On a Markov Game with Incomplete Information

Johannes Hörner [*], Dinah Rosenberg[†], Eilon Solan[‡] and Nicolas Vieille[§]

January 24, 2006

Abstract

We consider an example of a Markov game with lack of information on one side, that was first introduced by Renault (2002). We compute both the value and optimal strategies for a range of parameter values.

[*]MEDS Department, Kellogg School of Management, Northwestern University, *and* Département Finance et Economie, HEC, 1, rue de la Libération, 78 351 Jouy-en-Josas, France. e-mail: j-horner@kellogg.northwestern.edu

[†]Laboratoire d'Analyse Géométrie et Applications, Institut Galilée, Université Paris Nord, avenue Jean-Baptiste Clément, 93430 Villetaneuse, France; and laboratoire d'Econométrie de l'Ecole Polytechnique, 1, rue Descartes 75005 Paris, France . e-mail: dinah@zeus.math.univ-paris13.fr

[‡]MEDS Department, Kellogg School of Management, Northwestern University, *and* the School of Mathematical Sciences, Tel Aviv University, Tel Aviv 69978, Israel. e-mail: eilons@post.tau.ac.il

[§]Département Finance et Economie, HEC, 1, rue de la Libération, 78 351 Jouy-en-Josas, France. e-mail: vieille@hec.fr

[¶]The research of the third author was supported by the Israel Science Foundation (grant No. 69/01-1).

1 Introduction

Zero-sum repeated games with incomplete information on one side were introduced by Aumann and Maschler (1968, 1995) in a seminal work. In such games, two players play repeatedly a zero-sum matrix game, whose payoff matrix depends on a state of nature that is drawn prior to the beginning of the game. One of the players is informed about the outcome of the state of nature, while the other is not. During the game, the players' moves are observed, but not the corresponding payoff. Such games model long-term interactions with asymmetric information. The main issue is to what extent the informed player should use his private information, and how. Aumann and Maschler prove the existence of the (uniform) value and characterize both the value and the optimal strategies of the players.

Recently, Renault (2002) extended this model to include situations in which the state of nature follows a Markov chain, with exogenous transition function, thereby allowing for situations where the private information of the informed player gets renewed at random times. For such games, Renault proved the existence of the uniform value. However, no explicit formula for the value nor a characterization of optimal strategies is available. Neyman (2004) provided an alternative proof to Renault's results, using a reduction to the case of repeated games with incomplete information on one side a la Aumann and Maschler.

We here provide a number of results on a specific simple game due to Renault (2002).

2 Model and Main Results

2.1 The game

We first define the game. There are two possible states of nature, $\underline{s}$ and $\overline{s}$. The payoff matrices in the two states are given by:

$\underline{s}$

	L	R
T	1	0
B	0	0

$\overline{s}$

	L	R
T	0	0
B	0	1

We denote by $g_s(i,j)$ the payoff in state s induced by the action pair (i,j). The actual state s_n in stage $n \geq 1$ follows a Markov chain with transition function $\mathbf{P}(s_{n+1} = s_n | s_n) = p$, where $p \in [0,1]$ is a parameter of the game. That is, irrespective of the current state, there is a probability $1-p$ that the game will move to the other state at the next stage. We denote by θ_1 the probability that s_1 is $\underline{s}$.

The game proceeds in infinitely many stages. At each stage $n \geq 1$, player 1 and player 2 choose simultaneously a row $i_n \in \{T, B\}$ and a column $j_n \in \{L, R\}$. When doing so, both players are fully informed of the actions $i_1, j_1, \ldots, i_{n-1}, j_{n-1}$ that were chosen in previous stages. In addition, player 1 is fully informed of past and current states $s_1, \ldots, s_n$.

We denote this game by Γ_p. Observe that, for $p = 1/2$, the states (s_n) are independent variables. In effect, the two players then face a sequence of independent, identical, one-shot games with incomplete information. At the other extreme, for $p = 1$, the current state is fixed throughout the game, and thus Γ_1 coincides with the leading example of Chapter 1 in Aumann and Maschler (1995).

2.2 Main results

2.2.1 Background definitions

A strategy of the (uninformed) player 2 is a function $\tau : \cup_{n \in \mathbf{N}} (I \times J)^{n-1} \to [0,1]$, where $\tau(h)$ is the probability assigned to L after the sequence h of actions. We describe a strategy of player 1 as a function

$$\sigma : \cup_{n \in \mathbf{N}} (I \times J \times S)^{n-1} \to [0,1] \times [0,1],$$

with the interpretation that σ assigns to each sequence of past actions and past states a *pair* of mixed moves, to be used in states $\underline{s}$ and $\overline{s}$ respectively: that is, denoting $(x_n, y_n) = \sigma(s_1, i_1, j_1, \ldots, s_{n-1}, i_{n-1}, j_{n-1})$, x_n (resp. y_n) is the probability assigned to T if $s_n = \underline{s}$ (resp. $s_n = \overline{s}$).

Given a strategy pair (σ, τ), we denote by $\mathbf{P}_{\sigma,\tau}$ the induced probability measure over the set of infinite plays, and we denote by $\mathbf{E}_{\sigma,\tau}$ the corresponding expectation operator.

For a given stage $N \in \mathbf{N}$,

$$\gamma_N(\sigma, \tau) = \frac{1}{N} \mathbf{E}_{\sigma,\tau} \left[\sum_{n=1}^{N} g_{s_n}(i_n, j_n) \right]$$

is the average payoff in the first N stages. We denote by v_N the value of the N-stage game – i.e., the value of the game with payoff function γ_N.

A real number v is the (uniform) value if there are strategies that (approximately) guarantee v_N in all long games. To be specific:

Definition 1 *The real number v is the* value *of the game if, for every $\varepsilon > 0$ there is $N_0 \in \mathbf{N}$ and a pair of strategies (σ^*, τ^*) such that for every $N \geq N_0$*

$$\gamma_N(\sigma^*, \tau) + \varepsilon \geq v \geq \gamma_N(\sigma, \tau^*) - \varepsilon, \quad \forall \sigma, \tau. \tag{1}$$

Renault (2002) proves that the game Γ_p has a value v_p for each $p \in [0,1]$ and that moreover, v_p is independent of the initial distribution θ_1. It is easy to check that, in the present game, σ^* and τ^* can be chosen independently of ε.

For various values of p we will exhibit strategies σ^* and τ^* and a real number v such that $\liminf \gamma_{N\to\infty}(\sigma^*, \tau) \geq v$ and $\limsup \gamma_{N\to\infty}(\sigma, \tau^*) \leq v$, for each σ and τ. Such strategies will be called *optimal* strategies. It can be checked that the strategies σ^* and τ^* satisfy the uniformity condition (1), but we will leave this issue out of the present note.

2.2.2 Main Results

We first introduce strategies for both players, that will turn to be optimal strategies for a range of parameter values.

We start with player 1, and let $\theta \leq 1/2$. It is readily checked that the mixed action α_θ of player 1 that assigns probability $(1-\theta)/\theta$ to T in state $\underline{s}$, and probability 1 to B in state $\overline{s}$, is an optimal strategy of player 1 in the *one-shot* game with incomplete information in which player 1 is informed of the state of nature, while player 2 is not.[1]

$\frac{1-\theta}{\theta}$	1	0
$1-\frac{1-\theta}{\theta}$	0	0

θ

1	0	0
	0	1

$1-\theta$

For $\theta \geq 1/2$, by symmetry, we define α_θ to assign probability 1 to T in state $\underline{s}$, and $\theta/(1-\theta)$ to B in state $\overline{s}$.

[1] If $\theta < 1/2$, it is not the unique optimal strategy. However, it is the only one that renders player 2 indifferent between playing L or R.

We define σ^* as the strategy that plays the mixed move α_{θ_n} in stage n, where θ_n is the conditional probability that $s_n = \underline{s}$, given past actions of player 1. In other words, θ_n is the belief held by player 2 in stage n, when facing the strategy σ^*.[2] To be specific, letting $\alpha_{s,\theta_n}(i)$ be the probability assigned by α_{θ_n} to move i in state s, one has by Bayes' rule,

$$\theta_{n+1} = p \times \frac{\alpha_{\underline{s},\theta_n}(i)\theta_n}{\alpha_{\underline{s},\theta_n}(i)\theta_n + \alpha_{\overline{s},\theta_n}(i)(1-\theta_n)} + (1-p) \times \frac{\alpha_{\overline{s},\theta_n}(i)(1-\theta_n)}{\alpha_{\underline{s},\theta_n}(i)\theta_n + \alpha_{\overline{s},\theta_n}(i)(1-\theta_n)}.$$

The first fraction is the conditional probability that $s_n = \underline{s}$, given that player 1 just played i. In that event, there is probability p that $s_{n+1} = \underline{s}$. The second fraction is the conditional probability that $s_n = \overline{s}$, given that player 1 just played i. In that event, there is probability $1-p$ that $s_{n+1} = \underline{s}$.

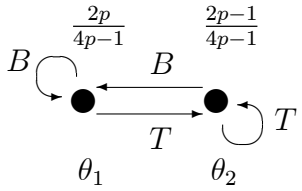
In effect, σ^* plays optimally in a myopic way. It maximizes the *current* payoff under the constraint that player 2 be indifferent between L and R, not taking into account the information on the current state that is transmitted to player 2.

We now describe a strategy τ^* of player 2. It is an automaton with two states, labelled θ_0, θ_1. Transitions are given by:

$$(\theta_0, T) \to \theta_1, \qquad (\theta_1, T) \to \theta_1,$$
$$(\theta_0, B) \to \theta_0, \qquad (\theta_1, B) \to \theta_0.$$

That is, the automaton moves to θ_0 (resp. to θ_1) whenever B (resp. T) is played. In state θ_0 (resp. in state θ_1), the automaton plays L (resp. R) with probability $\frac{2p}{4p-1}$.

Graphically, τ^* looks as follows.



[2] There is no circularity here, since the computation of θ_n involves only the strategy of player 1 in the first $n-1$ stages.

Figure 1: The strategy of player 2

In this figure, below each state appears its name, and above the state appears the probability to play L at that state. Transitions are indicated by arrows.

We now summarize our main results.

Theorem 2 *The following results hold:*

1. *One has $v_p = v_{1-p}$ for each $p \in [0,1]$;*
2. *$v_p = \frac{p}{4p-1}$, for $p \in [1/2, 2/3]$, and $v_p \leq \frac{p}{4p-1}$ for $p \geq 2/3$.*
3. *Let p^* be the unique real solution of $9p^3 - 13p^2 + 6p - 1 = 0$. ($p \sim 0.75891$). One has $v_{p^*} = \frac{p}{1-3p+6p^2} \backsimeq 0.348291466$.*
4. *The strategy σ^* is optimal for all $p \in [1/2, 2/3]$.*
5. *The strategy τ^* is optimal for all $p \in [1/2, 2/3]$.*

Actually, the second statement provides an upper bound on v_p. In Section 6, we will provide a lower bound as well.

Independently of us, and using different tools, Marino (2005) proved that $v_p = \frac{p}{4p-1}$ for $p \in [1/2, 2/3]$.

The proof is organized as follows. We start below by establishing the first claim in Theorem 2. In Section 3, we recall the so-called Average Cost Optimality Equality from the theory of MDP, and we derive a number of implications for the game under study. In Section 4, we derive the properties of τ^*. In Section 5, we prove the claims relative to player 1. Section 6 provides a lower bound on v_p.

2.3 $v_p = v_{1-p}$

Here we argue that the symmetries in the game imply $v_p = v_{1-p}$ for every $p \in [0,1]$. Thus, it is enough to study v_p for $p \in [1/2, 1]$. Given a strategy σ of player 1, we define a mirrored strategy σ', obtained by mirroring actions and states at *even* stages. That is, given a finite history h, we first construct h' by changing at all even stages every appearance of T (resp. B, L, R, $\underline{s}$, $\overline{s}$) to B (resp. T, R, L, $\underline{s}$, $\overline{s}$), and we define $\sigma'(h)$ to be $\sigma(h')$.

Given a strategy τ for player 2, we define its mirrored version τ' in a similar way. It is immediate to check that the average payoff induced by (σ', τ') in the game Γ_{1-p} is equal to the average payoff induced by (σ, τ) in Γ. This readily implies our claim.

3 The Average Cost Optimality Equality

The following Proposition is a version of the well-known Average Cost Optimality Equation (ACOE) for general Markov Decision Problems (MDP's). For a proof, see e.g. Feinberg and Shwartz (2002).

Proposition 3 *Let (S, A, r, q) be an MDP with a compact metric space S, compact action set A, continuous payoff function $r : S \times A \to \mathbf{R}$, and transition rule q: $S \times A \to \Delta(S)$ such that $q(\cdot \mid s, a)$ has finite support for every $(s, a) \in S \times A$.*

If there is a bounded function $V : S \to \mathbf{R}$ such that

$$v + V(s) = \max_{a \in A} \left(r(s,a) + \sum_{s' \in S} q(s' \mid s, a) V(s') \right) \text{ for each } s \in S, \quad (2)$$

then v is the value of the MDP, for each initial state $s \in S$. Moreover, a stationary strategy $\alpha = (\alpha(s))$ is optimal as soon as, for every $s \in S$, $\alpha(s)$ attains the maximum in (2).

On the other hand, if, for some $s^* \in S$, the sequence $n\,(v_n(\cdot) - v_n(s^*))$ has a point-wise limit V, and if the value v is independent of the initial state, then

$$v + V(s) = \sup_{a \in A} \left(r(s,a) + \sum_{s' \in S} q(s' \mid s, a) V(s') \right) \text{ for each } s \in S. \quad (3)$$

In this proposition, $\Delta(S)$ stands for the set of probability distributions over S. The value v is the supremum over all policies ϕ of the payoff function $\gamma(\phi) := \liminf \mathbf{E}_\phi[\frac{1}{N} \sum_{n=1}^{N} r(s_n, a_n)]$. Observe that the function V is determined up to an additive constant.

We will apply this result in two ways. Let first a strategy τ be given, that can be implemented with a finite automaton with state space Θ and deterministic transitions. The best-reply of player 1 when facing τ reduces to a Markov decision problem with state space $\Theta \times \{\underline{s}, \overline{s}\}$. Indeed, in any

stage, player 1 will be able to infer from the past history the current state of player 2's automaton – and knows in addition the current state. We will use this remark to prove that τ^* guarantees $p/(4p-1)$ in the game Γ_p.

Proposition 3 can be used to analyze player 1's optimal behavior in the game Γ_p. By Renault (2002) player 1 has an optimal strategy that does not depend on player 2's actions. Facing such a strategy σ, at any stage player 2 will simply compute her posterior belief, and play the action that yields the minimal current payoff, given her posterior belief and the mixed moves x and y used by player 1 in states $\underline{s}$ and $\overline{s}$. Hence, the sup inf of Γ_p can be computed under the assumption that player 2 always plays a best reply to player 1's current mixed move, given player 2's posterior belief over the current state. In other words, the value of the game Γ_p coincides with the value of the following auxiliary Markov Decision Problem, both for the finite and infinite horizon versions:

- the current state $\theta_n \in [0,1]$ corresponds to player 2's current posterior belief;
- the decision maker chooses a pair (x,y) of mixed moves;
- the current payoff is $\min\{x\theta_n, (1-y)(1-\theta_n)\}$;
- with probability $x\theta_n + y(1-\theta_n)$, (resp. $(1-x)\theta_n + (1-y)(1-\theta_n)$) the decision maker plays T (resp. B), and the next state is the conditional distribution θ_{n+1} of s_{n+1}, given the action played by the decision maker.

Moreover, the value v_n of the n-stage version of the above MDP coincides with the value of the n-stage version of the game Γ_p and it can be checked that the sequence $n\left(v_n(\cdot) - v_n(\theta^*)\right)$ has a point-wise limit V, for each choice of the initial state. It is known that the value v_n of the n-stage version of Γ_p is concave w.r.t. the initial distribution θ. Thus, V is also concave, hence continuous on $(0,1)$. Therefore, for each $\theta \in (0,1)$, one has

$$
\begin{aligned}
v_p + V(\theta) = \max_{x,y\in[0,1]} \{ & \min\left(x\theta, (1-y)(1-\theta)\right) \\
& + (\theta x + (1-\theta)y) \times V\left(p\frac{x\theta}{x\theta + (1-\theta)y} + (1-p)\frac{y(1-\theta}{x\theta + (1-\theta)y}\right) \\
& + (\theta(1-x) + (1-\theta)(1-y)) \\
& \qquad \times V\left(p\frac{(1-x)\theta}{(1-x)\theta + (1-\theta)(1-y)} + (1-p)\frac{(1-y)(1-\theta}{(1-x)\theta + (1-\theta)(1-y)}\right)\Big\}.
\end{aligned}
$$

The first term is the current payoff. In the second term, $\theta x + (1-\theta)y$ is the probability that T will be played, while the argument of V is the induced posterior belief. We will denote by $W_{x,y}$ the term between braces in the right-hand side.

Lemma 4 *There is an optimal strategy α_θ such that player 2 is always indifferent between L and R.*

Proof. We will prove that, for each θ, the maximum of $W_{x,y}$ is achieved for some $x_\theta, y_\theta \in [0,1]$ such that $\theta x_\theta = (1-\theta)(1-y_\theta)$. W.l.o.g., we assume that $\theta \leq 1/2$. Let $x, y \in [0,1]$ be a mixed action pair such that $\theta x \neq (1-\theta)(1-y)$. We will exhibit a pair x', y', with $\theta x' = (1-\theta)(1-y')$ and $W_{x',y'} \geq W_{x,y}$.

We will treat the case in which $\theta x > (1-\theta)(1-y)$. The discussion of the other case is similar and therefore omitted. Assume first that $y \geq x$. Set $x' = y' = 1-\theta$ so that $\theta x' = (1-\theta)(1-y')$. Observe that

$$W_{x',y'} = \theta x' + V(\theta).$$

Note that $\min\{\theta x', (1-\theta)(1-y')\} > \min\{\theta x, (1-\theta)(1-y)\}$. By concavity of V, it therefore follows that $W_{x',y'} > W_{x,y}$.

Assume now that $y < x$. Set $y' = y$ and decrease x to x' that satisfies $\theta x' = (1-\theta)(1-y')$. The current payoff is not affected: $\min\{\theta x', (1-\theta)(1-y')\} = (1-\theta)(1-y)$. On the other hand, less information about the current state is transmitted when using (x', y') than when using (x, y). By concavity of V, this will imply that $W_{x',y'} \geq W_{x,y}$. To be specific, the sum of the last two terms in $W_{x,y}$ is a convex combination $a_{x,y}V(\theta^T_{x,y}) + (1-a_{x,y})V(\theta^B_{x,y})$, with $a_{x,y}\theta^T_{x,y} + (1-a_{x,y})\theta^B_{x,y} = \theta$, and a similar observation holds for (x', y'). It is readily verified that $\theta^B_{x,y} < \theta^B_{x',y'} \leq \theta \leq \theta^T_{x',y'} < \theta^T_{x,y}$. The claim follows. ■

By the previous lemma, it follows that the function $V : [0,1] \to \mathbf{R}$ satisfies $V(\theta) = V(1-\theta)$ for each θ and the functional equation below, for each $\theta \leq 1/2$:

$$v_p + V(\theta) = \max_{x \in [0,1]} \left\{ \theta x + (1-\theta) V\left(1 - p + (2p-1)\frac{\theta}{1-\theta}x\right) + \theta V\left(p - (2p-1)x\right) \right\}. \tag{4}$$

Besides, any function x_θ that achieves the maximum in (4) for each θ yields an optimal stationary strategy for player 1.

As an illustration, we provide below the graphs of V for three different values of p, obtained by numerical simulation.

4 Player 2

We here prove the results relative to player 2.

Lemma 5 *The strategy τ^* of player 2 guarantees $p/(4p-1)$ for all $p \geq 1/2$.*

Proof. As player 1 knows the state of the automaton of player 2 and the true state of the game, in effect he faces a MDP with four states: $\mathcal{S} = \{(s_0,\theta_0),(s_1,\theta_0),(s_0,\theta_1),(s_1,\theta_1)\}$. In each one of these four states he has two available actions, T and B, and the payoff is:

$$\begin{aligned} r((s_0,\theta_0);T) = \frac{2p}{4p-1}, r((s_0,\theta_0);B) = 0, \\ r((s_1,\theta_0);T) = 0, r((s_1,\theta_0);B) = \frac{2p-1}{4p-1}, \\ r((s_0,\theta_1);T) = \frac{2p-1}{4p-1}, r((s_0,\theta_1);B) = 0, \\ r((s_1,\theta_1);T) = 0, r((s_1,\theta_1);B) = \frac{2p}{4p-1}. \end{aligned}$$

By the ACOE, $v = \frac{p}{4p-1}$ is the value if and only if there is a function $V : \mathcal{S} \to \mathbf{R}$ that satisfies:

$$\begin{aligned} v + V(s_0,\theta_0) = \max \Bigg\{ & \frac{2p}{4p-1} + pV(s_0,\theta_1) + (1-p)V(s_1,\theta_1), \\ & pV(s_0,\theta_0) + (1-p)V(s_1,\theta_0) \Big\}, \qquad (5) \\ v + V(s_1,\theta_0) = \max \{ & pV(s_1,\theta_1) + (1-p)V(s_0,\theta_1), \\ & \frac{2p-1}{4p-1} + pV(s_1,\theta_0) + (1-p)V(s_0,\theta_0) \Bigg\}, \qquad (6) \\ v + V(s_1,\theta_1) = \max \Bigg\{ & \frac{2p}{4p-1} + pV(s_1,\theta_0) + (1-p)V(s_0,\theta_0), \\ & pV(s_1,\theta_1) + (1-p)V(s_0,\theta_1) \Big\}, \qquad (7) \\ v + V(s_0,\theta_1) = \max \{ & pV(s_0,\theta_0) + (1-p)V(s_1,\theta_0), \\ & \frac{2p-1}{4p-1} + pV(s_0,\theta_1) + (1-p)V(s_1,\theta_1) \Bigg\}. \qquad (8) \end{aligned}$$

These equations are symmetric. One can verify that these equations imply that $V(s_0,\theta_1) = V(s_1,\theta_0)$ and $V(s_1,\theta_1) = V(s_0,\theta_0)$, and so Eqs. (**??**) to (5) can be omitted.

One can now verify that the following is a solution to these equations:

$$\begin{aligned} v &= \frac{p}{4p-1}, \\ V(s_0,\theta_0) = V(s_1,\theta_1) &= 0, \\ V(s_1,\theta_0) = V(s_0,\theta_1) &= -\frac{1}{4p-1}. \end{aligned}$$

■

5 Player 1

We here prove the results relative to player 1. We will first prove that σ^* guarantees $p/(4p-1)$ whenever $p \in [1/2, 2/3]$. Together with the results of the previous section, this yields $v_p = p/(4p-1)$ and implies that σ^* is indeed optimal for that range of values of p.

Next, we analyze the case where $p = p^* \backsimeq 0.75891$. We will prove that σ^* is optimal, and we also exhibit an optimal strategy for player 2.

Finally, we prove that the range of values of p for which σ^* is an optimal strategy of Γ_p, is an interval, thereby completing the proof of Theorem 2.

5.1 The case $p \in [1/2, 2/3]$

We here prove that σ^* is optimal in Γ_p, for all $p \in [1/2, 2/3]$. Under σ^*, and when the posterior probability of $\underline{s}$ is $\theta \leq \frac{1}{2}$, player 1 plays as follows:

$\underline{s}$

	L	R
1	1	0
0	0	0

θ

$\overline{s}$

	L	R
$1-\frac{\theta}{1-\theta}$	0	0
$\frac{\theta}{1-\theta}$	0	1

$1-\theta$

Observe that the probability that the action B is played is θ, so that the probability that the action T is played is $1-\theta$.

Lemma 6 *The strategy σ^* guarantees $p/(4p-1)$, for each $p \in [1/2, 2/3]$.*

Proof. Recall the optimality equation (4). In order to prove that the strategy σ^* yields $p/(4p-1)$, we need to find a function V, symmetric around

1/2, such that the equality

$$\frac{p}{4p-1} + V(\theta) = \theta + (1-\theta)V\left(1 - p + (2p-1)\frac{\theta}{1-\theta}\right) + \theta V(1-p)$$

holds for each $\theta \leq 1/2$.

Observe that one has $1 - p + (2p-1)\frac{\theta}{1-\theta} \geq 1/2$, whenever $p \in [1/2, 2/3]$: under σ^*, the posterior probability is either p, $1-p$, or jumps from one half of the posterior space $[1-p, p]$ to the other.

Therefore, we need only find a function $V : [1-p, 1/2] \to \mathbf{R}$, such that

$$\frac{p}{4p-1} + V(\theta) = \theta + (1-\theta)V\left(p - (2p-1)\frac{\theta}{1-\theta}\right) + \theta V(1-p).$$

One can verify that the function $V(\theta) = \dfrac{\theta}{4p-1}$ is a solution. The result follows. ■

5.2 The case $p = p^* \backsimeq 0.75891$

We here analyze the case where $p = p^*$, the unique real solution of the equation

$$9p^3 - 13p^2 + 6p - 1 = 0.$$

We show that $v_p = \frac{p}{1-3p+6p^2} \backsimeq 0.34829$ by showing that it is an equilibrium payoff. Since the game is zero-sum, the equilibrium consists of optimal strategies.

We are going to show that the following pair of strategies is an equilibrium:

- For player 1 we take the strategy σ that is defined in Section 2.2.
- For player 2 we take the following strategy that can be implemented by the following automaton with four states, $1-p$, $1-\theta$, θ and p.

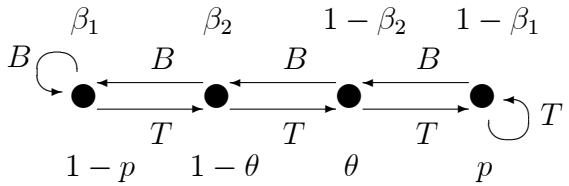


Figure 3: The strategy of player 2

In Figure 2, below each state appears its name, and above it appears the probability that the action L is played.

Denote $\theta = \frac{p}{3p-1} \backsimeq 0.5944$. One can verify that the following equations are satisfied.

$$\begin{aligned}\theta = \phi(p \mid B) &= \frac{2p-1}{p}p + \frac{(1-p)^2}{p} = p - (2p-1) \times \frac{1-p}{p}, \\ 1-\theta = \phi(\theta \mid B) &= \frac{2\theta-1}{\theta}p + (1-p)\left(1 - \frac{2\theta-1}{\theta}\right) = p - (2p-1) \times \frac{1-\theta}{\theta}.\end{aligned} \tag{9}$$

Whenever player 1 plays B, the posterior belief of player 2 evolves as follows: (i) from p it moves to θ, (ii) from θ it moves to $1-\theta$ whereas (iii) from either $1-\theta$ or $1-p$ and player 1 plays B, it moves to $1-p$. By symmetry arguments, this implies that the transitions of the automaton of player 2 mimic the evolution of his posterior belief, provided player 1 follows σ^*.

When player 1 follows σ^*, player 2 is always indifferent between playing T or B. Since σ^* is independent of player 2's moves, the payoff is independent of player 2's strategy. In particular, player 2's automaton is a best reply to σ^*.

We now turn to the optimization problem of player 1. In effect he faces a MDP with 8 states. Each state is composed of the actual state of nature (2 alternatives) and the state of the automaton of player 2 (4 alternatives). So that σ^* is the best response against the strategy of player 2, and γ is the value, the ACOE implies that there should be a function V that satisfies the following.

$$\begin{aligned}v + V(s_0, 1-p) &= \beta_1 + pV(s_0, 1-\theta) + (1-p)V(s_1, 1-\theta) \\ &\geq pV(s_0, 1-p) + (1-p)V(s_1, 1-p), \\ v + V(s_0, 1-\theta) &= \beta_2 + pV(s_0, \theta) + (1-p)V(s_1, \theta) \\ &\geq pV(s_0, 1-p) + (1-p)V(s_1, 1-p), \\ v + V(s_0, \theta) &= 1 - \beta_2 + pV(s_0, p) + (1-p)V(s_1, p) \\ &= pV(s_0, 1-\theta) + (1-p)V(s_1, 1-\theta), \\ v + V(s_0, p) &= 1 - \beta_3 + pV(s_0, p) + (1-p)V(s_1, p) \\ &= pV(s_0, \theta) + (1-p)V(s_1, \theta).\end{aligned}$$

Since V is determined up to an additive constant, we can set $V(s_0, p) = 0$, and then this system contains 6 equations in 6 variables. The unique solution

is

$$\begin{aligned}
v &= \frac{p}{1-3p+6p^2} \backsimeq 0.348291466,\\
\beta_1 &= \frac{4p-1}{1-3p+6p^2} \backsimeq 0.934232129,\\
\beta_2 &= \frac{2p}{1-3p+6p^2} \backsimeq 0.696582932,\\
V(s_0, 1-p) &= \frac{6p-2}{1-3p+6p^2} \backsimeq 1.171881327,\\
V(s_0, 1-\theta) &= \frac{2p}{1-3p+6p^2} \backsimeq 0.696582932,\\
V(s_0, \theta) &= \frac{2p-1}{1-3p+6p^2} \backsimeq 0.237649197,\\
V(s_0, p) &= 0.
\end{aligned}$$

One can verify that the incentive constraints are satisfied in this case.

6 A lower bound on v_p

Item 2 in Theorem 2 provides an upper bound on v_p for $p \geq 2/3$. We here illustrate how to compute a lower bound on v_p, and numerically investigate the tightness of these bounds.

We will obtain a lower bound by computing the payoff γ_p induced by σ^*. By the ACOE, γ_p is the unique real number such that there is a function V (symmetric around $1/2$) that satisfies

$$\gamma_p + V(\theta) = \theta + (1-\theta)V\left(1-p+(2p-1)\frac{\theta}{1-\theta}\right) + \theta V(1-p) \qquad \theta \leq 1/2$$

$$\gamma_p + V(\theta) = 1-\theta+\theta V\left(p-(2p-1)\frac{1-\theta}{\theta}\right) + (1-\theta)V(p) \qquad \theta \geq 1/2.$$

As V can be determined up to an additive constant, set $V(1-p) = 0$. Set $\theta_0 = 1-p$, and for each k

$$\theta_{k+1} = \min\left\{1-p+(2p-1)\frac{\theta_k}{1-\theta_k}, p-(2p-1)\frac{\theta_k}{1-\theta_k}\right\}.$$

Then $\theta_k \in [1-p, 1/2]$, and

$$\gamma_p + V(\theta_k) = \theta_k + (1-\theta_k)V(\theta_{k+1}) + \theta_k V(1-p).$$

Using $V(1-p) = 0$, and given the boundedness of V, it follows that:

$$\gamma_p = \frac{\theta_0 + (1-\theta_0)\theta_1 + (1-\theta_0)(1-\theta_1)\theta_2 + \cdots}{1 + (1-\theta_0) + (1-\theta_0)(1-\theta_1) + \cdots}.$$

It is convenient to introduce the sequence (u_n), with $u_0 = 1$ and $u_{n+1} = 1 - \theta_n$, so that (u_n) satisfies the recursive equation:

$$u_{n+1} = \max\left\{3p - 1 - \frac{2p-1}{u_n}, 2 - 3p + \frac{2p-1}{u_n}\right\}.$$

With these notations, the payoff γ_p is

$$\begin{aligned}\frac{1}{\gamma_p} &= \frac{u_0 + u_0u_1 + u_0u_1u_2 + \cdots}{1 - u_1 + u_1(1-u_2) + u_1u_2(1-u_3) + \cdots} \\ &= u_0 + u_0u_1 + u_0u_1u_2 + \cdots\end{aligned}$$

This last formula provides a method of computing γ_p for any p. We do not know how to explicitly solve for the sequence (u_n) in general, and no simple explicit formula γ_p seems to hold. For some values of p however, a closed formula can be obtained. Observe indeed that the recurrence equation on (u_n) writes

$$u_{n+1} = \max\{\psi(u_n), 1 - \psi(u_n)\},$$

where $\psi(u) := 3p - 1 - \dfrac{2p-1}{u}$ is increasing. Let u^* be a solution to $u = 1 - \psi(u)$. It is immediate to check that $u^* \geq 1/2$, hence $\psi(u^*) \leq u^*$, for otherwise, the inequality $2u^* < \psi(u^*) + 1 - \psi(u^*) = 1$ would hold. In particular, if $u_N = u^*$ for some N, then the sequence (u_n) is stationary from that stage on.

Next, consider the sequence (w_n) defined by $w_0 = 1 = u_0$ and

$$w_{n+1} = 3p - 1 - \frac{2p-1}{w_n}. \tag{10}$$

We claim that if $w_N = u^*$ (and $w_n \neq u^*$ for $n < N$), then $u_n = w_n$ for each $n < N$ and $u_n = u^*$ for each $n \geq N$. To prove this claim, it is enough

to check that $1-\psi(w_n) \leq \psi(w_n)$ for $n = 1, 2, \cdots, N-1$. To see why this holds, observe first that the sequence (w_n) is decreasing. We argue by contradiction, and assume that $w_{k+1} = \psi(w_k) < 1-\psi(w_k)$ for some $k < N$. Since $w_k > w_{k+1} \geq w_N$, this yields $w_N < 1-\psi(w_k)$. Since ψ is increasing, one obtains $w_N < 1-\psi(w_N)$ – a contradiction.

As a result, the computation of γ_p is easy for those values of p with the property that $w_n = u^*$ for some n.

The solution to (10) is

$$w_n = \sqrt{2p-1}\frac{\cos\left((n+1)\,\sigma-\lambda\right)}{\cos\left(n\sigma-\lambda\right)}, \tag{11}$$

where

$$\tan\sigma = \frac{\sqrt{(1-p)\,(9p-5)}}{(3p-1)}, \tan\lambda = \frac{3(1-p)}{\sqrt{(1-p)\,(9p-5)}}.$$

Using standard manipulations, the following appears. Given N, there exists a unique p such that N is the smallest integer for which $u_{N-1} = u_N$. This p solves the following polynomial equation:

$$\tan((N+1)\sigma) = \sqrt{\frac{9p-5}{1-p}} \times \frac{2p-1}{1-3p+\sqrt{p(9p-4)}}.$$

For instance, (i) $p \backsimeq .7589$ for $N = 2$, which is the special case already studied, (ii) $p \backsimeq .8583$ for $N = 3$, (iii) $p \backsimeq .9073$ for $N = 4$, (iv) $p \backsimeq .9348$ for $N = 5$, etc.

From the expression of w_n, one deduces

$$w_0 \cdots w_n = (2p-1)^{(n+1)/2}\frac{\cos\left((n+1)\,\sigma-\lambda\right)}{\cos\lambda}$$

It follows that

$$\frac{1}{\gamma_p} = 2\left(1 - \frac{(2p-1)^{(N+1)/2}\sqrt{2}(3p-1-\sqrt{p(9p-4)})}{\sqrt{p(1-p)}(3\sqrt{p}-\sqrt{9p-4})(5p-2-\sqrt{p(9p-4)})}\right).$$

For instance, $\gamma_p \backsimeq .2880$ for $p \backsimeq .8583$ (more precisely, γ_p is the unique real root of $-162+1737x-7279x^2+15002x^3-15276x^4+6169x^5 = 0$ for p the unique real root of $-4+37x-136x^2+248x^3-225x^4+81x^5 = 0$), and $\gamma_p \backsimeq .2460$ for $p \backsimeq .9073$, etc. It is readily verified that the equation $\gamma_p = \frac{p}{1-3p+6p^2}$, valid for $p \backsimeq .7589$, is not valid on any open interval.

We draw below a graph that features simultaneously the upper bound $p/(4p-1)$, and the functions v_p and γ_p. The latter two graphs were obtained numerically.

References

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

[2] Feinberg E.A. and Shwartz A. (2002) Handbook of Markov Decision Processes. Kluwer.

[3] Marino A. (2005) The Value and Optimal Strategies of a Particular Markov Chain Game. Preprint.

[4] Neyman A. (2004) Existence of Optimal Strategies in Markov Games with Incomplete Information, preprint

[5] Renault J. (2002) The Value of Markov Chain Games with Lack of Information on One Side, preprint