האוניברסיטה העברית בירושלים
THE HEBREW UNIVERSITY OF JERUSALEM

EXISTENCE OF OPTIMAL STRATEGIES IN MARKOV GAMES WITH INCOMPLETE INFORMATION

by

ABRAHAM NEYMAN

Discussion Paper # 413 December 2005

מרכז לחקר הרציונליות

CENTER FOR THE STUDY OF RATIONALITY

Feldman Building, Givat-Ram, 91904 Jerusalem, Israel
PHONE: [972]-2-6584135 FAX: [972]-2-6513681
E-MAIL: ratio@math.huji.ac.il
URL: http://www.ratio.huji.ac.il/

Existence of Optimal Strategies in Markov Games with Incomplete Information

Abraham Neyman[1]

December 29, 2005

[1]Institute of Mathematics and Center for the Study of Rationality, Hebrew University, 91904 Jerusalem, Israel. aneyman@math.huji.ac.il www.ratio.huji.ac.il/neyman.
This research was supported in part by Israeli Science Foundation grants 382/98 and 263/03, and by the Zvi Hermann Shapira Research Fund.

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 one 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 one 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.

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 [2] 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 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 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 $\Gamma = \langle M, Q, q_0, I, J, g\rangle$, or $\Gamma(q_0)$ for short, 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 [2] proved that the game $\Gamma(q_0)$ has a value $v(q_0)$ and player 2 has an optimal strategy. The present paper 1) shows that Renault's result follows[1] from the classical results of repeated games with incomplete information [1]; and 2) proves the existence of an optimal strategy for player 1. Thus,

Theorem 1 *The game $\Gamma(q_0)$ has a value $v(\Gamma(q_0))$ and both players have optimal strategies.*

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

Section 2 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 $\Gamma(q_0)$. Section 3 couples the Markov chain with stochastic processes that consist of essentially independent blocks of Markov chains. Section 4 contains the proof of Theorem 1.

Section 5 extends the model and the results to Markov games with incomplete information on on 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. In order to simplify the notation and the exposition, albeit at the cost of some repetition, we introduce the games with signals only after completing the proof of Theorem 1.

2 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

[1] I would have hoped that a reference to the theory of repeated games with incomplete information accompanied by a short sketch would have sufficed. However, as one expert failed to realize the derivation, it may be helpful here to put it in writing.

probability p, and stage game G^k. The support of a probability distribution $k \in \Delta(M)$ is denoted $S(k)$.

Let m be a positive integer such that all ergodic classes of the Markov chain with state space M and transition matrix Q^m are aperiodic. In what follows $Q^n_{z,z'}$ stands for the more explicit $(Q^n)_{z,z'}$. Let $K \subset \Delta(M)$ be the set of all[2] invariant distributions of an ergodic class of Q^m. Obviously, every two distinct elements of K have disjoint support. For every $k \in K$, the subspace $\mathbb{R}^{S(k)}$ of $\mathbb{R}^M$ is invariant under the linear transformation Q^m and therefore the event $z_{nm+1} \in S(k)$ is a subset of the event $z_{(n+1)m+1} \in S(k)$. Therefore, for every $k \in K$, $P(z_{nm+1} \in S(k))$ is monotonic nondecreasing in n. Define $p \in \Delta(K)$ by $p(k) = \lim_{n\to\infty} P(z_{nm+1} \in S(k))$.

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 $t \leq \ell$ he is informed of z_t, but player 2 is not informed of z_t. Stage actions are observable.[3] 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, $1 \leq t \leq \ell$, 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) induces a probability distribution on the plays $(z_1, i_1, j_1, \ldots, z_\ell, i_\ell, j_\ell)$. The payoff of the game G^k equals $G^k(a, b) = E^k_{a,b} \frac{1}{\ell} \sum_{t=1}^{\ell} g(z_t, i_t, j_t)$.

2.1 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 $k(z_1) \prod_{1\leq t<\ell} Q_{z_t,z_{t+1}}$, 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

[2]The set K is defined here independently of q_0. For a given initial distribution q_0 there may exist ergodic classes $k \in K$ such that $P(z_{nm+1} \in S(k)) = 0$. In that case we can have carried out our analysis by means of the repeated game with incomplete information where the set of states equals $\{k \in K : \exists n \text{ s.t. } P(z_{nm+1} \in S(k)) > 0\}$.

[3]The case of imperfect monitoring where each player observes a signal that depends stochastically on the current state and actions is covered in Section 5.

(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 $(z_1, i_1, j_1, \ldots, z_\ell, i_\ell, j_\ell)$.

The value of $\Gamma(p, \ell)$ exists by [1, 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 2 Section 4 that player 1 can guarantee $\bar{v}(p)$ and player 2 can guarantee $\bar{v}(p)$. Thus $\bar{v}(p) = \underline{v}(p)$ is the value of $\Gamma(q_0)$ (Corollary 2). Lemma 3, respectively Lemma 4, demonstrates the existence of an optimal strategy of player 2, respectively, player 1.

3 Auxiliary coupled processes

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 $x[i] = (x_{n_i+1}, \ldots, x_{\bar{n}_i})$.

3.1 A Coupling result

Let $(n_i)_{i\geq 1}$ and $(\bar{n}_i)_{i\geq 1}$ be an admissible pair of sequences with $(n_i - \bar{n}_{i-1})_{i>1}$ nondecreasing 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.

Define the $(\mathcal{F}_i)_i$-adapted stopping time T with $T \geq 1$ by

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

Lemma 1 *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 we have $P(A^i_{kz} \cap B^{i-1}_k) = P(B^{i-1}_k)k(z) = (1-2\varepsilon_i)p(k)k(z)$, 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).

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 $\delta > 0$ be such that for every sufficiently large positive integer j, for every $k \in K$, and $y, z \in S(k)$, we have $Q^{jm}(y,z) \geq (1-\delta^j)k(z)$. Let $\delta : \mathbb{N} \to \mathbb{R}_+$ be defined by[4] $1 - \delta(\ell) = \inf_{j \geq \ell} \min_{k \in K, y, z \in S(k)} Q^{jm}(y,z)/k(z)$. Set $\ell_i = (n_i - \bar{n}_{i-1})/m$ and $\delta(\ell_i m) = \delta_i$. Note that for sufficiently large ℓ_i we have $\delta_i \leq \delta^{\ell_i}$. Let B_i be the event $Y_i \leq (1-\delta_i)k(z)/Q^{\ell_i m}(y,z)$ where $z = z_{n_i+1} \in S(k)$ and $y = z_{\bar{n}_{i-1}+1}$.

[4] As each $k \in K$ is invariant under Q^m, $\min_{k \in K, y, z \in S(k)} Q^{jm}(y,z)/k(z)$ is monotonic nondecreasing in j and thus the inf appearing in the definition of $\delta(\ell)$ is in fact redundant.

Lemma 2 *There exists a stochastic process $(z_t^*)_t$ with values $z_t^* \in M^*$ such that for $n_i < t \le \bar{n}_i$ the (auxiliary) state z_t^* is a (deterministic) function of $z_1, \dots, z_t$ and $X_1, Y_1, \dots, X_i, Y_i$ such that*

i) $\forall \bar{n}_{i-1} < t \le n_i$ *and* $\forall t \le n_T$, $z_t^* = *$

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

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

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

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

Proof. $\forall \bar{n}_{i-1} < t \le n_i$ and $\forall t \le n_T$, 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 $i \ge 1$ 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 $Q^{\ell_j m}(y,z)(1-\delta_j)k(z)/Q^{\ell_j m}(y,z) = (1-\delta_j)k(z)$, where $j = T+i$. Note the 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 $1-\delta_j$, 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 $n_i < t \le \bar{n}_i$ the (auxiliary) state $\bar{z}_t$ is a (deterministic) function of $z_1, \dots, z_t$ and $X_1, Y_1, \dots, X_i, Y_i$ 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 $i \ge 1$, *conditional on* T, $\bar{z}[T], \dots, \bar{z}[T+i-1]$, *the process* $\bar{z}[T+i]$ *is a Markov chain with initial probability* k *and transition* Q

1.3 $\Pr(\bar{z}[T+i] = z[T+i]) \ge 1 - \delta_i$

Proof. Let $\mathbf{k}$ and $\bar{z}[k,i]$, $k \in K$ and $i \geq 1$, be independent random variables such that $\Pr(\mathbf{k} = k) = p(k)$ and each random variable $\bar{z}[k,i]$ 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 $\bar{z}[T+i] = z[k, T+i]$ on $z^*[T+i] = *[T+i]$ and $z_{n_T+1} \in S(k)$. ∎

4 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 3 *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 $i \geq 1$, $\bar{n}_i = i(\ell + \ell^2)m + \bar{\ell}$ and $n_i = \bar{n}_{i-1} + \ell m + \bar{\ell}$ where $\bar{\ell}$ is[5] 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 $\Gamma(q_0)$ 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 $h[i,0]$ stand for the empty string).

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

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

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

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

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

and therefore, as the density of the set of stages $\cup_i\{t : \bar{n}_{i-1} < t \le n_i\}$ 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 $t \le \bar{n}_1$ and in stages $\bar{n}_i + t$ with $1 \le t \le \ell m$ the strategy τ^* plays a fixed action $j^* \in J$. In stage $n_i + t$ with $i > 1$ and $1 \le t \le \ell^2 m$ the strategy τ^* plays the action $\tau(h^2[2], \ldots, h^2[i-1], h^2[i, t-1])$ (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 $\underline{v}(p)$.[6] ∎

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

Lemma 4 *Player 2 has an optimal strategy.*

Proof. Recall that the 5ε-optimal strategy τ^* appearing in the proof of Lemma 3 depends on the positive integer ℓ, the strategy τ of player 2 in $\Gamma(p, \ell^2 m)$, and the auxiliary nonnegative integer $\bar{\ell}$.

Fix a sequence $\ell_j \uparrow \infty$ with $v(p, \ell_j^2 m) < \underline{v}(q_0) + 1/j$ and strategies τ_j of player 2 that are $1/j$-optimal in $\Gamma(p, \ell_j^2 m)$. Let $d_j \ge j$ be a sequence of

[6] 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 Section 5.

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

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

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 $\bar{d}_j \geq d_j + jd_{j+1}m\ell_{j+1}^2/\ell_j^2$ 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 $N_{j-1} + d_j(\ell_j^2 + \ell_j)m < T \leq N_j$,

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

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

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

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

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

and therefore τ^* is an optimal strategy of player 2. ∎

Lemma 5 *Player 1 has an optimal strategy.*

Proof. By [1], 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 j-th 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 1. 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[7] 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[8] action i^*.

The conditional probability that $z^*[j] = z[j]$, given $T \le j$, is $1-\delta_j$. Therefore, it follows from the definition of σ that for every strategy τ of

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

[8]Same comment as in footnote 7.

player 2 and every j we have on $T \leq j$

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

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

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

■

5 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 $z_1, i_1, j_1, \ldots, z_t, i_t, j_t$, 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 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 $(a_t)_{1 \le t \le \ell m}$ 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 $(b_t)_{1 \le t \le \ell m}$ 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,[9] 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 $z \in S(k)$ $\Pr(z_{\ell m+1} = z) \ge (1 - \varepsilon)p(k)k(z)$.

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 $n_{ij} < t \le \bar{n}_{ij}$. For every $j \ge j_n$ consider the

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

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^n, 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 (sub-) 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

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

and therefore

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

Thus, if $i(T)$ is the minimal i such that $N_i \geq T$, then for sufficiently large 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 or equal $\frac{i(T)(i(T)+1)}{2}\ell m(\underline{v}+2\varepsilon+3\varepsilon) + j_n i(T)\ell m$. 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. If the model is modified so that the state process is a mixture of Markov chains the results about the existence of a value and optimal strategies for the uninformed player remain intact. However, the informed player need not have an optimal strategy.

References

[1] Aumann, R. J. and Maschler, M. (1995) *Repeated Games with Incomplete Information*, MIT Press.

[2] Renault, J. (2004) The value of Markov chain games with lack of information on one side. *Mathematics of Operations Research*, forthcoming.