Repeated Games with Public Signal and Bounded Recall

Jérôme Renault
CEREMADE
Université Paris Dauphine
Pl. du Marechal de Lattre de Tassigny
F–75775 Paris Cedex 16, France
renault@ceremade.dauphine.fr

Marco Scarsini
Dipartimento di Statistica
Università di Torino
Piazza Arbarello 8
I–10122 Torino, Italy
marco.scarsini@unito.it

Tristan Tomala
CEREMADE
Université Paris Dauphine
Pl. du Marechal de Lattre de Tassigny
F–75775 Paris Cedex 16, France
tomala@ceremade.dauphine.fr

February 2006

Abstract

This paper studies repeated games with public signals, symmetric bounded recall and pure strategies. Examples of equilibria for such games are provided and the convergence of the set of equilibrium payoffs is studied as the size of the recall increases. Convergence to the set of equilibria of the infinitely repeated game does not hold in general but for particular signals and games. The difference between private and public strategies is relevant and the corresponding sets of equilibria behave differently.

Key words: folk theorem, de Bruijn sequence, imperfect monitoring, uniform equilibrium.

MSC 2000 Subject Classification: Primary 91A20. Secondary 91A18, 91A43.

OR/MS subject classification: Primary: Games/group decisions, noncooperative.

1 Introduction

Repeated games with complete information are known to have multiple equilibria. The prominent result in this direction is the folk theorem which asserts that in games with perfect monitoring and perfectly rational players, every feasible and individually rational payoff can be sustained by an equilibrium of the repeated game. A more realistic model to study involves games with imperfect monitoring, where players observe imperfectly other players' actions, and bounded rationality, where players have limited information processing abilities. Typically these two problems have been studied separately in the literature. A notable exception is a recent paper by Cole and Kocherlakota (2005).

The literature on games with imperfect monitoring seeks to characterize the set of equilibrium payoffs (see e.g., Lehrer (1988, 1992a,b), Abreu et al. (1990), Fudenberg and Levine (1994), Tomala (1998), Renault and Tomala (2004)), and the literature on games with bounded rationality examines whether equilibrium payoffs of the unrestricted repeated game can be approximated by equilibrium payoffs of the repeated game with bounded rationality (see e.g., Rubinstein (1986), Abreu and Rubinstein (1988), Kalai and Stanford (1988), Lehrer (1988, 1994) Ben-Porath (1990, 1993), Sabourian (1998), Neyman (1998), Bavly and Neyman (2005)).

The aim of the present paper is to blend these two approaches. The question addressed here is the following. In games with imperfect monitoring, does the set of equilibrium payoffs of the game with bounded recall converge to the set of equilibrium payoffs of the unrestricted repeated game? This paper considers a model of repeated games with public signals and pure strategies for which a characterization of the set of equilibrium payoffs is available (Tomala (1998)). Players will be restricted to strategies with bounded recall, the size of the recall being the same for each player (see Sabourian (1998), who uses a similar model with perfect monitoring and proves a Folk theorem-like result).

A similar approach can be found in Cole and Kocherlakota (2005). It must be remarked that the model studied in this paper differs from the one studied by Cole and Kocherlakota (2005), who consider players with unbounded recall and equilibria implemented by strategies that depend only on the most recent k observations. Furthermore they use mixed strategies and a randomizing device.

In this paper only pure strategies will be considered. The type of results obtained here are a general study of public and private equilibria in games with imperfect monitoring and bounded recall. It will be shown that the set of public equilibria with bounded recall is a subset of the set of public equilibria with unbounded recall, but in general no convergence of these sets is guaranteed as the size of recall diverges. The set of private equilibria with bounded recall has even worse properties, in that it is not even a subset of the set of private equilibria with unbounded recall.

The example that will be examined more extensively in the paper refers to a minority game. It will be proved that for this game convergence of the set of public equilibria actually holds. This class of games originates from an idea of Arthur (1994, 1999), and has been studied in several articles in the physics literature (see, e.g., the recent books by Challet et al. (2005) and Coolen (2005), and references therein). This literature looks at the behavior of agents with bounded recall and bounded rationality, hence it neglect the strategic aspects of interaction. A strategic analysis of these games with infinite recall can be found in Renault et al. (2005). Analyzing this game under bounded recall brings this paper closer to the original spirit of physicists.

The paper is organized as follows. Section 2 describes the model. Section 3 examines the set of equilibrium payoffs for games with bounded recall and studies its convergence as the recall size diverges. Section 4 deals with minority games. Section 5 considers games with bounded recall and trivial monitoring where the players differ either in the size of recall or the cardinality of their action spaces. Finally Section 6 contains the proofs of the results.

2 The model

2.1 Description of the model

Consider a stage game

$$G = \langle N, (A^i)_{i\in N}, (g^i)_{i\in N}\rangle. \tag{2.1}$$

In this setting N is a set of players, for each $i \in N$, A^i is the set of actions available to player i, $A := \times_{i\in N} A^i$ is the set of action profiles, and the map $g^i : A \to \mathbb{R}$ is the payoff function for player i. Denote by $g : A \to \mathbb{R}^N$ the vector payoff function. For every $i \in N$, put $A^{-i} = \times_{j\in N, j\neq i} A^j$, therefore $a^{-i} \in A^{-i}$ will be a shortcut for $(a^j : j \neq i) \in \times_{j\in N,j\neq i} A^j$. Consider then a set of signals U and a mapping $\ell : A \to U$. In the whole paper the sets N, A^i, U are assumed nonempty and finite.

This game is repeated over time. At each round $t = 1, 2, \dots,$ players choose actions and if $a_t \in A$ is the action profile at stage t, they observe a public signal $u_t = \ell(a_t)$ before proceeding to the next stage. The set of histories of length $t \geq 0$ for player i is $\mathcal{H}^i_t := (A^i \times U)^t$, $\mathcal{H}^i_0$ being a singleton, and $\mathcal{H}^i = \cup_{t\geq 0}\mathcal{H}^i_t$ is the set of all histories for player i.

If all players perfectly observe, at the end of each stage, the actions played by the other players, then $U = A$ and ℓ is the identity mapping on A. If the players observe no signal, then the function ℓ is constant. These two cases will be referred to as *perfect monitoring* and *trivial monitoring*, respectively.

A strategy for player i is a mapping $\sigma^i : \mathcal{H}^i \to A^i$. The set of strategies for player i is denoted by Σ^i, and similar conventions are adopted for actions: $\Sigma = \times_{j\in N} \Sigma^j$, $\Sigma^{-i} = \times_{j\in N,j\neq i} \Sigma^j$. A profile of strategies $\sigma = (\sigma^i)_{i\in N}$ generates a unique history $(a_t(\sigma), u_t(\sigma))_{t\geq 1} \in (A \times U)^\infty$, where for each t, $u_t(\sigma) = \ell(a_t(\sigma))$. In the whole paper only pure strategies are considered.

Given a strategy profile σ, the average payoff for player i up to time T is $\gamma^i_T(\sigma) = \frac{1}{T}\sum_{t=1}^T g^i(a_t(\sigma))$, and $\gamma^i(\sigma) = \lim_{T\to\infty} \gamma^i_T(\sigma)$, when the limit exists.

Let Γ_∞ be the infinitely repeated game. Next definition recalls the concept of uniform equilibrium.

Definition 2.1. A strategy profile σ is a *uniform equilibrium* of Γ_∞ if

(a) for all $i \in N$, $\gamma^i(\sigma)$ exists.

(b) for all $\epsilon > 0$ there exists T_0 such that for all $T \geq T_0$, for all $i \in N$, for all $\tau^i \in \Sigma^i$, $\gamma^i_T(\tau^i, \sigma^{-i}) \leq \gamma^i_T(\sigma) + \epsilon$.

Denote by E_∞ the set of uniform equilibrium payoffs of Γ_∞, i.e., the set of vectors $(\gamma^i(\sigma))_{i \in N}$, where σ is a uniform equilibrium of Γ_∞.

2.2 Public strategies

Definition 2.2. Let $i \in N$. The strategy $\sigma^i \in \Sigma^i$ is called *public* if for all $t \geq 1$, and for all histories of length t, $h = (a^i_1, u_1, \ldots, a^i_t, u_t)$ and $h' = (b^i_1, v_1, \ldots, b^i_t, v_t)$,

$$(\forall s \in \{1, \ldots, t\},\ u^i_s = v^i_s) \implies \sigma^i(h) = \sigma^i(h').$$

In words a public strategy depends only on public signals. The set of public strategies of player i is denoted by $\widehat{\Sigma}^i$. A strategy profile σ is a *public equilibrium* if it is a uniform equilibrium and each player's strategy is public. The corresponding set of equilibrium payoffs is denoted by $\widehat{E}_\infty$. In the case of perfect monitoring, any strategy is public, since the public history contains all the past.

In repeated games with unbounded recall, every pure strategy is equivalent to a public strategy. Knowing her own strategy and the history of public signals, a player can deduce the actions she played in the past (see e.g. Tomala (1998)). More precisely the following lemma holds.

Lemma 2.3. *For every $\sigma^i \in \Sigma^i$, there exists $\widehat{\sigma}^i \in \widehat{\Sigma}^i$ such that for all $\tau^{-i} \in \Sigma^{-i}$ and for each stage t*

$$a_t(\sigma^i, \tau^{-i}) = a_t(\widehat{\sigma}^i, \tau^{-i}).$$

The proof is straightforward: the action played by sigma at the first stage depends on σ^i only, therefore the action played at the second stage depends only on σ^i and on the first public signal and so on, by induction.

Corollary 2.4. $\widehat{E}_\infty = E_\infty$.

To emphasize the dependence on the player's own past actions, a strategy that is not public will be called *private*. As it will be seen in the sequel, in games with bounded recall, considering public or private strategies makes a big difference.

2.3 Bounded recall

Consider now players who recall only recent observations. Informally, a strategy has recall k, if the player who uses it remembers only what happened on the k previous stages, and plays in a stationary way, i.e., this player has no clock and relies on her recall, but not on time. The formal definition is the following.

Definition 2.5. Given an integer $k \in \mathbb{N}$, the strategy $\sigma^i \in \Sigma^i$ has *recall* k if there exists a mapping $f : (A^i \times U)^k \to A^i$ such that for all $t > k$ and for all histories $h = (a^i_1, u_1, \ldots, a^i_t, u_t) \in \mathcal{H}^i_t$

$$\sigma^i(h) = f(a^i_{t-k+1}, u_{t-k+1}, \ldots, a^i_t, u_t).$$

By convention, a strategy that has recall 0 is a constant mapping on $\mathcal{H}^i$.

Lehrer (1988, 1992a,b) and Bavly and Neyman (2005) use a somewhat different definition in that for them, a bounded recall strategy is the choice of an initial recall plus the mapping f. This implies that whenever the initial recall appears during the course of the game, the player will play in the same way as at early stages. In the definition given here, the player plays as she wishes before stage k and then uses the stationary rule f. We believe that asymptotic results are unlikely to differ using one or another definition, however for small values of k, the initialization phase might be critical. Also note that Sabourian (1998) uses the same definition as the one given above.

The set of strategies for player i that have recall k is denoted by Σ^i_k and $\Sigma_k := \times_{i \in N} \Sigma^i_k$. Since the game is finite, for each $\sigma \in \Sigma_k$, the sequence $a_t(\sigma)$ is periodic, which implies the existence of $\gamma^i(\sigma)$. The normal form game $\Gamma_k = \langle N, (\Sigma^i_k), (\gamma^i) \rangle$ is thus well defined and the set of Nash equilibrium payoffs of Γ_k in pure strategies is denoted by E_k.

Let $\widehat{\Sigma}^i_k = \widehat{\Sigma}^i \cap \Sigma^i_k$ be the set of public strategies with recall k, $\widehat{\Gamma}_k = \langle N, (\widehat{\Sigma}^i_k), (\gamma^i) \rangle$ be the public-strategy game with recall k, and $\widehat{E}_k$ be the set of its (pure) Nash equilibrium payoffs.

Remark 2.6. In games with bounded recall, considering public strategies is a true restriction. From Lemma 2.3, every pure strategy σ^i is equivalent to a public strategy $\widehat{\sigma}^i$ but the bounded recall property is not preserved. It might be that σ^i has recall k but $\widehat{\sigma}^i$ does not. For example, consider trivial monitoring (the mapping ℓ is constant). Given any recall k, there is only one history of public signals, thus a public strategy with bounded recall is a constant strategy. By contrast, a private strategy (of recall 1) can simply alternate between two actions. The equivalent public strategy alternates between the two actions according to time and thus is not a public strategy with bounded recall according to Definition 2.5.

Remark 2.7. In the game with recall 0, strategies are constant and thus $\widehat{E}_0 = E_0$. This set further coincides with the set of pure Nash equilibrium payoffs of the stage game.

3 Equilibrium payoffs with bounded recall

In this section properties of the sets E_k and $\widehat{E}_k$ will be studied.

The following proposition holds.

Proposition 3.1. (a) *If a strategy σ is an equilibrium of $\widehat{\Gamma}_k$, then σ is a uniform equilibrium of $\widehat{\Gamma}_\infty$. Thus, $\widehat{E}_k \subset \widehat{E}_\infty$.*

(b) *If a strategy σ is an equilibrium of $\widehat{\Gamma}_k$, then σ is an equilibrium of Γ_k. Thus, $\widehat{E}_k \subset E_k$.*

(c) *If a strategy σ is an equilibrium of $\widehat{\Gamma}_k$, then σ is an equilibrium of $\widehat{\Gamma}_{k+1}$. Thus, $\widehat{E}_k \subset \widehat{E}_{k+1}$.*

These properties are somehow expected from such a model and are proved by arguments that are quite standard in the literature. As will be seen in the sequel of the paper, these are about the most general properties one can get for these games. In particular, the sequence E_k need neither be monotonic nor included in E_∞.

3.1 Convergence of $\widehat{E}_k$

The aim of this section is to establish whether the monotonic sequence $\widehat{E}_k$ converges to E_∞. Two remarkable cases are *trivial monitoring*, i.e. the mapping ℓ is constant and *perfect monitoring*, i.e. the mapping ℓ is one-to-one.

3.1.1 Trivial monitoring

Convergence of $\widehat{E}_k$ to E_∞ does not hold in general as is easily seen by considering trivial monitoring.

Proposition 3.2. *If the mapping ℓ is constant then,*

(a) *the set E_∞ is the convex hull of the set of Nash equilibrium payoffs of the one-shot game, whereas,*

(b) *for each k, $\widehat{E}_k$ is the set of Nash equilibrium payoffs of the one-shot game.*

3.1.2 Perfect monitoring

As mentioned before, with perfect monitoring any strategy can be represented by a public strategy, since the public history contains all the past. So $E_k = \widehat{E}_k$ for each k. The following example shows that even in this case, the convergence of $(\widehat{E}_k)_k$ to E_∞ may fail to hold.

There are three players, $N = \{1, 2, 3\}$. Player 3 only has one action, player 1 chooses the line, player 2 chooses the column, and as usual the first (resp. second, resp. third) coordinate of the vector payoff corresponds to the first (resp. second, resp. third) player.

	L	R
T	$\pi, -\pi, 1$	$-1, 1, 1$
B	$0, 0, 0$	$0, 0, 0$

Proposition 3.3. (a) *The payoff* $(0, 0, 1) \in E_\infty$.

(b) *The set* $\widehat{E}_k = E_k = \{(0, 0, 0)\}$.

The idea of the proof is that to get $(0, 0, 1)$ in equilibrium, player 1 plays T and player 2 alternates L and R with the correct frequencies. Since π is irrational, these frequencies must also be irrational but players with bounded recall eventually enter a cycle on a finite number of actions, generating rational frequencies.

3.2 Study of E_k

As the previous example shows, $(E_k)_k$ may not converge to E_∞, even in the case of perfect monitoring.

It is plain that $E_0 \subset E_\infty$ always holds and one may wonder if for $k \geq 1$, $E_k \subset E_\infty$. The following example shows that it is not always so. One may also wonder if $\lim_k E_k$ (set as $\cup_K \cap_{k \geq K} E_k$ to avoid existence problems) is a subset of E_∞. The answer is negative: the following example shows a point in $\cap_k E_k \backslash E_\infty$.

Consider the following two-player game, with $A^1 = \{T, M, B_1, B_2\}$, $A^2 = \{L, R\}$ and $U = \{u, v\}$. The payoffs and the public signals are indicated below.

	L	R
T	$0, 0,$	$0, 0$
M	$2, 2$	$0, 3$
B_1	$2, 1$	$0, 0$
B_2	$2, 1$	$0, 0$

payoffs

	L	R
T	u	u
M	u	u
B_1	u	v
B_2	u	v

signals

Proposition 3.4. *In the above game, for each* $k \geq 1$*, the payoff* $(2, 4/3) \in E_k$*, but* $(2, 4/3) \notin E_\infty$.

4 A minority game

This subsection describes a game, called *minority game* (MG), that serves as leading example in the paper.

In this game three players have to choose simultaneously one of two rooms: L (left) or R (right). For each profile of action $a = (a^1, a^2, a^3) \in \{L, R\}^3$, call *minority* *room* the less crowded room and *majority room* the most crowed room. Player i's payoff is then 1 if she chooses the minority room and 0 otherwise. Hence the payoff matrix of the MG is as follows, where player 1 chooses the row, player 2 the column, and player 3 the matrix.

$$\begin{array}{c|c|c|} & L & R \\ \hline L & 0,0,0 & 0,1,0 \\ \hline R & 1,0,0 & 0,0,1 \\ \hline \end{array} \qquad \begin{array}{c|c|c|} & L & R \\ \hline L & 0,0,1 & 1,0,0 \\ \hline R & 0,1,0 & 0,0,0 \\ \hline \end{array}$$

$$L \qquad\qquad\qquad\qquad R$$

The profile where one player chooses L and the two other players choose R is a Nash equilibrium. All pure Nash equilibria of this game are obtained by permutation of players and rooms. Denote by C be the convex hull of payoff vectors generated by these equilibria. If $e(i) \in \mathbb{R}^3$ is the vector whose i-th component is 1 and the other components are 0, then

$$C = \operatorname{conv}\{e(i) : i \in \{1,2,3\}\} = \left\{ x \in [0,1]^3 : \sum_{i=1}^{3} x^i = 1 \right\}.$$

It is worth noticing that this is also the set of Pareto-efficient payoffs in the game.

Consider now the repeated game where the majority room is publicly observed. At each stage $t = 1, 2, \ldots$, players choose their room and before stage $t+1$, the majority room is publicly announced: $U = \{L, R\}$, and

$$\ell(a) = \begin{cases} L & \text{if } \#\{i : a^i = L\} \geq 2, \\ R & \text{if } \#\{i : a^i = R\} \geq 2. \end{cases}$$

The following Folk-theorem-like result holds.

Proposition 4.1. *In the minority game $E_\infty = C$.*

4.1 de Bruijn graphs

In order to prove the results that follow some combinatorial concepts will be needed.

An oriented graph T_k is considered, where each of the 2^k nodes is labeled by a k-letter word written with the alphabet $\{L, R\}$. For $i \in \{1, \ldots, k\}$ let $x_i \in \{L, R\}$. The word $x = (x_1, \ldots, x_k)$ precedes the word $y = (y_1, \ldots, y_k)$ if $(x_2, \ldots, x_k) = (y_1, \ldots, y_{k-1})$. The word y succeeds x whenever x precedes y. Hence each node (i. e. the word associated to it) precedes only two nodes. Such a graph is called de Bruijn graph (see e. g. de Bruijn (1946) and Yoeli (1962) for some properties of these graphs). The following figure shows a de Bruijn graph T_3 based on sequences written with the alphabet $\{L, R\}$.

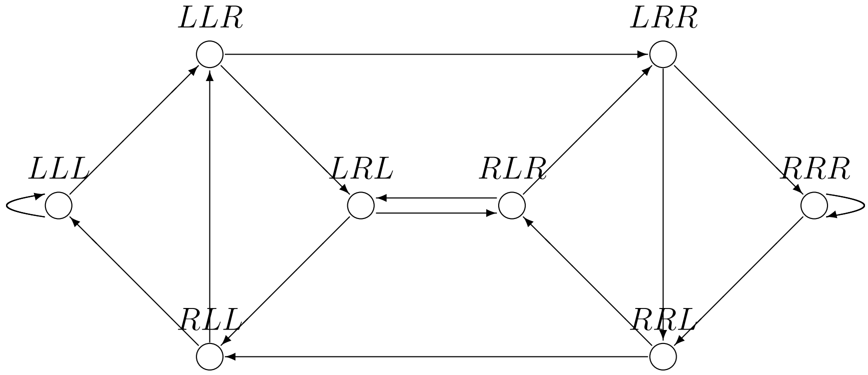

Figure 1. de Bruijn graph T_3

A proof of the following result can be found in Yoeli (1962)(see Lempel (1971) for a generalization to any finite alphabet).

Proposition 4.2. *For every p in $\{1, \ldots, 2^k\}$, there exists in the de Bruijn graph T_k a cycle with length p.*

A public k-recall strategy profile in a minority game corresponds to a cycle in the de Bruijn graph T_k, where each node in the cycle is assigned to one player or to none of them. The idea is that once the players have a public recall that corresponds to a node in the graph, they choose an action, whose consequence is that at most one player gains 1. The node is then assigned to that player. Furthermore the actions of the players at node x determine which of the two nodes that succeed x in the graph is chosen. If node x is assigned to player i, this means that i is alone in a room, therefore had she chosen the other room, she would have changed her payoff, but not the signal. Therefore if y succeeds x in the cycle of a certain strategy, and x is assigned to player i, then i cannot force the path to go to the other node z that succeeds x. Any other player, different from i can change the signal in x. A strategy is an equilibrium if there is no node in the cycle of this strategy where one of the players can change the signal and force a different cycle that gives her a higher payoff.

4.2 Some bounded-recall equilibria

We describe now some public equilibrium payoffs.

Lemma 4.3. (a) *For any $k \geq 0$, $\left(\frac{k}{k+1}, \frac{1}{k+1}, 0\right) \in \widehat{E}_k$.*

(b) *For any $k \geq 2$, $\left(\frac{k-2}{k}, \frac{1}{k}, \frac{1}{k}\right) \in \widehat{E}_k$.*

(c) *For any $k \geq 2$, $\left(\frac{k-2}{k}, \frac{2}{k}, 0\right) \in \widehat{E}_k$.*

As shown by the proofs of these results, properties of de Bruijn graphs are fundamental to determine bounded-recall equilibria in minority games. The constructions will identify in the graphs cycles that represent the equilibrium play, and will correctly assign a player to each node. These partial results allow to describe completely the set of public equilibrium payoffs for small values of k.

Proposition 4.4. (a) $E_0 = \widehat{E}_0 = \{(1,0,0),(0,1,0),(0,0,1)\}$.

(b) $E_1 = \widehat{E}_1 = \widehat{E}_0 \cup \left\{\left(\frac{1}{2},\frac{1}{2},0\right),\left(\frac{1}{2},0,\frac{1}{2}\right),\left(0,\frac{1}{2},\frac{1}{2}\right)\right\}$.

(c) $\widehat{E}_2 = \widehat{E}_1 \cup \left\{\left(\frac{1}{3},\frac{2}{3},0\right),\left(\frac{1}{3},0,\frac{2}{3}\right),\left(0,\frac{1}{3},\frac{2}{3}\right),\left(\frac{2}{3},\frac{1}{3},0\right),\left(\frac{2}{3},0,\frac{1}{3}\right),\left(0,\frac{2}{3},\frac{1}{3}\right)\right\}$.

Proposition 4.5. *E_2 differs from $\widehat{E}_2$ since $\left(\frac{1}{3},\frac{1}{3},\frac{1}{3}\right) \in E_2 \setminus \widehat{E}_2$.*

Note that for $k \leq 2$, all public equilibrium payoffs are on the boundary of the triangle C. A direct consequence of Lemma 4.3 is that $\left(\frac{1}{3},\frac{1}{3},\frac{1}{3}\right) \in \widehat{E}_3$, so when the recall is $k \geq 3$ there exists a public equilibrium payoff in the interior C.

4.3 Convergence of $\widehat{E}_k$ for the minority game

When the minority game is repeated with public monitoring of the majority room, then convergence holds.

Theorem 4.6. *In the repeated minority game where the public signal is given by the majority room,*

$$\lim_{k\to+\infty} \widehat{E}_k = E_\infty = C.$$

The construction, like standard Folk-theorems, uses a main path and punishments. Players agree on a cycle over the set of stage-Nash equilibria leading approximately to the target payoff. Since only stage-equilibria are played, a deviation that does not modify the signals is not profitable. When players see unexpected signals, they punish the deviator by staying for a long time in the same room they were in at the deviation stage. The punishment is effective since only a player who gets a zero payoff (i.e., is not alone in a room) can modify the signal. Before the deviation signal leaves the public recall, players re-write it in the recall. Two players can do so by playing the same action thus controlling the public signal. The detailed construction is given in Section 6.

Remark 4.7. Theorem 4.6 easily extends to a $2n+1$-player minority game (each player has to choose between L and R and receives a payoff of 1 if she is in the minority room and zero otherwise). However, the proof heavily relies on the specific properties of the game and signal function. Since convergence of $\widehat{E}_k$ to E_∞ is not always guaranteed, a challenging and open problem is to characterize $\lim_k \widehat{E}_k$.

4.4 Private equilibria and minority game

While for the minority game with public monitoring of the majority room, the set of public equilibrium payoffs $(\widehat{E}_k)_k$ monotonically converges to E_∞, not much is known about the behavior of E_k. Monotonicity of this sequence or convergence of E_k to E_∞ are still open problems. The following proposition shows that for some k, E_k is not a subset of E_∞.

Proposition 4.8. *In the minority game with public monitoring of the majority room,* $(3/7, 3/7, 0) \in E_3$ *and thus* $E_3 \not\subset E_\infty$.

This proposition is proved by constructing explicitly an equilibrium $\sigma = (\sigma^1, \sigma^2, \sigma^3)$ of Γ_3 with payoff $(3/7, 3/7, 0)$. The proof is quite lengthy and involved and seems to indicate that more general results in this direction are quite hard to obtain.

5 A guessing game

This section considers a game where player 1 wants to guess the action of player 2 in a repeated game with trivial monitoring. Player 1's payoffs are given by:

	L	R
T	1	0
B	0	1

and player's 2 payoffs are always 0. This ensures that equilibrium exists in pure strategies, an alternative way is to consider the zero-sum game and look at the $\min_2 \max_1$ value.

Under trivial monitoring if both players have the same recall, then $E_k = E_\infty = (1, 0)$: in zero-sum terminology, the $\min_2 \max_1$ of the game is 1. In fact, a strategy σ^2 of player 2 is just a cyclic sequence of L and R, and player 1 needs only to play a sequence σ^1 that mimics σ^2 with T replacing L and B replacing R.

The aim of this section is to show how giving a bit more complexity to player 2 dramatically affects equilibrium payoffs, or equivalently decreases the $\min_2 \max_1$ of the game. This is in contrast with usual results on zero-sum games with bounded complexity (Lehrer (1988), Ben-Porath (1993)) where players of approximately the same complexity, more or less guarantee the value of the game. This contrast is essentially due to the use of pure strategies.

Two variations of the above game will be examined. In Subsection 5.1, both players have the same recall, but the actions available to player 2 are duplicated. This may be viewed as follows: in the above game player 2 is allowed to choose, in addition to her action, a letter from the alphabet $\{1, 2\}$ and write it down on a notepad without affecting the payoffs whatsoever. In the repeated game with unbounded recall and pure strategies, allowing a player to write notes does not enlarge her strategy set, but it certainly does in the bounded recall case. The game studied in subsection 5.1 can thus be viewed as a modification of the one-shot guessing game or as a modification of the set of strategies available in the repeated game.

In subsection 5.2, player 2 is be given an extra bit of recall with respect to player 1.

Since monitoring is trivial, the sequences of actions played by the players evolve independently. Consider one player with trivial monitoring, two actions, L and R, and bounded recall k. Using a private strategy with recall k leads this player to play a periodic sequence of L and R which corresponds to a cycle in the de Bruijn graph T_k. By Proposition 4.2, its period can be any integer between 1 and 2^k. When the period is exactly 2^k the sequence is called a de Bruijn sequence of recall k.

5.1 Player 2 has same the recall but more actions than player 1

Consider the following two-player game, where: $A^1 = \{T, B\}$, $A^2 = \{L_1, L_2, R_1, R_2\}$, monitoring is trivial, player 2 has payoff 0 and and the payoffs for player 1 are given by:

	L_1	L_2	R_1	R_2
T	1	1	0	0
B	0	0	1	1

Proposition 5.1. *In the above game*

(a) $E_\infty = \{(1, 0)\} = E_0$,

(b) *for* $k \geq 1$, $(1/2, 0) \in E_k$.

The idea of the proof is the following: the most complex sequence player 1 can predict (or play) is a de Bruijn sequence of recall k on $\{L, R\}$. Using the labels of her actions L_1, L_2, player 2 can follow a more complicated pattern, namely, cyclically play one after another a deBruijn sequence of recall k on $\{L_1, L_2\}$ and a de Bruijn sequence of recall k on $\{R_1, R_2\}$. Hence player 1 cannot guess correctly this sequence more than half of the time.

5.2 Player 2 has as many actions but larger recall than player 1

Consider the game with trivial monitoring where player 2 has payoff 0 and the payoffs for player 1 are given by:

	L	R
T	1	0
B	0	1

Assume that player 1 has recall k and that player 2 has recall k'. How well can player 1 guess the actions of player 2? The worst equilibrium payoff for player 1 is given by

$$v_{k,k'} := \min_{\sigma^2 \in \Sigma^2_{k'}} \max_{\sigma^1 \in \Sigma^1_k} \gamma^1(\sigma^1, \sigma^2).$$

It is plain that $v_{k,k'}$ is non-decreasing in k and non-increasing in k', and $\frac{1}{2} \leq v_{k,k'} \leq 1$. If $k \geq k'$, player 1 will be able to mimic the sequence played by player 2, hence $v_{k,k'} = 1$. Focus now on $k' = k + 1$, and define

$$f(k) := v_{k,k+1} = \min_{\sigma^2 \in \Sigma^2_{k+1}} \max_{\sigma^1 \in \Sigma^1_k} \gamma^1(\sigma^1, \sigma^2).$$

Piccione and Rubinstein (2003, Section 5, Footnote 5) noticed that if player 2 plays a de Bruijn sequence of recall $k + 1$, then player 1 with recall k must "have a frequency of mistakes of at least $1/(2(k + 1))$." This implies that

$$f(k) \leq 1 - \frac{1}{2(k+1)}.$$

Notice that in the definition of $f(k)$ player 2 may use any strategy with recall $k + 1$, not necessarily de Bruijn sequence of recall $k + 1$.

Lemma 5.2. $f(0) = f(1) = 1/2$, *and* $f(2) = 4/7$.

Given the above lemma, intuition would lead to conjecture that f is non-increasing, and that $(f(k))_k$ converges to a limit, that would represent the asymptotic advantage for player 2 to possess a recall strictly larger than the recall of player 1. The following proposition disproves this intuition.

Proposition 5.3.

$$\lim_{k \to \infty} f(k) = 1/2.$$

The proof uses some arithmetic results. Since there exists a prime number p, $2^k < p < 2^{k+1}$ and a cycle of length p in T_{k+1}, player 2 chooses in the de Bruijn graph with recall $k + 1$ a cycle of length p. For every sequence chosen by player 1 of period $q \leq 2^k$, gcd(p, q)=1 so the period of the joint sequence of actions is pq which is large with respect to q. This implies that player 1 cannot coordinate too often with player 2.

6 Proofs

Section 3. Equilibrium payoffs with bounded recall

Proof of Proposition 3.1. (a) This kind of result is common in the literature on games with bounded complexity (see e.g., Neyman (1998), Ben-Porath (1993), Lehrer (1988, 1994)) and relies on a usual dynamic programming argument. Let σ be an equilibrium of $\widehat{\Gamma}_k$. For each player i, finding a best-reply in Σ^i to σ^{-i} amounts to solve a dynamic programming problem, where the state space is U^k, the set of public histories of length k, the action space is A^i, the payoff in state $h = (u_1, \ldots, u_k)$, given action a^i is $g^i(a^i, \sigma^{-i}(h))$, and the new state is $(u_2, \ldots, u_k, \ell(a^i, \sigma^{-i}(h)))$. It is well known (see Blackwell (1962)) that there exists a stationary optimal strategy. Thus, the best reply of player i to a profile of public strategies with recall k is a public strategy with recall k (see Abreu and Rubinstein (1988, Lemma 1)). Therefore, σ is a uniform equilibrium of Γ_∞.

(b) This follows directly from the previous point. The game $\widehat{\Gamma}_k$ is a *subgame* of Γ_k in the sense that the set of strategies of each player in $\widehat{\Gamma}_k$ is a subset of the set of strategies of this player in Γ_k. Let then σ be a strategy profile in $\widehat{\Gamma}_k$, if σ is not an equilibrium of Γ_k, then a player i has a profitable deviation in $\Sigma_k^i \subset \Sigma^i$, thus σ is not a uniform equilibrium contradicting the previous point.

(c) The argument is similar to the one used for point (b), $\widehat{\Gamma}_k$ is a subgame of $\widehat{\Gamma}_{k+1}$: any strategy with recall k can be played in the game with recall $k+1$. So, if a strategy profile σ in $\widehat{\Gamma}_k$ is not an equilibrium of $\widehat{\Gamma}_{k+1}$, then some player i has a profitable deviation in $\widehat{\Sigma}_{k+1}^i \subset \Sigma^i$, thus σ is not a uniform equilibrium contradicting point (a).

□

Proof of Proposition 3.2. In the game with unbounded recall, players may get the convex hull of E_0 by playing stage-Nash equilibria with the appropriate frequencies. Since monitoring is trivial, deviations go unnoticed, thus, at equilibrium, stage-Nash are played at (almost) all stages. Of course, this convexification of E_0 cannot be obtained by stationary strategies.

When signals are constant, for each k there is a unique public history of length k. Therefore $\widehat{\Gamma}_k$ is the game where each player is restricted to always play the same action. This game is thus identical (in strategies and payoffs) to the one-shot game.

□

Proof of Proposition 3.3. (a) Note that $(0, 0, 1) \in E_\infty$ by the classical folk theorem. This payoff is feasible by a sequence of action profiles where player 1 plays T at each stage and player 2 alternates between L and R with respective frequencies $\left(\frac{1}{1+\pi}, \frac{\pi}{1+\pi}\right)$. Player 1 punishes deviations from player 2 by playing B and player 2 punishes by playing R.

(b) Consider now bounded recall strategies. Take k in $\mathbb{N}$ and (x, y, z) in E_k $(= \widehat{E}_k)$. Since the play induced by a pure strategy profile with bounded recall is periodic, the average frequencies λ_1, λ_2, λ_3, λ_4 of the pure action profiles (respectively of (T, L), (T, R), (B, L), (B, R)) are non-negative rational numbers, summing up to one. We have:

$$
\begin{aligned}
x &= \lambda_1 \pi - \lambda_2, \\
y &= -\lambda_1 \pi + \lambda_2, \\
z &= \lambda_1 + \lambda_2.
\end{aligned}
$$

By individual rationality $x \geq 0$ and $y \geq 0$, so $\lambda_1 \pi = \lambda_2$. Since π is irrational, this implies $\lambda_1 = \lambda_2 = 0$. So $(x, y, z) = (0, 0, 0)$. Since $(0, 0, 0)$ is an equilibrium payoff, one obtains $E_k = \{(0, 0, 0)\}$ for each k.

□

Proof of Proposition 3.4. Consider a pure equilibrium of the repeated game with unbounded recall. If at some stage (M, L) is played, then player 2 may play R at this stage and get a payoff of 3 instead of 2, without any further consequence because the signal induced by (M, R) is the same as the signal induced by (M, L). Thus, in equilibrium, (M, L) cannot be played with positive frequency, and therefore $E_\infty \subset \{(x, y) \in \mathbb{R}^2, x + y \leq 3\}$.

Fix a positive integer k and define $\sigma = (\sigma^1, \sigma^2) \in \Sigma_k$ as follows.

- σ^2 plays L at each stage whatever happens.
- σ^1 plays B_1 at stage 1, and is defined via a main phase and a transition phase. After stage 1, player 1 using σ^1 says that she is in the main phase if and only if (all public signals in her recall equal u, and the last action played by player 1 is not T). If this condition is not satisfied, then player 1 says that she is in the transition phase.
  - In the transition phase, player 1 plays B_1 if her last k actions all equal T, and plays T otherwise.
  - In the main phase, player 1 induces the following periodic sequence of actions

$$\underbrace{B_1 B_1 \dots B_1}_{k\text{times}} \underbrace{B_2 B_2 \dots B_2}_{k\text{times}} \underbrace{MM \dots M}_{k\text{times}} \underbrace{B_1 B_1 \dots B_1}_{k\text{times}} \underbrace{B_2 B_2 \dots B_2}_{k\text{times}} \underbrace{MM \dots M}_{k\text{times}} \dots$$

    That is, player 1 plays B_2 (resp. M, resp. B_1) if her last k actions are all B_1 (resp. B_2, resp. M), and repeats her last action otherwise.

  This ends the definition of σ.

Under σ, the play remains forever in the main phase, inducing the payoff

$$\frac{1}{3}\, g(B_1, L) + \frac{1}{3}\, g(B_2, L) + \frac{1}{3}\, g(M, L) = \frac{1}{3}(2, 1) + \frac{1}{3}(2, 1) + \frac{1}{3}(2, 2) = \left(2, \frac{4}{3}\right).$$

It is now necessary to check that σ is an equilibrium of Γ_k.

The strategy σ^1 is obviously a best response against σ^2 because the maximal payoff for player 1 is 2. Let now τ^2 be any strategy of player 2 with recall k and assume that $\gamma^2(\sigma^1, \tau^2) > 1$. There must exist some first stage $\bar{t}$ where player 1 plays M. Then necessarily the following happened:

stage $\rightarrow$	$\bar{t} - 2k$	$\dots$	$\bar{t} - (k+1)$	$\bar{t} - k$	$\dots$	$\bar{t} - 1$	$\bar{t}$
action P1	B_1	$\dots$	B_1	B_2	$\dots$	B_2	M
public signal	u	$\dots$	u	u	$\dots$	u	

This implies that player 2 has played L and the signal was u at every stage $\bar{t}-2k,\ldots,\bar{t}-1$. Since $2k > k$ and τ^2 has recall k, player 2 using τ^2 will play L at every stage $t \geq \bar{t}-2k$. So $\gamma^2(\sigma^1,\tau^2) = 4/3 = \gamma^2(\sigma)$. Consequently σ^2 is a best response against σ^1 in the game with private strategies and recall 3, and $(2,4/3) \in E_k$. Hence $(2,4/3) \in \cap_{k\geq 1} E_k \backslash E_\infty$. □

Section 4. A minority game

Proof of Proposition 4.1. This follows directly from the characterization given in Tomala (1998, Theorem 5.1, page 104), , but a simple direct proof will be provided. First note that, since C is the convex hull of Nash equilibrium payoffs of the one-shot game, then $C \subset E_\infty$. Given any point x in C, one can find a sequence of Nash equilibria $(a_t)_t$ of the minority game, such that the average payoff vector along this sequence converges to x. Then, the strategy profile such that for each player i and stage t, player i plays a_t^i at stage t, irrespective of the history, is clearly a uniform equilibrium with payoff x.

To get the converse, note that there are two types of action profiles: either two players are in the same room and the profile is an equilibrium of the MG, or the three players are in the same room. In the latter case, each player has a profitable deviation (she prefers to switch room) and further this deviation does not change the majority room, i.e., the public signal. If at a strategy profile the three players are in the same room on a non-negligible set of stages, then player 1 can switch rooms at these stages. This increases her payoff at these stages without affecting public signals, hence without affecting the behavior of the other players. Such a strategy profile cannot be a uniform equilibrium and therefore $E_\infty \subset C$. □

Consider a minority game with recall k, and its associated de Bruijn graph T_k. Call m-cycle a cycle of length m, call *stable* the cycles where all the nodes have the same number of L's, and call *main* all the stable cycles containing the strings $L\cdots LR\cdots R$ or $R\cdots RL\cdots L$. There are $k-1$ main k-cycles and 2 main 1-cycles.

Proof of Lemma 4.3. (a) Consider the $(k+1)$-cycle that contains $R\cdots R$ and all the nodes of the main k-cycle that contains $R\cdots RL$. In equilibrium, players cycle on this $(k+1)$-cycle and elsewhere they go to this cycle as fast as possible.

Assign node $R\cdots R$ to player 2 and all the other nodes in the graph to player 1.

Player 1 can deviate only on $R\cdots R$, and she has no incentive to do it, because that would induce a cycle on the node RRR, that is assigned to player 2. Player 2 can deviate anywhere else, but she has no incentive to do it, since she cannot find a cycle that contains $R\cdots R$ and is shorter than the equilibrium cycle.

Player 3 can always deviate, but, since she would get a zero payoff anyway, she has no incentive to deviate.

The following figure shows the above equilibrium for $k=3$.

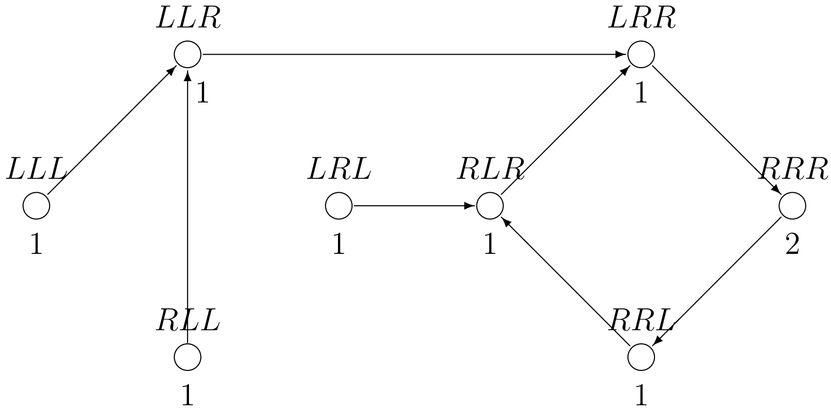

Figure 2. Equilibrium with 3-recall and payoff $\left(\frac{3}{4}, \frac{1}{4}, 0\right)$

(b) In each of the main cycles assign node $R \cdots RL \cdots L$ and $L \cdots L$ to player 2, nodes $L \cdots LR \cdots R$ and $R \cdots R$ to player 3, and the other nodes to player 1. In equilibrium players cycle on the main k-cycles and elsewhere they move to them fast. Player 1 has no incentive to deviate on the nodes assigned to 2 (or 3), because she would move to another node of 2 (or 3), and she would increase the distance to a node that is assigned to herself. For instance, if $k = 3$, deviating in LLL (RRR) would bring back to LLL (RRR), and hence increase the distance to the next 1-node, deviating in LLR (RRL) would increase the distance to the next 1-node from 1 to at least 3, deviating in RLL (LRR) would increase the distance to the next 1-node from 2 to at least 3.

Given the disposition on the main cycles of the nodes assigned to her, on the nodes assigned to 1 or 3, player 2 cannot find a shorter cycle that could induce her to deviate. The same is true for player 3.

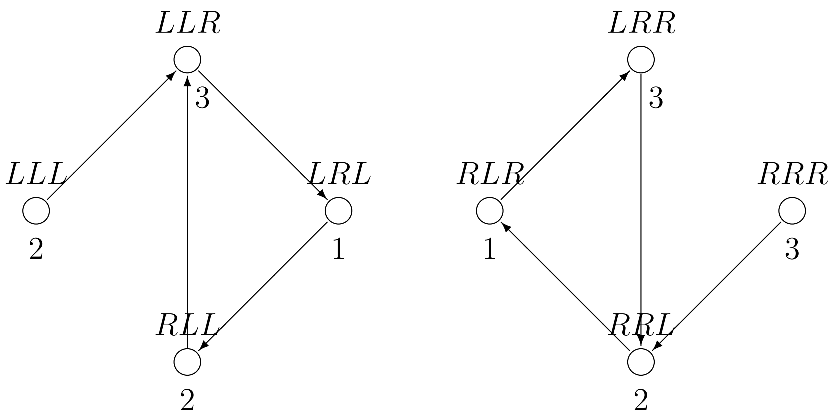

Figure 3. Equilibrium with 3-recall and payoff $\left(\frac{1}{3}, \frac{1}{3}, \frac{1}{3}\right)$

(c) As above. Just assign to player 2 the nodes that were assigned to 3, and repeat the argument.

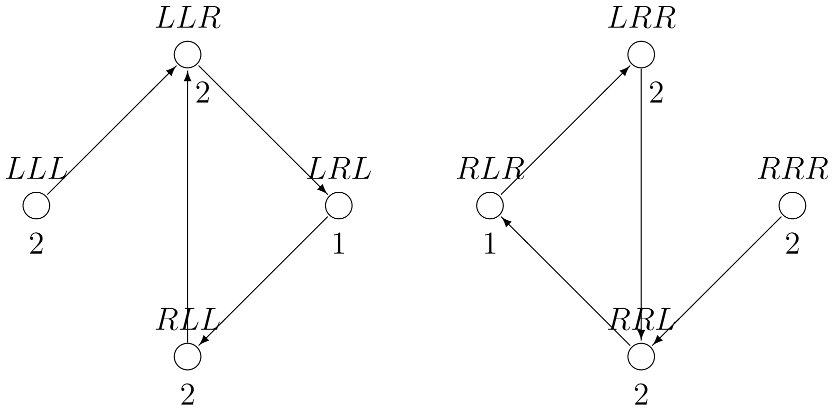

Figure 4. Equilibrium with 3-recall and payoff $\left(\frac{1}{3}, \frac{2}{3}, 0\right)$

$\square$

Proof of Proposition 4.4. (a) The only possible Nash equilibria of both E_0 and $\widehat{E}_0$ are repetitions of the same Nash equilibrium of the stage game.

(b) By Proposition 3.1, $\widehat{E}_0 \subset \widehat{E}_1$. Furthermore by Lemma 4.3(c) the payoffs $\left(\frac{1}{2}, \frac{1}{2}, 0\right), \left(\frac{1}{2}, 0, \frac{1}{2}\right), \left(0, \frac{1}{2}, \frac{1}{2}\right) \in \widehat{E}_1$. No other equilibrium payoff can be obtained with recall 1, since the maximal length of a cycle in the de Bruijn graph T_1 is 2.

(c) By Proposition 3.1, $\widehat{E}_1 \subset \widehat{E}_2$. Furthermore by Lemma 4.3(c) the payoffs $\left(\frac{1}{3}, \frac{2}{3}, 0\right), \left(\frac{1}{3}, 0, \frac{2}{3}\right), \left(0, \frac{1}{3}, \frac{2}{3}\right), \left(\frac{2}{3}, \frac{1}{3}, 0\right), \left(\frac{2}{3}, 0, \frac{1}{3}\right), \left(0, \frac{2}{3}, \frac{1}{3}\right) \in \widehat{E}_2$.

No other equilibrium payoff can be obtained with recall 2.

First it will be proved that $\left(\frac{1}{3}, \frac{1}{3}, \frac{1}{3}\right) \notin \widehat{E}_2$. In fact the maximal length of a cycle in the de Bruijn graph T_2 is 4. Hence, in order to obtain such a payoff in equilibrium, the players would have to cycle on a 3-cycle of T_2, and each node should be assigned to a different player. There are only two such cycles. Take for instance the cycle $LL \to LR \to RL$, and assume that these nodes are assigned to players 1, 2, and 3, respectively. Then player 2 will want to deviate in RL, and player 3 will want to deviate in LR. An analogous argument can be used for the cycle $LR \to RR \to RL$.

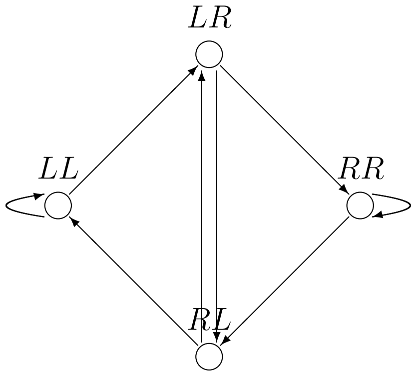

Figure 5. de Bruijn graph T_2

Consider the only 4-cycle in the graph T_2, namely, $LL \to LR \to RR \to RL$. This graph cannot give a payoff $\left(\frac{1}{2}, \frac{1}{4}, \frac{1}{4}\right)$ or its permutation. In fact if (i_1, i_2, i_3, i_4) denotes the strategy that assigns LL to player i_1, LR to player i_2, etc., then none of the configuration that give a payoff $\left(\frac{1}{2}, \frac{1}{4}, \frac{1}{4}\right)$ is an equilibrium:

- $(1, 1, 2, 3)$ is not an equilibrium, since player 3 would deviate in LR,
- $(1, 1, 3, 2)$ is not an equilibrium, since player 2 would deviate in LR,
- $(1, 2, 1, 3)$ is not an equilibrium, since player 3 would deviate in LR, and player 2 would deviate in RL,
- $(1, 3, 1, 2)$ is not an equilibrium, since player 2 would deviate in LR, and player 3 would deviate in RL,
- $(1, 2, 3, 1)$ is not an equilibrium, since player 1 would deviate in LR,
- $(1, 3, 2, 1)$ is not an equilibrium, since player 1 would deviate in LR.

$\square$

Proof of Proposition 4.5. First remark that with private strategies player i can cycle on RLL by using a strategy that relies on her own actions only. Therefore consider the strategy profile obtained by cycling:

$$\begin{array}{ccc} R & L & L \\ L & R & L \\ L & L & R \end{array}$$

where the i-th row indicates the strategy of the i-th player. This is clearly an equilibrium of Γ_2, since it is a repetition of one-stage Nash equilibria, and its payoff is $\left(\frac{1}{3}, \frac{1}{3}, \frac{1}{3}\right)$. Furthermore since the strategy of each player is based on her own actions, a deviation of one player does not change the actions of the other players. By Proposition 4.4(c) the payoff $\left(\frac{1}{3}, \frac{1}{3}, \frac{1}{3}\right)$ is not in $\widehat{E}_2$. $\square$

Proof of Theorem 4.6. Let $m \geq 2$ be an integer and let C_m be the set of vectors of $x \in C$ with rational components of the form $x^i = m^i/m$ with $m^i \geq 2$ integers. Then C_m converges to C as m goes to infinity i.e. $\sup_{x\in C} \inf_{y\in C_m} \|x - y\|$ goes to 0 as m goes to infinity. Therefore Theorem 4.6 follows from Lemma 6.1 below. $\square$

Lemma 6.1. *For every integers $m \geq 2$ and $K \geq 2m$, $C_m \subset \widehat{E}_k$ for $k = Km$.*

The following terminology will be used in the proof of Lemma 6.1. Call *word* any finite sequence of signals. Given two words $u = (u_1, \ldots, u_p)$ and $v = (v_1, \ldots, v_q)$, denote by uv the concatenated word $uv = (u_1, \ldots, u_p, v_1, \ldots, v_q)$. A word of length k is called a *public recall*. Given a public recall M, a word u of length $l \leq k$ is called a sub-word of M if there exist two words v, w such that $M = vuw$. The word consisting of $L \ldots L$, q times is denoted L^q. If u is a sub-word of M, define the *position* of u in M as the rank of the first letter of u. For instance, if M begins with u, then u has position 1; if M ends with u, then u has position $k - l + 1$.

Proof of Lemma 6.1. Let $m \geq 2$ be an integer and $x \in C_m$. The aim is to construct a strategy profile σ with payoff x which is an equilibrium of $\widehat{\Gamma}_k$ for $k = Km$, with $K \geq 2m$. The strategy construction is in a folk-theorem spirit. First the right payoff is obtained by playing an adapted main path. In case of detected deviation, punishments have to be performed. Because of finite recall, the evidence that a deviation occurred may disappear from the recall. To get a deviating player to be punished forever, players are asked to rewrite periodically a word in the public recall, indicating that a deviation has occurred and which actions should be used to punish. This construction relies heavily on properties of the minority game and the minority room as a signal. The following properties will be used extensively.

- A player who is in the minority room at some stage cannot change the signal at that stage. This implies that a player who gets a payoff of 1 at a given stage has no incentive to deviate at that stage since it can only decrease the stage-payoff and has no impact whatsoever on the future.

- The main path is be constructed so that at each stage a Nash equilibrium of the one-shot game is played. Thus at each stage there is one player in the minority room and the other two players are in the majority room, both receiving a payoff of zero. If the signal changes, that means that one of the two players in the majority room deviated, but the public signal does not tell who did. A simple way to punish the deviating player without knowing her identity is to apply the following policy: "If I see a wrong signal at stage t, then I remain in the room where I was at stage t." This insures that the deviating player, who was in the majority room when the deviation was detected, remains in the majority room as long as the punishment phase lasts.

- Any payoff vector can be obtained by two actions profiles giving different public signals (just exchange L and R).

- Two players can write any word in the public recall, whatever the behavior of the third player is.

Pick now a point $x = (x^i)_i \in C_m$. Then $x = \sum_i x^i e(i)$, where for each $i \in N$, $x^i = m^i/m$, with $m^i \geq 2$, so $x^i \geq 2/m$. Let $H = (a_1^*, \ldots, a_m^*) \in A^m$ be a sequence of action profiles of length m such that

1. the average payoff along H is

$$x = \frac{1}{m} \sum_{t=1}^{m} g(a_t^*),$$

2. the public history $(\ell(a_1^*), \ldots, \ell(a_m^*))$ associated to H is $L \ldots L$, m times.

Such a sequence exists, it suffices to play a sequence of Nash equilibria of the MG such that player i gains 1 exactly m^i times and the majority room is always L. For each room $r \in \{L, R\}$, let $\bar{r}$ be the other room, and, if a is an action profile, let $\bar{a}$ be the action profile where every player has switched room. Let $\bar{H} \in A^m$ be the sequence obtained from H by switching rooms: $\bar{H} = (\bar{a}_1^*, \ldots, \bar{a}_m^*)$. The main path will be the periodic repetition of the sequence $H\bar{H}$. Here is how to construct a profile of strategies of recall k that generates this periodic sequence of action profiles.

Let $W := L^m$ be the word induced by H. A word w is a sub-word of W if $w = L^q$ with $0 \leq q \leq m$. If a periodic repetition of $H\bar{H}$ is played, at each stage the public recall ends by a word of the type $\bar{W}w$ or $W\bar{w}$ with w sub-word of W (possibly of length 0). Call such words *end-words*. An end-word writes either $L^m R^q$ or $R^m L^q$, $0 \leq q < m$. The aim is to play a periodic repetition of $H\bar{H}$. In order to do that, at each stage knowledge of the end-word is sufficient to know what action profile should be played at the next stage. Thus, letting E be the set of end-words, there exists a mapping f which maps E to pure Nash equilibria of the MG and such that for each end-word e, $f(e) = (f^i(e))_{i \in N}$ is the action profile that follows e in the periodic repetition of $H\bar{H}$.

Consider now deviations. After each end-word e, $f(e)$ should be played. On the main path $f(e)$ induces a winning player $i(e)$ and a signal $r(e)$. If $\bar{r}(e)$ is observed, then some player $j \neq i(e)$ has deviated. Let us call *deviation-word*, a word of the type $e\bar{r}(e)$: a deviation word writes either $L^m R^q L$ or $R^m L^q R$, $0 \leq q < m$. If a deviation-word $e\bar{r}(e)$ appears in the recall, the strategy prescribes to keep on playing $f(e)$ as long as the position of $e\bar{r}(e)$ is greater than $2m$. During this punishing phase, the signal is completely controlled by the punished player, this player could then write in the recall *another* deviation-word $e'\bar{r}(e')$. To prevent other end-words to appear in the recall, if L^{m-1} (resp. R^{m-1}) appears, all players must play R (resp. L). Finally, when the position of $e\bar{r}(e)$ becomes less than or equal to $2m$, the players must rewrite this word in the recall by all playing the same actions for an appropriate number of times.

The exact definition of the strategy profile σ is given now.

- **Initialization**. At each stage $t \leq k$, each player plays L. At each stage $t > k$, apply the following points.
- **Main path**. If the recall contains no deviation-word and ends by the end-word e, each player i plays $f^i(e)$.
- **Early punishments**.
  - If the recall contains a deviation-word $e\bar{r}(e)$ whose position is greater than $2m$, and if the recall does not end by L^{m-1} or by R^{m-1}, then each player i plays $f^i(e)$.
  - If the recall contains a deviation-word $e\bar{r}(e)$ whose position is greater than $2m$, and if the recall ends by L^{m-1}, then each player i plays R.
  - If the recall contains a deviation-word $e\bar{r}(e)$ whose position is greater than $2m$, and if the recall ends by R^{m-1}, then each player i plays L.
- **Late punishments**. If the recall contains a deviation word $e\bar{r}(e) = L^m R^q L$ with $0 \leq q < m$, let p be its position.
  - If $m < p \leq 2m$, then each player i plays L.
  - If $m - q < p \leq m$, then each player i plays R.
  - If $p = m - q$, then each player i plays L.
- **Other memories**. For all other memories, each player plays L.

It remains to prove that the above-defined strategy profile σ has payoff x and is an equilibrium of $\widehat{\Gamma}_k$.

If all players play this strategy, the public recall after stage k is L^k, thus it ends by an end-word e. The next action profile is then $f(e)$ and the public recall still ends by an end-word so the strategy uses f again. By construction of f, this strategy profile generates the periodic repetition of $H\bar{H}$ and the payoff is indeed x.

Suppose that player i deviates. If the deviation never changes the signals, then player i changes action only at stages where she was in the minority room. Therefore she loses payoff at these stages and does not affect the behavior of other players. Such a deviation is thus not profitable.

Suppose now that player i changes the signal at some stage, therefore i is in the majority room at this stage. This generates a deviation-word $e\bar{r}(e)$. As long as the position of $e\bar{r}(e)$ is greater than $2m$, the other players play $f(e)$ so player i receives a payoff of zero, except if she generates words of the type L^{m-1} or R^{m-1}. In such cases, the other players will play both R or both L. Such situations appear at most every m stages. So, the only opportunities to player i to gain a payoff of 1 are when other players rewrite the deviations word (at most $2m$ stages), and once every m stages for $k-2m$ stages. The average payoff for player i is thus no more than

$$\begin{aligned}\frac{2m+\frac{k-2m}{m}}{k} &= \frac{2m+K-2}{km} \\ &\le \frac{2}{k}+\frac{1}{m} \\ &\le x^i,\end{aligned}$$

since $x^i \ge 2/m$, and $K \ge 2m$. □

Proof of Proposition 4.8. We construct an equilibrium $\sigma = (\sigma^1, \sigma^2, \sigma^3)$ of Γ_3 with payoff $(3/7, 3/7, 0)$. Given strategies of recall 3, the action played by a player at some stage depends only on her last 3 actions and on the last 3 public signals. The last $3 \wedge t$ actions or signals at time t will be called *available*.

The profile σ is defined as follows:

(a) If at least one available public signal is R, then σ recommends to each player to switch room, i.e. to play L if she played R at the previous stage, and vice-versa.

(b) Assume now that all available public signals are L.

(b1) Regarding the first three stages, as long as the public signal is L, σ recommends to play as follows:

stage →	1	2	3
P1	L	L	R
P2	R	L	L
P3	L	R	L

For example, the symbol R in line P3 means that at stage 2, σ^3 asks player 3 to play R if the public signal of stage 1 was L.

(b2) At every stage $t \ge 3$, if the last 3 public signals are L, each player $i \in \{1,2,3\}$ plays the action $f^i(a^i_{t-3}, a^i_{t-2}, a^i_{t-1}) \in \{L, R\}$ where $a^i_{t'}$ denotes the action played by player i at stage t' and the functions f^1, f^2, f^3 are described below.

last own actions	P1	P2	P3
LLL	R	R	L
LLR	R	L	L
LRL	L	R	L
LRR	L	L	L
RLL	L	L	L
RLR	L	R	L
RRL	R	L	L
RRR	L	L	L

Figure 6

At the intersection of column P2 and line RLL, the symbol L means that $f^2(RLL) = L$, i.e., at any stage $t \geq 3$, if the last 3 public signals were L, and the last actions played by player 2 were R (at stage $t-3$), L (at stage $t-2$), and L (at stage $t-1$), then player 2 following σ^2 should play L. This ends the definition of σ.

The proof is complete once Lemma 6.2 below is proved. □

Lemma 6.2. (a) *The payoff induced by σ is* $(3/7, 3/7, 0)$.

(b) *The strategy σ is an equilibrium of* Γ_3.

Proof. (a) Assume that σ is played. The induced play can be represented as follows.

stage →	1	2	3	4	5	6	7	8	9	10	11	12	13	
action P1	L	L	R	R	L	R	L	L	L	R	R	L	R	...
action P2	R	L	L	L	R	L	R	R	L	L	L	R	L	...
action P3	L	R	L	L	L	L	L	L	L	L	L	L	L	...
public signal	L	L	L	L	L	L	L	L	L	L	L	L	L	...

Figure 7

The action of a player in the minority room, if any, is emphasized with a circle. The public signal is L at every stage, the induced play eventually has period 7 (one can see a period from stage 3 to stage 9), and the induced payoff is (3/7,3/7,0).

(b) This part is a direct consequence of the next three lemmata, where the best reply condition is checked for every player.

□

Lemma 6.3. *In Γ_3, σ^3 is a best reply against σ^{-3}.*

Proof. Let τ^3 be any strategy of player 3 in Σ_3^3. It is necessary to prove that $\gamma^3(\tau^3, \sigma^{-3}) \leq \gamma^3(\sigma) = 0$. Assume in the sequel that (τ^3, σ^{-3}) is played, and distinguish two cases.

Case 1. Assume that the sequence of public signals never contains the symbol R. Then the sequence of actions played by players 1 and 2 is the same as in Figure 7. So at the stages 3,4,5,6, player 3 is playing L (otherwise the public signal will be R at some stage). Since τ^3 has recall 3, it implies that player 3 will play L at every stage $t \geq 3$. Since L is at each stage the majority room, $\gamma^3(\tau^3, \sigma^{-3}) = 0$.

Case 2. Assume that at some stage the public signal is R. Consider the first stage $\bar{t}$ where this happens. Up to stage $\bar{t}$, the actions played by player 1 and 2 correspond to

Figure 7, so at stage $\bar{t}$ it is not possible that both players 1 and 2 play R. Consequently, at stage $\bar{t}$: either (players 1 and 3 play R and player 2 plays L), or (players 2 and 3 play R and player 1 plays L). Recall now that σ^1 and σ^2 ask players 1 and 2 to change rooms whenever one of the available signals is R.

As long as one of the available public signals is R, players 1 and 2 will exchange rooms at each stage and, since players 1 and 2 are not in the same room, the payoff for player 3 will be zero. So to get out of this punishment phase, player 3 has to play three consecutive times L in order to induce three consecutive signals L. So it is possible to assume w.l.o.g. that there exists a stage t where the situation is as follows:

stage $\rightarrow$	t	$t+1$	$t+2$	$t+3$
action P1	L	R	L	$L^{(a)}$
action P2	R	L	R	$R^{(b)}$
action P3	L	L	L	
public signal	L	L	L	

or

stage $\rightarrow$	t	$t+1$	$t+2$	$t+3$
action P1	R	L	R	$L^{(c)}$
action P2	L	R	L	$R^{(d)}$
action P3	L	L	L	
public signal	L	L	L	

Figure 8

$^{(a)}$ because $f^1(L,R,L)=L$ (see Figure 6),
$^{(b)}$ because $f^2(R,L,R)=R$,
$^{(c)}$ because $f^1(R,L,R)=L$,
$^{(d)}$ because $f^2(L,R,L)=R$.

If player 3 plays R at stage $t+3$, then at this stage (players 1 and 3 play R and player 2 plays L) or (players 2 and 3 play R and player 1 plays L), and player 3 does not get out of the punishment phase where players 1 and 2 exchange rooms at each stage, and player 3's payoff is zero at each stage.

So let us assume that player 3 plays L at stage $t+3$. But since τ^3 has recall 3, player 3 will continue to play L as long as the public signal is L. The situation at the end of stage $t+2$ is similar to the situation at the end of stage 7 (left table) or stage 6 (right table) of Figure 8, and from this stage on player 3 will be in the majority room (the L room) hence will also have payoff zero. So $\gamma^3(\tau^3,\sigma^{-3})=0$. □

Lemma 6.4. *In Γ_3, σ^1 is a best reply against σ^{-1}.*

Proof. Let τ^1 be a strategy profile of player 1 in Σ_3^1. It is necessary to to prove that $\gamma^1(\tau^1,\sigma^{-1})\leq\gamma^1(\sigma)=3/7$. Assume that (τ^1,σ^{-1}) is played. Two cases are possible.

Case 1. Assume that at each stage the public signal is L. Then the situation is as follows:

stage $\rightarrow$	1	2	3	4	5	6	7	8	9	10	11	
action P1	L	L	X	Y	L	Z	L	L				
action P2	R	L	L	L	R	L	R	R	L	L	L	
action P3	L	R	L	L	L	L	L	L	L	L	L	...
public signal	L	L	L	L	L	L	L	L	L	L	L	...

with X, Y, Z in $\{L, R\}$.

If $(X,Y) = (L,L)$, then player 1 only plays L since σ^1 has recall 3. And $\gamma^1(\tau^1,\sigma^{-1}) = 0 \leq 3/7$. So it is possible to assume w.l.o.g. that $(X,Y) \neq (L,L)$. The same argument shows that $Z = R$.

If $(X,Y) = (L,R)$, then the actions played by player 1 are $LLLRLRLL$, which is are achievable with recall 3: since signals are L at each stage, player 1 relies on her actions only. If $(X,Y) = (R,L)$, then player 1 plays $LLR\ LLR\ LLR\ LLR\ldots$. But then at some stage the public signal will be R, yielding a contradiction.

The last case to consider is $(X,Y) = (R,R)$. In such a case:

stage $\rightarrow$	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	
action P1	L	L	R	R	L	R	L	L	T	U							
action P2	R	L	L	L	R	L	R	R	L	L	L	R	L	R	R	L	
action P3	L	R	L	L	L	L	L	L	L	L	L	L	L	L	L	L	...
public signal	L	L	L	L	L	L	L	L	L	L	L	L	L	L	L	L	...

If $T = R$, then player 1 plays the following sequence with period 6: $LLRRLR\ LLRRLR\ LLRRLR\ldots$. Since player 2 plays a sequence with period 7 and $\gcd(6,7) = 1$, at some stage the signal will be R, yielding a contradiction. So $T = L$, and necessarily $U = R$. This is exactly in the case of Figure 7, and $\gamma^1(\tau^1,\sigma^{-1}) = 3/7$.

Case 2. Assume that there exists some stage where the public signal is R. It is possible to proceed as in the proof of Lemma 6.3 (Case 2). Since $f^2(L,R,L) = f^2(R,L,R) = R$ and $f^3(L,R,L) = f^3(R,L,R) = L$, also in this case $\gamma^1(\tau^1,\sigma^{-1}) \leq 3/7$. $\square$

Lemma 6.5. *In Γ_3, σ^2 is a best reply against σ^{-2}.*

Proof. Let τ^2 in Σ_3^2 be a strategy of player 2. It is necessary to show that $\gamma^2(\tau^2,\sigma^{-2}) \leq 3/7 = \gamma^2(\sigma)$. Assume for the sake of contradiction that $\gamma^2(\tau^2,\sigma^{-2}) > 3/7$.

Claim. It cannot happen that at some stage, both players 1 and 3 play R.

Assume on the contrary that there exists a first stage $\bar{t}$ where both player 1 and player 3 play R. Necessarily $\bar{t} \geq 3$ and since player 3 plays R at $\bar{t}$, $\bar{t}$ cannot be the first stage where the signal is R. So there exists some stage $\widehat{t} < \bar{t}$ such that the signal at stage $\widehat{t}$ is R, and the signal at every stage t, $\widehat{t} < t < \bar{t}$ is L.

Since player 3 plays R at $\bar{t}$, then $\bar{t} \leq \widehat{t}+3$. By definition of $\bar{t}$, at stage $\widehat{t}$: the signal is R, either player 1 or player 3 play L, and player 2 plays R. So after stage $\widehat{t}$, players 1 and 3 start to exchange rooms and this contradicts the fact that both player 1 and player 3 play R at $\bar{t}$.

Two cases, and several sub-cases are possible.

Case 1. Assume that eventually the sequence of signals only contains L. There exists $\bar{t}$ with $u_t(\tau^2,\sigma^{-2}) = L$ for all $t \geq \bar{t}$.

Then for each stage $t \geq \bar{t}+3$, player 3 will play L (see Figure 6), and given the definition of f^1, player 1 will eventually play the following sequence with period 7: $LLLRRLR\ LLLRRLR\ LLLRRLR\ldots$

Since it was assumed that $\gamma^2(\tau^2, \sigma^{-2}) > 3/7$, there must exist 7 consecutive stages among which player 2 is in the minority room for at least 4 stages. Since the majority room should be L at each large enough stage, the sequence played by player 2 in this case should also have period 7 (prime number). One checks easily that there must exist $t \geq \bar{t}$ such that the play is:

stage $\rightarrow$	t	$+1$	$+2$	$+3$	$+4$	$+5$	$+6$	$+7$	$+8$	$+9$	$+10$	$+11$	$+12$	$+13$
action P1	L	L	L	R	R	L	R	L	L	L	R	R	L	R
action P2	R	R	R	L	L	R	L	R	R	R	L	L	R	L
action P3	L	L	L	L	L	L	L	L	L	L	L	L	L	L
public signal	L	L	L	L	L	L	L	L	L	L	L	L	L	L

The point is that, if the majority room is L at each stage, then τ^2 plays the periodic sequence $RRRLLRL$ $RRRLLRL$ $RRRLLRL\ldots$. This sequence will be denoted by ω in the sequel.

Subcase 1.a. Assume that all signals are L. Then the situation is as follows.

stage $\rightarrow$	1	2	3	4
action P1	L	L	R	R
action P2	X	L	L	L
action P3	L	R	L	L
public signal	L	L	L	L

It must be $X = R$ otherwise player 2 only plays L and $\gamma^2(\tau^2, \sigma^{-2}) = 0$. So player 2, at stage 4, plays L after RLL. This is not compatible with the sequence ω.

Subcase 1.b. Assume that there exists a last stage $\bar{t}$ where the public signal is R. Since player 1 and player 3 never play R at the same time, two possibilities can occur at stage $\bar{t}$.

Subsubcase 1.b.1. If player 1 plays R at stage $\bar{t}$, then

stage $\rightarrow$	$\bar{t}$	$+1$	$+2$	$+3$	$+4$	$+5$	$+6$	$+7$
action P1	R	$L^{(a)}$	$R^{(a)}$	$L^{(a)}$	$L^{(b)}$	$L^{(c)}$	$R^{(d)}$	
action P2	R	$L^{(e)}$	$L^{(e)}$	$L^{(e)}$	X	Y	$L^{(e)}$	
action P3	L	$R^{(a)}$	$L^{(a)}$	$R^{(a)}$	L	L	L	L
public signal	R	L	L	L	L	L	L	L

$^{(a)}$ player 1 and player 3 change rooms after a public signal R,
$^{(b)}$ because $f^1(L, R, L) = L$,
$^{(c)}$ because $f^1(R, L, L) = L$,
$^{(d)}$ because $f^1(L, L, L) = R$,
$^{(e)}$ by assumption, the signal has to be L at every stage $\geq \bar{t} + 1$.

If $X = L$, then player 2 will always play L and have a payoff of zero. So $X = R$. Then $Y = L$ because of the periodic sequence ω. But using ω again, at stage $\bar{t} + 6$ player 2 should play R, yielding a contradiction.

Subsubcase 1.b.2. If player 3 plays R at stage $\bar{t}$, then

stage $\rightarrow$	$\bar{t}$	$+1$	$+2$	$+3$	$+4$	$+5$	$+6$	$+7$	$+8$	$+9$
action P1	L	R	L	R	L	L	L	R	R	L
action P2	R	L	L	L	X	Y	Z	L	L	
action P3	R	L	R	L	L	L	L	L	L	
public signal	R	L	L	L	L	L	L	L	L	L

It must be that $X = R$, otherwise player 2 will always play L after $\bar{t}$. The sequence ω then gives $Y = L$, and $Z = R$. But by ω again at stage $\bar{t} + 7$, player 2 should play R, yielding a contradiction.

Case 2. It remains to consider the case with an infinite number of stages where the public signal is R.

Take any interval of stages $\{t_1, \ldots, t_2\}$, where $t_1 < t_2$, $u_{t_1}(\tau^2, \sigma^{-2}) = u_{t_2}(\tau^2, \sigma^{-2}) = R$, and for every $t \in \{t_1 + 1, \ldots, t_2 - 1\}$, $u_t(\tau^2, \sigma^{-2}) = L$. To conclude the proof, it is sufficient to show that the average payoff of player 2 at the stages $t_1, \ldots, t_2 - 1$ is at most $3/7$.

Assume by contradiction that it is not the case, i.e., assume that the average payoff of player 2 at the stages $t_1, \ldots, t_2 - 1$ is greater than $3/7$. Since player 1 and player 3 never play R at the same stage, at stage t_1, either (players 1 and 2 play R, player 3 plays L) or (players 3 and 2 play R, player 1 plays L). In each case, players 1 and 3 are going to exchange rooms at stages $t_1 + 1$, $t_1 + 2$, $t_1 + 3$, so the payoff of player 2 is zero at each stage t in $\{t_1, t_1 + 1, t_1 + 2, t_1 + 3\}$. It was assumed that the average payoff of player 2 between stage t_1 and stage $t_2 - 1$ is greater than $3/7$. This implies that $t_2 \geq t_1 + 8$. So the signal at the stages $t_1 + 1, \ldots, t_1 + 7$ is L. Two cases are possible.

Subcase 2.a. At stage t_1, player 3 plays L.

stage $\rightarrow$	t_1	$+1$	$+2$	$+3$	$+4$	$+5$	$+6$	$+7$	$+8$	$\ldots$	t_2
action P1	R	L	R	L	L	L	R	R	L	$\ldots$	R
action P2	R	L	L	L	X	Y	L	L	Z	$\ldots$	R
action P3	L	R	L	R	L	L	L	L	L	$\ldots$	L
public signal	R	L	L	L	L	L	L	L	L	$\ldots$	R

By a standard argument $X = R$ (otherwise player 2 plays only L and gets 0). If $Y = L$, then, since player 2 has recall 3

	t_1	$+1$	$+2$	$+3$	$+4$	$+5$	$+6$	$+7$	$+8$	$+9$	$+10$	$+11$	$+12$	$+13$	$+14$	$+15$	t_2
P1	R	L	R	L	L	L	R	R	L	R	L	L	L	R	R	L	R
P2	R	L	L	L	R	L	L	L	R	L	L	L	R	L	L	L	R
P3	L	R	L	R	L	L	L	L	L	L	L	L	L	L	L	L	L
signal	R	L	L	L	L	L	L	L	L	L	L	L	L	L	L	L	R

Then $t_2 = t_1 + 16$, and the average payoff of player 2 is 3/16. So to conclude subcase 2.a., it remains to consider the case when $Y = R$.

stage $\rightarrow$	t_1	+1	+2	+3	+4	+5	+6	+7	+8	+9
action P1	R	L	R	L	L	L	R	R	L	R
action P2	R	L	L	L	R	R	L	L	Z	T
action P3	L	R	L	R	L	L	L	L	L	L
public signal	R	L	L	L	L	L	L	L	L	

Since player 2 has recall 3, then necessarily $Z = L$, $T = R$ and $t_2 = t_1 + 9$. And the average payoff of player 2 is at most 3/9.

Subcase 2.b. At stage t_1, player 1 plays L.

stage $\rightarrow$	t_1	+1	+2	+3	+4	+5	+6	+7	+8	+9	+10
action P1	L	R	L	R	L	L	L	R	R	$L^{(d)}$	$R^{(f)}$
action P2	R	L	L	L	$R^{(a)}$	Y	Z	L	$L^{(b)}$		
action P3	R	L	R	L	L	L	L	L	L	$L^{(d)}$	$L^{(f)}$
public signal	R	L	L	L	L	L	L	L	$L^{(c)}$	$L^{(e)}$	

$^{(a)}$ standard argument because player 2 has recall 3,
$^{(b)}$ the only possibility is L otherwise there is no chance for the average payoff of player 2 to be greater than 3/7. Furthermore $^{(b)}$ implies $^{(c)}$, $^{(c)}$ implies $^{(d)}$, $^{(d)}$ implies $^{(e)}$, and $^{(e)}$ implies $^{(f)}$.

Now, $(Y, Z) = (L, L)$ is not possible because player 2 would play $LLLL$ at stages $t_1 + 5, t_1 + 6, t_1 + 7, t_1 + 8$. The case $(Y, Z) = (L, R)$ also is not possible, because player 2 would have to play the same action at both stages $t_1 + 6$ and $t_1 + 8$.

Assume that $(Y, Z) = (R, L)$. Then

stage $\rightarrow$	t_1	+1	+2	+3	+4	+5	+6	+7	+8	+9	+10
action P1	L	R	L	R	L	L	L	R	R	L	R
action P2	R	L	L	L	R	R	L	L	L	R	R
action P3	R	L	R	L	L	L	L	L	L	L	L
public signal	R	L	L	L	L	L	L	L	L	L	R

Here $t_2 = t_1 + 10$. The average payoff for player 2 at the stages t_1, $t_1 + 1, \ldots, t_2 - 1$ is only 3/10. The last case to consider is $(Y, Z) = (R, R)$.

stage $\rightarrow$	t_1	+1	+2	+3	+4	+5	+6	+7	+8	+9	+10
action P1	L	R	L	R	L	L	L	R	R	L	R
action P2	R	L	L	L	R	R	R	L	L	X'	Y'
action P3	R	L	R	L	L	L	L	L	L	L	L
public signal	R	L	L	L	L	L	L	L	L	L	

Necessarily $Y' = R$, and $t_2 = t_1 + 10$. The average payoff for player 2 is then at most 4/10 ($< 3/7$). $\square$

Section 5. A guessing game

Proof of Proposition 5.1. (a) Player 1 would like to play T when player 2 plays L_1 or L_2, and player 1 would like to play B when player 2 plays R_1 or R_2. So player 1 would like to guess at each stage whether the action played by player 2 is of type L (i.e. L_1 or L_2) or is of type R (i.e. R_1 or R_2). If player 1 has infinite recall, knowing the strategy of player 2 she can compute the next action player 2 is going to play, so player 1 in best response will have a payoff of 1. Hence $E_\infty = \{(1,0)\} = E_0$.

(b) Assume now that both players use strategies with recall k, where k is a fixed positive integer. It will be shown that $(1/2, 0) \in E_k$, i.e., that in the game with recall k player 2 can force player 1's payoff to be no more than 1/2.

More precisely a strategy $\sigma^2 \in \Sigma^2_k$ will be constructed, such that *for every* $\sigma^1 \in \Sigma^1_k$, $\gamma^1(\sigma^1, \sigma^2) = 1/2$.

Consider the alphabet $\{L_1, L_2\}$, and a de Bruijn sequence $L_{i(1)} \dots L_{i(n)}\ L_{i(1)} \dots L_{i(n)}\ L_{i(1)} \dots L_{i(n)} \dots$ of recall k, and call $n = 2^k$ it period (for every t, $i(t) \in \{1, 2\}$).

Such a sequence exists by Proposition 4.2. Now, $R_{i(1)} \dots R_{i(n)}\ R_{i(1)} \dots R_{i(n)}\ R_{i(1)} \dots R_{i(n)} \dots$ is a de Bruijn sequence with recall k on the alphabet $\{R_1, R_2\}$. The strategy σ^2 is defined to play the following periodic sequence with period $2n = 2^{k+1}$

stage	1	2	...	n	$n+1$	$n+2$	...	$2n$	$2n+1$	...	$3n$	$3n+1$
action	$L_{i(1)}$	$L_{i(2)}$	...	$L_{i(n)}$	$R_{i(1)}$	$R_{i(2)}$	...	$R_{i(n)}$	$L_{i(1)}$	...	$L_{i(n)}$	$R_{i(1)}$

Note that this sequence can be played with recall k. This ends the definition of σ^2.

Let now σ^1 be any strategy with recall k for player 1. After a finite number of stages, σ^1 induces a periodic sequence of T and B with period $q \leq 2^k$. Write $q = 2^\alpha r$, where $\alpha \in \{0, \dots, k\}$, and r is an odd positive integer.

Assume that (σ^1, σ^2) is played. The joint period of the action profiles is then $2^{k+1} r = q 2^{k+1-\alpha}$. Denote this period by p. Since $k + 1 - \alpha \geq 1$, then $p \geq 2q$. Fix a stage number t large enough that $t - 1$ is a multiple of $2n$, and consider the joint period $\{t, t+1, \dots, t+p-1\}$. During this period, player 2 plays n actions of "type" L, then n actions of "type" R, then n actions of type L, and so on. Denote by a_0, $a_1, \dots$, a_{q-1} the actions played by player 1 at the stages t, $t+1, \dots$, $t+q-1$. During the period $\{t, t+1, \dots, t+p-1\}$, player 1 plays a_0, a_1, ..., a_{q-1}, a_0, $a_1, \dots$, a_{q-1}, ..., a_0, a_1, ..., a_{q-1}.

Compute now the average payoff for player 1 in the period. Fix i in $\{0, \dots, q-1\}$, and consider the stages corresponding to i modulo q, i.e., consider the stages $t + i + lq$, where $l \in \{0, \dots, 2^{k+1-\alpha} - 1\}$. It will be shown that among those stages, player 2 plays half of the time an action of type L and half of the times an action of type R.

Denote by $\mathbb{Z}_{2n}$ the ring of residue classes modulo $2n$. For every l in $\{0,\ldots,2^{k+1-\alpha}-1\}$, denote by $x_l \in \mathbb{Z}_{2n}$ the congruence class of $i+lq$ modulo $2n$. If $x_l \in \{0,\ldots,n-1\}$, player 2 plays an action of type L at stage $t+i+lq$, and if $x_l \in \{n,\ldots,2n-1\}$, player 2 plays an action of type R at stage $t+i+lq$.

The point is that

$$i+(l+2^{k-\alpha})q = i+lq+2^{k-\alpha}q = i+lq+n2^{-\alpha}q = i+lq+nr.$$

Since r is odd, $x_{l+2^{k-\alpha}} = x_l + n$ in $\mathbb{Z}_{2n}$. So ($x_l \in \{0,\ldots,n-1\}$) if and only if ($x_{l+2^{k-\alpha}} \in \{n,\ldots,2n-1\}$), and consequently

$$\frac{1}{2^{k+1-\alpha}}\left|\{l\in\{0,\ldots,2^{k+1-\alpha}-1\}, x_l\in\{0,\ldots,n-1\}\}\right| = 1/2.$$

So for any value of a_i, the average payoff of player 1 among the stages $\{t+i+lq, l\in\{0,\ldots,2^{k+1-\alpha}\}\}$ is $1/2$. And thus $\gamma^1(\sigma^1,\sigma^2)=1/2$. Hence player 1 cannot do better than playing T at each stage.

□

Proof of Lemma 5.2. Start with the case $k=0$. Player 2 with recall 1 can play the following sequence with period 2: $TL\,TL\,TL\,TL\ldots$, and player 1 with recall 0 cannot guess player 2's action more than half of the stages. This proves that $f(0)=1/2$.

Consider now the case $k=1$. Player 2 with recall 2 can play the following sequence with period 4: $LLRR\;LLRR\;LLRR\;LLRR\ldots$. Player 1 with recall 1 will eventually play a constant sequence, or will alternate between L and R after each stage. In each case, her payoff will be $1/2$, so $f(1)=1/2$.

Put now $k=2$. It will first be shown that $f(2)\leq 4/7$.

Let σ^2 be the 3-recall strategy for player 2 that plays, starting from stage 1, the periodic sequence $LLLRRLR\;LLLRRLR\;LLLRRLR\ldots$. This sequence has period 7. Let now σ^1 be any strategy with recall 2 for player 1. The strategy σ^1 eventually induces a periodic sequence of T and B with period $q\leq 4=2^2$. Since 7 is a prime number, $\gcd(7,q)=1$, and the joint period of the action profiles is the product $7q$. This implies that every action of player 1 corresponding to one stage in the period of player 1, will face 4 times L and three times R during a joint period of length $7q$. So player 1 cannot guess the action of player 2 more than 4 times out of 7, and $\gamma^1(\sigma^1,\sigma^2)\leq 4/7$. So $f(2)\leq 4/7$.

It will be proved now that $f(2)\geq 4/7$. Assume that player 2 plays a strategy σ^2 with recall 3, and assume for the sake of contradiction that $\gamma^1(\sigma^1,\sigma^2)<4/7$ for every strategy σ^1 in Σ^1_2. Player 2 will eventually play a sequence of L and R with minimal period $p\leq 8(=2^3)$. If this sequence does not contain the same number of L and R, then player 1 can have a payoff of at least $4/7$, either by playing constantly L, or by playing constantly R. This is a contradiction, so the sequence of player 2 contains the same number of L and R, and p is even. It is easy to see that $p = 2$ is not possible, so three possibilities are left: $p = 4$, $p = 6$ or $p = 8$.

The case $p = 4$ corresponds here to the sequence: *LLRR LLRR LLRR*..., but then player 1 with recall 2 can guess at each stage the action of player 2. Hence a contradiction. Assume now that $p = 6$. Enumerate each case for the period of player 2, i.e., enumerate all periods of length 6 containing 3 L and 3 R. It is possible to assume w.l.o.g. that the first element of the period is L. The periods of player 2 are ordered lexicographically (with L preceding R). Ten cases are possible (how to select 2 stages among 5 for the last 2 L), even if some cases correspond to minimal periods lower than 6. In each case a periodic sequence is mentioned that can be played by player 1 with recall 2, and gives her a good payoff.

case number	1	2	3	4	5
period P2	*LLLRRR*	*LLRLRR*	*LLRRLR*	*LLRRRL*	*LRLLRR*
sequence P1	*LRLRLR*	*RLRLRL*	*LRLRLR*	*RLRLRL*	*LRLRLR*
payoff P1	4/6	4/6	4/6	4/6	4/6

case number	6	7	8	9	10
period P2	*LRLRLR*	*LRLRRL*	*LRRLLR*	*LRRLRL*	*LRRRLL*
sequence P1	*LRLRLR*	*LRLRLR*	*LRLRLR*	*RLRLRL*	*LRLRLR*
payoff P1	1	4/6	4/6	4/6	4/6

In each case, player 1 can have a payoff greater than 4/7.

The last case to consider is $p = 8$. Player 2 is playing a de Bruijn sequence with recall 3. There are only two such sequences, and by symmetry between L and R only the following sequence can be considered: *RRRLRLLL RRRLRLLL RRRLRLLL*.... Assume that player 1 plays the following sequence with period 4: *RRLL RRLL RRLL*.... Player 1's payoff is here $6/8 = 3/4 > 4/7$. Therefore $f(2) \geq 4/7$. □

Proof of Proposition 5.3. Bertrand's postulate, first proved by Chebyshev, states that for every integer $n \geq 2$, there exists a prime number p such that $n < p < 2n$ (see e.g., Nagell (1964)).

Fix a positive integer k. By Bertrand's postulate there exists a prime number p such that $2^k < p < 2^{k+1}$. By Proposition 4.2, one can construct a periodic sequence of L's and R's with period p that can be played by player 2 having recall $k+1$. This defines σ^2 in Σ^2_{k+1}: play according to this sequence.

Denote by ω this sequence, and let $L(\omega)$ and $R(\omega)$ be the respective numbers of L and R in a period of ω. Obviously $L(\omega) + R(\omega) = p$, and it can be assumed w.l.o.g. that $L(\omega) \geq R(\omega)$.

Fix now $\sigma^1 \in \Sigma^1_k$. Player 1 using σ^1 will eventually play a periodic sequence of T and B with period q, and $q \leq 2^k$. Since p is prime, $\gcd(p, q) = 1$, and this implies that the joint period of the action profiles is the product pq. Within this joint period, every action corresponding to one stage in a period of player 1, will thus face $L(\omega)$ times the action L of player 2 and $R(\omega)$ times the action R of player 2. So $\gamma^1(\sigma^1, \sigma^2) \leq L(\omega)/p$.

Recall that ω is a periodic sequence of L and R, with more L than R, and period p. It will be necessary to look for an upper bound of $L(\omega)/p$, or equivalently for a lower bound of $R(\omega)/p = 1 - L(\omega)/p$. For $i \geq 1$, denote by $A_i \in \{L, R\}$ the action played by player 2 at stage i. Then

$$\omega = A_1 A_2 \dots A_p \ A_1 A_2 \dots A_p \ A_1 A_2 \dots A_p \dots$$

with $A_{i+p} = A_i$, for each $i \geq 1$.

For each $i \geq k+2$, denote by $M_i = (A_{i-(k+1)}, A_{i-k}, \dots, A_{i-1})$ the recall of player 2 before stage i, and denote by $R(M_i) = |\{j \in \{i-(k+1), \dots, i-1\}, A_j = R\}|$ the number of R appearing in the vector M_i. The sequence $(M_i)_{i \geq k+2}$ is periodic with period p, and

$$\frac{1}{p} R(\omega) = \frac{1}{p} |\{i \in \{1, \dots, p\}, A_i = R\}| = \frac{1}{p} \left(\sum_{i=k+2}^{p+k+1} \frac{1}{k+1} R(M_i) \right).$$

The point is that $M_{k+2}, M_{k+3}, \dots, M_{p+k+1}$ are *distinct* elements of $\{L, R\}^{k+1}$, and p is large: $p > 2^k$. So more than half of the elements of $\{L, R\}^{k+1}$ are considered, and a lower bound for the average number of R is needed.

• Imagine first that k is even: $k = 2a$, with a in $\mathbb{N}$. Then exactly half of the elements in $\{L, R\}^{k+1}$ contain more L than R, and the worst case is obtained by selecting the p elements with fewer R. It is certainly even worse if only the $2^k = 2^{2a}$ elements with less R than L are taken and the average on these elements is computed.

$$\frac{R(\omega)}{p} > \frac{1}{2^{2a}} \sum_{l=0}^{a} \frac{l}{2a+1} \binom{2a+1}{l} =: F(a).$$

It is not difficult to compute $F(a)$. Since

$$\begin{aligned}
\sum_{l=0}^{a} l \binom{2a+1}{l} &= (2a+1) \sum_{l=1}^{a} \frac{(2a)!}{(l-1)! \ (2a+1-l)!} \\
&= (2a+1) \sum_{l=1}^{a} \binom{2a}{l-1} \\
&= (2a+1) \sum_{l=0}^{a-1} \binom{2a}{l} \\
&= (2a+1) \left(2^{2a-1} - \frac{1}{2} \binom{2a}{a} \right),
\end{aligned}$$

then

$$F(a) = 1/2 - \frac{1}{2^{2a+1}} \binom{2a}{a}.$$

In this case with even k,

$$\gamma^1(\sigma^1, \sigma^2) < 1 - F(a) = 1/2 + \frac{1}{2^{2a+1}} \binom{2a}{a}.$$

So for every non negative integer a

$$\frac{1}{2} \le f(2a) \le \frac{1}{2} + \frac{1}{2^{2a+1}} \binom{2a}{a}.$$

- Assume now that $k = 2a+1$ is odd. Proceeding the same way,

$$\frac{R(\omega)}{p} > \frac{1}{2^{2a+1}(2a+2)} \left(\sum_{l=0}^{a} l \binom{2a+2}{l} + 1/2(a+1) \binom{2a+2}{a+1} \right),$$

and one can check that the RHS of this inequality is nothing but

$$\frac{1}{2} + \frac{1}{2^{2a+3}} \binom{2a+2}{a+1} - \frac{1}{2^{2a+1}} \binom{2a+1}{a+1}.$$

Hence

$$\frac{1}{2} \le f(2a+1) \le \frac{1}{2} + \frac{1}{2^{2a+1}} \binom{2a+1}{a+1}.$$

The proof is concluded by noticing that both

$$\frac{1}{2^{2a}} \binom{2a}{a} \quad \text{and} \quad \frac{1}{2^{2a+1}} \binom{2a+1}{a+1}$$

go to zero as a goes to infinity. So

$$\lim_{k \to \infty} f(k) = \frac{1}{2}.$$

□

Acknowledgment

The work of Marco Scarsini was partially supported by MIUR-COFIN. This paper was written while he was visiting CEREMADE, whose nice hospitality is gratefully acknowledged.

References

Abreu, D., Pearce, D., and Stacchetti, E. (1990) Toward a theory of discounted repeated games with imperfect monitoring. *Econometrica* **58**, 1041–1063.

Abreu, D. and Rubinstein, A. (1988) The structure of Nash equilibrium in repeated games with finite automata. *Econometrica* **56**, 1259–1281.

Arthur, B. W. (1994) Inductive reasoning and bounded rationality. *Amer. Econom. Rev. Papers Proc.* **84**, 406–411.

Arthur, B. W. (1999) Complexity and the economy. *Science* **284**, 107–109.

Bavly, G. and Neyman, A. (2005) Online concealed correlation by bounded rationality. Mimeo.

Ben-Porath, E. (1990) The complexity of computing a best response automaton in repeated games with mixed strategies. *Games Econom. Behav.* **2**, 1–12.

Ben-Porath, E. (1993) Repeated games with finite automata. *J. Econom. Theory* **59**, 17–32.

Blackwell, D. (1962) Discrete dynamic programming. *Ann. Math. Statist.* **33**, 719–726.

de Bruijn, N. G. (1946) A combinatorial problem. *Nederl. Akad. Wetensch., Proc.* **49**, 758–764 = *Indagationes Math.* **8**, 461–467 (1946).

Challet, D., Marsili, M., and Zhang, Y.-C. (2005) *Minority Games.* Oxford University Press, Oxford.

Cole, H. L. and Kocherlakota, N. R. (2005) Finite memory and imperfect monitoring. *Games Econom. Behav.* **53**, 59–72.

Coolen, A. C. C. (2005) *The Mathematical Theory of Minority Games.* Oxford University Press, Oxford.

Fudenberg, D. and Levine, D. K. (1994) Efficiency and observability with long-run and short-run players. *J. Econom. Theory* **62**, 103–135.

Kalai, E. and Stanford, W. (1988) Finite rationality and interpersonal complexity in repeated games. *Econometrica* **56**, 397–410.

Lehrer, E. (1988) Repeated games with stationary bounded recall strategies. *J. Econom. Theory* **46**, 130–144.

Lehrer, E. (1992a) Correlated equilibria in two-player repeated games with nonobservable actions. *Math. Oper. Res.* **17**, 175–199.

LEHRER, E. (1992b) Two-player repeated games with nonobservable actions and observable payoffs. *Math. Oper. Res.* **17**, 200–224.

LEHRER, E. (1994) Finitely many players with bounded recall in infinitely repeated games. *Games Econom. Behav.* **7**, 390–405.

LEMPEL, A. (1971) m-ary closed sequences. *J. Combinatorial Theory Ser. A* **10**, 253–258.

NAGELL, T. (1964) *Introduction to Number Theory.* Second edition. Chelsea Publishing Co., New York.

NEYMAN, A. (1998) Finitely repeated games with finite automata. *Math. Oper. Res.* **23**, 513–552.

PICCIONE, M. and RUBINSTEIN, A. (2003) Modeling the economic interaction of agents with diverse abilities to recognize equilibrium patterns. *J. European Econom. Assoc.* **1**, 212–223.

RENAULT, J., SCARLATTI, S., and SCARSINI, M. (2005) A folk theorem for minority games. *Games Econom. Behav.* **53**, 208–230.

RENAULT, J. and TOMALA, T. (2004) Communication equilibrium payoffs in repeated games with imperfect monitoring. *Games Econom. Behav.* **49**, 313–344.

RUBINSTEIN, A. (1986) Finite automata play the repeated prisoner's dilemma. *J. Econom. Theory* **39**, 83–96.

SABOURIAN, H. (1998) Repeated games with M-period bounded memory (pure strategies). *J. Math. Econom.* **30**, 1–35.

TOMALA, T. (1998) Pure equilibria of repeated games with public observation. *Internat. J. Game Theory* **27**, 93–109.

YOELI, M. (1962) Binary ring sequences. *Amer. Math. Monthly* **69**, 852–855.

February 13, 2006