Int J Game Theory (2008) 37:581–596
DOI 10.1007/s00182-008-0134-5

ORIGINAL PAPER

Existence of optimal strategies in Markov games with incomplete information

Abraham Neyman

Accepted: 1 July 2008 / Published online: 1 August 2008

Abstract The existence of a value and optimal strategies is proved for the class of two-person repeated games where the state follows a Markov chain independently of players' actions and at the beginning of each stage only Player 1 is informed about the state. The results apply to the case of standard signaling where players' stage actions are observable, as well as to the model with general signals provided that Player 1 has a nonrevealing repeated game strategy. The proofs reduce the analysis of these repeated games to that of classical repeated games with incomplete information on one side.

Keywords Repeated games · Repeated games with incomplete information · Markov chain games

1 Introduction

The class of two-person zero-sum repeated games where the state follows a Markov chain independently of players' actions, and at the beginning of each stage only Player 1 is informed about the state, and players' stage actions are observable, is termed in Renault (2006) Markov chain games with incomplete information on one side.

The play of a Markov chain game with incomplete information on one side proceeds as follows. Nature chooses the initial state z_1 in the finite set of states M according

This research was supported in part by Israeli Science Foundation grants 382/98, 263/03, and 1123/06, and by the Zvi Hermann Shapira Research Fund.

A. Neyman (✉)
Institute of Mathematics and Center for the Study of Rationality,
Hebrew University, 91904 Jerusalem, Israel
e-mail: aneyman@math.huji.ac.il
URL: www.ratio.huji.ac.il/neyman

to an initial probability q_0. At stage t Player 1 observes the current state $z_t \in M$ and chooses an action i_t in the finite set of actions I and (simultaneously) Player 2 (who does not observe the state z_t) chooses an action j_t in the finite set of actions J. Both players observe the action pair (i_t, j_t). The next state z_{t+1} depends stochastically on z_t only; i.e., it depends neither on t, nor on current or past actions, nor on past states. Thus the states follow a Markov chain with initial distribution q_0 and transition matrix Q on M. The payoff at stage t is a function g of the current state z_t and the actions i_t and j_t of the players.

Formally, the game Γ is defined by the 6-tuple $\langle M, Q, q_0, I, J, g\rangle$ where M is the finite set of states, Q is the transition matrix, q_0 is the initial probability of $z_1 \in M$, I and J are the state-independent action sets of Player 1 and Player 2, respectively, and $g : M \times I \times J \to \mathbb{R}$ is the stage payoff function.

The transition matrix Q and the initial probability q_0 define a stochastic process on sequences of states by $P(z_1 = z) = q_0(z)$ and $P(z_{t+1} = z \mid z_1, \ldots, z_t) = Q_{z_t,z}$.

A pure, respectively, behavioral, strategy σ of Player 1 in the game Γ that is defined by $\langle M, Q, q_0, I, J, g\rangle$ is a sequence of functions $\sigma_t : (M \times I \times J)^{t-1} \times M \to I$ $(\sigma_t : (z_1, i_1, j_1, \ldots, i_{t-1}, j_{t-1}, z_t) \mapsto I)$, respectively $\mapsto \Delta(I)$ (where for a finite set D we denote by $\Delta(D)$ all probability distributions on D). A pure, respectively behavioral, strategy τ of Player 2 is a sequence of functions $\tau_t : (I \times J)^{t-1} \to J$, respectively $\mapsto \Delta(J)$.

A pair σ, τ of pure (mixed, or behavioral) strategies (together with the initial distribution q_0) induces a stochastic process with values $z_1, i_1, j_1, \ldots, z_t, i_t, j_t, \ldots$ in $(M \times I \times J)^\infty$, and thus a stochastic stream of payoffs $g_t := g(z_t, i_t, j_t)$.

A strategy σ^* (respectively, τ^*) of Player 1 (respectively, 2) *guarantees* v if for all sufficiently large n, $E^{q_0}_{\sigma^*,\tau} \frac{1}{n} \sum_{t=1}^n g_t \geq v$ (respectively, $E^{q_0}_{\sigma,\tau^*} \frac{1}{n} \sum_{t=1}^n g_t \leq v$) for every strategy τ (respectively, σ) of Player 2 (respectively, 1). We say that Player 1 (respectively, 2) *can guarantee* v in $\Gamma(q_0)$ if for every $\varepsilon > 0$ there is a strategy of Player 1 (respectively, 2) that guarantees $v - \varepsilon$ (respectively, $v + \varepsilon$).

The game has a *value* v if each player can guarantee v. A strategy of Player 1 (respectively, 2) that guarantees $v - \varepsilon$ (respectively, $v + \varepsilon$) is called an *ε-optimal strategy*, and a strategy that is ε-optimal for every $\varepsilon > 0$ is called an *optimal strategy*.

Renault (2006) proved that the Markov chain game Γ has a value v and Player 2 has an optimal strategy. The present paper (1) shows that Renault's result follows from the classical results of repeated games with incomplete information (Aumann and Maschler 1995); and (2) proves the existence of an optimal strategy for Player 1. Thus,

Theorem 1 *The Markov chain game Γ has a value and both players have optimal strategies.*

In addition, these results are extended in the present paper to the model with signals.

Section 2 presents a proof of Renault's results Renault (2006) that the Markov chain game Γ has a value and that Player 2 has an optimal strategy, and sketches the proof of the existence of an optimal strategy of Player 1. Section 3 introduces a class of auxiliary repeated games with incomplete information that serves in the proof of Theorem 1 as well as in approximating the value of Γ. Section 4 couples the Markov chain with stochastic processes that enable us to reduce the analysis of a Markov chain

game to that of a classical repeated game with incomplete information on one side. Section 5 contains the proof of Theorem 1.

Section 6 extends the model and the results to Markov games with incomplete information on one side and signals, where players' actions are unobservable and each player only observes a signal that depends stochastically on the current state and actions. The proof for the model with signals requires only minor modification. For simplicity of notation and exposition, albeit at the cost of some repetition, we introduce the games with signals only after completing the proof of Theorem 1.

2 Informal proofs

The proofs are based on the observation that if $(z_t)_t$ is a Markov chain then for properly chosen sequences $n_i < \bar{n}_i < n_{i+1}$, the Markov chain has with probability close to 1 entered at stage $n_1 + 1$ a communicating class C, and, conditional on the entered communicating class C, the processes $z[i] = z_{n_i+1}, \ldots, z_{\bar{n}_i}$, $i \geq 1$, are almost independent, and the distributions of the initial states in the ith block of stages z_{n_i+1}, $i \geq 1$, are almost identical.

Therefore, a slight alternation of the process (z_t) leads to a process $(\bar{z}_t)_t$ such that $\bar{z}_{n_1+1}$ is in one of the communicating classes C, and, conditional on $\bar{z}_{n_1+1} \in C$, $\bar{z}[1], \bar{z}[2], \ldots$ are independent and $\bar{z}_{n_i+1}$ identically distributed. The alternation is such that Player 1 can compute the state $\bar{z}_{n_i+1}$ as a function of $z_1, \ldots, z_{n_i+1}$ and a private lottery X, and therefore can play in the Markov chain game as if the process of states follows the altered process $(\bar{z}_t)_t$. Note that the altered process is not a Markov chain.

If the states z_t follow a general (not necessarily a Markov chain) stochastic process $(z_t)_t$ with $z_t \in M$ we can define the game $\Gamma((z_t)_t)$ as follows. Nature chooses an infinite sequence $(z_1, z_2, \ldots)$ according to the law of the process. At the beginning of each stage only Player 1 is informed about the state, and players' stage actions are observable.

The proofs assign to the Markov chain process $(z_t)_t$ (and $\varepsilon > 0$) a stochastic process $\bar{z}[1], \bar{z}[2], \ldots$, where each $\bar{z}[i]$ is a finite sequence of states, such that $\Gamma(\bar{z}[1], \bar{z}[2], \ldots)$ has a value and both players have optimal strategies, and there are natural maps $\sigma \mapsto \sigma^*$ and $\tau \mapsto \tau^*$ from strategies in the game $\Gamma(\bar{z}[1], \bar{z}[2], \ldots)$ to strategies in the Markov chain game $\Gamma((z_t)_t)$ so that if the strategy σ of Player 1 (respectively, τ of Player 2) guarantees v in $\Gamma(\bar{z}[1], \bar{z}[2], \ldots)$ then σ^* (respectively, τ^*) garantees $v - \varepsilon$ (respectively, $v + \varepsilon$) in $\Gamma((z_t)_t)$.

For the proofs that Markov games have a value and that Player 2 has an optimal strategy, the finite sequences $\bar{z}[i]$ will all have the same length, and the game $\Gamma(\bar{z}[1], \bar{z}[2], \ldots)$ will be a classical repeated game with incomplete information on one side. For the proof of the existence of an optimal strategy of (the informed) Player 1, the lengths of the finite sequences $\bar{z}[i]$ will converge to infinity, and the proof that Player 1 has an optimal strategy in $\Gamma(\bar{z}[1], \bar{z}[2], \ldots)$ and in $\Gamma((z_t)_t)$ will rely also on the structure of approximate optimal strategies of the informed player in repeated games with incomplete information on one side.

The finite sequence of states $\bar{z}[i]$ will be a minor (stochastic) alternation of the sequence $z[i] = z_{n_i+1}, \ldots, z_{\bar{n}_i}$ of states of the Markov chain in stages $n_i < t \leq \bar{n}_i$,

where $n_i < \bar{n}_i < n_{i+1}$. For notational convenience we define the process $(\bar{z}_t)_t$ and set $\bar{z}[i] = \bar{z}_{n_i+1}, \ldots, \bar{z}_{\bar{n}_i}$.

The stochastic process $(\bar{z}_t)_t$ is a function of the process $(z_t)_t$ and an independent lottery X, with $\bar{z}_t$ being a function of X and $z_1, \ldots, z_t$. The random variable X can be viewed as a private lottery of Player 1 in the game $\Gamma((z_t)_t)$. Therefore a (pure) strategy of Player 1 in the game $\Gamma((\bar{z}_t)_t)$ defines a (mixed) strategy of Player 1 in the Markov chain game $\Gamma((z_t)_t)$. The sets of strategies of (the uninformed) Player 2 in both games are identical. In addition, the construction of $(\bar{z}_t)_t$ will be such that for most stages t the probability that $z_t = \bar{z}_t$ is close to one. Therefore a strategy of Player 1 (respectively, 2) that guarantees v in the game $\Gamma((\bar{z}_t)_t)$ guarantees $v - \varepsilon$ (respectively, $v + \varepsilon$) in the Markov chain game $\Gamma((z_t)_t)$.

The natural lifting of a strategy in the game $\Gamma(\bar{z}[1], \bar{z}[2], \ldots)$ to a strategy in $\Gamma((\bar{z}_t)_t)$ (and thus to a strategy in $\Gamma((z_t)_t)$) is obtained by considering stages $t \leq n_1$ and $\bar{n}_i < t \leq n_{i+1}$ redundant and playing nonrevealingly in these redundant stages.

Now we turn to the details of the construction. First, we recall basic terminology and facts regarding (stationary/homogeneous) Markov chains with a finite state space M and transition matrix Q. For a positive integer n, an $M \times M$ matrix Q, and $z, z' \in M$, we denote by $Q^n_{z,z'}$ (or $Q^n(z, z')$) the (z, z')-th entry of the matrix Q^n; if Q is a transition matrix then $Q^n_{z,z'}$ is the probability that we move from z to z' in n steps when the single-step transition probabilities are defined by Q.

Fix a finite transition matrix Q. A state $z \in M$ is *recurrent* if $\sum_n Q^n_{z,z} = \infty$, equivalently, if the Markov chain that starts at z returns to z with probability 1. A state z *communicates* with a set z' if there are positive integers n and m such that $Q^n_{z,z'} > 0$ and $Q^m_{z',z} > 0$. A set of states C is a *communicating class* (or *ergodic set*) if every state in C communicates with any other state of C and no state of C communicates with a state outside C. Every state in a communicating class is recurrent. The *period* of a state z is the greatest common devisor of all n such that $Q^n_{z,z} > 0$. A state is *aperiodic* if its period is 1. Obviously, if $Q_{z,z} > 0$ then z is aperiodic. All states in the same communicating class have the same period.

A probability distribution $k \in \Delta(M)$ is *Q-invariant* if for every $z \in M$ we have $k(z) = \sum_{z' \in M} k(z') Q_{z',z}$. Every communicating class C has a unique invariant distribution $k_C \in \Delta(C)$ that is Q-invariant, and if $z \in C$ is aperiodic, then for every $z, z' \in C$ the limit of $Q^n_{z,z'}$ exists and equals $k_C(z')$. The set $\{k_C : C \text{ a communicating class}\}$ depends obviously on the transition matrix Q and is denoted $K(Q)$. Equivalently, $K(Q)$ is the (nonempty finite) set of the extreme point of the (polytope of) Q-invariant probability distribution.

Next, we state and prove a simple lemma regarding Markov chains.

Lemma 1 *Let M be a finite state space. There is a positive integer m such that for every $M \times M$ transition matrix Q,* (1) *all recurrent states of the transition matrix Q^m are aperiodic, and, moreover,* (2) *for every two states z, z' in the support of a distribution $k \in K(Q^m)$ we have $Q^m_{z,z'} > 0$.*

Proof Let R be the set of all recurrent states of the Markov chain with state space M and transition matrix Q. For every recurrent $z \in R$ there is $0 < n(z) \leq |M|$ such that $Q^{n(z)}_{z,z} > 0$. Therefore, there is $n > 0$ (e.g., $n = |M|!$ or the least common multiple of

$\{n(z) : z \in R\}$) such that $Q^n_{z,z} > 0$ for every $z \in R$. Let C be a communicating class of the transition matrix Q^n. Note that if $z \in C$ and $Q^{mn}_{z,z'} > 0$ for some $0 < m \leq |M|$ then $Q^{m'n}_{z,z'} > 0$ for every $m' \geq m$, and there is $m' \leq |M|$ such that $Q^{m'n}_{z',z} > 0$ (otherwise z is not a recurrent state of the transition matrix Q^n). Therefore, there is m (e.g., $m = |M|$) such that $Q^{mn}_{z,z'} > 0$ for some $z \in C$ implies that ($z' \in C$ and) $Q^{mn}_{z',z} > 0$. In particular, all reccurent states of the transition matrix Q^{mn} are aperiodic (with respect to the transition matrix Q^{mn}). □

Let $K = K(Q^m)$ where m is given by Lemma 1, and let $p(k)$ be the limit (as $\ell \to \infty$) of the probability that $z_{\ell m+1}$ is in the support $S(k)$ of the invariant probability k. (The limit exists because $\{z_{\ell m+1} \in S(k)\} \subset \{z_{(\ell+1)m+1} \in S(k)\}$ and therefore $P(z_{\ell m+1} \in S(k))$ is monotonic nondecreasing.) For sequences (n_i) and $(\bar{n}_i)$ with $n_i < \bar{n}_i \leq n_{i+1}$, set $z[i] = z_{n_i+1}, \ldots, z_{\bar{n}_i}$.

Fix $\varepsilon > 0$ and a uniform $[0, 1]$-valued random variable X (equivalently, a sequence $X_1, X_2, \ldots$ of independent uniform $[0, 1]$-valued random variable) that is independent of the process $(z_t)_t$. If n_i and $\bar{n}_i$ are multiples of m, and n_1 and $\min_i(n_{i+1} - \bar{n}_i)$ are sufficiently large, we can define a new process $(\bar{z}_t)$ such that (1) $\bar{z}_t$ is a function of $z_1, \ldots, z_t$ and X, (2) on $\bar{z}_{n_1+1} \in S(k)$ $(k \in K)$, $(\bar{z}[i] := \bar{z}_{n_i+1}, \ldots, \bar{z}_{n_i+\ell m})_i$ is a sequence of independent Markov chains with initial probability k and transition matrix Q, and (3) the probability that $\bar{z}[i] = z[i]$ (where $z[i] := z_{n_i+1}, \ldots, z_{n_i+\ell m+1}$) is $\geq 1 - 2\varepsilon$. For example, for $k \in K$ and $z \in S(k)$, we denote by A^i_{kz} the event $z_{n_i+1} = z$ and $A^i_k = \cup_{z \in S(k)} A^i_{zk}$. Let $\bar{A}^1_{kz}$ denote the intersection of the events A^1_{kz} and $X_1 \leq (1 - \varepsilon)p(k)k(z)/P(z_{n_1+1} = z) \leq 1$, and $\bar{A}^1_k = \cup_{z \in S(k)} \bar{A}^1_{kz}$. For $i > 1$ we denote by $\bar{A}^i_{kz}$ the intersection of the events $z_{n_i+1} = z$, $\bar{A}^1_k$, and $X_i \leq (1 - \varepsilon)p(k)k(z)/Q^{n_i-\bar{n}_{i-1}}_{z',z} \leq 1$, where $z_{\bar{n}_{i-1}+1} = z'$. Note that for sufficiently large n_1 we have $P(\bar{A}^1_{kz}) = (1 - \varepsilon)p(k)k(z)$, and for sufficiently large $\min_i(\bar{n}_i - n_i)$ we have $P(\bar{A}^i_{kz} \mid \bar{A}^1_k) = (1 - \varepsilon)k(z)$. On $\bar{A}^i_{kz}$ we set $\bar{z}[i] = z[i]$. It is now easy to complete the definition of the sequence $\bar{z}[1], \bar{z}[2], \ldots$ (on those parts of the probability space where it is not defined by the above rules) so that that the process $(\bar{z}_t)$ obeys (1)–(3). For example, if $(\tilde{z}_t)_t$ is a process that is independent of X and $(z_t)_t$, and where $P(\tilde{z}_{n_1+1} = z) = p(k)k(z)$ and on $\tilde{z}_{n_1+1} \in S(k)$, $(\tilde{z}[i])_i$ is a sequence of independent Markov chains with initial probability k and transition matrix Q, set $\bar{z}[i] = \tilde{z}[i]$ on the complement of $\bar{A}^i := \cup_{k \in K, z \in S(k)} \bar{A}^i_{kz}$.

In order to prove that Γ has a value and that Player 2 has an optimal strategy we set $n_i = i(\ell m + \ell' m)$ and $\bar{n}_i = n_i + \ell m$, where $\ell' \ll \ell$ are sufficiently large. Then, on $\bar{z}_{n_1+1} \in S(k)$, $\bar{z}[1], \bar{z}[2], \ldots$, are in addition identically distributed. Therefore, the game $\Gamma(\bar{z}[1], \bar{z}[2], \ldots)$ (defined formally in Sect. 3 and denoted $\Gamma(p, \ell m)$) that follows the states $\bar{z}[1], \bar{z}[2], \ldots$ and where at the beginning of each stage only Player 1 is informed about the state, is a classical repeated game with incomplete information (where each stage is an ℓm-stage game in extensive form), and thus has a value $v(p, \ell m)$. Each player can follow his optimal strategy in this auxiliary repeated game in stages $n_i < t \leq n_i + \ell m$ of the Markov game (where Player 1 computes the state $\bar{z}_t$ as a function of the process and the private signal/lottery X and plays nonrevealingly in the other stages) to guarantee in the Markov game a payoff within $O(\varepsilon + \ell'/\ell)$ of $v(p, \ell m)$. Therefore, the limit $\lim_{\ell \to \infty} v(p, \ell m)$ exists and

equals the value of the Markov game. Note that Player 2 can start following his auxiliary repeated game strategy at any stage $n_i + 1$. Therefore Player 2 can paste his ε-optimal strategies in the games $\Gamma(p, \ell m)$ into an optimal strategy in the Markov game.

The sketched proof above provides an alternative proof to the results of Renault (2006) that a Markov game with standard signaling has a value and that Player 2 has an optimal strategy. □

Patching ε-optimal strategies of Player 1 into an optimal strategy is more involved, and relies on a more detailed description and properties of approximate optimal strategies in $\Gamma(p, \ell m)$.

The additional needed care stems from the irreversibility of the revelation of information about the process, and the fact that information about the aperiodic class that is revealed to Player 2 when Player 1 plays an optimal strategy in $\Gamma(p, \ell m)$ depends on ℓ. Therefore, the constructed optimal strategy of Player 1 has the following characteristics. First, the starting time $n_1 + 1$ of Player 1 using/revealing his information about the Markov chain is a random time, which Player 1 can compute as a function of past states $z_1, \ldots, z_{n_1+1}$ of the chain and the auxiliary private lottery X and with the property that for every z in the support of $k \in K$ we have $P(z_{n_1+1} = z) = k(z)p(k)$. Second, the length of the ith stage of the auxiliary game is $\ell_i^2 m$ and whatever information revealed eventually by the optimal strategy of Player 1 can be communicated to Player 2 before Player 1 makes his mixed action choice at stage $n_1 + 1$.

Let us recall a few basic facts about repeated games with incomplete information. Let G_ℓ^q be the ℓ-stage repeated game where nature chooses a Markov chain $z_1, \ldots, z_\ell$ with transition matrix Q and initial probability $P(z_1 = z) = q(k)k(z)$ for z in the support of $k \in K$. Equivalently, nature chooses $k \in K$ with probability $q(k)$ and then a Markov chain with transition matrix Q and initial distribution k. The state z_t is revealed to Player 1 just before the play at stage t. At stage t the players choose an action $i_t \in I$ and an action $j_t \in J$, and following the play at stage t Player 2 observes a stochastic signal s^2 whose conditional distribution given the past is a function of the triple (z_t, i_t, j_t). (In the case of standard signaling $s^2 = (i_t, j_t)$.) The payoff to Player 1 of the play $z_1, i_1, j_1, \ldots, z_\ell, i_\ell, j_\ell$ is the average of the stage payoffs $g(z_t, i_t, j_t)$. Let $u_\ell(q)$ denote the maxmin of G_ℓ^q where Player 1 maximizes over his nonseparating strategies. It is known (Aumann and Maschler 1995) that the value of the game $\Gamma(p, \ell)$ is the maximum over all convex combinations $\sum_{i=0}^{|K|} \alpha(i) u_\ell(p(i))$ where $p = \sum_{i=0}^{|K|} \alpha(i) p(i)$.

We select a sequence of $\ell_j \uparrow \infty$ so that (1) the values $v(p, \ell_j^2 m)$ of $\Gamma(p, \ell_j^2 m)$ converge to $\bar{v}(p)$, and (2) the values $v(p, \ell_j^2 m)$ are approximately a convex combination of the form $\sum_{i=0}^{|K|} \alpha(i) u_{\ell_j^2 m}(p(i))$ with $p = \sum_{i=0}^{|K|} \alpha(i) p(i)$. Note that $\alpha(i)$ and $p(i)$ are independent of j and hence the need for an approximation. The fact that $\alpha(i)$ and $p(i)$ are independent of j enables us to patch together approximate optimal strategies of the games $\Gamma(p, \ell_j^2 m)$ and obtain an optimal strategy of Player 1 in the Markov game.

Define $n_0 = 0$ and for $i \geq 0$ set $n_{i+1} = n_i + \ell_i^2 m + \ell_i m$ and $\bar{n}_i = n_i + \ell_i^2 m$. The optimal strategy of Player 1 will use the information of the states only in stages

$n_i < t \le \bar{n}_i$. Set $z[i] = z_{n_i+1}, \ldots, z_{\bar{n}_i}$. We couple the process $(z_t)_t$ with a process $(\bar{z}_t)_t$ so that for some positive-integer-valued function T, where the event $T = i$ is a function of $z_1, \ldots, z_{n_i+1}$ and the coupling enabling $[0, 1]$-valued random variable X (which is independent of the process (z_t)), we have (1) z_{n_T+1} is a recurrent state of the Markov chain with transition Q^m, (2) conditional on $z_{n_T+1} \in S(k)$ the process $\bar{z}[T + i] = \bar{z}_{n_{T+i}+1}, \ldots, \bar{z}_{n_{T+i}}$ is a Markov chain with initial probability k and transition Q, (3) conditional on $z_{n_T+1} \in S(k)$ the processes $\bar{z}[T], \ldots, \bar{z}[T + i], \ldots$ are independent, and (4) the probability that $z[T + i] = \bar{z}[T + i]$ converges to 1 as $i \to \infty$.

The properties of the auxiliary coupled process $(\bar{z}_t)_t$ enables us to patch the approximate optimal strategies of Player 1 in the games $G^p_{\ell^2_j m}$ into an optimal strategy in the Markov game.

3 The auxiliary repeated games $\Gamma(p, \ell)$

The analysis of the game $\Gamma(q_0)$ is by means of auxiliary repeated games with incomplete information on one side, with a finite state space K, initial probability p, and stage game G^k.

The stage game $G^{k,\ell}$, or G^k for short, is a game in extensive form. More explicitly, it is an ℓ-stage game with incomplete information on one side. Nature chooses $r = (z_1 = z, \ldots, z_\ell) \in M^\ell$ where $z \in M$ is chosen according to the probability k, and $z_1 = z, \ldots, z_\ell$ follow the law of the Markov chain with transition matrix Q; before Player 1 takes his action at stage $t \le \ell$ he is informed of z_t, but Player 2 is not informed of z_t. Stage actions are observable.[1] Note that G^k is a finite game with finite strategy sets A for Player 1 and B for Player 2. An element $a \in A$, respectively, $b \in B$, is a sequence of functions a_t, respectively, b_t, $1 \le t \le \ell$, where $a_t : (z_1, i_1, j_1, \ldots, i_{t-1}, j_{t-1}, z_t) \mapsto I$, respectively, $b_t : (i_1, j_1, \ldots, i_{t-1}, j_{t-1}) \mapsto J$. The triple (r, a, b) defines a play $(z_1, i_1, j_1, \ldots, z_\ell, i_\ell, j_\ell)$. Therefore, the triple (k, a, b) defines a probability distribution on the plays $(z_1, i_1, j_1, \ldots, z_\ell, i_\ell, j_\ell)$ where $P(z_1 = z) = k(z)$, $P(z_{t+1} = z \mid z_1, \ldots, z_t, i_t, j_t) = Q_{z_t,z}$, $i_t = a(z_1, i_1, j_1, \ldots, i_{t-1}, j_{t-1}, z_t)$, and $j_t = b(i_1, j_1, \ldots, i_{t-1}, j_{t-1})$. The payoff of the game G^k equals $G^k(a, b) = E^k_{a,b} \frac{1}{\ell} \sum_{t=1}^{\ell} g(z_t, i_t, j_t)$ where the expectation is with respect to the probability defined by (k, a, b).

3.1 The game $\Gamma(p, \ell)$

Nature chooses $k \in K$ with probability $p(k)$. Player 1 is informed of k; Player 2 is not. The play proceeds in stages. In stage n, nature chooses $r = (z_1, \ldots, z_\ell) \in M^\ell$ with probability $k(z_1) \prod_{1 \le t < \ell} Q_{z_t,z_{t+1}}$, Player 1 chooses $a \in A$, and Player 2 chooses $b \in B$. The payoff to Player 1 is $G^k(a, b)$.

The signal s^2 to Player 2 is the function s^2 that assigns to the triple (r, a, b) the sequence of realized stage actions $i_1, j_1, \ldots, i_\ell, j_\ell$. The signal s^1 to Player 1 is the function s^1 that assigns to the triple (r, a, b) the play $(z_1, i_1, j_1, \ldots, z_\ell, i_\ell, j_\ell)$.

[1] The case of imperfect monitoring where each player observes a signal that depends stochastically on the current state and actions is covered in Sect. 6.

The value of $\Gamma(p,\ell)$ exists by Aumann and Maschler (1995, Theorem C, p. 191), and is denoted by $v(p,\ell)$. Set $\bar{v}(p) := \limsup_{\ell\to\infty} v(p,\ell m)$ and $\underline{v}(p) := \liminf_{\ell\to\infty} v(p,\ell m)$. Obviously $\bar{v}(p) \geq \underline{v}(p)$. We will show in Lemma 4 (Sect. 5) that (in the Markov chain game Γ) Player 1 can guarantee $\bar{v}(p)$ and Player 2 can guarantee $\underline{v}(p)$. Thus $\bar{v}(p) = \underline{v}(p)$ is the value of Γ (Corollary 2). Lemma 3, respectively Lemma 4, demonstrates the existence of an optimal strategy of Player 2, respectively, Player 1.

4 Auxiliary coupled processes

Let m, $K = K(Q^m)$, and $p \in \Delta(K)$, as defined in Sect. 2. Recall that the support of a probability distribution $k \in \Delta(M)$ is denoted $S(k)$.

An *admissible pair of sequences* is a pair of increasing sequences, $(n_i)_{i\geq 1}$ and $(\bar{n}_i)_{i\geq 1}$, with $n_i < \bar{n}_i < n_{i+1}$ and such that n_i and $\bar{n}_i$ are multiples of m. For a given admissible pair of sequences and a stochastic process (x_t) we use the notation $x[i] = (x_{n_i+1}, \ldots, x_{\bar{n}_i})$.

4.1 A coupling result

Let $(n_i)_{i\geq 1}$ and $(\bar{n}_i)_{i\geq 1}$ be an admissible pair of sequences with $(n_i - \bar{n}_{i-1})_{i>1}$ non-decreasing and with n_1 sufficiently large so that for every $k \in K$ and $z \in S(k)$ we have $P(z_{n_1+1} = z) \geq p(k)k(z)/2$ (and thus $P(z_{n_1+1} \in S(k)) \geq p(k)/2$). Let $X, X_1, Y_1, X_2, Y_2 \ldots$ be a sequence of iid random variables that are uniformly distributed on $[0,1]$ and so that the process $(z_t)_t$ (that follows the Markov chain with initial distribution q_0 and transition matrix Q) and the random variable $(X, X_1, Y_1, \ldots)$ are independent. Let $\mathcal{F}_i$ denote the σ-algebra of events generated by $X_1, \ldots, X_i$ and $z_1, \ldots, z_{n_i+1}$.

For $k \in K$ and $z \in S(k)$ the event $z_{n_i+1} = z$ is denoted A^i_{kz}. Let A^i_k be the event that $z_{n_i+1} \in S(k)$, i.e., $A^i_k = \cup_{z\in S(k)} A^i_{kz}$, and $A^i = \cup_{k\in K} A^i_k$. As $P(A^i_{kz}) \to p(k)k(z)$ and $P(A^1_{kz}) > p(k)k(z)/2$ by assumption, there exists a strictly decreasing sequence $\varepsilon_j \downarrow 0$ such that $P(A^i_{kz}) \geq (1-\varepsilon_i)p(k)k(z)$ for every $k \in K$ and $2\varepsilon_1 < 1$. Moreover, as each $k \in K$ is invariant under Q^m, we can choose such a sequence for any $\varepsilon_1 > 1 - \inf_{k\in K,\, z\in S(k)} \frac{P(A^1_{kz})}{p(k)k(z)}$ and thus we can assume that $\varepsilon_1 = \varepsilon_1(n_1) \to_{n_1\to\infty} 0$.

A positive-integer-valued random variable T such that for every $i \geq 1$ the event $\{T = i\}$ is $\mathcal{F}_i$-measurable is called an $(\mathcal{F}_i)_i$-adapted stopping time. We will define an $(\mathcal{F}_i)_i$-adapted stopping time T with $T \geq 1$ such that conditional on $\{T = i\}$ the process $z[T+j]$ $(j \geq 0)$ is with probability $p(k)$ a Markov chain with initial probability k and transition Q. Because the distribution k is invariant under Q^m and $n_{i+1} - \bar{n}_i$ is a multiple of m, it suffices to guarantee that for every $k \in K$ and every $z \in S(k)$ the probability that $z_{n_i+1} = z$, conditional on $T = i$, equals $p(k)k(z)$. In addition, our construction is such that T is finite with probability 1 (equivalently, $P(T \leq i) \to_{i\infty} 1$). In fact, by requiring in addition that $P(T \leq i) = 1 - 2\varepsilon_i$ the stopping time T is defined as follows.

Define the $(\mathcal{F}_i)_i$-adapted stopping time T with $T \geq 1$ by defining the event $\{T = i\}$ recursively:[2]

$$T = \begin{cases} 1 & \text{on } z_{n_1+1} = z \in S(k) \text{ and } X_1 \leq \frac{(1-2\varepsilon_1)p(k)k(z)}{P(A^1_{kz})} \\ i & \text{if } T \geq i > 1,\ z_{n_i+1} = z \in S(k) \text{ and } X_i \leq \frac{(2\varepsilon_{i-1}-2\varepsilon_i)p(k)k(z)}{P(A^i_{kz})-(1-2\varepsilon_{i-1})p(k)k(z)}. \end{cases}$$

Lemma 2 (i) $\forall k \in K$ *and* $\forall z \in S(k)$, $\Pr(z_{n_T+1} = z \mid T) = p(k)k(z)$ (*and thus* $\Pr(z_{n_T+1} \in S(k) \mid T) = p(k)$);

(ii) *Conditional on* $z_{n_T+1} \in S(k)$, *for every fixed* $i \geq 0$ *the process* $z[T+i]$ *is a Markov chain with initial probability* k *and transition* Q;

(iii) $\Pr(T \leq i) = 1 - 2\varepsilon_i$.

Proof For $k \in K$ and $z \in S(k)$ let B^i_{kz} denote the event that $T \leq i$ and $z_{n_i+1} = z \in S(k)$, and $B^i_k := \cup_{z\in S(k)} B^i_{kz}$. It follows that $P(B^1_{kz}) = P(A^1_{kz})(1-2\varepsilon_1)p(k)k(z)/P(A^1_{kz}) = (1-2\varepsilon_1)p(k)k(z)$ and thus $P(B^1_k) = \sum_{z\in S(k)}(1-2\varepsilon_1)p(k)k(z) = (1-2\varepsilon_1)p(k)$ and $P(T=1) = \sum_{k\in K}(1-2\varepsilon_1)p(k) = 1-2\varepsilon_1$. By induction on i it follows that $P(B^i_{kz}) = (1-2\varepsilon_i)p(k)k(z)$ and $P(T \leq i) = 1-2\varepsilon_i$; indeed, as the distribution k is invariant under Q^m we have $P(A^i_{kz} \cap B^{i-1}_k) = P(B^{i-1}_k)k(z) = (1-2\varepsilon_{i-1})p(k)k(z)$, and thus for $i > 1$ we have $P(B^i_{kz}) = P(B^{i-1}_k)k(z) + P(A^i_{kz} \setminus B^{i-1}_k)\frac{(2\varepsilon_{i-1}-2\varepsilon_i)p(k)k(z)}{P(A^i_{kz})-(1-2\varepsilon_{i-1})p(k)k(z)}$. As $P(A^i_{kz} \setminus B^{i-1}_k) = P(A^i_{kz}) - (1-2\varepsilon_{i-1})p(k)k(z)$ we deduce that $P(B^i_{kz}) = (1-2\varepsilon_i)p(k)k(z)$. In particular, $P(z_{n_i+1} = z \in S(k) \mid T = i) = p(k)k(z)$. Set $B^i = \cup_{k\in K} B^i_k$ and note that $P(B^i) = 1-2\varepsilon_i$. This completes the proof of (i) and (iii).

Obviously, $z[T+i]$ is a Markov chain with transition Q. As k is invariant under Q^m we deduce that for every $i \geq 0$ we have $\Pr(z_{n_T+i+1} = z \in S(k) \mid z_{n_T+1} \in S(k)) = k(z)$, which proves (ii). □

The next lemma couples the process $(z_t)_t$ with a process $(z^*_t)_t$ where the states z^*_t are elements of $M^* = M \cup \{*\}$ with $* \notin M$. Given $i \geq 1$ we denote by $*[i]$ the sequence of $*$s of length $\bar{n}_i - n_i$. Let $0 < \delta < 1$ be such that for every $k \in K$, and $y, z \in S(k)$, we have $Q^m(y,z) \geq (1-\delta)k(z)$. As k is Q^m-invariant, it follows by induction on $j \geq 1$ that $Q^{jm}(y,z) \geq (1-\delta^j)k(z)$. Indeed, $Q^{jm}(y,z) = \sum_{z'} Q^{(j-1)m}(y,z')Q^m(z',z) = \sum_{z'}((1-\delta^{j-1})k(z') + Q^{(j-1)m}(y,z') - (1-\delta^{j-1})k(z'))Q^m(z',z) \geq (1-\delta^{j-1})k(z) + \delta^{j-1}(1-\delta)k(z) = (1-\delta^j)k(z)$.

Let $\ell_i = (n_i - \bar{n}_{i-1})/m$, and let B_i be the event $Y_i \leq (1-\delta^{\ell_i})k(z)/Q^{\ell_i m}(y,z)$ where $y = z_{\bar{n}_{i-1}+1} \in S(k)$ and $z = z_{n_i+1} \in S(k)$.

Lemma 3 *There exists a stochastic process* $(z^*_t)_t$ *with values* $z^*_t \in M^*$ *such that for* $n_i < t \leq \bar{n}_i$ *the (auxiliary) state* z^*_t *is a (deterministic) function of* $z_1, \ldots, z_t$ *and* $X_1, Y_1, \ldots, X_i, Y_i$ *such that*

(i) $\forall \bar{n}_{i-1} < t \leq n_i$ *and* $\forall t \leq n_T$, $z^*_t = *$

(ii) *Everywhere, either* $z^*[i] = z[i]$ *or* $z^*[i] = *[i]$

[2] Note that the event $\{T \geq i\}$ is the complement of the event $\{T < i\}$.

(iii) $z^*[T] = z[T]$ *and thus* $\Pr(z^*_{n_T+1} = z \mid T) = p(k)k(z)$

(iv) $\Pr(z^*[T+i] = z[T+i] \mid T) = 1 - \delta^{\ell_{T+i}} \geq 1 - \delta^{\ell_i}$

(v) *For* $i \geq 1$, *conditional on* T, $z^*[T], \ldots, z^*[T+i-1]$, *the process* $z^*[T+i]$ *on* B_{T+i} (*and thus with probability* $= 1 - \delta^{\ell_{T+i}}$) *is a Markov chain with initial probability* k *and transition* Q, *and on the complement of* B_{T+i} (*and thus with conditional probability* $= \delta^{\ell_{T+i}}$) *it is* $*[T+i]$.

Proof $\forall \bar{n}_{i-1} < t \leq n_i$ and $\forall t \leq n_T$, set $z^*_t = *$; in particular, $z[i] = *[i]$ for $i < T$.

Define $z^*[T] = z[T]$ and thus (iii) holds, and for $i > T$ set $z^*[i] = z[i]$ on B_i and $z^*[i] = *[i]$ on the complement B_i^c of B_i. It follows that everywhere, either $z^*[i] = z[i]$ or $z^*[i] = *[i]$ and thus (ii) holds. For $i \geq 1$ the conditional probability that $z_{n_{T+i}+1} = z$ given T and $z_{\bar{n}_{T+i-1}+1} = y \in S(k)$ equals $Q^{\ell_j m}(y, z)(1 - \delta^{\ell_j})k(z)/Q^{\ell_j m}(y, z) = (1 - \delta^{\ell_j})k(z)$, where $j = T + i$. Note that this conditional probability is independent of y. Therefore, the conditional probability that $z^*[T+i] = z[T+i]$ given T and $z_{\bar{n}_T+1} \in S(k)$ equals $1 - \delta^{\ell_j}$ $(\geq 1 - \delta^{\ell_i})$, which proves (iv) and (v). □

Corollary 1 *There exists a stochastic process* $(\bar{z}_t)_t$ *with values* $\bar{z}_t \in M$ *such that for* $n_i < t \leq \bar{n}_i$ *the* (*auxiliary*) *state* $\bar{z}_t$ *is a* (*deterministic*) *function of* $z_1, \ldots, z_t$ *and* $X_1, Y_1, \ldots, X_i, Y_i$ *such that*

1.1 *The probability that* $\bar{z}_{n_T+1} = z$ *equals* $p(k)k(z)$ *for* $z \in S(k)$

1.2 *For* $i \geq 1$, *conditional on* T, $\bar{z}[T], \ldots, \bar{z}[T+i-1]$, *the process* $\bar{z}[T+i]$ *is a Markov chain with initial probability* k *and transition* Q

1.3 $\Pr(\bar{z}[T+i] = z[T+i]) \geq 1 - \delta^{\ell_{T+i}} \geq 1 - \delta^{\ell_i}$

Proof Let $\mathbf{k}$ and $\tilde{z}[k, i]$, $k \in K$ and $i \geq 1$, be independent random variables such that $\Pr(\mathbf{k} = k) = p(k)$ and each random variable $\tilde{z}[k, i]$ is a Markov chain of length $\bar{n}_i - n_i$ with initial distribution k and transition matrix Q. W.l.o.g. we assume that $\mathbf{k}$ and $\bar{z}[k, i]$, $k \in K$ and $i \geq 1$, are deterministic functions of X.

Set $\bar{z}_t = z_t$ for $t \leq n_T$ and for $\bar{n}_i < t \leq n_{i+1}$. Define $\bar{z}[T+i] = z[T+i]$ on $z^*[T+i] = z[T+i]$, and $\bar{z}[T+i] = \tilde{z}[k, T+i]$ on $z^*[T+i] = *[T+i]$ and $z_{n_T+1} \in S(k)$. □

5 Existence of the value and optimal strategies in $\Gamma(q_0)$

Assume without loss of generality that all payoffs of the stage games $g(z, i, j)$ are in $[0, 1]$.

Lemma 4 *Player 1 can guarantee* $\bar{v}(p)$ *and Player 2 can guarantee* $\underline{v}(p)$.

Proof Note that for $\ell < \ell'$ we have $v(p, \ell') \geq v(p, \ell)\ell/\ell'$ and therefore $\bar{v}(p) = \limsup_{\ell \to \infty} v(p, \ell^2 m)$. Similarly, $\underline{v}(p) = \liminf_{\ell \to \infty} v(p, \ell^2 m)$. Fix $\varepsilon > 0$. Let ℓ be sufficiently large with $v(p, \ell^2 m) > \bar{v}(p) - \varepsilon$, respectively $v(p, \ell^2 m) < \underline{v}(p) + \varepsilon$, $1/\ell < \varepsilon$, and so that $\delta^{\ell m} < \varepsilon$ and $\Pr(z_{\ell m+1} = z) \geq (1 - \varepsilon)p(k)k(z)$ for every $k \in K$ and $z \in S(k)$.

Set $\bar{n}_0 = 0$, and for $i \geq 1, \bar{n}_i = i(\ell+\ell^2)m + \bar{\ell}m$ and $n_i = \bar{n}_{i-1} + \ell m + \bar{\ell}m$ where $\bar{\ell}$ is[3] a nonnegative integer. Let $(z_t^*)_t$ be the auxiliary stochastic process obeying 1.1, 1.2, and 1.3 of Corollary 1. Define $g_t^* = g(z_t^*, i_t, j_t)$ (and recall that $g_t = g(z_t, i_t, j_t)$).

Let σ be a $\frac{1}{\ell}$-optimal (and thus an ε-optimal) strategy of Player 1 in $\Gamma(p, \ell^2 m)$ and let σ^* be the strategy in (the Markov game) Γ defined as follows. Set $h[i,t] = z^*_{n_i+1}, i_{n_i+1}, j_{n_i+1}, \ldots, z^*_{n_i+t}, i_{n_i+t}, j_{n_i+t}$, and $h[i] = h[i, \ell^2 m]$. In stages $\bar{n}_i < t \leq n_{i+1}$ $(i \geq 0)$ and in all stages on $T > 1$, the strategy σ^* plays a fixed action $i^* \in I$. On $T = 1$, in stage $n_i + t$ with $1 \leq t \leq \ell^2 m$ the strategy σ^* plays the mixed action $\sigma(h[1], \ldots, h[i-1], h[i, t-1], z^*_{n_i+t})$ (where $h[i, 0]$ stands for the empty string).

The definition of σ^*, together with the ε-optimality of σ and the properties of the stochastic process $z^*[1], z^*[2], \ldots$, implies that for all sufficiently large $i > 1$ and every strategy τ of Player 2 we have

$$E_{\sigma^*,\tau} \sum_{j=1}^{i} \sum_{n_j < t \leq \bar{n}_j} g_t^* \geq i\ell^2 m(\bar{v}(p) - 2\varepsilon - \Pr(T > 1))$$

On $z^*[j] = z[j]$, we have $\sum_{n_j < t \leq \bar{n}_j} g_t^* = \sum_{n_j < t \leq \bar{n}_j} g_t$. Therefore,

$$E_{\sigma^*,\tau} \sum_{j=1}^{i} \sum_{n_j < t \leq \bar{n}_j} g_t \geq i\ell^2 m(\bar{v}(p) - 4\varepsilon)$$

and therefore, as the density of the set of stages $\cup_i \{t : \bar{n}_{i-1} < t \leq n_i\}$ is $\ell/(\ell+\ell^2) < \varepsilon$, we deduce that σ^* guarantees $\bar{v}(p) - 5\varepsilon$ and therefore Player 1 can guarantee $\bar{v}(p)$.

Respectively, if τ is an ε-optimal strategy of Player 2 in the game $\Gamma(p, \ell^2 m)$, we define the strategy τ^* $(= \tau^*[\ell, \tau, \bar{\ell}])$ of Player 2 in $\Gamma(q_0)$ as follows. Set $h^2[i,t] = i_{n_i+1}, j_{n_i+1}, \ldots, i_{n_i+t}, j_{n_i+t}$, and $h^2[i] = h^2[i, \ell^2 m]$. In stages $t \leq \bar{n}_1$ and in stages $\bar{n}_i + t$ with $1 \leq t \leq \ell m$ the strategy τ^* plays a fixed action $j^* \in J$. In stage $n_i + t$ with $i \geq 1$ and $1 \leq t \leq \ell^2 m$ the strategy τ^* plays the action $\tau(h^2[1], \ldots, h^2[i-1], h^2[i, t-1])$ (where $h^2[n, 0]$ stands for the empty string).

The definition of τ^*, together with the ε-optimality of τ and the properties of the stochastic process $z^*[1], z^*[2], \ldots$ and $z[1], z[2], \ldots$, implies that τ^* guarantees $\underline{v}(p) + 5\varepsilon$ and therefore Player 2 can guarantee $\underline{v}(p)$.[4] □

Corollary 2 *The game $\Gamma(q_0)$ has a value $v(\Gamma(q_0)) = \underline{v}(p) = \bar{v}(p)$.*

Lemma 5 *Player 2 has an optimal strategy.*

Proof Recall that the 5ε-optimal strategy τ^* appearing in the proof of Lemma 4 depends on the positive integer ℓ, the strategy τ of Player 2 in $\Gamma(p, \ell^2 m)$, and the auxiliary nonnegative integer $\bar{\ell}$.

[3] The dependence on $\bar{\ell}$ enables us to combine the constructed ε-optimal strategies of Player 2 into an optimal strategy of Player 2.

[4] An alternative construction of a strategy σ^* of Player 1 that guarantees $\bar{v}(p) - \varepsilon$ is provided later in this section, and an alternative construction of a strategy τ^* that guarantees $\underline{v}(p) + \varepsilon$ is given in Sect. 6.

Fix a sequence $\ell_j \uparrow \infty$ with $v(p, \ell_j^2 m) < \underline{v}(p) + 1/j$ and strategies τ_j of Player 2 that are $1/j$-optimal in $\Gamma(p, \ell_j^2 m)$. Let $d_j \geq j$ be a sequence of positive integers such that for every strategy σ_j of Player 1 in $\Gamma(p, \ell_j^2 m)$ and every $d \geq d_j$ we have

$$E^p_{\sigma_j,\tau_j} \sum_{s=1}^{d} G^k(a(s), b(s)) \leq d v(p, \ell_j^2 m) + d/j.$$

Let $N_0 = 0$, $N_j - N_{j-1} = \bar{d}_j(\ell_j^2 + \ell_j)m$ where $\bar{d}_j > d_j$ is an integer, and $(j-1)d_j \ell_j^2 m \leq N_{j-1}$. E.g., choose integers $\bar{d}_j \geq d_j + j d_{j+1} m \ell_{j+1}^2/\ell_j^2$ and let $N_0 = 0$ and $N_j = N_{j-1} + \bar{d}_j(\ell_j^2 + \ell_j)m$.

By setting $\bar{n}_0^j = 0$, $\bar{n}_i^j = N_{j-1} + i(\ell_j + \ell_j^2)$ for $i \geq 1$, $n_1^j = N_{j-1} + \ell_j m$, and $n_i^j = \bar{n}_i^j - \ell_j^2 m$, we construct strategies $\tau^*[j] = \tau^*[\ell_j, \tau_j, \bar{\ell}_j = N_{j-1} + \ell_j m]$ such that if τ^* is the strategy of Player 2 that follows $\tau^*[j]$ in stages $N_{j-1} < t \leq N_j$ we have for every positive integer T with $N_{j-1} + d_j(\ell_j^2 + \ell_j)m < T \leq N_j$,

$$E_{\sigma,\tau^*} \sum_{t=N_{j-1}+1}^{T} g_t \leq (T - N_{j-1})(\underline{v}(p) + 2/j)$$

and therefore for every every positive integer T with $N_{j-1} < T \leq N_j$ we have

$$E_{\sigma,\tau^*} \sum_{t=1}^{T} g_t \leq T\underline{v}(p) + \sum_{i<j}(N_i - N_{i-1})2/i) + (T - N_{j-1})2/j + d_j(\ell_j^2 + \ell_j).$$

For every $\varepsilon > 0$ there is j_0 such that for $j \geq j_0$ we have $\frac{1}{N_{j-1}} \sum_{i<j}(N_i - N_{i-1})2/i) < \varepsilon$, $2/j < \varepsilon$, and $\frac{1}{N_{j-1}} d_j(\ell_j^2 + \ell_j) < \varepsilon$. Thus for $T > N_{j_0}$ we have

$$E_{\sigma,\tau^*} \frac{1}{T} \sum_{t=1}^{T} g_t \leq \underline{v}(p) + 3\varepsilon$$

and therefore τ^* is an optimal strategy of Player 2. □

Lemma 6 *Player 1 has an optimal strategy.*

Proof By Aumann and Maschler (1995), for every ℓ there exists $p(0, \ell), \ldots, p(|K|, \ell) \in \Delta(K)$ and a probability vector $\alpha(0, \ell), \ldots, \alpha(|K|, \ell)$ (i.e., $\alpha(i, \ell) \geq 0$ and $\sum_{i=0}^{|K|} \alpha(i, \ell) = 1$) such that $\sum_{i=0}^{|K|} \alpha(i, \ell) p(i, \ell) = p$ and $v(p, \ell^2 m) = \sum_{i=0}^{|K|} \alpha(i, \ell) u_\ell(p(i, \ell))$ where $u_\ell(q)$ is the max min of $G_\ell^q := \Gamma_1(q, \ell^2 m)$ where Player 1 is maximizing over all nonseparating strategies in G_ℓ^q, and Player 2 minimizes over all strategies.

Let $\ell_j \uparrow \infty$ such that $\lim_{j\to\infty} v(p, \ell_j^2 m) = \limsup_{\ell\to\infty} v(p, \ell^2 m)$, and the limits $\lim_{j\to\infty} \alpha(i, \ell_j)$, $\lim_{j\to\infty} p(i, \ell_j)$ and $\lim_{j\to\infty} u_{\ell_j}(p(i, \ell_j))$ exist and equal $\alpha(i)$, $p(i)$ and $u(i)$ respectively. Then

$$\limsup_{\ell\to\infty} v(p,\ell^2 m) = \sum_{i=0}^{|K|} \alpha(i)u(i).$$

Let $\bar{p}(i,\ell_j)[k] = p(i,\ell_j)[k]/\sum_{k\in S(p(i))} p(i,\ell_j)[k]$ if $k \in S(p(i))$, and $\bar{p}(i,\ell_j)[k] = 0$ if $k \notin S(p(i))$. Note that $\bar{p}(i,\ell_j) \to_{j\to\infty} p(i)$.

By the definition of a nonseparating strategy it follows that a nonseparating strategy in $\Gamma_1(q,\ell)$ is a nonseparating strategy in $\Gamma_1(q',\ell)$ whenever the support of q' is a subset of the support of q. Therefore, $u(i) \le \liminf_{j\to\infty} u_{\ell_j}(\bar{p}(i,\ell_j)) = \liminf_{j\to\infty} u_{\ell_j}(p(i))$. Let $\theta_i \to_{i\to\infty} 0$ with $u_{\ell_j}(p(i)) > u(i) - \theta_i$.

By possibly replacing the sequence ℓ_j by another sequence where the jth element of the original sequence, ℓ_j, repeats itself L_j (e.g., ℓ_{j+1}^2) times, we may assume in addition that $\ell_{j+1}^2/\sum_{i\le j}\ell_i^2 \to_{j\to\infty} 0$.

Let σ^{ji} be a nonseparating optimal strategy of Player 1 in the game $\Gamma_1(p(i),\ell_j^2 m)$. Set $\bar{n}_j = \sum_{r\le j}(\ell_r^2 + \ell_r)m$ and $n_j = \bar{n}_j - \ell_j^2 m$.

We couple the process $(z_t)_t$ with a process $(z_t^*)_t$ that satisfies conditions (i)–(v) of Lemma 3. Player 1 can construct such a process $(z_t^*)_t$ as z_t^* is a function of the random variables $X, X_1, Y_1, \ldots$ and $z_1, \ldots, z_t$.

Define the strategy σ of Player 1 as follows. Let $\beta(k,i) := p(i)[k]\alpha(i)/p(k)$ for $k \in K$ with $p(k) > 0$. Note that $\sum_i \beta(k,i) = 1$ for every k, and $\alpha(i) = \sum_k p(k)\beta(k,i)$. Conditional on $z_{N_T+1} \in S(k)$, choose i with probability $\beta(k,i)$ and in stages $n_j < t \le \bar{n}_j$ with $j \ge T$ and $z_{n_j+1}^* = z_{n_j+1}$ (equivalently, $z^*[j] = z[j]$) play according to σ^{ij} using the states of the process $z[j]$ $(= z^*[j])$, i.e., by setting $h[j,t] = z_{n_j+1}, i_{n_j+1}, j_{n_j+1}, \ldots, i_{n_j+t-1}, j_{n_j+t-1}, z_{n_j+t}$,

$$\sigma(z_1,\ldots,z_{n_j+t}) = \sigma^{ij}(h[j,t]).$$

In all other cases, σ plays a fixed[5] action i^*, i.e., in stages $t \le \bar{n}_T$ and in stages $\bar{n}_{j-1} < t \le n_j$ as well as in stages $n_j < t \le \bar{n}_j$ with $z^*[j] = *[j]$ σ plays a fixed[6] action i^*.

The conditional probability that $z^*[j] = z[j]$, given $T \le j$, is $1 - \delta^{\ell_{j-1}^2}$. For simplicity of the notations below, set $\delta_j = \delta^{\ell_{j-1}^2}$. It follows from the definition of σ that for every strategy τ of Player 2 and every j we have on $T \le j$

$$\begin{aligned} E_{\sigma,\tau}\left(\sum_{t=1}^{\ell_j^2 m} g_{n_j+t} \mid T\right) &\ge \ell_j^2 m \sum_i \alpha(i) u_{\ell_j}(p(i)) - \ell_j^2 m\delta_j \\ &\ge \ell_j^2 m \sum_i \alpha(i)u(i) - \ell_j^2 m(\theta_j + \delta_j). \end{aligned}$$

[5] In the model with signals this is replaced by the mixed action $x_{z_t}^*$.

[6] Same comment as in footnote 5.

As $P(T > j) = 2\varepsilon_j$, we have

$$E_{\sigma,\tau} \sum_{t=1}^{\ell_j^2 m} g_{n_j+t} \geq \ell_j^2 m \bar{v}(p) - \ell_j^2 m(\theta_j + 2\varepsilon_j + \delta_j)$$

and thus for $\bar{n}_j < n \leq \bar{n}_{j+1}$ we have

$$E_{\sigma,\tau} \sum_{t=1}^{n} g_t \geq n\bar{v}(p) - \sum_{s \leq j} \ell_s^2 m(\theta_s + 2\varepsilon_{s-1} + \delta_s + 1/\ell_s) - (n - \bar{n}_j).$$

As $(\theta_s + \varepsilon_{s-1} + \delta_s + 1/\ell_s) \to_{s\to\infty} 0$ we deduce that $\sum_{s\leq j} \ell_s^2 m(\theta_s + \varepsilon_s + \delta_s)/\bar{n}_j \to_{j\to\infty} 0$. In addition, $(\bar{n}_{j+1} - \bar{n}_j)/\bar{n}_j \to_{j\to\infty} 0$. Thus for every $\varepsilon > 0$ there is N sufficiently large such that for every $n \geq N$ and for every strategy τ of Player 2, we have

$$E_{\sigma,\tau} \frac{1}{n} \sum_{t=1}^{n} g_t \geq \bar{v}(p) - \varepsilon.$$

□

6 Markov chain games with incomplete information on one side and signals

The game model Γ with signals is described by the 7-tuple

$$\langle M, Q, q_0, I, J, g, R \rangle$$

where $\langle M, Q, q_0, I, J, g \rangle$ is as in the model without signals and observable actions and $R = (R^z_{i,j})_{z,i,j}$ describes the distribution of signals as follows. For every $(z, i, j) \in M \times I \times J$, $R^z_{i,j}$ is a probability distribution over $S_1 \times S_2$.

Following the play z_t, i_t, j_t at stage t, a signal $s_t = (s_t^1, s_t^2) \in S_1 \times S_2$ is chosen by nature with conditional probability, given the past $z_1, i_1, j_1, s_1, \ldots, z_t, i_t, j_t$, that equals $R^{z_t}_{i_t,j_t}$, and following the play at stage t Player 1 observes s_t^1 and z_{t+1} and Player 2 observes s_t^2.

Assume that for every $z \in M$ Player 1 has a mixed action $x_z^* \in \Delta(I)$ such that for every $j \in J$ the distribution of the signal s_2 is independent of z; i.e., for every $j \in J$ the marginals on S_2 of $\sum_i x_z^*(i) R^z_{i,j}$ are constant as a function of z.

Define m, $K = K(Q^m)$, $p \in \Delta(K)$, and the games $\Gamma(p, \ell)$ as in the basic model but with the natural addition of the signals. Let $v(p, \ell)$ be the value of $\Gamma(p, \ell)$. Set $\bar{v} = \limsup_{\ell\to\infty} v(p, \ell m)$ and $\underline{v} = \liminf_{\ell\to\infty} v(p, \ell m)$.

Let A and B denote the pure strategies of Player 1 and Player 2 respectively in $\Gamma_1(p, \ell m)$. A pure strategy $a \in A$ of Player 1 in $\Gamma_1(p, \ell m)$ is a sequence of functions $(a_t)_{1\leq t\leq \ell m}$ where $a_t : (M \times S_1)^{t-1} \times M \to I$. A pure strategy $b \in B$ of Player 2 in $\Gamma_1(p, \ell m)$ is a sequence of functions $(b_t)_{1\leq t\leq \ell m}$ where $b_t : (S_2)^{t-1} \to J$. A triple

$(x,k,b) \in \Delta(A) \times K \times B$ induces a probability distribution, denoted $s^2(x,k,b)$, on the signal in $S_2^{\ell m}$ to Player 2 in $\Gamma_1(p,\ell m)$. For every $q \in \Delta(K)$ we define $NS(q)$ as the set of nonseparating strategies of Player 1 in $\Gamma_1(p,\ell m)$, i.e., $x \in NS(q)$ iff for every $b \in B$ the distribution $s^2(x,k,b)$ is independent across all k with $q(k) > 0$.

Theorem 2 *The game Γ has a value and both players have optimal strategies. The limit of $v(p,\ell m)$ as $\ell \to \infty$ exists and equals the value of Γ.*

Proof The proof that Player 1 has a strategy σ^* that guarantees $\bar{v} - \varepsilon$ for every $\varepsilon > 0$ is identical to the proof (in the basic model) that Player 1 has an optimal strategy.

Next, we prove that Player 2 can guarantee $\underline{v}$. Let γ_n, or ε for short,[7] be a positive number with $0 < \varepsilon < 1/2$, and let ℓ_n, or ℓ for short, be a sufficiently large positive integer such that (1) for every $k \in K$ and $z, z' \in S(k)$ we have $Q^{\ell m}_{z,z'} > (1-\varepsilon)k(z')$, (2) $v(p,\ell m) < \underline{v} + \varepsilon$, and (3) for every $k \in K$ and $z \in S(k)$ $\Pr(z_{\ell m+1} = z) \geq (1-\varepsilon)p(k)k(z)$.

Let τ be an optimal strategy of Player 2 in $\Gamma(p,\ell m)$. Fix a positive integer j_n and construct the following strategy $\tau^*[n]$, or τ^* for short, of Player 2 in Γ. Set $N_i = \frac{i(i+1)}{2}\ell m$ and $n_{ij} = N_i + (j-1)\ell m$ and $\bar{n}_{ij} = n_{ij} + j\ell m$. Let B_i^j be the block of ℓm consecutive stages $n_{ij} < t \leq \bar{n}_{ij}$. For every $j \geq j_n$ consider the sequence of blocks $B_j^j, B_{j+1}^j, \ldots$, as stages of the repeated game $\Gamma(p,\ell m)$ and play in these blocks according to the strategy τ; formally, if $\hat{s}_i^j$ is the sequence of signals to Player 2 in stages $n_{ij} < t \leq \bar{n}_{ij}$, then play in stages $n_{ij} < t \leq \bar{n}_{ij}$ the "stage" strategy $\tau(\hat{s}_j^j, \ldots, \hat{s}_{i-1}^j)$. (In stages $t \notin \cup_{i \geq j} B_i^j$ τ^* plays a fixed action.) Note that for every j, $n_{i+1,j} - \bar{n}_{ij} = i\ell m$, and therefore there is an event C_j with probability $\geq 1 - \varepsilon - \frac{\varepsilon^j}{1-\varepsilon} > 1 - 3\varepsilon$ such that on C_j, the stochastic process $z[j,j], z[j+1,j], \ldots, z[i,j], \ldots$, where $z[i,j] := z_{n_{ij}+1}, \ldots, z_{\bar{n}_{ij}}$ $(i \geq j)$, is a mixture of iid stochastic processes of length ℓm: with probability $p(k)$ the distribution $z[i,j]$ is the distribution of a Markov chain of length ℓm with initial distribution $k(z)$ and transition matrix Q.

It follows that τ^* $(= \tau^*[n])$ guarantees $\underline{v} + 2\varepsilon + 3\varepsilon + \varepsilon$. Indeed, the definition of τ^* implies that for every sufficiently large $i' \geq j$ we have

$$E_{\sigma,\tau^*}\left(\sum_{i=j}^{i'} \sum_{t \in B_i^j} g_t \mid C_j\right) \leq (i'-j+1)\ell m(\underline{v} + 2\varepsilon)$$

and therefore

$$E_{\sigma,\tau^*} \sum_{i=j}^{i'} \sum_{t \in B_i^j} g_t \leq (i'-j+1)\ell m(\underline{v} + 2\varepsilon + 3\varepsilon)$$

[7] The dependence on n enables us to combine the ε-optimal strategies into an optimal strategy.

Thus, if $i(T)$ is the minimal i such that $N_i \geq T$, then for a sufficiently large positive integer T we have

$$E_{\sigma,\tau^*} \sum_{i=j}^{i(T)} \sum_{t \in B_i^j} g_t \leq (i(T) - j + 1)\ell m(\underline{v} + 2\varepsilon + 3\varepsilon)$$

and therefore $E_{\sigma,\tau^*} \sum_{t=1}^{T} g_t$ is $\leq E_{\sigma,\tau^*} \sum_{j=j_n}^{i(T)} \sum_{i=j}^{i(T)} \sum_{t \in B_i^j} g_t$, which is less than or equal to $\frac{i(T)(i(T)+1)}{2}\ell m(\underline{v}+2\varepsilon+3\varepsilon)+j_n i(T)\ell m$. As $i(T) = o(T)$ and $\frac{i(T)(i(T)+1)}{2}\ell m - T < i(T)\ell m$, the strategy τ^* guarantees $\underline{v} + 6\varepsilon$.

Choose a sequence $0 < \gamma_n \to 0$ and a corresponding sequence $\ell_n \uparrow \infty$. By properly choosing an increasing sequence T_n ($T_0 = 0$) and a sequence j_n with $\frac{j_n(j_n+1)}{2}\ell_n m + (j_n - 1)\ell_n m \geq T_{n-1}$ and playing in stages $T_{n-1} < t \leq T_n$ the strategy $\tau^*[n]$ we construct an optimal strategy of Player 2. □

Remarks 1. The value is independent of the signaling to Player 1.

2. The existence[8] of a nonrevealing mixed action x_z^* enables Player 1 to play nonrevealingly in the prefaced play, before the process enters into a communicating class, as well as in the "remixing" stages t.
3. If the model is modified so that the state process is a mixture of Markov chains (namely, nature chooses a pair (z, Q) according to a commonly known probability α on the product of the set of states M and the set of transition matrices, and conditional on the choice of (z, Q) the stochastic process of states obeys $z_1 = z$ and $P(z_t = z' \mid z_1, \ldots, z_{t-1}) = Q_{z_{t-1},z'}$ for $t > 1$) the results about the existence of a value and optimal strategies for the uninformed Player 2 hold. However, since Player 1 is not informed of the choice of Q, the informed Player 1 need not have an optimal strategy.

References

Aumann RJ, Maschler M (1995) Repeated games with incomplete information. MIT Press, Cambridge, MA, USA

Kemeny JG, Snell JL (1976) Finite Markov chains. Springer, New York

Renault J (2006) The value of Markov chain games with lack of information on one side. Math Oper Res 31:490–512

[8] This assumption is not needed in the classical results about repeated games with incomplete information on one side and signals.