6

Probabilistic pragmatics: A dialogical perspective

Bill Noble, Vladislav Maraev, and Ellen Breitholtz

6.1 Introduction

Much of communication relies on the implicit inferences that language users make as they interact. This is, broadly speaking, the focus of research in pragmatics. In recent years, pragmatic theories inspired by models of rational action and cooperation have extended the reach and precision of predictions that can be made about pragmatic behavior. Many conversational phenomena that are hard to account for using traditional pragmatic theories can potentially be covered by introducing a probabilistic component (Benz and van Rooij, 2007, Franke, 2009).

However, these efforts cast linguistic communication as discrete sequences of action/reaction, rather than comprehensive interaction events. They also often (though not always) assume a high degree of cooperation between actors in spite of the fact that pragmatic inference is possible in situations like argumentation where speakers have divergent objectives (Franke et al., 2012). In this paper we will argue that probabilistic approaches are relevant not only for modeling isolated instances of pragmatic reasoning, but also for highly contextualised reasoning situations which require agents to coordinate their linguistic resources, reason about previous utterances, and make use of common sense assumptions.

Probabilistic Approaches to Linguistic Theory.
Jean-Philippe Bernardy, Rasmus Blanck,
Stergios Chatzikyriakidis, Shalom Lappin,
Aleksandre Maskharashvili (eds.).

In Section 6.2 we will briefly introduce some key concepts in pragmatics before moving on to Section 6.3 in which we will summarise existing probabilistic approaches to pragmatics that have bearing on the type of examples we will consider further on. This includes work in game theoretic and Bayesian pragmatics with applications to social interaction as well as sociolinguistics. In Section 6.4 we argue that the approaches presented in the previous section are ill-equipped to explain pragmatic phenomena arising from natural dialogue. To remedy this, we suggest that pragmatic theories be integrated in a formal model of dialogue in which participants draw on shared principles of reasoning called *topoi*. Finally, in Section 6.5, we present a sketch of a probabilistic model of an instance of joke telling in which the humorous effect is related to the fact that there are several possible interpretations, which depending on context and the participants' respective take on the state of the interaction are more or less salient.

6.2 What is pragmatics?

The word "pragmatics" is derived from the Greek word for "act" or "do", and is thus intended to cover language use as opposed to linguistic meaning in the abstract. The concept traditionally covers a number of key topics, such as Deixis, Speech act theory, Presupposition, and Implicature.[1]

6.2.1 Pragmatic inference

One of the central issues in pragmatics is how implicit meaning is conveyed in speech or text. This kind of inference is usually discussed in terms of *presupposition* or *implicature*. Presupposition is often defined as an inference which survives embedding under negation (see for example Strawson, 1950).

...a proposition that P presupposes that Q iff Q must be true in order that P have a truth-value at all.

(Stalnaker, 1974 p.48)

According to this definition, the utterance "The Queen of England is bald" because it presupposes that there is a Queen of England; this needs to be true for the sentence to be understood as either true or false. Another type of inference that is common in linguistic interaction is *conversational implicature*. Grice (1975) formulated a theory of implicature that attempts to systematically describe how language users can convey (and mean) more than the truth-conditional

[1] For a thorough introduction to Pragmatics, see for example Levinson (1983), Verschueren and Verschueren (1999), Birner (2012)

content of an utterance. Grice distinguishes between *what is said* and *what is implicated*. The former corresponds to the truth-conditional meaning of an utterance and the latter to what a speaker conveys by uttering a certain string of words in a certain context. In this theory, implicature relies on agents who assume (though perhaps are unaware of) the *cooperative principle* elaborated as four *maxims* of rational and efficient communication. According to the maxims contributions should be truthful, as specific as necessary but not too specific, relevant, and presented in an orderly manner. In the exchange below, from Grice (1975), **B**'s reply that there is a garage around the corner would not be very helpful if **B** knew the garage to be closed, not to sell petrol, etc..

(A) 1. **A**: I am out of petrol
2. **B**: There is a garage around the corner

A expects **B**'s utterance to be a relevant, truthful and complete reply, based on background knowledge and an assumption that **B** is being cooperative.

By adhering to, or blatantly ignoring (*flouting*) the maxims and the cooperative principle, a speaker may express a lot more than the truth-conditional content of his/her utterance. Grice's theory has been extended and developed in a number of different ways, see for example Horn (1984), Levinson (2000) (Neo-Gricean) and Sperber and Wilson (1995), Carston and Hall (2012) (Relevance theory).

Pragmatic theories of inference explain how assumptions are added to the discourse context without being explicitly mentioned. This phenomenon was discussed by Stalnaker (1974) and Karttunen (1974) in the context of presupposition, but the term *accommodation* was coined by Lewis (1979) to describe a particular kind of inference. A simple example of presupposition accommodation can be found in the sentence "The Queen of England is bald", mentioned above, where the definite noun phrase "The Queen of England" presupposes the existence of a Queen of England. If a language user were to hear the sentence uttered, and not already be aware that England has a Queen, they would accommodate this presupposition and the fact that there is a Queen of England would be integrated into their information state.

This instance of accommodation is very straightforward. However, there are many cases where inferences involving a much higher degree of uncertainty are accommodated. Lewis considers accommodation as a general process not necessarily limited to presuppositions. Other types of accommodation are discussed for example in Cooper and Larsson (2010) (questions) and Breitholtz (2020) (topoi). We will return to these in Section 6.4.1.

6.3 Probabilistic approaches to pragmatics

Since pragmatics deals with what is implicit in linguistic communication, it is naturally tied up with uncertainty and especially *inference under uncertainty*.[2]

Recall the exchange in (A), in which **A** infers from **B**'s statement that there is a garage around the corner at which one can get petrol. If **A** had reason to think **B** might be lying, unknowledgable, or otherwise untrustworthy, or if they were unsure of the surface-level meaning of **B**'s utterance (for example, because of a dialectal difference in the meaning of the word *garage*), then **A** would have to take their uncertainty into account in the inference that petrol is available around the corner. Mathematically, it is natural to formalize uncertainty with probabilistic notions — probability distributions, random variables (and dependencies between them), priors, posteriors and so on.

While all of the formal pragmatic models we'll consider in this chapter deal with probability in some way, they differ in the *theoretical* role that probability plays. When uncertainty is modeled as a probability distribution over alternatives, what claim does the theorist make about the relationship between the mathematical model and the phenomena being modelled? There are at least a few different possible answers, which we characterize as follows:

- **Strong probabilistic cognitivism:** Some aspect of the linguistic agent's cognitive state effectively *is* (or is isometric to) a probability distribution. For example, the agent consciously or unconsciously assigns numerical probabilities to certain events, interpretations, etc.
- **Weak probabilistic cognitivism:** The behavior of individual agents conforms to the predictions made by a probabilistic model, whether or not there are cognitive causal factors that themselves correspond to mathematical structures from probability theory.
- **Aggregate probabilistic modeling:** Such models don't claim to capture individual cognition or behavior but rather, probability distributions models the *data* (which may be aggregated over many utterances and/or dialogues and/or speakers).

[2]Aside from uncertainty, another aspect of linguistic meaning for which probabilistic modeling is naturally suited is *vagueness*, however our focus in this chapter is on pragmatic phenomena that specifically arise from the interaction between speakers and the uncertainty that entails. For more on vagueness, see Chapter 4 of this volume.

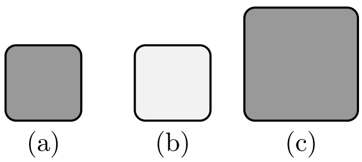
It isn't always explicit which of these modeling goals probabilistic models of pragmatics seek to achieve. Naturally, this can make results difficult to interpret from a psycholinguistic perspective—readers should be wary, for example, of taking model fit on data aggregated over multiple speakers and utterances as evidence for hypotheses stated in terms of probabilistic cognitivism. Even when the modeling goal is not explicit, it can be helpful to have these different possible goals in mind.

In the remainder of this section, we give an overview of several different research traditions in probabilistic pragmatics. Each of these strains of research are interconnected and influence each other, so there is no one natural starting place, but we begin with referring expression generation because it is the longest-running of the fields we will mention.

6.3.1 Referring expression generation

A *referring expression* is any linguistic expression that is intended to pick out a particular entity in a given context. Referring expression generation (REG) is the task of automatically generating successful referring expressions, given a perceptual scene (e.g., an image) and a target entity.

Referring expressions may use names, adjectival and noun phrases, pronouns, determiners, and various other constructions to achieve the goal of uniquely identifying the target entity. Although language users tend to be very good at coming up with successful referring expression in a given context, the procedure is far from trivial since it depends heavily on context. Suppose a speaker wanted to refer to entity a in the following scene:



They might produce the referring expression *the small dark square.* This uniquely picks out (a) since (b) is lightly shaded and (c) is larger than (a). Already we see that there is some context dependence, since the descriminativity of *small* is afforded by comparison with the other entities in the scene.

Reference also depends on dialogue context (Clark, 1996). Suppose that the same speaker-listener pair is subsequently confronted with another scene:

The small dark square would probably no longer successfully refer to (a) since there is another square which more emphatically fits that description. The speaker might now refer to (a) as *the small medium-shaded square*, but what if they wanted to refer to (d)? Without the previous dialogue context, *the small dark square* might work, but if this same expression had previously been used to refer to (a), confusion may arise. So it is that referring expressions can depend on both visual *and* discourse context. A strong dependence on context and the ability to manipulate which kinds of context are important make REG a popular test case for computational pragmatics.

Indeed, the relationship between REG and Gricean maxims was recognized from the very earliest work on REG (Appelt, 1985, Dale and Reiter, 1995), where it is noted that expressions that over-specify the target entity may, as a result of violating the maxim of quantity, generate a conversational implicature that confuses the listener.[3] Strictly following Gricean maxims, however, would not produce the most human-like expressions—on the contrary, experimental evdience suggests that humans routinely violate the maxim of quantity by over-describing referants, and that listeners have no problem with these descriptions (Engelhardt et al., 2006). In the example above, it would not be unusual for a human speaker to referto (b) as "the small orange square" even though "the orange one" would fully disambiguate it.

However, most work in REG is not explicitly concerned with the linguistic processes by which humans generate referring expressions, but rather has the goal of automatically generating expressions similar to those that humans produce under similar circumstances, or that maximize some other goal such as discriminativity. Nevertheless, REG algorithms are often inspired by pragmatic principles. Indeed, some studies have explicitly set out to use REG as a way of testing pragmatic hypotheses (see 6.3.3 for some examples).

6.3.2 Game theoretic pragmatics

Game theory is the study of a certain kind of idealized model of strategic interaction between agents. From a game theoretic perspective, interesting situations are those in which the outcome of

[3]See Krahmer and van Deemter (2012) for a survey of REG that includes some of the earliest work on the subject.

a given action depends on the actions of the other agents involved. In such situations, agents may take into account what they believe about the likely actions of other agents, as well as their preferences among the possible outcomes. Game theory seeks to give an account of the *strategies* available to agents; that is, the principles by which they select an action. In linguistic applications, the actions available to game theoretic agents represent possible utterances (for speakers) and responses to those utterances (for listeners). Listener actions are typically reducible to the semantic interpretation of the speaker's utterance, but are presented in terms of subsequent actions that depend on the listener's interpretation. These subsequent actions account for the utility received by speaker and listener as a result of the exchange.

In what follows, we will briefly introduce some concepts of game theory before moving on to discuss how they have been applied in pragmatics, paying particularly close attention to tho probabilistic aspects of those models.[4]

A key assumption in game theory is that agents are *rational*; that is, that their actions intend to bring about outcomes they prefer while minimizing costs associated with acting. As is common practice in economics, the cost of an action and the value (positive or negative) that agents derive from an outcome are measured in terms of *utility*, which is given a real number value in some abstract unit, often called *utils*. Rationality, then, is operationalized as *utility maximization* — rational agents are those who act in such a way that seeks to maximize the utility they receive from the outcome of their actions. In other words, given an agent with a set of possible actions $\mathcal{A}$, and utility function $U : \mathcal{A} \to \mathbb{R}$, the agent is assumed to select an action satisfying:[5]

$$U(a) = \max\{U(b) \mid b \in \mathcal{A}\}$$

The utility derived from a given action is not always certain. Epistemic uncertainty is uncertainty about the situation in which the action takes place, as in a poker game where the player doesn't know what cards their opponent holds. There may also be stochastic uncertainty about the outcome of an action, as in the uncertainty involved in drawing a card or rolling a die. Due to uncertainty, actions are, in general, not associated with a unitary outcome and its utility, but with a probability distribution over *possible* outcomes and their

[4]For more complete introductions to game theory with a focus on its applications to linguistics, see Benz and van Rooij (2007), Franke (2009, 2013).

[5]In cases where more than one action satisfies this criterion, either is considered rational.

associated utilities. Thus maximization of an action's *expected utility*, $E[U(a)]$ operationalizes rationality where uncertainty is involved.

The rationality assumption rules out actions whose expected utility is strictly lower than another action, regardless of the actions of other agents. Such an action is referred to as *strictly dominated* by the alternative action.

This is exemplified by the well-known *prisioner's dilemma*, which is usually introduced with a story that goes something like this: Two people are suspected of a crime. They are detained by police and questioned separately with no way to communicate. The police have offered each of the suspects a deal, which leaves them with a choice: they can *betray* the other and talk to the police, or *stay silent*. If one of the suspects chooses betrayal while the other stays silent, the one who talks to the police will be set free without spending any more time in jail, but the other suspect will spend three years in prison. If both suspects decide to betray the other, they will each be locked up for two years, but if they both stay silent, there is still enough evidence to send them each away for one year. We assume that while they can't communicate, each suspect understands the options available to everyone, and the results of each of the four possible outcomes, making it a game of *complete information*.[6]

The exact story isn't important in game theoretic terms, since what makes one game different from another is the structure of the actions available to the players and the relative utility of the different possible outcomes. This structure can be represented in so-called *normal form* as follows:

	b_1	b_2
a_1	$(-1,-1)$	$(-3,0)$
a_2	$(0,-3)$	$(-2,-2)$

In this table, $A_a = \{a_1, a_2\}$ are the actions available to agent a (*stay quiet* and *betray*, respectively) and likewise for agent b. Each cell of the table gives the utility for a and b if agent a performs action a_i and agent b performs b_j; that is, $(U_a(a_i, b_j), U_b(a_i, b_j))$.

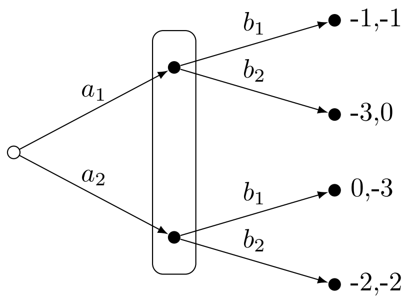
By looking at this table, we can easily see that for agent a, (sometimes referred to as the *row player*, since their action chooses between rows in the normal form table) action a_2 strictly dominates a_1 since $U_a(a_2, b_j) > U_a(a_1, b_j)$ for both $j = 1$ and $j = 2$. Since the game is symmetric, this same reasoning applies to b: each suspect derives more utility from reporting on the other, regardless of which action the

[6]Not to be confused with a game of *perfect information*, which it is not, since neither player knows what the other will do.

other one takes. This leaves a_2/b_2 as the only *Nash equilibrium*; that is, it is the only outcome in which neither player could increase their utility by unilaterally changing their strategy.[7]

The apparent paradox presented by the prisoner's dilemma is that if both players follow that course of action (which the rationality assumption implies they should), it will result in a longer cumulative prison sentence than if both players had remained silent.

The prisoner's dilemma is a classic introduction to game theory and the concept of strict domination, but since there is no communication involved, it is not particularly interesting to linguists. For modeling communication, a specific kind of game known as *sequential games* are particularly useful. In sequential games, each player acts with knowledge of the previous player's action. While it is possible to represent sequential games in normal form, the normal form representation is more naturally suited to games with one round of actions that all players perform simultaneously. For sequential games, *extensive form* representation is more common. For the sake of explication, the extensive form representation of the prisoner's dilemma is as follows:



The nodes of the tree represent possible game states, and the arrows between them are actions. Thus, if player a performs a_1, b still has a choice between actions b_1 and b_2. The simultaneous play of the prisoners dilemma setup is represented here by an *information set* in the second round of play; the box drawn around the nodes indicates that those two game states are indistinguishable by b. In game theoretic terms,

[7]The Nash equilibrium is what is known in game theory as a *solution concept*. In general, a solution concept may not necessarily result in a single solution to a given game, but stronger solution concepts eliminate more outcomes. As we will see, different solution concepts have been applied in game theoretic accounts of pragmatics.

this is equivalent to both players choosing their action simultaneously, but as we will see later, the possibility of multiple information sets is important for modeling communication.

A final concept we will introduce in this section is the distinction between *cooperative* versus *competitive* games. In a cooperative game, also known as a *coordination game*, the agent's incentives are aligned, meaning that they would like to coordinate their actions to bring about an outcome that is beneficial for both players. In a *pure coordination game*, the player's payoffs are identical in each outcome. On the other end of the spectrum, a *purely competitive* game, or *zero-sum game*, is one in which the payoffs sum to zero (or equivalently to another constant) in every outcome, meaning that the players are essentially competing for the same pool of utility. Naturally, there are many games that lie somewhere on the spectrum between cooperative and competitive. Which kind of game is most suited to modeling linguistic communication depends on how one thinks of communication more generally and the particularities of the situation being modeled.

Signalling games

Lewis (1969) introduced a class of game theoretic models called *signalling games* as a way of justifying convention as the basis for linguistic meaning. Signalling games (as they were first introduced) are collaborative sequential games of two players: a *sender*, s, and a *receiver*, r. The sender starts out with some (randomly selected) private information, $t \in T$, conventionally referred to as the sender's *type*. Intuitively, the type is the information or propositional content that the sender wants to communicate as their *type*. The sender also has a set M of messages they can send, and the receiver has a set A of actions they can perform. The actions could be a set of interpretations, in which case it may make sense to assume that $A = T$, though in the more general case the action space could be different from the set of sender types.

Since the game is sequential, the message the sender selects can depend on their t and the action the receiver chooses may depend on the message, m. Signalling games is collaborative because, at least in the original description, $U_s(t, a) = U_r(t, a)$ for all $t \in T$ and $a \in A$.

While signalling games were developed as a model of linguistic communication, the Lewisian formulas has actually been more influential in economics, to explain the role that signalling has in certain economic behaviors, such as the advertising of credentials by job-seekers (Spence, 1973), and the public actions of business owners ahead of their company's initial public offering (Brealey et al., 1977).

Evolutionary game theory

Evolutionary game theory relaxes some of the rationality assumptions imposed by the premise of traditional game theory. Instead of computing the maximum expected utility for every action, players act according to rules that are gradually updated over many rounds of play. Sometimes, as in their application to evolutionary biology, the rounds are conceived of as generations in some population of agents. In other cases, the rounds can be thought of as iterated play by a single agent, for example to model an agent learning. It is not always made explicit in linguistic applications which of these two interpretations is intended.

The first works to apply evolutionary game theory to the linguistic setting (Young, 1993, Blume et al., 1993) can be seen as an extension of Lewis (1969)'s work on convention. They sought to show how evolutionary pressures can encourage the development of systematic communication, even without (initially) conventional word meanings.

Later, this class of models received some empirical support when Jäger (2007) showed that the case marking systems in various languages correspond to different evolutionarily stable solutions in a game where communicative effort is shared by speaker with costly utterances and a hearer, who must disambiguate between possible interpretations. Expanding on this work, Jäger (2008) characterize the class of evolutionary games that will lead to stable solutions.

Game-theoretic explanations of pragmatic inferences are not necessairily incompatible with Grice's (1975) own treatment of such phenomena, since he maintains that the proposed maxims should be derivable from principles of rational communication. For example, van Rooij (2004) used evolutionary game theory to explain why "(un)marked expressions typically get an (un)marked interpretation", a phenomenon discussed by Horn (1984). Traditional Gricean accounts for such inferences rely on the *maxim of manner*, In the game theoretic model, the inference can be derived from conventions that result from repeated playing of a certain signalling game by rational agents.

Benz and van Rooij (2007) similarly use signalling games to develop a general theory of conversational implicature: By assuming that connective tissue, the receiver can work backwards to extract information that is not part of the literal content of a message.

As van Rooij and de Jager (2012) point out, the evolutionary game theoretic solution concept is strictly stronger than that of the Nash equilibrium in signaling games — many of the signalling games interpretations of Gricean pragmatic phenomena also have interpretations in evolutionary game theory.

Games of partial information

Another line of work uses *games of partial information.* These games are similar to signalling games, but situation regarding the receiver's ignorance of the speaker's type can be more complex. Parikh (2006) argues that games of partial information are strictly stronger than signalling games because the signals available to the speaker depend on the speaker's intention. The following is an example of a game of partial information game (reproduced from (Parikh, 1992); also analysed in (Parikh, 2006)).

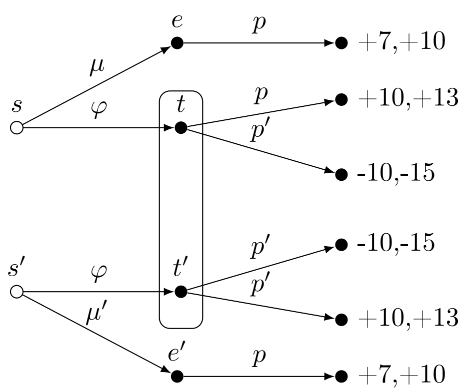
> Assume that **A**, dressed in business attire with briefcase in hand, on a street crowded with pedestrians rushing to work in midtown Manhattan at 8.30 a.m. on a Tuesday, meets a good friend, **B**, who asks him where he is headed. **A** responds with the sentence below:

(B) φ: I'm going to the bank.

This utterance, φ, is contrasted two other possible utterances:

(C) μ: I'm going to the financial bank.
μ': I'm going to the river bank.

The situation can be represented as follows:



Here, s is the situation where **A** is going to the financial bank, and p is the corresponding proposition (or more precisely, the act of **B** believing that proposition to be true). Likewise s' the situation where he is going to the river bank and p' is the corresponding belief by **B**.

Another difference with Lewisian signalling games is that different utterances are presumed to have different costs, which is why the outcomes of the μ and μ' branches give less utility for **A**: The more specific utterances are considered for some reason to be more costly.[8]

In the game above, the Nash equilibrium solution concept gives

$$\{(s, \varphi), (s', \mu'); (\{t, t'\}, p)\}$$

as the solution, meaning that unless **A** is actually going to the riverbank, they will utter the simpler but literally ambiguous φ, and B will interpret φ (correctly) as p.

In addition to conversational implicature, games of partial information have been used to analyse utterance (as opposed to sentence) meaning (Parikh, 1991), illocutionary force, miscommunication, and other pragmatic phenomena (Parikh, 2001).

Game theory for sociolinguistics

Although pragmatics and sociolingusitics are considered different disciplines, their subjec tmatter often overlaps — both pragmatics and sociolinguistics are concerned with how interactive dynamics affect interpretation in a real-world setting. For that reason, sociolinguistics is also naturally suited to game theoretic analysis. Burnett (2017, 2019) develops *social meaning games*, which seek to explain how an utterance can carry both a literal meaning and a social meaning, which conveys something about the speaker in relation to the social context.

6.3.3 Rational speech act models

In this section, we likewise attempt to provide an introduction to RSA for readers unfamiliar with the framework, but we also tread new territory by trying to systematically lay out the modeling choices RSA permits and assumptions it makes, thereby characterizing the space of pragmatic theories that RSA models well, and those that it excludes.[9]

The motivation behind rational speech acts (RSA) (Frank and Goodman, 2012) begins in the same place as Austin (1962)'s speech act theory, namely, with the observation that speech *is* action.[10] Whereas Austin's next move is to consider the different *kinds* of action that can be taken in the form of speech, RSA takes speech-as-action as a reason to apply a *rational actor* model in explaining the behavior of speakers. This move makes it possible to bring game theoretic modeling

[8]This could be, for example, because they are longer, though equating message cost to utterance length is a controversial assumption.

[9]For more extensive introductions to RSA we refer the reader to Yuan et al. (2018), Scontras et al. (2018, 2021).

[10]In this chapter, we shall distinguish between the modeling framework, which we call RSA and the original RSA model presented by Frank and Goodman (2012), which we refer to as the pragmatic listener/pragmatic speaker model.

(traditionally more common in fields like behavioral economics) to bear in the analysis of linguistic behavior (and ultimately, pragmatic meaning).

While the RSA framework can be used to model a wide array of situations in which speakers decide what to say and how to interpret their interlocutor's utterances, it is nevertheless not theory neutral. RSA makes the assumption that the source of pragmatic meaning is rational agents who reason about each other's linguistic behavior in order to communicate, and that those rational agents possess introspective transparency with respect to their own intentions. By way of exception, we note that White et al. (2020) considers the possibility of a more distant relationship between social reasoning and pragmatic behavior, developing a model in which the pragmatic behavior is first reasoned about and later routinized. The model does not strictly fit into the RSA framework, however, as it includes a neural network (the routinized part) that learns from a Bayesian model (the RSA part).

In the next section, we will try to give a general description of the RSA framework by noting where the RSA framework permits choices that may reflect different linguistic and behavioral theories. First, however, we present the original RSA model by Frank and Goodman (2012) as a place to start, from which we can consider various deviations.[11]

Vanilla RSA

The model imagines a reference game scenario, in which a speaker attempts to communicate some world state, s, with an utterance, u.[12]

The model stipulates a *literal listener*, who assigns a probability over a set of interpretations (states), given an utterance as follows:

$$P_{L_0}(s \mid u) \propto [\![u]\!](s) \cdot P(s) \tag{6.40}$$

Here, $[\![u]\!]$ is understood to be the literal meaning of u. In a classical truth theoretic semantics, this would be a function from states to truth values — indeed, this is how Frank and Goodman (2012) uses it, letting $[\![u]\!](s) = 1$ just in case u is true of s. In general, however, other interpretations are possible—Grove et al. (2021), for example, interprets $[\![u]\!]$ as a probability distribution over propositions as a way of accounting for ambiguity in the literal meaning of u. In either case, (6.40) says that the literal listener assigns probability to a given

[11]See Scontras et al. (2021) for a recent review of applications of RSA.

[12]We use the framing of *states* and *utterances* as in Scontras et al. (2021)'s excellent introduction to RSA, since it is more general than the original framing of *referents* and *words* employed by Frank and Goodman (2012).

interpretation, s, in proportion to its inclusion in the meaning of u, as well as the prior probability of s.

Next, the *pragmatic speaker* chooses an utterance based on its utility:

$$P_{S_1}(u \mid s) \propto exp(\alpha \cdot U_{S_1}(s; u)) \tag{6.41}$$

Here, U_{S_1} is the utility that the speaker derives from a literal listener's interpretation of having utterance u interpreted as s, where $C(u)$ gives the cost of utterance u:

$$U_{S_1} = logP_{L_0}(s \mid u) - C(u) \tag{6.42}$$

Unlike the literal listener, the pragmatic speaker takes their interlocutor into account when choosing an action — the utility assigned to a given utterance depends on the likelihood that they estimate the literal listener will assign to the intended meaning.

Finally, the pragmatic listener imagines that their interlocutor chose an utterance as a pragmatic speaker and assigns probability to interpretations based on yet another Bayesian inference:

$$P_{L_1}(s \mid u) \propto P_{S_1}(u \mid s) \cdot P(s) \tag{6.43}$$

Like the pragmatic speaker, the pragmatic listener takes their interlocutor's decision process into account, choosing an interpretation s in proportion to the likelihood their interlocutor would have uttered u if they intended s. Unlike the pragmatic speaker, the pragmatic listener assumes that their interlocutor is also pragmatic.

RSA variants

Deviating from the vanilla RSA presented above, there are many modeling choices that the RSA framework presents, including:

The depth of social recursion This can include choosing different depths of social recursion for the actual speaker and listener versus the mental model used by their interlocutor. In vanilla RSA, for example, the rational speaker uses the depth 1 rational listener as their mental model, which is also the model used for the *actual* listener. But it could be that the actual listener is instead modeled as a depth 2 listener.

The α parameter The α parameter is sometimes characterised as the *degree of rationality*, but it might be more accurately described as modeling an *aspect* of rationality, namely the extent to which an agent's choices are probabilistically proportional to versus determined by their relative utility.

The utility function The predictions of the RSA model may differ depending on the goals of the speaker.[13] Goodman and Stuhlmüller (2013) takes speaker uncertainty into account, using expected utility, as the objective. Potts et al. (2016) uses a cost function that depends on lexical frequency and/or the presence of alternative formulations. Finally, Qing and Franke (2015) systematically explores different utility functions.

There is some vagueness in what distinguishes RSA from other game theoretic approaches to pragmatics, but in general, the RSA tradition tends to combine game theoretic modeling with experimental pragmatics. In contrast to the theoretical work in game theoretic pragmatics (Section 6.3.2), RSA has placed less focus on identifying the appropriate solution concept for communication games, and has restricted its attention to quite simple toy scenarios in which a referring expression is selected from a limited set of alternatives. One criticism of the RSA paradigm is that the laboratory games typically used in RSA experimental work may tend to elicit rational puzzle-solving strategies in the subjects of experiments, rather than the kind of reasoning that is typical for natural dialogue. Going beyond these simple answer-response scenarios to consider situations that are more genuinely dialogical may reveal different patterns of reasoning and pragmatic inference on part of the interacting agents.

6.3.4 Challenges for probabilistic pragmatics

As described in Section 6.2, Grice (1975) proposed a set of conversational maxims as a basis for explaining pragmatic implicature. These maxims are derived from the assumption that linguistic communication takes place between agents who are perfectly rational and cooperative. Pragmatic theories in the Gricean tradition appeal to particular maxims to explain why an implicature does or does not arise in a given situation. Such explanations are *ad hoc* in the sense that there is often no meta-level theory for why a maxim should apply in one situation but not another, or what to expect when maxims are in conflict.

The approaches discussed in this section aspire to make the intuitions captured in Gricean pragmatics precise by connecting instances of linguistic behaviour more directly to the assumptions behind the maxims. In doing so, game theoretic models have demonstrated that many implicatures are preserved under conditions of imperfect cooperativity (see e.g., Franke et al., 2012). With regard to rationality,

[13] See Goodman and Frank (2016, box 1) for more examples of work that manipulates the speaker's utility function.

RSA and game theoretic models have ways of tweaking how rationality is expressed, by use of different game theoretic solution concepts van Rooij (2004), Parikh (2001) and through the α parameter and depth of recursion in the case of RSA (Zaslavsky et al., 2021). However, these models still fundamentally rely on the assumption that agents act in their own self-interest and that these interests are transparent to the agents themselves.

If the goal of pragmatic theory is to explain the behaviour of actual speakers (and not that of *homo economicus*), theories that relax or do away with the rationality assumption are necessary.[14] Game theoretic and RSA models can, however, be instrumental as *normative* models of linguistic behavior; that is, a model that predicts what an agent *should* do if they *were* acting rationally (according to a particular definition of rationality). These models may be used to represent a causal factor that is predictive (though not determinate of) the behaviour of actual speakers, and their fit to aggregate data (aggregated over time or populations) can be taken as evidence for that causal factor. Such considerations motivate the use of evolutionary game theory in pragmatics, as well as some work on RSA (White et al., 2020).[15]

The models discussed in this section are successful at doing away with Gricean maxims, instead deriving implicatures from the assumptions that underlie those maxims. However, they fall short of providing a unified meta-level theory. Instead, both game theory and RSA can be seen as frameworks through which a researcher can express a very specific linguistic situation — often restricted to a single utterance or very short exchange with limited options. It is out of scope for these frameworks to explain how the agents' incentives and potential actions arise or what effect an utterance has on the incentives and potential actions available in the *next* exchange.

Without this connective tissue, probabilistic models of pragmatics remain every bit as *ad hoc* as the Gricean approaches that preceded them.

[14] *Homo economicus* is a tongue-in-cheek moniker for the artificial subject of economic study—humans in contrast are not perfectly self-interested, introspective, or rational. Some branches of encomomics such as behavioral economics seek to study the consequences of relaxing rationality assumptions.

[15] The authors would like to thank Julian Grove for insightful conversations on the role of normative modeling in linguistic theory.

6.4 A Dialogical approach to pragmatics

Many pragmatic phenomena, such a Gricean implicatures and speech acts, have often been studied without considering a wider notion of dialogue context. The context of dialogue brings on additional constraints and considerations. For instance ones related to the goals of a particular conversational genre, such as negotiation or quarrel. In such cases cooperativeness can be violated even on the basic level (Castelfranchi, 1992). Other examples of absence of cooperativeness may include dialogues with artificial agents or neuroatypical people: even though the conversation assumes the absence of cooperation and mind-reading skills, the other participant can still interpret what has been said, or, if needed, take action to clarify it.

Gricean accounts of pragmatics assume that to understand a speaker's utterance, the hearer should recognise the *intention* that the speaker expressed by producing that utterance. One counterexample to this requirement is garden-path humour, e.g. "Should a person stir his coffee with his right hand or his left hand? Neither. He should use a spoon."[16], where the initial question posed is clearly not uttered with its bona fide intention, and if the sneaky intention is recognised correctly by the hearer, the humorous clash would not take place.

6.4.1 Reasoning in dialogue

One common problem for approaches to pragmatic reasoning in the context of dialogue modeling is that they don't pay enough attention to context. One reason for this, as suggested by Breitholtz et al. (2017), is that they are extensions of models developed to account for the truth value of sentences in terms of possible worlds rather than the meaning of utterances in interaction. Exceptions to this are for example Ginzburg (2012), Cooper (Fourthc.), Breitholtz (2020), Larsson and Traum (2000) who employ information state models of language that make it possible to account for how coordination of the dialogue progresses with successive utterances, and provides a structured characterisation of the information available to dialogue participants and offers a principled way in which asymmetries in shared knowledge can be represented.

Reasoning in dialogue often involves non-logical common-sense inferences, often referred to as *enthymematic*. An enthymeme is an argument which appeals to what is in the listener's mind. This means that unlike a syllogism, the inference presented as the conclusion of such argument is negotiable, cancellable. In (D), **A** presents an argument

[16] From Esar (1952)

that she cannot make it to a party because she is going to a wedding, but since the bride is pregnant she might be able to come later on.

(D) 1. **A**: Oh! I'm invited to a wedding that night.
2. **A**: But the bride is pregnant,
3. **A**: so I might drop by in the wee hours. (Breitholtz, 2014a, p1)

Thus, A communicates that the bride being pregnant is a *reason* for her being able to come to the party after all. To be efficient, or even understandable, this argument requires some underpinning notion or rule which sanctions it.

Such underpinning principles of reasoning have been discussed at length in the literature on rhetoric and argumentation (e.g. Toulmin, 2003, a.o.). However, the idea of rules of thumb available to language users, which justify statements, suggestions or other types of utterances goes back to early classical times. The Sophists taught their students to argue based on *topoi* — principles that had proved successful in similar types of arguments in the past (Jarratt, 1998). In modern times, the concept of topoi was introduced in linguistics as a theory of linguistic meaning with parts of discourse being connected by topoi (Ducrot, 1988). On this view the set of topoi accessible to an individual do not constitute a monolithic logical system, but represents a set of resources at the disposal of a dialogue participant for producing and interpreting arguments.

In the philosophy of language the type of reasoning involved in enthymematic arguments has been discussed in terms of implicatures (Grice, 1975, Sperber and Wilson, 1995), which are reached via assumptions of rationality and relevance. The necessity of background knowledge or common ground is not denied, but its role in a theory of pragmatic inference is often not described in a precise way. In the literature on non-monotonic logic the principles warranting conclusions are called *default rules* that is rules that are true if there is nothing to contradict them.

One of the advantages of using topoi as the underpinning for arguments, rather than default rules, is that the set of topoi of one agent does not need to be consistent or lead to consistent conclusions even within one model or domain (Breitholtz, 2020). This ability to follow various strains of reasoning — including inconsistent ones — seems to be a prerequisite for the complex type of interactive language understanding and problem solving that humans master so well.

6.4.2 Formal theories of dialogue

There are several approaches which aim at formal analyses of pragmatic inferences in a dialogical context.

KoS (Ginzburg, 2012) provides among the most detailed theoretical treatments of domain-general conversational relevance, especially for query responses — see Purver (2006) on Clarification Requests, Łupkowski and Ginzburg (2017) for a general account — which ties into the KoS treatment of non-sentential utterances. These utterances, which are often treated as performance data in standard syntactic and semantic theory, are crucial for modeling naturalistic dialogue. In this domain KoS has among the most detailed analyses (Fernández et al., 2007, Ginzburg, 2012).

Like in Lewis (1979) KoS likens language to a game, containing players (interlocutors), goals and rules. KoS represents language interaction in terms of the dynamically changing context. The meaning of an utterance is how it changes the context. In contrast to most approaches (e.g. Roberts, 2012), which represent a single context for both dialogue participants, KoS keeps a separate representation for each participant. The information state comprises a private part and a public part. The latter is called *Dialogue Game Board* (DGB) and it represents information that has been publicised in the current interaction. It tracks things that are assumed to be shared, such as visual space, moves (utterances, their form, content and illocutionary force), and questions under discussion.

KoS is based on the information-state update (ISU) approach, following several authors, including Larsson (2002) and Ginzburg (2012). In this view we present the information available to each participant of the dialogue (either a human or an artificial agent) in a rich information state. Being rich entails that the information state contains a hierarchy of facts, including the ones that are thought to be shared and the ones that have not been yet publicised.

The formal framework used in KoS is TTR, a Type Theory with Records (Cooper, 2005b, Cooper and Ginzburg, 2015). There has been a wide range of work in this formalism including the modeling of intentionality and mental attitudes (Cooper, 2005a), generalised quantifiers (Cooper, 2013), co-predication and dot types in lexical innovation, frame semantics for temporal reasoning, reasoning in hypothetical contexts (Cooper, 2011), spatial reasoning (Dobnik and Cooper, 2017), enthymematic reasoning (Breitholtz, 2014b), clarification requests (Purver, 2006, Ginzburg, 2012), negation (Cooper and Ginzburg, 2012), non-sentential utterance resolution (Fernández et al., 2007, Ginzburg, 2012) and iconic gesture (Lücking, 2016). In the case study of joke interpretation in Section 6.5 we will use a version of KoS formalised in TTR.

A brief introduction to TTR TTR is based on the notion that agents perceive an individual object that exists in the world in terms of being *of a particular type*. Such basic judgements performed by agents can be denoted as "a : Ind", meaning that a is an individual, in other words a is a *witness* of (the type) Ind(ividual). This is an example of a *basic* type in TTR, namely types that are not constructed from other types. An example of a more complex type in TTR is a *ptype* which is constructed from predicates, e.g. $\mathit{fresher\_than}(a, b)$, "$a$ is fresher than b". A witness of such a type can be a situation, a state or an event. To represent a more general event, such as "one individual item is fresher than another individual item" *record types* are used. Record types consist of a set of fields, which are pairs of unique labels and types. The record type which will correspond to the aforementioned sentence is the following:

$$\begin{bmatrix} \mathrm{x} & : & \mathit{Ind} \\ \mathrm{y} & : & \mathit{Ind} \\ \mathrm{c}_{\mathrm{fresh}} & : & \mathit{fresher\_than}(\mathrm{x}, \mathrm{y}) \end{bmatrix} \tag{6.44}$$

The witnesses of record types are *records*, consisting of a set of fields which are pairs of unique labels and values. In order to be of a certain record type, a record must contain at least the same set of labels as the record type, and the values must be of a type mentioned in the corresponding field of the record type. The record may contain additional fields with labels not mentioned in the record type. For example, the record (6.45) is of a type in (6.44) iff $a : Ind$, $b : Ind$, $s : \mathit{fresher\_than}(a, b)$ and q is of an arbitrary type.

$$\begin{bmatrix} \mathrm{x} & = & a \\ \mathrm{y} & = & b \\ \mathrm{c}_{\mathrm{fresh}} & = & s \\ \mathrm{c}_{\mathrm{price}} & = & q \end{bmatrix} \tag{6.45}$$

TTR also defines a number of type construction operations. Here we mention only the ones that are used in this chapter.

1. *List types*: if T is a type, then $[T]$ is also a type — the type of lists each of whose members is of type T. The list $[a_1, \ldots, a_n] : [T]$ iff for all i, $a_i : T$. Additionally, we use a type of non-empty lists, written as ${}_{ne}[T]$, which is a subtype of $[T]$ where $1 \leq i \leq n$. We

assume the following operations on lists: constructing a new list from an element and a list (cons), taking the first element of list (head), and taking the rest of the list (tail).

2. *Function types*: if T_1 and T_2 are types, then so is $(\lambda r : T_1.T_2)$, the type of functions from records of type T_1 to record type T_2. Additionally, T_2 may *depend* on the parameter (the witness of type T_1 passed to the function).

We model pragmatic arguments (enthymemes) and principles of reasoning (topoi) as function types from a situation of one type to another type of situation. For example, the function below has an antecedent type (the domain of the function) representing a type of situation where there are two objects and one of them is fresher than the other, and a result type (the co-domain of the function) representing a type of situation where the fresher object is better.

$$\lambda r : \begin{bmatrix} \text{x} & : \mathit{Ind} \\ \text{y} & : \mathit{Ind} \\ \text{c}_{\text{fresher}} & : \mathit{fresher\_than}(\text{x}, \text{y}) \end{bmatrix} . \begin{bmatrix} \text{c}_{\text{better}} : \mathit{better\_than}(r.\text{x}, r.\text{y}) \end{bmatrix} \tag{6.46}$$

Following Ginzburg (2012) and Larsson (2002) we will model the progress of dialogues in terms of the *information states* of the dialogue participants. In our analysis we will focus on the part of a dialogue participant's information state that is shared. That is, what has in some way been referred to in the dialogue, or what is necessary to integrate in the information state for a dialogue contribution to be interpreted in a relevant way. We will refer to this shared part of an interlocutor's information state as the DGB of that participant. We are particularly interested in how individual agents draw on individual (and sometimes distinct) resources. We will therefore use separate DGBs for each agent, rather than letting the DGB represent a God's eye notion of context. For example, although a topos may be of central relevance in the dialogue, it does not appear on the DGB until it has been made explicit, or until something has been said which has caused it to be accommodated. However, we also assume that a dialogue participant has access to a set of *rhetorical resources* which are not necessarily assumed to be shared. These are topoi that the speaker whose information state we model has access to, but that cannot be expected to be shared yet based on what has been said in the dialogue. This means of course that many topoi related to common knowledge and common sense are not considered shared in the dialogue, i. e. part of the DGB, until they have been introduced explicitly or by means of accommodation. We model the information state of an agent as a record type where the label DGB is associated with a record type of features that are assumed to be shared (these could be "latest utterance", "question under discussion" etc. — here we will only spell out the features that are focused in our analysis). The label "rhet_resources" represents the private part of an agent's information state where the agent can find topoi to invent arguments and warrant inferences.

$$\begin{bmatrix} \text{rhet\_resources} : \begin{bmatrix} \text{topoi} : [\, \textit{Topos}] \end{bmatrix} \\ \text{dgb} : \begin{bmatrix} \text{eud} : [\textit{Enthymeme}] \\ \text{topoi} : [\, \textit{Topoi}] \end{bmatrix} \end{bmatrix} \tag{6.47}$$

6.5 Case study: Humour

In Section 6.3 we discussed the main lines of research that apply probabilistic models to pragmatic phenomena. These theories focus on accounting for isolated instances of reasoning that are not embedded in dialogical context. This is also true of probabilistic models of sociolinguistic phenomena. Such models look to statistically account for particular linguistic choices made by an agent in the context of a language user at a specific point in time (Burnett, 2019). In Section 6.4, we argued that probabilistic models of pragmatic reasoning must be situated in a dialogical framework to incorporate a richer account of context — one that includes the interactive situation and background of the participants, and which is incrementally refined over the course of an interaction. In this section we will consider examples of reasoning in interaction where one speaker is engaging the other in a joke. Humour is a linguistic activity that relies on both dialogicity and reasoning under uncertainty. In humour, different aspects of context and reasoning are highly integrated, and the context changes incrementally as a result of the interaction. Moreover, it is an activity where the participants' priors affect how likely they are to "get the joke", and where each addition of new information to the discourse model adjust these priors.

6.5.1 Humour and inference

There is a long tradition of analysing humour from a linguistic perspective, most notably in the Semantic-Script Theory of Verbal Humour, SSTH (Raskin, 1985). Interpreting humour requires inherent dialogicity and multiple perspective taking, which has been described using the notion of different "story worlds" available to the interlocutors (Ritchie, 2018) and formally addressed using enthymematic reasoning drawing on topoi (Breitholtz and Maraev, 2019, Maraev et al., 2021).

A humorous effect is often created when one inference gives way to an inference arising from an alternate interpretation. The shift in interpretations can be explained by accommodating different topoi. Let us consider an example:

(E) "Is the doctor at home?" the patient asked in his bronchial whisper. "No," the doctor's young and pretty wife whispered in reply. "Come right in!" (Raskin, 1985, p. 109)

This joke (which is widely discussed in the literature on humour) takes advantage of the fact that there are two different topoi that are evoked at different stages of its progression. After the first part we are told that there is a man who is speaking with a whispering voice who is looking for a doctor. A reasonable explanation for this would be that the man is in need of a doctor, which is invoked by the use of "bronchial" which has disease associations, like "bronchitis". An underpinning topos at this stage would be something like "if the doctor is at home, he will help the patient". However, when we learn that the doctor is not at home and that his wife asks the guy at the door to come in, a more applicable topos might be something along the lines of "if the husband is not at home his wife may carry on an affair". This interpretation of the situation might become more likely when the listener learns that the person who answers the door is the doctor's wife and that she is young and pretty. However, the real shift happens when we learn that the doctor *not* being at home is presented as a reason for the man to come inside. We argue that in order for a joke like this to work there must be at least two ways of interpreting the situation, and one of them needs to be more likely to the person who is presented with the joke in order for them to be somewhat surprised by the eventual shift to the other interpretation.

The relevance of probabilities for the interpretation of jokes has been noted by Kao et al. (2013), who present probabilistic model of sentence comprehension accounting for the interpretation of jokes.[17]Crucially, a probabilistic model for this explanation of humor cannot rely on a population-level statistics, since the humorous effect is achieved in the individual speakers.

6.5.2 Probabilistic rhetorical resources

In the previous section we showed how competing available topoi which invite very different inferences in a particular situation contribute to humour. We will now move on to consider how uncertainty is

[17]However, their study is restricted to homophone puns, and does not assume an interactive perspective.

relevant to these inferences and how the humorous effect is in fact often reached by adding information that makes one topos more likely to be accommodated, rendering other acceptable topoi less likely and making the punchline more effective.

The humour arising from (E) can be explained by a shift from accommodating one topos to accommodating an entirely different topos. However, other jokes rely on some lingering uncertainty in the respective inferences.

Imagine a conversation such as the one below where **A** is telling **B** a joke. This joke elicits a humorous effect in **B** in two places — once after turn 3, where **B** produces a light chuckle and once (presumably) after turn 7, which might be considered to be the joke's ultimate punchline.

(F)

1	**A**	How do you put an elephant into a fridge?
2	**B**	Hmm, I don't know?
3	**A**	Open the door, put the elephant inside, close the door.
4	**B**	Haha okay
5	**A**	How do you put a giraffe into the fridge?
6	**B**	Open the door, put the giraffe inside, close the door?
7	**A**	Wrong! Open the door, get the elephant out, put the giraffe inside, close the door.

Breitholtz and Maraev (2019) analyse this joke in terms of salient and non-salient topoi. However, this analysis considers topoi as either applicable or non-applicable, whereas in fact the priors associated with the topoi are not likely to be 1 or 0 but somewhere in between. An account that considers this kind of gradience is more realistic and can also explain why some people may respond in a way that ruins the joke — it is because their priors are not the ones expected in the set up of the joke.

The success of the joke relies not only on **A**'s utterances but also on **B** participating in and contributing to the interaction. For this reason there are ways in which the joke might fail to have a humorous effect. For example, the joke relies on **B** making certain inferences rather than others based on different principles of reasoning, or topoi, which are all available to **B**. The success of the joke depends on the salience of these topoi. Furthermore, the humorous effect of turn 7 relies on the setup given by the rest of the dialogue. We want to explain how this setup affects which principles of reasoning are most salient and how they are subverted by turn 7 — in other words, we want to explain why **A** expects **B** to find the punchline funny.

The topic of the first question "how do you put an elephant into a fridge?" evokes a set of topoi in the mind of the listener (in the information state of our model these appear in the field "rhet_resources"). These are basic principles based on spatial relations such as "if you open a container and put an object in, the object will be in the container", "in order to fit an object in a container there must be enough space" and also topoi concerning fridges and elephants such as "a fridge is smaller than an elephant". We represent these principles of reasoning merged in one topos τ_c encompassing all of them. They act as constraints represented as *ptypes*. For instance, in the field $\mathrm{c_{space}}$ of (6.48). However, all of the constraints introduced by the relevant topoi will not be regarded as equally important. To capture this we assign weights to the constraints in the topos.

$$\tau_c = \lambda r : \begin{bmatrix} x & : \mathit{Ind} \\ y & : \mathit{Ind} \\ z & : \mathit{Ind} \\ p_1 & : \mathit{Float} \\ p_2 & : \mathit{Float} \\ \mathrm{c_{obj}} & : \mathit{object}(z) \\ \mathrm{c_{space}} & : \mathit{has\_enough\_space\_for}(x, z, p_1, p_2) \\ \mathrm{c_{fridge}} & : \mathit{fridge}(\mathrm{x}) \\ \mathrm{c_{agent}} & : \mathit{agent}(\mathrm{y}) \\ \mathrm{c_{open}} & : \mathit{open}(\mathrm{y}, \mathrm{x}) \\ \mathrm{c_{put}} & : \mathit{put\_in}(\mathrm{y}, \mathrm{z}, \mathrm{x}) \end{bmatrix} . \quad (6.48)$$
$$\big[\mathrm{s} : \mathit{in}(r.\mathrm{z}, r.\mathrm{x})\big]$$

We can assign additional parameters p_1 and p_2 to a constraint which determine probabilities, correspondingly, that the constraint holds and that it is relevant for the situation. For current purposes we treat these probabilities as independent. Given the context of the question "How do you put an elephant into a fridge?" we can say that the issue whether the container (a fridge) has enough space for fitting an object (an elephant) is important for answering it and that in most cases fridges will not have enough space for elephants. An aspect of this imaginary situation would look as follows:

$$\big[\ \mathrm{s_{space}} \ = \ \mathit{has\_enough\_space\_for}(f, e, 0.001, 0.988)\ \big],$$

where f and e are imaginary instances of a fridge and an elephant.

The applicability of a topos to a situation can be derived from the probabilities assigned to its constraints. For instance, one can characterise a topos as applicable if probabilities for its constraints are above a certain threshold (say, 0.5 for this example). Therefore, τ_c is not applicable.

In contrast, in some other context, either a narrative context of a fiction or a different cultural context, the priors for the c_{space} constraint can be different. For instance, if the joke is told in a context of a fairy tale where things can shrink in their size all the time, the likelihood of this constraint to hold would be much higher, e.g.

$$\left[\ \mathrm{s_{space}} \quad = \quad has\_enough\_space\_for(f, e, 0.643, 0.988)\ \right].$$

In this case the topos τ_c can be applied and the listener can just answer something along the lines of "Well, just put the elephant inside", which would ruin the joke. Similarly if the doctor's wife joke (E) is told in the context where affairs and lovers were just recently discussed, it might raise the probability of the lover topos from the beginning and diminish the surprise effect of the punchline, hence making the joke less funny.

6.6 Conclusion

In this paper we have considered the use of probabilistic methods in pragmatics, particularly with respect to the pragmatics of dialogue. First, we gave a brief account of the type of problems treated in pragmatics, focusing on inference. We then moved on to consider probabilistic approaches to pragmatics, focusing on RSA models and game theoretic models. Models in these traditions give a more rigorous account of pragmatic phenomena than traditional Gricean analyses. By taking seriously the idea that speech is action (Austin, 1962), they are able to apply mathematical tools that have been developed to model human behavior in disciplines such as economics. However, these models still fundamentally rely on the assumption that agents act in their own self-interest and that these interests are transparent to the agents themselves. On the contrary, much of human linguistic interaction is not particularly goal-directed.

Structurally, RSA and game-theoretic models don't resemble dialogue, which is the most basic and ubiquitous of all linguistic activity — in particular, they are non-incremental, only considering linguistic inferences that happen between discrete utterances selected from a finite action space. They also typically don't consider interactions beyond a single speaker-listener exchange and while extensions to multiple utterances are possible, they quickly become intractable, rendering them unsuitable for the analysis of a full dialogue. While RSA and game-theoretic models have a lot of explanatory power as a tool for analysing data collected in controlled experimental settings where parameters are reduced to a minimum, they do not handle natural dialogue very well.

When dialogue participants reason, their inferences tend to be based on common sense assumptions that are not universally rational. However, the assumptions, or topoi, could be more or less rational and thus more or less probable in the context of a particular situation involving a particular agent. This may lead her to assume one salient topos rather than another when drawing inferences. One type of dialogue which relies on the perceived probabilities of salient topoi is dialogue involving humour, such as dialogical jokes. In the final section of the paper we sketched a model of how probabilities can be introduced in a model of dialogue with a rich representation of context. In our sketch we focused on how the priors of a topos change during the course of the interaction, accounting for how a topos that was highly salient at the outset of a joke ends up having a very low probability towards its end. Although our model is merely a sketch, we believe it suggests that probabilistic approaches are indeed useful also in accounting for natural "messy" dialogue. However, more work — empirical as well as formal — is required to develop this model to integrate more aspects of natural dialogue while retaining the coverage of well-studied pragmatic phenomena.

References

Appelt, Douglas E. 1985. *Planning English Sentences*. Studies in Natural Language Processing. Cambridge: Cambridge University Press. ISBN 978-0-521-43803-2.

Austin, J. L. 1962. *How to Do Things with Words*. Cambridge: Harvard University Press.

Benz, Anton and Robert van Rooij. 2007. Optimal assertions, and what they implicate. A uniform game theoretic approach. *Topoi* 26(1):63–78.

Birner, Betty J. 2012. *Introduction to pragmatics*, vol. 38. John Wiley & Sons.

Blume, Andreas, Yong-Gwan Kim, and Joel Sobel. 1993. Evolutionary Stability in Games of Communication. *Games and Economic Behavior* 5(4):547–575.

Brealey, Richard, Hayne E. Leland, and David H. Pyle. 1977. Informational Asymmetries, Financial Structure, and Financial Intermediation. *The Journal of Finance* 32(2):371–387.

Breitholtz, Ellen. 2014a. *Enthymemes in Dialogue: A micro-rhetorical approach*. Ph.D. thesis, University of Gothenburg.

Breitholtz, Ellen. 2014b. Reasoning with topoi–towards a rhetorical approach to non-monotonicity. In *Proceedings of the 50th anniversary convention of the AISB, 1st–4th April 2014, Goldsmiths, University of London*.

Breitholtz, Ellen. 2020. *Enthymemes and topoi in dialogue: The use of common sense reasoning in conversation*, vol. 41 of *Current Research in the Semantics/Pragmatics*. Brill.

Breitholtz, Ellen, Christine Howes, and Robin Cooper. 2017. Incrementality all the way up. In *Computing Natural Language Inference Workshop at the International Conference on Computational Semantics (IWCS)*.

Breitholtz, Ellen and Vladislav Maraev. 2019. How to put an elephant in the title: Modeling humorous incongruity with topoi. In *Proceedings of the 23rd Workshop on the Semantics and Pragmatics of Dialogue - Full Papers*. London, United Kingdom: SEMDIAL.

Burnett, Heather. 2017. Sociolinguistic interaction and identity construction: The view from game-theoretic pragmatics. *Journal of Sociolinguistics* 21(2):238–271.

Burnett, Heather. 2019. Signalling games, sociolinguistic variation and the construction of style. *Linguistics and Philosophy* .

Carston, Robyn and Alison Hall. 2012. Implicature and explicature. In H.-J. Schmid, ed., *Cognitive Pragmatics*, chap. 3, pages 47–84. Berlin/Boston: de Gruyter.

Castelfranchi, Cristiano. 1992. No more cooperation, please! In search of the social structure of verbal interaction. In *Communication from an artificial intelligence perspective*, pages 205–227. Springer.

Clark, Herbert H. 1996. *Using Language*. Cambridge University Press.

Cooper, Robin. 2005a. Austinian truth, attitudes and type theory. *Research on Language and Computation* 3(4):333–362.

Cooper, Robin. 2005b. Records and record types in semantic theory. *Journal of Logic and Computation* 15(2):99–112.

Cooper, Robin. 2011. Copredication, quantification and frames. In S. Pogodalla and J.-P. Prost, eds., *Logical Aspects of Computational Linguistics (LACL 2011)*. Springer.

Cooper, Robin. 2013. Clarification and generalized quantifiers. *Dialogue and Discourse* 4:1–25.

Cooper, Robin. Fourthc. From perception to communication: An analysis of meaning and action using a theory of types with records (TTR). https://sites.google.com/site/typetheorywithrecords/drafts.

Cooper, Robin and Jonathan Ginzburg. 2012. Negative inquisitiveness and alternatives-based negation. In *Logic, Language and Meaning*, pages 32–41. Springer.

Cooper, Robin and Jonathan Ginzburg. 2015. Type theory with records for natural language semantics. *Handbook of Contemporary Semantic Theory, The* pages 375–407.

Cooper, Robin and Staffan Larsson. 2010. Accommodation and reaccommodation in dialogue. *R. Bäuerle, U. Reyle, & TE Zimmermann (Hg.): Presuppositions and discourse. Essays offered to Hans Kamp. Bingley, UK: Emerald Group Publishing. S* pages 101–124.

Dale, Robert and Ehud Reiter. 1995. Computational Interpretations of the Gricean Maxims in the Generation of Referring Expressions. *Cognitive Science* 19(2):233–263.

Dobnik, Simon and Robin Cooper. 2017. Interfacing language, spatial perception and cognition in type theory with records. *Journal of Language Modelling* 5(2):273–301.

Ducrot, Oswald. 1988. Topoï et formes topique. *Bulletin d'études de la linguistique française* 22:1–14.

Engelhardt, Paul E, Karl GD Bailey, and Fernanda Ferreira. 2006. Do speakers and listeners observe the gricean maxim of quantity? *Journal of memory and language* 54(4):554–573.

Esar, Evan. 1952. *The treasury of humorous quotations*. Phoenix House.

Fernández, Raquel, Jonathan Ginzburg, and Shalom Lappin. 2007. Classifying ellipsis in dialogue: A machine learning approach. *Computational Linguistics* 33(3):397–427.

Frank, Michael C. and Noah D. Goodman. 2012. Predicting pragmatic reasoning in language games. *Science (New York, N.Y.)* 336(6084):998.

Franke, M. 2009. *Signal to Act: Game Theory in Pragmatics*. AmsterdamInstitute for Logic, Language and Computation. ISBN 978-90-5776-202-4.

Franke, Michael. 2013. Game Theoretic Pragmatics. *Philosophy Compass* 8(3):269–284.

Franke, Michael, Tikitu De Jager, and Robert Van Rooij. 2012. Relevance in cooperation and conflict. *Journal of Logic and Computation* 22(1):23–54.

Ginzburg, Jonathan. 2012. *The interactive stance*. Oxford University Press.

Goodman, Noah D. and Michael C. Frank. 2016. Pragmatic Language Interpretation as Probabilistic Inference. *Trends in Cognitive Sciences* 20(11):818–829.

Goodman, Noah D. and Andreas Stuhlmüller. 2013. Knowledge and Implicature: Modeling Language Understanding as Social Cognition. *Topics in Cognitive Science* 5(1):173–184.

Grice, H. P. 1975. Logic and Conversation. In P. Cole, ed., *Speech Acts*, no. 3 in Syntax and Semantics, pages 44–55. New York: Acad. Pr, 5th edn. ISBN 978-0-12-785423-6.

Grove, Julian, Jean-Philippe Bernardy, and Stergios Chatzikyriakidis. 2021. From compositional semantics to Bayesian pragmatics via logical inference. In *Proceedings of the 1st and 2nd Workshops on Natural Logic Meets Machine Learning (NALOMA)*, pages 60–70. Groningen, the Netherlands (online): Association for Computational Linguistics.

Horn, Laurence. 1984. Toward a new taxonomy for pragmatic inference: Q-based and r-based implicature. *Meaning, form, and use in context* 42.

Jäger, Gerhard. 2007. Evolutionary Game Theory and Typology: A Case Study. *Language* 83(1):74–109.

Jäger, Gerhard. 2008. Evolutionary stability conditions for signaling games with costly signals. *Journal of Theoretical Biology* 253(1):131–141.

Jarratt, Susan C. 1998. *Rereading the sophists: Classical rhetoric refigured*. SIU Press.

Kao, Justine T, Roger Levy, and Noah D Goodman. 2013. The funny thing about incongruity: A computational model of humor in puns. In *Proceedings of the Annual Meeting of the Cognitive Science Society*, vol. 35.

Karttunen, Lauri. 1974. Presupposition and linguistic context. *Theoretical linguistics* 1(1):181–194.

Krahmer, Emiel and Kees van Deemter. 2012. Computational Generation of Referring Expressions: A Survey. *Computational Linguistics* 38(1):173–218.

Larsson, Staffan. 2002. *Issue-based dialogue management*. Ph.D. thesis, University of Gothenburg.

Larsson, Staffan and David R Traum. 2000. Information state and dialogue management in the trindi dialogue move engine toolkit. *Natural language engineering* 6(3-4):323–340.

Levinson, Stephen C. 1983. *Pragmatics*. Cambridge University Press.

Levinson, Stephen C. 2000. *Presumptive meanings: The theory of generalized conversational implicature*. MIT Press.

Lewis, David. 1979. Scorekeeping in a language game. *Journal of Philosophical Logic* 8(1):339–359.

Lewis, David K. 1969. *Convention: A Philosophical Study*. Wiley-Blackwell.

Lücking, Andy. 2016. Modeling co-verbal gesture perception in type theory with records. In *Proceedings of the 2016 Federated Conference on Computer Science and Information Systems*, pages 383–392.

Łupkowski, Paweł and Jonathan Ginzburg. 2017. Query responses. *Journal of Language Modelling* 4(2):245–292.

Maraev, Vladislav, Ellen Breitholtz, Christine Howes, Staffan Larsson, and Robin Cooper. 2021. Something old, something new, something borrowed, something taboo: Interaction and creativity in humour. *Frontiers in Psychology* 12:1443.

Parikh, Prashant. 1991. Communication and strategic inference. *Linguistics and Philosophy* 14(5):473–514.

Parikh, Prashant. 1992. A game-theoretical account of implicature. In Y. Vardi, ed., *Proceedings of the 4th Conference on Theoretical Aspects of Rationality and Knowledge*, pages 85–94. Monterey, California.

Parikh, Prashant. 2001. *The Use of Language*. CSLI Lecture Notes. Stanford, Calif: CSLI Publications. ISBN 978-1-57586-354-2 978-1-57586-355-9.

Parikh, Prashant. 2006. Pragmatics and Games of Partial Information. In A. Benz, G. Jäger, and R. van Rooij, eds., *Game Theory and Pragmatics*, Palgrave Studies in Pragmatics, Language and Cognition, pages 101–122. London: Palgrave Macmillan UK. ISBN 978-0-230-28589-7.

Potts, Christopher, Daniel Lassiter, Roger Levy, and Michael C. Frank. 2016. Embedded Implicatures as Pragmatic Inferences under Compositional Lexical Uncertainty. *Journal of Semantics* 33(4):755–802.

Purver, M. 2006. CLARIE: Handling clarification requests in a dialogue system. *Research on Language & Computation* 4(2):259–288.

Qing, Ciyang and Michael Franke. 2015. Variations on a Bayesian Theme: Comparing Bayesian Models of Referential Reasoning. In H. Zeevat and H.-C. Schmitz, eds., *Bayesian Natural Language Semantics and Pragmatics*, Language, Cognition, and Mind, pages 201–220. Cham: Springer International Publishing. ISBN 978-3-319-17064-0.

Raskin, Victor. 1985. *Semantic mechanisms of humor*. Synthese language library, 24. Dordrecht: Reidel. ISBN 9027718210.

Ritchie, Graeme. 2018. *The Comprehension of Jokes: A Cognitive Science Framework*. Routledge.

Roberts, Craige. 2012. Information structure: Towards an integrated formal theory of pragmatics. *Semantics and Pragmatics* 5:6–1.

Scontras, Gregory, Tessler Michael Henry, and Michael Franke. 2018. Probabilistic language understanding: An Introduction to the Rational Speech Act Framework. https://www.problang.org/.

Scontras, Gregory, Michael Henry Tessler, and Michael Franke. 2021. A practical introduction to the Rational Speech Act modeling framework. *arXiv:2105.09867 [cs]* .

Spence, Michael. 1973. Job Market Signaling. *The Quarterly Journal of Economics* 87(3):355–374.

Sperber, D. and D. Wilson. 1995. *Relevance: Communication and cognition*. Blackwell, 2nd edn.

Stalnaker. 1974. Pragmatic presuppositions. In M. K. Munitz and P. Unger, eds., *Semantics and Philosophy*. New York University Press.

Strawson, Peter F. 1950. On referring. *Mind* 59(235):320–344.

Sutton, Peter. 2022. The origins of vagueness. In J.-P. Bernardy, R. Blanck, S. Chatzikyriakidis, S. Lappin, and A. Maskharashvili, eds., *Probabilistic Approaches to Linguistic Theory*. CSLI Publications.

Toulmin, Stephen E. 2003. *The uses of argument*. Cambridge University Press.

van Rooij, Robert. 2004. Signalling Games Select Horn Strategies. *Linguistics and Philosophy* 27(4):493–527.

van Rooij, Robert and Tikitu de Jager. 2012. Explaining Quantity Implicatures. *Journal of Logic, Language and Information* 21(4):461–477.

Verschueren, Jef and Jan Verschueren. 1999. *Understanding pragmatics*. Oxford University Press.

White, Julia, Jesse Mu, and Noah D. Goodman. 2020. Learning to refer informatively by amortizing pragmatic reasoning. *CogSci* .

Young, H. Peyton. 1993. The Evolution of Conventions. *Econometrica* 61(1):57.

Yuan, Arianna, Will Monroe, Yu Bai, and Nate Kushman. 2018. Understanding the Rational Speech Act model. *CogSci* pages 2759–2764.

Zaslavsky, Noga, Jennifer Hu, and Roger Levy. 2021. A Rate–Distortion view of human pragmatic reasoning. *Proceedings of the Society for Computation in Linguistics* 4(1):347–348.