On knowledge games

J. M. LASRY, J.M. MOREL AND S. SOLIMINI

ABSTRACT. We give a formalization of the "knowledge games" which allows to study their decidability and convergence as a problem of mathematics. Our approach is based on a metalemma analogous to those of Von Neumann and Morgenstern at the beginning of Game Theory. We are led to definitions which characterize the knowledge games as objects in standard set theory. We then study rigorously the most classical knowledge games and, although we also prove that the *common knowledge* in these games may be uncomputable, show their decidability in a simple way.

RESUMEN. Proponemos una formalización de los "juegos de saber" que permite el estudio matemático de su decidabilidad y convergencia. Nuestra construcción se basa en un metalema análogo a los que Von Neumann y Morgenstern desarrollaron para fundar la teoría de los juegos. Los juegos de saber quedan definidos como objetos sencillos en el marco de la teoría de conjuntos. Estudiamos luego rigurosamente los juegos de saber más clásicos y, aunque el "saber común" puede ser no calculable, demostramos la decidabilidad de estos juegos con técnicas elementales.

1. INTRODUCTION

Our aim in these notes is to propose very simple decision procedures for treating processes where several subjects exchange information in such a way that reasoning about the knowledge of the other subjects is necessary for arriving at a solution. Classical examples of such processes are given below. The characteristics of these problems, (or "puzzles", or "games") is to require a

---

very formal and strict reasoning, but which cannot easily be endowed in classical set theory because of the very specific hypothesis about the subjects: As in game theory, one has to suppose that the subjects involved in the processes are perfectly rational, that they all know the theory of the process or game if such a theory exists. More precisely, the theory must be common knowledge among the players. (One says that something is *common knowledge* among the players if everyone knows it and if everyone knows that everyone knows it, and so on).

All these features require a formalization, and in the litterature about the subject, one can find at least three directions. A first one (see for instance (ST), (MDH), (GS)) is to reason informally but rigorously about those processes, which we shall call "knowledge games". However, as these games become less intuitive, as in the case of the "sum game" described below, one feels the need for a general formulation. This may lead to the logical approach, which has been extensively and widely studied in recent years. The basic idea, often attributed to Lewis, is to define modal operators of knowledge and to fix their meaning by appropriated axiomatics. One obtains logical knowledge theories for which consistency and completeness proofs can be given, as in classical modal logic, by Kripke semantics. This has led to developments in temporal logic for the description of discrete time temporal processes (MP, BKP), and in epistemic logic for the analysis of games and distributed reasoning (H, P, HM, FHV, L, FH, CM,...) These modal systems are simple enough to allow completely formal proofs in some particular games, as the "three wise men puzzle" (AS). Now, as is well-known, to generate formal proofs in modal systems is a NP-complete problem. That means (up to now) that logical formalization can help for verifying the correctness of some intuitive reasonings, but gives no efficient way of constructing decision procedures in particular problems. In fact, as some particular problem is treated, like byzantine agreement, one needs some *ad hoc* models coming from graph theory for finding mathematical impossibility proofs (FLM).

A new difficulty arises when the knowledge of the "players" has to contain a significant part of mathematics, as in the case of the "sum games" described below: ones wishes to have simple methods for dealing with these problems as set theory problems. In other terms, one seeks for models for these problems which endow them into classical mathematics. This can be done in the spirit of classical game theory: game theories always begin by metalemmas about the theory, which allow a reduction to set theory. The first example of such a metalemma is due to Von Neumann and Morgenstern in the study of the two players game with zero sum: it asserts that if there exists a rational theory for this game, and if this theory is common knowledge to both players, then the game matrix must verify the relation:

$$\min_i \max_j a(i,j) = \max_i \min_j a(i,j).$$

After giving a formal definition of knowledge games, we shall obtain in part 2 such a metalemma in the particular case of these games. This lemma asserts that at each step of a knowledge game, the "solutions that player i cannot exclude a priori" is a set which can easily be computed recursively from the initial data. We deduce that every finite knowledge game is decidable and obtain a necessary and sufficient condition in order that a game "converges", that is, finishes after a finite number of steps. Note that our formalization in the concrete case of knowledge games is very close to those of Aumann(A) and Tan and Werlang (TW). (These last authors have proposed a bayesian approach to Aumann's *common knowledge* notion). We do not need to know the common knowledge at each step in order to compute each player's knowledge. This is fortunate, since *computing the solutions "not excluded in the common knowledge" requires an infinite recursion at each step.* What can be finitely computed is a greater set, which is interpreted by Séré and Torlai (ST) as follows: it is the set of all solutions not excluded by an extra player, an "observer", whose only information would be the negative answers of the players. We shall see in a simple example that the set of solutions not excluded by the observer can be much greater than the common knowledge.However,we prove in our metalemma 1 that this set S_k, extremely simple to compute at each step k, is enough to determine the zero level knowledge of each player, that is the set of the solutions which he cannot exclude.

In part 3, we apply the theory to the games described in the introduction. We prove that the sum games are convergent if the number of players is greater than or equal to the number of sums. We prove the general decidability of these games and give a complete study of the case of two players and three sums. We now give a brief description of the games which will be discussed here.

The disk game: n players $1,2,...,n$ are given a coloured disk which can be white or black. Since the disk of each player is placed behind him, he cannot see it, but he can observe the other ones disks. It has been announced to the players that the number of black disks cannot exceed n-1. Then the players are asked (one by one, circularly or all together repeatedly) wether they are able to deduce the colour of their disk. The game finishes as some of the players answers positively. (See Lasry (La)).

Dirty children puzzle. An equivalent version of the previous game is the so called *dirty children puzzle,* whose description we take from Parikh [P]: "Suppose there are n children, some of whom have a dirty forehead. Each child can see everyone else's forehead, but not his own. A teacher makes the announcement "one of you has a dirty forehead". Then, as in the preceding game, the children are asked repeatedly: "is your forehead dirty?" The question is to know which ones of the childrens will be able to answer the question and at which time.

Cheating wifes puzzle. A variant of the preceding "games" is the *cheating wives puzzle* (Gamow and Stern (GS)). Suppose there is a village, in which the population is composed of *n* men and their respective wives. A very strict law oblies every husband which can deduce his wife's infidelity to kill her in the subsequent night. Now, the fact that any particular husband is cheated by his wife is perfectly known to all the other husbands. However, no wife is murdered until a missionary arrives to the village. Since he knows the many infidelities in the village, he thinks he has to speak against that. But wishing to avoid any murder, he only says publicly that there are some unfaithful wife in the village, fact which is perfectly known to all the assistants. However, after some days, all the cheating wives are murdered.

Sum game: This game seems to be due to David Gale and John C. Conway (G). We found their description in Lasry (La). The *n* players 1,2,...,*n* are given *n* positive real numbers $a_1, a_2,..., a_n$. Each of the players knows the numbers of the other ones, but ignores his own. Moreover, it is common knowledge among the players that the sum $a_1 + a_2 + ... + a_n$ is equal to one of *k* given numbers $s_1, s_2,...,s_k$. As in the preceding games, the players are asked in a fixed circular order wether they are able to deduce their own number. The game finishes as some of the players is able to answer "yes".

2. GENERAL THEORY OF KNOWLEDGE GAMES

In the following, we shall distinguish two kinds of definitions and results: those which correspond to intuitive views on the knowledges games and those which belong to the formal theory of these games. The first ones we shall call *metadefinitions* and *metalemmas*. They are essential in order to prove that the formal theory is well adapted to its intuitive aims.

Metadefinition 1. *A "knowledge game" G is given by* $G = (n, (I_k)_k, E, S)$ *where n is an entire number, I_k is a sequence of subsets of $I = \{1,2,...,n\}$ such that for any i in I, there exists infinitely many k such that i is in I_k. E is a given set and S a subset of E^n.*

Of course, *n* represents the number of "players", each *i* in *I* represents a player, and I_k is the set of the players who are asked at time *k*. *E* is the set of the values which can be a priori attributed to each player (colours, numbers, dirty forehead or not, etc...). $S \subset E^n$ is the set of all the attributions of values to the players which are effectively allowed by the rules of the game. So the generic knowledge game proceeds as follows.

At time $t=0$, a value a_i belonging to *E* is given to each player *i*, with the condition that $a=(a_1,a_2,...,a_n)$ must belong to *S*. Every player knows all the other player's values, but ignores his own. At time $t=k(k=1,2, ...)$, all the

players in I_k are simultaneously asked wether they are able to deduce their own value. The game stops as some player answers positively. Of course, the rules of the game, that is essentially the description of G, are assumed to be common knowledge among the players. Since each player is assumed to answer before the other ones, his answer becomes also immediately common knowledge.

Metadefinition 2. *A knowledge game $G = (n, (I_k)_k, E, S)$ is convergent if for every initial attribution of values $a \in S$, there exists a time k such that some player of I_k is able to deduce his own value.*

Notations

In what follows, we denote by $a = (a_1, a_2, ..., a_n)$ the values effectively given to the players at time $t = 1$, and x represents the (for the players) unknown solution. This is just a commodity of notation: for instance, a player is able to answer the question of whether he knows the solution if and only if "he knows that $x = a$". Similary, by "player i knows that $x \neq b$", we mean that player i knows that the solution is not b. To say that some proposition P is common knowledge among the players at time $t = k$, we shall write: $CK_k(P)$. Denote $E^i(y) = \{z \in E^n | \forall j \neq i, \ y_j = z_j\}$. For instance player i can observe $E^i(a)$, which means that he knows that $x_j = a_j$ for each $j \neq i$. If $F \subset E$, we denote similarly by $E^i(F)$ the set

$$E^i(F) = \bigcup_{y \in F} E^i(y).$$

We define inductively the following sets, which will be crucial in the proofs of decidability and convergence of the knowledge games.

$$S_1 = S$$
$$T_k^i = \{y \text{ in } S_k, \ \text{Card } S_k \cap E^i(y) = 1\} = \{y \text{ in } S_k, \ S_k \cap E^i(y) = \{y\}\}$$
$$T_k = \bigcup_{i \in I_k} T_k^i$$
$$S_{k+1} = S_k \backslash T_k.$$

We now state the main lemma of this work. Its proof is based on intuitive evidence and it has to be considered as a "metalemma": we mean that it might be true in any rational theory of decision for the "knowledge games" defined above. Moreover, since we use in it the informal notion of "possibilities which cannot be excluded by a player", its role in a formal theory would be of course that of a simple name given to the set $S_k \cap E^i(a)$.

Metalemma 1. *Assume that the game G with initial condition "a" has not finished before $t = k$. Then for any i in I, $S_k \cap E^i(a)$ is exactly the set of the x in E^n for which i does not know that $x \neq a$. In other words, $S_k \cap E^i(a)$ is the set of possibilities for the initial value that player i cannot exclude at time k (It also*

can be considered as the minimal set W_k for which player i knows at time k that a is in W_k). Moreover, T_k is the set of all initial values a for which the game finishes at time k.

Proof. Obvious for $k=1$, because everything that player i knows is that the solution x is in $S=S_1$ and that x belongs to $E^i(a)$. Thus no element of $S_1 \cap E^i(a)$ can be excluded at this time by player i. Assume the lemma proved at time $t=k$ for every initial condition a, and therefore common knowledge among the players. *It is clear that a player i can deduce at time $t = k$ what is a if and only if the set of non excluded values for him is of cardinality 1.* Now, by the induction hypothesis, this is equivalent to:

$$\text{Card } S_k \cap E^i(a) = 1,$$

which is equivalent to $a \in T_k^i$. Therefore, the fact that the game G does not stop at time k is equivalent to:

$$a \notin T_k = \bigcup_{i \in I_k} T_k^i.$$

Now the points x not excluded by player i at time $k+1$ are the points not excluded at time k, minus those which are newly excluded, that is those contained in T_k. Thus the points non excluded by player i at time $k+1$ in the play with initial condition "a" are exactly:

$$S_k \cap E^i(a) \backslash T_k = (S_k \backslash T_k) \cap E^i(a) = S_{k+1} \cap E^i(a).$$

Remark: Why S_k is not the "common knowledge". Since the preceding lemma is common knowledge among the players, the fact that x is in S_k at time k (if the game did not end before) must be common knowledge among the playes. Denote by S'_k the minimal set such that $CK_k(x \in S'_k)$. Then we can assert that $S_k \subset S'_k$. Are both sets equal? The answer is generally negative and one can give examples with finite games. Let us give one in the case of the sum game with two players and two sums: assume that both sums are rational numbers and that the numbers a_i given to the players are irrational. Then at the first step of the game each player deduces immediately that his own number is irrational, and therefore that the other player is in the same situation of seeing an irrational number. Thus the fact that a is made of irrationals is clearly common knowledge among the players and therefore the common knowledge at time $k=1$ is made of rational numbers. Now, this is not the case for S_1, which clearly contains intervals.

A consequence of Metalemma 1 is that we can now give a precise sense to the notions of decidability and convergence of a knowledge game. (In the following definition we use the notion of algorithm in a formal way and according (e.g.) to the theory of Alan Turing).

Definition 3. *A game $G = (n, (I_k)_k, E, S)$ is decidable if there is an explicit algorithm determining for every a in S, every i in I and for every k such that I_k contains i, whether $S_k \cap E^i(a)$ is of cardinality 1 and, if that occurs, giving this unique element. G is convergent if it verifies one of the following equivalent properties:*

(a) $S = \bigcup_k T_k$
(b) $\bigcap_k S_k = \varnothing$

By metalemma 1, T_k is the set of all initial values "a" for which the game finishes at time k. Thus, intuitively, the game finishes for every initial datum if and only if $S = \bigcup_k T_k$. The formal definition 3 corresponds to a formal translation of the metadefinition 2. In the following we shall always use the term "convergent" in the formal way.

Note that if the set E is finite and S can be given by the enumeration of its elements, one can obviously make an algorithm for giving explicitly the sets T_k and S_k. So we obtain a first generic result of decidability.

Theorem 1. *Every finite knowledge game is decidable.*

We now wish to obtain some clear and useful criteria of convergence for the knowledge games. We begin with two simple necessary conditions:

We shall say that the game G' is a **subgame** of the game G if $S' \subset S$ and both games differ at most by the fact that $S \neq S'$. It is intuitively evident that if G is convergent, so must every stronger game, and we shall prove it now:

Proposition 1. *If G' is a subgame of G, and if G is convergent, then so is G'.*

Proof. We show by induction that $S'_k \subset S_k$: it is true if $k = 1$. Assume it is true at time k and let $x \in T_k \cap S'_k$. Thus Card $(S_k \cap E^i(x)) = 1$ for some i in I_k and Card $(S'_k \cap E^i(x)) \geq 1$. Since $S'_k \subset S_k$, this implies Card $(S'_k \cap E^i(x)) = 1$. Thus we obtain $(T_k \cap S'_k) \subset T'_k$ and therefore $S'_{k+1} = S'_k \backslash T'_k \subset S'_k \backslash (T_k \cap S'_k) = S'_k \backslash T_k \subset S_k \backslash T_k = S_{k+1}$.

Assume that a knowledge game verifies the following property: there is no initial condition a for which some player i is able to find his own value a_i after having seen the other a_j. That means that $T_1 = \varnothing$, and since the structure of S is common knowledge among the players, it is also common knowledge that every player will answer "I don't know" at the first time where he is asked. Thus no progress (in knowledge, knowledge about knowledge and so on) is made after each player has been asked, and it is intuitive that the game will remain stable and nether end. *In our formalism, the proof of this property is trivial:* if $T_1 = \varnothing$, one gets $S_2 = S_1$ and by induction $S_k = S_1$. Thus the preceding

remark reduces to the fact that if a game is convergent, one has $T_1 \neq \emptyset$. Applying this property and proposition 1 to every subgame G' of G, we get the necessary condition:

Proposition 2. *A game is convergent only if it verifies the following property:*
(c) *For every $\emptyset \neq S' \subset S$ there exists $i \in I$ and $x \in S'$ such that* $Card\,(S' \cap E^i(x)) = 1$.

The condition (c) is in fact "almost sufficient" in order that a game is convergent. As an example, let us consider the case where the game has some finiteness property.

Theorem 2. *Assume that S is finite (or more generally that for every i in I and y in S, $Card(E^i(y) \cap S) < \infty$). Then the game G is convergent if and only if property (c) is verified.*

Proof. The necessary condition is proved in proposition 2. Conversely, assume that the game verifies property (c). Assume by contradiction that the game is not convergent. Then $S' = \bigcap_k S_k$ is nonempty. Then by (c) there exists i in I and y in S such that $Card\,(S' \cap E^i(y)) = 1$. The finite sets $S_k \cap E^i(y)$ forming a nonincreasing sequence, one still has $Card\,(S_k \cap E^i(y)) = 1$ for some k. Thus y is in the first successive T_k such that i belongs to I_k which is a contradiction.

Remark. The last part of the proof shows that one can check (c) by considering only the case $S' = \bigcap_k S_k$. This will be convenient in the applications.

3. APPLICATIONS TO PARTICULAR KNOWLEDGE GAMES

3.1. Decidability. We begin with some considerations about the decidability of the knowledge games described in the introduction. By using Theorem 1, one sees immediately that these games, including the sum game if the data are entire numbers, are decidable. Let us now briefly explain why the general sum game is also decidable; being assumed of course that the initial data, and the sums are computable numbers (to simplify, assume they are rational numbers). We prove by induction on k that if S_k is computable, so is S_{k+1}. More precisely, assume that S_k is a finite union of simplexes, each being defined e.g. by a finite set of explicit inequalities. Then S_{k+1} is finite union of simplexes obtained by:

1) Projecting all the simplexes of S_k on the hyperplanes $\{x_i = 0\}$ which is achieved by removing the terms in x_i in the inequalities defining the simplexes.

2) Making all possible intersections between the projected simplexes. Each

one of these projections is also a simplex and in order to intersect two simplexes,it is enough to realise the union of their sets of inequalities.

3) $S_k \backslash T_k^i$ is then the union of all possible intersections between the simplexes of S_k and the cylindric simplexes whose basis are the simplexes of $\{x_i = 0\}$ obtained in 2).

4) One obtains $S_k \backslash T_k$ as the intersection of the $S_k \backslash T_k^i$ for i in I_k. Deciding which simplexes are empty can be achieved by some entire linear programming algorithm.

3.2. Game convergence

We begin with a useful criterion for deciding wether a given game G is convergent.

Proposition 3. *Let $G = (n, (I_k)_k, E, S)$ be a knowledge game. Assume that E is contained in $\mathbb{R}$ and that S is a compact subset of $\mathbb{R}^n$, then a sufficient condition in order to obtain property (c) for closed S' is that there do not exist x, y in S with:*

1) $x = (x_1, x_2, ..., x_n)$, $y = (y_1, y_2, ..., y_n)$, and $x_i > y_i$ for every $i = 1, 2, ..., n$.

2) for $i = 1, 2, ..., n-1$ one has $x^i = (x_1, ..., x_i, y_{i+1}, ..., y_n) \in S$.

Proof. Let $\varnothing \neq S' \subset S$ and assume S' be closed: we shall construct x in S' such that Card $(S' \cap E^i(x)) = 1$ for some i in I. Define recursively an element $y = (y_1, y_2, ..., y_n)$ by:

$$y_n = \min \{y | (x_1, x_2, ..., , y) \in S' \text{ for some } (x_1, x_2, ..., x_{n-1})\}$$

$$y_{n-1} = \min \{y | (x_1, x_2, ..., y, y_n) \in S' \text{ for some } (x_1, x_2, ..., x_{n-2})\}$$

....

$$y_1 = \min \{y | (y, y_2, ..., , y_n) \in S'\}$$

Set $x^0 = (y_1, y_2, ..., y_n)$. If $E^i(x^0) \cap S'$ contains no other element than x^0, we have finished. If not, choose $x_1 > y_1$ such that $x^1 = (x_1, y_2, ..., y_n) \in S'$. If now $E^2(x^1) \cap S'$ contains no other element than x^1, we have again finished. If not we can find $x^2 = (x_1, x_2, y_3, ..., y_n) \in S'$ with $x_2 > y_2$. We iterate this process. Now, by assumption, it must end before arriving to define x^n, and therefore provides some x^i such that Card $(E^{i+1}(x^i) \cap S') = 1$.

Remark. Of course the previous statement holds for an arbitrary space E endowed whith an order relation which makes every closed set have a minimum.

Corollary 1. *Every sum game for which the number m of given sums does not exceeds the number n of players is convergent.*

Remark. We have stated this result with informal terms corresponding to the metadefinition 2, but we shall give it a rigorous sense by defining S for the sum games and proving the convergence in the formal sense of definition 3.

Proof. Define S by: $x = (x_1, x_2, ..., x_n) \in S$ if and only if it verifies

$$x_i \geqslant 0, \ i = 1,2,...,n; \ \sum_{i=1}^{n} x_i = s_j \quad \text{for some } j \text{ in } 1,2,...,m$$

S is obviously compact and we can prove by induction that so are the S_k. Indeed, assume by induction that S_k is compact. Then the functions defined on S by $x \to \text{Card} \ (E^i \ (x) \cap S_k)$ are clearly upper semi-continuous since two distinct elements in $E^i(x) \cap S_k$ have a distance bigger or equal than $\text{Min}_{i \neq j} |s_i - s_j|$. Thus the set of all x in S_k such that $\text{Card} \ (E^i(x) \cap S_k) = 1$ is open in S. We obtain that T_k is open in S and therefore S_{k+1} is also compact. We now observe that the finiteness assumption of Theorem 2 is easily checked and we prove that the criterion (c) for the convergence of the game is verified. The remark under Theorem 2 allows us to restrict ourselves to the case that $S' = \bigcap_k S_k$ and therefore in particular that S' is closed. Hence we are in a position to apply proposition 3 and we can assume by contradiction that there exists a sequence $x^0, ..., x^i, ..., x^n$ in S' with

$$x^i = (x_1, ..., x_i, y_{i+1}, ..., y_n) \text{ and } x_i < y_i \text{ for } i = 1,2,...,n$$

Denote by $s(x)$ the sum of the coordinates of x. Then $s(x^0) < s(x^1) ... < s(x^n)$, but these numbers must belong to $\{s_1, s_2, ..., s_m\}$, which is impossible if $m \leqslant n$.

Remark. The preceding result can be generalized in the following way: In the situation of Proposition 3, the game is convergent provided S has no subset with more than n elements which is totally ordered by the product ordering in E^n.

Analysis of the dirty children game and of the disk game. First notice that these puzzles are equivalent to a knowledge game, where $E = \{0, 1\}$ and the set S is defined by the fact that its elements $(x_1, x_2, ..., x_n) \in E^n$ verify the relation:

$$\sum_{i=1}^{n} x_i \in \{0, 1, ..., n-1\}$$

To see that, it is enough, for the disk game, to set $x_i = 1$ if the disk of i is black and $x_i = 0$ otherwise. For the dirty children game, we set $x_i = 0$ if the

child is dirty and $x_i = 1$ otherwise. So both games are identical and equivalent to a sum game with the special restriction on initial data that $x_i \in \{0, 1\}$. As a consequence of proposition 1 we obtain immediately the convergence of the disk and dirty children games. Now, the intuitive theory of these games explained in (La) provides not only this convergence, but also predicts which players find out their colour, and at which time. If our approach is valid, it must give the same result.

Let us compute the sets T_k, S_k: we have seen that $S = S_1$ can be defined as:

$$S = \{(x_1, x_2, ..., x_n) \in \{0,1\}^n, x_1 + x_2 + ... + x_n \leqslant n - 1\}.$$

To simplify the notations we shall assume that $I_k = I$, but what follows can be easily adapted to other cases. To evaluate T_1, it is enough to remark that if $x_1 + x_2 + ... + x_n < n - 1$, one obtains a new element of S_1 if one changes the value of an arbitrary x_i. Conversely, if $x_1 + x_2 + ... + x_n = n - 1$, there is some i such that $x_i = 0$ and this coordinate cannot be changed in S_1. Thus $x = (x_1, x_2, ..., x_n)$ is in T_1^i. We obtain that:

$$T_1 = \{x \in \{0,1\}^n, x_1 + x_2 + ... + x_n = n - 1\} \quad \text{and therefore:}$$

$$S_2 = \{x \in \{0,1\}^n, x_1 + x_2 + ... + x_n \leqslant n - 2\}.$$

Similary, one sees that T_2^i is the set of all x in S_2 whose i − th coordinate is zero and which verify $x_1 + x_2 + x_n = n - 2$. Hence:

$$T_2 = \{x \in \{0,1\}^n, x_1 + x_2 + ... + x_n = n - 2\},$$

and by an obvious iteration:

$$S_k = \{x \in \{0,1\}^n, x_1 + x_2 + ... + x_n \leqslant n - k\},$$

$$T_k = \{x \in \{0,1\}^n, x_1 + x_2 + ... + x_n = n - k\}.$$

Recall that the strategy of the player i is to compute its set of nonexcluded values $S_k^i = S_k \cap E^i(a)$, the game finishing as some of the players has a set of nonexcluded values of cardinality 1. Assume player i sees $n - r - 1$ black disks. Then $E^i(a)$ contains exactly two elements, whose sums are $n - r$ and $n - r - 1$. Since $S_{k+1}^i = S_k^i \backslash T_k$, one sees that $S_k^i = S_1$ for every $k \leqslant r$, and S_{r+1}^i is of cardinality 1. One concludes easily that if $n - r - 1$ black disks have been given, the play finished at time $r + 1$, as all white players conclude together that they are white.

Analysis of the cheating wives puzzle. Set $x_i = 1$ if the husband i is not cheated, $x_i = 0$ otherwise. Then the declaration of the missionary is equivalent to the fact that:

$$\sum_{i=1}^{n} x_i \in \{0, 1, \ldots, n-1\}.$$

Thus the only difference whith the disk or dirty children game is that each morning, as some husband has not killed the wife, this fact does not mean that he cannot decide how is his wive's behaviour, but only that he is not sure that she is unfaithful. Thus the cheating wifes puzzle does not a priori belongs to the class of knowledge games defined in section 2. However, the analysis of the previous game shows that only a cheated husband can end the game. Therefore the fact that a wife is not killed is a posteriori equivalent to the information that her husband cannot decide whether she is unfaithful or not.

The fact that the disk game is always ended by some white player can be expressed by the following fixed point property:

Proposition 4. *Let $f:E \to E$ and define for each i in I the function $f_i: E^n \to E^n$ by $(f_i(x))_j = x_j$ if $j \neq i$ and $(f_i(x))_j = f(x_i)$ if $j = i$. Assume f_i sends S into S. Then player i can end the game only if $f(a_i) = a_i$.*

Indeed, one proves easily by induction that $f(S_k) \subset S_k$. To apply the proposition to the disk game, it is enough to set $f(\text{black}) = \text{white}$ and $f(\text{white}) = \text{white}$.

3.3. Further analysis of the sum games

We have seen in 3.1 that the sum games (with computable data) are always decidable. In the well posed cases, namely when the number m of sums is not greater that the number n of players, we know that the sum game is convergent. If $m > n$, the situation is much more complex, the computation of the T_k being related in many aspects to ergodic theory. We shall begin with some general convergence and divergence results:

Proposition 4. *Let $n \geqslant 2$, $m \geqslant n$ and assume $s_{i+1} > (n/n-1)s_i$. Let, for simplicity, $I_k = I$ for every k. Then the sum game is convergent.*

Proof. One shows by induction that $\{a, \sum_i a_i = s_k\} \subset T_{m-k+1}$. Indeed, if $k = m$, one clearly has $a \in T_1^i$ if $\sum_{i \neq j} a_i > s_{m-1}$. Hence $a \notin T_1$ implies $\sum_j \sum_{i \neq j} a_i \leqslant n s_{n-1}$, that is $\sum_i a_i \leqslant (n/n-1) s_{m-1} < s_m$. Thus $\{a, \sum a_i = s_m\} \subset T_1$. The other steps are similar, and we obtain that $S_{m+1} = \emptyset$.

Proposition 5. *If $m = n+1$, one can find, $s_1, \ldots, s_{n+1}$ such that the associated game is convergent.*

Proof. Take $s_2 - s_1 = \ldots = s_n - s_{n-1} = d > 0$. Then the hypercube $S' = \{s_1, s_1 + d\} \times \{0, d\}^{n-1}$ is enclosed in S and therefore the necessary condition of convergence (c) in proposition 2 is not fulfilled.

A complete analysis of the sum game in the case m = 3, n = 2

Proposition 6. *Assume the number* n *of players is 2 and the number* m *of sums is 3. Then the sum game associated with the sums* $s_1 < s_2 < s_3$ *converges if and only if the following condition is verified:*

(d) *There exist* p *and* q *in* $\mathbb{Z}$ *such that* $p(s_2 - s_1) + q(s_3 - s_2) \in \,]s_1, \, s_3 - s_2[$.

Proof. We proceed in several steps.

Step 1: Preliminary remarks. One sees easily that $T_1 = \{(a_1, a_2) \text{ in } (\mathbb{R}^+)^2, a_1 > s_2 \text{ or } a_2 > s_2\}$, that consequently $S_2 = S \cap [0, s_2]^2$, $T_2 = \{(a_1, a_2), a_1 + a_2 = s_2 \text{ and } s_1 < a_1 < s_3 - s_2\}$. Thus if $s_1 \geq s_3 - s_2$, $T_2 = \varnothing$ and therefore $S_3 = S_2$ and the game is not convergent. In this case, since property (d) is not verified, the proposition 6 is true. So we may assume that $s_1 < s_3 - s_2$. A peculiarity of the game is that for any $a^0 = (a_1^0, a_2^0)$ in $S_2 \setminus T_2$ there is exactly one element of S_2, a^{-1} such that $a_1^{-1} = a_1^0$ and one element a^1 such that $a_2^1 = a_2^0$. Hence to any element a^0 of s_2 one can associate a unique sequence $(a^k)_{k \in K}$ such that $K = \mathbb{Z}$ if $a \in \bigcap_k S_k$ and $K = [p, q] \subset \mathbb{Z}$ if min $(|p|, |q|) = k$ is in T_{k-2}. Moreover, one clearly has for any k, $a_1^{k+1} = a_1^k$ or $a_2^{k+1} = a_2^k$ and $|a_1^{k+1} - a_1^k| + |a_2^{k+1} - a_2^k| \in \{s_2 - s_1, s_3 - s_2\}$ and $a_1^k + a_2^k = s_2$ if and only if $a_1^{k+1} + a_2^{k+1} = s_1$ or s_3. We shall call the sequence (a^k) "d-sequence associated whith a^0". A d-sequence is infinite on both sides if and only if its elements are in $S' = \bigcap_k S_k$.

Step 2: If the game converges, property (d) is true. Indeed, if the game is convergent, $a^0 = (0, s_2)$ is in some T_l, and there exists a sequence a^k of the preceding type with $a^0 = (0, s_2)$ and $a^{l-2} \in T_2$. Therefore one obtains

$$\sum_{k=1}^{l-2} (a_1^k - a_1^{k-1}) = a_1^{l-2} \in \,]s_1, \, s_3 - s_2[$$

and since $|a_1^k - a_1^{k-1}| \in \{0, s_2 - s_1, s_3 - s_2\}$ for any k, we obtain:

$p(s_2 - s_1) + q(s_3 - s_2) \in \,]s_1, \, s_3 - s_2[$ *for some* p *and* q *in* $\mathbb{Z}$.

Step 3: If the game does not converge, (that is $S' = \bigcap_k S_k = \varnothing$**), then** $(0, s_2)$ **is in** S'**.** Indeed, we know that S' is closed. Let a^0 verifying $a_1^0 = \text{Min} \{a_1, a \in S'\}$ and consider its associated sequence $(a_k)_{k \in \mathbb{Z}}$. Since the sequence (a_k) is infinite, there are elements in (a_k) such that $a_1^k + a_2^k = s_l$. Then $a_1^0 \leq s_1$ and since $s_1 < s_3 - s_2$ we may choose a^0 (see the remarks of Step 1) such that $a_1^0 + a_2^0 = s_2$, $a_1^0 \leq a_1^k$ and $a_2^k \leq a_2^0$ for any k in $\mathbb{Z}$. Setting $b^k = (a_1^k - a_1^0, a_1^0 + a_2^k)$, we obtain a new infinite d-sequence in S_2, associated with $b^0 = (0, s_2)$.

Step 4: $b^0 = (0, s_2)$ *is in* $S' = \bigcap_k S_k$ *only if* there is no (p, q) in $\mathbb{Z}^2$ such that $p(s_2 - s_1) + q(s_3 - s_2) \in \,]s_1, \, s_3 - s_2[$. Indeed, assume that the least relation is true for some (p, q) in $\mathbb{Z}^2$. We shall use the following lemma:

Lemma. *Let r, s, t in* $\mathbb{R}$ *with* $r + t \le s$ *and assume that* $pr + qt \in [0,s]$ *for some* (p,q) *in* $\mathbb{Z}^2$. *Then there exists a sequence* $(p_k, q_k)_{0 \le k \le l}$ *in* $\mathbb{Z}^2$ *such that* $(p_0, q_0) = (0,0), |p_k| \uparrow |p|, |q_k| \uparrow |q|, |p_k - p_{k-1}| + |q_k - q_{k-1}| = 1$ *and* $p_k r + q_k t \in [0, s]$.

The proof is straightforward by induction on $|p| + |q|$. Apply this lemma to $r = s_2 - s_1$, $t = s_3 - s_2$, $s = s_2$ and set

$$b^{2k} = (p_k(s_2 - s_1) + q_k(s_3 - s_2), \; s_2 - p_k(s_2 - s_1) + q_k(s_3 - s_2))$$

$$b^{2k+1} = (p_{k+1}(s_2 - s_1) + q_{k+1}(s_3 - s_2), \; s_2 - p_k(s_2 - s_1) + q_k(s_3 - s_2)).$$

The d-sequence $(b_k)_{0 \le k \le l}$ verifies $b^0 = (0, s_2)$ and $b^{2l} \in T_2$. Thus b^0 is not in $S' = \bigcap_k S_k$.

References

(AS) ATTARDI, G. and SIMI, M. *Metalanguage and reasoning across view-points.* Proc. of the European Conference on Artificial Inteligence, T.O'Shea editor, Elsevier Science (1984).

(A) AUMANN, R.J. *Agreeing to Disagree,* The Annals of Statistics, Vol. 4 N.° 6, (1976) 1236-1239.

(BKP) BARRINGER, H., KUIPER, R. and PNUELI, A. *Now you may compose temporal logic specifications.* Preprint, INRIA (1984).

(TW) CHIN-CHIU TAN, T. and RIBEIRO da COSTA WERLANG, S. *On Aumann's notion of Common Knowledge. An alternative approach.* Preprint, Gradua-te School of Business, University of Chicago, 1986 and Proccedings of the conference on theoretical aspects of reasoning about knowledge (ed. J.H.Halpern), Morgan Kaufman, (1986).

(FH) FAGIN, R., and HALPERN, J.Y. *Belief, awareness and limited reasoning.* Proc. 9th International Joint Conf. on Artifical Intelligence, Los Angeles (1985) pp. 491-501.

(FHV) FAGIN, R., HALPERN, J.Y. and VARDI, M.Y. *A model-theoretic analy-sis of knowledge.* Proc 25th IEEE Symp. on Foundations of Computer Science, West Palm Beach Florida (1984) pp. 268-278.

(FLM) FISHER, M.J., LYNCH, N. and MERRIT, M. *Easy impossibility proofs for distributed consensus problems.* Proc. of the fourth ACM symposium on Principles of Distributed Computing (1985).

(G) GALE, D. *Personal communication.*

(GS) GAMOW, G. and STERN, M. *Forty unfaithful wives,* Puzzle Math., The Viking Press, New York, (1958), pp. 20-23.

(HM) HALPERN, J. H. and MOSES, Y. *Knowledge and Common Knowledge in a distributed environment.* Proc. of the third ACM Symp. on Principles of Distributed Computing (1984) pp. 50-61.

(H) HALPERN, J.H. (editor). *Proceedings of the conference on theoretical as-pects of reasoning about knowledge,* Morgan Kaufman, (1986).

(La) LASRY, J.M. *Le "common knowledge".* Ornicar n.° 30, Editions du Seuil, Champ Freudien, pp. 75-93, (1984).

(L) LEHMAN, D. *Knowledge, Common Knowledge and related puzzles.* Third ACM Symp. on Principles of Distributed Computing (1984) pp. 62-67.

(MJ) MANI CHANDY, K. and JAYADEV, MISRA. *How processes learn.* Proc. of the fourth ACM Symposium on Principles of distributed Computing (1985) pp. 204-214.

(MP) MANNA, Z. and PNUELI, A. *Verification of concurrent programs: the temporal framework.* In "The correctness problem in computer science", eds Boyer and Moore, Academic Press London (1982).

(MDH)MOSES, Y., DOLEV, D. and HALPERN, J.Y. *Cheating husbands and other stories: a case study of knowledge, action and communication.* Proc. of the fourth ACM Symposium on Principles of Distributed Computing (1985).

(P) PARIKH, R. *Levels of knowledge in distributed computing.* IEEE n.° CH2321-8|86|0000|0314 (1986).

(P) PARIKH, R. *Propositional Game Logic.* IEEE Symposium on Foundation of Computer Science (1983) pp. 195-200.

(ST) SÉRÉ, ERIC and TORLAI OLIVIER. *Etude d'un jeu à p joueurs.* Rapport de stage, Ecole Polytechnique (1987).

CEREMADE
Université Paris-Dauphine
Place du Maréchal de Lattre de Tassigny
75775 Paris Cedex 16
FRANCE

SISSA
325 Viale Miramare
34014 Trieste
ITALY