Signaling Games and Gricean Pragmatics

Paul Daniell

Autumn 2009

Pragmatics has been, for a considerable time now, a central project in both linguistic cognitive science and the philosophy of language: practitioners in both fields, using the techniques *de rigueur* of their respective disciplines, have sought to formalize and represent how speakers *use* expressions in certain contexts and further understand the ways in which contexts can sanction an interpretation of the expression by interlocutors which, in day-to-day language, diverge from the expression's *conventional* or *literal* meaning.

The situations in their study are commonplace to most of us who are not cognitive scientists or philosophers. When Harry and Sally, preparing for their dinner party, hear the doorbell and Sally declares "Get the door Hal! It's the catering or the musicians," we think it perfectly natural for Harry to infer that his wife is trying to communicate that it is *exclusively* one group or the other that is before their front door but not both. This is true even though the "or" of the sentential calculus (which we might regard as coinciding with the *conventional meaning* of "or") permits that Sally's declaration is true in the *inclusive* case as well, where both the catering and musicians are before the door. Has Sally's *use* of the expression in this context sanctioned an elimination of the inclusive interpretation? Or are we forced to accept the thesis that natural language itself is simply full of semantical irregularities, such as the ambiguity between the inclusive and exclusive "or"? If we wish to hold the former and avoid the latter, can we *systematically* explain how such sanctioned alternate interpretations arise in everyday conversation?

This last question is the central preoccupation of H. Paul Grice's famous 1975 paper, *Logic and Conversation.* Grice begins by proposing a Cooperative Principle, which he says all rational and cooperative interlocutors are expected to observe, other things being equal: MAKE YOUR CONVERSATIONAL CONTRIBUTION SUCH AS IS REQUIRED, AT THE STAGE AT WHICH IT OCCURS, BY THE ACCEPTED PURPOSE OR DIRECTION OF THE TALK EXCHANGE IN WHICH YOU ARE ENGAGED.

In his analysis, Grice includes four additional Maxims or Rules of Thumb (Quality, Quantity, Relevance and Manner) which he argues are common knowledge to rational and cooperative interlocutors and which, for the most part, they also observe. These maxims are stated informally, as in the Maxim of Quantity: Make your contribution as informative as is required but not more informative than required.

Grice's Cooperative Principle and his Maxims help explain the central and distinguished notion of a *conversational implicature.* As in Sally's utterance to Harry at their dinner party, implicatures are involved in cases where an utterance serves to convey more than its conventional meaning. For Grice, a speaker's utterance gives rise to a conversational implicature when a hearer must infer something additional to its *literal meaning* in order to maintain that the speaker is obeying the Cooperative Principle though he or she is in violation of a Gricean Maxim.

Consider, for example, the following conversation, which might be overheard between two lawyers at the esteemed firm of Dewey, Cheatem, and Howe:

(1) a. PETER: Are you leaving early today?
    b. BILL: I've got to finish the Fishman case.

What Bill, a rational and cooperative speaker, must mean here is captured by the following sentence:

(2) Bill is not leaving early today because he's working late on a case.

We immediately observe that this situation satisfies one necessary condition of a Gricean conversational implicature: the literal meaning given in (1b) is in fact *irrelevant* to Peter's question (thereby violating the Maxim of Relevance). Bill's finishing the Fishman case simply does not answer whether he will leave early today. However if (2) is implicated and also part of *what is meant* by Bill's uttering (1b), then we have reason to believe that he is acting under the legislation of the Cooperative Principle.

Grice's schematic, concise idea has generated a massive literature; indeed, it is at the root of a growing tree of variant theories. One main branch has put emphasis on the need for modifying, altering, or pruning the set of Maxims. I will here be concerned with another cluster of variants, which have sought to alter the insights found in the Maxims and the Cooperative Principle, in order to make them align with other major theoretical techniques and abstractions found in the other social sciences, especially theoretical economics. Such theorists derive much consolation from Grice's own admission in a series of lectures given at Harvard, that the Maxims are themselves not unique to the study of language, but have analogues in other transactional behavior. Indeed, Grice says like near the beginning of his essay:

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

This insight has set the stage for the decision-theoretic and game-theoretic treatment of Gricean pragmatics in Franke (2008) and Van Rooij (2008). These studies in particular have focused on a type of strategic interaction between players known as a *signaling game.* Such games are not new to linguistic pragmatics; David Lewis made them central to his own analysis of the conventionality of language in his 1969 book, *Convention.* The recent studies I have cited, however, have attempted to reconcile signaling games with the Gricean conversational implicature. In particular they have focused on two distinct projects: first, they have attempted to show how the formalism of the signaling game can represent a special kind of well-studied implicature called a scalar implicature. Second, they have attempted to show how solution concepts to signaling games, once certain constraints are placed on the players, yield player strategies which coincide substantially with the predictions made by Grice's program.

My project here will be to suggest that though the first of these projects offers some promise in precisely and formally modeling scalar implicatures (especially where quantifiers form the scales), the second of these projects requires principles of rationality which are far too stringent for Grice's original *descriptive* project. I will attempt to argue that game theorists are forced to such stringent standards of rationality because they wish to identify player strategy profiles which are *uniquely* at equilibrium. The argument requires some setup, however. In what follows, I will review what scalar implicatures are in the first section and then describe something I call the *cost of equilibrium problem* in the second. In the third section, I will review the the form of the Lewisian signaling game and also what standards of rationality game theorists require to provide an analysis of *scalar implicatures* which coincides with Grice's story in *Logic and Conversation*. In the fourth section, I provide some arguments about why those standards of rationality may be incompatible with the story of conversational implicature in *general*. I conclude with some thoughts on what parts of the game theoretic project may be salvaged and re-purposed. First, however, we have to turn to the preliminaries.

1 Scalar Implicatures

Scalar implicatures (sometimes called "quantity implicatures") are a sort of conversational implicature generated by *semantically weak* utterances. What is implicated by such an utterance is the negation of one or more of its *semantically stronger* counterparts. Consider, for example the following conversation between Peter and Bill:

(3) a. Peter: Did you eat anything at Karen's birthday party?
    b. Bill: I ate some of the cake.

Two semantically stronger alternatives to Bill's response in (3b) are sentences reporting the two following facts:

(4) Bill ate most of the cake.

(5) Bill ate the entire cake.

Sentence (5) is stronger than (4) since saying (5) *semantically entails* (4). That is, if I have eaten all of something, I have also eaten most of it. Similarly if I have eaten most of this meal before me, I have eaten some of it. For this reason, the quantifiers in (3b), (4), and (5) lie on a quantificational *scale* increasing in semantic strength.

In what sense can we say that Bill's utterance of (3b) implicates the negation of (4) or (5)? The reasoning to this implicature can be drawn out schematically in two steps. I show this for the negation of (5) with the negation of (4) being similar, *mutatis mutandis*:

Negation Phase. Since Peter believes Bill to be following the Cooperative Principle and the Gricean Maxims, and since (5) includes a quantifier which is semantically stronger than the one present in (3b), the Maxim of Quantity would require that if (5) were true, Bill would have had to utter something other than (3b), since (3b) is not a semantic alternative to (5). For this reason, (5) must be false since Bill is following the Cooperative Principle.

Implicature Phase. Since Bill has uttered (3b) instead of

(6) I ate some but not all of the cake.

Bill has violated the Maxim of Quantity because he has failed to make his utterance "as informative as is required". Since Peter believes that Bill, his rational interlocutor, is still acting under the legislation of the Cooperative Principle, (6) must be *implicated* in Bill's uttering (3b).

Importantly, such scalar implicatures may arise along any scale of semantic phenomena ordered by strength. For example, a speaker's uttering

(7) It is possible that the Democrats pass Health Care Reform.

might implicate, in appropriate circumstances, the negation of the semantically stronger necessity claim in (8):

(8) It is not necessary that the Democrats pass Health Care Reform.

2 Games, Equilibrium, and Rationality

Games are mathematical objects which model strategic interactions between individuals. The most famous 20th century analysis of such games was given by John von Neumann and Oskar Morgenstern in their 1944 book *Theory of Games and Economic Behavior*. Von Neumann and Morgenstern put special emphasis on static games, in which each player makes a *single* decision, *simultaneously* and *independently*. Each player bases his or her decision on the *utility functions* of the agents in the game (including his or her own) in addition to purportedly weak, common knowledge assumptions about their *rationality*. One such principle can be stated as follows:

MU. In pairwise comparisons of consequences, prefer the one which maximizes utility.

The normal-form representation of static 2-player games is a table. Table 1, for example, models a simple game in which a Row Player must decide between R1 and R2 and a Column Player must decide between C1 and C2. Each cell in the table contains a vector which represents the utility payoff for the Row Player as the first component and the Column Player as second.

Table 1: A Coordination Game

	C_1	C_2
R_1	$\langle 100, 100 \rangle$	$\langle 1, 1 \rangle$
R_2	$\langle 1, 1 \rangle$	$\langle 10, 10 \rangle$

Formally, this game can be represented as a 4-tuple, including the sets of actions available to each player A_R, A_C and two utility functions, one for each player which maps pairs of actions into the real numbers : U_{Row}: $A_R \times A_C \mapsto \mathbb{R}$ and U_{Column}: $A_R \times A_C \mapsto \mathbb{R}$.

A player's *strategy* is a plan of action for any situation that may arise in a game. In static, two player games like the one represented in Table 1, these are just *moves*. For example, the Row Player may choose (i) R1 unconditionally; or (ii) R2 unconditionally. A complete set of strategies for every individual in the game is called a *strategy profile*. One principle project of game theory, starting with Von Neumann and Morgenstern, has been to investigate when games, in concert with a set of assumptions about rationality, yield stable predictions about a viable subset of strategy profiles or, in special cases, unique strategy profiles. One notion central to this project is the *Nash Equilibrium*, named after the mathematician John Nash. A strategy profile is in this kind of *equilibrium* if no agent can improve his or her payoff situation by unilaterally changing his or her own strategy. In our example, [R2, C2] is a Nash Equilibrium: the Column player would reduce his or her payoff if the selected profile is [R2, C1]. This is also true for the Row Player, if he or she would seek to deviate to [R1, C2].

Construed as a principle of rationality, MU does not here "produce" a *unique* Nash equilibrium. This is because [R1, C1] is also at Nash Equilibrium. We also immediately observe that [R1, C1] is Pareto efficient: this strategy makes all the players in the game strictly better off in comparison to any other strategy. What modifications to the our assumptions would be required to ensure the Pareto efficient strategy profile? We could add two, more restrictive though perhaps controversial conditions on rationality. The first involves agents' subjective probabilities concerning the actions of other players. The principle was first suggested by Pierre Simon, the Marquis de Laplace and because of this has come to be called Laplace's Principle of Indifference:

LPI. When actions lead to uncertain outcomes, assume that the outcomes are equally likely.

Introducing this principle requires a slight modification to the formal structure of the game. In particular it requires two additional probability distributions Pr_R and Pr_C, which represent each player's degree of belief about what the other will do. Laplace's Principle of Indifference makes the controversial claim that because the player's are acting under risk, they must assume that each of the other player's possible actions has an equal chance of being brought about. Nevertheless, with the introduction of probability distributions over the risky prospects involved for each player, we can determine the *expected utility* of each action. This allows us to introduce our second principle of rationality, the Maxim of Expected Utility:

MEU. Choose actions which maximize expected utility.

Since the expected utility of R1 for the Row Player and C1 is 50.5, which supersedes the expected utility of both R2 and C2, we have shown that LPI and MEU both yield a *unique* Nash Equilibrium in the strategy profile [R1, C1]. However we have done so only in a highly controversial way. We can understand this controversy in two different ways. First, both LPI and MEU are highly disputable principles of rationality. Proponents of the theory of subjective probability developed in the tradition of Ramsey, Savage, and De Finetti have charged that LPI must be based on unsupported metaphysical assumptions concerning symmetry; instead they have argued that the only normative restrictions on subjective prior probability are weaker constraints on the *coherence* of partial beliefs. In particular, that probabilities in an event space must be nonnegative, sum to the unit measure, and that the probability of the disjunction of two ore more events must be reflected in the *additivity* of their real-valued individual probabilities. Even more controversial is the second principle MEU, a form of which has been at the heart of a debate between consequentialists and their opponents for several centuries.

The second kind of controversy concerns whether or these games and their associated principles of rationality are being employed to *describe* human behavior ceteris paribus or whether they are objects to model *normative* decision making in situations described in a mathematically formal way. Some game theorists have held the latter. For example, Luce and Raiffa declare in a famous passage in *Games and Decisions*:

> We feel that it is crucial that the social scientists recognize that game theory is not *descriptive*, but rather (conditionally) *normative*. It states neither how people do behave or how they should behave in an absolute sense, but how they should behave if they wish to achieve certain ends.

But this is of course to make nonsense of Von Neumann and Morgenstern's stated purpose on the first page of *Theory of Games and Economic Behavior* (italics mine):

> The purpose of this book is to present a discussion of some fundamental questions of economic theory which require a different from that which they have found thus far in the literature. The analysis is concerned with some basic problems arising from a study of economic behavior which have been the center of attention of economist for a long time. They have their origin in the attempts to find *an exact description* of the endeavor of the individual to obtain a maximum of utility, or, in the case of the entrepreneur, a maximum of profit.

There is clearly much more to be said about this debate, but it is obvious to which camp game theoretic Gricean pragmatics must belong. Indeed, Gricean conversational implicatures and the Cooperative Principle would have no theoretical importance if they were not descriptive of human behavior. Conversational implicatures describe the inferences listeners regularly make in order to sustain the belief that most of their interlocutors are cooperative.

Though I do not wish here to take a stake in the debate between normative and descriptive game theory, it is clear that the *descriptive* project, with respect to Gricean pragmatics, faces two major competing desiderata. The first is that the constraints of rationality must be sufficiently weak so that the majority of competent speakers can satisfy them. The second is that the games must model situations of decision-making which have unique equilibria, without which the game formalism (in concert with its associated principles of rationality) loses all explanatory power. The competition between these two desiderata could be called the *cost of equilibrium* problem: if the requirements of rationality are too strong, then it will be unfeasible to say that they describe linguistic behavior; if they are too weak, they will not sanction the appropriate implicatures. My main project in the remainder of this paper will be to show that the proposed game theoretic treatment of Gricean implicature suggested by Van Rooij (2008) and Franke (2008) cannot satisfy these requirements. But first, we will have to get clear on the game formalism which both of these linguists use, which is the Lewisian Signaling Game.

3 Lewisian Signaling Games and Scalar Implicature

Unlike the static games discussed in section II, the game theoretic treatment of Gricean implicature presented in Van Rooij (2008) and Franke (2008) discuss *dynamic games*, in which players make a series of moves in sequence. Such games more naturally model the sequential nature of interactions between speakers and hearers. As mentioned earlier, the type of dynamic game that has been at the forefront of much work in linguistic pragmatics has been the signaling game proposed by David Lewis in *Convention.* Such games model circumstances of asymmetric information between a sender S and a receiver R. Here, we assume that S can perfectly observe one among a set of states of the world $\mathcal{T}$. He or she must also choose to send one among a set of messages $\mathcal{M}$ to R. After receiving some $m \in \mathcal{M}$, R must decide which action to take in the set $\mathcal{A}$. In all such games $|\mathcal{T}| = |\mathcal{A}|$ and $|\mathcal{A}| \leq |\mathcal{M}|$.

In conversational implicature interpretations of the Lewisian signaling game, this action is what kind of implicature to infer from the speaker's message. As with the static games discussed in the previous section, both the sender and the receive have utility functions, U_S and U_R. Similar to the two-person, two-strategy game presented in the previous section, Lewisian signaling games have a coordinated payoff matrix, in which each player's utility is similar or identical. In such games, when players coordinate their behavior, they receive greater payoffs which are usually found on the diagonal of the payoff matrix.

Additionally, van Rooij's treatment assumes that there is an empirical probability distribution Pr over the states of the world. We can therefore state the game as the tuple

$$\langle \{S, R\}, \mathcal{T}, Pr, \mathcal{M}, \mathcal{A}, U_R, U_S \rangle \tag{1}$$

Unlike the simple unconditional strategies for player described in section II, strategies here describe complete plans of action whatever the circumstance. A sender strategy is therefore a function σ: $\mathcal{T} \mapsto \mathcal{M}$. Likewise, a receiver strategy is the function ρ: $\mathcal{M} \mapsto \mathcal{A}$.

The expected utility of a particular strategy profile, for either player $i \in \{S, R\}$ is given by the following equation:

$$\mathrm{EU_i}(\sigma, \rho) = \sum_{\mathrm{t \in T}} \mathrm{Pr(t)} \times \mathrm{U_i(t}, \sigma(\mathrm{t}), \rho(\sigma(\mathrm{t}))) \tag{2}$$

As in our static game example, a Nash Equilibrium is a sort of strategy profile in which one player cannot benefit by unilateral movement. Formally, the pair $\langle \sigma^*, \rho^* \rangle$ is at Nash Equilibrium iff $\neg\exists\sigma : \mathrm{EU_S}(\sigma, \rho^*) > \mathrm{EU_S}(\sigma^*, \rho^*)$ and $\neg\exists\rho : \mathrm{EU_R}(\sigma^*, \rho) > \mathrm{EU_R}(\sigma^*, \rho^*)$.

How can we use such a game formalism to model the sort of scalar implicatures described in section I? Let us recall the typical sort of conversation involved in a scalar implicature, where (9b) and (9c) can be viewed as alternative responses and therefore members of $\mathcal{M}$:

(9) a. RECEIVER: Did you eat anything at the birthday party?
b. $\boldsymbol{m_{some}}$ SENDER : I ate some of the cake.
c. $\boldsymbol{m_{all}}$ SENDER: I ate all of the cake.

The states of the world in question, $\mathcal{T}$, can be captured by two different propositions (10) and (11)

(10) $\boldsymbol{t_{sbna}}$: SENDER ate some but not all of the cake.

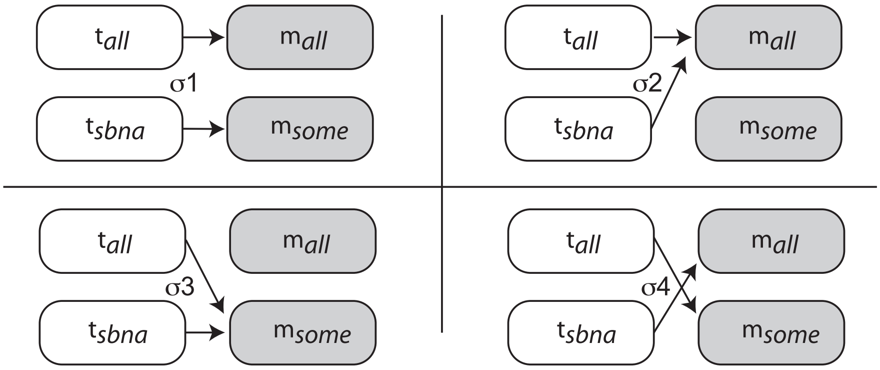

Figure 1: Strategies σ for the Sender (What message to send)

(11) $\boldsymbol{t_{all}}$: SENDER ate all of the cake.

This may sound complicated, but there are only four strategies in the Lewisian signaling game for the sender, which I've called σ_1 through σ_4. These are the states of the world the sender observes in combination with the message he or she chooses to send. They are illustrated in Figure 1.

Likewise, the actions (or implicatures which the receiver makes), $\mathcal{A}$, can be captured by the following propositions:

(12) $\boldsymbol{a_{sbna}}$ RECEIVER infers: SENDER ate some but not all of the cake.

(13) $\boldsymbol{a_{all}}$ RECEIVER infers: SENDER ate all of the cake.

The corresponding four strategies for the receiver are illustrated in Figure 2.

Since there are four strategies for each player, there are 16 strategy profiles. Because this is game has a coordinated payoff matrix, let us say that for a given state of the world t_x (i.e. how much cake the speaker ate) and the implicature made a_y, $\mathrm{U(t_x, a_y)} = 1$ when $x = y$ and 0 otherwise. This embodies mathematically Grice's Cooperative Principle: the speakers share certain aims or ends and benefit when they coordinate. Let us say, finally that $Pr(t_{sbna}) = x$ and $Pr(t_{all}) = 1 - x$. Without loss of any generality, we will say for the moment that $x > 1 - x$. We can therefore, calculate the

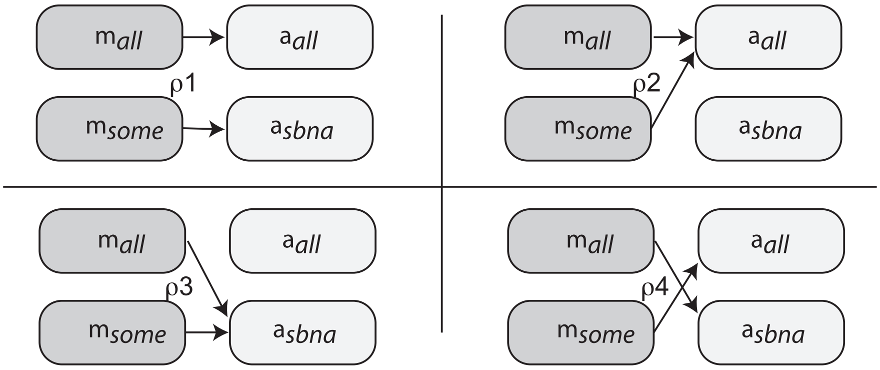

Figure 2: Strategies ρ for the Receiver (What implicature to infer)

payoff matrix for the 16 strategies in the Scalar Implicature game, which are listed in Table 2. (Here the first components of the payoff vector are for the receiver, and the second for the sender.)

Table 2: Payoffs for the Implicature Game

	σ_1	σ_2	σ_3	σ_4
ρ_1	$\boldsymbol{\langle 1, 1 \rangle}$	$\langle 1-x, 1-x \rangle$	$\langle x, x \rangle$	$\langle 0, 0 \rangle$
ρ_2	$\langle 1-x, 1-x \rangle$	$\langle 1-x, 1-x \rangle$	$\langle 1-x, 1-x \rangle$	$\langle 1-x, 1-x \rangle$
ρ_3	$\langle x, x \rangle$	$\boldsymbol{\langle x, x \rangle}$	$\boldsymbol{\langle x, x \rangle}$	$\langle x, x \rangle$
ρ_4	$\langle 0, 0 \rangle$	$\langle x, x \rangle$	$\langle 1-x, 1-x \rangle$	$\boldsymbol{\langle 1, 1 \rangle}$

Van Rooij (2008) shows that the the two-player scalar implicature game has multiple Nash equilibria, if we assume that rationality requires that players maximize expected utility. These equilibria, shown in bold, are $\langle \rho_1, \sigma_1 \rangle$, $\langle \rho_3, \sigma_2 \rangle$, $\langle \rho_3, \sigma_3 \rangle$, and $\langle \rho_4, \sigma_4 \rangle$. Within these, we can distinguish between those instances in which the sender's strategy involves sending only a single message and those in which different messages are sent. Naturally we think that communication between the speaker and the hearer has only occurred in the latter case, in which different messages lead to different actions or consequences. These strategy profiles are, for this reason, called *separating* *equilibria*. When single message strategies lead to different consequences at Nash equilibrium, we say the strategy profile is a *pooling equilibrium*.

Table 2 illustrates, in part, what I am calling the cost of equilibrium problem; since the game, so described, has produced too many equilibria, more restrictions must be placed on the set of strategy profiles to select only the desirable separating equilibria (i.e., $\langle \rho_1, \sigma_1 \rangle$) . Van Rooij suggests that weak and intuitive constraints on the rationality of the interlocutors can be introduced to do just that. This is what the study suggests specifically:

> Farell (1993) and others have shown that we can restrict this set of equilibria considerably, if we assume that messages have an exogenously given conventional meaning. Now we can demand of messages that they should be send truthfully and that equilibria should be Neologism-Proof.

To accommodate this requirement, van Rooij suggests that the form of the game be modified to include a semantic interpretation function $[\![\cdot]\!] : \mathcal{M} \mapsto 2^T$. The interpretation functions shows what states of the world are compatible with the conventional meaning of the set of messages. In our scalar implicature game, this is straightforward: $[\![m_{all}]\!] = \{t_{all}\}$ and $[\![m_{some}]\!] = \{t_{all}, t_{sbna}\}$. We can eliminate one competing pooling equilibrium in Table 2 (i.e. $\langle \rho_4, \sigma_4 \rangle$) if we make the requirement that the sender choose his or her strategy truthfully. More formally we can say that $\forall t \in T$ and $\forall \sigma, t \in [\![\sigma(t)]\!]$.

The second and more controversial constraint is the requirement that all equilibria be Neologism-Proof. Here, again, are van Rooij's words precisely:

> Farrell proposes that an equilibrium is *Neologism-Proof* if in no situation the speaker has an incentive to use an available (unused) credible message. The intuition behind this notion is that if there exists a credible message m_t that is not used by the sender in the equilibrium play of the game, and the sender of type t would be better off if she would have sent that message...then this equilibrium is not Neologism-Proof...If there there is chance that the

> speaker is off type t_{all}, or is in situation t_{all}, it is always better for a sender of that type to send the credible message m_{all} instead of the weaker m_{some}.[1]

In essence, the Neologism-Proofness places an antecedent restriction on the speaker that he or she may not introduce "neologistic" language arbitrarily (the precise sense of neologism will be made clearer in the next section). Because m_{some} can only be sent in the cases when the state of the world is t_{sbna} and m_{all} can only be sent for t_{all}, the sender strategy is part of the unique Nash Equilibrium $\langle \rho_1, \sigma_1 \rangle$. Thus van Roiij concludes:

> For this reason, we might call the inverse sender strategy $\sigma{*}^{-1}$ of this unique equilibrium applied to a message m its *pragmatic* interpretation function. But then, we see that 'John ate some of the cookies' pragmatically entails that John did not eat all of the cookies, i.e. the scalar implicature.

In the remaining, I wish to explore whether these two constraints are indeed plausible principles of rationality or whether they merely constitute an unwarranted gerrymandering the signaling game in an effort to make it coincide with accepted kinds of scalar implicature. It will be unfeasible to answer this question in a broad and unconditional way. I wish instead to consider whether credibility and Neologism-proofness preserve other sorts of implicatures Grice sanctioned in *Logic and Conversation*. Do all other Gricean conversational implicatures require speaker truthfulness and general neologism-proofness? If so, we may give credence to these principles as underlying and general parts of the Gricean program. If not, however, we have reason to suspect that they have been engineered to deal in a singular and specific way with the case of scalar implicatures. They would not be, it would seem, the general principles of rationality we are seeking in order to formalize the idea of conversational implicature through game theory. These are the issues to which I here turn.

[1]Van Rooij uses the letter f for his messages. I have changed these to my letter m.

4 Credibility and Neologism-Proofness

One important consideration which lends some weight to the game theoretic approach, van Rooij argues, is that the requirement of Credibility itself appears in Grice's Maxims, in the form of the Maxim of Quality. Here is what Grice has to say about this Maxim:

> Under the category falls a supermaxim – "Try to make your contribution one that is true" – and two more specific maxims:
>
> 1. Do not say what you believe to be false.
> 2. Do not say that which you lack adequate evidence.

Though van Rooij does not explicitly make this claim explicitly, we could also make an analogous claim about avoiding Neologistic uses of language. In particular we could say that Neologism-Proofness is captured by the Maxim of Manner; in particular the precept encoded in Grice's submaxim "Avoid obscurity of expression." That is, when the interlocutor utters the true sentence "I ate some of the cake" in order to communicate the fact that he ate all of it, his neologism is simply a violation of the Manner Maxim because it is merely an obscure or roundabout manner of communicating that fact.

Though it is true that the requirements of Credibility and Neologism-Proofness have analogues in Grice's Maxims, the Maxims themselves are not presented in the Gricean program as necessary conditions of rationality. In fact, if they Maxims were strict necessary conditions on rationality, then no rational agents could generate conversational implicatures at all. This is because a condition of generating a conversational implicature is that the speaker violate, flout, exploit, or opt out of the maxim. The only unfailing requirements of rationality in the theory of conversational implicature are the Cooperative Principle and its constraints that each participant of the conversation recognizes a common set of purposes or a mutually accepted direction.

Consider, for example, the analogous case in which the Maxim of Relation is construed not as a rule of thumb but as a requirement of rationality. When we observe the following conversation

(14) a. A: How is C is getting on in his job?
b. B: Well he hasn't been to prison yet.

we are forced into the incongruous position of accounting for the irrelevance of B's response by suggesting that he is irrational. Grice's own account of the implicature maintains that B is acting rationally even though he has flouted one of the maxims:

> In a suitable setting A might reason as follows: '(1) B has apparently violated the maxim 'Be relevant' and so may be regarded as having flouted one of the maxims conjoining perspicuity, yet I have no reason to suppose he is opting out from the operation of CP; (2) given the circumstances, I can regard his irrelevance as only apparent, if, and only if, I suppose him to think that C is potentially dishonest; (3) B knows that I am capable of working out step (2). So B implicates that C is potentially dishonest.'

So we can conclude that one serious difficulty with the game theoretic approach to Gricean pragmatics proposed by van Roiij is that the rational requirements of Truthfulness and Neologism-Proofness used to identify unique separating equilibria are in fact too strict, in general, to allow conversational implicatures to do their theoretical work. Such strict constraints of rationality would be in the unenvious position of saying that interlocutors like those in (14) are either irrational or that most rational interlocutors just simply don't have such conversations.

A theorist wishing to modify the game theoretic account of Gricean implicature to repair such problems might suggest an alternate account of where the constraints of Credibility and Neologism-Proofness are applied. Such a theorists might say that Credibility and Neologism-Proofness are not constraints on the rationality of the speaker, bur rather constraints on what a listener works out or could work out upon hearing an utterance. But such a modification doesn't seem to gain us anything. Consider again another Gricean example, in which A is discussing X and it is common knowledge that X has recently betrayed him. Now suppose A declares

(15) X is a fine friend.

If it were a constraint on his audience that they could not infer that A was flouting the maxim of quality, then it seems that the whole project of providing a theory of conversational implicature has been self-defeating. It would fail to explain why communication happens even when listeners infer that speakers are not uttering truths.

The second set of concerns about the game theoretic approach to Gricean pragmatics concerns the requirement of Neologism-Proofness specifically. Up until this point, I have treated this constraint as an analogue of the Maxim of Manner. However, the constraint is in fact both much more specific and strict. Recall that the conventional meaning of messages like $[\![m_{all}]\!]$ and $[\![m_{some}]\!]$ are given, in the game context model, by the codomain of the semantic interpretation function $[\![\cdot]\!]$. We assigned $[\![m_{some}]\!]$ to the set of states $\{t_{all}, t_{sbna}\}$ whereas we assigned $[\![m_{all}]\!]$ the singleton $\{t_{all}\}$. Note that in these cases, m_{all} is stronger than m_{some} since m_{all} *semantically entails* m_{some}: $[\![m_{all}]\!] \subset [\![m_{some}]\!]$.

Neologism-Proofness does not merely state that we should avoid being obscure or ambiguous; instead it requires that the semantically strongest message, if it exists, must *always be sent*. In our scalar implicature game, this doesn't seem burdensome because the scale of messages is antecedently defined and contains a finite (in fact handful) of options. In the case of numerical measures, however, the requirements of Neologism-Proofness become unreasonable. For example, if I am speaking with my mother after getting home from the grocery store, Neologism-Proofness requires that I must utter "The groceries cost thirty-one dollars and twenty-seven sense." even when I may just say "The groceries were about thirty bucks." Without modification, this principle would also require us to abandon the scientific practice of rounding to significant digits and opt instead to use the semantically most precise version of any measure. On the face of it, this seems to be recipe for irrational and burdensome linguistic labor.

Is it accurate to say that an sufficiently strong analogue to this principle can be found in Grice's Maxims? Though it does seem to resemble the Manner submaxims 'Avoid obscurity' and 'Avoid ambiguity,' none of Grice's precepts of cooperative conversation are specified technically or as strictly as the requirements of Neologism-Proofness.

Some practical concerns about theoretical parsimony may motivate us to avoid principles specified in strict set-theoretic relations. For example, Dale and Reiter (1995) attempt to formulate the Maxim of Quantity in strict set-theoretic terms for the purpose of generating referring expressions in a computational blocks world. They conclude however that the any procedure which always arrives at a maximally brief referring expression can be reduced to a minimal set cover problem, which is known to be NP-Hard in its computational complexity. Since such algorithm is mostly too slow for other experimental purposes, they suggest a variety of heuristical algorithms which occasionally produce suboptimal (i.e. not maximally brief) expressions. They argue that such algorithms not only have more practical promise but also more closely reflect the observable suboptimalities in natural language.

Perhaps it is a dubious matter whether we should, at this stage, be concerned with practical matters. However, there is another reason to worry about the set-theoretical specification of the Neologism-Proofness. This derives principally from Grice's mandate that all conversational implicatures must be able worked out by listeners in the conversation. The requirement of *calculability* is a necessary condition for all Gricean conversational implicature:

> The presence of a conversational implicature must be a capable of being worked out; for even if it can in fact be intuitively grasped, unless the intuition is replaceable by an argument, the implicature (if present at all) will not count as a conversational implicature; it will be a conventional implicature.

Stating the maxims in set theoretic terms thus casts the calculability of conversational implicatures in which they are flouted into certain peril. It is only when the maxims are stated without any special technical or mathematical apparatus that the story about listeners actually making such implicatures becomes convincing. This is just another reason to suspect that the requirements of Neologism-Proofness are too steep.

5 Conclusion

I will not attempt here a full summary of everything discussed in this paper. I hope to have shown that Lewisian signaling games, when used to model Gricean pragmatics, face serious hurdles by way of the cost of equilibrium problem. When such games are used as context models for conversation and only purported weak constraints are placed decision-making, like the maxim of expected utility, we are faced with a superfluence of Nash Equilibria. Selecting unique equilibria then comes at the cost of highly controversial principles of rationality.

This said, I do not think the project of game theoretic Gricean pragmatics is doomed beyond rescue. The formal apparatus of game theory allows conversational contexts to be modeled with mathematical precision. What does seem beyond rescue is the view that the dynamic Lewisian signaling games describe the pragmatic behavior of rational and linguistically competent interlocutors. There may be much salvageable in the approach if any claims about behavior are made in what Luce and Raiffa called a "conditionally normative" way. In such an altered approach, conclusions about linguistic behavior are not made in a universal and descriptive way, like "Every rational speaker in situation ϕ will ψ." Instead conditionally normative claims will make claims of the form "If the speaker is in situation ϕ and he wants the the listener to believe ξ, then he should do or say ϕ." Clearly this would be a significant shift from the approach taken by contemporary theorists like van Rooij and Franke. I leave it as a suggestion for further research.

References

Dale, R. and Reiter, E. (1995). Computational interpretations of the Gricean maxims in the generation of referring expressions. *Cognitive Science*, 19(2):233–263.

Franke, M. (2008). What is game theoretic pragmatics? Unpublished Manuscript, Universiteit van Amsterdam.

Grice, H. (1975). Logic and conversation. In *Logic of Grammar*, pages 64–75. Dickenson, Encinco, CA.

Lewis, D. (1969). *Convention: A philosophical study.* Harvard University Press.

Luce, R. and Raiffa, H. (1957). *Games and decisions.* Wiley New York.

Van Rooij, R. (2008). Games and quantity implicatures. *Journal of Economic Methodology*, 15(3):261–274.

Von Neumann, J. and Morgenstern, O. (1944). *Theory of games and economic behavior.* Princeton University Press, Princeton, New Jersey.