ABSORBING GAMES WITH A CLOCK AND TWO BITS OF MEMORY

KRISTOFFER ARNSFELT HANSEN, RASMUS IBSEN-JENSEN,
AND ABRAHAM NEYMAN

Abstract. An absorbing game is a two-person zero-sum repeated game. Some of the entries are "absorbing" in the sense that, following the play of an absorbing entry, with positive probability all future payoffs are equal to that entry's payoff. The outcome of the game is the long-run average payoff. We prove that a two-person zero-sum absorbing game, with either finite or compact action sets, has, for each $\varepsilon > 0$, ε-optimal strategies with finite memory. In fact, we show that there is an ε-optimal strategy that depends on the clock and three states of memory.

Keywords: Absorbing Games; Finite memory; Markov Strategies; Stochastic Games; Compact action sets

Declarations of interest: None

Financial support and the role of the sponsor: No financial support (besides affiliation) and they had no involvement in this paper.

Date: April 5, 2020.

1. Introduction

We consider a game in which the players repeatedly play a strategic (normal-form) game. Each action profile determines the payoff, the probability of absorption, and the absorbing payoff. Following absorption, all future payoffs are equal to the absorption payoff. Such games are called *absorbing games*.

Early studies of absorbing games [4, 3] focused on two-person zero-sum absorbing games with finitely many actions, deterministic transitions, i.e., where the absorption probabilities are either 0 or 1, and where the payoff of each action pair coincides with the absorbing payoff of that action pair. Such a game is represented by the payoff matrix where the absorbing entries are marked with a $*$. The classic example of an absorbing game, the Big Match, introduced by Gillette [4], is represented by the matrix

$$\begin{bmatrix} 1 & -1 \\ -1^* & 1^* \end{bmatrix}.$$

An absorbing game (with probabilistic transitions) with finitely many actions (or compact action sets) is represented by a matrix (or a strategic game) whose entries (or the outcome of players' actions) are triples (r, q, a) of a payoff r, a probability $0 \leq q \leq 1$, and a payoff a. Following the play of an entry (r, q, a), the current payoff is r, with probability q all future payoffs are a, and with probability $1 - q$ the players continue playing the same absorbing game.

The game model with compact action sets assumes that the payoff r, the absorption probability q, and the absorbing payoff a depend continuously on the actions.

We are mainly interested in such games where the outcome of the game is the long-run average payoff. By “long-run” we refer to the average payoff (per stage) of either the infinitely repeated game or a finitely repeated game with a long but unknown finite duration.

In the two-person zero-sum case where the sets of actions are finite, Kohlberg [7] proved that (1) the limit of the values of the λ-discounted game as λ goes to 0 exists, (2) the limit of the n-stage game as n goes to ∞ exists, (3) that both limits coincide, and (4) that this limit v is the value (of the undiscounted game) in the following strong sense: for every $\varepsilon > 0$, each player has an ε-optimal strategy, i.e., a strategy that guarantees him a payoff of v up to an error of ε in all the n-stage games with n sufficiently large, in all λ-discounted games with λ sufficiently small, and in the limiting-average game.[1]

[1] In the limiting-average game the payoff to player 1 (respectively player 2) is the expectation of the liminf (respectively limsup) of the average payoff in the first n stages as n goes to infinity.

These results of Kohlberg [7] were extended to the more general class of two-person zero-sum stochastic games with finitely many states and actions [2, 9] and to absorbing games with compact action sets [10].

The ε-optimal strategies constructed in [3, 7, 9, 10] depend in each stage on a statistic of past history, and this statistic (e.g., the sum of the past payoffs) takes value in an infinite set; hence, it requires an infinite memory.

The question that arises is how much dependence on past history is needed for an ε-optimal strategy. This dependence is formalized using the following concept.

A *memory-based* strategy in an absorbing game[2] is a strategy where the conditional probability of the action in a given round depends on the current memory state and the clock (i.e, the stage number) and the memory state is updated as a stochastic function of the current memory, the actions of the players in the previous stage, and the clock.

The ε-optimal strategies in [3, Theorem 2] are memory-based, and those in [3, Theorem 1], [7, Theorems 2.1 and 3.4], [9], and [10] are memory-based and clock-independent; i.e., the action in a given round and the memory update do not depend on the clock.

The value of the Big Match is zero. But no memory-based strategy of Player 1 that has a finite set of memory states and that is either clock-independent [1] or has a deterministic memory update function [5] can guarantee Player 1 strictly more than -1 in the Big Match. Recently it has been shown that for every $\varepsilon > 0$ there is a memory-based strategy for Player 1 with two memory states that is ε-optimal in the Big Match [6]. Obviously, given the previously mentioned impossibility results, the [6] ε-optimal strategy relies on stochastic and clock-dependent memory updating.

The question that arises is whether or not a finite memory suffices for an ε-optimal strategy in any absorbing game.

The present paper proves that an absorbing game, with either finite or compact action sets, has, for each $\varepsilon > 0$, ε-optimal strategies with finite memory. In fact we show that there is an ε-optimal strategy that depends on the clock and three states of memory.

The importance of absorbing games stems from the fact that advances in the theory of absorbing games, which are a subclass of stochastic games, serve as an important building block in the study of stochastic games as well as other models of dynamic games, and from their intimate relation to repeated games with symmetric incomplete information. In fact, there is a natural translation of the value (and equilibria) and optimal (and equilibria) strategies between the two. See, e.g., [8, 11, 12].

[2]The general concept of a memory-based strategy in the more general model of a stochastic game is given in the next section.

2. The model and related results

Absorbing games are an important subclass of stochastic games, and their study served as an important step in the analysis of stochastic games.

2.1. Stochastic games.

2.1. Stochastic games. A *finite two-person zero-sum stochastic game* Γ, henceforth, a *finite stochastic game*, is defined by a tuple (Z, I, J, r, p, z_1), where Z is a finite state space, I and J are the finite actions sets of Players 1 and 2 respectively, $r : Z \times I \times J \to \mathbb{R}$ is a payoff function, $p : Z \times I \times J \to \Delta(Z)$ is a transition function, and z_1 is the initial state.

A *compact stochastic game* is defined analogously, like the definition of a stochastic game, but the actions sets I and J are compact topological spaces, and the payoff function and the transitions depends continuously on the actions.

A state $z \in Z$ is called an *absorbing state* if $p(z, \cdot, \cdot) = \delta_z$, where δ_z is the Dirac measure on z. An *absorbing game* is a stochastic game with only one non-absorbing state, which is its initial state.

A *play* of the stochastic game is an infinite sequence $z_1, \ldots, z_t, i_t, j_t, \ldots$, where $(z_t, i_t, j_t) \in Z \times I \times J$. A play up to stage t is the finite sequence $h_t = (z_1, i_1, j_1, \ldots, z_t)$. The payoff r_t in stage t is $r(z_t, i_t, j_t)$ and the average of the payoffs in the first n stages, $\frac{1}{n}\sum_{t=1}^{n} r_t$, is denoted by $\bar{r}_n$.

The initial state of the multi-stage game is $z_1 \in Z$. In the t-th stage players simultaneously choose actions $i_t \in I$ and $j_t \in J$.

A behavioral strategy of Player 1, respectively Player 2, is a function σ, respectively τ, from the disjoint union $\dot\cup_{t=1}^{\infty}(Z \times I \times J)^{t-1} \times Z$ to $\Delta(I)$, respectively to $\Delta(J)$. The restriction of σ, respectively τ, to $(Z \times I \times J)^{t-1} \times Z$ is denoted by σ_t, respectively τ_t. In what follows, σ denotes a strategy of Player 1 and τ denotes a strategy of Player 2.

A strategy pair (σ, τ) defines a probability distribution $P_{\sigma,\tau}$ on the space of plays as follows. The conditional probability of $(i_t = i, j_t = j)$ given a play h_t up to stage t is the product of $\sigma(h_t)[i]$ and $\tau(h_t)[j]$. The conditional distribution of z_{t+1} given h_t, i_t, j_t is $p(z_t, i_t, j_t)$. The expectation w.r.t. $P_{\sigma,\tau}$ is denoted by $E_{\sigma,\tau}$

A stochastic game has a value $v = (v(z))_{z \in Z}$ if, for every $\varepsilon > 0$, there are strategies σ_ε and τ_ε such that for some positive integer n_ε

$$\varepsilon + E_{\sigma_\varepsilon,\tau}\bar{r}_n \geq v(z_1) \geq E_{\sigma,\tau_\varepsilon}\bar{r}_n - \varepsilon \quad \forall \sigma, \tau, n \geq n_\varepsilon, \tag{1}$$

and

$$\varepsilon + E_{\sigma_\varepsilon,\tau} \liminf_{n\to\infty} \bar{r}_n \geq v(z_1) \geq E_{\sigma,\tau_\varepsilon} \limsup_{n\to\infty} \bar{r}_n - \varepsilon \quad \forall \sigma, \tau. \tag{2}$$

It is known that all finite absorbing games [7] and, more generally, all finite stochastic games [9], and all compact absorbing games [10], have a value.

A strategy σ_ε (respectively τ_ε) that satisfies the left-hand (respectively, right-hand) inequality (1) is called *uniform ε-optimal*. A strategy σ_ε (respectively τ_ε) that satisfies the left-hand (respectively, right-hand) inequality (2) is called *limiting-average ε-optimal*.

If σ_ε and τ_ε are uniform ε-optimal strategies, then for every $\varepsilon' > \varepsilon$ there is $\lambda_{\varepsilon'} > 0$ such that for every τ and $0 < \lambda < \lambda_{\varepsilon'}$,

$$(3)\quad \varepsilon' + E_{\sigma_\varepsilon,\tau} \sum_{t=1}^{\infty} \lambda(1-\lambda)^{t-1} r_t \geq v(z_1) \geq E_{\sigma,\tau_\varepsilon} \sum_{t=1}^{\infty} \lambda(1-\lambda)^{t-1} r_t - \varepsilon' \quad \forall \sigma, \tau.$$

A strategy σ_ε that satisfies both left-hand inequalities (1) and (2) is called *ε-optimal*.

The definition of an absorbing game in this section generalizes the one given in the introduction. However, the analysis of the two-player zero-sum case is essentially the same in both models of an absorbing game. The reason is that once an absorbing state is reached, the players will play the optimal strategies of the stage game, resulting in a stage payoff whose expectation equals the value of the stage game. Therefore, one can replace each entry of the nonabsorbing state by the absorbtion probability, i.e., the probability of departure from the nonabsorbing state, and the absorbing payoff being the conditional expected value of the next absorbing state.

2.2. **Memory-based strategies.** A *memory-based strategy* σ generates a random sequence of memory states $m_1, \ldots, m_t, m_{t+1}, \ldots$, where the memory is updated stochastically in each stage, and selects its action i_t according to a distribution that depends on just the current time t, its current memory m_t, and the current state z_t. Explicitly, the conditional distribution of i_t, given $h_t^m := (z_1, m_1, i_1, j_1, \ldots, z_t, m_t)$, is a function σ_α of (t, z_t, m_t) and the conditional distribution of m_{t+1}, given $(h_t^m, i_t, j_t, z_{t+1})$, is a function σ_m of (t, z_t, m_t, i_t, j_t) (i.e., it depends on just the time t and the tuple (z_t, m_t, i_t, j_t)).

A memory-based strategy σ is *clock-independent* if the functions σ_α and σ_m are independent of t.

A *k-memory strategy* is a memory-based strategy in which the memory states m_t take values in a set with (at most) k elements. Note that a strategy is a Markov strategy if and only if it is a one-memory strategy, and a strategy is a stationary strategy if and only if it is a one-memory clock-independent strategy. A strategy *uses finite memory* if it is a k-memory strategy where k is finite. A strategy that uses finite memory is called a *finite-memory strategy*. The set of all k-memory strategies is denoted by $\mathcal{M}_k$.

The long-standing natural open problem that motivates the present paper is whether for every stochastic game there are ε-optimal strategies that use finite memory. This has recently been settled affirmatively for the Big Match [6], and the present paper settles this problem affirmatively for all absorbing games.

3. The result

The main result of the present paper is that, in every absorbing game, there is a finite-memory strategy that is ε-optimal and moreover that is a three-memory strategy.

Theorem 1. *Every absorbing game with value v has, for every $\varepsilon > 0$, a 3-memory strategy σ of Player 1 and n_ε such that for every strategy τ of Player 2,*

$$E_{\sigma,\tau} \liminf_{n\to\infty} \bar{r}_n \geq v - \varepsilon, \tag{4}$$

and

$$E_{\sigma,\tau}\bar{r}_n \geq v - \varepsilon \qquad \forall n \geq n_\varepsilon. \tag{5}$$

Similarly for Player 2.

4. A few preliminaries to the proof

It is sufficient to prove the result for a compact absorbing game whose initial state z_1 is the non-absorbing state and $v(z_1) = 0$. Fix a compact absorbing game whose initial state z_1 is the non-absorbing state and $v(z_1) = 0$.

The set of stages $t = 1, 2, \ldots$ of the infinite game is partitioned into consecutive epochs, indexed by $i = 1, 2, \ldots$, where the number of stages of the i-th epoch is s_i. The number of stages in the first n epochs equals $\sum_{i=1}^n s_i$ and is denoted by S_n.

The definition of the duration of the epochs is not identical to the definition in [6], but the proofs of the properties of the epochs' duration are very similar to those in [6]. In this section, we state these properties, and in order to make the present paper independent we include here also the slightly different proofs.

The epochs' duration. We now define the sequence (s_i) of durations of epochs. The sequence of durations depends on a fixed sufficiently small $0 < \varepsilon < 1/2$, and a sufficiently small $\lambda > 0$. For notational convenience we set $p = e^{-\lambda'}$, where $\lambda' = \lambda/(1-\lambda)$. The condition that λ is sufficiently small will guarantee, in particular, that p is sufficiently large so that[3] $1/2 < 1 - \varepsilon < p < 1$ and hence, $1/p < 2$. As $1 + \varepsilon > 1$, $\sum_{i=1}^\infty \frac{2}{i^{1+\varepsilon}} < \infty$. Let i_ε be a sufficiently large positive integer so that

$$\sum_{i=i_\varepsilon+1}^{\infty} \frac{1}{i^{(1+\varepsilon)}} < \varepsilon. \tag{6}$$

[3] In [6] the parameter p was an explicit function of ε, $p = 1 - \varepsilon$.

The duration of the i-th epoch, s_i, is the largest integer such that $p^{-s_i} \leq i^{1+\varepsilon}$ if $i > i_\varepsilon$, and $s_i = 1$ if $i \leq i_\varepsilon$. Recall that the sum of the durations of the first n epochs is denoted by $S_n = \sum_{i=1}^n s_i$.

Lemma 1. *The sequence* (s_i) *satisfies*

$$s_{i+1} \geq s_i \geq 1\,\forall i \quad \text{and } s_n/S_n \to_{n\to\infty} 0. \tag{7}$$

Proof. In short, the definition of s_i implies that the sequence s_i is non-decreasing and that $s_n = \Theta(\ln n)$ and hence $S_n = \Theta(n \ln n)$, and therefore $s_n/S_n \to_{n\to\infty} 0$.

For completeness, we spell out the details. Recall that $p > 1 - \varepsilon > 1/2$. Note that $1 < p^{-1} \leq i^{1+\varepsilon}$ for every $i > 1$; hence, $s_i \geq 1$ for every $i > i_\varepsilon$, and recall that $s_i = 1$ for every $1 \leq i \leq i_\varepsilon$. For $i > i_\varepsilon$, $p^{-s_i} \leq i^{1+\varepsilon}$ by the definition of s_i, and $i^{1+\varepsilon} < (i+1)^{1+\varepsilon}$. Hence, by the definition of s_{i+1}, we have $s_{i+1} \geq s_i$. We conclude that $1 \leq s_i \leq s_{i+1}$ for every i.

For $i > i_\varepsilon$, the definition of s_i implies that $p^{-s_i} \leq i^{1+\varepsilon} \leq p^{-s_i - 1}$; hence, $\frac{1+\varepsilon}{-\ln p} \ln i \geq s_i \geq -1 + \frac{1+\varepsilon}{-\ln p} \ln i$. Therefore, $s_n = \Theta(\ln n)$ and $S_n = \sum_{i=1}^n s_i = \Theta(n \ln n)$ as $n \to \infty$, and therefore $s_n/S_n \to_{n\to\infty} 0$. □

Lemma 2. *There exists a constant* K *such that for all positive integers* i *and* n *with* $i \leq n$*,*

$$\frac{s_i}{S_n} \leq \frac{s_n}{S_n} \leq K n^{-1} \leq K n^{-\varepsilon^2} i^{\varepsilon^2 - 1}. \tag{8}$$

Proof. In short, this lemma follows from the following properties: s_i is non-decreasing, $s_n = \Theta(\ln n)$, $S_n = \Theta(n \ln n)$, and $n^{-1} = n^{-\varepsilon^2} n^{\varepsilon^2 - 1} \leq n^{-\varepsilon^2} i^{\varepsilon^2 - 1}$.

For completeness, we spell out an explicit derivation of these inequalities. For $i > i_\varepsilon$, $s_i \geq -1 + \frac{1+\varepsilon}{-\ln p} \ln i$; hence, for $n > 2(i_\varepsilon + 1)$, $S_n \geq \sum_{n/2-1 \leq i \leq n} s_i \geq -n + \frac{n}{2} \frac{1+\varepsilon}{-2\ln p} \ln n$. For $n > i_\varepsilon$, $s_n \leq \frac{1+\varepsilon}{-\ln p} \ln n$; hence, $s_n \leq \frac{1+\varepsilon}{-\ln p} \ln n \leq K\, n^{-1} S_n$ for a sufficiently large K. Hence, for $n > 2(i_\varepsilon + 1)$, $s_n/S_n < K n^{-1}$ for a sufficiently large K, and therefore there is a positive constant K such that for every n we have $s_n/S_n < K n^{-1}$. □

The payoff to Player 1 in the j-th round of epoch i is denoted by r_j^i. Note that the j-th round of epoch i is the $(S_{i-1} + j)$-th stage of the game. Therefore, $\sum_{t=1}^{S_n} r_t = \sum_{i=1}^n \sum_{j=1}^{s_i} r_j^i$. Hence, for n sufficiently large and $S_{n-1} < T \leq S_n$, $\frac{1}{T} \sum_{t=1}^T r_t \geq \frac{1}{S_n} \sum_{i=1}^n \sum_{j=1}^{s_i} r_j^i - \varepsilon$.

Therefore, in order to prove the theorem it suffices to exhibit a strategy $\sigma \in \mathcal{M}_3$ of Player 1 and n_ε such that for every pure strategy τ of Player 2, we have

$$E_{\sigma,\tau} \liminf_{n\to\infty} \frac{1}{S_n} \sum_{i=1}^n \sum_{j=1}^{s_i} r_j^i \geq -5\varepsilon, \tag{9}$$

and

$$E_{\sigma,\tau}\frac{1}{S_n}\sum_{i=1}^{n}\sum_{j=1}^{s_i} r_j^i \geq -11\varepsilon \quad \forall\, n \geq n_\varepsilon. \tag{10}$$

5. The proof

We start with recalling a few results from the theory of finite two-person zero-sum stochastic games and the theory of compact absorbing games.

The payoff in stage t, $t = 1, 2, \ldots$, is denoted by r_t. The λ-discounted game, $0 < \lambda \leq 1$, is the game where the payoff is $\sum_{t=1}^{\infty} \lambda(1-\lambda)^{t-1} r_t$. It is known that in a two-player zero-sum λ-discounted compact stochastic games, and in particular, in an absorbing game with compact action sets, each player has a stationary optimal strategy. The value of the λ-discounted game is a function of the initial state and is denoted by v_λ.

If σ_λ is a stationary optimal strategy in the λ discounted game, then for every strategy τ of player 2, $E_{\sigma_\lambda,\tau}(\lambda r_t + (1-\lambda) v_\lambda(z_{t+1}) \mid \mathcal{H}_t) \geq v_\lambda(z_t)$, where $\mathcal{H}_t$ is the history of play up to stage t (including the state z_t). Hence,

$$E_{\sigma_\lambda,\tau}(\lambda'(r_t - v_\lambda(z_t)) + v_\lambda(z_{t+1}) - v_\lambda(z_t) \mid \mathcal{H}_t) \geq 0, \quad \text{where } \lambda' = \frac{\lambda}{(1-\lambda)}. \tag{11}$$

It is known that in a finite two-person zero-sum stochastic game and in a compact absorbing game, v_λ converges to a limit as λ goes to 0. We denote this limit by v. Note that for every absorbing state z, $v_\lambda(z)$ is independent of λ and (hence) is equal to $v(z)$.

We continue with the proof that a compact absorbing game has, for every $\varepsilon > 0$, a 3-memory strategy that is ε-optimal.

Recall that we fixed a compact absorbing game with $v(z_1) = 0$. Fix $0 < \varepsilon < 1/2$. W.l.o.g. we assume that $-1 < -1 + \varepsilon \leq r \leq 1 - \varepsilon < 1$. Hence, $-1 < -1 + \varepsilon \leq v(z) \leq 1 - \varepsilon < 1$ and $-1 < -1 + \varepsilon \leq v_\lambda(z) \leq 1 - \varepsilon < 1$ for every state z and discount rate $0 < \lambda \leq 1$.

The strategy σ consists of patching together strategies σ_{s_i}, which will be defined later, where σ_{s_i} is a strategy in the i-th epoch.

The strategy uses two mixed actions, C and A, which we term the careful mixed action and the adventurous mixed action respectively.

The careful action C is a limit point of the mixed action of a stationary optimal strategy in the λ-discounted game as λ goes to 0. The adventurous mixed action A is the mixed action of a stationary optimal strategy in the λ-discounted game, where λ is sufficiently small, and A is sufficiently close to the careful mixed action C.

The careful definition of the epoch strategies and the duration of the epochs will guarantee that the sequence of random variables v_i, where

$$v_i := v(z_{S_i+1}),$$

obeys the following two properties. For every pure strategy τ of Player 2,

$$E_{\sigma,\tau} \liminf \frac{1}{S_n} \sum_{i=1}^{n} s_i v_{i-1} \geq v_0 - \varepsilon, \tag{12}$$

and

$$E_{\sigma,\tau}\, v_i \geq v_0 - \varepsilon \ \ \forall i. \tag{13}$$

Remark. This remark's role is to explain the necessity of conditions (12) and (13). Condition (12) is essentially necessary for the strategy σ to obey (9): One can show that if σ is a strategy of Player 1 for which there is a pure strategy τ of Player 2 such that the left hand side of inequality (12) is $< -5\varepsilon$, then there is a pure strategy τ^* of Player 2 for which inequality (9) does not hold.[4] Condition (13) is essentially necessary for the strategy σ to obey (10): One can show that if σ is a strategy of Player 1 for which there is i and a pure strategy τ of Player 2 such that the left hand side of inequality (13) is $< -11\varepsilon$, then there is a pure strategy τ^* of Player 2 for which inequality (10) does not hold for all sufficiently large n. *This concludes the remark.*

In addition, the strategy σ will satisfy the following two properties. Set $y_t = r_t - v(z_t) + \varepsilon$, $y^i_j = y_{S_{i-1}+j}$, and $\beta_i = \frac{-1}{s_i} \sum_{j=1}^{s_i} y^i_j$. Note that $-2 < \beta_i < 2$. For every pure strategy τ of Player 2,

$$\lim_{n\to\infty} \frac{1}{S_n} \sum_{i=1}^{n} s_i 1_{\{\beta_i \geq 3\varepsilon\}} = 0 \quad P_{\sigma,\tau}\text{-a.e.} \tag{14}$$

and for a sufficiently large n_ε, for every $n \geq n_\varepsilon$ and every pure strategy τ of Player 2,

$$E_{\sigma,\tau} \frac{1}{S_n} \sum_{i=1}^{n} s_i 1_{\{\beta_i \geq 3\varepsilon\}} \leq \varepsilon. \tag{15}$$

Lemma 3. *Inequality (9) holds whenever inequalities (12) and (14) hold, and inequality (10) holds whenever inequalities (13) and (15) hold.*

Proof. Using the definition of β_i, the equality $-s_i\beta_i = \sum_{j=1}^{s_i}(r^i_j - v(z^i_j) + \varepsilon)$, and the inequality $\sum_{j=1}^{s_i}(v(z^i_j) - v(z^i_1)) \leq s_1 1_{\{z^i_{s_i} \neq z^i_1\}}$, we have

$$\begin{aligned} \sum_{j=1}^{s_i} r^i_j &= \sum_{j=1}^{s_i} (r^i_j - v_{i-1} + v_{i-1}) \geq s_i v_{i-1} - s_i\beta_i - s_i\varepsilon - s_i 1_{\{z^i_{s_i} \neq z^i_1\}} \\ &= s_i v_{i-1} - s_i\beta_i 1_{\{\beta_i \geq 3\varepsilon\}} - s_i\beta_i 1_{\{\beta_i < 3\varepsilon\}} - s_i\varepsilon - s_i 1_{\{z^i_{s_i} \neq z^i_1\}} \\ &\geq s_i v_{i-1} - 2 s_i 1_{\{\beta_i \geq 3\varepsilon\}} - 3\varepsilon s_i - \varepsilon s_i - s_i 1_{\{z^i_{s_i} \neq z^i_1\}}. \end{aligned}$$

[4]And even the weaker inequality, where the lim inf is replaced in (9) by lim sup, does not hold.

Hence, by using the monotonicity of the sequence s_i, which implies $s_n \geq s_i \, \forall i \leq n$, and the inequality $\sum_i 1_{\{z^i_{s_i} \neq z^i_1\}} \leq 1$, we deduce that

$$\sum_{i=1}^{n} \sum_{j=1}^{s_i} r^i_j \geq \sum_{i=1}^{n} s_i v_{i-1} - 2 \sum_{i=1}^{n} s_i 1_{\{\beta_i \geq 3\varepsilon\}} - 4\varepsilon S_n - s_n.$$

Therefore, if (12) and (14) hold, then inequality (9) holds, and if (13) and (15) hold, then, for n_ε sufficiently large so that $s_n/S_n < \varepsilon$, inequality (10) holds. □

The epoch strategy σ_s. We now define the epoch strategy σ_s. We illustrate the update function in Figure 1 and next describe it formally. If $s = 1$ then $\sigma_s = \sigma_1$ plays the careful mixed action C. Recall that the careful mixed action C is a limit point, as $\lambda \to 0+$, of A_λ, where A_λ is a stationary optimal strategy in the λ-discounted game. We proceed with the definition of the strategy σ_s for $s > 1$.

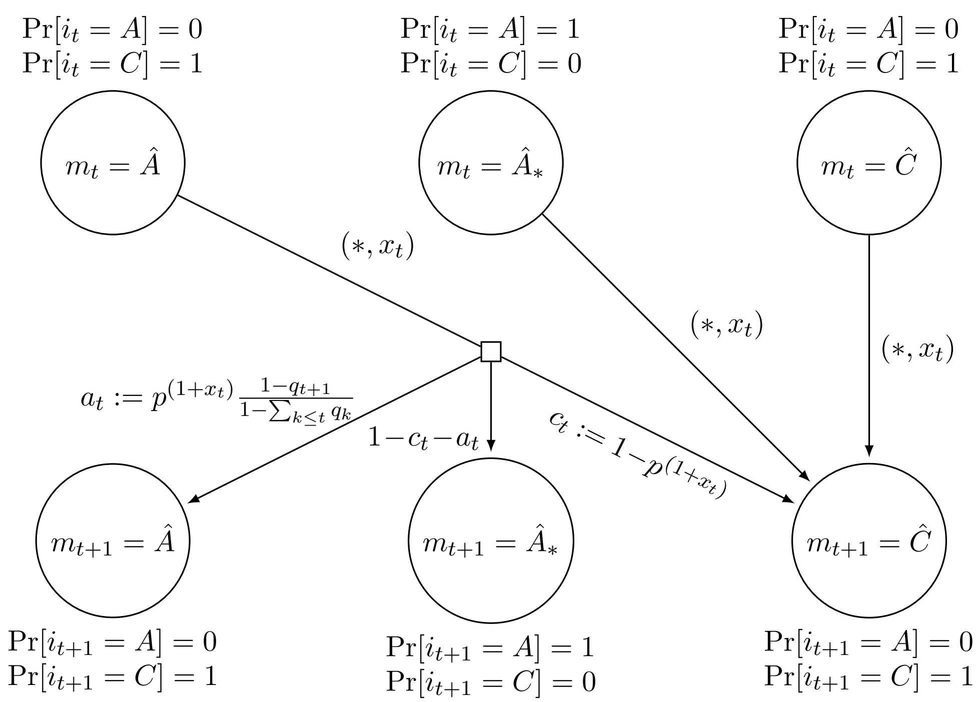

FIGURE 1. The memory update function for σ_s. The initial memory m_1 is $\hat{A}_*$ with pr. $\epsilon_1 p^s$ and otherwise $\hat{A}$

Let $0 < \lambda < 1$ be sufficiently small so that A_λ is sufficiently close to C,

$$\lambda' := \lambda/(1-\lambda) < \varepsilon/2,$$

v_λ is within $\varepsilon/2$ of v, i.e.,

$$|v_\lambda(z_1) - v(z_1)| < \varepsilon/2, \tag{16}$$

and, by setting $A := A_\lambda$, $E_{A,b} r(z,a,b) := \int r(z,a,b)\, A(a)$ is within ε of $E_{C,b} r(z,a,b)$ for any state z and action b of Player 2, i.e.,

$$|E_{A,b}\, r(z,a,b) - E_{C,b}\, r(z,a,b)| < \varepsilon \;\; \forall z, b. \tag{17}$$

Recall that $p = e^{-\lambda'}$, and let $s > 1$ be an integer.

The three states of memory of the strategy σ_s are $\widehat{C}$ (for continuing throughout with the careful mixed action C) and $\widehat{A}$ (for possible *future* play of the adventurous mixed action A) and $\widehat{A}_*$ (for playing the adventurous mixed action A).

The strategy plays the careful mixed action C when the memory state is either $\widehat{A}$ or $\widehat{C}$, and it plays the adventurous mixed action A when the memory state is $\widehat{A}_*$.

Fix $0 < \varepsilon_1 \leq 1 - p$. The initial state of memory, m_1, is $\widehat{A}_*$ with probability $q_1 = \varepsilon_1 p^s$ and $\widehat{A}$ with probability $1 - q_1$.

If the current memory state is either $\widehat{A}_*$ or $\widehat{C}$, then the next memory state is $\widehat{C}$.

We will define for each round number $1 \leq j < s$ an auxiliary random variable x_j with $x_j \geq -1$. The random variable x_j, which will be defined later as a function of the state and action of Player 2 in round j, is used by Player 1 in the definition of the stochastic memory updating. The distribution of the next memory state m_{j+1} as a function of the current memory state m_j and x_j is defined as follows.

The conditional probability that the next memory state m_{j+1} is $\widehat{A}_*$, given x_j and that the current memory state m_j is $\widehat{A}$, is

$$p^{(1+x_j)} \frac{q_{j+1}}{1 - \sum_{k \leq j} q_k},$$

where $q_k := \varepsilon_1 p^{s-(k-1)}$. Note that the condition $\varepsilon_1 \leq 1 - p$ guarantees that $1 - \sum_{k \leq j} q_k > q_{j+1} > 0$ for every $1 \leq j \leq s$, and the condition $x_j \geq -1$ guarantees that $p^{(1+x_j)} \leq 1$. Hence the above displayed formula is indeed a probability.

The conditional probability that $m_{j+1} = \widehat{A}$, given x_j and that the memory state is $m_j = \widehat{A}$, is

$$p^{(1+x_j)} \left(1 - \frac{q_{j+1}}{1 - \sum_{k \leq j} q_k}\right) = p^{(1+x_j)} \frac{1 - \sum_{k \leq j+1} q_k}{1 - \sum_{k \leq j} q_k}.$$

(Hence, the conditional probability that $m_{j+1} = \widehat{C}$, given x_j and that the current memory state m_j is $\widehat{A}$, is $1 - p^{(1+x_j)}$.)

This completes the definition of the strategy σ_s.

Recall that for every $1 \leq j < s$, $m_{j+1} = \widehat{C}$ whenever $m_j = \widehat{A}_*$ or $m_j = \widehat{C}$; hence, there is at most one value of $1 \leq j \leq s$ such that $m_j = \widehat{A}_*$.

Let τ be a strategy of Player 2. Let $p_j(x_1, \ldots, x_{j-1})$, or p_j for short, be the $P_{\sigma,\tau}$ conditional probability that $m_j = \widehat{A}_*$, given $x_1, \ldots, x_{j-1}$.

Lemma 4. *$p_j = \varepsilon_1 p^{s+\sum_{k<j} x_k}$; hence, $E_{\sigma_s,\tau} 1_{\{m_j=\widehat{A}_*\}} = E_{\sigma_s,\tau}(\varepsilon_1 p^{s+\sum_{k<j} x_k})$.*

Proof. $p_1 = \Pr(m_1 = \widehat{A}_*) = q_1 = \varepsilon_1 p^s$, which proves the statement of the lemma for $j = 1$. For $j > 1$, the event $m_j = \widehat{A}_*$ corresponds to a sequence of $j-1$ memory states $\widehat{A}$ followed by the memory state $\widehat{A}_*$. Hence, by the definition of the memory transitions,

$$\begin{aligned} p_j &= (1-q_1)\left(\prod_{1\le t<j-1} p^{(1+x_t)} \frac{1-\sum_{i\le t+1} q_i}{1-\sum_{i\le t} q_i}\right) p^{(1+x_{j-1})} \frac{q_j}{1-\sum_{t<j} q_t} \\ &= q_j \prod_{k<j} p^{(1+x_k)} = \varepsilon_1 p^{s+\sum_{k<j} x_k}. \end{aligned}$$

□

Consider the auxiliary game with $s+1$ stages, where the transitions of states and the stage payoffs follow the rules of the absorbing game and the players are active only in the first s stages j, $j = 1, \ldots, s$.

Recall that z_1 is the non-absorbing state. Define a function ω on the state space by

$$\omega(z) = v(z) - 2\varepsilon 1_{\{z=z_1\}}.$$

We continue with the explicit definition of x_j:

$$x_j := E_{A,b_j} r(z_j, a, b_j) - v(z_j) + \varepsilon.$$

The next lemma provides an inequality on the expectation of $\omega(z_{s+1}) - \omega(z_1)$, w.r.t. the probability $P_{\sigma_s,\tau}$ that is defined by the strategy σ_s of Player 1 and a (pure) strategy τ of Player 2.

Lemma 5. *For every strategy τ of Player 2,*

$$E_{\sigma_s,\tau}\omega(z_{s+1}) - \omega(z_1) \geq E_{\sigma_s,\tau}\varepsilon_1 p^{s+\sum_{j=1}^{s} x_j} - \varepsilon_1 p^s. \tag{18}$$

Proof. Let $\mathcal{H}_j$ be the σ-algebra of events defined by plays up to round j. If either $m_j = \widehat{C}$, or $m_j = \widehat{A}$, then σ plays the careful mixed action; hence, $E_{\sigma_s,\tau}(\omega(z_{j+1}) - \omega(z_j) \mid \mathcal{H}_j, m_j) \geq 0$. Hence,

$$E_{\sigma_s,\tau}(\omega(z_{j+1}) - \omega(z_j) \mid \mathcal{H}_j, m_j) \geq 1_{\{m_j=\widehat{A}_*\}} E_{\sigma_s,\tau}(\omega(z_{j+1}) - \omega(z_j) \mid \mathcal{H}_j, m_j).$$

If z_j is an absorbing state then $z_{j+1} = z_j$ and $x_j \geq 0$. Hence, on z_j being an absorbing state (and $m_j = \widehat{A}_*$), $\omega(z_{j+1}) = \omega(z_j)$, and $E_{\sigma_s,\tau}(\omega(z_{j+1}) - \omega(z_j) \mid \mathcal{H}_j, m_j) \geq E_{\sigma_s,\tau}(p^{x_j} - 1 \mid \mathcal{H}_j, m_j)$.

On z_j being the non-absorbing state and $m_j = \widehat{A}_*$,

$$
\begin{aligned}
& E_{\sigma_s,\tau}(\omega(z_{j+1}) - \omega(z_j) \mid \mathcal{H}_j, m_j) \\
\geq\ & E_{\sigma_s,\tau}(v_\lambda(z_{j+1}) - v_\lambda(z_j) \mid \mathcal{H}_j, m_j) \\
=\ & E_{\sigma_s,\tau}(\lambda'(r_j - v_\lambda(z_j)) + v_\lambda(z_{j+1}) - v_\lambda(z_j) \mid \mathcal{H}_j, m_j) \\
& -E_{\sigma_s,\tau}(\lambda'(r_j - v_\lambda(z_j)) \mid \mathcal{H}_j, m_j) \\
\geq\ & -E_{\sigma_s,\tau}(\lambda'(r_j - v(z_j) + \varepsilon + v(z_j) - v_\lambda(z_j) - \varepsilon) \mid \mathcal{H}_j, m_j) \\
\geq\ & -E_{\sigma_s,\tau}(\lambda'(x_j - \varepsilon/2) \mid \mathcal{H}_j, m_j) \\
\geq\ & E_{\sigma_s,\tau}(e^{-\lambda' x_t} - 1 \mid \mathcal{H}_j, m_j) = E_{\sigma_s,\tau}(p^{x_t} - 1 \mid \mathcal{H}_j, m_j).
\end{aligned} \tag{19}
$$

The first inequality follows from the inequality $\omega(z_{j+1}) - \omega(z_j) \geq v_\lambda(z_{j+1}) - v_\lambda(z_j)$, which follows from the definition of ω along inequality (16) and the equality $v(z) = v_\lambda(z)$ which holds whenever z is an absorbing state. The second inequality follows from the definition of the adventurous mixed action A along (11). The third inequality follows from the definition of x_j and the inequality $v(z_j) - v_\lambda(z_j) \leq \varepsilon/2$. The last inequality follows from the inequalities $\lambda' < \varepsilon/2$ and $e^{-\lambda' x_t} - 1 \leq -\lambda' x_t + (\lambda')^2 \leq -\lambda' x_t + \lambda' \varepsilon/2$.

Recall that x_j is a function of z_j and b_j. Hence, the conditional distribution of x_j, given $\mathcal{H}_j$ and m_j, is independent of m_j, and recall that $E_{\sigma_s,\tau}(1_{\{m_j = \widehat{A}_*\}} \mid \mathcal{H}_j) = \varepsilon_1 p^{s + \sum_{k<j} x_k}$. Therefore,

$$
\begin{aligned}
& E_{\sigma_s,\tau}(1_{\{m_j = \widehat{A}_*\}} E_{\sigma_s,\tau}(p^{x_j} - 1 \mid \mathcal{H}_j, m_j)) \\
=\ & E_{\sigma_s,\tau}(E_{\sigma_s,\tau}(1_{\{m_j = \widehat{A}_*\}} \mid \mathcal{H}_j) E_{\sigma_s,\tau}(p^{x_j} - 1 \mid \mathcal{H}_j)) \\
=\ & E_{\sigma_s,\tau}(\varepsilon_1 p^{s + \sum_{k<j} x_k} E_{\sigma_s,\tau}(p^{x_j} - 1 \mid \mathcal{H}_j)) \\
=\ & E_{\sigma_s,\tau}(E_{\sigma_s,\tau}(\varepsilon_1 p^{s + \sum_{k<j} x_k}(p^{x_j} - 1) \mid \mathcal{H}_j)) \\
=\ & E_{\sigma_s,\tau}(\varepsilon_1 p^{s + \sum_{k<j} x_k}(p^{x_j} - 1)).
\end{aligned}
$$

In the second to the last equality we used the fact that $\varepsilon_1 p^{s + \sum_{k<j} x_k}$ (is a function of h_j and thus) is measurable with respect to $\mathcal{H}_j$, and in the last equality we used the fact that the expectation is the expectation of the conditional expectation.

Therefore,

$$
\begin{aligned}
E_{\sigma_s,\tau}(\omega(z_{j+1}) - \omega(z_j)) \ &\geq\ E_{\sigma_s,\tau}(\varepsilon_1 p^{s + \sum_{k<j} x_k}(p^{x_j} - 1)) \\
&=\ \varepsilon_1 E_{\sigma_s,\tau}(p^{s + \sum_{k \leq j} x_k}) - \varepsilon_1 E_{\sigma_s,\tau}(p^{s + \sum_{k<j} x_k}).
\end{aligned}
$$

Summing the inequalities over $j = 1, \ldots, s$, we have

$$
E_{\sigma_s,\tau}\omega(z_{s+1}) - \omega(z_1) \geq E_{\sigma_s,\tau}\varepsilon_1 p^{s + \sum_{k \leq s} x_k} - \varepsilon_1 p^s.
$$

□

Define a function v on plays of the auxiliary $(s+1)$-stage game by $v = \omega(z_{s+1}) - \omega(z_1)$.

Lemma 6. *Let* $\alpha(x) = -\sum_{j=1}^{s} x_j/s$. *Then*

$$E_{\sigma_s,\tau}v \geq \varepsilon_1 E_{\sigma_s,\tau} p^{(1-\alpha(x))s} - \varepsilon_1 p^s \tag{20}$$

$$\geq \varepsilon_1 p^{(1-\theta)s} E_{\sigma_s,\tau} 1_{\{\alpha(x)\geq\theta\}} - \varepsilon_1 p^s \quad \forall\theta > 0. \tag{21}$$

Proof. Inequality (20) follows directly from Lemma 5.

The function $\alpha \mapsto p^{(1-\alpha)s}$ is nonnegative and monotonic increasing in α, and $p^{(1-\theta)s} \geq p^{(1-\theta)s} 1_{\{\alpha\geq\theta\}}$. Therefore, equality (20) implies inequality (21), which completes the proof of the lemma. ☐

The strategy σ. We proceed with the definition of the 3-memory strategy σ of Player 1. In short, the strategy σ follows the strategy σ_{s_i} in the i-th epoch. In order to see that σ is a 3-memory strategy, we define explicitly its memory states, its action function α_σ, and its memory updating function β_σ. The three states of memory of σ are $\widehat{C}$, $\widehat{A}$, and $\widehat{A_*}$. The action function of strategy σ in stage $t = S_{i-1} + j$, $1 \leq j \leq s_i$, namely, $\alpha_\sigma(t, m)$, coincides with the action function of σ_i in round j, i.e., $\alpha_\sigma(t, m) = \alpha_{\sigma_{s_i}}(j, m)$. The memory updating function of the strategy σ in stage $t = S_{i-1}+j$, $1 \leq j < s_i$, namely, $\beta_\sigma(t, m, a, b)$, coincides with the memory updating function of the strategy σ_i in round j, namely, $\beta_\sigma(t, m, a, b) = \beta_{\sigma_{s_i}}(j, m, a, b)$. The memory updating function of the strategy σ in stage $t = S_{i-1} + s_i = S_i$, namely, $\beta_\sigma(t, m, a, b)$, is such that the distribution of the memory state in stage $t+1$ (is independent of m, a, b and) coincides with the distribution of the initial state in round 1 of the strategy $\sigma_{s_{i+1}}$ (of the $(i+1)$-th epoch).

The strategy σ obeys properties (12), (13), (14), and (15). The next lemma introduces an auxiliary sequence of random variables, whose properties are used in the following lemma that shows that the strategy σ obeys properties (12), (13), (14), and (15). Set $\alpha_i = 0$ if $s_i = 1$ or if the i-th epoch starts with an absorbing state, and $\alpha_i = -\sum_{j=1}^{s_i} x_j^i$. ($*_j^i$ refers to the $*$-th entry in the j-th round of epoch i).

Let $v_i := \omega(z_{S_i+1})$. In other words, $v_i = \omega(z_1^{i+1})$, where z_j^i is the state in the j-th round of the i-th epoch.

Lemma 7. *The sequence of random variables* $(Y_i)_{i\geq1}$ *that is defined by*

$$Y_i = v_i - \sum_{k>\max(i,i_\varepsilon)}^{\infty} \frac{2}{k^{(1+\varepsilon)}}$$

obeys $Y_i - v_i \to_{i\to\infty} 0$, $Y_i \leq v_i \leq Y_i + \varepsilon$, $-1 < Y_i < 1$, *and for every pure strategy* τ *of Player 2,*

$$E_{\sigma,\tau}(Y_i - Y_{i-1} \mid \mathcal{H}_i) \geq \varepsilon_1 E_{\sigma,\tau}(i^{\varepsilon^2-1} 1_{\{\alpha_i\geq\varepsilon\}} \mid \mathcal{H}_i), \tag{22}$$

where $\mathcal{H}_i$ *is the history of play up to the start of the* i*-th epoch.*

Proof. By the definition of v_i and (6), $|Y_i| < 1$, and, as $Y_i - Y_{i-1} = v_i - v_{i-1}$ for $i < i_\varepsilon$ and $Y_i - Y_{i-1} = v_i - v_{i-1} + \frac{1}{i^{1+\varepsilon}}$ for $i \geq i_\varepsilon$, $Y_i - v_i \to_{i\to\infty} 0$ and $Y_i \leq v_i \leq Y_i + \varepsilon$. It remains to prove inequality (22).

Inequality (21) along with the definition of α_i implies that for $i \geq i_\varepsilon$,

$$\begin{aligned} E_{\sigma,\tau}(v_i - v_{i-1} \mid \mathcal{H}_i) &\geq \varepsilon_1 E_{\sigma,\tau} p^{(1-\varepsilon)s_i} 1_{\{\alpha_i \geq \varepsilon\}} - \varepsilon_1 p^{s_i} \\ &\geq \varepsilon_1 i^{\varepsilon^2 - 1} E_{\sigma,\tau} 1_{\{\alpha_i \geq \varepsilon\}} - \frac{\varepsilon_1/p}{i^{1+\varepsilon}}, \\ &\geq \varepsilon_1 i^{\varepsilon^2 - 1} E_{\sigma,\tau} 1_{\{\alpha_i \geq \varepsilon\}} - \frac{1}{i^{1+\varepsilon}}, \end{aligned} \tag{23}$$

Therefore, (22) holds for $i \geq i_\varepsilon$. For $1 \leq i < i_\varepsilon$, $E_{\sigma,\tau}(v_i - v_{i-1} \mid \mathcal{H}_i) \geq 0$ (because $s_i = 1$ for such i and thus σ plays C in such epochs) and $\alpha_i = 0$ (by definition), and (22) holds for $i < i_\varepsilon$. We conclude that (22) holds for every i. □

An implication of the lemma is that $(Y_i)_{i > i_\varepsilon}$ is a bounded submartingale and therefore converges a.e. (namely, with $P_{\sigma,\tau}$ probability 1) to a limit Y_∞. As $v_i - Y_i \to_{i\to\infty} 0$, v_i converges to Y_∞ as i goes to infinity.

Lemma 8. *The strategy σ obeys properties (12) and (13).*

Proof. Let τ be a pure strategy of Player 2. As Y_{i-1} is a function of the play up to the start of the i-th epoch, inequality (22) shows that the sequence of random variables $(Y_i)_{i\geq 0}$ is a submartingale (with respect to the probability distribution $P_{\sigma,\tau}$ on plays). In addition, $Y_0 \geq v_0 - \varepsilon$ and $v_i \geq Y_i$. Therefore, $E_{\sigma,\tau} v_i \geq E_{\sigma,\tau} Y_i \geq E_{\sigma,\tau} Y_0 \geq v_0 - \varepsilon$, which proves (13).

As Y_i is a bounded submartingale, it converges a.e. to a limit Y_∞ and $E_{\sigma,\tau} Y_\infty \geq Y_0$. As $Y_i - v_i \to_{i\to\infty} 0$, we have $v_i \to_{i\to\infty} Y_\infty$.

As $v_i \to_{i\to\infty} Y_\infty$, $\frac{s_i}{S_n} \to_{n\to\infty} 0$ for each fixed i, and $S_n = \sum_{i=1}^n s_i$, we have

$$\frac{1}{S_n} \sum_{i=1}^{n} s_i v_{i-1} \to_{n\to\infty} Y_\infty \ P_{\sigma,\tau}\text{-a.e.} \tag{24}$$

Hence, $E_{\sigma,\tau} \lim_{n\to\infty} \frac{1}{S_n} \sum_{i=1}^n s_i v_{i-1} = E_{\sigma,\tau} Y_\infty \geq Y_0 \geq v_0 - \varepsilon$, which proves (12). □

Lemma 9. *The strategy σ obeys properties (14) and (15).*

Proof. Note that (as $-1 < Y_i < 1$) $Y_i - Y_j < 2$. Taking the expectations in inequality (22), we deduce that $E_{\sigma,\tau}(Y_i - Y_{i-1}) \geq E_{\sigma,\tau}\, i^{\varepsilon^2-1} 1_{\{\alpha_i \geq \varepsilon\}}$. Summing these inequalities over all i such that $1 \leq i \leq n$, we deduce that

$$2 > E_{\sigma,\tau}(Y_n - Y_0) \geq E_{\sigma,\tau} \sum_{i=1}^{n} i^{\varepsilon^2-1} 1_{\{\alpha_i \geq \varepsilon\}}. \tag{25}$$

By the monotone convergence theorem, inequality (25) implies that $2 \geq E_{\sigma,\tau} \sum_{i=1}^\infty i^{\varepsilon^2-1} 1_{\{\alpha_i \geq \varepsilon\}}$. Hence, $\sum_{i=1}^\infty i^{\varepsilon^2-1} 1_{\{\alpha_i \geq \varepsilon\}}$ is finite a.e. Hence, using

(8), for every pure strategy τ of Player 2,

$$(26)\quad 0 \leq \frac{1}{S_n}\sum_{i=1}^{n} s_i 1_{\{\alpha_i \geq \varepsilon\}} \leq K\, n^{-\varepsilon^2}\sum_{i=1}^{n} i^{\varepsilon^2-1} 1_{\{\alpha_i\geq\varepsilon\}} \to_{n\to\infty} 0 \quad P_{\sigma,\tau}\text{-a.e.,}$$

Next we prove that

$$(27)\qquad 0 \leq \frac{1}{S_n}\sum_{i=1}^{n} s_i 1_{\{\beta_i \geq \alpha_i + 2\varepsilon\}} \to_{n\to\infty} 0 \quad P_{\sigma,\tau}\text{-a.e.}$$

Set $\gamma_i = \frac{1}{s_i}\sum_{j=1}^{s_i}(E_{\sigma,\tau}(r^i_j \mid \mathcal{H}^i_j) - v(z^i_j) + \varepsilon)$. As the strategy σ plays either the mixed action C or the mixed action A, inequality (17) implies that $\gamma_i \leq \alpha_i + \varepsilon$, and hence $1_{\{\beta_i\geq\alpha_i+2\varepsilon\}} \leq 1_{\{\beta_i\geq\gamma_i+\varepsilon\}}$. Note that $\beta_i - \gamma_i = \frac{1}{s_i}\sum_{j=1}^{s_i}(r^i_j - E_{\sigma,\tau}(r^i_j \mid \mathcal{H}^i_j))$. The sequences $(r^i_j - E_{\sigma,\tau}(r^i_j \mid \mathcal{H}^i_j))$, $j = 1,\ldots,s_i$, and $\frac{1}{s_i}\sum_{j=1}^{s_i}(r^i_j - E_{\sigma,\tau}(r^i_j \mid \mathcal{H}^i_j))$, $i \geq 1$, are sequences of bounded martingale differences. Therefore, by the weak law of large numbers (or the large deviations inequality) for bounded differences of a martingale, $P_{\sigma,\tau}(\beta_i \geq \gamma_i + \varepsilon) \to_{n\to\infty} 0$ and $\frac{1}{n}\sum_{i=1}^{n} 1_{\{\beta_i\geq\gamma_i+\varepsilon\}} \to_{n\to\infty} 0$. Therefore, using the monotonicity of the sequence s_i and the equality $\frac{ns_n}{S_n} = \Theta(1)$, we have

$$\frac{1}{S_n}\sum_{i=1}^{n} s_i 1_{\{\beta_i\geq\alpha_i+2\varepsilon\}} \leq \frac{1}{S_n}\sum_{i=1}^{n} s_i 1_{\{\beta_i\geq\gamma_i+\varepsilon\}} \leq \frac{ns_n}{S_n}\frac{1}{n}\sum_{i=1}^{n} 1_{\{\beta_i\geq\gamma_i+\varepsilon\}} \to_{n\to\infty} 0 \quad P_{\sigma,\tau}\text{-a.e.,}$$

which proves (27). As $1_{\{\beta_i\geq 3\varepsilon\}} \leq 1_{\{\alpha_i\geq\varepsilon\}} + 1_{\{\beta_i\geq\alpha_i+2\varepsilon\}}$, we deduce, using (26) and (27), that

$$(28)\qquad 0 \leq \frac{1}{S_n}\sum_{i=1}^{n} s_i 1_{\{\beta_i\geq 3\varepsilon\}} \to_{n\to\infty} 0,$$

which proves (14).

We proceed to prove (15). Let n_ε be a sufficiently large integer so that $K\, n_\varepsilon^{-\varepsilon^2} < \varepsilon/4$ (and an additional condition will be imposed later). Hence, $\frac{s_i}{S_n} \leq i^{\varepsilon^2-1}\varepsilon/4$ for every $n \geq n_\varepsilon$ and $i_\varepsilon < i \leq n$. Then, using inequality (25), we have

$$(29)\quad E_{\sigma,\tau}\frac{1}{S_n}\sum_{i=1}^{n} s_i 1_{\{\alpha_i\geq\varepsilon\}} \leq E_{\sigma,\tau}\sum_{i=1}^{n} i^{\varepsilon^2-1} 1_{\{\alpha_i\geq\varepsilon\}}\varepsilon/4 < \varepsilon/2 \quad \forall n \geq n_\varepsilon.$$

Next we prove that

$$(30)\qquad E_{\sigma,\tau}\frac{1}{S_n}\sum_{i=1}^{n} s_i 1_{\{\beta_i\geq\alpha_i+2\varepsilon\}} \leq \varepsilon/2 \quad \forall n \geq n_\varepsilon.$$

Recall that $s_i \to_{i\to\infty} \infty$. Hence, by the weak law of large numbers (or the large deviations inequality) for bounded differences of a martingale,

$P_{\sigma,\tau}(\beta_i \geq \gamma_i + \varepsilon) \to_{n\to\infty} 0$ and the rate of convergence is independent of τ. Hence, for a sufficiently large n_ε,

$$E_{\sigma,\tau} \frac{1}{S_n} \sum_{i=1}^{n} s_i 1_{\{\beta_i \geq \gamma_i + \varepsilon\}} \leq \varepsilon/2 \quad \forall n \geq n_\varepsilon.$$

As $1_{\{\beta_i \geq \alpha_i + 3\varepsilon\}} \leq 1_{\{\alpha_i \geq \varepsilon\}} + 1_{\{\beta_i \geq \alpha_i + 2\varepsilon\}} \leq 1_{\{\alpha_i \geq \varepsilon\}} + 1_{\{\beta_i \geq \gamma_i + \varepsilon\}}$, we deduce that

$$E_{\sigma,\tau} \frac{1}{S_n} \sum_{i=1}^{n} s_i 1_{\{\beta_i \geq 3\varepsilon\}} < \varepsilon \quad \forall n \geq n_\varepsilon. \tag{31}$$

which proves (15). □

End of proof. The constructed strategy σ is in $\mathcal{M}_3$. The results of lemmas 3, 8, and 9 shows that σ satisfies for every strategy τ of Player 2 inequalities (9) and (10). This complete the proof of the main result.

6. Open problems

The main open problem is whether or not in any finite stochastic game each player has a finite-memory strategy that is ε-optimal. We do not know if Player 1 has an ε-optimal strategy in the following stochastic game, due to Beweley and Kohlberg, with two non-absorbing states, see Figure 2. We suspect that finding a finite-memory ε-optimal strategy for this game will lead to a finite-memory strategy for all finite stochastic games.

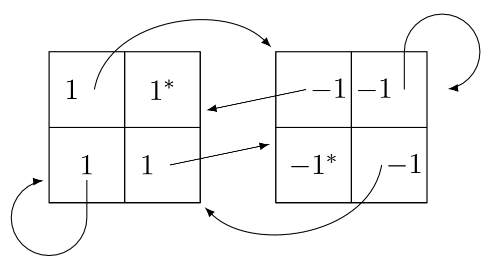

Figure 2. A game due to Beweley and Kohlberg for which we do not know a finite-memory clock-dependent strategy

References

[1] M. Amitai. Stochastic games with automata. Master's thesis, Hebrew University, Jerusalem, 1989. (in Hebrew).

[2] T. Bewley and E. Kohlberg. The asymptotic theory of stochastic games. *Mathematics of Operations Research*, 1:197–208, 1976.

[3] D. Blackwell and T. S. Ferguson. The big match. *The Annals of Mathematical Statistics*, 39(1):159–163, 1968.

[4] D. Gillette. Stochastic games with zero-stop probabilities. In *Contributions to the Theory of Games III*, volume 39 of *Ann. Math. Studies*, pages 179–187. Princeton University Press, 1957.

[5] K. A. Hansen, R. Ibsen-Jensen, and M. Koucký. The big match in small space (extended abstract). In M. Gairing and R. Savani, editors, *Proceedings of 9th International Symposium on Algorithmic Game Theory, SAGT 2016*, volume 9928 of *Lecture Notes in Computer Science*, pages 64–76. Springer, 2016.
[6] K. A. Hansen, R. Ibsen-Jensen, and A. Neyman. The big match with a clock and a bit of memory. In *Proceedings of the 2018 ACM Conference on Economics and Computation*, EC '18, page 149–150. Association for Computing Machinery, 2018.
[7] E. Kohlberg. Repeated games with absorbing states. *The Annals of Statistics*, 2(4):724–738, 1974.
[8] E. Kohlberg and S. Zamir. Repeated games of incomplete information: the symmetric case. *Annals of Statistics*, 2(5):1040–1041, 1974.
[9] J.-F. Mertens and A. Neyman. Stochastic games. *Int. J. of Game Theory*, 10(2):53–66, 1981.
[10] J.-F. Mertens, A. Neyman, and D. Rosenberg. Absorbing games with compact action spaces. *Mathematics of Operations Research*, 34:257–262, 2009.
[11] A. Neyman and S. Sorin. Equilibria in repeated games with incomplete information: The deterministic symmetric case. In T. Parthasaraty, editor, *Game Theoretic Applications to Economiccs and Operations Research*, pages 129–131. Kluwer Academic Publishers, 1997.
[12] A. Neyman and S. Sorin. Equilibria in repeated games with incomplete information: The general symmetric case. *International Journal of Game Theory*, 27:201–210, 1998.

Kristoffer Arnsfelt Hansen, Aarhus University
Email address: `arnsfelt@cs.au.dk`

Rasmus Ibsen-Jensen, University of Liverpool
Email address: `R.Ibsen-Jensen@liverpool.ac.uk`

Abraham Neyman, Hebrew University
Email address: `aneyman@huji.ac.il`