to always go out for banana splits and replant the lemon tree that keeps dying, and I never ever want to miss a pizza night. And that's how I know I want to marry you. If something as simple as pizza night is the highlight of my week. But not without the kids. Cody, Astor, you guys are my family and I'm gonna hang onto you for dear life. Please, say yes?
  Rita (crying overjoyed): Yes! Yes, we will marry you!

the behavior of those listeners in future interactions with the speaker, and planning for those future interactions is, of course, a rational thing to do.

There is a sizable literature on rationality in *repeated games* (see for instance Axelrod, 1984; Mailath and Samuelson, 2006 as two representative book-length treatments). This format has drawn a lot of attention because in many repeated games, *cooperative* and even *altruistic behavior*, which would be irrational in one-shot interactions, can be shown to be rational if the expected outcomes of future interactions are properly taken into account. As a general lesson, this line of investigation has shown that rationality in one-time interactions and rationality in repeated interactions may diverge considerably. We take it that the reputation-building and face-saving effects that underlie many indirect speech acts and instances of polite behavior can be integrated into the overall perspective of rational communication if a theory of repeated interactions is taken into account. An attempt at a formal implementation would go beyond the scope of this article though.

7 Conclusion

Natural language pragmatics can and should be studied at several levels of description simultaneously. Probabilistic pragmatics focuses on the *level of reasons*, to provide justifications for maxims, principles and constraints, while abstracting over specific cognitive processes. Operating at the level or reasons, probabilistic pragmatics is a conglomerate of converging approaches from different traditions (such as "Bayesian" psycholinguistics and game theoretic and decision theoretic pragmatics) that revolve around functional explanations of pragmatic phenomena as rational or optimally adapted for a conversational purpose. We have tried to characterize probabilistic pragmatics here, hoping that these considerations may inspire conceptual reflection about pragmatic theory in general. Our main points were these.

- Formal pragmatics can benefit from applying **probabilistic** techniques. There are two aspect to this. First and foremost, probability calculus is the basis of statistics. As such it is essential for the interpretation of quantitative empirical data such as experimental results and corpus studies. However, probability theory is relevant for pragmatics also at a deeper conceptual level. Under a Bayesian-in-a-weak-sense approach, we use probabilities to represent degrees of epistemic certainty of language users. As illustrated in our case studies, this affords fine-grained analyses of pragmatic phenomena that are not easily replicable in more traditional frameworks (especially Section 6). Moreover, probability models allow integration of many sources of uncertainty and **joint inferences** about several uncertain contextual parameters in a holistic fashion. In other words, no matter what pragmatic phenomenon we are interested, given the natural uncertainty involved in contextual language use, if our aim is more than to describe an idealized, abstract picture of sentence meanings or enumerate diverse factors that influence contextual language use in some way or other; i.e., if our aim is to get a predictive but integrative and holistic theory of pragmatic phenomena, probability models may be the only generally applicable and certainly the most practicable tool on the market.

- At the level of reasons, pragmatic inference is **interactive**, i.e., it involves reasoning about the beliefs and intentions of the other interlocutor. This makes game theory, as an established mathematical framework for strategic decision making, a useful tool for formal pragmatics. As mentioned previously, only part of contemporary formal pragmatics is interactive in this sense. Characterizing pragmatic readings of sentences in terms of algebraic meaning operators, for instance, is (usually) not.

- Pragmatic behavior can be conceived as a form of **rational** interaction. This means that behavioral patterns can be explained by the assumption that there is a quantity —call it *utility*— that the interlocutors strive to maximize. Utilities can be justified by general considerations. It is worth pointing out that by taking a rationalistic stance, formal pragmatics follows a similar trajectory as other fields of (psycho-)linguistics such as the study of human sentence processing (cf. Crocker, 2005; Levy, Reali, and Griffiths, 2009). Being probabilistic, interactive and rationalistic, the approach sketched here is also **Bayesian** in the strong sense that it models interpretation as "reverse production" via Bayes' rule (Zeevat, 2014). Again, many current proposals in formal pragmatics are not rationalistic in this sense, but use algebraic operators, principles and constraints to characterize pragmatic readings of sentences.

- To state a truism, theories of formal pragmatics are to be formulated in a mathematically precise way. Less trivially, a formal specification should make it amenable to a **computational** implementation, both when comparing model predictions to empirical data and when exploring consequences of a theory via computer simulations.

- Last but not least, probabilistic pragmatics is **data-oriented**, i.e., it confronts purely conceptual theorizing with observations that go beyond the traditional introspective method. Repeated experimental observations or corpus counts are necessay in all but the most trivial situations when we want to assess non-introspective quantitative aspects of relevant data questions. There can be many reasons why any given data observation has gradient aspects; it is the job of a suitable model or theory to explain where gradience and variability come from, and any two models might disagree about strength and origin of quantitative observations (see for instance the two competing models for the "typicality data" in Section 5).

References

Anderson, John R. (1990). *The Adaptive Character of Thought*. Hillsdale, NJ: Lawrence Erlbaum.

— (1991). "Is human cognition adaptive?" In: *Behavioral and Brain Sciences* 14.3RA, pp. 471–517.

Anscombre, Jean-Claude and Oswald Ducrot (1983). *L'argumentation dans la langue*. Brussels: Mardaga.

Atlas, Jay David and Stephen Levinson (1981). "It-clefts, Informativeness, and Logical Form". In: *Radical Pragmatics*. Ed. by Peter Cole. Academic Press, pp. 1–61.

Axelrod, Robert (1984). *The Evolution of Cooperation: Revised Edition*. New York: Basic Books.

Baumann, Peter, Brady Clark, and Stefan Kaufmann (2014). "Overspecification and the Cost of Pragmatic Reasoning about Referring Expressions". In: *Proceedings of CogSci*.

Benz, Anton (2012). "Errors in Pragmatics". In: *Journal of Logic, Language and Information* 21.1, pp. 97–116.

Bergen, Leon, Roger Levy, and Noah D. Goodman (2012). "That's what she (could have) said: How alternative utterances affect language use". In: *Proceedings of the 34th Annual Meeting of the Cognitive Science Society.*

— (2014). "Pragmatic Reasoning through Semantic Inference". unpublished manuscript.

Blume, Andread and Oliver Board (2014). "Intentional Vagueness". In: *Erkenntnis* 79.4, pp. 855–899.

Blutner, Reinhard (1998). "Lexical Pragmatics". In: *Journal of Semantics* 15, pp. 115–162.

— (2000). "Some Aspects of Optimality in Natural Language Interpretation". In: *Journal of Semantics* 17, pp. 189–216.

Blutner, Reinhard and Henk Zeevat (2009). “Optimality-Theoretic Pragmatics”. In: *Papers in Pragmasemantics*. Ed. by Anton Benz and Reinhard Blutner. Vol. 51. ZAS Papers in Linguistics. Center for General Linguistics (ZAS), pp. 1–25.

Brown, Penelope and Stephen C. Levinson (1987). *Politeness: Some Universals of Language Usage*. Cambridge University Press.

Carstensen, Alexandra, Elizabeth Kon, and Terry Regier (2014). “Testing a rational account of pragmatic reasoning: The case of spatial language”. In: *Paul Bello and Marcello Guarini and Marjorie McShane and Brian Scassellati*. Ed. by Paul Bello, Marcello Guarini, Marjorie McShane, and Brian Scassellati, pp. 2009–2013.

Chemla, Emmanuel and Raij Singh (2014). “Remarks on the Experimental Turn in the Study of Scalar Implicature (Part I & II)”. In: *Language and Linguistics Compass* 8.9, pp. 373–386, 387–399.

Chierchia, Gennaro, Danny Fox, and Benjamin Spector (2012). “Scalar Implicature as a Grammatical Phenonenon”. In: *Semantics. An International Handbook of Natural Language Meaning*. Ed. by Claudia Maienborn, Klaus von Heusinger, and Paul Portner. Berlin: de Gruyter, pp. 2297–2332.

Chomsky, Noam (2011). “Language and Other Cognitive Systems: What is Special About Language?” In: *Language Learning and Development* 7, pp. 263–278.

Ciardelli, Ivano, Jeroen Groenendijk, and Floris Roelofsen (2013). “Inquisitive Semantics: A New Notion of Semantic Meaning”. In: *Language and Linguistics Compass* 7.9, pp. 459–476.

Clark, Robin (2012). *Meaningful Games*. Cambridge, MA: MIT Press.

Cobreros, Pablo, Paul Égré, David Ripley, and Robert van Rooij (2012). “Tolerant, Classical, Strict”. In: *Journal of Philosophical Logic* 41.2, pp. 347–385.

Cohen, Ariel (2009). “Probability in Semantics”. In: *Language and Linguistics Compass* 3.1, pp. 265–281.

Cresswell, M. J. (1977). “The Semantics of Degree”. In: *Montague Grammar*. Ed. by Barbara Partee. New York: Academic Press, pp. 261–292.

Crocker, Matthew (2005). “Rational models of comprehension: addressing the performance paradox”. In: *Twenty-First Century Psycholinguistics: Four Cornerstones*. Ed. by Ann Cutler. Hillsdale, NJ: Lawrence Erlbaum Associates.

Cummins, Chris (2014). “Typicality made familiar: A commentary on Geurts and van Tiel (2013)”. In: *Semantics & Pragmatics* 7, pp. 1–15.

Dalrymple, Mary, Makoto Kanazawa, Yookyung Kim, Sam Mchombo, and Stanley Peters (1998). “Reciprocal Expressions and the Concept of Reciprocity”. In: *Linguistics and Philosophy* 21.2, pp. 159–210.

Degen, Judith (2015). “Investigating the distribution of *some* (but not *all*) implicatures using corpora and web-based methods”. In: *Semantics & Pragmatics* 8.11, pp. 1–55.

Degen, Judith and Michael Franke (2012). “Optimal Reasoning About Referential Expressions”. In: *Proceedings of SemDial 2012 (SeineDial): The 16th Workshop on the Semantics and Pragmatics of Dialogue*. Ed. by Sarah Brown-Schmidt, Jonathan Ginzburg, and Staffan Larsson, pp. 2–11.

Degen, Judith, Michael Franke, and Gerhard Jäger (2013). “Cost-Based Pragmatic Inference about Referential Expressions”. In: *Proceedings of the 35th Annual Meeting of the Cognitive Science Society*. Ed. by Markus Knauff, Michael Pauen, Natalie Sebanz, and Ipke Wachsmuth. Austin, TX: Cognitive Science Society, pp. 376–381.

Degen, Judith and Noah D. Goodman (2014). “Lost your Marbles? The Puzzle of Dependent Measures in Experimental Pragmatics”. In: *Proceedings of CogSci*. Ed. by Paul Bello, Marcello Guarini, Marjorie McShane, and Brian Scassellati. Austin, TX: Cognitive Science Society.

Degen, Judith and Michael K. Tanenhaus (2011). "Making Inferences: The Case of Scalar Implicature Processing". In: *Proceedings of the 33rd Annual Conference of the Cognitive Science Society*. Ed. by L. Carlson, C. Hölscher, and T. Shipley, pp. 3299–3304.

— (2015). "Processing Scalar Implicatures: A Constraint-Based Approach". In: *Cognitive Science* 39, pp. 667–710.

Degen, Judith, Michael Henry Tessler, and Noah D. Goodman (2015). "Wonky worlds: Listeners revise world knowledge when utterances are odd". In: *Proceedings of CogSci*.

Deo, Ashwini (2015). "The Semantic and Pragmatic Underpinnings of Grammaticalization Paths: The Progressive to Imperfective Shift". to appear in *Semantics & Pragmatics*.

Doran, Ryan, Rachel E. Baker, Yaron McNabb, Meredith Larson, and Gregory Ward (2009). "On the Non-Unified Nature of Scalar Implicature: An Empirical Investigation". In: *International Review of Pragmatics* 1.2, pp. 211–248.

Fox, Danny (2007). "Free Choice and the Theory of Scalar Implicatures". In: *Presupposition and Implicature in Compositional Semantics*. Ed. by Uli Sauerland and Penka Stateva. Hampshire: Palgrave MacMillan, pp. 71–120.

Frank, Michael C. and Noah D. Goodman (2012). "Predicting Pragmatic Reasoning in Language Games". In: *Science* 336.6084, p. 998.

— (2014). "Inferring word meanings by assuming that speakers are informative". In: *Cognitive Psychology* 75, pp. 80–96.

Franke, Michael (2009). "Signal to Act: Game Theory in Pragmatics". PhD thesis. Universiteit van Amsterdam.

— (2011). "Quantity Implicatures, Exhaustive Interpretation, and Rational Conversation". In: *Semantics & Pragmatics* 4.1, pp. 1–82.

— (2012). "On Scales, Salience & Referential Language Use". In: *Amsterdam Colloquium 2011*. Ed. by Maria Aloni, Floris Roelofsen, and Katrin Schulz. Lecture Notes in Computer Science. Berlin, Heidelberg: Springer, pp. 311–320.

— (2013). "Game Theoretic Pragmatics". In: *Philosophy Compass* 8.3, pp. 269–284.

— (2014a). "Pragmatic Reasoning about Unawareness". In: *Erkenntnis* 79.4, pp. 729–767.

— (2014b). "Typical use of quantifiers: A probabilistic speaker model". In: *Proceedings of CogSci*. Ed. by Paul Bello, Marcello Guarini, Marjorie McShane, and Brian Scassellati. Austin, TX: Cognitive Science Society, pp. 487–492.

Franke, Michael and Judith Degen (2015). "Reasoning in Reference Games: Individual- vs. Population-Level Probabilistic Modeling". under review, *Cognitive Science*.

Franke, Michael and Gerhard Jäger (2014). "Pragmatic back-and-forth reasoning". In: *Pragmatics, Semantics and the Case of Scalar Implicature*. Ed. by Salvatore Pistoia Reda. Palgrave Macmillan, pp. 170–200.

Frazee, Joey and David Beaver (2010). "Vagueness is Rational under Uncertainty". In: *Amsterdam Colloquium 2009*. Ed. by Maria Aloni and Katrin Schulz. LNAI 6042. Heidelberg: Springer.

Gatt, Albert, Roger P. G. van Gompel, Kees van Deemter, and Emiel Kramer (2013). "Are we Bayesian referring expression generators?" In: *Proceedings of the 35th Annual Meeting of the Cognitive Science Society*. Ed. by Markus Knauff, Michael Pauen, Natalie Sebanz, and Ipke Wachsmuth.

Gazdar, Gerald (1979). *Pragmatics: Implicature, Presupposition, and Logical Form*. New York: Academic Press.

Geurts, Bart (2010). *Quantity Implicatures*. Cambridge, UK: Cambridge University Press.

Goeree, Jacob K., Charles A. Holt, and Thomas R. Palfrey (2008). "Quantal Response Equilibrium". In: *The New Palgrave Dictionary of Economics*. Ed. by Steven N. Durlauf and Lawrence E. Blume. Basingstoke: Palgrave Macmillan.

Goodman, Noah D. and Daniel Lassiter (2014). "Probabilistic Semantics and Pragmatics: Uncertainty in Language and Thought". Chapter draft, Stanford University.

Goodman, Noah D. and Andreas Stuhlmüller (2013). "Knowledge and Implicature: Modeling Lanuage Understanding as Social Cognition". In: *Topics in Cognitive Science* 5, pp. 173–184.

Grice, Paul Herbert (1975). "Logic and Conversation". In: *Syntax and Semantics, Vol. 3, Speech Acts*. Ed. by Peter Cole and Jerry L. Morgan. New York: Academic Press, pp. 41–58.

Groenendijk, Jeroen and Martin Stokhof (1984). "Studies in the Semantics of Questions and the Pragmatics of Answers". PhD thesis. Universiteit van Amsterdam.

Hahn, Ulrike and Michael Oaksford (2007). "The rationality of informal argumentation: A Bayesian approach to reasoning fallacies". In: *Psychological Review* 114.3, pp. 704–732.

Halpern, Joseph Y. (2003). *Reasoning about Uncertainty*. MIT Press.

Horn, Laurence R. (1972). "On the Semantic Properties of Logical Operators in English". PhD thesis. UCLA.

— (1984). "Towards a New Taxonomy for Pragmatic Inference: Q-based and R-based Implicature". In: *Meaning, Form, and Use in Context*. Ed. by Deborah Shiffrin. Washington: Georgetown University Press, pp. 11–42.

— (1989). *A Natural History of Negation*. Chicago: Chicago University Press.

Hurford, James R. (1974). "Exclusive or Inclusive Disjunction". In: *Foundations of Language* 11, pp. 40–411.

de Jaegher, Kris (2003). "A Game-Theoretic Rationale for Vagueness". In: *Linguistics and Philosophy* 26.5, pp. 637–659.

de Jaegher, Kris and Robert van Rooij (2014). "Game-Theoretic Pragmatics Under Conflicting and Common Interests". In: *Erkenntnis* 79, pp. 769–820.

Jäger, Gerhard (2012). "Game Theory in Semantics and Pragmatics". In: *Semantics. An International Handbook of Natural Language Meaning*. Ed. by Claudia Maienborn, Paul Portner, and Klaus von Heusinger. Berlin: de Gruyter, pp. 2487–2516.

Kao, Justine T., Leon Bergen, and Noah D. Goodman (2014). "Formalizing the Pragmatics of Metaphor Understanding". In: *Proceedings of CogSci*. Ed. by Paul Bello, Marcello Guarini, Marjorie McShane, and Brian Scassellati. Austin, TX: Cognitive Science Society.

Kao, Justine T., Jean Y. Wu, Leon Bergen, and Noah D. Goodman (2014). "Nonliteral Understanding of Number Words". In: *PNAS* 111.33, pp. 12002–12007.

Kennedy, Christopher and Louise McNally (2005). "Scale Structure, Degree Modification, and the Semantics of Gradable Predicates". In: *Language* 81.2, pp. 345–381.

Lassiter, Daniel and Noah D. Goodman (2014). "Context, Scale Structure, and Statistics in the Interpretation of Positive-Form Adjectives". In: *Proceedings of SALT 23*.

Lauer, Sven (2013). "Towards a Dynamic Pragmatics". PhD thesis. Stanford University.

— (2014). "Mandatory implicatures in Gricean pragmatics". In: *Proceedings of* Formal & Experimental Pragmatics 2014. Ed. by Judith Degen, Michael Franke, and Noah D. Goodman, pp. 21–28.

Lee, James J. and Steven Pinker (2010). "Rationales for indirect speech: the theory of the strategic speaker." In: *Psychological review* 117.3, pp. 785–807.

Levinson, Stephen C. (1983). *Pragmatics*. Cambridge, UK: Cambridge University Press.

— (2000). *Presumptive Meanings. The Theory of Generalized Conversational Implicature*. Cambridge, Massachusetts: MIT Press.

Levy, Roger, Florencia Reali, and Thomas L. Griffiths (2009). "Modeling the effects of memory on human online sentence processing with particle filters". In: *Proceedings of the 22nd Conference on Neural Information Processing Systems (NIPS)*.

Luce, Duncan R. (1959). *Individual Choice Behavior: A Theoretical Analysis*. New York: Wiley.
Mailath, George J. and Larry Samuelson (2006). *Repeated Games and Reputations: Long-Run Relationships*. Oxford: Oxford University Press.
Marr, David (1982). *Vision: A Computational Investigation into the Human Representation and Processing of Visual Information*. New York: Freeman.
McFadden, Daniel (1976). "Quantal Choice Analysis: A Survey". In: *Annals of Economic and Social Measurement* 5.4, pp. 363–390.
Meibauer, Jörg and Markus Steinbach, eds. (2011). *Experimental Pragmatics/Semantics*. Amsterdam, Philadelphia: John Benjamins.
Merin, Arthur (1999). "Information, Relevance, and Social Decisionmaking: Some Principles and Results of Decision-Theoretic Semantics". In: *Logic, Language and Computation*. Ed. by Lawrence C. Moss, Jonathan Ginzburg, and Maarten de Rijke. Vol. 2. CSLI Publications, pp. 179–221.
Mühlenbernd, Roland (2013). "Signals and the Structure of Societies". PhD thesis. University of Tübingen.
Myung, In Jae (2003). "Tutorial on Maximum Likelihood Estimation". In: *Journal of Mathematical Psychology* 47, pp. 90–100.
Noveck, Ira A. and Dan Sperber, eds. (2004). *Experimental Pragmatics*. Hampshire: Palgrave MacMillan.
Oaksford, Michael and Nick Chater (1994). "A Rational Analysis of the Selection Task as Optimal Data Selection". In: *Psychological Review* 101.4, pp. 608–631.
Parikh, Prashant (2001). *The Use of Language*. Stanford University: CSLI Publications.
Pavan, Sascia (2014). "Rationality in game-theoretic pragmatics". In: *Linguistics and Philosophy* 37.3, pp. 257–261.
Pinker, Steven (2007). "The evolutionary social psychology of off-record indirect speech acts". In: *Intercultural pragmatics* 4.4, pp. 437–461.
— (2008). *The Stuff of Thought: Language as a Window into Human Nature*. Penguin Publishing Group.
Pinker, Steven, Martin A. Nowak, and James L. Lee (2008). "The logic of indirect speech". In: *Proceedings of the National Academy of Sciences* 105.3, pp. 833–838.
Potts, Christopher, Daniel Lassiter, Roger Levy, and Michael C. Frank (2015). "Embedded implicatures as pragmatic inferences under compositional lexical uncertainty". Manuscript.
Qing, Ciyang and Michael Franke (2014). "Gradable Adjectives, Vagueness, and Optimal Language Use: A Speaker-Oriented Model". In: *Proceedings of SALT*. Ed. by Jessi Grieser, Todd Snider, Sarah D'Antonio, and Mia Wiegand. elanguage.net, pp. 23–41.
— (2015). "Variations on a Bayesian Theme: Comparing Bayesian Models of Referential Reasoning". In: *Bayesian Natural Language Semantics and Pragmatics*. Ed. by Henk Zeevat and Hans-Christian Schmitz. Language, Cognition and Mind. Berlin: Springer.
van Rooij, Iris, Cory D. Wright, Johan Kwisthout, and Todd Wareham (2014). "Rational Analysis, Intractability, and the Prospects of 'as if'-Explanations". In: *Synthese*.
van Rooij, Robert and Katrin Schulz (2004). "Exhaustive Interpretation of Complex Sentences". In: *Journal of Logic, Language and Information* 13, pp. 491–519.
Rothschild, Daniel (2013). "Game Theory and Scalar Implicatures". In: *Philosophical Perspectives* 27.1, pp. 438–478.
Rotstein, Carmen and Yoad Winter (2004). "Total Adjectives vs. Partial Adjectives: Scale Structure and Higher-Order Modifiers". In: *Natural Language Semantics* 12.3, pp. 259–288.
Rubinstein, Ariel (2000). *Economics and Language*. Cambridge University Press.

Russell, Benjamin (2012). “Probabilistic Reasoning and the Computation of Scalar Implicatures”. PhD thesis. Brown University.

Sauerland, Uli (2012). “The Computation of Scalar Implicatures: Pragmatic, Lexical or Grammatical”. In: *Language and Linguistics Compass* 6.1, pp. 36–49.

Schulz, Katrin and Robert van Rooij (2006). “Pragmatic Meaning and Non-monotonic Reasoning: The Case of Exhaustive Interpretation”. In: *Linguistics and Philosophy* 29, pp. 205–250.

Solt, Stephanie (2011). “Notes on the Comparison Class”. In: *Vagueness in Communication*. Ed. by Robert van Rooij, Uli Sauerland, and Hans-Christian Schmitz. Berlin: Springer, pp. 189–206.

Spector, Benjamin (2014). “Global Positive Polarity Items and Obligatory Exhaustivity”. In: *Semantics & Pragmatics* 7.11, pp. 1–61.

Sperber, Dan and Deirde Wilson (1995). *Relevance: Communication and Cognition (2nd ed.)* Oxford: Blackwell.

— (2004). “Relevance Theory”. In: *Handbook of Pragmatics*. Ed. by Laurence R. Horn and Gregory Ward. Oxford: Blackwell, pp. 607–632.

Stalnaker, Robert (2005). “Saying and Meaning, Cheap Talk and Credibility”. In: *Game Theory and Pragmatics*. Ed. by Anton Benz, Gerhard Jäger, and Robert van Rooij. Palgrave MacMillan, pp. 83–100.

von Stechow, Arnim (1984). “Comparing Semantic Theories of Comparison”. In: *Journal of Semantics* 3.1, pp. 1–77.

— (2006). “Times as Degrees: Früher, Später, and Phase Adverbs”. manuscript.

von Stechow, Arnim and Thomas Ede Zimmermann (1984). “Term Answers and Contextual Change”. In: *Linguistics* 22.1, pp. 3–40.

Stiller, Alex, Noah D. Goodman, and Michael C. Frank (2011). “Ad-Hoc Scalar Implicature in Adults and Children”. In: *Proceedings of the Thirty-Third Annual Conference of the Cognitive Science Society*.

Sutton, Richard S. and Andrew G. Barto (1998). *Reinforcement Learning*. MIT Press.

Tenenbaum, Joshua B., Thomas L. Griffiths, and Charles Kemp (2006). “Theory-Based Bayesian Models of Inductive Learning and Reasoning”. In: *Trends in Cognitive Sciences* 10.7, pp. 309–318.

Tenenbaum, Joshua B., Charles Kemp, Thomas L. Griffiths, and Noah D. Goodman (2011). “How to Grow a Mind: Statistics, Structure, and Abstraction”. In: *Science* 331, pp. 1279–1285.

van Tiel, Bob (2014). “Quantity Matters: Implicatures, Typicality, and Truth”. PhD thesis. Radboud Universiteit Nijmegen.

van Tiel, Bob and Bart Geurts (2013). “Truth and Typicality in the Interpretation of Quantifiers”. talk held at Sinn und Bedeutung 18.

van Tiel, Bob, Emiel van Miltenburg, Natalia Zevakhina, and Bart Geurts (2014). “Scalar Diversity”. Manuscript.

Train, Kenneth E. (2009). *Discrete Choice Methods with Simulation*. Cambridge, MA: Cambridge University Press.

Winter, Yoad (2001). “Plural Predication and the Strongest Meaning Hypothesis”. In: *Journal of Semantics* 18, pp. 333–365.

Zeevat, Henk (2014). *Language Production and Interpretation: Linguistics meets Cognition*. Current Research in the Semantics/Pragmatics Interface. Leiden: Brill.