The Big Match with a Clock and a Bit of Memory

Kristoffer Arnsfelt Hansen, Aarhus University
Rasmus Ibsen-Jensen, University of Liverpool
Abraham Neyman, Hebrew University

The Big Match (Gillette, 1957)

- Game played in stages.
- Player 1 may *Continue* (C) or *Absorb* (A).
- Player 2 can play 1 or -1.
- If player 1 plays Continue, he receives as payoff the action of Player 2 in the current stage.
- If player 1 plays Absorb, he receives as payoff the negative of the action of Player 2 in the current stage and **all future stages**.

Matrix representation as **absorbing game**:

	1	-1
C	1	-1
A	-1*	1*

Stochastic Games

A finite two-person zero-sum stochastic game Γ is given by:

- A finite state space, Z.
- A pair of finite actions spaces, I and J.
- A payoff function,

$$r : Z \times I \times J \to \mathbb{R}\,.$$

- A transition function,

$$p : Z \times I \times J \to \Delta(Z)\,.$$

The Big Match is a stochastic game with 3 states (of which only one is non-absorbing).

How to Play the Big Match

Assume the number of stages is finite and fixed in advance.

	1	-1
C	1	-1
A	-1*	1*

Player 2 – easy:

Play each action with probability $1/2$.

Player 1 – (relatively) easy:

Play A with probability $^1/_{k+1}$ when k stages remain.

Observation:

- Strategy for Player 1 is unique.
- If number of stages is unknown or infinite, Player 1 has no optimal strategy!

Does Sunk Cost Principle Apply?

- Any strategy for Player 1 must make a decision about the stopping stage.
- Specifying a probability of each stage being stopping (Markov strategy) in advance is **worthless**: Cannot guarantee more than $-1 + \varepsilon$ points on average, for any $\varepsilon > 0$.

- Principle of sunk suggest that optimizing from the current stage onwards should be independent of the past.

Blackwell and Ferguson, 1968

There exit ε-optimal strategies for Player 1!

- At stage t, let $k_t = \sum_{j=1}^{t-1} r_j$ be the sum of rewards in previous stages.
- Play the absorbing action with (conditional) probability

$$\frac{1}{(N + k_t)^2}$$

where $N = N_\varepsilon$ is sufficiently large.

Implementing the strategy:

- Infinitely many **memory states** required, but can be made **public**.
- Does not depend on the current stage number (the **clock**).

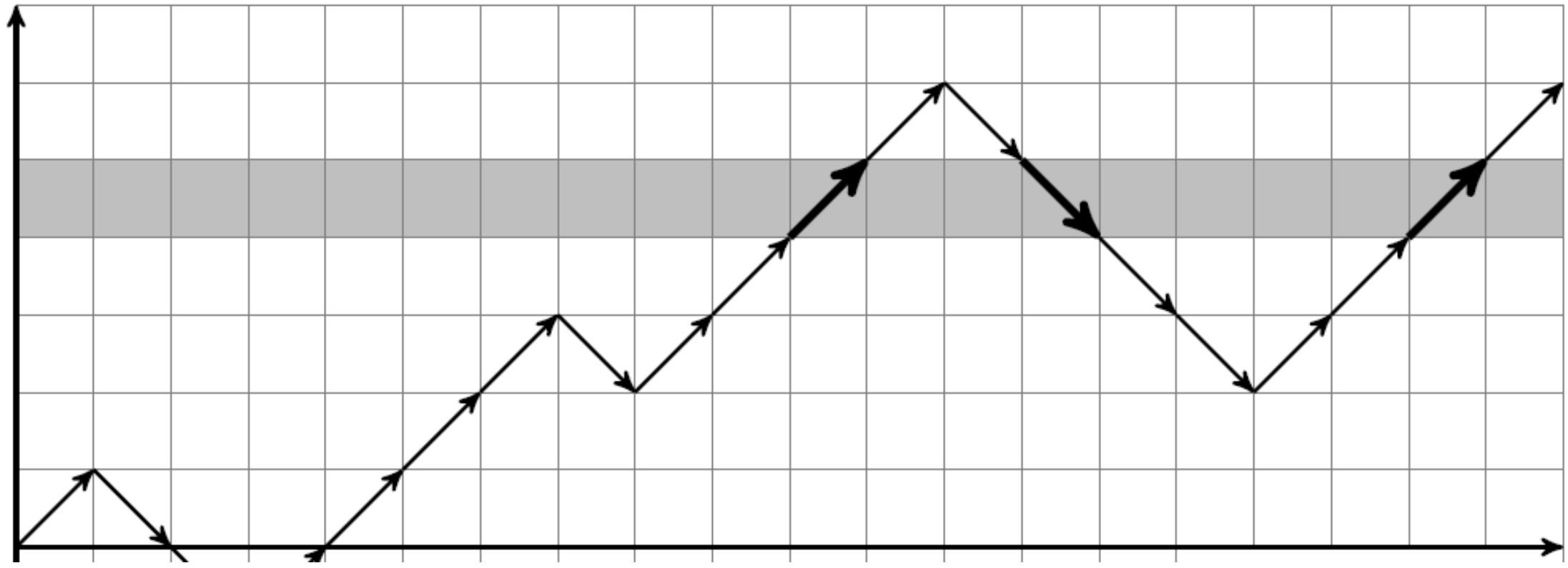
Blackwell and Ferguson, 1968

Alternative strategy:
In stage t play absorbing action with probability

(conditioned on reaching stage t without absorbing)

$\epsilon^2(1-\epsilon)^{\max(k_t,0)}$, $k_t = \sum_{j=1}^{t-1} r_j$

where r_i is the reward in stage i.



ε-optimal strategies in zero-sum stochastic games

Strategy σ_ε can satisfy:

- For all τ and $n \geq n_\varepsilon$: $E_{\sigma_\varepsilon,\tau}\bar{r}_n \geq v - \varepsilon$

Uniform ε-optimal

- For all τ: $E_{\sigma_\varepsilon,\tau} \liminf_{n\to\infty} \bar{r}_n \geq v - \varepsilon$

Limiting average ε-optimal

- For all τ: $E_{\sigma_\varepsilon,\tau} \limsup_{n\to\infty} \bar{r}_n \geq v - \varepsilon$

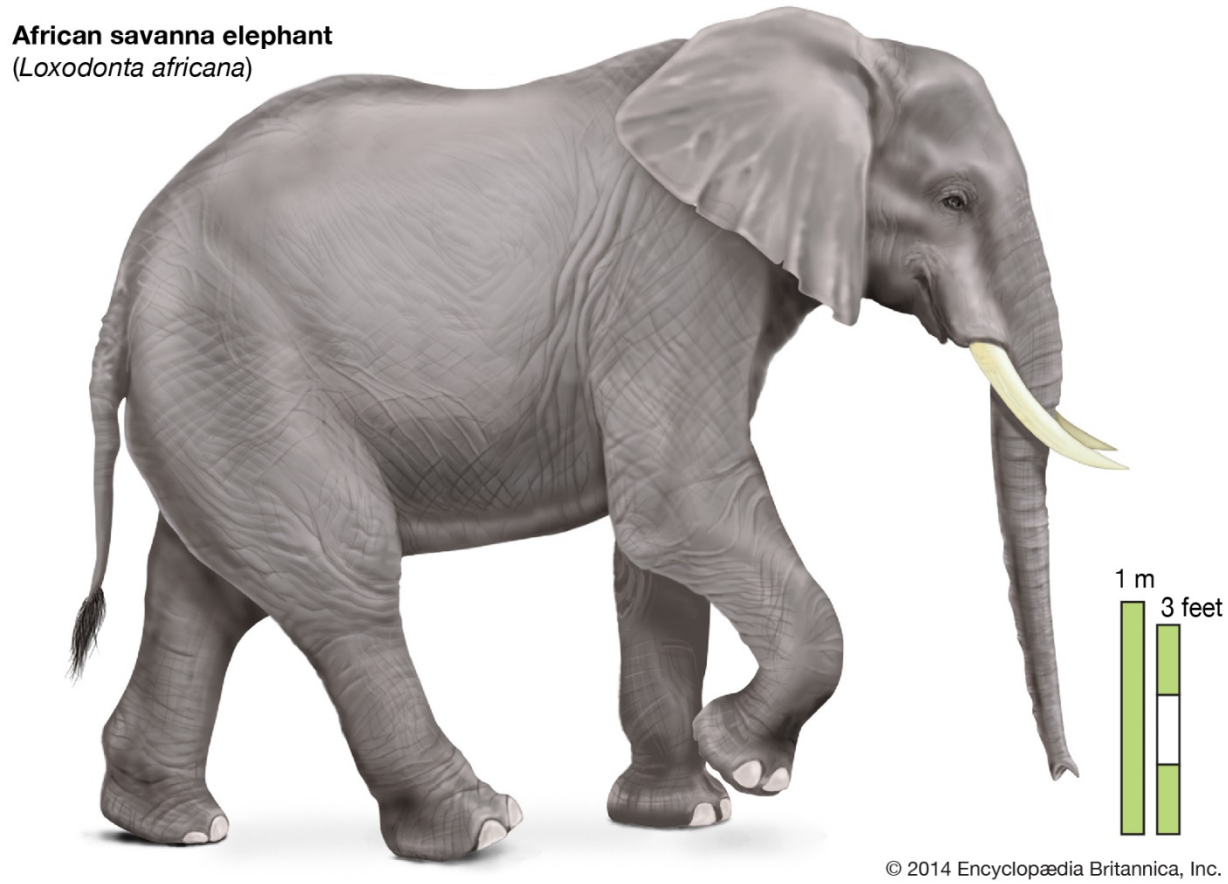
ε-optimal strategies: both uniform and limiting average ε-optimal.

Games that have ε-optimal strategies:

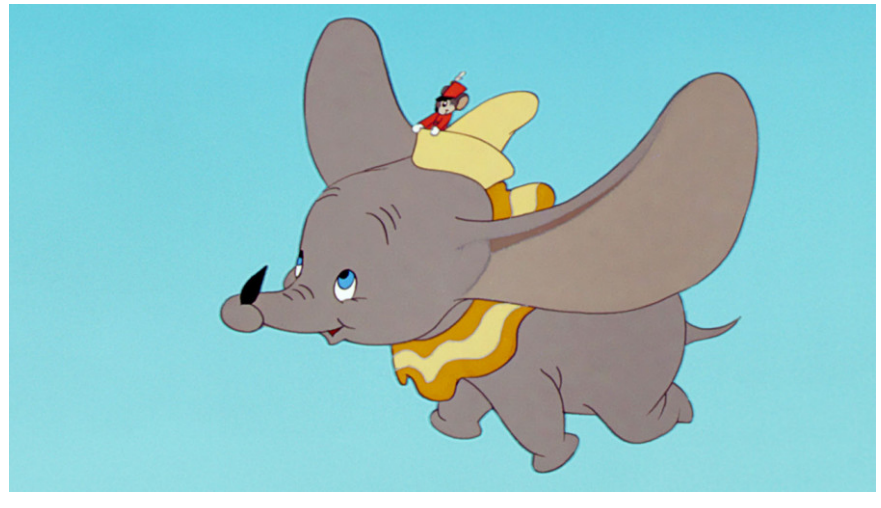
- All finite absorbing games (Kohlberg)
- All finite stochastic games (Mertens and Neyman).
- Absorbing games with compact action spaces (Mertens, Neyman, and Rosenberg)

The strategies depend on the entire history

- Who can play such strategies??



- Hansen, Ibsen-Jensen, Koucký (2016):

What about  ?

Memory-based Strategies

Strategy is k-memory if $m_t \in$ size k set.

A state of memory $m_t \in \mathbb{N}$ is maintained and updated by a pair of functions.

- σ_m (memory update function).
  - Input: t, m_t, c
  - Output: Distribution of m_{t+1}.

Update is *deterministic* if output is Dirac measure.

- σ_a (action function).
  - Input: t, m_t
  - Output: Probability of absorbing action.

The use of t determines clock-dependence/independence.

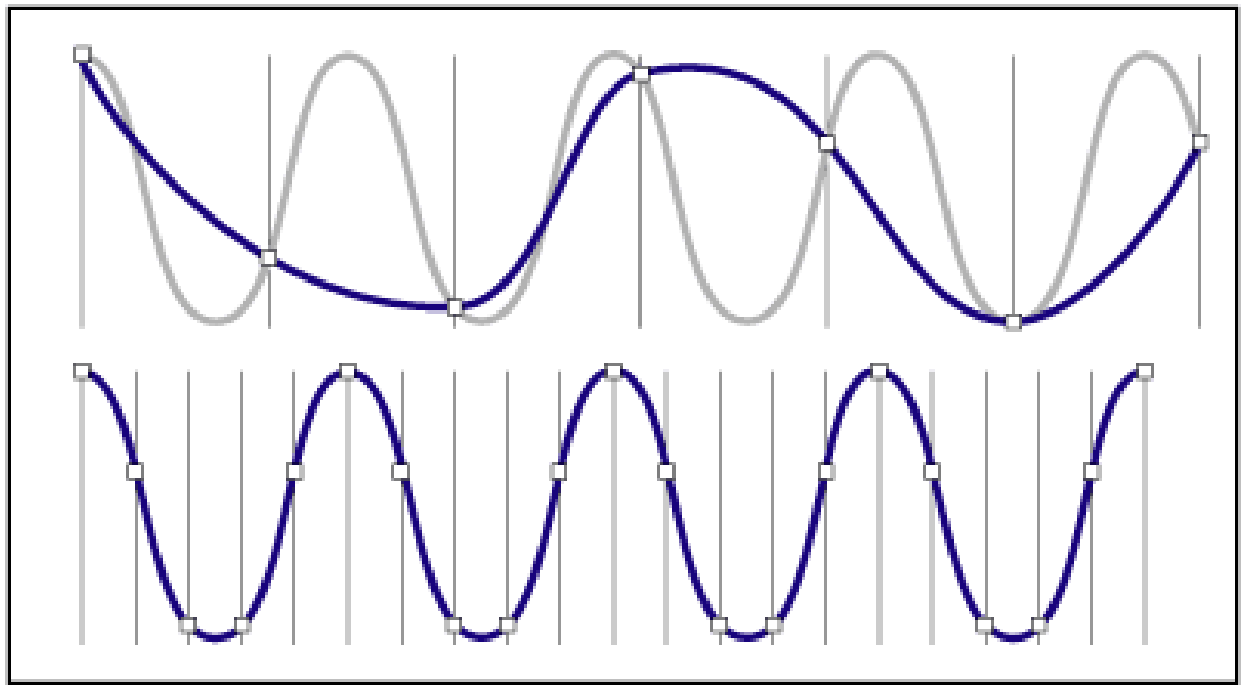
Small-space strategies (Hansen, Ibsen-Jensen, Koucký)

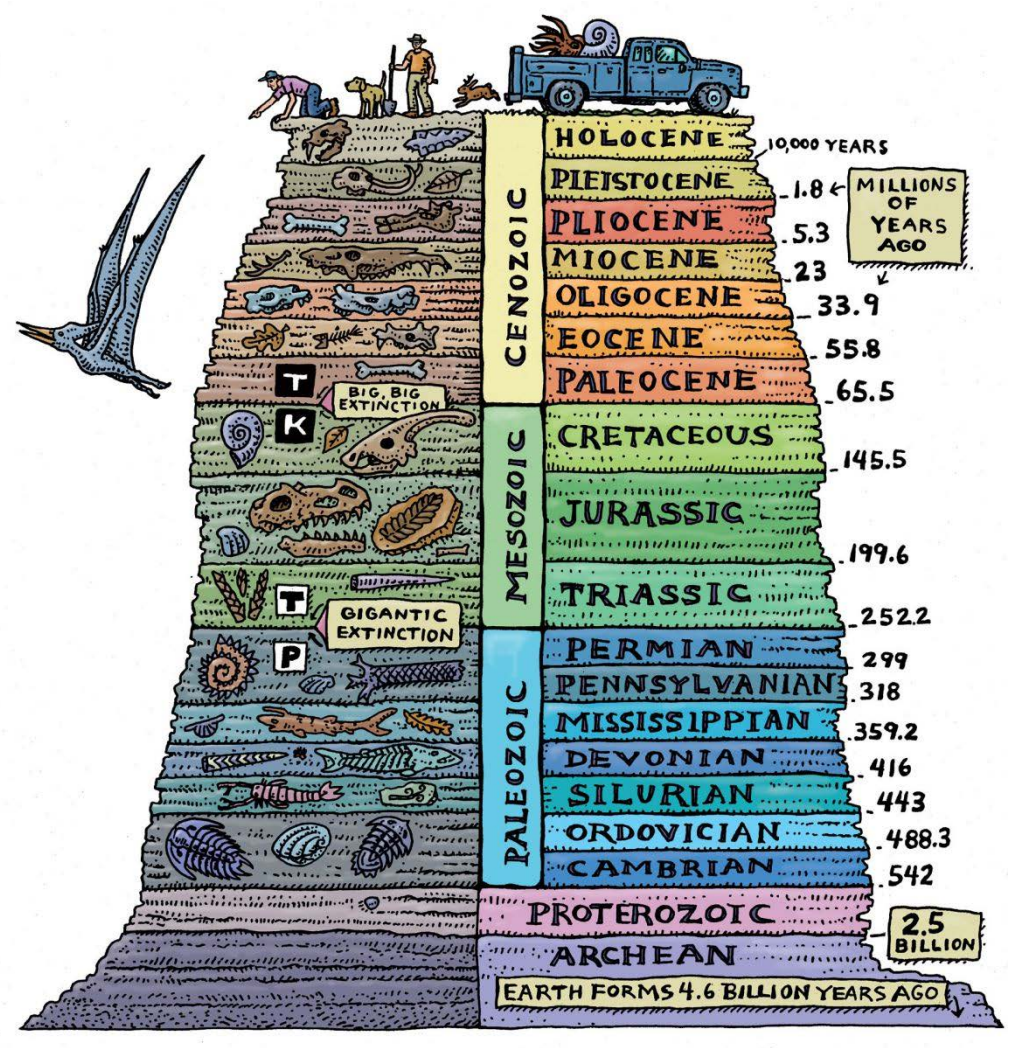
Theorem For all $\varepsilon > 0$, there is a limiting average ε-optimal strategy in the Big Match that for any $\delta > 0$ with probability $\geq 1 - \delta$ the uses $(\log n)^{O(1)}$ memory states in round n. (i.e. $\mathrm{O}(\log \log n)$ bits of memory).

Theorem For all $\varepsilon > 0$, and **any** non-decreasing unbounded function s, there is a limiting average **limsup** ε-optimal strategy in the Big Match that for any $\delta > 0$ with probability $\geq 1 - \delta$ the uses $O(s(n))$ memory states in round n.

Theorem No good finite-memory clock-dependent strategy with **deterministic** memory update function.

Two main ideas of the strategies





Sampling and Epoching

- Subsample the sequence of actions observed with a fixed sample rate.

- Send the samples to an **inner** Blackwell & Ferguson style strategy and take actions accordingly. Play Continue when not sampling.

- Split stages into epochs, restarting the inner strategy each time, and adjust sample rate to the desired expected epoch length. (Epoch terminates after a determined number of samples).

- For first theorem, sample rate is high enough to detect large dips in the average reward, and make the inner strategy react. In the second theorem large dips can go unnoticed due to low sample rate.

Importance of the Clock

Playing the Big Match with a Clock and a Bit of Memory

Theorem

For all $\varepsilon > 0$, there is a (clock-dependent) 2-memory strategy σ that is optimal for Player 1.

That is, there exist n_ε such that for any strategy τ of Player 2:

1.

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

2.

$$E_{\sigma,\tau}\, \bar{r}_n \geq -\varepsilon, \qquad n \geq n_\varepsilon \ ,$$

where $\bar{r}_n$ is the average payoff in the first n stages.

Comparison with Negative Results

The following strategies are worthless in the Big Match:

- Stationary strategies (Gillette, 1957).
  (=1-memory, clock-independent)
- Markov strategies (Blackwell and Ferguson, 1968).
  (=1-memory, clock-dependent)
- Finite memory, clock-independent strategies (Amitai, 1989).
- Finite memory, deterministic memory update (Hansen, Ibsen-Jensen, and Koucký, 2016).

The Strategy

- Split stages into epochs, with i-th epoch of size
$$s_i \approx (1+\epsilon)\log_{1/(1-\epsilon)} i$$
- In round t of epoch j play absorbing action with probability (conditioned on reaching epoch j)
$$\epsilon(1-\epsilon)^{s_j+\sum_{i=1}^{t-1} r_i}$$

where r_i is the outcome in round i of epoch j.

Crucial insight: This can be done with 1 bit together with the clock!

Comparison to Blackwell and Ferguson

- Blackwell and Ferguson:
  In stage t play absorbing action with probability (conditioned on reaching stage t)
  $$\epsilon^2(1-\epsilon)^{\max(\sum_{i=1}^{t-1} r_i, 0)}$$
  where r_i is the outcome in stage i.

- Our strategy:
  In round t of epoch j play absorbing action with probability (conditioned on reaching epoch j)
  $$\epsilon(1-\epsilon)^{s_j+\sum_{i=1}^{t-1} r_i}$$
  where r_i is the outcome in round i of epoch j.

Implementation with 2 Memory States

Goal: Play absorbing action with probability $\epsilon(1-\epsilon)^{s_j+\sum_{i=1}^{t-1} r_i}$

Memory states: $\hat{C}$ and $\hat{A}$ ("continuing" and "possible absorption")

Memory update function σ_m:

- In memory state $\hat{A}$, switch to $\hat{C}$ with probability $1-(1-\epsilon)^2$ if round payoff is 1.

Action function σ_a:

- In memory state $\hat{C}$, play non-absorbing action.
- In memory state $\hat{A}$, play absorbing action with round dependent probability.

3 equivalent views of the strategy

1. Start in $\hat{A}$. When $m_t = \hat{A}$ and $r_t = 1$ let to $m_{t+1} = \hat{C}$ with probability $1 - (1-\varepsilon)^2$. Play action A with probability $q_t/(1 - \sum_{i<t} q_i)$ when $m_t = \hat{A}$, where $q_i = \varepsilon(1-\varepsilon)^{s_j - i}$.

2. Choose a potential stopping time ℓ, such that $\ell = i$ with probability q_i. Sample each action of Player 2 with probability $1 - (1-\varepsilon)^2$. In round ℓ play the absorbing action if and only if all sampled actions were -1.

3. In round t play absorbing action with unconditional probability $\epsilon(1-\epsilon)^{s_j + \sum_{i=1}^{t-1} r_i}$

Further Results

ε-optimal strategies with finite memory for:

- Zero-sum absorbing games with finite or compact actions sets.

- Zero-sum repeated games with symmetric incomplete information with either deterministic or random signals.

Open Problems

- Minimal size of public memory needed for ε-optimal strategies?

- Minimal *recall* needed for ε-optimal strategies?

- Generalization to stochastic games?