Philosophy Compass 8/3 (2013): 269–284, 10.1111/phc3.12015

Game Theoretic Pragmatics

Michael Franke*
University of Amsterdam

Abstract
Game theoretic pragmatics is a small but growing part of formal pragmatics, the linguistic subfield studying language use. The general logic of a game theoretic explanation of a pragmatic phenomenon is this: (i) the conversational context is modelled as a *game* between speaker and hearer; (ii) an adequate *solution concept* then selects the to-be-explained behavior in the game model. For such an explanation to be convincing, both components, game model and solution concept, should be formulated and scrutinized as explicitly as possible. The article demonstrates this by a concise overview of both *evolutionary* and *non-evolutionary* approaches to game theoretic pragmatics, arguing for the use of *agent-based micro-dynamics* within evolutionary, and for the use of *epistemic game theory* within non-evolutionary approaches.

1. *The Theory of Games Meets Linguistic Pragmatics*

Game theory is a branch of applied mathematics that studies interactive decision making with many applications in sociology, political science, economics, biology, and, to not few people's surprise, also in linguistics. We will see why as soon as we understand what game theory is about. Game theory is about situations in which several agents (be it human beings, animals, institutions or whatever else is subject to investigation) must make decisions the outcome of which depends on what the other agents do. For example, if my twin brother and I go out for Lindy Hopping, we'd better not wear the same shoes, obviously, for otherwise no one would be able to tell us apart. Since mother got us both the same two pairs of shoes, namely Wellington boots and blue suede shoes, my brother and I have the same set of actions to choose from. The outcome of each of our individual choices depends on the choice of the other: wearing the same shoes would certainly be a terrible disaster, but additionally none of us would actually like to wear Wellington boots for Lindy Hop either. What is an everyday social dilemma of mediocre importance to the uninitiated, presents itself as a *strategic game* to the trained senses of the game theorist, which she would happily represent as such:

my choice	my brother's choice	
	a_{rubber}	a_{suede}
a_{rubber}	0,0	2,3
a_{suede}	3,2	1,1

My brother and I both have a choice between two actions: a_{rubber} and a_{suede}. I choose the row, my brother chooses the column. An outcome of the game is a pair of actions, one for me and one for my brother, and the table gives a numerical measure of the desirability of each possible outcome, the first number is for me (traditionally: the row player), the latter for my brother. The numbers are to a certain extent arbitrary.[1] In the present case, we model the inherited family preference of suede over rubber and the egocentric urge, often found in twins, to be adressed by the correct name.

A game model like the above is a specification of possible ways of behavior and of possible ways of forming beliefs about the choice situation. A game model does *not* specify what agents actually do or believe, or what they should do or believe. This is the task of so-called *solution concepts*. The most famous solution concept is that of a *Nash equilibrium*. A Nash equilibrium is a set of choices, one for each player, such that no player would be strictly better off when only she would deviate from the specified set of choices. There are two Nash equilibria (in pure strategies) in the rubber-or-suede game: one where I choose my blue suede shoes and my brother chooses his Wellingtons; and also the reverse pattern. Clearly, I prefer the former, which gives me a payoff of 3, to the latter, which gives me a payoff of 2. My brother has the exact opposite preference. Nash equilibrium does not yield a unique behavioral prediction in this case. Other solution concepts might, or else there might be a unique Nash equilibrium if we had construed the game in a different way. Much of game theory is exactly about this: studying formal representational models of interactive decision making, together with various solution concepts that select a certain subset of choices as, usually and in a to-be-specified sense, *optimal* behavior.[2]

Telling stories, giving commands and asking questions, also listening and figuring out what Joe meant when he said ''kind of'' – all of these linguistic activities have something essential in common with the rubber-or-suede game. Here and there, several agents have to make a decision how best to act given their preferences and their expectations about the others' behavior. As speakers we usually have a preference for getting our point across, or for persuading our hearers of our point of view. A speaker's action choices are naturally constituted by the set of possible utterances given by our grammatical competence, or some relevant subset thereof that interlocutors commonly attend to. Hearers may seek entertainment or information. The latter should preferably be relevant, or at least true. A hearer's action choices may be concrete ways of reacting to the words of a speaker, such as responding to a question, carrying out or denying a request. Or, they might be more abstract in the form of more or less conscious acts of forming an interpretation, i.e., an interpretation of what Joe meant when he said ''kind of''.

So what *did* Joe mean when he said ''kind of''? Recall that Joe uttered the sentence

(1) Sue and I, we are kind of married.

which is curious, because he could just have said

(2) Sue and I, we are married.

Why did he add ''kind of''? There are many possible explanations for his behavior. Still the most plausible ones among them share a common scaffolding which shows the marks of a game situation. The phenomenon is an instance of what is called *Horn's division of pragmatic labor*, in reference to Horn (1984).[3]

Had Joe uttered sentence (2), we would have come to believe that he is married to Sue fullstop. But he hasn't. He has gone out of his way and uttered something else, something more complex even. But then the extra effort for including ''kind of'' must

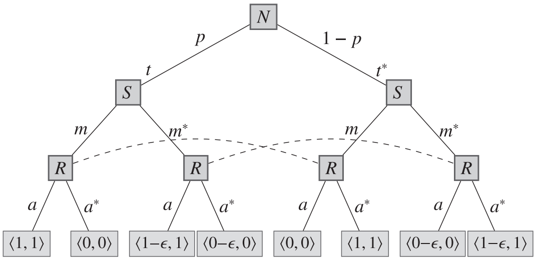

Fig. 1. Signaling game for Horn's division of pragmatic labor. Initially, Nature selects a state (with probabilities p and $1 - p$ respectively), then the sender selects a message, and finally the receiver selects an interpretation. The receiver does not know the actual state, which is indicated by the dotted line between those nodes in the game tree that the receiver cannot distinguish. Players share an interest in true communication, but use of the marked signal m^* incurs a small cost ϵ>0 for the sender.

have been for a reason. Clearly, Joe did not want us to come to believe simply that he is married to Sue. He wanted us to realize that there is something special about his relation to her. What's special about their relation cannot be deduced by reasoning about the speaker-hearer interaction alone, but must be abducted on the basis of further situational assumptions, a shared background of world knowledge and the like. Still, *that* there is something noteworthy about the relationship between Joe and Sue is an inference that is triggered by the use of ''kind of'' and cashed out by considering the speaker's and the hearer's choices, and expectations about their behavior.

Figure 1 gives one possible way of representing the utterance and interpretation of (1) by means of a game.[4] The game is a so-called *signaling game*, the kind of which was originally conceived by Lewis (1969). There are two players, the speaker and the hearer, which we will call sender and receiver for reasons of tradition. Unlike in the previous rubber-or-suede game, choices are made sequentially: first the speaker chooses an utterance, then the hearer interprets. Let's assume that the sender has a choice between two utterances: m and $m^\star$, corresponding to (2) and (1) respectively. We would also distinguish two states of affairs: one state t in which the relation between Joe and Sue is normal, and another state $t^\star$ in which it is abnormal in some unspecified way. Let's assume that the sender knows which state is actual, and that the receiver does not. This is modelled by assuming that the game begins with a chance move of nature the outcome of which only the sender can observe. To model stereotypicality we assume that a normal marriage situation is more likely than an abnormal one ($p > .5$). Since the receiver would like to know the true state that nature selected, his interpretation choices, a and $a^\star$, correspond one-to-one to the states that we distinguish. Players receive utilities for each possible sequence of play. The utilities shown in Fig. 1 implement the assumptions that (i) the receiver wants to know the true state, that (ii) the sender cooperatively shares the receiver's interest in successful communication, and that (iii) the sender also incurs a slight cost for sending the marked message $m^\star$.

There are two so-called *separating equilibria* for this game. These are Nash equilibria in which the sender uses different messages in different states, and the receiver correctly interprets the messages. Separating equilibria thus guarantee maximal communicative success for both states. (There are also equilibria that don't.) However, in one separating equilibrium, called the *Horn language* and represented in (3), the sender uses the marked message to communicate the marked state of affairs, whereas in the second separating equilibrium, called the *Anti-Horn language* and represented in (4), the reversed association of messages and states/actions holds.

(3)
$$
\begin{array}{l}
t \longrightarrow m \longrightarrow a \\[2ex]
t^{*} \longrightarrow m^{*} \longrightarrow a^{*}
\end{array}
$$

(4)
$$
\begin{tikzcd}
t \arrow[dr] & m \arrow[dr] & a \\
t^{*} \arrow[ur] & m^{*} \arrow[ur] & a^{*}
\end{tikzcd}
$$

The receiver's average payoff is a full optimal 1 for both equilibria. As for the sender, however, with $p > .5$, the Horn language is more efficient, with an average payoff of $p \times 1 + (1-p) \times (1-\epsilon)$, which is strictly bigger than the average payoff of $p \times (1-\epsilon) + (1-p) \times 1$ for the Anti-Horn language. Intuitively speaking, this is because (4) uses the costly message more often than (3) does.

Suppose now that I could convince you that due to its higher speaker efficiency the equilibrium in (3) is the *optimal* speaker-hearer behavior in the game in Fig. 1 and therefore the only one that we need to look at. Suppose also that you consent that the game in Fig. 1 is an adequate description of most, or maybe even all, relevant parameters that feed into the interpretation of an utterance of (1). In that case, I would have been able to offer to you an *explanation* for the pragmatic inference in question. I would have explained the inference with the help of (i) a representation of the linguistic context and (ii) a solution that uniquely selects the to-be-explained behavior given the context model. In general outline this is the logic of a game theoretic explanation of some linguistically interesting observation about language use and interpretation.

I am, however, absolutely convinced that you are not satisfied with the quality of the explanation so far. Indeed, you should not be. For a game theoretical explanation along these lines to be worth the readers' while, the assumptions underlying both the game models as well as the solution concepts need to be made explicit and justified in the context of the phenomenon that is to be explained. A basic distinction is to be made between *evolutionary* and *non-evolutionary approaches*. The latter explicitly model the speaker's and hearer's purported *reasoning* about language use; the former model diachronic processes of learning and language organization by gradual adaptation. Evolutionary and non-evolutionary approaches usually differ in how much rationality and representational ability we want to ascribe to pragmatic reasoners. I argue here that we should address this issue as explicitly as possible. Doing so entails shifting away from *static* solution concepts, like Nash equilibrium and evolutionary stability, to more *dynamic* solution concepts that address, not only what a stable outcome of playing a game is, but also *how* agents can arrive at a given outcome, be it by sophisticated learning, gradual adaptation or by explicit strategic reasoning. I suggest that the weapon of choice for the philosophically-minded and conceptually discriminating among us are: agent-based *micro-dynamics* for evolutionary models; and *epistemic game theory* for rationalistic models. The following two sections explain these notions against the background of a brief survey of the most influential theorizing within both branches of game theoretic pragmatics (cf. Jäger 2008a).

2. The Pragmatic Basis of Meaning

David Lewis was the first who applied game theoretic ideas to philosophical questions about language in his seminal work on *convention* (Lewis 1969). Inspired by the work of Schelling (1960), Lewis tried to give a non-circular naturalistic grounding of the notion of conventional meaning in terms of repeated plays of signaling games, thus answering challenges to conventionalism put forward by, among others, Russell (1921) and Quine (1936). Lewis' analysis of convention has been criticized and rectified in multiple ways (see Rescorla 2011, for overview). For instance, Vanderschraaf (1995) and Skyrms (1996, Chapter 4) argue for an analysis of convention in terms of *correlated equilibrium* (see Aumann 1974).[5] Sugden (1986) builds on Lewis' account, arguing to use *evolutionary stability* instead (cf. Skyrms 1996, Chapter 5).[6]

Indeed, Lewis' account of meaning evolution in terms of signaling games can be found in many later and independent approaches to language evolution (cf. Steels 1995, 1997; Nowak and Krakauer 1999; Lenaerts *et al.* 2005; Loreto *et al.* 2010) and is still widely considered the baseline model around which a lot of exciting current research revolves (see Skyrms 2010, for overview). As the simplest non-trivial case, Lewis considered a signaling game like the one in Fig. 1 where $p = .5$ and $\epsilon{=}0$. If this signaling game is played repeatedly (among two players or within a population of players), then, although messages might initially be entirely meaningless, meaningfulness may eventually evolve from the particular use of signals. In fact, the two patterns in (3) and (4) are the only *evolutionary stable strategies* (see Footnote 6) for this game and in both of these the use of signals bestows a meaning on signals. In (3), for instance, message m could be said to mean that the current state is t, while $m^{\star}$ means that the current state is $t^{\star}$, because these messages are used exclusively in these states. That both (3) and (4) are equally good candidates (when $p = .5$ and $\epsilon{=}0$) does justice to the intuition that linguistic meaning conventions are mostly arbitrary.[7]

None of this, however, says anything yet about how a population could arrive at a signaling convention either *vertically* by learning within a generation, or *horizontally* by some gradual process of adaptation across generations. For this, it is necessary to study so-called *evolutionary dynamics*. We should distinguish *macro-dynamics* from *micro-dynamics*. The former describe the dynamic change of the population as a whole, whereas the latter zooms in on each single agent, describing the dynamic change in terms of how each agent adapts her behavior over time.

The most prominent macro-dynamic studied in evolutionary game theory is certainly the *replicator dynamic*, an instance of so-called monotone-selection dynamics.[8] The general idea behind these kinds of dynamics is that the rate of change $\dot{B}_\tau$ in proportion B_τ of a particular kind of behavior B at time τ is proportional to how successful B is when playing against the rest of the population at time τ. (The replicator dynamic ensues if we assume that the *per capita* change $\dot{B}_\tau / B_\tau$ is equal to the success of B against the population minus the average success of all types of behavior against the population.)

There is quite some leeway in how to interpret a given macro-dynamic. The replicator dynamic, for instance, is usually conceptualized as follows. We assume that each individual in the given population follows an inborn behavioral pattern. For a signaling game like in Fig. 1, this is a pair of so-called pure strategies, one for the sender and one for the receiver. A pure sender (receiver) strategy is a function from states to messages (from messages to acts). We would furthermore assume that the population is (virtually) *infinite* and *homogenous*, such that all pairs of individuals are equally likely to interact with one another. Reproduction is asexual, mutation-free and the number of offspring of each individual is proportional to its *fitness*, which in turn is proportional to how well the inborn behavior of the individual in question fares against the *whole* of the population.

For the simple signaling game in Fig. 1, the behavior of the replicator dynamic is quite well understood. When p=.5 and ϵ=0, the situation can be summarized as follows (cf. Huttegger 2007). Only the separating equilibria (3) and (4) have substantial *basins of attraction*, i.e., population configurations close to (3) or (4) evolve towards (3) or (4). Although the equilibrium in (5), for instance, is also a *rest point* of the replicator dynamic (once there we do not move away), it has no substantial basin of attraction. When we look at a case with $p > .5$ and $\epsilon>0$, this changes. In that case, the behavioral pattern in (5), which is called *Smolensky language*, does have a non-negligible basin of attraction. Numerical simulations show that with $p = .75$ and ϵ=.2, for instance, ca. 15% of randomly chosen initial population configurations converge towards the Smolensky language in (5), ca. 58.5% evolve towards the more efficient Horn language (3), ca. 25.5% towards the Anti-Horn language (4).

(5)
$$\begin{tikzcd} t \arrow[r] & m \arrow[r] & a \\ t^* \arrow[ur] & m^* \arrow[ur] & a^* \end{tikzcd}$$

Notice how game theory in principle is able to also make *quantitative predictions* about how likely a particular linguistic pattern is to be observed (cf. Jäger 2007, for an application of this idea to differential case marking).

The cognitive abilities of agents required by (this interpretation of) the replicator dynamic are minimal. Agents need not be able to select optimal play or maintain a record of past encounters. In fact, agents need not even be able to conceptualize the game they are playing, as long as successful play affects the (asexual) reproduction rate. This is why the replicator dynamic (under this interpretation) seems more feasible as an explanation for the evolution of fairly basic signaling ability between non-cognizing agents, such as cells, or lower animals, than it is a model of meaning evolution among human agents. But even then, some assumptions of this interpretation might still seem dubious if not implausible, such as the deterministic asexual inheritance and homogeneity of the population.[9]

It is possible, however, to interpret the replicator dynamic also as a special case of a particular micro-dynamic, namely as the most likely path of evolution of a homogenous and large-enough society of agents each of which adapts her behavior by *conditional imitation* (Helbing 1996; Schlag 1998).[10] Intuitively speaking, an agent who updates her behavior by conditional imitation plays a pure strategy as before. Every now and then, she checks how well her neighbors fare with the strategy that they play. If some neighbor does better than she does, then she would imitate that neighbor's behavior with a probability proportional to how much better that neighbor is doing. This micro-level interpretation assumes slightly more sophisticated agents that are capable of tracing successes and also of distinguishing between the agents they interact with. Still, the agents are not fully rational, for otherwise they would play optimally against the observed behavior. (Imitation dynamics are *non-innovative*, while playing a best response might be strictly *innovative*: a rational best response to what all neighbors play might be a strategy that no neighbor plays, but only that which is present in the neighborhood can be imitated.) Although I certainly do not want to argue that conditional imitation entails exactly the right level of agent sophistication to account for meaning evolution, it should nonetheless be clear that it is a micro-level approach to evolutionary dynamics that offers enough grip on the agent-internal perspective to address issues of philosophical and linguistic importance within game theoretic models.

Moreover, using micro-level evolutionary dynamics we can also more readily dispense with the assumption that populations are homogenous. When the evolutionary dynamic is defined at the agent-level, it is easy to integrate more complex *interaction structures* that define which agents (are how likely to) interact with one another. The most straightforward implementation of an interaction structure is an undirected graph, whose nodes are associated with the agents and whose edges link those agents that may play against each other. Zollman (2005), for example, has shown that when agents are placed on a grid where every agent can communicate with his eight nearest neighbors (north, northwest, west ...), then a (non-conditional) imitation dynamic leads to stable patterns of *regional languages* where some connected regions of agents learned one language, while others learned another.[11] Using insights from sociology about the properties of friendship networks, it is easily possible to study micro-dynamics also on more realistic, yet algorithmically generated interaction structures, such as *small-world networks* (Watts and Strogatz 1998), *scale-free networks* (Barabási and Albert 1999), or even on network topologies taken, for example, from social network sites.[12] To give a concrete example of the influence of interaction structure for a variant of the game in Fig. 1, Table 1 compares the outcome of the replicator dynamic to that of the conditional imitation rule for 250 agents placed on a small-world network. We see that the micro-dynamic and the more realistic interaction structure lead to the most efficient Horn language in (3) more often, and less frequently to the non-separating Smolensky language in (5).

3. Rational Pragmatic Language Use

The pragmatic grounding of conventional meaning is not the only issue that game theoretic models have been applied to, of course. Many phenomena that concern linguistic

Table 1. Contingency table comparing the replicator dynamic (macro) with the conditional imitation rule (micro) for the game in Fig. 1 with $p = .75$ and $\epsilon = .2$. For each of 200 randomly generated initial strategy distributions, the simulation checked the outcome of both the macro- and the micro-level dynamic. The frequencies with which each pair of outcomes was observed is given in the table. For the micro-dynamic we used a small-world network of 250 agents as interaction structure and made sure that the randomly selected population distribution could be matched exactly by assigning pure strategies to 250 agents. The category ''none'' subsumes all cases where no (unique) language (for speaker and receiver role) was acquired. ''H'' is short for ''Horn'', ''AH'' for ''Anti-Horn'', and ''SM'' for ''Smolensky'' (see (3), (4), and (5) respectively).

	Micro				
Macro	**H**	**AH**	**SM**	**None**	**Total**
H	121	0	0	3	124
AH	3	40	0	0	43
SM	13	10	2	4	29
None	4	0	0	0	4
Total	141	50	2	7	

pragmatics in a more narrow sense have been investigated, too. The pioneer of this approach is Prashant Parikh, who started already in the late 1980s to use game theory for an explication of utterance meaning (Parikh 1991, 2000, 2001), for conversational implicatures (Parikh 1992, 2001), and many other pragmatic phenomena, such as jokes and misunderstandings (Parikh 2001), or discourse reference (Clark and Parikh 2007; Parikh 2010). In his most recent work (Parikh 2010), Parikh builds on game theory to present an all-encompassing framework of linguistic and non-linguistic communication, which he calls *equilibrium semantics*. This appellation highlights (i) the importance of the notion of equilibrium between the speaker's intended meaning and the hearer's interpretation for the notion of communication and (ii) the rather radical subsumption of all aspects of meaning, more traditionally divided into semantics or pragmatics, under one general framework.

It would go too far for this short overview article to discuss Parikh's innovative and multifaceted work in great detail. To get a general idea nonetheless, let us focus on a few important and reoccurring ideas of his theorizing (cf. Sally 2003; Allott 2006, for cogent criticism of Parikh's work). According to Parikh, the central task of the hearer in utterance interpretation is that of *disambiguation*, such as that of sentence (6) which is ambiguous between at least the two potential meanings in (6a) and (6b).

(6) Harvey is playing.
    a. The boy called ''Harvey'' is playing (with his friends or toys).
    b. The movie called ''Harvey'' is playing (at the theaters).

In a concrete situation where (6) has been uttered, the hearer needs to determine which of the two senses the speaker had intended to convey. Parikh assumes that speaker and hearer construct this disambiguation task, using shared situational and world knowledge, as, what he calls, a *game of partial information*, such as that in Fig. 2.[13] By making explicit what speaker and hearer know about the game and each other's behavior, Parikh argues that the notion of communication can be grounded in equilibrium analyses of these games (Parikh 1991, 2000, 2001).

What is crucial here, is that those games of partial information that Parikh considers have multiple equilibria, just like the games we looked at before. Parikh therefore uses *Pareto-dominance* as a second-order selection criterion among Nash equilibria.[14] For example, both pairs of strategies in (7) and (8) are equilibria of the game in Fig. 2, but with $p > .5$

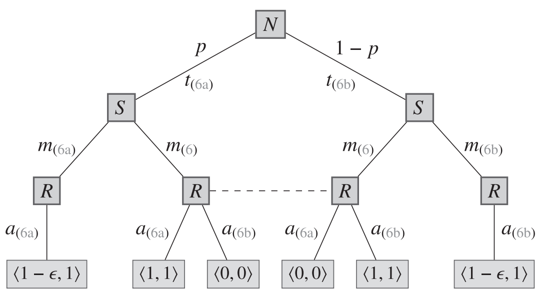

Fig. 2. Game of partial information for interpretation of (6). The states $t_{(6a)}$ and $t_{(6b)}$ are situations in which the speaker intends to convey (6a) and (6b) respectively. Indices on speaker choices point to the different corresponding sentences (6), (6a) and (6b). Finally, the receiver selects an interpretation of the observed message. Notice that games of partial information rule out untruthful use and interpretation of utterances as part of the game model.

(7) Pareto-dominates (8), because the sender is better off in (7) (she needs to pay the message costs ϵ only for the infrequent state) while the receiver is equally well off.

(7)
$$\begin{tikzcd}
t_{(6a)} \arrow[dr] & m_{(6a)} \arrow[r] & a_{(6a)} \\
t_{(6b)} \arrow[dr] & m_{(6)} \arrow[ur] & a_{(6b)} \\
& m_{(6b)} \arrow[ur] &
\end{tikzcd}$$

(8)
$$\begin{tikzcd}
t_{(6a)} \arrow[r] & m_{(6a)} \arrow[r] & a_{(6a)} \\
t_{(6b)} \arrow[r] & m_{(6)} \arrow[r] & a_{(6b)} \\
& m_{(6b)} \arrow[ur] &
\end{tikzcd}$$

Already this simple example shows some interesting problems with Parikh's use of Pareto-dominance.[15] Clearly, (7) Pareto-dominates (8) only if $p > .5$. (If $p < .5$ the dominance relation is reversed; if $p = .5$ Pareto-dominance does not select a unique equilibrium.) This means that the prior probabilities determine rather directly which outcome is predicted by the theory. As such this is not a problem, but it does emphasize the relevance of interpreting what the probabilities represent exactly. Allott (2006) argues against a frequentist interpretation of these probabilities, drawing on earlier work by Wilson and Matsui (1998). Firstly, if priors model the *frequency of the speaker intending this or that*, this should arguably not be independent of the utterance the speaker made, in which case the game theoretic explanation looses much of its bite, since we simply select the most probable speaker meaning associated with a given form. But if *frequency of occurrence of the described state* was relevant for disambiguation, then (9) should most likely be interpreted as (9a) and not as (9b), contrary to intuition.

(9) John wrote a letter.
a. John wrote a letter of the alphabet.
b. John wrote a letter of correspondence.

Allott concludes that prior probabilities in a game model as used by Parikh should rather "reflect the psycholinguistic activation and accessibility data" (Allott 2006, 134).[16]

But then, on the supposition that this is correct, the use of Pareto-dominance for equilibrium selection is cast into doubt, because only under a frequentist interpretation of the priors would the speaker actually experience any payoff difference. If the priors only encode strength of activation, the reason why (7) is preferred over (8) should also be given in psychological terms, like ease of activation or salience (e.g. in the sense of Schelling 1960), not in terms of payoff considerations alone.[17]

What this brief venture into the interpretation of priors and the sensibility of Pareto-dominance again shows is that game theoretic models benefit from explicitly modeling the agent-internal perspective. For non-evolutionary approaches, such a perspective is taken by *epistemic game theory*, which is a rather recent trend at the interface between game theory and formal epistemology (cf. Perea 2012). Epistemic game theory tries to determine which behavioral predictions follow from which assumptions about agents' (higher-order) beliefs, their rationality, their dispositions to act or revise beliefs and so on. This approach has two advantages. Firstly, if solution concepts are formulated explicitly in terms of agents' epistemic and behavioral dispositions, additional assumptions about the psychology of reasoners are more easily integrated. We will see an example of this presently. Secondly, we can identify the *epistemic conditions* under which the predictions of a given classical solution concept are obtained. For instance, with some due simplification, the predictions of Nash equilibrium ensue if all agents act rationally and know what everybody else is doing (cf. Stalnaker 1994; Aumann and Brandenburger 1995). The epistemic characterization of a solution concept thus indicates whether its application to a given scenario is sensible or not. As for the disambiguation of a sentence like (6), it seems implausible to assume that interlocutors *know* their behavior for certain, and therefore a mere equilibrium analysis, at least under this particular characterization, would seem inappropriate.

Of course, there are other epistemic conditions sufficient for selecting equilibrium behavior. One possibility, for example, is to formalize step-by-step reasoning towards a use-interpretation equilibrium in terms of interlocutors' beliefs and rationality, possibly mediated by additional psychological factors (cf. Benz 2006; Stalnaker 2006; Benz and van Rooij 2007; Jäger and Ebert 2009; Franke 2011; Jäger forthcoming). The main idea is that pragmatic reasoning starts by considering some salient, perhaps unstrategic behavior of either speaker or listener. Call this level-0 behavior. Level-(k + 1) behavior is then defined, in simplified terms, as a rational strategy against level-k behavior. For example, in the case of the game in Fig. 2, we could assume, in line with the idea that the prior probabilities $p > .5$ encode a behavioral disposition of the listener to first think of interpretation (6a) when hearing (6), that the interpretation strategy in (10) is a reasonable focal starting point for pragmatic reasoning.

(10)
$$\begin{tikzcd}
m_{(6a)} \arrow[r] & a_{(6a)} \\
m_{(6)} \arrow[ur] & a_{(6b)} \\
m_{(6b)} \arrow[ur] &
\end{tikzcd}$$

(11)
$$\begin{tikzcd}
t_{(6a)} \arrow[dr] & m_{(6a)} \\
t_{(6b)} \arrow[dr] & m_{(6)} \\
 & m_{(6b)}
\end{tikzcd}$$

Since $\epsilon>0$, the speaker's only rational behavior in response to the supposition that the listener plays (10) is (11). But since the only rational response to (11) is again (10) we have already reached a fixed point of this reasoning sequence. It can be shown that for certain classes of game models fixed points exist and are equilibria (e.g. Franke 2011), and indeed even Pareto-optimal under certain extra assumptions (Benz and van Rooij 2007). We could therefore think of this approach as a more epistemically explicit way of solving the problem of equilibrium selection for online pragmatic interpretation.

This kind of so-called *iterated best response* (IBR) reasoning not only works well for the simple disambiguation game in Fig. 2. It is not difficult to see that IBR reasoning also selects the Horn language (3) for the game in Fig. 1 (e.g. Franke 2009, 61–5). Still, the predominant application of the IBR approach so far has been to *quantity implicatures*, a special case of conversational implicature (Grice 1975).[18] Quantity implicatures are a prime case for game theoretic modeling, as they are thought to arise from reasoning about what other, logically stronger utterances a speaker could have made (cf. Gazdar 1979; Levinson 1983; Horn 1984; Geurts 2010). A simple example is the scalar inference that an utterance of the weaker (12a) normally implicates that the stronger (12b) is false.

(12) a. I occasionally listen to desert rock.
b. I frequently listen to desert rock.

Further applications include the use and interpretation of conditionals (Franke 2009, Chapter 5), or that of vague numerals (Jäger 2011).

Moreover, unlike many other formal approaches in pragmatics, the IBR model is not confined to cases of pure cooperation, and yields intuitively correct predictions when the usual Gricean assumption of pure cooperation is (partially or fully) leveled. In that case, the model explains our intuitions about (i) which aspects of conventional meaning are still *credible* (Rabin 1990; Farrell 1993; Farrell and Rabin 1996), as well as (ii) which pragmatic inferences that are normally attested in cooperative dialogue would still arise in non-cooperative settings (Franke 2010; Franke *et al.* 2012). For example, if the speaker of (12a) would actually not want to admit the truth of (12b), then the IBR model would predict that listeners who are naïve enough (level-1 interpreters) would draw the scalar inference that is generally attested in cooperative contexts, while more rational, higher-level reasoners would not, because they realize that the speaker has an incentive to make a strategic understatement (cf. Franke *et al.* 2012).

Other approaches to game theoretic pragmatics give up the assumption that conversation is mainly about cooperative exchange of relevant information even more radically. Building on work by Ducrot (1973) and Anscombre and Ducrot (1983), Merin (1999) demonstrates how pragmatic interpretation arises from conflicting interests of speaker and hearer to argue for or against a given hypothesis (cf. van Rooij 2004a, for overview). This *argumentative approach* to pragmatics is also taken, inter alia, by Rubinstein (2000) and Glazer and Rubinstein (2006). Non-cooperative language use has been studied intensively in economics, albeit not necessarily with a specifically linguistic interest (cf. Crawford and Sobel 1982; Myerson 1989; Rabin 1994). It may be expected that under the influence of game theoretic models linguistic pragmatics will also extend its scope beyond the special case of unshakable cooperativity in the future.

4. Conclusion

Game theoretic approaches within linguistics are tightly connected to use-based theories of meaning and language organization. Game theoretic models will appeal most straightforwardly to those who like to consider key features of language, on all its levels of organization, as an abstraction over intricate patterns of behavior (and its underlying cognitive processes) in which both psychological and social factors play a role, both at present, as well as during the adaptive processes that lead to the current state of affairs. In other words, game theory will seem like a good formal modeling tool to those who consider language a *complex adaptive system*, much in the sense of Beckner *et al.* (2009).

The role that game theory could play in such a view of language is similar to the role that logic plays in certain formal theories of natural language meaning. Game theory might be the benchmark model which to formulate a theory of language use in, or at least which to compare it to. But we could take the analogy with logic even further. If we consider game theory to operate on a high level of abstraction, as logic does too when seen in a Fregean anti-psychologistic light, we might want to take an *instrumentalist stance* towards our models, and say that the main benefit of the modeling approach is to learn from the moment of breakdown: we learn about the linguistic reality mostly when the model fails us; we see clearly that which the model does not capture only *because of* its failure. On the other hand, we might also be gradually more realistic about our models, in particular the assumed representations (game models) and processes (action choices, learning, belief formation ...). It is the more *cognitively-realistic stance* that promises to offer more interesting empirical predictions, because it does not shy away from the psycholinguistic laboratory. Yet, in order to make serious and testable predictions it is crucial to adopt the explicitly agent-internal perspective that this paper argued for.

Short Biography

Michael Franke's work lies at the junction between theoretical linguistics, philosophy, logic and the cognitive sciences. He holds a BSc in cognitive science from Osnabrück University, an MSc in Logic and a PhD in Philosophy from the University of Amsterdam. Most of his previous work was on applications of formal models from game and decision theory to linguistics. After spending two years as a researcher at the Department of Linguistics of the University of Tübingen, he is currently a researcher at the Institute for Logic, Language and Computation in Amsterdam where he is working on the formation of graded concepts by linguistic pressures from an evolutionary point of view.

Notes

* Correspondence: Michael Franke, FNWI ILLC, Universiteit van Amsterdam, P.O. Box 94242, 1090 GE, Amsterdam, the Netherlands. Email: m.franke@uva.nl

[1] The precise significance of the numerical information actually depends on the solution concept (see below) that would be used. In standard game theory, where utilities combine with probabilities to form a total preference relation over expected outcomes, any linear transformation of utilities may be considered equivalent. This need not be so, however, for certain evolutionary solution concepts. This issue is conceptually interesting and highly relevant for sound applications of the formal models, but, for reasons of space, we will have to put it aside here.

[2] There are many excellent introductions to game theory, e.g., Myerson (1991), Gibbons (1992), Osborne and Rubinstein (1994), Osborne (2004) or Heap and Varoufakis (2004). A good introduction with a linguistic focus is the first chapter of Benz *et al.* (2006).

[3] Alternatively, the label *M*-implicature is used (Levinson 2000) in allusion to Grice's Maxim of Manner (Grice 1975). It is Horn's division of pragmatic labor that has motivated the notion of *weak* bidirectional optimality (Blutner 1998, 2000).

[4] Accounts of Horn's division of pragmatic labor in terms of context models given by (something like) these signaling games have been proposed, among others, by van Rooij (2004c, 2006), Benz and van Rooij (2007), de Jaegher (2008), Franke (2009), Mühlenbernd (2011), Franke and Jäger (2012), and Jäger (forthcoming).

[5] Roughly speaking, a correlated equilibrium for the rubber-or-suede game would consists in me and my brother tossing a fair coin to decide who is allowed to wear suede. (Notice that this yields a better average outcome than randomizing independently, which still may lead to both wearing the same shoes.)

[6] With due oversimplification, an evolutionary stable strategy is a strategy that if played by a population, that population will not be taken over by small amounts of invading mutants (see Maynard Smith 1982). In role-conditioned games, like signaling games, evolutionary stable strategies are exactly the strict Nash equilibria, i.e., the sets of pure strategies such that each agent would do strictly worse deviating from the set if all others conformed (Selten 1980).

[7] Characterizing which sets of strategies are stable is, of course, more complicated when sender and receiver may have different interests than perfect communication (cf. Crawford and Sobel 1982), when we have more than two states and/or more than two messages (cf. Blume *et al.* 1993), and also when $p \neq .5$ and $\epsilon \neq 0$ (cf. Jäger 2008b).

[8] The replicator dynamic is due to Taylor and Jonker (1978). See Weibull (1997), Cressman (2003), Hofbauer and Sigmund (2003) or Gintis (2009) for a general introduction to evolutionary game theory. Huttegger (2007) and Huttegger *et al*. (2010) analyze Lewisean signaling games from a dynamic evolutionary point of view (see also Skyrms 2010).

[9] There are, of course, also non-deterministic macro-dynamics, such as the *mutator-replicator dynamic*, and also macro-dynamics modeling sexual reproduction (see Page and Nowak 2002; Nowak 2006a, for accessible overview).

[10] Szabó and Fáth (2007) and Roca *et al*. (2009) provide an overview of agent-based dynamics within evolutionary game theory and also cover the special case of the replicator dynamic.

[11] Wagner (2009) studied the same dynamic on small-world networks (see Footnote 12), looking also at cases with non-uniform priors. Mühlenbernd (2011) investigated a grid structure with more complex update dynamics, namely reinforcement learning and best-response dynamics also for signaling games with costs and skewed priors. Mühlenbernd and Franke (2012) studied these latter dynamics on small-world networks.

[12] A small-world is a network in which, intuitively speaking, (i) there are relatively few steps necessary to get from any one agent to any other via a path of links in the network, and (ii) for most pairs of ''friends'' (connected nodes) of a given ''person'' (node), these friends are also friends of each other. A scale-free network has both of these properties, too. Additionally the distribution given by the number of nodes with 1, 2, 3 ... connections follows a power law distribution (at least for some interval of natural numbers).

[13] A game of partial information is subtly different from standard signaling games (see Parikh 2006, for discussion). This is in large part due to the fact that these games are linked explicitly to *situation theory* (Barwise and Perry 1983; Barwise and Seligman 1997). The difference shows, for instance, in the distinction between *local games* that the hearer represents to himself after hearing a *particular* utterance, and *global games* that the speaker represents to herself involving many, possibly unrelated utterances, and their associated local games. For ease of exposition, Fig. 2 only gives a local game with a simplified payoff structure (for more detail, see Parikh 2010, 87–97).

[14] A pair of strategies, one for each player, is *Pareto-dominant* if there is no other pair of strategies such that one player does better, and the other player does not do worse under the latter than under the former.

[15] Parikh (2010) shows that for the subclass of games he considers most relevant Pareto-dominance is equivalent to *risk dominance*, which was proposed by Harsanyi and Selten (1988) as an equilibrium selection criterion. An equilibrium risk-dominates another, if it is less risky, in the sense that the players do not have to fear severe losses, should other players unexpectedly *not* conform to the equilibrium.

[16] A similar point is made by Franke (2009), 129–33.

[17] In a similar spirit, Sally (2003) demonstrates where and how psychological insights about how subjects play coordination games could be applicable within linguistic pragmatics. In general, the idea of focal point reasoning appears to have a lot more potential for further linguistic applications (cf. Potts 2008; Franke 2012).

[18] There are many other game-theoretic reconstructions of conversational implicatures (cf. van Rooij 2006, 2008, for overview). Next to Parikh (1992, 2001), we find de Jager and van Rooij (2007) who give an equilibrium-based account of quantity implicatures. Asher *et al*. (2001), van Rooy (2003), van Rooij (2004b), Scott-Phillips (2010) and Grim (2011) all give game theoretic underpinnings of various (post-)Gricean conversational principles. From the latter point of view, also the more general problem of why cooperation is prevalent in linguistic interaction is relevant (see Nowak 2006b, for overview).

References

Allott, Nicholas. 'Game Theory and Communication.' In *Game Theory and Pragmatics*. Eds Anton Benz. Gerhard, Jäger and Robert van Rooij. Hampshire: Palgrave MacMillan, 2006. 123–51.

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

Asher, Nicholas, et al. 'Game Theoretical Foundations for Gricean Constraints.' In *Proceedings of the 13th Amsterdam Colloquium*. Eds Robert van Rooy and Martin Stokhof. 2001. 31–7.

Aumann, Robert. 'Subjectivity and Correlation in Randomized Strategies.' *Journal of Mathematical Economics* 1.1 (1974): 67–96.

Aumann, Robert, and Adam Brandenburger. 'Epistemic Conditions for Nash Equilibrium.' *Econometrica* 63.5 (1995): 1161–80.

Barabási, Alber-László, and Réka Albert. 'Emergence of Scaling in Random Networks.' *Science* 286. 5439 (1999): 509–12.

Barwise, Jon, and John Perry. *Situations and Attitudes*. Cambridge, MA: MIT Press, 1983.

Barwise, Jon, and Jerry Seligman. *Information Flow: The Logic of Distributed Systems*. Cambridge, UK: Cambridge University Press, 1997.

Beckner, Clay, et al. 'Language Is a Complex Adaptive System: Position Paper.' *Language Learning* 59 (2009): 1–26.

Benz, Anton, et al. *Game Theory and Pragmatics*. Hampshire: Palgrave MacMillan, 2006.

Benz, Anton. 'Utility and Relevance of Answers.' In *Game Theory and Pragmatics*. Eds Anton Benz. Gerhard, Jäger and Robert van Rooij. Hampshire: Palgrave, 2006. 195–219.

Benz, Anton, and Robert van Rooij. 'Optimal Assertions and What They Implicate.' *Topoi* 26 (2007): 63–78.

Blume, Andreas, et al. 'Evolutionary Stability in Games of Communication.' *Games and Economic Behavior* 5.5 (1993): 547–75.
Blutner, Reinhard. 'Lexical Pragmatics.' *Journal of Semantics* 15 (1998): 115–62.
Blutner, Reinhard. 'Some Aspects of Optimality in Natural Language Interpretation.' *Journal of Semantics* 17 (2000): 189–216.
Clark, Robin, and Prashant Parikh. 'Game Theory and Discourse Anaphora.' *Journal of Logic, Language and Information* 16 (2007): 265–82.
Crawford, Vincent P., and Joel Sobel, 'Strategic Information Transmission.' *Econometrica* 50 (1982): 1431–51.
Cressman, Ross. *Evolutionary Dynamics and Extensive Form Games*. Cambridge, MA: MIT Press, 2003.
Ducrot, Oswald. *La preuve et le dire*. Paris: Maison Mame, 1973.
Farrell, Joseph. 'Meaning and Credibility in Cheap-Talk Games.' *Games and Economic Behavior* 5 (1993): 514–31.
Farrell, Joseph, and Matthew Rabin. 'Cheap Talk.' *The Journal of Economic Perspectives* 10.3 (1996): 103–18.
Franke, Michael. *Signal to Act: Game Theory in Pragmatics*. PhD thesis, Universiteit van Amsterdam, 2009.
Franke, Michael. 'Semantic Meaning and Pragmatic Inference in Non-cooperative Conversation.' In *Interfaces Explorations in Logic, Language and Computation*. Eds Thomas Icard and Reinhard Muskens. Lecture Notes in Artificial Intelligence. Berlin, Heidelberg: Springer-Verlag, 2010. 13–24.
Franke, Michael. 'Quantity Implicatures, Exhaustive Interpretation, and Rational Conversation.' *Semantics & Pragmatics* 4.1 (2011): 1–82.
Franke, Michael. 'On Scales, Salience & Referential Language Use.' In *Amsterdam Colloquium 2011*. Eds Maria Aloni, Floris Roelofsen and Katrin Schulz. Lecture Notes in Computer Science. Berlin, Heidelberg: Springer, 2012. 311–20.
Franke, Michael, and Gerhard Jäger. 'Bidirectional Optimization from Reasoning and Learning in Games.' *Journal of Logic, Language and Information* 21.1 (2012): 117–39.
Franke, Michael, et al. 'Relevance in Cooperation and Conflict.' *Journal of Logic and Computation* 22.1 (2012): 23–54.
Gazdar, Gerald. *Pragmatics: Implicature, Presupposition, and Logical Form*. New York: Academic Press, 1979.
Geurts, Bart. *Quantity Implicatures*. Cambridge, UK: Cambridge University Press, 2010.
Gibbons, Robert. *A Primer in Game Theory*. New York: Harvester Wheatsheaf, 1992.
Gintis, Herbert. *Game Theory Evolving: A Problem-Centered Introduction to Evolutionary Game Theory*. Princeton, NJ: Princeton University Press, 2009.
Glazer, Jacob, and Ariel Rubinstein. 'A Game Theoretic Approach to the Pragmatics of Debate.' In *Game Theory and Pragmatics*. Eds Anton Benz. Gerhard, Jäger and Robert van Rooij. Hampshire: Palgrave MacMillan, 2006. 248–62.
Grice, Paul Herbert. 'Logic and Conversation.' In *Syntax and Semantics, Vol. 3, Speech Acts*. Eds Peter Cole and Jerry L. Morgan. New York: Academic Press, 1975. 41–58.
Grim, Patrick. 'Simulating Grice: Emergent Pragmatics in Spatialized Game Theory.' In *Language, Games, and Evolution*. Eds. Anton Benz. Gerhard, Jäger and Robert van Rooij. Berlin, Heidelberg: Springer, 2011. 134–59.
Harsanyi, John C., and Reinhard Selten. *A General Theory of Equilibrium Selection*. Cambridge, MA: MIT Press, 1988.
Heap, Shaun P. Hargreaves, and Yanis Varoufakis. *Game Theory – A Critical Text*. 2nd ed. London: Routledge, 2004.
Helbing, Dirk. 'A Stochastic Behavioral Model and a 'Microscopic' Foundation of Evolutionary Game Theory.' *Theory and Decision* 40.2 (1996): 149–79.
Hofbauer, Josef, and Karl Sigmund. 'Evolutionary Game Dynamics.' *Bulletin of the American Mathematical Society* 40.4 (2003): 479–519.
Horn, Laurence R. Towards a New Taxonomy for Pragmatic Inference: Q-based and R-based Implicature.' In *Meaning, Form, and Use in Context*. Ed. Deborah Shiffrin. Washington: Georgetown University Press, 1984. 11–42.
Huttegger, Simon M. 'Evolution and the Explanation of Meaning.' *Philosophy of Science* 74 (2007): 1–27.
Huttegger, Simon M., et al. 'Evolutionary Dynamics of Lewis Signaling Games: Signaling Systems vs. Partial Pooling.' *Synthese* 172.1 (2010): 177–91.
de Jaegher, Kris. 'The Evolution of Horn's Rule.' *Journal of Economic Methodology* 15.3 (2008): 275–84.
Jäger, Gerhard. 'Evolutionary Game Theory and Typology: A Case Study.' *Language* 83.1 (2007): 74–109.
Jäger, Gerhard. 'Applications of Game Theory in Linguistics.' *Language and Linguistics Compass* 2.3 (2008a): 406–21.
Jäger, Gerhard. 'Evolutionary Stability Conditions for Signaling Games with Costly Signals.' *Journal of Theoretical Biology* 253.1 (2008b): 131–41.
Jäger, Gerhard. 'Game-Theoretical Pragmatics.' In *Handbook of Logic and Language*. Eds Johan van Benthem and Alice ter Meulen. Amsterdam: Elsevier, 2011. 467–91.
Jäger, Gerhard. 'Game Theory in Semantics and Pragmatics.' to appear In *Handbook of Semantics*. Eds C. Maienborn, P. Portner and K. von Heusinger, forthcoming.
Jäger, Gerhard, and Christian Ebert. 'Pragmatic Rationalizability.' In *Proceedings of Sinn und Bedeutung 13*. Eds Arndt Riester and Torgrim Solstad. 2009. 1–15.

de Jager, Tikitu, and Robert van Rooij, 'Explaining Quantity Implicatures.' In *Proceedings of the 11th conference on Theoretical Aspects of Rationality and Knowledge*. Ed. Dov Samet New York: ACM, 2007. 193–202.

Lenaerts, Tom, et al. 'The Evolutionary Language Game: An Orthogonal Approach.' *Journal of Theoretical Biology* 235 (2005): 566–82.

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

Levinson, Stephen C. *Presumptive Meanings. The Theory of Generalized Conversational Implicature*. Cambridge, MA: MIT Press, 2000.

Lewis, David. *Convention. A Philosophical Study*. Cambridge, MA: Harvard University Press, 1969.

Loreto, Vittorio, et al. 'Mathematical Modeling of Language Games.' In *Evolution of Communication and Language in Embodied Agents*. Eds Stefano Nolfi and Marco Mirolli. Berlin, Heidelberg: Springer-Verlag, Chap. 15, 2010. 263–81.

Maynard Smith, John. *Evolution and the Theory of Games*. Cambridge, UK: Cambridge University Press, 1982.

Merin, Arthur. 'Information, Relevance, and Social Decisionmaking: Some Principles and Results of Decision-Theoretic Semantics.' In *Logic, Language and 16 Computation*, Vol. 2. Eds Lawrence C. Moss, Jonathan Ginzburg and Maarten de Rijke. Stanford: CSLI Publications, 1999. 179–221.

Mühlenbernd, Roland. 'Learning with Neighbors: Emergence of Convention in a Society of Learning Agents.' *Synthese*. Online first. 183 (2011): 87–109.

Mühlenbernd, Roland, and Michael Franke. 'Signaling Conventions: Who Learns What Where and When in a Social Network.' *The Evolution of Language: Proceedings of 9th International Conference (EvoLang 9)*. Eds Thomas C. Scott- Phillips, Monica Tamariz, Erica A. Cartmill and James R. Hurford. Singapore: World Scientific, 2012. 242–9.

Myerson, Roger B. 'Credible Negotiation Statements and Coherent Plans.' *Journal of Economic Theory* 48.1 (1989): 264–303.

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

Nowak, Martin A. *Evolutionary Dynamics: Exploring the Equations of Life*. Cambridge, MA: Harvard University Press, 2006a.

Nowak, Martin A. 'Five Rules for the Evolution of Cooperation.' *Science* 314 (2006b): 1560–3.

Nowak, Martin A., and David C. Krakauer. 'The Evolution of Language.' *PNAS* 96 (1999): 8028–33.

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

Osborne, Martin J., and Ariel Rubinstein. *A Course in Game Theory*. Cambridge, MA: MIT Press, 1994.

Page, Karen M., and Martin A. Nowak. 'Unifying Evolutionary Dynamics.' *Journal of Theoretical Biology* 219 (2002): 93–8.

Parikh, Prashant. 'Communication and Strategic Inference.' *Linguistics and Philosophy* 14 (1991): 473–514.

Parikh, Prashant. 'A Game-Theoretic Account of Implicature.' In *TARK' 92: Proceedings of the 4th conference on Theoretical aspects of reasoning about knowledge*. Ed. Yoram Moses. San Francisco, CA: Morgan Kaufmann Publishers Inc., 1992. 85–94.

Parikh, Prashant. 'Communication, Meaning, and Interpretation.' *Linguistics and Philosophy* 23.2 (2000): 185–212.

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

Parikh, Prashant. 'Pragmatics and Games of Partial Information.' In *Game Theory and Pragmatics*. Eds Anton Benz. Gerhard, Jäger and Robert van Rooij. Hampshire: Palgrave MacMillan, 2006. 101–22.

Parikh, Prashant. *Language and Equilibrium*. Cambridge, MA: MIT Press, 2010.

Perea, Andres. *Epistemic Game Theory: Reasoning and Choice*. Cambridge, UK: Cambridge University Press, 2012.

Potts, Christopher. *Interpretive Economy, Schelling Points, and Evolutionary Stability*. Manuscript, UMass: Amherst, 2008.

Quine, Willard Van Orman. 'Truth by Convention.' In *Philosophical Essays for Alfred North Whitehead*. London: Longmans, Green and Co. 1936. 90–124.

Rabin, Matthew. 'Communication between Rational Agents.' *Journal of Economic Theory* 51 (1990): 144–70.

Rabin, Matthew. 'A Model of Pre-Game Communication.' *Journal of Economic Theory* 63.2 (1994): 370–91.

Rescorla, Michael. 'Convention.' In *The Stanford Encyclopedia of Philosophy*. Ed. Edward N. Zalta Spring, 2011.

Roca, Carlos P., et al. 'Evolutionary Game Theory: Temporal and Spatial Effects Beyond Replicatory Dynamics.' *Physics of Life Reviews* 6 (2009): 208–9.

van Rooij, Robert. 'Cooperative versus Argumentative Communication.' *Philosophia Scientiae* 8.2 (2004a): 195–209.

van Rooij, Robert. 'Evolution of Conventional Meaning and Conversational Principles.' *Synthese* 139.2 (2004b): 331–66.

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

van Rooij, Robert. *Optimality-Theoretic and Game-Theoretic Approaches to Implicatures*. Stanford Encyclopedia of Philosophy, 2006.

van Rooij, Robert. 'Games and Quantity Implicatures.' *Journal of Economic Methodology* 15.3 (2008): 261–74.

van Rooy, Robert. 'Quality and Quantity of Information Exchange.' *Journal of Logic, Language and Computation* 12 (2003): 423–51.

Rubinstein, Ariel. *Economics and Language*. Cambridge, UK: Cambridge University Press, 2000.

Russell, Bertrand. *The Analysis of Mind*. London: Unwin Brothers Ltd, 1921.
Sally, David. 'Risky Speech: Behavioral Game Theory and Pragmatics.' *Journal of Pragmatics* 35 (2003): 1223–45.
Schelling, Thomas C. *The Strategy of Conflict*. Cambridge, MA: Harvard University Press, 1960.
Schlag, Karl H. 'Why Imitate, and If So, How?' *Journal of Economic Theory* 78.1 (1998): 130–56.
Scott-Phillips, Thomas C. 'The Evolution of Relevance.' *Cognitive Science* 34 (2010): 585–601.
Selten, Reinhard. 'A Note On Evolutionarily Stable Strategies in Asymmetric Animal Conflicts.' *Journal of Theoretical Biology* 84 (1980): 93–101.
Skyrms, Brian. *Evolution of the Social Contract*. Cambridge, UK: Cambridge University Press, 1996.
Skyrms, Brian. *Signals*. Oxford: Oxford University Press, 2010.
Stalnaker, Robert. 'On the Evaluation of Solution Concepts.' *Theory and Decision* 37.1 (1994): 49–73.
Stalnaker, Robert. 'Saying and Meaning, Cheap Talk and Credibility.' In *Game Theory and Pragmatics*. Eds Anton Benz. Gerhard, Jäger and Robert van Rooij. Hampshire: Palgrave MacMillan, 2006. 83–100.
Steels, Luc. 'A Self-Organizing Spatial Vocabulary.' *Artificial Life* 2.3 (1995): 319–32.
Steels, Luc. 'The Synthetic Modeling of Language Origins.' *Evolution of Communication* 1.1 (1997): 1–35.
Sugden, Robert. *The Economics of Rights, Cooperation and Welfare*. Oxford: Blackwell, 1986.
Szabó, György, and Gábor Fáth. 'Evolutionary Games on Graphs.' *Physics Reports* 446 (2007): 97–216.
Taylor, Peter D., and Leo B. Jonker. 'Evolutionary Stable Strategies and Game Dynamics.' *Mathematical Bioscience* 40.1–2 (1978): 145–56.
Vanderschraaf, Peter. 'Convention as Correlated Equilibrium.' *Erkenntnis* 42 (1995): 65–87.
Wagner, Elliott. 'Communication and Structured Correlation.' *Erkenntnis* 71.3 (2009): 377–93.
Watts, Duncan J., and Steven H. Strogatz. 'Collective Dynamics of 'Small-World' Networks.' *Nature* 393.4 (1998): 440–2.
Weibull, Jörgen W. *Evolutionary Game Theory*. Cambridge, MA: MIT Press, 1997.
Wilson, Deirde, and Tomoko Matsui. 'Recent Approaches to Bridging: Truth, Coherence and Relevance.' *UCL Working Papers in Linguistics* 10 (1998): 173–200.
Zollman, Kevin J. S. 'Talking to Neighbors: The Evolution of Regional Meaning.' *Philosophy of Science* 72 (2005): 69–85.