Submitted to *Mathematics of Operations Research*
manuscript MOR-2018-230

The Big Match with a clock and a bit of memory

Kristoffer Arnsfelt Hansen
Aarhus University, arnsfelt@cs.au.dk,

Rasmus Ibsen-Jensen
University of Liverpool, rij@liverpool.ac.uk,

Abraham Neyman
Hebrew University, aneyman@huji.ac.il,

The Big Match is a multi-stage two-player game. In each stage Player 1 hides one or two pebbles in his hand, and his opponent has to guess that number; Player 1 loses a point if Player 2 is correct, and otherwise he wins a point. As soon as Player 1 hides one pebble, the players cannot change their choices in any future stage.

The undiscounted Big Match has been much-studied. Blackwell and Ferguson (1968) give an ε-optimal strategy for Player 1 that hides, in each stage, one pebble with a probability that depends on the entire past history. Any strategy that depends on just the clock *or* just a finite memory is worthless (i.e., cannot guarantee strictly more than the least reward). The long-standing natural open problem has been whether every strategy that depends on just the clock *and* a finite memory is worthless.

The present paper proves that there is such a strategy that is ε-optimal. In fact, we show that just two states of memory are sufficient.

Key words: Stochastic games; Markov strategies; Bounded memory
MSC2000 subject classification: Primary: 91A15; secondary: 91A05
OR/MS subject classification: Primary: Games/group decisions: Stochastic

---

1. Introduction The game of Odd and Even (Latin: *Par Impar Ludere*, Greek: ἀρτιασμός) has been popular since ancient Greek and Roman times. It is played by two players, Player 1 and Player 2. Player 1 hides (e.g., in his hands) a number of pebbles or other items (e.g., beans, nuts, almonds, astragali, or coins), and his opponent, Player 2, has to guess whether the number of hidden items is odd or even. Player 1 then reveals the number. If Player 2 is right, Player 1 loses a point; otherwise, Player 1 wins a point (from Player 2).

Player 1 can guarantee that he gets (at least) zero points on average by hiding an odd or even number of items with equal probability. Player 2 can guarantee that Player 1 gets (at most) zero points on average by guessing odd or even with equal probability.

The repeated Odd and Even game is the same game repeated many times. Each player can still guarantee himself zero points on average (per stage and hence also in total) by playing, independently, in each stage as before.

The Big Match is also a multi-stage game. It is a variant of the repeated Odd and Even game. In each stage Player 1 hides one or two pebbles. In each stage, Player 1 wins or loses a point. As long as Player 1 hides two pebbles, Player 1 wins a point iff Player 2 guesses odd in that stage. The first stage in which Player 1 hides one pebble is called the stopping stage. In the stopping

stage Player 1 wins a point iff Player 2 guesses even. In each subsequent stage, he wins a point iff he won a point in the stopping stage.

The Big Match was introduced by Gillette [3] and has been much studied, in part due to its arguably being the most basic game model that illustrates the difficulty of balancing the trade-off between short- and long-term strategic considerations.

In the Big Match, Player 2 can still guarantee that Player 1 gets zero points on average, independently of the number of stages, by guessing odd or even with equal probability and independently in each stage. Executing such a strategy does not require that Player 2 know the past history, the number of stages, or the stage number. However, the situation of Player 1 is completely different! Henceforth, unless otherwise mentioned, a strategy refers to a strategy of Player 1.

If Player 1 knows the number of stages, T, in advance, he can guarantee himself (at least) zero points on average. To guarantee this, he *must* hide one pebble with probability $\frac{1}{k+1}$ when k stages remain. Thus, for example, in the last stage he hides one or two pebbles with equal probability, and in the first stage he hides one pebble with probability $\frac{1}{T+1}$. Executing such a strategy requires that Player 1 know the stage number and the number of stages, but it does not require that Player 1 know the past history.

It follows from the above that if Player 1 does not know the number of stages T in advance, then he has no way of guaranteeing himself (at least) zero points (per stage) on average. This has led researchers to look for strategies that guarantee close to zero per stage on average in all sufficiently long Big Match games.

Any strategy in the Big Match has to decide on the stopping stage. A natural possibility is just to specify in advance the probability of each stage being the stopping stage. Such a strategy is called a Markov strategy. It has long been known, and it is easy to verify, that any Markov strategy in the Big Match is *worthless*; i.e., for any $\varepsilon > 0$ it does not guarantee more than $-1+\varepsilon$ points (per stage) on average in any sufficiently long Big Match game.

The principle of sunk cost seems to imply that optimizing from any point onwards should be independent of the past, and hence any optimization of the long-run average of the rewards can be achieved by a Markov strategy. Since any Markov strategy is worthless, one might erroneously conclude that any strategy is worthless.

However, this is not the case! Blackwell and Ferguson [2] introduced worthy (i.e., not worthless) strategies that prescribe the choice in each stage as a function of the past history. Moreover, [2] introduced, for every $\varepsilon > 0$, a strategy that is *ε-optimal*; namely, it guarantees at least $-\varepsilon$ points (per stage) on average in all sufficiently long games.[1]

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

A *memory-based* strategy in the Big Match is a strategy in which the conditional probability of hiding one pebble depends on the current memory state and the clock (i.e., the stage number). The memory state is updated as a stochastic function of the current memory and of the guess of Player 2 in the previous stage, as well as of the clock.

The ε-optimal strategies in [2, Theorem 2] are memory-based, and those in [2, Theorem 1] are memory-based and clock-independent; i.e., the hiding and memory updating do not depend on the clock. The memory state is simply the difference between the number of odd and even guesses; hence, up to stage T it takes integer values in the interval $[-T, T]$.

The ε-optimal strategy in [4] is memory-based and clock-independent. The memory state can be encoded so that, with high probability, up to stage T it takes integer values in $[0, \ln^c T]$, for some constant c (and $T > 3$).

[1] Recall that Player 2 has a strategy that ensures himself zero points per stage on average.

On the other hand, all memory-based strategies that have a finite set of memory states and either are clock-independent (see, e.g., [8]) or have a deterministic memory update function [4] are worthless in the Big Match.

It has been a long-standing natural open problem whether there exists an ε-optimal memory-based strategy that has a finite set of memory states (or even just a worthy one).

The present paper answers this question positively. We show that, for every $\varepsilon > 0$, there is such a strategy that is ε-optimal. Moreover, it is a two-memory strategy; namely, it has a two-element memory set.

Our ε-optimal strategy is also ε-optimal in the infinite game. In the infinite game, the average win per stage need not be well defined, as the average number of wins over the first T stages need not converge. Nonetheless, the ε-optimality of our strategy in the infinite game result is as strong as possible: Our ε-optimal strategy guarantees that for any strategy of Player 2 the expectation of the limit inferior of the averages of the stage payoffs is at least $-\varepsilon$.

2. The model and related results

2.1. Stochastic games A *finite two-person zero-sum stochastic game* Γ, henceforth, a *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 a starting state.

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 nonabsorbing state.

A *play* of the stochastic game is an infinite sequence $(z_1, i_1, j_1, \ldots, z_t, i_t, j_t, \ldots) \in (Z \times I \times J)^\infty$, where $(z_t, i_t, j_t) \in Z \times I \times J$. The set of all plays is denoted by H_∞. A play up to stage T is the finite sequence $h_T = (z_1, i_1, j_1, \ldots, z_T) \in (Z \times I \times J)^{T-1} \times Z$. The payoff r_t in stage t is $r(z_t, i_t, j_t)$ and the average of the payoffs in the first T stages, $\frac{1}{t}\sum_{t=1}^{T} r_t$, is denoted by $\bar{r}_T$.

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 the 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 T_ε

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

and

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

It is known that all absorbing games [5] and, more generally, all stochastic games [7] have a value.

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

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

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.[2] 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, or even just the Big Match, there are ε-optimal strategies that use finite memory.

In order to present previous results, we introduce a concept of a memory-based strategy with an infinite set of memory states, but where there is a bound on the growth of the number of memories in the first t stages of the game, for all t. The bound on this growth is represented by a nondecreasing function $f : N \to N$ and a nonincreasing function $\gamma : N \to [0, 1]$. An (f, γ)-memory strategy is a memory based strategy σ whose set of memory states is N and for every strategy τ and positive integer $t' \in N$, the $P_{\sigma,\tau}$-probability that $m_t \leq f(t)$ for all $t \geq t'$ is at least $1 - \gamma(t')$. Note that for the constant function $f = k$ and the zero function $\gamma = 0$, an $(f = k, \gamma = 0)$-memory strategy is a k-memory strategy.

2.3. The Big Match The Big Match, introduced in [3], is a highly inspiring stochastic game. The state space Z is $\{-1, 0, 1\}$.

Each state $z \in \{-1, 1\}$ is absorbing and the payoff function (to Player 1) in an absorbing state z is $r(z, \cdot, \cdot) = z$. The state $z = -1$ is called the *losing state of Player 1* and the state $z = 1$ is called the *winning state of Player 1*.

The action sets I and J are $\{0, 1\}$, and the payoff function in the nonabsorbing state 0 is

$$r(0, i, j) = \begin{cases} 1 & \text{if } j \neq i, \\ -1 & \text{if } j = i. \end{cases}$$

The transition distributions from the nonabsorbing state 0 are given by

$$p(0, i, j) = \begin{cases} \delta_0 & \text{if } i = 0 \\ \delta_{-1} & \text{if } i = j = 1 \\ \delta_1 & \text{if } i = 1 \neq j. \end{cases}$$

Blackwell and Ferguson [2] shows that the value of the Big Match is 0 by introducing, for every $\varepsilon > 0$, an ε-optimal strategy (which is, in addition, a clock-independent $(f, 0)$-memory strategy where $f(t) = 2t - 1$).

Hansen, Ibsen-Jensen and Koucký [4] introduces, for the Big Match (and also for any absorbing game), an ε-optimal, clock-independent, (f, γ)-memory strategy, where $f(t) = (\log t)^{O(1)}$ and γ is converging to 0.

[2] The present paper focuses on a k-memory strategy where k is a finite integer. However, the definition applies to any cardinal number k, which could be necessary for more general classes of games.

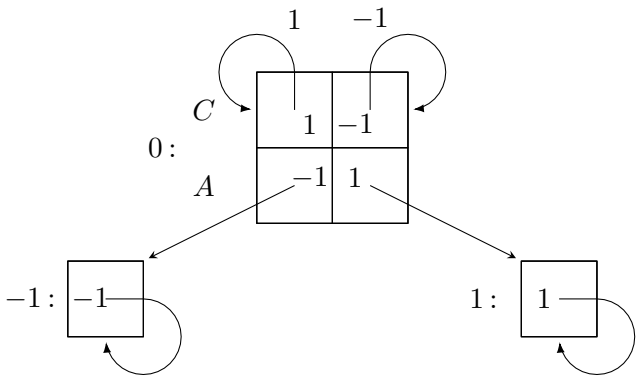

FIGURE 1. The Big Match

Fix $\varepsilon < 1$. It is known that there is neither a limiting-average nor a uniform ε-optimal strategy that is a finite-memory strategy that uses a deterministic memory-updating function σ_m; see [4] for the limiting-average case. Moreover, there is no ε-optimal mixed strategy that is a mixture of finitely many finite-memory strategies that each use a deterministic memory-updating function σ_m.

It is also known that there is neither a limiting-average nor a uniform ε-optimal strategy that is a clock-independent finite-memory strategy; see, e.g., [8] for the limiting-average case. Moreover, there is no mixed strategy that is a mixture of clock-independent finite-memory strategies that is ε-optimal [1].

3. The result The main result of the present paper is that, in the Big Match, there is a finite-memory strategy that is ε-optimal and moreover it is a two-memory strategy.

THEOREM 1. *For every $\varepsilon > 0$ there is a two-memory strategy σ of Player 1 and T_ε such that for every strategy τ of Player 2,*

$$E_{\sigma,\tau} \liminf_{T\to\infty} \bar{r}_T \geq -\varepsilon, \tag{3}$$

and

$$E_{\sigma,\tau} \bar{r}_T \geq -\varepsilon \qquad \forall T \geq T_\varepsilon. \tag{4}$$

4. The proof We label the nonabsorbing action of Player 1 by C and the absorbing action of Player 1 by A. The actions of Player 2 are labelled by -1 and 1 according to the resulting payoff to Player 1 when the nonabsorbing action is chosen by Player 1. The Big Match is depicted in Figure 1.

A pure strategy of Player 2 is identified with a sequence $x = (x_t)_{t=1}^{\infty}$ where x_t is the action of Player 2 in stage t conditional on no absorption.

In order to prove the theorem it suffices to define for every $\varepsilon > 0$ a strategy $\sigma \in \mathcal{M}_2$ of Player 1 and T_ε such that for every pure strategy x of Player 2 and $T \geq T_\varepsilon$, we have

$$E_{\sigma,x} \liminf_{T\to\infty} \frac{1}{T} \sum_{t=1}^{T} r_t \geq -4\varepsilon \tag{5}$$

and

$$E_{\sigma,x} \frac{1}{T} \sum_{t=1}^{T} r_t \geq -5\varepsilon. \tag{6}$$

Fix $0 < \varepsilon < 1$. 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 $s_{i,\varepsilon}$, or s_i for short. The number of stages in the first n epochs equals $\sum_{i=1}^{n} s_i$ and is denoted by S_n.

Two of the properties of the (to-be-defined) sequence of the number of stages in the i-th epoch, s_i, are

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

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 define a strategy $\sigma \in \mathcal{M}_2$ of Player 1 and n_ε such that for every pure strategy x of Player 2 and $n \geq n_\varepsilon$, we have

$$E_{\sigma,x} \liminf_{n\to\infty} \frac{1}{S_n} \sum_{i=1}^{n} \sum_{j=1}^{s_i} r_j^i \geq -2\varepsilon \tag{8}$$

and

$$E_{\sigma,x} \frac{1}{S_n} \sum_{i=1}^{n} \sum_{j=1}^{s_i} r_j^i \geq -4\varepsilon. \tag{9}$$

The strategy σ consists of patching together epoch strategies σ_{s_i}, which will be defined later. The epoch strategy σ_{s_i} is a strategy in the Big Match with s_i stages. It defines the strategy σ in the i-th epoch. The i-th epoch strategy σ_{s_i} depends on the number of stages s_i in the epoch and the fixed positive number ε.

The strategy σ_s will be defined for every positive integer s. However, its essential role in the proof is for large values of s.

Recall that a pure strategy of Player 2 is identified with the sequence x of actions of Player 2 when Player 1 plays constantly the nonabsorbing action. Let x_j^i be the coordinate of x that corresponds to the j-th round of the i-th epoch.

Remark. This remark serves to motivate the objectives and the desired properties of the epoch strategies σ_{s_i} and the strategy σ that is obtained by patching together the epoch strategies.

The epoch strategy σ_{s_i} tests (in an epoch that starts with the nonabsorbing state) the average $\bar{x}_i$ of x_j^i, $j = 1, \ldots, s_i$. The objective of the i-th epoch strategy σ_{s_i} is to ensure a good expected outcome if the absorbing action is played (see (1) in the next paragraph). The objective of the strategy σ is to eventually play the absorbing action if the upper density of the stages in epochs i with $\bar{x}_i \leq -\varepsilon$ is positive (see (2) in the next paragraph), and to ensure that for all sufficiently large n, the expectation of the density of the stages in epochs $i \leq n$ with $\bar{x}_i \leq -\varepsilon$ and no absorption yet is small (see (3) in the next paragraph)

The strategy σ of Player 1 controls two processes: the process of the values of the states (where the value of the nonabsorbing state is 0, the value of the losing state for Player 1 is -1, and the value of the winning state for Player 1 is 1) and the process of the actual payoffs. The strategy σ of Player 1 will guarantee (1) that an approximation of the process of the values of the states is a submartingale, and thus moves "upwards," and that the process of the payoffs obeys the following two properties. Set $\alpha_i = -\bar{x}_i$ if the i-th epoch starts in the nonabsorbing state, and $\alpha_i = 0$ otherwise. (2) for any strategy of Player 2, the fraction of the number of stages that are in epochs $i \leq n$ with $\alpha_i \geq \varepsilon$, namely, $\frac{1}{S_n}\sum_{i\leq n} s_i 1_{\{\alpha_i \geq \varepsilon\}}$, goes to 0 as $n \to \infty$, and (3) for all sufficiently large n, for any strategy of Player 2, the expectation of $\frac{1}{S_n}\sum_{i\leq n} s_i 1_{\{\alpha_i \geq \varepsilon\}}$ is $\leq \varepsilon$.

The control of the process of the values of the states along the control of the process of the actual payoffs guarantees the ε-optimality of the strategy σ. *This concludes the remark.*

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

$$v_i = \begin{cases} 0 & \text{if the state in stage } S_i+1 \text{ is the nonabsorbing state} \\ \varepsilon-1 & \text{if the state in stage } S_i+1 \text{ is the losing state for Player 1} \\ 1 & \text{if the state in stage } S_i+1 \text{ is the winning state for Player 1} \end{cases}$$

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

$$E_{\sigma,x} \liminf \frac{1}{S_n} \sum_{i=1}^{n} s_i v_{i-1} \geq v_0 - \varepsilon, \tag{10}$$

and

$$E_{\sigma,x}\, v_i \geq v_0 - \varepsilon \ \forall i. \tag{11}$$

Remark. This remark's role is to explain the necessity of conditions (10) and (11). Condition (10) is essentially necessary for the strategy σ to obey (8): one can show that if σ is a strategy of Player 1 for which there is a pure strategy x of Player 2 such that the left-hand side of inequality (10) is $< -3\varepsilon$, then there is a pure strategy x^* of Player 2 for which inequality (8) does not hold.[3] Condition (11) 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 i and a pure strategy x of Player 2 such that the left-hand side of inequality (11) is $< -4\varepsilon$, then there is a pure strategy x^* of Player 2 for which inequality (9) does not hold for all sufficiently large n. *This concludes the remark.*

In addition, the strategy σ will satisfy the following two properties. Set $\alpha_i = -\bar{x}_i$ if $v_{i-1} = 0$ (i.e., before absorption) and $\alpha_i = 0$ if $v_{i-1} \neq 0$ (i.e., after absorption). For every pure strategy x of Player 2,

$$\lim_{n\to\infty} \frac{1}{S_n} \sum_{i=1}^{n} s_i 1_{\{\alpha_i \geq \varepsilon\}} = 0 \quad P_{\sigma,x}\text{-a.e.} \tag{12}$$

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

$$E_{\sigma,x} \frac{1}{S_n} \sum_{i=1}^{n} s_i 1_{\{\alpha_i \geq \varepsilon\}} \leq \varepsilon. \tag{13}$$

Remark. This remark's purpose is to explain the role of conditions (12) and (13), and serve as an informal and sketchy proof of Lemmas 2 and 3. Note that whenever $v_{i-1} \neq 0$ the payoffs in the i-th epoch are constant: $r^i_j = 1$ if $v_{i-1} = 1$ and $r^i_j = -1 = v_{i-1} - \varepsilon$ if $v_{i-1} = \varepsilon - 1$. If $v_{i-1} = 0 = v_i$ then $-\alpha_i$ is the average of the payoffs to Player 1 in the i-th epoch. Therefore, whenever $v_{i-1} \neq 0$ or $v_{i-1} = 0 = v_i$, and $\alpha_i \leq \varepsilon$, the average of the payoffs to Player 1 in the i-th epoch is at least $v_{i-1} - \varepsilon$. Note that there is at most one epoch i such that $v_{i-1} = 0 \neq v_i$. Hence the impact of the payoffs in such an epoch is immaterial for the long-term averages of the payoffs. Therefore, if the durations of the epochs satisfy (7), a strategy σ that obeys (10) along (12) satisfies (8), and a strategy σ that obeys (11) and (13) satisfies (9). *This concludes the remark.*

We now return to the formal proof. We will prove that if the sequence (s_i) satisfies (7) then a strategy σ that satisfies (10) and (12) satisfies (8) and a strategy σ that satisfies (11) and (13) satisfies (9). We will use the following simple lemma.

LEMMA 1. *For every n and for any strategy pair, the sequence of rewards satisfies*

$$\sum_{i=1}^{n} \sum_{j=1}^{s_i} r^i_j \geq \sum_{i=1}^{n} s_i (v_{i-1} - \varepsilon) - \sum_{i=1}^{n} s_i 1_{\{\alpha_i \geq \varepsilon\}} - \sum_{i=1}^{n} s_i 1_{\{v_{i-1} \neq v_i\}}. \tag{14}$$

[3] And even the weaker inequality, $E_{\sigma,x} \limsup \frac{1}{S_n} \sum_{i=1}^{n} s_i v_{i-1} \geq v_0 - 3\varepsilon$, does not hold.

Proof. Note that

$$\sum_{j=1}^{s_i}(r_j^i - v_{i-1}) \geq \begin{cases} -\varepsilon s_i \geq -\varepsilon s_i - s_i 1_{\{\alpha_i \geq \varepsilon\}} & \text{if } v_{i-1} = \varepsilon - 1 \\ 0 \geq -\varepsilon s_i - s_i 1_{\{\alpha_i \geq \varepsilon\}} & \text{if } v_{i-1} = 1 \\ -\alpha_i s_i \geq -\varepsilon s_i - s_i 1_{\{\alpha_i \geq \varepsilon\}} & \text{if } v_i = v_{i-1} = 0 \\ -s_i = -s_i 1_{\{v_i \neq v_{i-1}\}} & \text{if } v_i \neq v_{i-1} = 0. \end{cases} \tag{15}$$

Therefore,

$$\sum_{j=1}^{s_i}(r_j^i - v_{i-1}) \geq -\varepsilon s_i - s_i 1_{\{\alpha_i \geq \varepsilon\}} - s_i 1_{\{v_{i-1} \neq v_i\}}. \tag{16}$$

Summing these inequalities over $1 \leq i \leq n$ yields inequality (14). □

Note that as $\sum_{i=1}^n 1_{\{v_i \neq v_{i-1}\}} \leq 1$, (7) implies that

$$\frac{1}{S_n}\sum_{i=1}^n s_i 1_{\{v_i \neq v_{i-1}\}} \leq \frac{s_n}{S_n} \to_{n\to\infty} 0. \tag{17}$$

Lemma 2. *If the sequence (s_i) satisfies (7) then a strategy σ that satisfies (10) and (12) satisfies inequality (8).*

Proof. Fix a pure strategy x of Player 2. As σ satisfies (10),

$$E_{\sigma,x} \liminf_{n\to\infty} \frac{1}{S_n}\sum_{i=1}^n s_i(v_{i-1} - \varepsilon) \geq v_0 - 2\varepsilon.$$

By (12), we have

$$E_{\sigma,x} \liminf_{n\to\infty} \frac{1}{S_n}\sum_{i=1}^n -s_i 1_{\{\alpha_i \geq \varepsilon\}} = 0.$$

By (17), we have

$$E_{\sigma,x} \liminf_{n\to\infty} \frac{1}{S_n}\sum_{i=1}^n -s_i 1_{\{v_i \neq v_{i-1}\}} \geq \liminf_{n\to\infty} -s_n/S_n = 0.$$

By Lemma 1, and by summing the three inequalities displayed above, we conclude that

$$\begin{aligned} & E_{\sigma,x} \liminf_{n\to\infty} \frac{1}{S_n}\sum_{i=1}^n \sum_{j=1}^{s_i} r_j^i \\ \geq\ & E_{\sigma,x} \liminf_{n\to\infty} \left(\frac{1}{S_n}\sum_{i=1}^n s_i(v_{i-1} - \varepsilon) - \frac{1}{S_n}\sum_{i=1}^n s_i 1_{\{\alpha_i \geq \varepsilon\}} - \frac{1}{S_n}\sum_{i=1}^n s_i 1_{\{v_{i-1} \neq v_i\}} \right) \\ \geq\ & E_{\sigma,x} \liminf_{n\to\infty} \frac{1}{S_n}\sum_{i=1}^n s_i(v_{i-1} - \varepsilon) + E_{\sigma,x} \liminf_{n\to\infty} \frac{1}{S_n}\sum_{i=1}^n -s_i 1_{\{\alpha_i \geq \varepsilon\}} \\ & + E_{\sigma,x} \liminf_{n\to\infty} \frac{1}{S_n}\sum_{i=1}^n -s_i 1_{\{v_{i-1} \neq v_i\}} \\ =\ & E_{\sigma,x} \liminf_{n\to\infty} \frac{1}{S_n}\sum_{i=1}^n s_i(v_{i-1} - \varepsilon) \geq v_0 - 2\varepsilon. \end{aligned}$$

□

LEMMA 3. *If the sequence (s_i) satisfies (7) then a strategy σ that satisfies (11) and (13) satisfies inequality (9).*

Proof. Assume that the strategy σ satisfies (11) and (13). Fix a pure strategy x of Player 2. By (11), we have that for every n,

$$E_{\sigma,x}\frac{1}{S_n}\sum_{i=1}^{n} s_i(v_{i-1}-\varepsilon) \geq E_{\sigma,x}\frac{1}{S_n}\sum_{i=1}^{n} s_i(v_0-2\varepsilon) = v_0-2\varepsilon.$$

By (13), we have that for every $n \geq n_\varepsilon$,

$$E_{\sigma,x}\frac{1}{S_n}\sum_{i=1}^{n} -s_i 1_{\{\alpha_i \geq \varepsilon\}} \geq -\varepsilon.$$

W.l.o.g. we assume that n_ε is sufficiently large so that $s_n/S_n < \varepsilon$ for every $n \geq n_\varepsilon$. Hence, for $n \geq n_\varepsilon$ we have

$$E_{\sigma,x}\frac{1}{S_n}\sum_{i=1}^{n} -s_i 1_{\{v_i \neq v_{i-1}\}} \geq -s_n/S_n \geq -\varepsilon.$$

By Lemma 1, and by summing the three inequalities displayed above, we conclude that for $n \geq n_\varepsilon$,

$$E_{\sigma,x}\frac{1}{S_n}\sum_{i=1}^{n}\sum_{j=1}^{s_i} r_j^i \geq v_0-4\varepsilon.$$

□

The epoch strategies σ_s. We now define the epoch strategies σ_s. If $s=1$ then σ_s plays the nonabsorbing action in the sole round of the epoch. We proceed with the definition of the strategy σ_s for $s>1$.

Let $s>1$ be an integer. We define a two-memory strategy σ_s of Player 1 in the s-stage Big Match. The two states of memory of the strategy σ_s are $\widehat{C}$ (for continuing throughout) and $\widehat{A}$ (for possible future absorption). The initial state of memory, m_1, is $\widehat{A}$.

To make several of the formulas below easier to read, we set $p=1-\varepsilon$. Recall that the payoff in stage t, r_t, is either $+1$ or -1.

We continue with the definition of the probabilistic memory-updating function. If $m_t=\widehat{A}$ and $i_t=A$, then $m_{t+1}=\widehat{C}$ (i.e., the conditional probability of $m_{t+1}=\widehat{C}$ is 1). If $m_t=\widehat{A}$, $i_t=C$, and $r_t=1$, then the conditional probability that $m_{t+1}=\widehat{A}$ is $p^2=p^{1+r_t}$. In all other cases, $m_{t+1}=m_t$. (In particular, if $m_t=\widehat{A}$, $i_t=C$, and $r_t=1$, the conditional probability that the memory state does not change, i.e., that $m_{t+1}=\widehat{A}$, is $1-p^{1+r_t}$.) Figure 2 illustrates the probabilistic memory-updating function.

We continue with the definition of the selection of the mixed action as a function of stage t and memory state m_t. The strategy plays action C if it is in memory state $\widehat{C}$. In order to define the mixed action in round j if it is in memory state $\widehat{A}$, we set

$$q_j := \varepsilon p^{s-j}.$$

Note that $q_j \geq 0$ and for $j \leq s$ we have $\sum_{k\leq j} q_k \leq \varepsilon\sum_{\ell=0}^{\infty} p^\ell = \varepsilon/(1-p)=1$. Hence, $0 \leq q_j/(1-\sum_{k<j} q_k) \leq 1$.

When the memory state of strategy σ_s in round j is $\widehat{A}$, it plays the absorbing action A (and moves to memory state $\widehat{C}$) with a conditional probability that depends on round j. This conditional probability is given by

$$\frac{q_j}{(1-\sum_{k<j} q_k)}.$$

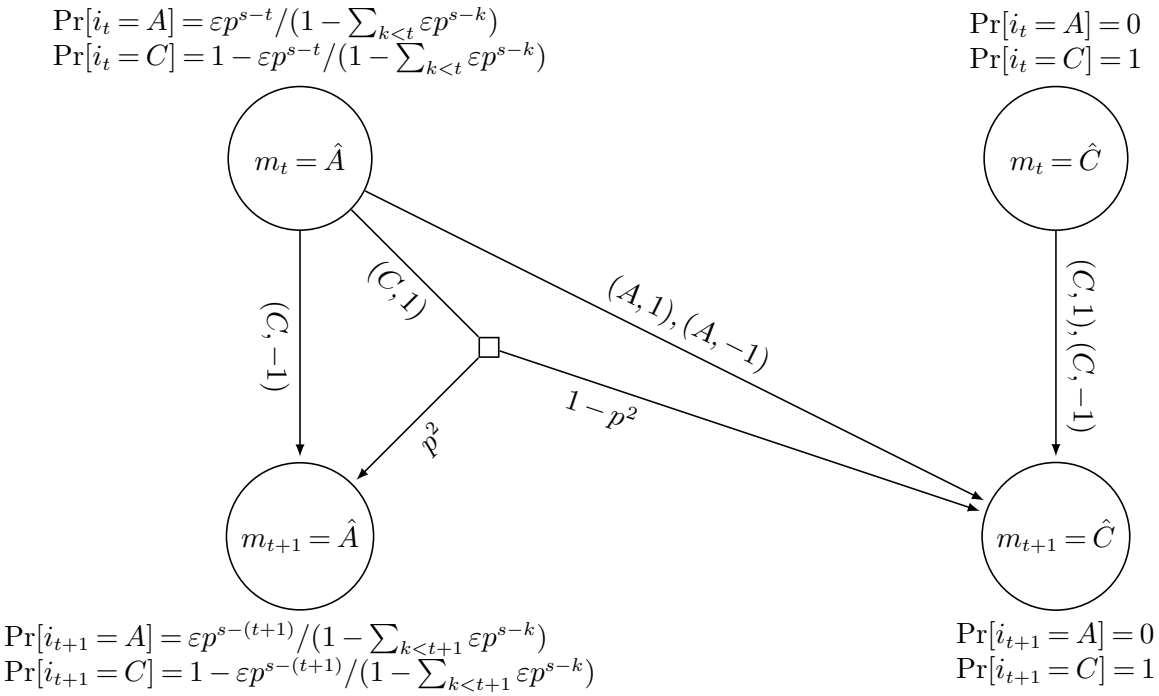

FIGURE 2. The figure illustrate how the memory is updated in the epoch-strategy. Initially, in each epoch, we have that $m_1 = \hat{A}$

Note that this conditional probability is, given parameters p and s that are specified by the strategy, a function of round j.

It follows that when it is in memory state $\widehat{A}$ it plays the nonabsorbing action C with the conditional probability (that depends on round j)

$$1 - \frac{q_j}{1-\sum_{k<j} q_k} = \frac{1-\sum_{k\le j} q_k}{1-\sum_{k<j} q_k}.$$

A pure strategy of Player 2 in the auxiliary game with $s+1$ stages is identified with the sequence $(x_k)_{1\le k\le s}$ of actions of Player 2; equivalently, with the sequence of payoffs to Player 1 conditional on no absorption. Note that if there was no absorption before round k then $x_k = r_k$. Fix a strategy $x = (x_k)_{1\le k\le s}$ for Player 2.

LEMMA 4. *Let $p_j(x)$, or p_j for short, be the (unconditional) probability that the strategy σ_s of Player 1 plays the absorbing action in stage $1 \le j \le s$. Then,*

$$p_j = \varepsilon p^{s-1+\sum_{k<j} x_k}.$$

Proof. Note that $p^2 1_{\{x_k=1\}} + p^0 1_{\{x_k=-1\}} = p^{1+x_k}$ is the conditional probability that $m_{k+1} = \widehat{A}$, given that $m_k = \widehat{A}$ and the action of Player 1 in round k is C.

Round j is the first round in which Player 1 plays the absorbing action iff the play follows the following pattern: Player 1 plays the absorbing action for the first time in the j-th round and $m_k = \widehat{A}$ for every $k \le j$.

The conditional probability, given $x_1, x_2, \ldots$, that the play follows this pattern is

$$\frac{q_j}{1-\sum_{k<j} q_j} \prod_{i<j} \frac{1-\sum_{k\le i} q_k}{1-\sum_{k<i} q_k} \prod_{k<j} (p^2 1_{\{x_k=1\}} + p^0 1_{\{x_k=-1\}})$$

$$= q_j \prod_{k<j} p^{1+x_k}$$
$$= q_j p^{j-1+\sum_{k<j} x_k} = \varepsilon p^{s-j} p^{j-1+\sum_{k<j} x_k} = \varepsilon p^{s-1} p^{\sum_{k<j} x_k}.$$

In the first equality we used the telescopic product $\prod_{i<j} \frac{1-\sum_{k\le i} q_k}{1-\sum_{k<i} q_k} = 1 - \sum_{k<j} q_j$ and the equality $p^2 1_{\{x_k=1\}} + p^0 1_{\{x_k=-1\}} = p^{1+x_k}$. □

Remark: *An alternative description of the epoch strategy.* We introduce here an alternative and equivalent definition of our epoch strategy. We do so to make the sampling aspect of our epoch strategy σ_s clearer. The alternative description is as follows: select a positive integer ℓ, where for each $1 \le j \le s$, the probability that the selected ℓ equals j is q_j. Sample the action of Player 2 in each round with probability $1-p^2$ and let the sampling of the different rounds be independent. Play the absorbing action in round j iff $j = \ell \le s$ and the payoff in each of the previously sampled rounds is -1.

Let σ_s^* be the strategy defined by the above alternative description. Let $p_j^*(x)$ be the $P_{\sigma_s^*,x}$-probability, i.e., the probability that is defined by the strategy σ_s^* of Player 1 and the pure strategy $x = (x_k)_{1\le k\le s}$ of Player 2, that the action A is played for the first time in round j. In order to show that the strategy σ_s^* equals σ_s, it suffices to show that for every pure strategy x of Player 2, $p_j^*(x) = p_j(x)$. Note that $p_j^*(x)$ equals q_j times the probability of no sampling in rounds $k<j$ with $x_k = 1$, i.e., times the product $\Pi_{k<j} p^{1+x_k}$. Hence, $p_j^*(x) = \varepsilon p^{s-j} \Pi_{k<j} p^{1+x_k} = p_j(x)$. Therefore, the strategy σ_s^* equals the strategy σ_s. *This concludes the remark.*

Consider the auxiliary games with $s+1$ stages, where dynamics and stage payoffs follow the rules of the Big Match and the players are active only in the first s stages j, $j = 1, \ldots, s$.

Let $\sigma = \sigma_s$. We study the distribution of the state in the last period, $s+1$, as a function of the strategy σ of Player 1 and a pure strategy τ of Player 2.

Let τ be a pure strategy of Player 2. Recall that we labelled the left-column action of Player 2 by -1 and the right-column action of Player 2 by 1; hence, the pure strategy τ of Player 2 is identified with the sequence of actions $x = x(\tau) = (x_1, \ldots, x_s)$.

Define a function v on plays of the auxiliary $(s+1)$-stage game as follows. If the play is absorbed in the winning state for Player 1, then $v = 1$. If the play is absorbed in the losing state for Player 1, then $v = \varepsilon - 1$, and otherwise $v = 0$.

LEMMA 5. *Let* $\alpha(x) = -\sum_{j=1}^{s} x_j/s$. *Then*

$$E_{\sigma,x} v = p^{(1-\alpha(x))s} - p^s \tag{18}$$
$$\ge p^{(1-\delta)s} 1_{\{\alpha(x)\ge\delta\}} - p^s \quad \forall \delta > 0 \tag{19}$$
$$\ge -p^s. \tag{20}$$

Proof. [4] For every integer c let J_c^+ be the set of all indices $1 \le j \le s$ such that $c = -\sum_{k<j} x_k$ and $x_j = -1$, and let J_c^- be the set of all indices $1 \le j \le s$ such that $c+1 = -\sum_{k<j} x_k$ and $x_j = 1$. Obviously, for each integer c, the sets of indices J_c^+ and J_c^- are disjoint, and the set of integers $\{1, \ldots, s\}$ is the disjoint union $\cup_c (J_c^+ \cup J_c^-)$. There is an illustration of the sets J_c^+ and J_c^- in Figure 3.

[4] The proof is broadly similar to [5, Lemma 2.6].

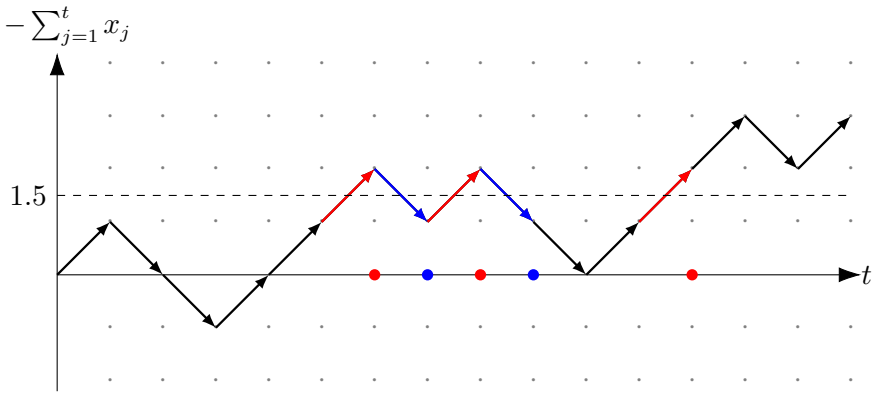

FIGURE 3. To illustrate the sets J_c^+ and J_c^- arcs are drawn between the points $(t, -\sum_{j=1}^{t} x_j)$. The set J_c^+ is then given by the arcs crossing the horizontal line at height $c+0.5$ upward and the set J_c^- is given by the arcs crossing the line downward. Shown here are the sets J_1^+ (red dots) and J_1^- (blue dots).

Obviously, if $j \in J_c^+$ and $j' \in J_c^-$ then $p^{c+1} = p^{1+\sum_{k<j} x_k} = p^{\sum_{k<j'} x_k}$. Therefore, using Lemma 4, we have

$$\begin{aligned} E_{\sigma,x} v &= \sum_{j=1}^{s} 1_{\{x_j=-1\}} p_j - \sum_{j=1}^{s} 1_{\{x_j=1\}} p_j p \\ &= \sum_{c} \sum_{j \in J_c^+} \varepsilon p^{s-c-1} - \sum_{c} \sum_{j \in J_c^-} \varepsilon p^{s-c-1}. \end{aligned}$$

Note that

$$|J_c^+| = \begin{cases} |J_c^-| & \text{if } \alpha > 0 \text{ and } \quad c \notin \{0,1,\ldots,\alpha s - 1\} \\ |J_c^-| + 1 & \text{if } \alpha > 0 \text{ and } \quad c \in \{0,1,\ldots,\alpha s - 1\} \\ |J_c^-| & \text{if } \alpha = 0 \\ |J_c^-| & \text{if } \alpha < 0 \text{ and } \; -c \notin \{1,2,\ldots,-\alpha s\} \\ |J_c^-| - 1 & \text{if } \alpha < 0 \text{ and } \; -c \in \{1,2,\ldots,-\alpha s\}. \end{cases}$$

Therefore,

$$E_{\sigma,x} v = \begin{cases} \sum_{j=0}^{\alpha s - 1} \varepsilon p^{s-j-1} = p^{(1-\alpha)s} - p^s \text{ if } \alpha > 0, \\ \\ \sum_{j=1}^{-\alpha s} \varepsilon p^{s+j-1} = p^{(1-\alpha)s} - p^s \text{ if } \alpha \le 0. \end{cases} \tag{21}$$

This completes the proof of equality (18).

The function $\alpha \mapsto p^{(1-\alpha)s}$ is nonnegative and monotonic increasing in α, and $p^{(1-\alpha)s} \ge p^{(1-\delta)s} 1_{\{\alpha \ge \delta\}} \ge 0$. Therefore, equality (18) implies inequalities (19) and (20), which completes the proof of the lemma. □

The epochs' duration. We now define the sequence (s_i), i.e., the sequence of durations of epochs. W.l.o.g. we assume that $0 < \varepsilon < 1/2$; 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

$$\sum_{i=i_\varepsilon+1}^{\infty} \frac{2}{i^{(1+\varepsilon)}} < \varepsilon. \tag{22}$$

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

LEMMA 6. *The sequence (s_i) satisfies (7).*

Proof. In short, the definition of s_i implies that the sequence s_i is nondecreasing 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 $1 < p^{-1} \le i^{1+\varepsilon}$ for every $i > 1$; hence, $s_i \ge 1$ for every $i > i_\varepsilon$, and recall that $s_i = 1$ for every $1 \le i \le i_\varepsilon$. For $i > i_\varepsilon$, $p^{-s_i} \le i^{1+\varepsilon}$ by the definition of s_i, and $i^{1+\varepsilon} < (i+1)^{1+\varepsilon}$. Hence, by the definition of s_{i+1}, we have $s_{i+1} \ge s_i$. We conclude that $1 \le s_i \le s_{i+1}$ for every i.

For $i > i_\varepsilon$, the definition of s_i implies that $p^{-s_i} \le i^{1+\varepsilon} \le p^{-s_i-1}$; hence, $\frac{1+\varepsilon}{-\ln p} \ln i \ge s_i \ge -1 + \frac{1+\varepsilon}{-\ln p} \ln i$. 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 7. *There exists a constant K such that for all positive integers i and n with $i \le n$,*

$$\frac{s_i}{S_n} \le \frac{s_n}{S_n} \le K n^{-1} \le K n^{-\varepsilon^2} i^{\varepsilon^2 - 1}. \tag{23}$$

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

For completeness, we spell out an explicit derivation of these inequalities. For $i > i_\varepsilon$, $s_i \ge -1 + \frac{1+\varepsilon}{-\ln p} \ln i \ge -1 + \frac{1+\varepsilon}{2\varepsilon} \ln i$ (using the inequality $2\varepsilon \ge -\ln(1-\varepsilon) = -\ln p$ for $\varepsilon < 1/2$); hence, for $n > 2(i_\varepsilon + 1)$, $S_n \ge \sum_{n/2-1 \le i \le n} s_i \ge -n + \frac{n}{2} \frac{1+\varepsilon}{4\varepsilon} \ln n$. For $n > i_\varepsilon$, $s_n \le \frac{1+\varepsilon}{-\ln p} \ln n \le \frac{1+\varepsilon}{\varepsilon} \ln n$ (using the inequality $\varepsilon \le -\ln(1-\varepsilon) = -\ln p$ for $0 < \varepsilon < 1$); hence, $s_n \le \frac{1+\varepsilon}{\varepsilon} \ln n \le K n^{-1} S_n$ 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 $s_n/S_n < K n^{-1}$.

□

We proceed with the definition of the strategy σ of Player 1. The strategy σ plays in the i-th epoch the strategy σ_{s_i}.

The next lemma introduces an auxiliary sequence of random variables, whose properties are used in the following lemma that show that the strategy σ obeys properties (10), (11), (12), and (13).

LEMMA 8. *The sequence of random variables $(Y_i)_{i \ge 1}$ that is defined by*

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

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

$$E_{\sigma,x}(Y_i - Y_{i-1} \mid h_i) \ge i^{\varepsilon^2 - 1} 1_{\{\alpha_i \ge \varepsilon\}} \tag{24}$$

for any history of play h_i up to the start of the i-th epoch.

Proof. By the definition of v_i and (22), $|Y_i| < 1$ and $Y_i \le v_i \le Y_i + \varepsilon$. As $\sum_{k > \max(i, i_\varepsilon)}^{\infty} \frac{2}{k^{(1+\varepsilon)}} \to_{i\to\infty} 0$, $Y_i - v_i \to_{i\to\infty} 0$.

Let x be a pure strategy of Player 2 and let $x^i = (x_1^i, \ldots, x_{s_i}^i)$ be the sequence of actions of Player 2 in epoch i assuming no absorption, and recall that

$$\alpha_i = \begin{cases} -\sum_{j=1}^{s_i} x_j^i / s_i & \text{if } v_{i-1} = 0 \text{ and } i > i_\varepsilon, \\ 0 & \text{otherwise.} \end{cases} \tag{25}$$

For $1 \le i \le i_\varepsilon$, $v_i = 0$ and $\alpha_i = 0$. Therefore, inequality (24) holds for every $1 \le i \le i_\varepsilon$. In addition, inequality (24) holds whenever $v_{i-1} \ne 0$. Hence it remains to prove that inequality (24) holds for $i > i_\varepsilon$ and $v_{i-1} = 0$.

Assume that $v_{i-1}=0$ and that $i>i_\varepsilon$. Inequality (19) along with the definition of α_i and s_i implies that for $i\geq i_\varepsilon$,

$$\begin{aligned} E_{\sigma,x}(v_i-v_{i-1}\mid h_i) &\geq p^{(1-\varepsilon)s_i}1_{\{\alpha_i\geq\varepsilon\}}-p^{s_i} \\ &\geq i^{\varepsilon^2-1}1_{\{\alpha_i\geq\varepsilon\}}-\frac{2}{i^{1+\varepsilon}} \end{aligned} \tag{26}$$

for any history of play h_i up to the start of the i-th epoch. As $Y_i-Y_{i-1}=v_i-v_{i-1}+\frac{2}{i^{1+\varepsilon}}$ for $i>i_\varepsilon+1$, we get that inequality (24) holds for every $i>i_\varepsilon$. □

LEMMA 9. *The strategy σ obeys properties (10) and (11).*

Proof. Let x 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 (24) shows that the sequence of random variables $(Y_i)_{i\geq 0}$ is a submartingale (with respect to the probability distribution $P_{\sigma,x}$ on plays). In addition, $Y_0\geq v_0-\varepsilon$ and $v_i\geq Y_i$. Therefore, $E_{\sigma,x}v_i\geq E_{\sigma,x}Y_i\geq E_{\sigma,x}Y_0\geq v_0-\varepsilon$, which proves (11).

As Y_i is a bounded submartingale, it converges a.e. to a limit Y_∞ and $E_{\sigma,x}Y_\infty\geq Y_0$. 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

$$\frac{1}{S_n}\sum_{i=1}^n s_i v_{i-1}\to_{n\to\infty}Y_\infty \ P_{\sigma,x}\text{-a.e.} \tag{27}$$

Hence, $E_{\sigma,x}\lim_{n\to\infty}\frac{1}{S_n}\sum_{i=1}^n s_iv_{i-1}=E_{\sigma,x}Y_\infty\geq Y_0\geq v_0-\varepsilon$, which proves (10). □

LEMMA 10. *The strategy σ obeys properties (12) and (13).*

Proof. Note that (as $-1<Y_i<1$) $Y_i-Y_j<2$. Taking the expectations in inequality (24), we deduce that $E_{\sigma,x}(Y_i-Y_{i-1})\geq E_{\sigma,x}\,i^{\varepsilon^2-1}1_{\{\alpha_i\geq\varepsilon\}}$. Summing these inequalities over all i such that $1\leq i\leq n$, we deduce that

$$2>E_{\sigma,x}(Y_n-Y_0)\geq E_{\sigma,x}\sum_{i=1}^n i^{\varepsilon^2-1}1_{\{\alpha_i\geq\varepsilon\}}. \tag{28}$$

By the monotone convergence theorem, inequality (28) implies that $2\geq E_{\sigma,x}\sum_{i=1}^\infty i^{\varepsilon^2-1}1_{\{\alpha_i\geq\varepsilon\}}$. Hence, $\sum_{i=1}^\infty i^{\varepsilon^2-1}1_{\{\alpha_i\geq\varepsilon\}}$ is finite a.e. Hence, using (23), for every pure strategy x of Player 2,

$$0\leq\frac{1}{S_n}\sum_{i=1}^n s_i1_{\{\alpha_i\geq\varepsilon\}}\leq K\,n^{-\varepsilon^2}\sum_{i=1}^n i^{\varepsilon^2-1}1_{\{\alpha_i\geq\varepsilon\}}\to_{n\to\infty}0\ \ P_{\sigma,x}\text{-a.e.}, \tag{29}$$

which proves (12).

We proceed to prove (13). Let n_ε be a sufficiently large integer so that $K\,n_\varepsilon^{-\varepsilon^2}<\frac{\varepsilon}{2}$. Hence, $\frac{s_i}{S_n}\leq i^{\varepsilon^2-1}\varepsilon/2$ for every $n\geq n_\varepsilon$ and $i_\varepsilon<i\leq n$. Then, using inequality (28), we have

$$E_{\sigma,x}\frac{1}{S_n}\sum_{i=1}^n s_i1_{\{\alpha_i\geq\varepsilon\}}\leq E_{\sigma,x}\sum_{i=1}^n i^{\varepsilon^2-1}1_{\{\alpha_i\geq\varepsilon\}}\varepsilon/2<\varepsilon\ \ \forall n\geq n_\varepsilon, \tag{30}$$

which proves (13).

□

5. Open problems The main open problem is whether or not in any stochastic game each player has a finite-memory strategy that is ε-optimal.

In the remainder of this section we introduce several additional open problems. These open problems are of independent interest and a few of them may turn out to be building blocks toward the solution of the main open problem.

5.1. Private versus public memory states The ε-optimal two-memory strategy in our proof uses private memory states (i.e., Player 2 does not observe the memory states of the strategy of Player 1).

We say that the memory states m_t are *public* if they are observed by all players and a strategy is *public* if the memory states are. For example, the memory states of the Blackwell and Ferguson [2] strategy in the Big Match, which are the possible differences between the number of odd and even guesses of Player 2, are public. Also, the Mertens and Neyman [7] ε-optimal strategies are public in a stochastic game, and so are any memory-based strategy in which the memory-updating functions are deterministic. The ε-optimal strategies that are introduced in Hansen et al. [4] are not public.

All the above-mentioned ε-optimal strategies are memory-based strategies with an infinite set of memory states. One can generalize the proof of [4, Theorem 6] to show that in the Big Match any public finite-memory strategy is worthless.

A natural question that arises is what is the minimal size of the public memory (as a function of t) that is needed for an ε-optimal strategy in a stochastic game. In order to state this problem formally, we recall the reader of the notion of an (f,γ)-memory strategy, where $f : N \to N$ is a nondecreasing function and $\gamma : N \to [0,1]$ is a nonincreasing function.

An (f,γ)-memory strategy is a memory-based strategy σ whose set of memory states is N and for every strategy τ and positive integer $t' \in N$, the $P_{\sigma,\tau}$-probability that $m_t \leq f(t)$ for all $t \geq t'$ is at least $1-\gamma(t')$.

The question is then for which functions f,γ does there exists a public (f,γ)-memory strategy for the Big Match and for stochastic games in general?

Note that we distinguish between a public finite-memory strategy and a mixed strategy that is a mixture of such strategies. In fact, a general mixing principle implies that in any stochastic game, any k-memory strategy (even if all memory states are private) is equivalent to a mixed strategy that is a mixture of (uncountably many) public k-memory strategies.

This principle follows from the following construction of a mixture of public k-memory strategies. Let σ be a k-memory strategy with memory states $(m_t)_{t\in\mathbb{N}}$, action function σ_α, and memory-updating function σ_m. For any sequence of permutations of $[k] := \{1,\ldots,k\}$, $\pi = (\pi_t)_{t=1}^{\infty}$, we define the public k-memory strategy $\pi\sigma$ that follows strategy σ and that makes public the memory states renamed according to π_t in round t, for each t.

The mixture of $\pi\sigma$, where the sequence of random permutations π_t, $t = 1,2\ldots$, is a sequence of i.i.d. permutations of $[k]$ and each π_t is uniformly distributed over all $k!$ permutations, is equivalent to the k-memory strategy σ.

5.2. Recall-based strategies The definitions in this section apply to a general stochastic game. A few of the open problems in this section concern some specific stochastic game.

A *recall-based strategy* is a memory-based strategy in which the memory state m_t is simply an encoding of $z_{t-k_t}, i_{t-k_t}, j_{t-k_t}, \ldots, z_{t-1}, i_{t-1}, j_{t-1}, z_t$, where $k_t < t$. As it is a memory-based strategy it follows that $k_{t+1} \leq k_t + 1$. A *k-recall strategy* is a recall-based strategy where the recall size k_t equals k. A *finite-recall strategy* is a k-recall strategy for some fixed finite k.

In a recall-based strategy the memory states are public and the memory-updating function is deterministic. Therefore, it follows from [4, Theorem 6] that in the Big Match, Player 1 has no worthy strategy that is a finite-recall strategy.

A natural question that arises is, what is the minimal recall (as a function of t) that is needed for an ε-optimal strategy in a stochastic game? In order to state this problem formally, we introduce the concept of f-recall strategies, where $f : \mathbb{N} \to \mathbb{N}$ is a nondecreasing function with $f(t) < t$ and $f(t+1) \leq f(t)+1$.

An *f-recall strategy* is a memory-based strategy in which the memory state m_t is an encoding of $z_{t-f(t)}, i_{t-f(t)}, j_{t-f(t)}, \ldots, z_{t-1}, i_{t-1}, j_{t-1}, z_t$.

The question that arises is, what are the functions f for which there is an f-recall strategy that is ε-optimal? The question applies to a general stochastic game as well as to the special case of the Big Match.

It is worthwhile to note that the ε-optimal strategy in the Big Match that is introduced in the present paper is an f-recall strategy with $f(t) \leq \frac{K \log t}{\varepsilon}$ for some positive constant K.

Remark. A counterpart of a recall-based strategy is a strategy with information lag. A strategy with information lag corresponds to a strategy in the model where a player observe the actions of her opponent after some time-dependent delay. An *f-delay strategy*, where $f : \mathbb{N} \to \mathbb{N}$ is a nondecreasing function with $f(t) < t$ and $f(t+1) \leq f(t) + 1$, specifies the action in stage n as a function of the state in stage n and the history of the play up to stage $n - f(n)$.

Levy [6, Theorem 3.1.1] assets that for any finite stochastic game and $\varepsilon > 0$, for any $\beta > 1$ Player 1 has an ε-optimal strategy which is an f-delay strategy, where $f(n) = O(\frac{n}{(\log n)^\beta})$, and [6, Proposition 3.2.2] asserts that if $\frac{n}{\log(\log n)} = o(f(n))$, then, for every $\varepsilon > 0$, Player 2 has a strategy τ_ε such that for every f-delay strategy σ of Player 1, $E_{\sigma,\tau_\varepsilon} \liminf_{n\to\infty} \frac{1}{n} \sum_{t=1}^{n} r_t \leq \varepsilon - 1$.

Acknowledgments. We are grateful to Elon Kohlberg and anonymous reviewers for their helpful comments on an earlier version of the paper.

References

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

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

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

[4] Hansen KA, Ibsen-Jensen R, Koucký M (2016) The big match in small space - (extended abstract). Gairing M, Savani R, eds., *Proceedings of 9th International Symposium on Algorithmic Game Theory, SAGT 2016*, volume 9928 of *Lecture Notes in Computer Science*, 64–76 (Springer).

[5] Kohlberg E (1974) Repeated games with absorbing states. *The Annals of Statistics* 2(4):724–738.

[6] Levy Y (2012) Stochastic games with information lag. *Games and Economic Behavior* 74(1):243 – 256.

[7] Mertens J, Neyman A (1981) Stochastic games. *Int. J. of Game Theory* 10(2):53–66.

[8] Sorin S (2002) *A First Course on Zero Sum Repeated Games* (Springer).