Game Theory and Scalar Implicatures

Daniel Rothschild*

September 4, 2013

1   Introduction

Much of philosophy of language and linguistics is concerned with showing what is special about language. One of Grice's (1967/1989) contributions, against this tendency, was to treat speech as a form of rational activity, subject to the same sorts of norms and expectations that apply to all such activity. This general perspective has proved very fruitful in pragmatics. However, it is rarely explicitly asked whether a particular pragmatic phenomenon should be understood entirely in terms of rational agency or whether positing special linguistic principles is necessary. This paper is concerned with evaluating the degree to which a species of simple pragmatic inferences, scalar implicatures, should be viewed as a form of rational inference. A rigorous answer to this last question requires using the theoretical resources of game theory. I show that *weak-dominance reasoning*, a standard form of game-theoretic reasoning, allows us to cash out the derivation of simple scalar implicatures as a form of rational inference. I argue that this account of the scalar implicatures is more principled and robust than other explanations in the game theory and pragmatics literature. However, we can still see that deriving Gricean implicatures on the basis of decision and game-theoretic tools is not nearly as straightforward as we might hope, and the ultimate tools needed are disquietingly powerful.

2 Deriving scalar implicatures

Consider this paradigmatic example of a scalar implicature:

Some/All Case There was a small cake in the refrigerator the night before. Mary says to Kay, the next morning, "I ate some of the cake last night." There is an obvious suggestion that Mary conveys the information that she ate some *but not all* of the cake last night.

For the example to work, we need to make certain background assumptions, but when these are in place (and it is not hard to imagine them in place) then Mary's words will carry the meaning suggested. Despite the seeming triviality of this phenomenon, scalar implicatures play a lively role in current contemporary linguistics, both formal and experimental (e.g. Horn, 1972; Chierchia, 2004; Noveck, 2001), .

It might seem that the kind of reasoning in the Some/All Case and others like it is simply rational inference. Let us try to tease out this thought: We can think of communicative situations of this kind as ones in which there is *asymmetric* information: Mary knows something Kay does not, namely how much cake she (Mary) ate. Mary wants Kay to know what she knows. Now suppose Mary, wanting to speak only once, has to choose between uttering the following two sentences:

(1)   a.   I ate some of the cake.
        b.   I ate all of the cake.

Let us assume, as is standard, that (1-b) strictly entails (1-a): if someone ate all the cake, then *ipso facto* they ate some of the cake. Consider the choice of utterance from Mary's perspective: If Mary ate all of the cake then she should assert (1-b) rather than (1-a) as that would convey more information to Kay, given the entailment relations. On the other hand, if she ate only some of the cake then (1-b) would give Kay *false* information which is undesirable. So, if Mary is reasonable she will utter (1-a) if she ate just some of the cake, and (1-b) if she ate all of the cake. Now think of this from Kay's perspective: she can also reason, as we did, about Mary's behavior, so she can conclude that Mary will only utter (1-a) if and only if she only ate some of the cake. So, when Mary utters (1-a) Kay will be able to work out that that Mary only ate some of the cake, and hence that (1-b) is false. This last inference, is the scalar implicature.

The informal reasoning, plausible as it may sound, is flawed. To see this note that in the reasoning above, when Mary considers what information she is conveying to Kay she is not taking account of the fact that Kay may read more into her utterance than *just* the literal message. If she does take this into account, the entailment relation between (1-a) and (1-b) does not carry over to an entailment relation between the information conveyed by each of the sentences. In other words, Mary cannot assume that just because one sentence has a stronger literal meaning than another that it will convey more information. But if we cannot assume that, the informal reasoning above does not go through.

To put the point another way: All that we can assume about Mary's preferences is that she wants to convey as much (true) information as possible. We cannot determine which of (1-a) or (1-b) will convey more information based on their literal meaning alone since we need to assume that sentences may convey more than their literal meaning. So, from the fact that Mary prefers to convey as much information as possible and the fact that (1-b) is literally weaker than (1-a), it does not follow that Mary will say (1-b) rather than (1-a). Something more is needed.

What we have in a communicative situation like the Some/All Case is a decision problem involving two rational agents. Such problems are known to be complex, and there is a special branch of economics and decision theory, game theory, devoted to analyzing them. Most of this paper is concerned with making the case that game theory allows us view scalar implicatures, like the one above, as a form of rational inference. There is already a substantial literature on this question, but my approach and conclusions differ from the standard ones in that literature, and I will try to mark these differences as I go along.

3   Plan

Here is the plan of this (lengthy) paper: The next section, §4, discusses Grice's maxims and the treatment of scalar implicatures in contemporary linguistics. §5–§8 presents the game-theoretic background necessary for the rest of the paper. In §5, I introduce the *signaling game* which is the standard model for communicative situations in game theory. In §6, I discuss the Nash equilibrium notion and the standard refinements of it used for thinking about games like the

3

signaling game. §7 discusses what constraints, outside of standard equilibrium notions, there are on how rational players should play games. §8 discusses what we should hope for in a game-theoretic model of scalar implicatures. §9 introduces a barebones game for modeling scalar implicatures and argues that standard equilibrium concepts cannot be used to argue that scalar implicatures are a rational inference. In §10, I argue that *iterative weak-dominance* reasoning can capture the derivation of scalar implicatures. §11 briefly considers (and rejects) the proposal that Gricean reasoning should be captured by means of the Pareto-Nash equilibrium. §12 discusses and criticizes what I call *reasoning-based accounts*, popular models in the current literature for explaining the derivation of scalar implicatures. In sections §13–§16, I discuss how to apply weak-dominance arguments to different types of scalar and relevance implicatures. §17 discusses *disclosure* cases from the economics literature and assesses their connection to scalar implicatures.

There are two main aims of this paper, one theoretical and one methodological. The theoretical aim is to show that iterative weak dominance provides a good way of explaining the derivation of scalar implicatures within a game-theoretic framework. To some extent this vindicates the commonly held view that scalar implicatures—in idealized cases—are a form of rational inference.

My methodological aim is less specific. There has been a significant amount of work in game-theoretic pragmatics in the last decade or so.[1] This work provides a variety of different techniques for deriving implicatures of many types. As I argued in the introduction, this is an important project because simple explanations of even the most basic implicatures in terms of rationality face serious conceptual hurdles. We must stay aware, though, that game theory, as a collection of different techniques for modeling situations, is a very powerful tool. Merely providing *some* model in game theory that makes a set of behavioral predictions is of little interest, since, for any precisely characterized and vaguely reasonable behavior, we should antecedently expect to be able to provide *some* game-theoretic model. We need to evaluate game-theoretic models of pragmatic reasoning along a variety of dimensions, asking questions like: How robust is the model against slight changes in one's description of the situation? How standard is the sort of reasoning being attributed to the players? Can

---

[1]Parikh (1991, 2001) provides the first game-theoretic account of implicatures, including scalar implicatures.

these models extend to parallel non-linguistic cases? I argue here that many of the prominent theories in the literature do not fare well in the face of this scrutiny. I should note, however, that this paper is by no means intended as a *review* of the current literature, and many important contributions go without mention.[2]

4   Maxims and scalar implicatures

Grice (1967/1989) first systematically identified and analyzed conversational phenomena along the line of the Some/All Case. Grice attempted to account for these by positing a set of maxims governing conversation. One of these maxims, Quantity, enjoins the speaker to "Make your contribution as informative as is required (for the current purposes of exchange)." Another maxim, Quality, enjoins the speaker to tell the truth. Implicatures consist, essentially, of inferences people can make about what the speaker must have meant to convey given that he was following Grice's maxims. So, in the Some/All case when we assume that Mary obeys Quantity and Quality and that Kay knows this, we seem to be able to get the right inference: Quantity enjoins Mary to say she ate all of the cake if she did while Quality prevents her from doing so if she doesn't. So if Mary says she ate some of the cake, the speaker can reasonably infer that she didn't eat all of it. This inference is the implicature.

Grice wanted to ground these maxims in human rationality rather than to merely put them forward as behavioral generalizations. He discusses the possibility that the maxims are valid because "it is just a well-recognized empirical fact that people do behave in these ways" but suggests that he would prefer to have them more solidly grounded in human rationality. Grice writes,

> So I would like to be able to show that observance of the Cooperative Principle and maxims is reasonable (rational) along the following lines: that anyone who cares about the goals that are central to conversation/communication (such as giving and receiving information, influencing and being influenced by others) must be expected to have an interest, given suitable circumstance, in participation in talk exchanges that will be profitable only on the as-

---

[2]Most notably, as my interest is in reconstructing inferences speakers can make, I do not discuss attempts to explain implicatures in terms of evolutionary game theory.

sumption that they are conducted in general accordance with the Cooperative Principle and the maxims.

Grice's views here are more nuanced than this one quote might suggest, but I think it is clear that Grice was interested in trying to ground either the maxims themselves (or the behavior they require) in human rationality and, more specifically, in the commonality of interests in speaker and audience in certain communicative situations.

Grice himself did not consider cases of *exactly* the form of the Some/All Case—though he considered very similar ones in his discussion of disjunction. Horn (1972) is widely regarded as the first person to give a systematic consideration of scalar implicatures.[3] Horn suggests that certain classes of linguistic expressions can be arrayed on a scale of strength. Examples include *some/all*, *one/two/three. . .* , and *few/none*. The critical feature of such paradigmatic scales is that the scale is ordered by logical strength: the higher members of the scale entail the lower members. So, *I ate all the cake* entails *I ate some of the cake*, *I ate three hotdogs* entails *I ate two hotdogs*, *I saw none of the students* entails *I saw few students*. This excludes, for instance, *short* and *tall* from being on a scale together since *I am tall* does not entail *I am short* or vice versa (similarly with *few* and *all*—though *a few* is on a scale with *all*).

Horn observed that the use of scalar terms systematically leads to the derivation of scalar implicatures parallel to that in the Some/All Case. This can be explained by supposing that choice between scalar terms is governed by the maxims of quantity and quality. Since Grice and Horn, the exact nature and logic of these implicatures has been developed extensively (e.g., Gazdar, 1979; Soames, 1982; Sauerland, 2004; Spector, 2006). Some linguists have recently argued that some aspects of the phenomenon are best captured by grammatical devices for generating implicatures rather than simply making reference to the Gricean maxims (Chierchia, 2004; Chierchia et al., 2012; Fox, 2006). Others have argued that the Gricean maxims, properly formalized, can capture most observed scalar implicatures (Sauerland, 2004; Schulz and van Rooij, 2006; Spector, 2006).

What I want to highlight is that in this debate over the nature of scalar implicatures, it is generally assumed by both parties that an explanation in

---

[3]There are also substantial non-Gricean—though Gricean-inspired—schools for explaining implicature such as relevance theory (Sperber and Wilson, 1986).

terms of the Gricean maxims would, all else equal, be more principled than one that makes reference to other special linguistic rules. Those who argue for special grammatical rules think the Gricean story is not capable of capturing all the relevant linguistic data. I take it the default preference for the Gricean view of implicatures gets some of its plausibility from the common belief that the maxims are not special linguistics rules, but rather follow from basic rationality assumptions. As I suggested earlier, game theory allows us to explore this assumption directly.

5   Signaling games

In this section and the follow two, I will review the basic game theory necessary for the paper. In most instances, I will try to be relatively informal in my presentation, though the details are easily found elsewhere and will be provided here when relevant. Those already familiar with basic game theory, signaling games, and standard equilibrium refinements, may skip these sections.

David Lewis's *Convention* (1969) gave the basic model of the signaling game. These games were introduced to the economics literature by Spence (1973) and have been studied since (see Sobel, 2009, for a recent review) as an important tool for modeling situations with asymmetric information.

The basic situation of a (two-person) signaling game can be described as follows: there are two players, a speaker and a hearer. The speaker has private information. We can describe this private information by saying there are different *types* the speaker might have and that the speaker's type is assigned with some probability distribution by nature. Each type of speaker has available to him the option to send a message in a set M. Each hearer, having seen the message, chooses an an actions from a set A. For each triplet of type, message and action we get a payoff U.[4]

Signaling games are *extensive-form* games, which is to say they are represented as a series of sequential moves by different players. Figure 1 gives a standard representation of a simple signaling game. This chart can be explained as follows: N represents "nature" which is taken to decide the type of

---

[4]So we can represent a (finite) signaling game as a $\{T, p, M, A, U\}$ where T is the set of types p is a probability distribution over them, M is the set of message, A is a function from messages to actions, and U is a function from triplets of types, messages and actions to pairs of payoffs from the speaker and hearer.
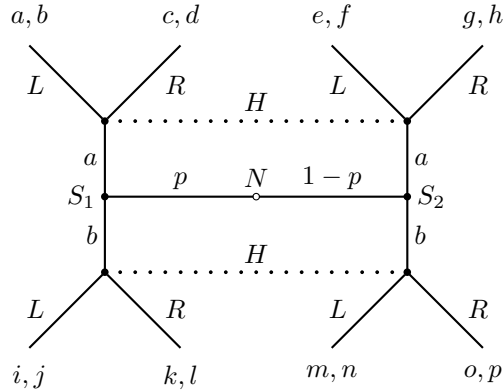
Figure 1: A signaling game

the sender S with probability p for type 1 and $1 - p$ for type 2. The speaker of either type has two message available to him a or b. The dashed line represents the information state of the hearer: indicating that he cannot tell what the speaker type is but only what signal the speaker used. L and R represent the two moves available to the hearer in any given state. The fact that the two speaker types are not connected by dashed lines means that the speaker knows which type he is. The pair of numbers at the end of each final branch represent the speaker and hearer payoff respectively. Note that we will discuss variations on this game where there are limitations on S or H's action based either on the type (in the case of S) or the message received (in the case of H).

The normal way of thinking about *strategies* in extensive-form games is as fully specific dispositions to play the game in a particular way (no matter how the other person plays). So, for example, in the game above we can identify the following four *pure* (i.e. non-probabilistic) speaker strategies: $1a2a$ (S_1 plays a and S_2 plays a), $1a2b$ (S_1 plays a and S_2 plays b), $1b2a$, $1b2b$. The pure hearer strategies, on the other hand, are responses to the different signals: $aLbL$ (H plays L in response to a and L in response to B), $aLbR$, $aRbL$, $aRbR$. Note that hearer strategies cannot depend on the type of speaker because the hearer does not have that information available to him. *Mixed* strategies are ones in which responses are chosen with a probability distribution rather than deterministically. We can also talk about the expected payoff for a pair of strategies in the normal way.

A natural question is how extensive-form games, such as signaling games,

relate to the standard representation of games by tables, such as the prisoner's dilemma game represented in figure 2. We call representations these tabular

	C	D
C	3, 3	1, 4
D	4, 1	2, 2

Figure 2: Prisoner's Dilemma

representations, *normal* or *strategic* forms. Corresponding to each extensive-form game is a normal-form game, in which we think of each player as (at once) choosing complete pure strategies. I will give an example of a normal-form representation of an extensive-form game in the next section. For a signaling game the normal form can be given as a mapping from all the possible combinations of pure strategies to the *expected* payoff (given the probability distribution over states of nature).[5] Note that the correspondence between extensive-form games and normal-form games is not one-to-one: for any normal-form game there are many extensive-form games compatible with it.

6   Nash equilibria and refinements

Game theory provides various *solution concepts* for assessing how players should play games. The most standard solution concept is the Nash equilibrium. A little notation: let $U_i(s_1, S_2)$ be the expected payoff to player i of playing strategy s_1 against an opponent who plays strategy s_2. A Nash equilibrium is a pair of strategies for player 1 and player 2, (s_1, s_2), such that for all strategies s_1' possible for player 1, $U_1(s_1, s_2) \geq U_1(s_1', s_2)$ and for all strategies s_2' possible for player 2, $U_2(s_1, s_2) \geq U_2(s_1, s_2')$. In other words, a Nash equilibrium is pair of strategies that neither player has an incentive to deviate alone from. In the prisoner's dilemma the only Nash equilibrium is the one where both players defect: there is an incentive for both players to deviate together but not for just one to.

In extensive-form games with sequential moves, such as the signaling game there are Nash equilibria that are not very plausible. To see, this consider the simple multi-stage game in figure 3. In this game the first player chooses be-

---

[5]Alternatively, but less standardly, we can treat nature as another "player" also and have a three-person game with its normal form.
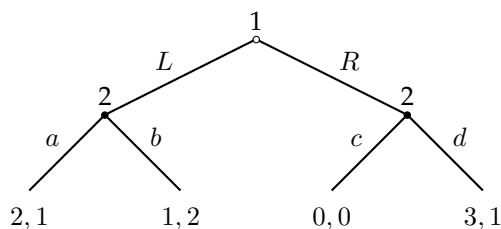
9

Figure 3: Extensive-form game

tween L and R, so he has two pure strategies (L and R), whereas the second player has a different choice of move depending on what the first player has done. Thus, there are four full, non-probabilistic strategies for him: choose a in response to L, and c in response to R (which we'll write $LaRc$), $LaRd$, $LbRc$ and $LbRd$. We can use this listing of strategies to give the normal-form representation of the game in figure 4. The strategy pair in which player 1 plays

	$LaRc$	$LaRd$	$LbRc$	$LbRd$
L	$2,1$	$2,1$	$1,2$	$1,2$
R	$0,0$	$3,1$	$0,0$	$3,1$

Figure 4: Normal form of game in figure 3

strategy L and player 2 plays $LbRc$ is a Nash equilibrium. Note, however, that if this equilibrium is played, player 2 will never have a chance to play c in response to R. However, the strategy pair being a Nash Equilibrium depends on this disposition on his part, as otherwise player 1 would prefer to play R. It seems that, in fact, were player 1 to play R, player 2, if he were rational (in the sense of utility maximizing), would have to play d in response to maximize his payoff. However, if player 1 knows that, and is himself rational, then he will know that he can maximize his payoff by playing R. So while $(L, LbRc)$ is a Nash equilibrium, it is not one which we should expect rational agents to play.

There are a set of refinements of the Nash equilibrium concept designed to deal with this inadequacy. Indeed, there are different refinements for games with perfect information such as the one in figure 3 from those for games with imperfect information such as signaling games. Since we will focus on games of imperfect information, I will only present refinements that apply to them.[6]

---

[6]Thus I will not discuss subgame perfection, the most common refinement to deal with games like that of figure 3.

What we can do is ensure, relative to a pair of strategies, rationality of the strategy at every point where a player *might* make a decision. The most standard equilibrium refinement for games of imperfect information that captures this sequential rationality is the perfect Bayesian equilibrium (PBE). As one might guess from the name, a PBE is modeled on the idea that both players are Bayesian-updaters trying to maximize expected utility. When a strategy is a PBE it means, essentially, that it could be played by two players exhibiting such Bayesian rationality who each believe the other is playing the equilibrium strategy. Informally, a pair of strategies is a PBE if at every decision point, each move at that point maximizes expected utility relative to the credence function of the player, which itself is constrained by the belief that the other player is playing the PBE and whatever other information is available at that point given the game structure. If the belief that the other player is playing the PBE is not possible at a decision point, then the only condition put on the players move is it maximize utility relative to *some* belief about how the other player is acting. In the game in figure 3, the move Rc fails to maximize utility on *any* beliefs and so is ruled out.

We can give the definition of a PBE for signaling games more explicitly. There are two conditions for a pair of strategies, (s_S, s_H), to be a PBE: one condition on the speaker strategy and one condition on the hearer strategy. The speaker strategy must simply be such that for each type it maximizes expected utility on the assumption that the hearer plays s_H. The hearer strategy must meet a more complex condition. We assume the hearer starts (before the game play) with a graded belief about the speaker's type based on the probability distribution associated with N's move. We then assume that in response to each message the hearer updates his belief about the speaker's type in accordance with Bayes' rule based on the assumption that the speaker is playing s_S if this is possible. If this is not possible, we simply assume he forms some consistent belief about how the hearer will play. The condition on the hearer's response to a message m is that there is a credence function satisfying the conditions above on which the hearer's response maximizes his expected utility. Essentially this is two separate conditions: for so-called "on the path" messages—messages that given s_S and the game structure, the speaker has a positive probability of receiving—Bayes' rule determines hearers credences after receiving the message and the equilibrium condition is that the response

is rational on for those credences. For so-called "off-the-path" messages—
messages the hearer would assigns probability 0 to given the game structure
and that the speaker is playing s_S—the hearer's response must be optimal for
some belief about the speaker's type that is consistent with the hearer receiving
an off-the-path message. Refinements of the PBE typically take the form of a
restriction on beliefs in response to off-the-path messages.[7]

7 Rationality and dominance

The equilibrium notions discussed above are just that: pairs of strategies that
are self-reinforcing. Once you know you are in an equilibrium you don't have
any incentive to get out of it. However, as many have noted, when we think of
games as one-off events without prior coordination these notions tell us little
about how players should or will play: for in these cases the players cannot
be assumed to have gotten into an equilibrium already. There are two related
problems: 1) Even if there is only one Nash (or perfect Bayesian) equilibrium
available in a game we have not explained why rational agents should play
it. 2) Even if we could argue that rational agents should only play strategies
that form part of some Nash equilibrium, most games provide more than one
equilibria so we still are left with the question of how players should choose
an equilibrium.[8]

In many cases, there simply is nothing to say about what rational agents
should do. As an example consider the simple coordination game in figure 5.
In this game, the players have a mutual interest in meeting at one of the two
pure Nash equilibria: the upper left (AC) and the lower right squares (BD). It

	C	D
A	1, 1	0, 0
B	0, 0	1, 1

Figure 5: Coordination game

should be obvious that there is no privileged rational way to play the game.

To summarize: Nash equilibria (and their refinements) are strategy pairs
that one player does not want to *deviate* from if he knows the other player will

---

[7]Cho and Kreps (1987) give a notable example of such a refinement.

[8]There is a large literature on equilibrium selection (most notably, Harsanyi and Selten, 1988),
but the considerations there do not obviously connect up with simple considerations of rationality.

play his part. The existence of a Nash equilibria does not mean a player *should* play it: after all there can be multiple Nash equilibria. In certain cases, equilibrium refinements may help eliminate some equilibria that are in some way defective, but in symmetric situations like figure 5 it follows a priori that equilibrium concepts will not dictate a unique way rational players should play.

There are, also, clear cases where we can say how rational agents should play the game. To delineate these cases there is a tradition of using non-equilibrium notions, such as dominance reasoning to capture the ways in which simple rationality might dictate play. Dominance relations, by definition, obtain between strategies in a normal-form representation. A given strategy, s, for player, i, *strictly dominates* another strategy, s', if no matter what s guarantees i a strictly higher payoff than s'. By *no matter what*, I mean no matter what strategy the other player chooses and no matter what the state of nature is. We can reasonably assume that rational agents will not play strictly dominated strategies, based on the simple idea that rational agents want to maximize expected utility. The notion of strict dominance alone, in some cases, determine how rational agents will play a game. For instance, in the prisoner's dilemma, since defection is a strictly dominant move for both players, we expect rational agents to defect.[9]

Another form of dominance reasoning that will be particularly important for us is *weak dominance*. A given strategy s for player i *weakly dominates* another strategy s' if s guarantees i at least as high a payoff as s' and for some state of affairs (i.e. some possible opponent strategy and/or state of nature) s gives i a strictly higher payoff then s'. Consider, for instance, the game in figure 6. Here,

	C	D
A	1,1	1,0
B	1,1	1,1

Figure 6: Simple game

it might seem that a rational player 2 will choose C over D as that guarantees him at least as high a payoff as D no matter what player 1 does and a higher pay off in one case. Note that weakly dominated strategies, unlike strongly

---

[9]As a note, it is easy to show that if there is a strategy pair determined by strict domination it is the unique Nash equilibrium. Also it's worth noting that we can view survival of rounds of dominance reasoning as an equilibrium refinement. However, I call these non-equilibrium notions since they don't depend on considering an equilibrium.

dominates strategies *can* form part of a Nash equilibrium (as well as PBEs). In the game in figure 6 there are three pure Nash equilibria: $(A, C), (B, C), (B, D)$ and the last of these includes the weakly dominated move D.

In both normal-form and extensive-form games we can also consider *iterative* applications of dominance reasoning. The informal idea is this: player 1 and player 2 are both rational and both know they know this, and know they know they know this, and so on. In other words, they have common knowledge of rationality. In this case, it would seem, we can iteratively eliminate dominated strategies: i.e. in player 1's own strategic reasoning he can assume that player 2 will not play any strictly dominated strategies, and evaluate dominance on *that* assumption. A simple example is in figure 7. Note that no move

	C	D
A	1, 1	2, 2
B	2, 2	1, 3

Figure 7: Solvable by iterative dominance

is dominant for player 1 without any assumptions about player 2's behavior. However, for player 2, D strictly dominates C. If player 1 knows that player 2 will not play strictly dominated strategies, then, on this assumption, move A strictly dominates B for player 1.

Iterative dominance reasoning, in general, works as follows: we eliminate a strategy or move based on the fact that it is dominated by another move or strategy. Then we update our understanding of the game to reflect this elimination by considering a *new* game in which the dominated strategy or move is not allowed. We continue until there are no more dominated strategies in the game we have left.

With respect to weak dominance such iterative reasoning can be problematic. Iterative weak-dominance reasoning is possible, but in many cases the result of such reasoning depends on the order in which it is done. Take, for, example the case in figure 8. Here different orders of elimination of weakly

	L	R
A	2, 3	3, 3
B	1, 0	0, 1
C	0, 1	1, 0

Figure 8: Order of weak dominance elimination matters

dominated strategies result in either (A,L) or (A,R) being the strategy pair left after elimination. What this shows is that merely demonstrating that some strategy pair is be reached by a chain of iterated dominance reasoning *cannot* be an argument in itself that rational players should play the strategies in the pair. Despite this, there are some results that in a restricted class of games the order of iterated weak dominance does not affect the outcome (at least in terms of payout).[10] In particular, in games with identical payoffs the order of iterative elimination by weak dominance does not affect strategic reasoning. This will be useful for us, as all the games we will consider have identical payoffs for both players. Another way of ensuring that iterated weak-dominance reasoning results in a unique result is to ensure that at each stage of reasoning one eliminates *all* strategies/moves that are weakly dominated. This special form of iterative weak dominance is called *iterative admissibility*.

Given that the order of eliminating strategies matters, it cannot be a requirement on rationality that players do not play iteratively dominated strategies (for this could eliminate all strategies, as in 8). However many authors, most influentially Kohlberg and Mertens (1986), posit that rational players should not play a strategy unless it survives *some* complete series of iterative elimination of weakly-dominant strategies.[11] However, this position is far from universally held by game theorists.

Iterative-weak dominance reasoning is a much more powerful tool than strict dominance reasoning for extensive-form games. Consider, for instance, the extensive-form game in figure 3. Here there is not sufficient *strict* domination of any strategy over another to determine uniquely what we think of as rational play. For instance, the strategy for second player of playing b if the first player plays L and c if the first player plays R is not strictly dominated. For this strategy will do just as well as playing b and d as long as the first player plays L. Nonetheless $LbRd$ weakly dominates $LbRc$ since there it always does as well, but sometimes does better. Thus, in this game, as in many extensive-form games, weak-dominance reasoning is necessary to single out rational lines of play. However, while in the game in figure 3 weak dominance reasoning might seem attractive in other games, such as the centipede game

---

[10]The earliest such result (which is good enough for our purposes) is Rochet (1980) (see also Marx and Swinkles, 1997).

[11]By 'complete series' I mean a series of eliminations at end of which no further elimination by weak dominance is possible.

Rosenthal (1981) weak-dominance reasoning gives very unintuitive results as it validates backward induction. This and other consideration has led many to think weak-dominance reasoning should not be taken to guide rational play.

As a note, in extensive-form games, we can speak of a move rather than a strategy being weakly dominated, by which we will mean that every strategy that includes that move is weakly dominated. For instance, in the game in figure 3 player 2's moves a and c are weakly dominated, since any strategy that includes those moves will be weakly dominated.

What I have done here is present two related ways of reasoning about games *in absence* of an equilibrium: iterative strict and weak dominance. Both of these ways of reasoning about games apply to extensive-form games through their normal form.[12] On the surface, both seem plausible as ways rational agents *should* think about games: after all, why should one ever play a dominated strategy?

However, spelling out a formal concept of rationality and common knowledge of rationality that justifies these forms of reasoning (in both extensive and normal-form games) is a non-trivial task. Since the eighties many game-theorists have tried to spell out constraints on rational strategic reasoning that capture *inter alia* these forms of dominance reasoning.[13] I will not review this extensive and complex literature on the *epistemic foundations* of game theory.[14] However, it is worth noting that conceptions of common knowledge (or belief) of rationality that require players to play only those strategies reached by iterative elimination of strictly dominated strategies have long been known, while conceptions of common knowledge of rationality the force iterative elimination of weakly dominated strategies have, more recently, been explored (Brandenburger, 2007; Brandenburger et al., 2008). For our purposes it will be enough to say that iterative dominance reasoning (in both forms) is arguably a form of rational inference in games where common knowledge of rationality obtains.

---

[12]Of course, strictly speaking reduction of games by strict and weak dominance are just algebraic ways of reducing games, I call them methods of reasoning as they closely connect to intuitive ways of thinking of a chain of reasoning.

[13]The starting point is the notion of rationalizable strategies independently proposed by Pearce (1984) and Bernheim (1984).

[14]See Battigalli and Bonanno (1999) and Brandenburger (2007) for introduction to program.

8 Game-theoretic account of implicatures

The goal of this paper is to provide an analysis of the communicative situation in which scalar implicatures occur that explains the derivation of the implicatures as a sort of rational inference. The hope would be that, at least for idealized cases, we could view the speaker's following of the Gricean maxims as well as the hearer's derivation of the scalar implicatures as simply rationally compelling behavior, given the set-up of the game. If we succeed, we show why in idealized but still useful models rational agents should behave as the Gricean maxims dictate. If we can show this then we need not view the maxims as mere useful empirical generalizations but rather as generalizations about what how rational agents should act—at least for idealized cases.

Any attempt at carrying out this project will include two elements (not always entirely separate): a model of the communicative situation (i.e. the game itself) and an analysis of why rational agents should choose the Gricean reasoning. The plausibility of such a model will depend on both these components.

The model itself is meant to describe in a formally tractable way, the communicative problem involved in using scalar terms. To do this, we need to abstract away from much of the complexity of real-world communication. Thus, the models we use will be idealized and simplistic. This itself is not a problem if the idealizations and simplifications made do not change the basic structure of the communicative problem we are focusing on. Modeling is a delicate art because the space of possibilities is so wide. It's not interesting to find *some* model that vindicates Gricean reasoning, one needs to show that that the model is, at least, a simple, plausible representation of the real-world situation.

In this section I will discuss some basic components of the scalar implicature situation that a game-theoretic account should include and give some indications about how signaling games can capture these.[15]

8.1 Asymmetric information

The classic Gricean situations include asymmetric information: the speaker knows something that the hearer does not. Signaling games are a natural way of modeling this asymmetry, of course, which is why they are often used to

---

[15]Here I am much in debt to the extensive literature on game-theoretic pragmatics, in particular Franke (2009).

model language in a game-theoretic setting.

8.2 Cooperativeness

Gricean reasoning works by assuming that the speaker and hearer want to help each other, and they both know this. The natural way of modeling this in a game-theoretic setting is to say that the *payoffs* for both players are aligned. Games with this property are called games of *coordination* (Schelling, 1960; Lewis, 1969).

8.3 Relevance

It is well known that scalar implicatures are only possible when the differences between the different states of affairs is *relevant* to the hearer. The most natural way to model such relevance is to have the payoffs dependent on the hearer action in such a way that the hearer benefits from knowing the private information of the speaker. If we are interested in pure communication then the hearer "action" is only a sort of nominal aspect of the game intended to capture the fact that it is useful for the hearer (and hence the speaker) for him to know what the speaker knows.

8.4 Background beliefs

It is standard in game theory to assume that speaker and hearer have common knowledge of the structure of the game. This will be a useful assumption for us to incorporate as well: we need to assume that the payoffs are identical for both players and that they know this (and that they know they know this...).

8.5 Message costs

We assume that there are not high costs for uttering one sentence rather than another: after all, different sentences require only slightly different muscle movements. In game theory, a signaling game with no message costs is called a *cheap talk* game.[16]

---

[16] The classic paper on cheap talk games is Crawford and Sobel (1982), see also Farrell and Rabin (1996).

8.6 Meaning

The most vexed issue about modeling implicatures using signaling games is how to build in the Gricean account of the literal meaning of sentences. On the Gricean view, the sentence *I ate some of the cake* is literally compatible with eating all of the cake. It is only once the implicature is drawn that it comes to have the stronger meaning which is incompatible with eating all of the cake. In order to explain this inference, we need to provide a model in which we can say that the sentence has this weaker literal meaning. This is a considerable theoretical challenge, as we will see.

Standardly in the economics literature on signaling games it is assumed that all messages (in a cheap talk game) are "inherently" meaningless signals which only get meaning in the context of an equilibrium. This view is largely inherited from Lewis's seminal treatment of signaling games in *Convention*.

What I will call a Lewis signaling game is *very* similar to the ones we want to discuss. Lewis also assume costless messages and identity of payoff between speaker and hearer. However, Lewis focused on the question of how messages get meanings to start with, not the question of how, *given the message meaning*, extra inferences can be inferred. Lewis's suggestion is that meaning derives from the use of messages in certain repeated signaling games where the same equilibrium was repeatedly played. Take for instance a simple Lewis signaling game like that in figure 9. There are the two *separating* equilibria where the
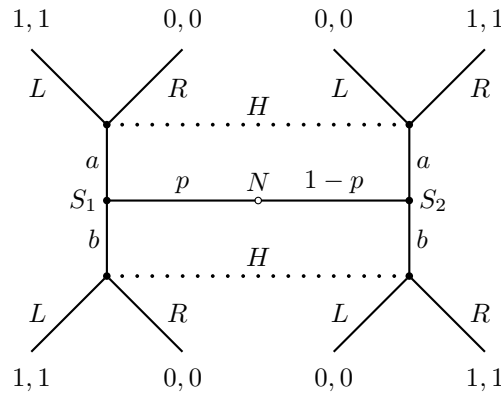
Figure 9: A Lewis game

speaker plays either $1a2b$ or $1b2a$ and the hearer responds either $aLbR$ or $aRbL$, respectively. These equilibria result in real communication, in Lewis's picture.

(Other equilibria such as the *babbling* equilibria where speakers choose signals randomly and *pooling* equilibria where speakers always use the same message do not result in any communication in the usual sense.) Within each separating equilibria we can speak of the meaning of a or b, but the notion of the meaning of the signals is only defined relative to the equilibria that is being played.[17] The meaning of a signal in an equilibrium is just defined by what types of speakers uses that signal (or what actions it induces, if the meaning is viewed imperatively). This is the notion of meaning most often implicitly or explicitly assumed in the game theory literature.

That this conception of meaning will not be adequate in a model of scalar implicatures should be obvious. The entire point of the Gricean reasoning is that with implicatures there is a divergence between how a sentence is used and what its literal meaning is. So defining the meaning of a term by how it is actually used in the game model will preclude the possibility of capturing scalar implicatures within the model.

The only way we can use Lewis's equilibrium conception of meaning is by considering an equilibrium as a sort of *starting point* and thinking of implicatures as rationally motivated deviations from the equilibrium. In a sense, this is exactly the approach taken in much of the game theory and pragmatics literature such as Benz (2006); Benz and van Rooij (2007); Franke (2009, 2011), theories which I will discuss later. I think there is some promise to this approach, but it is conceptually difficult. If people do not play some according to some equilibrium in the end, and if it can be derived by some sort of reasoning that they will not, then it is irrational to take as one's starting point in rational deliberation the proposition that they will play this way. Franke (2009, 2011) embraces the irrationality and claims his model is a model of *bounded* rationality. This is an interesting tact, but it is a bit strange to think that we can only model literal meaning in a model of bounded rationality as it would be desirable to model meaning in a way that does not constitutively depend on bounded rationality assumptions.[18]

One obvious way of modeling meaning is by the constraint that speakers can only use messages when they are literally true (Parikh, 1991, 2001). To

---

[17]Lewis was essentially trying to give a game-theoretic reconstruction of Carnap's notion of truth-in-model.

[18]Of course, Franke's motivation is partly to capture non-cooperative uses of language, which cannot be captured in the way outlined below.

effect this we can use non-standard signaling games where speakers are restricted to use only those messages that are literally true given their type. For examples supposing in the Lewis game in figure 9 that the signal a literally means the speaker is in state 1, whereas the signal b literally means the speaker is in state 1 or in state 2. If it is common knowledge that speakers only use messages that are literally true then we can model this situation with the game in figure 10. The critical point is that the meaning assumptions are not strong
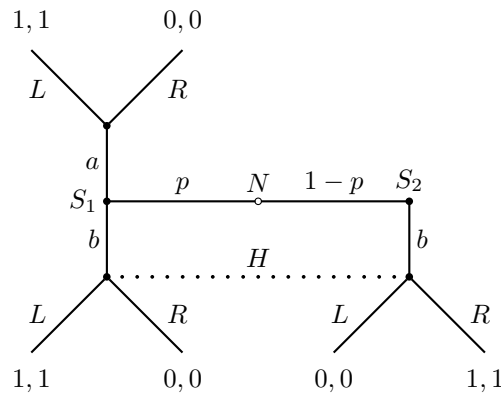

Figure 10: Lewis game with built-in meanings

enough to fully determine speaker actions. Thus, how the speaker acts within the confines of these assumptions may allow the hearer to make strategic inferences.

Restricting speakers to send only true messages is certainly the most natural and most common way of treating Gricean literal meaning in a game-theoretic setting. However, there are some problems with using this technique. One is that it seems to exclude from the start the idea that speech may be either non-literal (i.e. metaphorical in some way) or intentionally deceitful. This does not seem like a serious criticism. The game is a *model* of how a speaker and hearer conceptualize their situation. This does not mean that every assumption in the model needs to be considered an unrevisable assumption of the speaker or hearer. The speaker or hearer do not need to *always* conceptualize speech-act situations as constraining speech to literal meaning, but for the purposes of scalar implicatures (where speakers *are* assumed to speak truly, in the standard Gricean model) this seems like a natural assumption. Criticizing this model for not allowing for non-literal or strategically deceitful utterances is simply

21

criticizing the model for not doing something it is not intended to do.

Franke (2009, pp. 35–37) motivates his choice not to stipulate truthful utterances in an account of scalar implicatures by two further arguments. First, he argues, for reasons we will see below, that there is no unique equilibrium requiring Gricean play in these models. So these models simply are not structured correctly to account for scalar implicature. Most of this paper is concerned with trying to demonstrate that we can make a good argument for Gricean behavior in this model. Besides the methodological problem this argument has of assuming that we *should* be able to derive Gricean play in our model, I also think this point is substantively wrong. The main argument of this paper is that such models do allow derivation of Gricean play by using weak-dominance reasoning, rather than standard equilibrium refinements.

Second, Franke (2009) suggests that there is a conceptual problem with this model of literal meaning. He writes as follows:

> I can very well say whatever I like, whenever I like to whomever I like. I may have to face social or even legal consequences from time to time, but it is not as if the semantics of my language restricts the muscles of my jaw and vocal track, regulating what I possibly can and what I cannot utter.

I think this argument is not compelling: Game models surely do not need to provide moves corresponding to all physically possible actions. We use models to capture players *assumptions* about how certain situations are structured—what reasonable possibilities players consider. Even if the assumptions turn out to be false in some instances it does not mean that speakers and hearers do not make them.

This is not to say I think that forcing speakers to say only true messages is the only way to model natural language meanings in a game-theoretic setting. The literature on *credibility* in game theory provides some interesting other possibilities (Crawford and Sobel, 1982; Farrell, 1993; Rabin, 1990; Stalnaker, 2005). I hope to address these other approaches and their relations to scalar implicatures on another occasion.

8.7 Rationality

Besides the model itself, a game-theoretic grounding for scalar implicatures will include an argument that rational players in the situation modeled should play in the Gricean way. The plausibility of such an argument depends on the use of appropriate tools. The game theory literature contains a plethora of equilibrium refinements and more ad hoc refinements can be invented. Merely showing that there is *some* refinement which justifies Gricean behavior does not show much about the *rationality* of Gricean reasoning. We do not necessarily need to use standard game-theoretic tools to explain the rationality of Gricean implicatures, but whatever tools we use, we need to convincingly argue for their appropriateness.

8.8 Robustness

Whatever model of the game and of player rationality we use its plausibility and explanatory value depends on it lacking arbitrary restrictions. A robust model should not depend, in order to get results, on relatively arbitrary assumptions. For instance, a signaling game model that depended on a very specific payoff structure to yield the Gricean result would not provide a robust account of scalar implicatures. Similarly, very strong constraints on the reasoning patterns of players will not give a plausible grounding of Gricean reasoning.

Many of the game-theoretic derivations of implicatures that one can find in the current literature are not robust against small changes. For instance, as I will argue, some versions of the Iterated Best Response model and the Optimal Answer model need strong assumptions about players beliefs about the probability of various states in the game.

8.9 Relation to non-linguistic reasoning

Game theory was developed to deal with strategic interaction generally rather than language use. Indeed, systematic treatment of linguistic communication in mainstream game theory is a relative recent phenomenon. One hope in using game theory to model pragmatic inferences is to relate the underling reasoning driving these inferences to reasoning in non-linguistic cases. (This idea

of relating pragmatics to non-linguistic behavior was also one of Grice's major goals.)

The explanatory value of a model of scalar implicatures, thus, depends to some degree on whether it is sufficiently general to also capture non-linguistic behavior with similar structures to that of pragmatic inferences. If we can do this, we can provide an argument that pragmatic inferences are grounded in general practical rationality rather than some specific linguistic mechanism. I will argue, in section 17 that the particular reasoning I use here extends to parallel cases from the economics literature.

9   Simple Gricean game

In this section I present and discuss a very simple model of the Some/All Case in a game-theoretic framework that is meant to capture its essential features in accord with the principles about modeling given above.[19] I then show—as is well known in the literature—that the standard equilibrium concept for signaling games, perfect Bayesian equilibrium (PBE), fails to single out the Gricean strategy in this game.

In this basic model, nature determines whether we are in a *some* situation or an *all* situation with a certain probability distribution. Intuitively we think of the *some* situation as the one where the speaker ate just some of the cake and knows it and the *all* situation as the one where he ate all of the cake and knows it. If the speaker is in the *some* situation (he is of type S_s) then he can and must send the message m_s. If he is in the *all* situation (S_a) then he can send either the message m_s or m_a. It follows that if S sends message m_a then H will knows the speaker-type directly by knowledge of the game structure, but if S sends message m_s then H does not know the speaker type directly, but rather must infer it. If the hearer makes this inference then this corresponds to deriving the scalar implicature. Figure 11 represents this partial specification of the game.

We have not yet specified what H does once S has sent a message, nor what the payoffs are. How we do this will affect our assessment of the rationality of different strategies. One natural way of modeling the fact that the S's type matters to H is to suppose that the H makes some sort of choice after hearing

---

[19]This basic model is very similar to Parikh's original model, though he chooses not to use the standard signaling game to describe the situation (Parikh, 1991).
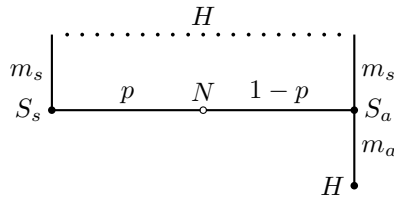
Figure 11: Simple Gricean game

S's message, and that H wants to make the choice one way if S is of type S_s and a different way if S is of type S_a. We assume a cooperative situation in which the speaker and hearer's interests are aligned so that the payoffs are identical. I represent this situation with the complete extensive game in figure 12.
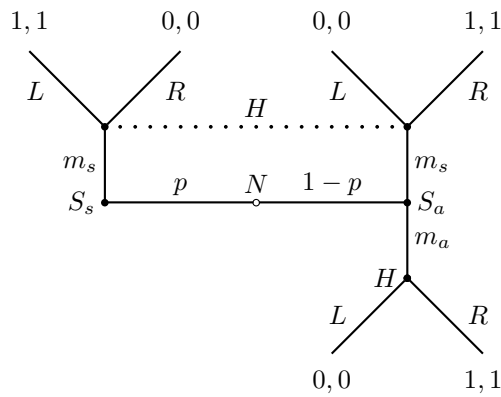

Figure 12: Simple Gricean game with hearer response

I will call strategies *Gricean* when they intuitively accord with Grice's maxims and Grice's posited implicatures. The Gricean strategy profile for this game is clear: for S it is to send m_S when in state S_s and m_a when in state S_a, for H it is the best response to this, i.e., to play L in response to m_s and to play R in response to m_a. This combination of strategies will ensure payoffs of 1 in each play, the best that can be hoped for, as indicated by the normal-form description of the game in figure 13. Since both players are guaranteed their highest possible payoffs if they play this pair of strategies, the pair is a PBE.

Unfortunately, the Gricean strategy pair is not the only PBE. Suppose $p \leq .5$.

	m_sLm_aR	m_sLm_aL	m_sRm_aL	m_sRm_aR
$S_sm_sS_am_a$	1	p	0	$1-p$
$S_sm_sS_am_s$	p	p	$1-p$	$1-p$

Figure 13: Normal form of Gricean game

Consider the *pooling* strategy for both players: S sends m_s no matter what and H responds R no matter. It is easy to see that this pair of strategies is also a PBE: neither type of S has any incentive to change his behavior (type S_a gets his maximum payoff, whereas S_s has no other options) and H, getting no usable information from the S, is strictly maximizing his *expected* payoff by playing R. For similar reasons, the pooling equilibrium also satisfies many standard equilibrium refinements such as the Intuitive Criterion of Cho and Kreps (1987).[20]

I think it can be safely said that at an intuitive level the Gricean strategy seems compelling. It seems like *even without prior discussion* either player can safely assume the other player will play according to it. The question is how we cash out the intuitive rationality of the strategy by means of plausible game-theoretic tools. All we have seen so far is that the standard equilibrium concepts cannot do this.[21]

10   Dominance arguments

We saw above how the Simple Gricean Model captures the basic conversational situation in which the speaker chooses between 'some' and 'all'. However, the standard equilibrium notions do not single out the Gricean strategy. Here I'll show that iterative weak-dominance reasoning does single it out.

If we look at the normal-form representation in figure 13 we can immediately read off the weak dominance relations. Note first that any hearer strategy which includes the response of L to m_a is weakly dominated. If we eliminate these two hearer strategies we get the normal form representation in

---

[20]Depending on the probabilities their can also be babbling PBEs in this game: one's where speaker S chooses messages randomly (when he has a choice) and H ignores the message.

[21]This conclusion might seem in tension with van Rooij's claims that Gricean behavior in various games arises because it is the only Nash equilibrium (van Rooij, 2009; de Jager and van Rooij, 2007). Van Rooij's work uses substantive assumption about the structure of the games beyond the ones here. For this reason I do not find his results robust enough to support the conclusion that scalar implicatures are *generally* derived because Gricean behavior is the unique Nash equilibrium.

figure 14. Here each player has two choices of strategies: the Gricean strategy and the pooling strategy. In this reduced game the speaker's pooling

	m_sLm_aR	m_sRm_aR
$S_sm_sS_am_a$	1	$1-p$
$S_sm_sS_am_s$	p	$1-p$

Figure 14: Normal form after first removal of dominated strategies

strategy ($S_sm_sS_am_s$) is weakly dominated by the speaker's Gricean strategy ($S_sm_sS_am_a$). So we can eliminate that pooling strategy to get the game in figure 15. A final application of strict dominance eliminates the hearer's pooling

	m_sLm_aR	m_sRm_aR
$S_sm_sS_am_a$	1	$1-p$

Figure 15: Normal form after second removal of dominated strategies

strategy, and we are left with the Gricean speaker strategy.[22]

It is not very intuitive to think of Gricean games by way of their normal form: we normally think of players choosing moves as they go rather than opting for total strategies. For this reason it will be helpful to redo the reasoning using the extensive-form representation. We can think of each of our applications of weak dominance as an elimination of one possible move from the extensive form. We start with the representation in figure 11. Since the move L in response to message m_a will always result in a lower payoff it can be safely assumed that H will not make that move. This gives us the reduced game tree in figure 16. In this tree, the speaker in state S_a will guarantee himself the highest payoff by sending m_a, so this move is weakly dominant. If we adjust the tree to reflect this we now get the game in figure 17. In this tree a rational H can only play L in response to m_s, so we can conclude that the full Gricean strategic behavior is what we should expect. Intuitively, these successive removal of weakly dominated moves are justified on the assumption that it is common knowledge that both players will not play weakly dominated moves.[23]

---

[22]Note that the dominance argument does not depend on choosing the normal-form representation of signaling games in which nature is factored in by expectations rather than treating it as a separate player. The same reasoning works if we treat the Simple Gricean game as a three person game with nature being the third player, however in this version the last step is another instance of weak dominance not strict dominance.

[23]As I noted earlier, actually modeling this kind of common knowledge of rationality in a explicit way is difficult. Brandenburger et al. (2008) discuss how to overcome these conceptual problems.
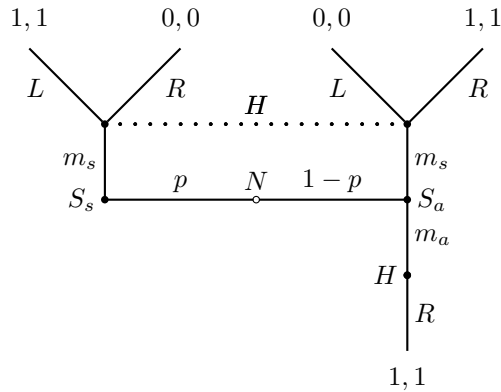
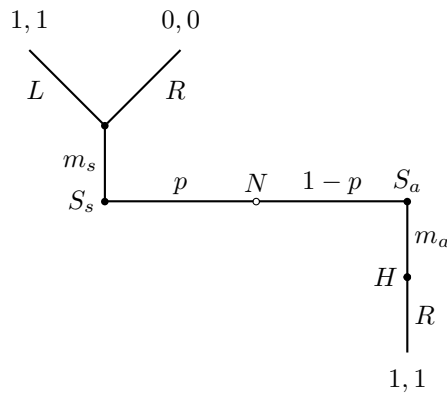Figure 16: Simple Gricean game after first removal of dominated strategies

Figure 17: Simple Gricean game after second removal of dominated strategies

To summarize: for the Simple Gricean game, iterative elimination of weakly dominated straggles leaves only the Gricean strategy pair. So, in this model it seems that we do not need to stipulate Gricean maxim over and above player rationality. In the next two sections, I discuss and criticize alternative ways of capturing scalar implicatures from the recent game theory and pragmatics literature.

11   Pareto-Nash equilibrium and payoff dominance

In his well-known work on game theory and pragmatics Parikh (1991, 2001) argues that we should capture Gricean reasoning using the Pareto-Nash-Equilibrium solution concept. That is we should assume that rational agents (in these sorts

of games) will only play strategies that are on Nash equilibria which provide at least as high payoffs as any other Nash equilibria.[24] In the Simple Gricean game there are only two perfect Bayesian equilibria the Gricean one and the Pooling one. The Gricean one gives higher expected payoffs then the pooling one so it is the only Pareto-Nash Equilibrium.

The critical question here is whether choosing Pareto-Nash Equilibria are really a legitimate constraint on equilibrium choice. Parikh notes that without such constraints we cannot explain obviously compelling behavior in simple games with coordinative pay-off structure. For example consider the coordination game in figure 18. It is arguable that two rational agents playing this

	C	D
A	1000, 1000	0, 0
B	0, 0	1, 1

Figure 18: Lopsided coordination game

game would choose the A and C equilibrium.[25] Standard equilibrium refinements or conceptions of rationality cannot capture this.

It is not clear, however, that small payoff-dominance has such a stark effect. For instance, if the difference is just a small one, it is not clear that we should argue that rationality compels players to choose that one: the goal of each player is just to play what the other plays. In most cases, payoff dominance makes one solution *salient* in the sense of Schelling (1960) and Lewis (1969), and since it also pays more it is *natural* way to play (and to assume that the other play will too). The Gricean strategy is also salient in being the unique Pareto-Nash equilibrium in the Simple Gricean game (in addition to whatever other salience it may have) its salience may well supports its choice as an equilibrium. But the conclusion that the Gricean strategy is played because it is salient, would not seem to be a very satisfying account of its basis in *rationality*, and in game theory it is rarely assumed that rational players only play strategies that are part of payoff dominant equilibria.

---

[24]In other words, in the terminology of Harsanyi and Selten (1988) rational players never choose *payoff-dominated* equilibria.

[25]Of course, I am assuming that the ratios between payoffs rather than just the strict ordering matter, not a standard assumption. Regardless I am just making the intuitive point that where there are great differences in payoffs the Pareto-Nash equilibrium seems quite compelling.

12 Reasoning-based accounts

For comparative purposes, in this section, I will discuss the treatment of the simple Gricean game by the reasoning-based strategies. These strategies, which include the Iterated Best Response models of Jäger (2007) and Franke (2009) and the Optimal Answer model of Benz (Benz, 2006; Benz and van Rooij, 2007), take as inspiration the idea that a certain form of reasoning may lead to Gricean behavior. They differ from the treatment above in that less needs to be built into the game structure itself and more into the assumptions about how the players will play the game.

12.1 Iterated Best Response models

The idea behind Iterated Best Response models is that strategic reasoning involves a hierarchy of increasingly sophisticated thinking terminating at the point where further strategic sophistication is otiose, i.e. a fixed point, if such a point can be reached. In the game model itself we make no constraints based on the literal meaning, as in figure 19. This game is, thus, a simple coordination
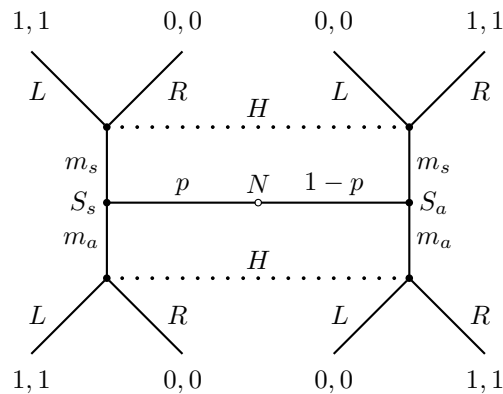

Figure 19: Model with no built-in meaning

game. Literal meanings instead of being part of the game structure, go rather into the reasoning of the players.[26]

I will analyze a variation of Franke's model.[27] Here is a simplified version

---

[26]This feature is for the cases we consider here inessential: We could use a more articulated game structure, forcing truthfulness, and start the chain of reasoning within that structure.

[27]The reader should consult FrankeDiss, FrankeQuantity for more details. Franke's model and Jäger's model are quite similar, though Jäger intended his model as an evolutionary one, whereas

Type	Strategy	
S_0	some $\rightarrow s$	all $: .5 \rightarrow a .5 \rightarrow s$
H_1	$s \rightarrow R$	$a \rightarrow L$
S_2	some $\rightarrow s$	all $\rightarrow a$
H_3	$s \rightarrow R$	$a \rightarrow L$
$\dots$		

Figure 20: Hierarchy of speaker and hearer types, $p = .5$

of the model: The most basic first assumption is that there is a sort of *default* or *focal* behavior for S that consist in simply sending random signals *that accord with the literal meaning*. Call an S who behaves this way an S_0. Let H_1 be an H who plays the best strategy he can *on the assumption that he is playing against an S_0*. Let an S_2 be an S who plays the best strategy he can assuming he is playing against an H_1. And so on.[28] This gives us a hierarchy of types of plays for both S and H.

These definition are not quite as precise as Franke's but they are a good starting point. Let us see what they do for us in the game in figure 19. The behaviors for the hierarchy of types, when $p = .5$ is in figure 20. For any $n > 1$, if n is odd then $H_n = H_1$ and n is even then $S_N = S_2$. So, the series reaches a fixed point immediately. According to the IBR theory (in this simple form) players will play those fixed points. This is perfectly Gricean play, so the model predicts the basic scalar implicatures.

As Franke (2009, 2011) makes clear this is a theory of *bounded* rationality. We are considering a hierarchy of increasingly sophisticated players playing against each other. However, *none* of the reasoning used in determining the hierarchy itself is consistent with genuine rationality in the usual sense. It is not rational—in the usual sense—to suppose that the person you are playing against is less sophisticated than you.[29] Of course, ultimately IBR is just a formal specification of games and solutions, and does not need to conform to any

---

Franke's is meant to be a model of bounded rationality. An evolutionary model does not give a *rational* grounding for Gricean implicatures in the usual sense, and thus is outside the purview of my discussion here.

[28]I am keeping things quicker by not discussing the parallel sequence starting with H_0 as an H who has a literal interpretation of the message from S.

[29]This is not to say that it is not rational to play the strategies that are the result of such reasoning: it is just that the reasoning itself is not easily characterized as what we should expect from two players with common knowledge of rationality.

Type	Strategy	
S_0	$some \to s$	$all : .5 \to a.5 \to s$
H_1	$s \to R$	$a \to R$
S_2	any	any
$\cdots$		

Figure 21: Hierarchy of speaker and hearer types, $p = .1$

intuitive notion of rationality, bounded or not, it is just inspired by them.

The predictions of this model differ depending on the underlying probabilities in the game structure. Consider, for instance, the game in figure 19 in which $p = .1$. The Iterated Best Response sequence is given in figure 21. The problem is that when H_1 is responding to S_0, the m_s message does not sufficiently alter his beliefs to make H_1 want to act any differently from how he would in the absence of any information. This means that there is no unique best response for S_2 to H_1, since H_1 ignores the message given. Thus, the basic characterization of the model I gave above does not extend to cover this case since there is no unique best response. We can, though, easily extend the model to handle cases in which there is not a unique best response. For instance we can assume that the player chooses randomly among eligible responses (as Franke does) or we can allow sets of responses into the model (as Jager does).[30] Either way, in the case considered here, we will not get a sequence that converges to Gricean strategy. Thus, the simple IBR model fails to deal with a large class of Some/All Cases.

I should note that Franke (2009, §2.2.4, 3.1) admits to some of the shortcomings of the IBR approach for some initial probabilities. Franke argues that the probabilities in his models should not be thought of as real-world probabilities but rather as "condensed and simplified representations of generally accessible meaning associations." In other words, we should think of some aspects of the representation of the scalar implicature situation not as relating to a real-world situation but rather as something that comes as part of the meaning of the words 'some' and 'all'. Moves like this, however, make the model seem like less of a rational reconstruction and more of a substantive psychological

---

[30]Franke's choice here is essential in order to capture free-choice inferences.

account of meaning.[31]

12.2   Optimal Answer model

A related model is the *Optimal Answer* model of Benz (2006) and Benz and van Rooij (2007). This model is, in essence, a version of the IBR model, though it is framed rather differently. Essentially S's action is predicated on the assumption that H simply update his beliefs by means of the literal meaning of the message. Then, H's *actual* response is based on the belief that S will act in the way just stated (and a faith in S's expertise). Without going into detail it is worth noting that a model with this structure makes very strong bounded rationality assumptions. For the model to work S needs to solve a decision problem based on the idea that H will act in a naïve way in which he does not actually act (i.e. S thinks that H will just update his beliefs according to the literal message).

There are predictive problems which are similar in character to those facing the IBR model. In essence, predictive success requires H to have the right background beliefs about the probability of the different states. As in the IBR model, if the H thinks that the *some* state is very unlikely and that S is merely speaking literally then H's action will not be affected by S's message. So, in this case, S, with his assumption of a naïve H, will not have any incentive to use the Gricean strategy.

I should note that these worries (for both IBR and Optimal Answer models) can be assuaged by structuring the payoffs in a different way. Indeed, many of the derivations in the literature within Optimal Answer model depend on payoffs being determined by degrees of H's belief in the true state of the S, rather than a discrete choice on H's part. This treatment of payoffs may eliminate the models failings in the face of skewed probabilities, but we must then take it on as yet another substantive assumption. Iterative weak dominance reasoning will work with *either* way of modeling payoffs.

---

[31]Franke also suggests an entirely different way of dealing with the problem: by assuming that speakers and hearers always assign low probabilities to any available strategies. This will work in some cases, but it is not clear how to get this to cover scales with more than two points in an empirically satisfactory way.

12.3 Critique of reasoning-based models

Reasoning-based accounts, both the IBR and the Optimal Answer models, do not provide a credible reconstruction of scalar implicatures as a species of rational inference (indeed it is not clear that they are intended to). Let me summarize the two (related) points that lead me to this conclusion:

Non-Standard Framework Both these accounts use non-standard frameworks: they propose particular, intricate chains of reasoning to account for scalar implicatures. It is important to note that these models are *not* given as a heuristic to get a result also achievable by assumptions of real rationality. In that sense the title of bounded *rationality* is misleading. In fact, on these models literal meanings are *defined* in terms of theses chains of reasoning, so it is not even in principle possible to reconstruct a non-bounded, truly rational model. Of course, the models are ultimately just formal characterizations of solution concepts, and they need not cohere with any particular conception of human rationality (bounded or not) even if they are inspired by them.

Once we accept this, though we face a more general problem in evaluating these models. Clearly what should support them is empirical data. The problem here is that the empirical bar for theories *constructed* to handle a set of data is very high. Given the large space of parameters that can be tinkered with in creating a new equilibrium concept, we should *expect* wide empirical coverage. But models such as IBR are specifically constructed to work with a limited set of linguistic situations, so are resistant to the sort of wide-empirical support they would need.

Parameter Setting I noted that, unlike weak dominance reasoning, both the IBR model and the Optimal Answer model only works over a limited range of assumptions about the prior probabilities. Both also rely on a particular specification of what naive behavior is to get the reasoning going (i.e. the starting point of the level-k reasoning embodied in the hierarchy of speaker and hearer types). Not only do we need to assume a of specific pattern of reasoning, but we also need to assume a particular initial set-up not common to all (real-world) cases of scalar implicature derivation. This is not to say this assumption is unpalatable: after all, perhaps we *do* as a matter of fact assume a kind of even distribution of

probabilities of underlying states when calculating scalar implicatures. Even if the assumption has some plausibility it is a substantive assumption about how these situations are structured. The more such assumptions one makes, the less robust the model is.

For these reasons it seems to me that we must view the reasoning-based models not as credible explanations of why implicatures are rational inferences but rather as strong empirical hypotheses (stated within a game-theoretic framework) about how implicatures are calculated. To make this a credible hypothesis these models need to provide more empirical coverage than standard Gricean theories can.[32] This is an interesting line of research but I think it is important to distinguish what we get from them from the project here of trying to capture scalar-implicature derivation as rational inference.

13   Non-expertise

Effectively the Simple Gricean game assumes speaker *expertise*. The two types of speakers both have all of the relevant information about the state of the world: the speaker either knows that some of the cake was eaten or that all of the cake was eaten. However, we can imagine that speakers might only have partial information.

Let us model the two complete situations as worlds: w_a is the world where all was eaten, $w_{s\neg a}$ is the world where just some was eaten. Now consider the simple Some/All game where the speaker can have any level of knowledge compatible with truthfully saying 'some' or 'all'. Figure 22 lists the three types of speakers in this situation, classified according to their states of knowledge. Ignoring payoffs, but assuming that the speaker can only utter sentences he believes to be true, we can draw this as an extensive-form game in figure 23. Assuming that the payoffs are aligned and H benefits from having confidence in the true state of the world, any strategy for S_a that does not have him playing m_a with probability one will be weakly dominated by the variant in which S_a

---

[32]Franke (2009), in particular, tries to cover a range of cases beyond simple scalar implicatures including manner implicatures and free-choice implicatures with his model. Weak-dominance reasoning will not derive free-choice implicatures, but this might not be a bad result as there is considerable empirical evidence that they do not pattern with normal scalar implicatures (Chemla, 2009).

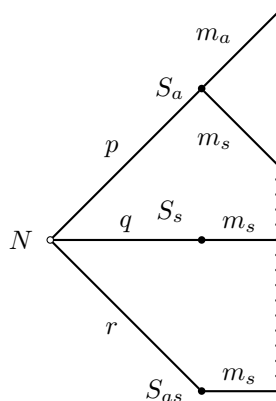| Type	Knowledge
S_a	$\{w_a\}$
S_s	$\{w_{s\neg a}\}$
S_{as}	$\{w_a, w_{s\neg a}\}$

Figure 22: Speaker types

Figure 23: Simple Gricean game with non-expert

sends m_a with probability 1. If H knows that S will not play weakly dominated strategies, then H can infer that if he receives signal m_s then S is not of type t_a.

Thus, weak-dominance reasoning tells us that the hearer can infer that if the message 'some' is received then the speaker does not know the world is w_a (though he might not have ruled it out either). This is the standard 'epistemic' inference assumed in the literature on pragmatics.[33] So, merely adding uncertainty to the simple model does not cause any problem for the kind of reasoning we were using before.

14 Three-point scales

So far, for simplicity's sake, we have concentrated on a *two-point* scale, 'some'/'all'. However, classic Gricean reasoning based on the maxim of Quantity extends

---

[33]The point that epistemic inferences are the only one's licensed without knowledge assumption was first emphasized in a formal framework in Soames' (1982) critique of Gazdar's (1979) treatment of implicatures.

to scales of arbitrary size. Consider, for instance, a three-point scale, such as {'some', 'most', 'all'}, where 'all' entails 'most' which entails, in turn, 'some'— e.g. 'I ate all the cake' → 'I ate most of the cake' → 'I ate some of the cake'. If a speaker chooses between these three expression guided by the the Maxim of Quantity then he will only say 'some' when he does not know 'most' or 'all', and he will only say 'most' when does not know 'all'. Thus, the hearer can infer from an utterance of 'most' that the speaker does not know 'all', and from an utterance of 'some' that the speaker does not know 'most' or 'all'. (Assuming, as always, that these differences are known to be relevant.)

Let us consider the case in which we are dealing with an expert speaker, and thus limit ourselves to three speaker types. The extensive-form representation (without H's responses) is given in figure 24. The full set of possible pure strategies for S is in figure 25. In this game the hearer has three possible responses A, B, C and the payoffs (which are dependent just on S's type and H's response) are coordinative: both players are best off if the hearer can determine the speaker's type and act appropriately. Figure 26 gives all the possible pure hearer strategies in this game (with the first round of weakly-dominated strategies removed).

In this game there will always be more than one PBE. While the Gricean strategies will always be a PBE, either the pair (PP2,NG3) or (PP2, NG4) will also be a PBE (which one depends on the initial probabilities of the different speaker types).

Despite there being at least two possible PBEs, iterative weak-dominance reasoning will again pick out the Gricean strategy pair. First, consider the speaker of type S_a. For him the use of any message but m_a is weakly dominated: m_a will always get him the highest payoff no matter what strategy the hearer chooses, whereas using and m_m and m_s will not always do so. If the hearer knows that the speaker will always use m_a when he is of type S_a, then he can infer that if he receives another message the speaker cannot be of type S_a. If the speaker knows this, in turn, then for the speaker of type S_m sending message m_m weakly dominates sending message m_s, its only alternative. Thus, the hearer will send that message, which means he will play the Gricean strategy in all instances. If the hearer, in turn, knows this then the only strategy left is the pure Gricean interpretation strategy. So iterative weak dominance reasoning forces Gricean behavior. It should be clear, too, that the reasoning
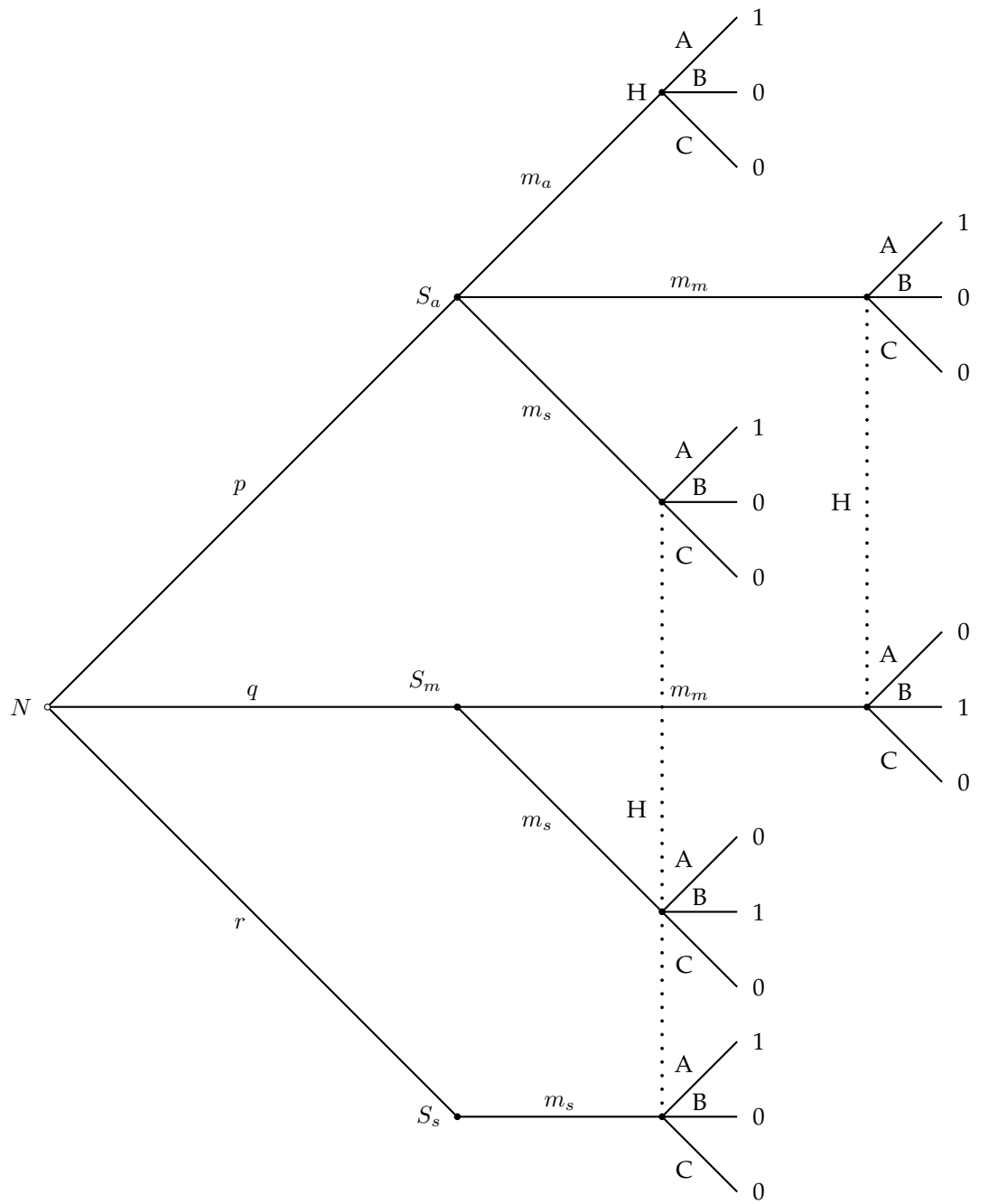
Figure 24: Three-point scale

38

Type	S_a	S_m	S_s
Grice	m_a	m_m	m_s
PP1	m_m	m_m	m_s
PP2	m_m	m_s	m_s
PP3	m_s	m_m	m_s
Pooling	m_s	m_s	m_s

Figure 25: Speaker pure strategies

	m_a	m_m	m_s
Grice	A	B	C
NG1	A	B	B
NG2	A	B	A
NG3	A	A	C
NG4	A	A	B
NG5	A	A	A

Figure 26: Hearer pure strategies

here generalizes to n-point scales for any finite n.[34]

15 Expanding alternative utterances

A persistent criticism of the Gricean account of scalar implicatures is it depends on a very narrow limitation on the number of messages a speaker is able to send. A version of this problem, recently dubbed "the symmetry problem", can be explained as follows: Suppose we have an expert speaker in in the Some/All problem.[35] Recall the informal reasoning that gets us from the fact that the speaker says 'some' to the implicature that he knows that the state of the world is the *some* state and not the *all* state: If the speaker had known that the 'all' sentence was true he would have said it. He did not, so he must not know it to be true. Therefore, he knows that the speaker knows 'some' is true (since he is an expert).

Unfortunately, we can give a symmetric line of reasoning if the speaker has utterance 'some but not all' available to him. If the speaker had known that 'some but not all' was true he should have said it (by Quantity). Since he

---

[34]Generalizing beyond a two-point scale is not to be taken for granted, there are various strategies that might work for a two-point scale that fail to generalize, such as "naive unravelling" an approach proposed by Franke et al. (2012).

[35]The problem has been known for a long time, but was formalized by Irene Heim and Kai von Fintel in lecture notes (see e.g. Fox, 2006).

didn't say it he doesn't know it. Therefore he must know that 'all' is true (since he is an expert).

In truth, then, an expert speaker who has available to him the utterances 'some', 'some but not all', and 'all' who is governed just by the maxim of Quantity (and truthfulness constraints) should always say either 'some but not all' or 'all': he should not ever say 'some' alone. This is a bad prediction: in real life, expert speakers often use 'some' alone.

A parallel problem faces a game-theoretic treatment once we allow speakers to use a 'some but not all' message in the Simple Gricean game. This is represented in figure 27. As the symmetry of the game there makes clear, there is no way to argue on rational grounds in this model that m_s should be interpreted in any particular way. Thus we again seem to require that well-informed speakers should never say 'some' as opposed to 'some but not all'.
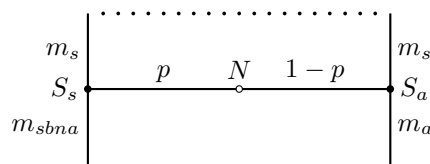

Figure 27: Gricean game with some-but-not-all

I will discuss two ways of dealing with this problem. One way is to restrict the alternatives available in a given speech-act situation. The other way is to add message costs.

15.1 Restrictions on alternatives

It is common to simply posit that there are a restricted set of messages that we consider when we evaluate scalar implicatures, and that these messages do not include 'some but not all'. If we do this we can simply keep our previous models in which 'some but not all' was not a recognized speaker option. This may not be as ad hoc as it seems. We might think that there are constraints

in the grammar on which lexical items compete against each other, and that these constraints facilitate Gricean inferences. Since Horn (1972) there has been some effort in linguistics to give principled conditions for two lexical items to compete against each other (Matsumoto, 1995; Fox and Katzir, 2009).

15.2 Message costs

A trick common in the game theory and pragmatics literature for dealing with this problem is to suppose that the 'some but not all' has a small cost.[36] This cost should not be of the same magnitude as the cost of failing to convey information: cooperative speakers do not sacrifice relevant communicative content by saying something shorter (rather monosyllabic teenagers do this to display uncooperativeness). There are different ways of modeling such small costs in game theory, but for here simply using small numbers will be sufficient.

Take the case of an expert speaker who has available to him three messages m_a, m_s, and m_{sbna}. We will assume that m_{sbna} incurs a small cost. Here, there are only two types of speakers but three messages: the basic game structure is in figure 28. Assuming the S will not use weakly dominate strategies we
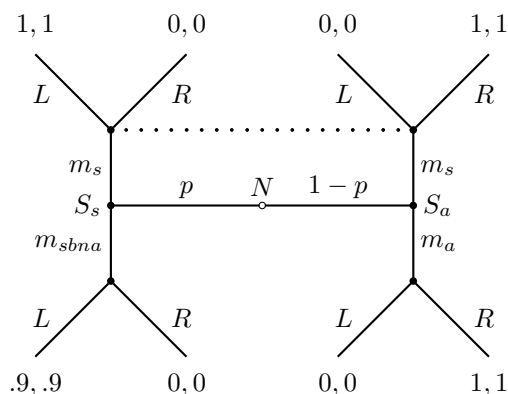
Figure 28: Some but not all with costs

can prune the game to get the new game in figure 29. Now, of course, m_{sbna} becomes a dominated move for the S. Thus, assuming speaker expertise and marginal cost to saying the longer form, we get the desired result that m_s is used to indicate the speaker knows some.

---

[36]Parikh (1991, 2001) first suggested this strategy. For examples in the IBR tradition see Franke (2009, 2011) and Jäger and Ebert (2008).
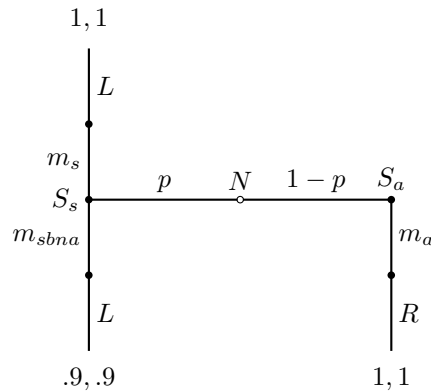
Figure 29: Some but not all with costs

Note that this reasoning has limited generality. As the reader can verify, we still are forced to make the prediction that a speaker who is not assumed to be an expert will never say 'some' to mean 'some but not all'. To overcome this limitation we need to make more substantive (and less plausible) assumptions.

16   Relevance implicatures

So far, I have only considered cases of scalar implicatures. However, in this section I will suggest that another kind of implicature, a relevance implicature, might be captured using weak-dominance reasoning.

A famous example from Grice goes as follows. The hearer stops in the car and asks the speaker if there's anywhere to get petrol. The speaker responds that there's petrol station down the road. This response takes this to implicate that the station is open and operating.

We can give a model of this situation in which the expert speaker has two possible states of knowledge, 1) that he knows where the petrol station is and knows it is open (call a speaker who knows this a S_o) , 2) that he knows where one is but knows it is closed (call a speaker who knows this a S_c). This is a rather crude simplification but a game-theoretic model of this situation will need to make some sort of assumption like this. We can also think of two signaling options available to him: 1) giving information about the location of the petrol station (g), 2) not giving information ($\neg g$). If the hearer is told where the petrol station is he has the choice of either going to the station (action

G), or continuing his journey without going there (action C). If he does not receive that information he must continue his journey. The payoffs are ordered as follows: the payoffs are highest if the gas station is open and the hearer goes there. The medium payoff is that the hearer just continues his search (regardless of whether the gas station is there or not). And the lowest payoff is if the gas station is closed and the hearer goes there. Figure 30 is the extensive-form representation of this game.
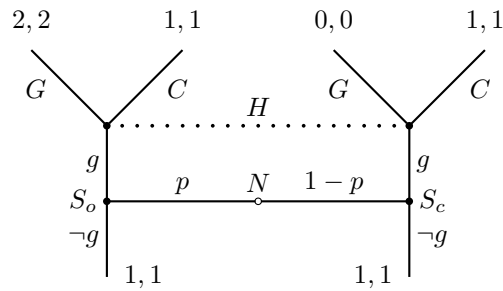

Figure 30: Petrol station game

It is immediately clear from this model that for a speaker (S_c) who knows the petrol station is closed not providing any information is a weakly dominant move. For doing so ensures that the speaker continues his journey which is the highest payoff possible. If this is common knowledge, then then it is also common knowledge that if the the speaker gives the location of the gas station then he is of type S_o. In this case, when hearer receives the location, he will know that it is open (and thus go to it to maximize his utility).

So in this simple model of Grice's example, weak dominance reasoning again yields the correct implicature. This is a very open-ended kind of example however, so it is hard to say if this is the right way of modeling it. My main point here is just to demonstrate that the iterative weak dominance reasoning plausibly extends to cover other examples of implicatures besides scalar implicatures. However the open-endedness inherent in how we model such examples makes it hard to give a convincing argument that this is how we *should* think of such implicatures.

17   A non-linguistic case

In this section, I will consider a situation, analogous to a commonly discussed one in the economics literature, which has the same structure as some of our scalar implicature cases.[37] I show that the iterative weak-dominance reasoning provides a plausible explanation of how people act in this case, just as it does in the scalar implicature cases. This provides further support for the idea that Gricean reasoning need not be thought of as a special sort of reasoning: rather the exact same patterns of reasoning can be observed in non-linguistic contexts.

First, I will sketch the standard model in the economics literature on disclosure of verifiable information (Grossman, 1981; Milgrom, 1981). In this situation, a seller has some asset, which he knows the value of, and the seller has the option to release evidence that proves the asset is at least a certain worth. For simplicity assume there are n (total) pieces of evidence the seller can release e_0 to e_{n-1}. Assume the buyer knows there are n possible states of value the asset can have $v_0 < v_2 \ldots < v_{n-1}$. In state v_0 there is no evidence the seller can release, in state v_0, the seller can release e_0, and in state v_1 the seller can release either e_0 or e_1, and so on. The situation works as follows: first the seller the releases some chosen piece of evidence, and then the buyer performs an action (i.e. makes purchase or a bid for a purchase). We assume the payoffs are such that the buyer can make the optimal decision if only if he knows the value of the asset. The seller, on the other hand, prefers buyer actions corresponding to higher beliefs about the value, regardless of the actual value.

Regardless of the buyer's initial beliefs about the likelihood of the asset having each of the values (as long as he thinks each value is possible), there is a direct (informal) argument that the seller will release all the information he can: i.e. the seller with an asset of value v_{n-1} will release e_{n-1} and the seller with an asset of value v_{n-2} will release e_{n-2} and so on. The argument for this view goes as follows: the seller of an asset v_{n-1} guarantees himself his highest possible return by releasing e_{n-1}, so he will do this. Since this is known to the buyer, then the seller of an asset with value v_{n-2} will guarantee himself his highest return by releasing e_{n-2}, and so on. This argument is formalized and applied to certain variations on the situation above by Grossman (1981); Milgrom (1981), in what is sometimes known as the full-disclosure theorem.

---

[37]Dirk Bergemann first suggested this connection to me; Sobel (2010) and Franke et al. (2012) discusses this connection.

	buy	don't
v_0	$1, -1$	$0, 0$
v_1	$1, 1$	$0, 0$

Figure 31: Two state disclosure

However, as Franke et al. (2012) notes, that there are very simple situations of this form where there are PBEs in which full-disclosure does not occur. In particular just consider a simple case in which there are two states, v_0 and v_1 and two pieces of evidence e_0 and e_1. The only choice the seller makes is whether to release e_1 or e_0 when the asset has value v_1 (if the value is v_0 he must release e_0). Assume they buyer has two options: to buy or not to buy, and prefers buying only if the value is v_1. We could then have the payoffs as in Figure 31. As long as the the prior distribution is sufficiently skewed to v_1 there will be a PBE in which the seller always releases e_0 and the buyer always buys. But we can escape this problem, and get full disclosure by iterative elimination of weakly-dominated strategies.

I take it as a distinct virtue of the iterative dominance approach that it supports such structural analogies between linguistic problems and non-linguistic problems. This analogy is one the reasoning-based approaches I discussed above cannot capture: For those views, as I argued, are sensitive to the initial probability distribution. But in the disclosure case, the behavior is not contingent on this distribution. Rather we should think that regardless of the initial probability distribution we expect the "Gricean" behavior on the part of seller and buyer.

18   Conclusion

The main purpose of my paper has been to show that for a variety of idealized cases we can give a principled game-theoretic account of the derivation of scalar implicatures. The assumptions and cost of this model were high, however. We assumed discrete models taken out of context with mutually known and understood payoffs—something we rarely see in real speech-act situations. We also assumed completely cooperativeness which we modeling by absolute identity of payoffs. We also hard-wired into the model the semantic conventions by refusing to let speakers say sentences they do not believe

to be true. Even with these assumptions the most standard solution concepts, such as Nash equilibrium or perfect Bayesian equilibrium, do not serve to pick out the Gricean behavior. However, we saw that in many cases iterative elimination of weakly dominated strategies serves to single out the Gricean strategy.

References

Battigalli, Pierpaolo and Bonanno, Giacomo (1999). Recent results on belief, knowledge and the epsitemic foundations of game theory. *Research in Economics*, 53:149–225.

Benz, Anton (2006). Utility and relevance of answers. In Benz Anton, Gerhard Jäger, and Robert van Rooij (eds.), *Game Theory and Pragmatics*. Breheney.

Benz, Anton and van Rooij, Robert (2007). Optimal assertions and what they implicate. *Topoi*, 26:63–78.

Bernheim, B. Douglas (1984). Rationalizable strategic behavior. *Econometrica*, 52:1007–1028.

Brandenburger, Adam (2007). The power of paradox: some recent developments in interactive epistemology. *International Journal of Game Theory*, 35:465–492.

Brandenburger, Adam, Friedenberg, Amanda, and Keisler, H. Jerome (2008). Admissibility in games. *Econometrica*, 76:307–352.

Chemla, Emmanuel (2009). Universal implicatures and free choice effects: Experimental data. *Semantics and Pragmatics*, 2:1–33.

Chierchia, Gennaro (2004). Scalar implicatures, polarity phenomenon, and the syntax/pragmatic interface. In Andrea Belleti (ed.), *Structures and Beyond*. Oxford University Press.

Chierchia, Gennaro, Fox, Danny, and Spector, Benjamin (2012). Scalar implicature as a grammatical phenomenon. In Paul Portner, Claudia Maienborn, and Klaus von Heusinger (eds.), *Handbook of Semantics*, volume 3, pages 2297–2332. de Gruyter.

Cho, In-Koo and Kreps, David M. (1987). Signaling games and stable equilbiria. *Quarterly Journal of Economics*, 102:179–221.

Crawford, Vincent and Sobel, Joel (1982). Strategic information transmission. *Econometrica*, 50:1431–1451.

Farrell, Joseph (1993). Meaning and credibility in cheap-talk games. *Games and Economic Behavior*, 5:514–531.

Farrell, Joseph and Rabin, Matthew (1996). Cheap talk. *Journal of Economic Perspectives*, 10:103–118.

Fox, Danny (2006). Free choice and the theory of scalar implicatures. Unpublished manuscript.

Fox, Danny and Katzir, Roni (2009). On the characterization of alternatives. Unpublished manuscript, MIT.

Franke, Michael (2009). *Signal to Act*. Ph.D. thesis, Institute for Logic, Language and Computation, University of Amsterdam.

— (2011). Quantity implicatures, exhaustive interpretation, and rational conversation. *Semantics and Pragmatics*, 4(1).

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

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

Grice, Paul (1967/1989). Logic and conversation. In *Studies in the Ways of Words*. Harvard University Press.

Grossman, Sanford (1981). The informational role of warranties and private disclosure about product quality. *Journal of Law and Economics*, 24:461–483.

Harsanyi, John and Selten, Reinhard (1988). *A General Theory of Equilibrium Selection in Games*. MIT Press.

Horn, Laurence (1972). *The Semantics of the Logical Operators in English*. Ph.D. thesis, UCLA.

Jäger, Gerhard (2007). Game dynamics connects semantics with pragmatics. In Ahti-Veikko Pietarinen (ed.), *Game Theory and Linguistic Meaning*. Elsevier.

Jäger, Gerhard and Ebert, Christian (2008). Pragmatic rationalizability. In Arndt Riester and Torgim Solstad (eds.), *Proceedings of SuB13*.

de Jager, Tikitu and van Rooij, Robert (2007). Explaining quantity implicatures. In *Proceedings of the 11th conference on theoretical aspects of rationality and knowledge*.

Kohlberg, Elon and Mertens, Jean-François (1986). On the strategic stability of equilibria. *Econometrica*, 54:1003–1037.

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

Marx, Leslie and Swinkles, Jeroen (1997). Order independence for iterated weak dominance. *Games and Economic Behavior*, 18:219–245.

Matsumoto, Yo (1995). The conversational condition on horn scales. *Linguistics and Philosophy*, 18:21–60.

Milgrom, Paul (1981). Good news and bad news: Representation theorems and applications. *Bell Journal of Economics*, 12:380–391.

Noveck, Ira (2001). When children are more logical than adults: experimental investigations of scalar implicature. *Cognition*, 78:165–188.

Parikh, Prashant (1991). Communication and strategic inference. *Linguistics and Philosophy*, 14(473–531).

— (2001). *The Use of Language*. CSLI.

Pearce, David (1984). Rational strategic behavior and the problem of perfection. *Econometrica*, 52:1029–1050.

Rabin, Matthew (1990). Communication between rational agents. *Journal of Economic Theory*, 51:144–170.

Rochet, Jean-Charles (1980). Selection of unique equilibrium payoff for extensive games with perfect information. Mimeo.

van Rooij, Robert (2009). Optimal-theoretic and game-theoretic approaches to implicature. In Edward N. Zalta (ed.), *Stanford Encyclopedia of Philosophy*. Spring 2009 edition.

Rosenthal, Robert (1981). Games of perfect information, predatory pricing, and the chain store. *Journal of Economic Theory*, 25:92–100.

Sauerland, Uli (2004). Scalar implicatures in complex sentences. *Linguistics and Philosophy*, 27:367–391.

Schelling, Thomas C. (1960). *The Strategy of Conflict*. Harvard University Press.

Schulz, Katrin and van Rooij, Robert (2006). Pragmatic meaning and non-monotonic reasoning. *Linguistics and Philosophy*, 29(2):205–250.

Soames, Scott (1982). How presuppositions are inherited: A solution to the projection problem. *Linguistic Inquiry*, 13(3):483–545.

Sobel, Joel (2009). Signaling games. In *Encyclopedia of Complexity and Systems Science*. 8125–8139.

— (2010). Giving and recieving advice. Working Paper UCSD.

Spector, Benjamin (2006). *Aspects de la pragmatique des opérateurs logiques*. Ph.D. thesis, Universite Paris 7, Denis Diderot.

Spence, Michael (1973). Job market signaling. *The Quarterly Journal of Economics*, 87:355–374.

Sperber, Dan and Wilson, Deirdre (1986). *Relevance: Communication and Cognition*. Blackwell.

Stalnaker, Robert (2005). Saying and meaning, cheap talk and credibility. In Anton Benz, Gerhard Jäger, and Robert van Rooij (eds.), *Game Theory and Pragmatics*. Palgrave MacMillan.