Mean-Payoff-Parity and Lifting Strategies from MDPs to 2-Player Stochastic Games

Mohan Dantam
Department of Computer Science, University of Oxford, UK

Richard Mayr
School of Informatics, University of Edinburgh, UK

Abstract

We consider the strategy complexity (i.e., memory and randomization) of optimal strategies in turn-based 2-player zero-sum stochastic games. Results in [25, 34] show how to lift optimal memoryless strategies for shift-invariant inverse-submixing objectives from MDPs to 2-player stochastic games with an exponential increase in the number of memory modes. We show the corresponding lower bound, i.e., the extra exponential memory is required in general, even for randomized strategies.

Moreover, we solve the strategy complexity of the well-studied mean-payoff-parity objective ($\texttt{MP} > 0 \cap \texttt{EPAR}$) in 2-player stochastic games. This objective is also shift-invariant inverse-submixing, but easier than the worst case for this class. In MDPs, Maximizer has optimal *memoryless randomized* strategies, while optimal *deterministic* strategies require *exponential* memory. However, in stochastic games, optimal *randomized* strategies require, at least and at most, *linear memory* (equal to the number of even colors).

Finally, we show that the different construction in [28, 3] for lifting memoryless (resp. finite-memory) *deterministic* strategies from MDPs (resp. 1-player games) to 2-player games cannot be generalized even to memoryless *randomized* strategies. We construct a shift-invariant objective where Max and Min each have optimal memoryless randomized strategies in all MDPs, but optimal (randomized) Max strategies still require *infinite* memory in deterministic 2-player games.

2012 ACM Subject Classification Computing methodologies → Stochastic games

Keywords and phrases MDPs, Stochastic Games, Parity, Mean-payoff, Strategy complexity

Digital Object Identifier 10.4230/LIPIcs.CONCUR.2026.30

Related Version *Full Version*: `https://arxiv.org/abs/2606.19324` [19]

1 Introduction

Background on strategy complexity. We consider 2-player turn-based perfect information stochastic games (here denoted as SGs) played on finite graphs. They are also called $2\frac{1}{2}$-*player games* [14, 13] or *competitive Markov decision processes* [24]. Introduced by Shapley [38] in 1953, they have since played a central role in the solution of many problems, e.g., synthesis of reactive systems [37, 36] and formal specification and verification [22, 23, 1]. Every state either belongs to one of the players (Max or Min) or is a random state. In each round of the game the player who owns the current state chooses the successor state along the game graph and in random states the successor is chosen according to a predefined distribution. Given a start state and strategies of Max and Min, this yields a distribution over induced infinite plays. Objectives are defined via a measurable reward function that assigns rewards to plays (possibly just an indicator function of a measurable set), and the players try to maximize (resp. minimize) the expected reward.

We consider the *strategy complexity*, i.e., the required memory and randomization of optimal strategies in SGs for a given objective. In particular we study the question how the strategy complexity in SGs compares to the strategy complexity in Markov decision processes (MDPs).


© Mohan Dantam and Richard Mayr;
licensed under Creative Commons License CC-BY 4.0

Lifting strategies from MDPs to SGs for shift-invariant inverse-submixing objectives. *Lifting* strategies from MDPs to SGs means adapting them to work even in the more general context of games with a strategic opponent [25, 34]. Results about lifting typically address the strategy complexity, e.g., if optimal Max strategies for a certain objective require only memory of size x in MDPs then optimal Max strategies in SGs require only memory of some size $y > x$, for a certain y, possibly depending on the objective and on the size/structure of the SG.

For shift-invariant inverse-submixing objectives, optimal and ε-optimal finite-memory (randomized or deterministic) Max strategies can be lifted from MDPs to SGs with a *doubly exponential increase* in the number of memory modes (using n extra bits of public memory in binary-branching games where n states are Min-controlled plus an additional $2^n \cdot p$ bits where p bits are required for optimal strategies in MDPs) [25, Theorem 1.2]. (Recall that x bits of memory means 2^x memory modes.) A restricted subcase of this was observed in [34, Theorem 6].

Contribution 1: When memoryless randomized strategies suffice for Max in MDPs, the lifting above incurs an exponential increase in the number of memory modes ($p = 0$, n extra bits). In Section 3, we show the corresponding exponential lower bound for the step from MDPs to games (even deterministic games) in this case. This partly solves the open question in [25, end of Sec. 6.1]. We construct a class of examples for the multi-dimensional $\mathtt{MP} > \vec{0}$ objective (which is shift-invariant inverse-submixing). Though Max has optimal memoryless randomized strategies in all MDPs, optimal Max strategies in deterministic 2-player games require an exponential number of memory modes, even if the strategies may use randomization and even if Max's memory is private (hidden from the Min player).

Strategy complexity of mean-payoff-parity. The mean-payoff-parity objective ($\mathtt{MP} > 0 \cap \mathtt{EPAR}$) is to attain a strictly positive mean payoff while also satisfying a parity objective. This combines quantitative performance with a qualitative correctness condition (encoded by parity), and it has frequently been studied, e.g., [20, 8, 12, 9, 11]. $\mathtt{MP} > 0 \cap \mathtt{EPAR}$ is shift-invariant inverse-submixing, and hence suitable for the lifting of Max strategies from MDPs to SGs as in [25, Theorem 1.2]. With deterministic strategies, Max already requires an exponential number of memory modes even in MDPs [26, Fig. 1], and hence the extra memory used in the lifting construction ($n + 2^n \cdot p$ extra bits of public memory, where p grows polynomially in the size of the game) yields a doubly exponential upper bound on the number of memory modes required in SGs.

Contribution 2: We show that the situation is different for randomized strategies. In Section 4 we show that Max has optimal *memoryless randomized* strategies in MDPs. However, the lifting construction of [25, Theorem 1.2] would still yield exponentially many memory modes for Max strategies in SGs. We show that optimal Max strategies for $\mathtt{MP} > 0 \cap \mathtt{EPAR}$ in SGs (and also in deterministic games) require, at least and at most, just *linearly* many memory modes, equal to the number of even colors. I.e., $\mathtt{MP} > 0 \cap \mathtt{EPAR}$ is easier than the exponential worst case for the lifting construction demonstrated in our lower bound in Section 3. This is an interesting example where randomization in strategies drastically reduces the amount of memory required, but does not eliminate the need for memory entirely.

Note however that none of this applies to the *different* mean-payoff-parity objective with a *non-strict* inequality $\mathtt{MP} \geq 0 \cap \mathtt{EPAR}$. Optimal strategies for this objective require infinite memory even in deterministic 1-player games (and thus also in MDPs/SGs) [12], due to a pathological case where infrequent negative rewards at increasing intervals (e.g., -1 at times 2^n for all $n \in \mathbb{N}$ and rewards zero otherwise) can still yield a mean-payoff of zero.

Lifting deterministic strategies from MDPs to SGs. A different lifting construction with orthogonal preconditions was described in [28, Theorem 9]. Given an objective, if the players have optimal memoryless *deterministic* strategies in all maximizing (resp. minimizing) MDPs, then they also have optimal memoryless deterministic strategies in all SGs. Under mild assumptions, this can be generalized to finite-memory deterministic strategies [3, Theorem 4.1] (in stochastic or non-stochastic arenas).

Contribution 3: We show that such results do *not* generalize to *randomized* strategies, even under very strong assumptions. In Section 5 we present stronger counterexamples than the ones given in [3, 39]. E.g., we construct a shift-invariant objective where Max and Min each have optimal memoryless randomized strategies in all MDPs, but optimal (randomized) Max strategies still require *infinite* memory in deterministic 2-player games.

2 Preliminaries

Let $\mathbb{Z}$ (resp. $\mathbb{Z}_+, \mathbb{N}, \mathbb{Q}$) denote the set of integers (resp. positive, non-negative integers, rationals). The "size" of an integer n or a rational $q = \frac{m}{n}$ (where $\gcd(m,n) = 1$) is defined in a natural way assuming binary representation. Let $\|n\| \stackrel{\text{def}}{=} \lceil \log_2(|n|+1) \rceil + 1$ and $\|q\| \stackrel{\text{def}}{=} \|m\| + \|n\| + 1$. A *probability distribution* over a countable set S is a function $f : S \to [0,1]$ with $\sum_{s \in S} f(s) = 1$. Let $\mathsf{supp}(f) \stackrel{\text{def}}{=} \{s \mid f(s) > 0\}$ denote the support of f and $\mathcal{D}(S)$ is the set of all probability distributions over S. If S is finite and $f(S) \subseteq \mathbb{Q}$, we define the *bit size* of f to be $\text{bits}(f) \stackrel{\text{def}}{=} \sum_{s \in S}(\|f(s)\|)$. Likewise for probabilistic functions with rational probabilities $p : S \to \mathcal{D}(T)$ with finite S and T: $\text{bits}(p) \stackrel{\text{def}}{=} \sum_{s \in S} \text{bits}(p(s))$.

Games, MDPs and Markov chains. We consider finite-state 2-player turn-based perfect-information stochastic games (here abbreviated as SGs) $\mathcal{G} = (S, (S_\Box, S_\Diamond, S_\bigcirc), E, P)$ where the finite set of states S is partitioned into the states $S_\Box$ of the player $\Box$ (*Maximizer*, for short Max), states $S_\Diamond$ of player $\Diamond$ (*Minimizer*, for short Min), and chance vertices (aka random states) $S_\bigcirc$. If $S_\bigcirc = \emptyset$ then it is called a deterministic 2-player game. Let $E \subseteq S \times S$ be the transition relation. We write $s \to s'$ if $(s,s') \in E$ and assume that $\mathsf{Succ}(s) \stackrel{\text{def}}{=} \{s' \mid sEs'\} \neq \emptyset$ for every state s. The *probability function* P assigns each random state $s \in S_\bigcirc$ a distribution over its successor states, i.e., $P(s) \in \mathcal{D}(\mathsf{Succ}(s))$. We extend the domain of P to $S^*S_\bigcirc$ by $P(\rho s) \stackrel{\text{def}}{=} P(s)$ for all $\rho s \in S^+S_\bigcirc$. An *MDP* is a game where one of the two players does not control any states. An MDP is *maximizing* (resp. *minimizing*) iff $S_\Diamond = \emptyset$ (resp. $S_\Box = \emptyset$). A *Markov chain* is a game with only random states, i.e., $S_\Box = S_\Diamond = \emptyset$.

Strategies. A *play* is an infinite sequence $s_0 s_1 \ldots \in S^\omega$ such that $s_i \to s_{i+1}$ for all $i \geq 0$. A *path* is a finite prefix of a play. Let $\mathit{Plays}(\mathcal{G}) \stackrel{\text{def}}{=} \{\rho = (s_i)_{i \in \mathbb{N}} \mid s_i \to s_{i+1}\}$ denote the set of all possible plays. A *strategy* of the player $\Box$ ($\Diamond$) is a function $\sigma : S^*S_\Box \to \mathcal{D}(S)$ ($\pi : S^*S_\Diamond \to \mathcal{D}(S)$) that assigns to every path $ws \in S^*S_\Box$ ($\in S^*S_\Diamond$) a probability distribution over the successors of s. If these distributions are always Dirac then the strategy is called *deterministic* (aka pure), otherwise it is called *randomized*. The set of all strategies of player $\Box$ and $\Diamond$ in $\mathcal{G}$ is denoted by $\Sigma^\mathcal{G}$ and $\Pi^\mathcal{G}$, respectively. A play/path $s_0 s_1 \ldots$ is compatible with a pair of strategies (σ, π) if $s_{i+1} \in \mathsf{supp}(\sigma(s_0 \ldots s_i))$ whenever $s_i \in S_\Box$ and $s_{i+1} \in \mathsf{supp}(\pi(s_0 \ldots s_i))$ whenever $s_i \in S_\Diamond$.

Memory-based strategies can be defined via *Mealy Machines* [35] as follows. A strategy for a player $\odot \in \{\Box, \Diamond\}$ is a tuple $\tau = (\mathsf{M}, \mathsf{m}_0, \mathsf{upd}, \mathsf{nxt})$ where M is the set of memory modes, m_0 is the initial memory mode, the next/successor function $\mathsf{nxt} : \mathsf{M} \times S_\odot \to \mathcal{D}(S)$ chooses a (distribution over) successor states based on the current memory mode and state,

and the update function $\mathsf{upd}\colon \mathsf{M} \times E \to \mathcal{D}(\mathsf{M})$ updates the memory mode upon observing a transition. Let $\tau[\mathsf{m}]$ denote the strategy τ starting in memory mode m instead of m_0. Finite-memory Max (resp. Min) strategies use a *finite* set M of memory modes and are denoted by $\Sigma_f^{\mathcal{G}}$(resp. $\Pi_f^{\mathcal{G}}$). Randomized memory updates are strictly more expressive than deterministic memory updates for finite-memory strategies [31]. Finite-memory strategies with deterministic/randomized next function and deterministic/randomized update function are called FDD, FRD, FDR, FRR, respectively, i.e., the first D/R refers to randomization in the next function and second D/R refers to randomization in the memory update function [31, Figure 1.1] (adapted to our setting since we do not consider randomization over the initial memory mode). Note that, under standard definitions of behavioral strategies in game theory, any memory in memory-based strategies is *private by default*, since internal memory modes of the strategy of one player are not part of the history of the play, and thus not visible to the other player(s). Only in the special case of *deterministic* strategies there is no difference between public memory (observable by the other player) and private memory, since the other player can then accurately infer the current memory mode from the known deterministic strategy and the observed history. However, it can make a big difference for randomized strategies, where fixing a private-memory strategy of one player yields only a partially observable MDP; see, e.g., the Big Match, where the difference is crucial [29].

Strategies with memory $|\mathsf{M}| = 1$ are called *memoryless*. Memoryless deterministic (resp. randomized) strategies are called MD (resp. MR).

The memory-size of a finite-memory strategy $\tau = (\mathsf{M}, \mathsf{m}_0, \mathsf{upd}, \mathsf{nxt})$ is defined as $\|\tau\| \stackrel{\text{def}}{=} |\mathsf{M}|$. The total bit size of τ is defined as the size of the rational probabilities used in τ, i.e., $\text{bits}(\tau) \stackrel{\text{def}}{=} \text{bits}(\mathsf{nxt}) + \text{bits}(\mathsf{upd})$.

Measure. A game $\mathcal{G}$ with initial state s_0 and strategies (σ, π) yields a probability space $(s_0 S^\omega, \mathcal{F}_{s_0}, \mathcal{P}^{\mathcal{G}}_{\sigma,\pi,s_0})$ where $\mathcal{F}_{s_0}$ is the σ-algebra generated by the cylinder sets $s_0 s_1 \dots s_n S^\omega$ for $n \geq 0$. The probability measure $\mathcal{P}^{\mathcal{G}}_{\sigma,\pi,s_0}$ is first defined on the cylinder sets. For $\rho = s_0 \dots s_n$, let $\mathcal{P}^{\mathcal{G}}_{\sigma,\pi,s_0}(\rho) \stackrel{\text{def}}{=} 0$ if ρ is not compatible with σ, π and otherwise $\mathcal{P}^{\mathcal{G}}_{\sigma,\pi,s_0}(\rho S^\omega) \stackrel{\text{def}}{=} \prod_{i=0}^{n-1} \tau(s_0 \dots s_i)(s_{i+1})$ where τ is σ or π or P depending on whether $s_i \in S_\Box$ or $S_\Diamond$ or $S_\bigcirc$, respectively. By Carathéodory's extension theorem [2], this defines a unique probability measure on the σ-algebra.

Objectives and payoff functions. General objectives are defined by real-valued measurable functions $v\colon s_0 S^\omega \to \mathbb{R}$, and we write $\mathcal{E}(\cdot)$ for the expectation w.r.t. $\mathcal{P}$ and v. For event-based objectives, v is just the indicator function of a measurable set of plays $\mathtt{O} \subseteq S^\omega$, i.e., we consider the probability that plays belong to $\mathtt{O}$.

An objective v is called *shift-invariant* iff for all finite paths ρ and plays $\rho' \in S^\omega$, we have $v(\rho\rho') = v(\rho')$. It has also been called *prefix-independent* [27, Section 4] or *uniform* [17, Definition 3] in prior literature, but we follow the notation of [25] and use shift-invariant to avoid any ambiguities. Objective v is called *submixing* iff for all sequences of finite non-empty words u_0, w_0, u_1, $w_1 \dots$ we have $v(u_0 w_0 u_1 w_1 \dots) \leq \max(v(u_0 u_1 \dots), v(w_0 w_1 \dots))$, and *inverse-submixing* iff $v(u_0 w_0 u_1 w_1 \dots) \geq \min(v(u_0 u_1 \dots), v(w_0 w_1 \dots))$ [25].

We assume standard notions about LTL, reachability and parity objectives [16]; cf. Appendix A. $\mathtt{EPAR}$ denotes max-even parity.

Reward based objectives. Let $r\colon E \to \{-R, \dots, 0, \dots, R\}^d$ be a bounded function that assigns (possibly multi-dimensional) finite rewards to transitions. If $s \to s'$ and $r((s,s')) = c$, we write $s \xrightarrow{c} s'$. Let $\rho = s_0 \xrightarrow{c_0} s_1 \xrightarrow{c_1} \dots$ be a play. We say that ρ satisfies *mean-payoff* $\mathtt{MP} > \vec{c}$ for some constant $\vec{c} \in \mathbb{R}^d$ iff $\left(\liminf_{n\to\infty} \frac{1}{n} \sum_{i=0}^{n-1} c_i\right) > c$ (component-wise). The *mean-payoff-parity* objective is defined as $\mathtt{MP} > 0 \cap \mathtt{EPAR}$. Parity and mean payoff objectives are shift-invariant, but reachability is not. Parity is submixing and inverse-submixing.

For finite-state SGs there exist optimal MD Max strategies for the reachability $\mathsf{F}\,T$ objective [18], EPAR objective [40], and MP > 0 objective for dimension one [5, Prop. 7].

The size of a game $\mathcal{G}$ with an objective v can be thought of as the number of bits required to describe the states, transitions and probabilities used by $\mathcal{G}$ along with the description of v.

Determinacy. Given an objective v and a game $\mathcal{G}$, state s *has value* (w.r.t. v) iff

$$\sup_{\sigma\in\Sigma^{\mathcal{G}}}\inf_{\pi\in\Pi^{\mathcal{G}}}\mathcal{E}^{\mathcal{G}}_{\sigma,\pi,s}(v)=\inf_{\pi\in\Pi^{\mathcal{G}}}\sup_{\sigma\in\Sigma^{\mathcal{G}}}\mathcal{E}^{\mathcal{G}}_{\sigma,\pi,s}(v).$$

If s has value then $\mathtt{val}^{\mathcal{G}}_v(s)$ denotes the value of s defined by the above equality. A game with an objective is called *weakly determined* if every state has value. Stochastic games with Borel objectives are weakly determined [32, 33]. Our objectives above are Borel, hence any boolean combination of them is also weakly determined [12]. For $\varepsilon > 0$ and state s, a strategy

1. $\sigma \in \Sigma^{\mathcal{G}}$ is ε-optimal (maximizing) iff $\mathcal{E}^{\mathcal{G}}_{\sigma,\pi,s}(v) \geq \mathtt{val}^{\mathcal{G}}_v(s) - \varepsilon$ for all $\pi \in \Pi^{\mathcal{G}}$.
2. $\pi \in \Pi^{\mathcal{G}}$ is ε-optimal (minimizing) iff $\mathcal{E}^{\mathcal{G}}_{\sigma,\pi,s}(v) \leq \mathtt{val}^{\mathcal{G}}_v(s) + \varepsilon$ for all $\sigma \in \Sigma^{\mathcal{G}}$.

A 0-optimal strategy is called *optimal*. An MD strategy is called *uniformly* ε-optimal (resp. uniformly optimal) if it is so from every start state. Given an event-based objective $\mathtt{O}$, an optimal strategy for player $\Box$ from state s is *almost surely* winning if $\mathtt{val}^{\mathcal{G}}_{\mathtt{O}}(s) = 1$. By $\mathtt{AS}^{\mathcal{G}}_{\Box}(\mathtt{O})$ we denote the set of states that have an almost surely winning strategy for $\Box$ for objective $\mathtt{O}$. We drop subscripts and superscripts wherever possible if they are clear from the context.

3 Lifting Strategies from MDPs to 2-Player Stochastic Games for Shift-Invariant Inverse-Submixing Objectives

The *finite memory transfer theorem* of Gimbert & Kelmendi [25, Theorem 1.2] shows that finite-memory Min (resp. Max) strategies can be lifted from MDPs to 2-player stochastic games if the payoff function is both shift-invariant and submixing (resp. inverse-submixing). This is a consequence of the following slightly stronger theorem. (Adapted to our notation, because we consider objectives from Max's point of view.)

▶ **Theorem 1** ([25, Theorem 6.1]). *Let f be a shift-invariant and inverse-submixing payoff function. If, for all $\varepsilon > 0$, Max has an ε-optimal strategy with finite memory in every finite-state MDP, then in every finite-state turn-based two-player stochastic game he has an ε-subgame-perfect strategy that has finite memory.*

The statement also holds for $\varepsilon = 0$, that is: if Max has a finite-memory optimal strategy in every game controlled by himself, then in every two-player game he has an optimal subgame-perfect strategy with finite memory.

Subgame-perfect strategies are a stronger notion than optimal strategies and it suffices to view them as optimal strategies for our results. The proof of [25, Theorem 6.1] also shows that deterministic (resp. randomized) Max strategies in MDPs are lifted to deterministic (resp. randomized) Max strategies in the game. While [25] only consider strategies with deterministic memory updates, their construction can also lift (from MDPs to games) strategies with randomized memory updates. However, in [25, Theorem 6.1], the *extra* memory bits used by the constructed Max strategy in the game are always updated deterministically. Thus this Max strategy always uses *public* memory, since the memory in the MDP strategies was trivially public (by the absence of an opponent).

The size of the memory used by Max's strategy in the 2-player game can be described as follows. For each MD strategy choice for Min, Max fixes this strategy and computes the optimal strategy in this derived MDP and correspondingly reserves the required memory

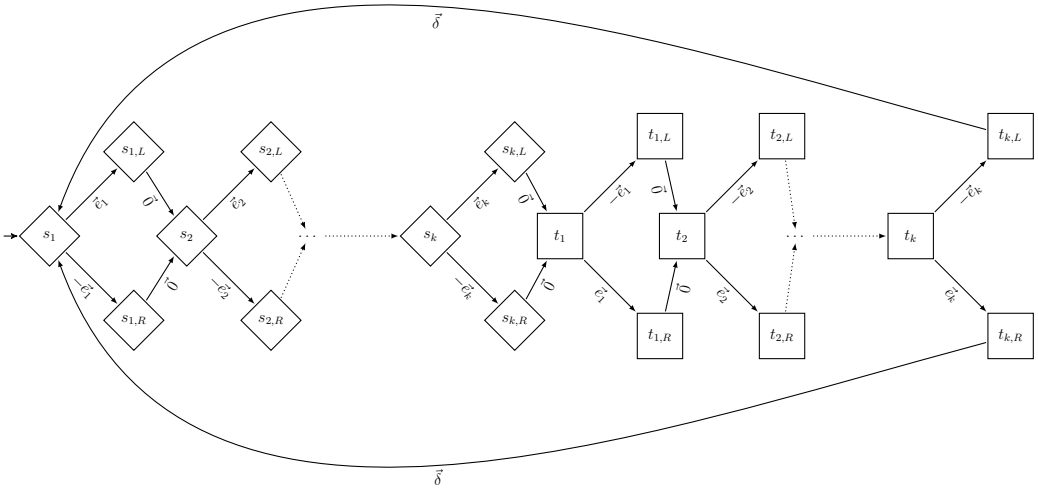

Figure 1 In the game $\mathcal{G}_k$, optimal Max strategies for $\mathtt{MP} > \vec{0}$ for $2k$ dimensions require at least 2^k memory modes. We have $2k$-dimensional reward vectors $\vec{e_i}$ such that $\vec{e_i}[2i-1] = +1$, $\vec{e_i}[2i] = -1$ and 0 elsewhere. The special vector $\vec{\delta} = (\delta, \ldots, \delta)$ has value $\delta \stackrel{\text{def}}{=} 2^{-2k}$ in every dimension.

space to reliably play this strategy (think of this as corresponding to remembering how to play when Min eventually fixes to this strategy). In addition to this, Max also remembers the choice Min played last at each of its states. Assuming k is the number of Min-controlled states, d the maximal out-degree of these states and p to be the maximum number of memory bits needed to play in any of the derived MDPs, then the total number of bits used by the described strategy would be of the order $k \cdot \lceil \log_2 d \rceil + d^k \cdot p$. Thus, in general, Max uses a *doubly exponential* number of memory modes in the size of the game, even if his good strategies in all MDPs use just polynomially many memory modes; cf. [25, Sec. 6.1].

An interesting subcase is to look at the construction when $p = 0$, i.e., when optimal strategies for Max in MDPs are memoryless. In this case, the above construction implies a memory bound of the order $k \cdot \lceil \log_2 d \rceil$ which is only exponential w.r.t. the game size. The end of [25, Sec. 6.1] claims that this is unnecessary, since a different lifting result [28] implies that memoryless strategies suffice, even in games. We show that this is true only for deterministic strategies and does not generalize to strategies that are memoryless randomized. In particular, we show the matching exponential lower bound in Theorem 2. Even if Max has optimal memoryless randomized strategies in all MDPs, in general he still needs exponential memory in the 2-player game, even if the memory is private. This attempts to partly tighten the lower bound at [25, end of Sec. 6.1] by showing that some parts of the construction are necessary when considering randomized strategies. On the other hand, for some particular objectives, the construction in [25, Theorem 6.1] uses more memory than necessary. In Section 4 we show that for the $\mathtt{MP} > 0 \cap \mathtt{EPAR}$ objective Max requires, at least and at most, just a *linear* number of memory modes in SGs (and the lower bound holds even for deterministic games).

The multi-dimensional $\mathtt{MP} > \vec{0}$ objective is shift-invariant (by definition) and inverse-submixing (by [6, Lemma 6] for dimension 1, and the fact that event-based inverse-submixing objectives are closed under intersection). Moreover, by [4, Prop. 5.1], almost surely winning strategies for $\mathtt{MP} > \vec{0}$ in MDPs can be chosen as memoryless randomized. However, the games in Figure 1 (similar to [15, Fig. 4] but not identical) show that even randomized Max strategies need exponentially many memory modes.

▶ **Theorem 2.** *There is a family of deterministic games $\mathcal{G}_k$ with $6k$ states as in Figure 1, where every state is almost surely winning for the $2k$-dimensional* $\texttt{MP} > \vec{0}$ *objective for Max, but any randomized Max strategy with $< 2^k$ private memory modes cannot win almost surely.*

Proof. Consider the games $\mathcal{G}_k$ from Figure 1 with $6k$ states. The dimension of the rewards $2k$ is split into k blocks of 2 dimensions each and the reward vector $\vec{e_i}$ is $(+1, -1)$ on the i-th block and zero elsewhere. First Min makes k decisions, choosing between $\vec{e_i}$ and $-\vec{e_i}$ for each $i = 1, 2, \ldots, k$. Then Max makes k decisions, choosing between $-\vec{e_i}$ and $\vec{e_i}$ for each $i = 1, 2, \ldots, k$. Max can win $\texttt{MP} > \vec{0}$ surely (from s_1 and thus from every other state) by exactly copying Min's choices such that these rewards cancel out, which leaves just the strictly positive reward vector $\vec{\delta} = (2^{-2k}, \ldots, 2^{-2k}) > \vec{0}$ between each visit to s_1. (Since $\vec{\delta}$ can be described with a number of bits that is polynomial in k, $\mathcal{G}_k$ has polynomial size.)

Intuitively, Max needs at least 2^k memory modes to remember which of the 2^k possible choices Min made, in order to copy it. While this is easy to prove for deterministic Max strategies, we show a stronger result: Even randomized Max strategies with $< 2^k$ private memory modes cannot win almost surely.

Let σ be an arbitrary randomized Max strategy with $m < 2^k$ private memory modes. Overall, for the k choices in the states $s_1, \ldots, s_k$, Min has 2^k different possible options, which we denote by index numbers $j \in \{1, \ldots, 2^k\}$. (Similarly for Max's options in states $t_1, \ldots, t_k$.) Let π_j be the deterministic Min strategy that always plays option j forever. Let π be the randomized Min strategy which initially picks a $j \in \{1, \ldots, 2^k\}$, with equal probability 2^{-k}, and then plays π_j forever. (Note that π does not care about Max's strategy at all, and thus neither about Max's memory modes, so the memory can be private.) Now we show that $\mathcal{P}^{\mathcal{G}_k}_{\sigma,\pi,s_1}\left(\texttt{MP} > \vec{0}\right) < 1$.

Let M_k^j be the Markov chain derived from $\mathcal{G}_k$ by fixing σ and π_j. If M_k^j is irreducible and aperiodic then let P be the $2^k \times m$ matrix such that $P(j, i)$ is the probability, in the steady-state, that σ is in memory mode $i \in \{1, \ldots, m\}$ in state t_1. (It is also possible that M_k^j is reducible, since it has a memory component from the strategy σ as part of its state. If σ is such that it would never visit certain memory modes again eventually, then this makes M_k^j reducible. In this case we consider an irreducible subchain that is reached with maximal probability. If M_k^j is periodic then consider the long-run average frequency of being in memory mode i in state t_1 instead of the probability in the following arguments.) Let Q be the $m \times 2^k$ matrix where $Q(i, j)$ is the probability that σ will play option j in states $t_1, \ldots, t_k$ if it is in memory mode i in state t_1.

The probability that Max will play option j in M_k^j in the long run is then given by $(PQ)(j, j)$, the j-th diagonal value in the product matrix. Since $m < 2^k$, the $2^k \times 2^k$ matrix PQ does not have full rank. Thus at least one of its eigenvalues is 0. Moreover, since both matrices P and Q are row-stochastic, the absolute value of the eigenvalues of PQ are upper-bounded by 1. So the sum of the eigenvalues of PQ is upper-bounded by $2^k - 1$. The Cayley–Hamilton theorem implies that the trace is the sum of the eigenvalues. This yields an upper bound on the trace of PQ, namely $Tr(PQ) = \sum_{j=1}^{2^k} (PQ)(j, j) \leq 2^k - 1$. Hence there exists a j' such that $(PQ)(j', j') \leq 1 - 2^{-k}$. So, in $M_k^{j'}$, the long-run probability that Max will pick some option different from j' is at least 2^{-k}. Hence there exists at least one option $j'' \neq j'$ that is picked by Max in the long run with probability $\geq 2^{-k} 2^{-k} = 2^{-2k}$. Since $j'' \neq j'$, there exists at least one block of two dimensions, say x and $x + 1$, where Max choosing j' has effect $(+1, -1)$ and j'' has the opposite effect $(-1, +1)$ (or vice-versa; this case is symmetric). Thus, between two visits to state s_1, the expected reward in dimension x is $\leq -2 \cdot 2^{-2k} + \delta = -2^{-2k} < 0$. Therefore, in $M_k^{j'}$, the $\texttt{MP} > \vec{0}$ objective is satisfied with probability zero, since it almost surely fails in dimension x. I.e., $\mathcal{P}^{M_k^{j'}}_{\sigma,\pi_{j'},s_1}\left(\texttt{MP} > \vec{0}\right) = 0$.

Since Min's strategy π initially decides to become $\pi_{j'}$ with probability 2^{-k}, we obtain that $\mathcal{P}^{\mathcal{G}_k}_{\sigma,\pi,s_1}\left(\mathtt{MP} > \vec{0}\right) \leq 1 - 2^{-k}(1 - \mathcal{P}^{M_k^{j'}}_{\sigma,\pi_{j'},s_1}\left(\mathtt{MP} > \vec{0}\right)) = 1 - 2^{-k} < 1$. ◀

As a side note, the assumption of a shift-invariant inverse-submixing Max objective implies that Min always has optimal memoryless deterministic (MD) strategies in all MDPs/SGs by [25, Theorem 1.1], since this objective is then shift-invariant and submixing for Min.

4 Strategy Complexity of $\mathtt{MP} > 0 \cap \mathtt{EPAR}$ With Randomization

In this section, unless otherwise stated, we consider the one-dimensional $\mathtt{MP} > 0 \cap \mathtt{EPAR}$ objective, where Max needs to attain a strictly positive one-dimensional mean payoff while also satisfying a max-even parity condition. These are defined via a reward function r and a coloring function Col where rewards are in $[-R, R]$ and colors in $\{0, 1, \ldots, d\}$.

For almost surely winning strategies in MDPs $\mathcal{M}$, *deterministic* Max strategies require $\mathcal{O}(\exp(\|\mathcal{M}\|))$ memory modes, i.e., exponential memory is both necessary and sufficient [26, Theorem 5]. We show that *randomized* Max strategies require less memory: none in MDPs and just *linearly* many memory modes in stochastic games.

▶ **Theorem 3.** *In maximizing MDPs, almost surely winning strategies for the multi-dimensional* $\mathtt{MP} > \vec{0} \cap \mathtt{EPAR}$ *objective can be chosen memoryless randomized (MR).*

Proof. We use the fact that, for any strategy, almost surely all the runs eventually end up in an end component [21, Theorem 3.2]. In a finite-state MDP, there can only be a finite (albeit exponential) number of end components. Let $\mathcal{M}_1 \stackrel{\text{def}}{=} \left(S_1, S^1_\Box, S^1_\bigcirc, E_1, P_1\right)$ be one such end component of $\mathcal{M}$. We say that $\mathcal{M}_1$ is "winning" iff the highest color is even and $S_1 \subseteq \mathtt{AS}^{\mathcal{M}_1}_\Box\left(\mathtt{MP} > \vec{0}\right)$. Given an almost surely winning strategy σ^*, except for a null set, all induced runs must eventually stay inside a winning end component. Thus from every almost surely winning state it is possible to almost surely reach a winning end component.

▷ Claim 4. Let $\mathcal{M}_1$ be a winning end component. Then there is a memoryless randomized strategy σ^* which is almost surely winning for $\mathtt{MP} > \vec{0} \cap \mathtt{EPAR}$ from every state in $\mathcal{M}_1$.

Proof. By definition of a winning end component, we know that $S_1 \subseteq \mathtt{AS}^{\mathcal{M}_1}_\Box\left(\mathtt{MP} > \vec{0}\right)$. From the proof of [4, Prop. 5.1 (ii)] we obtain that there exists a memoryless randomized strategy ξ_ε for some $\varepsilon > 0$, such that ξ_ε almost surely wins $\mathtt{MP} > \vec{0}$ from any state in S_1 and $\|\varepsilon\|$, the size of the probabilities used by ξ_ε, is polynomial in $\|\mathcal{M}_1\|$. Furthermore, every edge in E_1 is used with some positive probability. This implies that ξ_ε also satisfies $\mathtt{EPAR}$ almost surely, since the highest color in $\mathcal{M}_1$ is even. ◁

Consider the union of all winning end components $\mathcal{C}$ in $\mathcal{M}$. We obtain an almost surely winning MR strategy for $\mathtt{MP} > \vec{0} \cap \mathtt{EPAR}$ from every state in $\mathcal{M}$ as follows. Outside of $\mathcal{C}$ it plays an almost surely winning uniform MD strategy for the reachability objective $\mathsf{F}\,\mathcal{C}$. Inside each maximal winning end component $\mathcal{M}_1$ in $\mathcal{C}$ it plays the MR strategy ξ_ε from Claim 4. ◀

Since both $\mathtt{MP} > 0$ and $\mathtt{EPAR}$ are shift-invariant and inverse-submixing (by [6, Lemma 6] and [25, Prop. 3.1]), the same holds for the conjunction $\mathtt{MP} > 0 \cap \mathtt{EPAR}$. For randomized strategies, even with the improved upper bound in MDPs of Theorem 3, the construction in Theorem 1 ([25, Sec. 6.1]) still only yields an exponential upper bound on the memory of Max strategies in stochastic games. We show that this is excessive in general. Optimal randomized Max strategies for $\mathtt{MP} > 0 \cap \mathtt{EPAR}$ in stochastic games require, at least and at most, #(distinct even colors) many memory modes, i.e., just linear memory.

▶ **Theorem 5.** *Consider a game $\mathcal{G} = (S, (S_\Box, S_\Diamond, S_\bigcirc), E, P)$, coloring function Col with the highest color d, reward function r and objective* $\texttt{MP} > 0 \cap \texttt{EPAR}$ *for player Max.*

1. *If Max can win almost surely from some state s then there also exists an almost surely winning randomized Max strategy from s that uses at most k public memory modes and total bit size $\mathcal{O}\left(\|\mathcal{G}\|^{d+c}\right)$ where $k = \#(\text{distinct even colors})$, d the highest color, and c is a constant independent of $\mathcal{G}$.*
2. *The bound on the number of required memory modes for almost surely winning randomized Max strategies is tight. There is a family of deterministic games $\{\mathcal{G}_n \mid n \geq 2\}$ as in Definition 6 and Figure 2 such that*
   - *$\|\mathcal{G}_n\| = \Theta(n)$, $\mathcal{G}_n$ contains n even colors and every state is almost surely winning for Max.*
   - *For every $n \geq 2$, any Max strategy with $< n$ (private) memory modes is worthless, i.e., it cannot guarantee* $\texttt{MP} > 0 \cap \texttt{EPAR}$ *with any positive probability.*

The idea of using randomization to reduce memory complexity is not new and appears, e.g., in [10, 7, 30]. However, it is interesting to note that the Max strategy in SGs still requires a small amount of memory even in the presence of randomization and cannot be made memoryless, unlike in MDPs (Theorem 3).

Table 1 Worst case number of memory modes required for almost surely winning Max strategies for objective $\texttt{MP} > 0 \cap \texttt{EPAR}$ in MDPs and stochastic games.

Strategy \ Arena	MDPs $\mathcal{M}$	Stochastic games $\mathcal{G}$
Deterministic	$\Theta(\exp(\|\mathcal{M}\|))$ [26]	$\mathcal{O}(2^{-\exp(\|\mathcal{G}\|)})$ [26, 25]
Randomized	1	$\#(\text{distinct even colors}) = \Theta(\|\mathcal{G}\|)$

Table 1 highlights the strategy complexity of almost surely winning Max strategies in MDPs and games for the $\texttt{MP} > 0 \cap \texttt{EPAR}$ objective. Note that while there is no explicit result for the complexity of deterministic winning strategies, the results from [26] and lifting this strategy using Theorem 1 ([25, Sec. 6.1]) provide a doubly exponential upper bound. The entries in the randomized row are our contributions. The bit size of these randomized Max strategies is $\mathcal{O}(\|\mathcal{M}\|^c)$ and $\mathcal{O}\left(\|\mathcal{G}\|^{d+c_1}\right)$ in MDPs and games respectively, where $k = \#(\text{distinct even colors})$, d is the highest color, c and c_1 are constants.

Upper bound. The upper bound in Theorem 5 Item 1 is shown by induction on the number of colors, with a case distinction whether the highest color is odd or even. Max's optimal strategy needs to switch between different modes, and it uses randomized memory updates with small probabilities to implement a "probabilistic clock", in order to stay in one mode for a very long time (in expectation), but not forever; cf. [19, Appendix B].

Outline of the proof. We first outline the proof of the upper bound in Theorem 5, Item 1. (Full proof in [19, Appendix B].) It follows the induction argument of [11, Lemma 2,3] along with strategy complexity analysis for $\texttt{MP} > 0 \cap \texttt{EPAR}$ instead of $\texttt{MP} \geq 0 \cap \texttt{EPAR}$. There are two cases, depending on whether the maximum color in the game is odd or even.

If the maximum color d is odd, then the EPAR part of $\texttt{MP} > 0 \cap \texttt{EPAR}$ requires that this color must eventually never been seen any more (except in a null set of the plays). Using a ranking argument, we show that it is possible to partition the state space into layers $Z_1, \cdots, Z_\ell$ such that

- Max can force the win in any one of these layers if the play stays there forever.
- Min cannot infinitely often switch between the layers, except in a null set of the plays.

The above two facts can be combined to build a winning Max strategy. Interestingly, this strategy does not use any additional memory, compared to the inductive case with one less color. It can be seen as a combination of memoryless attractor strategies on certain states, combined with almost surely winning strategies in other states in subgames with at least one less color (using the IH).

If the maximum color d is even, Max tries wherever possible to reach a state of this color but does so sufficiently infrequently, so that this does not compromise the satisfaction of the positive mean-payoff objective. This is achieved by operating in phases with Max playing either for $\mathtt{MP} > 0 \cap \mathsf{F}(S(d))$ or for $\mathtt{MP} > 0 \cap \mathtt{EPAR}$ in a subgame with at least 1 less color.

Max's strategy. Let S denote the set of states in the game $\mathcal{G}$, and $X \stackrel{\text{def}}{=} \mathtt{Attr}_{\Box}(S(d))$ the positive attractor of states with color d in $\mathcal{G}$, $Y \stackrel{\text{def}}{=} S \setminus X$. Since $\mathcal{G}[Y]$ has at least 1 less color, by induction hypothesis let σ^*_Y denote the almost surely winning strategy for Max in $\mathcal{G}[Y]$. Let $\sigma^*_{\mathtt{MP}}$ denote the optimal MD strategy for $\mathtt{MP} > 0$ in $\mathcal{G}$ and $\sigma^*_{\mathtt{Attr}}$ denote the positive attractor strategy in states in $X \cap S_{\Box}$. Then the optimal randomized Max strategy σ^* works as follows.

Phase-1. If the current state is in X, play the mixed strategy $\varepsilon_0\sigma^*_{\mathtt{Attr}} + (1-\varepsilon_0)\sigma^*_{\mathtt{MP}}$. I.e., play $\sigma^*_{\mathtt{Attr}}$ with some sufficiently small probability $\varepsilon_0 > 0$ and play $\sigma^*_{\mathtt{MP}}$ with probability $1-\varepsilon_0$. Else, play according to $\sigma^*_{\mathtt{MP}}$. At every step, there is a sufficiently small chance $\varepsilon_1 > 0$ to stop this phase. This is done by changing Max's memory mode with probability ε_1. So Phase-1 stops eventually almost surely, where the expected time to stopping depends on ε_1. This serves as a makeshift probabilistic clock (since this strategy does not use a real clock). After this phase is over, if the play is in X, restart Phase-1, else move to Phase-2.

Phase-2. Play according to σ^*_Y while the play is in Y. If the play ever moves to X, switch to Phase-1.

One has to choose $\varepsilon_0, \varepsilon_1 > 0$ so that the overall strategy is almost surely winning. Intuitively, $\varepsilon_0 > 0$ is chosen so small that the mean-payoff is strictly positive in Phase-1, but this holds only in the long run. Moreover, while Phase-2 also yields a strictly positive mean-payoff in the long run, it might switch back to Phase-1 early while the accumulated reward is still negative. (However, this possible negative reward can be bounded, in expectation.) Therefore $\varepsilon_1 > 0$ is chosen so small that Phase-1 is played for a very long time (in expectation). Hence Phase-1 closely approximates its long run behavior with strictly positive mean-payoff, and thus it also compensates for any possible negative reward obtained during a temporary switch to Phase-2.

Also note that this strategy uses randomization in both the memory updates and the next move. This corresponds to DRR randomized strategy class in the notation of [31, Figure 1.1]. Furthermore, the strategy needs 1 additional memory mode (compared to strategy σ^*_Y) in order to know the current phase. This is necessary, because it needs to play different strategies in Y: $\sigma^*_{\mathtt{MP}}$ in Phase-1 and σ^*_Y in Phase-2.

See [19, Appendix B] for the full proof.

Lower bound. Towards proving the lower bound in Theorem 5 Item 2, we construct the following class of games in Figure 2. The idea is that Max needs to remember Min's last move from state s to some state x, since, in order to win, he needs to imitate it from state t by going to state $x-1$.

The following definition formally defines the class of games in Figure 2.

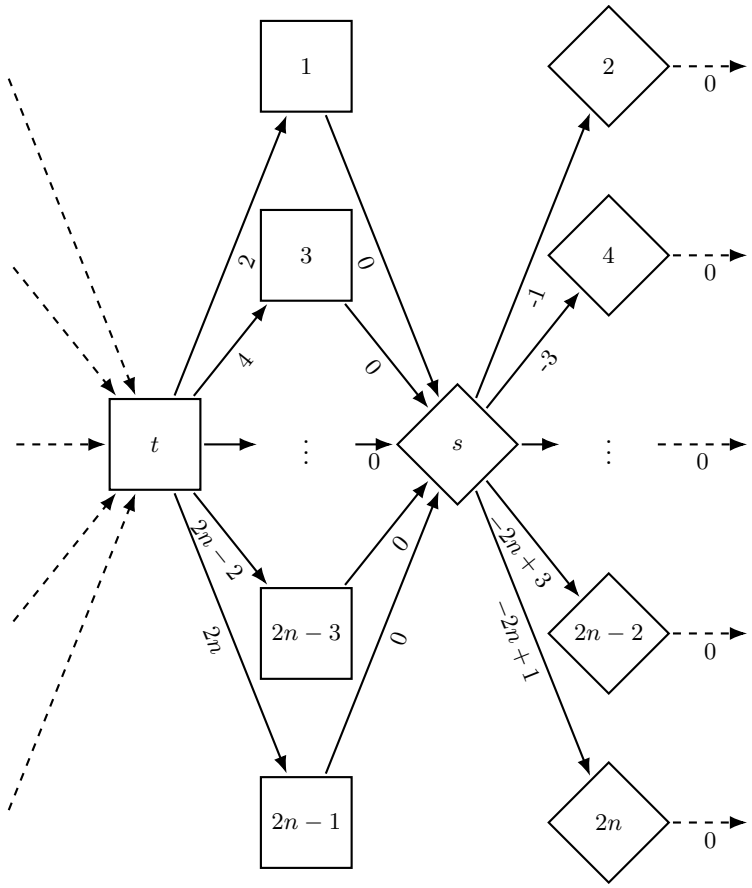

Figure 2 Game $\mathcal{G}_n$ from Definition 6 with odd states up to $2n-1$ and even states up to $2n$.

▶ **Definition 6.** *For each* $n \in \mathbb{Z}_+$ *let* $Odd^n \stackrel{def}{=} \{x \in \mathbb{Z}_+ \mid x \leq 2n \wedge x \mod 2 = 1\}$ *and* $Even^n \stackrel{def}{=} \{x \in \mathbb{Z}_+ \mid x \leq 2n \wedge x \mod 2 = 0\}$ *and* $\mathcal{G}_n \stackrel{def}{=} \left(S^n, \left(S^n_\square, S^n_\diamond, S^n_\bigcirc\right), E^n, P^n\right)$ *where*

1. $S^n_\square \stackrel{def}{=} \{t\} \uplus Odd^n$, $S^n_\diamond \stackrel{def}{=} \{s\} \uplus Even^n$, $S^n_\bigcirc \stackrel{def}{=} \emptyset$, *consequently* P^n *is trivial.*
2. $E^n \stackrel{def}{=} t \times Odd^n \cup Odd^n \times s \cup s \times Even^n \cup Even^n \times t$

For $n \in \mathbb{Z}_+$, *state* n *has color* n *and* $Col(s) = Col(t) \stackrel{def}{=} 1$. *The reward function* r^n *on the edges is defined as* $r^n((t,i)) \stackrel{def}{=} i+1, \quad r^n((i,s)) \stackrel{def}{=} 0, \quad r^n((s,j)) \stackrel{def}{=} -j+1, \quad r^n((j,t)) \stackrel{def}{=} 0.$

Proof of Item 2 (Theorem 5). It is clear from Definition 6 that $\|\mathcal{G}_n\| = \Theta(n)$ and there are n even colors in $\mathcal{G}_n$ for every $n \geq 2$. Min (resp. Max) controls only one state s (resp. t), where it has a non-trivial choice. Max can win $\texttt{MP} > 0 \cap \texttt{EPAR}$ surely from state s (and thus from every other state) by the strategy that, at state t, copies Min's most recent observed choice at state s. I.e., if Min's last step was $s \to x$ for some even x then Max chooses $t \to (x-1)$. (This Max strategy requires n memory modes.) All induced plays trivially satisfy $\texttt{EPAR}$. Moreover, the total reward between consecutive visits to state s is always $-(x-1)+x = 1$, and thus $\texttt{MP} > 0$ is also satisfied surely.

However, we show that all Max strategies with $< n$ private memory modes are worthless.

Towards a contradiction, assume that there exists a Max strategy σ from state s in $\mathcal{G}_n$ with a set of private memory modes M where $|\mathsf{M}| < n$ and $\inf_\pi \mathcal{P}^{\mathcal{G}_n}_{\sigma,\pi,s}(\texttt{MP} > 0 \cap \texttt{EPAR}) > 0$.

First we show, by induction on k, the following property $P(k)$ for every $1 \leq k \leq n-1$: There exists a subset M_k of the memory modes, such that $|\mathsf{M}_k| \geq k$ and for every $\mathsf{m} \in \mathsf{M}_k$, the support of $\sigma[\mathsf{m}](t)$ is limited to the k lowest options, i.e., $\subseteq \{1, 3, \ldots, 2k-1\}$.

Base case $k = 1$. Let π be the Min strategy that always chooses the lowest option 2. If, for every $\mathsf{m} \in \mathsf{M}$, $\sigma[\mathsf{m}](t)$ includes options *different from* 1, then $\mathcal{P}^{\mathcal{G}_n}_{\sigma,\pi,s}(\mathtt{MP} > 0 \cap \mathtt{EPAR}) \leq \mathcal{P}^{\mathcal{G}_n}_{\sigma,\pi,s}(\mathtt{EPAR}) = 0$, a contradiction. Hence there must exist at least one memory mode $\mathsf{m} \in \mathsf{M}_1$ such that the support of $\sigma[\mathsf{m}](t)$ is limited to option 1.

Induction step $k-1$ to k. Let π be the Min strategy that always chooses the option $2k$. Consider the long-run behavior of the finite-state Markov chain induced by $\mathcal{G}_n$, σ and π. In the runs where Max's memory mode at t is eventually always contained in M_{k-1}, the total payoff between consecutive visits to s is ≤ -1 and therefore $\mathtt{MP} < 0$ and thus these runs are losing for $\mathtt{MP} > 0 \cap \mathtt{EPAR}$. Hence there must exist a non-null set of runs where Max's memory mode at t is infinitely often in $\mathsf{M} \setminus \mathsf{M}_{k-1}$. Suppose that for all $\mathsf{m} \in \mathsf{M} \setminus \mathsf{M}_{k-1}$ the support of $\sigma[\mathsf{m}](t)$ is *not* limited to the k lowest options $\{1, \ldots, 2k-1\}$. Then $\mathcal{P}^{\mathcal{G}_n}_{\sigma,\pi,s}(\mathtt{MP} > 0 \cap \mathtt{EPAR}) \leq \mathcal{P}^{\mathcal{G}_n}_{\sigma,\pi,s}(\mathtt{EPAR}) = 0$, a contradiction. Therefore there exists at least one $\mathsf{m} \in \mathsf{M} \setminus \mathsf{M}_{k-1}$ such that the support of $\sigma[\mathsf{m}](t)$ is limited to the k lowest options $\{1, \ldots, 2k-1\}$. Thus $\mathsf{M}_k \supseteq \{\mathsf{m}\} \uplus \mathsf{M}_{k-1}$ and, by induction hypothesis, $|\mathsf{M}_k| \geq 1 + |\mathsf{M}_{k-1}| \geq 1 + (k-1) = k$. This concludes the induction step.

For $k = n-1$, property $P(n-1)$ yields the required contradiction. Since $|\mathsf{M}| \leq n-1$, we have $\mathsf{M} = \mathsf{M}_{n-1}$. Thus the support of $\sigma(t)$ never includes the highest option $2n-1$. Let π be the Min strategy that always chooses the highest option $2n$. Then $\mathcal{P}^{\mathcal{G}_n}_{\sigma,\pi,s}(\mathtt{MP} > 0 \cap \mathtt{EPAR}) \leq \mathcal{P}^{\mathcal{G}_n}_{\sigma,\pi,s}(\mathtt{MP} > 0) = 0$. ◀

5 Counterexamples for Lifting Randomized Strategies

A different method to lift strategies from MDPs to SGs was described in [28, Theorem 9]. Unlike [25, Theorem 6.1], it does not require that the objective is shift-invariant inverse-submixing, but instead has stronger prerequisites about the strategy complexity in MDPs (resp. 1-player games). Given an objective, if both players have optimal memoryless *deterministic* strategies in all maximizing (resp. minimizing) MDPs, then they also have optimal memoryless deterministic strategies in all SGs. Under mild assumptions this can be generalized to finite-memory deterministic strategies [3, Theorem 4.1], in stochastic or non-stochastic arenas.

Such results do *not* generalize to *randomized* strategies by the counterexamples below.

However, first note that there cannot exist any counterexample where the objective is shift-invariant and submixing. In that case [25, Theorem 1.1] implies that Max has optimal MD strategies in SGs. This leaves the question whether there are counterexamples that satisfy other strong properties, e.g., shift-invariant and inverse-submixing.

One counterexample was discussed in [3, Section 4.4]. For the $\mathtt{MP} = 0$ objective, Max (resp. Min) have optimal MR (resp. MD) strategies in deterministic 1-player games, but optimal Max strategies require at least 1 bit of memory in deterministic 2-player games, even if randomization is allowed. [3, Section 4.4] does not analyze the strategy complexity of this objective in MDPs, but it is easy to show that Max (resp. Min) has optimal MR (resp. MD) strategies even in MDPs. It is also easy to extend this example with multiple rewards, such that optimal Max strategies require more (but still finite) memory. However, there remains one downside. While the $\mathtt{MP} = 0$ objective is shift-invariant, it is neither submixing nor inverse-submixing. The plays/sequences constructed in the proof of [6, Lemma 9] (only in the full version on arXiv) provide counterexamples to either property.

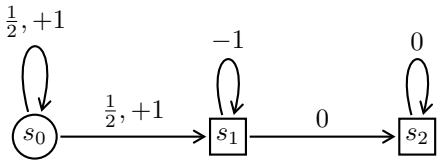

Figure 3 An MDP where every optimal Max strategy for the objective W requires infinite memory. In the random state s_0 the successor is either s_0 or s_1, each with probability $1/2$, and the reward is $+1$. In the controlled state s_1 Max chooses between staying in s_1 with reward -1 or going to s_2 with reward 0. State s_2 is a sink with reward 0.

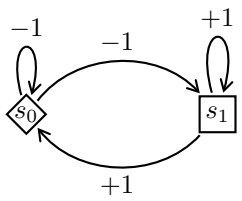

Figure 4 A deterministic 2-player game where every optimal Max strategy for the objective $\texttt{O}$ with $X \stackrel{\text{def}}{=} \{s_0\}$ requires infinite memory. State s_0 (resp. s_1) belongs to Min (resp. Max) with reward -1 (resp. $+1$). The players choose between staying in the current state or switching.

Our lower bounds for the one-dimensional $\texttt{MP} > 0 \cap \texttt{EPAR}$ objective in Section 4 show a slightly stronger property. Max (resp. Min) have optimal MR (resp. MD) strategies even in MDPs, but Max strategies in deterministic 2-player games can require arbitrarily many memory modes (equal to the number of even colors), even if randomization is allowed. Moreover, the $\texttt{MP} > 0 \cap \texttt{EPAR}$ objective is shift-invariant and inverse-submixing.

The multi-dimensional $\texttt{MP} > \vec{0}$ objective considered in Section 3 is also shift-invariant and inverse-submixing and Max (resp. Min) have optimal MR (resp. MD) strategies even in MDPs. However, optimal Max strategies in deterministic 2-player games can require an *exponential* (in the dimension) number of memory modes, even if randomization is allowed.

If one drops the requirement that Min has optimal MD strategies in MDPs, then there exist even stronger counterexamples where Max requires infinite memory in deterministic 2-player games.

An example in [39, Proposition 3.1.3] considers the objective $W \stackrel{\text{def}}{=} W_1 \cup W_2$, where $W_1 \stackrel{\text{def}}{=} \{c_1 c_2 \cdots \mid \liminf_{n\to\infty} \sum_{i=1}^{n} c_i = +\infty\}$ and $W_2 \stackrel{\text{def}}{=} \{c_1 c_2 \cdots \mid \sum_{i=1}^{n} c_i = 0 \text{ for infinitely many } n\}$. Both players have optimal finite-memory deterministic strategies in deterministic 1-player games, but Max needs infinite memory in deterministic 2-player games. However, note that the objectives W_2 and $W = W_1 \cup W_2$ are not shift-invariant. Moreover, [39] does not discuss the strategy complexity of W in MDPs, and the example in Figure 3 shows that optimal Max strategies for W already require infinite memory in MDPs. I.e., W is not a counterexample for lifting strategies from MDPs to SGs.

▶ **Proposition 7.** *Given the MDP in Figure 3, Max has an optimal strategy that attains the objective W with probability 1, but every FRR Max strategy is not optimal.*

Proof. First, since s_0 is left almost surely, W_1 is a null set under all Max strategies.

An optimal Max strategy plays as follows. When s_1 is reached with total reward $+n$ (which happens with probability 2^{-n}), then it loops n times in s_1 and then goes to s_2 where the total reward stays 0 forever. This satisfies W_2 (and thus W) with probability 1.

Now consider an FRR Max strategy with m memory modes. Since m is finite, there exists at least one memory mode m and numbers $n_2 > n_1 > 0$ such that the two events of entering s_1 with memory mode m and total rewards n_1 and n_2 each have a nonzero probability. Thus with nonzero probability either the state s_2 is reached with a total reward $\neq 0$ or s_2 is never reached, and hence W is not satisfied almost surely. ◀

We now present a stronger counterexample where the objective is shift-invariant and both Max and Min each have optimal MR (or alternatively, FDD) strategies in all MDPs, but optimal Max strategies still require infinite memory (even if randomization is allowed) in deterministic 2-player games.

Let $\mathtt{O}_1 \stackrel{\text{def}}{=} \{c_1 c_2 \cdots \mid \limsup_{n\to\infty} \sum_{i=1}^{n} c_i > -\infty\}$. Let $X \subseteq S$ be a subset of the states, e.g., defined via a coloring function. Our objective $\mathtt{O}$ combines $\mathtt{O}_1$ with Büchi and co-Büchi objectives wrt. X. Let $\mathtt{O} \stackrel{\text{def}}{=} \mathsf{FG}X \ \vee \ ((\mathsf{GF}X) \wedge \mathtt{O}_1)$. The objective $\mathtt{O}$ is shift-invariant, but neither submixing nor inverse-submixing.

▶ **Proposition 8.** *Max has optimal MR (and FDD) strategies for* $\mathtt{O}$ *in maximizing MDPs.*

Proof. Given a maximizing MDP, we can consider the end components wrt. an optimal strategy. For each winning end component E there are several possible cases.

If $\mathsf{FG}X$ is satisfied almost surely in E then an MD strategy is sufficient inside E, since that is a co-Büchi objective.

Otherwise, if it is possible to satisfy $((\mathsf{GF}X) \wedge \mathtt{O}_1)$ almost surely in E, then there are two cases. In the first case, it is possible to satisfy even the stronger objective $((\mathsf{GF}X) \wedge \mathtt{MP} > 0)$ almost surely in E. This objective is a special case of $\mathtt{MP} > 0 \cap \mathtt{EPAR}$ and thus an MR (or alternatively FDD) strategy is sufficient by Theorem 3. In the second case, it is possible to almost surely satisfy $((\mathsf{GF}X) \wedge \mathtt{O}_1)$ but not $((\mathsf{GF}X) \wedge \mathtt{MP} > 0)$ inside E. There must exist a state $x \in X$ inside E, such that one can satisfy $((\mathsf{GF}x) \wedge \mathtt{O}_1)$. Since E is finite, it must be possible to satisfy $((\mathsf{GF}x) \wedge \mathtt{MP} = 0)$ almost surely. Thus, there must exist a strategy from state x such that x is re-visited almost surely with an expected total reward $= 0$. Playing such a strategy repeatedly is also sufficient to satisfy $((\mathsf{GF}x) \wedge \mathtt{O}_1)$ almost surely. Thus it suffices to play an MD strategy for the optimal expected total reward (paid out upon visiting x) inside E.

Hence inside every winning end component an MD strategy or an MR (resp. FDD) strategy suffices. Elsewhere, we can fix an MD strategy that maximizes the chance of reaching a winning end component. Altogether this yields an optimal MR strategy (resp. FDD strategy). ◀

▶ **Proposition 9.** *Min has optimal MR (and FDD) strategies for* $\mathtt{O}$ *in minimizing MDPs.*

Proof. It suffices to show the existence of optimal Max MR (and FDD) strategies for the complement objective $\overline{\mathtt{O}} = \mathsf{GF}\overline{X} \ \wedge \ ((\mathsf{FG}\overline{X}) \vee \overline{\mathtt{O}_1}) = (\mathsf{FG}\overline{X}) \vee (\mathsf{GF}\overline{X} \wedge \overline{\mathtt{O}_1})$.

Like in the proof of Proposition 8, we can consider the strategies in winning end components E. For the co-Büchi objective $\mathsf{FG}\overline{X}$, an MD strategy in E suffices. Otherwise, in finite-state MDPs, we can win $(\mathsf{GF}\overline{X} \wedge \overline{\mathtt{O}_1})$ almost surely iff we can win $(\mathsf{GF}\overline{X} \wedge \mathtt{MP} < 0)$ almost surely. For this an MR (or alternatively FDD) strategy is sufficient by Theorem 3.

Hence inside every winning end component an MD strategy or an MR (resp. FDD) strategy suffices. Elsewhere, we can fix an MD strategy that maximizes the chance of reaching a winning end component. Altogether this yields an optimal MR strategy (resp. FDD strategy). ◀

However, in deterministic 2-player games, optimal Max strategies for $\mathtt{O}$ require infinite memory.

▶ **Proposition 10.** *Consider the deterministic 2-player game* $\mathcal{G}$ *from Figure 4, and objective* $\mathtt{O}$ *with* $X \stackrel{\text{def}}{=} \{s_0\}$. *Max can win objective* $\mathtt{O}$ *surely, but every FRR Max strategy cannot guarantee any positive probability for* $\mathtt{O}$.

Proof. Towards the first part, we define an optimal Max strategy that keeps track of the total reward and plays as follows. Whenever s_1 is reached with some negative total reward $-n$, Max loops $n-1$ times at s_1 and then goes back to s_0, which makes the total reward 0. This ensures objective $\mathtt{O}$ against any Min strategy as follows. In plays where Min eventually stays in s_0 forever we have $\mathsf{FG}X$ and thus $\mathtt{O}$. Otherwise, s_0 and s_1 are both visited infinitely often and the lim sup of the total reward is 0, and thus $\mathtt{O}$ holds.

Towards the second part, we consider an FRR Max strategy σ with m private memory modes, starting in memory mode m_0. We will construct a Min strategy π such that $\mathcal{P}^{\mathcal{G}}_{\sigma,\pi,s_0}(\mathtt{O}) = 0$.

For every memory mode m let $p(\mathsf{m})$ be probability that, in state s_1 and memory mode m, in the next step σ returns to s_0. Let $p \stackrel{\text{def}}{=} \min\{p(\mathsf{m}) \mid p(\mathsf{m}) > 0\} > 0$. Let π be the MR Min strategy that in s_0 goes to s_1 with probability $p/2$ and to s_0 with probability $1-p/2$. Fixing the strategies σ, π yields a Markov chain with $2m$ states and initial state (s_0, m_0).

Let E be the event that s_1 is visited when σ is in some memory mode m with $p(\mathsf{m}) = 0$.

Conditioned under E, the properties $\mathsf{FG}X$ and $\mathsf{GF}X$ are not satisfied and thus $\mathtt{O}$ is not satisfied, i.e., $\mathcal{P}^{\mathcal{G}}_{\sigma,\pi,s_0}(\mathtt{O} \cap E) = 0$. If $\overline{E}$ is a null set then this shows the required property.

Otherwise, conditioned under $\overline{E}$, both states s_0 and s_1 are visited infinitely often almost surely, since $p > p/2 > 0$. In the steady state, the probability α of being in s_0 satisfies $\alpha \geq \alpha(1-p/2) + (1-\alpha)p$, and therefore $\alpha \geq 2/3$. Hence the (conditional under $\overline{E}$) expected mean payoff is $\leq \alpha(-1) + (1-\alpha) \leq -1/3$. It follows that $\mathcal{P}^{\mathcal{G}}_{\sigma,\pi,s_0}(\mathtt{MP} < 0 \mid \overline{E}) = 1$ and thus $\mathcal{P}^{\mathcal{G}}_{\sigma,\pi,s_0}(\mathtt{O}_1 \mid \overline{E}) = 0$. Moreover have have $\mathcal{P}^{\mathcal{G}}_{\sigma,\pi,s_0}(\mathsf{FG}X) = 0$, since $p/2 > 0$ and thus $\mathcal{P}^{\mathcal{G}}_{\sigma,\pi,s_0}(\mathtt{O} \cap \overline{E}) = 0$.

Finally, $\mathcal{P}^{\mathcal{G}}_{\sigma,\pi,s_0}(\mathtt{O}) = \mathcal{P}^{\mathcal{G}}_{\sigma,\pi,s_0}(\mathtt{O} \cap E) + \mathcal{P}^{\mathcal{G}}_{\sigma,\pi,s_0}(\mathtt{O} \cap \overline{E}) = 0$. ◀

6 Conclusion and Open Questions

While finite-memory Max strategies for shift-invariant inverse-submixing objectives can be lifted from MDPs to 2-player stochastic games by [25, Theorem 6.1], this yields doubly exponentially many extra memory modes in general and exponentially many extra memory modes when lifting memoryless randomized strategies. In the latter case, this exponential extra memory cannot be avoided in general (Theorem 2).

An open question concerns the exact trade-off between memory and attainment, i.e., how good can randomized strategies with a sub-optimal number of memory modes be in the worst case, relative to strategies with sufficient memory.

The mean-payoff-parity objective $\mathtt{MP} > 0 \cap \mathtt{EPAR}$ in Section 4 provides an interesting example where randomization in strategies drastically reduces the amount of memory required (from doubly exponential to linear), but does not eliminate the need for memory entirely.

An open question is whether low-memory strategies for $\mathtt{MP} > 0 \cap \mathtt{EPAR}$ require randomized memory updates, or whether just randomized actions and deterministic memory updates would suffice.

References

1 Rajeev Alur, Thomas A. Henzinger, and Orna Kupferman. Alternating-time temporal logic. *Journal of the ACM*, 49(5):672–713, 2002. doi:10.1145/585265.585270.

2 Patrick Billingsley. *Probability and measure*. John Wiley & Sons, 2008.

3 Patricia Bouyer, Youssouf Oualhadj, Mickael Randour, and Pierre Vandenhove. Arena-independent finite-memory determinacy in stochastic games. *Logical Methods in Computer Science*, 19, 2023. doi:10.46298/lmcs-19(4:18)2023.

4 Tomás Brázdil, Václav Brožek, Krishnendu Chatterjee, Vojtěch Forejt, and Antonín Kučera. Markov decision processes with multiple long-run average objectives. *Logical Methods in Computer Science*, 10, 2014.

5 Tomás Brázdil, Václav Brozek, and Kousha Etessami. One-Counter Stochastic Games. In Kamal Lodaya and Meena Mahajan, editors, *IARCS Annual Conference on Foundations of Software Technology and Theoretical Computer Science (FSTTCS 2010)*, volume 8 of *Leibniz International Proceedings in Informatics (LIPIcs)*, pages 108–119, Dagstuhl, Germany, 2010. Schloss Dagstuhl – Leibniz-Zentrum für Informatik. Full version at `http://arxiv.org/abs/1009.5636`. `doi:10.4230/LIPIcs.FSTTCS.2010.108`.

6 Tomás Brázdil, Václav Brozek, and Kousha Etessami. One-counter stochastic games. *CoRR*, abs/1009.5636, 2010. (Cited for good reason. LIPIcs version of this paper does not contain the relevant material.). `arXiv:1009.5636`.

7 Krishnendu Chatterjee. Optimal strategy synthesis in stochastic Müller games. In *International Conference on Foundations of Software Science and Computational Structures (FoSSaCS)*, volume 4423 of *Lecture Notes in Computer Science*, pages 138–152. Springer, 2007. `doi:10.1007/978-3-540-71389-0_11`.

8 Krishnendu Chatterjee and Laurent Doyen. Energy and mean-payoff parity Markov decision processes. In *International Symposium on Mathematical Foundations of Computer Science (MFCS)*, volume 6907, pages 206–218, 2011. `doi:10.1007/978-3-642-22993-0_21`.

9 Krishnendu Chatterjee and Laurent Doyen. Games and Markov decision processes with mean-payoff parity and energy parity objectives. In *Mathematical and Engineering Methods in Computer Science (MEMICS)*, volume 7119 of *LNCS*, pages 37–46. Springer, 2011. `doi:10.1007/978-3-642-25929-6_3`.

10 Krishnendu Chatterjee, Laurent Doyen, Hugo Gimbert, and Thomas A. Henzinger. Randomness for free. *Information and Computation*, 245:3–16, 2015. `doi:10.1016/J.IC.2015.06.003`.

11 Krishnendu Chatterjee, Laurent Doyen, Hugo Gimbert, and Youssouf Oualhadj. Perfect-information stochastic mean-payoff parity games. In *International Conference on Foundations of Software Science and Computational Structures (FoSSaCS)*, volume 8412 of *LNCS*, 2014. `doi:10.1007/978-3-642-54830-7_14`.

12 Krishnendu Chatterjee, Thomas A. Henzinger, and Marcin Jurdziński. Mean-payoff parity games. In *Logic in Computer Science (LICS)*, pages 178–187, 2005. `doi:10.1109/LICS.2005.26`.

13 Krishnendu Chatterjee, Marcin Jurdziński, and Thomas A. Henzinger. Simple stochastic parity games. In *Computer Science Logic (CSL)*, volume 2803 of *LNCS*, pages 100–113. Springer, 2003. `doi:10.1007/978-3-540-45220-1_11`.

14 Krishnendu Chatterjee, Marcin Jurdziński, and Thomas A. Henzinger. Quantitative stochastic parity games. In *ACM-SIAM Symposium on Discrete Algorithms (SODA)*, pages 121–130. SIAM, 2004. URL: `http://dl.acm.org/citation.cfm?id=982792.982808`.

15 Krishnendu Chatterjee, Mickael Randour, and Jean-François Raskin. Strategy synthesis for multi-dimensional quantitative objectives. *Acta Informatica*, 51(3-4):129–163, 2014. `doi:10.1007/S00236-013-0182-6`.

16 E.M. Clarke, O. Grumberg, and D. Peled. *Model Checking*. MIT Press, December 1999.

17 Thomas Colcombet and Damian Niwinski. On the positional determinacy of edge-labeled games. *Theoretical Computer Science*, 352(1-3):190–196, 2006. `doi:10.1016/J.TCS.2005.10.046`.

18 Anne Condon. The complexity of stochastic games. *Information and Computation*, 96(2):203–224, 1992. `doi:10.1016/0890-5401(92)90048-K`.

19 Mohan Dantam and Richard Mayr. Mean-payoff-parity and lifting strategies from MDPs to 2-player stochastic games, 2026. `arXiv:2606.19324`.

20 Laure Daviaud, Marcin Jurdziński, and Ranko Lazić. A pseudo-quasi-polynomial algorithm for mean-payoff parity games. In *Logic in Computer Science (LICS)*, pages 325–334, 2018.

21 Luca De Alfaro. *Formal verification of probabilistic systems*. PhD thesis, Stanford University, 1997.

22 Luca De Alfaro and Thomas A. Henzinger. Interface automata. *ACM SIGSOFT Software Engineering Notes*, 26(5):109–120, 2001. `doi:10.1145/503209.503226`.

23 David L. Dill. *Trace theory for automatic hierarchical verification of speed-independent circuits*, volume 24. MIT press Cambridge, 1989.

24 J. Filar and K. Vrieze. *Competitive Markov Decision Processes*. Springer, 1997.

25 H. Gimbert and E. Kelmendi. Submixing and shift-invariant stochastic games. *International Journal of Game Theory*, 52:1179–1214, 2023. doi:10.1007/s00182-023-00860-5.

26 Hugo Gimbert, Youssouf Oualhadj, and Soumya Paul. Computing optimal strategies for Markov decision processes with parity and positive-average conditions. Working paper or preprint, 2011. URL: https://hal.science/hal-00559173/en/.

27 Hugo Gimbert and Wiesław Zielonka. Games where you can play optimally without any memory. In *International Conference on Concurrency Theory (CONCUR)*, volume 3653 of *Lecture Notes in Computer Science*, pages 428–442. Springer, 2005. doi:10.1007/11539452_33.

28 Hugo Gimbert and Wiesław Zielonka. Pure and Stationary Optimal Strategies in Perfect-Information Stochastic Games. Working paper or preprint, December 2009. URL: https://hal.archives-ouvertes.fr/hal-00438359.

29 Kristoffer Arnsfelt Hansen, Rasmus Ibsen-Jensen, and Abraham Neyman. The big match with a clock and a bit of memory. *Mathematics of Operations Research*, 48, 2022.

30 Florian Horn. Random fruits on the Zielonka tree. In *International Symposium on Theoretical Aspects of Computer Science (STACS)*, volume 3 of *LIPIcs*, pages 541–552. Schloss Dagstuhl – Leibniz-Zentrum für Informatik, Germany, 2009. doi:10.4230/LIPIcs.STACS.2009.1848.

31 James C. A. Main and Mickael Randour. Different strokes in randomised strategies: Revisiting Kuhn's theorem under finite-memory assumptions. In *CONCUR*, volume 243 of *LIPIcs*, pages 22:1–22:18. Schloss Dagstuhl – Leibniz-Zentrum für Informatik, 2022. doi:10.4230/LIPIcs.CONCUR.2022.22.

32 A. Maitra and W. Sudderth. Stochastic games with Borel payoffs. In *Stochastic Games and Applications*, pages 367–373. Kluwer, Dordrecht, 2003.

33 Donald A. Martin. The determinacy of Blackwell games. *Journal of Symbolic Logic*, 63(4):1565–1581, 1998. doi:10.2307/2586667.

34 Richard Mayr, Sven Schewe, Patrick Totzke, and Dominik Wojtczak. Simple stochastic games with almost-sure energy-parity objectives are in NP and coNP. In *Proc. of Fossacs*, volume 12650 of *LNCS*, 2021. Extended version on arXiv. arXiv:2101.06989.

35 George H. Mealy. A method for synthesizing sequential circuits. *The Bell System Technical Journal*, 34(5):1045–1079, 1955. doi:10.1002/j.1538-7305.1955.tb03788.x.

36 Amir Pnueli and Roni Rosner. On the synthesis of a reactive module. In *Annual Symposium on Principles of Programming Languages (POPL)*, pages 179–190, 1989. doi:10.1145/75277.75293.

37 Peter J. Ramadge and W. Murray Wonham. Supervisory control of a class of discrete event processes. *SIAM journal on control and optimization*, 25(1):206–230, 1987.

38 Lloyd S. Shapley. Stochastic games. *Proceedings of the national academy of sciences*, 39(10):1095–1100, 1953.

39 Pierre Vandenhove. *Strategy complexity of zero-sum games on graphs*. PhD thesis, Université Paris-Saclay; Université de Mons, 2023. URL: https://theses.hal.science/tel-04095220.

40 Wiesław Zielonka. Infinite games on finitely coloured graphs with applications to automata on infinite trees. *Theoretical Computer Science*, 200(1-2):135–183, 1998.

A Appendix for Section 2

Reachability. We use the syntax and semantics of the LTL operators [16] F (eventually) and G (always) to specify some conditions on plays. A *reachability objective* is defined by a set of target states $T \subseteq S$. A play $\rho = s_0 s_1 \dots$ belongs to $\mathsf{F}\,T$ iff $\exists i \in \mathbb{N}\, s_i \in T$. Similarly, ρ belongs to $\mathsf{F}^{\leq n}T$ (resp. $\mathsf{F}^{\geq n}T$) iff $\exists i \leq n$ (resp. $i \geq n$) such that $s_i \in T$. Dually, the *safety* objective $\mathsf{G}\,T$ consists of all plays which never leave T. We have $\mathsf{G}\,T = \neg\mathsf{F}\neg T$.

Parity. A parity objective is defined via a bounded function $Col : S \to \mathbb{N}$ that assigns non-negative priorities (aka colors) to states. Given an infinite play $\rho = s_0 s_1 \ldots$, let $\mathrm{Inf}(\rho)$ denote the set of numbers that occur infinitely often in the sequence $Col(s_0)Col(s_1)\ldots$. A play ρ satisfies *even parity* w.r.t. Col iff the maximum [1] of $\mathrm{Inf}(\rho)$ is even. Otherwise, ρ satisfies *odd parity*. The objective even parity is denoted by $\mathtt{EPAR}(Col)$ and odd parity is denoted by $\mathtt{OPAR}(Col)$. Most of the time, we implicitly assume that the coloring function is known and just write $\mathtt{EPAR}$ and $\mathtt{OPAR}$. Observe that, given any coloring Col, we have $\overline{\mathtt{EPAR}} = \mathtt{OPAR}$ and $\mathtt{OPAR}(Col) = \mathtt{EPAR}(Col + 1)$ where $Col + 1$ is the function which adds 1 to the color of every state. This justifies to consider only one of the even/odd parity objectives, but, for the sake of clarity, we distinguish these objectives wherever necessary. For any subset T and color c, we denote by $T(c)$, states in T with color c.

Attractors, traps and subgames. A non-empty set $H \subseteq S$ defines a *subgame* iff for all $s \in H \cap S_{\bigcirc}$, $\mathsf{Succ}(s) \subseteq H$ and every state in H has at least one successor in H. The resulting subgame which is well-defined is denoted by $\mathcal{G}[H]$. For a set $T \subseteq S$, the *positive attractor* for player $\odot \in \{\Box, \Diamond\}$ in game $\mathcal{G}$, denoted by $\mathtt{Attr}_{\odot}(T, \mathcal{G})$ is the set of states s where $\odot$ has a strategy to ensure that $\mathsf{F}T$ is satisfied with positive probability when starting from s. For any given T, the positive attractor set and a uniform MD strategy $\tau_{\mathtt{Attr}}$ which satisfies this can be computed in polynomial time. A set $U \subseteq S$ is a *trap* for Min iff for every Min/random state in U, the successors are always in U and there is at least one successor in U for every Max state. One can similarly define a trap for Max. Note that by definition a trap for either player is a subgame. Also, for any set T, $S \setminus \mathtt{Attr}_{\odot}(T, \mathcal{G})$ is a trap for $\odot$ if it is non-empty.

[1] The maximum exists since Col is bounded, even in the general case where S is not finite. However, in this paper, we only consider finite sets of states S.