DEPARTMENT OF ECONOMICS

UNIVERSITY OF MILAN - BICOCCA

WORKING PAPER SERIES

Rational Learning in Imperfect Monitoring Games

Mario Gilli

No. 46 - March 2002

Dipartimento di Economia Politica
Università degli Studi di Milano - Bicocca
http://dipeco.economia.unimib.it

Rational Learning in Imperfect Monitoring Games

Mario Gilli*
Università di Milano-Bicocca - Dipartimento di Economia Politica
Piazza dell'Ateneo Nuovo, 1 - 20126 Milano - Italy
tel: ++39 02 461916
fax: ++39 02 700439225
email: mario.gilli@unimib.it

March 8, 2002

JEL: C72, D83, D82
Keywords: rational learning, active and passive learning, stationarity, consistency, merging.

Abstract

This paper provides a general framework to analyze rational learning in strategic situations where the players have private information and update their private priors collecting data through optimal experimentation. The theory of statistical inference for stochastic processes and of Markovian dynamic programming is applied to study players asymptotic behavior in the context of repeated and recurring games, proving convergence towards Conjectural equilibria, an opportune generalization of Nash equilibria for this kind of strategic situations. Since the main bulk of the literature on rational learning regards convergence towards equilibria of repeated games, the main contribution of this paper is to argue for rational learning in recurring games, providing dynamic foundations for equilibria of the one-shot game. The analysis focuses on the problem of non stationary environment and on the problem of the correct specification of the stochastic law which regulates players' observations. In this way the paper shows both the limitations and the possibilities of rational learning models in game theory, in particular explaining when and why consistency rather than merging is the correct notion of learning in games.

*This paper is a revised version of chapter 3 of my Ph.D. dissertation at Cambridge University. I am grateful for their helpful comments to the participants at the seminars in Bari, Bocconi, Cambridge, Genova and Roma universities. I benefit from discussions with Pier Paolo Battigalli, David Canning, Giacomo Costa, David Easley, Bob Evans, Dean Foster, Frank Hahn, Fioravante Patrone and especially Rupert Gatti and Hamid Sabourian. Finally I would like to thank an anonymous referee and the editor for their useful suggestions.

"If a man will begin with certainties,
he shall end in doubts,
but if he will be content to begin with doubts,
he shall end in certainties"
Francis Bacon "The Advancement of Learning", Book I, Ch. 5.

1 Introduction

1.1 Background

Can decision-makers learn to make equilibrium choices through repeated experiences? How can be described the behavioral patterns associated to long-run outcomes of dynamic processes of learning? What can be learned from non-equilibrium explanations of equilibrium? Questions like these form the research agenda and provide motivation to study learning models. These basic problems are not distinctive of strategic settings, game theory however provides powerful techniques to study agents out-of-equilibrium behavior.

The research agenda of this work is the analysis of rational learning models versus bounded rational learning models. The primary goal of this paper is to construct a general framework to discuss rational learning models in strategic situations where the players have private information and update their private priors collecting data through optimal experimentation. Its core is the explicit derivation from a model of repeated strategic interaction of the suitable setting to study rational learning, focusing the analysis both on Bayesian updating and on optimal experimentation. This task is pursued using powerful tools taken from the asymptotic theory of statistical inference for stochastic processes and from Markovian dynamic programming, so that the proof of convergence to equilibria is straightforward and natural. The work is organized in two parts: the first is negative and through an example shows the difficulties of bounded rational learning models, the second is positive and argues for rational learning models pointing out how to solve the problems faced by this approach. The paper shows how to construct stationary environments even when the players are learning and then (both for repeated and for recurring games) proves the convergence of rational learning processes to steady states characterized as Conjectural equilibria (Gilli [29]). The construction of the model, the results and their proof allow a comprehensive discussion and understanding of when and why rational learning is a fruitful hypothesis. In this way it is shown when and why consistency rather than merging is the correct notion of learning in games: when convergence towards equilibria of the repeated game is considered, then merge of opinions in the Blackwell and Dubins sense is crucial (theorem 1), when the focus is on equilibria of the stage game, then consistency in the statistical sense is crucial (theorem 2). The main bulk of the literature on rational learning regards convergence to equilibria of the repeated game, but very few papers analyze the processes of convergence towards equilibria of the stage game, usually assuming either myopic players or incomplete information. The hypothesis of myopia is particularly ill suited to learning contexts, since myopic agents have no reason to learn, while the limitation to incomplete information games obscures the crucial role of stationarity and of consistency in recurring games, as shown in section 5 and 6 of this paper. The main contribution of this paper then is to argue for rational learning in models of recurring games, providing dynamic foundations for equilibria of the one-shot game, via an explicit model of repeated games.

1.2 Outline

How to model learning is the basic question faced by the literature in this field. An obvious answer is through Bayesian updating. This means that

1. the agents have a well specified statistical model, which means
   (a) a set of possible states of nature (s.o.n.)
   (b) a prior on this set
   (c) a set of signals correlated to s.o.n.
   (d) a likelihood function which regulates the probability of the signals given the true s.o.n.
   (e) a set of possible choices
2. the priors are updated using Bayes rule.

As Jordan [39] e Nyarko [53] have shown, the hypothesis of Bayesian learning is not actually restrictive in the sense that any stochastic process of beliefs can be generated as the outcome of Bayesian updating choosing the statistical model in an opportune way. Instead, the truly important distinction is between rational (or sophisticated Bayesian) and bounded rational (or naive Bayesian) learning models. This crucial dichotomy depends on the subjective specification of the likelihood function: in rational learning models the decision-makers use the correct function to update their beliefs, while in bounded rational learning models the agents have an incorrect view of the random process that governs beliefs' updating. The relationship between the actual situation and its subjective representation by the learning agents is crucially important since the asymptotic properties of stochastic processes, and thus the effectiveness of any learning algorithm, are highly situation dependent (see e.g. Yamada [63]).

A classic example of naive Bayesian learning is fictitious play: Bayesian learning results in fictitious play dynamics when the players have beta prior distributions and analyze past observations as if their opponents' play were governed by a fixed, albeit unknown, probability distribution, even if actually the players' behavior is changing through time just because of learning (see e.g. Eichberger [17] and Fudenberg and Levine [25] chapter 2). This case illustrates some of the problems with this approach, in particular lack of convergence and repeated foolish behavior, since the players do not change their assumptions on the stochastic behavior of their opponents even when they face overwhelming evidence that their model of the stochastic setting is wrong. Indeed in naive Bayesian learning models the agents are updating their probability evaluations but they are not learning since they are not changing their theories even when they face increasing falsifying evidence.

In rational learning models the agents are learning about parameters of a distribution through repeated application of Bayes theorem using a correctly specified likelihood function. The rational learning approach faces two main difficulties. A first basic problem is that attempts to predict opponents' behavior can change the probability of future strategies and this influences agents' learning possibilities, in particular the likelihood functions. Therefore the stochastic process of strategies (and thus of signals) may be non stationary: in a learning process there is every reason to suppose that the relationship between observable and payoff relevant variables will not be stationary, even if the underlying environment is stationary. And in a non-stationary environment it is not clear if sequential revision of beliefs can lead to more accurate predictions and thus to equilibria. This will be called the **problem of non stationary statistical model**. The second crucial problem is that the agents are given likelihood functions which are a correct description of the data generating process resulting when these likelihood functions are actually used. Since each agent's specification of the likelihood function is correct given her own specification and those of all other agents, these situations are Nash equilibria of a grander game. Thus rational learning models need to justify the assumption of perfect understanding of the stochastic environment by the learning agents. Actually on one hand the best argument for rational learning is its optimality, but on the other hand this is also the best argument against it since it doesn't explain how this optimal procedure is discovered by the players. This will be called the **problem of correct theory**.

The generality of the model used in this work allows to tackle the above two main problems of rational learning. This paper provides sufficient but tight conditions to obtain a stationary statistical model where it is possible to apply the classic tools of stochastic dynamic programming (see e.g. Hinderer [32]) and of statistical inference for stochastic processes (see e.g. Basawa and Rao [4]). The problem of correct theory is more tricky to face since by definition the explanation of learning of the rational learning procedure is outside its domain. Therefore it is not possible to offer such explanation explicitly, but this paper provides a model where the players' comprehension of their stochastic environment does not look very demanding.

The results of this work allow to conclude that the reasons for using rational learning models are essentially two. First, at the moment there exists no formal precise definition of what is learning in a substantial sense: at the best there is just a theory of beliefs' updating, but naive Bayesian learning is unconvincing. Second, the knowledge of possible rest points of rational learning processes is useful when the precise characterization of substantial learning is unknown. As I show in this and in related papers (Gilli [30] and [31]), such steady states are situations where no player receives information contradicting her conjectures, called Conjectural equilibria (CE). This is a notion of stable behavior for game theoretic frameworks, which I have argued is the correct notion of equilibrium for games with signals (see Gilli [29], and for related considerations Battigalli [5], Gilli [26], Battigalli and Guaitoli [6], Battigalli et al. [7], Dekel et al. [14], Kalai and Lehrer [41], Fudenberg and Levine [24], Rubinstein and Wolinsky 1995, Sandroni and Smorodinsky [59], Sorin [62]). The results of this paper provide independent reasons to rely on CE as the right equilibrium notion for games with private information.

1.3 Related literature

The problems outlined before are studied considering Imperfect Monitoring Games (IMG). These are an effective generalization of strategic and extensive form games, where the information received by each player depends on the strategy profile played, so that there is a role for active learning policies. These strategic models actually are normal form games with signal functions which represent players' private signals as function of players' strategies. An overview of recent developments on strategic settings with private signals is Kandori [43], while the relationship between IMG and extensive form games is the object of Gilli [27].

The most direct predecessors of this paper are Kalai and Lehrer and Nyarko works on learning to play Nash and correlated equilibria (Kalai and Lehrer [40], Nyarko [52]), although I consider a more general setting. Other related papers are Jackson and Kalai [34] and [35], Jackson et al. [36], Jordan [37] and [38], Koutsougeras and Yannelis [45], Nyarko [50], [51], [54] and [55]. These works consider (incomplete information) strategic form games and (private) priors defined on the set of types and outcomes, showing when and how opportune stochastic processes converge to Correlated or Nash equilibria of the (true) stage game. The approach I follow in this paper is different since I allow for imperfect and private monitoring, while I do not explicitly consider the players' types: this would only complicate the setting without adding any generality since incomplete information games with private priors can easily be incorporated in my setting. Recently there have been many important contributions which from different perspectives stress the restrictions of the rational learning models à la Kalai and Lehrer: Miller and Sanchirico [47], Nachbar [48] and [49], Sandroni and Smorodinsky [60], Young and Foster [65]. Since these papers refer to Kalai and Lehrer [40] rational learning model, their focus is on the hypothesis of absolute continuity of players' beliefs and on repeated games. In particular they can apply their results to recurring games only assuming a zero discount factor, which implies myopic behavior and thus eliminates any incentive to experiment. My contribution in understanding restrictions and generality of rational learning is different and complementary, since I focus especially on recurring games and consequently on the assumptions which guarantee consistency of asymptotic beliefs, not on merging and thus on absolute continuity.

The paper is organized as follows. Section 2 illustrates the main points by means of an example, showing the problems with bounded rational learning models. Section 3 provides the basic definitions and the model of repeated strategic interaction. Section 4 adapts Kalai and Lehrer main result to this more general setting. Section 5 provides the detailed construction of the recurring game. Section 6 studies the convergence towards equilibria of the stage game. Section 7 concludes with final remarks.

2 An Example

The aims of this example are to show the role of Imperfect Monitoring Games and of Conjectural Equilibria to study learning and to illustrate the problems with Bounded Rational Learning Models

Consider the extensive form game of figure 1:

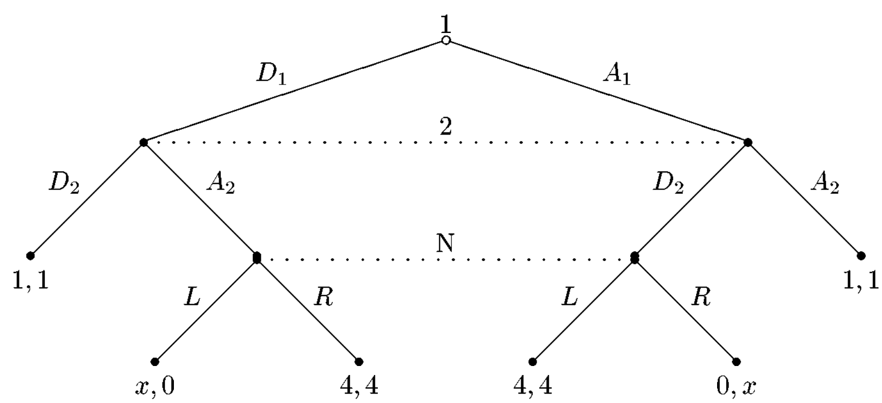

Figure 1

This situation can easily be represented as an IMG, where nature chooses between the matrix L and R and each player receives a signal as a function of strategy profile played. To simplify suppose that the signals coincide with the outcomes, which are public knowledge (figures 2 and 3).

L

	A_2	D_2
A_1	**1, 1**	**4, 4**
D_1	x, **0**	**1, 1**

R

	A_2	D_2
A_1	**1, 1**	**0,** x
D_1	**4, 4**	**1, 1**

Figure 2

L

	A_2	D_2
A_1	m^6	m^4
D_1	m^2	m^1

R

	A_2	D_2
A_1	m^6	m^5
D_1	m^3	m^1

Figure 3

Suppose $x = 0$ and consider the strategy profile (D_1, D_2, ω), where $\omega \in \{L, R\}$ is the unknown state of nature: this situation is never a Bayes-Nash equilibrium since with common prior at least one of the players must choose A_i. However, suppose that player 1 believes $\omega = L$ with probability less than $1/4$ and $s_2 = D_2$ with probability 1, that player 2 believes $\omega = L$ with probability greater than $3/4$ and $s_1 = D_1$ with probability 1. Then for both players is rational to play D_i. Moreover the signals received by the players when they play (D_1, D_2) do not contradict their conjectures on the true state of nature. Therefore this situation is a (Conjectural) equilibrium of the IMG: each player is rational and there is no incentive to change behavior. Moreover the infinitely repeated strategies (D_1, D_2) are part of a Conjectural equilibrium also for the repeated IMG, even if there are many other Conjectural equilibria for positive discount factor. Note that in equilibrium the players agree to disagree on their subjective probability evaluations of the true state of nature (see Aumann [3]).

Now suppose that $x = \frac{8}{3} - \epsilon$ with $\epsilon = o(n^2)$ and that the players are interested in the outcomes of the IMG and not of the repeated game, i.e. consider a recurring game.

Assume that the players are uncertain about the opponent's mixed strategy λ_j and the true state of nature and that **they believe that the opponents' play corresponds to a sequence of i.i.d. random variables with a fixed but unknown probability distribution** $\lambda_j \in \Delta\{A_j, D_j\}$), i.e. the players believe that the probability law which regulates the opponent's behavior is stationary. Therefore using Bayesian inference the players' posterior has a density with respect to the prior, proportional to the likelihood function they use:

$$\frac{\lambda_j(A_j)^a(1 - \lambda_j(A_j))^{n-a}}{\int_0^1 \lambda_j(A_j)^a(1 - \lambda_j(A_j))^{n-a}\, \mu_i(d\lambda_j)} \tag{1}$$

where a is the number of times A_j has been observed by i in n observations, $\lambda_j(A_j) \in [0, 1]$ is the probability j gives to strategy A_j, and μ_i are i's beliefs on opponent's mixed strategy $\lambda_j \in \Delta(A_j, D_j)$. In particular suppose each player has the following subjective probability evaluations

$$\mu_i(\lambda_j(A_j) \leq p) = p \qquad \mu_1(L) = 3/4 \qquad \mu_2(L) = 1/4,$$

i.e. each player has an uniform distribution on opponents' mixed behavior (full support) and different prior on the true state of nature. From these uniform distribution we derive the following players' beliefs on opponents' pure strategies (the assessments using Fudenberg and Kreps [20] and [21] terminology):

$$\beta_i(A_j) = \int_0^1 \lambda_j(A_j)\, d\mu_j(\lambda_j) = \int_0^1 \lambda_j(A_j)\, d\lambda_j = 1/2.$$

Suppose that both players have a zero discount factor. Then both players will play A_i, which is the unique myopic best response to the above evaluations. Consequently they observe m^6 from which they infer the opponent's behavior but not the true state of nature. Hence they update $\mu_i(\lambda_j) \in \Delta(\{A_j, D_j\})$ using (1) and therefore the second period beliefs are such that

$$\beta_i(A_j) = 2/3 \qquad \mu_1(L) = 3/4 \qquad \mu_2(L) = 1/4,$$

Then both players will play D_i, which is the unique myopic best response to the above evaluations. Consequently they observe m^1 from which they infer the opponent's behavior but not the true state of nature. Then the third period beliefs are such that both players will play A_i, starting the above iteration again. Then for any $n \in \mathbf{N}$, the players' behavior is a perfectly correlated alternation between (A_1, A_2) and (D_1, D_2), such that the players' will never agree on the probability of the true state of nature. Two facts are worth noting. First, the players' strategies are perfectly correlated: even if they choose simultaneously and independently, the correlation may endogenously develop as result of correlated observations and because of imperfect monitoring it can asymptotically persist even if the agents play independently (see Lehrer 1991). Second, the players behavior does not correspond to a sequence of i.i.d. random variables, i.e. the players have a misspecified model of the stochastic law that regulates the stochastic process of the signals they are observing. Note that in the long run the finite sequence of observed signals $\{m^1, m^6, m^1, m^6, \cdots\}$ has negligible probability according to the supposed stochastic model with i.i.d. random draws. This notwithstanding the players do not change their model persisting in their foolish behavior.

In this imperfect monitoring game, a player may try to discover the opponent's behavior and the true state of nature through experimentation. In other words a player with a strictly positive discount factor can rationally give up some utility today in order to have useful information for utility maximization tomorrow (see section 5). Surely this behavior is extremely important in learning contexts, however there are very few works where rational players use active learning strategies to learn about the component game, since all the papers mentioned in the introduction consider strategic form games where by definition there is perfect monitoring and thus no role for experimentation. Two exceptions are Jackson and Kalai [34] and [35], however they limit their analysis to Bayesian games. Fudenberg and Kreps [20] and [21] and Fudenberg and Levine [23] consider experimentation in extensive form games, however they use bounded rational learning models.

3 The Model

3.1 The Stage Game

Let $\Delta(\cdot)$ be the set of all probability measure on $(\cdot)$ and $\otimes_{i\in N}\Delta(Y_i)$ is the set of all independent probability measure on $\times_{i\in N}Y_i$. The players repeatedly play a fixed stage game, described by an imperfect monitoring game.

Definition 1 *An* **Imperfect Monitoring Game** *(IMG) is defined as follows:*

$$G(\eta) := (N, S_i, u_i, \eta_i)$$

where:

- *N is the set of players,*
- *S_i is the set of player i's pure strategies; moreover $S := \times_{i\in N}S_i$, $\Lambda_i := \Delta(S_i)$ and $\Lambda := \otimes_{i\in N}\Lambda_i$;*
- *$u_i : \Delta(S) \to \mathbf{R}$ is player i's von Neumann-Morgenstern utility function, so that: $u_i(\mu) = E_\mu[u_i(s)], \quad \mu \in \Delta(S)$;*
- *$\eta_i : S \to M_i$ is player i's signal function: $\eta_i(\bar{s})$ is the signal privately received by i when the strategy profile $\bar{s}$ played. The information on opponents' behavior revealed by a signal m_i depends on the functional form of η_i and on the strategy played by i. This means that the players have an incentive to experiment to discover new information (active learning).*

As usual the subscript $-i$ denotes the js different from i and $(-i, i)$ a complete profile.

To simplify the analysis I make the following **structural assumptions**:

Assumption 1 *The analysis is restricted to the subclass of finite IMGs, that is the sets N, S_i, M_i are finite.*

Let $\rho_i[\mu] \in \Delta(M_i)$ be the probability measure induced on M_i by $\mu \in \Delta(S)$: then because of assumption 1 $\quad \rho_i[\mu](m_i) := \sum_{\{s|\eta_i(s)=m_i\}} \mu(s)$.

Assumption 2 *The signal is defined to contain all of the information player i receives about opponents' choices. Therefore*

$$\rho_i[s_i, \mu_{-i}] = \rho_i[s_i, \mu'_{-i}] \quad \Rightarrow \quad u_i(s_i, \mu_{-i}) = u_i(s_i, \mu'_{-i}).$$

This assumption means that each player receives her payoff after the stage game and that each player knows her own move.

3.2 The Repeated Game

To allow learning and to analyze asymptotic behavior, suppose that the IMG is played infinitely many times and that at the end of every period t each player i observes a stochastic outcome m_i^t, which is drawn from the finite set M_i according to the probability distribution $\rho_i[\mu]$, where $\mu \in \Delta(S)$ is the unknown players' behavior. To analyze such situation, I need to generalize the usual notions defined for repeated games.

A **history** for player i is an infinite sequence of elements of i's signals; the set of such sequences will be denoted by H_i. Formally:

$$H_i := \bigcup_{t=0}^{\infty} H_i^t \quad where$$

$$H_i^0 := \{h_i^0\}, \quad \text{and for any } t \geq 1 \;\; H_i^t := H_i^0 \times M_i^{(t)}$$

where the superscripts t and (t) denote period t and the t-fold Cartesian product of the sets respectively. Hence a history at time t for player i, h_i^t, is the private information received by player i in the periods up to t.

A pure superstrategy for player i specifies the strategy to select after each possible history. Therefore the **set of pure superstrategies** for player i, F_i, is so defined: $f_i \in F_i$ if and only if $f_i := \{f_i^t(h_i^{t-1})\}_{t=1}^{\infty}$ with $f_i^t : H_i^{t-1} \to S_i$. Define F_i^t as the set of player i times t superstrategies: $F_i^t := \{f_i^t | f_i^t : H_i^{t-1} \to S_i\}$. Thus $F_i = \times_{t \in \mathbf{N}} F_i^t$, or $f_i \in F_i$ if and only if $f_i : H_i \to S_i$. The definition of mixed and behavior strategies in this context requires some care, as was first noticed by Aumann [2]. In fact the set of pure superstrategies, being the Cartesian product of countably many copies of S_i, has the cardinality of the continuum (see Kolmogorov and Fomin [44]). If we think of a mixed superstrategy as a random device for choosing a pure superstrategy, then the following construction follows immediately (see Aumann [2]): a mixed superstrategy is a random variable from a sample space into the space of pure strategies. Consider an abstract probability space $(\Omega, \mathcal{A}, \gamma)$, then player i's **set of mixed superstrategies** Φ_i is defined as follows: $\phi_i \in \Phi_i$ if and only if $\phi_i = (\phi_i^t)_{t=1}^{\infty}$ with $\phi_i^t : \Omega \times H_i^{t-1} \to S_i$ where $\phi_i^t(\cdot, h_i^{t-1}) : \Omega \to S_i$ is measurable. The **set of behavior superstrategies** for a player i, B_i, is similarly defined, asking however for an additional restriction: $b_i \in B_i$ if and only if $b_i = (b_i^t)_{t=1}^{\infty}$ with $b_i^t : \Omega \times H_i^{t-1} \to S_i$ where $b_i^t(\cdot, h_i^{t-1}) : \Omega \to S_i$ is measurable and $b_i^t(\cdot, h_i^{t-1})$, $b_i^{\tau}(\cdot, h_i^{\tau-1})$ are mutually independent random variables $\forall t \neq \tau$, $\forall h_i^t, h_i^{\tau}$. Therefore $B_i \subset \Phi_i$.

The **outcome at time t for player i**, $O_i^t(f)$, is the message received by i at t as a function of the strategies played at t according to the superstrategy profile f. The definition is inductive: $O_i^0(f) := h_i^0$ and

$$\forall t \geq 1 \;\; O_i^t(f) := \eta_i[f^t(O^0(f), \cdots, O^{t-1}(f))] \in M_i.$$

The sequence of outcomes for player i up to period t determines the **outcome path at time t for player i**, $P_i^t(f)$:

$$\forall f \in F, \quad \forall t \geq 0 \quad P_i^t(f) := \{O_i^\tau(f)\}_{\tau=0}^t \in H_i^0 \times M_i^{(t)}.$$

Finally, the infinite sequence of private histories faced by i defines the **outcome path for player** i:

$$P_i(f) := \{O_i^t(f)\}_{t=0}^\infty \in M_i^{(\infty)} := H_i^0 \times M_i^1 \times M_i^2 \times \cdots .$$

When f is omitted, these expressions should be interpreted as "realization" of the mapping considered. For example $P_i \in M_i^{(\infty)}$ is a realization of the outcome path $P_i(f)$.

The **payoff function** $U_i : \Delta(F) \to \mathbf{R}$ for this repeated game is:

$$U_i(x) := E_x \sum_{t=1}^{\infty} \delta_i^t u_i(f^t(P^{t-1}(f))) \qquad for \;\; x \in \Delta(F) \quad \text{and} \quad \delta_i \in [0,1).$$

Summing up, a **Repeated Imperfect Monitoring Game** (RIMG) is defined by

$$G^\infty(\delta, \eta) = (N, F_i, U_i).$$

Now suppose that the players' horizon time is finite and denoted by T. Then the previous notation should be changed substituting T for ∞, and $G^T(\delta,\eta) := (F_{i,T}, U_{i,T}, N)$ denotes the finitely repeated IMG. In the following analysis I will also study players' beliefs and behavior as T goes to infinity. To this aim it is useful (see lemma 2 in the appendix) to consider an increasing sequence of finitely repeated IMGs, obtained increasing the horizon time T: $G^T(\delta,\eta) \subseteq G^{T+1}(\delta,\eta) \subseteq \cdots \subseteq G^\infty(\delta,\eta)$. To obtain this increasing sequence, I make the following assumption:

Assumption 3 *There exists a null element, denoted by $\emptyset$, which is added to the elements previously defined to obtain: $F_{i,T} \subseteq F_i$, and $U_{i,T} = U_i$. This means that $\emptyset \in M_i$, $\emptyset \in S_i$, $\eta_i(\emptyset) = \emptyset$, that the generic element of the superstrategies sets becomes: $f_{i,T} = (f_i^0, \cdots, f_i^T, \emptyset, \emptyset, \cdots) \in F_i$ and that $u_i(\emptyset) = 0$.*

Remark: assumption 3 obviously implies $\Phi_{i,T} \subseteq \Phi_i$ and $B_{i,T} \subseteq B_i$

Before further constructions, I sum up the notation introduced so far.

NOTATION	
Expression	**Meaning**
N	set of players
S_i	set of player i's pure strategies
$\Delta(\cdot)$	set of probability measures on $\cdot$
$u_i : \Delta(S) \to \mathbf{R}$	player i' utility function
$\eta_i : S \to M_i$	player i' signal function
$H_i = \cup_{t=0}^\infty H_i^t$	set of possible histories for player i
$F_i = \times_{t \in \mathbf{N}} F_i^t$	set of possible superstrategies for player i
$\Phi_i, \quad B_i$	set of i's mixed, behavioral superstrategies
$O_i^t(f), \quad P_i^t(f)$	outcome, outcome path at time t for player i
$U_i : \Delta(F) \to \mathbf{R}$	i's utility function for the repeated game.

3.3 The Stochastic Dynamics of Beliefs and Behavior

Let $(F, \mathcal{F}, x)$ be a probability space , where F is the set of pure superstrategies, $\mathcal{F}$ the Borel σ-algebra of F and $x \in \Delta(F)$ a generic probability measure on F. Assumption 1 and the Tychonov product theorem (see e.g. Kuratowski [46]) imply that F is a compact metric space in the product topology and thus $\Delta(F)$ is a compact metric space if endowed with the weak topology and with the metric being the Prohorov metric (see Billingsley [8]). Then consider the probability space $(M_i^{(\infty)}, \mathcal{H}_i, \mathbf{P}_x^i)$. The construction of this probability space involves some steps. Let $P_i \in M_i^{(\infty)}$ be a possible outcome path for player i and define for each $t \in \mathbf{N}$ a mapping $Z_t : M_i^{(\infty)} \to M_i$ such that $Z_t(P_i) := O_i^t$, that is Z_t is the projection of P_i on its t element. Consider the class $\mathcal{H}_i^C$ consisting of the cylinders, that is of the sets of the form $\{P_i \in M_i^{(\infty)} | (Z_{t_1}(P_i), \cdots, Z_{t_k}(P_i)) \in C\}$, where k is an integer, $(t_1, \cdots, t_k)$ is a k-tuple in $\mathbf{N}$ and C belongs to the Borel σ-algebra generated by $M_i^{(k)}$. Then it is possible to prove (see e.g. Billingsley [9]) that $\mathcal{H}_i^C$ is a field such that $\mathcal{H}_i$ is the σ-field generated by it. Therefore, since the Z_t are measurable functions on $(M_i^{(\infty)}, \mathcal{H}_i)$, if $\mathbf{P}$ is a probability measure on $\mathcal{H}_i$, then $\{Z_t\}_{t \in \mathbf{N}}$ is a stochastic process on $(M_i^{(\infty)}, \mathcal{H}_i, \mathbf{P})$. Now consider the probability distribution inductively defined according to the following rules: $\mathbf{P}_x^{i,(0)}(h_i^0) = 1$ and

$$\forall t \geq 1 \quad \mathbf{P}_x^{i,(t)}(h_i^{t-1}, m_i) = \mathbf{P}_x^{i,(t-1)}(h_i^{t-1}) \times [\int_{\{f | P_i^t(f_i, f_{-i}) = h_i^{t-1}, m_i\}} x(df)].$$

It is immediate to check that $\mathbf{P}_x^{i,(t_1)}, \cdots, \mathbf{P}_x^{i,(t_k)}$ are a system of probability distributions satisfying the Kolmogorov's consistency conditions. Therefore there exists a probability measure $\mathbf{P}_x^i$ on $\mathcal{H}_i$ such that the stochastic process $\{Z_t\}_{t \in \mathbf{N}}$ on $(M_i^{(\infty)}, \mathcal{H}_i, \mathbf{P}_x^i)$ has the $\mathbf{P}_x^{i,(t_1)}, \cdots, \mathbf{P}_x^{i,(t_k)}$ as its finite-dimensional distributions.

Note that by definition of outcome path to calculate $\mathbf{P}_x^{i,(t)}$ player i should know the IMG she is playing: the superstrategies actually played by $-i$ depend on the private information they receive during the play, which in turn is a "structural" function of the superstrategies played. If i is uncertain about the game she is playing, then $\mathbf{P}_x^{i,(t)}$ can be thought of as a conditional probability given G, and a probability measure on $\mathcal{G}$, the set of all IMGs satisfying assumptions 1 and 2, can be introduced. Because of assumption 1, for a fixed number of players N and of pure strategies S, $\mathcal{G} = \mathbf{R}^{N \times S} \times \mathbf{R}^{N \times S}$, where for $(y', y'') \in \mathcal{G}$, $y'(i, s)$ is the payoff to player i under strategy profile s and $y''(i, s)$ is the signal to player i under strategy profile s. Therefore I can consider a probability measure $p_i \in \Delta(\mathcal{G})$ by referring to the probability space $(\mathbf{R}^n, \mathcal{B}^n, p_i)$, where $\mathcal{B}^n$ is the family of Borel sets of $\mathbf{R}^n$. Then I can derive the marginal probability of h_i^t, given p_i: $\mathbf{P}_x^{i,(t)}(h_i^t \mid p_i) = \int_{\mathcal{G}} \mathbf{P}_x^{i,(t)}(h_i^t \mid G)\, p_i(dG)$. Therefore $\mathbf{P}_x^{i,(t)} \in \Delta(H_i^t)$ includes both the structural uncertainty about the imperfect monitoring game i is playing, and the strategic uncertainty about opponents' behavior. In what follows I will omit p_i since I intend to focus on strategic uncertainty, but the analysis could trivially include structural uncertainty at the cost of further complicating the notation.

Summing up, the **objective situation as faced by player** i is described by a stochastic process $\{O_i^t(f)\}_{t=0}^{\infty}$ with probability law $\mathbf{P}_x^i$.

Now, consider the **subjective situation of player** i, given this stochastic environment. Player i wish to maximize U_i w.r.t. $\phi_i \in \Phi_i$, but she is uncertain about opponents' behavior $\phi_{-i} \in \Phi_{-i}$. To maximize utility the uncertainty relative to the opponents' random behavior ϕ_{-i} is equivalent to the uncertainty about opponents' pure superstrategies: see the example of section 2 and Pearce [57] lemma 2. Therefore from i's point of view the set of the states of the word is represented by F_{-i} and thus a Bayesian player i is endowed with a prior belief $\xi_i \in \Delta(F_{-i})$, where the probability space $(F_{-i}, \mathcal{F}_{-i}, \xi_i)$ is constructed deriving the marginal distributions from $(F, \mathcal{F}, x)$. This subjective assessment may exhibit correlation, but this does not contradict the fact that the actual strategy choices are independent. This correlation is due to i's uncertainty: even if i believes that the opponents choose their strategies independently, she may feel that they have common characteristics which partially resolve the strategic uncertainty. Moreover as shown in example 2, the correlation may endogenously develop as the result of correlated observations and because of imperfect monitoring it can asymptotically persist even if the agents play independently (see Lehrer 1991 and the example of section 2).

Consider the information player i collects by playing. Her beliefs are updated at time t using this information, i.e.

$$P_i^t(f_{-i}, f_i). \tag{2}$$

Note that this information depends on f_i, i.e. on i's behavior. Moreover consider f_{-i}: even the opponents' strategic choices depend on player i superstrategy since

$$f_{-i} \;=\; \{f_{-i}^t(h_{-i}^{t-1})\}_{t=1}^{\infty} \;=\; \{f_{-i}^t(P_{-i}^t(f_{-i}, f_i))\}_{t=1}^{\infty}. \tag{3}$$

As a consequence of expressions (2) and (3), **player i's beliefs depend on** f_i for two different reasons:

1. f_i takes part in determining the information that i receives at each stage, i.e. P_i^t is a function of f_i, as shown by expression (2). This aspect regards the **"informative links between periods"**, that generate the possibility of experimentation, i.e. of active learning behavior;

2. f_i takes part in determining the information that i's opponents receive at each stage, i.e. P_{-i}^t is a function of f_i, as shown by expression (3). I will refer to the second aspect using the label **"strategic links between periods"** since it is connected to players' behavior in repeated games.

During the play, i is refining her information about opponents' behavior (passive learning), but the actual amount of information obtained depends on the superstrategy followed (active learning). For a fixed f_i construct the natural filtration of the stochastic process given by the outcome path:

$$\mathcal{F}^t_{-i}(f_i) := \sigma(P^t_i(f_{-i}, f_i)),$$

where $\sigma(X(\omega))$ denotes the σ-algebra generated by the random variable $X(\omega)$. Intuitively $\mathcal{F}^t_{-i}(f_i)$ represents all the possible information about opponents' superstrategy that i could collect at t following the dynamic superstrategy f_i. In fact $\sigma(X(\omega))$ consists precisely of those events A for which, for each and every ω, player i can decide whether or not A has occurred, i.e. whether or not $\omega \in A$, on the basis of the observed value of the random variable X. Formally a filtration $\{\mathcal{F}^t_{-i}\}$ is an increasing sequence of sub-σ-algebras of $\mathcal{F}_{-i}$, i.e. $\mathcal{F}^1_{-i}(f_i) \subseteq \mathcal{F}^2_{-i}(f_i) \subseteq \cdots \subseteq \mathcal{F}_{-i}$, and the natural filtration of a stochastic process $\{O^t_i(f_i)\}_t$ is the filtration generated by it in the sense that $\mathcal{F}^t_{-i}(f_i) := \sigma(O^0_i, \cdots, O^t_i)$. Finally, define $\mathcal{F}^\infty_{-i}(f_i) := \sigma(\bigcup_{t\in\mathbf{N}} \mathcal{F}^t_{-i}(f_i)) \subseteq \mathcal{F}_{-i}$. Then, for a fixed f_i, $P_i(f_{-i}, f_i) := \{O^t_i(f_{-i}, f_i)\}_{t\in\mathbf{N}}$ is a stochastic process adapted to the natural filtration $\{\mathcal{F}^t_{-i}(f_i)\}$, because by definition $O^t_i(f_{-i}, f_i)$ is $\mathcal{F}^t_{-i}(f_i)$-measurable. Therefore for every t and for every $A \in \mathcal{F}_{-i}$ there exists a version of the conditional expectation $E[\chi_A(f_{-i})|\mathcal{F}^t_{-i}(f_i)]$, where χ_A is the indicator function for the set A. Indicate such a version with $\xi^t_i[f_i](A)$; then $\xi^t_i[f_i] \in \Delta(F_{-i})$ is a regular conditional probability distribution (Theorem 8.1 of Parthasarathy [56]). As the notation stress, *such a probability measure depends on* f_i. This probability measure represents the updated beliefs of player i at time t, given that she is following the superstrategy f_i.

The discussion of this section can be summmarized in the following assumption, where $SUPP[\cdot]$ denotes the support of a probability measure:

Assumption 4 Bayesian rationality: *In the RIMG*

1. *every player* $i \in N$ *chooses* $\phi_i \in \Phi_i$ *if and only if*

$$\exists \xi_i \in \Delta(F_{-i}): \ \ \forall f_i \in SUPP[\phi_i] \quad f_i \in arg\,max_{f'_i \in F_i} E_{\xi_i} U_i(f'_i, f_{-i}).$$

2. *every player* $i \in N$ *updates her beliefs* ξ_i *according to the following expression:*

$$\forall f_i \in F_i, \quad \forall A \in \mathcal{F}_{-i}, \quad \forall t \in \mathbf{N} \qquad \xi^t_i[f_i](A) = E[\chi_A(f_{-i})|\mathcal{F}^t_{-i}(f_i)].$$

Remark: part 1 of assumption 4 is meaningful since U_i is continuous and F_i compact, part 2 because of the existence of a regular conditional probability.

4 Convergence to Conjectural Equilibria of the Repeated Imperfect Monitoring Game

The result presented in this section is a simple extension to RIMG of Kalai and Lehrer [40] main result, and it is included for completeness.

The notion of Conjectural equilibrium (CE) is meant to model a situation where rational players have learnt everything they can learn given their access to information in equilibrium. Rational players could be wrong in their conjectures on opponents' behavior because of lack of observations, but each player's conjecture on what is observable is correct and, although they have private information and private priors, players' behavior and conjectures are mutually consistent in the sense that what actually happen is coherent with each player's experience (see the example). A comprehensive discussion of this concept is in Gilli [29]. Now consider the specification of this general notion for RIMGs.

Definition 2 *A* **Conjectural Equilibrium** *for a RIMG $G^\infty(\delta,\eta)$ is a superstrategy profile $\hat{\phi} \in \Phi$ such that for each player $i \in N$, there exists a belief $\xi_i \in \Delta(F_{-i})$ such that for any superstrategy in the support of $\hat{\phi}_i$ the following two conditions hold*

$$\hat{f}_i \in arg\ max_{f_i \in F_i}\ E_{\xi_i}\ U_i(f_i, f_{-i}) \tag{4}$$

$$\mathbf{P}^i_{(\xi_i,\hat{f}_i)} = \mathbf{P}^i_{(\hat{\phi}_{-i},\hat{f}_i)} \tag{5}$$

where (x,y) denotes the product of the probability measures x and y.

Remarks:

1. (4) is a condition of dynamic Bayesian rational behavior, therefore very different forms of behavior are coherent with this condition: learning and teaching, rewarding and punishing, building reputation, and so on;

2. (5) is the equilibrium condition: each player has perfect foresight of the stochastic process of her own signals as induced by the players' equilibrium behavior, while nothing is assumed about forecasts of out-of-equilibrium signals;

3. a Nash equilibrium is a particular case of CE where there is perfect foresight of opponents' behavior in and out-of-equilibrium;

4. CE in repeated games are called Subjective equilibria (Kalai and Lehrer [41]).

Given two probability measures α and β on the same σ-field $\mathcal{H}$, α is absolutely continuous with respect to β if and only if $\forall A \in \mathcal{H} \quad \beta(A) = 0 \Rightarrow \alpha(A) = 0$. Moreover let $||\cdot||$ be the norm of the sup in the space of all probability measures on the same σ-algebra $\mathcal{H}_i$: $||\mathbf{P}^i_1 - \mathbf{P}^i_2|| := sup_{P_i \in \mathcal{H}_i} |\mathbf{P}^i_1(P_i) - \mathbf{P}^i_2(P_i)|$.

Theorem 1 *Let ϕ and $(\xi_1, \cdots, \xi_N)$ be vectors of strategies representing the actual choice and the beliefs of the players. Suppose that for every player i, ϕ is absolutely continuous with respect to ξ_i. Then players' behavior will converge to a Conjectural Equilibrium of the RIMG, in the sense that for every player i*

$$||\mathbf{P}^i_{(\xi_i,f_i)}(\cdot|P^t_i) - \mathbf{P}^i_{(\phi_{-i},f_i)}(\cdot|P^t_i)|| \to_{t\to\infty} 0 \quad \mathbf{P}^i_{(\phi_{-i},f_i)} - a.e.; \tag{6}$$

PROOF: it follows from Blackwell and Dubins [10] main result and it is a version of Kalai and Lehrer [40] main result for this more general setting: the details are in Gilli [31]. ♡

Remark: the hypothesis of absolute continuity is quite strong, but it is a condition of **rational learning**, since when not satisfied players' subjective view of their stochastic environment is misspecified. In Kalai and Lehrer [42] there is a comprehensive discussion of this assumption in connection with results about merging. Recently there have been many contributions explaining why this assumption is restrictive: see Miller and Sanchirico [47], Nachbar [48] and [49], Nyarko [55], Sandroni [58], Young and Foster [65]. The problem is that in this context absolute continuity is endogenous since we are asking for a property which connects beliefs and behavior, which in turn are connected because of optimization: there is kind of fixed point argument involved in the hypothesis of absolute continuity of theorem 1, as expected in an assumption of rational learning.

5 The Construction of a Recurring Game

In this section I want to study the inferential problem faced by a player who wishes to learn opponents' behavior **in the stage game**. To this aim, I need to sterilize the strategic links that characterize the repeated games. This is obtained constructing a **recurring game** (RG), a type of repeated interaction involving different players at each time. We shall think of the stage game as having N roles, and for each role $i \in N$ ther exists a population C_i of agents, who are eligible to play that role (see Jackson and Kalai [34] and Young [64]).

In this paper to construct a RG, I consider a specific model of anonymous repeated interaction with random matching adapted from Fudenberg and Levine [23], but the framework can easily be used to study similar models. In this section I consider a **finite** players' horizon T and consequently I truncate the repeated game at T. The reason to assume a finite T will become apparent in the following pages: if T is infinite, then proposition 2 does not hold and it would be impossible to specify a rational learning model for recurring games. In this section T will be omitted to avoid further complications in notation, but the reader must be aware that in this section every variable depends on the length T of players' horizon. By contrast note that t refers to a period t in the life of the player, i.e. $t \in \{0, \cdots, T\}$.

To study the convergence, if any, to equilibria of the stage game we need to explain why the strategic links are absent. Consider a RIMG where a population with a continuum of identical agents with total mass one is associate to each player. In every period there is a random matching of the agents randomly drawn from different populations. Each agent recalls what happened in her previous encounters without knowing anything about the experiences of her current opponents. Thus within each population we can distinguish different types according to their personal past history. Because of these hypotheses the probability to be matched with the same opponents, and therefore the possibility of influencing their behavior, is negligible and this very unlikely event is not detectable. Therefore there are no strategic links, while the informative links are still present. Let $G(\eta)$ be the component game.

The precise characteristics of the model (the matching model and proposition 2 are borrowed from Fudenberg and Levine [23]) are the following:

1. A population with a continuum of agents with mass equal to 1 corresponds to each player of $G(\eta)$, that is each population is isomorphic to $[0, 1]$;

2. there exists a double infinite sequence of periods: $\cdots, -1, 0, 1, \cdots$;

3. in each time in every population there are $1/T$ new agents, in each generation there are $1/T$ agents while $1/T$ agents of age T die: every agent lives T periods and the mass of population is stationary;

4. in every period each agent is randomly and independently extracted from her population, and with probability $1/T$ is matched with agents of age t.

5. each agent of population i in every period observes a private signal m_i, which is a function of the strategy profile played in that particular matching, i.e. $m_i = \eta_i(s)$;

6. the agents' behavior can be described by the following function:

$$\forall i \in N \quad \forall j_i \in [0, 1] \qquad f_{j_i} : H_{j_i} \to S_i,$$

   where j_i denotes the agent $j \in [0, 1]$ of population $i \in N$. The agents choose a pure dynamic strategy, if more than one superstrategy is optimal, then one is selected according to some unspecified mechanism. This notwithstanding the average population's behavior is described by a mixed superstrategy.

In this model the players rationally ignore the strategic links between periods, so that the only temporal connections are the informative links. Consequently the superstrategy followed by i's opponents $f_{-i} = \{f_{-i}^t\}_{t=1}^T$ is seen by i as a sequence of the stage game strategies, chosen by players acting on the basis of information that i can not influence. Therefore from i's point of view $f_{-i} = \{s_{-i}^t\}_{t=1}^T$. In this RG the behavior of player i's opponents can be represented as the following sequence:

$$f_{j_{-i}} = \{s_{j_{-i}}^t\}_{t=0}^T, \tag{7}$$

where $j_{-i} := [j_k]_{k \neq i}$, with $j_k \in [0, 1]$ and $k \in N \setminus \{i\}$. (An agent of) player i knows that the sequence of opponents' strategies is a function of their specific past experience, but she can not distinguish the history characterising opponents' types, because of the hypotheses on the information the agents have when they are playing. Otherwise player i would be interested in $h_{j_{-i}}$ also, and thus in opponents' superstrategies. In other words, player i's strategic uncertainty regards the opponents' choice of a superstrategy, that is the choice of a $f_{j_k} \in F_k$ for all $k \neq i$. But i can not distinguish among different agents in the same population: i's actual opponents are the $-i$'s populations and thus i's uncertainty should regard the average behavior of opponents' populations, because it is the opponents' **average** behavior which is payoff relevant. Averaging among $j_k \in [0,1]$ for each $k \neq i$, the sequence (7) of opponents' strategic choices becomes $\{\lambda_{-i}^t\}_{t=0}^T \in \Lambda_{-i}^{(T+1)}$.

Therefore in this setting the set of the states of the word for player i is $\Lambda_{-i}^{(T+1)}$: player i faces the sequence of realizations $\{s_{-i}^t\}_{t=0}^T$, that she imperfectly observes through the signal function η_i, and from the succession of signals $\{m_i^t\}_{t=0}^T$ she want to derive information on the true state of the world, i.e. on the true unknown probability law $\{\lambda_{-i}^t\}_{t=0}^T \in \Lambda_{-i}^{(T+1)}$

governing opponents' population behavior. But, and this is the crucial problem, this probability law is **non stationary**, because in general λ_{-i}^t depends on t. This fact poses a very serious inferential problem to the player facing the time series of the signals trying to predict the future behavior of the stochastic process $\{s_{-i}^t\}_{t=0}^T$. Eventually the problem could be unsolvable, since there could be too few observations: the true unknown λ_{-i}^t depends on t, but any agent of player i has only a single observation for each period and if there are no links between past and future the problem is statistically unsolvable. In other words there would be only $T+1$ observations to infer $T+1$ stochastic variables. This is a very serious problem for rational learning models which use correctly specified statistical models: if the likelihood functions are not stationary, then there is no correct way to solve the statistical inferential problem and thus there exists no rational learning model. To allow the players to solve this inferential problem either we introduce some intertemporal restriction on the time series in order to reduce its time heterogeneity or we give up the construction of rational learning models considering misspecified but plausible learning processes, as usually done. Using Jackson et al. [36] terminology, rational learning in RG requires a representation of the stochastic process of players' behavior as a learnable pattern, even if not necessarily stationary. To obtain this result, I borrow the use of steady state assumptions from Fudenberg and Levine [23]. Proposition 1 shows that the assumption of a steady state in the updating process of each population's average history is sufficient to generate a stochastic process which is extremely well behaved: the succession of opponents' strategies faced by i becomes an exchangeable sequence of random variables, and thus De Finetti theorem applies and the stochastic process of opponents' behavior has an unknown but time invariant probability law $\lambda_{-i} \in \Lambda_{-i}$.

At each time t it is possible to associate to each history h_k the percentage of agents in populations k having experienced that particularly history. If we indicate such percentages by $\theta_k^t(h_k)$, with $\theta_k^t \in \Delta(H_k)$, this gives the proportion of types h_k^t at period t in population $k \in N$. Note that $\theta^t \in \otimes_{i \in N} \Delta(H_i)$ depends on the ϕ played and thus I will write $\theta^t[\phi]$.

Proposition 1 *For every $\phi \in \Phi$, if $\theta^t[\phi]$ is independent from t, then the true λ^t is also independent of t.*

PROOF: see the appendix. ♡

Consider the dynamic law $d : \otimes_{i \in N} \Delta(H_i) \to \otimes_{i \in N} \Delta(H_i)$, which regulates the updating of the average experience in the populations for a given super-strategy profile ϕ_T, where $d[\theta]_i(h_i)$ indicates the fraction of population i with history h_i after the matching, given that the populations' average history was θ: $d[\theta]_i(h_i^0) = 1/T$

$$d[\theta]_i(h_i, \eta_i(s_i, s_{-i})) = 0, \quad \text{if} \quad \forall f_i \in SUPP[\phi_i] \qquad s_i \neq f_i(h_i)$$

$$d[\theta]_i(h_i, m_i) = \theta_i(h_i) \sum_{\{f_i | f_i(h_i) = s_i\}} \phi_i(f_i) \sum_{\{s_{-i} | \eta_i(f_i(h_i), s_{-i}) = m_i\}} [\prod_{k \neq i} \theta'_k(s_k)] \quad \text{otherwise,}$$

where $\theta'_k(s_k)$ is defined according to equation (8). Note that a steady state of the updating process of populations' average experience is a fixed point of d. Now it is possible to conclude that such a steady state exists.

Proposition 2 *In RG, for every finite T and for every $\phi_T \in \Phi_T$, there exists a steady state $\bar{\theta}_T \in \Delta(H_T)$ of the updating process of the proportion of private experiences, and therefore a corresponding stationary $\bar{\theta}'_T[\phi_T] \in \Lambda$.*

PROOF: see the appendix. ♡

Remark: this result requires T finite, otherwise the fixed point does not exist; for this reason I should first consider T finite, and then take the limit for $T \to \infty$ before of considering i's asymptotic behavior.

Now consider **player i's conditional beliefs** at time t, $\xi_i^t[f_i] : H_i^t \to \Delta(F_{-i})$. According to lemma 1 (see the appendix) it is possible to state that for every player's beliefs $\xi_i^t[f_i]$ there exists a sequence $\{\beta_i^{t,\tau}[f_i]\}_{\tau=0}^T$ such that

$$\forall t, \quad \forall \tau \quad \beta_i^{t,\tau}[f_i] : H_i^t \times H_{j_{-i}}^\tau \to \Delta(S_{j_{-i}}). \tag{8}$$

Averaging as j_k takes values in $[0,1]$ and knowing that the agents play independently, equation (9) becomes

$$\forall t \quad \forall \tau \quad \beta_i^{t,\tau}[f_i] : H_i^t \times \Delta(H_{-i}^\tau) \to \Delta(\Lambda_{-i}).$$

This means that at time t player i's forecasting of opponents' behavior at τ depends on their average experience at τ. Now if we consider a steady state in the space $\Delta(H_{-i})$, i.e. a constant $\bar{\theta}_{-i} \in \Delta(H_{-i})$, *and the players are aware of it*, then it is possible to model player i's beliefs as a sequence of probability measure on opponents mixed strategies, conditional to i's information :

$$\mu_i^t[f_i] : H_i^t \to \Delta(\Lambda_{-i}), \quad where \ \forall \tau \ \mu_i^t[f_i] := \beta_i^{t,\tau}[f_i](\bar{\theta}_{-i}).$$

This means that in a steady state the average behavior of a population at time τ does not depend on its average history at that particular time, and thus when the players are aware of this their beliefs on this average behavior do not depend on opponents' experiences but only on personal information. Note that even in this stationary case the informative links between periods do not disappear.

This discussion can be summed up in the following assumption, which according to proposition 2 is meaningful:

Assumption 5 Rational learning: *in the recurring game the players solve their inferential problem about other populations' strategic choice believing to be, and indeed being, in a steady state $\bar{\theta}_T \in \Delta(H_T)$. Thus player i's beliefs are modelled through the following sequence of conditional probability measures*

$$\mu_i^t[f_i] : H_i^t \to \Delta(\Lambda_{-i}), \quad where \quad \forall \tau \;\; \mu_i^t[f_i] := \beta_{-i}^{t,\tau}[f_i](\bar{\theta}_{-i}).$$

These beliefs are updated using the correct likelihood function, i.e. the multinomial distribution.

Remark: Assumption 5 is the hypothesis of rational learning for the recurring game. In particular according to this assumption the players believe to face, and indeed are facing, a fixed time invariant probability distribution of opponents' strategies $\bar{\theta}'_{-i,T} \in \Lambda_{-i}$, being uncertain of what the true distribution is. In other words this assumption guarantees that the stochastic process of signals observed by each player is exchangeable and thus learnable: the players apply De Finetti representation theorem to make statistical inferences. Note that this assumption actually has two parts: the first states that the stochastic process of opponents' strategies is stationary, the second that the players are aware of this stationarity. Both parts are necessary to define players beliefs as $\mu_i^t \in \Delta(\Lambda_{-i})$ and to update them using the multinomial distribution.

6 Convergence to Conjectural Equilibria of the Imperfect Monitoring Game

To study players' asymptotic behavior I should let players' lifetime to go to infinity, and then consider players' behavior in the last periods of their life. The analysis is complicated by the fact that the existence of a steady state - and thus of a well defined inferential problem - depends on T finite, while the analysis of players' asymptotic behavior requires a T big enough; this is the reason to consider a double limit: first $T \to \infty$, then $t \to \infty$. This approach is justified by lemma 2 in the appendix, which uses the fact that by assumption 3 $\{G^T(\delta, \eta)\}_T$ is an increasing sequence of games or, using Fudenberg and Levine [22] definition 3.2, that $G^T(\delta, \eta)$ is a restriction of $G^\infty(\delta, \eta)$. Lemma 2 implies that the limit game and the infinitely repeated game are "equivalent". As a consequence I can identify $G^\infty(\delta, \eta)$ with the limit game for $T \to \infty$ and I will omit the index ∞ to mean a limit point for $T \to \infty$. Therefore when I speak of a limit point of a sequence, I mean as $t \to \infty$, given that T has gone to infinity.

At time t player i has observed t signals $m_i^1, \cdots, m_i^t$, which is a partial realization of the stochastic process $\{\eta_i(s_{-i}^t, s_i^t)\}_{t=0}^T$, that depends on a fixed time invariant probability distribution of opponents' strategies. The aim of the following constructions is to reduce this complex inferential problem to a simpler standard statistical decision problem with a set of possible states of nature, a set of actions, an observable random variable and a well defined ordering on the set of possible outcomes. In what follows I consider each probability space endowed with the topology of the weak convergence in measure. Let construct in the usual way the probability space $(\Lambda_{-i}, \mathcal{B}_{-i}, \mu_i)$ where $\mathcal{B}_{-i}$ is the Borel σ-field generated by Λ_{-i}. Player i at each period t observes a signal which can be thought of as the realization of the stochastic process formed by the following sequence of random variable defined on an abstract measurable space $(\Omega, \mathcal{A})$: $\hat{m}_i^t : \Omega \times S_i \to M_i$, where $\hat{m}_i^t(\omega, s_i)$ is player i's signal received at t when she plays s_i. The probability law of this stochastic process is determined by the following random variable:

$$\rho_i : \Lambda_{-i} \times S_i \to \Delta(M_i), \quad i.e. \quad \mathbf{P}(\hat{m}_i^t = m_i) = \rho_i(m_i | s_i, \lambda_{-i}).$$

If we fix player i's strategy s_i, then the following stochastic process is obtained: $\hat{m}_i^t[s_i] : \Omega \to M_i$ with probability law determined by the following random variable: $\hat{\rho}_i : \Lambda_{-i} \to \Delta(M_i)$, where $\hat{\rho}_i$ is defined as follows:

$$\forall \lambda_{-i} \in \Lambda_{-i} \quad \forall m_i \in M_i \qquad \hat{\rho}_i(\hat{m}_i^t[s_i] = m_i | \lambda_{-i}) = \sum_{\{s_{-i} | \eta_i(s_i, s_{-i}) = m_i\}} \lambda_{-i}(s_{-i}).$$

Therefore $\Delta(M_i)$ is the "parameter space", such that for every $\hat{\rho}_i \in \Delta(M_i)$ there is a probability measure $\mathbf{P}_{\hat{\rho}_i}$ on $\mathcal{A}$ under which the $\{\hat{m}_i^t\}$ are independent with common distribution $\mathbf{P}_{\hat{\rho}_i}(\{\omega \in \Omega | \hat{m}_i^t[s_i](\omega) = m_i\}) = \hat{\rho}_i(m_i | \lambda_{-i})$. Now construct the probability space $(\Delta(M_i), \mathcal{S}_{-i}, \hat{\mu}_i)$ where $\mathcal{S}_{-i}$ is the Borel σ-field generated by $\Delta(M_i)$ and $\forall A \in \mathcal{S}_{-i} \quad \hat{\mu}_i(A) := \mu_i(\{\lambda_{-i} | \hat{\rho}_i(m_i[s_i] | \lambda_{-i}) \in A\})$. Lemma 3 in the appendix shows that $\hat{\mu}_i \in \Delta(M_i)$ is a well defined probability measure. Define $\mathbf{P}_{\hat{\mu}_i}$ as the probability measure induced on the sample path by the prior $\hat{\mu}_i \in \Delta(M_i)$ such that $\forall A \in \mathcal{A} \quad \mathbf{P}_{\hat{\mu}_i}(A) = \int_{\Delta(M_i)} \mathbf{P}_{\hat{\rho}_i}(A)\, \hat{\mu}_i(d\hat{\rho}_i)$. Finally to construct a statistical decision problem I need a well defined ordering on the set of possible outcomes. According to assumption 2 the signal contain all of the relevant information, hence the problem of player i's optimal behavior can be addressed considering just her uncertainty on $\Delta(M_i)$. For any $\alpha_{-i} \in \Delta(S_{-i})$, define $v_i(s_i, \rho_i(\cdot | s_i, \alpha_{-i})) := u_i(s_i, \alpha_{-i})$. Thus $v_i : S_i \times \Delta(M_i) \to \mathbf{R}$, and this payoff function is well defined since assumption 2 guarantees that $u_i(s_i', \alpha_{-i}') \neq u_i(s_i'', \alpha_{-i}'') \quad \Rightarrow \quad \rho_i(\cdot | s_i', \alpha_{-i}') \neq \rho_i(\cdot | s_i'', \alpha_{-i}'')$. In this way *to each player i at each time t can be coupled a well defined statistical decision problem* (see Ferguson [18]), where:

- Λ_{-i} is the set of states of nature,
- S_i is the set of actions,
- $S_i \times \Delta(M_i)$ is the set of outcomes,
- v_i is the utility function defined on the set of outcomes,
- M_i is the set of observable random variable,
- $\hat{\rho}_i \in \Delta(M_i)$ is the unknown probability law of the random variable $\hat{m}_i^t[s_i] \in M_i$, that depends on the true unknown state of nature λ_{-i},

- $f_i^t : H_i^t \to S_i$ is the set of strategies.

If we consider the dynamic problem then each player i faces a *sequential statistical decision problem*, where at each time she should decide whether and how to experiment in order to collect information about the true unknown state of nature $\bar{\theta}'_{-i} \in \Lambda_{-i}$. This problem can be solved using the techniques of Markovian stochastic dynamic programming (see Hinderer [32]). Preliminary is the construction of the relevant filtration for the Bayesian inferential problem. Let $\bar{f}_i = \{\bar{s}_i^t\}$ be fixed and consider the following stochastic process: $\{\hat{m}_i^t[\bar{s}_i^t]\}$.

Denote the natural filtration generated by this stochastic process by $\{\mathcal{S}_{-i}^t(\bar{f}_i)\}$ where $\mathcal{S}_{-i}^t(\bar{f}_i) := \sigma(\hat{m}_i^\tau[\bar{s}_i^\tau] \ : \ \tau = 1, \cdots, t)$. By construction, $\{\hat{m}_i^t[\bar{s}_i^t]\}$ is a stochastic process adapted to the filtration $\{\mathcal{S}_{-i}^t(\bar{f}_i)\}$. Then, for every fixed superstrategy f_i and $\forall t$, there exists a version of the conditional expectation

$$\hat{\mu}_i^t[f_i](A) := E[\chi_A(\rho_i)|\mathcal{S}_{-i}^t(f_i)] \qquad \forall A \in \mathcal{S}_{-i}. \tag{9}$$

Note that (10) defines a regular conditional probability distribution (see Parthasarathy [56]). Moreover by definition $\hat{\mu}_i^t[f_i]$ is a function of f_i, because the information player i can collect during the play depends on her dynamic superstrategy (active learning). With the obvious abuse of notation, the objective function of player i can be written as follows:

$$E_{\mu_i} U_i(f_i, \lambda_{-i}) \ = \ E_{\hat{\mu}_i} \sum_{t=1}^{T} \delta^t v_i(f_i^t(P_i^t(f)), \rho_i(\cdot|f_i^t(P_i^t(f)), \lambda_{-i}) =: E_{\hat{\mu}_i} V_i(f_i, \rho_i).$$

In our model of recurring games individual actions provide information about the opponents' behavior as well as current reward. Therefore usually there is a trade-off between current reward and information which may be useful in the future. This problem can be described using the techniques of stochastic dynamic programming. Since the only connection between periods occurs through the beliefs, the decision problem of player i can be formulated as a Markovian stochastic dynamic programming problem with state space $\Delta(\Delta(M_i))$. Therefore the value function $V_i(\cdot)$ is the solution of the classic Bellman equation (see Blackwell [11], Theorem 7 and in general Hinderer [32]):

$$V_i(\hat{\mu}_i) \ = \ max_{s_i \in S_i} \ [(1-\delta) E_{\hat{\mu}_i} v_i(s_i, \rho_i(\cdot|s_i, \theta'_{-i})) + \delta \sum_{m_i} \rho_i(m_i|s_i, \theta'_{-i}) V_i(\hat{\mu}_i(\cdot|m_i))]$$

where $\hat{\mu}_i(\cdot|m_i)$ is the posterior given the signal m_i. Two of the most influential papers on this topic for situations with one decision maker are Easley and Kiefer [16] and Aghion et al. [1]. Unfortunately their results are not directly applicable to game theoretic contexts (the following notation is mine). In particular Easley and Kiefer assume (Assumption A.4) that the decision maker belief has a density continuous in m_i, s_i, λ_{-i} and that the supports of these densities are the same for all $(s_i, \lambda_{-i}) \in S_i \times \Lambda_{-i}$, hypotheses generally not satisfied in a game theoretic framework. On the other hand Aghion et al. assume (assumption A.2) that the signal function is continuous in (s_i, λ_{-i}), which is not true in my model.

Luckily using the formal tools of their appendix it is not difficult to show that their approach can be applied to the problems I study in this paper, as I do in lemma 6 in the appendix proving that asymptotically a rational agent will maximize her short run utility. According to this theorem the dynamically optimal choices of the players, as t goes to infinity, are going to maximize their best forecast of their short run utility. This happens because, while the benefit to increase information is diminishing, the experimentation costs in terms of unobtained short run payoff remains constant. Therefore eventually the short run utility maximization becomes predominant.

Then using these tools I can prove the convergence of players' behavior to a CE of the component game, using a classic result on consistency of Bayes estimates.

A strategy profile is a CE of an IMG if the signals induced by such a profile do not contradict the players' beliefs that rationalize their choices.

Definition 3 *A* **Heterogeneous Conjectural Equilibrium** *for an imperfect monitoring game $G(\eta)$ is a strategy profile $\lambda' \in \Lambda$ such that for each player $i \in N$, for any strategy in the support of λ'_i there exists a belief $\mu_i \in \Delta(\Lambda_{-i})$ such that the following two conditions hold*

$$s'_i \in \ arg\ max_{s_i} u_i(s_i, \mu_i) \tag{10}$$

$$\forall m_i \in M_i \quad \mu_i(\{\lambda_{-i} \in \Lambda_{-i} | \rho_i(m_i|s'_i, \lambda_{-i}) = \rho_i(m_i|s'_i, \lambda'_{-i})\}) = 1. \tag{11}$$

Remark: This definition of CE is weaker than the usual one, since it allows different beliefs to be used to rationalize each pure strategy in the support of σ_i. As theorem 2 will show, this is the opportune notion for model of recurring games where players are randomly matched with one another and observe only the results of their own match.

Theorem 2 *In the recurring game if the players' prior beliefs gives strictly positive probability to any possible opponents' strategy, then with $\mathbf{P}_{\hat{\mu}_i}$ probability one all limit points of the players' dynamic choices belong to the support of an Heterogeneous Conjectural Equilibrium of the component game.*

PROOF: see appendix B. ♡

Remarks:

1. in each population the agents play pure superstrategies, nevertheless the convergence is to HCE in mixed strategies. According to the random matching model, different agents in the same population play different pure strategies because of different experiences, even if any individual pure strategy is justified by a belief that is not falsified in equilibrium; thus an HCE is obtained averaging among the agents in each population;

2. the weak topology used in this paper has the fewest open sets of any natural topology, so it is fairly easy for a Bayes' estimates to be consistent;

3. the results on the consistency property of Bayes' estimates applies to the recurring game without further conditions both because the priors have full support and because the game is finite. Hence there are not the inconsistency problems of the infinite dimensional case: if the underlying probability mechanism allows an infinite number of possible outcomes, then the Bayes' estimates can be inconsistent, even if the true state of nature belongs to the prior's support (see Diaconis and Freedman [15] for a comprehensive discussion on the consistency of Bayes estimates);

4. for recurring games the correct notion of learning is consistency, while for repeated game is merging. For a thoughtful discussion of the relations between these two notions of interpersonal consensus see Schervish and Seidenfeld [61].

7 Concluding Remarks

The main results of this paper regard the convergence of rational learning processes towards Conjectural Equilibria. But I don't believe the relevance of this paper regards this equilibrium concept. I believe, instead, that the core of the paper is the generality of the approach in the construction of rational learning models. This generality has many interesting implications. First, it allows to specialize the results to more specific settings. For example, the theorems of this paper show that rational learning processes converge to Self-Confirming Equilibria in extensive form games and to Nash Equilibria in normal form games. Second, it clears up the crucial distinction between Bayesian learning (assumption 4) and rational learning (assumption 5). Finally, it makes possible to discuss the main problems of rational learning, i.e. the problems of non stationary environment and of correct theory. This is the specific topic of these concluding remarks.

The first half of assumption 5 regards the problem of non stationary environment. Although the assumption is restrictive, it is not unreasonable in this context. Indeed the inferential problem posed here is a classical induction problem, for revision of beliefs is just the process of learning from data and it is well known that induction is sensible only if the universe is stationary: "If ... the past may be no rule for the future, all experience becomes useless and can give rise to no inference or conclusion" (Hume [33]). I am not claiming that this assumption is realistic or that this is the true way used by the agents to learn. What I want to emphasize is that it is necessary to make this kind of assumptions if we want to study rational learning in this context. Otherwise the inferential problem is statistically unsolvable and the same assumption of rational learning would prevent the players from learning, giving rise to a mutually contradictory model. As Hume himself argues, it is not possible to show empirically that the assumption of stationary environment is true, it must be assumed *a priori*: "It is impossible ... that any arguments from experience can prove this resemblance of the past to the future, since all these arguments are founded on the supposition of that resemblance" (ibidem). In other words, if we refuse to assume stationary stochastic processes at some level and at a certain degree, we need to change the way the players learn, but until inductive and rational learning is considered, some form of intertemporal restriction is required. Otherwise, if the original opinions are based on independence of signals any possibility of learning through experience is excluded, because the latter, by definition, requires that the original opinions will not be modified on the basis of any observation of results. This general principle is at work in the context of repeated games also: interdependence of observations is obtained since opponents' superstrategies do not change over time and the players recognize it since we assume absolute continuity of actual behavior with respect to players' beliefs.

The second half of assumption 5, however, leaves us with the uncomfortable hypothesis of players' perfect comprehension of their stochastic setting, the problem of correct theory. In other words the assumption that the stochastic environment is stationary and that this is known, leaves it open to ask how it can be learned. But this is not a question that can be solved explicitly within a rational learning approach since by definition rational learning takes for granted that the players perfectly understand their stochastic environment. Note moreover that even naive Bayesian learning does suffer from the same problem, since the bounded rational approach doesn't explain how the agents have constructed their model of their stochastic environment, which is anyway hopelessly incorrect and systematically falsified by the available evidence. Anyway the problem is open for sure, although I think that it is more reasonable to assume that the players know the general characteristics of their stochastic environment, rather than the exact specific behavior of their opponents as in Nash equilibria, or that they are systematically wrong in their perceptions of the qualitative characteristics of their environment as in bounded rational learning models.

To provide a provisional answer to the general question of when rational learning is the correct approach to use, I propose the following interpretation. Under assumption 5 the world described by a recurring game can be thought of as a society where a standard of behavior for each possible role (player/population) has evolved through time, and the members of this society wish to learn these standards in order to maximize their lifetime utility. Even if the members of this society would know the actual laws that generate these standards through the individual self interested behavior, they can not have all the information required to calculate them at each temporal stage: their only chance of learning then is the past as a guide for the future. Therefore they approach this problem in a inductive way. Personally I think that it is more reasonable to assume that the players perfectly know their stochastic setting when this is stationary, than to assume they have an incorrect model which is never revised even when they face overwhelming falsifying evidence. More generally, the model of this paper shows that a rational man must have some theoretical views about the nature of the things he is learning about, for example whether the stochastic process of observable events is exchangeable. Then the problem is how to model the revision of these theories, a topic for future works. Until such model is not available, I think the best approach is to avoid clearly incorrect theories, i.e. to use a rational learning approach.

References

[1] Aghion, P., Bolton, P., Harris, C., Jullien, B.: Optimal Learning by Experimentation. Review of Economic Studies **58**, 621-654 (1991).

[2] Aumann, R.: Mixed and Behavioral Strategies in Infinite Extensive Games. In Dresher, M., Shapley, L., Tucker, A. (eds.) Advances in Game Theory. Princeton: Princeton University Press 1964.

[3] Aumann, R.: Agreeing to Disagree. Annals of Statistics **4**, 1236-1239 (1976).

[4] Basawa. I.V., Prakasa Rao, B.L.S.: Statistical Inference for Stochastic Processes. New York: Academic Press 1980.

[5] Battigalli, P.: Comportamento razionale ed equilibrio nei giochi e nelle situazioni strategiche. Unpublished dissertation, Bocconi University 1987.

[6] Battigalli, P., Guaitoli, D.: Conjectural Equilibria and Rationalizability in a Macroeconomic Game with Incomplete Information. Working Paper 1988-6, Bocconi university (1988).

[bgm] attigalli, Gilli, M., Molinari, C.: Learning and Convergence to Equilibrium in Repeated Strategic Interactions: an Introductory Survey. Ricerche Economiche **XLVI**, 335-378 (1991).

[7] Billingsley, P.: Convergence of Probability Measure. New York: Wiley 1968.

[8] Billingsley, P.: Probability and Measure. New York: Wiley 1986.

[9] Blackwell, D., Dubins, L.: Merging of Opinions with Increasing Information. Annals of Mathematical Statistics **38**, 882-886 (1962).

[10] Blackwell, D.: Discounted Dynamic Programming. Annals of Mathematical Statistics **36**, 226-235 (1965).

[11] Chung, K.: A Course in Probability Theory. New York: Academic Press 1974.

[12] Chakrabarti, S.: Equilibrium in Behavior Strategies in Infinite Extensive Form Games with Imperfect Information. Economic Theory **2**, 481-494 (1992).

[13] Dekel, E., Fudenberg, D., Levine, D.: Learning to Play Bayesian Games, mimeo (2001).

[14] Diaconis, P., Freedman, D.: On the Consistency of Bayes Estimates. Annals of Mathematical Statistics **14**, 1-26 (1986).

[15] Easley, D., Kiefer, N.: Controlling a Stochastic Process with Unknown Parameters. Econometrica **56**, 1054-1064 (1988).

[16] Eichberger, J.: Bayesian Learning in Repeated Normal Form Games. Games and Economic Behavior **11**, 254-278 (1995).

[17] Ferguson, T.: Mathematical Statistics. New York: Academic Press 1967.

[18] Freedman, D.: On the Asymptotic Behavior of Bayes Estimates in the Discrete Case I. Annals of Mathematical Statistics **36**, 1386-1403 (1963).

[19] Fudenberg, D., Kreps, D.: Learning in Extensive Games: Experimentation and Nash Equilibrium. Mimeo, Harvard University, (1995a).

[20] Fudenberg, D., Kreps, D.: Learning in Extensive Games: Self-Confirming Equilibrium. Games and Economic Behavior **8**, 20-55, (1995b).

[21] Fudenberg, D., Levine, D.: Limit Games and Limit Equilibria. Journal of Economic Theory **38**, 261-279, (1986).

[22] Fudenberg, D., Levine, D.: Steady State Learning and Nash Equilibrium. Econometrica **61**, 547-574 (1993a).

[23] Fudenberg, D., Levine, D.: Self-Confirming Equilibrium. Econometrica **61**, 523-546, (1993b).

[24] Fudenberg, D., Levine, D.: The Theory of Learning in Games. Cambridge, Mass. : The MIT Press 1998.

[25] Gilli, M.: Equilibrio ed aspettative nella teoria dei giochi e nella teoria economica. Unpublished dissertation, Bocconi University 1987.

[26] Gilli, M.: Modelling Strategic Interaction: a Case for Imperfect Monitoring Games. Working paper, 1995-3, Bocconi University 1995.

[27] Gilli, M.: Equilibrium and Learning in Game Theory. Unpublished doctoral dissertation, Cambridge University 1995.

[28] Gilli, M.: On Non-Nash Equilibria. Games and Economic Behavior **27**, 184-203 (1999a).

[29] Gilli, M.: Adaptive Learning in Imperfect Monitoring Games. Review of Economic Dynamics **2**, 472-485 (1999b).

[30] Gilli, M.: A General Approach to Rational Learning in Games. Bulletin of Economic Research **4**, 275-303 (2001).

[31] Hinderer, K. : Foundations of Non-stationary Dynamic programming with Discrete Time Parameter. New York: Springer-Verlag 1970.

[32] Hume, D. : An Enquiry Concerning Human Understanding. Chicago: William Benton 1952 (1 ed. 1748).

[33] Jackson, M. O., Kalai, E.: Social Learning in Recurring Games. Games and Economic Behavior **21**, 102-134 (1997).

[34] Jackson, M. O., Kalai, E.: Reputation versus Social Learning. Journal of Economic Theory **88**, 40-59 (1999).

[35] Jackson, M. O., Kalai, E., Smorodinsky, R.: Bayesian Representation of Stochastic Processes under Learning: De Finetti Revised. Econometrica **67**, 875-893 (1999).

[36] Jordan, J.: Bayesian Learning in Normal Form Games. Games and Economic Behavior **3**, 60-81 (1991).

[37] Jordan, J.: Bayesian Learning in Repeated Games. Games and Economic Behavior **9**, 8-20 (1995).

[38] Jordan, J.: Bayesian Learning in Games: a non-Bayesian Perspective. In Bicchieri, C., Jeffrey, R., Skyrms, B. (eds.) The Dynamics of norms, pp. 149-174. Cambridge: Cambridge University Press 1997.

[39] Kalai, E., Lehrer, E.: Rational Learning Leads to Nash Equilibrium. Econometrica **61**, 1019-1045 (1993a).

[40] Kalai, E., Lehrer, E.: Subjective Equilibrium in Repeated Games. Econometrica **61**, 1231-1240 (1993b).

[41] Kalai, E., Lehrer, E.: Weak and Strong Merging of Opinions. Journal of Mathematical Economics **23**, 73-86 (1994).

[42] Kandori, M.: Introduction to Repeated Games with Private Monitoring, forthcoming in Journal of Economic Theory (2001).

[43] Kolmogorov, A.N., Fomin, S.V.: Introductory real analysis. New York: Dover Publications 1975.

[44] Koutsougeras L.C., Yannelis, N.C.: Convergence and Approximation Results for Non-cooperative Bayesian Games: Learning Theorems. Economic Theory **6**, 843-858 (1994).

[45] Kuratowski, K.: Topology. New York: Academic Press 1968.

[46] Miller R.I., Sanchirico, C.W.: The Role of Absolute Continuity in "Merging of Opinions" and "Rational learning". Games and Economic Behavior **29**, 170-190 (1999).

[47] Nachbar, J. H.: Prediction, Optimization, and Learning in Repeated Games. Econometrica **65**, 275-309 (1997).

[48] Nachbar, J. H.: Rational Bayesian Learning in Repeated Games. mimeo (1998).

[49] Nyarko, Y.: The convergence of Bayesian beliefs hierarchies. C.V. Starr Center Working Paper N. 91-50, New York University (1991).

[50] Nyarko, Y.: Bayesian learning without common priors and convergence to Nash equilibria. Mimeo, New York University 1992.

[51] Nyarko, Y.: Bayesian Learning Leads To Correlated Equilibria in Normal Form Games. Economic Theory **6**, 821-842 (1994).

[52] Nyarko, Y.: Savage-Bayesian agents play a repeated game. In Bicchieri, C., Jeffrey, R., Skyrms, B. (eds.) The Dynamics of norms, pp. 175-198. Cambridge: Cambridge University Press 1997.

[53] Nyarko, Y.: Bayesian Learning and convergence to Nash Equilibria without Common Prior. Economic Theory **11**, 643-655 (1998a).

[54] Nyarko, Y.: The Truth is in the Eye of the Beholder: or Equilibrium in Beliefs and Rational Learning in Games. Mimeo, New York University (1998b).

[55] Parthasarathy, K. R.: Probability Measures on Metric Spaces. New York: Academic Press 1967.

[56] Pearce, D.: Rationalizable Strategic Behavior and the Problem of Perfection. Econometrica **42**, 1029-1050 (1984).

[57] Sandroni, A.: Does Rational Learning Lead to Nash Equilibria in Finitely Repaetd Games. Journal of Economic Theory, **78**, 195-218 (1998).

[58] Sandroni, A., Smorodinsky, R.: Dynamic Foundations for Equilibrium in Games - A Second Look. Mimeo (1998).

[59] Sandroni, A., Smorodinsky, R.: The Speed of Rational Learning. International Journal of Game Theory, **28**, 199-210, (1999).

[60] Schervish, M.J., Seidenfeld, T.: An Approach to Consensus and Certainty with Increasing Evidence. Journal of Statistical Planning and Inference **25**, 401-414 (1990).

[61] Sorin, S.: Merging, Reputation, and repeated Games with incomplete Information. Games and Economic Behavior **29**, 274-308 (1999).

[62] Yamada, K.: Asymptotic Behavior of Posterior Distributions for Random Processes under Incorrect Models. Journal of Mathematical Analysis and Applications **56**, 294-308 (1976).

[63] Young, P. H.: Individual Strategy and Social Structure. Princeton: Princeton University Press 1998.

[64] Young, P. H., Foster, D. P.: When Rational Learning Fails. mimeo (1999).

A Convergence to Conjectural Equilibria of the Imperfect Monitoring Game

First consider the two propositions that allow to construct a learnable pattern for the recurring game.

Proposition 3 *For every $\phi \in \Phi$, if $\theta^t[\phi]$ is independent from t, then the true λ^t is also independent of t.*

PROOF: fix ϕ. At each time t the statistical distribution of possible plays of s_k^t for population $k \in N$ is given by $\theta_k'^t[\phi_k]$, which is uniquely defined as follows:

$$\theta_k'^t[\phi_k](s_k) = \sum_{\{f_k | f_k(\bar{h}_k)=s_k\}} \phi_k(f_k) \sum_{\{\bar{h}_k | f_k(\bar{h}_k)=s_k\}} \theta_k^t(\bar{h}_k), \tag{12}$$

with $\theta_k'^t[\phi_k] \in \Lambda_k$. For T finite, then also F_T and H_T are finite and thus the summations in (9) are well defined.

By definition in such a context $\theta_k'^t[\phi_k](s_k)$ is the true probability of s_k being played at time t in a match of the anonymous repeated game with random matching, given that the population of player k is following the average/mixed superstrategy ϕ_k. Therefore equation (9) implies that if $\forall k \quad \theta_k^t$ is independent from t, then even the actual probability distribution of play $\theta'^t[\phi] \in \Lambda$ is independent from t. ♡

Now we prove that such a steady state exists.

Proposition 4 *In RG, for every finite T and for every $\phi_T \in \Phi_T$, there exists a steady state $\bar{\theta}_T \in \Delta(H_T)$ of the updating process of the proportion of private experiences, and therefore a corresponding stationary $\bar{\theta}'_T[\phi_T] \in \Lambda$.*

PROOF: let T be finite and remember that by assumption 1 $G(\eta)$ is finite. The steady state is a fixed point of the function d, but this function is polynomial and thus trivially continuous. Moreover by definition $\Delta(H_T)$ is a compact and convex set. Therefore the usual fixed point theorems imply that there exists a fixed point of d and thus a steady state $\bar{\theta}_T \in \Delta(H_T)$. Then proposition 1 implies that $\forall T$ finite, $\forall \phi_T$, there exists a stationary $\bar{\theta}'_T(\phi_T) \in \Lambda$. ♡

Lemma 1 provides a useful alternative representation of mixed strategies.

Lemma 1 *Every time t mixed strategy $\phi_i^t : \Omega \times H_i^{t-1} \to S_i$ can be represented by a function $\phi_i^t : H_i^{t-1} \to \Delta(S_i)$, with the obvious abuse of notation.*

PROOF: see Chakrabarti [13], lemma 3.2 ♡
Now a result crucial to connect the limit game with the infinitely repeated game.

Lemma 2 *The following expressions hold:*

1. $\bigcup_{T=1}^{\infty} H_{i,T} = H_i, \quad \bigcup_{T=1}^{\infty} F_{i,T} = F_i, \quad \bigcup_{T=1}^{\infty} \Phi_{i,T} = \Phi_i;$

2. $\lim_{T\to\infty} U_{i,T} = U_i$;

3. *for each limit point* θ'_∞ *of* $\{\theta'_T\}_T$, $\theta'_\infty \in \Lambda$;

4. *for each limit point* $\mu_{i,\infty}$ *of* $\{\mu_{i,T}\}_T$, $\mu_{i,\infty} \in \Delta(\Lambda_{-i})$;

5. *for each limit point* $\hat{\mu}_{i,\infty}$ *of* $\{\hat{\mu}_{i,T}\}_T$, $\hat{\mu}_{i,\infty} \in \Delta(\Delta(M_i))$;

6. *if* ϕ_T *are* ϵ_T*-equilibria of* $G^T(\delta,\eta)$ *with* $\epsilon_T \to \epsilon$ *and* $\phi_T \to \phi$*, then* ϕ *is an* ϵ*-equilibrium of* $G^\infty(\delta,\eta)$*. Moreover if* ϕ *is an* ϵ*-equilibrium of* $G^\infty(\delta,\eta)$*, there exists sequences* $\{\phi_T\}$ *and* $\{\epsilon_T\}$*, with* $\epsilon_T \to \epsilon$ *and* $\phi_T \to \phi$*, such that* ϕ_T *is an* ϵ_T*-equilibrium of* $G^T(\delta,\eta)$.

PROOF: expressions 1 and 2 are an immediate consequence of assumption 3. Moreover each limit point considered in the lemma exists, because all the sets involved are compact. Therefore expressions 3, 4 and 5 follow trivially from Λ, $\Delta(\Lambda_{-i})$ and $\Delta(\Delta(M_i))$ compactness (in the weak topology). Finally note that $U_{i,T}$ is continuous at infinity in the sense of Fudenberg and Levine [22] definition 4.1, that is: $\lim_{T\to\infty} sup|U_{i,T}(f_T) - U_{i,T}(\bar{f}_T)| = 0$ for all $f_T, \bar{f}_T$ with the first T strategies equal. Therefore proposition 6 of the lemma follows from Fudenberg and Levine [22] Limit Theorem. ♡

Now a result of logical consistency for the construction of the proof of convergence.

Lemma 3 *$\hat{\mu}_i$ is a well defined probability measure.*

PROOF: $\hat{\mu}_i$ is a well defined probability measure if

$$\forall A \in \mathcal{S}_{-i} \quad \{\lambda_{-i} | \hat{\rho}_i(m_i[s_i]|\lambda_{-i}) \in A\} \in \mathcal{B}_{-i}$$

but this is implied by the facts that $\mathcal{S}_{-i}$ and $\mathcal{B}_{-i}$ are the Borel σ-fields generated respectively by $\Delta(M_i)$ and Λ_{-i} and that $\hat{\rho}_i$ is linear and continuous in λ_{-i}. ♡

The following two lemmas are well known and therefore their proof is omitted (see e.g. Chung [12]).

Lemma 4 *For every $f_i \in F_i$ and for every $A \in \mathcal{S}_{-i}$, the stochastic process $\{\hat{\mu}_i^t[f_i](A)\}_t$ is a dominated martingale relative to $(\{\mathcal{S}_{-i}^t(f_i)\}, \hat{\mu}_i)$.*

Lemma 5 *The following results hold for all $i \in N$ and for each $f_i \in F_i$:*

$$\forall A \in \mathcal{S}_{-i} \qquad \hat{\mu}_i^t[f_i](A) \to_{t\to\infty} \hat{\mu}_i^\infty[f_i](A) \qquad \mathbf{P}_{\hat{\mu}_i}\ a.e.$$

$$\forall A \in \mathcal{S}_{-i} \qquad \hat{\mu}_i^\infty[f_i](A) = E[\chi_A|\mathcal{S}_{-i}^\infty(f_i)] \qquad \mathbf{P}_{\hat{\mu}_i}\ a.e.$$

$$\hat{\mu}_i^t[f_i] \to_{t\to\infty} \hat{\mu}_i^\infty[f_i] \quad weakly \quad and \quad \mathbf{P}_{\hat{\mu}_i}\ a.e.$$

Then the crucial lemma for convergence to equilibria of the stage game.

Lemma 6 *For all players $i \in N$ let $f_i = \{s_i^t\}$. Then with $\mathbf{P}_{\hat{\mu}_i}$ probability one all limit points of $\{f_i\}$ maximise* $E_{\hat{\mu}_i^\infty[f_i]}\, v_i(s_i, \rho_i(\cdot|s_i, \theta'_{-i}))$.

PROOF: see Aghion et al [1], theorem 2.5. Note that it is possible to apply this theorem because in this setting v_i is continuous in (s_i, ρ_i) and assumption (B) of p.644 is also satisfied because F_i is compact in the product topology (by Tychonov product theorem) and U_i is continuous in f_i. Now consider the signal function: η_i is defined on $S_i \times S_{-i}$ but it can be extended to the domain $S_i \times \Delta(M_i)$ in the following way:

$$\hat{\eta}_i(s_i, \rho_i) = \begin{cases} \eta_i(\bar{s}) & \text{if } (s_i, \rho_i) = (\bar{s}_i, \rho_i(\cdot|\bar{s}_i, \delta_{\bar{s}_{-i}})) \\ X & \text{otherwise} \end{cases}$$

where $\delta_{\bar{s}_{-i}}$ is the probability measure assigning point mass at $\bar{s}_{-i}$ and X is a null element. Then it is immediate to see that $\hat{\eta}_i$ is Borel measurable in (s_i, ρ_i) and it is easy to check that this is all I need for the proof of theorem 2.5 in Aghion et al. [1] (see theorem A.4 and A.6 in the appendix). ♡

Finally the main result on convergence to CE of the IMG

Theorem 3 *If $\hat{\mu}_i \in \Delta^\circ(\Delta(M_i))$, then in the anonymous RIMG with random matching with $\mathbf{P}_{\hat{\mu}_i}$ probability one all limit points of the players' pure super-strategies belong to the support of an Anonymous Conjectural Equilibrium of the component game.*

PROOF: fix a generic $i \in N$, consider an optimal $f_i \in F_i$ and let $\bar{s}_i$ be a limit point of f_i. Because of assumption 1, there exists a finite time $\hat{t}$ such that, possibly along a subsequence, $\forall t_n \geq \hat{t} \quad s_i^{t_n} = \bar{s}_i$. Consider a subsequence converging to $\bar{s}_i$: under the assumptions done, for each possible $\theta'_{-i} \in \Lambda_{-i}$, $\{\hat{m}_i^{t_n}[s_i^{t_n}]\}_{t_n \geq \hat{t}}$ is a sequence of independent random variable with unknown common distribution $\hat{\rho}_i$. But then we are in the classical case of Bayesian inference: $(\Lambda_{-i}, \mathcal{B}_{-i})$ is the space of probability related to the state of nature, $\{\hat{m}_i^{t_n}[s_i^{t_n}]\}_{t_n \geq \hat{t}} = \{\hat{m}_i^{t_n}[\bar{s}_i]\}_{t_n \geq \hat{t}}$ is a sequence of finite valued random variable ("the observations") defined on an abstract space $(\Omega, \mathcal{A})$. The probability law of $\hat{m}_i^{t_n} \quad \forall t_n \geq \hat{t}$ is the following unknown random variable: $\hat{\rho}_i : \Lambda_{-i} \to \Delta(M_i)$, where $\Delta(M_i)$ is the "parameter" space, such that for every $\hat{\rho}_i \in \Delta(M_i)$ there is a probability measure $\mathbf{P}_{\hat{\rho}_i}$ on $\mathcal{A}$ under which the $\{\hat{m}_i^t\}$ are independent with common distribution

$$\mathbf{P}_{\hat{\rho}_i}(\{\omega \in \Omega | \hat{m}_i^t[s_i](\omega) = m_i\}) = \hat{\rho}_i(m_i|\lambda_{-i}).$$

Therefore under each $\lambda_{-i} \in \Lambda_{-i}$ the $\{\hat{m}_i^{t_n}[s_i^{t_n}]\}_{t_n \geq \hat{t}}$ are identically and independently distributed with common distribution $\hat{\rho}_i(\cdot|\lambda_{-i})$ and $\hat{\mu}_i$ is the prior on the Borel σ-field of $\Delta(M_i)$, $\mathcal{S}_{-i}$.

Now define $\bar{\rho}_i := \rho_i(\cdot|\bar{\theta}'_{-i})$ for the true $\bar{\theta}'_{-i} \in \Lambda_{-i}$. Then we know from the classical theorems on the consistency of the Bayes' estimates (see e.g. Freedman [19] theorem 1) that if M_i is finite, then $(\bar{\rho}_i, \hat{\mu}_i)$ is consistent if and only if $\bar{\rho}_i \in SUPP[\hat{\mu}_i]$. Consistency here means that the conditional probability $\hat{\mu}_i^t[f_i]$ will converge weakly to a $\hat{\mu}_i^\infty[f_i]$ (lemmas 4 and 5) such that

$$\hat{\mu}_i^\infty[f_i](\{\hat{\rho}_i = \bar{\rho}_i\}) = 1 \quad \mathbf{P}_{\bar{\rho}_i} a.e.$$

By assumption $\hat{\mu}_i \in \Delta^\circ(\Delta(M_i))$ and thus $(\bar{\rho}_i, \hat{\mu}_i)$ is consistent. But this means that a.s. each limit point $\bar{s}_i$ of f_i belong to the support of a CE because by definition of $\hat{\mu}_i$ there exists a $\mu_i \in \Delta(\Lambda_{-i})$ corresponding to $\hat{\mu}_i^\infty[f_i]$ such that:

$$\mu_i(\{\lambda_{-i} | \rho_i(\cdot|\bar{s}_i, \lambda_{-i}) = \rho_i(\cdot|\bar{s}_i, \bar{\theta}'_{-i})\}) = 1 \qquad \mathbf{P}_{\bar{\rho}_i} a.e.$$

Moreover lemma 6 and the definition of v_i imply that with $\mathbf{P}_{\hat{\mu}_i}$ probability one

$$\bar{s}_i \in arg\,max\; u_i(\cdot, \mu_i).$$

Finally note that since by definition $\mathbf{P}_{\hat{\mu}_i}(A) = \int_A \mathbf{P}_{\hat{\rho}_i}\, \hat{\mu}_i(d\hat{\rho}_i)$, then $\mathbf{P}_{\bar{\rho}_i}(A) = 1$ implies $\mathbf{P}_{\hat{\mu}_i}(A) = 1$ ♡