HAL open science

Asymmetric Warfare

Xavier Bressaud, Anthony Quas

To cite this version:

Xavier Bressaud, Anthony Quas. Asymmetric Warfare. 2014. hal-00956182

HAL Id: hal-00956182

https://hal.science/hal-00956182v1

Preprint submitted on 6 Mar 2014

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

ASYMMETRIC WARFARE

XAVIER BRESSAUD AND ANTHONY QUAS

We study a simple two player dynamic game with asymmetric information introduced by Renault in [2] and studied by Hörner, Rosenberg, Solan and Vieille in [1].

At each stage, the system is in one of two state $\underline{s}$ and $\bar{s}$, given by matrices as follows:

$$\underline{s} : \begin{pmatrix} 1 & 0 \\ 0 & 0 \end{pmatrix} \qquad \text{and} \qquad \bar{s} : \begin{pmatrix} 0 & 0 \\ 0 & 1 \end{pmatrix}$$

The players simultaneously make a choice: Player 1 plays T(top) or B(bottom), while Player 2 plays L(left) or R(right). Player 1 receives the corresponding entry of the matrix describing the current state of the system, so that if the system is in state $\bar{s}$ and Players 1 and 2 play top and left respectively, then Player 1 receives \$1, whereas otherwise he receives nothing.

A crucial aspect of the game is that Player 1 is aware of the state *before* choosing his move, whereas Player 2 is never told of the state. Each player sees the moves of the other, but is not informed of the winnings (although Player 1 can deduce this information from what is known to him, whereas Player 2 cannot). The winnings go to Player 1. Thus player 1's job is to maximize his gain, whereas Player 2's job is to minimize the gain.

Player 1 thus faces a tradeoff between short term (he has sufficient information to optimize his expected payoff in the current turn) versus long term (if he always plays so as to optimize his payoff in the current turn, then he reveals the state of the system to Player 2, who can then use this information to minimize Player 1's winnings).

The state of the system is assumed to undergo Markov evolution, where the system stays in its current state between moves with fixed probability $p \geq \frac{1}{2}$, or switches with probability $1 - p$. The dynamics governing the switching is known to both players.

The general theory of zero sum games ensures (see e.g. [2]) that there exist a value v and strategies σ for Player 1 and τ for Player 2, such that whenever Player 2 uses strategy τ, Player 1's long-term expected winnings are at most v; whereas whenever Player 1 uses strategy σ, his long-term expected winnings are at least v. Thus any strategy

for Player 1 gives a lower bound for the value of the game (by taking the infimum of the expected long-term gain over all possible counter-strategies by Player 2). Similarly any strategy for Player 2 gives an upper bound for the value of the game.

As usual in game theory, the best strategies are often mixed strategies. That is, given all of the information available to a player, his strategy returns a probability vector distributing mass to the available moves. Since we intend to work in the limiting case, we identify the following spaces describing the past: $X_1 = \{T, B\}^{\mathbb{Z}^-}$, $X_2 = \{L, R\}^{\mathbb{Z}^-}$ and $X_S = \{\underline{s}, \bar{s}\}^{\mathbb{Z}^-}$. A strategy for Player 1 can then be formally described as a map σ from $X_1 \times X_2 \times X_S$ to $[0,1]^2$, where $(\sigma(x,y,z))_1$ gives the probability of playing T if Player 1's past moves were x, Player 2's past moves were y and the system is currently in state $\underline{s}$, while $(\sigma(x,y,z))_2$ gives the probability of playing T if the system is in state $\bar{s}$. Similarly, a strategy for Player 2 is a map τ from $X_1 \times X_2$ to $[0,1]$, where $\tau(x,y)$ gives the probability of playing L if Player 1's past moves were x and Player 2's past moves were y.

Our goal, of course, is essentially to find v and the optimal strategies σ and τ. These, as one expects, drastically depend on p. The answers for $p = \frac{1}{2}$ and $p = 1$ are straightforward: in the first case, Player 1 always plays as if he were facing a one-shot game and wins with probability $\frac{1}{2}$ while in the other case he can not use his information and has to play randomly as if he did not have any advantage and wins only with probability $\frac{1}{4}$. In [1], the authors exhibit a strategy σ^* for Player 1 (defined properly in Section 2) and prove that it is optimal for all $\frac{1}{2} \leq p \leq \frac{2}{3}$. In this setting, they give a closed formula for the value v_p of the game and also provide an optimal strategy τ^* for Player 2 (based on a two state automaton). They compute the value of the strategy σ^* as the sum of a series for all values of the parameter, hence providing a lower bound for the value of the game (useful when larger than $\frac{1}{4}$), while an upper bound is given by the value of the strategy τ^*. They compute this lower bound explicitly for specific values of the parameter p larger than $\frac{2}{3}$. In the very special case $p = p^*$ solving $9x^3 - 13x^2 + 6x - 1 = 0$ ($p^* \simeq 0.7589$), they observe that σ^* is still optimal. In this case they also exhibit an optimal strategy for Player 2 (more tricky but still based on a finite automaton). Finally, they raise the question of the optimality of σ^* for instance at $p = \frac{3}{4}$. We provide a negative answer and prove:

Theorem 1. *The strategy σ^* is optimal for $p < 0.719$ and not optimal for some $p < 0.733$.*

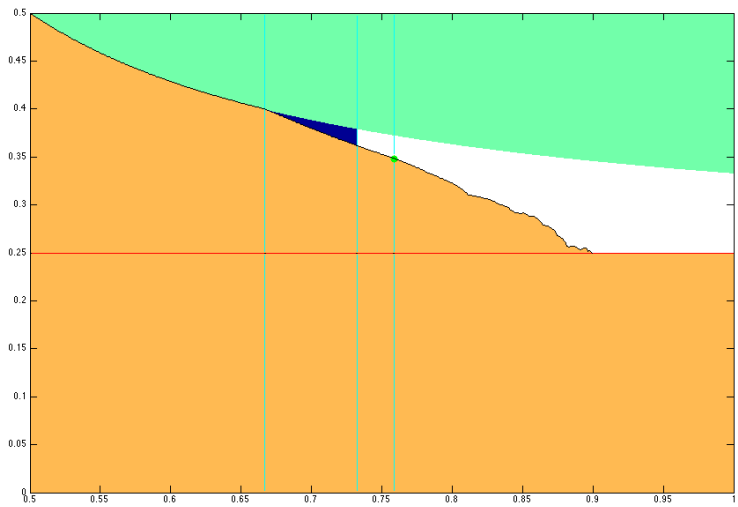

FIGURE 1. Bounds on the value of the game as a function of the parameter p : it is known from [1] that the value is between the green and the orange zones: the upper bound (green) is given by $p/(4p-1)$; it coincides with the lower bound (orange) given by the value of strategy σ^* up to $p = \frac{2}{3}$; the best known lower bound is the value of σ^* (numerically computed) as long as it is greater than 0.25 which is the easy lower bound provided by the uniform random strategy; note the special point $p^* \simeq 0.7589$ for which the lower bound is exactly the value. The blue (dark) zone indicates the improvement we make to the upper bound: σ^* is optimal up to about 0.72.

Indeed, on the one hand, we prove that σ^* is optimal and exhibit an optimal strategy for Player 2 under a condition which is valid up to some critical value p_c which we prove to be larger than 0.719, and, on the other hand, we exhibit a strategy whose value is higher than that of σ^* for some p smaller than 0.733 (so that $p_c < 0.733$). In both cases, the proofs are based on checking that a certain finite set of inequalities is satisfied.

The fact that $p_c \geq \frac{2}{3}$ was established in [1]. Experimentation strongly suggests that $p_c > 0.732$, but we have not been able to show this rigorously. The methods in this article give, for each n, a family (C_n) of inequalities, such that if p satisfies (C_n) for any n, then σ^* is optimal for p. The proof that σ^* is optimal up to 0.719 proceeds by considering

two intervals of parameters and showing that on both intervals, (C_9) is satisfied for all parameters in the interval. Further, if one picks values of p randomly in the range $[0.719, 0.732]$ and then tests (C_n) for $n = 50$, $n = 100$, ..., $n = 500$, an experiment showed that out of 10000 randomly selected p values, for each, at least one of the collections of sufficient conditions for optimality of σ^* was satisfied. It seems likely that for any $p_0 < p_c$, there is an n such that the level n conditions are uniformly satisfied by all $p \in [\frac{2}{3}, p_0]$. Unsurprisingly the first value of n for which the collection of inequalities is satisfied becomes larger as p approaches the conjectured $p_c \in (0.732, 0.733)$ and at the same time, the number of intervals of p into which the range must be sub-divided is expected to grow exponentially with n. Accordingly it is evident that one can go beyond $p = 0.719$, but continuation requires an increasing amount of effort for a decreasing amount of improvement.

We conjecture the existence of a really critical point beyond which σ^* would in general not be optimal. "In general", because as already noticed, and our scheme clearly shows, there are still special values beyond. Quite surprisingly, we had to introduce tools from dynamical systems (thermodynamic formalism) to show the optimality of σ^*. The strategy of Player 2 to which σ^* is the optimal response turns out to be a strategy that takes into account *all* the past moves of Player 1, making it crucial to closely control the behaviour of the orbit of a certain dynamical system. However, the result relies, above all, on standard tools of game theory.

The paper is laid out as follows: in Section 1, we introduce classical tools from game theory. In Section 2, we define properly the strategy σ^*, prove its basic properties and compute its value for all p. In Section 3, we search for a strategy for Player 2 to which σ^* would be an optimal response. We give a system of equations that has solutions for p smaller than 0.78. Solutions really yield the desired strategies only if they satisfy a set of inequalities. We find in Section 4 a necessary and sufficient condition for these inequalities to hold in terms of the pressure of a potential. We show in Section 5 that the pressure condition is satisfied for all p less than 0.719023. This ends the proof of the first part of Theorem 1. In Section 6 we exhibit a strategy for Player 1 with a larger gain than σ^* for certain values of p; the smallest such value of p that we found is smaller than 0.733. This will finish the proof of Theorem 1.

1. Tools from game theory

The technical framework that we use to prove these statements is the study of Markov Decision Processes (MDP). A Markov decision process is one in which the system moves around a compact state space Ω, influenced by an agent who can, at each step, choose from one of a compact (in our case, finite) set of transition probabilities on the state space, each one with a given one-step payoff. The *value* of the process is the maximal long-term expected value of the gain.

The basic theorem for these processes is the following:

Theorem 2. *Let a Markov decision process have compact state space Ω. Let Π be a compact set of probability transition measures on Ω (indexed by Ω) and let $A\colon \Pi \to \mathbb{R}$ be a continuous function. Then there exists $v \in \mathbb{R}$ and a function $V\colon \Omega \to \mathbb{R}$ such that*

$$V(\omega) + v = \max_{\mu \in \Pi} \left(A(\mu(\omega, .)) + \mathbb{E}_{\mu(\omega,.)} V \right). \tag{1}$$

We, in fact, need only the (simpler) converse, which we state in the following form.

Theorem 3. *Suppose a Markov decision process has state space Ω (not necessarily compact). Let Π be a compact set of probability transition measures on Ω (indexed by Ω) and let $A\colon \Pi \to \mathbb{R}$. Suppose there exist $v \in \mathbb{R}$ and a* ***bounded*** *function $V\colon \Omega \to \mathbb{R}$ such that (1) is satisfied.*

Then for any strategy σ, the expected long-term gain (that is the limit of the average of the A's in a sequence of moves) is at most v. If the strategy is such that for each $\omega \in \Omega$, the strategy always picks a μ such that the maximum in (1) is attained, then the strategy has long-term gain given by v.

To prove this, we define the *adapted gain* by time n to be $\tilde{G}_n = V(\omega_n) + G_n$, where G_n is the gain over the first n steps. The condition guarantees, no matter which strategy is chosen $\mathbb{E}\tilde{G}_{n+1} \leq \mathbb{E}\tilde{G}_n + v$. If a strategy is chosen for which (1) is satisfied, then we have $\mathbb{E}\tilde{G}_{n+1} = \mathbb{E}\tilde{G}_n + v$. The fact that V is bounded ensures that G_n/n has the same asymptotic behaviour as $\tilde{G}_n/n$.

We interpret $V(\omega)$ as the relative score of the position ω. This is there in order to take long-term effects into account. This can be thought of as answering the question *What is the long-term total difference between starting at some fixed ω_0 and starting at ω?* This will be finite under suitable continuity and contractivity assumptions.

The MDP equation therefore says that if I choose the kernel μ giving the maximum on the right, then my expected gain plus difference in V values is v.

The way we use this is as follows. Suppose (for example) Player 2 is looking for an optimal response to a strategy σ for Player 1. From Player 2's point of view, the state space is $\Omega = X_1 \times X_2$ (the information to which he has access).

Let us suppose that $v \in \mathbb{R}$ and $V \colon \Omega \to \mathbb{R}$ satisfies (1).

Player 2 is then trying to decide between playing L and R. Since he knows σ, he can compute the expected immediate payoff if he plays either L or R (Player 2 computes the expected payoff during the round based on his knowledge of the past, and on the assumption that Player 1 is using σ). An optimal strategy (there may be many) versus σ is any strategy that always picks the option attaining the minimum expectation.

We now turn to another frequently used idea in game theory:

Principle. Suppose that

(1) τ is an optimal response to σ; and
(2) σ is an optimal response to τ

Then σ is an optimal strategy for Player 1. Similarly τ is an optimal strategy for Player 2.

Proof. Let $v(s,t)$ denote the average long-term gain if Player 1 plays strategy s and Player 2 plays strategy t and let $v_0 = v(\sigma, \tau)$

Since we have assumed that τ is an optimal response to σ, the hypothesis implies that $v(\sigma, \tau') \geq v_0$ for any τ'. In this sense, σ is an optimal strategy for Player 1: it guarantees at least the equilibrium gain against any opposing strategy. The strategy τ is an optimal strategy for Player 2: it guarantees that Player 1 can not exceed the equilibrium gain, no matter what strategy he plays. $\square$

We exploit this principle repeatedly in the remainder of this note.

A symmetry argument explained in [1] shows that for $0 \leq p \leq \frac{1}{2}$, $v_p = v_{1-p}$. Hence, in what follows we consider the case $\frac{1}{2} \leq p \leq 1$. We will be looking mainly at the strategy σ^* introduced in [1].

In what follows, if Player 1 is assumed to be playing using the strategy σ^* (to be defined below), we frequently refer to Player 2's *belief* that the system is in state $\underline{s}$. This is also sometimes referred to also as the *fictitious probability that the system is in state* $\underline{s}$. More formally, this is just the conditional probability that the system is in the state $\underline{s}$ given all the information available to Player 2 (that is the sequence of past moves made by both players), given that Player 1 is using σ^*.

Of course since Player 1 is able to see the same information as Player 2 (and more), he can also calculate Player 2's belief that the system is in state $\underline{s}$.

2. The strategy σ^*

The strategy σ^* that we study for Player 1 is apparently circular in that the moves that Player 1 makes are based on Player 2's belief of the system's state under the assumption that Player 1 is playing σ^*! One way to resolve this circularity was introduced in the paper of Hörner, Rosenberg, Solan and Vieille [1]. Here, the authors used a finite sequence of moves, rather than an infinite past. In that case, Player 1 could calculate Player 2's belief at the nth stage based on the $n-1$ moves that preceded it, and could make the nth move on the basis of that.

We use an alternative way of resolving the circularity. We introduce a quantity θ, define its evolution and show retrospectively that θ is exactly the same as Player 2's belief assuming strategy σ^*.

We define two maps as follows:

$$f_B(\theta) = \begin{cases} p\frac{2\theta-1}{\theta} + (1-p)\frac{1-\theta}{\theta} & \text{if } \theta \geq \frac{1}{2}; \\ 1-p & \text{if } \theta \leq \frac{1}{2}. \end{cases}$$

$$f_T(\theta) = \begin{cases} p & \text{if } \theta \geq \frac{1}{2}; \\ p\frac{\theta}{1-\theta} + (1-p)\frac{1-2\theta}{1-\theta} & \text{if } \theta \leq \frac{1}{2}. \end{cases}$$

Notice that $f_B(\theta) = 1 - f_T(1-\theta)$. We define a function Φ by setting $\Phi(x)$ to be $f_B(x)$ if $x \geq \frac{1}{2}$ and $f_T(x)$ otherwise. We set $p_n = \Phi^n(p)$ for all $n \geq 0$.

The strategy σ^* is then defined as follows: Player 1 maintains an internal 'state', θ.

$$\mathbb{P}(\text{playing T}) = \begin{cases} 1 & \text{if the system is in state } \underline{s} \text{ and } \theta \leq \frac{1}{2}; \\ \frac{1-2\theta}{1-\theta} & \text{if the system is in state } \bar{s} \text{ and } \theta \leq \frac{1}{2}; \\ \frac{1-\theta}{\theta} & \text{if the system is in state } \underline{s} \text{ and } \theta \geq \frac{1}{2}; \\ 0 & \text{if the system is in state } \bar{s} \text{ and } \theta \geq \frac{1}{2}. \end{cases}$$

Otherwise, he plays B. After playing, his state is updated by the rule $\theta_{\text{new}} = f_T(\theta)$ if he plays T and $\theta_{\text{new}} = f_B(\theta)$ if he plays B.

Lemma 4. *Suppose that Player 1 is playing strategy σ^* and that Player 2 is playing an arbitrary strategy τ. If prior to a move Player 1's state matched Player 2's belief that the system was in state $\underline{s}$, then the same holds after the move also.*

Observe that the Player 2's belief would also be the belief of an external observer aware only of the choices of the players.

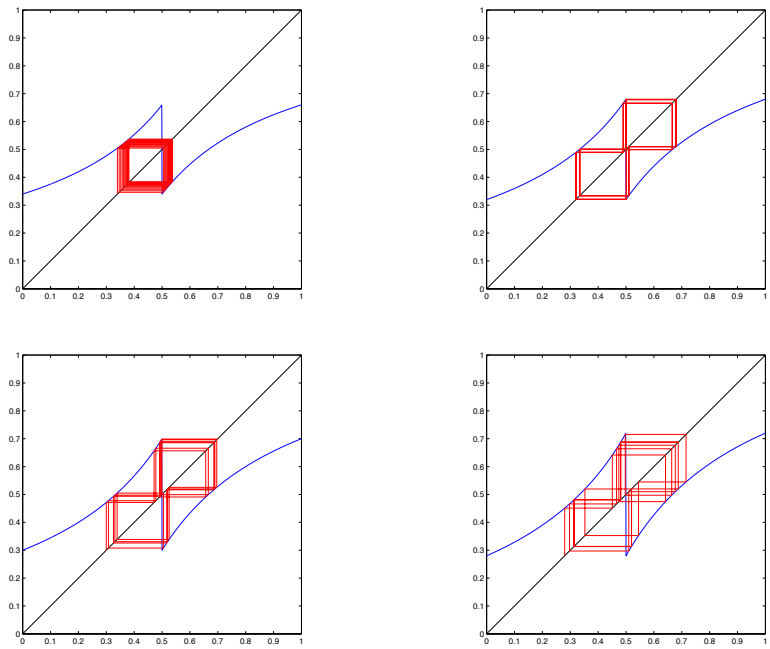

FIGURE 2. The graphs of $\theta \mapsto \Phi(\theta)$ and first points of the orbit of $1-p$, for values of p ranging from $p = 0.66$ to $p = 0.72$ in steps of 0.02.

Proof. Suppose without loss of generality that Player 2's belief prior to a move that the system is in state $\underline{s}$ is $\theta \geq \frac{1}{2}$. By assumption, this is also Player 1's state.

Player 2 then anticipates the following possibilities:

- The system is in state $\underline{s}$, and Player 1 plays T with probability $\theta \times (1-\theta)/\theta = 1-\theta$
- The system is in state $\underline{s}$, and Player 1 plays B with probability $\theta \times (2\theta - 1)/\theta = 2\theta - 1$.
- The system is in state $\bar{s}$ and Player 1 plays B with probability $(1-\theta) \times 1 = 1-\theta$.

If Player 1 plays T, then Player 2 deduces that the system was in state $\underline{s}$ (as if the system were in state $\bar{s}$, he would not play B). Accordingly the probability that the system is in state $\underline{s}$ at the next step is p.

Conversely, if Player 1 plays B, then conditioned on this, Player 2 concludes that the system was in state $\underline{s}$ with probability $(2\theta-1)/\theta$ and in state $\bar{s}$ with probability $(1-\theta)/\theta$. Hence his updated belief that the system is in state $\underline{s}$ is $p(2\theta-1)/\theta + (1-p)(1-\theta)/\theta$. In either case, we see that his updated belief matches the new state of Player 1.

□

The critical feature of σ^* that we make use of is the fact that the expected gain for Player 1 if he plays σ^* is the same no matter which strategy is used by Player 2. We demonstrate this in the lemma below. In view of the Principle above, if one can find a strategy τ, to which σ^* is the optimal response, then σ^* and τ are optimal strategies for Player 1 and Player 2 respectively.

Lemma 5. *The expected gain for Player 1 when playing strategy σ^* is independent of the strategy played by Player 2.*

Proof. We consider the Markov decision process for Player 2. The state of the process will be just his fictitious probability, θ, that the system is in the state $\underline{s}$. His action has no effect on the evolution of the state, and so his chosen move will just be the one with the lower expected payoff.

Suppose without loss of generality that $\theta \geq \frac{1}{2}$. Then if Player 2 plays R, then Player 1 gains if the system was in state $\bar{s}$ (if $\theta \geq \frac{1}{2}$ then Player 1 always plays B if the system is in state $\bar{s}$). The expected gain for Player 1 from this strategy is therefore $1 - \theta$. Similarly, if Player 2 plays L, then Player 1 gains if the system was in state $\underline{s}$ and Player 1 chose to play T. This happens with probability $\theta \times (1 - \theta)/\theta = 1 - \theta$.

Similarly, if $\theta < \frac{1}{2}$, the expected gain for Player 1 is θ, independently of any move played by Player 2.

Hence any strategy has the same expected one-step gain from any position. The next position attained by the system is independent of the move made by Player 2. □

Consider the evolution of Player 2's beliefs. These always belong to the set $\bigcup_{n\geq 0} \Phi^n\{p, 1-p\}$. Notice that even though p is not explicitly written, the value of $f_B(x)$ and $f_T(x)$ depend on p. Since for $x \geq \frac{1}{2}$, we have $f_B(1-x) = 1 - f_T(x)$, we have $\Phi^n(1-p) = 1 - \Phi^n(p)$ for all n.

When $\theta \geq \frac{1}{2}$, the belief returns to p when the system is in state $\underline{s}$ and Player 1 selects T. If $\theta > \frac{1}{2}$ and the system is in state $\bar{s}$ (i.e. there is a mismatch between Player 2's belief and the state of the system), Player 1 never selects T. When $\theta \leq \frac{1}{2}$, the belief returns to $1-p$ when the system is in state $\bar{s}$ and Player 1 selects B.

We view this as a ladder with base $\{p, 1-p\}$ and rungs $\{p_n, 1-p_n\}$, for $n \geq 1$, on which the belief follows a Markov chain: at each step, one either ascends one level, or falls down to the base. Falling off corresponds to making the choice that returns the state to p or $1-p$.

Lemma 6. *If Player 1 plays strategy σ^*, then his long-term expected gain is equal to the proportion of time spent at the base of the ladder, irrespective of the strategy played by Player 2.*

We can therefore deduce an explicit lower bound (in the form of an infinite sum) for the value of the game as a function of the parameter p.

Proof. Consider the evolution of Player 2's beliefs. These always belong to the set $\bigcup_{n\geq 0} \Phi^n\{p, 1-p\}$.

Recall from Lemma 5 that if Player 2's belief is θ, the expected payoff for Player 1 is given by is $\min(\theta, 1-\theta)$ independently of the strategy played by Player 2.

On the other hand, the probability of returning to p or $1-p$ from θ or $1-\theta$ is also $\min(\theta, 1-\theta)$. We verify this in the case $\theta \geq \frac{1}{2}$. The belief returns to p only if the system is in state $\underline{s}$ and Player 1 selects T. The probability of this is $1-\theta = \min(\theta, 1-\theta)$ as required.

Hence from the nth rung of the ladder, the probability of falling off is $\min(\Phi^n(p), 1-\Phi^n(p))$. This is the same as the expected payoff from that state. That is, in any position, the expected payoff from the next turn is equal to the probability of falling off the ladder at the next turn. We let $u_n = \max(\Phi^n(p), 1-\Phi^n(p))$ be the complementary probability: the probability of continuing up the ladder from the nth stage.

One can check that for this Markov chain, the stationary distribution gives level n probability

$$\pi_n = \frac{u_0 \dots u_{n-1}}{1 + u_0 + u_0 u_1 + u_0 u_1 u_2 + \dots}.$$

We do not specify any initial measure, but the renewal structure of the chain shows that on the long term the gain is described by the invariant measure, independently of the initial conditions: after a random but finite amount of time, Player 1 will play so that θ becomes p (or $1-p$).

Since in any state, the expected gain is the same as the probability of 'falling off the ladder', we see that the expected gain for Player 1 if he plays σ^* is given by

$$v = \frac{1}{1 + u_0 + u_0 u_1 + u_0 u_1 u_2 + \dots},$$

irrespective of Player 2's strategy, where we recall that the quantities $(u_i)_{i\geq 0}$ are functions of p. Observe that this expression was already derived in [1]. □

We give an alternative expression for $1/v$ in the form of a sum of matrix products. This is not strictly necessary for what follows, but it

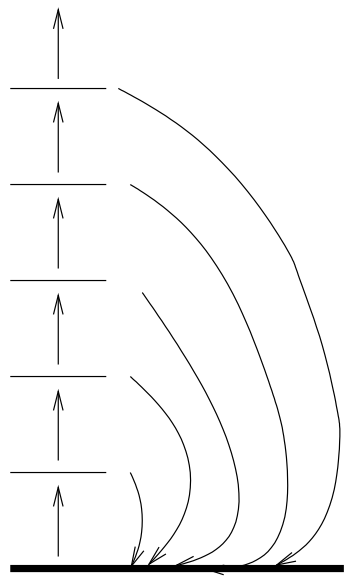

FIGURE 3. Player 2's fictitious probability that the system is in state $\underline{s}$ can be modeled by a ladder: if Player 1 plays B while $\theta < \frac{1}{2}$; or T while $\theta > \frac{1}{2}$, then the fictitious probability becomes $1-p$ or p respectively, corresponding to the bottom rung of the ladder. Note that the nth rung of the ladder corresponds both to $\Phi^n(p)$ and $\Phi^n(1-p)$

is here as it is reassuring to compare this with an expression that arises later for $1/v$.

We will write $p_n = \Phi^n(p)$ as a quotient of two polynomials in p: $p_n = a_n/b_n$, so that $p_0 = p/1$. Also write $\epsilon_n = 1$ if $p_n \geq \frac{1}{2}$ and 0 otherwise.

If $\epsilon_n = 1$, we have $p_{n+1} = f_B(p_n)$, while if $\epsilon_n = 0$, we have $p_{n+1} = f_T(p_n)$.

If $\epsilon_n = 1$, we have $u_n = p_n = a_n/b_n$ and

$$\begin{aligned}\frac{a_{n+1}}{b_{n+1}} &= f_B(a_n/b_n) = \frac{p(2a_n - b_n)/b_n + (1-p)(b_n - a_n)/b_n}{a_n/b_n} \\ &= \frac{a_n(3p-1) - b_n(2p-1)}{1a_n + 0b_n}.\end{aligned}$$

Similarly if $\epsilon_n = 0$, we have $u_n = 1 - p_n = (b_n - a_n)/b_n$ and

$$\begin{aligned}\frac{a_{n+1}}{b_{n+1}} = f_T(a_n/b_n) &= \frac{pa_n/b_n + (1-p)(b_n - 2a_n)/b_n}{(b_n - a_n)/b_n} \\ &= \frac{(3p-2)a_n + (1-p)b_n}{-a_n + b_n}.\end{aligned}$$

In both cases, we see that $u_n = b_{n+1}/b_n$. Introducing matrices $U_1 = \begin{pmatrix} 3p-1 & -(2p-1) \\ 1 & 0 \end{pmatrix}$ and $U_0 = \begin{pmatrix} 3p-2 & 1-p \\ -1 & 1 \end{pmatrix}$, we have

$$\begin{pmatrix} a_{n+1} \\ b_{n+1} \end{pmatrix} = U_{\epsilon_n} \begin{pmatrix} a_n \\ b_n \end{pmatrix}.$$

Now, taking the product of the u_n's, we obtain *par téléscopage* $u_0 \cdots u_n = b_{n+1}/b_0 = b_{n+1}$. Hence we get the expression

$$u_0 u_1 \cdots u_n = b_{n+1} = \begin{pmatrix} 0 & 1 \end{pmatrix} U_{\epsilon_n} \dots U_{\epsilon_0} \begin{pmatrix} p \\ 1 \end{pmatrix}.$$

Summing over n, we obtain another expression for the average long-term gain that will accrue to Player 1 if he plays σ^*.

$$\frac{1}{v} = \begin{pmatrix} 0 & 1 \end{pmatrix} (I + U_{\epsilon_0} + U_{\epsilon_1} U_{\epsilon_0} + U_{\epsilon_2} U_{\epsilon_1} U_{\epsilon_0} + \dots) \begin{pmatrix} p \\ 1 \end{pmatrix} \tag{2}$$

3. Strategies for Player 2

In [1], the authors showed that σ^* is optimal for $p \in [\frac{1}{2}, \frac{2}{3}]$ and for a specific value $p^* \approx 0.758$ that is the unique value of p for which $p_1 > \frac{1}{2}$ and $p_1 = p_3$. In both cases, they exhibit a strategy for Player 2 based on a finite state automaton where transitions in the automaton are governed by actions of Player 1 and then show that σ^* is an optimal response to this strategy. For $p > \frac{2}{3}$, we are going to proceed along the same lines, except that strategies for Player 2 will be based on a countable state automaton rather than a finite one. In this section, we identify strategies for Player 2 that are candidates for this purpose. The proof that they have the correct property (that σ^* is an optimal response to the strategies τ_p that we construct) is in the next two sections.

As follows from Lemma 5, any strategy of Player 2 is an optimal response to σ^*.

In the case $\frac{1}{2} \leq p \leq \frac{2}{3}$, one can check that the range of f_B is in $[1-p, \frac{1}{2}]$, while the range of f_T is in $[\frac{1}{2}, p]$. Thus the last move of Player 1 (assuming that Player 1 is using strategy σ^*) is sufficient to

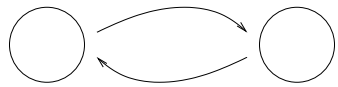

FIGURE 4. For $p < \frac{2}{3}$, Player 2's automaton has two states, capturing whether he believes it's more likely the system is in $\underline{s}$ or $\bar{s}$. Which state he is in depends solely on Player 1's last move.

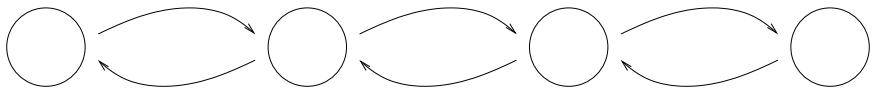

FIGURE 5. For $p = p^*$, Player 2's automaton has 4 states. The states correspond to different values of Player 2's belief that the system is in state $\underline{s}$. Transitions between states are completely determined by Player 1's moves.

determine whether Player 2 believes that it is more likely that the system is in state $\underline{s}$ or $\bar{s}$. The strategy τ^* proposed for Player 2 is a mixed strategy, playing L with probability $(2p-1)/(4p-1)$ and R with probability $2p/(4p-1)$ if $\theta > \frac{1}{2}$ and with the reverse probabilities otherwise (see Figure 4). In [1], is shown that σ^* is an optimal response to τ^* hence (σ^*, τ^*) is an equilibrium.

In the case $p = p^*$, it turns out there are only 4 possible values attained by Player 2's fictitious probability that the system is in state $\underline{s}$. Namely, we have $1-p < f_T(1-p) < f_B(p) < p$ and f_T maps $1-p$, $f_T(1-p)$, $f_B(p)$ and p to $f_T(1-p)$, $f_B(p)$, p and p respectively. Similarly f_B maps $1-p$, $f_T(1-p)$, $f_B(p)$ and p to $1-p$, $1-p$, $f_T(1-p)$ and $f_B(p)$ respectively. [1] shows that σ^* is an optimal response to a strategy τ^{**} (and hence an equilibrium strategy), given by a 4 state automaton corresponding to these four values of θ together with rules corresponding to the above: if Player 1 plays T, then the system moves one step to the right; if Player 1 plays B, then system moves one step to the left (see Figure 5). In each state of the automaton, there is an associated probability distribution on Player 2's choice of L or R, which they exhibit explicitly.

Our results are based on exhibiting strategies for Player 2 for which he plays L and R with non-zero probabilities that depend solely on his belief that the system is in state $\underline{s}$. Since Player 2's beliefs evolve in a manner that only depends on the actions of Player 1, we may once again describe his strategy by an automaton. The principal differences are: (1) the automaton generally has a countably infinite number of

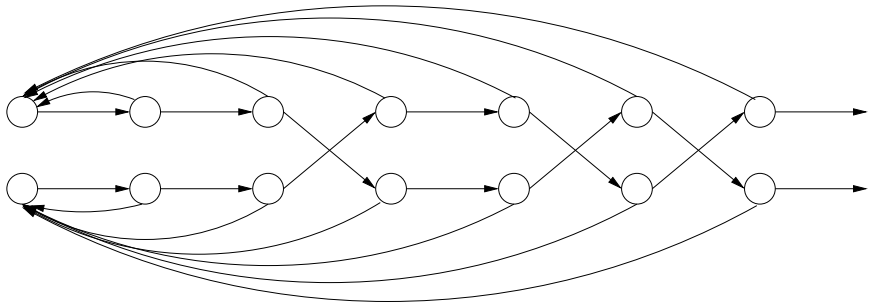

FIGURE 6. A typical automaton for a strategy of Player 2. The states on the left of the diagram are those where the belief of Player 2 is p or $1-p$. Those in the upper half of the diagram are those where Player 2 believes it is more likely the system is in state $\underline{s}$. For states in the upper half of the diagram, if Player 1 plays T, the state returns to $p_0 = p$, while if Player 1 plays B, the state advances to the right. In the lower half of the diagram, if Player 1 plays B, the state returns to $1-p$, while it advances if Player 1 plays T. The pattern of which arrows switch sides and which continue depends on p.

states; and (2) the entire structure of the automaton depends on p. An example of such an automaton is shown in Figure 6.

The pattern of arrows is completely determined by p. The description of the strategy will be complete once we specify for each state, the probability of playing L. Recall that the states are labelled by $(p_n)_{n\geq 0}$ and $(1-p_n)_{n\geq 0}$. If the automaton is in state θ, we will define $x(\theta)$ to be the probability that Player 2 chooses L. In this case, we will say that $(x(\theta))_{\theta\in[1-p,p]}$ is the strategy that Player 2 is playing.

As mentioned above, to show that σ^* is optimal, it suffices to find a strategy $x(\theta)$, to which σ^* is the best response. We therefore suppose that a particular strategy $x(\theta)$ has been selected by Player 2, and we ask whether σ^* is a best response for Player 1. We will show that for certain p and $x(\theta)$, we can solve the equations (1) for Player 1.

The state space that we use for Player 1 will consist of a pair (θ, s), where θ is Player 2's fictitious probability that the system is in state $\underline{s}$ and $s \in \{\underline{s}, \bar{s}\}$ is the state of the system.

If Player 1 plays T starting from the state $(\theta, \underline{s})$, his expected gain is $x(\theta)$ and the system will move to $(f_T(\theta), \underline{s})$ or $(f_T(\theta), \bar{s})$ with probability p and $1-p$ respectively. If Player 1 plays B starting from $(\theta, \underline{s})$, his expected gain is 0 and the system moves to $(f_B(\theta), \underline{s})$ or $(f_B(\theta), \bar{s})$ with

probability p and $1-p$ respectively. The transitions from $(\theta, \bar{s})$ are, of course, similar.

In order for the strategy σ^* to be optimal, we are looking for a value v and a relative score function $V(s,\theta)$ satisfying (1). We define $G(\theta) = V(\underline{s},\theta)$ and $H(\theta) = V(\bar{s},\theta)$. We will look for solutions respecting the symmetry of the game, that is such that $G(\theta) = H(1-\theta)$. Similarly, we impose the condition $x(\frac{1}{2}) = \frac{1}{2}$. Since there is an arbitrary additive constant, we use it to impose the condition $G(p) = -H(p)$, and we define $G(p) = -Z$. We have chosen the negative sign because we interpret the relative score to be a measure of Player 2's lack of knowledge of the state of the system. The greater the distance between his belief and the actual state of the system, the easier we expect it to be for Player 1 to exploit this mismatch.

From the form of the strategy σ^*, for σ^* to be optimal, we require that both B and T have equal expected *gain plus relative score* when θ matches the state (that is when $\theta > \frac{1}{2}$ and $s = \underline{s}$; or $\theta < \frac{1}{2}$ and $s = \bar{s}$). Further this value of expected gain plus relative score should match v plus the current relative score.

When there is a mismatch (or $\theta = \frac{1}{2}$), the v plus the current relative score should be equal to the expected gain plus relative score if Player 1 plays according to the actual state of the system and this value should be at least as large as the expected gain plus relative score if Player 1 were to play the opposite move.

In equations, we require if $\theta > \frac{1}{2}$,

$$\begin{aligned} G(\theta) + v &= x(\theta) + pG(p) + (1-p)H(p) \\ &= pG(f_B(\theta)) + (1-p)H(f_B(\theta)); \\ H(\theta) + v &= (1 - x(\theta)) + pH(f_B(\theta)) + (1-p)G(f_B(\theta)) \\ &\geq pH(p) + (1-p)G(p). \end{aligned} \tag{3}$$

If $\theta < \frac{1}{2}$, we require

$$\begin{aligned} G(\theta) + v &= x(\theta) + pG(f_T(\theta)) + (1-p)H(f_T(\theta)) \\ &\geq pG(1-p) + (1-p)H(1-p); \\ H(\theta) + v &= 1 - x(\theta) + pH(1-p) + (1-p)G(1-p) \\ &= pH(f_T(\theta)) + (1-p)G(f_T(\theta)). \end{aligned} \tag{4}$$

Finally, if $\theta = \frac{1}{2}$, then the strategy σ^* tells Player 1 to play T if the state is $\underline{s}$ and B if the state is $\bar{s}$. In either case, the expected gain is $\frac{1}{2}$.

Hence we have

(5)
$$G(\tfrac{1}{2}) + v = \tfrac{1}{2} + pG(1-p) + (1-p)H(1-p) \geq pG(p) + (1-p)H(p);$$
$$H(\tfrac{1}{2}) + v = \tfrac{1}{2} + pH(p) + (1-p)G(p) \geq pH(1-p) + (1-p)G(p).$$

Using the above, we express $x(\theta)$ in terms of the other unknown quantities:

$$x(\theta) = \begin{cases} G(\theta) + v + \gamma Z & \text{if } \theta < \frac{1}{2}; \\ 1 - (H(\theta) + v + \gamma Z) & \text{if } \theta > \frac{1}{2}; \\ \frac{1}{2} & \text{if } \theta = \frac{1}{2}. \end{cases} \tag{6}$$

Given this, solutions to (3),(4) and (5) are in correspondence with solutions to the following matrix equations together with some inequalities that we list below.

$$\begin{pmatrix} G(\theta) \\ H(\theta) \end{pmatrix} = A_\epsilon \begin{pmatrix} G(\Phi(\theta)) \\ H(\Phi(\theta)) \end{pmatrix} - v \begin{pmatrix} 1 \\ 1 \end{pmatrix} + (1-\gamma Z) b_\epsilon \text{ for } \theta \neq \tfrac{1}{2} \tag{7}$$

$$G(\tfrac{1}{2}) = H(\tfrac{1}{2}) = \tfrac{1}{2} - v - \gamma Z, \tag{8}$$

where ϵ is 1 if $\theta > \frac{1}{2}$ and 0 if $\theta < \frac{1}{2}$. $A_0 = \begin{pmatrix} \gamma & -\gamma \\ 1-p & p \end{pmatrix}$, $A_1 = \begin{pmatrix} p & 1-p \\ -\gamma & \gamma \end{pmatrix}$, $b_0 = \begin{pmatrix} 1 \\ 0 \end{pmatrix}$, and $b_1 = \begin{pmatrix} 0 \\ 1 \end{pmatrix}$.

Exploiting the symmetries in the definitions of G and H, the inequalities that need to be satisfied from (3), (4) and (5) and to ensure $x(\theta)$ lies in the range $[0,1]$ are

$$\begin{aligned} G(\theta) &\geq \gamma Z - v \text{ for } \theta < \tfrac{1}{2} \\ -\gamma Z - v \leq G(\theta) &\leq 1 - \gamma Z - v \text{ for } \theta > \tfrac{1}{2} \\ 4\gamma Z &\leq 1. \end{aligned} \tag{9}$$

These inequalities come from (4), (6) and (5) respectively. A crude calculation shows that $A_0(p)$ and $A_1(p)$ are contracting for $p < 0.788$ and so we can iterate (7) to obtain expressions for $G(\theta)$ and $H(\theta)$ in terms of which side of the interval iterates of $\Phi^n(\theta)$ lie in. Hence provided (9) are satisfied, it follows that σ^* is the optimal response to the strategy for Player 2 given by $x(\theta)$.

We will show that the inequalities in (9) are satisfied in certain parameter ranges by showing that function G is monotone decreasing

(and hence the function H is monotone increasing). Then to establish the estimates on ranges of θ values, one can simply show that the estimates hold at the appropriate endpoints.

As a consequence of the above equalities, we have

$$\begin{aligned}\begin{pmatrix} G(\theta) \\ H(\theta) \end{pmatrix} &= -v(I + A_{\eta_0} + A_{\eta_0}A_{\eta_1} + \ldots)\begin{pmatrix} 1 \\ 1 \end{pmatrix} \\ &\quad + (1-\gamma Z)(b_{\eta_0} + A_{\eta_0}b_{\eta_1} + A_{\eta_0}A_{\eta_1}b_{\eta_2} + \ldots),\end{aligned} \tag{10}$$

where $(\eta_n)_{n\geq 0}$ is the sequence of sides of the interval that θ lands in when iterating Φ. Noticing that $b_\eta + A_\eta \begin{pmatrix} 1 \\ 1 \end{pmatrix} = \begin{pmatrix} 1 \\ 1 \end{pmatrix}$ for $\eta = 0, 1$, we see that the contractions $\vec{y} \mapsto A_\eta \vec{y} + b_\eta$ have a common fixed point, so that applying the contractions in any sequence to any point, one has convergence to $\begin{pmatrix} 1 \\ 1 \end{pmatrix}$. It follows that $b_{\eta_0} + A_{\eta_0}b_{\eta_1} + A_{\eta_0}A_{\eta_1}b_{\eta_2} + \ldots = \begin{pmatrix} 1 \\ 1 \end{pmatrix}$. It readily follows that we can rewrite (10) :

$$\begin{pmatrix} G(\theta) \\ H(\theta) \end{pmatrix} = -v(I + A_{\eta_0} + A_{\eta_0}A_{\eta_1} + \ldots)\begin{pmatrix} 1 \\ 1 \end{pmatrix} + (1-\gamma Z)\begin{pmatrix} 1 \\ 1 \end{pmatrix}. \tag{11}$$

Recall that (ϵ_n) is the sequence of sides of the interval that the particular point p lands in under iteration, rather than an arbitrary point θ. Define $\vec{w}$ by

$$\vec{w} = (I + A_{\epsilon_0} + A_{\epsilon_0}A_{\epsilon_1} + \ldots)\begin{pmatrix} 1 \\ 1 \end{pmatrix}.$$

Recall that $G(p) = -Z$ and $H(p) = Z$, so that substituting this and the above observations and definitions into (10), we obtain

$$\begin{pmatrix} -Z \\ Z \end{pmatrix} = -v\begin{pmatrix} w_1 \\ w_2 \end{pmatrix} + (1-\gamma Z)\begin{pmatrix} 1 \\ 1 \end{pmatrix}$$

Finally, we rewrite this as

$$\begin{aligned}(2p-2)Z + vw_1 &= 1 \\ 2pZ + vw_2 &= 1.\end{aligned}$$

These equations can be solved to give expressions for v and Z as:

$$\begin{aligned}v &= 1/(pw_1 + (1-p)w_2) \\ Z &= (w_1 - w_2)v/2.\end{aligned} \tag{12}$$

At this stage, we have proved that

Lemma 7. *If $\frac{1}{2} \leq p \leq 0.78$, Equation (6) together with (11) and (12) define a function $(x(\theta))_{\theta\in[0,1]}$. If it satisfies Inequalities (9) then*

$(x(\theta))_{\theta\in[p,1-p]}$ *yields a strategy for Player 2 to which* σ^* *is a best response.*

Notice that we now have a second, apparently independent equation for v. For reassurance, we verify that the expressions are, in fact, equal. Starting from this second expression, we have

$$\begin{aligned} 1/v &= \begin{pmatrix} p & 1-p \end{pmatrix} (I + A_{\epsilon_0} + A_{\epsilon_0}A_{\epsilon_1} + A_{\epsilon_0}A_{\epsilon_1}A_{\epsilon_2} + \ldots) \begin{pmatrix} 1 \\ 1 \end{pmatrix} \\ &= \begin{pmatrix} 1 & 1 \end{pmatrix} (I + A^T_{\epsilon_0} + A^T_{\epsilon_1}A^T_{\epsilon_0} + A^T_{\epsilon_2}A^T_{\epsilon_1}A^T_{\epsilon_0} + \ldots) \begin{pmatrix} p \\ 1-p \end{pmatrix} \end{aligned}$$

Notice that

$$\begin{pmatrix} 1 & 0 \\ 1 & 1 \end{pmatrix} A^T_\epsilon \begin{pmatrix} 1 & 0 \\ -1 & 1 \end{pmatrix} = U_\epsilon,$$

for $\epsilon \in \{0, 1\}$.

Accordingly, we can rewrite the expression for $1/v$ as

$$\begin{aligned} &\begin{pmatrix} 1 & 1 \end{pmatrix} \begin{pmatrix} 1 & 0 \\ -1 & 1 \end{pmatrix} (I + U_{\epsilon_0} + U_{\epsilon_1}U_{\epsilon_0} + U_{\epsilon_2}U_{\epsilon_1}U_{\epsilon_0} + \ldots) \begin{pmatrix} 1 & 0 \\ 1 & 1 \end{pmatrix} \begin{pmatrix} p \\ 1-p \end{pmatrix} \\ &= \begin{pmatrix} 0 & 1 \end{pmatrix} (I + U_{\epsilon_0} + U_{\epsilon_1}U_{\epsilon_0} + U_{\epsilon_2}U_{\epsilon_1}U_{\epsilon_0} + \ldots) \begin{pmatrix} p \\ 1 \end{pmatrix}. \end{aligned}$$

This expression matches the one that we found in (2).

4. Conditions for monotonicity

To prove that the inequalities (9) are satisfied (in a range of values of p) we are going to show that G (and H) are monotonic, and we control the boundary values. In this section, we show that they follow from a *pressure* condition for a given potential for the dynamical system Φ.

Write $|\cdot|$ for the Euclidean norm on $\mathbb{R}^2$. We can check that for $p \lesssim 0.788$, the matrices A_0 and A_1 are contractions of $\mathbb{R}^2$. Let $\alpha < 1$ be the contraction constant, so that $|A_\epsilon x| \leq \alpha|x|$

Define the space, $L^\infty = L^\infty([1-p, p], \mathbb{R}^2)$, of measurable $\mathbb{R}^2$-valued functions on $[1-p, p]$ with norm given by $\|X\| = \text{ess sup}_{\theta\in[1-p,p]} |X(\theta)|$.

We then define an operator, $\mathcal{L}$, on L^∞ by

$$\mathcal{L}X(\theta) = A_{\epsilon(\theta)}X(\Phi(\theta)) - v \begin{pmatrix} 1 \\ 1 \end{pmatrix} + (1 - \gamma Z)b_{\epsilon(\theta)} \tag{13}$$

Notice that $\|\mathcal{L}X - \mathcal{L}Y\| \leq \alpha\|X - Y\|$, so that $\mathcal{L}$ is a contraction of L^∞, a complete metric space. Hence, by the Banach contraction mapping theorem, $\mathcal{L}$ has a unique fixed point, X^*. Let $X_n = \mathcal{L}^n 0$.

Write $\Delta f(\theta)$ for $\lim_{t\to\theta^+} f(t) - \lim_{t\to\theta^-} f(t)$. By symmetry, we see that $\Delta X_n(\frac{1}{2})$ is a multiple of $\left(\begin{smallmatrix} -1 \\ 1 \end{smallmatrix}\right)$ for all n. From the contraction

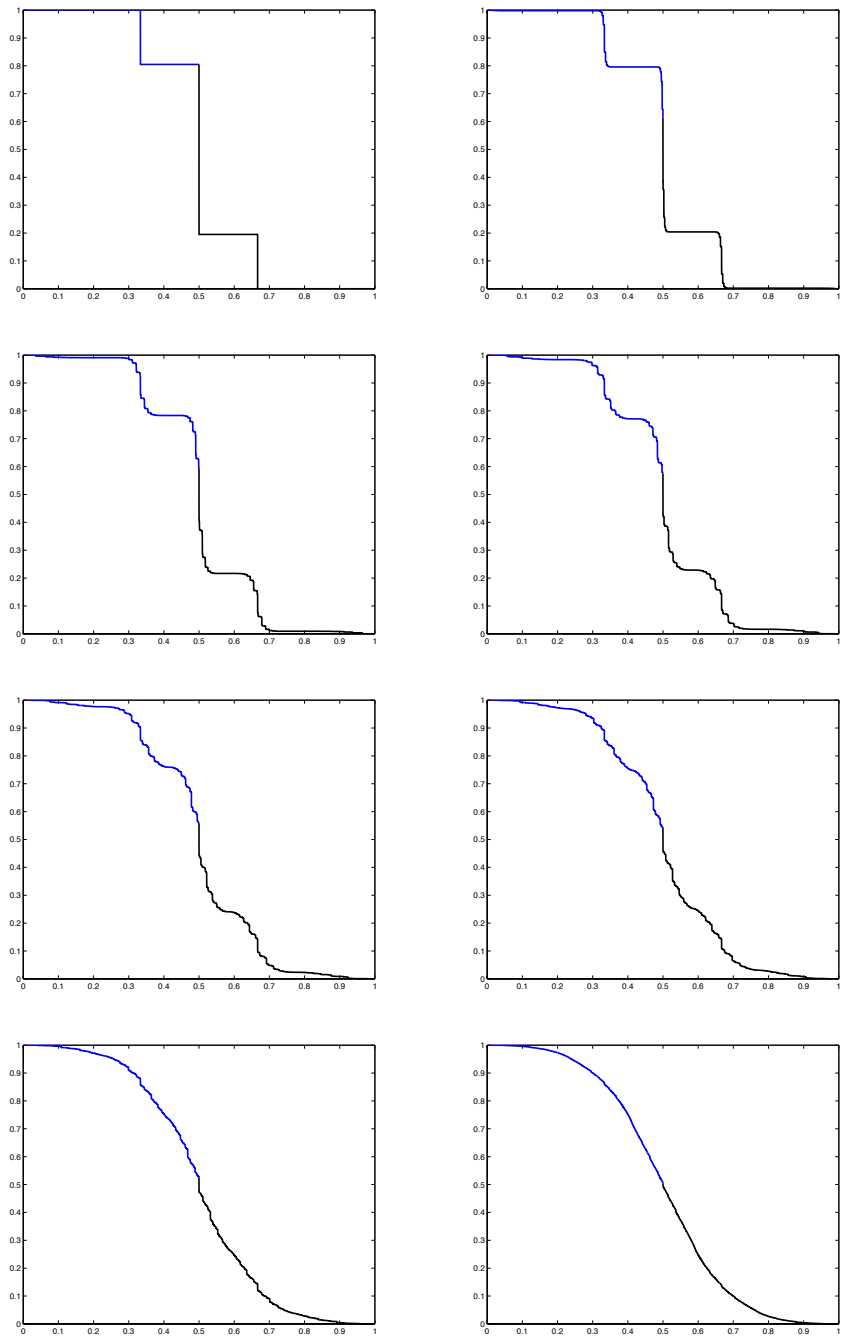

FIGURE 7. The graphs of $\theta \mapsto x(\theta)$ for values of p ranging from $p = 0.6625$ to $p = 0.7325$ in steps of 0.01.

mapping theorem, there exists $M > 0$ such that $|\Delta X_n(\frac{1}{2})| \leq M$ for all n. From (13), we observe that for $\theta \neq \frac{1}{2}$,

$$\Delta X_n(\theta) = A_{\epsilon(\theta)} \Delta X_{n-1}(\Phi(\theta)). \tag{14}$$

Taking 0 to be the zero function, we have $\|\mathcal{L}^n 0 - X^*\| \to 0$. Notice that X_n has discontinuities only at pre-images of $\frac{1}{2}$ of order at most n and is piecewise constant between discontinuities.

We now use ideas originating from Ruelle's thermodynamic formalism (see Walters' book [3] for an accessible introduction) – our presentation is, in fact, self-contained – it should be readable without any knowledge of thermodynamic formalism. We show that if a certain series is convergent (this corresponds to a thermodynamic pressure being negative), then X^* is of pure jump type. At the same time, we show that G is monotonically decreasing and H is monotonically increasing.

We use a technical result independent of our specific context:

Lemma 8. *Let (a_k) be an sequence in $[0,1]$ with $a_k \neq a_{k'}$ for all $k \neq k'$ and let b_k be a summable sequence of non-negative numbers. Suppose that (f_n) is a sequence of real-valued functions defined on $[0,1]$, each of pure jump type. Suppose further that the only discontinuities of (f_n) occur at the $a_k's$ and that $|\Delta f_n(a_k)| \leq b_k$ for each k and n. If $\|f_n - f\|_\infty \to 0$, then f is of pure jump type with discontinuities only at the a_k's. The magnitude of the discontinuity of f at a_k is bounded above by b_k.*

Proof. Denote Vf the total variation of f and $V_I f$ the variation of f on the interval I. Notice that the f_n have uniformly bounded variation and hence f has bounded variation also. Hence if it has a unique (up to additive constants) Lebesgue decomposition as a sum $f_d + f_c$ where f_c is continuous and f_d has only jump-type discontinuities. For any $\epsilon > 0$, there exists a K such that $\sum_{k \geq K} b_k < \epsilon$. Letting $I_1, \ldots, I_M$ be any disjoint collection of intervals avoiding the a_k's with $k < K$, we see that $\sum_{i=1}^M V_{I_i} f < \epsilon$. In particular, we deduce $Vf_c < \epsilon$ for arbitrary ϵ so that f_c has pure jump type. We also deduce that f_d cannot have any jump discontinuities other than at the a_k's and the result is proven. □

Suppose that $\theta > \frac{1}{2}$. Then we have $\Phi(\theta) = f_B(\theta) = (3p-1) - \gamma/\theta$ and we calculate

$$A_1 \begin{pmatrix} \Phi(\theta) - 1 \\ \Phi(\theta) \end{pmatrix} = \frac{\gamma}{\theta} \begin{pmatrix} \theta - 1 \\ \theta \end{pmatrix}.$$

Similarly, if $\theta < \frac{1}{2}$, we have

$$A_0 \begin{pmatrix} \Phi(\theta) - 1 \\ \Phi(\theta) \end{pmatrix} = \frac{\gamma}{1-\theta} \begin{pmatrix} \theta - 1 \\ \theta \end{pmatrix}.$$

If $t \in \Phi^{-n}(\frac{1}{2})$, let $\theta_i = \Phi^{n-i}(t)$ and let $\epsilon_i = \epsilon(\theta_i)$. The above shows

$$A_{\epsilon_i}\begin{pmatrix}\theta_{i-1}-1\\ \theta_{i-1}\end{pmatrix} = \frac{\gamma}{\max(\theta_i, 1-\theta_i)}\begin{pmatrix}\theta_i-1\\ \theta_i\end{pmatrix}.$$

Combining these equalities gives

$$A_{\epsilon_n}\cdots A_{\epsilon_1}\begin{pmatrix}-\frac{1}{2}\\ \frac{1}{2}\end{pmatrix} = \prod_{i=1}^{n}\frac{\gamma}{\max(\theta_i, 1-\theta_i)}\begin{pmatrix}\theta_n-1\\ \theta_n\end{pmatrix}. \tag{15}$$

Since $1-p \le \theta_n \le p$ for all possible choices of t and using (14), one obtains

$$\big|\Delta X_k(\theta_n)\big| \le C\prod_{i=1}^{n}\frac{\gamma}{\max(\theta_i, 1-\theta_i)}.$$

In order to apply Lemma 8, we therefore need to establish the inequality

$$\sum_n \sum_{\theta\in\Phi^{-n}(\frac{1}{2})}\prod_{i=0}^{n-1}\frac{\gamma}{\max(\Phi^i(\theta), 1-\Phi^i(\theta))} < \infty. \tag{16}$$

Defining Z_n to be $\sum_{\theta\in\Phi^{-n}(\frac{1}{2})}\prod_{i=0}^{n-1}(\gamma/\max(\Phi^i(\theta), 1-\Phi^i(\theta)))$, the pressure (of the function $f(\theta) = \log\big(\gamma/\max(\theta, 1-\theta)\big)$ with respect to the transformation Φ) is precisely $\limsup_{n\to\infty}(1/n)\log Z_n$. In particular, the negativity of the pressure is a sufficient condition (and is very close to a necessary condition) to ensure that X^* is of pure jump type. The jumps satisfy $\Delta X^*(\theta) = A_{\epsilon(\theta)}\Delta X^*(\Phi(\theta))$. By (15), they are all of the same sign.

Now provided the pressure is negative and p is not a pre-image of $\frac{1}{2}$, we check using (7) that $G(\frac{1}{2}^+) = \gamma Z - v$, $G(\frac{1}{2}^-) = 1 - 3\gamma Z - v$ so that $\Delta G(\frac{1}{2}) = 4\gamma Z - 1$. On the other hand, the total of all discontinuities (all of the same sign) is $-2Z$. In order for these to have the same sign, one sees that $Z > 0$ and $4\gamma Z < 1$. The function $G(\theta)$ is therefore a decreasing function.

Now to check (9), it suffices to show that $G(\frac{1}{2}^-) \ge \gamma Z - v$, $G(\frac{1}{2}^+) \le 1 - \gamma Z - v$ and $G(p) \ge -\gamma Z - v$. The first two of these follow from the fact that $Z > 0$ and $4\gamma Z < 1$. To verify the last inequality, we note that the above contraction argument works outside the range $[1-p, p]$, so that G is monotonic on all of $[0, 1]$. Since $\Phi(1) = p$, we apply (7) to see that $G(1) = -\gamma Z - v$, so that the third inequality is satisfied by monotonicity.

Assembling the facts above, we arrive at the following.

Corollary 9. *Suppose p is such that the condition* (16) *is satisfied. Then σ^* is an optimal strategy for Player 1.*

Reviewing the steps above, we see that we made use of the assumption that $\Phi^n(p) \neq \frac{1}{2}$ for all n. (If p is a preimage of $\frac{1}{2}$, then the expressions for $G(\frac{1}{2}^+)$ and $G(\frac{1}{2}^-)$ are not valid as G and H are discontinuous at p and $1-p$).

In the case where p *is* a preimage of $\frac{1}{2}$, the essential modification is to show that $G(\frac{1}{2}^+) - G(\frac{1}{2}) = G(\frac{1}{2}) - G(\frac{1}{2}^-)$. The matrix equalities (7) then ensure that at preimages, x say, of $\frac{1}{2}$, one has that $G(x)$ is the averages of $G(x^-)$ and $G(x^+)$ (and similarly for H). This allows us to deduce monotonicity and verify inequalities on entire intervals by checking at a finite collection of points as before.

5. Pressure bounds

In this section, we find ranges of p where (16) is satisfied.

Since the map Φ satisfies the symmetry $\Phi^n(1-t) = 1 - \Phi^n(t)$, we will study instead the map $\alpha(t) = \max(\Phi(t), 1-\Phi(t))$ as a map from $[\frac{1}{2}, 1]$ to itself. Notice that $\Phi(t) < \frac{1}{2}$ if and only if $t < \frac{2}{3}$. The map can therefore be expressed as:

$$\alpha_p(t) = \begin{cases} 2 - 3p + \gamma/t & \text{if } t < \frac{2}{3}; \\ 3p - 1 - \gamma/t & \text{if } t \geq \frac{2}{3}. \end{cases}$$

We denote the left branch by f and the right branch by g. As previously noted γ, α, Φ, f and g all depend on p. Let $\psi(t) = \gamma/t$ and write $\psi^{(n)}(t) = \psi(t)\psi(\alpha(t)) \cdots \psi(\alpha^{n-1}t)$. We then check that the summability condition (16) is equivalent to

$$\sum_n Z_n(p) < \infty; \quad \text{where} \tag{17}$$

$$Z_n(p) = \sum_{t \in \alpha^{-n}(\frac{1}{2})} \psi^{(n)}(t). \tag{18}$$

We do this by showing $\limsup \frac{1}{n} \log Z_n(p) < 0$ in various ranges of p (or in thermodynamic language, the function $\log \psi$ has negative pressure with respect to α in those ranges of p).

We partition $[\frac{1}{2}, p]$ into sub-intervals, counting possible transitions between pairs of intervals, and over-estimating ψ on the intervals to give a rigorous, finitely-calculable estimate for the pressure in various ranges of p.

In principle, this should work by taking finer and finer partitions up to $p_c \approx 0.7321$. We give a proof on a smaller range of p-values, subject to checking a manageable number of polynomial inequalities.

It turns out that for p in the range $[\frac{2}{3}, p_c]$, the map α is renormalizable. Letting $p_i = \alpha^i(p)$ (so that $p_0 = p$), one can check that $\frac{1}{2} < p_1 < p_2 < p$ and $\alpha([\frac{1}{2}, p_1]) = [p_2, p]$ and $\alpha([p_2, p]) = [\frac{1}{2}, p_1]$.

The map $t \mapsto \alpha(t)$ is strictly decreasing on $[\frac{1}{2}, \frac{2}{3}]$ and strictly increasing on $[\frac{2}{3}, p]$, and satisfies $\alpha(\frac{2}{3}) = \frac{1}{2}$. To demonstrate the above renormalization, one needs to verify the inequalities

$$p_1 \leq \tfrac{2}{3} \tag{19}$$

$$p_3 \leq p_1. \tag{20}$$

Suppose that $[\frac{1}{2}, 1]$ is partitioned into intervals $J_0, \ldots, J_{k-1}$. Let $\beta_i = \max_{x \in J_i} \psi(x)$. Let $m_{ij} = \max_{y \in J_j} \#\{x \in J_i \colon \alpha(x) = y\}$. Let A be the $k \times k$ matrix with entries $a_{ij} = \beta_i m_{ij}$. Then we claim that

$$\limsup_{n \to \infty} \tfrac{1}{n} \log Z_n(p) \leq \log \rho(A),$$

where $\rho(A)$ denotes the spectral radius. Hence in order to establish that $Z(p) < \infty$, it suffices to exhibit a finite partition such that the corresponding matrix A has spectral radius less than 1.

To establish the claim, notice that there are at most $m_{i_0 i_1} m_{i_1 i_2} \ldots m_{i_{n-1} i_n}$ nth order preimages x of a point y in J_{i_n} with the property that $\alpha^t(x) \in J_{i_t}$ for each $0 \leq t < n$. For each such preimage, the largest possible contribution to the sum is $\beta_{i_0} \ldots \beta_{i_{n-1}}$, so that we see

$$Z_n(p) = \sum_{t \in \alpha_p^{-n}(\frac{1}{2})} \psi_p^{(n)}(t) \leq \sum_i (A^n)_{ij},$$

where j is the index of the interval containing $\frac{1}{2}$. This establishes the claim.

5.1. **The range (2/3,0.70237758).** In this range, we are able to give a particularly simple bound. We divide $[\frac{1}{2}, p]$ into three sub-intervals: $J_0 = [\frac{1}{2}, p_1]$, $J_1 = [p_1, p_2]$ and $J_2 = [p_2, p]$.

By the above, we have $\alpha_p(J_0) \subset J_2$ and $\alpha_p(J_2) \subset J_0$. One checks that $\frac{2}{3} \in J_2$.

Given this, we see that $\alpha(J_1) \subset J_0 \cup J_1$. Since α_p is unimodal, and the critical point lies outside J_1, we see the restriction of α_p to J_1 is monotonic.

Recall that $\psi(t) = \gamma/t$ and notice that ψ is bounded above by $2\gamma = 4p - 2$ which is strictly less than 1 for $p \leq p_c \leq \frac{3}{4}$.

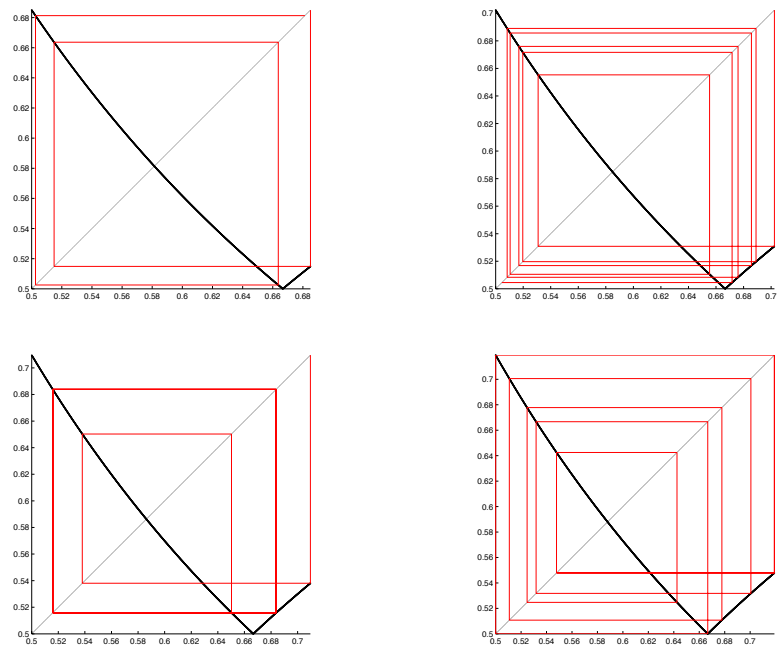

FIGURE 8. The graphs of $\theta \mapsto \alpha(\theta)$ and first points of the orbit of p, for $p = 0.685, p = 0.7023..., p = 0.709...$ and $p = 0.719....$ The renormalizablity of $\theta \mapsto \alpha(\theta)$ may be seen from the fact that in each of the graphs points to the right of the fixed point are mapped to the left of the fixed point and vice versa.

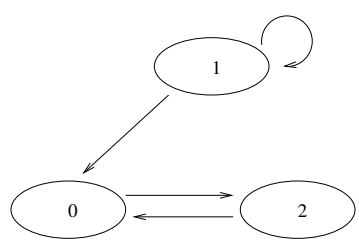

FIGURE 9. Transitions with 3 intervals in the range $\frac{2}{3} < p < 0.70237758$.

Since α is monotonic on J_0 and J_1, we have $m_{02} = m_{10} = m_{11} = 1$, so that A_{11} and A_{10} are both given by γ/p_1, while $A_{02} = \gamma/\frac{1}{2} = 2\gamma$. Finally, $m_{20} = 2$ and $A_{20} = 2\gamma/p_2$.

It's well known that the spectral radius of a matrix is the maximum of the spectral radius of its communicating components. The component containing 1 has spectral radius $A_{11} = \gamma/p_1$, while the component

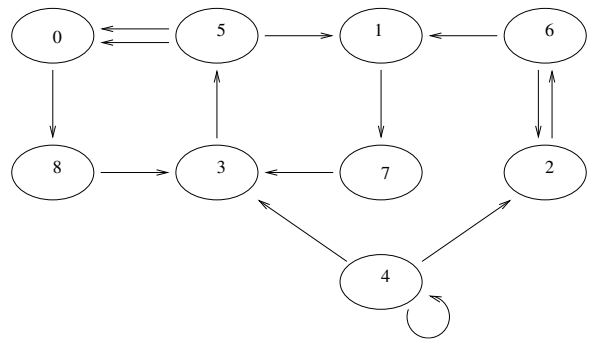

FIGURE 10. Full 9 interval transition diagram for $\frac{2}{3} < p < 0.709637$. The double arrow signifies that $m_{50} = 2$

.

containing 0 and 2 has spectral radius $(A_{02}A_{20})^{1/2} = 2\gamma/\sqrt{p_2}$. We have $A_{11} < 1$ for all p, and $2\gamma/\sqrt{p_2} < 1$ in the range $\frac{2}{3} < p < 0.70237758$. (At the endpoint of the interval, $2\gamma/\sqrt{p}_2$ becomes 1).

5.2. **The range (2/3, 0.709636).** Here, and in the next range, we divide $[\frac{1}{2}, p]$ into 9 sub-intervals. In this range, we check that the following inequalities are satisfied:

$$\tfrac{1}{2} < p_7 < p_3 < p_5 < p_9 < p_1 < p_2 < \tfrac{2}{3} < p_6 < p_4 < p_8 < p.$$

We divide the interval $[\frac{1}{2}, p]$ into subintervals $J_0, \ldots, J_8$ as follows: $J_0 = [\frac{1}{2}, p_7]$; $J_1 = [p_7, p_3]$; $J_2 = [p_3, p_5]$; $J_3 = [p_5, p_1]$; $J_4 = [p_1, p_2]$; $J_5 = [p_2, p_6]$; $J_6 = [p_6, p_4]$; $J_7 = [p_4, p_8]$; and $J_8 = [p_8, p]$.

The transitions between the intervals are shown in Figure 10.

There are three connected components, one (the interval J_4 by itself) with radius γ/p_1, one (the intervals J_2 and J_6) with radius $\gamma/\sqrt{p_3 p_6}$. Both of these are less than 1 by the above argument. The third component is illustrated in Figure 11 and consists of two loops of period 4 sharing a common pair of vertices. The spectral radius of this component is the fourth root of the sum of the product of the multipliers around the two loops. That is, the spectral radius of this component is given by

$$\gamma \left(\frac{1}{p_5 p_2} \left(\frac{2}{\frac{1}{2} p_8} + \frac{1}{p_7 p_4} \right) \right)^{1/4}.$$

This quantity is less than 1 in the given range.

Notice that the principal component has period 4 because the original map is twice renormalizable.

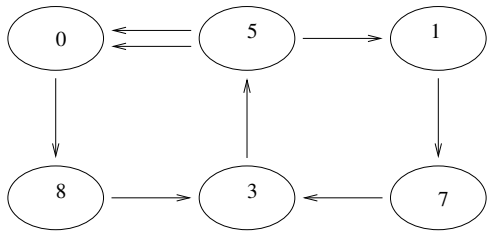

FIGURE 11. Principal component for $\frac{2}{3} < p < 0.709637$

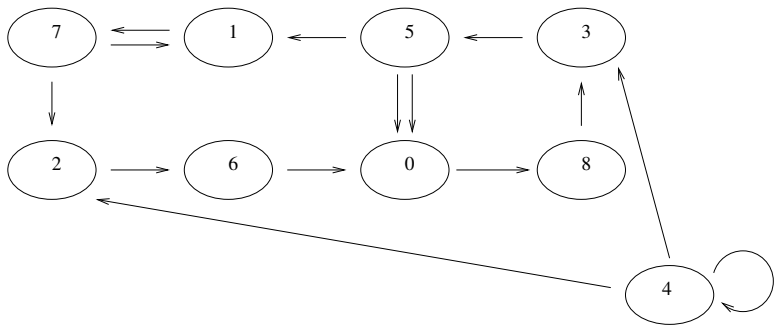

FIGURE 12. The transitions in the range $0.709637 < p < 0.719023$.

5.3. **The range [0.709637,0.719023].** In this parameter range, the map is only once renormalizable. At 0.709636979, there is a coincidence $p_3 = p_5$ (so that all odd iterates beyond the third coincide; all even iterates beyond the fourth coincide).

The right end point of the interval, 0.7190233023, occurs when p_9 hits $\frac{1}{2}$. On the parameter interval $[0.709636979, 0.7190233023]$, the functions $p \mapsto p_i$ are monotone for each $1 \leq i \leq 9$. The graphs of the functions do not cross.

In this range, we have $\frac{1}{2} < p_9 < p_5 < p_3 < p_7 < p_1 < p_2 < \frac{2}{3} < p_8 < p_4 < p_6 < p$.

Again, we use these points (excluding p_9 and $\frac{2}{3}$) to define a collection of intervals: $J_0 = [\frac{1}{2}, p_5]$, $J_1 = [p_5, p_3]$, $J_2 = [p_3, p_7]$, $J_3 = [p_7, p_1]$, $J_4 = [p_1, p_2]$, $J_5 = [p_2, p_8]$, $J_6 = [p_8, p_4]$, $J_7 = [p_4, p_6]$ and $J_8 = [p_6, p]$. The transitions are $0 \to 8$; $1 \to 7$; $2 \to 6$; $3 \to 5$; $4 \to 2, 3, 4$; $5 \to 0, 0, 1$; $6 \to 0$; $7 \to 1, 2$; and $8 \to 3$ (where repeated transitions correspond to values of m that exceed 1).

This is illustrated in Figure 12.

The single component consisting of J_4 always has multiplier less than 1. The transition matrix of the principal component is given by

$$\gamma \begin{pmatrix} 0 & 0 & 0 & 0 & 0 & 0 & 0 & 2 \\ 0 & 0 & 0 & 0 & 0 & 0 & q_5 & 0 \\ 0 & 0 & 0 & 0 & 0 & q_3 & 0 & 0 \\ 0 & 0 & 0 & 0 & q_7 & 0 & 0 & 0 \\ 2q_2 & q_2 & 0 & 0 & 0 & 0 & 0 & 0 \\ q_8 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & q_4 & q_4 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & q_6 & 0 & 0 & 0 & 0 \end{pmatrix}$$

where $q_i = 1/p_i$.

We check that q_4, q_5 and q_8 are increasing in the parameter range, while q_3, q_7 and q_6 are decreasing. Substituting the maximum values of each of these quantities in the range and also using the maximal value of γ, we obtain a matrix whose spectral radius is 0.9773, giving the required bound on the pressure in this range.

In principle it should be possible to extend by smaller and smaller intervals as long as the pressure remains negative. For example, the matrix based on using 230 iterates shows that the pressure is negative for $p = 0.7321$.

Probably a computer lower bound would show that the pressure is positive for some value of p not much larger than this.

At this stage, we have proved that

Lemma 10. *Strategy σ^* for Player 1 and Strategy $(x(\theta))_{\theta\in[p,1-p]}$ for Player 2 are optimal strategies if $\frac{2}{3} \leq p \leq 0.719023$.*

6. Better than σ^* after the critical point

Beyond the critical point, we suspect that the strategy σ^* is often not optimal, especially when the orbit of $1-p$ comes close to $\frac{1}{2}$. Indeed, we propose strategies — far from optimal — which make better than σ^* for specific values of p ; we prove this claim completely for $\frac{3}{4}$ (which was an explicit open question); we also show the computation for the value $p = 0.73275300915$..

Let p be large enough so that we can expect σ^* not to be optimal. We choose k_0 so that $\tilde{\theta} = \Phi^{k_0}(1-p) < \frac{1}{2}$ is close to $\frac{1}{2}$. We also let $\epsilon > 0$ be a small real number.

We modify slightly σ^* to a strategy $\sigma_{k_0,\epsilon}$ in the following way: if $\theta \neq \tilde{\theta}, 1-\tilde{\theta}$, then Player 1 plays following σ^*. But if $\theta = \tilde{\theta}$ (recall that $\tilde{\theta} < \frac{1}{2}$), then Player 1 "perturbs" his reaction by ϵ: he plays T with probability:

- $1-\epsilon$ if $s = \underline{s}$,

- $1-(1-\epsilon)\frac{\tilde{\theta}}{1-\tilde{\theta}} = \frac{1-(2-\epsilon)\tilde{\theta}}{1-\tilde{\theta}}$ if $s=\overline{s}$,

and, if $\theta = 1-\tilde{\theta}$, he plays T with probability

- $(1-\epsilon)\frac{\tilde{\theta}}{1-\tilde{\theta}}$ if $s=\underline{s}$,
- ϵ if $s=\overline{s}$;

in the case $\theta=\tilde{\theta}$, the belief is updated as:

- if Player 1 plays T, it becomes : $\tilde{\theta}_\epsilon := p\frac{(1-\epsilon)\tilde{\theta}}{1-\tilde{\theta}} + (1-p)\frac{1-(2-\epsilon)\tilde{\theta}}{1-\tilde{\theta}}$,
- if Player 1 plays B, it becomes : $1-p_\epsilon := (1-p)(1-\epsilon)+p\epsilon$,

and in the case $\theta = 1-\tilde{\theta}$,

- if Player 1 plays T, it becomes : $p_\epsilon = p(1-\epsilon)+(1-p)\epsilon$,
- if Player 1 plays B, it becomes : $p\frac{(2-\epsilon)(1-\tilde{\theta})-1}{1-\tilde{\theta}}+(1-p)(1-\epsilon)\frac{1-\tilde{\theta}'}{\tilde{\theta}'}$.

The critical aspect in this choice of perturbation of the strategy is that it enjoys the same property as σ^*, namely that the expected gain of Player 1 is unaffected by Player 2's choice of moves. It is this that will allow us to compute the expected long-term gain of the new strategy.

We shall compare the value of this strategy $\sigma_{k_0,\epsilon}$ with the value of σ^*. We are going to prove that

Lemma 11. *$v_p(\sigma_{k_0,\epsilon}) > v_p(\sigma^*)$ if and only if*

$$(1-\tilde{\theta})\left(w(\Phi(\tilde{\theta}))-w(\tilde{\theta}_\epsilon)\right) > \tilde{\theta}\left(w(p_\epsilon)-(1-\epsilon)w(p)\right), \tag{21}$$

where w stands for $w(\theta)=\sum_{k=0}^{\infty}\max\left(\Phi^k(\theta),1-\Phi^k(\theta)\right)$.

It is straightforward to compute the expected (one step) payoff (given θ) with this new strategy. When $\theta\neq\tilde{\theta}$ and $\theta\neq 1-\tilde{\theta}$, the computation is the same as for σ^*, i.e. $\min(\theta,1-\theta)$. Otherwise a similar computation yields an expected payoff: $(1-\epsilon)\tilde{\theta}$.

Observe that with the strategy $\sigma_{k_0,\epsilon}$, when $\theta=\tilde{\theta}$, the one-step expected payoff is a bit smaller than with the strategy σ^*. However, the update of the belief is slightly different and one may hope that this new belief puts Player 1 in a better position for the future: in a sufficiently improved position to compensate for the loss in the one-step expected payoff. The objective is to show that this is possible for some values of p.

For this purpose, we have to find an expression for the long term expected payoff. Whatever Player 2 plays, he sees a Markov chain on the beliefs (governed by the random changes of the state and the values of Player 1's choices). The belief may take the values p and $1-p$ and values in the k_0 first terms of the orbits of p and $1-p$; when it reaches $\tilde{\theta}$, it may jump to the values of the belief after $\tilde{\theta}$; namely $\tilde{\theta}_\epsilon$ or $1-p_\epsilon$

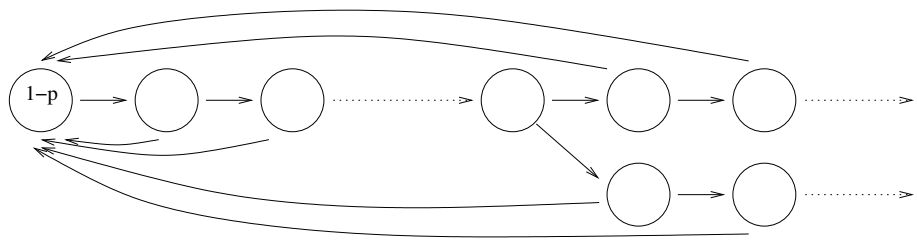

FIGURE 13. Schematic depiction of the (symmetrized) Markov chain. At each state other than $\tilde{\theta}$, one choice leads back to the base, and the other goes to the right.

and then continue on their orbits for some random time and then go back to $1-p$ or p. It is convenient to further assume that neither $\tilde{\theta}$ nor $1-\tilde{\theta}$ belong to the orbits of p_ϵ and $\tilde{\theta}_\epsilon$. We observe that the symmetry $\theta \mapsto 1-\theta$ does not affect either the transitions or the payoff so it suffices to follow the orbits modulo the symmetry about $\frac{1}{2}$.

6.1. **Invariant measure for the Markov Chain.** Let us set $\Theta_k(\theta) = \max(\Phi^k(\theta), 1-\Phi^k(\theta))$, $\Theta_k = \Theta_k(1-p)$, $\Theta_k^1 = \Theta_k(\theta_0^\epsilon)$ and $\Theta_k^2 = \Theta_k(1-p_\epsilon)$. That is,

$$\begin{aligned}
\Theta_k &= \max(\Phi^k(1-p), 1-\Phi^k(1-p)) \\
\Theta_k^1 &= \max(\Phi^k(\tilde{\theta}_\epsilon), 1-\Phi^k(\tilde{\theta}_\epsilon)) \\
\Theta_k^2 &= \max(\Phi^k(1-p_\epsilon), 1-\Phi^k(1-p_\epsilon))
\end{aligned}$$

We consider a Markov chain on the countable state space $\{\Theta_k, 0 \le k \le k_0; \Theta_k^1, \Theta_k^2, k \ge 0\}$ with transition probabilities :

- If $k < k_0$, $\Theta_k \to \Theta_{k+1}$ with probability Θ_k and $\Theta_k \to \Theta_0$ with probability $1-\Theta_k$.
- If $k = k_0$ $\Theta_{k_0} \to \Theta_0^1$ with probability $1-\tilde{\theta}$ and $\Theta_{k_0} \to \Theta_0^2$ with probability $\tilde{\theta}$.
- For all $k \ge 0$, for $i = 1, 2$, $\Theta_k^i \to \Theta_{k+1}^i$ with probability Θ_k^i and $\Theta_k^i \to \Theta_0$ with probability $1-\Theta_k^i$.

It is straightforward to compute the invariant measure for this chain. We denote Π_n^i the probability to be at Θ_n^i (with $\Theta^0 = \Theta$). Invariance implies it must satisfiy:
For $1 \le n \le k_0$

$$\Pi_n = \Pi_0 \prod_{k=0}^{n-1} \Theta_k.$$

For all $n \geq 0$

$$\Pi_n^1 = \Pi_{k_0}(1-\tilde{\theta}) \prod_{k=0}^{n-1} \Theta_k^1,$$

and

$$\Pi_n^2 = \Pi_{k_0}\tilde{\theta} \prod_{k=0}^{n-1} \Theta_k^2.$$

The balance in Θ_0^0 implies

$$\Pi_0 = \sum_{n=0}^{k_0-1}(1-\Theta_n)\Pi_n + \sum_{n=0}^{\infty}(1-\Theta_n^1)\Pi_n^1 + \sum_{n=0}^{\infty}(1-\Theta_n^2)\Pi_n^2. \tag{22}$$

Since it is a probability measure, it also must satisfy :

$$\sum_{n=0}^{k_0-1}\Pi_n + \Pi_{k_0} + \sum_{n=0}^{\infty}\Pi_n^1 + \sum_{n=0}^{\infty}\Pi_n^2 = 1. \tag{23}$$

Equality (22) expresses the fact that the return time to the 'base' $(1-p)$ is almost surely finite. We introduce notation $Q = \prod_{k=0}^{k_0-1}\Theta_k = \Pi_{k_0}/\Pi_0$ and $w(\theta) = \sum_{n=0}^{\infty}\prod_{k=0}^{n-1}\Theta_k(\theta)$. This latter quantity gives the ratio of all weights in the sub-tree rooted at θ to the weight of θ. Using this notation, we can write equality (23) as

$$\Pi_0\left(\sum_{n=0}^{k_0}\prod_{k=0}^{n-1}\Theta_k + Q(1-\tilde{\theta})w(\Theta_0^1) + Q\tilde{\theta}w(\Theta_0^2)\right) = 1.$$

Hence

$$\Pi_0 = \left[\sum_{n=0}^{k_0}\prod_{k=0}^{n-1}\Theta_k + Q((1-\tilde{\theta})w(\Theta_0^1) + \tilde{\theta}w(\Theta_0^2))\right]^{-1}.$$

6.2. **Expected payoff.** The expected payoff can be written as the sum of the expected payoff (given the state) weighted by the probability of the state ; namely,

$$v_p(\sigma_{k_0,\epsilon}) = \sum_{n=0}^{k_0-1}(1-\Theta_n)\Pi_n + (1-\epsilon)\tilde{\theta}\Pi_{k_0} + \sum_{n=0}^{\infty}(1-\Theta_n^1)\Pi_n^1 + \sum_{n=0}^{\infty}(1-\Theta_n^2)\Pi_n^2.$$

Using (22), we obtain

$$v_p(\sigma_{k_0,\epsilon}) = \Pi_0 + (1-\epsilon)\tilde{\theta}\Pi_{k_0} = \Pi_0(1+(1-\epsilon)\tilde{\theta}Q).$$

We want to show that for well chosen p, k_0 and ϵ, $v_p(\sigma_{k_0,\epsilon}) > v_p(\sigma^*)$. We recall that $v(\sigma^*) = 1/w(p)$. Hence the desired inequality can be written in the following sequence of equivalent forms:

$$\Pi_0\left(1+(1-\epsilon)Q\tilde{\theta}\right) > 1/w(p);$$

$$\left(1+(1-\epsilon)\tilde{\theta}Q\right)w(p) > \sum_{n=0}^{k_0}\prod_{k=0}^{n-1}\Theta_k + Q\left((1-\tilde{\theta})w(\Theta_0^1)+\tilde{\theta}w(\Theta_0^2)\right);$$

$$\left(1+(1-\epsilon)\tilde{\theta}Q\right)w(p) > w(p)-Q(1-\tilde{\theta})w(\Theta_{k+1})+Q\left((1-\tilde{\theta})w(\Theta_0^1)+\tilde{\theta}w(\Theta_0^2)\right).$$

Cancelling the $w(p)$'s and dividing through by the common factor Q, we obtain as a conclusion, that $v_p(\sigma_{k_0,\epsilon}) > v_p(\sigma^*)$ if and only if

$$(1-\tilde{\theta})\left(w(\Phi(\tilde{\theta}))-w(\tilde{\theta}_\epsilon)\right)-\tilde{\theta}\left(w(1-p_\epsilon)-(1-\epsilon)w(1-p)\right) > 0. \tag{24}$$

This completes the proof of Lemma 11.

6.3. **The case p=3/4.** When p takes the value $\frac{3}{4}$, the symbolic dynamic of $(1-p)$ starts with 00101010 and $\Phi^7(1-p) = \Phi^7(1/4) = 1085/2244 \simeq .4835....$ We shall set $k_0 = 7$ and $\tilde{\theta} = 1085/2244$.

Next we estimate $w(\theta)$ for the relevant values of θ. First we do it for $1-p$ and for $\Phi(\tilde{\theta})$. Recall that $w(\theta) = \sum_{n=0}^{\infty}\prod_{k=0}^{n-1}\Theta_k(\theta)$. The general term is positive. As soon as $k \geq 1$, $\frac{1}{2} \leq \Theta_k \leq p$. Hence, the remainder of the sequence is bounded by

$$\begin{aligned}\sum_{n\geq N}\prod_{k=0}^{n-1}\Theta_k(\theta) &\leq \left(\prod_{k=0}^{N-1}\Theta_k(\theta)\right)\sum_{n\geq 0}p^n \\ &\leq \frac{\left(\prod_{k=0}^{N-1}\Theta_k(\theta)\right)}{1-p} \leq \frac{p^N}{1-p} \leq 4\left(\frac{3}{4}\right)^N.\end{aligned}$$

We do the computation with $N = 50$, so the bound on the error is smaller than 10^{-10} (and the obvious bound p^{-N} is itself of order 10^{-7}). We obtain with this approximation $w(1-p) \simeq 2.8354$ and $w(\Phi(\tilde{\theta})) \simeq 2.7432$.

Then numerical experimentation (see Figure 14) suggests taking $\epsilon = 0.01$. For this value of ϵ, we also compute $w(\tilde{\theta}_\epsilon) \simeq 2.7305$ and $w(1-p_\epsilon) \simeq 2.8203$. This is sharp enough to see the difference between

$$\tilde{\theta}\big(w(1-p_\epsilon)-(1-\epsilon)w(1-p)\big) \simeq 0.0064$$

and

$$(1-\tilde{\theta})\big(w(\Phi(\tilde{\theta})-w(\tilde{\theta}_\epsilon))\big) \simeq 0.0065.$$

We conclude that

$$(1-\tilde{\theta})\big(w(\Phi(\tilde{\theta}))-w(\tilde{\theta}_\epsilon)\big)-\tilde{\theta}\big(w(1-p_\epsilon)-(1-\epsilon)w(1-p)\big) > 10^{-5},$$

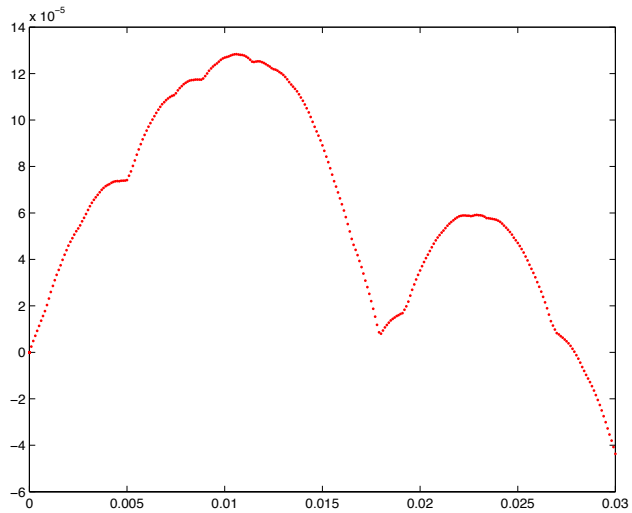

FIGURE 14. Numerical approximation of the graph of the left side of (21) (vertical axis) as a function of ϵ (horizontal axis)

so that, according to Lemma 11, we have shown that σ^* is not optimal for $p = 3/4$. The expected payoff of the alternative strategy can be computed: we obtain $v_{\frac{3}{4}}(\sigma^*) = 0.35267910...$ and $v_{\frac{3}{4}}(\sigma_{7,0.01}) = 0.35267964...$, showing a difference between the values of

$$v_{\frac{3}{4}}(\sigma_{7,0.01}) - v_{\frac{3}{4}}(\sigma^*) \simeq 5 \times 10^{-7}.$$

6.4. **The case p=0.73275300915.** The same computations with $k_0 = 57$ (so that $\theta_0 \simeq 0.49999805...$) yield : $w(1-p) \simeq 2.76648483$ and $w(\Phi(\tilde{\theta})) \simeq 2.766474044$. Taking for instance $\epsilon = 0.00015$, we obtain $w(\tilde{\theta}_\epsilon) \simeq 2.766266543$ and $w(1-p_\epsilon) \simeq 2.766277325$. Hence,

$$\tilde{\theta}\big(w(1-p_\epsilon) - (1-\epsilon)w(1-p)\big) \simeq 1.0373 \times 10^{-4}$$

and

$$(1-\tilde{\theta})\big(w(\Phi(\tilde{\theta}) - w(\tilde{\theta}_\epsilon))\big) \simeq 1.0375 \times 10^{-4}.$$

Finally, we compute

$$(1-\tilde{\theta})\big(w(\Phi(\tilde{\theta})) - w(\tilde{\theta}_\epsilon)\big) - \tilde{\theta}\big(w(1-p_\epsilon) - (1-\epsilon)w(1-p)\big) \simeq 1.72 \times 10^{-8}.$$

We conclude that, for $p = 0.73275300915$, $v_p(\sigma_{57,0.00015}) > v_p(\sigma^*)$. A direct computation indeed shows that $v_p(\sigma^*) \simeq 0.361469540454542$ while $v_p(\sigma_{57,0.00015}) \simeq 0.361469540454504$, i.e. a difference larger than 3×10^{-14}. This concludes the proof of Theorem 1.

References

[1] J. Hörner, D. Rosenberg, E. Solan, and N. Vieille. On a Markov game with one-sided incomplete information. *Oper. Res.*, 58:1107–1115, 2010.

[2] J. Renault. The value of Markov chain games with lack of information on one side. *Math. Oper. Res.*, 31:490–512, 2006.

[3] P. Walters. *An introduction to ergodic theory*. Springer, 1982.