Disambiguation Games in Extended and Strategic Form

Sascia Pavan
s.pavan@tele2.it

Abstract

The aim of this paper is to pursue the line of research initiated by Prashant Parikh which gives content and rigour to the intuitive idea that speaking a language is a rational activity. He employs the most promising tool to that end, namely game theory. I consider one of his examples as a sample case, and the model I build is a slight modification of that developed by him. I argue that my account has some advantage, yet many of the key ideas employed are left unchanged. I analyse this model in detail, describing some of its formal features. I conclude raising a problem that has not been yet, sketching a plausible solution.

1. Introduction

The case I want to analyse concerns sentences like

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

This sentence has two readings, one is that there is a certain man in New York, either very unlucky, or reckless, or masochist, that is mugged every ten minutes. The other reading is that every ten minutes, some man or other, not necessarily the same, gets mugged in New York. Imagine an actual conversation where (1) is uttered, the problem is, how can the hearer decide what is the reading originally intended by the speaker? As for (1), we can hardly imagine a situation where the reading intended by the speaker is the first one – namely the unlucky, reckless, masochist interpretation – and where this is the reading selected by the hearer. A relevant feature of (1) is that one of the two possible readings entails the other, in this case the second reading is a logical consequence of the first. We can think of sentences sharing this same feature with (1), but such that they can be employed in a conversation where the intended reading is the logically stronger one. Consider

All of my graduate students love a Finnish student in my Game-Theory class. (2)

Suppose that (2) is uttered by a professor in Amsterdam. I do not know how many Finnish students studying game theory there are in Amsterdam. Assume there are very few of them. My intuition is that in most situations the hearer would infer that there is a unique Finnish student in the speaker's class that all graduate students love.

I will use game theory to analyse those conversations that involve sentences that, like (1) and (2), can be interpreted in two different ways, such that one reading is a logical consequence of the other. My starting point will be the account proposed by Prashant Parikh in several works [6, 7, 8].

If modelled in game-theoretic terms,[1] conversations like these involve two players, 1 and 2, where the set of 2's possible moves contains two elements, say A and B, corresponding to two alternative interpretations of some ambiguous sentence ϕ. As is customary in game theory, I will imagine that player 1 is male, and player 2 female. In Parikh's model, player 1 has some private information, unknown to player 2. Parikh defines this basic unknown as the speaker's intended meaning. Player 2 has some beliefs about what this private information is, hence about what message player 1 wants to convey, and these beliefs can be expressed as subjective probabilities. Here lies the main shortcoming of Parikh's model. The hearer's task in a conversation is to guess the speaker's intended meaning, or, better, to select a set of possible alternatives, and assess a probability value for these. Therefore, if we model disambiguation as a game, the probability of the intended meaning should not be one of the primitives of the game, rather we have to explain how the hearer can infer this value from the structure of the game. In other terms, if the task of player 2 is to guess what the intended meaning is, and if she already knows which alternative is more likely to be true, then there is not much to be done anymore, she only needs to multiply the subjective probability of each alternative by the payoffs that the moves available to her would yield in each of these alternatives. Suppose that p is the prior probability that player 2 assigns to the belief that player 1 wants to convey the meaning corresponding to A; and that $1-p$ is the probability of the belief that he wants to convey the meaning

[1] Most of the notation and terminology employed is borrowed from Roger Myerson [4], chapters 2-5.

B. Let g_a be the gain for player 2 if she selects the interpretation A when player 1 really wants to convey A, and let m_a be her gain if she selects A when 1's intended meaning is B. Similarly, let g_b be her gain if she correctly selects B, and m_b her gain when she wrongly selects B. If we describe the situation in this way, her task is very simple, she must select A whenever $p \times g_a + (1-p) \times m_a > p \times m_b + (1-p) \times g_b$ and B whenever $p \times g_a + (1-p) \times m_a < p \times m_b + (1-p) \times g_b$. Once we know that she is able to assign a probability value to the belief that 1's intended meaning is A – no matter how she could accomplish this – there is nothing more to be explained, and hence no more need to appeal to game theory to give an account of her behaviour. But, presumably, we need game theory to explain how she could assess this probability.

This is why I claim that the content of player 1's private information has to be something more basic, and therefore that player 2's prior probabilities have to concern what player 1 actually knows. With this modelling of the game, the speaker's intention to convey a given message can be derived from facts with a minor degree of intentionality, namely his knowledge. To paraphrase Willard Van Quine [9], it reduces the grade of *intentional involvement*. Just consider the questions 'What does player 1 know?' and 'What does player 1 want to say?'. We are not always able to provide definite answers to the questions of the first kind, but, at least, we can assess the probability of the answers, just considering what we know about the player's sources of information. Of course, we can also assess the probability of the answers of the questions of the second kind, but the data to be considered include all those relevant for the first kind, and something else, at least this person's goals. In other words, any reasoning behind an answer to a question of the first kind is conceptually simpler than that required by the second kind.

2. Extensive Form

What is now the shape of our model? If A and B are the only legitimate interpretations of an ambiguous utterance ϕ, then either he believes that A or he believes that B. But in the case we are examining, one of the two readings is a logical consequence of the other, for example we can assume that B logically entails A. If this is true, then if 1 believes that B, he necessarily believes that A. Then, as far as player 2 knows, there are two possibilities:

alternative a: 1 knows that A and it is not the case that he knows that B (either because he knows that not B, or because he does not know whether B);

alternative b: he knows that A and B.

This imposes some restrictions on the payoffs of the game. If a is the real situation, then, if 2 selects A when 1 utters ϕ, she will acquire some new and reliable true knowledge, let us name 'g_a' the value that this outcome has for her. But, if in the same situation she chooses B, instead, she gets a false or at least unreliable new belief and hence some bad result, let us name 'm_b' the value of this outcome. If b is the real situation, then the choice of B will yield some new knowledge, and let be g_b the value she puts on it. But since in this situation the information corresponding to A is true and reliable as well, if she chooses A she does not get some bad payoff, her gain should again be g_a. Let us now use 'p' to refer to the prior probability of situation a, so that $1 - p$ is the prior probability of b. Which moves are available to player 1? One of them is of course the uttering of the ambiguous sentence ϕ. But, he could also choose to convey the message he has in mind using some longer but unambiguous sentence, μ_a if he is in situation a, μ_b if he is in situation b. When player 1's choice is one of these two, player 2 does not have to consider alternative interpretations, hence, in game-theoretic terms, she has no opportunity to move. In this case, no misunderstanding is possible.

We have to imagine that he is sincere and honest, that she believes what he says, and that this is common knowledge. For simplicity, imagine also that both of them are interested exclusively in the pure flow of information and no further aims. This is unrealistic, of course, but it is just an idealization not more problematic than the physicist's speculations on frictionless planes. Following Parikh, I will construct my model as a coordination game where the players have the same payoffs, which are determined by the net value of the information minus a 'cost' which is proportional to the length of the sentence. Since the two players have identical payoffs, this is a game of 'pure coordination'. The rationale for this choice is that when honest and rational agents communicate, they all aim at successful communication. Of course there are commonly cases where this is not true, most notably when people lie. But we can legitimately focus attention on those benign cases, especially because the very possibility of lying presupposes the existence of honest communication.

The payoff g_a has to be equal to the value of the true information provided by A, call it v_a, minus the cost c involved by ϕ. If player 1 utters μ_a in situation a, there is no possibility of a misunderstanding, but its cost is higher. Hence this combination yields a value $g'_a = v_a - c'$, where $c' > c$. Similarly, if we call 'v_b' the net value of the true information provided by B, we have that $g_b = v_b - c$. And if player 1 utters μ_b in situation b, then the payoff will be $g'b = vb - c'$, if, for the sake of simplicity, we assume that the cost involved by μ_a and μ_b is analogous. Moreover, since B logically entails A, while A does not entail B, we should have that $v_b > v_a$, and this entails that $g_b > g_a$, and $g'_b > g'_a$.

What is the best choice for player 2? Can we say again that she has only to check whether $p \times g_a + (1-p) \times g_a > p \times m_b + (1-p) \times g_b$, i.e. whether $p > (g_b - g_a)/(g_b - m_b)$, or whether $g_a < p \times m_b + (1-p) \times g_b$? Assume that this is the case, and imagine, for example, that $g_a > p \times m_b + (1-p) \times g_b$. What happens if b, not, a, is the real situation, and 1 wants to convey message B? He would probably utter the longer but unambiguous sentence μ_b. This behaviour is not outright irrational, we shall see that it corresponds to an equilibrium in our model, but it is not always the best outcome that player 1 and player 2 can get, in other words it might be inefficient. Moreover, from the mere fact that there is a probability p that 1 knows that A and not that B nothing follows, from a conceptual point of view, about what he means when he says something. Now she really needs to consider also his goals in order to be able to guess which alternative he wants her to choose.

We can conceive of cases where an unambiguous sentence is so much longer than the corresponding ambiguous one, that a cheap misunderstanding can be preferable to an unambiguous but demanding speech act. We can also imagine situations where the speakers choose ambiguous and potentially misleading messages because they do not want other people to acquire some confidential information. Just think of two spies involved in a telephone conversation, both knowing that their line has been tapped. Sometimes a leak can do more harm than a misunderstanding. I will assume that this is not the case in the conversation we are considering, and that in this case the cost of an utterance is relatively small when compared to the net value of information. Hence, the model employed here requires the following ordering relations: $g_b > g'_b > g_a > g'_a > m_b$, $g'_a - m_b > g_b - g'_b$, $g'_b - g_a > g_a - g'_a$.

But maybe the set of moves available to player 1 is incomplete. Perhaps we should also consider the possibility of uttering μ_a in situation b, and μ_b in situation a. Of course if player 1 uttered μ_a knowing that A is false, he would be lying, and, under the assumption that we are trying to analyse a case of patently honest communication this move would yield a bad outcome for both. But the other case cannot be dismissed so easily, remember that A is true in situation b. The payoff would actually be g'_a. The fact is that whatever the choice of 2, the gain would be higher if player 1 chose μ_b or ϕ. This means that, according to the model presented here, it is never rational for player 1 to choose to utter μ_a in situation b. In technical terms, any strategy where the speaker utters μ_a in situation b or μ_b in situation a is strongly dominated, and can be eliminated from the game. In this case the model simply predicts the existence of a scalar implicature, to the effect that if 1 utters μ_a, then 2 infers that it is not the case that 1 knows that B. Of course the ordering among payoffs that was depicted above presupposes that if 1 knew that B, then he would not conceal this information to the hearer. In situations not covered by this analysis, the speaker could utter μ_a knowing that B, if he did not want 2 to know.

Similarly, we could include a pair of 'don't say anything' moves for player 1. Of course, when he chooses one of these additional moves, she has no possibility to move, and the payoff should be equal to 0 for both players. I will assume that both g'_a and g'_b are strictly positive. If this is the case, then, again, any strategy involving one of these additional moves is strongly dominated, hence I will ignore this possible variant of the game. Yet, this shows that I have not mentioned a fact which is implicitly presupposed by our model, namely the fact that, for example, at node a, player 1 knows that A and also *wants* to convey this information. Maybe this is what is meant by Parikh when he says that the chance nodes 'represent [player 1's] intention to convey' A or B [7].[2] If this is the case, the objections I raised in section 1 miss the mark. But, even in this case, his original models should be modified. As I have already said, with cases like (1) or (2), the payoff obtained when player 2 rightly chooses the stronger interpretation, should be higher than that obtained when she rightly chooses the weaker one.

Now we have all the elements to build our game. I will first construct it as a game of imperfect information in extensive form, which will be called Γ^e. It has the structure of a tree, as is shown in Figure 1.

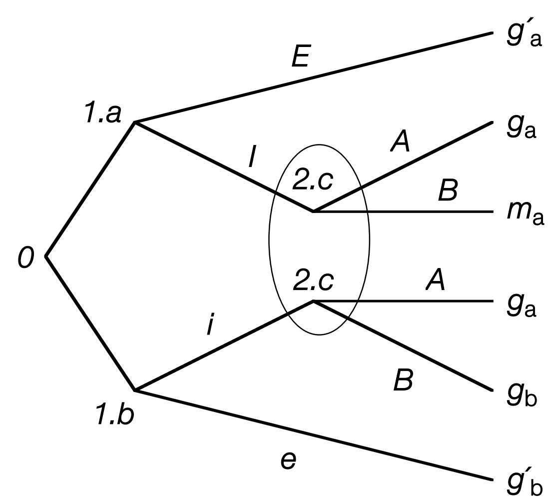

Figure 1. The game in extensive form

[2] P. 27.

This is a deviation from Parikh's path. In the model presented here, I imagine that the first event in the game is a chance move made by 'Nature', which determines whether 1 knows that A and does not know that B, or knows that A and B. At this point 1 can make his move. As usual, I also imagine that the whole structure of the game is common knowledge. Parikh's game in extensive form is not a tree but a pair of trees, since he argues that player 2 cannot construct anything before 1's utterance [7],[3] this is why he proposes the notion of a game of *partial information*. I have the impression that this is an unnecessary – but harmless – deviation from more traditional notions, unless we want our model to mirror the actual mental processes of speaker and hearer. I observe that if Parikh is right in his claim that these disambiguation games should not be treated as ordinary games of imperfect information, the same would hold, for example, for Spence's 'model of education' [12].

Then, a and b are chance events with prior probability p and $1-p$, respectively, where $1 > p > 0$, as usual. If player 1 is in situation a, he can utter either ϕ or μ_a, and we can label these two moves 'I' and 'E', respectively – where 'I' stands for 'implicit' and 'E' for 'explicit'. If he is in situation b, he can choose between ϕ and μ_b, and we can call these alternative moves 'i' and 'e'. Player 2 has a chance to move only if the game is in one of the states labelled '$2.c$', where the identucal labels and the ellipsis manifest the fact that she is not able to distinguish them, technically speaking they belong to the same *information set*. Her options are the two moves A and B.

The fact that there are only two alternative states in 2's information set follows from the characteristic features of the examples considered, namely the fact that one of the two readings is entailed by the other. It is not even necessary that this be a logical entailment – like it is in our example – but the entailment has to be common knowledge. If the two alternative interpretations were mutually exclusive, we would build another game with two epistemic possibilities, but there would be a difference in the ordering of the outcomes. The choice of A when 1 means B, for example, would lead to a bad result. If the two alternatives readings were logically and conceptually unrelated, player 2 would have an information set containing three elements. And of course we can conceive of cases where an ambiguous sentence admits of more than two readings.

3. Nash Equilibria in Strategic Form

The normal representation of our game is the set $\Gamma = \{N, C_1, C_2, u\}$, where $N = \{1, 2\}$ is the set of players, $C_1 = \{Ei, Ee, Ii, Ie\}$ and $C_2 = \{A, B\}$ are the sets of their pure strategies, and u is their payoff function, hence a function from $C_1 \times C_2$ to the real line $\mathbb{R}$. It satisfies the pattern shown in Table 1.

	A	B
Ei	$p \times g'_a + (1-p) \times g_a$	$p \times g'_a + (1-p) \times g_b$
Ee	$p \times g'_a + (1-p) \times g'_b$	$p \times g'_a + (1-p) \times g'_b$
Ii	g_a	$p \times m_b + (1-p) \times g_b$
Ie	$p \times g_a + (1-p) \times g'_b$	$p \times m_b + (1-p) \times g'_b$

Table 1. The game in strategic form

I will now establish a few properties of the model.

Theorem 3.1 *Ii is strongly dominated in Γ*

Proof. Ii is strongly dominated if and only if $\exists \sigma_1 \in \Delta(C_1)$ such that

$$\begin{aligned} & u(Ii, A) < \sigma_1(Ei)u(Ei, A) + \\ & \sigma_1(Ee)u(Ee, A) + \sigma_1(Ie)u(Ie, A) + \\ & (1 - \sigma_1(Ei) - \sigma_1(Ee) - \sigma_1(Ie))u(Ii, A) \end{aligned} \tag{3}$$

and

$$\begin{aligned} & u(Ii, B) < \sigma_1(Ei)u(Ei, B) + \\ & \sigma_1(Ee)u(Ee, B) + \sigma_1(Ie)u(Ie, B) + \\ & (1 - \sigma_1(Ei) - \sigma_1(Ee) - \sigma_1(Ie))u(Ii, B) \end{aligned} \tag{4}$$

Inequalities (3) and (4) are equivalent to

$$\frac{\sigma_1(Ei) + \sigma_1(Ee)}{\sigma_1(Ie) + \sigma_1(Ee)} < \frac{(1-p)(g'_b - g'_a)}{p(g_a - g'_a)} \tag{5}$$

and

$$\frac{\sigma_1(Ei) + \sigma_1(Ee)}{\sigma_1(Ie) + \sigma_1(Ee)} > \frac{(1-p)(g_b - g'_b)}{p(g'_a - m_b)} \tag{6}$$

respectively. Since $g'_b > g_a$, we have that $(g'_b - g'_a)/(g_a - g'_a) > 1$. Moreover, we stated above that $g'_a - m_b > g_b - g'_b$, therefore $1 > (g_b - g'_b)/(g'_a - m_b)$. This entails

$$\frac{g'_b - g'_a}{g_a - g'_a} > \frac{g_b - g'_b}{g'_a - m_b}$$

and hence

$$\frac{(1-p)(g'_b - g'_a)}{p(g_a - g'_a)} > \frac{(1-p)(g_b - g'_b)}{p(g'_a - m_b)}$$

At this point it is an easy task to find values for $\sigma_1(Ei)$, $\sigma_1(Ie)$, and $\sigma_1(Ee)$ that satisfy inequalities (5) and (6). QED

Observe that Theorem 3.1 entails that no strategy profile τ where $\tau_1(Ii) > 0$ is a Nash equilibrium.

[3] P. 83.

Theorem 3.2 *There is no equilibrium in Γ where both Ei and Ie have strictly positive probability.*

Proof. Assume that σ is such an equilibrium. Then the following inequalities have to be true:

$$\sum_{c_2 \in C_2} \sigma_2(c_2)u(Ei, c_2) \geq \sum_{c_2 \in C_2} \sigma_2(c_2)u(Ee, c_2)$$
$$\sum_{c_2 \in C_2} \sigma_2(c_2)u(Ie, c_2) \geq \sum_{c_2 \in C_2} \sigma_2(c_2)u(Ee, c_2)$$

They are equivalent to

$$\sigma_2(A) \leq \frac{g_b - g'_b}{g_b - g_a}$$
$$\sigma_2(A) \geq \frac{g'_a - m_b}{g_a - m_b}$$

respectively. But this cannot be. In fact, given the ordering among payoffs, $(g'_a - m_b)(g'_b - g_a) > (g_b - g'_b)(g_a - g'_a)$, $(g'_a - m_b)(g_b - g'_b) + (g'_a - m_b)(g'_b - g_a) > (g'_a - m_b)(g_b - g'_b) + (g_b - g'_b)(g_a - g'_a)$, $(g'_a - m_b)(g_b - g_a) > (g_b - g'_b)(g_a - m_b)$, and hence

$$\frac{g'_a - m_b}{g_a - m_b} > \frac{g_b - g'_b}{g_b - g_a} \tag{7}$$

QED

Theorem 3.3 *There is no equilibrium Γ where both Ie and Ee have strictly positive probability.*

Proof. Assume that σ is such an equilibrium. Then the following equation has to be true

$$\sum_{c_2 \in C_2} \sigma_2(c_2)u(Ie, c_2) = \sum_{c_2 \in C_2} \sigma_2(c_2)u(Ee, c_2)$$

which amounts to

$$\sigma_2(A) = \frac{g'_a - m_b}{g_a - m_b}$$

This means that $1 > \sigma_2(A) > 0$, hence in this equilibrium player 2 is indifferent between strategies A and B, and this means

$$\sum_{c_1 \in C_1} \sigma_1(c_1)u(c_1, A) = \sum_{c_1 \in C_1} \sigma_1(c_1)u(c_1, B) \tag{8}$$

Since $\sigma_1(Ei) = 0$ and $\sigma_1(Ii) = 0$, (8) becomes $g_a = m_b$, which is impossible. QED

Theorem 3.4 *There is no equilibrium where both Ei and Ee have strictly positive probability.*

Proof. Analogous to the preceding one. QED

How many equilibria are there? Of course there are two equilibria in pure strategies, namely $\eta = ([Ie], [A])$ and $\theta = ([Ei], [B])$, but there is also an infinite set of mixed equilibria.

Theorem 3.5 *If*

$$\pi_1(Ee) = 1 \tag{9}$$

and

$$\frac{g'_a - m_b}{g_a - m_b} \geq \pi_2(A) \geq \frac{g_b - g'_b}{g_b - g_a} \tag{10}$$

then π is a Nash equilibrium.

Proof. Consider a modified game $\Gamma^* = \{N, C_1^*, C_2, u^*\}$ where $C_1^* = \{Ei, Ie, Ee\}$, and u^* is just u after its domain has been restricted accordingly. Since Ii is strongly dominated, 3.1, every equilibrium of Γ^* is an equilibrium of Γ, and vice versa. Suppose that π is a strategy profile that satisfies conditions (9) and (10). Define ω as $p(g'_a - g'_b) + g'_b$, which is the expected payoff of both players under π. Since player 2 is clearly indifferent between A and B when player 1's strategy is $[Ee]$, in order to show that π is an equilibrium, we only need to prove the following statements:

$$\omega \geq \sum_{c_2 \in C_2} \pi_2(c_2)u(Ei, c_2) \tag{11}$$
$$\omega \geq \sum_{c_2 \in C_2} \pi_2(c_2)u(Ie, c_2) \tag{12}$$

But the conjunction of conditions (11) and (12) is equivalent to (10). Hence π is a Nash equilibrium of Γ^* and therefore of Γ as well. QED

Theorems 3.1, 3.2, 3.3, and 3.4 entail that there are no other equilibria.

4. Efficiency

Summing up, there are two equilibria in pure strategies, namely η and θ, and many mixed equilibria π. All these mixed equilibria are somehow equivalent, since they yield the same expected payoff, and they all amount to the fact that player 1 goes for the costly but unambiguous option, and player 2 has no opportunity to move. These mixed equilibria are the least efficient ones. As for the equilibria in pure strategies, η is the unique Pareto efficient equilibrium iff

$$p > \frac{g_b - g'_b}{g_b - g'_b + g_a - g'_a}$$

and θ is the unique Pareto efficient equilibrium iff

$$p < \frac{g_b - g'_b}{g_b - g'_b + g_a - g'_a}$$

Parikh's account predicts that the players will tend to converge on the more efficient equilibrium. Robert Van Rooij rejects this solution concept, claiming that it is 'unusual' [10]. This claim is quite odd. On the one hand, there is some agreement among some scholars on the view that we should expect rational players to converge on efficient equilibria in cooperative games [3, 4]. And we should bear in mind that a rational justification of a solution concept is perhaps always desirable, but not strictly necessary, as long as any included profile is (at least) a (Nash) equilibrium and is empirically adequate [1].

Yet, even if I hold that Parikh's relying on Pareto efficiency is probably the most natural choice, I will propose a line of defence which is rejected by him. Imagine that the players were allowed some preplay communication [4], before the beginning of the game, hence before player 1 has access to his private information. Since they are given the opportunity to reach an agreement over the strategy to adopt during the game, they will presumably agree to converge on the equilibrium that is the most profitable one for both, namely on the uniquely Pareto efficient one. Of course an actual occurrence of this kind of communication is unrealistic, but the players do not have to be really engaged in it in order to know what would happen in such a counterfactual situation, because this can be inferred from the structure of the game, it is a feature of the game, which is common knowledge. According to Parikh this argument is untenable for two reasons. First, if you explain successful communication in terms of preplay communication you fall into an infinite regress. Second, 'even if such an infinite regress were avoidable, the solution would certainly require a great deal of effort suggesting that languages aren't quite so efficient as they in fact are' [7].[4] I argue that both of these tenets can be rejected. The model presented here is an account of disambiguation, which is a particular phenomenon occurring in communication. I claimed that our two players could converge on a unique equilibrium, if they considered what would have happened if they had had the opportunity to reach an agreement over a coordinated plan. If this imaginary preplay communication is conceived as involving only unambiguous sentences, there seems to be no danger of an infinite regress, yet the response is the same: they would have agreed to converge on the unique Pareto efficient equilibrium. The second point is less clear to me, since the kind of reasoning that we come to attribute to our players does not seem to involve a great deal of computational effort, compared to the construction of the model itself.

The main shortcoming the Pareto-Nash solution concept that I borrowed from Parikh is that it does not explain what should happen in the limit case where

$$p = \frac{g_b - g_b'}{g_b - g_b' + g_a - g_a'}$$

and therefore both η and θ are (weakly) Pareto efficient. Being uncertain over which course of action should be chosen, our players could end up converging on one of the mixed equilibria. The reasoning is as follows. Resume the argument from preplay communication of the preceding paragraph. The upshot of a counterfactual conversation like the one described would be indeterminate, in this case, they cannot tell what they would have agreed on, just considering the structure of the game. They know for sure that they would have agreed to converge on one of the two weakly efficient equilibria, but they do not know which one, they are equally probable. Therefore, the beliefs that player 1 will end up choosing $[Ei]$ and that he will end up choosing $[Ie]$ are equally probable for player 2

player 2 deems to be equally probable the belief that player 1 will end up choosing $[Ei]$ and the belief that he will end up choosing $[Ie]$. In other words, he comes to believe that he will choose the mixed strategy σ where $\sigma_1(Ei) = \sigma_1(Ie) = \frac{1}{2}$. But this expectation is self-refuting, since, by theorem 3.2, it is not part of a Nash equilibrium. But a similar reasoning would lead 1 to expect that 2 will choose the mixed strategy $\tau_2(A) = \tau_2(B) = \frac{1}{2}$. This is nothing more than the belief that 2 does not know what to do, and hence that she will choose at random. This belief is not self-refuting, since it belongs to one of the mixed equilibria by theorem 3.5, because

$$\frac{g_a' - m_b}{g_a - m_b} > \frac{1}{2} > \frac{g_b - g_b'}{g_b - g_a}$$

And the very same reasoning that can lead 1 to form this belief can lead 2 to believe that 1 has this belief, and so on. This argument is rather unorthodox, and I take it to bee an interim solution to the problem I raised. Another line of reasoning which was promising, at least *a priori*, proved to be a dead end. I will present it anyway in the following two sections, since it gives the occasion to analyze some features of the model which are interesting in themselves, and it will help to meet a possible objection.

5. Trembling Hand Perfect Equilibria

One might hope to select a unique equilibrium arguing that in our analysis player 2 does not exploit all the evidence she has at her disposal, since in order to make a rational choice she must consider not the prior probability of a and b, but the conditional probability of those events, given that player 1 decided to utter ϕ. This suggests that we consider this a sequential game.

[4] P. 39n.

The multiagent representation [4] – also called agent-normal form [11], and agent strategic [5] – is a way to represent games in extensive form as games in strategic form, alternative to the normal representation. In the multiagent representation, there is a player, called (temporary) agent, for every information set of every player. Hence, as far as our game is concerned, player 1 is represented by two agents in the multiagent representation, say a and b. While there is only one agent for player 2, say c.

A *behavioural strategy profile* of a game in extensive form is a mixed strategy profile of its multiagent representation. A generic behavioural strategy profile of our game is $(\sigma_a, \sigma_b, \sigma_c)$, and it specifies a probability distribution for every agent of every player. The behavioural strategy profile $([I],[e],[A])$ corresponds to our Nash equilibrium η in an intuitive way, so that it can be called its *behavioural representation* [4]. Since there should not be any danger of misunderstanding, until the end of this section, I will use the names of the strategy profiles of (the normal representation of) our original game to refer to their behavioural representations. Hence, I will set $\eta = (\eta_a, \eta_b, \eta_c) = ([I],[e],[A])$, and similarly for the other equilibria.

Definition 5.1 A trembling hand perfect equilibrium of a game in extensive form is a trembling hand perfect equilibrium of its multiagent representation [4, 5]. $\lhd$

Searching for trembling hand perfect equilibria of games in extensive form, is precisely a way to search for strategy profiles that are coherent if we consider the sequential nature of a game.

Theorem 5.2 *η is a trembling hand perfect equilibrium of Γ^e*

Proof. Recall that η is a perfect equilibrium iff there exists a sequence $(\eta^k)_{k=1}^{\infty}$ such that each η^k is a perturbed behavioural strategy profile where every move gets positive probability, and, moreover

(i)

$$\lim_{k\to\infty} \eta_s^k(d_s) = \eta_s(d_s) \quad \forall s \in S \quad \forall d_s \in D_s$$

(ii)

$$\eta_s \in \text{argmax}_{\tau_s \in \Delta(D_s)} \sum_{d\in D} \left(\prod_{r\in N-s} \eta_r^k(d_r)\right) \tau_s(d_s) u(d) \quad \forall s \in S$$

where $S = (a,b,c)$ is the set of all information states of all players, and, for each $s \in S$, D_s is the set of moves available to the relevant player in state s, and $D = \times_{s\in S} D_s$. It is not difficult to find a sequence satisfying these criteria. Set

$$\xi = \frac{(1-p)(g_b - g_a)}{p(g_a - m_b)}$$

Then $\forall k \in (1,2,3,...)$, if $\xi \geq 1$,

$$\eta_a^k(I) = \frac{2k-1}{2k} \quad \eta_b^k(i) = \frac{1}{2k\xi} \quad \eta_c^k(A) = 1 - \frac{g_a - g_a'}{k(g_a - m_b)}$$

If $\xi < 1$, instead, set

$$\eta_b^k(i) = \frac{1}{2k}$$

and the rest as before. You can see at a glance that these sequences satisfy condition (i). Consider now the expected payoff for player 1 when he is in state a and is planning to make move $\tau_a \in \Delta(D_a)$, and all moves at all other states are made according to scenario η^k. It is equal to

$$\sum_{d-a\in D-a} \left(\prod_{r\in N-a} \eta_r^k(d_r)\right) \times \left[\tau_a(I)u(d_{-a}, I) + (1-\tau_a(I))u(d_{-a}, E)\right] \tag{13}$$

We can consider (13) as a function of $\tau_a(I)$, and if we calculate the derivative of this function we get

$$p[\eta_c^k(A)(g_a - m_b) + m_b - g_a']$$

As you can easily verify, this value is either null or positive for all k, and this means that, since $\eta_a(I) = 1$

$$\eta_a \in \text{argmax}_{\tau_a \in \Delta(D_a)} \sum_{d\in D} \left(\prod_{r\in N-a} \eta_r^k(d_r)\right) \tau_a(d_a) u(d)$$

Similarly, if you consider the corresponding expected payoff for player 1 when he is in state b, i.e.

$$\sum_{d-b\in D-b} \left(\prod_{r\in N-b} \eta_r^k(d_r)\right) \times \left[\tau_b(i)u(d_{-b}, i) + (1-\tau_b(i))u(d_{-b}, e)\right]$$

regard it as a function of $\tau_b(i)$, and calculate its derivative, you get

$$(1-p)[\eta_c^k(A)(g_a - g_b) + g_b - g_b']$$

which is either null or negative for all k, because of inequality (7), and this means that, since $\eta_b(i) = 0$,

$$\eta_b \in \text{argmax}_{\tau_b \in \Delta(D_b)} \sum_{d\in D} \left(\prod_{r\in N-b} \eta_r^k(d_r)\right) \tau_b(d_b) u(d)$$

Finally, if you calculate the expected payoff for player 2, you have

$$\sum_{d-c \in D-c} \left(\prod_{r \in N-c} \eta_r^k(d_r) \right) \times$$
$$[\tau_c(A)u(d_{-c}, A) + (1 - \tau_c(A))u(d_{-c}, B)]$$

whose derivative is

$$\eta_a^k(I)p(g_a - m_b) + \eta_b^k(i)(1-p)(g_a - g_b)$$

which is either null or positive for all k, and this entails

$$\eta_c \in \text{argmax}_{\tau_c \in \Delta(D_c)}$$
$$\sum_{d \in D} \left(\prod_{r \in N-c} \eta_r^k(d_r) \right) \tau_c(d_c)u(d)$$

QED

The case of θ is completely analogous.

Theorem 5.3 *θ is a trembling hand perfect equilibrium of Γ^e*

Proof. A suitable sequence is

$$\theta_a^k(I) = \frac{1}{2k} \quad \theta_b^k(i) = \frac{2k-1}{2k} \quad \theta_c^k(A) = \frac{g_b - g_b'}{k(g_b - g_a)}$$

if $\xi \geq 1$, and

$$\theta_a^k(I) = \frac{\xi}{2k} \quad \theta_b^k(i) = \frac{2k-1}{2k} \quad \theta_c^k(A) = \frac{g_b - g_b'}{k(g_b - g_a)}$$

if $\xi < 1$. QED

As for the mixed equilibria the case is simpler.

Theorem 5.4 *The mixed equilibria π are trembling hand perfect in the extensive form of the game*

Proof. Since $1 > \pi > 0$, we can set $\pi_c^k(A) = \pi_c(A)$, and

$$\eta_a^k(I) = \frac{1}{2k} \quad \eta_b^k(i) = \frac{1}{2k\xi}$$

whenever $\xi \geq 1$, and

$$\eta_a^k(I) = \frac{\xi}{2k} \quad \eta_b^k(i) = \frac{1}{2k}$$

otherwise. QED

Theorems 5.2, 5.3, and 5.4 show that all of the Nash equilibria can be legitimately regarded as tenable, even if take into account the fact that the players do not act simultaneously, and hence that player 2 must update her beliefs upon the evidence that player 1 chose the ambiguous utterance ϕ. This is indeed trivial for the pure equilibria, but not for the mixed ones. The question is, what does 2 believe when she sees that 1 has uttered ϕ? But an answer to this question is relative to a strategy profile. For example, under η, if 1 utters ϕ, she knows for sure that she is in her upper node, since the conditional probability of this event, upon the evidence that he has chosen either I or i, is equal to 1. But what does she believe in the same situation under one of the mixed equilibria, where all the nodes in her information set have null prior probability, and hence the traditional Bayesian theory leaves the corresponding conditional probability undefined? Many of the refinements of the Nash equilibrium concept are an attempt to give an answer to this question. The reasoning behind the notion of trembling hand perfect equilibrium in extensive games is this: when a player comes to know that an event with null prior probability has actually occurred, she believes that this was due to a mistake made by one of the other players, when performing his strategy. Then she updates her beliefs, assuming that all the possible mistakes had an infinitesimal prior probability.

6. Proper Equilibria

Summing up, theorems 5.2, 5.3, and 5.4 strengthen the overall strategy of this paper, which amounts to the claim that the players will converge on the unique Pareto efficient equilibrium, whenever there is one, and on one of the mixed equilibria otherwise. For completeness, I will mention another feature of the model which is less reassuring. The concept of proper equilibrium builds on the idea behind trembling hand perfect equilibrium, and adds the restriction that less costly mistakes have a higher probability than more dangerous ones [4].

Definition 6.1 A mixed strategy profile σ is an **ε-proper equilibrium** iff all pure strategies get strictly positive probability and, for every player i and any pair of pure strategies c_i and e_i in C_i,

$$\text{if} \quad u_i(\sigma_{-i}, [c_i]) < u_i(\sigma_{-i}, [e_i]), \quad \text{then} \quad \sigma_i(c_i) \leq \varepsilon \sigma_i(e_i)$$

◁

Definition 6.2 A mixed strategy profile σ is a **proper equilibrium** iff there is a sequence $(\varepsilon^k, \sigma^k)_{k=1}^{\infty}$ such that

$$\lim_{k\to\infty} \varepsilon^k = 0, \quad \lim_{k\to\infty} \sigma_i^k(c_i) = \sigma_i(c_i), \quad \forall i \in N, \forall c_i \in C_i,$$

and, for every k, σ^k is an ε^k-proper equilibrium. ◁

Proper equilibria are usually applied to the normal representation of games in extensive form. It is probably evident that η and θ are proper equilibria, the interesting case is the following one.

Theorem 6.3 *A mixed strategy profile π such that $\pi_1(Ee) = 1$ is a proper equilibrium of Γ iff*

$$\pi_2(A) = \frac{p(g'_a - m_b + g'_b - g_b) + g_b - g'_b}{p(g_a - m_b + g_a - g_b) + g_b - g_a}$$

I will not provide the proof here, I only remark that for every game, there are three and only three equilibria, and that the suitable value for $\pi_2(A)$ is always included in the open interval

$$\left(\frac{g_b - g'_b}{g_b - g_a}, \frac{g'_a - m_b}{g_a - m_b}\right)$$

I admit that this fact is not welcome, since it weakens the strategy adopted so far. I just take it as a reason for not adopting proper equilibrium as a solution concept in disambiguation games.

7. Conclusion

Summing up, the substance of this work is a new game-theoretic analysis of the capacity humans have to communicate using ambiguous expressions. The background hypothesis is that these tasks are accomplished because humans are rational creatures, and, when two people are involved in a conversation, they crucially capitalize on this fact, assuming that it is common knowledge. I built on ideas first developed by Prashant Parikh, raising some objections that led me to modify his models.

I built a game of imperfect information in extensive form, where a hearer and a speaker are the two players, the speaker has some private information, and his task is to convey this piece of information to the hearer. Here lies the main difference between my analysis and Parikh's, since, in his model, the relevant private information of the speaker is the intended meaning of his speech act. I argued that my reform renders the theory more natural and conceptually simpler.

The examples I chose as sample cases were simpler to analyze than more general cases, because of the structural features of the resulting model. In the end I retain Parikh's conclusion that speakers tend to focus on efficient equilibria, but I also proposed a solution to a problem that had been left open, namely, the strategy adopted by the speakers when there is not a unique efficient equilibrium. I argued that, in this case, the speaker goes for the ambiguous expression, which is costly, but safe. The argument I used to back both of these tenets hinges on the idea that the players are able to guess the joint strategy they would agree on, were they allowed some preplay communication before the beginning of the game. This kind of argument is not new. It is crucial that the players do not really need to entertain this kind of communication in order to know what would ensue from it. Yet, I acknowledged that my argument is partially unorthodox, from the point of view of the existing literature.

I also showed that the relevant equilibria are plausible even if we consider that a conversation is sequential in nature, proving that they are trembling hand perfect. And I ended stating, omitting the proof, that not all the equilibria are proper, which I take to be an unwelcome result.

Now the task is to extend this analysis to other, more general and more complex cases, and check whether the claims that have been put forward here have a wider application.

References

[1] N. Allott. Game theory and communication. In A. Benz, editor, *Game Theory and Pragmatics*. Macmillan, Palgrave, 2006.

[2] J. Harsanyi and R. Selten. *A General Theory of Equilibrium Selection in Games*. MIT Press, Cambridge Massachusetts, 1988.

[3] R. Myerson. *Game Theory: Analysis of Conflict*. Harvard University Press, Cambridge Massachusetts, 1991.

[4] M. Osborne and A. Rubinstein. *A Course in Game Theory*. MIT Press, Cambridge Massachusetts, 1994.

[5] P. Parikh. A game-theoretic account of implicature. In Y. Moses, editor, *Theoretical Aspects of Reasoning about Knowledge*, 1992.

[6] P. Parikh. *The Use of Language*. CSLI, Stanford CA, 2001.

[7] P. Parikh. Radical semantics: A new theory of meaning. *Journal of Philosophical Logic*, 35:349–391, 2006.

[8] W. Quine. *The Ways of Paradox and other Essays*. Harvard University Press, Cambridge Massachusetts, 2 edition, 1976.

[9] R. V. Rooij. Signalling games select horn strategies. *Linguistics and Philosophy*, 27:493–527, 2004.

[10] R. Selten. Reexamination of the perfectness concepts for equilibrium points in extensive games. *International Journal of Game Theory*, 4:25–55, 1975.

[11] M. Spence. Job market signaling. *The Quarterly Journal of Economics*, 87:355–374, 1975.