MATHEMATICS OF OPERATIONS RESEARCH
Vol. 31, No. 1, February 2006, pp. 13–30
ISSN 0364-765X | EISSN 1526-5471 | 06 | 3101 | 0013
DOI 10.1287/moor.1050.0174

Empirical Distributions of Beliefs Under Imperfect Observation

Olivier Gossner
Paris-Jourdan Sciences Economiques, UMR CNRS-EHESS-ENS-ENPC 8545, 48 Boulevard Jourdan, 75014 Paris, France, and MEDS, Kellogg School of Management, Northwestern University, Evanston, Illinois 60208-2009, USA,
gossner@ens.fr, http://www.enpc.fr/ceras/gossner/

Tristan Tomala
CEREMADE, UMR CNRS 7534 Université Paris-Dauphine, Place de Lattre de Tassigny, 75016 Paris, France
tomala@ceremade.dauphine.fr, http://www.ceremade.dauphine.fr/~tomala/

Let $(\mathbf{x}_n)_n$ be a process with values in a finite set X and law P, and let $\mathbf{y}_n = f(\mathbf{x}_n)$ be a function of the process. At stage n, the conditional distribution $\mathbf{p}_n = P(\mathbf{x}_n \mid \mathbf{x}_1, \ldots, \mathbf{x}_{n-1})$, element of $\Pi = \Delta(X)$, is the belief that a perfect observer, who observes the process online, holds on its realization at stage n. A statistician observing the signals $\mathbf{y}_1, \ldots, \mathbf{y}_n$ holds a belief $\mathbf{e}_n = P(\mathbf{p}_n \mid \mathbf{x}_1, \ldots, \mathbf{x}_n) \in \Delta(\Pi)$ on the possible predictions of the perfect observer. Given X and f, we characterize the set of limits of expected empirical distributions of the process $(\mathbf{e}_n)$ when P ranges over all possible laws of $(\mathbf{x}_n)_n$.

Key words: stochastic process; signals; entropy; repeated games
MSC2000 subject classification: Primary: 60G35, 94A17; secondary: 91A20
OR/MS subject classification: Primary: probability, entropy; secondary: games
History: Received February 28, 2003; revised January 26, 2004, and June 14, 2005.

1. Introduction. We study the gap in predictions made by agents that observe different signals about some process $(\mathbf{x}_n)_n$ with values in a finite set X and law P. Assume that a perfect observer observes $(\mathbf{x}_n)_n$, and a statistician observes a function $\mathbf{y}_n = f(\mathbf{x}_n)$. At stage n, $\mathbf{p}_n = P(\mathbf{x}_n \mid \mathbf{x}_1, \ldots, \mathbf{x}_{n-1})$, element of $\Pi = \Delta(X)$ the set of probabilities on X, is the prediction that a perfect observer of the process makes on its next realization. To a sequence of signals $\mathbf{y}_1, \ldots, \mathbf{y}_n$ corresponds a belief $\mathbf{e}_n = P(\mathbf{p}_n \mid \mathbf{y}_1, \ldots, \mathbf{y}_{n-1})$ that the statistician holds on the possible predictions of the perfect observer. The information gap about the future realization of the process at stage n between the perfect observer and the statistician is seen in the fact that the perfect observer knows $\mathbf{p}_n$, whereas the statistician knows only the law $\mathbf{e}_n$ of $\mathbf{p}_n$ conditional to $\mathbf{y}_1, \ldots, \mathbf{y}_{n-1}$.

We study the possible limits of expected empirical distributions of the process $(\mathbf{e}_n)$ when P ranges over all possible laws of $(\mathbf{x}_n)_n$.

We call experiments the elements of $E = \Delta(\Pi)$ and experiment distributions the elements of $\Delta(E)$. We say that an experiment distribution δ is achievable if there is a law P of the process for which δ is the limiting expected empirical distributions of $(\mathbf{e}_n)$. To an experiment e, we associate a random variable $\mathbf{p}$ with values in Π and with law e. Let $\mathbf{x}$ be a random variable with values in X such that, conditional on the realization p of $\mathbf{p}$, $\mathbf{x}$ has law p. Let then $\mathbf{y} = f(\mathbf{x})$. We define the entropy variation associated to e as

$$\Delta H(e) = H(\mathbf{p}, \mathbf{x} | \mathbf{y}) - H(\mathbf{p}) = H(\mathbf{x} | \mathbf{p}) - H(\mathbf{y}).$$

This mapping measures the evolution of the uncertainty for the statistician on the predictions of the perfect observer.

Our main result is that an experiment distribution δ is achievable if and only if $\mathbf{E}_\delta(\Delta H) \geq 0$.

This result has applications both to statistical problems and to game-theoretic ones.

Given a process $(\mathbf{x}_n)$ with law P, consider a repeated decision problem, where at each stage an agent has to take a decision and gets a stage payoff, depending on his action and the realization of the process at that stage. We compare the optimal payoff for an agent observing the process online and for an agent observing only the process of signals. At each stage, each agent maximizes his conditional expected payoff given his information. His expected payoff at stage n thus writes as a function of the beliefs he holds at stage $n-1$ on the next stage's realization of the process. Then, the expected payoff at stage n to each agent conditional to the past signals of the statistician—the agent with least information—is a function of $\mathbf{e}_n$. Both long-run expected payoffs are thus functions of the long-run expected empirical distribution of the process $(\mathbf{e}_n)$. Our result allows to derive characterizations of the maximal value of information in repeated decision problems measured as the maximal (under possible laws P of the process) difference of long-run average expected payoffs between the perfect observer and the statistician in a given decision problem.

Information asymmetries in repeated interactions is also a recurrent phenomenon in game theory, and arise in particular when agents observe private signals, or have limited information processing abilities.

In a repeated game with private signals, each player observes at each stage of the game a signal that depends on the action profile of all the players. While public equilibria of these games (see, e.g., Abreu et al. [1] and Fudenberg et al. [8]) or equilibria in which a communication mechanism serves to resolve information asymmetries (see, e.g., Compte [6], Kandori and Matsushima [17], and Renault and Tomala [27]) are well characterized, endogenous correlation and endogenous communication give rise to difficult questions that have only been tackled for particular classes of signalling structures (see Lehrer [20], Renault and Tomala [26], and Gossner and Vieille [15]).

Consider a repeated game in which a team of players $1, \ldots, N-1$ with action sets $A_1, \ldots, A_{N-1}$ tries to minimize the payoff of player N. Let $X = A_1 \times \cdots \times A_{N-1}$. Assume that each player of the team perfectly observes the actions played, whereas player N only observes a signal on the team's actions given by a map f defined on X. A strategy σ for the team that doesn't depend on player N's actions induces a process over X with law P_σ, and the maximal payoff at stage n to player N given his history of signals is a function of the experiment $\mathbf{e}_n$, i.e., of player N's beliefs on the distribution of joint actions of other players at stage n. Hence the average maximal payoff to player N against such a strategy for the team is a function of the induced experiment distribution. Note, however, that the team is restricted in the choice of P_σ, since the actions of all the players must be independent conditional on the past play. This paper also provides a characterization of achievable experiment distributions when the transitions of the process P are restricted to belong to a closed set of probabilities C. This characterization can be used to characterize the minimax values in classes of repeated games with imperfect monitoring (see Gossner and Tomala [14]). Gossner et al. [13] elaborate techniques for the computation of explicit solutions and fully analyse an example of game with imperfect monitoring. Another example is studied by Goldberg [9].

Information asymmetries also arise in repeated games when agents have different information processing abilities: some players may be able to predict more accurately future actions than others. These phenomena have been studied in the frameworks of finite automata (see Ben Porath [3], Neyman [22], [23], Gossner and Hernández [12], Bavly and Neyman [2], and Lacôte and Thurin [18]), bounded recall (see Lehrer [19], [21], Piccione and Rubinstein [25], Bavly and Neyman [2], and Lacôte and Thurin [18]), and time-constrained Turing machines (see Gossner [10], [11]). We hope the characterizations derived in this paper may provide a useful tool for the study of repeated games with boundedly rational agents.

The next section presents the model and the main results, while the remainder of this paper is devoted to the proof of our theorem.

2. Definitions and main results.

2.1. Notations.

For a finite set S, $|S|$ denotes its cardinality.

For a compact set S, $\Delta(S)$ denotes the set of Borel regular probability measures on S and is endowed with the weak-$*$ topology (thus $\Delta(S)$ is compact).

If $(\mathbf{x}, \mathbf{y})$ is a pair of finite random variables—i.e., with finite range—defined on a probability space $(\Omega, \mathcal{F}, P)$, $P(\mathbf{x}|y)$ denotes the conditional distribution of $\mathbf{x}$ given $\{\mathbf{y} = y\}$ and $P(\mathbf{x}|\mathbf{y})$ is the random variable with value $P(\mathbf{x}|y)$ if $\mathbf{y} = y$.

Given a set S and x in S, the Dirac measure on x is denoted ϵ_x: this is the probability measure with support $\{x\}$.

If $\mathbf{x}$ is a random variable with values in a compact subset of a topological vector space V, $\mathbf{E}(\mathbf{x})$ denotes the barycenter of $\mathbf{x}$ and is the element of V such that for each continuous linear form φ, $\mathbf{E}(\varphi(\mathbf{x})) = \varphi(\mathbf{E}(\mathbf{x}))$.

If p and q are probability measures on two probability spaces, $p \otimes q$ denotes the product probability.

2.2. Definitions.

2.2.1. Processes and distributions.

Let $(\mathbf{x}_n)_n$ be a process with values in a finite set X such that $|X| \geq 2$ and let P be its law. A statistician observes the value of $\mathbf{y}_n = f(\mathbf{x}_n)$ at each stage n, where $f\colon X \to Y$ is a fixed mapping. Before stage n, the history of the process is $x_1, \ldots, x_{n-1}$ and the the history available to the statistician is $y_1, \ldots, y_{n-1}$. The conditional law of $\mathbf{x}_n$ given the history of the process is

$$\mathbf{p}_n(x_1, \ldots, x_{n-1}) = P(\mathbf{x}_n | x_1, \ldots, x_{n-1}).$$

This defines a $(\mathbf{x}_1, \ldots, \mathbf{x}_{n-1})$-measurable random variable $\mathbf{p}_n$ with values in $\Pi = \Delta(X)$. The statistician holds a belief on the value of $\mathbf{p}_n$. For each history $y_1, \ldots, y_{n-1}$, we let $\mathbf{e}_n(y_1, \ldots, y_{n-1})$ be the conditional law of $\mathbf{p}_n$ given $y_1, \ldots, y_{n-1}$:

$$\mathbf{e}_n(y_1, \ldots, y_{n-1}) = P(\mathbf{p}_n | y_1, \ldots, y_{n-1}),$$

i.e., for each $\pi \in \Pi$: $\mathbf{e}_n(y_1, \ldots, y_{n-1})(\pi) = P(\{\mathbf{p}_n = \pi | y_1, \ldots, y_n\})$. This defines a $(\mathbf{y}_1, \ldots, \mathbf{y}_{n-1})$-measurable random variable $\mathbf{e}_n$ with values in $E = \Delta(\Pi)$. Following Blackwell [4], [5], we call *experiments* the elements of E.

The empirical distribution of experiments up to stage n is

$$\mathbf{d}_n(y_1, \ldots, y_{n-1}) = \frac{1}{n} \sum_{m \le n} \epsilon_{\mathbf{e}_m(y_1, \ldots, y_{m-1})}.$$

So for each $e \in E$, $\mathbf{d}_n(y_1, \ldots, y_{n-1})(e)$ is the average number of times $1 \le m \le n$ such that $e_m(y_1, \ldots, y_{m-1}) = e$. The $(\mathbf{y}_1, \ldots, \mathbf{y}_{n-1})$-measurable random variable $\mathbf{d}_n$ has values in $D = \Delta(E)$. We call D the set of *experiment distributions*.

Definition 2.1. We say that the law P of the process *n-achieves* the experiment distribution δ if $\mathbf{E}_P(\mathbf{d}_n) = \delta$, and that δ is *n-achievable* if there exists P that n-achieves δ. D_n denotes the set of n-achievable experiment distributions.

We say that the law P of the process *achieves* the experiment distribution δ if $\lim_{n \to +\infty} \mathbf{E}_P(\mathbf{d}_n) = \delta$, and that δ is *achievable* if there exists P that achieves δ. D_∞ denotes the set of achievable experiment distributions.

Achievable distributions have the following properties:

Proposition 2.1. (i) *For* $n, m \ge 1$, $\frac{n}{n+m} D_n + \frac{m}{n+m} D_m \subset D_{m+n}$.
(ii) $D_n \subset D_\infty$.
(iii) D_∞ *is the closure of* $\bigcup_n D_n$.
(iv) D_∞ *is convex and closed.*

Proof. To prove (i) and (ii), let P_n and P'_m be the laws of processes $\mathbf{x}_1, \ldots, \mathbf{x}_n$ and $\mathbf{x}'_1, \ldots, \mathbf{x}'_m$ such that P_n n-achieves $\delta_n \in D_n$ and P'_m m-achieves $\delta'_m \in D_m$. Then, any process of law $P_n \otimes P'_m$ $(n+m)$-achieves $\frac{n}{n+m}\delta_n + \frac{m}{n+m}\delta'_m \in D_{m+n}$, and any process of law $P_n \otimes P_n \otimes P_n \otimes \ldots$ achieves $\delta_n \in D_\infty$. Point (iii) is a direct consequence of the definitions and of (ii). Point (iv) follows from (i) and (iii). □

Example 2.1. Assume f is constant, let $(\mathbf{x}_n)_n$ be the the process on $\{0, 1\}$ such that $(\mathbf{x}_{2n-1})_{n \le 1}$ are i.i.d. uniformly distributed and $\mathbf{x}_{2n} = \mathbf{x}_{2n-1}$. At odd stages $\mathbf{e}_{2n-1} = \epsilon_{(\frac{1}{2}, \frac{1}{2})}$ a.s. and at even stages $\mathbf{e}_{2n} = \frac{1}{2}\epsilon_{(1,0)} + \frac{1}{2}\epsilon_{(0,1)}$ a.s. Hence the law of $(\mathbf{x}_n)_n$ achieves the experiment distribution $\frac{1}{2}\epsilon_{e_1} + \frac{1}{2}\epsilon_{e_2}$.

Example 2.2. Assume again f constant, a parameter p is drawn uniformly in $[0, 1]$, and $(\mathbf{x}_n)_n$ is a family of i.i.d. Bernoulli random variables with parameter p. In this case, $p_n \to p$ a.s., and therefore e_n weak-$*$ converges to the uniform distribution on $[0, 1]$. The experiment distribution achieved by the law of this process is thus the Dirac measure on the uniform distribution on $[0, 1]$.

2.2.2. Measures of uncertainty. Let $\mathbf{x}$ be a finite random variable with values in X and law P. Throughout this paper, log denotes the logarithm with base 2. By definition, the entropy of $\mathbf{x}$ is

$$H(\mathbf{x}) = -\mathbf{E} \log P(\mathbf{x}) = -\sum_x P(x) \log P(x),$$

where $0 \log 0 = 0$ by convention. Note that $H(\mathbf{x})$ is nonnegative and depends only on the law P of $\mathbf{x}$ and we shall also denote it $H(P)$.

Let $(\mathbf{x}, \mathbf{y})$ be a couple of finite random variables with joint law P. The conditional entropy of $\mathbf{x}$ given $\{\mathbf{y} = y\}$ is the entropy of the conditional distribution $P(\mathbf{x}|y)$:

$$H(\mathbf{x}|y) = -\mathbf{E}[\log P(\mathbf{x}|y)].$$

The conditional entropy of $\mathbf{x}$ given $\mathbf{y}$ is the expected value of the previous

$$H(\mathbf{x}|\mathbf{y}) = \sum_y H(\mathbf{x}|y) P(y).$$

One has the following additivity formula:

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

Given an experiment e, let $\mathbf{p}$ be a random variable in Π with distribution e, $\mathbf{x}$ be a random variable in X such that the conditional distribution of $\mathbf{x}$ given $\{\mathbf{p} = p\}$ is equal to p and $\mathbf{y} = f(\mathbf{x})$. Note that since $\mathbf{x}$ is finite and since the conditional distribution of $\mathbf{x}$ given $\{\mathbf{p} = p\}$ is well defined, we can extend the definition of the conditional entropy by letting $H(\mathbf{x}|y) = \int H(p)\, de(p)$.

Definition 2.2. The *entropy variation* associated to e is

$$\Delta H(e) = H(\mathbf{x}|\mathbf{p}) - H(\mathbf{y}).$$

Remark 2.1. Assume that e has finite support (hence the associated random variable $\mathbf{p}$ also has finite support). From the additivity formula

$$H(\mathbf{p}, \mathbf{x}) = H(\mathbf{p}) + H(\mathbf{x}|\mathbf{p}) = H(\mathbf{y}) + H(\mathbf{p}, \mathbf{x}|\mathbf{y}).$$

Therefore $\Delta H(e) = H(\mathbf{p}, \mathbf{x}|\mathbf{y}) - H(\mathbf{p})$.

The mapping ΔH measures the evolution of the uncertainty of the statistician at a given stage. Fix a history of signals $y_1, \ldots, y_{n-1}$, consider the experiment $e = \mathbf{e}_n(y_1, \ldots, y_{n-1})$, and let $\mathbf{p} = \mathbf{p}_n$: e is the conditional law of $\mathbf{p}$ given the history of signals. Set also $\mathbf{x} = \mathbf{x}_n$ and $\mathbf{y} = \mathbf{y}_n$. The evolution of the process and of the information of the statistician at stage n is described by the following procedure:

- Draw $\mathbf{p}$ according to e;
- If $\mathbf{p} = p$, draw $\mathbf{x}$ according to p;
- Announce $\mathbf{y} = f(\mathbf{x})$ to the statistician.

The uncertainty—measured by entropy—for the statistician at the beginning of the procedure is $H(\mathbf{p})$. At the end of the procedure, the statistician knows the value of $\mathbf{y}$ and $\mathbf{p}, \mathbf{x}$ are unknown to him, the new uncertainty is thus $H(\mathbf{p}, \mathbf{x}|\mathbf{y})$. $\Delta H(e)$ is therefore the variation of entropy across this procedure. It also writes as the difference between the entropy added to $\mathbf{p}$ by the procedure: $H(\mathbf{x}|\mathbf{p})$, and the entropy of the information gained by the statistician: $H(\mathbf{y})$.

Lemma 2.1. *The mapping* $\Delta H\colon E \to \mathbb{R}$ *is continuous.*

Proof. $H(\mathbf{x}|\mathbf{p}) = \int H(p)\, de(p)$ is linear continuous in e, since H is continuous on Π. The mapping that associates to e the law of $\mathbf{y}$ is also linear continuous. □

2.3. Main results. We characterize achievable distributions.

Theorem 2.1. *An experiment distribution* δ *is achievable if and only if* $\mathbf{E}_\delta(\Delta H) \geq 0$.

We also prove a stronger version of the previous theorem in which the transitions of the process are restricted to belong to an arbitrary subset of Π.

Definition 2.3. The distribution $\delta \in D$ has support in $C \subset \Pi$ if for each e in the support of δ, the support of e is included in C.

Definition 2.4. Given $C \subset \Pi$ a process $(\mathbf{x}_n)_n$ with law P is a C-process if for each n, $P(\mathbf{x}_n|\mathbf{x}_1, \ldots, \mathbf{x}_{n-1}) \in C$, P-almost surely.

Remark 2.2. If P is the law of a C-process and P achieves δ, then δ has support in C. This observation follows readily from the previous definitions.

Theorem 2.2. *Let* C *be a closed subset of* Π. *The experiment distribution* δ *is achievable by the law of a* C*-process if and only if* δ *has support in* C *and* $\mathbf{E}_\delta(\Delta H) \geq 0$.

Remark 2.3. If C is closed, the set of experiment distributions that are achievable by laws of C-processes is convex and closed. The proof is identical as for D_∞, so we omit it.

2.4. Trivial observation. We say that the observation is trivial when f is constant.

Lemma 2.2. *If the observation is trivial, any* δ *is achievable.*

This fact can easily be deduced from Theorem 2.1. Since f is constant, $H(\mathbf{y}) = 0$ and thus $\Delta H(e) \geq 0$ for each $e \in E$. However, a simple construction provides a direct proof in this case.

Proof. By closedness and convexity, it is enough to prove that any $\delta = \epsilon_e$ with e of finite support is achievable. Let thus $e = \sum_k \lambda_k \epsilon_{p_k}$. Again by closedness, assume that the λ_ks are rational with common denominator 2^n for some n. Let $x \neq x'$ be two distinct points in X and $\mathbf{x}_1, \ldots, \mathbf{x}_n$ be i.i.d. with law $\frac{1}{2}\epsilon_x + \frac{1}{2}\epsilon_{x'}$, so that $(\mathbf{x}_1, \ldots, \mathbf{x}_n)$ is uniform on a set with 2^n elements. Map $(\mathbf{x}_1, \ldots, \mathbf{x}_n)$ to some random variable $\mathbf{k}$ such that $P(\mathbf{k} = k) = \lambda_k$. Construct then the law P of the process such that conditional on $\mathbf{k} = k$, $\mathbf{x}_{t+n}$ has law p_k for $t \geq 1$. P achieves δ.

2.5. Perfect observation. We say that information is perfect when f is one to one. Let E_d denote the set of Dirac experiments, i.e., measures on Π whose support are a singleton. This set is a weak-$*$ closed subset of E.

LEMMA 2.3. *If information is perfect, δ is achievable if and only if* $\operatorname{supp}\delta \subset E_d$.

We derive this result from Theorem 2.1.

PROOF. If $e \in E_d$, the random variable $\mathbf{p}$ associated to e is constant a.s., therefore $H(\mathbf{x}|\mathbf{p}) = H(\mathbf{x}) = H(\mathbf{y})$ since observation is perfect. Thus $\Delta H(e) = 0$, and $\mathbf{E}_\delta(\Delta H) = 0$ if $\operatorname{supp}\delta \subset E_d$. Conversely, assume $\mathbf{E}_\delta(\Delta H) \geq 0$. Since the observation is perfect, $H(\mathbf{y}) = H(\mathbf{x}) \geq H(\mathbf{x}|\mathbf{p})$, and thus $\Delta H(e) \leq 0$ for all e. So, $\Delta H(e) = 0$ δ-almost surely, i.e., $H(\mathbf{x}|\mathbf{p}) = H(\mathbf{x})$ for each e in a set of δ-probability one. For each such e, $\mathbf{x}$ and $\mathbf{p}$ are independent, i.e., the law of $\mathbf{x}$ given $\mathbf{p} = p$ does not depend on p. Hence e is a Dirac measure.

2.6. Example of a nonachievable experiment distribution.

EXAMPLE 2.3. Let $X = \{i, j, k\}$ and $f(i) = f(j) \neq f(k)$. Consider distributions of the type $\delta = \epsilon_e$.

If $e = \epsilon_{\frac{1}{2}\epsilon_j + \frac{1}{2}\epsilon_k}$, δ is achievable. Indeed, such δ is induced by an i.i.d. process with stage law $\frac{1}{2}\epsilon_j + \frac{1}{2}\epsilon_k$.

On the other hand, if $e = \frac{1}{2}\epsilon_{\epsilon_j} + \frac{1}{2}\epsilon_{\epsilon_k}$, under e the law of $\mathbf{x}$ conditional on $\mathbf{p}$ is a Dirac measure and thus $H(\mathbf{x}|\mathbf{p}) = 0$, whereas the law of $\mathbf{y}$ is the one of a fair coin and $H(\mathbf{y}) = 1$. Thus $\mathbf{E}_\delta(\Delta H) = \Delta H(e) < 0$ and from Theorem 2.1, δ is not achievable.

The intuition is as follows: if δ were achievable by P, only j and k would appear with positive density P-a.s. Since $f(j) \neq f(k)$, the statistician can reconstruct the history of the process given his signals, and therefore correctly guess $P(\mathbf{x}_n|x_1, \ldots, x_{n-1})$. This contradicts $e = \frac{1}{2}\epsilon_{\epsilon_j} + \frac{1}{2}\epsilon_{\epsilon_k}$, which means that at almost each stage, the statistician is uncertain about $P(\mathbf{x}_n|x_1, \ldots, x_{n-1})$ and attributes probability $\frac{1}{2}$ to ϵ_j and probability $\frac{1}{2}$ to ϵ_k.

3. Reduction of the problem. The core of our proof is to establish the next proposition.

PROPOSITION 3.1. *Let $\delta = \lambda\epsilon_e + (1-\lambda)\epsilon_{e'}$, where λ is rational, e, e' have finite support, and $\lambda\Delta H(e) + (1-\lambda)\Delta H(e') > 0$. Let $C = \operatorname{supp} e \cup \operatorname{supp} e'$. Then, δ is achievable by the law of a C-process.*

Sections 4–7 are devoted to the proof of this proposition. We now prove Theorem 2.2 from Proposition 3.1. Theorem 2.1 is a direct consequence of Theorem 2.2 with $C = \Pi$.

3.1. The condition $\mathbf{E}_\delta \Delta H \geq 0$ is necessary. We prove now that any achievable δ must verify $\mathbf{E}_\delta \Delta H \geq 0$.

PROOF. Let δ be achieved by P. Recall that $\mathbf{e}_n$ is a $(\mathbf{y}_1, \ldots, \mathbf{y}_{n-1})$-measurable random variable with values in E. $\Delta H(\mathbf{e}_n)$ is thus a $(\mathbf{y}_1, \ldots, \mathbf{y}_{n-1})$-measurable real-valued random variable and from the definitions

$$\Delta H(\mathbf{e}_m(y_1, \ldots, y_{m-1})) = H(\mathbf{p}_m, \mathbf{x}_m | y_1, \ldots, y_m) - H(\mathbf{p}_m | y_1, \ldots, y_{m-1})$$

Thus

$$\begin{aligned} \mathbf{E}_P \Delta H(\mathbf{e}_m) &= H(\mathbf{p}_m, \mathbf{x}_m | \mathbf{y}_1, \ldots, \mathbf{y}_m) - H(\mathbf{p}_m | \mathbf{y}_1, \ldots, \mathbf{y}_{m-1}) \\ &= H(\mathbf{x}_m | \mathbf{p}_m, \mathbf{y}_1, \ldots, \mathbf{y}_{m-1}) - H(\mathbf{y}_m | \mathbf{y}_1, \ldots, \mathbf{y}_{m-1}). \end{aligned}$$

Setting for each m, $H_m = H(\mathbf{x}_1, \ldots, \mathbf{x}_m | y_1, \ldots, \mathbf{y}_m)$, we wish to prove that $\mathbf{E}_P \Delta H(\mathbf{e}_m) = H_m - H_{m-1}$. To do this, we apply the additivity formula to the quantity

$$\bar{H} := H(\mathbf{x}_1, \ldots, \mathbf{x}_m, \mathbf{y}_m, \mathbf{p}_m | \mathbf{y}_1, \ldots, \mathbf{y}_{m-1})$$

in two different ways. First,

$$\begin{aligned} \bar{H} &= H_{m-1} + H(\mathbf{x}_m, \mathbf{y}_m, \mathbf{p}_m | \mathbf{x}_1, \ldots, \mathbf{x}_{m-1}, \mathbf{y}_1, \ldots, \mathbf{y}_{m-1}) \\ &= H_{m-1} + H(\mathbf{x}_m | \mathbf{p}_m) \end{aligned}$$

where the second equality holds since $\mathbf{y}_m$ is a deterministic function of $\mathbf{x}_m$, $\mathbf{p}_m$ is $\mathbf{x}_1, \ldots, \mathbf{x}_{m-1}$ measurable and the law of $\mathbf{x}_m$ depends on $\mathbf{p}_m$ only. Secondly,

$$\begin{aligned} \bar{H} &= H(\mathbf{y}_m | \mathbf{y}_1, \ldots, \mathbf{y}_{m-1}) + H(\mathbf{x}_1, \ldots, \mathbf{x}_m, \mathbf{p}_m | \mathbf{y}_1, \ldots, \mathbf{y}_m) \\ &= H(\mathbf{y}_m | \mathbf{y}_1, \ldots, \mathbf{y}_{m-1}) + H_m, \end{aligned}$$

where the second equality holds since again, $\mathbf{p}_m$ is $\mathbf{x}_1, \ldots, \mathbf{x}_{m-1}$ measurable. It follows:

$$\mathbf{E}_P \Delta H(\mathbf{e}_m) = H_m - H_{m-1},$$

and thus

$$\sum_{m\le n} \mathbf{E}_P \Delta H(\mathbf{e}_m) = H(\mathbf{x}_1,\dots,\mathbf{x}_n \mid \mathbf{y}_1,\dots,\mathbf{y}_n) \ge 0.$$

From the definitions

$$\mathbf{E}_\delta(\Delta H) = \lim_n \frac{1}{n} \sum_{m\le n} \mathbf{E}_P \Delta H(\mathbf{e}_n),$$

which gives the result. □

3.2. C-perfect observation. To prove that $\mathbf{E}_\delta \Delta H \ge 0$ is a sufficient condition for δ to be achievable, we first need to study the case of perfect observation in details.

Definition 3.1. Let C be a closed subset of Π. The mapping f is *C-perfect* if for each p in C, f is one to one on $\operatorname{supp} p$.

We let $E_{C,d} = \{\epsilon_p, p\in C\}$ be the set of Dirac experiments with support in C. $E_{C,d}$ is a weak-$*$ closed subset of E and $\{\delta\in D, \operatorname{supp}\delta\subset E_{C,d}\}$ is a weak-$*$ closed and convex subset of D.

Lemma 3.1. *If f is C-perfect, then the following 3 assertions are equivalent*:
(i) *The experiment distribution δ is achievable by the law of a C-process.*
(ii) $\operatorname{supp}\delta\subset E_{C,d}$.
(iii) $\mathbf{E}_\delta(\Delta H) = 0$.

Proof. Point (1) ⇔ Point (2). Let $(\mathbf{x}_n)$ be a C-process P, δ achieved by P, and p_1 be the law of $\mathbf{x}_1$. Since f is one to one on $\operatorname{supp} p_1$, the experiment $\mathbf{e}_2(y_1)$ is the Dirac measure on $p_2 = P(\mathbf{x}_2 \mid x_1)$. By induction, assume that the experiment $\mathbf{e}_n(y_1,\dots,y_{n-1})$ is the Dirac measure on $p_n = P(\mathbf{x}_n \mid x_1,\dots,x_{n-1})$. Since f is one to one on $\operatorname{supp} p_n$, y_n reveals the value of x_n and $\mathbf{e}_{n+1}(y_1,\dots,y_n)$ is the Dirac measure on $P(\mathbf{x}_n \mid x_1,\dots,x_n)$. We get that under P, at each stage the experiment belongs to $E_{C,d}$ P-a.s., and thus $\operatorname{supp}\delta\subset E_{C,d}$.

Conversely, let δ be such that $\operatorname{supp}\delta\subset E_{C,d}$. Since the set of achievable distribution is closed, it is sufficient to prove that for any $p_1,\dots,p_k$ in C, $n_1,\dots,n_k$ integers, $n=\sum_j n_j$, $\delta = \sum_j (n_j/n)\epsilon_{e_j}$ is feasible where $e_j = \epsilon_{p_j}$. But then, $P_n = p_1^{\otimes n_1} p_2^{\otimes n_2}\cdots p_k^{\otimes n_k}$ n-achieves δ.

Point (2) ⇔ Point (3). If $e\in E_{C,d}$, the random variable $\mathbf{p}$ associated to e is constant a.s., therefore $H(\mathbf{x}\mid\mathbf{p}) = H(\mathbf{x}) = H(\mathbf{y})$ since f is C-perfect. Thus $\Delta H(e) = 0$, and therefore $\mathbf{E}_\delta(\Delta H) = 0$ whenever $\operatorname{supp}\delta\subset E_{C,d}$.

Conversely, assume $\mathbf{E}_\delta(\Delta H) = 0$. Since f is C-perfect, for each e with support in C, $H(\mathbf{y}) = H(\mathbf{x}) \ge H(\mathbf{x}\mid\mathbf{p})$ implying $\Delta H(e)\le 0$. Thus $\Delta H(e) = 0$ δ-a.s., i.e., $H(\mathbf{x}\mid\mathbf{p}) = H(\mathbf{x})$ for each e in a set of δ-probability one. For each such e, $\mathbf{x}$ and $\mathbf{p}$ are independent, i.e., the law of $\mathbf{x}$ given $\mathbf{p}=p$ does not depend on p, hence e is a Dirac measure. Thus $\operatorname{supp}\delta\subset E_{C,d}$. □

3.3. The condition $\mathbf{E}_\delta \Delta H \ge 0$ is sufficient. According to Proposition 3.1, any $\delta = \lambda\epsilon_e + (1-\lambda)\epsilon_{e'}$ with λ rational, e, e' of finite support and such that $\lambda\Delta H(e) + (1-\lambda)\Delta H(e') > 0$ is achievable by the law of a C-process with $C = \operatorname{supp} e \cup \operatorname{supp} e'$. We apply this result to prove Theorem 2.2.

Proof. [Proof of Theorem 2.2 from Proposition 3.1]. Let $C\subset\Pi$ be closed, $E_C\subset E$ be the set of experiments with support in C, and $D_C\subset D$ be the set of experiment distributions with support in E_C. Take $\delta\in D_C$ such that $\mathbf{E}_\delta(\Delta H)\ge 0$.

Assume first that $\mathbf{E}_\delta(\Delta H) = 0$ and that there exists a weak-$*$ neighborhood V of δ in D_C such that for any $\mu\in V$, $\mathbf{E}_\mu(\Delta H)\le 0$. For $p\in C$, let $\nu = \epsilon_{\epsilon_p}$. There exists $0<t<1$ such that $(1-t)\delta + t\nu\in V$, and therefore $\mathbf{E}_\nu(\Delta H)\le 0$. Taking $\mathbf{x}$ of law p and $\mathbf{y} = f(\mathbf{x})$, $\mathbf{E}_\nu(\Delta H) = \Delta H(\epsilon_p) = H(\mathbf{x}) - H(\mathbf{y}) \le 0$. Since $H(\mathbf{x})\ge H(f(\mathbf{x}))$, we obtain $H(\mathbf{x}) = H(f(\mathbf{x}))$ for each $\mathbf{x}$ of law $p\in C$. This implies that f is C-perfect and the theorem holds by Lemma 3.1.

Otherwise, there is a sequence δ_n in D_C weak-$*$ converging to δ such that $\mathbf{E}_{\delta_n}(\Delta H) > 0$. Since the set of achievable distributions is closed, we assume $\mathbf{E}_\delta(\Delta H) > 0$ from now on. The set of distributions with finite support being dense in D_C (see, e.g., Parthasaraty [24, Theorem 6.3, p. 44]), again by closedness we assume

$$\delta = \sum_j \lambda_j \epsilon_{e_j}$$

with $e_j\in E_C$ for each j. Let S be the finite set of distributions $\{\epsilon_{e_j}; j\}$. We claim that δ can be written as a convex combination of distributions δ_k such that

- For each k, $\mathbf{E}_\delta(\Delta H) = \mathbf{E}_{\delta_k}(\Delta H)$.
- For each k, δ_k is the convex combination of two points in S.

This follows from the following lemma of convex analysis.

Lemma 3.2. *Let S be a finite set in a vector space and f be a real-valued affine mapping on $\operatorname{co} S$ the convex hull of S. For each $x \in \operatorname{co} S$, there exists an integer K, nonnegative numbers $\lambda_1, \ldots, \lambda_K$ summing to one, coefficients $t_1, \ldots, t_K$ in $[0,1]$, and points (x_k, x'_k) in S such that*
- $x = \sum_k \lambda_k (t_k x_k + (1-t_k)x'_k)$.
- *For each k,* $t_k f(x_k) + (1-t_k) f(x'_k) = f(x)$.

Proof. Let $a = f(x)$. The set $S_a = \{y \in \operatorname{co} S, f(y) = a\}$ is the intersection of a polytope with a hyperplane. It is thus convex and compact so by Krein-Milman's theorem (see, e.g., Rockafellar [28]), it is the convex hull of its extreme points. An extreme point y of S_a—i.e., a face of dimension 0 of S_a—must lie on a face of $\operatorname{co} S$ of dimension at most 1, and therefore is a convex combination of two points of S. □

We apply Lemma 3.2 to $S = \{\epsilon_{e_j}; j\}$ and to the affine mapping $\delta \mapsto \mathbf{E}_\delta(\Delta H)$. Since the set of achievable distributions is convex, it is enough to prove that for each k, δ_k is achievable. The problem is thus reduced to $\delta = \lambda \epsilon_e + (1-\lambda)\epsilon_{e'}$ such that $\lambda \Delta H(e) + (1-\lambda)\Delta H(e') > 0$. We approximate λ by a rational number and since C is closed, we may assume that the supports of e and e' are finite subsets of C. Proposition 3.1 now applies. □

4. Presentation of the proof of Proposition 3.1. Consider an experiment distribution of the form

$$\delta = \frac{N}{N+M}\epsilon_e + \frac{M}{N+M}\epsilon_{e'},$$

where $e, e' \in E$ have finite support, N, M are integers such that $N\Delta H(e) + M\Delta H(e') > 0$. Under δ, e and e' appear with respective frequencies $\frac{N}{N+M}$ and $\frac{M}{N+M}$. We present the idea of the construction of a process that achieves δ.

Fix some history of signals $(y_1, \ldots, y_n)$ and denote $\mathbf{u}_n = (\mathbf{x}_1, \ldots, \mathbf{x}_n)$ the (random) past history of the process. Conditional to $(y_1, \ldots, y_n)$, $\mathbf{u}_n$ has then law $P(\mathbf{u}_n | y_1, \ldots, y_n)$.

A first step is to prove that when $H(\mathbf{u}_n)$ is "large enough," and if the distribution of $\mathbf{u}_n$ is close to a uniform distribution—we say that $\mathbf{u}_n$ satisfies an asymptotic equipartition property (AEP)—one can map or *code*, $\mathbf{u}_n$ into another random variable $\mathbf{v}_n$ with values in Π^n whose law is close to $e^{\otimes N}$ (i.e., e i.i.d. N times). This allows to define the process at stages $n+1, \ldots, n+N$ as follows: given $\mathbf{v}_n = (\mathbf{p}_{n+1}, \ldots, \mathbf{p}_{n+N})$, define $(\mathbf{x}_{n+1}, \ldots, \mathbf{x}_{n+N})$ such that for each t, $n+1 \le t \le n+N$, given $\{\mathbf{p}_t = p\}$, $\mathbf{x}_t$ has conditional law p and is independent of all other random variables. Defined in this way, the process is such that for each stage between $n+1$ and $n+N$, the belief induced at that stage is close to e.

Consider now the history of signals $(y_1, \ldots, y_n, y_{n+1}, \ldots, y_{n+N})$ up to time $n+N$, and set $\mathbf{u}_{n+N} = (\mathbf{x}_1, \ldots, \mathbf{x}_{n+N})$ the (random) past history of the process with conditional law $P(\mathbf{u}_{n+N} | y_1, \ldots, y_n, y_{n+1}, \ldots, y_{n+N})$. We show that, for a set of large probability of sequences of signals, $H(\mathbf{u}_{n+N})$ is close to $H(\mathbf{u}_n) + N\Delta H(e)$ and $\mathbf{u}_{n+N}$ also satisfies an AEP. As before, if $H(\mathbf{u}_{n+N})$ is "large enough," one can code $\mathbf{u}_{n+N}$ into some random variable $\mathbf{v}_{n+N}$ whose law is close to $e'^{\otimes M}$. This allows to define as above the process during stages $n+N+1$ to $n+N+M$ such that the induced beliefs at those stages are close to e'.

Let $\mathbf{u}_{n+N+M}$ represent the random past history of the process given the signals past signals at stage $n+N+M$. Then, for a set of sequences of large probability, $H(\mathbf{u}_{n+N+M})$ is close to $H(\mathbf{u}_n) + N\Delta H(e) + M\Delta H(e) \ge H(\mathbf{u}_n)$, since $N\Delta H(e) + M\Delta H(e) > 0$ and $\mathbf{u}_{n+N+M}$ satisfies an AEP. The procedure can in this case be iterated.

The construction of the process begins by an initialization phase, which allows to get a "large" $H(\mathbf{u}_n)$.

Section 5 presents the construction of the process for one block of stages and establishes bounds on closeness of probabilities. In §6, we iterate the construction, and show the full construction of the process P. We terminate the proof by proving the weak-$*$ convergence of the experiment distribution to $\lambda e + (1-\lambda)e'$ in §7. In this last part, we first control the Kullback distance between the law of the process of experiments under P and an ideal law $Q = e^{\otimes n} \otimes e'^{\otimes m} \otimes e^{\otimes n} \otimes e'^{\otimes m} \otimes \ldots$, and finally relate the Kullback distance to weak-$*$ convergence.

5. The one block construction.

5.1. Kullback and absolute Kullback distance. For two probability measures with finite support P and Q, we write $P \ll Q$ when P is absolutely continuous with respect to Q, i.e., $(Q(x) = 0 \Rightarrow P(x) = 0)$.

Definition 5.1. Let K be a finite set and P, Q in $\Delta(K)$ such that $P \ll Q$, the Kullback distance between P and Q is

$$d(P\|Q) = \mathbf{E}_P\left[\log \frac{P(\cdot)}{Q(\cdot)}\right] = \sum_k P(k) \log \frac{P(k)}{Q(k)}.$$

We recall the absolute Kullback distance and its comparison with the Kullback distance from Gossner and Vieille [16] for later use.

DEFINITION 5.2. Let K be a finite set and P, Q in $\Delta(K)$ such that $P \ll Q$, the absolute Kullback distance between P and Q is

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

LEMMA 5.1. *For every P, Q in $\Delta(K)$ such that $P \ll Q$,*

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

See the proof of (Gossner and Vieille [16, Lemma 17, p. 223]).

5.2. Equipartition properties. We say that a probability P with finite support satisfies an equipartition property (**EP**) when all points in the support of P have close probabilities.

DEFINITION 5.3. Let $P \in \Delta(K)$, $n \in \mathbb{N}$, $h \in \mathbb{R}_+, \eta > 0$. P satisfies an **EP**(n, h, η) when

$$P\left\{k \in K, \left| -\frac{1}{n} \log P(k) - h \right| \leq \eta \right\} = 1.$$

We say that a probability P with finite support satisfies an **AEP** when all points in a set of large P-measure have close probabilities.

DEFINITION 5.4. Let $P \in \Delta(K)$, $n \in \mathbb{N}$, $h \in \mathbb{R}_+$, $\eta, \xi > 0$. P satisfies an **AEP**(n, h, η, ξ) when

$$P\left\{k \in K, \left| -\frac{1}{n} \log P(k) - h \right| \leq \eta \right\} \geq 1 - \xi.$$

REMARK 5.1. If P satisfies an **AEP**(n, h, η, ξ) and m is a positive integer, then P satisfies an **AEP**$(m, \frac{n}{m}h, \frac{n}{m}\eta, \xi)$.

5.3. Types. Given a set K and in integer n, we denote $\tilde{k} = (k_1, \ldots, k_n) \in K^n$ a finite sequence in K. The type of $\tilde{k}$ is the empirical distribution $\rho_{\tilde{k}}$ induced by $\tilde{k}$; that is, $\rho_{\tilde{k}} \in \Delta(K)$ and $\forall k$, $\rho_{\tilde{k}}(k) = \frac{1}{n}|\{i = 1, \ldots, n, k_i = k\}|$. The type set $T_n(\rho)$ of $\rho \in \Delta(K)$ is the subset of K^n of sequences of type ρ. Finally, the set of types is $\mathbb{T}_n(K) = \{\rho \in \Delta(K), T_n(\rho) \neq \emptyset\}$. The following estimates the size of $T_n(\rho)$ for $\rho \in \mathbb{T}_n(K)$ (see, e.g., Cover and Thomas [7, Theorem 12.1.3, p. 282]):

$$\frac{2^{nH(\rho)}}{(n+1)^{|\operatorname{supp}\rho|}} \leq |T_n(\rho)| \leq 2^{nH(\rho)}. \tag{1}$$

5.4. Distributions induced by experiments and by codifications. Let $e \in \Delta(\Pi)$ be an experiment with finite support and n be an integer.

NOTATION 5.1. *Let $\rho(e)$ be the probability on $\Pi \times X$ induced by the following procedure: First, draw $\mathbf{p}$ according to e, then draw $\mathbf{x}$ according to the realization of $\mathbf{p}$. Let $Q(n, e) = \rho(e)^{\otimes n}$.*

We approximate $Q(n, e)$ in a construction where $(\mathbf{p}_1, \ldots, \mathbf{p}_n)$ is measurable with respect to some random variable $\mathbf{l}$ of law $P_{\mathcal{L}}$ in an arbitrary set $\mathcal{L}$.

NOTATION 5.2. *Let $(\mathcal{L}, P_{\mathcal{L}})$ be a finite probability space and $\varphi\colon \mathcal{L} \to \Pi^n$. We denote by $P = P(n, \mathcal{L}, P_{\mathcal{L}}, \varphi)$ the probability on $\mathcal{L} \times (\Pi \times X)^n$ induced by the following procedure: Draw $\mathbf{l}$ according to $P_{\mathcal{L}}$, set $(\mathbf{p}_1, \ldots, \mathbf{p}_n) = \varphi(\mathbf{l})$, then draw $\mathbf{x}_t$ according to the realization of $\mathbf{p}_t$.*

We let $\tilde{P} = \tilde{P}(n, \mathcal{L}, P_{\mathcal{L}}, \varphi)$ be the marginal of $P(n, \mathcal{L}, P_{\mathcal{L}}, \varphi)$ on $(\Pi \times X)^n$.

To iterate such a construction, we relate properties of the "input" probability measure $P_{\mathcal{L}}$ with those of the "output" probability measure $P(\mathbf{l}, \mathbf{p}_1, \ldots, \mathbf{p}_n, \mathbf{x}_1, \ldots, \mathbf{x}_n | \mathbf{y}_1, \ldots, \mathbf{y}_n)$.

Propositions 5.1 and 5.2 exhibit conditions on $P_{\mathcal{L}}$ such that there exists φ for which $\tilde{P}(n, \mathcal{L}, P_{\mathcal{L}}, \varphi)$ is close to $Q(n, e)$, and with large probability under $P = P(n, \mathcal{L}, P_{\mathcal{L}}, \varphi)$, $P(\mathbf{l}, \mathbf{p}_1, \ldots, \mathbf{p}_n, \mathbf{x}_1, \ldots, \mathbf{x}_n | \mathbf{y}_1, \ldots, \mathbf{y}_n)$ satisfies an adequate **AEP**.

In Proposition 5.1, the condition on $P_{\mathcal{L}}$ is an **EP** property, thus a stronger input property than the output property, which is stated as an **AEP**. Proposition 5.2 assumes that $P_{\mathcal{L}}$ satisfies an **AEP** property only.

5.5. EP to AEP codification result. We now state and prove our coding proposition when the input probability measure $P_{\mathcal{L}}$ satisfies an **EP**.

PROPOSITION 5.1. *For each experiment e, there exists a constant $U(e)$ such that for every integer n with $e \in \mathbb{T}_n(\Pi)$ and for every finite probability space $(\mathcal{L}, P_{\mathcal{L}})$ that satisfies an* **EP**(n, h, η) *with $n(h - H(e) - \eta) \geq 1$, there exists a mapping $\varphi: \mathcal{L} \to \Pi^n$ such that letting $P = P(n, \mathcal{L}, P_{\mathcal{L}}, \varphi)$ and $\tilde{P} = \tilde{P}(n, \mathcal{L}, P_{\mathcal{L}}, \varphi)$:*
(i) $d(\tilde{P}||Q(n,e)) \leq 2n\eta + |\operatorname{supp} e| \log(n+1) + 1$
(ii) *For every $0 < \varepsilon < 1$, there exists a subset $\mathcal{Y}_\varepsilon$ of Y^n such that*
  (a) $P(\mathcal{Y}_\varepsilon) \geq 1 - \varepsilon$
  (b) *For $\tilde{y} \in \mathcal{Y}_\varepsilon$, $P(\cdot|\tilde{y})$ satisfies an* **AEP**$(n, h', \eta', \varepsilon)$
with $h' = h + \Delta H(e)$ and $\eta' = (U(e)/\varepsilon^2)(\eta + 1/\sqrt{n})$.

PROOF. [Proof of Proposition 5.1]. Set $\rho = \rho(e)$ and $\tilde{Q} = Q(n, e)$.
Construction of φ: Since $P_{\mathcal{L}}$ satisfies an **EP**(n, h, η),

$$2^{n(h-\eta)} \leq |\operatorname{supp} P_{\mathcal{L}}| \leq 2^{n(h+\eta)}.$$

From the previous and Equation (1), there exists $\varphi: \mathcal{L} \to T_n(e)$ such that for every $\tilde{p} \in T_n(e)$,

$$2^{n(h-\eta-H(e))} - 1 \leq |\varphi^{-1}(\tilde{p})| \leq (n+1)^{|\operatorname{supp} e|} 2^{n(h+\eta-H(e))} + 1. \tag{2}$$

Bound on $d(\tilde{P}||\tilde{Q})$: $\tilde{P}$ and $\tilde{Q}$ are probabilities over $(\Pi \times X)^n$, which are deduced from their marginals on Π^n by the same transition probabilities. It follows from the definition of the Kullback distance that the distance from $\tilde{P}$ to $\tilde{Q}$ equals the distance of their marginals on Π^n:

$$d(\tilde{P}||\tilde{Q}) = \sum_{\tilde{p} \in T_n(e)} \tilde{P}(\tilde{p}) \log \frac{\tilde{P}(\tilde{p})}{\tilde{Q}(\tilde{p})}.$$

Using Equation (2) and the **EP** for $P_{\mathcal{L}}$, we obtain that for $\tilde{p} \in T_n(e)$:

$$\tilde{P}(\tilde{p}) \leq (n+1)^{|\operatorname{supp} e|} 2^{n(2\eta - H(e))} + 2^{-n(h-\eta)}.$$

On the other hand, since for all $\tilde{p} \in T_n(e)$, $\tilde{Q}(\tilde{p}) = 2^{-nH(e)}$:

$$\frac{\tilde{P}(\tilde{p})}{\tilde{Q}(\tilde{p})} \leq (n+1)^{|\operatorname{supp} e|} 2^{2n\eta} + 2^{-n(h-\eta-H(e))}.$$

Part (i) of the proposition now follows since $H(e) \leq h - \eta$.
Estimation of $|d|(\tilde{P}(\cdot|\tilde{y})||\tilde{Q}(\cdot|\tilde{y}))$: For $\tilde{y} \in Y^n$ s.t. $\tilde{P}(\tilde{y}) > 0$, we let $\tilde{P}_{\tilde{y}}$ and $\tilde{Q}_{\tilde{y}}$ in $\Delta((\Pi \times X)^n)$ denote $\tilde{P}(\cdot|\tilde{y})$ and $\tilde{Q}(\cdot|\tilde{y})$, respectively. Direct computation yields

$$\sum_{\tilde{y}} \tilde{P}(\tilde{y}) d(\tilde{P}_{\tilde{y}}||\tilde{Q}_{\tilde{y}}) \leq d(\tilde{P}||\tilde{Q}).$$

Hence for $\alpha_1 > 0$:

$$P\{\tilde{y}, d(\tilde{P}_{\tilde{y}}||\tilde{Q}_{\tilde{y}}) \geq \alpha_1\} \leq \frac{2n\eta + |\operatorname{supp} e| \log(n+1) + 1}{\alpha_1}$$

and from Lemma 5.1,

$$P\{\tilde{y}, |d|(\tilde{P}_{\tilde{y}}||\tilde{Q}_{\tilde{y}}) \leq \alpha_1 + 2\} \geq 1 - \frac{2n\eta + |\operatorname{supp} e| \log(n+1) + 1}{\alpha_1}. \tag{3}$$

The statistics of $(\tilde{p}, \tilde{x})$ under $\tilde{P}$: We argue here that the type $\rho_{\tilde{p},\tilde{x}} \in \Delta(\Pi \times X)$ of $(\tilde{p}, \tilde{x}) \in (\Pi \times X)^n$ is close to ρ, with large P-probability. First, note that since φ takes its values in $T_n(e)$, the marginal of $\rho_{\tilde{p},\tilde{x}}$ on Π is e with P-probability one. For $(p, x) \in \Pi \times X$, the distribution under P of $n\rho_{\tilde{p},\tilde{x}}(p, x)$ is the one of a sum of $ne(p)$ independent Bernoulli variables with parameter $p(x)$. For $\alpha_2 > 0$, the Bienaymé-Chebyshev inequality gives

$$P(|\rho_{\tilde{p},\tilde{x}}(p,x) - \rho(p,x)| \geq \alpha_2) \leq \frac{\rho(p,x)}{n\alpha_2^2}.$$

Hence

$$P(\|\rho_{\tilde{p},\tilde{x}} - \rho\|_\infty \leq \alpha_2) \geq 1 - \frac{1}{n\alpha_2^2}. \tag{4}$$

The set of $\tilde{y} \in Y^n$ s.t. $\widetilde{Q}_{\tilde{y}}$ satisfies an AEP has large P-probability: For $(\tilde{p}, \tilde{x}, \tilde{y}) = (p_i, x_i, y_i)_i \in (\Pi \times X \times Y)^n$ s.t. $\forall i,\ f(x_i) = y_i$, we compute

$$\begin{aligned}
-\frac{1}{n}\log Q_{\tilde{y}}(\tilde{p},\tilde{x}) &= -\frac{1}{n}\left(\sum_i \log\rho(p_i,x_i) - \log\rho(y_i)\right)\\
&= -\sum_{(p,x)\in(\operatorname{supp} e)\times X} \rho_{\tilde{p},\tilde{x}}(p,x)\log\rho(p,x)\\
&\quad + \sum_{y\in Y}\rho_{\tilde{p},\tilde{x}}(y)\log\rho_{\tilde{p},\tilde{x}}(y)\\
&= -\sum_{(p,x)}\rho(p,x)\log\rho(p,x) + \sum_y \rho(y)\log\rho(y)\\
&\quad + \sum_{(p,x)}(\rho(p,x) - \rho_{\tilde{p},\tilde{x}}(p,x))\log\rho(p,x)\\
&\quad - \sum_y(\rho(y) - \rho_{\tilde{p},\tilde{x}}(y))\log\rho(y).
\end{aligned}$$

Since

$$-\sum_{(p,x)}\rho(p,x)\log\rho(p,x) = H(\rho)$$

and denoting $f(\rho)$ the image of ρ on Y:

$$\sum_y \rho(y)\log\rho(y) = -H(f(\rho))$$

letting $M_0 = -2|(\operatorname{supp} e)\times X|\log(\min_{p,x}\rho(p,x))$, this implies

$$\left| -\frac{1}{n}\log\widetilde{Q}_{\tilde{y}}(\tilde{p},\tilde{x}) - H(\rho) + H(f(\rho))\right| \le M_0\|\rho - \rho_{\tilde{p},\tilde{x}}\|_\infty. \tag{5}$$

Define

$$\begin{aligned}
A_{\alpha_2} &= \left\{(\tilde{p},\tilde{x},\tilde{y}),\ \left| -\frac{1}{n}\log\widetilde{Q}_{\tilde{y}}(\tilde{p},\tilde{x}) - H(\rho) + H(f(\rho))\right| \le M_0\alpha_2\right\}\\
A_{\alpha_2,\tilde{y}} &= A_{\alpha_2}\cap((\operatorname{supp} e)\times X\times\{\tilde{y}\}),\ \ \tilde{y}\in Y^n.
\end{aligned}$$

Equations (4) and (5) yield

$$\begin{aligned}
\sum_{\tilde{y}} P(\tilde{y})\widetilde{P}_{\tilde{y}}(A_{\alpha_2,\tilde{y}}) &= P(A_{\alpha_2})\\
&\ge 1 - \frac{1}{n\alpha_2^2}\\
\sum_{\tilde{y}} P(\tilde{y})(1 - \widetilde{P}_{\tilde{y}}(A_{\alpha_2,\tilde{y}})) &\le \frac{1}{n\alpha_2^2}.
\end{aligned}$$

Then, for $\beta > 0$,

$$P\{\tilde{y}, 1 - \widetilde{P}_{\tilde{y}}(A_{\alpha_2,\tilde{y}}) \ge \beta\} \le \frac{1}{n\alpha_2^2\beta}$$

and

$$P\{\tilde{y}, \widetilde{P}_{\tilde{y}}(A_{\alpha_2,\tilde{y}}) \ge 1 - \beta\} \ge 1 - \frac{1}{n\alpha_2^2\beta}. \tag{6}$$

Definition of $\mathscr{Y}_\varepsilon$ and verification of ii(a): Set

$$\begin{cases}
\alpha_1 = \dfrac{4n\eta + 2|\operatorname{supp} e|\log(n+1) + 2}{\varepsilon}\\[2ex]
\alpha_2 = \dfrac{2}{\varepsilon\sqrt{n}}\\[2ex]
\beta = \dfrac{\varepsilon}{2}
\end{cases}$$

and let

$$\begin{cases} \mathcal{Y}^1_\varepsilon = \{\tilde{y}, |d|(\tilde{P}_{\tilde{y}}||\tilde{Q}_{\tilde{y}}) \le \alpha_1 + 2\} \\ \mathcal{Y}^2_\varepsilon = \{\tilde{y}, \tilde{P}_{\tilde{y}}(A_{\alpha_2,\tilde{y}}) \ge 1-\beta\} \\ \mathcal{Y}_\varepsilon = \mathcal{Y}^1_\varepsilon \cap \mathcal{Y}^2_\varepsilon \end{cases}.$$

Equations (3) and (6) and the choice of α_1, α_2, and β imply

$$P(\mathcal{Y}_\varepsilon) \ge 1-\varepsilon.$$

Verification of ii(b): We first prove that $\tilde{P}_{\tilde{y}}$ satisfies an **AEP** for $\tilde{y} \in \mathcal{Y}_\varepsilon$. For such $\tilde{y}$, the definition of $\mathcal{Y}^1_\varepsilon$ and Markov inequality give

$$\tilde{P}_{\tilde{y}}\left\{|\log \tilde{P}_{\tilde{y}}(\cdot) - \log \tilde{Q}_{\tilde{y}}(\cdot)| \le (\alpha_1+2)\frac{2}{\varepsilon}\right\} \ge 1 - \frac{\varepsilon}{2}.$$

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

$$\tilde{P}_{\tilde{y}}\left\{\left|-\frac{1}{n}\log \tilde{Q}_{\tilde{y}}(\cdot) - H(\rho) + H(f(\rho))\right| \le M_0\alpha_2\right\} \ge 1 - \frac{\varepsilon}{2}.$$

The two above inequalities yield

$$\tilde{P}_{\tilde{y}}\left\{\left|-\frac{1}{n}\log \tilde{P}_{\tilde{y}}(\cdot) - H(\rho) + H(f(\rho))\right| \le \frac{2(\alpha_1+2)}{n\varepsilon} + M_0\alpha_2\right\} \ge 1 - \varepsilon. \tag{7}$$

Remark now that $P(l, \tilde{p}, \tilde{x}|\tilde{y}) = P_{\tilde{y}}(\tilde{p}, \tilde{x})P(l|\tilde{p})$. If $\varphi(l) \ne \tilde{p}$, $P(l|\tilde{p}) = 0$. Otherwise, $P(l|\tilde{p}) = P(l)/P(\tilde{p})$ and Equation (2) and the **EP** for $P_{\mathcal{L}}$ imply

$$\frac{2^{n(h-\eta)}}{2^{n(h+\eta)}((n+1)^{|\operatorname{supp} e|}2^{n(h-\eta-H(e))}+1)} \le \frac{P(l)}{P(\tilde{p})} \le \frac{2^{n(h+\eta)}}{2^{n(h-\eta)}(2^{n(h-\eta-H(e))}-1)}.$$

From this, we deduce using $n(h-\eta-H(e)) \ge 1$:

$$|\log P(l) - \log P(\tilde{p}) - n(H(e)-h)| \le 3n\eta + \log(n+1)|\operatorname{supp} e| + 1. \tag{8}$$

Let $P_{\tilde{y}}$ denote $P(\cdot|\tilde{y})$ over $\mathcal{L} \times (\Pi \times X)^n$. Setting

$$A := 3\eta + \frac{\log(n+1)|\operatorname{supp} e|}{n} + \frac{1}{n} + \frac{2(\alpha_1+2)}{n\varepsilon} + M_0\alpha_2$$

Equations (7) and (8) imply

$$P_{\tilde{y}}\left\{\left|-\frac{1}{n}\log P_{\tilde{y}}(\cdot) - (H(\rho) - H(f(\rho)) - H(e) + h)\right| \le A\right\} \ge 1-\varepsilon.$$

Using $\varepsilon < 1$, $\log(n+1) \le 2\sqrt{n}$ and $n \ge \sqrt{n}$ we deduce

$$A \le \frac{11\eta}{\varepsilon^2} + \frac{10|\operatorname{supp} e| + 9 + 2M_0}{\sqrt{n}\varepsilon^2}.$$

Since $\Delta H(e) = H(\rho) - H(e) - H(f(\rho))$, letting $U(e) = 19|\operatorname{supp} e| + 2M_0$, Equations (7), (8), and (9) yield

$$P_{\tilde{y}}\left\{\left|-\frac{1}{n}\log P_{\tilde{y}}(\cdot) - (h + \Delta H(e))\right| \le \frac{U(e)}{\varepsilon^2}\left(\eta + \frac{1}{\sqrt{n}}\right)\right\} \ge 1-\varepsilon,$$

which is the desired **AEP**. □

5.6. AEP to AEP codification result. Building on Proposition 5.1, we now can state and prove the version of our coding result in which the input is an **AEP**.

PROPOSITION 5.2. *For each experiment e, there exists a constant $U(e)$ such that for every integer n with $e \in \mathbb{T}_n(\Pi)$ and for every finite probability space $(\mathscr{L}, P_{\mathscr{L}})$ that satisfies an* $\mathbf{AEP}(n, h, \eta, \xi)$ *with $n(h - H(e) - \eta) \geq 2$ and $0 < \xi < 1/2^{12}$, there exists a mapping $\varphi\colon \mathscr{L} \to \Pi^n$ such that letting $P = P(n, \mathscr{L}, P_{\mathscr{L}}, \varphi)$ and $\tilde{P} = \tilde{P}(n, \mathscr{L}, P_{\mathscr{L}}, \varphi)$:*

(i) $d(\tilde{P} || Q(n, e)) \leq 2n(\eta + \xi \log|\operatorname{supp} e|) + |\operatorname{supp} e| \log(n+1) + 2$

(ii) *For every $0 < \varepsilon < 1$, there exists a subset $\mathscr{Y}_\varepsilon$ of Y^n such that*

 (a) $P(\mathscr{Y}_\varepsilon) \geq 1 - \varepsilon - 2\xi^{1/4}$

 (b) *For $\tilde{y} \in \mathscr{Y}_\varepsilon$, $P(\cdot|\tilde{y})$ satisfies an* $\mathbf{AEP}(n, h', \eta', \xi')$

with $h' = h + \Delta H(e)$, $\eta' = (U(e)/\varepsilon^2)(\eta + 1/\sqrt{n}) + (4/n)\xi^{1/12}$, and $\xi' = \varepsilon + 3\sqrt{\xi}$.

We first establish the following lemma.

LEMMA 5.2. *K is a finite set. Suppose that $P \in \Delta(K)$ satisfies an* $\mathbf{AEP}(n, h, \eta, \xi)$. *Let the typical set of P be*:

$$C = \left\{ k \in K, \left| -\frac{1}{n} \log P(k) - h \right| \leq \eta \right\}.$$

Let $P_C \in \Delta(K)$ be the conditional probability given C: $P_C(k) = P(k|C)$. Then, P_C satisfies an $\mathbf{EP}(n, h, \eta')$ *with $\eta' = \eta + 2(\xi/n)$ for $0 < \xi < \frac{1}{2}$.*

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

PROOF. [Proof of Proposition 5.2]. Set again $\rho = \rho(e)$ and $\tilde{Q} = Q(n, e)$. Let C be the typical set of $P_{\mathscr{L}}$. From Lemma 5.2, $P'_{\mathscr{L}} = P_{\mathscr{L}}(\cdot|C)$ satisfies an $\mathbf{EP}(n, h, \eta + 2(\xi/n))$. Since $n(h - H(e) - \eta) \geq 2$, $n(h - H(e) - \eta - 2(\xi/n)) \geq 1$. Applying Proposition 5.1 to e yields: a constant $U(e)$, a mapping $\varphi\colon C \to \Pi^n$, an induced probability P' on $\mathscr{L} \times (\Pi \times X)^n$, and subsets $(\mathscr{Y}'_\varepsilon)_\varepsilon$ of Y^n.

Choose $\bar{p} \in \arg\max e(p)$ and extend φ to $\mathscr{L}$ by setting it to $(\bar{p}, \ldots, \bar{p})$ outside C. With $P'' = P_{\mathscr{L}}(\cdot|C) \otimes (\epsilon_{\bar{p}} \otimes \bar{p})^{\otimes n}$, the probability induced by $P_{\mathscr{L}}$ and φ on $\mathscr{L} \times (\Pi \times X)^n$ is then $P = P_{\mathscr{L}}(C)P' + (1 - P_{\mathscr{L}}(C))P''$. Set $\tilde{P}$ as the marginal of P on $(\Pi \times X)^n$. To verify point (i), using that the Kullback distance in convex in both arguments, we write

$$\begin{aligned} d(\tilde{P} \| \tilde{Q}) &\leq P_{\mathscr{L}}(C) d(\tilde{P}' \| \tilde{Q}) + (1 - P_{\mathscr{L}}(C)) n d(\epsilon_{\bar{p}} \otimes \bar{p} \| \rho) \\ &\leq d(\tilde{P}' \| \tilde{Q}) + \xi n d(\epsilon_{\bar{p}} \| e) \\ &\leq 2n\left(\eta + \frac{2\xi}{n}\right) + |\operatorname{supp} e| \log(n+1) + 1 + \xi n \log(|\operatorname{supp} e|) \\ &\leq 2n(\eta + \xi \log|\operatorname{supp} e|) + |\operatorname{supp} e| \log(n+1) + 2. \end{aligned}$$

Let $\mathscr{Y} = \{\tilde{y}, P'(\tilde{y}) > \xi^{1/4} P''(\tilde{y})\}$ and $\mathscr{Y}_\varepsilon = \mathscr{Y}'_\varepsilon \cap \mathscr{Y}$. Then, $P'(\mathscr{Y}) \geq 1 - \xi^{1/4}$ and $P'(\mathscr{Y}'_\varepsilon) \geq 1 - \varepsilon$ so that $P'(\mathscr{Y}'_\varepsilon) \geq 1 - \xi^{1/4} - \varepsilon$ and $P(\mathscr{Y}_\varepsilon) \geq 1 - \xi^{1/4} - \xi - \varepsilon \geq 1 - \varepsilon - 2\xi^{1/4}$, which is point ii(a).

We now prove ii(b). For $\tilde{y} \in \mathscr{Y}_\varepsilon$, let $C(\tilde{y})$ be the $(n, h', (U(e)/\varepsilon^2)(\eta + 1/\sqrt{n}))$ typical set of $P'(\cdot|\tilde{y})$ and $A(\tilde{y}) = \{(l, \tilde{x}), P'(l, \tilde{x} \mid \tilde{y}) > \xi^{2/3} P''(l, \tilde{x} \mid \tilde{y})\}$. Then,

$$\begin{aligned} P(C(\tilde{y}) \cap A(\tilde{y}) \mid \tilde{y}) &= \frac{(P_{\mathscr{L}}(C)P' + (1 - P_{\mathscr{L}}(C))P'')(C(\tilde{y}) \cap A(\tilde{y}))}{(P_{\mathscr{L}}(C)P' + (1 - P_{\mathscr{L}}(C))P'')(\tilde{y})} \\ &\geq \frac{(1 - \xi)P'(C(\tilde{y}) \cap A(\tilde{y}))}{(1 + \xi^{3/4})P'(\tilde{y})} \\ &\geq (1 - \xi - \xi^{3/4}) P'(C(\tilde{y}) \cap A(\tilde{y}) \mid \tilde{y}) \\ &\geq (1 - \xi - \xi^{3/4})(1 - \varepsilon - \xi^{2/3}) \\ &\geq 1 - \varepsilon - 3\sqrt{\xi}, \end{aligned}$$

where the first inequality uses $P''(\tilde{y}) \leq \xi^{-1/4} P'(\tilde{y})$, and the third one uses $P'(A(\tilde{y})|\tilde{y}) \geq 1 - \xi^{2/3}$ and $P'(C(\tilde{y})|\tilde{y}) \geq 1 - \varepsilon$. This says that $C(\tilde{y}) \cap A(\tilde{y})$ will be the typical set for $P(\cdot|\tilde{y})$ and fixes the value of ξ'.

To estimate the parameter η', we evaluate the ratio $P(l,\tilde{x}\mid\tilde{y})/P'(l,\tilde{x}\mid\tilde{y})$. For $\tilde{y}\in\mathcal{Y}_\varepsilon$ and $(l,\tilde{x})\in C(\tilde{y})\cap A(\tilde{y})$, we obtain

$$\begin{aligned}\frac{P(l,\tilde{x}\mid\tilde{y})}{P'(l,\tilde{x}\mid\tilde{y})} &= \frac{(P_{\mathcal{L}}(C)P'+(1-P_{\mathcal{L}}(C))P'')(l,\tilde{x})}{(P_{\mathcal{L}}(C)P'+(1-P_{\mathcal{L}}(C))P'')(\tilde{y})}\frac{1}{P'(l,\tilde{x}\mid\tilde{y})}\\ &\geq \frac{(1-\xi)P'(l,\tilde{x})}{(1+\xi^{3/4})P'(\tilde{y})}\frac{1}{P'(l,\tilde{x}\mid\tilde{y})}\\ &\geq 1-\xi-\xi^{3/4}\\ &\geq 1-2\xi^{3/4}.\end{aligned}$$

On the other hand,

$$\begin{aligned}\frac{P(l,\tilde{x}\mid\tilde{y})}{P'(l,\tilde{x}\mid\tilde{y})} &= \frac{P_{\mathcal{L}}(C)P'(l,\tilde{x})+(1-P_{\mathcal{L}}(C))P''(l,\tilde{x})}{P_{\mathcal{L}}(C)P'(\tilde{y})(1+((1-P_{\mathcal{L}}(C))/P_{\mathcal{L}}(C))(P''(\tilde{y})/P'(\tilde{y})))}\frac{1}{P'(l,\tilde{x}\mid\tilde{y})}\\ &\leq \left(\frac{P'(l,\tilde{x})}{P'(\tilde{y})}+\frac{(1-P_{\mathcal{L}}(C))P''(l,\tilde{x})}{P_{\mathcal{L}}(C)P'(\tilde{y})}\right)\frac{1}{P'(l,\tilde{x}\mid\tilde{y})}\\ &\leq 1+\frac{\xi}{1-\xi}\frac{P''(l,\tilde{x})}{\xi^{1/4}P''(\tilde{y})}\frac{1}{P'(l,\tilde{x}\mid\tilde{y})}\leq 1+\frac{\xi}{1-\xi}\frac{1}{\xi^{1/4}\xi^{2/3}}\\ &\leq 1+2\xi^{1/12}.\end{aligned}$$

Hence

$$|\log P(l,\tilde{x}\mid\tilde{y})-\log P'(l,\tilde{x}\mid\tilde{y})|\leq -\log(1-2\xi^{1/12})\leq 4\xi^{1/12}.$$

Hence the result from ii(b) of Proposition 5.2. □

6. Construction of the process. Taking up the proof of Proposition 3.1, let λ rational, e, e' having finite support be such that $\lambda\Delta H(e)+(1-\lambda)\Delta H(e')>0$ and $C=\operatorname{supp} e\cup\operatorname{supp} e'$. We wish to construct a law P of a C-process that achieves $\delta=\lambda\epsilon_e+(1-\lambda)\epsilon_{e'}$. Again, by closedness of the set of achievable distributions, we assume w.l.o.g. $e\in\mathbb{T}_{n_0}(C)$, $e'\in\mathbb{T}_{n_0}(C)$ for some common n_0, $0<\lambda<1$, and $\lambda=\frac{M}{M+N}$ with M,N multiples of n_0.

Since $\lambda\Delta H(e)+(1-\lambda)\Delta H(e')>0$, we assume w.l.o.g. $\Delta H(e)>0$. Remark that for each $p\in\operatorname{supp} e$, $\Delta H(\epsilon_p)=H(\mathbf{x}|p)-H(\mathbf{y}|p)$, thus

$$\mathbf{E}_e(\Delta H(\epsilon_p))=H(\mathbf{x}|\mathbf{p})-H(\mathbf{y}|\mathbf{p})\geq H(\mathbf{x}|\mathbf{p})-H(\mathbf{y})=\Delta H(e)>0.$$

Therefore, there exists $p_0\in\operatorname{supp} e$ such that $\Delta H(\epsilon_{p_0})>0$ and we assume w.l.o.g. $\operatorname{supp} e'\ni p_0$. Hence $\max\{d(\epsilon_{p_0}\|e), d(\epsilon_{p_0}\|e')\}$ is well defined and finite.

We construct the process by blocks. For a block lasting from stage $T+1$ up to stage $T+M$ (resp. $T+N$), we construct $(\mathbf{x}_1,\ldots,\mathbf{x}_T)$-measurable random variables $\mathbf{p}_{T+1},\ldots,\mathbf{p}_{T+M}$ such that their distribution conditional to $\mathbf{y}_1,\ldots,\mathbf{y}_T$ is close to that of M (resp. N) i.i.d. random variables of law e (resp. e'). We then take $\mathbf{x}_{T+1}\ldots,\mathbf{x}_{T+M}$ of law $\mathbf{p}_{T+1},\ldots,\mathbf{p}_{T+M}$, and independent of the past of the process conditional to $\mathbf{p}_{T+1},\ldots,\mathbf{p}_{T+M}$.

We define the process $(\mathbf{x}_t)_t$ and its law P over $\overline{N}=N_0+L(M+N)$ stages, where (M,N) are multiples of (m,n), inductively over blocks of stages.

Definition of the blocks. The first block labeled 0 is an initialization phase that lasts from stage 1 to N_0. For $1\leq k\leq L$, the $2k$-th [resp. $2k+1$-th] block consists of stages $N_0+(k-1)(M+N)+1$ to $N_0+(k-1)(M+N)+M$ [resp. $N_0+(k-1)(M+N)+M+1$ to $N_0+k(M+N)$].

Initialization block. During the initialization phase, $\mathbf{x}_1,\mathbf{x}_2,\ldots,\mathbf{x}_{N_0}$ are i.i.d. with law p_0, inducing a law P_0 of the process during this block.

First block. Let S_0 be the set of $\tilde{y}_0\in Y^{N_0}$ such that $P_0(\cdot|\tilde{y}_0)$ satisfies an $\mathbf{AEP}(M,h_0,\eta_0,\xi_0)$. After histories in S_0 and for suitable values of the parameters h_0,η_0,ξ_0, applying Proposition 5.2 to $(\mathcal{L},P_{\mathcal{L}})=(X^{N_0},P_0(\cdot|\tilde{y}_0))$ allows to define random variables $\mathbf{p}_{N_0+1},\ldots,\mathbf{p}_{N_0+M}$ such that their distribution conditional to $\mathbf{y}_1,\ldots,\mathbf{y}_{N_0}$ is close to that of M i.i.d. random variables of law e. We then take $\mathbf{x}_{N_0+1},\ldots,\mathbf{x}_{N_0+M}$ of law $\mathbf{p}_{N_0+1},\ldots,\mathbf{p}_{N_0+M}$, and independent of the past of the process conditional to $\mathbf{p}_{N_0+1},\ldots,\mathbf{p}_{N_0+M}$. We let $\mathbf{x}_t$ be i.i.d. with law p_0 after histories not in S_0. This defines the law of the process up to the first block.

Second block. Let $\tilde{y}_1$ be a history of signals to the statistician during the initialization block and the first block. Proposition 5.2 ensures that, given $\tilde{y}_0\in S_0$, $P_1(\cdot|\tilde{y}_1)$ satisfies an $\mathbf{AEP}(M,h_0',\eta_0',\xi_0')$ with probability no less than $1-\varepsilon-2\xi_0^{1/4}$, where we set $h_0'=h_0+\Delta H(e)$, $\eta_0'=(U(e)/\varepsilon^2)(\eta_0+1/\sqrt{M})+(4/M)\xi_0^{1/12}$, and $\xi_0'=\varepsilon+3\sqrt{\xi_0}$.

But an $\mathbf{AEP}(M, h_0', \eta_0', \xi_0')$ is identical to an $\mathbf{AEP}(N, (M/N)h_0', (M/N)\eta_0', \xi_0')$. Since $M/N = \lambda/(1-\lambda)$, for each $\tilde{y}_0 \in S_0$, $P_1(\cdot|\tilde{y}_1)$ satisfies an $\mathbf{AEP}(N, h_1, \eta_1, \xi_1)$ with probability no less that $1-\varepsilon-2\xi_0^{1/4}$, with $h_1 = \lambda/(1-\lambda)(h_0+\Delta H(e))$, $\eta_1 = \lambda/(1-\lambda)\big[(U(e)/\varepsilon^2)(\eta_0+1/\sqrt{M})+(4/M)\xi_0^{1/12}\big]$, $\xi_1 = \varepsilon+3\sqrt{\xi_0}$. Thus the set S_1 of $\tilde{y}_1$ such that $P_1(\cdot|\tilde{y}_1)$ satisfies an $\mathbf{AEP}(N, h_1, \eta_1, \xi_1)$ has probability no less than $1-2\varepsilon-2\xi_0^{1/4}-2\xi_1^{1/4}$.

Inductive construction. We define inductively the laws P_k for the process up to block k and parameters h_k, η_k, ξ_k. We set $N_k = M$ if k odd and $N_k = N$ if k even. Let S_k be the set of histories $\tilde{y}_k$ for the statistician up to block k such that $P_k(\cdot|\tilde{y}_k)$ satisfies an $\mathbf{AEP}(N_k, h_k, \eta_k, \xi_k)$. After $\tilde{y}_k \in S_k$, define the process during block k to approximate e i.i.d. if k is odd, and e' i.i.d. if k is even. After $\tilde{y}_k \notin S_k$, let the process during block k be i.i.d. with law p_0 conditional to the past. Proposition 5.2 ensures that conditional on $\tilde{y}_k \in S_k$, $P_{k+1}(\cdot|\tilde{y}_{k+1})$ satisfies an $\mathbf{AEP}(N_{k+1}, h_{k+1}, \eta_{k+1}, \xi_{k+1})$ with probability no less than $1-\varepsilon-2\xi_k^{1/4}$, where h_{k+1}, η_{k+1}, and ξ_{k+1} are given by the recursive relations

$$\begin{cases} h_{k+1} = \dfrac{\lambda}{1-\lambda}(h_k+\Delta H(e)) \\ \eta_{k+1} = \dfrac{\lambda}{1-\lambda}\left(\dfrac{U(e)}{\varepsilon^2}\left(\eta_k+\dfrac{1}{\sqrt{M}}\right)+\dfrac{4}{M}\xi_k^{1/12}\right) \\ \xi_{k+1} = \varepsilon+3\sqrt{\xi_k} \end{cases}$$

if k is even, and

$$\begin{cases} h_{k+1} = \dfrac{1-\lambda}{\lambda}(h_k+\Delta H(e')) \\ \eta_{k+1} = \dfrac{1-\lambda}{\lambda}\left(\dfrac{U(e')}{\varepsilon^2}\left(\eta_k+\dfrac{1}{\sqrt{N}}\right)+\dfrac{4}{N}\xi_k^{1/12}\right) \\ \xi_{k+1} = \varepsilon+3\sqrt{\xi_k} \end{cases}$$

if k is odd.

The definition of the process for the $2L+1$ blocks is complete, provided for each k odd, $M(h_k-H(e)-\eta_k)\geq 2$, and for each k even, $N(h_k-H(e')-\eta_k)\geq 2$. We seek now conditions on $(\varepsilon, \eta_0, \xi_0, N_0, M, N, L)$ such that these inequalities are fulfilled. We first establish bounds on the sequences (ξ_k, η_k, h_k) and introduce some notations

$$a(\varepsilon) = \frac{1}{\varepsilon^2}\max\left(\frac{\lambda}{1-\lambda}U(e);\ \frac{1-\lambda}{\lambda}U(e')\right) \tag{9}$$

$$c(\varepsilon, M, N) = \max\left(\frac{\lambda}{1-\lambda}\frac{U(e)}{\varepsilon^2}\frac{1}{\sqrt{M}}+\frac{8}{M};\ \frac{1-\lambda}{\lambda}\frac{U(e')}{\varepsilon^2}\frac{1}{\sqrt{N}}+\frac{8}{N}\right). \tag{10}$$

Lemma 6.1. *For $k = 1, \ldots, 2L$:*
(i) $\xi_k \leq \xi_{\max} = 11((\varepsilon)^{2^{-2L}}+(\xi_0)^{2^{-2L}})$.
(ii) $\eta_k \leq \eta_{\max} = (a(\varepsilon))^{2L}\big[\eta_0 - c(\varepsilon, M, N)/(1-a(\varepsilon))\big]+c(\varepsilon, M, N)/(1-a(\varepsilon))$.
(iii) *$h_k \geq h_0$ for k even and $h_k \geq h_1$ for k odd.*

Proof. (1) Let θ be the unique positive number such that $\theta = 1+3\sqrt{\theta}$, one can check easily that $\theta < 11$ (numerically, $\theta \cong 10.91$). Using that for $x, y > 0$, $\sqrt{x+y} \leq \sqrt{x}+\sqrt{y}$ and for $0 < x < 1$, $x < \sqrt{x}$, one verifies by induction that for $k = 1, \ldots, 2L$:

$$\xi_k \leq \theta\varepsilon^{2^{-k}} + 3^{\sum_{j=0}^{k-1}2^{-j}}(\xi_0)^{2^{-k}}$$

and the result follows.

(2) One easily checks numerically that $\xi_{\max} < 22$ and $4\xi_{\max}^{1/12} < 8$. From the definition of the sequence (η_k), for each k:

$$\eta_{k+1} \leq a(\varepsilon)\eta_k + c(\varepsilon, M, N)$$

the expression of $\eta_{\max}$ follows.

(3) For k even, $h_{k+2} = h_k + (1/\lambda)(\lambda\Delta H(e)+(1-\lambda)\Delta H(e')) > h_k$, similarly for k odd and the proof is completed by induction. □

The starting entropy h_0 comes from the initialization block.

Lemma 6.2. *For all h_0, η_0, ξ_0, there exists $\overline{N}_0(h_0, \eta_0, \xi_0)$ such that for any (N_0, M) that satisfy the conditions*

$$N_0 \geq \overline{N}_0(h_0, \eta_0, \xi_0) \tag{11}$$

$$\left|\frac{N_0}{M}\Delta H(\epsilon_{p_0}) - h_0\right| \leq \frac{\eta_0}{3} \tag{12}$$

$$P(\{\tilde{y}_0, P_0(\cdot|\tilde{y}_0) \text{ satisfies an } \mathbf{AEP}(M, h_0, \eta_0, \xi_0)\}) \geq 1 - \xi_0.$$

Proof. Since $\mathbf{x}_1, \ldots, \mathbf{x}_{N_0}$ are i.i.d. with law p_0, the conditional distributions $P_0(\mathbf{x}_i|f(\mathbf{x}_i))$ are also i.i.d. and for each $i = 1, \ldots, N_0$, $H(\mathbf{x}_i|f(\mathbf{x}_i)) = \Delta H(\epsilon_{p_0})$. Let $\bar{h} = \Delta H(\epsilon_{p_0}) > 0$, $\bar{\eta} = (\bar{h}/h_0)(\eta_0/3)$ and for each N_0:

$$C_{N_0} = \left\{x_1, \ldots, x_{N_0}, \left| -\frac{1}{N_0}\log P(x_1, \ldots, x_{N_0}|f(x_1), \ldots, f(x_{N_0})) - \bar{h}\right| \leq \bar{\eta}\right\}.$$

By the law of large numbers there is n_0 such that for $N_0 \geq n_0$, $P(C_{N_0}) \geq 1 - \xi_0^2$. For each sequence of signals $\tilde{y}_0 = (f(x_1), \ldots, f(x_{N_0}))$, define

$$C_{N_0}(\tilde{y}_0) = \left\{x_1, \ldots, x_{N_0}, \left| -\frac{1}{N_0}\log P(x_1, \ldots, x_{N_0}|\tilde{y}_0) - \bar{h}\right| \leq \bar{\eta}\right\}$$

and set

$$S_0 = \{\tilde{y}_0, P_0(C_{N_0}(\tilde{y}_0)|\tilde{y}_0) \geq 1 - \xi_0\}.$$

Then, $P(C_{N_0}) = \sum_{\tilde{y}_0} P_0(\tilde{y}_0)P_0(C_{N_0}(\tilde{y}_0)|\tilde{y}_0) \leq P(S_0) + (1-\xi_0)(1-P(S_0))$, and therefore $P(S_0) \geq 1 - \xi_0$, which means

$$P(\{\tilde{y}_0, P_0(\cdot|\tilde{y}_0) \text{ satisfies an } \mathbf{AEP}(N_0, \bar{h}, \bar{\eta}, \xi_0)\}) \geq 1 - \xi_0.$$

Thus for each $\tilde{y}_0 \in S_0$, $P_0(\cdot|\tilde{y}_0)$ satisfies an $\mathbf{AEP}(M, (N_0/M)\bar{h}, (N_0/M)\bar{\eta}, \xi_0)$. Choose then (M, N_0) such that condition (12) is fulfilled and from the choice of $\bar{\eta}$, $P_0(\cdot|\tilde{y}_0)$ satisfies an $\mathbf{AEP}(M, h_0, \eta_0, \xi_0)$.

We give now sufficient conditions for the construction of the process to be valid.

Lemma 6.3. *If the following two conditions are fulfilled*:

$$M(h_0 - H(e) - \eta_{\max}) \geq 2 \tag{13}$$

$$N(h_1 - H(e') - \eta_{\max}) \geq 2, \tag{14}$$

then for $k = 0, \ldots, 2L$,

$$\begin{cases} M(h_k - H(e) - \eta_k) \geq 2 & \text{for } k \text{ odd} \\ N(h_k - H(e') - \eta_k) \geq 2 & \text{for } k \text{ even.} \end{cases}$$

Proof. Follows from Lemma 6.1. □

Summing up, we get:

Lemma 6.4. *Under conditions* (11), (12), (13), *and* (14), *the process is well defined.*

Note that the process so constructed is indeed a C-process, since at each stage n, the conditional law of $\mathbf{x}_n$ given $(x_1, \ldots, x_{n-1})$, belongs either to $\operatorname{supp} e$ or to $\operatorname{supp} e'$.

7. Bound on Kullback distance. Let P be the law of the process process $(\mathbf{x}_t)$ defined above. We estimate on each block the distance between the sequence of experiments induced by P with $e^{\otimes M}$ [resp $e'^{\otimes N}$]. Then, we show that these distances can be made small by an adequate choice of the parameters. Finally, we prove the weak-$*$ convergence of the distribution of experiments under P to $\lambda\epsilon_e + (1-\lambda)\epsilon_{e'}$.

Lemma 7.1. *There exists a constant $U(e, e')$ such that if* (11), (12), (13), *and* (14) *are fulfilled, then for all k odd,*

$$\begin{aligned}\mathbf{E}\, d\big(P(\mathbf{p}_{t_k+1}, \ldots, \mathbf{p}_{t_{k+1}}|\tilde{y}_{k-1})\big)\|e^{\otimes M} \\ \leq M \cdot U(e, e') \cdot \left(\eta_{\max} + \xi_{\max} + \frac{\log(M+1)}{M} + L\big(\varepsilon + 2\xi_{\max}^{1/4}\big)\right),\end{aligned}$$

and for all k even,

$$\mathbf{E}\, d\big(P(\mathbf{p}_{t_k+1},\ldots,\mathbf{p}_{t_{k+1}}|\tilde{y}_{k-1})\big)\|e'^{\otimes N}$$
$$\leq N\cdot U(e,e')\cdot\left(\eta_{\max}+\xi_{\max}+\frac{\log(N+1)}{N}+L\big(\varepsilon+2\xi_{\max}^{1/4}\big)\right),$$

where for each k, t_k denotes the last stage of the $(k-1)$th block.

Proof. Assume k odd, the even case being similar. For $\tilde{y}_{k-1}\in S_{k-1}$, Proposition 5.2 shows that

$$d\big(P(\mathbf{p}_{t_k+1},\ldots,\mathbf{p}_{t_{k+1}}|\tilde{y}_{k-1})\|e^{\otimes M}\big)$$
$$\leq 2M\big(\eta_{\max}+\xi_{\max}\log(|\operatorname{supp} e|)\big)+|\operatorname{supp} e|\log(M+1)+2.$$

For $\tilde{y}_{k-1}\notin S_{k-1}$,

$$d(P(\mathbf{p}_{t_k+1},\ldots,\mathbf{p}_{t_{k+1}}|\tilde{y}_{k-1})\|e^{\otimes M})=Md(\epsilon_{p_0}\|e).$$

The result follows, using $P\big(\bigcap_1^{2L}S_k\big)\geq 1-2L(\varepsilon+2\xi_{\max}^{1/4})$ and with $U(e,e')=\max\big\{2,2d(\epsilon_{p_0}\|e),2d(\epsilon_{p_0}\|e'),2|\operatorname{supp} e|,2|\operatorname{supp} e'|\big\}$.

Lemma 7.2. *For any L and any $\gamma>0$, there exists $(\varepsilon,\varepsilon_0,\eta_0)$, and $(\overline{M},\overline{N})$ such that for all $(M,N)>(\overline{M},\overline{N})$, conditions (13) and (14) are fulfilled and for all N_0 such that (11) and (12) hold, for all k odd,*

$$\mathbf{E}\, d\big(P(\mathbf{p}_{t_k+1},\ldots,\mathbf{p}_{t_{k+1}}|\tilde{y}_{k-1})\big)\|e^{\otimes M}\leq M\gamma$$

and for all k even,

$$\mathbf{E}\, d\big(P(\mathbf{p}_{t_k+1},\ldots,\mathbf{p}_{t_{k+1}}|\tilde{y}_{k-1})\big)\|e'^{\otimes N}\leq N\gamma.$$

Proof. We show how to choose the parameters to achieve the above result.

(i) Choose ε and ξ_0 such that $\xi_{\max}$ and $L\varepsilon$ are small.
(ii) Choose η_0 and (M_η,N_η) such that $(\log(M+1))/M$, $(\log(N+1))/N$, and $\eta_{\max}$ are small for all $(M,N)\geq(M_\eta,N_\eta)$.
(iii) Choose $N_0\geq N_0(h_0,\eta_0,\xi_0)$.
(iv) Choose $(\overline{M},\overline{N})\geq(M_\eta,N_\eta)$ such that (13) and (14) are satisfied for $(M,N)\geq(\overline{M},\overline{N})$.
(v) Choose $(M,N)\geq(\overline{M},\overline{N})$ such that (12) holds.

Applying Lemma 7.1 then yields the result. □

Lemma 7.3. *For any $\gamma>0$, there exists $(\varepsilon,\xi_0,\eta_0,M,N,N_0,L)$ that fulfill (11), (12), (13), and (14) and such that*

(i) *for k odd, $\mathbf{E}d(P(\mathbf{p}_{t_k+1},\ldots,\mathbf{p}_{t_{k+1}}|\tilde{y}_{k-1}))\|e^{\otimes M}\leq M\gamma$*
(ii) *for k even, $\mathbf{E}d(P(\mathbf{p}_{t_k+1},\ldots,\mathbf{p}_{t_{k+1}}|\tilde{y}_{k-1}))\|e'^{\otimes N}\leq N\gamma$*
(iii) *$N_0/\overline{N}\leq\gamma$.*

Proof. It is enough to use the previous lemma, where L is chosen a large constant times $1/\gamma$. Then, remark that for (M,N) large enough, (12) is fulfilled for N_0 of an order constant times M, hence of the order constant times $\overline{N}/L$. □

7.1. Weak-$*$ convergence. Lemma 7.3 provides a choice of parameters for each $\gamma>0$, hence a family of processes P_γ and a corresponding family $(\delta_\gamma)_\gamma$ of elements of D_∞.

Lemma 7.4. *δ_γ weak-$*$ converges to $\lambda\epsilon_e+(1-\lambda)\epsilon_{e'}$ as γ goes to 0.*

Proof. With $\delta'=(N_0/\overline{N})\epsilon_{p_0}+(1-(N_0/\overline{N}))(\lambda\epsilon_e+(1-\lambda)\epsilon_{e'})$, since $N_0/\overline{N}\leq\gamma$, δ' converges weakly to $\lambda\epsilon_e+(1-\lambda)\epsilon_{e'}$ as γ goes to 0. Let $g\colon E\to\mathbb{R}$ continuous, we prove that $|\mathbf{E}_{\delta'}g-\mathbf{E}_{\delta_\gamma}g|$ converges to 0 as γ goes to 0.

$$|\mathbf{E}_{\delta'}g-\mathbf{E}_{\delta_\gamma}g|\leq\frac{1}{\overline{N}-N_0}\sum_{k\text{ odd}}\sum_{t=t_k+1}^{t_{k+1}}\mathbf{E}|g(\mathbf{e}_t)-g(e)|$$
$$+\frac{1}{\overline{N}-N_0}\sum_{k\text{ even}}\sum_{t=t_k+1}^{t_{k+1}}\mathbf{E}|g(\mathbf{e}_t)-g(e')|.$$

By uniform continuity of g, for every $\bar{\varepsilon} > 0$, there exists $\bar{\alpha} > 0$ such that

$$\|e_1 - e_2\|_1 \leq \bar{\alpha} \implies |g(e_1) - g(e_2)| \leq \bar{\varepsilon}.$$

We let $e_k = e$ for k odd and $e_k = e'$ for k even and $\|g\| = \max_{e''} |g(e'')|$. For t in the k-th block

$$\begin{aligned}\mathbf{E}|g(\mathbf{e}_t) - g(e_k)| &\leq \bar{\varepsilon} + \frac{2\|g\|}{\bar{\alpha}} \mathbf{E}\|\mathbf{e}_t - e_k\|_1 \\ &\leq \bar{\varepsilon} + \frac{2\|g\|}{\bar{\alpha}} \sqrt{2\ln 2 \cdot \mathbf{E}d(\mathbf{e}_t \| e_k)},\end{aligned}$$

since $\|p - q\|_1 \leq \sqrt{2\ln 2 \cdot d(p\|q)}$ (Cover and Thomas [7, Lemma 12.6.1, p. 300]) and from Jensen's inequality. Applying Jensen's inequality again:

$$\frac{1}{N_k} \sum_{t=t_k+1}^{t_{k+1}} \mathbf{E}|g(\mathbf{e}_t) - g(e_k)| \leq \bar{\varepsilon} + \frac{2\|g\|}{\bar{\alpha}} \sqrt{\frac{2\ln 2}{N_k} \sum_{t=t_k+1}^{t_{k+1}} \mathbf{E}d(\mathbf{e}_t \| e_k)}.$$

Now,

$$\begin{aligned}\sum_{t=t_k+1}^{t_{k+1}} \mathbf{E}d(\mathbf{e}_t \| e_k) &= \sum_{t=t_k+1}^{t_{k+1}} \mathbf{E}d\big(P(\mathbf{p}_t | \tilde{y}_{k-1}, y_{t_k+1}, \ldots, y_{t-1}) \| e_k\big) \\ &\leq \sum_{t=t_k+1}^{t_{k+1}} \mathbf{E}d\big(P(\mathbf{p}_t | \tilde{y}_{k-1}, \mathbf{p}_{t_k+1}, \ldots, \mathbf{p}_{t-1}) \| e_k\big) \\ &= \mathbf{E}_{\tilde{y}_{k-1}} d\big(P(\mathbf{p}_{t_k+1}, \ldots, \mathbf{p}_{t_{k+1}} | \tilde{y}_{k-1}) \| e_k^{\otimes N_k}\big) \\ &\leq N_k \gamma,\end{aligned}$$

where the first inequality comes from the convexity of the Kullback distance. Reporting in the previous and averaging over blocks yields

$$|\mathbf{E}_{\delta'} g - \mathbf{E}_{\delta_\gamma} g| \leq \bar{\varepsilon} + \frac{2\|g\|}{\bar{\alpha}} \sqrt{2\ln 2 \cdot \gamma}.$$

Thus $|\mathbf{E}_{\delta'} g - \mathbf{E}_{\delta_\gamma} g|$ goes to 0 as γ goes to 0. □

References

[1] Abreu, D., D. Pearce, E. Stacchetti. 1990. Toward a theory of discounted repeated games with imperfect monitoring. *Econometrica* **58** 1041–1063.
[2] Bavly, G., A. Neyman. 2003. Online concealed correlation by boundedly rational players. Discussion Paper Series 336, Center for Rationality and Interactive Decision Theory, Hebrew University, Jerusalem, Israel.
[3] Ben Porath, E. 1993. Repeated games with finite automata. *J. Econom. Theory* **59** 17–32.
[4] Blackwell, D. 1951. Comparison of experiments. *Proc. Second Berkeley Sympos. Math. Statist. Probability*, University of California Press, Berkeley, CA, 93–102.
[5] Blackwell, D. 1953. Equivalent comparison of experiments. *Ann. Math. Statist.* **24** 265–272.
[6] Compte, O. 1998. Communication in repeated games with imperfect private monitoring. *Econometrica* **66** 597–626.
[7] Cover, T. M., J. A. Thomas. 1991. *Elements of Information Theory*. Wiley Series in Telecomunications. John Wiley and Sons, New York.
[8] Fudenberg, D., D. K. Levine, E. Maskin. 1994. The folk theorem with imperfect public information. *Econometrica* **62** 997–1039.
[9] Goldberg, Y. 2003. On the minmax of repeated games with imperfect monitoring: A computational example. Discussion Paper Series 345, Center for the Study of Rationality, Hebrew University, Jerusalem, Israel.
[10] Gossner, O. 1998. Repeated games played by cryptographically sophisticated players. Working Paper DP 9835, CORE, Louvain-la-neuve, Belgium.
[11] Gossner, O. 2000. Sharing a long secret in a few public words. Working Paper DP 2000-15, THEMA, Paris, France.
[12] Gossner, O., P. Hernández. 2003. On the complexity of coordination. *Math. Oper. Res.* **28** 127–141.
[13] Gossner, O., R. Laraki, T. Tomala. 2004. Maxmin computation and optimal correlation in repeated games with imperfect monitoring. Working Paper DP-2004-63, Cahiers du Ceremade, Université Paris Dauphine, Paris, France. *Math. Programming*.
[14] Gossner, O., T. Tomala. 2006. Secret correlation in repeated games with signals. *Math. Oper. Res.* Forthcoming.
[15] Gossner, O., N. Vieille. 2001. Repeated communication through the ‘and’ mechanism. *Internat. J. Game Theory* **30** 41–61.
[16] Gossner, O., N. Vieille. 2002. How to play with a biased coin? *Games Econom. Behavior* **41** 206–226.
[17] Kandori, M., H. Matsushima. 1998. Private observation, communication and collusion. *Rev. Econom. Stud.* **66** 627–652.
[18] Lacôte, G., G. Thurin. 2003. How to efficiently defeat strategies of bounded complexity. Mimeographed.
[19] Lehrer, E. 1988. Repeated games with stationary bounded recall strategies. *J. Econom. Theory* **46** 130–144.
[20] Lehrer, E. 1991. Internal correlation in repeated games. *Internat. J. Game Theory* **19** 431–456.

[21] Lehrer, E. 1994. Finitely many players with bounded recall in infinitely repeated games. *Games Econom. Behavior* **7** 390–405.
[22] Neyman, A. 1997. Cooperation, repetition, and automata. S. Hart, A. Mas-Colell, eds. *Cooperation: Game-Theoretic Approaches*. NATO ASI Series F, Vol. 155. Springer-Verlag, Berlin, Germany, Berlin, Germany, 233–255.
[23] Neyman, A. 1998. Finitely repeated games with finite automata. *Math. Oper. Res.* **23** 513–552.
[24] Parthasaraty, K. R. 1967. *Probability Measures on Metric Spaces*. Academic Press, New York.
[25] Piccione, M., A. Rubinstein. 2003. Modeling the economic interaction of agents with diverse abilities to recognize equilibrium patterns. *J. Eur. Econom. Association* **1** 212–223.
[26] Renault, J., T. Tomala. 1998. Repeated proximity games. *Internat. J. Game Theory* **27** 539–559.
[27] Renault, J., T. Tomala. 2004. Communication equilibrium payoffs of repeated games with imperfect monitoring. *Games Econom. Behavior* **49** 313–344.
[28] Rockafellar, R. T. 1970. *Convex Analysis.* Princeton University Press, Princeton, NJ.