Sascia Pavan
s.pavan@tele2.it

# DISAMBIGUATION GAMES IN EXTENDED AND STRATEGIC FORM

> The use of language is, indeed, *the* primary manifestation of our rationality: it is the rational activity *par excellence*.
> (Michael Dummett, *The seas of Language*, 104)

The aim of this paper is to pursue the line of research initiated by Prashant Parikh (1992; 2001; 2006) which gives content and rigour to the intuitive idea that speaking a language is a rational activity. Parikh 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 over his, yet many of the key ideas employed are left unchanged. I analyse this model in detail, describing some of its formal features. I conclude showing where the model proves to be wanting, sketching a promising direction for further research.

The case I want to analyse concerns sentences like

(1) Every ten minutes a man gets mugged in New York (Parikh 2001).

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), I think that 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

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

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.

My aim is 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. If modelled in game-theoretic terms, 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 $\varphi$. As is customary in game theory, I will imagine that player 1 is male, and player 2 is 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. I believe that here lies the main shortcoming of Parikh's model. The task of player 2 is to guess what the intended meaning is, therefore 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 $B$. 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 $p\times g_a + (1–p)\times m_a < p\times m_b + (1–p)\times g_b$.[1] 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*. 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 $A$.[2] 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 $B$.

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 (1976: 158-176), it reduces the grade of *intentional involvement*. 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, I guess that 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 $b$. I assume that $1>p>0$, otherwise the solution of the game would be trivial. How can she decide which is the best choice? 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 $g_a > p\times m_b + (1–p)\times g_b$, or whether $g_a < p\times m_b + (1–p)\times g_b$? No, because the fact

---

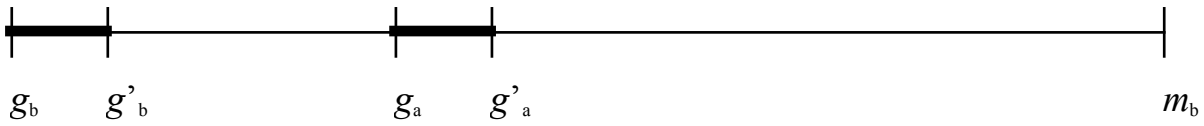
[1] One might say that payoffs have to be multiplied by *conditional* probabilities, not *prior* probabilities. In other words that the factor $p$ has to be the conditional probability she assigns to the event that he wants to convey $A$, based on the evidence provided by his actual move. This line of reasoning requires that our game be a sequential game (Myerson 1991: chapter 4). In any case this is not Parikh's approach. More on sequential games later on.

[2] Am I assuming that our players are logically omniscient? Well, of course I am.

that *a* has prior probability *p* does not, by itself, entail anything about the probability of the event that 1 wants to convey a certain message.

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, $\mu_a$ if he is in situation *a*, $\mu_b$ 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, there is no possibility of a misunderstanding. Following Parikh, I will assume that the payoff is given by the net value of the information minus a 'cost' which is proportional to the length of the sentence (Parikh 2001: 30-31). Hence, $g_a$ 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 $\mu_a$ 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 $\mu_b$ in situation *b*, then the payoff will be $g'_b=v_b-c'$, if, for the sake of simplicity, we assume that the cost involved by $\mu_a$ and $\mu_b$ 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 $g'_b>g'_a$.

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 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 true in most ordinary conversations, where the cost of an utterance is relatively small when compared to the net value of information.[3] The ordering among outcomes is represented by the following picture.




The model employed here requires the following ordering relations: $g_b > g'_b > g_a > g'_a > m_b$, $g'_a-m_b > g_b-g'_b$, and $g'_b-g_a > g_a-g'_a$.

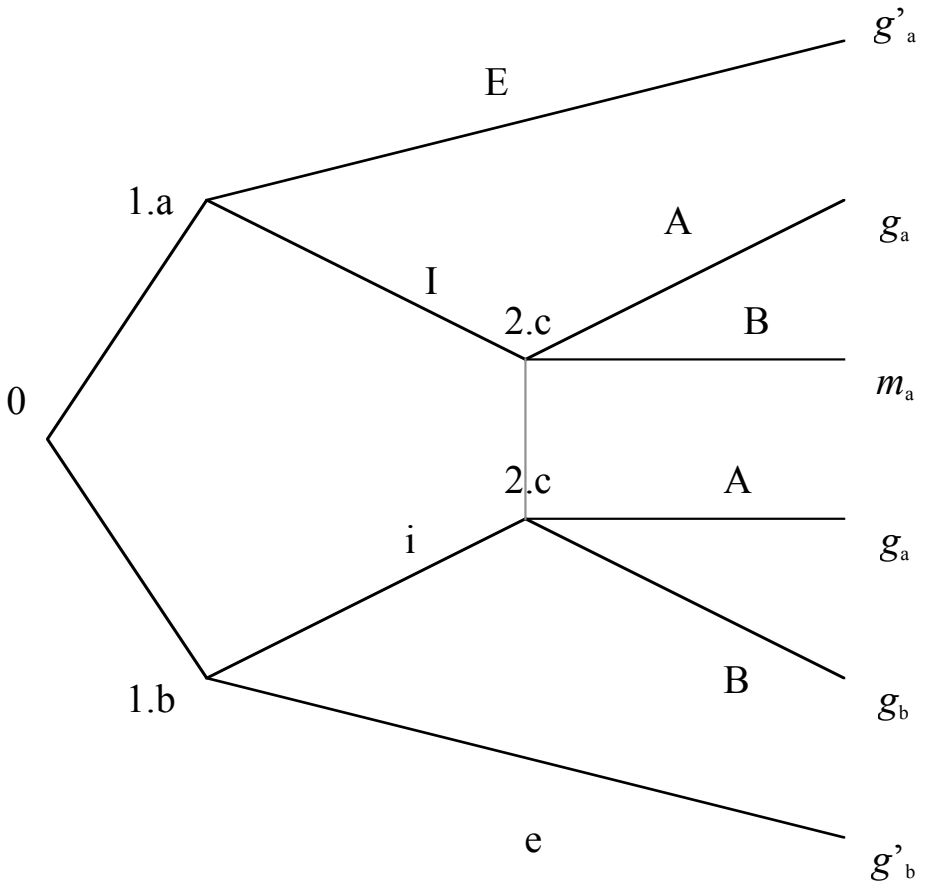
What about player 1's payoffs? I will follow Parikh on this respect as well, and I will construct my model as a coordination game where the players have the same payoffs (Parikh 2001: 29). The rationale for doing this is that when honest and rational agents communicate, they all aim at successful communication. 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.

But maybe the set of moves available to player 1 is incomplete. Perhaps we should also consider the possibility of uttering $\mu_a$ in situation *b*, and $\mu_b$ in situation *a*. Of course if player 1 uttered $\mu_a$ knowing that *A* is false, he would be lying, and, under the assumption that we are trying to analyse a case of patently honest communications – we can sensibly imagine that 2 knows that 1 wants to tell the truth, that 1 knows that 2 knows, and so on – 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

[3] See Parikh (2001: 30-31).

player 1 chose $\mu_b$ or $\varphi$. This means that, according to the model presented here, it is never rational for player 1 to choose to utter $\mu_a$ in situation *b*. In technical terms, any strategy profile where the speaker utters $\mu_a$ in situation *b* or $\mu_b$ 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 $\mu_a$, 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 $\mu_a$ in situations where he knows that *B*, but he does not want 2 to know.

Now we have all the elements to build our game. I will first construct it as a game of imperfect information in extensive form,[4] which has the structure of a tree, as is shown in the figure below.[5]




The root is a chance node, hence *a* and *b* are chance events with prior probability *p* and 1-*p*, respectively, and I will assume that 1>*p*>0. If player 1 is in situation *a*, he can utter either $\varphi$ or $\mu_a$, and we can label these two moves '*I*' and '*E*', respectively – were '*I*' stands for 'implicit' and '*E*' for 'explicit'. If he is in situation *b*, he can choose between $\varphi$ and $\mu_b$, and we can call these alternative

[4] Most of the notation and terminology employed in this paper is borrowed from Roger Myerson (1991: chapters 2-5).

[5] 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, since he argues that player 2 cannot construct anything before 1's utterance (2001: 83), 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, and I think we should not. And even Parikh seems to subscribe this view (2001: 83).

moves '*i*' and '*e*'. Player 2 has a chance to move only if the game is in one of the states labelled '2.c', which have the same label and are linked by a light grey line to manifest the fact that she is not able to distinguish them, technically speaking they belong to the same *information set* (Myerson 1991: 37-46). 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 f course we can conceive of cases where an ambiguous sentence admits of more than two readings.

The *normal representation* (Myerson 1991: 46-51) 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 $\mathbf{R}$. It satisfies the pattern shown in this table:

| | *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* | $p\times g_a + (1-p)\times g'_b$ | $p\times m_b + (1-p)\times g'_b$ |

I will first show that strategy *Ii* is strongly dominated, which entails that no strategy profile $\tau$ where $\tau_1(Ii)>0$ is a Nash equilibrium (Myerson 1991: 148).

According to the standard definition, *Ii* is strongly dominated if and only if $\exists\sigma_1\in\Delta(C_1)$ such that

(3)
$$u(Ii,A) < \sigma_1(Ei)u(Ei,A) + \sigma_1(Ee)u(Ee,A) + \sigma_1(Ie)u(Ie,A) + (1-\sigma_1(Ei)-\sigma_1(Ee)-\sigma_1(Ie))u(Ii,A)$$

and

(4)
$$u(Ii,B) < \sigma_1(Ei)u(Ei,B) + \sigma_1(Ee)u(Ee,B) + \sigma_1(Ie)u(Ie,B) + (1-\sigma_1(Ei)-\sigma_1(Ee)-\sigma_1(Ie))u(Ii,B)$$

Inequalities (3) and (4) are equivalent to

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

and

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

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). For instance, we can set $\sigma_1(Ee)=0$ and $\sigma_1(Ie)=1-\sigma_1(Ei)$ and then solve the equation

$$\frac{\sigma_1(Ei)}{1-\sigma_1(Ei)} = \frac{1}{2}\left(\frac{(1-p)(g'_b - g'_a)}{p(g_a - g'_a)} - \frac{(1-p)(g_b - g'_b)}{p(g'_a - m_b)}\right) + \frac{(1-p)(g_b - g'_b)}{p(g'_a - m_b)}.$$

Next I show that there is no equilibrium where both *Ei* and *Ie* have strictly positive probability. Assume that $\sigma$ is such an equilibrium. Then the following equation has to be true:

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

Since $\sigma_2(B)=1-\sigma_2(A)$, this is equivalent to

$$\sigma_2(A) = \frac{p(g'_b + g'_a - g_b - m_b) + g_b - g'_b}{p(2g_a - g_b - m_b) + g_b - g_a}.$$

Since $\sigma_2(A)$ cannot be greater than 1, this is true only if

$$p(2g_a - g_b - m_b) + g_b - g_a \geq p(g'_b + g'_a - g_b - m_b) + g_b - g'_b,$$

hence only if

$$p \geq \frac{g'_b - g_a}{g'_b - g_a - (g_a - g'_a)},$$

and this cannot be, since $1 > p$.

Next I show that there is no equilibrium where both *Ie* and *Ee* have strictly positive probability. Assume that $\sigma$ 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

$$\sum_{c_1 \in C_1} \sigma_1(c_1) u(c_2, A) = \sum_{c_1 \in C_1} \sigma_1(c_1) u(c_2, B) \quad (7)$$

Since $\sigma_1(Ei)=0$ and $\sigma_1(Ii)=0$, (7) becomes $g_a=m_b$, which is impossible. A similar proof shows that there is no equilibrium where both *Ei* and *Ee* have strictly positive probability.

How many equilibria are there? Of course there are two equilibria in pure strategies, namely $\eta=([Ie],[A])$ and $\theta=([Ei],[B])$, but there are also infinitely many mixed equilibria $\pi$ where

$$\pi_1(Ee)=1 \quad (8)$$

and

$$\frac{g'_a - m_b}{g_a - m_b} \geq \pi_2(A) \geq \frac{g_b - g'_b}{g_b - g_a} \quad (9)$$

The proof of this fact is as follows. First of all, observe that, 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

$$\frac{g'_a - m_b}{g_a - m_b} > \frac{g_b - g'_b}{g_b - g_a} \quad (10)$$

Next, we can 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, every equilibrium of $\Gamma^*$ is an equilibrium of $\Gamma$, and vice-versa. Suppose that $\pi$ is a strategy profile that satisfies conditions (8) and (9). Define $\omega$ as $p(g'_a-g'_b)+g'_b$, which is the expected payoff of both players under $\pi$. Since player 2 is clearly indifferent between *A* and *B* when player 1's strategy is [*Ee*], in order to show that $\pi$ is an equilibrium, we only need to prove the following statements:

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

$$\omega \geq \sum_{c_2 \in C_2} \pi_2(c_2) u(Ie, c_2) \quad (12).$$

But the conjunction of conditions (11) and (12) is equivalent to (9). Hence $\pi$ is an equilibrium of $\Gamma^*$ and therefore of $\Gamma$ as well.

All these 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. Therefore we can gather them together and imagine that there is a unique mixed equilibrium $\pi$, where $\pi_1(Ee)=1$ and $\pi_2(A)$ takes an indeterminate value satisfying condition (9).

Summing up, there are two equilibria in pure strategies, namely $\eta$ and $\theta$, and a mixed equilibrium $\pi$. It is quite natural to claim that in a coordination game like this one, the players will tend to converge on the more efficient equilibrium.[6] This is the solution concept adopted by Parikh in his works, yet the line of defence I will propose is rejected by him. Imagine that the players were allowed some preplay communication (Myerson 1991: 108-112), before the beginning of the game,

[6] The contrary claim made by Robert Van Rooy (2004: 506) is slightly odd. Myerson (1991: 485) and John Harsanyi and Reinhard Selten (1988: 356) adopt a different view.

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 preplay 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." (Parikh 2001: 39n). 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, which has to be accomplished anyway.

The least efficient equilibrium is always $\pi$. As for the other two, $\eta$ is the unique Pareto efficient equilibrium iff

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

and $\theta$ is the unique Pareto efficient equilibrium iff

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

The main shortcoming of this analysis 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 $\eta$ and $\theta$ are (weakly) Pareto efficient. I do not envisage any quick solution employing the usual refinements of the notion of equilibrium.[7] For example, 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 $\varphi$. This suggests that we consider this a *sequential* game, and check which strategies are rational in this enriched setting. More technically, we should consider our game in extensive form and find which

[7] All equilibria of these games are trembling hand perfect. We could prove this showing that none of *Ie*, *Ei*, and *Ee* is weakly dominated (Osborne and Rubinstein 1995: 248). The structure of these proofs closely resemble the proof that *Ii* is strongly dominated. See also Parikh (2001: 39n).

*behavioural-strategy profiles* are *sequential-equilibrium scenarios* (Myerson 1991: 154-176). A behavioural-strategy profile specifies a probability distribution over the set of moves for every information state of every player. You can easily check that for every strategy profile in our game, there is exactly one corresponding behavioural-strategy profile, therefore we can say that the latter is the *behavioural representation* of the former. Unfortunately, the behavioural representations of our three equilibria $\eta$, $\theta$, and $\pi$ are sequential-equilibrium scenarios of the game in extensive form, hence this solution concept does not provide a solution to this problem. But this fact also shows that our prediction that players will converge on the unique Pareto efficient equilibrium, whenever possible, remains plausible even if we consider the sequential nature of the game. We have to bear in mind that the notion of Nash equilibrium and most of its refinements are upper solution concepts since their aim is the inclusion of all reasonable predictions about the outcome of a game, but they do not always rule out all unreasonable predictions. Given a game and an equilibrium $\sigma$ of this game, a mixed strategy $\sigma_i$ for some player *i*, can be conceived as a plausible prediction of *i*'s behaviour on the part of the other players. In our game, for example, the fact that $\pi$ is an equilibrium amounts to the fact that when 1 believes that the probability that 2 will choose *A* is $\sigma_i(A)$, in any state his best choice is to go for the unambiguous expression; and that when 2 believes that this will be 1's behaviour, she has no preference over *A* and *B*.[8] But there is no reason to believe that in all situations, there is always only one rational expectation concerning the behaviour of other rational agents. Sometimes, some particular feature of an equilibrium could make it salient so that the players will tend to converge on this equilibrium. The argument dealing with counterfactual preplay communication developed above shows that this is what we should expect in our model. Hence that our two players will converge on the unique Pareto efficient equilibrium, if there is one. For example, suppose that $\eta$ is the unique Pareto efficient equilibrium in the game. It amounts to the prediction that player 1 will choose *I* if he happens to be in state *a*, and *E* otherwise. And that player 2 will choose *A*, if given the opportunity to move. The fact that $\eta$ corresponds to a sequential equilibrium scenario shows that these choices are the best choices for both players, given the beliefs that, according to $\eta$, they have when they have the opportunity to move. When the game begins, according to this equilibrium player 1 is certain that 2 will choose *A*, hence his best choice is to choose *I* if he is in *a*, and *b* otherwise. Now suppose that 2 is given the opportunity to move. Her prior probability that 1 is in *a* is equal to *p*. What is the conditional probability that he is in *a*, given the evidence that he chose to utter $\varphi$, hence that his choice has been either *I* or *i*? It has to be equal to $p\eta_1((I)/(p\eta_1(I)+(1-p)\eta_1(i))$, hence equal to 1. Given this posterior belief, *A* is the best choice for 2. This is a triviality, but it shows that the coordinated plan to converge on the unique Pareto efficient equilibrium cannot be ruled out by an appeal to the sequential nature of the game.[9]

We can show that the behavioural representations of our three equilibria $\eta$, $\theta$, and $\pi$ are sequential-equilibrium scenarios, proving that they are *perfect* equilibria of the *multiagent representation* of our game (Myerson 1991: 61-63, 217), which in general is a logically stronger solution concept, yet admits of a simpler proof. The multiagent representation – also called agent-normal form (Selten 1975) – is a way to represent games in extensive form as games in normal form. In the multiagent representation, there is a player, called (*temporary*) *agent*, for every information state of every player (Myerson 1991: 61). 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, namely *c*.

---

[8] See the discussion in Osborne and Rubinstein (1994: 43-44).

[9] This meets an objection raised by Van Rooy (2004: 512-514).

I will show that this fact holds for the strategy profile η, in other words I will show that the behavioural representation of η is a perfect equilibrium. The behavioural representation of η specifies a move for every information state of every player. Hence, if the strategy profile η is the pair $\eta=(\eta_1,\eta_2)=([Ie],[A])$, its behavioural representation will be the triple $([I],[e],[A])$. Since there should not be any danger of misunderstanding, we can retain the same symbol and set $\eta=(\eta_a,\eta_b,\eta_c)=([I],[e],[A])$. This is clearly a strategy profile of the multiagent representation of our game.

According to the definition provided by Roger Myerson (1991: 216), η is a perfect equilibrium iff there exists a sequence $(\eta^k)_{k=1}^{\infty}$ such that each $\eta^k$ is a perturbed behavioural strategy profile where every move gets positive probability, and, moreover,

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

(ii) $$\eta_s\in\arg\max_{\tau_s\in\Delta(D_s)}\sum_{d\in D}\left(\prod_{r\in N-s}\eta_r^k(d_r)\right)\tau_s(d_s)u(d)$$, $\forall s\in S$,

where $S$ is the set of all information states of all players, hence $S=(a, b, c)$, 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, for every $\forall k\in\{1,2,3,...\}$, 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 $\xi<1$,

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

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 $\eta^k$. It is equal to

(12) $$\sum_{d_{-a}\in D_{-a}}\left(\prod_{r\in N-a}\eta_r^k(d_r)\right)\left[\tau_a(I)u(d_{-a},I)+(1-\tau_a(I))u(d_{-a},E)\right].$$

We can consider (12) as a function of $\tau_a(I)$, and if we calculate the derivative of this function we get

$$p\left[\eta_c^k(A)(g_a-m_b)+m_b-g'_a\right].$$

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

$$\eta_a \in \underset{\tau_a \in \Delta(D_a)}{\arg\max} \sum_{d \in D} \left( \prod_{r \in N-a} \eta_r^k(d_r) \right) \tau_a(d_a) u(d) .$$

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

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

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

$$(1-p)\left[ \eta_c^k(A)(g_a - g_b) + g_b - g'_b \right],$$

which is either null or negative for all *k*, because of inequality (10), and this means that, since $\eta_b(i)=0$,

$$\eta_b \in \underset{\tau_b \in \Delta(D_b)}{\arg\max} \sum_{d \in D} \left( \prod_{r \in N-b} \eta_r^k(d_r) \right) \tau_b(d_b) u(d) .$$

Finally, construct the function that corresponds to (12) in state *c*, i.e.

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

its 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

$$\eta_c \in \underset{\tau_c \in \Delta(D_c)}{\arg\max} \sum_{d \in D} \left( \prod_{r \in N-c} \eta_r^k(d_r) \right) \tau_c(d_c) u(d) .$$

The case of θ is completely analogous. 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 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 $\xi < 1$. As for π, it is simpler because $1 > \pi_c(A) > 0$, hence 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.

Going back to the case where there is no unique Pareto efficient equilibrium, my intuition is that, in such a case, player 1 would adopt strategy [*Ee*], and player 2 would be indifferent between her two strategies, and this would yield equilibrium $\pi$. Unfortunately, I am not aware of a solution concept that could justify this intuition. I guess that the source of this problem is that this model is too simplistic. Player 2 will assign a probability value to *a* and *b*, but player 1 will only have an approximate knowledge of this value, and player 2 will only have an approximate knowledge of this knowledge of the other player, and so on. This consideration leads me to the conclusion that the model presented here is only an approximation that fares well when the players assign similar prior probabilities to *a* and *b*, and the prior probability of *a* is not 'too close' to the crucial value that makes both $\eta$ and $\theta$ Pareto efficient, and that a more accurate model should be framed as a Bayesian game of incomplete information (Harsanyi 1967-68; Myerson 1991: 67-74).

## REFERENCES


Harsanyi, J.C. 1967-68. "Games with Incomplete information Played by 'Bayesian' Players. Part I", *Management Science* 14: 159-182

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

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

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

Parikh, P. 1992. "A game-theoretic account of implicature", *Theoretical Aspects of Reasoning about Knowledge*, edited by Y. Moses, Morgan Kaufmann, California.

Parikh, P. 2001. *The Use of Language*, CSLI Publications, Stanford California

Parikh, P. 2006. "Radical Semantics: A New Theory of Meaning", *Journal of Philosophical Logic* 35: 349-391

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

Selten, R. 1975. "Reexamination of the Perfectness Concepts for Equilibrium Points in Extensive Games", *International Journal of Game Theory* 4: 25-55

Van Rooy, R. 2004. "Signalling Games Select Horn Strategies", *Linguistics and Philosophy* 27: 493-527