ANNUAL REVIEWS

*Annual Review of Linguistics*

# Game-Theoretic Approaches to Pragmatics

Anton Benz[1] and Jon Stevens[2]

[1]Leibniz-Centre General Linguistics (ZAS), Berlin 10117, Germany; email: benz@leibniz-zas.de

[2]Department of Linguistics, The Ohio State University, Columbus, Ohio 43210; email: stevens.400@osu.edu

Annu. Rev. Linguist. 2018. 4:173–91

First published as a Review in Advance on September 15, 2017

The *Annual Review of Linguistics* is online at linguist.annualreviews.org

https://doi.org/10.1146/annurev-linguistics-011817-045641



## Keywords

pragmatics, game theory, implicature, Grice, rational communication, Bayesian modeling

## Abstract

We present an overview and comparison of different game-theoretic approaches to Gricean pragmatics, including games of partial information, optimal answer models, error models, iterated best response models, and rational speech act models. We address phenomena of disambiguation, scalar implicature, and relevance implicature.

## 1. INTRODUCTION

Linguistic pragmatics studies modulations in meaning that result from interactions between speaker, hearer, and discourse context. The context dependence of meaning makes it difficult to precisely spell out these modulations. Pragmatics is, therefore, widely considered the fuzziest subbranch of linguistics. Game theory, in general, is a mathematical framework developed for studying decision making involving several agents. In applications to pragmatics, the agents are often, although not always, speaker and hearer. The speaker makes a decision about what linguistic form to produce, and the hearer about what interpretation to assign to the speaker's utterance. If they are cooperative, they have a common goal: Both want the hearer to arrive at the interpretation that the speaker had in mind. Game theory provides a precise framework for representing such problems, for thinking about them, and for finding solutions to them. In this review, we illustrate how game theory is applied to pragmatic problems, using examples of scope disambiguation, relevance implicature, and scalar implicature.

Game-theoretic pragmatics emerged as a research field after the turn of the millennium. The most influential work of the period before 2000 is arguably that by Lewis (1969), who developed an extremely influential model of conventional meaning in a game-theoretic framework. Other notable works from this period include those by Zaefferer (1977) and Merin (1999) on decision theory, R. Parikh (1996) on vagueness, and P. Parikh (1990, 1991, 1992), who presented the first comprehensive game-theoretic pragmatic framework. With the turn of the millennium, there was a sudden increase in publications from various fields, among them works by P. Parikh (2000, 2001), Rubinstein (2000), Dekker & van Rooij (2000), Asher et al. (2002), and van Rooij (2004b, circulating from 2001 on). In addition to Gricean pragmatics and disambiguation phenomena, leading problems were partial blocking, question–answer relations, and the evolution of meaning and grammatical regularities (Benz et al. 2006b).

In this article, we concentrate on game-theoretic approaches to disambiguation and conversational implicature. Many earlier approaches tried to explicate Grice's rather informal notion of relevance with the help of decision-theoretic relevance measures (Merin 1999, van Rooij 2004b, Schulz & van Rooij 2006). The idea was to explain implicature by generalizing the neo-Gricean account of scalar quantity implicature: If the speaker tries to maximize the relevance of his contributions, the hearer is entitled to infer that everything that would have been more relevant but was left unsaid is false according to the speaker.

In one of Grice's famous examples (Grice 1975), A and B are planning their summer holidays in France. They would like to visit C, an old friend of B. So A asks B, *Where does C live?* B answers, *Somewhere in the south of France.* Here, one can reason as follows. A more specific answer mentioning the city where C lives would have been more relevant; therefore, because B did not mention it, one can conclude that B does not know where C lives. However, the problems of relevance scale approaches can already be seen with another of Grice's famous examples, the out-of-petrol example. Assume that A says to B, *I am out of petrol*, and B answers, *There is a garage around the corner.* In this situation, according to standard relevance measures, an answer also mentioning that the garage is open would have been more relevant as it increases the expected success of going there. According to the logic of relevance maximization, the fact that the speaker did not mention this should entitle us to infer that the garage is closed, which is, of course, not the case. It can be very generally shown that such unintended consequences cannot be avoided in relevance scale approaches, because those approaches are not interactional—that is, they do not represent multiple agents' beliefs about one another (see Benz 2006, 2007 for a discussion).

This article serves as a guide to what we consider the most important approaches to Gricean pragmatics. For a more thorough introduction to game and decision theory for linguistics, we

refer the reader to Benz et al. (2006b). For game theory in general, there are a wide variety of textbooks available, including those by Myerson (1991), Fudenberg & Tirole (1991), Osborne & Rubinstein (1994), and Dixit et al. (2009). For an overview of topics in game-theoretic pragmatics, we recommend the collections edited by Benz et al. (2006a, 2011) and Pietarinen (2007). Recent survey articles on game-theoretic pragmatics include those by Jäger (2011), Franke & Jäger (2014), van Rooij & Franke (2015), and Franke (2017).

## 2. GAMES OF PARTIAL INFORMATION

One of the first, and perhaps one of the simplest, applications of game theory to pragmatics is Parikh's (2001) model of scope disambiguation (see also Clark 2012, Parikh 2010). We begin with scope disambiguation because it is a simple and clear case of what is perhaps the core problem of pragmatics: how to select the most likely intended interpretation of an utterance from among a well-defined set of candidate interpretations. In that sense, all pragmatic problems can be regarded as disambiguation problems. Parikh named the games with which he represented these problems games of partial information (GPIs).

### 2.1. Pragmatics as Disambiguation

Imagine hearing the following statistic (from Parikh 2001):

(1) Every 10 minutes, a man gets mugged in New York.

Semantically, this utterance is ambiguous, because the quantifiers *every* and *a* can differ in terms of their scope. If *every 10 minutes* scopes over *a man*, the interpretation is, 'For every 10 minute interval $j$, there is some man $x$ such that $x$ gets mugged during $j$ in New York.' If *a man* scopes over *every 10 minutes*, the interpretation is, 'There is some man $x$ such that for every 10 minute interval $j$, $x$ gets mugged during $j$ in New York.' (An unlucky fellow, indeed.) How the ambiguity is resolved depends on context—a priori, it seems almost certain that *every 10 minutes* takes higher scope, but it is possible for a context to favor the other reading, as in: "I just watched a movie about people who lead extremely unlucky lives; in it, a woman from LA crashes her car 10 times in one week, and...", followed by utterance 1. We can model the resolution of this ambiguity by modeling the interaction between speaker and hearer as a game. Typically, the games considered in pragmatics are variants and extensions of so-called signaling games (Lewis 1969). GPIs are just one such variant. The game has two players, whom we call speaker $S$ and hearer $H$. The game starts with a weighted coin flip that assigns the speaker a type $t$ with probability $P(t)$. In game theory, a type represents a player's private information; it corresponds to what linguists and philosophers call a person's knowledge or information state. It is assumed that the probability $P(t)$ with which $t$ is chosen is shared knowledge between $S$ and $H$. The game continues with $S$, who knows her type $t$, choosing a message $m$. Finally, $H$, who knows $m$ but not $t$, chooses an interpretation (i.e., a guess as to $S$'s type) $i$. With this move, the game ends. The outcome of the game can be identified with the sequence $(t, m, i)$. $S$ and $H$ have preferences over outcomes that can be represented by numerical values. For example, if $S$ and $H$ prefer successful over unsuccessful communication, their preference can be represented by a utility function $U$, for which $U(t, m, i) = 0$ if $t \neq i$ and $U(t, m, i) = 1$ if $t = i$. Sometimes it is useful to distinguish $S$'s and $H$'s preferences. In this case, one needs to consider two utility functions, one for the speaker $U_S$ and one for the hearer $U_H$. We assume, unless stated otherwise, that $S$ and $H$ share a semantics and a lexicon, and that $S$ can send only truthful messages. We also assume that this game structure is commonly known to both $S$ and $H$.

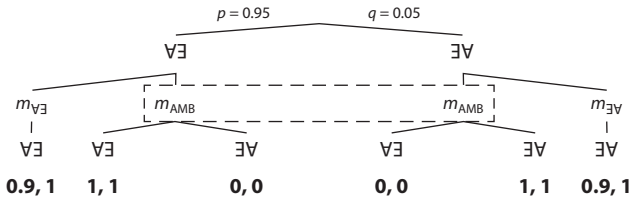


**Figure 1**

Parikh's disambiguation game. The left branch of the tree represents a type ∀∃ speaker, who could send an unambiguous message with cost 0.1, or an ambiguous message at no cost. The right branch represents a type ∃∀ speaker, who can also unambiguously signal her type at a cost, or else send the same ambiguous message as type ∀∃. The ambiguity of the ambiguous message is represented by a dashed box around the two states that are possible given that message.

Let us now build a game representation for utterance 1. The possible types in the game are the two different scope meanings, which we label ∀∃ (*every* scopes over *a*) and ∃∀ (*a* scopes over *every*). We can use the same labels for *H*'s interpretations—*H* guesses that *S* intends either ∀∃ or ∃∀. Consider the following possible speaker messages:

- $m_{\text{AMB}}$: ambiguous scope message, as observed in utterance 1.
- $m_{\forall\exists}$: more effortful message conveying ∀∃, e.g., 'Every 10 minutes, some man or other is mugged in New York.'
- $m_{\exists\forall}$: more effortful message conveying ∃∀, e.g., 'Every 10 minutes, a particular man is mugged in New York.'

This game is a coordination game (Bacharach 2006; see also Schelling 1960) in the sense that if one player does well, the other does well, too. In a pure coordination game, the utility functions for the players are identical. But in this signaling game, it may seem more natural to choose different utility functions for *S* and *H*, because we need to factor in the cost of, or effort to produce, *S*'s message. *H*'s utility is not affected by message costs, so we may assume that $U_H$ is equal to the utility function defined above representing an interlocutor who is interested only in communicative success:[1]

(2) $U_H(t, m, i) := 1$ if $t = i$; else 0.

As $m$ does not influence *H*'s utilities, we can shorten $U_H(t, m, i)$ to $U_H(t, i)$. For *S*, by contrast, $m$ directly affects utility because $m$ comes with a cost, $C(m)$, which is deducted from *S*'s utility:

(3) $U_S(t, m, i) := U_H(t, i) - C(m)$.

As mentioned above, it is assumed that *S*'s message cannot be false—that is, that $m_{\forall\exists}$ cannot be uttered by a speaker of type ∃∀, and that $m_{\exists\forall}$ cannot be uttered by a speaker of type ∀∃. Moreover, *S* and *H* also share knowledge of the likelihood of the two types.

We can represent this game in its entirety as a tree diagram, termed an extensive form game (**Figure 1**). The root node of the tree branches into two possibilities representing the two possible speaker types. The type is assigned via a weighted coin flip. Let us assume for the sake of concreteness that utterance 1 occurs in a context where the prior probability of type ∀∃ ($p$) is 0.95. That means that the probability of type ∃∀ ($q$, equal to $1 - p$) is 0.05. Let us also assume that the cost of sending one of the longer unambiguous messages is 0.1, whereas the cost for sending the

[1] For reasons we cannot address because of space constraints, whether or not costs are represented in *H*'s utilities has no effect on the analysis of the games.

**Table 1 Possible strategies in Parikh's (2001) disambiguation game**

| | $S_1$ | $S_2$ | $S_3$ | $S_4$ | | $H_1$ | $H_2$ |
|---|---|---|---|---|---|---|---|
| ∀∃ | $m_{\forall\exists}$ | $m_{\forall\exists}$ | $m_{\text{AMB}}$ | $m_{\text{AMB}}$ | $m_{\text{AMB}}$ | ∀∃ | ∃∀ |
| ∃∀ | $m_{\exists\forall}$ | $m_{\text{AMB}}$ | $m_{\exists\forall}$ | $m_{\text{AMB}}$ | $m_{\forall\exists}$ | ∀∃ | ∀∃ |
| | | | | | $m_{\exists\forall}$ | ∃∀ | ∃∀ |

shorter ambiguous message is 0. Once $S$'s type is assigned by the weighted coin flip, $S$ chooses a message. If $S$ unambiguously signals her type via $m_{\forall\exists}$ or $m_{\exists\forall}$, then $H$ has no choice but to guess correctly. In that event, $H$ receives utility 1 (the maximum), but $S$ receives only 0.9 due to the cost of the message. If $S$ sends $m_{\text{AMB}}$, by contrast, $H$ does not know whether she is in a state where a type ∀∃ speaker sent $m_{\text{AMB}}$ or where a type ∃∀ speaker sent $m_{\text{AMB}}$. If $H$ guesses correctly, both players receive the maximum utility of 1. But if $H$ guesses incorrectly, both players get utility 0, the least desired outcome.

What can be predicted about how players behave or should behave in this game? One of the core assumptions of classic game theory is that players are rational, utility-maximizing agents. In situations with probabilistic uncertainty about the state of the world, this means that they will maximize the expected utility of their actions. Expected utility (von Neumann & Morgenstern 1944) is the weighted average utility for all possible outcomes given what the player has observed so far. In decision theory, one considers the case in which the outcome of an action depends only on the unknown state of the world. Thus, the expected utility $EU(a)$ of action $a$ is defined as $\sum_w P(w)\, U(w, a)$. In game theory, the situation is compounded by the fact that the outcome depends not only on one's own actions but also on the actions of others. Whether a speaker and a hearer have success depends on both the speaker's strategy of choosing messages and the hearer's interpretation strategy. (A player's strategy is a complete specification of what that player should do in any state of gameplay.) This interdependence means that predictions about behavior have to be predictions about speaker behavior and hearer behavior simultaneously. Much of game theory is concerned with the characterization of strategy pairs on which rational agents can converge. The most foundational concept here is the Nash equilibrium (Nash 1950), which is a set of strategies in which no individual player could gain any utility by unilaterally deviating from his or her strategy. Here, we introduce a refinement of Nash equilibria that is directly relevant to signaling games.[2]

A Nash equilibrium is a set of strategies, one for each player, in which each player's strategy maximizes that player's expected utility, given the strategies of the other players. Let us approach this concept step by step. First, what are the strategies in Parikh's disambiguation game? A strategy tells us for each possible information state of a player how he or she reacts to it. The speaker's information state is his type $t$, and the hearer's is the message $m$ she received from the speaker. In general, strategies can be probabilistic, but here we consider only the case of pure strategies—that is, strategies that are functions from information states into actions. **Table 1** shows all pure speaker and hearer strategies for Parikh's disambiguation game.

Next, we calculate the expected utilities of strategies. As explained above, in game theory the success of a strategy depends on the strategies of others. Thus, a player's expected utility can be calculated only by assuming fixed strategies for the other players. Let us therefore consider each strategy pair $(S, H)$ with the assumption that the speaker follows strategy $S$ and the hearer follows strategy $H$. As the speaker knows his type and knows that the hearer will follow $H$, his

[2]The canonical refinement for signaling games is that of a perfect Bayesian equilibrium (Harsanyi 1968, Fudenberg & Tirole 1991).

**Table 2 Expected utilities of strategy pairs in Parikh's (2001) disambiguation game**

| $EU(S\|H)$ | $S_1$ | $S_2$ | $S_3$ | $S_4$ | $EU(H\|S)$ | $S_1$ | $S_2$ | $S_3$ | $S_4$ |
|---|---|---|---|---|---|---|---|---|---|
| $H_1$ | 0.9 | 0.855 | 0.995 | 0.95 | $H_1$ | 1 | 0.95 | 1 | 0.95 |
| $H_2$ | 0.9 | 0.905 | 0.045 | 0.05 | $H_2$ | 1 | 1 | 0.05 | 0.05 |

expected utility of message $m$ is simply the utility $U(t, m, H(m))$, where $H(m)$ denotes the hearer's interpretation given message $m$ under strategy $H$. $S(t)$ denotes the speaker's message given type $t$ under strategy $S$. Therefore, the expected utility of the speaker's strategy $S$ as a whole, given hearer strategy $H$, is

$$(4) \quad EU(S|H) = \sum_t P(t) \times U_S(t, S(t), H(S(t))).$$

Accordingly, the hearer's expected utility of $H$ given $S$ is

$$(5) \quad EU(H|S) = \sum_t P(t) \times U_H(t, H(S(t))).$$

Thus equipped, we can systematically calculate the expected utilities of all strategy pairs. For example, $EU(S_2|H_1) = P(\forall\exists) \times U_S(\forall\exists, m_{\forall\exists}, \forall\exists) + P(\exists\forall) \times U_S(\exists\forall, m_{\text{AMB}}, \forall\exists) = 0.95 \times 0.9 + 0.05 \times 0 = 0.855$. **Table 2** shows the expected utilities for Parikh's GPIs.

Overall, the strategy pair $(S_3, H_1)$ yields the highest expected utility for both speaker and hearer, and in particular, none has an interest in switching to another strategy while the other one plays his or her strategy. However, there exists another Nash equilibrium, the strategy pair $(S_2, H_2)$. If the speaker plays $S_2$, the hearer wants to play $H_2$, and if the hearer plays $H_2$, the speaker wants to play $S_2$. Thus, $(S_2, H_2)$ is a stable equilibrium. Only if the interlocutors can make sure that they switch to $(S_3, H_1)$ at the same time can they improve their situation. Parikh (1990, 2001) proposed the following criterion for solving GPIs: Interlocutors always converge at the equilibrium with the highest expected utilities.[3] This is $(S_3, H_1)$. In this equilibrium, the speaker produces the ambiguous $m_{\text{AMB}}$ for the more probable type $\forall\exists$, and the unambiguous but more complex $m_{\exists\forall}$ for the less frequent type $\exists\forall$. The hearer correctly interprets $m_{\text{AMB}}$ as $\forall\exists$. This is the intuitive solution to this disambiguation problem.

The idea that pragmatic phenomena can be viewed and explained as disambiguation problems has been explored by a number of researchers. The attraction of this approach is due in part to its conceptual simplicity. Among the phenomena that have been analyzed in this way are the choice of referring expressions (Clark & Parikh 2007, Mayol 2006, Mayol & Clark 2010), illocutionary force disambiguation (Parikh 2001), conversational implicature (Parikh 1992, 2001; Ross 2006), resolution of underspecification (Parikh 1990), partial blocking (Ross 2006), politeness and implicature (Clark 2012, chapter 8), and prototypes (Clark 2012, chapter 9). Horn's principle of division of pragmatic labor (Horn 1989) and Levinson's M-principle (Levinson 2000) can be justified in this framework. There were also early attempts to ground bidirectional Optimality Theory (Blutner 2000, 2004; Benz & Mattausch 2011) in game-theoretic principles (Dekker & van Rooij 2000, van Rooij 2004a, Ross 2006).

Despite their simplicity, GPIs have become less prominent in game-theoretic pragmatics, for several reasons. First, one is interested not only in the equilibria upon which interlocutors eventually agree but also in the reasoning process that leads there. Second, the models tend to overgenerate ambiguities, as every difference in the complexity of a message is predicted to

[3] This equilibrium is called the Pareto Nash equilibrium.

lead to meaning differentiation. Third, most of the research in this area was done before the rise of experimental pragmatics, and hence has not been applied to empirical data, which are increasingly important today. Finally, it has been debated whether GPIs capture the relevant features of dialogue situations and, in particular, whether pragmatic problems should really be considered disambiguation problems (Benz 2012a).

We discuss these reasons in more detail along with the examples in Sections 3 and 4, below. Next, we consider a disambiguation game for scalar implicature. This example also provides us with an opportunity to introduce a solution concept based on iterative reasoning of interlocutors about one another.

## 2.2. Disambiguation and Scalar Implicature

We consider a simple scalar implicature:

(6) Some of the students passed the exam.

The classic account of these examples is that *some* ($m_{\exists}$) is compatible with a reading in which all of the students passed the exam but that in most contexts an implicature arises that not all of them passed, because if all of the students had passed, presumably the speaker would have used the more specific form *all* ($m_{\forall}$). In the case that the speaker wants to communicate that some but not all passed, she can choose between the ambiguous *some* ($m_{\exists}$) and a literal description *some but not all* ($m_{\exists\neg\forall}$). **Figure 2** shows the GPI for this situation. It is structurally identical to the scope ambiguity example except that the unambiguous alternative for communicating $\forall$ is no more complex than the ambiguous message. The same calculation as for **Figure 1**, above, shows that the strategy pair for which the speaker chooses $m_{\forall}$ for type $\forall$ and $m_{\exists}$ for $\exists\neg\forall$, and for which the hearer interprets $m_{\exists}$ as $\exists\neg\forall$, is the one with highest expected utility independently of the probabilities $p$ and $q$, provided that they are both greater than zero.

We now consider a solution to this game that leads to the same equilibrium but begins from a situation in which each interlocutor is uncertain about the strategy of the other. Suppose the speaker is in a situation in which she wants to communicate $\forall$ (**Figure 2**). The speaker may reason as follows: If the hearer interprets $m_{\exists}$ as $\forall$, then both $m_{\exists}$ and $m_{\forall}$ lead to success and the same utility, but if the hearer interprets $m_{\exists}$ as $\exists\neg\forall$, then he had better choose $m_{\forall}$. Therefore, the choice of $m_{\forall}$ is never worse than $m_{\exists}$, and sometimes better. In this case, $m_{\forall}$ is said to weakly dominate $m_{\exists}$. The principle of elimination of weakly dominated strategies states that both players can figure out that the weakly dominated alternative can be ruled out as a possible choice. This transforms the game shown in

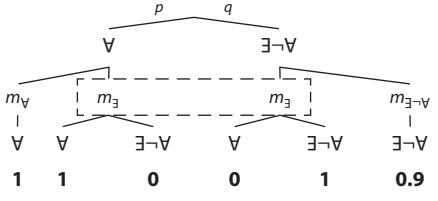


**Figure 2**

The disambiguation game for simple scalar implicatures. In the first node on the left branch, the speaker is in a state where she wants to communicate $\forall$, and does so unambiguously. In the second node of that branch, the speaker wants to communicate $\forall$, but uses a message that is ambiguous between $\forall$ and $\exists\neg\forall$. In the first node of the right branch, the speaker wants to communicate $\exists\neg\forall$, but uses the ambiguous message. In the second node of the right branch, the speaker unambiguously signals type $\exists\neg\forall$, but at a cost of 0.1.

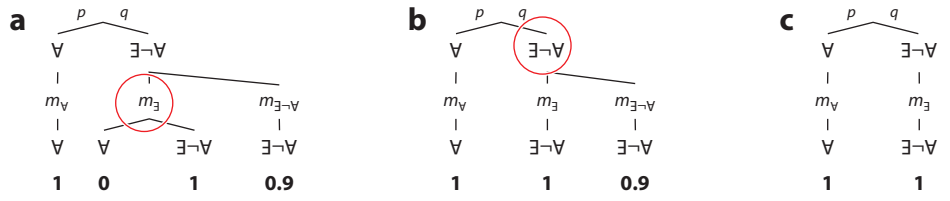


**Figure 3**

Iterated elimination of weakly dominated alternatives. (*a*) The red circle represents a situation in which the speaker produces $m_∃$ for type ∃¬∀. (*b*) The red circle represents a situation in which the speaker wants to communicate ∃¬∀. (*c*) The solution to this game is as follows: The speaker chooses $m_∀$ for type ∀ and $m_∃$ for type ∃¬∀.

**Figure 2** into the game shown in **Figure 3*a***. The understanding of this transformation is that the transformation leaves the solutions unchanged—that is, both games have the same solutions.

Let us turn to the hearer. How should she interpret $m_∃$? She may reason as follows: If the speaker produces the unambiguous $m_{∃¬∀}$ for type ∃¬∀, then interpreting $m_∃$ as ∃¬∀ cannot hurt, because it will not be produced anyway. If, however, the speaker produces $m_∃$ for type ∃¬∀ (**Figure 3*a***), interpreting $m_∃$ as ∃¬∀ is clearly better than interpreting it as ∀. Therefore, interpretation ∃¬∀ weakly dominates interpretation ∀ such that the latter interpretation can be eliminated from the game. This transforms the game shown in **Figure 3*a*** into that shown in **Figure 3*b***. Finally, we return to the speaker and consider the situation in which he wants to communicate ∃¬∀ (**Figure 3*b***). Here, the choice of $m_∃$ strongly dominates the choice of $m_{∃¬∀}$; in other words, for all remaining hearer strategies, $m_∃$ is preferred over $m_{∃¬∀}$. After the elimination of $m_{∃¬∀}$, the game shown in **Figure 3*c*** has a trivial solution: The speaker chooses $m_∀$ for type ∀ and $m_∃$ for type ∃¬∀, and the hearer interprets accordingly. This solution, based on iterated elimination of weakly dominated strategies, was proposed by Pavan (2013) and Rothschild (2013).

The final equilibrium is the same in both analyses. What is gained by the account based on iterated elimination of weakly dominated strategies? First, it leads stepwise from a situation in which speaker and hearer have complete uncertainty about one another's strategy to the final solution, thereby explaining how the equilibrium is adopted by speaker and hearer. Iterated elimination of weakly dominated strategies imposes strong requirements on the reasoning capabilities of the interlocutors involved. An interesting question, therefore, concerns how much reasoning is necessary for establishing certain equilibria. A drawback of this approach is that the criterion of weak dominance has only limited applications.

## 3. OPTIMAL ANSWER AND ERROR MODELS

In this section, we further discuss the reasoning processes leading to optimal behavior. We introduce the optimal answer (OA) model (Benz 2006, 2011; Benz & van Rooij 2007) and error models (Benz 2009; 2012a,b).

### 3.1. Optimal Answer Models

OA models account for the causal role of propositional content of utterances in determining implicatures while minimizing the interlocutors' reasoning about one another. We begin with a discussion of the famous out-of-petrol example (Grice 1975):

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

Taken as a dialogue between two participants, A and B, it is clear that B's utterance conversationally implicates that the garage is, to the best of B's knowledge, open and selling petrol. The reason is that, were the implicature known to be false, B's utterance would no longer be relevant to A's problem, namely finding petrol for her car.

Because we are analyzing B's utterance, B is the speaker ($S$) and A is the hearer ($H$) in this example. In contrast to previous examples, the primary task of the hearer is to make a decision not about the truth-conditional interpretation of the utterance but rather about an action choice. Let us simplify the situation and assume that A can begin her search for petrol at any location $l$ in town and, if there is no petrol, continue to the next location. A possible location is the place around the corner, $l_c$. If we assume $S$'s utterance to be literally true, then there are two salient possible world states: (*a*) The garage around the corner is open and selling petrol, and thus useful to $H$ (denoted by $l_c^+$), or (*b*) the garage is not open or not selling petrol, and thus useless to $H$ (denoted by $l_c^-$). Let $\varepsilon$ be the expected utility of a random search starting at a place different from $l_c$: $0 < \varepsilon < 1$. Then, we assume that $H$'s utility $U_H$ is as follows:

- 1 if there is petrol at $l_c$ and $H$ does go there,
- $\varepsilon$ if there is petrol at $l_c$ and she does not go there,
- 0 if there is no petrol at $l_c$ and $H$ does go there, or
- $\varepsilon$ if there is no petrol at $l_c$ and $H$ does not go there.

If we assume a fully cooperative speaker, $U_S$ can be calculated in the same way as above, where a cost of $c$ is deducted for longer messages. Clearly, the additional costs of uttering *There is a garage around the corner + that is open and sells petrol* are almost negligible in comparison to starting the random search at the wrong place, or going to a closed garage. Thus, we assume that $c$ is nominal—that is, positive but very small. We again simplify and consider only the following messages: $m_{l_c}$ ('There is a garage around the corner'), $m_{l_c^+}$ ('There is a garage around the corner that is open and sells petrol'), and $m_{l_c^-}$ ('There is a garage around the corner, but it is closed right now'). Putting this together gives us the extensive form game shown in **Figure 4**.

The solution to this game depends on the value of $p$—but something about that doesn't seem quite right. For this example, we have the sense that the propositional content of the utterance $m_{l_c}$ ('There is a garage around the corner') would cause the addressee to go to the garage in virtue of its propositional content, independently of $p$. An important question arises: How does the hearer utilize propositional content to make decisions? We need to encode the fact that the hearer's knowing there is a garage around the corner makes that location a better bet for her than any other location in town, given that she is unfamiliar with the area. This is true (under reasonable assumptions) regardless of how likely the garage is to be open. In other words, if the hearer is told by the speaker that there is a garage around the corner, then the hearer can reason as follows: The speaker gave me information that induces me to look around the corner in search for petrol; the speaker can figure out that I will do so, and she is cooperative and competent; thus, her only reason for telling me $m_{l_c}$ is that, to the best of her knowledge, petrol is available

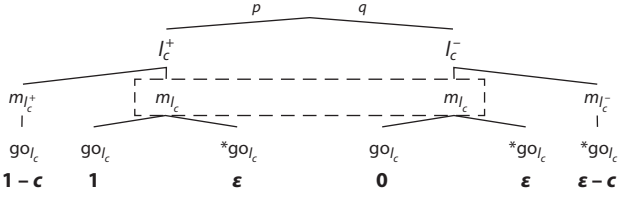


**Figure 4**

Partial game tree of the out-of-petrol example.

now. This reasoning translates into a form of backward induction:[4] For each message, calculate how the hearer will decide on the basis of the pure propositional content; then, for each possible world $w$, select all those messages that induce the hearer to choose an action that is optimal in $w$. This process provides all admissible messages. Thus, if $B(m)$ is the set of all actions for which $EU(a|[\![m]\!])$ is maximal, and $B(w)$ is the set of actions with maximal utility in $w$, then the admissible messages in world $w$ are the set $\mathrm{Adm}(w) = \{m|B(m) \subseteq B(w)\}$. Thus, the hearer can infer from an utterance of message $m$ that $m$ is admissible, and therefore that the actual world is an element of $\{w \in [\![m]\!]|B(m) \subseteq B(w)\}$. This is the optimal answer (OA) model of implicature (Benz & van Rooij 2007, Benz 2011).[5] In the out-of-petrol example, the set of actions with maximal expected utility $B(m_{l_c})$ has only one element, $\mathrm{go}_{l_c}$. The action $\mathrm{go}_{l_c}$ is optimal in a world $w$, in other words, $B(m_{l_c}) = \{\mathrm{go}_{l_c}\} \subseteq B(w)$, if and only if in $w$ petrol is available around the corner.

There are some implicit assumptions here. First, we assume that every proposition can be expressed, and that differential costs do not play a role. This means that the OA model is basically a model of content selection. Second, we assume that in every possible world there is an action that allows the addressee to achieve her goal, and that the speaker is an expert who knows an optimal action. For example, if the speaker knows only that there is a garage but not whether it is open, then clearly an utterance of $m_{l_c}$ no longer implies that the garage is open. If partial speaker knowledge is a possibility, then the game tree in **Figure 4** is no longer an appropriate representation of the dialogue situation. If in the actual world there is no place where petrol is available, then the OA model based on **Figure 4** predicts that all answers are equally good. If not all propositions can be expressed, then the speaker may be forced to choose an underinformative message because a more informative one is not available.

## 3.2. Error Models

In all of the models considered so far, communication proceeds in only one direction. The speaker produces a signal, and the hearer makes a choice based on the signal. Then the game ends. Ambiguities are resolved by choosing the interpretation with the highest expected utility. However, at least in face-to-face communication, there is another option concerning how to react to uncertainty about interpretation, namely asking for clarification. In this section, we explore the consequences of this possibility and introduce, in particular, error models (EMs) (Benz 2009; 2012a,b) that extend standard signaling games by allowing efficient clarification requests.

We first reconsider Parikh's scope disambiguation problem (utterance 1) with the analysis shown in **Figure 1**. The following example is structurally identical to Parikh's example. It contains an ambiguous sentence $m_X$ and two equally complex alternatives $m_A$ and $m_B$:

Doctor's appointment

Background: John is known to regularly consult two different doctors, physicians A and B. He consults A more often than B. Then *S* utters one of the following sentences.

(8a) John has a doctor's appointment at 4 pm. He requests you to pick him up afterwards. ($m_X$)

[4] In general, backward induction is an appropriate method for finding a solution if the game is a game of complete and perfect information; that is, all previous moves in the game must be known. This is not the case for signaling games. Therefore, the claim that backward induction is appropriate here is a nontrivial one.

[5] In the out-of-petrol example we can assume an implicit question: 'Where can A get petrol now?' Admissible messages can be considered optimal answers to this question. The model was first developed for content selection of answers (Benz 2006).

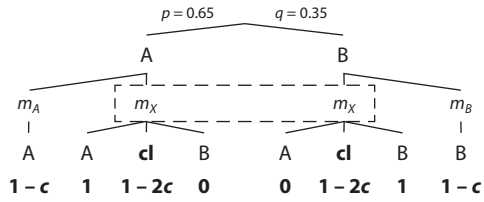


**Figure 5**

The error model for the doctor's appointment problem. $c$ represents the costs of uttering a more complex alternative. **cl** represents a clarification request. $2c$ represents the costs that are higher for a successful communication than for the unambiguous utterance. The dashed box represents the state of the hearer after an utterance of $m_X$.

(8b) John has a doctor's appointment at A's practice at 4 pm. He requests you to pick him up afterwards. ($m_A$)

(8c) John has a doctor's appointment at B's practice at 4 pm. He requests you to pick him up afterwards. ($m_B$)

As $m_A$ and $m_B$ are equally complex, and $m_A$ is more probable than $m_B$, the same analysis that predicts that the sentence *Every 10 minutes, a man gets mugged in New York* communicates that every 10 minutes one or the other man gets mugged (∀∃) also predicts that the sentence $m_X$ communicates $m_A$. This is obviously not the case. Instead, the natural reaction of the addressee is to assume that the speaker forgot to mention the location and to ask for clarification.

**Figure 5** shows a game tree representing this situation. The costs of uttering a more complex alternative are represented by $c$. The costs of uttering an additional clause necessary to make it unambiguous are certainly negligible compared with the costs of waiting in vain at the wrong practice. Thus, it is assumed that $c$ is nominal—that is, positive but very, very small. A clarification request, represented by **cl**, will incur additional costs, but can be assumed to lead to an answer that resolves the ambiguity of the utterance; thus, **cl** will lead to successful communication minus costs that are higher (represented by $2c$) than those for the unambiguous utterance.

Let us consider the hearer after an utterance of $m_X$. If the hearer is certain that the speaker produces $m_X$ in only one of the two branches—say, in situation A—then she can safely interpret $m_X$ as A and achieve utility 1. However, if there is even the slightest uncertainty, for example, due to speaker error, then the clarification request **cl** strictly dominates the immediate interpretation moves A and B. For the speaker, if uncertainty is a possibility, then the hearer will always respond to $m_X$ with **cl** with utility $1 - 2c$. Thus, the literal alternative $m_A$ with utility $1 - c$ strictly dominates $m_X$ with expected utility $1 - 2c$. The net result is that the speaker always uses the literal alternatives.

This example shows the dramatic effect of efficient clarification requests and noisy, or error-prone, speaker strategies. The use of ambiguous utterances seem to be pragmatically ruled out, which is contradicted by everyday experience. For Parikh's scope disambiguation problem (utterance 1), one may argue that the two interpretations *all–some* (∀∃) and *some–all* (∃∀) are such that ∃∀ has a probability of practically zero, so there is no practically relevant ambiguity. However, in the case of scalar implicature, for example, there is a real ambiguity that cannot be explained away by one alternative interpretation having practically zero probability. How can this conflict be solved? Key is the following bus ticket example:

Bus ticket

(9) An email was sent to all employees that bus tickets for a joint excursion have been bought and are ready to be picked up. By mistake, no contact person was named. Hence, $H$ asks one of the secretaries:

H: Where can I get the bus tickets for the excursion?

S: Ms. Müller is sitting in office 2.07. ($m(M, 2.07)$)

$\rightsquigarrow$ Bus tickets are available from Ms. Müller.

The secretary's answer does not semantically resolve the question of whether Ms. Müller has the tickets or not. In contrast to the out-of-petrol example, here the pure propositional content is not sufficient to induce the hearer to go to office 2.07 in search of tickets. For example, suppose the hearer finds a list of employees and reads there that $m(M, 2.07)$; contrast this scenario with one in which the hearer reads on a map that there is a garage around the corner. How, then, can this example be explained? Imagine that there is a second person who might have the ticket, Mr. Schmidt, and that either he or Ms. Müller may sit in 2.07 or in 3.11. We can now generate four possible utterances that would answer $H$'s question unambiguously, each of the form *X has the bus tickets; X is sitting in Y*, where $X$ is either Ms. Müller or Mr. Schmidt and $Y$ is either 2.07 or 3.11. The given answer $m(M, 2.07)$ is a substring of only one of these alternatives, namely of the one that says Ms. Müller has the tickets. Therefore, the hearer can unambiguously recover from answer $m(M, 2.07)$ that Ms. Müller has the bus tickets.

EMs show the value of considering the hearer's ability to consider more detailed messages in order to recover missing information. This also extends to scalar implicatures. A different way to look at the problem of scalar implicature is to begin, as in the bus ticket example, with a speaker who chooses a literal description and then simplifies it according to some rules. Such a speaker can produce $m_{\forall}$ for $\forall$ and $m_{\exists\neg\forall}$ for $\exists\neg\forall$. The semantically ambiguous utterance $m_{\exists}$ is a reduced form of $\exists\neg\forall$ and, hence, can occur only for $\exists\neg\forall$. Therefore, the game that has to be solved is not that shown in **Figure 2** but rather that shown in **Figure 3*a***. This game can be solved without iterated elimination of weakly dominated strategies—the structure of the game itself predicts that *some* should come to mean 'some but not all' in contexts where the interlocutors want to maximize informativity. Note that the costs incurred by complex utterances should, in the spirit of EMs, be considered to be nominal. Benz (2012a,b) and N. Gotzner & A. Benz (manuscript in revision) further developed this model to account for scalar implicature of complex sentences.

## 4. ITERATED BEST RESPONSE MODELS

OA models have the attractive property that they include a method for deriving a Nash equilibrium, in contrast to Parikh's GPIs, in which equilibria must be determined by checking whether each pair of strategies meets the equilibrium criteria. But OA models are limited in their application, in that they are restricted to models of content selection. That is, OA models are about choosing what information to convey, and not about how to convey that information. Franke (2009, 2011) developed an influential method for determining a Nash equilibrium in games where the speaker's choice is between alternative forms that can be used to present the same information. This is the iterated best response (IBR) model. Jäger & Ebert (2009) and Jäger (2011) also made important contributions; an earlier work (Jäger 2007) was an important precursor based on evolutionary best response dynamics. Scalar implicature is the canonical example, although the IBR method has been applied to other phenomena, including the selection of prosodic focus patterns (Stevens 2016). IBR models provide a simple, intuitive algorithm for determining equilibria in games for

which certain assumptions can be made. The simplest instantiation of IBR assumes the following: (*a*) The speaker and hearer have a common semantics, (*b*) the hearer assumes a priori that the speaker is being truthful, (*c*) the different speaker types are assumed to be a priori equiprobable, and (*d*) the hearer responds to surprise messages by selecting randomly from among types for which the heard message would be truthful.

IBR has its roots in the so-called level-*n* reasoning tradition,[6] which holds that rational agents reason iteratively about others' beliefs (e.g., 'I think that you think that I think that. . .'). The IBR algorithm proceeds by first defining a default strategy for one of the players—we begin with a default hearer strategy, which we denote $H_0$—and then iteratively reasoning about how the other player best responds to each possible action assuming the other player's strategy. In most cases, $H_0$ is a "literal" strategy—the hearer simply chooses interpretations randomly from among those that are semantically compatible with the speaker's message. The speaker can employ a strategy $S_1$ that maximizes the likelihood of communicative success given $H_0$. In turn, the hearer can employ a strategy $H_2$ that maximizes the likelihood of success given $S_1$, and so on, until the speaker and hearer converge on a stable pair of strategies $(S_n, H_{n+1})$. The strategies considered in IBR models are, in general, mixed strategies; in other words, the speaker's strategy $S(.\,|t)$ is a probability distribution over the set of all messages that are true given type $t$, and the hearer's strategy $H(\,.\,|m)$ is a probability distribution over the set of all interpretations that are consistent with message $m$. The IBR algorithm proceeds as follows.

- $H_0(m)$: Guess randomly from $[\![m]\!]$.
- $S_1(t)$: Select the least costly $m$ that maximizes $H_0(t|m)$; if there are several of them, choose any of them with equal probability.
- $H_2(m)$: Select the $t$ that maximizes $S_1(m|t)$; if there are several of them, choose any of them with equal probability.
- $S_3(t)$: Select the least costly $m$ that maximizes $H_2(t|m)$; if there are several of them, choose any of them with equal probability.
- . . .
- Stop if and only if iteration converges to a stable mapping between types and messages.

We now apply this process to example 6, reproduced below as example 10:

(10) Some of the students passed the exam.

Consider only the "cheap" messages $m_\exists$ and $m_\forall$. We can derive an equilibrium via IBR in the following way:

- $H_0$: $m_\exists \rightarrow \{\exists\neg\forall, \forall\}$; $m_\forall \rightarrow \{\forall\}$.
- $S_1$: $\exists\neg\forall \rightarrow \{m_\exists\}$; $\forall \rightarrow \{m_\forall\}$.
- $H_2$: $m_\exists \rightarrow \{\exists\neg\forall\}$; $m_\forall \rightarrow \{\forall\}$.

The literal hearer strategy assumes the meaning 'all' for the message $m_\forall$, and guesses either 'some but not all' or 'all' for the message $m_\exists$. Knowing this, $S_1$, if of type $\forall$, would do better to use the unambiguous $m_\forall$ and, if of type $\exists\neg\forall$, must use $m_\exists$, as $m_\forall$ is false for that type. This sets up an unambiguous mapping between messages at $H_2$, whereupon convergence occurs to a stable pair of strategies.

This is the simplest instantiation of IBR. A problem with it arises when we add the costlier unambiguous message $m_{\exists\neg\forall}$:

[6]The *p*-beauty contest game, first presented by Moulin (1986), served as an important inspiration for IBR models. For an overview, see Camerer (2003, section 5.2).

- $H_0$: $m_{\exists} \rightarrow \{\exists\neg\forall, \forall\}$; $m_{\forall} \rightarrow \{\forall\}$; $m_{\exists\neg\forall} \rightarrow \{\exists\neg\forall\}$.
- $S_1$: $\exists\neg\forall \rightarrow \{m_{\exists\neg\forall}\}$; $\forall \rightarrow \{m_{\forall}\}$.
- $H_2$: $m_{\exists} \rightarrow \{\exists\neg\forall, \forall\}$; $m_{\forall} \rightarrow \{\forall\}$; $m_{\exists\neg\forall} \rightarrow \{\exists\neg\forall\}$.

Clearly, the speaker would always do better to simply unambiguously signal her type with $m_{\exists\neg\forall}$. One way around this problem would be to introduce a large cost into the IBR method to exclude this option. This choice is pursued by several approaches that we consider in the next section. However, it is not realistic, because, as they say, "talk is cheap"—we have no evidence that slightly more effortful utterances present insurmountable costs to speakers of language. Franke (2009) restricted alternatives a priori to the standard Horn alternatives, such that $m_{\exists\neg\forall}$ was not a possible message. In order to allow the unambiguous alternative without positing high costs for extra syllables, one requires an approach akin to EMs, wherein 'some,' when it is off-equilibrium (i.e., unexpected), is taken to be a truncation of 'some but not all.'

IBR works beyond these simple cases, and makes interesting predictions about sentences with multiple quantifiers; for instance, example 11*a* should have the same meaning as in example 11*b*:

(11a) Some of the students passed some of the exams.

(11b) i. Some of the students passed some but not all of the exams.

ii. Some of the students passed none of the exams.

iii. None of the students passed all of the exams.

Franke (2009, 2011) worked this approach out in great detail, addressing a wide range of sentences with embedded scalar implicatures, including disjunctions, conditionals, and free-choice readings. This constituted the first game-theoretic model that covered as wide a range of phenomena as competing semantic approaches, such as those by Chierchia (2004), Sauerland (2004), and Fox (2007).

## 5. THE EMPIRICAL TURN

All of the models that we have encountered so far are designed to explain analytically how a pragmatic phenomenon could have arisen. They were developed independently of experimental data and computational applications. Although experimental pragmatics was developed concurrently with game-theoretic pragmatics (going back to about 2000), only after 2012 did the two research fields come closer together. A key advance in this area was the so-called rational speech act (RSA) model (Frank & Goodman 2012). This line of research originated in cognitive science and was developed independently of the game-theoretic approaches to pragmatics. The RSA model applies economic models of decision making with discrete choices and bounded rationality (Train 2003). As with the OA and IBR models, strategies are determined by backward calculation, beginning with a naïve hearer who takes only literal information into account and continuing with a speaker who chooses utterances such that the expected success of the naïve hearer is maximized. However, contrary to those models, speaker and hearer do not maximize expected utility in the pure sense. For one, the speaker uses a softmax function (see equation 12, below) to choose utterances proportionally to his or her expected utility. The hearer chooses an interpretation proportionally to the probability of that interpretation being correct given the speaker's softmax strategy. This means that, for both interlocutors, suboptimal choices are not ruled out; rather, they are only less frequent than the optimal ones. This accounts for the fact that people do not always choose the optimal action in real-life language tasks. The speaker is assumed to take the hearer's strategy into account, but as mentioned above, the speaker's strategy $P_S(m|t)$ is defined such that the expected utility of uttering $m$ when referring to $t$—here assumed to be $P(t|[\![m]\!]) - C(m)$, where

$C(m)$ is the message cost—is embedded in the so-called softmax function with a free parameter $\lambda$, a "rationality parameter" that modulates the variance of the speaker's strategy and, therefore, the frequency with which the speaker will deviate from an optimal communicative strategy. The listener's strategy $P_L(t|m)$ is calculated from $P_S(m|t)$ via Bayes's rule:

$$(12) \quad P_S(m|t) = \frac{e^{\lambda\,(P(t|[\![m]\!])-C(m))}}{\sum_{m'} e^{\lambda\,(P(t|[\![m']\!])-C(m'))}}, \quad P_L(t|m) = \frac{P_S(m|t)\;P(t)}{\sum_v P_S(m|v)\;P(v)}.$$

To use these equations to model experimental data, one must determine the prior probabilities of interpretations $P(t)$. The cost parameter and the "rationality parameter" $\lambda$ are usually fitted to the data post hoc. If $\lambda$ increases to infinity, then the strategy always chooses the utterance with maximal expected utility, as in classical game theory; if $\lambda$ is 0, then the speaker chooses randomly. Therefore, $\lambda$ represents the extent to which the speaker behaves rationally. Finally, the hearer may again reason about the speaker and choose according to her expectations about the speaker's strategy. Again, it is assumed that she does this proportional to expected utility—in other words, that her interpretation strategy $P_L(t|m)$ is proportional to the prior probability of $t$ multiplied by the speaker's probability of uttering $m$ in $t$ (see equation 12). One would expect that the speaker and hearer's reasoning about one another would continue, as in the IBR model, until an equilibrium is reached. However, RSA models are strongly motivated by the idea that human reasoning capabilities are limited. The speaker and hearer's reasoning about one another in RSA models, therefore, generally stops after the hearer reasons about the speaker—in IBR terms, after the sequence $H_0$–$S_1$–$H_2$.

An important application of these models is the reference game (Frank & Goodman 2012, Degen & Franke 2012, Qing & Franke 2015). In this game, a speaker and a hearer are given a small number of objects, such as colored squares and circles; the speaker then has to choose a word, and the hearer has to guess which object the speaker wants to refer to. For example, the speaker may say *green* or *circle*, and the hearer chooses among a green circle, a blue circle, and a green square. In such experiments, subjects do not behave uniformly. A certain percentage will choose *green*, another percentage *circle*. Also, hearers will choose referents with a certain probability, not categorically. All previous models—error, disambiguation, OA, and IBR—predict stable equilibria that leave no room for the suboptimal choices that can be observed in experiments. Because they generally predict pure strategies, they are not able to explain probabilistic behavior, namely behavior that cannot be represented by strategies with probabilities restricted to 0 and 1. Another feature that sets RSA models apart from all other models is the fact that they can be fitted to data. For example, in a reference game, $m$ and $t$ in equation 12 stand for words and referents. If the prior probability of referents $P(t)$ is given, the cost parameter and the rationality parameter $\lambda$ can be optimized. Beyond reference games, this approach has found significant applications to scalar implicatures (Goodman & Stuhlmüller 2013, Degen & Goodman 2014, Degen et al. 2015), grammatical properties of adjectival scales (Lassiter & Goodman 2014, Qing & Franke 2014), and implicatures in complex sentences (Bergen et al. 2016, Potts et al. 2016).

## 6. CONCLUSION

We have provided a survey of the most important game-theoretic approaches to Gricean pragmatics. The models differ from one another on several dimensions: Is there a reasoning process leading to equilibria? (This is the case for IBR, OA, and RSA.) If there is one, how long can it be? (Both OA and RSA introduce limits on the number of iterations.) Is observed behavior assumed to be in equilibrium at all? (RSA does not assume this.) Do they take feedback in the form of clarification requests into account? (Only EM does this.) Do they assume classic rationality? (RSA does not.) Can they be fitted to data? (RSA is particularly suited to this.)

What we have not addressed in this article is the question of the explanatory value of the different approaches. To have a free parameter in a model that can be fitted to data does not necessarily make the model better than models that only predict pure strategies. Currently, there is no established criterion that would enable an objective comparison of models. Our own tendency is to propose communicative success in controlled dialogue as such a criterion. This issue, however, must be left for future discussions.

### SUMMARY POINTS

1. Games of partial information (GPIs) analyze pragmatic phenomena as disambiguation problems.
2. Error models (EMs) account for clarification requests, in opposition to purely disambiguation-based analyses.
3. Iterated best response (IBR) models establish equilibria of behavior via processes of agents' iterated reasoning about one another.
4. Optimal answer (OA) and rational speech act (RSA) models assume that each interlocutor's sequence of reasoning about the other is short.
5. RSA models provide parameterized models that can be fitted to data.

## DISCLOSURE STATEMENT

The authors are not aware of any affiliations, memberships, funding, or financial holdings that might be perceived as affecting the objectivity of this review.

## ACKNOWLEDGMENTS

The writing of this review was supported by the Bundesministerium für Bildung und Forschung (grant 01UG1411), by the Deutsche Forschungsgemeinschaft (grants BE 4348/3-1 and BE 4348/3-2), and by the American Council of Learned Societies.

## LITERATURE CITED

Asher N, Sher I, Williams M. 2002. Game theoretic foundations for Gricean constraints. In *Proceedings of the 2001 Amsterdam Colloquium on Formal Semantics*, ed. R van Rooy, M Stokhof, pp. 31–37. Amsterdam: Inst. Log. Lang. Comput.

Bacharach M. 2006. *Beyond Individual Choice: Teams and Frames in Game Theory*. Princeton, NJ: Princeton Univ. Press

Benz A. 2006. Utility and relevance of answers. See Benz et al. 2006a, pp. 195–214

Benz A. 2007. On relevance scale approaches. In *Proceedings of Sinn und Bedeutung 11*, ed. E Puig-Waldmüller, pp. 91–105. Barcelona: Univ. Pompeu Fabra

Benz A. 2009. Implicatures of irrelevant answers and the principle of optimal completion. In *Lecture Notes in Artificial Intelligence*, vol. 5422: *Proceedings of the 7th International Tbilisi Symposium on Logic, Language, and Computation. Revised Selected Papers*, ed. P Bosch, D Gabelaia, J Lang, pp. 95–109. Berlin: Springer

Benz A. 2011. How to set up normal optimal answer models. See Benz et al. 2011, pp. 14–39

Benz A. 2012a. Errors in pragmatics. *J. Log. Lang. Inf.* 21:97–116

Benz A. 2012b. Implicatures of complex sentences in error models. In *Practical Theories and Empirical Practice*, ed. A Schalley, pp. 273–306. Amsterdam: Benjamins

Benz A, Ebert C, Jäger G, van Rooij R, ed. 2011. *Lecture Notes in Artificial Intelligence*, vol. 6207: *Language, Games, and Evolution. Trends in Current Research on Language and Game Theory*. Heidelberg, Ger.: Springer
Benz A, Jäger G, van Rooij R, ed. 2006a. *Game Theory and Pragmatics*. Basingstoke, UK: Palgrave Macmillan
Benz A, Jäger G, van Rooij R. 2006b. An introduction to game theory for linguists. See Benz et al. 2006a, pp. 1–82
Benz A, Mattausch J, ed. 2011. *Linguistics Today*, vol. 180: *Bidirectional Optimality Theory*. Amsterdam: Benjamins
Benz A, van Rooij R. 2007. Optimal assertions and what they implicate: a uniform game theoretic approach. *Topoi Int. Rev. Philos.* 27:63–78
Bergen L, Levy R, Goodman ND. 2016. Pragmatic reasoning through semantic inference. *Semant. Pragmat.* 9:1–83
Blutner R. 2000. Some aspects of optimality in natural language interpretation. *J. Semant.* 17:189–216
Blutner R. 2004. Pragmatics and the lexicon. In *The Handbook of Pragmatics*, ed. L Horn, G Ward, pp. 488–514. Oxford, UK: Blackwell
Camerer CF. 2003. *Behavioral Game Theory: Experiments in Strategic Interaction*. Princeton, NJ: Princeton Univ. Press
Chierchia G. 2004. Scalar implicatures, polarity phenomena, and the syntax/pragmatics interface. In *Structures and Beyond*, ed. A Belletti, pp. 39–103. Oxford, UK: Oxford Univ. Press
Clark R. 2012. *Meaningful Games: Exploring Language with Game Theory*. Cambridge, MA: MIT Press
Clark R, Parikh P. 2007. Game theory and discourse anaphora. *J. Log. Lang. Inf.* 16:265–82
Degen J, Franke M. 2012. Optimal reasoning about referential expressions. In *Proceedings of the 16th Workshop on the Semantics and Pragmatics of Dialogue*, ed. S Brown-Schmidt, J Ginzburg, S Larsson, pp. 2–11. Paris: Univ. Paris-Diderot
Degen J, Goodman ND. 2014. Lost your marbles? The puzzle of dependent measures in experimental pragmatics. In *Proceedings of the 36th Annual Conference of the Cognitive Science Society*, ed. P Bello, M Guarini, M McShane, B Scassellati, pp. 397–402. Cambridge, MA: Cogn. Sci. Soc.
Degen J, Tessler MH, Goodman ND. 2015. Wonky worlds: Listeners revise world knowledge when utterances are odd. In *Proceedings of the 37th Annual Conference of the Cognitive Science Society*, ed. DC Noelle, R Dale, AS Warlaumont, J Yoshimi, T Matlock, et al., pp. 548–53. Cambridge, MA: Cogn. Sci. Soc.
Dekker P, van Rooij R. 2000. Bi-directional optimality theory: an application of game theory. *J. Semant.* 17:217–42
Dixit AK, Skeath S, Reiley DH. 2009. *Games of Strategy*. New York: Norton
Fox D. 2007. Free choice and the theory of scalar implicatures. In *Presupposition and Implicature in Compositional Semantics*, ed. U Sauerland, P Stateva, pp. 71–120. Basingstoke, UK: Palgrave Mcmillan
Frank MC, Goodman ND. 2012. Predicting pragmatic reasoning in language games. *Science* 336:998
Franke M. 2009. *Signal to act: game theory in pragmatics*. PhD thesis, Univ. Amsterdam, Neth.
Franke M. 2011. Quantity implicatures, exhaustive interpretation, and rational conversation. *Semant. Pragmat.* 4:1–82
Franke M. 2017. Game theory in pragmatics: evolution, rationality, and reasoning. In *Oxford Research Encyclopedia of Linguistics*. **https://doi.org/10.1093/acrefore/9780199384655.013.202**
Franke M, Jäger G. 2014. Pragmatic back–and–forth reasoning. In *Pragmatics, Semantics and the Case of Scalar Implicatures*, ed. SP Reda, pp. 170–200. New York: Palgrave Macmillan
Fudenberg D, Tirole J. 1991. *Game Theory*. Cambridge, MA: MIT Press
Goodman ND, Stuhlmüller A. 2013. Knowledge and implicature: modeling language understanding as social cognition. *Top. Cogn. Sci.* 5:173–84
Grice HP. 1975. Logic and conversation. In *Syntax and Semantics*, ed. P Cole, JL Morgan, 3:41–58. New York: Academic
Harsanyi JC. 1968. Games of incomplete information played by 'Bayesian' players. Part II. *Manag. Sci.* 14:320–34
Horn LR. 1989. *A Natural History of Negation*. Chicago: Univ. Chicago Press
Jäger G. 2007. Game dynamics connects semantics and pragmatics. In *Game Theory and Linguistic Meaning*, ed. AVJ Pietarinen, pp. 89–102. Amsterdam: Elsevier
Jäger G. 2011. Game-theoretical pragmatics. In *Handbook of Logic and Language*, ed. J van Benthem, A ter Meulen, pp. 467–91. Amsterdam: Elsevier

Jäger G, Ebert C. 2009. Pragmatic rationalizability. In *Proceedings of Sinn und Bedeutung 13*, ed. A Riester, T Solstad, pp. 1–15. Stuttgart, Ger.: Univ. Stuttgart

Lassiter D, Goodman ND. 2014. Context, scale structure, and statistics in the interpretation of positive-form adjectives. In *Proceedings of the 23rd Semantics and Linguistic Theory Conference (SALT 23)*, ed. T Snider, pp. 587–610. Ithaca, NY: CLC

Levinson SC. 2000. *Presumptive Meanings: The Theory of Generalized Conversational Implicatures*. Cambridge, MA: MIT Press

Lewis D. 1969. *Convention*. Cambridge, MA: Harvard Univ. Press

Mayol L. 2006. On pronouns in Catalan and game theory. In *Proceedings of the European Summer School in Logic, Language, and Information (ESSLLI) Workshop on Ambiguity in Anaphora*, ed. R Artstein, M Poesio, pp. 73–82. Málaga, Spain: FoLLI

Mayol L, Clark R. 2010. Pronouns in Catalan: games of partial information and the use of linguistic resources. *J. Pragmat.* 42:781–99

Merin A. 1999. Information, relevance, and social decisionmaking: some principles and results of decision-theoretic semantics. In *Logic, Language, and Information*, ed. L Moss, J Ginzburg, M de Rijke, 2:179–221. Stanford, CA: Cent. Study Lang. Inf.

Moulin H. 1986. *Game Theory for Social Sciences*. New York: NYU Press

Myerson RB. 1991. *Game Theory: Analysis of Conflict*. Cambridge, MA: Harvard Univ. Press

Nash JF. 1950. *Non-cooperative games*. PhD thesis, Princeton Univ., Princeton, NJ

Osborne MJ, Rubinstein A. 1994. *A course in game theory*. Cambridge, MA: MIT Press

Parikh P. 1990. Situations, games, and ambiguity. In *Situation Theory and Its Applications*, ed. R Cooper, K Mukai, J Perry, 1:449–69. Stanford, CA: Cent. Study Lang. Inf.

Parikh P. 1991. Communication and strategic inference. *Linguist. Philos.* 14:473–514

Parikh P. 1992. A game-theoretic account of implicature. In *Proceedings of the 4th Conference on Theoretical Aspects of Reasoning About Knowledge*, ed. Y Moses, pp. 85–94. Monterey, CA: Morgan Kaufmann

Parikh P. 2000. Communication, meaning, and interpretation. *Linguist. Philos.* 23:185–212

Parikh P. 2001. *The Use of Language*. Stanford, CA: Cent. Study Lang. Inf.

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

Parikh R. 1996. Vagueness and utility: the semantics of common nouns. *Linguist. Philos.* 17:521–35

Pavan S. 2013. Quantity implicatures and iterated admissibility. *Linguist. Philos.* 36:261–90

Pietarinen AVJ, ed. 2007. *Game Theory and Linguistic Meaning*. Amsterdam: Elsevier

Potts C, Lassiter D, Levy R, Frank MC. 2016. Embedded implicatures as pragmatic inferences under compositional lexical uncertainty. *J. Semant.* 33:755–802

Qing C, Franke M. 2014. Gradable adjectives, vagueness, and optimal language use: a speaker-oriented model. In *Proceedings of the 24th Semantics and Linguistic Theory Conference (SALT 24)*, ed. T Snider, S D'Antonio, M Weigand, pp. 23–41. Ithaca, NY: CLC

Qing C, Franke M. 2015. Variations on a Bayesian theme: comparing Bayesian models of referential reasoning. In *Bayesian Natural Language Semantics and Pragmatics*, ed. H Zeevat, HC Schmitz, pp. 201–20. Heidelberg, Ger.: Springer

Ross I. 2006. *Games interlocutors play: new adventures in compositionality and conversational implicature*. PhD thesis, Univ. Pa., Philadelphia

Rothschild D. 2013. Game theory and scalar implicatures. *Philos. Perspect.* 27:438–78

Rubinstein A. 2000. *Economics and Language*. Cambridge, UK: Cambridge Univ. Press

Sauerland U. 2004. Scalar implicatures in complex sentences. *Linguist. Philos.* 27:367–91

Schelling T. 1960. *The Strategy of Conflict*. Cambridge, MA: Harvard Univ. Press

Schulz K, van Rooij R. 2006. Pragmatic meaning and non-monotonic reasoning: the case of exhaustive interpretation. *Linguist. Philos.* 29:205–50

Stevens J. 2016. Focus games. *Linguist. Philos.* 39:395–441

Train KE. 2003. *Discrete Choice Methods with Simulation*. Cambridge, UK: Cambridge Univ. Press

van Rooij R. 2004a. Relevance and bidirectional optimality theory. In *Optimality Theory and Pragmatics*, ed. R Blutner, H Zeevat, pp. 173–210. Basingstoke, UK: Palgrave Macmillan

van Rooij R. 2004b. Utility of mention—some questions. *Res. Lang. Comput.* 2:401–16

van Rooij R, Franke M. 2015. Optimality-theoretic and game-theoretic approaches to implicature. In *The Stanford Encyclopedia of Philosophy*, ed. EN Zalta. **https://plato.stanford.edu/entries/implicature-optimality-games/**

von Neumann J, Morgenstern O. 1944. *Theory of Games and Economic Behavior*. Princeton, NJ: Princeton Univ. Press

Zaefferer D. 1977. Understanding misunderstandings: a proposal for an explanation of reading choices. *J. Pragmat.* 1:329–46