ENTROPY AND CODIFICATION IN REPEATED GAMES WITH IMPERFECT MONITORING

OLIVIER GOSSNER AND TRISTAN TOMALA

Abstract. We characterize the min max values of a class of repeated games with imperfect monitoring. Our result relies on the optimal trade-off for the team formed by punishing players between optimization of stage-payoffs and generation of signals for future correlation. Amounts of correlation are measured through the entropy function. Our theorem on min max values stems from a more general characterization of optimal strategies for a class of optimization problems.

1. Introduction

The central result in the theory of repeated games with perfect monitoring is the Folk Theorem (see Aumann and Shapley [AS94], Rubinstein [Rub77] for the non-discounted case, Fudenberg and Maskin [FM86] for the discounted case, and Benoît and Krishna [BK85] and Gossner [Gos94] for finitely repeated games), which characterizes the set of equilibrium payoffs of these repeated games. Perfect monitoring refers to the assumption that after each stage, the actions taken by all players are publicly announced. In many situations, this assumption is unrealistic and has to be replaced by a signalling structure, under which each player gets to observe a private signal, which depends stochastically on the actions taken by all players.

Two main strands of literature extend the folk theorem to games with imperfect monitoring. The first (see [FLM94] and [RT98] among others) seeks conditions on the signalling structure under which all equilibria of the repeated game with imperfect monitoring remain equilibrium payoffs of the repeated game with imperfect monitoring. The second fixes a signalling structure, and seeks to characterize the set of all equilibrium payoffs (see [Leh91] [Leh89], [RT00] for instance).

As pointed out by Lehrer [Leh91], the set of equilibrium payoffs of a repeated game with imperfect monitoring can differ from the set of payoffs of the same game with perfect monitoring in two aspects. The most obvious is that imperfect monitoring can make it more difficult to monitor a player's behavior (e.g. if actions are not observable by all), and this tends to reduce the set of equilibrium payoffs. On the other hand, in some cases a group of players may use the signalling structure of the game in order to generate a correlation device, and this can result in an extension of the set of equilibria.

Possibilities of correlation arising from signals naturally pose the question of the characterization of minmax payoffs, which is often necessary to obtain a characterization of the set of equilibrium payoffs. The minmax payoff of a player in a repeated game with signals always lies between this player's minmax payoff of the one-shot game in mixed strategies and in correlated strategies. An example, already presented in [RT98], is recalled in section 2.2.

We provide a characterization of the minmax payoffs of a given player I in a class of repeated games with signals. Let $\{1, \dots, I\}$ be the set of players, A^i player i's

Date: February 2003.

finite set of actions, $\Delta(A^i)$ player i's set of mixed strategies (i.e. the set of probability distributions over A^i). Player I's stage payoff function is given by $r\colon \Pi_i A^i \to \mathbb{R}$. After each stage, if $a = (a^i)_i$ is the action profile played, a signal b is drawn in a finite set B according to $\rho(b|a)$, where $\rho\colon \Pi_i A^i \to \Delta(B)$. Player I observes b and a^I, whereas other players observe b and a. We make two assumptions on ρ. The first, which is rather innocuous, is that b does not reveal a (namely, there exist two action profiles a, a' and a signal b such that $\rho(b|a).\rho(b|a') > 0$). The second is that the distribution of signals to I does not depend on I's actions, ρ is then a function of $(a_i)_{i\neq I}$. We view the game as zero-sum between the team formed by players $1,\dots,I-1$ and player I.

A history of length n for the team [resp. for player I] is an element h_n of $H_n = (B \times \Pi_i A^i)^n$, [resp. h_n^I of $H_n^I = (A^I \times B)^n$], and a (behavioral) strategy for a team player i [resp. for player I] is a sequence $(\sigma_n^i)_{n\geq 0}$ with $\sigma_t^i\colon H_n \to \Delta(A^i)$ [resp. $(\sigma_n^I)_{n\geq 0}$ with $\sigma_n^I\colon H_n^I \to \Delta(A^I)$]

The λ-discounted, n-stage, and uniform min max payoffs of player I, denoted v_λ^I, v_n^I, and v_∞^I are defined the usual way (see Mertens Sorin Zamir [MSZ94]).

Our characterization of v_∞^I relies on the following heuristics: given strategies of the team players $\sigma^{-I} = (\sigma^i)_{i\neq I}$ and $h_t^I \in H_t^I$, player I possesses a conditional probability on the sequence of histories $h_t \in H_t$ to the team, and to any such h_t corresponds a $I-1$ tuple of mixed actions at stage $t+1$. Hence, the actions of the team players at stage $t+1$ are correlated conditional to h_t^I, but independent conditional to h_t. This motivates the introduction of the concept of a correlation system $\mathbf{z}$, which is a I-tuple of random variables $\mathbf{z} = (\mathbf{k}, \mathbf{x}_1, \dots, \mathbf{x}_{I-1})$, where $\mathbf{k}$ takes values in an arbitrary finite set, and $\mathbf{x}_i$ in A^i, with the condition that the random variables $(\mathbf{x}_1,\dots,\mathbf{x}_{I-1})$ are independent conditional to $\mathbf{k}$. We measure the effect of playing according to a correlation system $\mathbf{z}$ both in payoffs and in information.

Let $D(\mathbf{z})$ be the distribution induced by $\mathbf{z}$ on $A^1 \times \cdots \times A^{I-1}$, and $\pi(\mathbf{z}) = \max_{a^I \in A^I} \mathbf{E}_{D(\mathbf{z})} r(\mathbf{a}^1, \dots, \mathbf{a}^{I-1}, a^I)$ be the payoff of player I induced by a best reply against $D(\mathbf{z})$.

Let also $\mathbf{t}$ be the random variable representing the signal of player I, drawn according to $\rho(\mathbf{x}_1,\dots,\mathbf{x}_{I-1})$. Playing according to $\mathbf{z}$ has an effect on future abilities of the team to correlate, which we measure in terms of entropies (see section 2.4.1 for a reminder). More precisely, the information gain induced by $\mathbf{z}$ is $\Delta H(\mathbf{z}) = H(\mathbf{x}_1,\dots,\mathbf{x}_{I-1},\mathbf{t}|\mathbf{k}) - H(\mathbf{t})$, defined as the difference between the entropy of the extra information received by the team and the entropy of the signal to player I. A simple, yet remarkable, property of $\Delta H(\mathbf{z})$ is the equality $\Delta H(\mathbf{z}) = H(\mathbf{z}|\mathbf{t}) - H(\mathbf{k})$, where $\Delta H(\mathbf{z})$ appears as the difference between the entropy of the information of the team which is not possessed by I after $\mathbf{z}$ is played and before.

Let V be the set of all pairs $(\Delta H(\mathbf{z}), \pi(\mathbf{z}))$ where $\mathbf{z}$ ranges over all correlation systems, and let $\operatorname{co} V$ denote its convex hull.

Consider an optimization problem in which the team can choose any correlation system $\mathbf{z}_n$ at stage n, provided the total entropy generated is non negative after any stage (for any n, $\sum_{m=1}^n \Delta H(\mathbf{z}_m) \geq 0$), the goal being to minimize the long term average of the sequence $(\pi(\mathbf{z}_n))_n$. A classical argument in optimization theory (see Aubin and Ekeland [AE76]) shows that the value of this optimization problem is $\min\{x_2 | (x_1,x_2) \in \operatorname{co} V, x_1 \geq 0\}$.

Even though there is no formal equivalence between the above mentioned optimization problem and the min max values of repeated games with imperfect monitoring, the former provides a good heuristics for the latter. In fact, we prove that the values of the two coincide. Namely:

Theorem 1.

$$v^I_\infty = \lim_{n\to\infty} v^I_n = \lim_{\lambda\to 1} v^I_\lambda = \min\{x_2 | (x_1, x_2) \in \operatorname{co} V, x_2 \geq 0\}$$

That $\lim_{n\to\infty} v^I_n$ and $\lim_{\lambda\to 1} v^I_\lambda$ exist and coincide follows from classical arguments. That $v^I_\infty \geq \min\{x_2 | (x_1, x_2) \in \operatorname{co} V, x_2 \geq 0\}$ follows from additivity properties of entropies, as in the argument of Neyman and Okada ([NO99],[NO00]).

To prove the other inequality, we rely on the fact that player I has no influence on signals, and consider strategies of the team which do not depend on I's actions (we are actually going to consider such strategies only, and show they can guarantee as much as general strategies). A best reply for player I against such strategies is to maximize payoffs at stage $t+1$ against the conditional distribution of actions of the team at stage $t+1$ given the team strategy and h^I_t. This remark leads us to formulate the problem faced by the team in a dynamic programming framework.

We view the problem of the team as a particular instance of the class of optimization problems given by:

- A finite set S;
- A choice set D which is a closed subset of $\Delta(S)$;
- A continuous mapping $g : \Delta(S) \to \mathbb{R}$;
- A partition T of S;
- A strategy ξ prescribes a choice in A given any finite sequence of elements of S;
- Given ξ and the induced probability P_ξ on all finite histories, the payoff at stage m after the sequence of signals $(t_1, ..., t_{m-1})$ is given by $g(P_\xi(.|t_1, ..., t_{m-1}))$.

When D is the set of distributions on $\Pi_{i\neq I} A^i \times B$ induced by independent strategies of the team ($D = \{d \otimes \rho(\,\cdot \mid d), d \in \otimes_{i\neq I}\Delta(A^i)\}$), T the signal to player I, g the best response expected payoff to player I, we claim that the optimization problem is equivalent to the min max problem of the initial repeated game with signals (see section 5.1).

The optimization problem also covers the class of zero-sum games studied by Gossner and Vieille [GV02] in which player 1 is restricted to play in pure strategies, but privately observes at each stage the outcome of an exogenously given random variable. In this case, D is the set of actions of player 1, S describes the action of player 1 and his/her observation of the random variable, and T contains the action of player 1 only. We show that the optimization problem covers a larger class of such *biased coin problems* in section 5.2.

We do not assume g concave, although it is if the goal of the team is to minimize the payoff of player I. By assuming g continuous, we also cover models in which player I is a sequence of short-run players maximizing one-stage payoffs against their beliefs on the team's actions. The situation need not be zero-sum between the team and the sequence of short-run players.

In any finite game in which each player i's payoff function ($i = 1, \ldots, I-1$) is the opposite of player I's payoff function r, a result due to von Stengel and Koller [vSK97] shows that there exists a Nash equilibrium payoff in which player I receives $\min\max r$. Furthermore, in case of multiplicity of Nash equilibria, this Nash payoff appears to be the most natural one. Our result relates the amount of information that can be shared by the team using the signals of the game with the the limit of these Nash payoffs for finite repetitions of the stage game with imperfect monitoring.

Our theorem essentially states that the main constraint on the sequence of systems $(\mathbf{z}_n)_n$ that may be used by the team is that this sequence does not use more

entropy than it generates. Indeed, at each stage n, we may have $\Delta H(\mathbf{z}_n) > 0$, in which case entropy is generated and may be used in the future, or $\Delta H(\mathbf{z}_n) < 0$ which means that some entropy is spent. We prove that if $n\Delta H(\mathbf{z}) + n'\Delta H(\mathbf{z}') \geq 0$, there exist strategies of the team that play approximately $\mathbf{z}$ and $\mathbf{z}'$ in proportions $\frac{n}{n+n'}$ and $\frac{n'}{n+n'}$ of the time.

The probabilistic tools we use for this proof belong to the class of coding techniques, which are fundamental in information theory (cf. Shannon [Sha48], Cover and Thomas [CT91]), and were already applied to a game-theoretic problem in Gossner and Vieille [GV02].

2. Model and results

Notations: When U is a finite set, $\Delta(U)$ represents the simplex of distributions over U and $|U|$ denotes the cardinality of U. Bold characters $(\mathbf{x}, \mathbf{y}, \ldots)$ represent random variables.

2.1. **Main definitions.** We study a discrete-time dynamic optimization problem given by the following data.

- A finite set of outcomes S;
- A closed set of decisions $D \subseteq \Delta(S)$;
- A continuous payoff function $g : \Delta(S) \to \mathbb{R}$;
- A partition T of S.

At each stage $n = 1, 2, \ldots$, a decision maker (d.m.) chooses a distribution d on S which belongs to D. An outcome s in S is then drawn at random according to d and observed by the d.m. An outside observer (o.o.) gets to observe a signal depending on s. If s is selected, the o.o. is informed of $t \in T$ such that $s \in t$.

We assume that the d.m. recalls all past outcomes. A strategy is a decision rule prescribing which decision to take after each finite sequence of outcomes. Formally, it is a mapping $\xi : \bigcup_{n\geq 1} S^{n-1} \to D$, S^0 being an arbitrary singleton. Let S^∞ be the set of infinite sequences of outcomes and $\mathcal{F}$ be its product σ-algebra. A strategy ξ induces a probability measure P_ξ on $(S^\infty, \mathcal{F})$.

Let ξ be a strategy. Let $\mathbf{s}_m$ be the random outcome and $\mathbf{t}_m$ be the random signal at stage m. Let us denote $\mathbf{s}^{m-1} = (\mathbf{s}_1, \ldots, \mathbf{s}_{m-1})$ and $\mathbf{t}^{m-1} = (\mathbf{t}_1, \ldots, \mathbf{t}_{m-1})$. Let $P_\xi(\cdot \mid \mathbf{t}^{m-1})$ be the distribution of $\mathbf{s}_m$ given $\mathbf{t}^{m-1}$. The payoff at stage m is $g(P_\xi(\cdot \mid \mathbf{t}^{m-1}))$ and the average expected payoff is defined, for each $n \geq 1$ and strategy ξ by:

$$\gamma_n(\xi) = \mathbf{E}_\xi[\frac{1}{n}\sum\nolimits_{m=1}^{n} g(P_\xi(\cdot \mid \mathbf{t}^{m-1}))]$$

where $\mathbf{E}_\xi$ denotes expectation with respect to P_ξ. For each discount factor $\lambda \in [0, 1)$, the discounted payoff is defined by:

$$\gamma_\lambda(\xi) = \mathbf{E}_\xi[\sum\nolimits_{m=1}^{\infty}(1-\lambda)\lambda^{m-1} g(P_\xi(\cdot \mid \mathbf{t}^{m-1}))]$$

Let $v_n = \sup_\xi \gamma_n(\xi)$ be the value of the n-stage problem and $v_\lambda = \sup_\xi \gamma_\lambda(\xi)$.

2.2. **An example.** *A coordination problem.* Consider the following three-person problem. There are two casinos in Los Juegos, C_1 and C_2. Gamblers 1 and 2 want to meet each other at the same casino but absolutely want to avoid spending the evening with gambler 3 whereas gambler 3 is willing to meet at least one of the two others. The payoffs for gamblers 1 and 2 are the same and equal 1 if gamblers 1 and 2 manage to meet each other without gambler 3 and 0 otherwise. The payoff for gambler 3 is the opposite. Seen as a three player game, this problem is represented by the following couple of matrices. The entries are payoffs for players 1 and 2, 1 is the row player, 2 is the column player and 3 chooses the matrix.

$$\begin{array}{c} \\ C_1 \\ C_2 \\ \\ C_1 \end{array} \begin{array}{c} \begin{array}{cc} C_1 & C_2 \end{array} \\ \begin{pmatrix} 0 & 0 \\ 0 & 1 \end{pmatrix} \\ \end{array} \quad \begin{array}{c} \begin{array}{cc} C_1 & C_2 \end{array} \\ \begin{pmatrix} 1 & 0 \\ 0 & 0 \end{pmatrix} \\ C_2 \end{array}$$

Formulated in our problem, gamblers 1 and 2 can be seen as a single entity: the decision maker. The payoff for the d.m. is the common payoff for them. The set of outcomes is $S = \{C_1, C_2\} \times \{C_1, C_2\}$. Gamblers 1 and 2 are allowed to choose their casinos at random but their choices are supposed to be independent: they cannot correlate their strategies. The action set for the d.m. is:

$$D = \{x \otimes y \mid x \in \Delta(\{C_1, C_2\}), y \in \Delta(\{C_1, C_2\})\}$$

$\otimes$ denotes the direct product of probabilities, and D is the subset of product probability distributions on $\{C_1, C_2\} \times \{C_1, C_2\}$. The d.m. assumes the worse: gambler 3 wants to minimize her payoff. Let σ be a probability distribution on $\{C_1, C_2\} \times \{C_1, C_2\}$. The payoff g is such that:

$$g(\sigma) = \min\{\sigma(C_1, C_1); \sigma(C_2, C_2)\}$$

Note that $\max\{g(d) \mid d \in D\} = \frac{1}{4}$ and is achieved with $d^* = x^* \otimes y^*$ such that $x^*(C_1) = y^*(C_2) = 0.5$.

Case 1: Perfect monitoring. Suppose that all gamblers have perfect observation i.e. each one observes the actions of all the others. Then, at each stage n, conditional on past signals, the next moves for gamblers 1 and 2 are independent. Stage payoffs are then at most $\frac{1}{4}$ which is the best they can guarantee by playing a^* at each stage.

Case 2: Gambler 3 receives no signal. Suppose that gamblers 1 and 2 have perfect observation whereas gambler 3 gets blank signals. Then gamblers 1 and 2 can perfectly correlate themselves from stage 2 on. At the first stage, each one chooses between C_1 and C_2 with equal probability. Then, if gambler 1 chose C_1 (resp. C_2) at the first stage, both choose C_1 (resp. C_2) at all subsequent stages. Conditional on gambler 3's information, the expected distribution on $\{C_1, C_2\} \times \{C_1, C_2\}$ at stage $n \geq 2$ is $0.5(1,0) \otimes (1,0) + 0.5(0,1) \otimes (0,1)$. The payoff is thus $\frac{1}{2}$ at those stages.

Case 3: Gambler 3 observes the actions of 2 only. Suppose now that gambler 3 observes the moves of gambler 2: if $s = (i, j)$ then $t = j$.

Consider the following strategy. The d.m. chooses d^* at each odd period. Let i_{2n-1} be the move of gambler 1 at stage $2n-1$. If $i_{2n-1} = C_1$ then at stage $2n$ the d.m. chooses $d = (1,0) \otimes (1,0)$: gamblers 1 and 2 meet at C_1 with probability 1 and if $i_{2n-1} = C_2$, at stage $2n$ the d.m. chooses $d = (0,1) \otimes (0,1)$: gamblers 1 and 2 meet at C_2 with probability 1. So at each even period, conditional on the information of the o.o., the d.m. chooses the action $0.5(1,0) \otimes (1,0) + 0.5(0,1) \otimes (0,1)$ and his payoff is 0.5. On the average (as n grows to infinity) this strategy yields 0.325. Our main theorem will show that the decision maker can actually guarantee more, characterize the maximum payoff that can be guaranteed, and show how to construct optimal strategies for the d.m. .

Case 4: No signals. We assume now that neither 1, 2, or 3 gets to observe the actions of the other players. If player 1 and 2 can correlate their actions at the beginning of the game according to the toss of a fair coin, they can guarantee an expected payoff of $\frac{1}{2}$ in the repeated game. If no correlation is possible between them at the beginning of the game, they can guarantee $\frac{1}{4}$ only.

2.3. **The convergence results.** Existence of $\lim_n v_n$ and $\lim_{\lambda \to 1} v_\lambda$ are obtained easily.

Proposition 2. *1. v_n converges to $\sup_n v_n$.*

2. $\lim_{\lambda\to 1} v_\lambda$ exists and equals $\sup_n v_n$ which is also equal to $\sup_\lambda v_\lambda$.

Proof. 1. We claim that $(nv_n)_n$ is superadditive, that is for each positive integers n, m:

$$v_{n+m} \geq \frac{n}{n+m} v_n + \frac{m}{n+m} v_m$$

Let ξ and ξ' be two strategies and n, m be two positive integers. Define the following strategy ξ'': play according to ξ until stage n and at stage $n+1$, forget all past outcomes and play ξ' as if the problem had started at this stage. We get then,

$$v_{n+m} \geq \gamma_{n+m}(\xi'') = \frac{n}{n+m}\gamma_n(\xi) + \frac{m}{n+m}\gamma_m(\xi')$$

Taking suprema over ξ and ξ' yields the inequality. Choose now n_0 that achieves the supremum up to $\varepsilon > 0$. Each subsequence of the form $(v_{kn_0+r})_k$ with $r = 1, \dots, n_0$ has all its cluster points above $\sup_n v_n - \varepsilon$. The result follows.

2. For each discount factor λ and strategy ξ, the discounted payoff is a convex combination of the finite averages. We have (see Lehrer and Sorin [LS92]) :

$$\gamma_\lambda(\xi) = \sum\nolimits_{n\geq 1} (1-\lambda)^2 n \lambda^{n-1} \gamma_n(\xi)$$

Since $\gamma_n(\xi) \leq \sup_n v_n$, $\gamma_\lambda(\xi) \leq \sup_n v_n$.

Take $n \geq 1$, $\varepsilon > 0$ and let ξ be a strategy such that $\gamma_n(\xi) \geq v_n - \varepsilon$. Define a cyclic strategy ξ' as follows: play ξ until stage n and restart this strategy every n stages. For each $m = 1, \dots, n$, set $y_m = \mathbf{E}_\xi \left[g(P_\xi(\cdot|\mathbf{t}^{m-1})) \right]$. Because ξ' is cyclic:

$$\gamma_\lambda(\xi') = \sum\nolimits_{m=1}^{n} (1-\lambda)\lambda^{m-1} y_m + \lambda^n \gamma_\lambda(\xi')$$

So,

$$\gamma_\lambda(\xi') = \sum\nolimits_{m=1}^{n} (1-\lambda) \frac{\lambda^{m-1}}{1-\lambda^n} y_m$$

Then, $\lim_{\lambda\to 1} \gamma_\lambda(\xi') = \frac{1}{n}\sum_{m=1}^{n} y_m \geq v_n - \varepsilon$ which ends the proof of the proposition. □

From now on, we shall denote by v_∞ the common value of $\lim v_n$ and $\lim v_\lambda$. Our main result is a characterization of v_∞.

2.4. A characterization of v_∞.

2.4.1. *A brief reminder on information theory.* Let $\mathbf{x}$ be a random variable with finitely many values $\{x_1, \dots, x_L\}$ and with law $p = (p_k)_{k=1}^L$. Set for each x, $p(x) = p(\mathbf{x} = x) = p_x$. Throughout the paper, we write log to denote the logarithm with base 2. By definition, the entropy of $\mathbf{x}$ is:

$$H(\mathbf{x}) = -\mathbf{E}[\log p(\mathbf{x})] = -\sum\nolimits_{k=1}^{L} p_k \log p_k$$

Note that $H(\mathbf{x})$ is non-negative and depends only on the law p of $\mathbf{x}$. One can therefore define the entropy of a probability distribution $p = (p_k)_{k=1,\dots,L}$ with finite support by $H(p) = -\sum_{k=1}^{L} p_k \log p_k$.

Let $(\mathbf{x}, \mathbf{y})$ be a couple of finite random variables with joint law p. For each x, y, define $p(x|y)$ as $p(\mathbf{x} = x|\mathbf{y} = y)$ if $p(\mathbf{y} = y) > 0$ and arbitrarily otherwise. The conditional entropy of $\mathbf{x}$ given $\mathbf{y}$ is:

$$H(\mathbf{x} \mid \mathbf{y}) = -\mathbf{E}[\log p(\mathbf{x} \mid \mathbf{y})] = -\sum\nolimits_{x,y} p(x,y) \log p(x \mid y)$$

The following is the fundamental relation of additivity of entropies:

$$H(\mathbf{x}, \mathbf{y}) = H(\mathbf{y}) + H(\mathbf{x} \mid \mathbf{y})$$

2.4.2. *Main definitions.* Let ξ be a strategy. At each stage n, after any observed sequence $\mathrm{t}^n = (t_1, \ldots, t_n)$, all relevant data about past outcomes is contained in the probability distribution $p_n = P_\xi(\cdot|\mathrm{t}^n)$. In particular, the next stage payoff depends on p_n and future payoffs depend on the evolution of this variable. Our optimization problem can thus be represented as a dynamic programming problem whose state space is a set of probability distributions with finite support and with p_n as the state variable. Suppose that at stage n, $t_1, \ldots, t_n$ has been observed by the o.o. The situation is as if an element $s_1, \ldots, s_n$ had been drawn from the finite set $\mathcal{S}^n$ with probability $P_\xi(s_1, \ldots, s_n \mid t_1, \ldots, t_n)$ and announced to the d.m. but not to the o.o. This motivates the following definition.

Definition 3.

- *A decision system is a pair $c = (p, d^{\mathcal{L}})$ where p is a probability distribution on a finite set $\mathcal{L}$ and $d^{\mathcal{L}}$ is an element of $D^{\mathcal{L}}$.*
- *Let $\mathcal{C}(D)$ be the set of all decision systems.*

A decision system will be feasible in state p_n if the associated probability distribution p can be obtained by compressing the information contained in the state, i.e. there is a mapping $\phi : S^n \to \mathcal{L}$ such that for each $k \in \mathcal{L}$, $p(k) = p_n(\phi^{-1}(k))$. We shall say that the strategy ξ plays according to $c = (p, d^{\mathcal{L}})$ at stage n after $t_1, \ldots, t_{n-1}$ occurred if:

- $\mathcal{L} = S^{n-1}$
- $\forall k \in \mathcal{L}$, $p(k) = P_\xi(k|t_1, \ldots, t_{n-1})$
- $d^{\mathcal{L}}$ is the restriction of ξ to sequences of length $n-1$.

Playing at some stage according to a system c has an effect on the uncertainty of the o.o. about past outcomes. At stage n, after $t_1, \ldots, t_n$ occurred, all the relevant data about this uncertainty is contained in the probability distribution p_n on S^n, $p_n = P_\xi(\cdot|t_1, \ldots, t_n)$. Suppose that at stage $n+1$, ξ plays according to c. The next *state* p_{n+1} is in $\Delta(S^{n+1})$ and equals $P_\xi(\cdot|t_1, \ldots, t_{n+1})$ with probability $P_\xi(t_{n+1}|t_1, \ldots, t_n)$.

When she chooses a decision system, the d.m. faces a trade-off between maximizing the stage-payoff and controlling the transitions to the next state on which future payoff depend. We choose now to measure the uncertainty contained in a finite probability distribution by its entropy. It will be a consequence of our result that controlling the real-valued variable $H(p_n)$ is enough to solve this problem.

Definition 4. *Let $c = (p, d^{\mathcal{L}})$ be a decision system.*

- *The payoff yielded by c is $\pi(c) = g(\sum_k p(k) d^{\mathcal{L}}(k))$.*
- *Let $\mathbf{k}_c$ be a random variable with law p. Denote by $\mathbf{s}_c$ the random signal induced by $\mathbf{k}_c$ and $d^{\mathcal{L}}$. The law of $\mathbf{s}_c$ is $\sum_k p(k) d^{\mathcal{L}}(k)$. Let also $\mathbf{t}_c$ be the random variable in T such that $\mathbf{s}_c \in \mathbf{t}_c$. The entropy variation of c is:*

$$\Delta H(c) = H(\mathbf{k}_c, \mathbf{s}_c \mid \mathbf{t}_c) - H(\mathbf{k}_c) = H(\mathbf{s}_c \mid \mathbf{k}_c) - H(\mathbf{t}_c)$$

The *entropy variation* is just the new entropy minus the initial entropy. Because of the properties of conditional entropy, the entropy variation can be written as the difference between the entropy gain and the entropy loss. The entropy gain is the additional uncertainty contained in $\mathbf{s}_c$; the entropy loss is the entropy of $\mathbf{t}_c$ which is observed by the o.o. .

We consider all vectors of the form $(\Delta H(c), \pi(c))$:

$$V = \{(\Delta H(c), \pi(c)) \mid c \in \mathcal{C}(D)\}$$

Lemma 5. *V is a compact subset of $\mathbb{R}^2$.*

Proof. Let L be a positive integer and $\mathcal{C}_L(D)$ be the set of elements of $\mathcal{C}(D)$ whose associated probability p has at most L points in its support. We set then:

$$V_L = \{(\Delta H(c), \pi(c)) \mid c \in \mathcal{C}_L(D)\}$$

We will prove that $V = V_L$ for some L. Since $\mathcal{C}_L(D)$ is a compact set and the mappings $\pi(\cdot)$ and $\Delta H(\cdot)$ are continuous on it, the result follows.

Note that for each $c \in \mathcal{C}(D)$, $\Delta H(c) = \sum_k p(k) H(d^{\mathcal{L}}(k)) - H(\mathbf{t}_c)$ and that $H(\mathbf{t}_c)$ depends only on the law of $\mathbf{s}_c$, $\sum_k p(k) d^{\mathcal{L}}(k)$. Thus, the vector $(\Delta H(c), \pi(c))$ is a function of the vector

$$\Big(\sum\nolimits_k p(k) d^{\mathcal{L}}(k), \sum\nolimits_k p(k) H(d^{\mathcal{L}}(k))\Big)$$

which belongs to:

$$\operatorname{co}\{\big(d, H(d)\big) \mid d \in D\}$$

The set $\{\big(d, H(d)\big) \mid d \in D\}$ lies in $\mathbb{R}^{|S|+1}$. From Carathéodory's theorem, any convex combination of elements of this set is a convex combination of at most $|S| + 2$ points. □

2.4.3. *Statement of the main result.* Our main theorem rules out the case in which the signal to the o.o. reveals the outcome. So:

Definition 6. *The problem has imperfect information structure if there is a decision d in D, and $s, s' \in t$ such that $\rho(s \mid d)\rho(s' \mid d) > 0$.*

Identify each d in D with a special $c_d = (p, d) \in \mathcal{C}(D)$ such that p is a Dirac measure. The problem has imperfect information structure if and only if there exists d such that $\Delta H(c_d) > 0$.

This condition is necessary and sufficient to get existence of a decision that allows to enter a state $p_n = P_\xi(\cdot|t_1, \dots, t_n)$ with $H(p_n) > 0$. In fact, under this condition, states with arbitrarily large entropy can be reached.

Our main result is the following:

Theorem 7. *If the problem has imperfect information structure, then:*

$$v_\infty = \sup\{x_2 \in \mathbb{R} \mid (x_1, x_2) \in \operatorname{co} V, x_1 \geq 0\}$$

Otherwise, for each n, $v_n = \max\{g(d) \mid d \in D\}$.

In words, if the problem has imperfect information structure, then the limiting value is the highest payoff associated to a convex combination of decision systems under the constraint that the entropy variation is non-negative. Observe that since V is compact and intersects the half-plane $x_1 \geq 0$, the supremum is indeed a maximum.

A more functional statement can be obtained. Define for each real number h:

$$u(h) = \max\{\pi(c) \mid c \in \mathcal{C}(D), \Delta H(c) \geq h\}$$

From the definition of V we have for each h:

$$u(h) = \max\{x_2 \mid (x_1, x_2) \in V, x_1 \geq h\}$$

Since V is compact, $u(h)$ is well defined. Let $\operatorname{cav} u$ be the least concave function greater than u. Then:

$$\sup\{x_2 \in \mathbb{R} \mid (x_1, x_2) \in \operatorname{co} V, x_1 \geq 0\} = \operatorname{cav} u(0)$$

Indeed, u is upper-semi-continuous, decreasing and the hypograph of u is the comprehensive set $V - \mathbb{R}^2_+$ associated to V. This implies that $\operatorname{cav} u$ is also decreasing, u.s.c. and its hypograph is $\operatorname{co}(V - \mathbb{R}^2_+)$.

2.4.4. *Sketch of the proof.* The proof of theorem 7 consists of two parts. First (lemma 19), we prove that the d.m. cannot guarantee more than $\sup\{x_2 \in \mathbb{R} \mid (x_1, x_2) \in \operatorname{co} V, x_1 \geq 0\}$. Given any strategy for the d.m. , we consider the (random) sequence of decision systems that it induces and prove that for any stage n, the average entropy variation (over all stages up to stage n and over all histories) is non-negative. From this we deduce that the pair (average entropy variation, average payoff) belongs to V, and conclude that the average payoff up to stage n can be no more than $\sup\{x_2 \in \mathbb{R} \mid (x_1, x_2) \in \operatorname{co} V, x_1 \geq 0\}$.

For the second part of the proof, we consider two decision systems c_1 and c_2 and $n_1, n_2 \in \mathbb{N}$ such that $n_1 \Delta H(c_1) + n_2 \Delta H(c_2) \geq 0$, and construct a strategy of the d.m. that guarantees a payoff $\frac{n_1}{n_1+n_2}\pi(c_1) + \frac{n_2}{n_1+n_2}\pi(c_2)$ (up to any $\varepsilon > 0$). The idea is to approximate an "ideal" strategy of the d.m. that would play cycles of c_1 for n_1 stages followed by c_2 for n_2 stages. Assume a strategy ξ has been defined up to state m, and that the history of signals t^m to the o.o. up to stage m is such that the distribution $P_\xi(\mathrm{s}^m \mid \mathrm{t}^m)$ of the history of outcomes s^m, is close to (up to some space isomorphism) the distribution of N r.v.'s i.i.d. $(\frac{1}{2}, \frac{1}{2})$. We then prove that, for $l \in \{1, 2\}$, ξ can be extended by a strategy at stages $m, \dots m + n_l$ in such a way that:

- At each stage $m < n \leq m + n_l$, $P_\xi(s_n \mid \mathrm{t}^{n-1})$ is close to the distribution induced by c_l;
- With large probability (on the set of histories of the o.o.), $P_\xi(\mathrm{s}^{m+n_l} \mid \mathrm{t}^{m+n_l})$ is close to the distribution of $N + n_l \Delta H(c_l)$ r.v.'s i.i.d. $(\frac{1}{2}, \frac{1}{2})$.

We then apply the above described result inductively to construct ε-optimal strategies of the d.m.

2.4.5. *Back to example 2.2, case 3.* Let us consider the cyclic strategy proposed in the example. It consists in playing alternatively two decisions systems. With c_{+1} the decision system in which p is a dirac measure and the d.m. chooses $d = (\frac{1}{2}, \frac{1}{2}) \otimes (\frac{1}{2}, \frac{1}{2})$, $\pi(c_{+1}) = 0.25$ and $\Delta H(c_{+1}) = +1$. Letting c_{-1} in which $p = \frac{1}{2}H + \frac{1}{2}T$ is the law of a fair coin and the d.m. chooses $(1, 0) \otimes (1, 0)$ if H, $(0, 1) \otimes (0, 1)$ if T, $\pi(c_{-1}) = 0.5$ and $\Delta H(c_{-1}) = -1$ since the move of gambler 2 reveals both the action of gambler 1 and the realization of the coin. Playing c_{+1} at odd stages and c_{-1} at even stages gives an average payoff of 0.325 and an average entropy variation of 0.

We now prove the existence of strategies for gamblers 1 and 2 that guarantee more than 0.325. To do this, we show the existence of a convex combination of two decision systems with average payoff larger than 0.325 and a non-negative average entropy variation, and apply theorem 7.

Define the decision system c_ϵ where p is a fair coin an where the d.m. chooses $(1 - \epsilon, \epsilon) \otimes (1, 0)$ if H, and $(\epsilon, 1 - \epsilon) \otimes (0, 1)$ if T. We have $\pi(c_\epsilon) = \frac{1-\epsilon}{2}$ and $\Delta H(c_\epsilon) = h(\epsilon) - 1$ where for $x \in]0, 1[$, $h(x) = -x \log(x) - (1 - x) \log(1 - x)$, $h(0) = h(1) = 0$. Using that $h'(0) = +\infty$, we deduce the existence of $\epsilon > 0$ such that $(\Delta H(c_\epsilon), \pi(c_\epsilon))$ lies above the line

$$\{\lambda(-1, 0.5) + (1 - \lambda)(1, 0.25), \lambda \in [0, 1]\}$$

For this ε, there exists $0 \leq \lambda \leq 1$ such that $\lambda \Delta H(c_\epsilon) + (1 - \lambda)\Delta H(c_{+1}) = 0$ and $\lambda \pi(c_\epsilon) + (1 - \lambda)\pi(c_{+1}) > 0.325$, which in turn implies that the gamblers can guarantee more than 0.325.

3. Further information theory tools

3.1. **Kullback distance and related tools.** We recall the classical notion of Kullback distance (see Cover and Thomas [CT91]) and some notions introduced in [GV02] which will be used in the proofs.

Definition 8. *Let X be a finite set and P, Q in $\Delta(X)$ such that $P \ll Q$: $Q(x) = 0 \Rightarrow P(x) = 0$, the Kullback distance between P and Q is,*

$$d(P||Q) = \mathbf{E}_P\left[\log \frac{P(\cdot)}{Q(\cdot)}\right] = \sum\nolimits_x P(x) \log \frac{P(x)}{Q(x)}$$

We recall some useful properties of the Kullback distance.

Lemma 9. (1) *$d(P||Q) \geq 0$ and $d(P||Q) = 0 \Rightarrow P = Q$.*

(2) *$d(P||Q)$ is a convex function of the pair (P, Q).*

(3) *Let $X = S^n$ with n a positive integer. Denote $\mathrm{s}^{\mathrm{m}-1} = (s_1, \ldots, s_{m-1})$ for $m = 1, \ldots, n$. Let $P(\cdot|\mathrm{s}^{\mathrm{m}-1})$ be the distribution of s_m given $\mathrm{s}^{\mathrm{m}-1}$ induced by P. Then,*

$$d(P||Q) = \sum\nolimits_{m=1}^{n} \mathbf{E}_P\left[d(P(\cdot|\mathbf{s}^{m-1})||Q(\cdot|\mathbf{s}^{m-1})\right]$$

(4)

$$\|P - Q\|_1 \leq f(d(P||Q))$$

with $\forall \delta \geq 0, f(\delta) = \sqrt{(2 \ln 2)\delta}$.

Proof. We refer to theorem 2.6.3 p.26, theorem 2.7.2 p.30, theorem 2.5.3 p.23 and lemma 12.6.1 p.300 in [CT91]. □

In our main strategy construction, we shall compare the induced probability P on S^n with some ideal distribution Q. Namely, P and Q will be close in Kullback distance. We argue now that this implies that payoffs will also be close. The ideal distribution Q turns out to be the independent product of its marginals. So let $q_1, \ldots, q_n$ be elements of $\Delta(S)$ and set $Q = q_1 \otimes \cdots \otimes q_n$. For each stage $m = 1, \ldots, n$ and sequence of signals $\mathrm{t}^{m-1} = (t_1, \ldots, t_{m-1})$, let $P(\cdot|\mathrm{t}^{m-1})$ be the distribution of s_m under P given t^{m-1}. Since the payoff function g is continuous, $\left|g(P(\cdot|\mathrm{t}^{m-1})) - g(q_m)\right|$ is small when $\left\|P(\cdot|\mathrm{t}^{m-1}) - q_m\right\|_1$ is small.

We denote also $\mathrm{s}^{m-1} = (s_1, \ldots, s_{m-1})$ and write $\mathrm{s}^{m-1} \in \mathrm{t}^{m-1}$ has a shorthand of $s_l \in t_l, l = 1, \ldots, m$. We give the following definition.

Definition 10. *Let n be a positive integer and $P, Q \in \Delta(S^n)$. The strategic distance from P to Q is:*

$$d^n_{\mathcal{S}}(P||Q) = \frac{1}{n} \sum\nolimits_{m=1}^{n} \mathbf{E}_P\left[\left\|P(\cdot|\mathbf{t}^{m-1}) - Q(\cdot|\mathbf{t}^{m-1})\right\|_1\right]$$

Note that this quantity depends on the partition T of S. Such a measure of distance was used in [GV02] in the case where T is the finest partition of S. The link with the Kullback distance is the following.

Proposition 11. *Let P, Q be distributions on S^n with $Q = q_1 \otimes \cdots \otimes q_n$.*

$$d^n_{\mathcal{S}}(P||Q) \leq f(\frac{1}{n} d(P||Q))$$

Proof. Applying Jensen's inequality to the concave function f of property 4 of lemma 9 yields:

$$d^n_{\mathcal{S}}(P||Q) \leq f\left(\frac{1}{n} \sum\nolimits_{m=1}^{n} \mathbf{E}_P\left[d(P(\cdot|\mathbf{t}^{m-1})||q_m)\right]\right)$$

Now, $P(\cdot|\mathrm{t}^{m-1}) = \sum_{\mathrm{s}^{m-1}\in\mathrm{t}^{m-1}} P(\mathrm{s}^{m-1}|\mathrm{t}^{m-1})P(\cdot|\mathrm{s}^{m-1})$. Since the Kullback distance is convex:

$$d(P(\cdot|\mathrm{t}^{m-1})||q_m) \leq \sum\nolimits_{\mathrm{s}^{m-1}\in\mathrm{t}^{m-1}} P(\mathrm{s}^{m-1}|\mathrm{t}^{m-1})d(P(\cdot|\mathrm{s}^{m-1})||q_m)$$

So by property 3 of lemma 9,

$$\sum\nolimits_{m=1}^{n} \mathbf{E}_P\left[d(P(\cdot|\mathbf{t}^{m-1})||q_m)\right] \leq \sum\nolimits_{m=1}^{n} \mathbf{E}_P\left[d(P(\cdot|\mathbf{s}^{m-1})||q_m)\right] = d(P||Q)$$

ending the proof. □

We recall the absolute Kullback distance from [GV02] for later use.

Definition 12. *Let X be a finite set and P, Q in $\Delta(X)$ such that $P \ll Q$, the absolute Kullback distance between P and Q is,*

$$|d|\,(P||Q) = \mathbf{E}_P\left|\log\frac{P(\cdot)}{Q(\cdot)}\right|$$

The following is proved in [GV02]:

Lemma 13. *For every P, Q in $\Delta(X)$ such that $P \ll Q$,*

$$d(P||Q) \leq |d|\,(P||Q) \leq d(P||Q) + 2$$

3.2. **Equipartition properties.** Let $(\mathbf{x}_n)_n$ be a sequence of i.i.d. random variables in a finite set X. The well-known Asymptotic Equipartition Property (see Cover and Thomas [CT91], chapter 3) asserts that for large n, almost all sequences $\mathrm{x}^n = (x_1, \ldots, x_n) \in X^n$ have approximately the same probability. More precisely, set $h = H(\mathbf{x}_1)$ and let $\eta > 0$. Let P be the probability measure on X^n defined by $(\mathbf{x}_1, \ldots, \mathbf{x}_n)$. Define the set of typical sequences:

$$C(n,\eta) = \left\{\mathrm{x}^n \in X^n, \left|-\frac{1}{n}\log P(\mathrm{x}^n) - h\right| \leq \eta\right\}$$

Since the variables are i.i.d., the weak law of large numbers ensures that for each $\varepsilon > 0$, there is n_ε such that $P(C(n,\eta)) \geq 1-\varepsilon$ for $n \geq n_\varepsilon$. We need to depart from the i.i.d. assumption and work with probability measures that verify some equipartition property.

Definition 14. *Let $P \in \Delta(X)$, $n \in \mathbb{N}$, $h \in \mathbb{R}_+$, $\eta, \epsilon > 0$. P verifies an* $\mathbf{AEP}(n, h, \eta, \varepsilon)$, *when*

$$P\{x \in X, |-\frac{1}{n}\log P(x) - h| \leq \eta\} \geq 1-\varepsilon$$

The proof of our main result will rely heavily on a stability property of **AEP**s given in proposition 15. We first state it informally.

Assume the d.m. would like to use a decision system (μ, σ) repeatedly i.i.d. for n stages, and this would induce an "ideal" probability Q over $(\mathcal{K}\times S)^n$. Assume that a distribution $P_{\mathcal{L}}$ satisfying an $\mathbf{AEP}(n, h, \eta, \varepsilon)$ is available, with h greater than $H(\mu) + \eta + 2\frac{\varepsilon}{n}$. Then, the d.m. can mimick a random variable distributed "close to" $\mu^{\otimes n}$ from a random variable with law $P_{\mathcal{L}}$, and play at each stage according to σ given the realization of the mimicked random variable. The first part of the lemma states that the induced law P over $(\mathcal{K}\times S)^n$ is close enough to Q. The second says that, for a large proportion (under P) over sequences of signals to the o.o. , a distribution $P'_{\mathcal{L}}$ satisfying an $\mathbf{AEP}(n, h + \Delta H(c), \eta', \varepsilon')$ is available to the d.m. after the n stages are played (conditions are provided on η' and ε').

Given a finite set $\mathcal{K}$ the type of $\mathrm{k} = (k_1, \ldots, k_n) \in \mathcal{K}^n$ is the empirical distribution of k. The type set of $\mu \in \Delta(\mathcal{K})$ is the subset of K^n of sequences of type μ. Finally,

the set of types is $\mathbb{T}_n(\mathcal{K}) = \{\mu \in \Delta(\mathcal{K}), T_n(\mu) \neq \emptyset\}$. The following estimates the size of $T_n(\mu)$ for $\mu \in \mathbb{T}_n(\mathcal{K})$ (Cover and Thomas [CT91] Theorem 12.1.3 page 282):

$$\frac{2^{nH(\mu)}}{(n+1)^{|\mathcal{K}|}} \leq |T_n(\mu)| \leq 2^{nH(\mu)} \tag{1}$$

Proposition 15. *Let $\mu \in \mathbb{T}_n(\mathcal{K})$, $\sigma\colon \mathcal{K} \to D$. Denote by $\rho = \mu \otimes \sigma$ and $Q = \rho^{\otimes n}$ the probabilities over $\mathcal{K} \times S$ and $(\mathcal{K} \times S)^n$ induced by μ and σ, and let ρ_T be the marginal of ρ on T. There exists constants $(a_\varepsilon)_{\varepsilon>0}$ such that for any $P_{\mathcal{L}} \in \Delta(\mathcal{L})$ that verifies an $\mathbf{AEP}(n, h, \eta, \varepsilon_P)$ with $h \geq H(\mu) + \eta + 2\frac{\varepsilon_P}{n}$ and $0 < \varepsilon_P < \frac{1}{16}$, there exists a mapping $\varphi\colon \mathcal{L} \to \mathcal{K}^n$ such that, letting $P \in \Delta(\mathcal{K} \times S)^n$ be induced by $P_{\mathcal{L}}$, φ, and σ:*

1. $d(P||Q) \leq 2n(\eta + \varepsilon_P \log|\mathcal{K}|) + |\mathcal{K}| \log(n+1) + 1$
2. *For every $\varepsilon > 0$, there exists a subset $\mathcal{U}_\varepsilon$ of T^n such that:*
   (a) $P(\mathcal{U}_\varepsilon) \geq 1 - \varepsilon - 2\sqrt{\varepsilon_P}$
   (b) *For $\mathrm{t} \in \mathcal{U}_\varepsilon$, $P(\cdot|\mathrm{t})$ verifies an $\mathbf{AEP}(n, h', \eta', \varepsilon + 3\sqrt{\varepsilon_P})$*

with $h' = H(\rho) - H(\rho_T)$ and $\eta' = a_\varepsilon(\eta + \frac{\log(n+1)}{n}) + 4\frac{\sqrt{\varepsilon_P}}{n}$.

3.3. **Proof of proposition 15.**

Definition 16. *Let $P \in \Delta(X)$, $n \in \mathbb{N}$, $h \in \mathbb{R}_+$,$\eta > 0$. P verifies an $\mathbf{EP}(n, h, \eta)$, when*

$$P\{x \in X, |-\frac{1}{n}\log P(x) - h| \leq \eta\} = 1$$

Lemma 17. *Suppose that P verifies an $\mathbf{AEP}(n, h, \eta, \varepsilon)$. Let the typical set of P be:*

$$C = \{x \in X, |-\frac{1}{n}\log P(x) - h| \leq \eta\}$$

Let $P_C \in \Delta(X)$ be the conditional probability given C: $P_C(x) = P(x|C)$. Then, P_C verifies an $\mathbf{EP}(n, h, \eta')$ with $\eta' = \eta + 2\frac{\varepsilon}{n}$ for $0 < \varepsilon < \frac{1}{2}$.

Proof. Follows immediately, since for $0 < \varepsilon < \frac{1}{2}$, $-\log(1-\varepsilon) \leq 2\varepsilon$. □

Before proving prop. 15, we establish a similar result when $P_{\mathcal{L}}$ verifies an **EP** (instead of an **AEP**).

Proposition 18. *Let $\mu \in \mathbb{T}_n(\mathcal{K})$, $\sigma\colon \mathcal{K} \to D$. Denote by $\rho = \mu \otimes \sigma$ and $Q = \rho^{\otimes n}$ the probabilities over $\mathcal{K} \times S$ and $(\mathcal{K} \times S)^n$ induced by μ and σ, and let ρ_T be the marginal of ρ on T. There exists constants $(a_\varepsilon)_{\varepsilon>0}$ such that for any $P_{\mathcal{L}} \in \Delta(\mathcal{L})$ that verifies an $\mathbf{EP}(n, h, \eta)$ with $h \geq H(\mu) + \eta$, there exists a mapping $\varphi\colon \mathcal{L} \to \mathcal{K}^n$ such that, letting $P \in \Delta(\mathcal{K} \times S)^n$ be induced by $P_{\mathcal{L}}$, φ, and σ:*

1. $d(P||Q) \leq 2n\eta + |\mathcal{K}| \log(n+1) + 1$
2. *For every $\varepsilon > 0$, there exists a subset $\mathcal{T}_\varepsilon$ of T^n such that:*
   (a) $P(\mathcal{T}_\varepsilon) \geq 1 - \varepsilon$
   (b) *For $\mathrm{t} \in \mathcal{T}_\varepsilon$, $P(\cdot|\mathrm{t})$ verifies an $\mathbf{AEP}(n, h', \eta', \varepsilon)$*

with $h' = H(\rho) - H(\rho_T)$ and $\eta' = a_\varepsilon(\eta + \frac{\log(n+1)}{n})$.

Proof of prop. 18.
Construction of φ: Let $\tilde{\mathcal{L}} = \{l \in \mathcal{L}, P_{\mathcal{L}}(l) > 0\}$. Since P_L verifies an $\mathbf{EP}(n, h, \eta)$,

$$2^{n(h-\eta)} \leq |\tilde{\mathcal{L}}| \leq 2^{n(h+\eta)}$$

Using the previous and equation (1), we can choose $\varphi\colon \tilde{\mathcal{L}} \to T_n(\mu)$ such that for every $\mathrm{k} \in \mathcal{K}^n$,

$$2^{n(h-\eta-H(\mu))} - 1 \leq |\varphi^{-1}(\mathrm{k})| \leq (n+1)^{|K|} 2^{n(h+\eta-H(\mu))} + 1 \tag{2}$$

Bound on $d(P||Q)$: P and Q are probabilities over $(\mathcal{K}\times S)^n$ which are deduced from their marginals on $\mathcal{K}^n$ by the same transition probabilities. Hence, letting $P_{\mathcal{K}}$ and $Q_{\mathcal{K}}$ denote their marginals on $\mathcal{K}^n$, $d(P||Q)=d(P_{\mathcal{K}}||Q_{\mathcal{K}})$. By definition of the Kullback distance:

$$d(P_{\mathcal{K}}||Q_{\mathcal{K}})=\sum\nolimits_{\mathrm{k}\in T_n(\mu)} P_{\mathcal{K}}(\mathrm{k})\log\frac{P_{\mathcal{K}}(\mathrm{k})}{Q_{\mathcal{K}}(\mathrm{k})}$$

Using equation 2 and the **EP** for $P_{\mathcal{L}}$, we get for $\mathrm{k}\in T_n(\mu)$

$$P_{\mathcal{K}}(\mathrm{k})\leq(n+1)^{|K|}2^{n(2\eta-H(\mu))}+2^{-n(h-\eta)}$$

On the other hand, since for all $\mathrm{k}\in T_n(\mu)$, $Q_{\mathcal{K}}(\mathrm{k})=2^{-nH(\mu)}$.

$$\frac{P_{\mathcal{K}}(\mathrm{k})}{Q_{\mathcal{K}}(\mathrm{k})}\leq(n+1)^{|K|}2^{2n\eta}+2^{-n(h-\eta-H(\mu))}$$

Part (1) of the proposition now follows since $H(\mu)\leq h-\eta$.

Estimation of $|d|(P(\cdot|\mathrm{t})||Q(\cdot|\mathrm{t}))$: For $\mathrm{t}\in T^n$ s.t. $P(\mathrm{t})>0$, we let P_{t} and Q_{t} in $\Delta((\mathcal{K}\times S)^n)$ denote $P(\cdot|\mathrm{t})$ and $Q(\cdot|\mathrm{t})$ respectively. Direct computation yields:

$$\sum\nolimits_{\mathrm{t}} P(\mathrm{t})d(P_{\mathrm{t}}||Q_{\mathrm{t}})=d(P||Q)$$

Hence for any value of a parameter $\alpha_1>0$ that we shall fix later:

$$P\left\{\mathrm{t},d(P_{\mathrm{t}}||Q_{\mathrm{t}})\geq\alpha_1\right\}\leq\frac{2n\eta+|\mathcal{K}|\log(n+1)+1}{\alpha_1}$$

and so from lemma 13,

$$P\left\{\mathrm{t},|d|(P_{\mathrm{t}}||Q_{\mathrm{t}})\leq\alpha_1+2\right\}\geq 1-\frac{2n\eta+|\mathcal{K}|\log(n+1)+1}{\alpha_1} \tag{3}$$

The statistics of (k,s) under P: We prove that the empirical distribution $\rho_{\mathrm{e}}\in\Delta(\mathcal{K}\times S)$ of $\mathrm{e}\in(\mathcal{K}\times S)^n$ is close to ρ, with large P-probability. Since φ takes its values on $T_n(\mu)$, the marginal of ρ_{e} on $\mathcal{K}$ is μ with P-probability one. Furthermore, for $(k,s)\in\mathcal{K}\times S$, the distribution under P of $n\rho_{\mathrm{e}}(k,s)$ is the one of a sum of $n\mu(k)$ Bernoulli independent trials having value 1 with probability $q_k(s)$ and 0 with probability $1-q_k(s)$. Hence, for $\alpha_2>0$ the Bienaymé-Chebyshev inequality gives:

$$P(|\rho_{\mathrm{e}}(k,s)-\rho(k,s)|\geq\alpha_2)\leq\frac{\rho(k,s)}{n\alpha_2^2}$$

Hence,

$$P\left(\|\rho_{\mathrm{e}}-\rho\|_\infty\leq\alpha_2\right)\geq 1-\frac{1}{n\alpha_2^2} \tag{4}$$

The set of $\mathrm{t}\in T^n$ s.t. Q_{t} verifies an AEP has large P-probability: When $\tau\in\Delta(\mathcal{K}\times S)$ and $t\in T$, we let $\tau(t)=\sum_{(k,s)\in\mathcal{K}\times S,s\in t}\tau(k,s)$: τ is then seen as an element of $\Delta(T)$. For $(\mathrm{e},\mathrm{t})=(\mathrm{k},\mathrm{s},\mathrm{t})=(k_i,s_i,t_i)_i\in(K\times\mathcal{S}\times\mathcal{T})^n$ s.t. $s_i\in t_i$ for every i, we compute:

$$\begin{aligned}
-\frac{1}{n}\log Q_{\mathrm{t}}(\mathrm{k},\mathrm{s}) &= -\frac{1}{n}(\sum\nolimits_i\log\rho(k_i,s_i)-\log\rho(t_i))\\
&= -\sum\nolimits_{(k,s)\in\mathcal{K}\times S}\rho_{\mathrm{e}}(k,s)\log\rho(k,s)+\sum\nolimits_{t\in T}\rho_{\mathrm{e}}(t)\log\rho_{\mathrm{e}}(t)\\
&= -\sum\nolimits_{(k,s)}\rho(k,s)\log\rho(k,s)+\sum\nolimits_t\rho(t)\log\rho(t)\\
&\quad+\sum\nolimits_{(k,s)}(\rho(k,s)-\rho_{\mathrm{e}}(k,s))\log\rho(k,s)-\sum\nolimits_t(\rho(t)-\rho_{\mathrm{e}}(t))\log\rho(t)
\end{aligned}$$

Now, since $-\sum_{(k,s)} \rho(k,s)\log\rho(k,s) = H(\rho)$ and $\sum_t \rho(t)\log\rho(t) = -H(Q_T)$:

$$|-\frac{1}{n}\log Q_{\mathrm{t}}(\mathrm{k},\mathrm{s}) - h'| \leq -2|K\times S|\log(\min_{k,s}\rho(k,s))\|\rho-\rho_{\mathrm{e}}\|_\infty \tag{5}$$

With $M = -2|K\times S|\log(\min_{k,s}\rho(k,s))$, define

$$\begin{aligned} A_{\alpha_2} &= \{(\mathrm{k},\mathrm{s},\mathrm{t}), |-\frac{1}{n}\log Q_{\mathrm{t}}(\mathrm{k},\mathrm{s}) - h'| \leq M\alpha_2\} \\ A_{\alpha_2,\mathrm{t}} &= A_{\alpha_2}\cap \mathcal{K}\times\mathcal{S}\times\{\mathrm{t}\},\ \mathrm{t}\in T^n \end{aligned}$$

Using equations 4 and 5 we deduce:

$$\begin{aligned} \sum\nolimits_{\mathrm{t}} P(\mathrm{t})P_{\mathrm{t}}(A_{\alpha_2,\mathrm{t}}) &= P(A_{\alpha_2}) \\ &\geq 1-\frac{1}{n\alpha_2^2} \end{aligned}$$

Thus, for $\beta>0$,

$$P\{\mathrm{t}, P_{\mathrm{t}}(A_{\alpha_2,\mathrm{t}}) \leq 1-\beta\} \leq 1-\frac{1}{n\alpha_2^2\beta} \tag{6}$$

Definition of $\mathcal{T}_\varepsilon$ and verification of point 2 of prop. 18: In order to complete the proof we now fix the parameters:

$$\begin{cases} \alpha_1 &= \frac{4n\eta+2|\mathcal{K}|\log(n+1)+2}{\varepsilon} \\ \alpha_2 &= \frac{1}{4n\varepsilon^2} \\ \beta &= \frac{\varepsilon}{2} \end{cases}$$

and let:

$$\begin{cases} \mathcal{T}_\varepsilon^1 &= \{\mathrm{t}, |d|(P_{\mathrm{t}}||Q_{\mathrm{t}}) \leq \alpha_1+2\} \\ \mathcal{T}_\varepsilon^2 &= \{\mathrm{t}, P_{\mathrm{t}}(A_{\alpha_2,\mathrm{t}}) \leq 1-\beta\} \\ \mathcal{T}_\varepsilon &= \mathcal{T}_\varepsilon^1\cap\mathcal{T}_\varepsilon^2 \end{cases}$$

Then, equations 3 and 6 imply

$$P(\mathcal{T}_\varepsilon) \geq 1-\varepsilon$$

We now prove that for $\mathrm{t}\in\mathcal{T}_\varepsilon$, P_{t} verifies an **AEP**. For such t, the definition of $\mathcal{T}_\varepsilon^1$ and equation 3 imply:

$$P_{\mathrm{t}}\left\{|\log P_{\mathrm{t}}(\cdot) - \log Q_{\mathrm{t}}(\cdot)| \leq \frac{2(\alpha_1+2)}{\varepsilon}\right\} \geq 1-\frac{\varepsilon}{2}$$

From the definition of $\mathcal{T}_\varepsilon^2$:

$$P_{\mathrm{t}}\left\{|-\frac{1}{n}\log Q_{\mathrm{t}}(\cdot) - h'| \leq M\alpha_2\right\} \geq 1-\frac{\varepsilon}{2}$$

The two above inequalities yield:

$$P_{\mathrm{t}}\left\{|-\frac{1}{n}\log P_{\mathrm{t}}(\cdot) - h'| \leq \frac{2(\alpha_1+2)}{n\varepsilon} + M\alpha_2\right\} \geq 1-\varepsilon$$

Hence the desired **AEP**. □

Proof of prop. 15. Let C be the typical set of $P_{\mathcal{L}}$ (as in lemma 17) and $P'_{\mathcal{L}} = P_{\mathcal{L}}(\cdot \mid C)$. From lemma 17, $P'_{\mathcal{L}}$ verifies an $\mathbf{EP}(n,h,\eta+2\frac{\varepsilon_P}{n})$. Applying prop. 15 to (μ,σ,n) yields constants $(a_\varepsilon)_\varepsilon$, and applied to $P'_{\mathcal{L}}$ yields $\varphi\colon C\to\mathcal{K}^n$, an induced probability P' on $(\mathcal{K}\times S)^n$, and subsets $(\mathcal{T}_\varepsilon)_\varepsilon$ of T^n. Choose $\bar{k}\in\arg\max\mu(k)$ and extend φ to $\mathcal{L}$ by setting it to $(\bar{k},\dots,\bar{k})$ outside C. The probability induced by $P_{\mathcal{L}}$ and φ on

$(\mathcal{K}\times S)^n$ is then $P = P_{\mathcal{L}}(C)P' + (1-P_{\mathcal{L}}(C))(\bar{k}\otimes\sigma)^{\otimes n}$ where we loosely indentify $\bar{k}$ with the unit mass on $\bar{k}$. To verify point 1, write:

$$\begin{aligned} d(P\|Q) &\leq P_{\mathcal{L}}(C)d(P'\|Q) + (1-P_{\mathcal{L}}(C))nd(\bar{k}\otimes\sigma(k)\|\rho) \\ &\leq d(P'\|Q) + \varepsilon_P n d(\bar{k}\|\mu) \\ &\leq d(P'\|Q) + \varepsilon_P n\log(|\mathcal{K}|) \end{aligned}$$

With $\mathcal{T} = \{\mathrm{t},\ P'(\mathrm{t}) > \sqrt{\varepsilon_P}(\bar{k}\otimes\sigma)^{\otimes n}(\mathrm{t})\}$, $P'(\mathcal{T}) \geq 1-\sqrt{\varepsilon_P}$ and then $P(\mathcal{T}) \geq 1-\varepsilon_P-\sqrt{\varepsilon_P}$. Let $\mathcal{U}_\varepsilon = \mathcal{T}_\varepsilon \cap \mathcal{T}$. We now prove that $\mathcal{U}_\varepsilon$ fulfills requirements 2a and 2b on $\mathcal{T}_\varepsilon$. Point 2a is straightforward. For 2b, for $\mathrm{t}\in\mathcal{U}_\varepsilon$, let $C(\mathrm{t})$ be the $(n, h', a_\varepsilon(\eta+\frac{\log(n+1)}{n}))$ typical set of $P(\cdot\mid\mathrm{t})$, and $A(\mathrm{t}) = \{(\mathrm{k},\mathrm{s}),\ P'(\mathrm{k},\mathrm{s}\mid\mathrm{t}) > \sqrt{\varepsilon_P}(\bar{k}\otimes\sigma)^{\otimes n}(\mathrm{k},\mathrm{s}\mid\mathrm{t})\}$. Then,

$$\begin{aligned} P(C(\mathrm{t})\cap A(\mathrm{t})\mid\mathrm{t}) &= \frac{(1-P_{\mathcal{L}}(C))(\bar{k}\otimes\sigma)^{\otimes n}(C(\mathrm{t})\cap A(\mathrm{t})) + P_{\mathcal{L}}(C)P'(C(\mathrm{t})\cap A(\mathrm{t}))}{(1-P_{\mathcal{L}}(C))(\bar{k}\otimes\sigma)^{\otimes n}(\mathrm{t}) + P_{\mathcal{L}}(C)P'(\mathrm{t})} \\ &\geq \frac{(1-\varepsilon_P)P'(C(\mathrm{t})\cap A(\mathrm{t})}{(\sqrt{\varepsilon_P}+1)P'(\mathrm{t})} \\ &\geq (1-2\sqrt{\varepsilon_P})P'(C(\mathrm{t})\cap A(\mathrm{t})\mid\mathrm{t}) \\ &\geq 1-3\sqrt{\varepsilon_P}-\varepsilon \end{aligned}$$

For $\mathrm{t}\in\mathcal{U}_\varepsilon$ and $(\mathrm{k},\mathrm{s})\in C(\mathrm{t})\cap A(\mathrm{t})$, a similar computation shows that

$$|\log P(\mathrm{k},\mathrm{s}\mid\mathrm{t}) - \log P'(\mathrm{k},\mathrm{s}\mid\mathrm{t})| \leq -\log(1-2\sqrt{\varepsilon_P}) \leq 4\sqrt{\varepsilon}$$

□

4. Proof of the main result

We prove the following two lemmas.

Lemma 19. *For each integer n and strategy ξ:*

$$\gamma_n(\xi) \leq \operatorname{cav} u(0)$$

Lemma 20. *For each $\varepsilon > 0$, there is an integer n and strategy ξ such that:*

$$\gamma_n(\xi) \geq \operatorname{cav} u(0) - \varepsilon$$

Proof of lemma 19. The argument used here is similar to that of Neyman and Okada ([NO99],[NO00]). Let ξ be a strategy and let the sequences $\mathbf{s}_n, \mathbf{t}_n$ of random signals associated to it. Define then the entropy variation at stage m as the r.v.

$$\Delta_m = H(\mathbf{s}_1,\ldots,\mathbf{s}_m\mid\mathbf{t}_1,\ldots,\mathbf{t}_m) - H(\mathbf{s}_1,\ldots,\mathbf{s}_{m-1}\mid\mathbf{t}_1,\ldots,\mathbf{t}_{m-1})$$

From the definition of u, at each stage m we have $g(P_\xi(\cdot|\mathbf{t}_1,\ldots,\mathbf{t}_{m-1})) \leq u(\Delta_m)$ a.s. Thus,

$$\gamma_n(\xi) \leq \mathbf{E}_\xi\left[\frac{1}{n}\sum\nolimits_{m=1}^n u(\Delta_m)\right] \leq \mathbf{E}_\xi\left[\operatorname{cav} u(\frac{1}{n}\sum\nolimits_{m=1}^n \Delta_m)\right]$$

Now $\sum_{m=1}^n \Delta_m = H(\mathbf{s}_1,\ldots,\mathbf{s}_n \mid \mathbf{t}_1,\ldots,\mathbf{t}_n) \geq 0$. Since $\operatorname{cav} u$ is decreasing, the result follows. □

Lemma 20 will follow from the following:

Lemma 21. *Let c, c' in $\mathcal{C}(A)$ and $\lambda\in[0,1]$ such that $\lambda\Delta H(c) + (1-\lambda)\Delta H(c') \geq 0$. For each $\bar{\varepsilon} > 0$, there is an integer $\bar{n}$ and strategy ξ such that:*

$$|\gamma_{\bar{n}}(\xi) - \lambda\pi(c) - (1-\lambda)\pi(c')| \leq \bar{\varepsilon}$$

Proof. **First approximations.**

Since the problem has imperfect information structure, fix d_0 such that $\Delta H(c_{d_0}) > 0$ (section 2.4.3). Denote $c = (\mu, d^{\mathcal{K}})$ and $c' = (\mu', d^{\mathcal{K}'})$ with $\mu \in \Delta(\mathcal{K})$, $\mu' \in \Delta(\mathcal{K}')$. We assume w.l.o.g. $\mu \in \mathbb{T}_{n_0}(\mathcal{K}), \mu' \in \mathbb{T}_{n_0}(\mathcal{K}')$ for some common n_0, $\lambda = \frac{m_1}{m_1+n_1}$ with m_1, n_1 multiples of n_0 and that d_0 is a.c. w.r.t. both c and c', since the convex combination λc, $(1-\lambda)c'$ can be approximated by convex combinations that satisfy these conditions and yield arbitrarily close payoffs and entropies. Hence $\delta_0 = \max\{d(d_0\|c), d(d_0\|c')\}$ is well defined and finite.

Idea of the construction.

The strategy ξ starts by an initialization phase during which d_0 is played for some N_0 stages.

Let (M, N) denote a multiple of (m, n). After the initialization phase, provided N_0 is large enough, an application of proposition 15 allows to construct ξ from stages $N_0 + 1$ to $N_0 + M$ in such a way that conditional to the information of the o.o. , the distribution of outcomes of the d.m. at stages N_0+1 to N_0+M is close to the one ν_c induced by c. We then apply the same proposition from stage N_0+M+1 to N_0+M+N, approximating the distribution of outcomes $\nu_{c'}$ induced by c' during these stages.

By some $2L$ applications of proposition 15, we construct ξ that approximates N_0 times d_0 followed by L cycles of M times c followed by N times c'. The total length will then be $\bar{n} = N_0 + L(M + N)$.

The distribution induced P by our constructed strategy ξ over $S^{\bar{n}}$ approximates Q, defined as the direct product of N_0 times the distribution of outcomes induced by d_0, followed by $(\nu_c^{\otimes M} \otimes \nu_{c'}^{\otimes N})^{\otimes L}$.

In the remaining of the proof, we complete the definition of ξ, define N_0, M, N in such a way that $d(P\|Q)$ is negligible compared to $\bar{n}$, and conclude that ξ yields a payoff close to $\lambda\pi(c) + (1-\lambda)\pi(c')$.

Definition of the strategy.

The strategy ξ is constructed on consecutive blocks of stages $B_0, B_1, \ldots, B_{2L}$, where L is an integer parameter. B_0 has length N_0, and B_k has length M for odd k, and N for $k > 0$ even. We define inductively the strategy over successive blocks, as well as sequences of parameters $(h_k, \eta_k, \varepsilon_k)$. Let $h_0 = \Delta H(c_{d_0}), \eta_0, \varepsilon_0$ be initialization values.

- Block B_0. Play d_0 at each stage of B_0. Set $h_1 = \frac{N_0}{M} h_0, \eta_1 = \frac{N_0}{M}\eta_0, \varepsilon_1 = \varepsilon_0$. Let τ^0 be the history of signals observed by the o.o. at B_0, σ^0 be the history of outcomes over B_0 and P_{τ^0} be the distribution of histories of outcomes conditional on τ^0. Declare B^0 *successful* if P_{τ^0} verifies an $\mathbf{AEP}(M, h_1, \eta_1, \varepsilon_1)$ and $\left|-\frac{1}{M}\log P_{\tau^0}(\sigma^0) - h_1\right| \leq \eta_1$ and declare it *failed* otherwise.
- Block B_k, $k > 0$. Let τ^{k-1} be the history of signals observed by the o.o. at blocks $B_0, \ldots, B_{k-1}$, σ^{k-1} be the history of outcomes over blocks $B_0, \ldots, B_{k-1}$ and $P_{\tau^{k-1}}$ be the distribution of histories of outcomes conditional on τ^{k-1}.
  - If B_{k-1} was declared failed, declare B_k failed and play d_0 at each stage of B_k.
  - Otherwise
    - If k is odd and $P_{\tau^{k-1}}$ verifies an $\mathbf{AEP}(M, h_k, \eta_k, \varepsilon_k)$ with $h_k \geq H(\mu) + \eta_k + 2\frac{\varepsilon_k}{M}$, apply proposition 15 to the data $(\mathcal{K}, \mu, \sigma = d^{\mathcal{K}}, P_{\tau^{k-1}})$ and define ξ as the strategy given in this proposition. Set:
      $h_{k+1} = (h_k + \Delta H(c))\frac{\lambda}{1-\lambda}$,
      $\eta_{k+1} = (a_{\varepsilon_0}(\eta_k + \frac{\log(M+1)}{M}) + 4\frac{\sqrt{\varepsilon_k}}{M})\frac{\lambda}{1-\lambda}$,
      $\varepsilon_{k+1} = \varepsilon_0 + 3\sqrt{\varepsilon_k}$.

Let τ^k be the history of signals observed by the o.o. at blocks $B_0, \ldots, B_k$, σ^k be the history of outcomes over blocks $B_0, \ldots, B_k$ and P_{τ^k} be the distribution of histories of outcomes conditional on τ^k. Declare B^k successful if P_{τ^k} verifies an $\mathbf{AEP}(N, h_{k+1}, \eta_{k+1}, \varepsilon_{k+1})$ and $\left|-\frac{1}{N}\log P_{\tau^k}(\sigma^k) - h_{k+1}\right| \leq \eta_{k+1}$ and declare it failed otherwise.

∗ If k is even and $P_{\tau^{k-1}}$ verifies an $\mathbf{AEP}(N, h_k, \eta_k, \varepsilon_k)$ with $h_k \geq H(\mu') + \eta_k + 2\frac{\varepsilon_k}{N}$, apply proposition 15 to the data ($\mathcal{K}'$, μ', $\sigma' = d'^{\mathcal{K}'}$, P_t^{k-1}) and define ξ as the strategy given in this proposition. Set then:
$h_{k+1} = (h_k + \Delta H(c'))\frac{1-\lambda}{\lambda}$,
$\eta_{k+1} = (a_{\varepsilon_0}(\eta_k + \frac{\log(N+1)}{N}) + 4\frac{\sqrt{\varepsilon_k}}{N})\frac{1-\lambda}{\lambda}$,
$\varepsilon_{k+1} = \varepsilon_0 + 3\sqrt{\varepsilon_k}$.
Let τ^k be the history of signals observed by the o.o. at blocks $B_0, \ldots, B_k$, σ^k be the history of outcomes over blocks $B_0, \ldots, B_k$ and P_{τ^k} be the distribution of histories of outcomes conditional on τ^k. Declare B^k successful if P_{τ^k} verifies an $\mathbf{AEP}(M, h_{k+1}, \eta_{k+1}, \varepsilon_{k+1})$ and $\left|-\frac{1}{M}\log P_{\tau^k}(\sigma^k) - h_{k+1}\right| \leq \eta_{k+1}$ and declare it failed otherwise.

The definition of ξ is now complete for given values of $(N_0, M, N, \eta_0, \varepsilon_0, L)$. In the sequel we prove that for an adequate choice of these parameters, the strategy has the desired properties.

Claim 22. There exists n_0 such that for $N_0 \geq n_0$,

$$P(\{\tau^0, P_{\tau^0} \text{ verifies an } \mathbf{AEP}(N_0, h_0, \eta_0, \varepsilon_0)\}) \geq 1 - \varepsilon_0$$

Proof. For each N_0 define:

$$C_{N_0} = \left\{\sigma^0, \left|-\frac{1}{n}\log P(\sigma^0 \mid \tau^0(\sigma^0)) - h_0\right| \leq \eta_0\right\}$$

where σ^0 is the sequence of outcomes at block B_0 and $\tau^0(\sigma^0)$ is the associated sequence of signals. Applying the Asymptotic Equipartition Property (see [CT91], thm 3.1.1, p. 51), there is n_0 such that for $N_0 \geq n_0$, $P(C_{N_0}) \geq 1 - \varepsilon_0^2$. Cut then C_{N_0} according to the values of τ^0: $C_{N_0}(\tau^0) = \{\sigma^0, \left|-\frac{1}{n}\log P(\sigma^0 \mid \tau^0) - h_0\right| \leq \eta_0\}$ and set:

$$B = \{\tau^0, P_{\tau^0}(C_{N_0}(\tau^0)) \geq 1 - \varepsilon_0\}$$

Then $P(C_{N_0}) = \sum_{\tau^0} P(\tau^0)P_{\tau^0}(C_{N_0}(\tau^0)) \leq P(B) + (1-\varepsilon_0)(1-P(B))$ and therefore $P(B) \geq 1 - \varepsilon_0$.

□

Claim 23. (1) $\forall k = 0, \ldots, 2L$, $\varepsilon_k \leq \varepsilon_{\max} = \theta(\varepsilon_0)^{2^{-2L}}$ with $\theta = 1 + 3\sqrt{\theta}$.

(2) If (M, N) are such that: $\max\left(\frac{\log(M+1)}{M} + \frac{4\sqrt{\varepsilon_{\max}}}{a_{\varepsilon_0}M}, \frac{\log(N+1)}{N} + \frac{4\sqrt{\varepsilon_{\max}}}{a_{\varepsilon_0}N}\right) \leq \eta_0$
Setting, $r_{\varepsilon_0} = a_{\varepsilon_0}\max(\frac{\lambda}{1-\lambda}, \frac{1-\lambda}{\lambda}) > 1$, we get:

$$\forall k = 0, \ldots, 2L,\ \eta_k \leq \eta_{\max} = \frac{r_{\varepsilon_0}^{2L}(2r_{\varepsilon_0} - 1) - r_{\varepsilon_0}}{r_{\varepsilon_0} - 1}\eta_0$$

(3) $\forall k = 1, \ldots, 2L$, $h_k \geq h_1$ for k odd and $h_k \geq h_2$ for k even.

Proof. Goes easily by induction, using for (3) that $M\Delta H(c) + N\Delta H(c') \geq 0$. □

Claim 24. There exists (N_0, M, N) such that for each $k = 0, \dots, 2L$,

$$\begin{cases} h_k \geq H(\mu) + \eta_k + \frac{2\varepsilon_k}{M} & \text{for } k \text{ odd} \\ h_k \geq H(\mu') + \eta_k + \frac{2\varepsilon_k}{N} & \text{for } k \text{ even} \end{cases}$$

Proof. From claim 23 and by definition of h_1 and h_2, it is enough to choose (N_0, M, N) such that:

$$N_0 h_0 \geq MH(\mu) + M\eta_{\max} + 2\varepsilon_{\max}$$

and

$$N_0 h_0 \geq NH(\mu') + N\eta_{\max} - N\frac{\lambda}{1-\lambda}\Delta H(c) + 2\varepsilon_{\max}$$

□

For $k = 1, \dots, 2L$, let $P^{k,\tau^{k-1}}$ be the distribution of sequences of outcomes at block B^k conditional on τ^{k-1}, the history of signals prior to block B_k. Let also Q^k be the distribution of sequences of outcomes at block B^k under Q. Set:

$$D(P\|Q) = \sum\nolimits_{k=1}^{2L} E_P\left[d(P^{k,\tau^{k-1}}\|Q^k)\right]$$

Claim 25. $\forall \bar{\delta} > 0$, there is a choice of the parameters $(N_0, M, N, \eta_0, \varepsilon_0)$ such that $D(P\|Q) \leq \bar{n}\bar{\delta}$.

Proof. For $k = 0, \dots, 2L$, let E_k be the event $\{$ B_k is successful $\}$ and let $p_k = P(E_k \mid E_0, \dots, E_{k-1})$. Given that $B_0, \dots, B_{k-1}$ have been successful, the conclusion of proposition 15 applied with $\varepsilon_P = \varepsilon_k$ and $\varepsilon = \varepsilon_0$ gives

$p_k \geq (1 - \varepsilon_0 - 2\sqrt{\varepsilon_k})(1 - \varepsilon_0 - 3\sqrt{\varepsilon_k}) \geq (1 - \varepsilon_{k+1})^2$

and thus $p_k \geq 1 - 2\varepsilon_{\max}$, and $P(E_k) \geq (1 - 2\varepsilon_{\max})^k$. Set $D_k = E_P\left[d(P^{k,\tau^{k-1}}\|Q^k)\right]$. By proposition 15, we get for k odd:

$$D_k \leq 2M(\eta_{\max} + \varepsilon_{\max}\log(|\mathcal{K}|)) + |\mathcal{K}|\log(M+1) + 1 + (1 - (1 - 2\varepsilon_{\max})^{2L})M\delta_0$$

and for k even,

$$D_k \leq 2N(\eta_{\max} + \varepsilon_{\max}\log(|\mathcal{K}'|)) + |\mathcal{K}'|\log(N+1) + 1 + (1 - (1 - 2\varepsilon_{\max})^{2L})N\delta_0$$

Given the values of (η_0, ε_0), choose (M, N) such that $\max(\frac{\log(M+1)}{M}, \frac{\log(N+1)}{N}) \leq \varepsilon_{\max}$. Setting $e = \max(2\log(|\mathcal{K}|) + |\mathcal{K}| + 1, 2\log(|\mathcal{K}'|) + |\mathcal{K}'| + 1)$ and summing over k we get:

$$D(P\|Q) \leq \bar{n}(2\eta_{\max} + e\varepsilon_{\max} + (1 - (1 - 2\varepsilon_{\max})^{2L})\delta_0)$$

For fixed L, choose η_0, ε_0 small enough to get $D(P\|Q) \leq \bar{n}\bar{\delta}$.

□

We are now in position to complete the proof of lemma 21.

Estimation of payoffs.

Q is an independent product of distributions on S. Let q_m be the distribution of the outcome at stage m under Q. The absolute difference of expected payoffs under P and under Q at stage m is:

$$\delta_m = \left|\mathbf{E}_P\left[g(P(\cdot|\mathbf{t}^{m-1}))\right] - g(q_m)\right|$$

Since g is uniformly continuous on $\Delta(S)$, for every $\bar{\varepsilon} > 0$, there is $\bar{\alpha} > 0$ such that:

$$\left\|P(\cdot|\mathbf{t}^{m-1}) - q_m\right\|_1 \leq \bar{\alpha} \Longrightarrow \left|g(P(\cdot|\mathbf{t}^{m-1})) - g(q_m)\right| \leq \frac{\bar{\varepsilon}}{2}$$

Then,

$$\delta_m \leq \frac{\bar{\varepsilon}}{2} P(\left\|P(\cdot|\mathbf{t}^{m-1}) - q_m)\right\|_1 \leq \bar{\alpha}) + 2\left\|g\right\| P(\left\|P(\cdot|\mathbf{t}^{m-1}) - q_m)\right\|_1 > \bar{\alpha})$$

where $\|g\|$ denotes $\max\{|g(d)|, d \in D\}$. Then using Markov's inequality:

$$\delta_m \leq \frac{\bar{\varepsilon}}{2} + \frac{2\left\|g\right\|}{\bar{\alpha}} \mathbf{E}_P\left[\left\|P(\cdot|\mathbf{t}^{m-1}) - q_m\right\|_1\right] \tag{7}$$

Let us denote by $\bar{\pi}_k(P^k_{\tau^{k-1}})$ (resp. $\bar{\pi}_k(Q^k)$) the average expected payoff on block k under $P^k_{\tau^{k-1}}$ (resp. Q^k) and by n_k the length of B_k. By averaging equation 7 on block k, we get:

$$\left|\bar{\pi}_k(P^k_{\mathrm{t}}) - \bar{\pi}_k(Q^k)\right| \leq \frac{\bar{\varepsilon}}{2} + \frac{2\left\|g\right\|}{\bar{\alpha}} d^{n_k}_{\mathcal{S}}(P^k_{\tau^{k-1}}||Q^k)$$

and from proposition 11,

$$\left|\bar{\pi}_k(P^k_{\mathrm{t}}) - \bar{\pi}_k(Q^k)\right| \leq \frac{\bar{\varepsilon}}{2} + \frac{2\left\|g\right\|}{\bar{\alpha}} f(\frac{1}{n_k} d(P^k_{\mathrm{t}}||Q^k)) \tag{8}$$

By averaging equation 8 on the set of blocks and taking expectation, and since P and Q coincide on B_0, we get:

$$\left|\gamma_{\bar{n}}(\xi) - \frac{1}{\bar{n}} \sum\nolimits_{m=1}^{\bar{n}} g(q_m)\right| \leq \frac{\bar{\varepsilon}}{2} + \frac{2\left\|g\right\|}{\bar{\alpha}} \sum\nolimits_k \frac{n_k}{\bar{n}} \mathbf{E}_P f(\frac{1}{n_k} d(P^k_{\mathrm{t}}||Q^k))$$

and applying Jensen to the concave function f yields:

$$\left|\gamma_{\bar{n}}(\xi) - \frac{1}{\bar{n}} \sum\nolimits_{m=1}^{\bar{n}} g(q_m)\right| \leq \frac{\bar{\varepsilon}}{2} + \frac{2\left\|g\right\|}{\bar{\alpha}} f(\frac{1}{\bar{n}} D(P\|Q))$$

By definition of Q we have:

$$\frac{1}{\bar{n}} \sum\nolimits_{m=1}^{\bar{n}} g(q_m) = \frac{N_0}{\bar{n}} g(d_0) + \frac{\bar{n} - N_0}{\bar{n}} (\lambda\pi(c) + (1-\lambda)\pi(c'))$$

Thus,

$$\left|\frac{1}{\bar{n}} \sum\nolimits_{m=1}^{\bar{n}} g(q_m) - (\lambda\pi(c) + (1-\lambda)\pi(c'))\right| \leq 2\left\|g\right\| \frac{N_0}{\bar{n}}$$

and,

$$|\gamma_{\bar{n}}(\xi) - (\lambda\pi(c) + (1-\lambda)\pi(c'))| \leq \frac{\bar{\varepsilon}}{2} + \frac{2\left\|g\right\|}{\bar{\alpha}} f(\frac{1}{\bar{n}} d(P\|Q)) + 2\left\|g\right\| \frac{N_0}{\bar{n}}$$

Fix now $\bar{\varepsilon}$ and choose the parameters of the strategy so as to ensure:

(1) $D(P\|Q) \leq \bar{n}\bar{\delta}$ with $\bar{\delta}$ such that $\frac{2\|g\|}{\bar{\alpha}} f(\bar{\delta}) \leq \frac{\bar{\varepsilon}}{4}$
(2) $2\left\|g\right\| \frac{N_0}{\bar{n}} \leq \frac{\bar{\varepsilon}}{4}$

For this very last point observe that $\frac{N_0}{\bar{n}} = \frac{1}{1+L\frac{M+N}{N_0}}$ so L must be chosen large enough. The proof is now complete.

□

5. Applications

5.1. **Minmax levels.** We now deduce theorem 1 from theorem 7. Take back the model of repeated games with imperfect monitoring developed in the introduction.

Proof of theorem 1. We first prove that v^I_∞, $\lim v^I_n$ and $\lim v^I_\lambda$ are no less than $\min\{x_2|(x_1,x_2)\in \operatorname{co} V, x_1\geq 0\}$ by establishing the equivalence between strategies of the d.m. in one instance of the optimization problem and a class of strategies for the team in the repeated game with imperfect monitoring. Call strategies of the team *oblivious* if they do not depend on player I's past moves. Against oblivious team strategies, and in any version of the game, I's best response is to play a stage-by-stage best response to the distribution of actions the team at that stage. This allows us to reduce the choice of best oblivious strategies of the team to the optimization setup given in section 2 by letting:

- $S=\Pi_{i\neq I}A_i\times B$.
- D be the set of probabilities on $\Pi_{i\neq I}A_i\times B$ which can be written as $p\otimes\rho$ with $p\in\otimes_{i\neq I}\Delta(A_i)$.
- For any distribution in $\Delta(S)$,the payoff depends on its marginal q on $\Pi_i A_i$ and we let $g(q)$ be defined as $g(q)=-\max_{a^I\in A^I} r(a^I,q)$, where we still denote by r the linear extension of r to probabilities on $\Pi_i A_i$.
- The partition T of S is such that if $((a^i)_{i\neq I},b)$ is selected, only b is observed by the o.o.

There is one-to-one correspondence between strategies of the team and strategies of the d.m. in the optimization problem, and the expected best reply payoff of player I at any stage t against oblivious strategies equals the d.m. payoff at stage t from the corresponding strategies. Hence, the team problem restricted to oblivious strategies and the d.m. problem are equivalent. Note also that the assumption that player I does not observe $(a_i)_{i\neq I}$ implies that the optimization problem has imperfect information structure (any tuple of mixed strategies with full support of the team in the one-shot game generates positive entropy). By theorem 7, the team can guarantee $\min\{x_2|(x_1,x_2)\in \operatorname{co} V, x_2\geq 0\}$.

To complete the proof of theorem 1, it remains to show that the team cannot guarantee less than $\min\{x_2|(x_1,x_2)\in \operatorname{co} V, x_2\geq 0\}$ using any possible (unrestricted) strategies. The point is that the argument of lemma 19 is still valid. Let σ^{-I} be a strategy profile for the team. Let σ^I be a strategy of player I that plays at each stage a best response to the expected distribution of actions of the team. Let h_t be the history at stage t and h^I_t be the history for player I. The moves of the team players at stage $t+1$ are independent conditional on h_t which, in the eye of player I is distributed according to $P_\sigma(\cdot|h^I_t)$. So for fixed h^I_t, this defines the correlation system that the team uses at stage $t+1$. Player I plays a stage-by-stage best response to $P_\sigma(\cdot|h^I_t)$. Let Δ_{t+1} be the entropy variation induced at stage $t+1$. The payoff at stage $t+1$ is thus $-g(P_\sigma(\cdot|h^I_t))$ and by definition of the function u, $-g(P_\sigma(\cdot|h^I_t))\geq -u(\Delta_{t+1})$. Averaging over stages, taking expectation and applying Jensen's inequality, we get the result as in lemma 19. □

5.2. **Biased coin problems.** A biased coin problem is a zero-sum repeated game where player 1 is restricted to choose pure actions but meanwhile observes exogenous random variables and can condition his actions on these observations. This model was introduced by Gossner and Vieille [GV02] in the case of i.i.d. exogenous random variables. We present a more general version of this problem now.

A biased coin problem is given by:

- A zero-sum game $G=(A^1,A^2,r)$, where A^1 and A^2 are finite sets of actions and $r:A^1\times A^2\to\mathbb{R}$ is a payoff function.
- A mapping Q from $A^1\times A^2$ to the set of distributions on a finite set U.

The game is played as follows. At each stage $t=1,2,\dots$, the game G is played. If (a^1,a^2) is chosen player 1 observes privately a random variable with distribution

$Q(\cdot|a^1, a^2)$. The actions are publicly monitored.

Player 1's is restricted to play in pure strategies. It is easy to find G such that this game has no value (take matching pennies) unless the random variables allow player 1 to mix his actions as if he were unrestricted. The issue is thus to find how much player 1 can guarantee, i.e. the maxmin of the repeated game. Gossner and Vieille [GV02] study i.i.d. random variables, i.e. $Q(\cdot|a^1, a^2)$ does not depend on (a^1, a^2). We treat now the case where the law of those random variables is controlled by player 1 only, i.e. we assume from now on: $Q(\cdot|a^1, a^2) = Q(\cdot|a^1)$.

Let $q \in \Delta(A^1)$ be a mixed action for player 1. Denote $h_{a^1} = H(Q(\cdot|a^1))$ for $a^1 \in A^1$ and $\Delta H(q) = \sum_{a^1} q(a^1)h_{a^1} - H(q)$, for $q \in \Delta(A^1)$. For $h \in \mathbb{R}$, set $u(h) = \max_{\Delta H(q) \geq h} \min_{a^2} r(q, a^2)$.

Theorem 26. *If* $\min_{a^1} h_{a^1} > 0$*, the minmax of the infinitely repeated game is* $\operatorname{cav} u(0)$.

Proof. We apply our main theorem to the following data.

- $S = A^1 \times U$.
- $A = \{\delta_{a^1} \otimes Q(\cdot|a^1) \mid a^1 \in A^1\}$
- For each distribution on S with marginal $q \in \Delta(A^1)$, the payoff depends on q only and is set as $g(q) = \min_{a^2} r(q, a^2)$.
- The partition T is such that the o.o. is informed of a^1 and not of u.

Assume now that player 1 is restricted to strategies that forget player 2's moves. Such a strategy of player 1 is exactly a strategy for the o.o. in the optimization problem. Now player 2 cannot control the behavior of player 1 and therefore can play stage-by-stage best replies, hence the definition of the payoff function g. A decision system for the d.m. in this set up can be identified with a mixed action and the entropy variation is just the one given above. This proves that player 1 can guarantee $\operatorname{cav} u(0)$.

On the other hand, take a strategy of player 1 and assume that player 2 plays stage-by-stage best replies. Consider again the entropy variation at stage $t+1$ to get that the stage payoff is less than $u(\Delta_{t+1})$. The conclusion follows as in lemma 19. □

References

[AE76] J.-P. Aubin and I. Ekeland. Estimates of the duality gap in non-convex optimization problems. *Mathematics of Operations Research*, 1:225–245, 1976.

[AS94] R. J. Aumann and L. S. Shapley. Long-term competition—A game theoretic analysis. In N. Megiddo, editor, *Essays on game theory*, pages 1–15. Springer-Verlag, New-York, 1994.

[BK85] J.-P. Benoît and V. Krishna. Finitely repeated games. *Econometrica*, 53(4):905–922, 1985.

[CT91] T. M. Cover and J. A. Thomas. *Elements of information theory*. Wiley Series in Telecomunications. Wiley, 1991.

[FLM94] D. Fudenberg, D. K. Levine, and E. Maskin. The folk theorem with imperfect public information. *Econometrica*, 62:997–1039, 1994.

[FM86] D. Fudenberg and E. Maskin. The folk theorem in repeated games with discounting or with incomplette information. *Econometrica*, 54:533–554, 1986.

[Gos94] O. Gossner. The folk theorem for finitely repeated games with mixed strategies. *International Journal of Game Theory*, 24:95–107, 1994.

[GV02] O. Gossner and N. Vieille. How to play with a biased coin? *Games and Economic Behaviour*, 41:206–226, 2002.

[Leh89] E. Lehrer. Nash equilibria of n player repeated games with semi-standard information. *International Journal of Game Theory*, 19:191–217, 1989.

[Leh91] E. Lehrer. Internal correlation in repeated games. *International Journal of Game Theory*, 19:431–456, 1991.

[LS92] E. Lehrer and S. Sorin. A uniform tauberian theorem in dynamic programming. *Mathematics of Operations Research*, 17:303–307, 1992.

[MSZ94] J.-F. Mertens, S. Sorin, and S Zamir. Repeated games. CORE discussion paper 9420-9422, 1994.

[NO99] A. Neyman and D. Okada. Strategic entropy and complexity in repeated games. *Games and Economic Behavior*, 29:191–223, 1999.

[NO00] A. Neyman and D. Okada. Repeated games with bounded entropy. *Games and Economic Behavior*, 30:228–247, 2000.

[RT98] J. Renault and T. Tomala. Coalition-proof correlated equilibrium: a definition. *International Journal of Game Theory*, 27:539–559, 1998.

[RT00] J. Renault and T. Tomala. Communication equilibrium payoffs of repeated games with imperfect monitoring. Cahier du Ceremade, N0034, 2000.

[Rub77] A. Rubinstein. Equilibrium in supergames. Center for Research in Mathematical Economics and Game Theory, Research Memorandum 25, 1977.

[Sha48] C. Shannon. A mathematical theory of communication. *Bell System Technical Journal*, 27:379–423 ; 623–656, 1948.

[vSK97] B. von Stengel and D. Koller. Team max min equilibria. *Games and Economics Behavior*, 21:309–321, 1997.

THEMA, UMR CNRS 7536, Université Paris 10 – Nanterre
E-mail address: Olivier.Gossner@u-paris10.fr

CEREMADE, UMR CNRS 7534 Université Paris 9 – Dauphine
E-mail address: tomala@ceremade.dauphine.fr