What is Game Theoretic Pragmatics?

Michael Franke

December 17, 2008

Abstract

Models and methods of rational choice theory naturally suggest themselves as excellent candidates for formal accounts of pragmatic inferences conceived in a Gricean fashion as the result of interpreting language use as rational human action. This paper spells out a particular way of using game theory in linguistic pragmatics and assesses weaknesses and strength of the suggested approach. The paper outlines the main research questions of game theoretic pragmatics and introduces (i) dynamic games of incomplete information, with emphasis on signaling games, as formal models of the context of utterance and (ii) perfect Bayesian equilibrium as a rationalistic solution concept for signaling games.

1 Linguistic Pragmatics & the Gricean Programme

It is a near-platitude that under normal circumstances we reliably learn more from observing the honest *utterance* of a declarative sentence[1] than we would learn from the direct observation of infallible evidence that the proposition expressed by that sentence was true. If John stands by the window and says

(1) It's raining

we learn more from his utterance than what we would learn from a glimpse of the wet street outside (assuming for the sake of argument that this counts as infallible evidence for rain). Of course, if John is honest and reliable, we do learn that it is raining from his utterance, just as we would from observation. But depending on the concrete circumstances, John's utterance, but certainly not the observation of the wet street outside, might also inform us that

[1] Although declarative sentences have received by far the most attention, similar remarks could be made about non-declarative sentences, phrases, words, gestures or any other kind of ostensive behavior with a sufficient history of preceding uses to bestow an element of commonly expected meaningfulness to it.

(2) a. John advises we should take an umbrella, or that

b. John (hereby) declares the picnic cancelled, or that

c. John is sick of living in Amsterdam.

These are non-trivial pieces of information that we might acquire as proficient interpreters that go way beyond the meaning of the sentence "It's raining." So where does this information come from? Why is such surplus information reliably inferred and communicated? What role does the conventional, semantic meaning of an utterance play in the process of fully understanding it? What features of the context of an utterance are important for its interpretation? These are the kind of questions that LINGUISTIC PRAGMATICS tries to raise, sharpen and answer.

Gricean Pragmatics. One way of approaching the difference between utterance and observation is to see an utterance clearly as an instance of human *action*, and as such subject it to commonsense conceptualization in terms of the speaker's beliefs, preferences and intentions. From this point of view, we may conceive of linguistic pragmatics as an investigation into the systematic relationship between the conventional, semantic meaning of a linguistic token and the overall significance that it may acquire when put to use in human action in a concrete context.[2]

It clearly has a certain appeal to distinguish aspects of meaning that belong to the meaningful sign proper and those that arise from the reasons and ends for which a meaningful sign is used. For instance, we would not want to hold that the sentence (1) itself contains ambiguously all the possible further shades of meaning it might acquire in special contexts. This is because the list of such special contextualized meanings would be enormous if not infinite and moreover fairly irregular so as to undermine any reasonable concept of semantic meaning. This is not only so for very context-dependent inferences like those in (2), but also for inferences that appear rather rule-like — inferences that are tied closely, for instance, to the use of a particular lexical item. A standard example here is the quantifier phrase "some." In most situations an utterance of the sentence (3a), may reliably convey the inference in (3b).[3]

[2]This view of pragmatics still resembles the distinction of semiotic subdisciplines into syntax, semantics and pragmatics which was introduced by Charles M. Morris: while syntax studies the relation between signs, and semantics the relation between signs and objects, pragmatics "deals with the origins, uses, and effects of signs within the total behavior of the interpreters of signs" (Morris 1946, p. 219).

[3]To be precise, the inference that sentence (3a) gives rise to has either a stronger or a weaker

(3) a. I saw some of your children today.

b. The speaker did not see *all* of the hearer's children today.

But would we want to say that "some" is semantically ambiguous between "some and possibly all" and "some and not all"? Preferably not, many philosophers of language have argued, because, among other things, the attested inference can be easily *cancelled* as in (4), whose consistency would be hard to explain if "some and not all" was a part of the semantic meaning of the phrase "some".

(4) I saw some of your children today, and maybe even all of them.

The case against a lexical ambiguity in the meaning of "some", for instance, has already been made by John Stuart Mill in the 19$^{\text{th}}$ century in a response to an ambiguity thesis proposed by William Hamilton:

> "No shadow of justification is shown (...) for adopting into logic a mere sous-entendu of common conversation in its most unprecise form. If I say to any one, 'I saw some of your children to-day', he might be justified in inferring that I did not see them all, not because the words mean it, but because, if I had seen them all, it is most likely that I should have said so: even though this cannot be presumed unless it is presupposed that I must have known whether the children I saw were all or not." (Mill 1867)

Roughly a century later, Herbert Paul Grice reiterated Mill's position in his William James Lectures presented at Harvard in 1967. In a condensed formulation that has become known as Grice's *Modified Occam's Razor* he demanded that "senses are not to be multiplied beyond necessity" (Grice 1989, p. 47).[4] Grice's main contribution to a defense of parsimony in logical semantics was

---

epistemic reading (Soames 1982):

(1) The speaker does not know/believe that she saw all of the hearer's children.

(2) The speaker does not know whether she saw all of the hearer's children.

I will come back to this issue only in section 3 where I briefly discuss how a simple game theoretic model that accounts for the inference in (3b) can be extended to handle the more refined epistemic case.

[4]The name of Grice's postulate is chosen in reference to 'Occam's Razor' a principle loosely attributed to the 14th century philosopher William of Occam (though not found in his writing), which pleads for ontological parsimony in theorizing: "entia non sunt multiplicanda praeter necessitatem."

the proof that the pragmatic inferences in question can be explained *systematically* based on certain assumptions about proper conduct of a conversation. Grice hypothesized that in most normal circumstances interlocutors share a common core of convictions about the purpose of a conversation and behave, in a sense, *rationally* towards this commonly shared end. This regularity in linguistic behavior explains, so Grice's conjecture, pragmatic inferences of the attested sort.

In particular, Grice proposed to view conversation as guided by an overarching Cooperative Principle, formulated as a rule of conduct for speakers:

Cooperative Principle: "Make your contribution such as it is required, at the stage at which it occurs, by the accepted purpose or direction of the talk exchange in which you are engaged." (Grice 1989, p. 26)

Subordinated to the Cooperative Principle, Grice famously gave a perspicuous set of guidelines for proper speaker conduct in his Maxims of Conversation:

Maxim of Quality: Try to make your contribution one that is true.

(i) Do not say what you believe to be false.
(ii) Do not say that for which you lack adequate evidence.

Maxim of Quantity:

(i) Make your contribution as informative as is required for the current purposes of the exchange.
(ii) Do not make your contribution more informative than is required.

Maxim of Relation:

(i) Be relevant.

Maxim of Manner: Be perspicuous.

(i) Avoid obscurity of expression.
(ii) Avoid ambiguity.
(iii) Be brief (avoid unnecessary prolixity).
(iv) Be orderly. (Grice 1989, p. 26–27)

Grice showed that hearers can reliably and systematically interpret utterances and infer additional information that goes beyond the semantic meaning of the uttered sentence, based on the assumption that the speaker obeys the Cooperative Principle and the Maxims of Conversation. The main idea of the Gricean Programme is thus to make pragmatic inference amenable to systematic investigation, and to find regularities and structure in conversational behavior and natural language interpretation. Indeed, this idea has had tremendous impact on the philosophy of language and linguistic pragmatics, inspiring and spawning a whole industry of literature on topics and problems raised by Grice's work.[5]

Conversational Implicatures. In particular, Grice's notion of conversational implicature still excites the community, a term of art which he coined for what was *meant* with an utterance but not *said* (see Levinson 1983; Horn 2004, for general overview). A certain kind of conversational implicature has attracted more attention than any other, namely scalar implicatures — so-called because of the particular role that ordered scales of alternative expressions play in the derivation of such inferences. We have already seen an example of scalar implicature in example (3). The inference from "some" to "some and not all" hinges on the idea that there is a semantically stronger lexical alternative "all" which the speaker has not, but could have used relevantly and informatively. Similar reasoning yields similar scalar implicatures for other examples of so-called Horn scales (after Horn 1984). Here are some further examples:[6]

(5) a. It's possible that Yuuki is coming late again.
$\rightsquigarrow$ It's not certain/necessary that Yuuki is coming late again.

b. Hanako sometimes listens to jazz.
$\rightsquigarrow$ Hanako does not often/always listen to jazz.

The reason why scalar implicatures attracted so much attention is presumably because of their relative context-independence, or better context-robustness. Although cases where a scalar item does not give rise to a scalar implicature exist, the inference nevertheless is fairly regular and persistent across contexts. Scalar implicatures are prime examples of what Grice had called generalized conversational implicatures. These are highly regular

[5] For more on the impact of Grice's philosophical work see Neale (1992) and Chapman (2005).

[6] I use the symbol $\rightsquigarrow$ to mark a possible candidate implicature that an utterance of a given sentence has or might have in a standard context of its use. I again gloss over the epistemic status of such implicatures.

and predictable implicatures, unlike PARTICULARIZED CONVERSATIONAL IMPLICATURES like (2) that arise only in very special contextual circumstances. That is why some theorists have argued that scalar implicatures have a special *default status* (Levinson 2000) or have made the case for integration of scalar implicature calculation into syntax (Chierchia 2004; Fox 2007; Chierchia, Fox, and Spector 2008). This conceptual debate is furthermore informed by a fast growing literature on empirical research into scalar implicature calculation (see Noveck and Sperber 2004, and reference therein). In order to put game theoretic pragmatics on the map, this paper will focus on a simple case of scalar implicature too.

Branches of Gricean Pragmatics. To say that Grice's contribution was heavily influential is not to imply that it was entirely uncontroversial. Even to those who wholeheartedly embarked on the Gricean Programme the exact formulation of the maxims seemed a point worth improvement. It was felt that —to say it with a slightly self-referential twist— the Gricean maxims did not do justice to themselves, in particular to the Maxim of Manner, being long-winded and too vague to yield precise predictions in a number of linguistically relevant cases. Over the years, many attempts have been made to *refine* and *reduce* the Gricean maxims.

A particularly prominent and successful strand of maxim reduction is found in the work of so-called NEO-GRICEANS (Horn 1972; Gazdar 1979; Atlas and Levinson 1981; Levinson 1983; Horn 1984). This work is largely in keeping with the Gricean assumption of cooperation in conversation and seeks to explain pragmatic inference foremost in a refined explication of the Maxim of Quantity. Another prosperous school of research that arose from a critique of Grice's maxims, is RELEVANCE THEORY (Sperber and Wilson 1995; Sperber and Wilson 2004), according to which the Maxim of Relation deserves the main role in a theory of interpretation. Crucially, relevance theory explicitly sees itself as a *cognitive theory*, rather than a mere addition to a logico-semantic account of meaning, and we may say that, in this and other respects, relevance theory is less Gricean than, for instance, the neo-Griceans. Relevance theorists sometimes refer to their position as POST-GRICEAN, clearly indicating that relevance theory abandons the Cooperative Principle and leaves behind the Maxims of Conversation in favor of an interpretation principle in terms of cognitive effects and processing efforts.

While neo-Griceans foreground the Maxim of Quantity in natural language interpretation, and while relevance theorists emphasize the role of a cognitively informed notion of communicative relevance, Grice himself held that the grounds for his communicative principles were to be found in human *rationality*. He wrote:

> "As one of my avowed aims is to see talking as a special case or variety of purposive, indeed rational, behaviour, it may be worth noting that the specific expectations or presumptions connected with at least some of the foregoing maxims have their analogues in the sphere of transactions that are not talk exchanges."
>
> (Grice 1989, p. 28)

Picking up Grice's conjecture about a rational foundation of his maxims, Kasher (1976) showed how to deduce Grice's maxims from a single postulate of human rationality in action.[7] For linguistic pragmatics, however, the question is not so much whether Grice's maxims can be reduced to rationality, but rather whether the pragmatic interpretation behavior we would like to explain in terms of the maxims can be feasibly described as rational. This is where a formal theory of rational human agency in the form of decision and game theory enters: in GAME THEORETIC PRAGMATICS, to which we will turn next, we would like to ask to what extent it is reasonable to explain pragmatic inference as a rationalistic interpretation of human action; and we would moreover like to address this question by using formal models of rational choice theory in order to be able to formulate problems and answers as sharply and succinctly as possible.[8]

2 Games as Context Models

A GAME in its technical sense is a mathematical structure that represents abstractly a decision situation of several agents, where the outcome of the decisions of each agent depends on the choices of the other agents.[9] Game theory distinguishes different kinds of games, traditionally classified along two dimension: whether the agents' choices are *simultaneous or in sequence*, and (ii) whether all agents have *complete or incomplete information* about the shared decision situation they are in (not counting the information what other players

[7] Some thirty years later, Rothschild (2008) addresses the related question whether Gricean maxims can be given a game-theoretic foundation.

[8] For further general assessments of applications of game theory in linguistics see Jäger (2008).

[9] For general introductions to game theory see Myerson (1991); Gibbons (1992); Osborne and Rubinstein (1994); Osborne (2004). A good introduction to game theory in a linguistic context is the introduction in Benz, Jäger, and van Rooij (2006).

are going to do in the future). Games where players move simultaneously are called STATIC GAMES (alt.: strategic games); games where players move in sequence are called DYNAMIC GAMES (alt.: sequential games).

Games as Models of the Utterance Context. Although static games with complete information are the easiest and most manageable kinds of games, they are unfortunately not the most natural choice for a model of utterance interpretation. Utterances and their pragmatic reception are rather to be modelled as *dynamic* games, because we would like to capture the sequential nature of utterance and subsequent reception/reaction and the natural asymmetry in information between interlocutors. Of course, different kinds of utterances would require different kinds of dynamic games. For instance, in modelling a run-of-the-mill case of an informative assertion the speaker should possess information that the hearer lacks, whereas in the case of a stereotypical information-seeking question we would like to refer to a game in which the speaker is uninformed about a particular contingency while the hearer is (possibly/partially) informed.

In general, I suggest that a game should be regarded as a reduced and idealized, but for certain purposes sufficient MODEL OF THE UTTERANCE CONTEXT: it represents a few (allegedly: *the most*) relevant parameters of a conversational context, viz., the interlocutors' beliefs, behavioral possibilities and preferences, in rather crude, idealized abstraction. This general, conceptual point will become clearer when we look at an easy example of a dynamic game and its interpretation as a context model.

Consider the following scenario: Alice is preparing dinner for her visitor Bob who would like to bring a bottle of wine. Depending on whether Alice prepares beef or fish, Bob would like to bring red or white wine respectively. Both Alice and Bob share the same interest in wine matching the dinner, but while Alice knows what she is preparing for dinner, Bob does not. However, we may assume that Bob does not need to guess what Alice is preparing because Alice can simply tell him by saying "I'm preparing beef/fish." Only then would Bob make his decision to bring either red or white wine.

This contrived scenario is perhaps the simplest possible example of a stereotypical informative assertion: the speaker (Alice) has some piece of information that the hearer (Bob) lacks but would like to have in order to make a well-informed decision; the speaker then utters a sentence (which we may assume has a semantic meaning already) and the hearer possibly changes his initial beliefs in some fashion and chooses his action subsequently. This idealized situation should then be modelled as a particular dynamic game with incomplete information.

Signaling Games. The crucial ingredients of the context of utterance of the previous example —such as Ann's knowledge of what she is preparing for dinner; Bob's uncertainty thereof; Ann's and Bob's available choices; their desires and preferences— all can be captured in a relatively simple game called SIGNALING GAME. A signaling game is a special kind of dynamic game with incomplete information that has been studied extensively in economics (Spence 1973), biology (Zahavi 1975; Grafen 1990), philosophy (Lewis 1969), and linguistics (Parikh 1992; Parikh 2001; van Rooij 2004*b*).[10] Formally, a signaling game (with meaningful signals) is a tuple

$$\langle \{S, R\}, T, \Pr, M, [\![\cdot]\!], A, \mathrm{U}_S, \mathrm{U}_R \rangle$$

where sender S and receiver R are the players of the game; T is a set of states of the world; $\Pr \in \Delta(T)$ is a full support probability distribution over T, which represents the receiver's uncertainty which state in T is actual;[11] M is a set of messages that the sender can send; $[\![\cdot]\!] : M \rightarrow \mathcal{P}(T)$ is a denotation function that gives the predefined semantic meaning of a message as the set of all states where that message is true; A is the set of response actions available to the receiver; and $\mathrm{U}_{S,R} : T \times M \times A \rightarrow \mathbb{R}$ are utility functions for both sender and receiver that give a numerical value for, roughly, the desirability of each possible play of the game.

Alice and Bob's wine-choice scenario can be represented as the signaling game given in figure 1. There are two possible states of nature (only one of which is actual, of course): in $t_{\texttt{beef}}$ Alice prepares beef, and in $t_{\texttt{fish}}$ she prepares fish. Alice knows which state is actual, but Bob does not and so his uncertainty is represented numerically in the probability distribution Pr;

[10]Parikh explicitly denies that he is using standard signaling game models. Though fairly similar to signaling games, his GAMES OF PARTIAL INFORMATION are not quite the same and also not standard in game theory. A discussion of differences and similarities, advantages and disadvantages would take us too far in the present context (but see Parikh 2006). For clarity and fairness though, it should be mentioned that the criticism discussed in section 3 that applies to signaling game models does not necessarily apply to Parikh's game models.

[11]As for notation, $\Delta(X)$ is the set of all probability distributions over set X, Y^X is the set of all functions from X to Y, $X : Y \rightarrow Z$ is alternative notion for $X \in Z^Y$, and $\mathcal{P}(X)$ is the power set of X. We say that a probability distribution $\delta \in \Delta(X)$ has FULL SUPPORT if for all $x \in X$ $\delta(x) > 0$. To ask for full support receiver beliefs is to require that the receiver does not rule out *a priori* that certain states are actual, which is fairly natural. Other than that, the receiver may be as opinionated as is imaginable.

	$\Pr(\cdot)$	U_S, U_R		M & $[\![\cdot]\!]$	
		$a_{\texttt{red}}$	$a_{\texttt{white}}$	$m_{\texttt{beef}}$	$m_{\texttt{fish}}$
$t_{\texttt{beef}}$	.6	1,1	0,0	$\surd$	–
$t_{\texttt{fish}}$	.4	0,0	1,1	–	$\surd$

Figure 1: Signaling game for coordination

according to the table in figure 1 then, Bob finds it just a little more likely that Alice prepares beef than that she prepares fish (perhaps because she has shown a tendency towards beef in the past). Alice can say either of two things $m_{\texttt{beef}}$ "I'm preparing beef" or $m_{\texttt{fish}}$ "I'm preparing fish" with the obvious semantic meaning as indicated by the check marks in figure 1. In turn, Bob can choose to bring red wine ($a_{\texttt{red}}$) or white wine ($a_{\texttt{white}}$). Both Alice and Bob value an outcome where the wine matches the food more than an outcome where it doesn't; beyond that, they have even identical preferences in the given example. (The table in figure 1 lists Alice's utilities first, then Bob's, as a function of t and a only, i.e., we assume for simplicity of representation that the agent's payoffs do not depend on the message that is sent, but only on the state and the response action.)

Signaling Games for Pragmatic Interpretation. Of course, in a certain sense the example situation and the signaling game in figure 1 are not particularly interesting for linguistic pragmatics. There is not much room for pragmatic inference in this toy example: commonsense has it that Alice would tell Bob that she is preparing beef if and only if she is indeed preparing beef, and Bob will bring red wine if and only if Alice tells him that she is preparing beef. A context model of a pragmatically more interesting situation is the signaling game in figure 2, which is intended to capture (again: in violent abstraction) the arguably simplest context of utterance in which we would expect a scalar implicature like the one in (3) to arise. The signaling game

	$\Pr(\cdot)$	$a_{\exists\neg\forall}$	$a_{\forall}$	$m_{\texttt{some}}$	$m_{\texttt{all}}$
$t_{\exists\neg\forall}$	p	1,1	0,0	$\surd$	–
$t_{\forall}$	$1-p$	0,0	1,1	$\surd$	$\surd$

Figure 2: The some-all game: a context model for scalar implicature

in figure 2 has two states $t_{\exists\neg\forall}$ and $t_{\forall}$, two messages $m_{\texttt{some}}$ and $m_{\texttt{all}}$ with semantic meaning as indicated and two receiver interpretation actions $a_{\exists\neg\forall}$ or $a_{\forall}$ which correspond one-to-one with the states. We could think of these actions either as concrete actions, as interpretations that the receiver wants to adopt or just as placeholders indicating what is relevant for the receiver in the given context. Also in this example sender and receiver payoffs are perfectly aligned in order to model the assumption that interlocutors cooperate and care to coordinate on proper interpretation.

In a situation modelled by the signaling game in figure 2, we would intuitively expect the sender and receiver to behave as follows:

(i) the sender sends $m_{\texttt{some}}$ in state $t_{\texttt{some}}$ and the message $m_{\texttt{all}}$ in state $t_{\texttt{all}}$;

(ii) the receiver responds to $m_{\texttt{some}}$ with $a_{\texttt{some}}$ and to $m_{\texttt{all}}$ with $a_{\texttt{all}}$.

This would correspond to the intuitive use of the corresponding natural language expressions and would, in a sense, explain how the semantic meaning of a scalar term like "some" is enriched by the presence of a stronger alternative in a cooperative signaling situation. However, we do not yet have any way of talking about the concrete behavior of agents. Games are just context models; they contain the set of all possible moves an agent can make, but they lack a specification of the agents' concrete behavior. This is what we turn to in section 4. But before we turn to this, I would like to assess some of the more obvious general merits and detriments of using dynamic games of incomplete information as formal models of the context of utterance.

3 Some Pros and Cons of Games as Context Models

The main benefit of using dynamic games of incomplete information as context models is that these structures are very rich and expressive. Proper use of such representational means allows for more generality, fine-grained distinctions and conceptual clarity than other less formal approaches to pragmatics are capable of. The questions we need to address are (i) where exactly the expressive power and precision of game models lie, (ii) whether our game models are really appropriate and sufficient for the purpose at hand, and (iii) to what extent we actually need the expressivity of such game models in linguistic pragmatics. Here are some critical thoughts on these matters.

Explicit representation of beliefs. First of all, let me stress that the class of games in question, dynamic games of incomplete information, is indeed very expressive. Although much attention has been devoted to signaling game models in pragmatics, this should not be taken to mean that this is the only game model we should use. In principle, we can represent *any* arbitrarily complex situation of nested beliefs about beliefs and preferences of agents in a dynamic game of incomplete information. In signaling games, for instance, the interlocutors' preferences and action alternatives are common knowledge. Also, the speaker knows exactly what state of the world is actual. But these assumption can easily be given up at the expense of a more complex game if necessary.

Explicit representations of individual preferences. Games can model very fine distinctions not only in beliefs, but also in the preferences of individual agents. These can be relevant for linguistic interpretation in diverse ways. For instance, under normal circumstances the answer to a question like in (6) is interpreted exhaustively as implicating that Bill did not come, but the answer to a question like that in (7) is not.

(6) a. Who, of John, Bill and Mary, came to the party?
  b. John and Mary did.
  c. $\rightsquigarrow$ Bill did not.

(7) a. Where can I get an Italian newspaper?
  b. At the reception.
  c. $\not\rightsquigarrow$ Not at the airport.

The reason for this difference in interpretation of answers intuitively lies in the relevance that certain information has for the questioner based on a practical decision he faces (see van Rooij 2003). To account for the structural commonality and differences of cases (6) and (7), models that represent an agent's individual preferences in a goal-oriented setting are advantageous if not necessary.

Another example where individual hearer preferences notably inform the interpretation of an utterance is in the formation of CONTEXTUAL RELEVANCE SCALES. As noted by Hirschberg (1985), scalar inferences are not confined to lexicalized scales such as "some" and "all", but can also occur with non-entailing expressions that are not necessarily lexically associated. Take for instance the following example:

(8) a. So, I hear Hans Mustermann took part in the last three Olympic games. But did he ever win Olympic gold?

b. He won silver in Beijing.
$\rightsquigarrow$He did not win gold in Athens or Sidney.
$\not\rightsquigarrow$ He did not win bronze in Athens or Sidney.

A detailed representation of preferences pins down what exactly is relevant for the conversationalists, independent from lexicalized scales (see Benz 2007, for more on relevance scales in rational choice models). The crucial point is that rational choice models reduce the notion of relevance in context to individual preferences, which seems very intuitive and appealing.[12]

Explicit representations of variable degrees of preference conflict. Games as context models not only include preferences of single agents, but crucially those of all interlocutors. This lets us model different levels of partial alignment or divergence of preferences of multiple agents. Grice's assumption of cooperation in conversation is easily integrated as a special case, but it is clear that the representative power of games provides much more generality. Game models let us represent arbitrary constellations of partially cooperative, partially adversary discourses. Predictions are not confined to cooperation only —as in traditional Gricean approaches— or to argumentation only —as for instance in the work of Ducrot (1973), Anscombre and Ducrot (1983) and Merin (1999)— and this makes game theoretic pragmatics much more general and systematically applicable than other approaches (see van Rooij 2004*a*; Benz 2006; Franke, de Jager, and van Rooij 2008, for further discussion).

Where does the game model come from? A serious problem for the application of games in pragmatics is the question where the particular ingredients of the context model are supposed to come from. Some modelling choices are needed when we define a game, and some of them heavily influence the predictions of the account. This may be more critical for some elements of games than for others.

Take for instance the set M in the signaling game model in figure 2. Why did we assume that there are *only* the two messages $m_{\texttt{some}}$ and $m_{\texttt{all}}$? In particular, why did we exclude a message with a semantic meaning "some but

[12]That individual preferences give rise to a notion of contextual relevance of information is one point in favor of using models from rational choice theory in linguistic pragmatics. A further favorable point is that a preference for informativity, as postulated in Grice's Maxim of Quantity and upheld by the neo-Griceans, falls out as a *special case* in preference-based approaches, just as it should. This argument is presented formally by Bernardo (1979) in a general form, and by van Rooij (2004*c*) in the context of natural language interpretation.

not all”? The availability of different forms will certainly affect the behavior and belief dynamics of agents under most solution concepts, so the choice of action alternatives is non-trivial and the success of accounting for a pragmatic phenomenon like scalar implicature hinges on the appropriate specification of in particular the set M. However, this problem is not new and it is exactly what prompted neo-Griceans to postulate lexicalized Horn scales, as introduced in section 1. So it might seem fair to say that game theory simply shares the problem of specifying suitable alternatives with the neo-Griceans.

A similar point could be raised with respect to the set T of states of the world as representations of different meanings that are to be distinguished in the context model. Also here the question arises for every concrete modelling choice how to justify the set T. Interestingly, this problem has not been addressed critically in other approaches to formal pragmatics where the question which possible meanings are to be considered is equally relevant, but is just never explicitly addressed. Thus conceived it is an advantage of a stringent formal context model in terms of a game to bring a necessary modelling choice to the foreground. Still, the question where the set T should come from —and what, for instance, its possible relation to the set M is— is one of the issues that Gricean pragmatics has to answer, whether under a game theoretic approach or under another suitably formal approach.

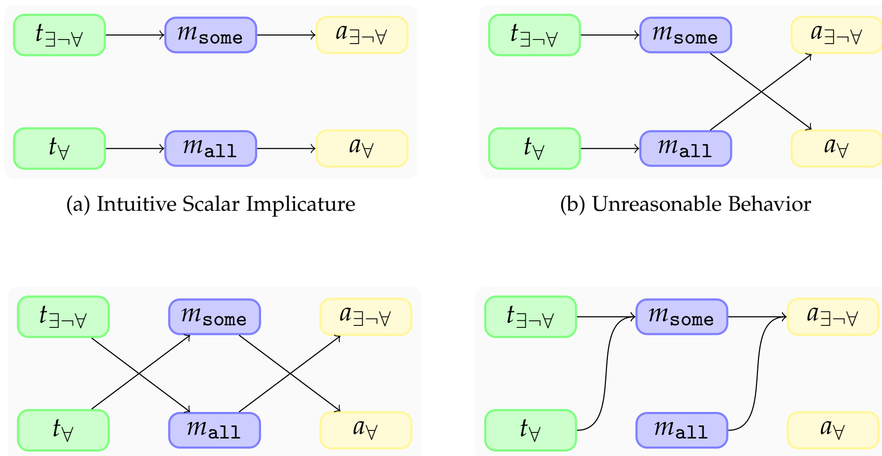
Complexity of game models. The final point I would like to address is an anticipated objection to the precision of the game models and the complexity that this precision entails. For one, it is not a drawback that game models are *too* precise in that their expressive power might exceed naïve intuition. It’s better to regard game models as tools for sharpening intuitive concepts. For another, it is not an argument against games as context models that such models get fairly complicated proportional to the degree that we suspend simplifying idealizations. Rather the situations that we try to model are fairly complicated *themselves* and the models again only help organize and characterize the subtleties of intricate conversational situations. Moreover, it is clear that we trade empirical or introspective realism in the model for (mathematical) complexity only because we conserve the models’ level of precision which guarantees a theory with clear predictions and a fair chance of being falsified. By this reasoning, it is decidedly *not* an advantage of informal accounts (like, e.g., relevance theory) not to show an increase in explanatory complexity when purportedly little, if no simplifying idealizations are operative.

Chapter Summary. Taken together, it's fair to say that games are very rich representational means which help pin down relevant aspects of a context of utterance in at least sufficient detail. There are certainly problems with the representation of contexts as games, but we should not be mistaken about the significance of these problems: they are not necessarily arguments against game theoretic pragmatics in general, but rather the very contribution of this formally ambitious approach; I'd like to say that some of the issues raised here as objections to game models of the utterance context are critical challenges for linguistic pragmatics as such.

4 Strategies, Solutions, Rationality & Equilibrium

Strategies as Representations of Agents' Behavior. Recall that games specify the general behavioral possibilities of agents, but do not specify further how agents do or should in fact behave. On top of the game model we therefore represent the behavior of players in terms of STRATEGIES which select possible moves for each agent for any of their choice points in the game. For signaling games, a PURE SENDER STRATEGY $\sigma \in M^T$ is a function from states to messages, because the sender has to decide what to say for each state that might be actual. A PURE RECEIVER STRATEGY $\rho \in A^M$ is a function from messages to actions, because the receiver knows only what message has been sent, but not what state is actual. A PURE STRATEGY PROFILE $\langle \sigma, \rho \rangle$ is then a complete characterization of the interlocutors' *behavior* in a given signaling game, and the set of all such pairs gives the set of all behavioral possibilities of our abstract conversationalists.

Figure 3 gives four (out of sixteen possible) examples for pure strategy profiles for the some-all game from section 2. Sender strategies —functions in M^T— are represented by the set of arrows leaving the state nodes on the left; receiver strategies —functions in A^M— are represented by the set of arrows leaving the message nodes in the middle. At present the four strategy profiles in figure 3 are just arbitrary examples of different strategies and their combination in strategy profiles. We will have a closer look at exactly this selection of strategy profiles no sooner than section 5. Still, notice already at this point that the our general definition of a sender strategy includes cases like that in figure 3c where the sender sends *untrue* messages: as far as semantics is concerned the message $m_{\texttt{all}}$ is *not* true in state $t_{\exists\neg\forall}$, but nonetheless the general definition of a sender strategy includes the possibility that the sender violates the Maxim of Quality, so to speak. We will come back to this issue at some



(a) Intuitive Scalar Implicature

(b) Unreasonable Behavior

(c) Permuted Meaning

(d) Pooling

Figure 3: Some pure strategy profiles for the some/all game

length later on in section 5.

Solution Concepts Yield the Model's Prediction of Behavior. Of course, we would like to have criteria why we expect certain (collective) behavior and not others to show in the situation that is modelled by a given game. For example, some suitable criterion on the behavior of agents should select the intuitive scalar implicature play in figure 3a and rule out all other strategy profiles. In general, it is a SOLUTION CONCEPT for a particular kind of game which should do that for us. A solution concept, in a sense, selects good from bad strategy profiles. For game theoretic pragmatics the situation is this: supposing that we have reached agreement that a particular game is a good representation of the context of utterance of a sentence whose use and pragmatic interpretation we would like to explain, an appropriate solution concept should then select all and only those strategy profiles that represent the intuitively or empirically attested data. We could then regard the pair consisting of the game-as-context-model and the solution concept as the explanation of the data.

So, which solution concept does the trick, in the some-all game and in general? This question is not easily answered, and in fact constitutes one of the main research questions of game theoretic pragmatics as we will see in section 5. Different solution concepts might not only yield different predictions for the same game, but might also have quite different conceptual motivations.

For instance, solution concepts in classical game theory incorporate an assumption of rationality of players, whereas solution concepts in evolutionary game theory, though possibly subjecting the very same game models, often do not appeal to any rationality constraint.[13] To provide a basis for the subsequent general discussion I suggest looking at what is possibly the easiest classical equilibrium concept for signaling games. This will be sufficient to assess some of the problems and chances of game theoretic pragmatics as a rationalistic approach to natural language interpretation (and will moreover turn out to be already complicated enough).

Rationality, Expected Utility and Beliefs. The notion of rationality in both classical decision and game theory where agents have to make decisions under uncertainty about the outcomes of their actions is Bayesian rationality. The idea behind Bayesian rationality is maximization of expected utility, which is a technical measure for the gain an action is subjectively expected to yield. Towards a general definition, fix a set of alternative actions A, and a set of states T that the outcome of performing an action depends on. We assume that our decision maker has preferences over all outcomes, i.e. pairs $T \times A$, which is given by the numerical utility function $\mathrm{U} : T \times A \to \mathbb{R}$. We also assume that she has beliefs about the actual state, which is given by a probability distribution over states $\mathrm{Pr} \in \Delta(T)$. The agent's *expected utility* of performing an action a as a function of belief Pr is then defined as

$$\mathrm{EU}(a, \mathrm{Pr}) = \sum_{t \in T} \mathrm{Pr}(t) \times \mathrm{U}(t, a).$$

This helps define Bayesian rationality as follows:

(9) **Bayesian Rationality** Given an agent's behavioral alternatives A, his beliefs Pr and preferences U, the agent is *rational* only if he chooses an action $a \in A$ which maximizes his expected utility (as given by Pr and U).[14]

This notion of Bayesian rationality can be implemented rather easily in a solution concept for signaling games. We already have the players' action

[13]See Weibull (1997) and the introduction in Benz, Jäger, and van Rooij (2006) for more on evolutionary game theory.

[14]Notice that the definition in (9) has only "only if", because, strictly speaking, an agent who chooses an act that maximizes expected utility need not *be* rational, although, in a sense, she definitely *behaves* rationally. With locution "behaves rationally" instead of "be rational" in (9) both directions of implication are true. However, since we always only reason *from* the assumption of an agent's *de facto* rationality, and not *to* it, we only need that rationality implies utility maximization in expectation.

alternatives —the set M for the sender and the set A for the receiver— and the agents' preferences over outcomes U_S and U_R. All we need to do is to specify the agents' beliefs in order to determine when their behavior would count as rational. So, what are the sender's and the receiver's beliefs in a signaling game?

The sender knows the actual state, whatever it turns out to be. So there is not much uncertainty involved. The sender actually has complete information about the game. But she does not have *perfect information*, as the game theorist would say, for she does not know what the receiver will do. We can characterize the sender's belief about what the receiver will do as a PROBABILISTIC RECEIVER STRATEGY $\rho \in (\Delta(A))^M$ that specifies a probability distribution over A for each m: $\rho(m)$ then gives the probabilistic beliefs of the sender about which action the receiver will play if he observes m.[15] Given the sender's belief ρ about the receiver's behavior, we can define the sender's *expected utility* of sending message m in state t as a function of her belief ρ:

$$\mathrm{EU}_S(m, t, \rho) = \sum_{a \in A} \rho(m, a) \times \mathrm{U}_S(t, m, a).$$

In line with Bayesian rationality, if S is rational and believes ρ she should send a message m in state t only if it maximizes her expected utility given belief ρ. We say that a pure sender strategy is rational just in case it selects an action which maximizes expected utility in all states, i.e., σ is a RATIONAL PURE SENDER STRATEGY given belief ρ if and only if for all t:

$$\sigma(t) \in \arg\max_{m \in M} \mathrm{EU}_S(m, t, \rho).$$

To characterize which receiver behavior counts as rational, we similarly have to specify a feasible description of R's beliefs. However, here the situation is a little more complicated, because the receiver not only has imperfect information —not knowing what the sender does—, but also incomplete information —not knowing what the actual state of the world is. In a sense, we could say that there are three things that the receiver is uncertain about (see Battigalli 2006):

(i) R has *prior uncertainty* about which state is actual before he observes a message; these PRIOR BELIEFS are specified by the distribution Pr in the signaling game;

[15]The use of the same notation for both pure and probabilistic strategies is sanctioned because pure strategies are easily construed as special cases of probabilistic strategies.

(ii) R also is uncertain about the sender's behavior; again we can characterize the receiver's beliefs about what the sender may do as a PROBABILISTIC SENDER STRATEGY, i.e., a function: $\sigma \in (\Delta(M))^T$ that gives a probability distribution over M for each t;

(iii) and finally R also has *posterior uncertainty* about which state is actual after he observes a message; for clarity, this is not because the actual state changes, but because the receiver's beliefs about the actual state may be influenced by the observation what message the sender has sent; these POSTERIOR BELIEFS can be described as a function $\mu \in (\Delta(T))^M$ that gives a probability distribution over T for each m.

The important component in the receiver's beliefs for a characterization of rational receiver behavior are, of course, the latter posterior beliefs: it's *after* observing a message that the receiver is called to act, so it's with respect to the beliefs he holds at that time that we should judge him rational or not. Therefore, given a posterior belief μ, we define R's expected utility of performing a after message m has been received as

$$\mathrm{EU}_R(a, m, \mu) = \sum_{t \in T} \mu(m, t) \times \mathrm{U}_R(t, m, a)$$

and say that ρ is a RATIONAL PURE RECEIVER STRATEGY if and only if for all m

$$\rho(m) \in \arg\max_{a \in A} \mathrm{EU}_R(a, m, \mu).$$

But, of course, although the posterior beliefs μ are crucial for judging R's behavior, μ should be derived, at least in part, from the other two components of R's uncertainty. What we need is a consistency criterion that the receiver's posterior beliefs fit his prior beliefs and his conjecture about the sender's behavior. Technically speaking, we want the posterior beliefs μ to be derived from Pr and σ by BAYESIAN CONDITIONALIZATION. We say that the receiver's posterior beliefs μ are CONSISTENT with his beliefs Pr and σ if and only if for all t in T and for all m in the image of σ we have:

$$\mu(m, t) = \frac{\Pr(t) \times \sigma(t, m)}{\sum_{t' \in T} \Pr(t') \times \sigma(t', m)}.$$

It's not crucial to understand Bayesian conditionalization for any concern raised in this paper, so I will not elaborate. But it will be critical in the subsequent discussion that consistency only applies to messages in the image of σ, i.e., to messages that are expected to be sent under the belief σ.

Perfect Bayesian Equilibrium. Having defined what behavior is rational for sender and receiver individually, we are able to characterize which strategy profiles $\langle \sigma, \rho \rangle$ satisfy a joint constraint on rationality. But since, as we saw, we need to supply a characterization of the beliefs of agents to tell rational from irrational behavior individually we also have to do so for the pair. So we could say that $\langle \sigma, \rho \rangle$ is a rational strategy profile just in case there is some belief ρ' for which σ is rational and some belief μ for which ρ is rational. This is close to what we will require, but we would like to be more specific. The idea of equilibrium for which John Nash and game theory is famous is, roughly, to require additionally to rationality that the beliefs of players be correct, i.e., derived from the strategy profile (at least as far as possible).[16] So here it is, our rationalistic equilibrium solution concept for signaling games: we say that a triple $\langle \sigma, \rho, \mu \rangle$ is a PERFECT BAYESIAN EQUILIBRIUM iff three conditions hold:

(i) σ is rational given the belief ρ;

(ii) ρ is rational given the belief μ;

(iii) μ is consistent with Pr and the belief σ.

Having thus defined a basic equilibrium solution concept for signaling games, the important question becomes to what extent this solution concept helps explain the intuitive conversational behavior. The next section will tend to this matter briefly.

5 Equilibrium Selection & Semantic Meaning

Pragmatics and the Problem of Equilibrium Selection. The concrete question to be addressed at the beginning of this section is: does perfect Bayesian equilibrium uniquely select the strategy profile in figure 3a? First of all, the strategy profile in figure 3a is indeed a perfect Bayesian equilibrium. This is also informally appreciated:

- let σ and ρ be the pure sender and receiver strategies as represented in figure 3a;
- given the belief that the receiver plays ρ, the only sender strategy which is rational given S's preferences is σ;

[16] It is interesting to note at that it is a fairly common and widespread misconception to think that equilibrium notions require common believe or knowledge of rationality. They don't (see Stalnaker 1994; Aumann and Brandenburger 1995, for a characterization of Nash equilibrium).

- the receiver's posterior beliefs are completely determined by the sender's strategy: the only belief μ consistent with any full support prior and the sender's strategy σ is the posterior belief that puts full credence, i.e. probability 1, on state $t_{\exists\neg\forall}$ after hearing $m_{\texttt{some}}$ and full credence on $t_{\forall}$ after hearing $m_{\texttt{all}}$;
- given this belief μ and the R's preferences, ρ is the only rational receiver strategy.

But what about the other strategy profiles in figure 3? Again a piece of good news is that the strategy profile in figure 3b is *not* a perfect Bayesian equilibrium: informally speaking, if the sender's strategy reveals the actual state, it is irrational given the receiver's payoffs to reverse the meaning of the signals. On the other hand, rather unfortunately, the strategy profiles in figures 3c and 3d, which represent intuitively unattested ways of conversational behavior, are also perfect Bayesian equilibria. The interested reader will quickly verify for herself that 3c is, and we will see presently why 3d is too. But that means that, sadly, perfect Bayesian equilibrium, as it stands, is not strong enough to rule out all unintuitive strategy profiles.

This problem is a concrete instance of the more general PROBLEM OF EQUILIBRIUM SELECTION, well-known and notorious in game theory. A whole branch of economics literature is dedicated to the search for appropriate refinements of standard equilibrium concepts, such as perfect Bayesian equilibrium. It may therefore appear fair to say that the most confronting problem of game theoretic pragmatics is, in a sense, a game theoretic one, namely the specification of an appropriate solution concept that yields intuitively acceptable predictions about conversational behavior in contexts represented by game models.

Equilibrium Selection by Enforcing Truthful Signaling in the Game Model. There is a very obvious idea that we should try in order to refine the predictions of perfect Bayesian equilibrium. There is something fishy about the use of *semantically meaningful messages* in the two strategy profiles in figures 3c and 3d that are, as of yet, not rule out by perfect Bayesian equilibrium. Certainly, the equilibrium in figure 3c is clearly distinct, because it not only employs messages in reversal to their semantic meaning, but has the speaker use message $m_{\texttt{all}}$ in a state where it is not true. In virtue of Grice's Maxim of Quality, we might want to rule that untrue signaling be excluded. So, suppose we restrict the underlying signaling game in such a way that the sender may only choose true signals in a given state. For clarity, this is then a restriction of the underlying signaling game: we change the context model, assuming that the speaker *cannot* speak untruthfully and that this is common knowledge between interlocutors. We will come back to this contentious assumption presently. For the time being, suffice it to note that enforced truthfulness obviously excludes the behavior in figure 3c from the set of feasible strategy profiles, and thus leaves perfect Bayesian equilibrium with one problem less, so to speak.

Does restriction to truthful signaling also help with the strategy profile in figure 3d? Strictly speaking, it does *not* and it is instructive to see why. First of all, the sender is not using any message untruthfully in this case. The only point that our semantic intuitions might object to is the interpretation of, or rather reaction to, the message $m_{\texttt{all}}$. The receiver responds to this message with action $a_{\exists\neg\forall}$, which is the action that is optimal in a state where the message $m_{\texttt{all}}$ is not true and hence could have never been used in. So there is something weird about this response, but still it is not immediately ruled out by the requirement that the sender is to speak the truth.

It pays to look more closely at the question why 3d is a perfect Bayesian equilibrium, even when truthful signaling is enforced, and why it should not be a solution intuitively. To address these questions, let's fix some terminology first. In the strategy profile 3d, the message $m_{\texttt{all}}$ is an unsent, so-called SURPRISE MESSAGE. What is "suprising" about surprise messages is that a receiver who believes that the sender sends signals in accordance with a given sender strategy will not expect such messages to be sent. In the 'some-all game' surprise messages exist if and only if the sender plays a POOLING STRATEGY, i.e., a strategy where the same message is sent in several states.[17] To fix terminology, we say that any strategy profile or equilibrium in which the sender plays a pooling strategy is called a POOLING STRATEGY PROFILE or a POOLING EQUILIBRIUM.

Now, perfect Bayesian equilibrium does not rule out the pooling strategy profile in 3d even if we restrict the signaling game to truthful signaling, because the solution concept does not restrict the receiver's posterior beliefs μ for surprise messages *at all*. The only requirement that perfect Bayesian equilibrium places on μ is that it be consistent with prior beliefs Pr and a belief in a sender strategy σ. Consistency, however, is a condition on non-surprise messages only; it does not restrict the receiver's COUNTERFACTUAL BELIEFS, de-

[17]This is still assuming that the receiver does not rule out states a priori, i.e., that Pr has full support and that we are looking at pure strategies only.

fined as those beliefs he holds after surprise messages.[18] In particular for the case 3d, we cannot exclude a belief μ^* that $t_{\exists\neg\forall}$ is more likely than $t_\forall$ after observing $m_{\texttt{all}}$, because this is consistent with any full support prior and the pooling sender strategy in question. But under belief μ^* it is indeed rational to respond with $a_{\exists\neg\forall}$ to $m_{\texttt{all}}$. Consequently, perfect Bayesian equilibrium does not rule out this pooling equilibrium.

This is not a good prediction and it makes apparent a glaring shortcoming of perfect Bayesian equilibrium: although the concept demands the receiver to act rational also on the basis of counterfactual beliefs if there are some, it does *not* require these counterfactual beliefs to reflect the structure of the game appropriately. To wit, if we forbid the sender to send semantically untrue messages, the receiver should know this and this knowledge should also be represented in any counterfactual beliefs. In particular, if the sender cannot send untrue messages, then based on the game structure the receiver should not believe that it is possible *at all* that the actual state is $t_{\exists\neg\forall}$ after the message $m_{\texttt{all}}$ is *or would be* observed.

The Proper Role of Semantics in a Model of Pragmatic Reasoning. These considerations show that perfect Bayesian equilibrium is not strong enough a solution concept to serve our intuitions well as a predictor of behavior in the some-all game. There are, however, various stronger solution concepts in the game theoretic literature that address similar predictive weaknesses of perfect Bayesian equilibrium. Without going into details, suffice it to mention that notions such as TREMBLING HAND PERFECT EQUILIBRIUM (Selten 1965; Selten 1975), or SEQUENTIAL EQUILIBRIUM (Kreps and Wilson 1982) are refinements of perfect Bayesian equilibrium that would indeed exclude all the unintuitive strategy profiles if we fix that the sender *has to* send messages truthfully.[19]

Be that as it may, the *conceptual* problem of equilibrium selection in game theoretic pragmatics, I want to argue, is not solved by (i) hard-wiring truthful sender behavior into the game model and (ii) resorting to more refined and more technical solution concepts. What is needed is a more general specification of the role that semantic meaning plays in pragmatic deliberation. The problem is that it is not reasonable to assume *in the context model* that

[18] I call these beliefs counterfactual because they give rise to beliefs of the following form in concert with a belief in a pooling sender strategy: "S does not send m, but if she *would*, the actual state would be t with probability p, t' with probability q etc."

[19] Yet another, but conceptually different possibility is to use NEOLOGISM-PROOFNESS as defined by Farrell (1993) as a refinement of equilibrium. This is the solution proposed by van Rooij (2008) to rule out the pooling equilibrium in the some-all game.

the speaker *cannot* —not even for fun, so to speak— use a signal that is not true. Granted, hearers may *assume* that utterances are true as a default, in most circumstances barring evidence to the contrary. And of course, if it is commonly believed that the conversation is a cooperative effort and interests in communication are shared, then there is indeed no reason whatsoever to expect that the speaker might lie or mislead with untrue signals. But then, this verdict is really one that should fall out of considerations about the sender's reasonable behavior. It should not be hard-wired into her behavioral possibilities. In principle, we can say whatever we want whenever we want. It is only that certain strategic considerations convince us that truthfulness and trust are reasonable strategies in many communicative situations. Hence, a more principled way of integrating semantic meaning into the solution concept, not the context model, is clearly desirable.[20]

Summary and Outlook. To sum up at this point, game theoretic pragmatics shares a problem with other applications of game theory, namely the need to specify an appropriate solution concept that *uniquely* yields the intuitively/empirically desirable predictions. This is indeed one of the main (open) research questions of game theoretic pragmatics: how to assign to semantic meaning a conceptually plausible role in a mathematically precise account of pragmatic reasoning. There does not appear to be any established game theoretic notion that we could merely plug off the shelf and apply to pragmatics.

However, this lacuna is perhaps more chance than doom, because it leaves research in game theoretic pragmatics with the freedom to define a feasible solution concept based on exactly those assumptions —preferably independently and empirically motivated— about human behavior and cognition that we deem relevant in natural language use and interpretation. There is no reason why we need to stick to traditional concepts of equilibrium, or rationality in its strong Bayesian form. Empirical results of experimental game theory and psycholinguistics should ideally inform the formalization of both context models and solution concept. Empirical research in game theory is blooming (see Camerer 2003), and applications of empirically informed applications of game theory to pragmatics should —and gradually are— following suit (see Sally 2003; de Jaegher, Rosenkranz, and Weitzel 2008).

In particular, EPISTEMIC GAME THEORY seems like a very promising plat-

[20] I will not pursue this matter any further in this paper, but see Stalnaker (2006) and Franke (2008) for more discussion of message credibility in a pragmatic context.

form to formally implement empirically motivated assumptions about the psychology of reasoners. Epistemic game theory explicitly models agents' epistemic states in order to explore the consequences of different kinds of (belief in) rationality and other assumptions about agents' interactive beliefs, mental architecture and reasoning capacities.

Concluding thus on a rather speculative note, I am tempted to say that it is to my mind less essential how much classical game theory ultimately survives in a cognitively adequate game theoretic model of pragmatic reasoning, as long as its ideals of mathematical precision and conceptual clarity be conserved.[21]

A References

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.

Aumann, Robert and Adam Brandenburger (1995). "Epistemic Conditions for Nash Equilibrium". In: *Econometrica* 63.5. Pp. 1161–1180.

Battigalli, Pierpaolo (2006). "Rationalization in Signaling Games: Theory and Applications". In: *International Game Theory Review* 8.1. Pp. 67–93.

Benz, Anton (2006). "Utility and Relevance of Answers". In: *Game Theory and Pragmatics*. Ed. by Anton Benz, Gerhard Jäger, and Robert van Rooij. Palgrave. Pp. 195–219.

— (2007). "On Relevance Scale Approaches". In: *Proceedings of Sinn und Bedeutung 11*. Ed. by Estela Puig-Waldmüller. Pp. 91–105.

Benz, Anton, Gerhard Jäger, and Robert van Rooij, eds. (2006). *Game Theory and Pragmatics*. Hampshire: Palgrave MacMillan.

Bernardo, José M. (1979). "Expected Information as Expected Utility". In: *The Annals of Statistics* 7.3. Pp. 686–690.

[21] This text grew out of a presentation held at the University of Osnabrück on May 28, 2008. I would like to thank the organizers, Saskia Nagel and Jan Slaby, for inviting me. For engaged discussion and warm-hearted hospitality, I would furthermore like to thank the audience of the talk, and in particular Peter Bosch, Judith Degen, Cornelia and Christian Ebert, Wulf Gärtner, Helmar Gust, Kai-Uwe Kühnberger and Achim Stephan. The paper directly profited from comments by and discussions with Judith Degen, Cornelia Ebert, Jacqueline Griego, Tikitu de Jager, Robert van Rooij and Marc Staudacher. Errors are my own.

Camerer, Colin F. (2003). *Behavioral Game Theory: Experiments in Strategic Interaction*. Princeton University Press.

Chapman, Siobhan (2005). *Paul Grice, Philosopher and Linguist*. Hampshire: Palgrave MacMillan.

Chierchia, Gennaro (2004). "Scalar Implicatures, Polarity Phenomena and the Syntax/Pragmatics Interface". In: *Structures and Beyond*. Ed. by Adriana Belletti. Oxford University Press. Pp. 39–103.

Chierchia, Gennaro, Danny Fox, and Benjamin Spector (2008). "The Grammatical View of Scalar Implicatures and the Relationship between Semantics and Pragmatics". Unpublished Manuscript.

Ducrot, Oswald (1973). *La preuve et le dire*. Maison Mame.

Farrell, Joseph (1993). "Meaning and Credibility in Cheap-Talk Games". In: *Games and Economic Behavior* 5. Pp. 514–531.

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

Franke, Michael (2008). "Meaning and Inference in Case of Conflict". In: *Proceedings of the 13th ESSLLI Student Session*. Ed. by Kata Balogh. Pp. 65–74.

Franke, Michael, Tikitu de Jager, and Robert van Rooij (2008). "Relevance in Cooperation and Conflict". Unpublished Manuscript, Universiteit van Amsterdam.

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

Gibbons, Robert (1992). *A Primer in Game Theory*. New York: Harvester Wheatsheaf.

Grafen, Alan (1990). "Biological Signals as Handicaps". In: *Journal of Theoretical Biology* 144. Pp. 517–546.

Grice, Paul Herbert (1989). *Studies in the Ways of Words*. Harvard University Press.

Hirschberg, Julia (1985). "A Theory of Scalar Implicature". PhD thesis. University of Pennsylvania.

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 I-based implicatures". In: *Meaning, Form, and Use in Context*. Ed. by Deborah Shiffrin. Washington: Georgetown University Press. Pp. 11–42.

— (2004). "Implicature". In: *The Handbook of Pragmatics*. Ed. by Laurence R. Horn and Gregory Ward. Blackwell. Pp. 3–28.

de Jaegher, Kris, Stephanie Rosenkranz, and Utz Weitzel (2008). “Economic Laboratory Experiment on Horn’s Rule”. Unpublished Manuscript, Utrecht School of Economics.

Jäger, Gerhard (2008). “Applications of Game Theory in Linguistics”. In: *Language and Linguistics Compass* 2/3. Pp. 406–421.

Kasher, Asa (1976). “Conversational Maxims and Rationality”. In: *Language in Focus: Foundations, Methods and Systems*. Ed. by Asa Kasher. Dordrech: Reidel. Pp. 197–216.

Kreps, David M. and Robert Wilson (1982). “Sequential Equilibria”. In: *Econometrica* 50.4. Pp. 863–894.

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

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

Lewis, David (1969). *Convention. A Philosophical Study*. Harvard University Press.

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.

Mill, John Stuart (1867). *An Examination of Sir William Hamilton’s Philosophy*. 3rd ed. London: Longman.

Morris, Charles M. (1946). *Signs, Language, and Behavior*. New York: Prentice Hall.

Myerson, Roger B. (1991). *Game Theory: Analysis of Conflict*. Harvard University Press.

Neale, Stephen (1992). “Paul Grice and the Philosophy of Language”. In: *Linguistics and Philosophy* 15. Pp. 509–559.

Noveck, Ira A. and Dan Sperber, eds. (2004). *Experimental Pragmatics*. Hampshire: Palgrave MacMillan.

Osborne, Martin J. (2004). *An Introduction to Game Theory*. New York: Oxford University Press.

Osborne, Martin J. and Ariel Rubinstein (1994). *A Course in Game Theory*. MIT Press.

Parikh, Prashant (1992). “A Game-Theoretic Account of Implicature”. In: *TARK ’92: Proceedings of the 4th conference on Theoretical aspects of reasoning about knowledge*. San Francisco: Morgan Kaufmann Publishers Inc. Pp. 85–94. ISBN: 1-55860-243-9.

Parikh, Prashant (2001). *The Use of Language*. Stanford University: CSLI Publications.

— (2006). “Pragmatics and Games of Partial Information”. In: *Game Theory and Pragmatics*. Ed. by Anton Benz, Gerhard Jäger, and Robert van Rooij. Palgrave MacMillan. Pp. 101–122.

van Rooij, Robert (2003). “Questioning to Resolve Decision Problems”. In: *Linguistics and Philosophy* 29. Pp. 727–763.

— (2004*a*). “Cooperative versus Argumentative Communication”. In: *Philosophia Scientiae* 8.2. Pp. 195–209.

— (2004*b*). “Signalling Games Select Horn-Strategies”. In: *Linguistics and Philosophy* 27. Pp. 493–527.

— (2004*c*). “Utility, Informativity and Protocols”. In: *Journal of Philosophical Logic* 33.4. Pp. 389–419.

— (2008). “Games and Quantity Implicatures”. In: *Journal of Economic Methodology* 15.3. Pp. 261–274.

Rothschild, Daniel (2008). “Grice, Utterance Choice, and Rationality”. Unpublished Manuscript, Columbia University.

Sally, David (2003). “Risky Speech: Behavioral Game Theory and Pragmatics”. In: *Journal of Pragmatics* 35. Pp. 1223–1245.

Selten, Reinhard (1965). “Spieltheoretische Behandlung eines Oligopolmodells mit Nachfrageträgheit”. In: *Zeitschrift für die die gesamte Staatswissenschaft* 121. Pp. 301–324.

— (1975). “Re-examination of the Perfectness Concept for Equilibrium Points in Extensive Games”. In: *International Journal of Game Theory* 4. Pp. 25–55.

Soames, Scott (1982). “How Presuppositions are Inherited: A Solution to the Projection Problem”. In: *Linguistic Inquiry* 13.3. Pp. 483–545.

Spence, Andrew Michael (1973). “Job market signaling”. In: *Quarterly Journal of Economics* 87. Pp. 355–374.

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 (1994). “On the Evaluation of Solution Concepts”. In: *Theory and Decision* 37. Pp. 49–73.

— (2006). “Saying and Meaning, Cheap Talk and Credibility”. In: *Game Theory and Pragmatics*. Ed. by Anton Benz, Gerhard Jäger, and Robert van Rooij. Hampshire: Palgrave MacMillan. Pp. 83–100.

Weibull, Jörgen W. (1997). *Evolutionary Game Theory*. MIT Press.

Zahavi, Amotz (1975). “Mate Selection - A Selection for a Handicap”. In: *Journal of Theoretical Biology* 53. Pp. 205–214.