

# Arena-Independent Finite-Memory Determinacy in Stochastic Games<sup>★</sup>

Patricia Bouyer<sup>1</sup>, Youssouf Oualhadj<sup>2</sup>, Mickael Randour<sup>3</sup>, and Pierre Vandenhover<sup>1,3</sup>

<sup>1</sup> LMF, Université Paris-Saclay, CNRS, ENS Paris-Saclay, France

<sup>2</sup> LACL – Université Paris Est Créteil, Créteil, France

<sup>3</sup> F.R.S.-FNRS & UMONS – Université de Mons, Belgium

**Abstract.** We study stochastic zero-sum games on graphs, which are prevalent tools to model decision-making in presence of an antagonistic opponent in a random environment. In this setting, an important question is the one of *strategy complexity*: what kinds of strategies are sufficient or required to play optimally (e.g., randomization or memory requirements)? Our contributions further the understanding of *arena-independent finite-memory (AIFM) determinacy*, i.e., the study of objectives for which memory is needed, but in a way that only depends on limited parameters of the game graphs. First, we show that objectives for which pure AIFM strategies suffice to play optimally also admit pure AIFM *subgame perfect* strategies. Second, we show that we can reduce the study of objectives for which pure AIFM strategies suffice in two-player stochastic games to the easier study of one-player stochastic games (i.e., Markov decision processes). Third, we *characterize* the sufficiency of AIFM strategies through two intuitive properties of objectives. This work extends a line of research started on deterministic games in [BLO<sup>+</sup>20] to stochastic ones.

**Keywords:** two-player games on graphs, stochastic games, Markov decision processes, finite-memory determinacy, optimal strategies

## 1 Introduction

*Controller synthesis* consists, given a system, an environment, and a specification, in automatically generating a controller of the system that guarantees the specification in the environment. This task is often studied through a game-theoretic lens: the system is a game, the controller is a player, the uncontrollable environment is its adversary, and the specification is a game objective [Ran13]. A *game on graph* consists of a directed graph, called an *arena*, partitioned into two kinds of vertices: some of them are controlled by the system (called *player 1*) and the others by the environment (called *player 2*). Player 1 is given a *game objective* (corresponding to the specification) and must devise a *strategy* (corresponding to the controller) to accomplish the objective or optimize an outcome. The strategy can be seen as a function that dictates the decisions to make in order to react to every possible chain of events. In case of uncertainty in the system or the environment, probability distributions are often used to model transitions in the game graph, giving rise to the *stochastic game* model. We study here *stochastic turn-based zero-sum games on graphs* [Con92], also called perfect-information stochastic games. We also discuss the case of *deterministic games*, which can be seen as a subcase of stochastic games in which no probability distributions are used in transitions.

*Strategy complexity.* A common question underlying all game objectives is the one of *strategy complexity*: how *complex* must optimal strategies be, and how *simple* can optimal strategies be? For each distinct game objective, multiple directions can be investigated, such as the need for randomization [CDGH10] (must optimal strategies make stochastic choices?), the need for memory [GZ05, GZ09, BLO<sup>+</sup>20] (how

<sup>★</sup> Research supported by F.R.S.-FNRS under Grant n° F.4520.18 (ManySynth), and ENS Paris-Saclay visiting professorship (M. Randour, 2019). Mickael Randour is an F.R.S.-FNRS Research Associate and Pierre Vandenhover is an F.R.S.-FNRS Research Fellow.

much information about the past must optimal strategies remember?), or what trade-offs exist between randomization and memory [CdAH04,CRR14,MPR20]. With respect to memory requirements, three cases are typically distinguished: *memoryless-determined objectives*, for which *memoryless* strategies suffice to play optimally; *finite-memory-determined objectives*, for which finite-memory strategies suffice (memory is then usually encoded as a deterministic finite automaton); and objectives for which infinite memory is required. High memory requirements (such as exponential memory and obviously infinite memory) are a major drawback when it comes to implementing controllers; hence specific approaches are often developed to look for *simple* strategies (e.g., [DKQR20]).

Many classical game objectives (reachability [Con92], Büchi and parity [CJH04], energy [BBE10], discounted sum [Sha53]...) are memoryless-determined, both in deterministic and stochastic arenas. Nowadays, multiple general results allow for a more manageable proof for most of these objectives: we mention [Kop06,BFMM11,AR17] for sufficient conditions in deterministic games, and [Gim07,GK14] for similar conditions in one-player and two-player stochastic games. One milestone for memoryless determinacy in deterministic games was achieved by Gimbert and Zielonka [GZ05], who provide *two characterizations* of it: the first one states two necessary and sufficient conditions (called *monotony* and *selectivity*) for memoryless determinacy, and the second one states that memoryless determinacy in both players' *one-player games* suffices for memoryless determinacy in two-player games (we call this result the *one-to-two-player lift*). Together, these characterizations provide a theoretical and practical advance. On the one hand, monotony and selectivity improve the high-level understanding of what conditions well-behaved objectives verify. On the other hand, only having to consider the one-player case thanks to the one-to-two-player lift is of tremendous help in practice. A generalization of the one-to-two-player lift to *stochastic games* was shown also by Gimbert and Zielonka in an unpublished paper [GZ09] and is about memoryless strategies that are *pure* (i.e., not using randomization).

*The need for memory.* Recent research tends to study increasingly complex settings — such as combinations of qualitative/quantitative objectives or of behavioral models — for which finite or infinite memory is often required; see examples in deterministic games [CD12,VCD<sup>+</sup>15,BFRR17,BMR<sup>+</sup>18,BHRR19], Markov decision processes — i.e., one-player stochastic games [RRS15,RRS17,CKK17,BRR17,BDOR20], or stochastic games [CFK<sup>+</sup>13,CD16,CP19,CKWW20,MSTW21]. Motivated by the growing memory requirements of these endeavors, research about strategy complexity often turns toward *finite-memory determinacy*. Proving finite-memory determinacy is sometimes difficult (already in deterministic games, e.g., [BHM<sup>+</sup>17]), and as opposed to memoryless strategies, there are few widely applicable results. We mention [LPR18], which provides sufficient conditions for finite-memory determinacy in Boolean combinations of finite-memory-determined objectives in deterministic games. Results for multi-player non-zero-sum games are also available [LP18].

*Arena-independent finite-memory.* An interesting middle ground between the well-understood memoryless determinacy and the more puzzling finite-memory determinacy was proposed for deterministic games in [BLO<sup>+</sup>20] with the formalization of *arena-independent finite-memory (AIFM) strategies*: strategies that may use memory, but given an objective, for which a *single memory structure* suffices to play optimally in any arena. In practice, this memory structure may depend on parameters of the objective (for instance, largest weight, number of priorities), but not on parameters intrinsically linked to the arena (e.g., number of states or transitions). In stochastic games, objectives for which AIFM strategies notably suffice include the aforementioned memoryless-determined objectives (a very simple memory structure with a single state suffices), generalized reachability games [CKWW20] (the memory depends on the number of objectives, but not on the size of the arena),  $\omega$ -regular games [CH12] (the memory depends on the objective, on the number of priorities). Still, there are objectives for which finite-memory strategies suffice, but with an underlying memory structure depending on parameters of the arena (examples are provided in [MSTW21, Theorem 6]).

Nevertheless, AIFM strategies have a remarkable feature: in deterministic arenas, AIFM generalizations of *both* characterizations from [GZ05] hold, including the one-to-two-player lift [BLO<sup>+</sup>20]. From a practical point of view, it brings techniques usually linked to memoryless determinacy to many finite-memory-determined objectives.

*Contributions.* We provide an overview of desirable properties of objectives in which pure AIFM strategies suffice to play optimally in *stochastic* games, and tools to study them. This entails:

- a proof of a specific feature of objectives for which pure AIFM strategies suffice to play optimally: for such objectives, there also exist pure AIFM *subgame perfect* strategies (Theorem 22), which is a stronger requirement than optimality;
- a more general one-to-two-player lift: we show the equivalence between the existence of pure AIFM optimal strategies in two-player games for both players and the existence of pure AIFM strategies in one-player games, thereby simplifying the proof of memory requirements for many objectives (Theorem 23);
- two intuitive conditions generalizing monotony and selectivity in the stochastic/AIFM case, which are equivalent to the sufficiency of pure AIFM strategies to play optimally in *one-player stochastic arenas* (Theorem 32) for all objectives that can be encoded as *real payoff functions*.

In practice, this last theorem can be used to prove memory requirements in one-player arenas, and then the second theorem can be used to lift these to the two-player case.

These results reinforce both sides on the frontier between AIFM strategies and general finite-memory strategies: on the one hand, objectives for which pure AIFM strategies suffice indeed share interesting properties with objectives for which pure memoryless strategies suffice, rendering their analysis easier, even in the stochastic case; on the other hand, our novel result about SP strategies does not hold for (arena-dependent) finite-memory strategies, and therefore further distinguishes the AIFM case from the finite-memory case.

The one-to-two-player lift for pure AIFM strategies in stochastic games is not surprising, as it holds for pure memoryless strategies in stochastic games [GZ09], and for AIFM strategies in deterministic games [BLO<sup>+</sup>20]. However, although the monotony/selectivity characterization is definitely inspired from the deterministic case [GZ05,BLO<sup>+</sup>20], it had not been formulated for stochastic games, even in the pure memoryless case — its proof involves new technical difficulties to which our improved understanding of subgame perfect strategies brings insight.

All our results can be called in various settings: they can be applied in an independent way for deterministic games and for stochastic games, and they can also consider optimality under restriction to different classes of strategies (allowing or not the use of randomization and of infinite memory).

The proof technique for the one-to-two-player lift shares a similar outline in [GZ05,GZ09,BLO<sup>+</sup>20] and in this paper: it relies on an induction on the number of edges in arenas to show the existence of memoryless optimal strategies. This edge-induction technique is frequently used in comparable ways in other works about memoryless determinacy [Kop06,Gim07,GK14,CD16]. In the AIFM case, the extra challenge consists of applying such an induction to the right set of arenas in order for a result about memoryless strategies to imply something about AIFM strategies. Work in [BLO<sup>+</sup>20] paved the way to neatly overcome this technical hindrance and we were able to factorize the main argument in Lemma 21.

*Outline.* We introduce our framework and notations in Section 2. We discuss AIFM strategies and tools to relate them to memoryless strategies in Section 3, which allows us to prove our result about subgame perfect strategies. The one-to-two-player lift is presented in Section 4, followed by the one-player characterization in Section 5. We provide an illustrative application of our results in Section 6.

## 2 Preliminaries

Let  $C$  be an arbitrary set of *colors*.

*Probabilities.* For a measurable space  $(\Omega, \mathcal{F})$  (resp. a finite set  $\Omega$ ), we write  $\text{Dist}(\Omega, \mathcal{F})$  (resp.  $\text{Dist}(\Omega)$ ) for the set of *probability distributions on*  $(\Omega, \mathcal{F})$  (resp. *on*  $\Omega$ ). For  $\Omega$  a finite set and  $\mu \in \text{Dist}(\Omega)$ , we write  $\text{Supp}(\mu) = \{\omega \in \Omega \mid \mu(\omega) > 0\}$  for the *support of*  $\mu$ .

*Arenas.* We consider stochastic games played by two players, called  $\mathcal{P}_1$  (for player 1) and  $\mathcal{P}_2$  (for player 2), who play in a turn-based fashion on *arenas*.

**Definition 1 (Arena).** A (two-player stochastic turn-based) arena is a tuple  $\mathcal{A} = (S_1, S_2, A, \delta, \text{col})$ , where:

- $S_1$  and  $S_2$  are two disjoint finite sets of states, respectively controlled by  $\mathcal{P}_1$  and  $\mathcal{P}_2$  — we denote  $S = S_1 \uplus S_2$  for the union of all states;
- $A$  is a finite set of actions;
- $\delta: S \times A \rightarrow \text{Dist}(S)$  is a partial function called probabilistic transition function;
- $\text{col}: S \times A \rightarrow C$  is a partial function called coloring function.

For a state  $s \in S$ , we write  $A(s)$  for the set of actions that are available in  $s$ , that is, the set of actions for which  $\delta(s, a)$  is defined. For  $s \in S$ , function  $\text{col}$  must be defined for all pairs  $(s, a)$  such that  $a$  is available in  $s$ . We require that for all  $s \in S$ ,  $A(s) \neq \emptyset$ .

The last condition ensures that there is at least one available action in every state (i.e., arenas are *non-blocking*). For  $s, s' \in S$  and  $a \in A(s)$ , we usually denote  $\delta(s, a, s')$  instead of  $\delta(s, a)(s')$  for the probability to reach  $s'$  in one step by playing  $a$  in  $s$ , and we write  $(s, a, s') \in \delta$  if and only if  $\delta(s, a, s') > 0$ . An interesting subclass of (stochastic) arenas is the class of *deterministic* arenas: an arena  $\mathcal{A} = (S_1, S_2, A, \delta, \text{col})$  is *deterministic* if for all  $s \in S$ ,  $a \in A(s)$ ,  $|\text{Supp}(\delta(s, a))| = 1$ .

Let  $\mathcal{A} = (S_1, S_2, A, \delta, \text{col})$  be an arena. A *play* of  $\mathcal{A}$  is an infinite sequence of states and actions  $s_0 a_1 s_1 a_2 s_2 \dots \in (SA)^\omega$  such that for all  $i \geq 0$ ,  $(s_i, a_{i+1}, s_{i+1}) \in \delta$ . The set of all plays starting in a state  $s \in S$  is denoted  $\text{Plays}(\mathcal{A}, s)$ . A prefix of a play is an element in  $S(AS)^*$  and is called a *history*; the set of all histories starting in a state  $s \in S$  is denoted  $\text{Hists}(\mathcal{A}, s)$ . For  $S' \subseteq S$ , we write  $\text{Plays}(\mathcal{A}, S')$  (resp.  $\text{Hists}(\mathcal{A}, S')$ ) for the unions of  $\text{Plays}(\mathcal{A}, s)$  (resp.  $\text{Hists}(\mathcal{A}, s)$ ) over all states  $s \in S'$ . For  $\rho = s_0 a_1 s_1 \dots a_n s_n$  a history, we write  $\text{out}(\rho)$  for  $s_n$ . For  $i \in \{1, 2\}$ , we write  $\text{Hists}_i(\mathcal{A}, s)$  and  $\text{Hists}_i(\mathcal{A}, S')$  for the corresponding histories  $\rho$  such that  $\text{out}(\rho) \in S_i$ . For  $s \in S$  (resp.  $S' \subseteq S$ ) and  $s' \in S$ , we write  $\text{Hists}(\mathcal{A}, s, s')$  (resp.  $\text{Hists}(\mathcal{A}, S', s')$ ) for the histories in  $\text{Hists}(\mathcal{A}, s)$  (resp.  $\text{Hists}(\mathcal{A}, S')$ ) such that  $\text{out}(\rho) = s'$ .

We write  $\widehat{\text{col}}$  for the extension of  $\text{col}$  to histories and plays: more precisely, for a history  $\rho = s_0 a_1 s_1 \dots a_n s_n$ , we write  $\widehat{\text{col}}(\rho)$  for the finite sequence  $\text{col}(s_0, a_1) \dots \text{col}(s_{n-1}, a_n) \in C^*$ ; for  $\pi = s_0 a_1 s_1 a_2 s_2 \dots$  a play, we write  $\widehat{\text{col}}(\pi)$  for the infinite sequence  $\text{col}(s_0, a_1) \text{col}(s_1, a_2) \dots \in C^\omega$ .

A *one-player arena* of  $\mathcal{P}_i$  is an arena  $\mathcal{A} = (S_1, S_2, A, \delta, \text{col})$  such that for all  $s \in S_{3-i}$ ,  $|A(s)| = 1$ . A one-player arena in our context corresponds to the notion of *Markov decision process (MDP)* often found in the literature [Put94, BK08].

For technical reasons that will be further justified later, we will usually work on arenas where the set of initial states is explicitly specified.

**Definition 2 (Initialized arena).** An initialized arena is a couple  $(\mathcal{A}, S_{\text{init}})$  such that  $\mathcal{A}$  is an arena and  $S_{\text{init}}$  is a non-empty subset of the states of  $\mathcal{A}$ , called the set of initial states. We assume w.l.o.g. that all states of  $\mathcal{A}$  are reachable from  $S_{\text{init}}$  following transitions with positive probabilities in the probabilistic transition function of  $\mathcal{A}$ .

If an initialized arena has only one initial state  $s \in S$ , we write  $(\mathcal{A}, s)$  for  $(\mathcal{A}, \{s\})$ .

We will often compare initialized arenas even if they are not formally defined on the same state space by using a natural definition of *isomorphism*: we say that two initialized arenas  $((S_1, S_2, A, \delta, \text{col}), S_{\text{init}})$  and  $((S'_1, S'_2, A', \delta', \text{col}'), S'_{\text{init}})$  are *isomorphic* if there exist a bijection  $\psi_S: S \rightarrow S'$  and for all  $s \in S$ , a bijection  $\psi_A^s: A(s) \rightarrow A'(\psi_S(s))$  such that  $\psi_S(S_1) = S'_1$ ,  $\psi_S(S_2) = S'_2$ ,  $\psi_S(S_{\text{init}}) = S'_{\text{init}}$ , and for all  $s_1, s_2 \in S$ ,  $a \in A$ , we have  $\delta(s_1, a)(s_2) = \delta'(\psi_S(s_1), \psi_A^{s_1}(a))(\psi_S(s_2))$  and  $\text{col}(s_1, a) = \text{col}'(\psi_S(s_1), \psi_A^{s_1}(a))$ .

We will consider sets (which we call *classes*) of initialized arenas, which are usually denoted by the letter  $\mathfrak{A}$ . Although our results often apply to more fine-grained classes of arenas, typical classes that we will consider consist of all one-player or two-player, deterministic or stochastic initialized arenas. We use *initialized* arenas throughout the paper for technical reasons, but all of our results can be converted to results using only the more classical notion of arena.

*Memory.* To play in games, players use strategies, which can sometimes be efficiently implemented with *finite memory*. We define a classical notion of memory based on complete deterministic automata on *colors*. The goal of using colors instead of states/actions for transitions of the memory is to allow to define memory structures independently of arenas, so that they can be used in all arenas.

**Definition 3 (Memory skeleton).** A memory skeleton is a tuple  $\mathcal{M} = (M, m_{\text{init}}, \alpha_{\text{upd}})$  where  $M$  is a set of memory states,  $m_{\text{init}} \in M$  is an initial state and  $\alpha_{\text{upd}}: M \times C \rightarrow M$  is an update function. We add the following constraint: for all finite sets of colors  $B \subseteq C$ , the number of states reachable from  $m_{\text{init}}$  with transitions provided by  $\alpha_{\text{upd}}|_{M \times B}$  is finite (where  $\alpha_{\text{upd}}|_{M \times B}$  is the restriction of the domain of  $\alpha_{\text{upd}}$  to  $M \times B$ ).

We slightly relax the usual finiteness constraint for the state space by simply requiring that whenever restricted to finitely many colors, the state space of the skeleton is finite. Memory skeletons with a finite state space are all encompassed by this definition, but this also allows some memory skeletons with infinitely many states. For example, if  $C = \mathbb{N}$ , the tuple  $(\mathbb{N}, 0, (m, n) \mapsto \max\{m, n\})$ , which remembers the greatest color seen, is a valid memory skeleton: for any finite  $B \subseteq C$ , we only need to use memory states up to  $\max B$ . However, the tuple  $(\mathbb{N}, 0, (m, n) \mapsto m + n)$  remembering the current sum of all colors seen is *not* a memory skeleton, as infinitely many states are reachable from 0, even if only  $B = \{1\}$  can be used.

We denote  $\widehat{\alpha_{\text{upd}}}: M \times C^* \rightarrow M$  for the natural extension of  $\alpha_{\text{upd}}$  to finite sequences of colors.

It will often be useful to use two memory skeletons in parallel, which is equivalent to using their *product*.

**Definition 4 (Product of skeletons).** Let  $\mathcal{M}^1 = (M^1, m_{\text{init}}^1, \alpha_{\text{upd}}^1)$  and  $\mathcal{M}^2 = (M^2, m_{\text{init}}^2, \alpha_{\text{upd}}^2)$  be two memory skeletons. We define their product  $\mathcal{M}^1 \otimes \mathcal{M}^2$  as the memory skeleton  $(M, m_{\text{init}}, \alpha_{\text{upd}})$  obtained as follows:  $M = M^1 \times M^2$ ,  $m_{\text{init}} = (m_{\text{init}}^1, m_{\text{init}}^2)$ , and, for all  $m^1 \in M^1$ ,  $m^2 \in M^2$ ,  $c \in C$ ,  $\alpha_{\text{upd}}((m^1, m^2), c) = (\alpha_{\text{upd}}^1(m^1, c), \alpha_{\text{upd}}^2(m^2, c))$ .

The update function of the product simply updates both skeletons in parallel.

**Definition 5 (Product initialized arenas).** Let  $(\mathcal{A} = (S_1, S_2, A, \delta, \text{col}), S_{\text{init}})$  be an initialized arena and  $\mathcal{M} = (M, m_{\text{init}}, \alpha_{\text{upd}})$  be a memory skeleton. We define the product initialized arena  $(\mathcal{A}, S_{\text{init}}) \times \mathcal{M}$  as the initialized arena  $((S'_1, S'_2, A, \delta', \text{col}'), S'_{\text{init}})$  where:

- $S'_{\text{init}} = S_{\text{init}} \times \{m_{\text{init}}\}$ ,
- $\delta': (S \times M) \times A \rightarrow \text{Dist}(S \times M)$  is such that for all  $(s, m) \in S \times M$  and  $a \in A$ ,  $\delta'((s, m), a)$  is defined if and only if  $\delta(s, a)$  is defined, in which case  $\delta'((s, m), a, (s', m'))$  is equal to  $\delta(s, a, s')$  if  $\alpha_{\text{upd}}(m, \text{col}(s, a)) = m'$ , and is 0 otherwise — this implies that  $A((s, m)) = A(s)$ ,
- $S'$  is the smallest subset of  $S \times M$  such that  $S'_{\text{init}} \subseteq S'$ , and for all  $(s, m), (s', m') \in S \times M$ ,  $a \in A$ , if  $(s, m) \in S'$  and  $\delta((s, m), a, (s', m'))$  is positive, then  $(s', m') \in S'$ ; we define  $S'_1 = S' \cap (S_1 \times M)$  and  $S'_2 = S' \cap (S_2 \times M)$ ,
- for all  $(s, m) \in S'$  and  $a \in A(s)$ ,  $\text{col}'((s, m), a) = \text{col}(s, a)$ .

A product initialized arena  $(\mathcal{A}, S_{\text{init}}) \times \mathcal{M}$  is an initialized arena with transitions obtained from  $\mathcal{A}$ , with state space enriched with extra information about the current memory state, which is initialized at  $m_{\text{init}}$ . We only keep states that are reachable from  $S'_{\text{init}}$  following transitions of  $\delta'$ , thereby enforcing that in initialized arenas, all states in the state space are reachable from the initial states. Even if memory skeletons have infinitely many states or transitions, product initialized arenas are always finite, as only finitely many colors appear in an initialized arena, and only these colors appear in the product initialized arena.

*Strategies.* We can now define strategies, which are functions describing what each player does in response to every possible scenario.

**Definition 6 (Strategy).** Given an initialized arena  $(\mathcal{A}, S_{\text{init}})$  and  $i \in \{1, 2\}$ , a strategy of  $\mathcal{P}_i$  on  $(\mathcal{A}, S_{\text{init}})$  is a function  $\sigma_i: \text{Hists}_i(\mathcal{A}, S_{\text{init}}) \rightarrow \text{Dist}(A)$  such that for all  $\rho \in \text{Hists}_i(\mathcal{A}, S_{\text{init}})$ ,  $\text{Supp}(\sigma_i(\rho)) \subseteq A(\text{out}(\rho))$ . For  $i \in \{1, 2\}$ , we denote by  $\Sigma_i^G(\mathcal{A}, S_{\text{init}})$  the set of strategies of  $\mathcal{P}_i$  on  $(\mathcal{A}, S_{\text{init}})$ .

The letter  $G$  in  $\Sigma_i^G(\mathcal{A}, S_{\text{init}})$  stands for “general” as it encompasses all possible strategies; we now discuss interesting subclasses of strategies.

A strategy  $\sigma_i$  of  $\mathcal{P}_i$  on  $(\mathcal{A}, S_{\text{init}})$  is *pure* if it does not resort to probability distributions to choose actions, that is, if for all  $\rho \in \text{Hists}_i(\mathcal{A}, S_{\text{init}})$ ,  $|\text{Supp}(\sigma_i(\rho))| = 1$ . If a strategy is not pure, then it is *randomized*.

A strategy  $\sigma_i$  of  $\mathcal{P}_i$  on  $(\mathcal{A}, S_{\text{init}})$  is *memoryless* if every distribution over actions it selects only depends on the current state of the arena, and not on the whole history, that is, if for all  $\rho, \rho' \in \text{Hists}_i(\mathcal{A}, S_{\text{init}})$ ,  $\text{out}(\rho) = \text{out}(\rho')$  implies  $\sigma_i(\rho) = \sigma_i(\rho')$ . A *pure memoryless* strategy of  $\mathcal{P}_i$  can be simply specified as a function  $S_i \rightarrow A$ .

A strategy  $\sigma_i$  of  $\mathcal{P}_i$  on  $(\mathcal{A}, S_{\text{init}})$  is *finite-memory* if it can be encoded as a *Mealy machine*  $\Gamma = (\mathcal{M}, \alpha_{\text{nxt}})$ , with  $\mathcal{M} = (M, m_{\text{init}}, \alpha_{\text{upd}})$  being a memory skeleton and  $\alpha_{\text{nxt}}: S_i \times M \rightarrow \text{Dist}(A)$  being the *next-action function*, which is such that for  $s \in S_i$ ,  $m \in M$ ,  $\text{Supp}(\alpha_{\text{nxt}}(s, m)) \subseteq A(s)$ . Strategy  $\sigma_i$  is encoded by  $\Gamma$  if for all histories  $\rho \in \text{Hists}_i(\mathcal{A}, S_{\text{init}})$ ,

$$\sigma_i(\rho) = \alpha_{\text{nxt}}(\text{out}(\rho), \widehat{\alpha_{\text{upd}}}(m_{\text{init}}, \widehat{\text{col}}(\rho))).$$

If  $\sigma_i$  can be encoded as a Mealy machine  $(\mathcal{M}, \alpha_{\text{nxt}})$ , we say that  $\sigma_i$  is *based on (memory)  $\mathcal{M}$* . If  $\sigma_i$  is *based on  $\mathcal{M}$*  and is *pure*, then the next-action function can be specified as a function  $S_i \times M \rightarrow A$ . Memoryless strategies correspond to finite-memory strategies based on the *trivial memory skeleton*  $\mathcal{M}_{\text{triv}} = (\{m_{\text{init}}\}, m_{\text{init}}, (m_{\text{init}}, c) \mapsto m_{\text{init}})$  that has a single state.

We denote by  $\Sigma_i^{\text{PFM}}(\mathcal{A}, S_{\text{init}})$  (resp.  $\Sigma_i^{\text{P}}(\mathcal{A}, S_{\text{init}})$ ,  $\Sigma_i^{\text{GFM}}(\mathcal{A}, S_{\text{init}})$ ,  $\Sigma_i^{\text{G}}(\mathcal{A}, S_{\text{init}})$ ) the set of pure finite-memory (resp. pure, finite-memory, general) strategies of  $\mathcal{P}_i$  on  $(\mathcal{A}, S_{\text{init}})$ . A *type of strategies* is an element  $X \in \{\text{PFM}, \text{P}, \text{GFM}, \text{G}\}$  corresponding to these subsets.

*Outcomes.* Let  $(\mathcal{A} = (S_1, S_2, A, \delta, \text{col}), S_{\text{init}})$  be an initialized arena. For  $\rho \in \text{Hists}(\mathcal{A}, S_{\text{init}})$ , we denote

$$\text{Cyl}(\rho) = \{\pi \in \text{Plays}(\mathcal{A}, S_{\text{init}}) \mid \rho \text{ is a prefix of } \pi\}$$

for the *cylinder of  $\rho$* , that is, the set of plays (which are infinite) starting with  $\rho$ . We denote by  $\mathcal{F}_{(\mathcal{A}, S_{\text{init}})}$  the smallest  $\sigma$ -algebra generated by all the cylinders of histories in  $\text{Hists}(\mathcal{A}, S_{\text{init}})$ . Hence,  $(\text{Plays}(\mathcal{A}, S_{\text{init}}), \mathcal{F}_{(\mathcal{A}, S_{\text{init}})})$  is a measurable space.

When both players have decided on a strategy and an initial state has been chosen, the generated object is a (finite or countably infinite) Markov chain, which induces a probability distribution on the plays. More precisely, for strategies  $\sigma_1$  of  $\mathcal{P}_1$  and  $\sigma_2$  of  $\mathcal{P}_2$  on  $(\mathcal{A}, S_{\text{init}})$  and  $s \in S_{\text{init}}$ , we denote  $\mathbf{P}_{\mathcal{A}, s}^{\sigma_1, \sigma_2}$  for the probability distribution on  $(\text{Plays}(\mathcal{A}, s), \mathcal{F}_{(\mathcal{A}, s)})$  induced by  $\sigma_1$  and  $\sigma_2$ , starting from state  $s$ . This distribution is defined on the set of cylinders as follows: for  $\rho = s_0 a_1 s_1 \dots a_n s_n \in \text{Hists}(\mathcal{A}, s)$ ,

$$\mathbf{P}_{\mathcal{A}, s}^{\sigma_1, \sigma_2}[\text{Cyl}(\rho)] = \prod_{j=1}^n \sigma_{i_j}(s_0 \dots a_{j-1} s_{j-1})(a_j) \cdot \delta(s_{j-1}, a_j, s_j),$$

where  $i_j = 1$  if  $s_{j-1} \in S_1$ , and  $i_j = 2$  if  $s_{j-1} \in S_2$ . This pre-measure can be uniquely extended to  $(\text{Plays}(\mathcal{A}, s), \mathcal{F}_{(\mathcal{A}, s)})$  by Carathéodory’s extension theorem [Dur19, Theorem A.1.3], as the class of cylinders is a semi-ring of sets that generates the whole  $\sigma$ -algebra.

Similarly, we define  $\mathcal{F}$  to be the smallest  $\sigma$ -algebra on  $C^\omega$  generated by the set of all cylinders on  $C$ . We can extend  $\widehat{\text{col}}$  to distributions over  $(\text{Plays}(\mathcal{A}, s), \mathcal{F}_{(\mathcal{A}, s)})$ : for  $\mu \in \text{Dist}(\text{Plays}(\mathcal{A}, s), \mathcal{F}_{(\mathcal{A}, s)})$ , we write  $\widehat{\text{col}}(\mu)$  for the distribution  $\mu \circ \widehat{\text{col}}^{-1} \in \text{Dist}(C^\omega, \mathcal{F})$ . In particular, every probability distribution  $\mathbf{P}_{\mathcal{A}, s}^{\sigma_1, \sigma_2}$  naturally induces a probability distribution  $\widehat{\text{col}}(\mathbf{P}_{\mathcal{A}, s}^{\sigma_1, \sigma_2})$  over  $(C^\omega, \mathcal{F})$  through the  $\widehat{\text{col}}$  function, which we denote  $\mathbf{Pc}_{\mathcal{A}, s}^{\sigma_1, \sigma_2}$ .

*Preferences.* To specify each player’s objective, or preference, we use the general notion of *preference relation*.

**Definition 7 (Preference relation).** A preference relation  $\sqsubseteq$  (on  $C$ ) is a total preorder over  $\text{Dist}(C^\omega, \mathcal{F})$ .



The idea is that  $\mathcal{P}_1$  favors the distributions in  $\text{Dist}(C^\omega, \mathcal{F})$  that are the largest for  $\sqsubseteq$ , and as we are studying *zero-sum* games,  $\mathcal{P}_2$  favors the distributions that are the smallest for  $\sqsubseteq$ . Equivalently,  $\mathcal{P}_2$ 's goal is to obtain the largest distribution for the *inverse preference relation*  $\sqsubseteq^{-1}$ , defined as  $\mu \sqsubseteq^{-1} \mu'$  if and only if  $\mu' \sqsubseteq \mu$ . For  $\sqsubseteq$  a preference relation and  $\mu, \mu' \in \text{Dist}(C^\omega, \mathcal{F})$ , we write  $\mu \sqsubset \mu'$  if  $\mu \sqsubseteq \mu'$  and  $\mu' \not\sqsubseteq \mu$ .

Depending on the context, it might not be necessary to define a preference relation as *total*: it is sufficient to order distributions that can arise as an element  $P_{\mathcal{A},s}^{\sigma_1,\sigma_2}$  in the context. For example, in the specific case of deterministic games in which only pure strategies are considered, all distributions that arise are always Dirac distributions on a single infinite word in  $C^\omega$ . In this context, it is therefore sufficient to define a total preorder over all Dirac distributions (which we can then see as infinite words, giving a definition of preference relation similar to [GZ05,BLO<sup>+</sup>20]). Another reasonable constraint comes from the fact that we mostly consider finite arenas, in which only finitely many colors appear: distributions that generate infinite words with infinitely many colors should therefore not be considered.

*Example 8.* We give three examples corresponding to three different ways to encode preference relations. First, a preference relation can be induced by an event  $W \in \mathcal{F}$  called a *winning condition*, which consists of infinite sequences of colors. The objective of  $\mathcal{P}_1$  is to maximize the probability that the event  $W$  happens. An event  $W$  naturally induces a preference relation  $\sqsubseteq_W$  such that for  $\mu, \mu' \in \text{Dist}(C^\omega, \mathcal{F})$ ,  $\mu \sqsubseteq_W \mu'$  if and only if  $\mu(W) \leq \mu'(W)$ . For  $C = \mathbb{N}$ , we give the example of the *weak parity* winning condition  $W_{\text{wp}}$  [Tho08], defined as

$$W_{\text{wp}} = \{c_1 c_2 \dots \in C^\omega \mid \max_{j \geq 1} c_j \text{ exists and is even}\}.$$

In finite arenas, the value  $\max_{j \geq 1} c_j$  always exists, as there are only finitely many colors that appear. This is different from the classical parity condition, which requires the maximal color *seen infinitely often* to be even, and not just the maximal color seen. In particular,  $W_{\text{wp}}$  is not prefix-independent.

A preference relation can also be induced by a Borel (*real*) *payoff function*  $f: C^\omega \rightarrow \mathbb{R}$ . For example, if  $C = \mathbb{Q}$  and  $\lambda \in (0, 1) \cap \mathbb{Q}$ , a classical payoff function [Sha53] is the *discounted sum*  $\text{Disc}_\lambda$ , defined for  $c_1 c_2 \dots \in C^\omega$  as

$$\text{Disc}_\lambda(c_1 c_2 \dots) = \lim_n \sum_{i=0}^n \lambda^i \cdot c_{i+1}.$$

The goal of  $\mathcal{P}_1$  is to maximize the expected value of  $f$ , which is defined for a probability distribution  $\mu \in \text{Dist}(C^\omega, \mathcal{F})$  as  $E_\mu[f] = \int f d\mu$ . A payoff function  $f$  naturally induces a preference relation  $\sqsubseteq_f$ : for  $\mu_1, \mu_2 \in \text{Dist}(C^\omega, \mathcal{F})$ ,  $\mu_1 \sqsubseteq_f \mu_2$  if and only if  $E_{\mu_1}[f] \leq E_{\mu_2}[f]$ . Payoff functions are more general than winning conditions: for  $W$  a winning condition, the preference relation induced by the indicator function of  $W$ , which is a payoff function, corresponds to the preference relation induced by  $W$ .

It is also possible to specify preference relations that cannot be expressed as a payoff function. An example is given in [CFK<sup>+</sup>12]: we assume that the goal of  $\mathcal{P}_1$  is to see color  $c \in C$  with probability *precisely*  $\frac{1}{2}$ . We denote the event of seeing color  $c$  as  $\diamond c \in \mathcal{F}$ . Then for  $\mu, \mu' \in \text{Dist}(C^\omega, \mathcal{F})$ , if  $\mu(\diamond c) \neq \frac{1}{2}$  and  $\mu'(\diamond c) = \frac{1}{2}$ ,  $\mu \sqsubset \mu'$ .  $\triangleleft$

Combining an initialized arena, describing how the players interact with each other, and a preference relation, describing both players' objectives, defines an *initialized game*.

**Definition 9 (Initialized game).** A (two-player stochastic turn-based zero-sum) initialized game is a tuple  $\mathcal{G} = (\mathcal{A}, S_{\text{init}}, \sqsubseteq)$ , where  $(\mathcal{A}, S_{\text{init}})$  is an initialized arena and  $\sqsubseteq$  is a preference relation.

*Optimality of strategies.* Let  $\mathcal{G} = (\mathcal{A}, S_{\text{init}}, \sqsubseteq)$  be an initialized game and  $X \in \{\text{PFM}, \text{P}, \text{GFM}, \text{G}\}$  be a type of strategies. For  $s \in S_{\text{init}}$ ,  $\sigma_1 \in \Sigma_1^X(\mathcal{A}, S_{\text{init}})$ , we define

$$\text{UCol}_{\sqsubseteq}^X(\mathcal{A}, s, \sigma_1) = \{\mu \in \text{Dist}(C^\omega, \mathcal{F}) \mid \exists \sigma_2 \in \Sigma_2^X(\mathcal{A}, s), P_{\mathcal{A},s}^{\sigma_1,\sigma_2} \sqsubseteq \mu\}.$$

The set  $\text{UCol}_{\sqsubseteq}^{\mathbf{X}}(\mathcal{A}, s, \sigma_1)$  corresponds to all the distributions that are at least as good for  $\sqsubseteq$  as a distribution that  $\mathcal{P}_2$  can induce by playing a strategy  $\sigma_2$  of type  $\mathbf{X}$  against  $\sigma_1$ ; this set is upward-closed w.r.t.  $\sqsubseteq$ . We can define a similar operator for strategies of  $\mathcal{P}_2$ : for  $s \in S_{\text{init}}$ ,  $\sigma_2 \in \Sigma_2^{\mathbf{X}}(\mathcal{A}, S_{\text{init}})$ ,

$$\text{DCol}_{\sqsubseteq}^{\mathbf{X}}(\mathcal{A}, s, \sigma_2) = \{\mu \in \text{Dist}(C^\omega, \mathcal{F}) \mid \exists \sigma_1 \in \Sigma_1^{\mathbf{X}}(\mathcal{A}, s), \mu \sqsubseteq \text{Pc}_{\mathcal{A}, s}^{\sigma_1, \sigma_2}\}.$$

For  $\sigma_1, \sigma'_1 \in \Sigma_1^{\mathbf{X}}(\mathcal{A}, S_{\text{init}})$ , we say that  $\sigma_1$  is *at least as good as  $\sigma'_1$  from  $s \in S_{\text{init}}$  under  $\mathbf{X}$  strategies* if

$$\text{UCol}_{\sqsubseteq}^{\mathbf{X}}(\mathcal{A}, s, \sigma_1) \subseteq \text{UCol}_{\sqsubseteq}^{\mathbf{X}}(\mathcal{A}, s, \sigma'_1).$$

This inclusion means that the best replies of  $\mathcal{P}_2$  against  $\sigma'_1$  yield an outcome that is at least as bad (w.r.t.  $\sqsubseteq$ ) as the best replies of  $\mathcal{P}_2$  against  $\sigma_1$ .

Symmetrically for  $\sigma_2, \sigma'_2 \in \Sigma_2^{\mathbf{X}}(\mathcal{A}, S_{\text{init}})$ , we say that  $\sigma_2$  is *at least as good as  $\sigma'_2$  from  $s \in S_{\text{init}}$  under  $\mathbf{X}$  strategies* if

$$\text{DCol}_{\sqsubseteq}^{\mathbf{X}}(\mathcal{A}, s, \sigma_2) \subseteq \text{DCol}_{\sqsubseteq}^{\mathbf{X}}(\mathcal{A}, s, \sigma'_2).$$

**Definition 10 (Optimal strategy).** Let  $\mathcal{G} = (\mathcal{A}, S_{\text{init}}, \sqsubseteq)$  be an initialized game and  $\mathbf{X} \in \{\text{PFM}, \text{P}, \text{GFM}, \text{G}\}$  be a type of strategies. A strategy  $\sigma_i \in \Sigma_i^{\mathbf{X}}(\mathcal{A}, S_{\text{init}})$  is  $\mathbf{X}$ -optimal in  $\mathcal{G}$  if it is at least as good under  $\mathbf{X}$  strategies as any other strategy in  $\Sigma_i^{\mathbf{X}}(\mathcal{A}, S_{\text{init}})$  from all  $s \in S_{\text{init}}$ .

When the considered preference relation  $\sqsubseteq$  is clear from the context, we often talk about  $\mathbf{X}$ -optimality in an initialized arena  $(\mathcal{A}, S_{\text{init}})$  to refer to  $\mathbf{X}$ -optimality in the initialized game  $(\mathcal{A}, S_{\text{init}}, \sqsubseteq)$ . Notice that given two *isomorphic* initialized arenas, there is an obvious bijection between the strategies on them, and the properties of the strategies (pure, memoryless, finite-memory,  $\mathbf{X}$ -optimal...) are preserved through this bijection.

Our goal will be to understand, given a preference relation, a class of arenas, and a type of strategies, what kinds of strategies are sufficient to play optimally. In the following definition, abbreviations *AIFM* and *FM* stand respectively for *arena-independent finite-memory* and *finite-memory*.

**Definition 11 (Sufficiency of strategies).** Let  $\sqsubseteq$  be a preference relation,  $\mathfrak{A}$  be a class of initialized arenas,  $\mathbf{X} \in \{\text{PFM}, \text{P}, \text{GFM}, \text{G}\}$  be a type of strategies, and  $\mathcal{M}$  be a memory skeleton.

- We say that pure memoryless strategies (resp. pure strategies based on  $\mathcal{M}$ ) suffice to play  $\mathbf{X}$ -optimally in  $\mathfrak{A}$  for  $\mathcal{P}_1$  if for all  $(\mathcal{A}, S_{\text{init}}) \in \mathfrak{A}$ ,  $\mathcal{P}_1$  has a pure memoryless strategy (resp. a pure strategy based on  $\mathcal{M}$ ) that is  $\mathbf{X}$ -optimal in  $(\mathcal{A}, S_{\text{init}})$ .
- We say that pure AIFM strategies suffice to play  $\mathbf{X}$ -optimally in  $\mathfrak{A}$  for  $\mathcal{P}_1$  if there exists a memory skeleton  $\mathcal{M}$  such that pure strategies based on  $\mathcal{M}$  suffice to play  $\mathbf{X}$ -optimally in  $\mathfrak{A}$  for  $\mathcal{P}_1$ .
- We say that pure FM strategies suffice to play  $\mathbf{X}$ -optimally in  $\mathfrak{A}$  for  $\mathcal{P}_1$  if for all  $(\mathcal{A}, S_{\text{init}}) \in \mathfrak{A}$ , there exists a memory skeleton  $\mathcal{M}$  such that  $\mathcal{P}_1$  has a pure strategy based on  $\mathcal{M}$  that is  $\mathbf{X}$ -optimal in  $(\mathcal{A}, S_{\text{init}})$ .

If  $\mathfrak{A}$  is clear in the context (typically all initialized deterministic or stochastic arenas), we often omit it. When no type of strategies is specified, it means that we consider optimality against all (general) strategies.

Since memoryless strategies are a specific kind of finite-memory strategies based on the same memory skeleton  $\mathcal{M}_{\text{triv}}$ , the sufficiency of pure memoryless strategies is equivalent to the sufficiency of pure strategies based on  $\mathcal{M}_{\text{triv}}$ , and is therefore just a specific case of the sufficiency of pure AIFM strategies. Notice the difference between the order of quantifiers for AIFM and FM strategies: the sufficiency of pure AIFM strategies implies the sufficiency of pure FM strategies, but the opposite is false, as we show in the following example.

*Example 12.* Let us consider the *energy parity* winning condition studied in deterministic arenas in [CD12]. We do not explain this winning condition in detail, but comment on its memory requirements and how it illustrates the difference between AIFM and FM strategies. For this objective, in



deterministic arenas, pure memoryless strategies suffice to play optimally for  $\mathcal{P}_2$ .<sup>4</sup> On the other hand,  $\mathcal{P}_1$  can play optimally with finite memory, but the memory needed *depends on the number of states of the arena*; it is not arena-independent. Therefore, pure FM strategies suffice in deterministic games for  $\mathcal{P}_1$ , but not pure AIFM strategies:  $\mathcal{P}_1$  needs to change its memory skeleton depending on the arena, and no single memory skeleton is sufficient to play optimally in all deterministic arenas (even with a fixed and finite number of colors).

Interestingly, it is shown in [MSTW17] that the same winning condition needs infinite memory in (even one-player) stochastic arenas for  $\mathcal{P}_1$ , which shows that the sufficiency of pure FM strategies in deterministic arenas does not imply the sufficiency of pure FM strategies in stochastic arenas.

Now let us reconsider the weak parity winning condition  $W_{wp}$  introduced in Example 8: the goal of  $\mathcal{P}_1$  is to maximize the probability that the greatest color seen is even. As will be proven formally in Section 6 thanks to the results of this article, to play optimally in any stochastic game, it is sufficient for both players to remember the greatest color already seen, which can be implemented by the memory skeleton  $\mathcal{M}_{\max} = (\mathbb{N}, 0, (m, n) \mapsto \max\{m, n\})$ . As explained above, this memory skeleton has an infinite state space, but as there are only finitely many colors in every (finite) arena, only a finite part of the skeleton is sufficient to play optimally in a given arena. The size of the skeleton used for a fixed arena depends on the appearing colors, but for a fixed number of colors, it does not depend on parameters of the arena (such as its state and action spaces). Therefore pure AIFM strategies suffice to play optimally for both players, and more precisely pure strategies based on  $\mathcal{M}_{\max}$  suffice for both players.  $\triangleleft$

We define a second stronger notion related to optimality of strategies, which is the notion of *subgame perfect strategy*: a strategy is *subgame perfect* in a game if it reacts optimally to all histories consistent with the arena, even histories not consistent with the strategy itself, or histories that only a non-rational adversary would play. To do so, we first need an extra definition.

**Definition 13 (Shifted distributions, strategies and preference relations).** For  $w \in C^*$ ,  $\mu \in \text{Dist}(C^\omega, \mathcal{F})$ , we define the shifted distribution  $w\mu$  as the distribution such that for an event  $E \in \mathcal{F}$ ,  $w\mu(E) = \mu(\{w' \in C^\omega \mid ww' \in E\})$ .

Let  $(\mathcal{A}, S_{\text{init}})$  be an arena, and  $\sigma_i \in \Sigma_i^G(\mathcal{A}, S_{\text{init}})$ . For  $\rho = s_0 a_1 s_1 \dots a_n s_n \in \text{Hists}(\mathcal{A}, S_{\text{init}})$ , we define the shifted strategy  $\sigma_i[\rho] \in \Sigma_i^G(\mathcal{A}, \text{out}(\rho))$  as the strategy such that, for  $\rho' = s_n a_{n+1} s_{n+1} \dots a_m s_m \in \text{Hists}_i(\mathcal{A}, \text{out}(\rho))$ ,  $\sigma_i[\rho](\rho') = \sigma_i(s_0 a_1 s_1 \dots a_m s_m)$ .

For  $\sqsubseteq$  a preference relation and  $w \in C^*$ , we define the shifted preference relation  $\sqsubseteq_{[w]}$  as the preference relation such that for  $\mu, \mu' \in \text{Dist}(C^\omega, \mathcal{F})$ ,  $\mu \sqsubseteq_{[w]} \mu'$  if and only if  $w\mu \sqsubseteq w\mu'$ .

**Definition 14 (Subgame perfect strategy).** Let  $\mathcal{G} = (\mathcal{A}, S_{\text{init}}, \sqsubseteq)$  be an initialized game and  $X \in \{\text{PFM}, \text{P}, \text{GFM}, \text{G}\}$  be a type of strategies. A strategy  $\sigma_i \in \Sigma_i^X(\mathcal{A}, S_{\text{init}})$  is  $X$ -subgame perfect ( $X$ -SP) in  $\mathcal{G}$  if for all  $\rho \in \text{Hists}(\mathcal{A}, S_{\text{init}})$ , shifted strategy  $\sigma_i[\rho]$  is  $X$ -optimal in the initialized game  $(\mathcal{A}, \text{out}(\rho), \sqsubseteq_{[\widehat{\text{col}}(\rho)])}$ .

Strategies that are  $X$ -SP are in particular  $X$ -optimal; the converse is not true in general.

For technical reasons, we will use the notion of *equilibrium* as a tool in proofs to show the existence of optimal strategies.

**Definition 15 (Nash, SP equilibrium).** Let  $\mathcal{G} = (\mathcal{A}, S_{\text{init}}, \sqsubseteq)$  be an initialized game and  $X \in \{\text{PFM}, \text{P}, \text{GFM}, \text{G}\}$  be a type of strategies. A pair of strategies  $(\sigma_1, \sigma_2) \in \Sigma_1^X(\mathcal{A}, S_{\text{init}}) \times \Sigma_2^X(\mathcal{A}, S_{\text{init}})$  is an  $X$ -Nash equilibrium ( $X$ -NE) in  $\mathcal{G}$  if for all  $\sigma'_1 \in \Sigma_1^X(\mathcal{A}, S_{\text{init}})$ , for all  $\sigma'_2 \in \Sigma_2^X(\mathcal{A}, S_{\text{init}})$ , for all  $s \in S_{\text{init}}$ ,

$$\text{Pc}_{\mathcal{A},s}^{\sigma'_1, \sigma_2} \sqsubseteq \text{Pc}_{\mathcal{A},s}^{\sigma_1, \sigma_2} \sqsubseteq \text{Pc}_{\mathcal{A},s}^{\sigma_1, \sigma'_2}.$$

We say that  $(\sigma_1, \sigma_2)$  is an  $X$ -subgame perfect equilibrium ( $X$ -SPE) if for all  $\rho \in \text{Hists}(\mathcal{A}, S_{\text{init}})$ , the pair of strategies  $(\sigma_1[\rho], \sigma_2[\rho])$  is an  $X$ -Nash equilibrium in  $(\mathcal{A}, \text{out}(\rho), \sqsubseteq_{[\widehat{\text{col}}(\rho)])}$ .

<sup>4</sup> Work on deterministic arenas only consider P-optimality, but as pure strategies suffice for Borel objectives [Mar75], this implies G-optimality.

A pair of strategies  $(\sigma_1, \sigma_2)$  is thus an X-NE if no player has any interest in unilaterally deviating from its strategy (using strategies of type X), as the induced probability distribution on the colors (or equivalently, the induced Markov chain) would not be better for this player than not changing the strategy. In the zero-sum context, if  $(\sigma_1, \sigma_2)$  is an X-NE (resp. X-SPE), then both  $\sigma_1$  and  $\sigma_2$  are X-optimal (resp. SP).

We say that a pair of strategies  $(\sigma_1, \sigma_2)$  is pure (resp. randomized, memoryless, based on  $\mathcal{M}$ ) if both  $\sigma_1$  and  $\sigma_2$  are pure (resp. randomized, memoryless, based on  $\mathcal{M}$ ). Thanks to the fact that we consider zero-sum games, we can use the following handy result about Nash equilibria.

**Lemma 16.** *Let  $\mathcal{G} = (\mathcal{A}, S_{\text{init}}, \sqsubseteq)$  be a game and  $X \in \{\text{PFM}, \text{P}, \text{GFM}, \text{G}\}$  be a type of strategies. Let  $(\sigma_1^a, \sigma_2^a), (\sigma_1^b, \sigma_2^b) \in \Sigma_1^X(\mathcal{A}, S_{\text{init}}) \times \Sigma_2^X(\mathcal{A}, S_{\text{init}})$  be two X-NE (resp. X-SPE) in  $\mathcal{G}$ . Then  $(\sigma_1^a, \sigma_2^b)$  is also an X-NE (resp. X-SPE) in  $\mathcal{G}$ .*

*Proof.* A very similar proof appears in [BLO<sup>+</sup>20, Lemma 2]. We do the proof for NE, and the result about SPE follows since its definition uses the notion of NE. We need to prove that for all  $\sigma'_1 \in \Sigma_1^X(\mathcal{A}, S_{\text{init}})$ , for all  $\sigma'_2 \in \Sigma_2^X(\mathcal{A}, S_{\text{init}})$ , for all  $s \in S_{\text{init}}$ ,

$$\text{Pc}_{\mathcal{A},s}^{\sigma'_1, \sigma'_2} \sqsubseteq \text{Pc}_{\mathcal{A},s}^{\sigma_1^a, \sigma_2^b} \sqsubseteq \text{Pc}_{\mathcal{A},s}^{\sigma_1^a, \sigma'_2}. \quad (1)$$

Since  $(\sigma_1^a, \sigma_2^a)$  is an X-NE, we know that

$$\text{Pc}_{\mathcal{A},s}^{\sigma_1^b, \sigma_2^a} \sqsubseteq \text{Pc}_{\mathcal{A},s}^{\sigma_1^a, \sigma_2^a} \sqsubseteq \text{Pc}_{\mathcal{A},s}^{\sigma_1^a, \sigma_2^b},$$

instantiating  $\sigma'_1$  and  $\sigma'_2$  as  $\sigma_1^b$  and  $\sigma_2^b$  in the definition of X-NE. Similarly, since  $(\sigma_1^b, \sigma_2^b)$  is an X-NE, we know that

$$\text{Pc}_{\mathcal{A},s}^{\sigma_1^a, \sigma_2^b} \sqsubseteq \text{Pc}_{\mathcal{A},s}^{\sigma_1^b, \sigma_2^b} \sqsubseteq \text{Pc}_{\mathcal{A},s}^{\sigma_1^b, \sigma_2^a},$$

instantiating  $\sigma'_1$  and  $\sigma'_2$  as  $\sigma_1^a$  and  $\sigma_2^a$  in the definition of X-NE.

One can see from the last two lines that all six probability distributions over sequences of colors are equivalent w.r.t.  $\sqsubseteq$  as the inequalities form a cycle. Now, let  $\sigma'_1 \in \Sigma_1^X(\mathcal{A}, S_{\text{init}})$  and  $\sigma'_2 \in \Sigma_2^X(\mathcal{A}, S_{\text{init}})$ . Since  $(\sigma_1^a, \sigma_2^a)$  and  $(\sigma_1^b, \sigma_2^b)$  are both X-NE, and since  $\text{Pc}_{\mathcal{A},s}^{\sigma_1^a, \sigma_2^b}$  is equivalent w.r.t.  $\sqsubseteq$  to both  $\text{Pc}_{\mathcal{A},s}^{\sigma_1^a, \sigma_2^a}$  and  $\text{Pc}_{\mathcal{A},s}^{\sigma_1^b, \sigma_2^b}$ , we obtain

$$\text{Pc}_{\mathcal{A},s}^{\sigma'_1, \sigma'_2} \sqsubseteq \text{Pc}_{\mathcal{A},s}^{\sigma_1^b, \sigma_2^b} \sqsubseteq \text{Pc}_{\mathcal{A},s}^{\sigma_1^a, \sigma_2^b} \sqsubseteq \text{Pc}_{\mathcal{A},s}^{\sigma_1^a, \sigma_2^a} \sqsubseteq \text{Pc}_{\mathcal{A},s}^{\sigma_1^b, \sigma_2^a},$$

thus (1) is verified.  $\square$

*Operations on arenas.* We introduce two operations on arenas that we will use multiple times through the course of this article.

For  $\mathcal{A}$  an arena,  $w \in C^*$ , and  $s$  a state of  $\mathcal{A}$ , we write

$$\mathcal{A}_{w \rightsquigarrow s} \quad (2)$$

for the *prefix-extended arena* that consists of arena  $\mathcal{A}$  with an extra “chain” of states leading up to  $s$  with the same colors as  $w$ . Formally, if  $\mathcal{A} = (S_1, S_2, A, \delta, \text{col})$ ,  $w = c_1 c_2 \dots c_n$ , and  $s \in S$ , we define  $\mathcal{A}_{w \rightsquigarrow s}$  as the arena  $(S'_1, S'_2, A', \delta', \text{col}')$  where  $S'_1 = S_1 \uplus \{s_0^w, \dots, s_{n-1}^w\}$ ,  $S'_2 = S_2$ ;  $A' = A \uplus \{a^w\}$ ,  $A'|_S = A$ , and for  $i$ ,  $0 \leq i \leq n-1$ ,  $A'(s_i^w) = \{a^w\}$ ;  $\delta'|_{S \times A} = \delta$ , for  $i$ ,  $0 \leq i < n-1$ ,  $\delta'(s_i^w, a^w, s_{i+1}^w) = 1$ , and  $\delta'(s_{n-1}^w, a^w, s) = 1$ ;  $\text{col}'|_{S \times A} = \text{col}$ , and for  $i$ ,  $0 \leq i \leq n-1$ ,  $\text{col}'(s_i^w, a^w) = c_{i+1}$ .

If  $\mu = \text{Pc}_{\mathcal{A},s}^{\sigma_1, \sigma_2}$  is induced by strategies  $\sigma_1$  and  $\sigma_2$  on some initialized arena  $(\mathcal{A}, s)$ , notice that the shifted distribution  $w\mu$  equals  $\text{Pc}_{\mathcal{A}_{w \rightsquigarrow s}, s_0^w}^{\sigma'_1, \sigma'_2}$ , where  $\sigma'_1$  play the only available action  $a^w$  until it reaches  $s$ , and then  $\sigma'_1$  and  $\sigma'_2$  play like  $\sigma_1$  and  $\sigma_2$ , ignoring they ever saw  $w$ .

For two arenas  $\mathcal{A}_1$  and  $\mathcal{A}_2$  with disjoint state spaces, if  $s_1$  and  $s_2$  are two states controlled by  $\mathcal{P}_1$  that are respectively in  $\mathcal{A}_1$  and  $\mathcal{A}_2$  with disjoint sets of available actions, we write

$$(\mathcal{A}_1, s_1) \sqcup (\mathcal{A}_2, s_2) \quad (3)$$

for the *merged arena* in which  $s_1$  and  $s_2$  are merged, and everything else is kept the same. The merged state which comes from the merge of  $s_1$  and  $s_2$  is usually called  $t$ . Formally, let  $\mathcal{A}_1 = (S_1^1, S_2^1, A^1, \delta^1, \text{col}^1)$ ,  $\mathcal{A}_2 = (S_1^2, S_2^2, A^2, \delta^2, \text{col}^2)$ ,  $s_1 \in S_1^1$ , and  $s_2 \in S_1^2$ . We assume that  $S^1 \cap S^2 = \emptyset$  and that  $A(s_1) \cap A(s_2) = \emptyset$ . We define  $(\mathcal{A}_1, s_1) \sqcup (\mathcal{A}_2, s_2)$  as the arena  $(S_1, S_2, A, \delta, \text{col})$  with  $S_1 = S_1^1 \sqcup S_1^2 \sqcup \{t\} \setminus \{s_1, s_2\}$ ,  $S_2 = S_2^1 \sqcup S_2^2$ ;  $A(t) = A^1(s_1) \sqcup A^2(s_2)$  and all the other available actions are kept the same as in the original arenas; for  $i \in \{1, 2\}$ ,  $\delta(t, a) = \delta^i(t, a)$  if  $a \in A(s_i)$  and all the other transitions are kept the same as in the original arenas (with transitions going to  $s_1$  or  $s_2$  being directed to  $t$ ); for  $i \in \{1, 2\}$ ,  $\text{col}(t, a) = \text{col}^i(t, a)$  if  $a \in A(s_i)$  and all the other colors are kept the same as in the original arenas. A symmetrical definition can be written if  $s_1$  and  $s_2$  are both controlled by  $\mathcal{P}_2$ .

In practice, we often consider classes  $\mathfrak{A}$  of initialized arenas that are *closed with respect to some operation* — we specify the exact meaning for each operation we will use here:

- for  $\mathcal{M}$  a memory skeleton,  $\mathfrak{A}$  is *closed by product with  $\mathcal{M}$*  if for all  $(\mathcal{A}, S_{\text{init}}) \in \mathfrak{A}$ ,  $(\mathcal{A}, S_{\text{init}}) \ltimes \mathcal{M} \in \mathfrak{A}$ ;
- $\mathfrak{A}$  is *closed by prefix-extension* if for all  $(\mathcal{A}, S_{\text{init}}) \in \mathfrak{A}$ , for all  $w \in C^*$ , for all states  $s$  of  $\mathcal{A}$ ,  $(\mathcal{A}_{w \rightsquigarrow s}, S_{\text{init}} \cup \{s_0^w\}) \in \mathfrak{A}$ .

Standard classes of arenas are all closed by these operations: we give as examples the classes of all initialized one-player deterministic arenas of  $\mathcal{P}_1$ , one-player stochastic arenas of  $\mathcal{P}_1$ , two-player deterministic arenas, and two-player stochastic arenas (corresponding respectively to the classes of 1-player,  $1\frac{1}{2}$ -player, 2-player,  $2\frac{1}{2}$ -player arenas often found in the literature). Throughout the article, we state results with the exact required closure properties for generality, but most applications use such standard classes.

### 3 Coverability and subgame perfect strategies

In this section, we establish a few key results about memory and playing optimally. The main tool is given by Lemma 21, which can be used to reduce questions about the sufficiency of AIFM strategies in reasonable classes of initialized arenas to the sufficiency of memoryless strategies in a subclass. We end the section by showing the use of Lemma 21 in the proof of our first main result (Theorem 22), which shows that the sufficiency of pure AIFM strategies implies the stronger existence of pure AIFM SP strategies in well-behaved classes of initialized arenas.

First, we restate the classical result linking playing optimally with memory  $\mathcal{M}$  in an initialized arena and playing optimally with a memoryless strategy in its product with  $\mathcal{M}$ .

**Lemma 17.** *Let  $\sqsubseteq$  be a preference relation,  $X \in \{\text{PFM}, \text{P}, \text{GFM}, \text{G}\}$  be a type of strategies, and  $\mathcal{M} = (M, m_{\text{init}}, \alpha_{\text{upd}})$  be a memory skeleton. Let  $\mathcal{G} = (\mathcal{A}, S_{\text{init}}, \sqsubseteq)$  be an initialized game, and let  $\sigma_i \in \Sigma_i^X(\mathcal{A}, S_{\text{init}})$  be a finite-memory strategy encoded by a Mealy machine  $\Gamma = (\mathcal{M}, \alpha_{\text{next}})$ . Then,  $\sigma_i$  is  $X$ -optimal in  $\mathcal{G}$  if and only if  $\alpha_{\text{next}}$  corresponds to a memoryless  $X$ -optimal strategy in game  $\mathcal{G}' = ((\mathcal{A}, S_{\text{init}}) \ltimes \mathcal{M}, \sqsubseteq)$ .*

A proof to a very similar result can be found in [BLO<sup>+</sup>20, Lemma 1]. This lemma can be restated for SP strategies, for NE and for SPE with a very similar proof.

*Proof.* This proof goes through multiple steps, which all rely on establishing a correspondence between properties of  $(\mathcal{A}, S_{\text{init}})$  and  $(\mathcal{A}, S_{\text{init}}) \ltimes \mathcal{M}$ . We first establish a bijection  $\mathcal{H}$  between their finite histories and then a bijection  $f$  between their strategies. This is sufficient to show that the  $\text{UCol}_{\sqsubseteq}^X$  operators are preserved through  $f$ , which shows that  $X$ -optimality is preserved through  $f$ . It is then left to show that  $f(\sigma_1)$  corresponds to  $\alpha_{\text{next}}$ .

We first define a bijection  $\mathcal{H}: \text{Hists}(\mathcal{A}, S_{\text{init}}) \rightarrow \text{Hists}((\mathcal{A}, S_{\text{init}}) \ltimes \mathcal{M})$ . Let  $\rho = s_0 a_1 s_1 \dots a_n s_n \in \text{Hists}(\mathcal{A}, S_{\text{init}})$ . We set  $m_0 = m_{\text{init}}$ , and for  $1 \leq j \leq n$ ,  $m_j = \alpha_{\text{upd}}(m_{j-1}, \text{col}(s_{j-1}, a_j))$ . We define  $\mathcal{H}(\rho) = (s_0, m_0) a_1 (s_1, m_1) \dots a_n (s_n, m_n)$ . Notice that  $\widehat{\text{col}}(\mathcal{H}(\rho)) = \widehat{\text{col}}(\rho)$ . Furthermore,  $\mathcal{H}$  is bijective; as the initial state of the memory  $m_{\text{init}}$  is fixed and the memory skeleton is deterministic, the memory states added to  $\rho$  to obtain  $\mathcal{H}(\rho)$  are uniquely determined.

We now show that there is a correspondence between strategies of  $\Sigma_i^X(\mathcal{A}, S_{\text{init}})$  and strategies of  $\Sigma_i^X((\mathcal{A}, S_{\text{init}}) \ltimes \mathcal{M})$ : intuitively, augmenting the arena with the skeleton allows some strategies to be played using less memory, but does not fundamentally change each player's possibilities. We define a function  $f: \Sigma_i^X(\mathcal{A}, S_{\text{init}}) \rightarrow \Sigma_i^X((\mathcal{A}, S_{\text{init}}) \ltimes \mathcal{M})$ . For  $\tau_i \in \Sigma_i^G(\mathcal{A}, S_{\text{init}})$ ,  $\rho' \in \text{Hists}_i((\mathcal{A}, S_{\text{init}}) \ltimes \mathcal{M})$ , we define  $f(\tau_i)(\rho') = \tau_i(\mathcal{H}^{-1}(\rho'))$  (exploiting that actions are the same in  $(\mathcal{A}, S_{\text{init}})$  as in  $(\mathcal{A}, S_{\text{init}}) \ltimes \mathcal{M}$ ). Function  $f$  is bijective (for  $\tau'_i \in \Sigma_i^G((\mathcal{A}, S_{\text{init}}) \ltimes \mathcal{M})$ , its inverse  $f^{-1}$  can be specified as  $f^{-1}(\tau'_i) = \tau'_i \circ \mathcal{H}$ ). Moreover, it preserves the pure/randomized and the finite-memory/infinite-memory features of the strategies.

We observe the following fact<sup>5</sup> about  $f$ : for all  $s \in S_{\text{init}}$ , for all  $\tau_1 \in \Sigma_1^X(\mathcal{A}, S_{\text{init}})$ ,  $\tau_2 \in \Sigma_2^X(\mathcal{A}, S_{\text{init}})$ ,

$$\text{PC}_{\mathcal{A}, s}^{\tau_1, \tau_2} = \text{PC}_{(\mathcal{A}, s) \ltimes \mathcal{M}}^{f(\tau_1), f(\tau_2)}. \quad (4)$$

It can easily be proven by induction that these probability distributions match on all cylinders (as they always induce the same distributions on the actions and on the colors after corresponding histories  $\rho$  and  $\mathcal{H}(\rho)$ ), hence they are equal.

Now let  $\tau_1 \in \Sigma_1^X(\mathcal{A}, S_{\text{init}})$ . We notice that for all  $s \in S_{\text{init}}$ ,

$$\begin{aligned} \text{UCol}_{\sqsubseteq}^X(\mathcal{A}, s, \tau_1) &= \{\mu \in \text{Dist}(C^\omega, \mathcal{F}) \mid \exists \tau_2 \in \Sigma_2^X(\mathcal{A}, S_{\text{init}}), \text{PC}_{\mathcal{A}, s}^{\tau_1, \tau_2} \sqsubseteq \mu\} \\ &= \{\mu \in \text{Dist}(C^\omega, \mathcal{F}) \mid \exists \tau_2 \in \Sigma_2^X(\mathcal{A}, S_{\text{init}}), \text{PC}_{\mathcal{A} \ltimes \mathcal{M}, (s, m_{\text{init}})}^{f(\tau_1), f(\tau_2)} \sqsubseteq \mu\} && \text{by (4)} \\ &= \{\mu \in \text{Dist}(C^\omega, \mathcal{F}) \mid \exists \tau'_2 \in \Sigma_2^X((\mathcal{A}, S_{\text{init}}) \ltimes \mathcal{M}), \text{PC}_{\mathcal{A} \ltimes \mathcal{M}, (s, m_{\text{init}})}^{f(\tau_1), \tau'_2} \sqsubseteq \mu\} \\ &= \text{UCol}_{\sqsubseteq}^X((\mathcal{A}, S_{\text{init}}) \ltimes \mathcal{M}, (s, m_{\text{init}}), f(\tau_1)), \end{aligned}$$

where the penultimate line holds by bijectivity of  $f$ . The property holds symmetrically for a strategy  $\tau_2 \in \Sigma_2^X(\mathcal{A}, S_{\text{init}})$ . Thus  $X$ -optimality of strategies is preserved through  $f$ .

Now remember that  $\sigma_i \in \Sigma_i^X(\mathcal{A}, S_{\text{init}})$  is a strategy encoded by a Mealy machine  $(\mathcal{M}, \alpha_{\text{next}})$ . We notice that  $f(\sigma_i)$  corresponds to  $\alpha_{\text{next}}$  interpreted over the product initialized arena and is thus memoryless. By the previous property, we have that  $\sigma_i$  is  $X$ -optimal in  $\mathcal{G}$  if and only if  $\alpha_{\text{next}}$  is  $X$ -optimal in  $\mathcal{G}'$ .  $\square$

We now define a property of initialized arenas called *coverability by  $\mathcal{M}$*  (for a memory skeleton  $\mathcal{M}$ ), which happens to characterize initialized arenas that are a product with  $\mathcal{M}$  (Lemma 19). Albeit intuitive, this is a key technical step, as the class of arenas covered by a memory skeleton is sufficiently well-behaved to support edge-induction arguments, whereas it is more difficult to perform such techniques directly on the class of product arenas: removing a single edge from a product arena makes it hard to express as a product arena, whereas it is clear that coverability is preserved.

**Definition 18 (Coverability by  $\mathcal{M}$ ).** *An initialized arena  $((S_1, S_2, A, \delta, \text{col}), S_{\text{init}})$  is covered by memory skeleton  $\mathcal{M} = (M, m_{\text{init}}, \alpha_{\text{upd}})$  if there exists a function  $\phi: S \rightarrow M$  such that for all  $s \in S_{\text{init}}$ ,  $\phi(s) = m_{\text{init}}$ , and for all  $(s, a, s') \in \delta$ ,  $\alpha_{\text{upd}}(\phi(s), \text{col}(s, a)) = \phi(s')$ .*

This property means that it is possible to assign a unique memory state to each arena state such that transitions of the arena always update the memory state in a way that is consistent with the memory skeleton. Note that isomorphism of initialized arenas preserves coverability by any memory skeleton. Also, every initialized arena is covered by  $\mathcal{M}_{\text{triv}}$ , which is witnessed by the constant function  $\phi$  associating  $m_{\text{init}}$  to every state.

Definitions close to our notion of *coverability by  $\mathcal{M}$*  were introduced for deterministic arenas in [Kop08, BLO<sup>+</sup>20]. If we restrict our definition to deterministic arenas, the definition of *adherence with  $\mathcal{M}$*  in [Kop08, Definition 8.12] is very similar, but does not distinguish initial states from the rest (neither in the arena nor in the memory skeleton) — the reason is that [Kop08] only considers *prefix-independent* objectives, for which selecting the right initial memory state is not as important (see [Kop08, Proposition 8.2]). Our property of  $(\mathcal{A}, S_{\text{init}})$  being *covered by  $\mathcal{M}$*  is also equivalent to  $\mathcal{A}$

<sup>5</sup> Remember that our preference relation is defined over distributions over *sequences of colors*, and Equation (4) compares two such distributions.

being both *prefix-covered* and *cyclic-covered* by  $\mathcal{M}$  from  $S_{\text{init}}$ , as defined in [BLO<sup>+</sup>20]. Distinguishing both notions gives insight in [BLO<sup>+</sup>20] as they are used at different places in proofs (*prefix-covered* along with *monotony*, and *cyclic-covered* along with *selectivity*). Here, we opt for a single concise definition, as most of our proofs do not mention *monotony* and *selectivity*.

We link products and coverability: first, product initialized arenas are covered; second, covered initialized arenas are exactly the ones that are isomorphic to their own product.

**Lemma 19.** *Let  $\mathcal{M}$  be a memory skeleton and  $(\mathcal{A}, S_{\text{init}})$  be an initialized arena. The product initialized arena  $(\mathcal{A}, S_{\text{init}}) \ltimes \mathcal{M}$  is covered by  $\mathcal{M}$ . Moreover,  $(\mathcal{A}, S_{\text{init}})$  is covered by  $\mathcal{M}$  if and only if  $(\mathcal{A}, S_{\text{init}})$  is isomorphic to  $(\mathcal{A}, S_{\text{init}}) \ltimes \mathcal{M}$ .*

*Proof.* Let  $\mathcal{M} = (M, m_{\text{init}}, \alpha_{\text{upd}})$ ,  $\mathcal{A} = (S_1, S_2, A, \delta, \text{col})$ . We show that  $(\mathcal{A}, S_{\text{init}}) \ltimes \mathcal{M}$  is covered by  $\mathcal{M}$ . Let  $\phi: S \times M \rightarrow M$  be the projection on  $M$  (that is,  $\phi(s, m) = m$  for all  $(s, m) \in S \times M$ ). This function witnesses that  $(\mathcal{A}, S_{\text{init}}) \ltimes \mathcal{M}$  is covered, by definition of product initialized arena: only transitions that are consistent with the memory structure are allowed.

Assume now  $(\mathcal{A}, S_{\text{init}})$  is covered by  $\mathcal{M}$ , witnessed by function  $\phi$ . We show that  $(\mathcal{A}, S_{\text{init}})$  is isomorphic to  $(\mathcal{A}, S_{\text{init}}) \ltimes \mathcal{M}$ . In  $(\mathcal{A}, S_{\text{init}}) \ltimes \mathcal{M}$ , it is not possible to reach two states  $(s, m)$ ,  $(s, m')$  with  $m \neq m'$  from  $S_{\text{init}} \times \{m_{\text{init}}\}$ : otherwise, this would contradict that  $(\mathcal{A}, S_{\text{init}})$  is covered by  $\mathcal{M}$ . Hence the function  $\psi_S: s \mapsto (s, \phi(s))$  is a bijection between states of  $(\mathcal{A}, S_{\text{init}})$  and states of  $(\mathcal{A}, S_{\text{init}}) \ltimes \mathcal{M}$  (remember that we only keep the reachable states of the product initialized arena). Moreover, all the actions, transitions and colors are preserved, by definition of product initialized arena. Hence  $(\mathcal{A}, S_{\text{init}})$  is isomorphic to its own product with  $\mathcal{M}$ . Now for the other direction, assume  $(\mathcal{A}, S_{\text{init}})$  is isomorphic to  $(\mathcal{A}, S_{\text{init}}) \ltimes \mathcal{M}$ . Thus,  $(\mathcal{A}, S_{\text{init}})$  can be expressed as a product with  $\mathcal{M}$  and is thus covered by  $\mathcal{M}$  by the first claim.  $\square$

This last lemma shows in some sense an equivalence between a product and a covered initialized arena: product initialized arenas are covered (a similar result for non-initialized arenas is discussed in [BLO<sup>+</sup>20, Lemma 8]) and conversely, covered initialized arenas can be written as a product. The latter implication requires the use of *initialized* arenas to be expressed in a concise way: if a game could always start from any state of an arena, taking the product with a memory skeleton would virtually always make the arena grow, and it could therefore not be isomorphic to its own product. That is one of the main reasons we resort to *initialized* arenas: we are therefore able to talk interchangeably about *being a product*, which is a technical property at the core of the idea of *playing with memory*, and about *coverability*, which is a more intuitive, easy-to-check condition that trivially benefits from nice closure properties.

We establish two easy consequences of the previous lemmas to have a better understanding of the links between covered and product initialized arenas.

**Corollary 20.** *Let  $\mathcal{M}_1$  and  $\mathcal{M}_2$  be two memory skeletons. An initialized arena  $(\mathcal{A}, S_{\text{init}})$  is covered by  $\mathcal{M}_1$  and by  $\mathcal{M}_2$  if and only if it is covered by  $\mathcal{M}_1 \otimes \mathcal{M}_2$ .*

*Proof.* Let  $(\mathcal{A}, S_{\text{init}})$  be covered by  $\mathcal{M}_1$  and by  $\mathcal{M}_2$ . It is thus isomorphic to  $((\mathcal{A}, S_{\text{init}}) \ltimes \mathcal{M}_1) \ltimes \mathcal{M}_2$  by applying Lemma 19 twice. Notice that  $((\mathcal{A}, S_{\text{init}}) \ltimes \mathcal{M}_1) \ltimes \mathcal{M}_2$  is isomorphic to  $(\mathcal{A}, S_{\text{init}}) \ltimes (\mathcal{M}_1 \otimes \mathcal{M}_2)$  (simply consider the bijection  $\psi_S: ((s, m_1), m_2) \mapsto (s, (m_1, m_2))$ ). Hence,  $(\mathcal{A}, S_{\text{init}})$  is isomorphic to  $(\mathcal{A}, S_{\text{init}}) \ltimes (\mathcal{M}_1 \otimes \mathcal{M}_2)$ , and by using Lemma 19 in the other direction, is covered by  $\mathcal{M}_1 \otimes \mathcal{M}_2$ . Following the arguments backwards yields the other direction of the implication.  $\square$

The following lemma sums up our main practical use of the idea of *coverability*, by proving an equivalence between optimal strategies with memory in initialized arenas and memoryless optimal strategies in covered initialized arenas, in classes of arenas with mild hypotheses.

**Lemma 21.** *Let  $\mathcal{M}$  be a memory skeleton, and  $X \in \{\text{PFM}, \text{P}, \text{GFM}, \text{G}\}$  be a type of strategies. Let  $\mathfrak{A}$  be a class of initialized arenas closed by product with  $\mathcal{M}$ . Then,  $\mathcal{P}_1$  has an  $X$ -optimal (resp.  $X$ -SP) strategy based on  $\mathcal{M}$  in all initialized arenas in  $\mathfrak{A}$  if and only if  $\mathcal{P}_1$  has a memoryless  $X$ -optimal (resp.  $X$ -SP) strategy in all initialized arenas covered by  $\mathcal{M}$  in  $\mathfrak{A}$ . Also, there is an  $X$ -NE (resp.  $X$ -SPE) based on  $\mathcal{M}$  in all initialized arenas in  $\mathfrak{A}$  if and only if there is a memoryless  $X$ -NE (resp.  $X$ -SPE) in all initialized arenas covered by  $\mathcal{M}$  in  $\mathfrak{A}$ .*



*Proof.* We first prove that products of initialized arenas in  $\mathfrak{A}$  with  $\mathcal{M}$  correspond exactly to initialized arenas of  $\mathfrak{A}$  covered by  $\mathcal{M}$ , that is,

$$\{(\mathcal{A}, S_{\text{init}}) \times \mathcal{M} \mid (\mathcal{A}, S_{\text{init}}) \in \mathfrak{A}\} = \{(\mathcal{A}, S_{\text{init}}) \in \mathfrak{A} \mid (\mathcal{A}, S_{\text{init}}) \text{ is covered by } \mathcal{M}\}. \quad (5)$$

We start with the left-to-right inclusion. Let  $(\mathcal{A}, S_{\text{init}}) \times \mathcal{M}$  be an initialized product arena, with  $(\mathcal{A}, S_{\text{init}}) \in \mathfrak{A}$ . Then,  $(\mathcal{A}, S_{\text{init}}) \times \mathcal{M}$  belongs to  $\mathfrak{A}$ , as  $\mathfrak{A}$  is closed by product with  $\mathcal{M}$ . Moreover,  $(\mathcal{A}, S_{\text{init}}) \times \mathcal{M}$  is covered by  $\mathcal{M}$  by Lemma 19. For the right-to-left inclusion, let  $(\mathcal{A}, S_{\text{init}}) \in \mathfrak{A}$  be covered by  $\mathcal{M}$ . Then, it is isomorphic to  $(\mathcal{A}, S_{\text{init}}) \times \mathcal{M}$  by Lemma 19. Therefore, it can be expressed as the product of an initialized arena of  $\mathfrak{A}$  with  $\mathcal{M}$ .

We now prove the main statements of the lemma. Player  $\mathcal{P}_1$  has an X-optimal (resp. X-SP) strategy based on  $\mathcal{M}$  in all initialized arenas of  $\mathfrak{A}$  if and only if  $\mathcal{P}_1$  has a memoryless X-optimal (resp. X-SP) strategy in all products of initialized arenas in  $\mathfrak{A}$  with  $\mathcal{M}$  (by Lemma 17) if and only if  $\mathcal{P}_1$  has a memoryless X-optimal (resp. X-SP) strategy in all initialized arenas covered by  $\mathcal{M}$  in  $\mathfrak{A}$  (by (5)).

Similarly, there is an X-NE (resp. X-SPE) based on  $\mathcal{M}$  in all initialized arenas in  $\mathfrak{A}$  if and only if there is a memoryless X-NE (resp. X-SPE) in all products of initialized arenas in  $\mathfrak{A}$  with  $\mathcal{M}$  (by Lemma 17) if and only if there is a memoryless X-NE (resp. X-SPE) in all initialized arenas covered by  $\mathcal{M}$  in  $\mathfrak{A}$  (by (5)).  $\square$

We conclude this section with one of our main results, which shows that when pure strategies based on the same memory skeleton  $\mathcal{M}$  are sufficient to play *optimally*, then pure SP strategies based on  $\mathcal{M}$  exist.

**Theorem 22.** *Let  $\mathcal{M}$  be a memory skeleton, and  $X \in \{\text{PFM}, \text{P}, \text{GFM}, \text{G}\}$  be a type of strategies. Let  $\mathfrak{A}$  be a class of initialized arenas closed by product with  $\mathcal{M}$  and by prefix-extension. If  $\mathcal{P}_1$  has pure X-optimal strategies based on  $\mathcal{M}$  in all initialized arenas of  $\mathfrak{A}$ , then  $\mathcal{P}_1$  has pure X-SP strategies based on  $\mathcal{M}$  in all initialized arenas of  $\mathfrak{A}$ . If there exist pure X-NE based on  $\mathcal{M}$  in all initialized arenas of  $\mathfrak{A}$ , then there exist pure X-SPE based on  $\mathcal{M}$  in all initialized arenas of  $\mathfrak{A}$ .*

*Proof.* We start by proving the first claim (going from X-optimal to X-SP strategies). As  $\mathfrak{A}$  is closed by product with  $\mathcal{M}$ , by Lemma 21, both the hypothesis and the thesis of this claim can be reformulated for pure memoryless strategies in initialized arenas covered by  $\mathcal{M}$ . We thus prove equivalently that  $\mathcal{P}_1$  has pure memoryless X-SP strategies in all initialized arenas covered by  $\mathcal{M}$  in  $\mathfrak{A}$ , based on the hypothesis that  $\mathcal{P}_1$  has pure memoryless X-optimal strategies in all initialized arenas covered by  $\mathcal{M}$  in  $\mathfrak{A}$ .

Let  $(\mathcal{A}_0, S_{\text{init}}^0) \in \mathfrak{A}$  be covered by  $\mathcal{M}$ . By hypothesis,  $\mathcal{P}_1$  has a pure memoryless X-optimal strategy  $\sigma_1^0$  on  $(\mathcal{A}_0, S_{\text{init}}^0)$ . If this strategy is X-SP, then we are done. If not, then that means that there exists  $\rho_0 \in \text{Hists}(\mathcal{A}_0, S_{\text{init}}^0)$ , with  $s_0 = \text{out}(\rho_0)$  and  $w_0 = \widehat{\text{col}}(\rho_0)$ , such that  $\sigma_1^0[\rho_0]$  is not X-optimal in  $(\mathcal{A}_0, s_0, \sqsubseteq_{[w_0]})$ . We extend arena  $\mathcal{A}_0$  to a new prefix-extended arena  $\mathcal{A}_1 = (\mathcal{A}_0)_{w_0 \rightsquigarrow s_0}$  (this notation was introduced at (2)) by “plugging” a copy of history  $\rho_0$  before  $s_0$ . We also fix  $S_{\text{init}}^1 = S_{\text{init}}^0 \uplus \{s_0^{w_0}\}$ , where  $s_0^{w_0}$  is the first state of the newly added chain with colors similar to  $w_0$ . Initialized arena  $(\mathcal{A}_1, S_{\text{init}}^1)$  is in  $\mathfrak{A}$  since  $\mathfrak{A}$  is closed by prefix-extension. We show that  $(\mathcal{A}_1, S_{\text{init}}^1)$  is covered by  $\mathcal{M}$ : the covering property holds from  $S_{\text{init}}^0$  because  $(\mathcal{A}_0, S_{\text{init}}^0)$  was already covered by  $\mathcal{M}$  and the newly added states are not reachable from  $S_{\text{init}}^0$ , and it holds from  $s_0^{w_0}$  because the colors up to  $s_0$  are the same as history  $\rho_0$  from  $S_{\text{init}}^0$ .

By hypothesis, there exists a pure memoryless X-optimal strategy  $\sigma_1^1$  on  $(\mathcal{A}_1, S_{\text{init}}^1)$ . We argue that  $\sigma_1^1$  is X-optimal in  $(\mathcal{A}_0, s_0, \sqsubseteq_{[w_0]})$  (i.e., after seeing  $\rho_0$ ); if it were not, then it would not be X-optimal from  $(\mathcal{A}_1, s_0^{w_0}, \sqsubseteq)$ , as in both cases, the sequence of colors  $w_0$  is seen before  $s_0$  is reached, and the same (memoryless) strategy is played from  $s_0$ . The restriction of strategy  $\sigma_1^1$  to  $\text{Hists}(\mathcal{A}_0, S_{\text{init}}^0)$  is therefore also X-optimal in  $(\mathcal{A}_0, S_{\text{init}}^0)$ , but it is better than  $\sigma_1^0$  after seeing  $\rho_0$ .

If the restriction of  $\sigma_1^1$  to  $\text{Hists}(\mathcal{A}_0, S_{\text{init}}^0)$  is X-SP in  $(\mathcal{A}_0, S_{\text{init}}^0)$ , then we are done. If not, then it means that some history  $\rho_1 \in \text{Hists}(\mathcal{A}_0, S_{\text{init}}^0)$  witnesses that  $\sigma_1^1$  is not X-SP in  $(\mathcal{A}_0, S_{\text{init}}^0)$ . Let  $w_1 = \widehat{\text{col}}(\rho_1)$ ,  $s_1 = \text{out}(\rho_1)$ . We can keep going and build an arena  $\mathcal{A}_2 = (\mathcal{A}_1)_{w_1 \rightsquigarrow s_1}$ , with initial states  $S_{\text{init}}^2 = S_{\text{init}}^1 \uplus \{s_0^{w_1}\}$ , which gives us a pure memoryless X-optimal strategy  $\sigma_1^2$  on  $(\mathcal{A}_2, S_{\text{init}}^2)$ .

We keep building initialized arenas  $(\mathcal{A}_i, S_{\text{init}}^i)$  and pure memoryless X-optimal strategies  $\sigma_1^i$  as long as the restrictions of the strategies to  $\text{Hists}(\mathcal{A}_0, S_{\text{init}}^0)$  are not X-SP in  $(\mathcal{A}_0, S_{\text{init}}^0)$ . We argue that this iteration



ends after a finite number of steps. The restriction of every strategy  $\sigma_1^i$  to  $\text{Hists}(\mathcal{A}_0, S_{\text{init}}^0)$  is necessarily different from the same restriction for all the previous strategies: for all  $j$ ,  $0 \leq j < i$ ,  $\sigma_1^i$  is better than  $\sigma_1^j$  after seeing history  $\rho^j$ , and can therefore not be equal to  $\sigma_1^j$ . Moreover, there are only finitely many pure memoryless strategies on  $(\mathcal{A}_0, S_{\text{init}}^0)$  (as this arena is finite), and there is a bijection between *pure memoryless* strategies of arenas  $(\mathcal{A}_i, S_{\text{init}}^i)$  and of arena  $(\mathcal{A}_0, S_{\text{init}}^0)$  (as building prefix-extensions does not provide more choices for memoryless strategies).

Combining that all strategies  $\sigma_1^i$  are different and the finiteness of the number of strategies shows that the iteration ends, and therefore, that, for some  $i \geq 0$ , the restriction of the pure memoryless strategy  $\sigma_1^i$  to  $\text{Hists}(\mathcal{A}_0, S_{\text{init}}^0)$  is X-SP in  $(\mathcal{A}_0, S_{\text{init}}^0)$ .

The proof to go from pure X-NE based on  $\mathcal{M}$  to pure X-SPE based on  $\mathcal{M}$  works in the same way, as there are also finitely many *pairs* of pure memoryless strategies.  $\square$

The facts that we consider *finite* arenas and that the hypothesis is about *pure AIFM strategies* are both crucial in the previous proof, as we need the finiteness of the type of strategies considered.

This result shows a major distinction between the sufficiency of AIFM strategies and the more general sufficiency of FM strategies: if a player can always play optimally with the same memory, then SP strategies may be played with the same memory as optimal strategies — if a player can play optimally but needs *arena-dependent* finite memory, then infinite memory may still be required to obtain SP strategies. One such example is provided in [LPR18, Example 16] for the *average-energy games with lower-bounded energy* objective in deterministic games:  $\mathcal{P}_1$  can always play optimally with pure finite-memory strategies [BHM<sup>+</sup>17, Theorem 13], but infinite memory is needed for SP strategies.

## 4 One-to-two-player lift

Our goal in this section is to obtain a practical tool to help study the memory requirements of two-player stochastic (or deterministic) games. This tool consists in reducing the study of the sufficiency of pure AIFM strategies for both players in *two-player* games to *one-player* games. We will first state our result, and the rest of the section is devoted to its proof. This result mentions two properties of classes of arenas called being *closed by subarena* and *closed by split*, which we will introduce later. In particular, it can be instantiated with  $\mathfrak{A}$  being the class of all initialized deterministic arenas or the class of all initialized stochastic arenas.

**Theorem 23 (Pure AIFM one-to-two-player lift).** *Let  $\sqsubseteq$  be a preference relation,  $\mathcal{M}_1$  and  $\mathcal{M}_2$  be two memory skeletons, and  $X \in \{\text{PFM}, \text{P}, \text{GFM}, \text{G}\}$  be a type of strategies. Let  $\mathfrak{A}$  be a class of initialized arenas that is closed by subarena, by split, and by product with  $\mathcal{M}_1$  and  $\mathcal{M}_2$ . Assume that*

- *in all initialized one-player arenas of  $\mathcal{P}_1$  in  $\mathfrak{A}$ ,  $\mathcal{P}_1$  can play X-optimally with a pure strategy based on memory  $\mathcal{M}_1$ ;*
- *in all initialized one-player arenas of  $\mathcal{P}_2$  in  $\mathfrak{A}$ ,  $\mathcal{P}_2$  can play X-optimally with a pure strategy based on memory  $\mathcal{M}_2$ .*

*Then all initialized two-player arenas in  $\mathfrak{A}$  admit a pure X-NE based on memory  $\mathcal{M}_1 \otimes \mathcal{M}_2$ . If  $\mathfrak{A}$  is moreover closed by prefix-extension, then all initialized two-player arenas in  $\mathfrak{A}$  admit a pure X-SPE based on memory  $\mathcal{M}_1 \otimes \mathcal{M}_2$ .*

The practical usage of this result can be summed up as follows: to determine whether pure AIFM strategies are sufficient for both players in stochastic (resp. deterministic) arenas to play X-optimally, it is sufficient to prove it for stochastic (resp. deterministic) one-player arenas. Studying memory requirements of one-player arenas is significantly easier than studying memory requirements of two-player arenas, as a one-player arena can be seen as a graph (in the deterministic case) or an MDP (in the stochastic case). Still, we will bring more tools to study memory requirements of one-player arenas in Section 5.

Our proof technique for Theorem 23 is able to deal in a uniform manner with stochastic arenas and with deterministic arenas, under different types of strategies. It borrows ideas from [GZ09] and

from [BLO<sup>+</sup>20] and extends them both: it extends [GZ09] by generalizing to a wider type of strategies (AIFM instead of memoryless) and it extends [BLO<sup>+</sup>20] by extending the class of arenas and preference relations considered (stochastic instead of deterministic). Thanks to Theorem 22, we also go further in our understanding of the optimal strategies: we are able to obtain the existence of X-SPE with almost the same constraints, instead of the weaker existence of X-NE.

#### 4.1 Proof scheme

In order to prove Theorem 23, we first establish a similar result about memoryless strategies. We carry out all our intermediate proofs with the concept of Nash equilibrium, and we will strengthen it to subgame perfect equilibria at the end, thanks to Theorem 22.

**Lemma 24 (Memoryless one-to-two-player lift).** *Let  $\sqsubseteq$  be a preference relation and  $X \in \{\text{PFM}, \text{P}, \text{GFM}, \text{G}\}$  be a type of strategies. Let  $\mathfrak{A}$  be a class of initialized arenas that is closed by subarena and by split. If both players have pure memoryless X-optimal strategies in the initialized one-player arenas in  $\mathfrak{A}$ , then all initialized arenas in  $\mathfrak{A}$  admit a pure memoryless X-NE.*

This result and its proof originate from [GZ09, Theorem 9], with the addition that we consider here *initialized* arenas, and that the strategies do not have to be optimal from all states, but only from the specified initial states. We explain why, albeit small, this is an important addition to obtain our result.

Lemma 24 can immediately be instantiated with  $\mathfrak{A}$  being the class of all deterministic or stochastic arenas to obtain an interesting result about pure memoryless strategies. Our goal will be to instantiate it, for some fixed memory skeleton  $\mathcal{M}$ , with the class of *initialized arenas covered by  $\mathcal{M}$* , so that we can later obtain results about pure AIFM strategies (through Lemma 21) instead of only using it for pure memoryless strategies, in a similar spirit to [BLO<sup>+</sup>20]. As we will see, the class of initialized arenas covered by  $\mathcal{M}$  happens to be closed by subarena and by split. This extension to pure AIFM strategies is one precise step where our notion of *initialized arenas* finds its use: in covered (or product) arenas, we are only interested in optimality from arena states associated to memory state  $m_{\text{init}}$ , and not from all states. Hence without restating and reproving [GZ09, Theorem 9] with an extra quantification on the initial states, it seems difficult to extend it in a straightforward way to a result about pure AIFM strategies.

#### 4.2 Proving Lemma 24

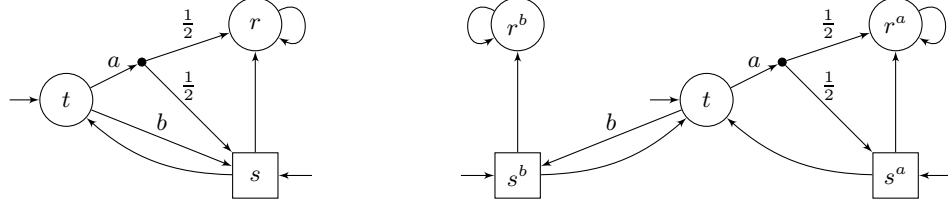
We recall the definitions of subarena and split from [GZ09], extending them in a natural way to initialized arenas.

**Definition 25 (Initialized subarena).** *Let  $(\mathcal{A} = (S_1, S_2, A, \delta, \text{col}), S_{\text{init}})$  be an initialized arena. An initialized subarena of  $(\mathcal{A}, S_{\text{init}})$  is an initialized arena  $((S_1, S_2, A', \delta', \text{col}'), S_{\text{init}})$  such that  $A' \subseteq A$ ,  $\delta' \subseteq \delta$  (that is, some states might lose a few available actions), and  $\text{col}'$  is the restriction of  $\text{col}$  to the pairs  $(s, a)$  such that  $s \in S$  and  $a \in A'(s)$ .*

An initialized subarena keeps the same state space and initial states as the original arena, but with fewer available actions. Remember that we assume that arenas are non-blocking, hence at least one available action should be kept in each state of the initialized subarena. We say that a class  $\mathfrak{A}$  of initialized arenas is *closed by subarena* if for all  $(\mathcal{A}, S_{\text{init}}) \in \mathfrak{A}$ , if  $(\mathcal{A}', S_{\text{init}})$  is a subarena of  $(\mathcal{A}, S_{\text{init}})$ , then  $(\mathcal{A}', S_{\text{init}}) \in \mathfrak{A}$ .

We now define the notion of *split on  $t \in S$*  of an arena: the main idea is that for some state  $t$  of the arena, the state space of the arena is augmented in such a way that players remember what was the last action played when leaving state  $t$ . In practice, for each action available in  $t$ , we make a copy of the arena where only this action is available in  $t$ , and we then merge all these arenas on state  $t$ . Every action  $a$  available in  $t$  leads to a copy of the arena in which states are labeled by  $a$ . Before introducing the formal definition, we provide an example of a split in Figure 1. There are two actions  $a$  and  $b$  available

in  $t$ , and we make a copy of the other states for each action. This way, when the game is for instance in  $s^a$ , we know that the last action that was chosen in  $t$  was  $a$  (which is not necessarily the case when in  $s$  in the original arena). The probabilities, colors and initial states are preserved in each copy of the initialized arena.



**Fig. 1.** Initialized arena with  $A(t) = \{a, b\}$  (omitting colors) (left) and its split on  $t$  (right). States controlled by  $\mathcal{P}_1$  (resp.  $\mathcal{P}_2$ ) are depicted by circles (resp. squares). The dot after playing action  $a$  represents a stochastic transition, with probability  $\frac{1}{2}$  to go to  $r$  and  $\frac{1}{2}$  to go to  $s$ .

**Definition 26 (Split).** Let  $(\mathcal{A} = (S_1, S_2, A, \delta, \text{col}), S_{\text{init}})$  be an initialized arena, and  $t \in S_1$ . For  $a \in A(t)$ , we denote  $(\mathcal{A}_a, S_{\text{init}}^a)$  as the initialized subarena of  $(\mathcal{A}, S_{\text{init}})$  in which only action  $a$  is available in  $t$  and in which all states are renamed  $s \mapsto s^a$ . The split on  $t$  of  $(\mathcal{A}, S_{\text{init}})$  is the initialized arena  $(\text{Split}(\mathcal{A}, t), S_{\text{init}}^t)$  where

$$\text{Split}(\mathcal{A}, t) = \bigsqcup_{a \in A(t)} (\mathcal{A}_a, t^a),$$

with merged state called  $t$ , and  $S_{\text{init}}^t = \bigcup_{a \in A(t)} \{s^a \mid s \in S_{\text{init}}\}$  (with  $t^a = t$  for all  $a \in A(t)$ ).

This could be defined symmetrically for a state  $t \in S_2$ . The merge operator  $\sqcup$  was introduced at (3). When considering a split arena on a state  $t$ , we use the convention that  $t^a = t$  for any action  $a$  available in  $t$ . Moreover, for  $a$  an action available in  $t$ , and  $S' \subseteq S$ , we write  $(S')^a$  for  $\{s^a \mid s \in S'\}$ . A class  $\mathfrak{A}$  of initialized arenas is *closed by split* if for all  $(\mathcal{A}, S_{\text{init}}) \in \mathfrak{A}$ , for all states  $t$  of  $\mathcal{A}$ ,  $(\text{Split}(\mathcal{A}, t), S_{\text{init}}^t) \in \mathfrak{A}$ .

It is possible to formulate a few intuitive results linking plays and strategies of an initialized arena and its split. These results are provided in a very similar context in [GZ09]; we recall them precisely in Appendix A and sketch their statements here.

Let  $(\mathcal{A} = (S_1, S_2, A, \delta, \text{col}), S_{\text{init}})$  be an initialized arena with a state  $t \in S_1$ . For all  $a \in A(t)$ , it is possible to build a natural bijection between strategies in  $\Sigma_i^G(\mathcal{A}, S_{\text{init}})$  and strategies on the split with restricted initial states  $\Sigma_i^G(\text{Split}(\mathcal{A}, t), S_{\text{init}}^a)$ . Intuitively, the available actions are the same at every step and the split does not offer any more possibilities (besides having more initial states — that is why we must restrict the initial states to have a bijection). More memory might be needed to play the corresponding strategy in  $(\mathcal{A}, S_{\text{init}})$  than in its split (since the information of the last action played in  $t$  is not explicitly given in  $\mathcal{A}$ ), but finite-memory strategies stay finite-memory in both directions. This bijection also preserves the “pure” feature of the strategy and preserves optimality and NE (Lemma 35).

Now if a pair of strategies  $(\sigma_1, \sigma_2)$  is an X-NE in  $(\text{Split}(\mathcal{A}, t), S_{\text{init}}^a)$  and  $\sigma_1$  is pure memoryless with  $\sigma_1(t) = a$ , only the part  $S^a$  of the split is ever reached during the play, which corresponds to the state space of  $\mathcal{A}$ . It is therefore possible to transform this X-NE into an X-NE  $(\sigma'_1, \sigma'_2)$  in  $(\mathcal{A}, S_{\text{init}})$  such that  $\sigma'_1$  is still pure and memoryless (Lemma 36).

If a pair of strategies  $(\sigma_1, \sigma_2)$  on  $(\text{Split}(\mathcal{A}, t), S_{\text{init}}^a)$  is such that  $\sigma_1$  is pure memoryless with  $\sigma_1(t) = a$ , then for  $s^a \in S_{\text{init}}^a$ , we can show that the distribution  $\text{Pc}_{\text{Split}(\mathcal{A}, t), s^a}^{\sigma_1, \sigma_2}$  is equal to  $\text{Pc}_{\mathcal{A}, s^a}^{\sigma_1^a, \sigma_2^a}$ , where  $\sigma_1^a$  and  $\sigma_2^a$  are simply restrictions of  $\sigma_1$  and  $\sigma_2$  to histories of the subarena  $(\mathcal{A}_a, S_{\text{init}}^a)$  (Lemma 37).

We are now ready to prove Lemma 24. The proof is by induction on the number of *choices* in arenas: for  $\mathcal{A} = (S_1, S_2, A, \delta, \text{col})$  an arena, the *number of choices in  $\mathcal{A}$*  is defined as

$$n_{\mathcal{A}} = \left( \sum_{s \in S} |A(s)| \right) - |S|. \quad (6)$$

When the number of choices in  $\mathcal{A}$  is 0, it means that there is exactly one available action in each state.

*Proof (of Lemma 24).* We proceed by induction on the number of choices in arenas. If an initialized arena  $(\mathcal{A}, S_{\text{init}}) \in \mathfrak{A}$  is such that  $n_{\mathcal{A}} = 0$ , then both players only have a single available strategy which is both pure and memoryless. Hence this pair of strategies correspond to a pure memoryless X-NE, which proves the base case. Now let  $n > 0$ : we assume that the result holds for every initialized arena  $(\mathcal{A}, S_{\text{init}}) \in \mathfrak{A}$  with  $n_{\mathcal{A}} < n$ , and let  $(\mathcal{A}, S_{\text{init}}) \in \mathfrak{A}$  be an initialized arena such that  $n_{\mathcal{A}} = n$ .

If  $\mathcal{P}_1$  has no choice (that is,  $|A(s)| = 1$  for all  $s \in S_1$ ), then  $(\mathcal{A}, S_{\text{init}})$  is an initialized one-player arena of  $\mathcal{P}_2$ , and  $\mathcal{P}_2$  has a pure memoryless X-optimal strategy by hypothesis from the statement of the theorem (and as  $\mathcal{P}_1$  has only one possible strategy which happens to be pure and memoryless, we have a pure memoryless X-NE). We now focus on the case where  $\mathcal{P}_1$  has at least one choice: let  $t \in S$  be such that  $|A(t)| \geq 2$ . For  $a \in A(t)$ , let  $(\mathcal{A}_a, S_{\text{init}}^a)$  be the initialized subarena of  $(\mathcal{A}, S_{\text{init}})$  with only action  $a$  available in  $t$ , and all states renamed  $s \mapsto s^a$ . This implies that  $(\mathcal{A}_a, S_{\text{init}}^a)$  is in  $\mathfrak{A}$ , as  $\mathfrak{A}$  is closed by subarena. Notice that  $n_{\mathcal{A}_a} < n_{\mathcal{A}}$ , as this is the same arena except that some available actions are removed in  $t$ . By induction hypothesis, there is thus a pure memoryless X-NE  $(\sigma_1^a, \sigma_2^a)$  in arena  $(\mathcal{A}_a, S_{\text{init}}^a)$ , for each  $a \in A(t)$ .

We now consider the split on  $t$  of  $(\mathcal{A}, S_{\text{init}})$ , which we denote  $(\text{Split}(\mathcal{A}, t), S_{\text{init}}^t)$ , and which belongs to  $\mathfrak{A}$  as  $\mathfrak{A}$  is closed by split. Consider the pure memoryless strategy  $\sigma_2 = \bigcup_{a \in A(t)} \sigma_2^a$  of  $\mathcal{P}_2$  defined on  $(\text{Split}(\mathcal{A}, t), S_{\text{init}}^t)$  which, when in  $S^a$  for some  $a \in A(t)$ , plays the same actions as the pure memoryless strategy  $\sigma_2^a$  (this prescribes a unique action to every state of  $\text{Split}(\mathcal{A}, t)$  controlled by  $\mathcal{P}_2$ ; the only overlap between the subarenas is  $t$ , but  $t$  belongs to  $\mathcal{P}_1$ ). We cannot straightaway define a similar strategy  $\sigma_1$  for  $\mathcal{P}_1$ , as it would not be well-defined in  $t$  — we first have to carefully choose the action played in  $t$ .

To do so, we consider the initialized one-player arena  $(\text{Split}(\mathcal{A}, t)_{\sigma_2}, S_{\text{init}}^t)$  of  $\mathcal{P}_1$  resulting from fixing the (pure memoryless) strategy  $\sigma_2$  of  $\mathcal{P}_2$  in  $\text{Split}(\mathcal{A}, t)$ . This initialized arena is an initialized subarena of  $(\text{Split}(\mathcal{A}, t), S_{\text{init}}^t)$  (actions of  $\mathcal{P}_2$  have been removed), and hence belongs to  $\mathfrak{A}$ . By hypothesis from the statement of the theorem,  $\mathcal{P}_1$  has a pure memoryless X-optimal strategy  $\tau_1$  in  $(\text{Split}(\mathcal{A}, t)_{\sigma_2}, S_{\text{init}}^t)$  since it is an initialized one-player arena. Let  $a^* \in A(t)$  be the action  $\tau_1(t)$  played by  $\mathcal{P}_1$  in  $t$ . We use this strategy  $\tau_1$  to define a pure memoryless strategy  $\sigma_1$  of  $\mathcal{P}_1$  on  $(\text{Split}(\mathcal{A}, t), S_{\text{init}}^t)$  that plays  $a^*$  in  $t$ , and for all  $a \in A(t)$ , that behaves like  $\sigma_1^a$  in  $S^a$ ; formally,

$$\sigma_1(t) = \tau_1(t) = a^*, \text{ and for } a \in A(t), s \in S_1 \setminus \{t\}, \sigma_1(s^a) = \sigma_1^a(s^a).$$

We now show that  $(\sigma_1, \sigma_2)$  is a pure memoryless X-NE in  $(\text{Split}(\mathcal{A}, t), S_{\text{init}}^t)$ . We restrict our attention to initial states  $S_{\text{init}}^{a^*}$  as these states are in the part of the arena that  $\sigma_1$  always goes back to, and that will help us convert  $\sigma_1$  into a corresponding pure memoryless strategy on  $(\mathcal{A}, S_{\text{init}})$  using Lemma 36. We show that for all  $s^{a^*} \in S_{\text{init}}^{a^*}$ , for all  $\sigma'_1 \in \Sigma_1^X(\text{Split}(\mathcal{A}, t), S_{\text{init}}^{a^*})$ ,  $\sigma'_2 \in \Sigma_2^X(\text{Split}(\mathcal{A}, t), S_{\text{init}}^{a^*})$ ,

$$\text{Pc}_{\text{Split}(\mathcal{A}, t), s^{a^*}}^{\sigma'_1, \sigma_2} \subseteq \text{Pc}_{\text{Split}(\mathcal{A}, t), s^{a^*}}^{\sigma_1, \sigma_2} \subseteq \text{Pc}_{\text{Split}(\mathcal{A}, t), s^{a^*}}^{\sigma_1, \sigma'_2}.$$

Let  $s^{a^*} \in S_{\text{init}}^{a^*}$ . The right-hand side inequality is clear, as the play always stays in  $\mathcal{A}_{a^*}$  by definition of  $\sigma_1$ , and the restriction of  $\sigma_1$  and  $\sigma_2$  to  $\text{Hists}(\mathcal{A}_{a^*}, S_{\text{init}}^{a^*})$  is  $(\sigma_1^{a^*}, \sigma_2^{a^*})$ , which is an X-NE in  $(\mathcal{A}_{a^*}, S_{\text{init}}^{a^*})$ .

For the left-hand side inequality, let  $\sigma'_1 \in \Sigma_1^X(\text{Split}(\mathcal{A}, t), S_{\text{init}}^{a^*})$  be a strategy of  $\mathcal{P}_1$ . We denote by  $\tau_1^{a^*}$  the restriction of  $\tau_1$  to  $\text{Hists}(\mathcal{A}_{a^*}, S_{\text{init}}^{a^*})$ .

We have

$$\begin{aligned} \text{Pc}_{\text{Split}(\mathcal{A}, t), s^{a^*}}^{\sigma'_1, \sigma_2} &\subseteq \text{Pc}_{\text{Split}(\mathcal{A}, t), s^{a^*}}^{\tau_1, \sigma_2} && \text{as } \tau_1 \text{ is X-optimal against } \sigma_2 \text{ in } (\text{Split}(\mathcal{A}, t), S_{\text{init}}^t), \text{ and } s^{a^*} \in S_{\text{init}}^t \\ &= \text{Pc}_{\mathcal{A}_{a^*}, s^{a^*}}^{\tau_1^{a^*}, \sigma_2^{a^*}} && \text{by Lemma 37} \\ &\subseteq \text{Pc}_{\mathcal{A}_{a^*}, s^{a^*}}^{\sigma_1^{a^*}, \sigma_2^{a^*}} && \text{as } (\sigma_1^{a^*}, \sigma_2^{a^*}) \text{ is an X-NE in } \mathcal{A}_{a^*} \\ &= \text{Pc}_{\text{Split}(\mathcal{A}, t), s^{a^*}}^{\sigma_1, \sigma_2} && \text{by Lemma 37.} \end{aligned}$$

Hence  $(\sigma_1, \sigma_2)$  is a pure memoryless X-NE in  $(\text{Split}(\mathcal{A}, t), S_{\text{init}}^t)$ . We can transform  $(\sigma_1, \sigma_2)$  into an X-NE  $(\sigma'_1, \sigma'_2)$  of  $(\mathcal{A}, S_{\text{init}})$  by Lemma 36, with  $\sigma'_1$  pure memoryless, but not necessarily  $\sigma'_2$ . We can however

perform the same proof for  $\mathcal{P}_2$ , and obtain a second X-NE  $(\sigma'_1, \sigma'_2)$  such that  $\sigma'_2$  is a pure memoryless strategy on  $(\mathcal{A}, S_{\text{init}})$ . Then we simply mix both X-NE (by Lemma 16), and obtain that  $(\sigma'_1, \sigma'_2)$  is a pure memoryless X-NE in  $(\mathcal{A}, S_{\text{init}})$ .  $\square$

### 4.3 From memoryless to AIFM

In this section, we show how to apply Lemma 21 to Lemma 24 to lift its results from the sufficiency of pure memoryless strategies to the sufficiency of pure AIFM strategies, which will imply Theorem 23.

**Lemma 27.** *Let  $\mathcal{M}$  be a memory skeleton, and  $\mathfrak{A}$  be a class of initialized arenas closed by subarena and by split. The class of all initialized arenas covered by  $\mathcal{M}$  in  $\mathfrak{A}$  is closed by subarena and by split.*

*Proof.* Let  $(\mathcal{A} = (S_1, S_2, A, \delta, \text{col}), S_{\text{init}}) \in \mathfrak{A}$  be an initialized arena covered by  $\mathcal{M} = (M, m_{\text{init}}, \alpha_{\text{upd}})$ . We show that its subarenas and its splits are still covered by  $\mathcal{M}$ . Let  $\phi: S \rightarrow M$  be the witness that  $(\mathcal{A}, S_{\text{init}})$  is covered by  $\mathcal{M}$ .

If we consider an initialized subarena  $(\mathcal{A}', S_{\text{init}})$  of  $(\mathcal{A}, S_{\text{init}})$ , the same function  $\phi$  will still be a witness that  $(\mathcal{A}', S_{\text{init}})$  is covered by  $\mathcal{M}$ , as the state space is the same, the condition to check is a universally quantified property over the transitions, and there are fewer transitions in  $\mathcal{A}'$  than in  $\mathcal{A}$ .

Let  $t \in S$ , and let  $(\text{Split}(\mathcal{A}, t), S_{\text{init}}^t)$  be the split on  $t$  of  $(\mathcal{A}, S_{\text{init}})$ . We define a function  $\hat{\phi}$  such that

$$\hat{\phi}(t) = \phi(t), \text{ and for } a \in A(t), s \in S \setminus \{t\}, \hat{\phi}(s^a) = \phi(s).$$

Function  $\hat{\phi}$  witnesses that  $(\text{Split}(\mathcal{A}, t), S_{\text{init}}^t)$  is covered by  $\mathcal{M}$  as every transition in  $\text{Split}(\mathcal{A}, t)$  corresponds to a transition in  $\mathcal{A}$  with the same color, and linking two states assigned to the same memory state in  $\mathcal{A}$  as in  $\text{Split}(\mathcal{A}, t)$ .  $\square$

We are now ready to prove Theorem 23, the main result of this section.

*Proof (of Theorem 23).* Note first that as  $((\mathcal{A}, S_{\text{init}}) \times \mathcal{M}_1) \times \mathcal{M}_2$  is isomorphic to  $(\mathcal{A}, S_{\text{init}}) \times (\mathcal{M}_1 \otimes \mathcal{M}_2)$ ,  $\mathfrak{A}$  is in particular closed by product with  $\mathcal{M}_1 \otimes \mathcal{M}_2$ .

Using Lemma 21, the hypotheses can be reformulated as follows: for  $i \in \{1, 2\}$ ,  $\mathcal{P}_i$  has a pure memoryless X-optimal strategy in all initialized one-player arenas in  $\mathfrak{A}$  that are covered by  $\mathcal{M}_i$ .

Now consider the subclass  $\mathfrak{A}' = \{(\mathcal{A}, S_{\text{init}}) \in \mathfrak{A} \mid (\mathcal{A}, S_{\text{init}}) \text{ is covered by } \mathcal{M}_1 \otimes \mathcal{M}_2\}$ . For  $i \in \{1, 2\}$ ,  $\mathcal{P}_i$  has a pure memoryless X-optimal strategy in all its initialized one-player arenas in  $\mathfrak{A}'$  (using that if an arena is covered by  $\mathcal{M}_1 \otimes \mathcal{M}_2$ , it is in particular covered by  $\mathcal{M}_i$  by Lemma 20). Moreover,  $\mathfrak{A}'$  is closed by subarena and by split by Lemma 27. Hence by Lemma 24, for all initialized arenas in  $\mathfrak{A}'$ , there exists a pure memoryless X-NE. Using Lemma 21 again in the other direction allows us to conclude that all initialized arenas in  $\mathfrak{A}$  admit a pure X-NE based on  $\mathcal{M}_1 \otimes \mathcal{M}_2$ .

By Theorem 22, using that  $\mathfrak{A}$  is closed by prefix-extension, the existence of pure X-NE based on  $\mathcal{M}_1 \otimes \mathcal{M}_2$  in all initialized arenas in  $\mathfrak{A}$  implies the existence of pure X-SPE based on  $\mathcal{M}_1 \otimes \mathcal{M}_2$  in all initialized arenas in  $\mathfrak{A}$ .  $\square$

Theorem 23 along with Lemmas 17 and 19 actually gives a bit more information about memory requirements in individual arenas than is strictly written. The way it is phrased shows that for an initialized arena  $(\mathcal{A}, S_{\text{init}})$ , memoryless strategies always suffice to play X-optimally in  $(\mathcal{A}, S_{\text{init}}) \times (\mathcal{M}_1 \otimes \mathcal{M}_2)$  (through Lemma 17). But if  $(\mathcal{A}, S_{\text{init}})$  is already covered by  $\mathcal{M}_1 \otimes \mathcal{M}_2$ , then as  $(\mathcal{A}, S_{\text{init}})$  is isomorphic to  $(\mathcal{A}, S_{\text{init}}) \times (\mathcal{M}_1 \otimes \mathcal{M}_2)$  by Lemma 19, memoryless strategies are actually sufficient directly in  $(\mathcal{A}, S_{\text{init}})$ . Memory  $\mathcal{M}_1 \otimes \mathcal{M}_2$  is thus an upper bound on the required memory, and studying coverability may show for some initialized arenas that less memory is sufficient: for example, if it is already covered by  $\mathcal{M}_1 \otimes \mathcal{M}_2$  (resp. by  $\mathcal{M}_1$ , by  $\mathcal{M}_2$ ), memoryless strategies (resp. strategies based on  $\mathcal{M}_2$ , on  $\mathcal{M}_1$ ) are sufficient. An application of Theorem 23 is provided in Section 6.

## 5 AIFM characterization

For this section, we fix  $\sqsubseteq$  a preference relation,  $X \in \{\text{PFM}, \text{P}, \text{GFM}, \text{G}\}$  a type of strategies, and  $\mathcal{M} = (M, m_{\text{init}}, \alpha_{\text{upd}})$  a memory skeleton. We distinguish only two classes of initialized arenas: the class  $\mathfrak{A}_{\mathcal{P}_1}^{\text{D}}$  of all initialized one-player deterministic arenas of  $\mathcal{P}_1$ , and the class  $\mathfrak{A}_{\mathcal{P}_1}^{\text{S}}$  of all initialized one-player stochastic arenas of  $\mathcal{P}_1$ . A class of arenas will therefore be specified by a letter  $Y \in \{\text{D}, \text{S}\}$ , which we fix for the whole section. Our aim is to give a better understanding of the preference relations for which pure strategies based on  $\mathcal{M}$  suffice to play  $X$ -optimally in  $\mathfrak{A}_{\mathcal{P}_1}^Y$ , by *characterizing* it through two intuitive conditions. All definitions and proofs are stated from the point of view of  $\mathcal{P}_1$ . We first introduce some more notations.

As we only work with one-player arenas in this section, we abusively write  $\text{Pc}_{\mathcal{A},s}^{\sigma_1}$  and  $\text{Pc}_{\mathcal{A},s}^{\sigma_1}$  for the distributions on plays and colors induced by a strategy  $\sigma_1$  of  $\mathcal{P}_1$  on  $(\mathcal{A}, s)$ , with no strategy for  $\mathcal{P}_2$ .

For  $\mathcal{A}$  a one-player arena of  $\mathcal{P}_1$  and  $s$  a state of  $\mathcal{A}$ , we write

$$[\mathcal{A}]_s^X = \{\text{Pc}_{\mathcal{A},s}^{\sigma_1} \mid \sigma_1 \in \Sigma_1^X(\mathcal{A}, s)\}$$

for the set of distributions over  $(C^\omega, \mathcal{F})$  induced by strategies of type  $X$  in  $\mathcal{A}$  from  $s$ .

For  $m_1, m_2 \in M$ , we write  $L_{m_1, m_2} = \{w \in C^* \mid \widehat{\alpha_{\text{upd}}}(m_1, w) = m_2\}$  for the language of words that are read from  $m_1$  up to  $m_2$  in  $\mathcal{M}$ . Such a language can be specified by the deterministic automaton that is simply the memory skeleton  $\mathcal{M}$  with  $m_1$  as the initial state and  $m_2$  as the unique final state.

We extend the *shifted distribution* notation introduced in Definition 13 to sets of distributions: for  $w \in C^*$ , for  $\Lambda \subseteq \text{Dist}(C^\omega, \mathcal{F})$ , we write  $w\Lambda$  for the set  $\{w\mu \mid \mu \in \Lambda\}$ .

Given  $\sqsubseteq$  a preference relation, we extend it to sets of distributions: for  $\Lambda_1, \Lambda_2 \subseteq \text{Dist}(C^\omega, \mathcal{F})$ , we write  $\Lambda_1 \sqsubseteq \Lambda_2$  if for all  $\mu_1 \in \Lambda_1$ , there exists  $\mu_2 \in \Lambda_2$  such that  $\mu_1 \sqsubseteq \mu_2$ ; we write  $\Lambda_1 \sqsubset \Lambda_2$  if there exists  $\mu_2 \in \Lambda_2$  such that for all  $\mu_1 \in \Lambda_1$ ,  $\mu_1 \sqsubset \mu_2$ . Notice that  $\neg(\Lambda_1 \sqsubseteq \Lambda_2)$  is equivalent to  $\Lambda_2 \sqsubset \Lambda_1$ . If  $\Lambda_1$  is a singleton  $\{\mu_1\}$ , we write  $\mu_1 \sqsubseteq \Lambda_2$  for  $\{\mu_1\} \sqsubseteq \Lambda_2$  (and similarly for  $\Lambda_2$ , and similarly using  $\sqsubset$ ). Notice that  $\mu_1 \sqsubseteq \mu_2$  is equivalent to  $\{\mu_1\} \sqsubseteq \{\mu_2\}$ , so this notational shortcut is sound. For two initialized arenas  $(\mathcal{A}_1, s_1)$  and  $(\mathcal{A}_2, s_2)$ , the inequality  $[\mathcal{A}_1]_{s_1}^X \sqsubseteq [\mathcal{A}_2]_{s_2}^X$  means that for every strategy of type  $X$  on  $(\mathcal{A}_1, s_1)$ , there is a strategy of type  $X$  on  $(\mathcal{A}_2, s_2)$  that induces a distribution that is at least as good.

We can now present the two properties of preference relations at the core of our characterization. These properties are called *X-Y-M-monotony* and *X-Y-M-selectivity*; they depend on a type of strategies  $X$ , a type of arenas  $Y$ , and a memory skeleton  $\mathcal{M}$ . The first appearance of the monotony (resp. selectivity) notion was in [GZ05], which dealt with deterministic arenas under pure strategies and memoryless strategies; their monotony (resp. selectivity) is equivalent to our  $\text{P-D-}\mathcal{M}_{\text{triv}}$ -monotony (resp.  $\text{P-D-}\mathcal{M}_{\text{triv}}$ -selectivity). In [BLO<sup>+</sup>20], these definitions were generalized to deal with the sufficiency of pure strategies based on  $\mathcal{M}$  in deterministic arenas; their notion of  $\mathcal{M}$ -monotony (resp.  $\mathcal{M}$ -selectivity) is equivalent to our  $\text{P-D-}\mathcal{M}$ -monotony (resp.  $\text{P-D-}\mathcal{M}$ -selectivity).

**Definition 28 (Monotony).** *We say that  $\sqsubseteq$  is X-Y-M-monotone if for all  $m \in M$ , for all  $(\mathcal{A}_1, s_1), (\mathcal{A}_2, s_2) \in \mathfrak{A}_{\mathcal{P}_1}^Y$ , there exists  $i \in \{1, 2\}$  such that*

$$\forall w \in L_{m_{\text{init}}, m}, w[\mathcal{A}_{3-i}]_{s_{3-i}}^X \sqsubseteq w[\mathcal{A}_i]_{s_i}^X.$$

The crucial part of the definition is the order of the last two quantifiers: of course, given a  $w \in L_{m_{\text{init}}, m}$ , as  $\sqsubseteq$  is total, it will always be the case that  $w[\mathcal{A}_1]_{s_1}^X \sqsubseteq w[\mathcal{A}_2]_{s_2}^X$  or that  $w[\mathcal{A}_2]_{s_2}^X \sqsubseteq w[\mathcal{A}_1]_{s_1}^X$ . However, we ask for something stronger: it must be the case that the set of distributions  $w[\mathcal{A}_i]_{s_i}^X$  is preferred to  $w[\mathcal{A}_{3-i}]_{s_{3-i}}^X$  for any word  $w \in L_{m_{\text{init}}, m}$ .

The original monotony definition [GZ05] states that when presented with a choice *once* among two possible continuations, if a continuation is better than the other one after some prefix, then this continuation is also at least as good after all prefixes. This property is not sufficient for the sufficiency of pure memoryless strategies as it does not guarantee that if the same choice presents itself multiple times in the game, the same continuation should always be chosen, as alternating between both continuations



might still be beneficial in the long run — this is dealt with by selectivity. If memory skeleton  $\mathcal{M}$  is necessary to play optimally, then it makes sense that there might be different optimal choices depending on the current memory state and that we should only compare prefixes that reach the same memory state. The point of taking into account a memory skeleton  $\mathcal{M}$  in our definition of X-Y- $\mathcal{M}$ -monotony is to distinguish classes of prefixes and to only compare prefixes that are read up to the same memory state from  $m_{\text{init}}$ .

**Definition 29 (Selectivity).** *We say that  $\sqsubseteq$  is X-Y- $\mathcal{M}$ -selective if for all  $m \in M$ , for all  $(\mathcal{A}_1, s_1), (\mathcal{A}_2, s_2) \in \mathfrak{A}_{\mathcal{P}_1}^Y$  such that for  $i \in \{1, 2\}$ ,  $\widehat{\text{col}}(\text{Hists}(\mathcal{A}_i, s_i, s_i)) \subseteq L_{m,m}$ , for all  $w \in L_{m_{\text{init}},m}$ ,*

$$w[(\mathcal{A}_1, s_1) \sqcup (\mathcal{A}_2, s_2)]_t^X \sqsubseteq w[\mathcal{A}_1]_{s_1}^X \cup w[\mathcal{A}_2]_{s_2}^X.$$

Our formulation of the selectivity concept differs from the original definition [GZ05] and its AIFM counterpart [BLO<sup>+</sup>20] in order to take into account the particularities of the stochastic context, even if it can be proven that they are equivalent in the pure deterministic case. However, the idea is still the same: the original selectivity definition states that when presented with a choice among multiple possible continuations after some prefix, if a continuation is better than the others, then as the game goes on, if the same choice presents itself again, it is sufficient to always pick the same continuation to play optimally; there is no need to alternate between continuations. This property is not sufficient for the sufficiency of pure memoryless strategies as it does not guarantee that for all prefixes, the same initial choice is always the one we should commit to — this is dealt with by monotony. The point of memory skeleton  $\mathcal{M}$  in our definition is to guarantee that every time the choice presents itself, we are currently in the same memory state  $m$ .

In both definitions, the point of X is to distinguish whether we allow all (including randomized) strategies, or only pure strategies; the point of Y is to quantify over a specific set of arenas.

An interesting property is that both notions are stable by product with a memory skeleton: if  $\sqsubseteq$  is X-Y- $\mathcal{M}$ -monotone (resp. X-Y- $\mathcal{M}$ -selective), then for all memory skeletons  $\mathcal{M}'$ ,  $\sqsubseteq$  is also X-Y- $(\mathcal{M} \otimes \mathcal{M}')$ -monotone (resp. X-Y- $(\mathcal{M} \otimes \mathcal{M}')$ -selective). The reason is that in each definition, we quantify universally over the class of all prefixes  $w$  that reach the same memory state  $m$ ; if we consider classes that are subsets of the original classes, then the definition still holds. This property matches the idea that playing with more memory is never detrimental.

Combined, it is intuitively reasonable that X-Y- $\mathcal{M}$ -monotony and X-Y- $\mathcal{M}$ -selectivity are equivalent to the sufficiency of pure strategies based on  $\mathcal{M}$  to play X-optimally in  $\mathfrak{A}_{\mathcal{P}_1}^Y$ : monotony tells us that when a single choice has to be made given a state of the arena and a memory state, the best choice is always the same no matter what prefix has been seen, and selectivity tells us that once a good choice has been made, we can commit to it in the future of the game. We formalize and prove this idea in Theorem 32. First, we add an extra restriction on preference relations which is useful when stochasticity is involved.

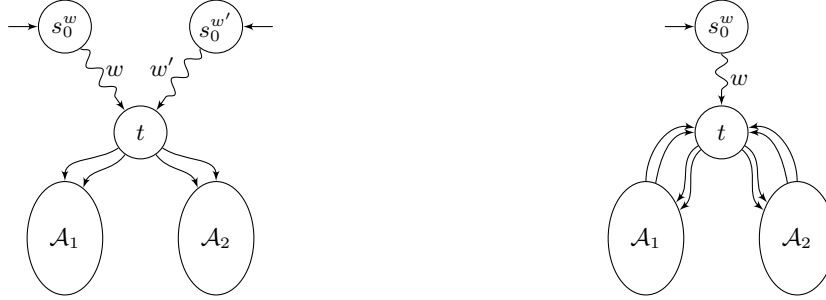
**Definition 30 (Mixing is useless).** *We say that mixing is useless for  $\sqsubseteq$  if for all sets  $I$  at most countable, for all reals  $(\lambda_i)_{i \in I}$  such that  $\sum_{i \in I} \lambda_i = 1$ , for all families  $(\mu_i)_{i \in I}, (\mu'_i)_{i \in I}$  of distributions in  $\text{Dist}(C^\omega, \mathcal{F})$ , if for all  $i \in I$ ,  $\mu_i \sqsubseteq \mu'_i$ , then  $\sum_{i \in I} \lambda_i \mu_i \sqsubseteq \sum_{i \in I} \lambda_i \mu'_i$ .*

That is, if we can write a distribution as a convex combination of distributions, then it is never detrimental to improve a distribution appearing in the convex combination.

*Remark 31.* All preference relations encoded as real payoff functions (as defined in Example 8) satisfy this property, thanks to properties of the expected value. The third preference relation from Example 8 (having a probability to reach  $c \in C$  that is precisely  $\frac{1}{2}$ ) does not satisfy this property: if  $\mu_1(\Diamond c) = 0$ ,  $\mu'_1(\Diamond c) = \frac{1}{2}$ , and  $\mu_2(\Diamond c) = 1$ , we have  $\mu_1 \sqsubset \mu'_1$  and  $\mu_2 \sqsubseteq \mu_2$ , but  $\frac{1}{2}\mu'_1 + \frac{1}{2}\mu_2 \sqsubset \frac{1}{2}\mu_1 + \frac{1}{2}\mu_2$ . In case we consider pure strategies and deterministic games, only Dirac distributions on infinite words occur as probability distributions induced by an arena and a strategy, so the requirement that mixing is useless is not needed.  $\triangleleft$

**Theorem 32.** Assume  $X \in \{P, PFM\}$  and  $Y = D$ , or that mixing is useless for  $\sqsubseteq$ . Then pure strategies based on  $\mathcal{M}$  suffice to play  $X$ -optimally in all initialized one-player arenas in  $\mathfrak{A}_{\mathcal{P}_1}^Y$  for  $\mathcal{P}_1$  if and only if  $\sqsubseteq$  is  $X$ - $Y$ - $\mathcal{M}$ -monotone and  $X$ - $Y$ - $\mathcal{M}$ -selective.

We start with the proof of the necessary condition of Theorem 32, which is the easiest direction. The main idea is to build the right arenas (using the arenas occurring in the definitions of monotony and selectivity) so that we can use the hypothesis about the existence of pure  $X$ -optimal strategies based on  $\mathcal{M}$  to immediately deduce  $X$ - $Y$ - $\mathcal{M}$ -monotony and  $X$ - $Y$ - $\mathcal{M}$ -selectivity. It is not necessary that mixing is useless for  $\sqsubseteq$  for this direction of the equivalence.



**Fig. 2.** Initialized arenas  $(\mathcal{A}_{\text{mon}}, \{s_0^w, s_0^{w'}\})$  (left) and  $(\mathcal{A}_{\text{sel}}, s_0^w)$  (right).

*Proof (of the necessary condition of Theorem 32).* We assume that pure strategies based on  $\mathcal{M}$  suffice to play  $X$ -optimally in  $\mathfrak{A}_{\mathcal{P}_1}^Y$  for  $\mathcal{P}_1$ .

We first prove that  $\sqsubseteq$  is  $X$ - $Y$ - $\mathcal{M}$ -monotone. Let  $m \in M$  and  $(\mathcal{A}_1, s_1), (\mathcal{A}_2, s_2) \in \mathfrak{A}_{\mathcal{P}_1}^Y$  be initialized one-player arenas of  $\mathcal{P}_1$ . If for all  $w \in L_{m_{\text{init}}, m}$ , both  $w[\mathcal{A}_1]_{s_1}^X \sqsubseteq w[\mathcal{A}_2]_{s_2}^X$  and  $w[\mathcal{A}_2]_{s_2}^X \sqsubseteq w[\mathcal{A}_1]_{s_1}^X$ , that is, if both sets of distributions are just as good as each other, then we can take either  $i = 1$  or  $i = 2$  and the definition is satisfied. If that is not the case, this means that there exists  $w' \in L_{m_{\text{init}}, m}$  such that, w.l.o.g.,

$$w'[\mathcal{A}_1]_{s_1}^X \sqsubset w'[\mathcal{A}_2]_{s_2}^X. \quad (7)$$

We take  $i = 2$ . It is left to show that for all  $w \in L_{m_{\text{init}}, m}$ ,

$$w[\mathcal{A}_1]_{s_1}^X \sqsubseteq w[\mathcal{A}_2]_{s_2}^X.$$

Let  $w \in L_{m_{\text{init}}, m}$ . For  $j \in \{1, 2\}$ , we assume w.l.o.g. that state  $s_j$  has no incoming transition in  $\mathcal{A}_j$ , and therefore cannot be reached after being left. If it is not the case, we can create a new one-player arena  $\mathcal{A}'_j$  by adding a new state  $s'_j$  mimicking the outgoing transitions of  $s_j$ , but without any ingoing transition, and we have  $[\mathcal{A}_j]_{s_j}^X = [\mathcal{A}'_j]_{s'_j}^X$ . We also assume w.l.o.g. that the state and action spaces of  $\mathcal{A}_1$  and  $\mathcal{A}_2$  are disjoint.

We consider the arena

$$\mathcal{A}_{\text{mon}} = (((\mathcal{A}_1, s_1) \sqcup (\mathcal{A}_2, s_2))_{w \rightsquigarrow t})_{w' \rightsquigarrow t}$$

where  $t$  is the state resulting from the merge of  $s_1$  and  $s_2$ . We consider two initial states  $s_0^w$  and  $s_0^{w'}$ , which are the states at the start of the “chains” corresponding respectively to  $w$  and  $w'$ . Arena  $\mathcal{A}_{\text{mon}}$  is depicted on Figure 2;  $\mathcal{A}_{\text{mon}}$  consists of two chains reading  $w$  and  $w'$  up to state  $t$ , and then a choice between going to  $\mathcal{A}_1$  or  $\mathcal{A}_2$ , with no possibility of ever going back to  $t$ .

The initialized arena  $(\mathcal{A}_{\text{mon}}, \{s_0^w, s_0^{w'}\})$  is in  $\mathfrak{A}_{\mathcal{P}_1}^Y$  as all operations used preserve the number of players and the deterministic/stochastic feature. By hypothesis,  $\mathcal{P}_1$  has a pure  $X$ -optimal strategy  $\sigma_1 \in \Sigma_1^{\text{PFM}}(\mathcal{A}_{\text{mon}}, \{s_0^w, s_0^{w'}\})$  encoded as a Mealy machine  $(\mathcal{M}, \alpha_{\text{next}})$ . Remember that both  $w$  and  $w'$  reach state  $m$  of the memory skeleton  $\mathcal{M}$  when read from  $m_{\text{init}}$ . Therefore, no matter whether the play starts

in  $s_0^w$  or  $s_0^{w'}$ , the action played in  $t$  by strategy  $\sigma_1$  is given by  $\alpha_{\text{next}}(t, m)$  (which cannot be a randomized choice, as  $\sigma_1$  is pure). Since  $\sigma_1$  is X-optimal in  $(\mathcal{A}_{\text{mon}}, s_0^{w'})$ , by (7), this action must necessarily be an action of  $\mathcal{A}_2$ . Now since  $\sigma_1$  is also X-optimal in  $(\mathcal{A}_{\text{mon}}, s_0^w)$ , this means that going to  $\mathcal{A}_2$  after  $w$  is at least as good as going to  $\mathcal{A}_1$ . In other words, we have

$$w[\mathcal{A}_1]_{s_1}^{\text{X}} \subseteq w[\mathcal{A}_2]_{s_2}^{\text{X}},$$

which ends the X-Y-M-monotony proof.

We now prove that  $\sqsubseteq$  is X-Y-M-selective. Let  $m \in M$  and  $(\mathcal{A}_1, s_1), (\mathcal{A}_2, s_2) \in \mathfrak{A}_{\mathcal{P}_1}^{\text{Y}}$  such that for  $i \in \{1, 2\}$ ,  $\text{Hists}(\mathcal{A}_i, s_i, s_i) \subseteq L_{m, m}$ . Let  $w \in L_{m_{\text{init}}, m}$ . We consider the arena

$$\mathcal{A}_{\text{sel}} = ((\mathcal{A}_1, s_1) \sqcup (\mathcal{A}_2, s_2))_{w \rightsquigarrow t}$$

where  $t$  is the state resulting from the merge of  $s_1$  and  $s_2$ . We consider an initial state  $s_0^w$ , which is the state at the start of the “chain” corresponding to  $w$ . Arena  $\mathcal{A}_{\text{sel}}$  is depicted on Figure 2;  $\mathcal{A}_{\text{sel}}$  consists of one chain reading  $w$  up to state  $t$ , and then has the ability to go either to  $\mathcal{A}_1$  and  $\mathcal{A}_2$ . Here, it is possible to visit  $t$  multiple times (as long as it was possible to go back to  $s_1$  in  $(\mathcal{A}_1, s_1)$  or to  $s_2$  in  $(\mathcal{A}_2, s_2)$ ).

The initialized arena  $(\mathcal{A}_{\text{sel}}, s_0^w)$  is in  $\mathfrak{A}_{\mathcal{P}_1}^{\text{Y}}$  as all operations used preserve the number of players and the deterministic/stochastic feature. By hypothesis,  $\mathcal{P}_1$  has a pure X-optimal strategy  $\sigma_1 \in \Sigma_1^{\text{PFM}}(\mathcal{A}_{\text{mon}}, s_0^w)$  encoded as a Mealy machine  $(\mathcal{M}, \alpha_{\text{next}})$ . By X-optimality of  $\sigma_1$ , we have that

$$w[(\mathcal{A}_1, s_1) \sqcup (\mathcal{A}_2, s_2)]_t^{\text{X}} \subseteq \text{Pc}_{\mathcal{A}_{\text{sel}}, s_0^w}^{\sigma_1}. \quad (8)$$

Since  $w$  is in  $L_{m_{\text{init}}, m}$  and for  $i \in \{1, 2\}$ ,  $\widehat{\text{col}}(\text{Hists}(\mathcal{A}_i, s_i, s_i))$  is a subset of  $L_{m, m}$ , we have that  $\widehat{\text{col}}(\text{Hists}(\mathcal{A}_{\text{sel}}, s_0^w, t))$  is a subset of  $L_{m_{\text{init}}, m}$ . Therefore, at each passage in  $t$ , strategy  $\sigma_1$  plays the action given by  $\alpha_{\text{next}}(t, m)$  (which cannot be a randomized choice, as  $\sigma_1$  is pure). Strategy  $\sigma_1$  thus commits to  $\mathcal{A}_1$  or  $\mathcal{A}_2$  forever, which means that

$$\text{Pc}_{\mathcal{A}_{\text{sel}}, s_0^w}^{\sigma_1} \in w[\mathcal{A}_1]_{s_1}^{\text{X}} \cup w[\mathcal{A}_2]_{s_2}^{\text{X}}.$$

By combining this last fact with (8), we obtain that

$$w[(\mathcal{A}_1, s_1) \sqcup (\mathcal{A}_2, s_2)]_t^{\text{X}} \subseteq w[\mathcal{A}_1]_{s_1}^{\text{X}} \cup w[\mathcal{A}_2]_{s_2}^{\text{X}},$$

which ends the X-Y-M-selectivity proof.  $\square$

We sketch the proof of the sufficient condition of Theorem 32. We first reduce the problem to the existence of pure memoryless strategies in initialized arenas covered by  $\mathcal{M}$ , using Lemma 21. We proceed with an induction on the number of choices in these arenas (as for Theorem 23). The base case is again trivial (as in an arena in which all states have a single available action, there is a single strategy which is pure and memoryless). For the induction step, we take an initialized arena  $(\mathcal{A}', S_{\text{init}}) \in \mathfrak{A}_{\mathcal{P}_1}^{\text{Y}}$  covered by  $\mathcal{M}$  with at least one choice, and we pick a state  $t$  with (at least) two available actions. A memory state  $\phi(t)$  is associated to  $t$  thanks to the coverability property. We consider the subarenas  $(\mathcal{A}'_a, S_{\text{init}})$  with a single action  $a$  available in  $t$ , to which we can apply the induction hypothesis and obtain a pure memoryless X-optimal strategy  $\sigma_1^a$  in each subarena. It is left to prove that one of these strategies is also X-optimal in  $(\mathcal{A}', S_{\text{init}})$  — this is where X-Y-M-monotony and X-Y-M-selectivity come into play.

The property of X-Y-M-monotony tells us that one of these subarenas  $(\mathcal{A}'_{a^*}, S_{\text{init}})$  is preferred to the others w.r.t.  $\sqsubseteq$  after reading any word in  $L_{m_{\text{init}}, \phi(t)}$ . We now want to use X-Y-M-selectivity to conclude that there is no reason to use actions different from  $a^*$  when coming back to  $t$ , and that  $\sigma_1^{a^*}$  is therefore also X-optimal in  $(\mathcal{A}', S_{\text{init}})$ . To do so, we take any strategy  $\sigma_1 \in \Sigma_1^{\text{X}}(\mathcal{A}', s)$  for  $s \in S_{\text{init}}$  and we condition distribution  $\text{P}_{\mathcal{A}', s}^{\sigma_1}$  over all the ways it reaches (or not)  $t$ , which gives a convex combination of probability distributions. We want to state that once  $t$  is reached, no matter how, switching to strategy  $\sigma_1^{a^*}$  is always beneficial. For this, we would like to use X-subgame-perfection of  $\sigma_1^{a^*}$  rather than simply X-optimality: this is why in the actual proof, our induction hypothesis is about X-SP strategies and not X-optimal

strategies. Luckily, Theorem 22 indicates that requiring subgame perfection is not really stronger than what we want to prove. We then use that mixing is useless for  $\sqsubseteq$  to replace all the parts that go through  $t$  in the convex combination by a better distribution induced by  $\sigma_1^{a^*}$  from  $t$ .

We need two (intuitive) technical lemmas, whose proofs can be found in Appendix B. We first define a similar notion to *shifted distributions* (Definition 13) for distributions on plays: for  $(\mathcal{A}, s)$  an initialized one-player arena, for  $\rho \in \text{Hists}(\mathcal{A}, s)$ , if  $\mu \in \text{Dist}(\text{Plays}(\mathcal{A}, \text{out}(\rho)), \mathcal{F}_{(\mathcal{A}, \text{out}(\rho))})$  is a distribution on plays, then for  $E \in \mathcal{F}_{(\mathcal{A}, s)}$  an event, we define

$$\rho\mu(E) = \mu(\{\pi \in \text{Plays}(\mathcal{A}, \text{out}(\rho)) \mid \rho\pi \in E\}).$$

We have used here an abuse of notation: if  $\rho = s_0a_1s_1 \dots a_ns_n$ , for  $\pi = s_na_{n+1}s_{n+1} \dots \in \text{Plays}(\mathcal{A}, \text{out}(\rho))$ , we write  $\rho\pi$  for the play  $s_0a_1s_1 \dots a_ns_na_{n+1}s_{n+1} \dots$ , with no repetition of  $s_n$ .

**Lemma 33.** *Let  $(\mathcal{A} = (S_1, S_2, A, \delta, \text{col}), s)$  be an initialized one-player arena and  $\rho \in \text{Hists}(\mathcal{A}, s)$ . Let  $\mu$  be a distribution on plays in  $\text{Dist}(\text{Plays}(\mathcal{A}, \text{out}(\rho)), \mathcal{F}_{(\mathcal{A}, \text{out}(\rho))})$ . We have*

$$\widehat{\text{col}}(\rho\mu) = \widehat{\text{col}}(\rho)\widehat{\text{col}}(\mu).$$

**Lemma 34.** *Let  $(\mathcal{A}, s) \in \mathfrak{A}_{\mathcal{P}_1}^Y$  be an initialized one-player arena and  $\sigma_1, \tau_1 \in \Sigma_1^G(\mathcal{A}, s)$  be two strategies. Let  $\rho = s_0a_1s_1 \dots a_ns_n \in \text{Hists}(\mathcal{A}, s)$ . We say that  $\sigma_1$  coincides with  $\tau_1$  on  $\rho$  if for each prefix  $\rho_i = s_0a_1s_1 \dots a_is_i$  of  $\rho$  with  $0 \leq i < n$ ,  $\sigma_1(\rho_i) = \tau_1(\rho_i)$ . If  $\sigma_1$  coincides with  $\tau_1$  on  $\rho$ , then*

$$P_{\mathcal{A}, s}^{\sigma_1}[\text{Cyl}(\rho)] = P_{\mathcal{A}, s}^{\tau_1}[\text{Cyl}(\rho)].$$

Let  $t$  be a state of  $\mathcal{A}$ . We write  $\neg\Diamond t$  for the event in  $\mathcal{F}_{(\mathcal{A}, s)}$  that consists of all the infinite plays that never visit  $t$ . Assume that for all  $\rho = s_0a_1s_1 \dots a_ns_n \in \text{Hists}(\mathcal{A}, s)$  such that for all  $i$ ,  $0 \leq i < n$ ,  $s_i \neq t$ ,  $\sigma_1$  coincides with  $\tau_1$  on  $\rho$ . Then

$$P_{\mathcal{A}, s}^{\sigma_1}[\neg\Diamond t] = P_{\mathcal{A}, s}^{\tau_1}[\neg\Diamond t] \text{ and, if } P_{\mathcal{A}, s}^{\sigma_1}[\neg\Diamond t] > 0, P_{\mathcal{A}, s}^{\sigma_1}[\cdot \mid \neg\Diamond t] = P_{\mathcal{A}, s}^{\tau_1}[\cdot \mid \neg\Diamond t].$$

We now have all the ingredients for the proof of the missing implication of Theorem 32.

*Proof (of the sufficient condition of Theorem 32).* We assume now that mixing is useless for  $\sqsubseteq$ , and that  $\sqsubseteq$  is X-Y- $\mathcal{M}$ -monotone and X-Y- $\mathcal{M}$ -selective. We prove that pure strategies based on  $\mathcal{M}$  suffice to play X-optimally in  $\mathfrak{A}_{\mathcal{P}_1}^Y$  for  $\mathcal{P}_1$ . Equivalently, thanks to Lemma 21, we show that for all initialized arenas covered by  $\mathcal{M}$  in  $\mathfrak{A}_{\mathcal{P}_1}^Y$ ,  $\mathcal{P}_1$  has a pure memoryless X-optimal strategy. We will actually prove something stronger, which is that for all initialized one-player arenas covered by  $\mathcal{M}$  in  $\mathfrak{A}_{\mathcal{P}_1}^Y$ ,  $\mathcal{P}_1$  has a pure memoryless X-SP strategy.

Let  $(\mathcal{A}, S_{\text{init}}) \in \mathfrak{A}_{\mathcal{P}_1}^Y$  be an initialized one-player arena covered by  $\mathcal{M}$ . Our proof proceeds by induction on the number of choices  $n_{\mathcal{A}'}$  of subarenas  $(\mathcal{A}', S_{\text{init}})$  of  $(\mathcal{A}, S_{\text{init}})$ . Our induction will prove the following property for subarenas  $(\mathcal{A}', S_{\text{init}})$  of  $(\mathcal{A}, S_{\text{init}})$ : there exists a pure memoryless strategy  $\sigma_1 \in \Sigma_1^{\text{PFM}}(\mathcal{A}', S_{\text{init}})$  such that for all  $\rho \in \text{Hists}(\mathcal{A}, S_{\text{init}})$ ,  $\sigma_1$  is X-optimal in the game  $(\mathcal{A}', \text{out}(\rho), \sqsubseteq_{[\widehat{\text{col}}(\rho)]})$ . We call this property *having a pure memoryless  $(\mathcal{A}, S_{\text{init}})$ -X-SP strategy*. There is a slight abuse of notation in the definition:  $\sigma_1$  is not necessarily well-defined from  $\text{out}(\rho)$ , but as it is pure memoryless, we simply interpret it as a function  $S \rightarrow A$ , and for  $\rho' \in \text{Hists}(\mathcal{A}', \text{out}(\rho))$ , we define  $\sigma_1(\rho') = \sigma_1(\text{out}(\rho'))$ . For subarenas  $(\mathcal{A}', S_{\text{init}})$ , having a pure memoryless  $(\mathcal{A}, S_{\text{init}})$ -X-SP strategy is stronger than having a pure memoryless X-SP strategy, as  $\text{Hists}(\mathcal{A}', S_{\text{init}})$  is a subset of  $\text{Hists}(\mathcal{A}, S_{\text{init}})$ . For arena  $(\mathcal{A}, S_{\text{init}})$ , having a pure memoryless  $(\mathcal{A}, S_{\text{init}})$ -X-SP strategy is equivalent to having a pure memoryless X-SP strategy, which is what we want to prove. Requiring SP strategies instead of simply optimal strategies may seem stronger than what we actually need, but by Theorem 22, it turns out being equivalent in this AIFM context; we use SP strategies in this case for technical reasons.

Let  $(\mathcal{A}' = (S_1, S_2, A', \delta', \text{col}'), S_{\text{init}})$  be a subarena of  $(\mathcal{A}, S_{\text{init}})$ . If  $n_{\mathcal{A}'} = 0$ , then  $\mathcal{P}_1$  has only one strategy which is in particular a pure memoryless  $(\mathcal{A}, S_{\text{init}})$ -X-SP strategy (notation  $n_{\mathcal{A}'}$  is defined

at (6)). Now let  $n > 0$ ; we assume that the property is true for all arenas  $(\mathcal{A}', S_{\text{init}})$  such that  $n_{\mathcal{A}'} < n$ , and we take  $(\mathcal{A}', S_{\text{init}})$  such that  $n_{\mathcal{A}'} = n$ . Since  $n > 0$ , there is a state  $t \in S_1$  such that  $|A'(t)| \geq 2$ .

For  $a \in A'(t)$ , let  $(\mathcal{A}'_a, S_{\text{init}}) \in \mathfrak{A}_{\mathcal{P}_1}^Y$  be the initialized subarena of  $(\mathcal{A}', S_{\text{init}})$  such that only action  $a$  is available in  $t$ . Initialized arena  $(\mathcal{A}'_a, S_{\text{init}})$  is covered by  $\mathcal{M}$  (Lemma 27). By induction hypothesis, for all  $a \in A'(t)$ ,  $\mathcal{P}_1$  has a pure memoryless  $(\mathcal{A}, S_{\text{init}})$ -X-SP strategy  $\sigma_1^a$  in  $(\mathcal{A}'_a, S_{\text{init}})$ .

Let  $m \in M$  be the memory state corresponding to  $t$  in  $(\mathcal{A}, S_{\text{init}})$ , that is, if  $\phi$  is the function witnessing that  $(\mathcal{A}, S_{\text{init}})$  is covered,  $m = \phi(t)$ . The same function  $\phi$  also witnesses that all the initialized subarenas of  $(\mathcal{A}, S_{\text{init}})$  are covered by  $\mathcal{M}$ . As  $\sqsubseteq$  is X-Y- $\mathcal{M}$ -monotone, there exists  $a^* \in A'(t)$  such that for all  $w \in L_{m_{\text{init}}, m}$ , for all  $a \in A'(t)$ ,

$$w[\mathcal{A}'_a]_t^X \sqsubseteq w[\mathcal{A}'_{a^*}]_t^X. \quad (9)$$

Notice that as  $(\mathcal{A}, S_{\text{init}})$  is covered by  $\mathcal{M}$ ,  $\widehat{\text{col}}(\text{Hists}(\mathcal{A}, S_{\text{init}}, t)) \subseteq L_{m_{\text{init}}, m}$ .

We now prove that the pure memoryless strategy  $\sigma_1^{a^*}$  is  $(\mathcal{A}, S_{\text{init}})$ -X-SP in  $(\mathcal{A}', S_{\text{init}})$ . Let  $\rho \in \text{Hists}(\mathcal{A}, S_{\text{init}})$ . We denote  $w = \widehat{\text{col}}(\rho)$  and  $s = \text{out}(\rho)$ .

Let  $\sigma_1$  be any strategy in  $\Sigma_1^X(\mathcal{A}', s)$ . Our goal is to show that  $\sigma_1^{a^*}$  is at least as good as  $\sigma_1$  in  $(\mathcal{A}', s, \sqsubseteq_{[w]})$ , i.e., that  $w\text{Pc}_{\mathcal{A}', s}^{\sigma_1} \sqsubseteq w\text{Pc}_{\mathcal{A}', s}^{\sigma_1^{a^*}}$ . We condition  $\text{P}_{\mathcal{A}', s}^{\sigma_1}$  over whether  $t$  is visited or not (we assume that  $t$  is both visited and not visited with a non-zero probability — otherwise, one of the terms of the following sum is simply 0). We denote by  $\diamond t$  the event of visiting state  $t$  and by  $\text{Hists}(\mathcal{A}', s, t!)$  the set of histories in  $\text{Hists}(\mathcal{A}', s, t)$  that visit  $t$  exactly *once* (at their last step). We have

$$\begin{aligned} \text{P}_{\mathcal{A}', s}^{\sigma_1} &= \text{P}_{\mathcal{A}', s}^{\sigma_1}[-\diamond t] \cdot \text{P}_{\mathcal{A}', s}^{\sigma_1}[\cdot \mid -\diamond t] + \text{P}_{\mathcal{A}', s}^{\sigma_1}[\diamond t] \cdot \text{P}_{\mathcal{A}', s}^{\sigma_1}[\cdot \mid \diamond t] \\ &= \text{P}_{\mathcal{A}', s}^{\sigma_1}[-\diamond t] \cdot \text{P}_{\mathcal{A}', s}^{\sigma_1}[\cdot \mid -\diamond t] + \sum_{\rho' \in \text{Hists}(\mathcal{A}', s, t!)} \text{P}_{\mathcal{A}', s}^{\sigma_1}[\text{Cyl}(\rho')] \cdot \text{P}_{\mathcal{A}', s}^{\sigma_1}[\cdot \mid \text{Cyl}(\rho')] \\ &= \text{P}_{\mathcal{A}', s}^{\sigma_1}[-\diamond t] \cdot \text{P}_{\mathcal{A}', s}^{\sigma_1}[\cdot \mid -\diamond t] + \sum_{\rho' \in \text{Hists}(\mathcal{A}', s, t!)} \text{P}_{\mathcal{A}', s}^{\sigma_1}[\text{Cyl}(\rho')] \cdot \rho' \text{P}_{\mathcal{A}', t}^{\sigma_1[\rho']}. \end{aligned}$$

By applying operator  $\widehat{\text{col}}$  to  $\text{P}_{\mathcal{A}', s}^{\sigma_1}$  and shifting the distribution with  $w$ , thanks to Lemma 33 and the previous equation, we have

$$w\text{Pc}_{\mathcal{A}', s}^{\sigma_1} = \text{P}_{\mathcal{A}', s}^{\sigma_1}[-\diamond t] \cdot w\widehat{\text{col}}(\text{P}_{\mathcal{A}', s}^{\sigma_1}[\cdot \mid -\diamond t]) + \sum_{\substack{\rho' \in \text{Hists}(\mathcal{A}', s, t!) \\ \text{s.t. } \widehat{\text{col}}(\rho') = w'}} \text{P}_{\mathcal{A}', s}^{\sigma_1}[\text{Cyl}(\rho')] \cdot ww'\text{Pc}_{\mathcal{A}', t}^{\sigma_1[\rho']}. \quad (10)$$

For  $\rho' \in \text{Hists}(\mathcal{A}', s, t!)$ ,  $w' = \widehat{\text{col}}(\rho')$ , let us focus on the distribution  $ww'\text{Pc}_{\mathcal{A}', t}^{\sigma_1[\rho']}$ . Notice that distribution  $\text{Pc}_{\mathcal{A}', t}^{\sigma_1[\rho']}$  can also be induced by some strategy in  $\Sigma_1^X(\text{Split}(\mathcal{A}', t), t)$  by Lemma 35. Therefore,

$$ww'\text{Pc}_{\mathcal{A}', t}^{\sigma_1[\rho']} \in ww'[\text{Split}(\mathcal{A}', t)]_t^X = ww'[\bigsqcup_{a \in A'(t)} (\mathcal{A}'_a, t)]_t^X.$$

Using the hypotheses, we get

$$\begin{aligned} ww'\text{Pc}_{\mathcal{A}', t}^{\sigma_1[\rho']} &\in ww'[\bigsqcup_{a \in A'(t)} (\mathcal{A}'_a, t)]_t^X \\ &\sqsubseteq \bigcup_{a \in A'(t)} ww'[\mathcal{A}'_a]_t^X && \text{by X-Y-}\mathcal{M}\text{-selectivity} \\ &\sqsubseteq ww'[\mathcal{A}'_{a^*}]_t^X && \text{by (9), which relied on X-Y-}\mathcal{M}\text{-monotony} \\ &\sqsubseteq ww'\text{Pc}_{\mathcal{A}'_{a^*}, t}^{\sigma_1^{a^*}} && \text{as } \sigma_1^{a^*} \text{ is pure memoryless } (\mathcal{A}, S_{\text{init}})\text{-X-SP in } (\mathcal{A}'_{a^*}, S_{\text{init}}). \end{aligned}$$

Therefore, by using this last equation in (10), thanks to the fact that mixing is useless for  $\sqsubseteq$  (or, if we consider pure strategies and deterministic arenas, that the sum contains a single term corresponding

to an infinite word), we obtain

$$w\text{Pc}_{\mathcal{A}',s}^{\sigma_1} \sqsubseteq \text{P}_{\mathcal{A}',s}^{\sigma_1}[\neg\Diamond t] \cdot w\widehat{\text{col}}(\text{P}_{\mathcal{A}',s}^{\sigma_1}[\cdot \mid \neg\Diamond t]) + \sum_{\substack{\rho' \in \text{Hists}(\mathcal{A},s,t!) \\ \text{with } \widehat{\text{col}}(\rho')=w'}} \text{P}_{\mathcal{A}',s}^{\sigma_1}[\text{Cyl}(\rho')] \cdot ww'\text{Pc}_{\mathcal{A}'_{a^*},t}^{\sigma_1^{a^*}}. \quad (11)$$

We show that the right-hand side of this inequation can be written as a distribution  $w\text{Pc}_{\mathcal{A}'_{a^*},s}^{\tau_1}$ , for a suitably chosen strategy  $\tau_1 \in \Sigma_1^X(\mathcal{A}'_{a^*}, s)$ .

Let  $\tau_1 \in \Sigma_1^X(\mathcal{A}'_{a^*}, s)$  be such that  $\tau_1$  starts playing like  $\sigma_1$  and then switches to  $\sigma_1^{a^*}$  as soon as  $t$  is visited; formally, for  $\rho'' \in \text{Hists}(\mathcal{A}'_{a^*}, s)$ ,

$$\tau_1(\rho'') = \begin{cases} \sigma_1(\rho'') & \text{if } \rho'' \text{ does not visit } t \\ \sigma_1^{a^*}(\text{out}(\rho'')) & \text{if } \rho'' \text{ visits } t. \end{cases}$$

Strategy  $\tau_1$  only plays action  $a^*$  in  $t$ , and is therefore a strategy on  $(\mathcal{A}'_{a^*}, s)$ . As  $\tau_1$  coincides with  $\sigma_1$  as long as  $t$  has not been visited, using Lemma 34, we have

$$\text{P}_{\mathcal{A}',s}^{\sigma_1}[\neg\Diamond t] = \text{P}_{\mathcal{A}'_{a^*},s}^{\tau_1}[\neg\Diamond t], \quad \text{P}_{\mathcal{A}',s}^{\sigma_1}[\cdot \mid \neg\Diamond t] = \text{P}_{\mathcal{A}'_{a^*},s}^{\tau_1}[\cdot \mid \neg\Diamond t], \quad \text{P}_{\mathcal{A}',s}^{\sigma_1}[\text{Cyl}(\rho')] = \text{P}_{\mathcal{A}'_{a^*},s}^{\tau_1}[\text{Cyl}(\rho')].$$

Moreover, for all  $\rho' \in \text{Hists}(\mathcal{A}', s, t!)$ ,

$$\text{Pc}_{\mathcal{A}'_{a^*},t}^{\sigma_1^{a^*}} = \text{Pc}_{\mathcal{A}'_{a^*},t}^{\tau_1[\rho']}$$

as  $t$  is immediately visited. We can therefore replace all terms of the right-hand side of (11) and obtain, using Lemma 33,

$$\begin{aligned} w\text{Pc}_{\mathcal{A}',s}^{\sigma_1} &\sqsubseteq \text{P}_{\mathcal{A}'_{a^*},s}^{\tau_1}[\neg\Diamond t] \cdot w\widehat{\text{col}}(\text{P}_{\mathcal{A}'_{a^*},s}^{\tau_1}[\cdot \mid \neg\Diamond t]) + \sum_{\substack{\rho' \in \text{Hists}(\mathcal{A},s,t!) \\ \text{with } \widehat{\text{col}}(\rho')=w'}} \text{P}_{\mathcal{A}'_{a^*},s}^{\tau_1}[\text{Cyl}(\rho')] \cdot ww'\text{Pc}_{\mathcal{A}'_{a^*},t}^{\tau_1[\rho']} \\ &\sqsubseteq w\widehat{\text{col}}(\text{P}_{\mathcal{A}'_{a^*},s}^{\tau_1}[\neg\Diamond t] \cdot \text{P}_{\mathcal{A}'_{a^*},s}^{\tau_1}[\cdot \mid \neg\Diamond t]) + \sum_{\rho' \in \text{Hists}(\mathcal{A},s,t!)} \text{P}_{\mathcal{A}'_{a^*},s}^{\tau_1}[\text{Cyl}(\rho')] \cdot \rho'\text{P}_{\mathcal{A}'_{a^*},t}^{\tau_1[\rho']} \\ &= w\widehat{\text{col}}(\text{P}_{\mathcal{A}'_{a^*},s}^{\tau_1}[\neg\Diamond t] \cdot \text{P}_{\mathcal{A}'_{a^*},s}^{\tau_1}[\cdot \mid \neg\Diamond t]) + \text{P}_{\mathcal{A}'_{a^*},s}^{\tau_1}[\Diamond t] \cdot \text{P}_{\mathcal{A}'_{a^*},s}^{\tau_1}[\cdot \mid \Diamond t]) \\ &= w\widehat{\text{col}}(\text{P}_{\mathcal{A}'_{a^*},s}^{\tau_1}) \\ &= w\text{Pc}_{\mathcal{A}'_{a^*},s}^{\tau_1}. \end{aligned}$$

Now since  $w$  is the sequence of colors corresponding to  $\rho \in \text{Hists}(\mathcal{A}, S_{\text{init}})$  and  $\sigma_1^{a^*}$  is  $(\mathcal{A}, S_{\text{init}})$ -X-SP in  $(\mathcal{A}'_{a^*}, S_{\text{init}})$ , we have  $w\text{Pc}_{\mathcal{A}'_{a^*},s}^{\tau_1} \sqsubseteq w\text{Pc}_{\mathcal{A}'_{a^*},s}^{\sigma_1^{a^*}}$ , which ends the proof.  $\square$

We provide an application of Theorem 32 in Section 6, proving that a preference relation admits pure AIFM optimal strategies in its one-player games. The literature provides some *sufficient* conditions for preference relations to admit pure memoryless optimal strategies in one-player stochastic games (for instance, in [Gim07]). Here, we obtain a full characterization when mixing is useless for  $\sqsubseteq$  (in particular, this is a full characterization for real payoff functions), which can deal not only with memoryless strategies, but also with the more general AIFM strategies. We do not believe that our characterization is necessarily easy to use in practice; when the sufficient conditions of [Gim07] are verified, they are arguably easier to verify in general. However, there are examples in which these sufficient conditions are not verified even though pure memoryless strategies suffice (one such example is provided in [BBE10]), and that is where our characterization can help.

It is interesting to relate the concepts of monotony and selectivity to other properties from the literature to simplify the use of our characterization. For instance, if a payoff function  $f: C^\omega \rightarrow \mathbb{R}$  is *prefix-independent*, then it is also X-Y-M-monotone for any X, Y, and M; therefore, the problem immediately reduces to analyzing selectivity in order to study the sufficiency of pure AIFM strategies with  $f$ . We use such a reasoning in the stochastic case of the application of Section 6.



## 6 Application

Let  $C = \mathbb{N}$ . We illustrate the use of our two main theorems (Theorems 32 and 23) to study the memory requirements of the *weak parity* [Tho08] winning condition

$$W_{\text{wp}} = \{c_1 c_2 \dots \in C^\omega \mid \max_{j \geq 1} c_j \text{ exists and is even}\},$$

which was introduced in Example 8, both in deterministic and in stochastic games. In this example, we abusively use  $W_{\text{wp}}$  for  $\sqsubseteq_{W_{\text{wp}}}$ . We say that a word  $w \in C^\omega$  is *winning* if  $w \in W_{\text{wp}}$ , and *losing* if  $w \notin W_{\text{wp}}$ . As this preference relation can be encoded as a payoff function (namely, the indicator function of  $W_{\text{wp}}$ ), we have that mixing is useless for  $W_{\text{wp}}$ .

*Deterministic games.* We first focus on deterministic games with pure strategies: we show that pure memoryless strategies are sufficient. To do so, we first consider one-player games — notice that reasoning about one-player games of  $\mathcal{P}_1$  and of  $\mathcal{P}_2$  is very similar, as the objective of  $\mathcal{P}_2$  can be rephrased as the objective of  $\mathcal{P}_1$  just by replacing all colors  $c$  by  $c+1$ . We thus only show arguments from the point of view of  $\mathcal{P}_1$ . We prove that the class  $\mathfrak{A}_{\mathcal{P}_1}^{\text{D}}$  of all initialized one-player deterministic arenas of  $\mathcal{P}_1$  admits pure memoryless P-optimal strategies (i.e., pure P-optimal strategies based on the trivial memory skeleton  $\mathcal{M}_{\text{triv}}$  with a single state) by proving that  $W_{\text{wp}}$  is P-D- $\mathcal{M}_{\text{triv}}$ -monotone and P-D- $\mathcal{M}_{\text{triv}}$ -selective.

We start with P-D- $\mathcal{M}_{\text{triv}}$ -monotony. Let  $(\mathcal{A}_1, s_1), (\mathcal{A}_2, s_2) \in \mathfrak{A}_{\mathcal{P}_1}^{\text{D}}$ . Notice that as we are restricted to pure strategies in deterministic arenas, notation  $[\mathcal{A}_i]_{s_i}^{\text{P}}$  refers to a set of (Dirac distributions on) infinite words. For  $i \in \{1, 2\}$  let

$$e_i = \max\{\max_{j \geq 1} c_j \mid c_1 c_2 \dots \in [\mathcal{A}_i]_{s_i}^{\text{P}} \cap W_{\text{wp}}\}$$

be the greatest *even* color reachable in  $(\mathcal{A}_i, s_i)$  without reaching any greater color (or  $-\infty$  if it is not possible to have an even maximal color).

We first deal with the case  $e_1 \neq -\infty$  or  $e_2 \neq -\infty$ . Assume w.l.o.g. that  $e_1 \leq e_2$ . Let  $\sigma_1^2 \in \Sigma_1^{\text{P}}(\mathcal{A}_2, s_2)$  be a pure strategy achieving a maximal color exactly  $e_2$ . We prove that for any word  $w \in C^*$  (that is, for any word  $w \in L_{m_{\text{init}}, m_{\text{init}}}$ , as the memory skeleton has only one state), we have

$$w[\mathcal{A}_1]_{s_1}^{\text{P}} \sqsubseteq w\text{Pc}_{\mathcal{A}_2, s_2}^{\sigma_1^2}. \quad (12)$$

Let  $w = c_1 c_2 \dots c_n \in C^*$  and  $n_w = \max_{1 \leq j \leq n} c_j$ . If  $n_w$  is even or  $n_w \leq e_2$ , then  $w\text{Pc}_{\mathcal{A}_2, s_2}^{\sigma_1^2}$  is a winning word and (12) holds. If not, it means that  $n_w$  is odd and  $n_w > e_2 \geq e_1$ , in which case all words in  $w[\mathcal{A}_1]_{s_1}^{\text{P}}$  are necessarily losing, and (12) also holds.

We now deal with the case  $e_1 = e_2 = -\infty$ , in which there is no way to obtain an even maximal color both in  $\mathcal{A}_1$  and in  $\mathcal{A}_2$ . For  $i \in \{1, 2\}$ , let

$$o_i = \min\{\max_{j \geq 1} c_j \mid c_1 c_2 \dots \in [\mathcal{A}_i]_{s_i}^{\text{P}}\}$$

be the minimal greatest color appearing along a play (which is necessarily odd, as an even greatest color is not possible). Assume w.l.o.g. that  $o_1 \geq o_2$ , and let  $\sigma_1^2 \in \Sigma_1^{\text{P}}(\mathcal{A}_2, s_2)$  be a pure strategy achieving a maximal color exactly  $o_2$ . We show again that for all  $w \in C^*$ ,

$$w[\mathcal{A}_1]_{s_1}^{\text{P}} \sqsubseteq w\text{Pc}_{\mathcal{A}_2, s_2}^{\sigma_1^2}. \quad (13)$$

Let  $w = c_1 c_2 \dots c_n \in C^*$  and  $n_w = \max_{1 \leq j \leq n} c_j$ . If all words in  $w[\mathcal{A}_1]_{s_1}^{\text{P}}$  are losing then (13) is true. If there is a winning word in  $w[\mathcal{A}_1]_{s_1}^{\text{P}}$ , it means that  $n_w$  is even and  $n_w > o_1 \geq o_2$ . Hence,  $w\text{Pc}_{\mathcal{A}_2, s_2}^{\sigma_1^2}$  is also winning and (13) also holds.

In both cases, we have  $w[\mathcal{A}_1]_{s_1}^{\text{P}} \sqsubseteq w\text{Pc}_{\mathcal{A}_2, s_2}^{\sigma_1^2} \sqsubseteq w[\mathcal{A}_2]_{s_2}^{\text{P}}$ . This proves P-D- $\mathcal{M}_{\text{triv}}$ -monotony.

We now turn to P-D- $\mathcal{M}_{\text{triv}}$ -selectivity. Let  $(\mathcal{A}_1, s_1), (\mathcal{A}_2, s_2) \in \mathfrak{A}_{\mathcal{P}_1}^{\text{D}}$ . Note that the requirement that  $\widehat{\text{col}}(\text{Hists}(\mathcal{A}_i, s_i, s_i)) \subseteq L_{m_{\text{init}}, m_{\text{init}}}$  does not bring information as with this particular memory skeleton, all words are in  $L_{m_{\text{init}}, m_{\text{init}}}$ . Let  $w = c_1 c_2 \dots c_n \in C^*$  and  $n_w = \max_{1 \leq j \leq n} c_j$ . We prove that

$$w[(\mathcal{A}_1, s_1) \sqcup (\mathcal{A}_2, s_2)]_t^{\text{P}} \sqsubseteq w[\mathcal{A}_1]_{s_1}^{\text{P}} \cup w[\mathcal{A}_2]_{s_2}^{\text{P}}. \quad (14)$$

If all words of  $w[(\mathcal{A}_1, s_1) \sqcup (\mathcal{A}_2, s_2)]_t^{\text{P}}$  are losing, then (14) is true. We now assume that there is a winning word  $ww' \in w[(\mathcal{A}_1, s_1) \sqcup (\mathcal{A}_2, s_2)]_t^{\text{P}}$ ; we show that we can also find a winning word in  $w[\mathcal{A}_1]_{s_1}^{\text{P}} \cup w[\mathcal{A}_2]_{s_2}^{\text{P}}$ .

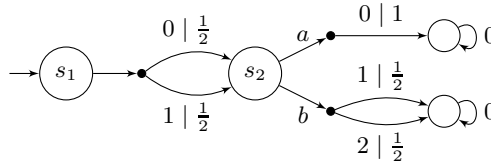
Assume word  $ww'$  sees its maximal color  $n$  in  $w'$ . If the play corresponding to  $w'$  in  $(\mathcal{A}_1, s_1) \sqcup (\mathcal{A}_2, s_2)$  comes back to  $t$  after seeing  $n$  for the first time, then  $n$  is the greatest color on some cycle on  $t$ . This means that the strategy repeatedly playing this cycle only takes actions in  $\mathcal{A}_1$  or in  $\mathcal{A}_2$  and also wins. If the play corresponding to  $w'$  does not come back to  $t$  after seeing  $n$  for the first time, then there is a suffix of the play fully in  $\mathcal{A}_1$  or in  $\mathcal{A}_2$  — this suffix can be played after the first visit to  $t$ , and this generates a winning play.

Now, assume  $ww'$  sees its maximal color  $n$  in  $w$ . If the play corresponding to  $w'$  comes back to  $t$ , the strategy repeatedly playing this cycle on  $t$  is winning, as no color greater than  $n$  is seen in this cycle. If there is no cycle on  $t$ , it means that  $w'$  is already an infinite word in  $[\mathcal{A}_1]_{s_1}^{\text{P}} \cup w[\mathcal{A}_2]_{s_2}^{\text{P}}$ .

We have therefore shown (14) in every case; this shows P-D- $\mathcal{M}_{\text{triv}}$ -selectivity.

We have proven that  $W_{\text{wp}}$  is P-D- $\mathcal{M}_{\text{triv}}$ -monotone and P-D- $\mathcal{M}_{\text{triv}}$ -selective; by Theorem 32, this implies that pure memoryless strategies are sufficient to play P-optimally in one-player deterministic arenas of  $\mathcal{P}_1$ . The same arguments holds from the point of view of  $\mathcal{P}_2$ . As we have shown that both players' one-player arenas admit pure memoryless P-optimal strategies, by Theorem 23, we conclude that both players have pure memoryless P-optimal (even P-SP) strategies in all two-player deterministic arenas.

*Stochastic games.* Interestingly, memory requirements of  $W_{\text{wp}}$  are larger in stochastic games but pure AIFM strategies still suffice and we can therefore still apply our results. An example of a one-player stochastic arena that requires memory is provided in Figure 3. Intuitively, in this case, memory is necessary for correct risk-assessment: it may sometimes be needed to attempt to get a greater color with a smaller probability, and that depends on the current maximal color. In this example, keeping in memory the greatest color seen is sufficient to play optimally.



**Fig. 3.** Initialized one-player stochastic arena that requires memory for winning condition  $W_{\text{wp}}$ .  $\mathcal{P}_1$  can win with probability  $\frac{3}{4}$  by taking the risk to play action  $b$  only if 1 has been seen before. Notation “ $c \mid p$ ” next to a transition means that color  $c$  is seen, and that this transition is taken with probability  $p$ .

We generalize this idea and prove that memory skeleton  $\mathcal{M}_{\text{max}} = (\mathbb{N}, 0, (m, n) \mapsto \max\{m, n\})$  suffices to play optimally in all stochastic arenas for both players (as argued earlier, although this skeleton is infinite, it is finite as soon as we restrict it to a finite set of colors).

We prove that the class  $\mathfrak{A}_{\mathcal{P}_1}^{\text{S}}$  of all initialized one-player stochastic arenas of  $\mathcal{P}_1$  admits pure memoryless G-optimal strategies based on  $\mathcal{M}_{\text{max}}$  by proving G-S- $\mathcal{M}_{\text{max}}$ -monotony and G-S- $\mathcal{M}_{\text{max}}$ -selectivity.

The weak parity winning condition is not prefix-independent but using the definition of  $\mathcal{M}_{\text{max}}$ , we prove the following related property: for all  $m \in \mathbb{N}$ , for all finite words  $w_1, w_2 \in L_{0, m}$ , for all infinite words  $w \in C^\omega$ ,

$$w_1 w \in W_{\text{wp}} \iff w_2 w \in W_{\text{wp}}. \quad (15)$$

That is, similar prefixes (in the sense that they reach the same state of the memory skeleton) have the same influence on the outcome; the winning continuations are the same. Let  $m \in \mathbb{N}$ ,  $w_1, w_2 \in L_{0,m}$ , and  $w = c_1 c_2 \dots \in C^\omega$ . Assume  $w_1 w \in W_{\text{wp}}$ . Let  $n_w = \max_{j \geq 1} c_j$ . If  $m \geq n_w$ , then it means  $m$  is even and  $w_2 w$  is therefore also winning. If  $m < n_w$ , then it means that  $n_w$  is even and  $w_2 w$  is also in  $W_{\text{wp}}$ .

Property (15) implies the following for distributions: for  $\mu \in \text{Dist}(C^\omega, \mathcal{F})$ , for all  $w_1, w_2 \in L_{0,m}$ ,  $w_1 \mu(W_{\text{wp}}) = w_2 \mu(W_{\text{wp}})$ . This implies G-S- $\mathcal{M}_{\text{max}}$ -monotony: let  $m \in \mathbb{N}$  and  $(\mathcal{A}_1, s_1), (\mathcal{A}_2, s_2) \in \mathfrak{A}_{\mathcal{P}_1}^S$ ; assume that for some  $w \in L_{0,m}$ , we have w.l.o.g.  $w[\mathcal{A}_1]_{s_1}^G \subseteq w[\mathcal{A}_2]_{s_2}^G$ . Then we automatically have that for all  $w' \in L_{0,m}$ , we have  $w'[\mathcal{A}_1]_{s_1}^G \subseteq w'[\mathcal{A}_2]_{s_2}^G$ , which proves G-S- $\mathcal{M}_{\text{max}}$ -monotony.

We now turn to G-S- $\mathcal{M}_{\text{max}}$ -selectivity. Let  $m \in \mathbb{N}$  and  $(\mathcal{A}_1, s_1), (\mathcal{A}_2, s_2) \in \mathfrak{A}_{\mathcal{P}_1}^S$  such that for  $i \in \{1, 2\}$ ,  $\widehat{\text{col}}(\text{Hists}(\mathcal{A}_i, s_i, s_i)) \subseteq L_{m,m}$ . Let  $w \in L_{0,m}$ . Let  $\mathcal{A}$  be the arena  $((\mathcal{A}_1, s_1) \sqcup (\mathcal{A}_2, s_2))_{w \rightsquigarrow t}$  with merged state  $t$ .

Thanks to the structure of the memory skeleton, we can make the following key observation: any play in  $\text{Plays}(\mathcal{A}, s_0^w)$  that visits  $t$  infinitely many times has a maximal color exactly  $m$ ; indeed,  $m$  is a color appearing in  $w$ , and if a color greater than  $m$  is seen, the memory state cannot go back down to  $m$ , so  $t$  cannot be visited again (it would contradict that every history from  $t$  to  $t$  is in  $L_{m,m}$ ).

Let  $\sigma_1 \in \Sigma_1^G(\mathcal{A}, t)$ . Our goal is to show that it is possible to do at least as well as  $w\text{Pc}_{\mathcal{A},t}^{\sigma_1}[W_{\text{wp}}]$  without the need to use actions both in  $\mathcal{A}_1$  and in  $\mathcal{A}_2$  at  $t$ .

We first assume that  $m$  is even: visiting  $t$  infinitely often is therefore winning for  $\mathcal{P}_1$ . If there is a strategy that, from  $t$ , comes back to  $t$  with probability 1, then  $\mathcal{P}_1$  can achieve the objective with probability 1 by repeatedly going back to  $t$ . The use of randomization at  $t$  is not necessary for this strategy: since it goes back to  $t$  with probability 1, every action it may play allows going back to  $t$  with probability 1. Thus, such a strategy does not need to use actions both in  $\mathcal{A}_1$  and in  $\mathcal{A}_2$ , as every time it leaves  $t$ , it can play the same action and repeat the strategy until it reaches  $t$  again. The G-S- $\mathcal{M}_{\text{max}}$ -selectivity is therefore satisfied, as  $w[\mathcal{A}_1]_{s_1}^G$  or  $w[\mathcal{A}_2]_{s_2}^G$  contains a strategy that wins with probability 1, which is at least as good as  $w\text{Pc}_{\mathcal{A},t}^{\sigma_1}[W_{\text{wp}}]$ .

Assume now that  $m$  is odd or that there is no strategy that comes back to  $t$  with probability 1. In the latter case, the probability to go back to  $t$  from  $t$  has a probability less than  $1 - \varepsilon$  for some  $\varepsilon > 0$  for all strategies; therefore, visiting  $t$  infinitely often necessarily has probability 0 for all strategies. We condition  $w\text{Pc}_{\mathcal{A},t}^{\sigma_1}[W_{\text{wp}}]$  over which part of the arena the play ends in: either it visits  $t$  infinitely often (event  $\square \diamond t$ ), or it sticks to  $\mathcal{A}_1$  or  $\mathcal{A}_2$  without visiting  $t$  from some point on (events  $\diamond \square \mathcal{A}_1 \setminus t$  and  $\diamond \square \mathcal{A}_2 \setminus t$ ).

We have

$$w\text{Pc}_{\mathcal{A},t}^{\sigma_1}[W_{\text{wp}}] = \text{P}_{\mathcal{A},t}^{\sigma_1}[\square \diamond t] \cdot w\text{Pc}_{\mathcal{A},t}^{\sigma_1}[W_{\text{wp}} \mid \square \diamond t] + \sum_{i \in \{1,2\}} \text{P}_{\mathcal{A},t}^{\sigma_1}[\diamond \square \mathcal{A}_i \setminus t] \cdot w\text{Pc}_{\mathcal{A},t}^{\sigma_1}[W_{\text{wp}} \mid \diamond \square \mathcal{A}_i \setminus t].$$

If  $m$  is odd, all infinite plays in  $\square \diamond t$  are losing; if all strategies visit  $t$  infinitely often with probability 0, then  $\text{P}_{\mathcal{A},t}^{\sigma_1}[\square \diamond t] = 0$ : in any case, the first term is 0.

We focus on the last two terms. For  $i \in \{1, 2\}$ , if the play stays in  $\mathcal{A}_i \setminus t$  from some point onward, as the value is independent from the actual prefix before the last visit to  $t$  (by property (15)), it means that it is possible to reach the same value while never going to  $\mathcal{A}_{3-i}$ . That is, there exists a strategy  $\sigma_1^i \in \Sigma_1^G(\mathcal{A}_i, t)$  such that

$$w\text{Pc}_{\mathcal{A},t}^{\sigma_1}[W_{\text{wp}} \mid \diamond \square \mathcal{A}_i \setminus t] \leq w\text{Pc}_{\mathcal{A}_i,t}^{\sigma_1^i}[W_{\text{wp}}].$$

We do not prove it formally; a very similar argument can be found in the proof of [Gim07, Theorem 4]: intuitively, it builds a strategy  $\sigma_1^i$  that induces a distribution on the projection of the plays of  $(\mathcal{A}, t)$  obtained by  $\sigma_1$  to plays of  $(\mathcal{A}_i, s_i)$  (by removing the cycles on  $t$  in  $(\mathcal{A}_{3-i}, s_{3-i})$ ). Thus, if we play the best strategy among  $\sigma_1^1$ , which obtains a value at least as good as the part that ends in  $\mathcal{A}_1 \setminus t$ , and  $\sigma_1^2$ , which obtains a value at least as good as the part that ends in  $\mathcal{A}_2 \setminus t$ , what we obtain is something at least as good as the value obtained by  $\sigma_1$ , without needing to consider actions both in  $\mathcal{A}_1$  and in  $\mathcal{A}_2$ .

We have proven that  $W_{\text{wp}}$  is G-S- $\mathcal{M}_{\text{max}}$ -monotone and G-S- $\mathcal{M}_{\text{max}}$ -selective; by Theorem 32, this implies that pure strategies based on  $\mathcal{M}_{\text{max}}$  are sufficient to play G-optimally in one-player stochastic arenas of  $\mathcal{P}_1$ . The same arguments with the same memory skeleton holds from the point of view of  $\mathcal{P}_2$ .

As we have shown that both players' one-player arenas admit pure G-optimal strategies based on  $\mathcal{M}_{\max}$ , by Theorem 23, we conclude that both players have pure G-optimal (even G-SP) strategies based on  $\mathcal{M}_{\max} \otimes \mathcal{M}_{\max}$  (which corresponds to  $\mathcal{M}_{\max}$ ) in all two-player stochastic arenas.

## 7 Conclusion

We have studied stochastic games and gave an overview of desirable properties of preference relations that admit *pure arena-independent finite-memory optimal strategies*. Our analysis provides general tools to help study memory requirements in stochastic games, both with one player (Markov decision processes) and two players, and links both problems. It generalizes both work on deterministic games [GZ05,BLO<sup>+</sup>20] and work on stochastic games [GZ09].

A natural question that remains unsolved is the link between memory requirements of a preference relation in deterministic and in stochastic games; our results can be called independently to study both problems, but do not describe a bridge to go from one to the other yet.

## References

- AR17. Benjamin Aminof and Sasha Rubin. First-cycle games. *Inf. Comput.*, 254:195–216, 2017.
- BBE10. 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, December 15-18, 2010, Chennai, India*, volume 8 of *LIPIcs*, pages 108–119. Schloss Dagstuhl - Leibniz-Zentrum für Informatik, 2010.
- BDOR20. Thomas Brihaye, Florent Delgrange, Youssef Oualhadj, and Mickael Randour. Life is random, time is not: Markov decision processes with window objectives. *Log. Methods Comput. Sci.*, 16(4), 2020.
- BFMM11. Alessandro Bianco, Marco Faella, Fabio Mogavero, and Aniello Murano. Exploring the boundary of half-positionality. *Ann. Math. Artif. Intell.*, 62(1-2):55–77, 2011.
- BFR17. Véronique Bruyère, Emmanuel Filiot, Mickael Randour, and Jean-François Raskin. Meet your expectations with guarantees: Beyond worst-case synthesis in quantitative games. *Inf. Comput.*, 254:259–295, 2017.
- BHM<sup>+</sup>17. Patricia Bouyer, Piotr Hofman, Nicolas Markey, Mickael Randour, and Martin Zimmermann. Bounding average-energy games. In Javier Esparza and Andrzej S. Murawski, editors, *Foundations of Software Science and Computation Structures - 20th International Conference, FOSSACS 2017, Held as Part of the European Joint Conferences on Theory and Practice of Software, ETAPS 2017, Uppsala, Sweden, April 22-29, 2017, Proceedings*, volume 10203 of *Lecture Notes in Computer Science*, pages 179–195, 2017.
- BHRR19. Véronique Bruyère, Quentin Hautem, Mickael Randour, and Jean-François Raskin. Energy mean-payoff games. In Wan Fokkink and Rob van Glabbeek, editors, *30th International Conference on Concurrency Theory, CONCUR 2019, August 27-30, 2019, Amsterdam, the Netherlands*, volume 140 of *LIPIcs*, pages 21:1–21:17. Schloss Dagstuhl - Leibniz-Zentrum für Informatik, 2019.
- BK08. Christel Baier and Joost-Pieter Katoen. *Principles of model checking*. MIT Press, 2008.
- BLO<sup>+</sup>20. Patricia Bouyer, Stéphane Le Roux, Youssef Oualhadj, Mickael Randour, and Pierre Vandenheve. Games where you can play optimally with arena-independent finite memory. In Konnov and Kovács [KK20], pages 24:1–24:22.
- BMR<sup>+</sup>18. Patricia Bouyer, Nicolas Markey, Mickael Randour, Kim G. Larsen, and Simon Laursen. Average-energy games. *Acta Informatica*, 55(2):91–127, 2018.
- BRR17. Raphaël Berthon, Mickael Randour, and Jean-François Raskin. Threshold constraints with guarantees for parity objectives in Markov decision processes. In Ioannis Chatzigiannakis, Piotr Indyk, Fabian Kuhn, and Anca Muscholl, editors, *44th International Colloquium on Automata, Languages, and Programming, ICALP 2017, July 10-14, 2017, Warsaw, Poland*, volume 80 of *LIPIcs*, pages 121:1–121:15. Schloss Dagstuhl - Leibniz-Zentrum für Informatik, 2017.
- CD12. Krishnendu Chatterjee and Laurent Doyen. Energy parity games. *Theor. Comput. Sci.*, 458:49–60, 2012.
- CD16. Krishnendu Chatterjee and Laurent Doyen. Perfect-information stochastic games with generalized mean-payoff objectives. In Martin Grohe, Eric Koskinen, and Natarajan Shankar, editors, *Proceedings of the 31st Annual ACM/IEEE Symposium on Logic in Computer Science, LICS '16, New York, NY, USA, July 5-8, 2016*, pages 247–256. ACM, 2016.

- CdAH04. Krishnendu Chatterjee, Luca de Alfaro, and Thomas A. Henzinger. Trading memory for randomness. In *1st International Conference on Quantitative Evaluation of Systems (QEST 2004), 27-30 September 2004, Enschede, The Netherlands*, pages 206–217. IEEE Computer Society, 2004.
- CDGH10. Krishnendu Chatterjee, Laurent Doyen, Hugo Gimbert, and Thomas A. Henzinger. Randomness for free. In Petr Hlinený and Antonín Kucera, editors, *Mathematical Foundations of Computer Science 2010, 35th International Symposium, MFCS 2010, Brno, Czech Republic, August 23-27, 2010. Proceedings*, volume 6281 of *Lecture Notes in Computer Science*, pages 246–257. Springer, 2010.
- CFK<sup>+</sup>12. Taolue Chen, Vojtech Forejt, Marta Z. Kwiatkowska, Aistis Simaitis, Ashutosh Trivedi, and Michael Ummels. Playing stochastic games precisely. In Maciej Koutny and Irek Ulidowski, editors, *CONCUR 2012 - Concurrency Theory - 23rd International Conference, CONCUR 2012, Newcastle upon Tyne, UK, September 4-7, 2012. Proceedings*, volume 7454 of *Lecture Notes in Computer Science*, pages 348–363. Springer, 2012.
- CFK<sup>+</sup>13. Taolue Chen, Vojtech Forejt, Marta Z. Kwiatkowska, Aistis Simaitis, and Clemens Wiltsche. On stochastic games with multiple objectives. In Krishnendu Chatterjee and Jiri Sgall, editors, *Mathematical Foundations of Computer Science 2013 - 38th International Symposium, MFCS 2013, Klosterneuburg, Austria, August 26-30, 2013. Proceedings*, volume 8087 of *Lecture Notes in Computer Science*, pages 266–277. Springer, 2013.
- CH12. Krishnendu Chatterjee and Thomas A. Henzinger. A survey of stochastic  $\omega$ -regular games. *J. Comput. Syst. Sci.*, 78(2):394–413, 2012.
- CJH04. Krishnendu Chatterjee, Marcin Jurdzinski, and Thomas A. Henzinger. Quantitative stochastic parity games. In J. Ian Munro, editor, *Proceedings of the Fifteenth Annual ACM-SIAM Symposium on Discrete Algorithms, SODA 2004, New Orleans, Louisiana, USA, January 11-14, 2004*, pages 121–130. SIAM, 2004.
- CKK17. Krishnendu Chatterjee, Zuzana Kretínská, and Jan Kretínský. Unifying two views on multiple mean-payoff objectives in Markov decision processes. *Log. Methods Comput. Sci.*, 13(2), 2017.
- CKWW20. Krishnendu Chatterjee, Joost-Pieter Katoen, Maximilian Weininger, and Tobias Winkler. Stochastic games with lexicographic reachability-safety objectives. In Shuvendu K. Lahiri and Chao Wang, editors, *Computer Aided Verification - 32nd International Conference, CAV 2020, Los Angeles, CA, USA, July 21-24, 2020, Proceedings, Part II*, volume 12225 of *Lecture Notes in Computer Science*, pages 398–420. Springer, 2020.
- Con92. Anne Condon. The complexity of stochastic games. *Inf. Comput.*, 96(2):203–224, 1992.
- CP19. Krishnendu Chatterjee and Nir Piterman. Combinations of qualitative winning for stochastic parity games. In Wan J. Fokkink and Rob van Glabbeek, editors, *30th International Conference on Concurrency Theory, CONCUR 2019, August 27-30, 2019, Amsterdam, the Netherlands*, volume 140 of *LIPICs*, pages 6:1–6:17. Schloss Dagstuhl - Leibniz-Zentrum für Informatik, 2019.
- CRR14. Krishnendu Chatterjee, Mickael Randour, and Jean-François Raskin. Strategy synthesis for multi-dimensional quantitative objectives. *Acta Inf.*, 51(3-4):129–163, 2014.
- DKQR20. Florent Delgrange, Joost-Pieter Katoen, Tim Quatmann, and Mickael Randour. Simple strategies in multi-objective MDPs. In Armin Biere and David Parker, editors, *Tools and Algorithms for the Construction and Analysis of Systems - 26th International Conference, TACAS 2020, Held as Part of the European Joint Conferences on Theory and Practice of Software, ETAPS 2020, Dublin, Ireland, April 25-30, 2020, Proceedings, Part I*, volume 12078 of *Lecture Notes in Computer Science*, pages 346–364. Springer, 2020.
- Dur19. Rick Durrett. *Probability: Theory and Examples*. Cambridge Series in Statistical and Probabilistic Mathematics. Cambridge University Press, 5th edition, 2019.
- Gim07. Hugo Gimbert. Pure stationary optimal strategies in Markov decision processes. In Wolfgang Thomas and Pascal Weil, editors, *STACS 2007, 24th Annual Symposium on Theoretical Aspects of Computer Science, Aachen, Germany, February 22-24, 2007, Proceedings*, volume 4393 of *Lecture Notes in Computer Science*, pages 200–211. Springer, 2007.
- GK14. Hugo Gimbert and Edon Kelmendi. Two-player perfect-information shift-invariant submixing stochastic games are half-positional. *CoRR*, abs/1401.6575, 2014.
- GZ05. Hugo Gimbert and Wiesław Zielonka. Games where you can play optimally without any memory. In Martín Abadi and Luca de Alfaro, editors, *CONCUR 2005 - Concurrency Theory, 16th International Conference, CONCUR 2005, San Francisco, CA, USA, August 23-26, 2005, Proceedings*, volume 3653 of *Lecture Notes in Computer Science*, pages 428–442. Springer, 2005.
- GZ09. Hugo Gimbert and Wiesław Zielonka. Pure and Stationary Optimal Strategies in Perfect-Information Stochastic Games with Global Preferences. Unpublished, 2009.

- KK20. Igor Konnov and Laura Kovács, editors. *31st International Conference on Concurrency Theory, CONCUR 2020, September 1-4, 2020, Vienna, Austria (Virtual Conference)*, volume 171 of *LIPIcs*. Schloss Dagstuhl - Leibniz-Zentrum für Informatik, 2020.
- Kop06. Eryk Kopczyński. Half-positional determinacy of infinite games. In Michele Bugliesi, Bart Preneel, Vladimiro Sassone, and Ingo Wegener, editors, *Automata, Languages and Programming, 33rd International Colloquium, ICALP 2006, Venice, Italy, July 10-14, 2006, Proceedings, Part II*, volume 4052 of *Lecture Notes in Computer Science*, pages 336–347. Springer, 2006.
- Kop08. Eryk Kopczyński. *Half-positional Determinacy of Infinite Games*. PhD thesis, Warsaw University, 2008.
- LP18. Stéphane Le Roux and Arno Pauly. Extending finite-memory determinacy to multi-player games. *Inf. Comput.*, 261(Part):676–694, 2018.
- LPR18. Stéphane Le Roux, Arno Pauly, and Mickael Randour. Extending finite-memory determinacy by boolean combination of winning conditions. In Sumit Ganguly and Paritosh K. Pandya, editors, *38th IARCS Annual Conference on Foundations of Software Technology and Theoretical Computer Science, FSTTCS 2018, December 11-13, 2018, Ahmedabad, India*, volume 122 of *LIPIcs*, pages 38:1–38:20. Schloss Dagstuhl - Leibniz-Zentrum für Informatik, 2018.
- Mar75. Donald A. Martin. Borel determinacy. *Annals of Mathematics*, pages 363–371, 1975.
- MPR20. Benjamin Monmege, Julie Parreaux, and Pierre-Alain Reynier. Reaching your goal optimally by playing at random with no memory. In Konnov and Kovács [KK20], pages 26:1–26:21.
- MSTW17. Richard Mayr, Sven Schewe, Patrick Totzke, and Dominik Wojtczak. MDPs with energy-parity objectives. In *32nd Annual ACM/IEEE Symposium on Logic in Computer Science, LICS 2017, Reykjavik, Iceland, June 20-23, 2017*, pages 1–12. IEEE Computer Society, 2017.
- MSTW21. Richard Mayr, Sven Schewe, Patrick Totzke, and Dominik Wojtczak. Simple stochastic games with almost-sure energy-parity objectives are in NP and coNP. *CoRR*, abs/2101.06989, 2021.
- Put94. Martin L. Puterman. *Markov Decision Processes: Discrete Stochastic Dynamic Programming*. Wiley Series in Probability and Statistics. Wiley, 1994.
- Ran13. Mickael Randour. Automated synthesis of reliable and efficient systems through game theory: A case study. In *Proc. of ECCS 2012*, Springer Proceedings in Complexity XVII, pages 731–738. Springer, 2013.
- RRS15. Mickael Randour, Jean-François Raskin, and Ocan Sankur. Variations on the stochastic shortest path problem. In Deepak D’Souza, Akash Lal, and Kim Guldstrand Larsen, editors, *Verification, Model Checking, and Abstract Interpretation - 16th International Conference, VMCAI 2015, Mumbai, India, January 12-14, 2015. Proceedings*, volume 8931 of *Lecture Notes in Computer Science*, pages 1–18. Springer, 2015.
- RRS17. Mickael Randour, Jean-François Raskin, and Ocan Sankur. Percentile queries in multi-dimensional Markov decision processes. *Formal Methods Syst. Des.*, 50(2-3):207–248, 2017.
- Sha53. L. S. Shapley. Stochastic games. *Proceedings of the National Academy of Sciences*, 39(10):1095–1100, 1953.
- Tho08. Wolfgang Thomas. Church’s problem and a tour through automata theory. In Arnon Avron, Nachum Dershowitz, and Alexander Rabinovich, editors, *Pillars of Computer Science, Essays Dedicated to Boris (Boaz) Trakhtenbrot on the Occasion of His 85th Birthday*, volume 4800 of *Lecture Notes in Computer Science*, pages 635–655. Springer, 2008.
- VCD<sup>+</sup>15. Yaron Velner, Krishnendu Chatterjee, Laurent Doyen, Thomas A. Henzinger, Alexander Moshe Rabinovich, and Jean-François Raskin. The complexity of multi-mean-payoff and multi-energy games. *Inf. Comput.*, 241:177–196, 2015.

## A Results on splits

We recall here technical results about *split arenas* (Definition 26) that are already present in [GZ09] with the slight difference that we consider *initialized* arenas.

Let  $(\mathcal{A} = (S_1, S_2, A, \delta, \text{col}), S_{\text{init}})$  be an initialized arena,  $t \in S_1$  be a state controlled by  $\mathcal{P}_1$ , and  $(\text{Split}(\mathcal{A}, t), S_{\text{init}}^t)$  be the split of  $(\mathcal{A}, S_{\text{init}})$  on  $t$ .

For all  $a \in A(t)$ , we build a natural bijection between plays of the arena and plays of its split: it is a function

$$\phi^a : \text{Hists}(\mathcal{A}, S_{\text{init}}) \rightarrow \text{Hists}(\text{Split}(\mathcal{A}, t), S_{\text{init}}^a).$$



This function simply labels the different states appearing along the play with the right action to make it a play of the split, starting arbitrarily with  $a$ : if  $t$  has never been visited, it picks action  $a$  by default; if  $t$  has been visited, it picks the last action played in  $t$ . Formally, let  $\rho = s_0 a_1 s_1 \dots a_n s_n \in \mathbf{Hists}(\mathcal{A}, S_{\text{init}})$ . We define  $\phi^a(\rho) = s_0^{a'_0} a_1 s_1^{a'_1} \dots a_n s_n^{a'_n}$  where we assume as usual that  $t^{a'_i} = t$ , and for  $i$ , with  $0 \leq i \leq n$  such that  $s_i \neq t$ ,

$$a'_i = \begin{cases} a & \text{if for all } j < i, s_j \neq t, \\ a_{k+1} & \text{if } k \text{ is the index of the visit to } t \text{ preceding } s_i \text{ in } \rho. \end{cases}$$

The history  $\phi^a(\rho)$  is a history of the split by construction. Function  $\phi^a$  has an inverse  $(\phi^a)^{-1}$  which associates to any history of the split starting in  $S_{\text{init}}^a$  the same history in which all the action labels have been removed. We can extend function  $\phi^a$  to a bijection on plays: let

$$\phi_\infty^a : \text{Plays}(\mathcal{A}, S_{\text{init}}) \rightarrow \text{Plays}(\text{Split}(\mathcal{A}, t), S_{\text{init}}^a)$$

be the function such that for  $\pi = s_0 a_1 s_1 a_2 s_2 \dots \in \text{Plays}(\mathcal{A}, S_{\text{init}})$ , if  $\rho_n = s_0 a_1 s_1 \dots a_n s_n$  is a prefix of  $\pi$ , then  $\phi_\infty^a(\pi) = \lim_{n \rightarrow \infty} \phi^a(\rho_n)$ .

For  $i \in \{1, 2\}$ , for all  $a \in A(t)$ , we build a natural bijection between strategies of the arena and strategies of its split: it is a function

$$\Phi_i^a : \Sigma_i^G(\mathcal{A}, S_{\text{init}}) \rightarrow \Sigma_i^G(\text{Split}(\mathcal{A}, t), S_{\text{init}}^a).$$

For  $\sigma_i \in \Sigma_i^G(\mathcal{A}, S_{\text{init}})$  a strategy of  $\mathcal{P}_i$ , we define  $\Phi_i^a(\sigma_i) = \sigma_i \circ (\phi^a)^{-1}$ . Similarly, this function has an inverse: for  $\sigma_i \in \Sigma_i^G(\text{Split}(\mathcal{A}, t), S_{\text{init}}^a)$ , we define  $(\Phi_i^a)^{-1}(\sigma_i) = \sigma_i \circ \phi^a$ .

We prove a few results about these bijections, which correspond to [GZ09, Proposition 10 and Lemma 12].

**Lemma 35.** *Let  $a \in A(t)$ ,  $s \in S_{\text{init}}$ ,  $\sigma_1 \in \Sigma_1^G(\mathcal{A}, S_{\text{init}})$ , and  $\sigma_2 \in \Sigma_2^G(\mathcal{A}, S_{\text{init}})$ . We have*

$$\text{Pc}_{\mathcal{A}, s}^{\sigma_1, \sigma_2} = \text{Pc}_{\text{Split}(\mathcal{A}, t), s^a}^{\Phi_1^a(\sigma_1), \Phi_2^a(\sigma_2)}.$$

*Let  $\sqsubseteq$  a preference relation and  $\mathbf{X} \in \{\text{PFM}, \text{P}, \text{GFM}, \text{G}\}$  be a type of strategies. For  $i \in \{1, 2\}$ , strategy  $\sigma_i$  is pure (resp. finite-memory) if and only if strategy  $\Phi_i^a(\sigma_i)$  is pure (resp. finite-memory). For  $i \in \{1, 2\}$ ,  $\sigma_i$  is  $\mathbf{X}$ -optimal in  $(\mathcal{A}, S_{\text{init}}, \sqsubseteq)$  if and only if  $\Phi_i^a(\sigma_i)$  is  $\mathbf{X}$ -optimal in  $(\text{Split}(\mathcal{A}, t), S_{\text{init}}^a, \sqsubseteq)$ . Moreover,  $(\sigma_1, \sigma_2)$  is an  $\mathbf{X}$ -NE in  $(\mathcal{A}, S_{\text{init}}, \sqsubseteq)$  if and only if  $(\Phi_1^a(\sigma_1), \Phi_2^a(\sigma_2))$  is an  $\mathbf{X}$ -NE in  $(\text{Split}(\mathcal{A}, t), S_{\text{init}}^a, \sqsubseteq)$ .*

*Proof.* We prove the first claim by showing the equality of two distributions in  $\text{Dist}(\text{Plays}(\mathcal{A}, s), \mathcal{F}_{(\mathcal{A}, s)})$ :

$$\text{P}_{\mathcal{A}, s}^{\sigma_1, \sigma_2}[\cdot] = \text{P}_{\text{Split}(\mathcal{A}, t), s^a}^{\Phi_1^a(\sigma_1), \Phi_2^a(\sigma_2)}[\phi_\infty^a(\cdot)]. \quad (16)$$

We prove the equality for cylinders  $\text{Cyl}(\rho)$  with  $\rho \in \mathbf{Hists}(\mathcal{A}, S_{\text{init}})$ . Notice that  $\phi_\infty^a(\text{Cyl}(\rho)) = \text{Cyl}(\phi^a(\rho))$ . Thanks to our construction of functions  $\Phi_1^a$  and  $\Phi_2^a$ , an easy induction on the length of  $\rho$  shows that

$$\text{P}_{\mathcal{A}, s}^{\sigma_1, \sigma_2}[\text{Cyl}(\rho)] = \text{P}_{\text{Split}(\mathcal{A}, t), s^a}^{\Phi_1^a(\sigma_1), \Phi_2^a(\sigma_2)}[\text{Cyl}(\phi^a(\rho))].$$

Since cylinders generate the  $\sigma$ -algebra, this proves (16).

Now notice that the bijection on plays  $\phi_\infty^a$  preserves the sequence of colors seen, i.e., for all  $\pi \in \text{Plays}(\mathcal{A}, S_{\text{init}})$ ,  $\widehat{\text{col}}(\pi) = \widehat{\text{col}}(\phi_\infty^a(\pi))$ . Using the definition of  $\text{Pc}$ , we can therefore conclude that

$$\text{Pc}_{\mathcal{A}, s}^{\sigma_1, \sigma_2} = \text{Pc}_{\text{Split}(\mathcal{A}, t), s^a}^{\Phi_1^a(\sigma_1), \Phi_2^a(\sigma_2)}.$$

The claims about  $\mathbf{X}$ -optimality and  $\mathbf{X}$ -NE follow from the first one: as  $\Phi_i^a$  is a bijection that preserves the induced distributions on colors, it also preserves  $\mathbf{X}$ -optimality and  $\mathbf{X}$ -NE.

Bijection  $\Phi_i^a$  clearly preserves the “pure” feature of strategies by construction, in both directions. Now if  $\sigma_i$  is finite-memory, then  $\Phi_i^a(\sigma_i)$  does not need any more memory to play as it has access to the same information and the last action played in  $t$ . In the other direction, if  $\Phi_i^a(\sigma_i)$  is finite-memory, then  $\sigma_i$  can play with the same memory *plus* extra information about the last action that was played in  $t$ . Therefore,  $\sigma_i$  might need more memory than  $\Phi_i^a(\sigma_i)$ , but that memory stays finite.<sup>6</sup>  $\square$

**Lemma 36.** *Let  $\sqsubseteq$  be a preference relation and  $X \in \{\text{PFM}, \text{P}, \text{GFM}, \text{G}\}$  be a type of strategies. Let  $(\mathcal{A} = (S_1, S_2, A, \delta, \text{col}), S_{\text{init}})$  be an initialized arena, and  $(\text{Split}(\mathcal{A}, t), S_{\text{init}}^t)$  be its split on  $t$  for some  $t \in S_1$ . Assume  $(\sigma_1, \sigma_2)$  is an  $X$ -NE in  $(\text{Split}(\mathcal{A}, t), S_{\text{init}}^{a^*})$  for some  $a^* \in A(t)$ . If  $\sigma_1$  is pure memoryless, and  $\sigma_1(t) = a^*$ , then there exists an  $X$ -NE  $(\sigma'_1, \sigma'_2)$  in  $(\mathcal{A}, S_{\text{init}})$  such that  $\sigma'_1$  is pure memoryless.*

*Proof.* We set  $(\sigma'_1, \sigma'_2) = ((\Phi_1^{a^*})^{-1}(\sigma_1), (\Phi_2^{a^*})^{-1}(\sigma_2))$ , which is an  $X$ -NE in  $(\mathcal{A}, S_{\text{init}})$  by Lemma 35. Moreover, Lemma 35 shows that  $\sigma'_1$  is pure. It is left to prove that in this particular case,  $\sigma'_1$  is memoryless. Let  $\rho \in \text{Hists}_1(\mathcal{A}, S_{\text{init}})$  be a history consistent with  $\sigma'_1$ . We know that if  $\text{out}(\rho) = t$ , then  $\sigma'_1(\rho) = a^*$  because this is the only possible action played in  $t$  by  $\sigma_1$  in  $\text{Split}(\mathcal{A}, t)$ . Now assume  $\text{out}(\rho) \neq t$ . Notice that since any action taken at  $t$  is necessarily  $a^*$  (and since we start in  $S_{\text{init}}^{a^*}$ ), any state appearing along  $\phi^{a^*}(\rho)$  (except  $t$ ) is necessarily labeled by  $a^*$ . Therefore, we have  $\sigma'_1(\rho) = \sigma_1(\phi^{a^*}(\rho)) = \sigma_1(s^{a^*})$ , which only depends on  $s$ .  $\square$

**Lemma 37.** *Let  $\sqsubseteq$  be a preference relation. Let  $(\mathcal{A} = (S_1, S_2, A, \delta, \text{col}), S_{\text{init}})$  be an initialized arena, and  $(\text{Split}(\mathcal{A}, t), S_{\text{init}}^t)$  be its split on  $t$  for some  $t \in S_1$ . Let  $\sigma_1 \in \Sigma_1^{\text{PFM}}(\text{Split}(\mathcal{A}, t), S_{\text{init}}^t)$  and  $\sigma_2 \in \Sigma_2^{\text{G}}(\text{Split}(\mathcal{A}, t), S_{\text{init}}^t)$ . Assume  $\sigma_1$  is pure memoryless, and let  $a^* = \sigma_1(t)$ . Let  $(\mathcal{A}_{a^*}, S_{\text{init}}^{a^*})$  be the initialized subarena of  $(\mathcal{A}, S_{\text{init}})$  in which only action  $a^*$  is available in  $t$  and states are renamed  $s \mapsto s^{a^*}$ . Let  $\sigma_1^{a^*}$  and  $\sigma_2^{a^*}$  be the restrictions of  $\sigma_1$  and  $\sigma_2$  to  $\text{Hists}(\mathcal{A}_{a^*}, S_{\text{init}}^{a^*})$ , which are strategies on  $(\mathcal{A}_{a^*}, S_{\text{init}}^{a^*})$ . Then for all  $s^{a^*} \in S_{\text{init}}^{a^*}$ ,*

$$\text{Pc}_{\text{Split}(\mathcal{A}, t), s^{a^*}}^{\sigma_1, \sigma_2} = \text{Pc}_{\mathcal{A}_{a^*}, s^{a^*}}^{\sigma_1^{a^*}, \sigma_2^{a^*}}.$$

*Proof.* Let  $s^{a^*} \in S_{\text{init}}^{a^*}$ . Notice that any play in  $\text{Split}(\mathcal{A}, t)$  consistent with  $\sigma_1$  starting in  $s^{a^*}$  only visits states among  $S^{a^*}$ , since it starts there and every action played in  $t$  is  $a^*$ . This shows that  $\sigma_1^{a^*}$  and  $\sigma_2^{a^*}$  are indeed well-defined strategies on  $(\mathcal{A}_{a^*}, S_{\text{init}}^{a^*})$ . Moreover, it shows that

$$\text{Pc}_{\text{Split}(\mathcal{A}, t), s^{a^*}}^{\sigma_1, \sigma_2} = \text{Pc}_{\mathcal{A}_{a^*}, s^{a^*}}^{\sigma_1^{a^*}, \sigma_2^{a^*}}$$

since every infinite play stays in the subarena of  $\text{Split}(\mathcal{A}, t)$  corresponding to  $\mathcal{A}_{a^*}$ .  $\square$

## B Missing proofs of Section 5

We prove the two technical lemmas from Section 5 that were stated without proof.

**Lemma 33.** *Let  $(\mathcal{A} = (S_1, S_2, A, \delta, \text{col}), s)$  be an initialized one-player arena and  $\rho \in \text{Hists}(\mathcal{A}, s)$ . Let  $\mu$  be a distribution on plays in  $\text{Dist}(\text{Plays}(\mathcal{A}, \text{out}(\rho)), \mathcal{F}_{(\mathcal{A}, \text{out}(\rho))})$ . We have*

$$\widehat{\text{col}}(\rho\mu) = \widehat{\text{col}}(\rho)\widehat{\text{col}}(\mu).$$

<sup>6</sup> Formally, our memory model is based on colors, and not on actions. However, we can easily enrich the game graph with a new color for each action available at  $t$ , and a memory skeleton (which reads colors) can then remember the last action played at  $t$ .

*Proof.* Let  $E \in \mathcal{F}$  be an event about infinite sequences of colors. Then,

$$\begin{aligned}
\widehat{\text{col}}(\rho\mu)(E) &= \rho\mu(\widehat{\text{col}}^{-1}(E)) && \text{by definition of } \widehat{\text{col}} \text{ on distributions} \\
&= \mu(\{\pi \in \text{Plays}(\mathcal{A}, \text{out}(\rho)) \mid \rho\pi \in \widehat{\text{col}}^{-1}(E)\}) && \text{by definition of } \rho\mu \\
&= \mu(\{\pi \in \text{Plays}(\mathcal{A}, \text{out}(\rho)) \mid \widehat{\text{col}}(\rho\pi) \in E\}) \\
&= \mu(\{\pi \in \text{Plays}(\mathcal{A}, \text{out}(\rho)) \mid \widehat{\text{col}}(\rho)\widehat{\text{col}}(\pi) \in E\}) \\
&= (\mu \circ \widehat{\text{col}}^{-1})(\{w' \in C^\omega \mid \widehat{\text{col}}(\rho)w' \in E\}) \\
&= \widehat{\text{col}}(\mu)(\{w' \in C^\omega \mid \widehat{\text{col}}(\rho)w' \in E\}) && \text{by definition of } \widehat{\text{col}} \text{ on distributions} \\
&= (\widehat{\text{col}}(\rho)\widehat{\text{col}}(\mu))(E) && \text{by definition of } \widehat{\text{col}}(\rho)\widehat{\text{col}}(\mu).
\end{aligned}$$

□

**Lemma 34.** Let  $(\mathcal{A}, s) \in \mathfrak{A}_{\mathcal{P}_1}^\vee$  be an initialized one-player arena and  $\sigma_1, \tau_1 \in \Sigma_1^G(\mathcal{A}, s)$  be two strategies. Let  $\rho = s_0a_1s_1 \dots a_ns_n \in \text{Hists}(\mathcal{A}, s)$ . We say that  $\sigma_1$  coincides with  $\tau_1$  on  $\rho$  if for each prefix  $\rho_i = s_0a_1s_1 \dots a_is_i$  of  $\rho$  with  $0 \leq i < n$ ,  $\sigma_1(\rho_i) = \tau_1(\rho_i)$ . If  $\sigma_1$  coincides with  $\tau_1$  on  $\rho$ , then

$$\mathbf{P}_{\mathcal{A},s}^{\sigma_1}[\text{Cyl}(\rho)] = \mathbf{P}_{\mathcal{A},s}^{\tau_1}[\text{Cyl}(\rho)].$$

Let  $t$  be a state of  $\mathcal{A}$ . We write  $\neg\Diamond t$  for the event in  $\mathcal{F}_{(\mathcal{A},s)}$  that consists of all the infinite plays that never visit  $t$ . Assume that for all  $\rho = s_0a_1s_1 \dots a_ns_n \in \text{Hists}(\mathcal{A}, s)$  such that for all  $i$ ,  $0 \leq i < n$ ,  $s_i \neq t$ ,  $\sigma_1$  coincides with  $\tau_1$  on  $\rho$ . Then

$$\mathbf{P}_{\mathcal{A},s}^{\sigma_1}[\neg\Diamond t] = \mathbf{P}_{\mathcal{A},s}^{\tau_1}[\neg\Diamond t] \text{ and, if } \mathbf{P}_{\mathcal{A},s}^{\sigma_1}[\neg\Diamond t] > 0, \mathbf{P}_{\mathcal{A},s}^{\sigma_1}[\cdot \mid \neg\Diamond t] = \mathbf{P}_{\mathcal{A},s}^{\tau_1}[\cdot \mid \neg\Diamond t].$$

*Proof.* Assume  $\mathcal{A} = (S_1, S_2, A, \delta, \text{col})$ . Using the definition of  $\mathbf{P}_{\mathcal{A},s}[\text{Cyl}(\rho)]$  and the hypothesis, we have

$$\mathbf{P}_{\mathcal{A},s}^{\sigma_1}[\text{Cyl}(\rho)] = \prod_{i=1}^n \sigma_1(\rho_{i-1})(a_i) \cdot \delta(s_{i-1}, a_i, s_i) = \prod_{i=1}^n \tau_1(\rho_{i-1})(a_i) \cdot \delta(s_{i-1}, a_i, s_i) = \mathbf{P}_{\mathcal{A},s}^{\tau_1}[\text{Cyl}(\rho)]$$

which proves the first claim.

Let  $H = \{\text{Cyl}(\rho) \mid \rho \in \text{Hists}(\mathcal{A}, s) \text{ and } \sigma_1 \text{ coincides with } \tau_1 \text{ on } \rho\}$ . By the first claim,  $\mathbf{P}_{\mathcal{A},s}^{\sigma_1}$  and  $\mathbf{P}_{\mathcal{A},s}^{\tau_1}$  are equal on  $H$ . As this class is closed by intersection, by the monotone class lemma,  $\mathbf{P}_{\mathcal{A},s}^{\sigma_1}$  and  $\mathbf{P}_{\mathcal{A},s}^{\tau_1}$  are also equal on the smallest  $\sigma$ -algebra generated by  $H$ .

For  $E \in \mathcal{F}_{(\mathcal{A},s)}$  an event, we denote by  $E^c = \text{Plays}(\mathcal{A}, s) \setminus E$  its complement; notice that

$$\begin{aligned}
\neg\Diamond t &= \left( \bigcup_{n \in \mathbb{N}} \underbrace{\text{Cyl}(S \setminus \{t\}, A, \dots, S \setminus \{t\}, A, t)}_{n \text{ times } "S \setminus \{t\}, A"} \right)^c \\
&= \bigcap_{n \in \mathbb{N}} \underbrace{\text{Cyl}(S \setminus \{t\}, A, \dots, S \setminus \{t\}, A, t)^c}_{n \text{ times } "S \setminus \{t\}, A"}
\end{aligned}$$

where  $\text{Cyl}(S \setminus \{t\}, A, \dots, S \setminus \{t\}, A, t)$  refers to the union over all cylinders of histories that go through a certain number of states that are not  $t$ , and then end in  $t$ . Event  $\neg\Diamond t$  can therefore be expressed with complements and countable intersection of cylinders in  $H$ , so  $\mathbf{P}_{\mathcal{A},s}^{\sigma_1}[\neg\Diamond t] = \mathbf{P}_{\mathcal{A},s}^{\tau_1}[\neg\Diamond t]$ , which proves the second claim.

We now assume that  $\mathbf{P}_{\mathcal{A},s}^{\sigma_1}[\neg\Diamond t] > 0$ . Using the definition of conditional probabilities, to prove that  $\mathbf{P}_{\mathcal{A},s}^{\sigma_1}[\cdot \mid \neg\Diamond t] = \mathbf{P}_{\mathcal{A},s}^{\tau_1}[\cdot \mid \neg\Diamond t]$ , it is left to prove that for all events  $E \in \mathcal{F}_{(\mathcal{A},s)}$ ,  $\mathbf{P}_{\mathcal{A},s}^{\sigma_1}[E \cap \neg\Diamond t] = \mathbf{P}_{\mathcal{A},s}^{\tau_1}[E \cap \neg\Diamond t]$ . We first prove this equality if  $E$  is a cylinder  $\text{Cyl}(\rho)$  for some  $\rho \in \text{Hists}(\mathcal{A}, s)$ . If  $\rho$  visits  $t$ , then  $E \cap \neg\Diamond t = \emptyset$  and  $\mathbf{P}_{\mathcal{A},s}^{\sigma_1}[E \cap \neg\Diamond t] = \mathbf{P}_{\mathcal{A},s}^{\tau_1}[E \cap \neg\Diamond t] = 0$ . If  $\rho$  does not visit  $t$ , then  $E$  is an element of  $H$ . Therefore,  $E \cap \neg\Diamond t$  can be expressed with complements and countable intersection of elements of  $H$ , so  $\mathbf{P}_{\mathcal{A},s}^{\sigma_1}[E \cap \neg\Diamond t] = \mathbf{P}_{\mathcal{A},s}^{\tau_1}[E \cap \neg\Diamond t]$ . Since measures  $\mathbf{P}_{\mathcal{A},s}^{\sigma_1}[\cdot \cap \neg\Diamond t]$  and  $\mathbf{P}_{\mathcal{A},s}^{\tau_1}[\cdot \cap \neg\Diamond t]$  are equal on all cylinders, they are also equal on all the events in  $\mathcal{F}_{(\mathcal{A},s)}$ . □