ANNALES D'ÉCONOMIE ET DE STATISTIQUE. – N° 25/26 — 1992

Information and Rationality: Some Comments

Sylvain SORIN *

ABSTRACT. – The purpose of this note is to analyze some properties of two concepts that are widely used in game theory, namely information and rationality, and more specifically to clarify their relation. Starting with the basic approach where rationality is "maximisation given the information", one is led, as soon as there are many players, to different models depending on the information space. We explicitely describe an "open model" (the knowledge of the players is not even public but coherent), a "closed model" (where either the game itself is common knowledge or even more the rationality of the players is common knowledge) and a "local model" (where at a certain state of the world rationality emerges).

Information et rationalité : quelques réflexions

RÉSUMÉ. – On propose ici quelques réflexions sur le rapport entre les concepts d'information et de rationalité. Dès qu'il y a plus d'un agent, l'information peut porter sur la « nature », les croyances ou la rationalité des autres agents. Cela conduit à dégager différents niveaux de modèles cohérents.

---

* S. SORIN : ENS, Département de Mathématiques et d'Informatique et SEGMI, Université Paris X. A first version of this work was presented in October 1990 at the "International Conference on Organisations and Games", CIRM, Luminy, France. I deeply benefited from the comments of the participants, especially Françoise FORGES. Some of the ideas introduced here have been independently developped in a much more formal way in a recent working paper by AUMANN and BRANDENBURGER [1991].

1 Introduction

The purpose of this note is to analyze some properties of two concepts that are widely used in game theory, namely *information* and *rationality*, and more specifically to clarify their relation. This topic has been recently studied in numerous papers, for a very broad survey see BINMORE and BRANDENBURGER [1989]. It should be clear that the comments I propose are deeply influenced by these works, especially AUMANN [1988] and BINMORE [1987-88]. Rather than adding a new original approach I will focus on the compatibility of different hypotheses. The analysis will be mainly devoted to the notion of equilibrium and we will consider several conditions under which it is meaningful.

The simplest way to define "rationality" is to start with the basic one-person case where to behave rationaly means to maximize given the "information". It is thus clear that the two concepts are dependent, or more precisely that rationality, in this very naive framework, can be defined only after information. But already in this context a first difficulty occurs since information does not simply mean "private information" (*i. e.* messages known to the player, this would amount to modify the strategy set) but information about the world to which the player belongs, *i. e.* within the model one has to specify the information of the player about the model; at this point at least two approaches can be defined, the first one being simply descriptive, the second one being bayesian. Note that the bayesian view point is a specific example of connection between information and rationality, since it is based on the following principle: a rational player should have a prior probability distribution on unknown parameters or facts – or at least on all those that can affect the result (Savage's small world axiom). This requirement contains in fact two aspects: one of "consistency" (the player knows what are the relevant parameters – there are no "hidden variables" –) to define a set, the other being the more usual existence or construction of a probability on it.

When dealing with more than one player new difficulties appear since, if one tries to apply the above analysis, one has to incorporate into the surrounding world the other players, hence their behavior or more generally their "types". One is thus led to a model whose description should include information hypotheses about rationality (following KANEKO [1987*b*] one could call *factual* the information about the basic data and *structural* the information about the nature of the players) and now information emerges as a second level concept, to be defined after rationality. We reach here a kind of circular approach. To know whether a player is rational, one has to specify his information and this information should include statements about the other players rationality. It follows that the right way to formalize a model may be rather global than deductive and then some requirement of coherence have to be satisfied, all this being very similar to the definition of a language, where there is no starting point. Within this very abstract scheme, one nevertheless sees the possibility of different levels or degrees of rationality (cf. MERTENS [1989, p. 2-3] related to different choices of information contexts.

Let us add a last preliminary remark: it is in a sense better to speak about information than about knowledge since often a knowledge operator is defined trough an information process and a rational behavior ("deduction"). I know the event "*a*" means very often: I read "*a*" in the newspaper "*x*" and "*x*" is a good reference. In particular an information can be wrong without involving any question of irrationality.

We will now examine several hypotheses under which the equilibrium concept makes sense. Three models will be considered that we call *open*, *closed* and *local* (BRANDENBURGER and BINMORE [1989, p. 53] introduced the terminology of "closed universe", the concept of "local" is defined in AUMANN [1988], *see* also KANEKO [1987*a*] who defines two formalizations corresponding to a "complete information" and a "naive interpretation"). Finally we shall add few comments on the consequences of these different approaches.

We will use the framework of a game form that allows for a more precise discussion. Hence each player i has a strategy set S_i and F is an outcome function from the product S of the S_i's to some outcome set Z. Each player i has also a preference relation over Z that we represented by a real function g_i and that induces a payoff function G_i on S. We shall denote by B_i the best reply correspondence of i, defined on $S_{-i} = \prod_{j \neq i} S_j$ with values in S_i. All the discussion will be in the framework of a one shot strategic form game, hence we do not distinguish between pure or mixed strategies, nor between "bayesian" or strategic equilibrium. An equilibrium is then a vector s with s_i in $B_i(s_{-i})$ for all i.

2 Open Model

The concept of equilibrium considered here is very similar to what AUMANN [1989] calls "population equilibrium". We assume that each player is rational in the sense that he maximizes given his information. Basically the strategies of the other are given but what player i needs to know, to behave rationally in using s_i, is only the value of the correspondence B_i at s_{-i}, not s_{-i} itself, nor g_i or F, *a fortiori* not G_{-i}. This corresponds to an adaptative behavior: s is proposed (or selected) and each player gets the minimal information that allows him to accept the proposal he received. Giving him some more information one can consider that, facing a situation, a player looks at the outcomes induced by his strategies, *i. e.* $F(s_{-i}, .)$, and knowing his own payoff uses a best answer. Formally there are two aspects: one is a decentralized process of information that allows each player to be "bayesian rational", the second is a coherence or compatibility condition between strategies and "signals".

It is worthwhile to remark that in order to play a "sure" (maxmin) strategy, a player needs to know more, *i. e.*, his own payoff function, but no coherence is needed on the information of the different players.

This model is descriptive in the sense that it exhibits a behavior without explaining where it comes from. It is deeply related to one of the basic arguments for equilibrium: if a "stable situation" is reached, it has to be an equilibrium. Here each player (animal?, plant?, cell?) adapts himself to the environment, supposed to be invariant. The only "information" corresponds to results of past experiments and memory or in a biological context, natural selection. (Recall that conversely such a process of adaptative behavior does not necessary lead to an equilibrium when there is no experimentation, see *e. g.* SELTEN'S horse with inconsistent priors, BINMORE [1987]).

Note that in this open model a player may even ignore that he is playing a game, namely that his opponents have strategical opportunities, this being reflected in the invariance aspect. It follows that no factual common knowledge is assumed, say about the game form, not even public knowledge of the strategy s but a coherence condition is needed (the information for each player is as if s was played).

Intermediary cases can also be observed like common knowledge of the game form, public knowledge of s, etc., but as long as G_{-i} is unknown no "justification" or "explanation" of the equilibrium is available, in particular from the point of view of a single player who basically ignores that s is actually an equilibrium.

3 Closed Model

We consider here a bayesian approach hence the system is closed: all relevant parameters are identified and there is a common reference (*consistency* in HARSANYI'S sense), so that we are in a small world in SAVAGE'S terminology. Nevertheless there are still two levels, say factual and structural.

3a) The first one corresponds to a common knowledge of the game parameters (strategies and payoffs) and of s. This implies that "s is an equilibrium" is common knowledge, contrary to the previous case. (Note the main difference for a player between: playing an equilibrium, knowing that he is playing an equilibrium and finally knowing that this fact is common knowledge). This model corresponds to the main approach to strategic games where it is often assumed that each player can compute the set of equilibria and that this procedure itself is common knowledge (in particular in all sequential models with bargaining about equilibria in the subgames). Note also that if s is only public knowledge, the fact that it is an equilibrium is also public knowledge, assuming factual public knowledge.

3 b) The second approach corresponds to a discussion on a proposal s. Even if s is an equilibrium, and even if this is common knowledge, the question whether s will be played remains open. Here hypotheses about the behavior (rationality) of the players and the information of each player on it have to be done and what is needed to "justify" the play of s, from the point of view of coherence (*see* section 1), is common knowledge of rationality. It is easy to construct examples where mutual knowledge of rationality would not suffice to enforce the play of a pre-specified equilibrium. Note that this concept of common knowledge of rationality itself is auto referential since it roughly means rational at each information level "y" that can be obtain from information level "x" and a rational deduction at this level "x", assuming the others doing the same. (This does not seem to be reducible to the usual AUMANN's framework of a set of events with private partitions or knowledge operators). The typical example of such a situation is the bayesian approach to (correlated) equilibrium of AUMANN [1987]: if a player i adopts a bayesian approach, is rational and assumes $-i$ to behave rationally, then his distribution on S corresponds to a correlated equilibrium. (I wrote (correlated) equilibrium since I believe that in this framework it has really to be considered as an equilibrium of the extended game – the situation is different when one studies the set of such extensions or the mechanisms that the players can construct to correlate their strategies).

Note that a more general model, incorporating hypotheses concerning rationality can be obtained with the framework of 3 *a*) by introducing "types" but the crucial hypotheses to close the model are common prior and common knowledge of it.

Finally one should remark that the notion of "subjective" correlated equilibrium seems incompatible with both an open and closed model. In the first case it corresponds to a lack of consistency, recall that the opponent's payoffs are not known and rationalizability cannot be used. In the second framework either the players know that their priors may differ and then they should build the usual hierarchy of beliefs, *see* MERTENS and ZAMIR [1985], or we are not in a closed model. In fact one has to be more precise since this procedure can reach an inconsistent state in the universal belief space and moreover the players may or may not realize this fact. Somehow the notion of equilibrium itself looses sense here, since basically the players are no longer playing the same game, from their own point of view: in a two person zero sum game both players can expect a strictly positive payoff. The situation is the same in the case of secrete information: player 1 knows the event A but player 2 doesn't know this fact.

4 Local Model

We follow here AUMANN [1988] in describing a situation like a closed model, but without coherence between information and rationality. Nevertheless one can obtain by considering specific states a situation similar to an equilibrium in an open model. It is clear in such a framework that the strategies of the others are not necessarily known, but the distribution on actions can be public knowledge like in the analysis in Part 2. Moreover the fact that an equilibrium is played is not usually common knowledge, but can be mutual knowledge of some order.

In the following example due to AUMANN [1988], the payoff matrix is on the left and the matrix on the right describes the possible states of the world (with an uniform distribution on ω_i, $i = 1, \ldots 6$).

	a	b
A	1,1	0,0
B	0,0	2,2

	a	a	b
A	ω_1	ω_2	
A	ω_3		ω_4
B		ω_5	ω_6

FIGURE 1

At ω_1 we observe an equilibrium in an open model. Moreover A a is public knowledge but player 1 does not that player 2 knows that player 1 plays a. Note that the model is not closed since player 1 is not rational at ω_3.

A similar situation can occur with a correlated equilibria.

In the game above, the two by two submatrix in the Left-Top corner corresponds to a correlated equilibrium in an open model.

Remark that in this current presentation, without coherence, there is no "canonical representation": looking only at the distribution induced on S by the strategies would prevent the analysis of the states where one player is not rational or affect the degree of irrationality of the information system.

	a	b
A	0,0	3,1
B	1,3	0,0

	a	b	a	b
A	ω_1	ω_2	ω_3	ω_4
B	ω_5			
A	ω_6			
B	ω_7			

FIGURE 2

5 Further Remarks

5a) We already noticed that to model the fact that players play in equilibrium or the fact that they moreover know it requires very different hypotheses, say 2 versus 3*a*. If, moreover, one wants the players to know that the equilibrium will be played, then 3*b* is necessary. In fact this is exactly the kind of hypotheses needed in two other different procedures, namely "backwards induction" and "rationalizability" (BERNHEIM [1984], PEARCE [1984]), when one assume that they are performed by all players and their result are common knowledge. (Obviously in any specific finite game mutual knowledge of some order will suffice). Here also a difference has to be done between the fact that a game is, say dominance solvable, and the fact that the players know it. In particular if one wants to say that in such a case the game is, from the point of view of the players, determined, one needs factual and structural common knowledge hypotheses.

5b) In particular in the context of games in extensive form and with perfect information, assuming rules and payoffs common knowledge, one can imagine a more specific description of the kind: at each node every player is rational or not and has some information about the behavior of his opponents in the subgame. In this framework it is known that rationality at every node and common knowledge of this fact can be incoherent: consider a subtree dominated for one player. Completely different conclusions would be drawn when dealing the with agent normal form.

In the following example T is never part of a subgame perfect equilibrium but is compatible with an equilibrium for $x \leqslant 1$.

One could imagine that at the initial node player 2 is considered as a rational player anticipating T and that his behavior if B depends on y.

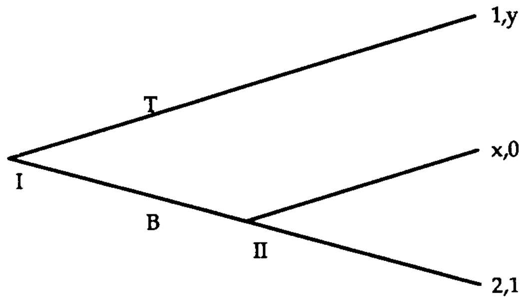

FIGURE 3

Note that an equilibrium induces a path compatible with rationality at each of its nodes and common knowledge of this fact and a subgame perfect equilibrium corresponds to a maximal such set of "rational nodes".

In both following examples there is only one maximal "rational" path (one node in the first, all in the second) that gives paradoxal results, but the interpretation changes if one assume "types" in example 4 (Rosenthal) or alternance of the same two players in example 5 (Kalai).

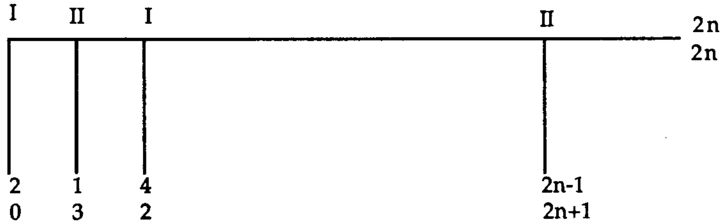

FIGURE 4

5c) These hypotheses of common knowledge (factual or structural) can be difficult to justify and this may sometimes weaken the use of the model, due to the discontinuity at the limit of the mutual knowledge operator, see for example NEYMAN [1989] or RUBINSTEIN [1989]. In the first case the players solve dependent maximization problems, hence a propagation effect arises but there is independence at the limit; in the second case the mass of the consistent probability disappear at ∞.

5d) In particular in situations where there is no common knowledge on the basic data, like the length of the game, and even when a player does not know his own payoff (uncertainty on the discount factor) other definitions, like uniform equilibria, appears to be much more robust and useful.

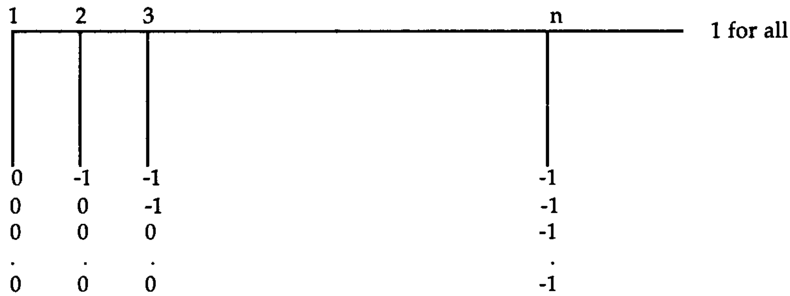

FIGURE 5

5 e) Similarly in some examples, the elimination of weakly dominated strategies seems very counter intuitive, when considering the underlying information.

1,1	0,0	0,1
0,0	1,1	0,1
1,0	1,0	0,0

FIGURE 6

Depending whether one player anticipates the other elimination, or both eliminates simultaneously all or some weakly dominated strategies the results are completely different namely:

1,1	0,1
1,0	0,0

0,1
0,1
0,0

0,0

FIGURE 7

5 f) In the same vein, recall there are situations where the only equilibrium is, especially in a closed model, difficult to justify (*see* AUMANN and MASCHLER [1972]).

1,0	1,1
0,1	2,0

FIGURE 8

The only equilibrium gives 1 to player 1 but requires him to play (1/2, 1/2) while T guarantees 1.

5g) The choice of a model corresponding to section 2 or 3 has important consequences on the interpretation of the results. Consider a case with unique equilibrium payoff (or asymptotically uniqueness like in AUMANN and SORIN [1989]): in the first framework one has a description of what will be observed, if an equilibrium emerges; in the second paradigm it is common knowledge that any equilibrium will lead to this payoff. In particular the theories leading to a selection of a unique equilibrium consider a closed model with moreover some common knowledge of "hyper rationality".

5h) Similarly the interpretation of a perturbed game (like in AUMANN and SORIN [1989]) can either be: there is common knowledge about the distribution on types and one analyzes the situation at the state of the world where both players are rational; or every player maximizes against some "empirical" or "prior" perturbed strategy of the opponent and there is some coherence.

5i) Finally the self-enforcing aspect of the equilibrium takes a different meaning according to the framework. In an open model, it just means coherence : s_{-i} is suggested to $-i$ and s_i to you, you agree; now s_{-i} is to be played, are you willing to play s_i? AUMANN [1990] criticism that requires, say in the two players, two by two case, not only s_1 and s_2 to be in equilibrium but t_1 to be better for 2 against t_2 (so that the fact that 2 wants 1 to play s_1 shows in fact his willingness to play s_2), cannot be expressed in an open model (1 does not know 2's payoff) where, from my point of view the notion of self-enforcement does still make sense.

• References

AUMANN, R.-J. (1987). – "Correlated Equilibria as an Expression of Bayesian Rationality", *Econometrica*, 55, pp. 1-18.

AUMANN, R.-J. (1988). – "Preliminary Notes on Irrationality in Game Theory", preprint.

AUMANN, R.-J. (1989). – "Perspectives on Bounded Rationality", preprint.

AUMANN, R.-J. (1990). – "Nash Equilibria are Not Self-Enforcing", *Economic Decision-Making: Games, Econometrics and Optimisation*, J. J. GABSZEWICZ, J.-F. RICHARD and L. A. WOLSEY eds, Elsevier Science Publishers.

AUMANN, R.-J. and BRANDENBURGER, A. (1991). – "Epistemic Conditions for Nash Equilibrium", preprint.

AUMANN, R.-J. and MASCHLER, M. (1972). – "Some Thoughts on the Minimax Principle", *Management Science*, **18/5** Part II, pp. 54-63.

AUMANN, R.-J. and SORIN, S. (1989). – "Cooperation and Bounded Recall", *Games and Economic Behavior*, 1, pp. 5-39.

BERNHEIM, B. D. (1984). – "Rationalizable Strategic Behavior", *Econometrica*, 52, pp. 1007-1028.

BINMORE, K. (1987-1988). – "Modeling rational Players I and II", *Economics and Philosophy*, 3, pp. 179-214 and 4, pp. 9-55.

BINMORE, K. and BRANDENBURGER, A. (1989). – "Common Knowledge and Game Theory", preprint.

KANEKO, M. (1987*a*). – "The Conventionnally Stable Set in Noncooperative Games with Limited Observations I: Definitions and Introductory Arguments", *Mathematical Social Sciences*, 13, pp. 93-128.

KANEKO, M. (1987*b*). – "Structural Common Knowledge and Factual Common Knowledge", preprint.

MERTENS, J.-F. (1989). – "Equilibrium and Rationality: Context and History-Dependence", CORE D.P. 8927.

MERTENS, J.-F. and ZAMIR, S. (1985). – "Formulation of Bayesian Analysis for Games with Incomplete Information", *International Journal of Game Theory*, 14, pp. 1-29.

NEYMAN, A. (1989). – "Games without Common Knowledge", preprint.

PEARCE, D. G. (1984). – "Rationalizable Strategic Behavior and the Problem of Perfection", *Econometrica*, 52, pp. 1029-1050.

RUBINSTEIN, A. (1989). – "The Electronic Mail Game: Strategic Behavior Under 'Almost Common Knowledge'", *The American Economic Review*, 79, pp. 385-391.