Partial-Observation Stochastic Games: How to Win when Belief Fails

Krishnendu Chatterjee, Laurent Doyen

Partial-Observation Stochastic Games: How to Win when Belief Fails

Krishnendu Chatterjee
IST Austria

Laurent Doyen
LSV, ENS Cachan & CNRS, France

***Abstract*—We consider two-player stochastic games played on finite graphs with reachability objectives where the first player tries to ensure a target state to be visited almost-surely (i.e., with probability 1), or positively (i.e., with positive probability), no matter the strategy of the second player.**

We classify such games according to the information and the power of randomization available to the players. On the basis of information, the game can be one-sided with either (a) player 1, or (b) player 2 having partial observation (and the other player has perfect observation), or two-sided with (c) both players having partial observation. On the basis of randomization, the players (a) may not be allowed to use randomization (pure strategies), or (b) may choose a probability distribution over actions but the actual random choice is external and not visible to the player (actions invisible), or (c) may use full randomization.

Our main results for pure strategies are as follows. (1) For one-sided games with player 1 having partial observation we show that (in contrast to full randomized strategies) *belief-based* (subset-construction based) strategies are not sufficient, and we present an exponential upper bound on memory both for almost-sure and positive winning strategies; we show that the problem of deciding the existence of almost-sure and positive winning strategies for player 1 is EXPTIME-complete. (2) For one-sided games with player 2 having partial observation we show that non-elementary memory is both necessary and sufficient for both almost-sure and positive winning strategies. (3) We show that for the general (two-sided) case finite-memory strategies are sufficient for both positive and almost-sure winning, and at least non-elementary memory is required.

We establish the equivalence of the almost-sure winning problems for pure strategies and for randomized strategies with actions invisible. Our equivalence result exhibits serious flaws in previous results of the literature: we show a non-elementary memory lower bound for almost-sure winning whereas an exponential upper bound was previously claimed.

***Keywords*-Partial-observation games, Stochastic games, Reachability and Büchi objectives, Positive and Almost-sure winning, Complexity, Memory bounds.**

I. Introduction

Games on graphs. Two-player games on graphs play a central role in several important problems in computer science, such as controller synthesis [33], [35], verification of open systems [2], realizability and compatibility checking [1], [21], [18], and many others. Most results about two-player games on graphs make the hypothesis of *perfect observation* (i.e., both players have perfect or complete observation about the state of the game). This assumption is often not realistic in practice. For example in the context of hybrid systems, the controller acquires information about the state of a plant using digital sensors with finite precision, which gives imperfect information about the state of the plant [20], [26]. Similarly, in a concurrent system where the players represent individual processes, each process has only access to the public variables of the other processes, not to their private variables [37], [2]. Such problems are better modeled in the more general framework of *partial-observation* games [36], [37], [38], [16], [6] and have been studied in the context of verification and synthesis [30], [2], [20], [45].

Partial-observation stochastic games and subclasses. In two-player partial-observation stochastic games on graphs with a finite state space, in every round, both players independently and simultaneously choose actions which along with the current state give a probability distribution over the successor states in the game. In a general setting, the players may not be able to distinguish certain states which are observationally equivalent for them (e.g., if they differ only by the value of private variables). The state space is partitioned into *observations* defined as equivalence classes and the players do not see the actual state of the game, but only an observation (which is typically different for the two players). The model of partial-observation games we consider is the same as the model of stochastic games with signals [6] and is a standard model in game theory [39], [41]. It subsumes other classical game models such as concurrent games [40], [19], probabilistic automata [34], [32], and partial-observation Markov decision processes (POMDPs) [31] (see also the recent decidability and complexity results for probabilistic automata [3], [4], [5], [9], [10], [11], [24] and for POMDPs [15], [3], [43]).

The special case of *perfect observation* for a player corresponds to every observation for this player being a singleton. Depending on which player has perfect observation, we consider the following *one-sided* subclasses of the general two-sided partial-observation stochastic games: (1) *player*-1 *partial and player*-2 *perfect* where player 2 has perfect observation, and player 1 has partial observation; and (2) *player*-1 *perfect and player*-2 *partial* where player 1 has perfect observation, and player 2 has partial observation. The case where the two players have perfect observation corresponds to the well-known perfect-information (perfect-observation) stochastic games [40], [17], [19].

Note that in a given (two-sided) game G, if player 1 wins in the setting of player-1 partial and player-2 perfect, then player 1 wins in the game G as well. Analogously, if player 1

cannot win in the setting of player 1 perfect and player 2 partial, then player 1 does not win in the game G either. In this sense, the one-sided games are conservative over- and under-approximations of two-sided games. In the context of applications in verification and synthesis, the conservative approximation is that the adversary is all powerful, and hence the games with player 1 partial and player 2 perfect games provide the important worst-case analysis of partial-observation games.

Objectives and qualitative problems. In this work we consider partial-observation stochastic games with *reachability* objectives where the goal of player 1 is to reach a set of target states, and games with *Büchi* objectives where the goal of player 1 is to visit some target state infinitely often. The study of partial-observation games is considerably more complicated than games of perfect observation. For example, in contrast to perfect-observation games, strategies in partial-observation games require both randomization and memory for reachability objectives; and the *quantitative* problem of deciding whether there exists a strategy for player 1 to ensure that the target is reached with probability at least $\frac{1}{2}$ can be decided in NP $\cap$ coNP for perfect-observation stochastic games [17], whereas the problem is undecidable even for partial-observation stochastic games with only one player [32]. Since the quantitative problem is undecidable, we consider the following *qualitative* problems: the *almost-sure* problem for reachability (resp. Büchi) objectives asks whether there exists a strategy for player 1 to ensure that the target set is reached (resp. visited infinitely often) with probability 1; the *positive* problem asks the same question, but requires positive probability instead of probability 1. For Büchi objectives, the positive problem is undecidable [3], and the almost-sure problem is polynomially equivalent to the almost-sure problem for reachability objectives [3]. Therefore, we discuss reachability objectives, and the results for Büchi objectives follow.

Classes of strategies. In general, randomized strategies are necessary to win with probability 1 in a partial-observation game with reachability objective [16]. However, there exist two types of randomized strategies where either (i) actions are visible, the player can observe the action he played [16], [6], or (ii) actions are invisible, the player may choose a probability distribution over actions, but the source of randomization is external and the actual (random) choice of the action is invisible to the player [25]. The second model is more general since the qualitative problems of randomized strategies with actions visible can be reduced in polynomial time to randomized strategies with actions invisible, by modeling the visibility of actions using the observations on states.

With actions visible, the almost-sure (resp. positive) problem was shown to be EXPTIME-complete (resp. PTIME-complete) for one-sided games with player 1 partial and player 2 perfect [16], and 2EXPTIME-complete (resp. EXPTIME-complete) in the two-sided case [6]. For the positive problem memoryless randomized strategies exist, and for the almost-sure problem *belief-based* strategies exist (strategies based on subset construction that consider the possible current states of the game). It was remarked (without any proof) in [16, p.4] that these results easily extend to randomized strategies with actions invisible for one-sided games with player 1 partial and player 2 perfect. It was claimed in [25] (Theorems 1 & 2) that the almost-sure problem is 2EXPTIME-complete for randomized strategies with actions invisible for two-sided games, and that belief-based strategies are sufficient for player 1. Thus it is believed that the two qualitative problems with actions visible or actions invisible are essentially equivalent.

Pure strategies and motivation. In this paper, we consider the class of *pure* strategies, which do not use randomization at all. Pure strategies arise naturally in the synthesis of controllers and processes that do not have access to any source of randomization, such as synchronizers for lock placement in concurrent programs [8], and controllers for robot planning [29]. Moreover we will establish deep connections between the qualitative problems for pure strategies and for randomized strategies with actions invisible, which on one hand exhibit major flaws in previous results of the literature (the remark without proof of [16] and the main results of [25]), and on the other hand show that the solution for almost-sure winning randomized strategies with actions invisible (which is the most general case) can be surprisingly obtained by solving the problem for pure strategies.

Contributions. The contributions of the paper are as follows.

1) *Player* 1 *partial and player* 2 *perfect.* We show that both for almost-sure and positive winning, belief-based pure strategies are not sufficient. This implies that the classical approaches relying on the belief-based subset construction cannot work for solving the qualitative problems for pure strategies. However, we present an optimal exponential upper bound on the memory needed by pure strategies (the exponential lower bound follows from the special case of non-stochastic games [7]). By a reduction to perfect-observation games of exponential size, we show that both the almost-sure and positive problems are EXPTIME-complete for one-sided games with perfect-observation for player 2. In contrast to the previous proofs of EXPTIME upper bound that rely either on subset constructions or enumeration of belief-based strategies, our correctness proof relies on a novel rank-based argument that works uniformly both for positive and almost-sure winning. The structure of this construction also provides symbolic antichain-based algorithms (see [22] for a survey of the antichain approach) for solving the qualitative problems that avoids the explicit exponential construction. Thus for the important special case of player 1 partial and player 2 perfect we establish optimal memory bound, complexity bound, and obtain symbolic algorithmic solutions for the qualitative problems.

	one-sided player 2 perfect		one-sided player 1 perfect		two-sided	
	Positive	Almost-sure	Positive	Almost-sure	Positive	Almost-sure
Randomized (actions visible)	Memoryless	Exponential (belief-based)	Memoryless	Memoryless	Memoryless	Exponential (belief-based)
Randomized (actions invisible)	Memoryless	**Exponential (belief is not sufficient)**	Memoryless	Memoryless	Memoryless	**Non-elem. low. bound** **Finite upp. bound**
Pure	**Exponential (belief is not sufficient)**	**Exponential (belief is not sufficient)**	**Non-elem. complete**	**Non-elem. complete**	**Non-elem. low. bound** **Finite upp. bound**	**Non-elem. low. bound** **Finite upp. bound**

TABLE I
MEMORY REQUIREMENT FOR PLAYER 1 AND REACHABILITY OBJECTIVE.

2) *Player* 1 *perfect and player* 2 *partial.*
   a) We show a very surprising result that both for positive and almost-sure winning, pure strategies for player 1 require memory of non-elementary size (i.e., a tower of exponentials). This is in sharp contrast with (i) the case of randomized strategies (with or without actions visible) where memoryless strategies are sufficient for positive winning, and with (ii) the previous case where player 1 has partial observation and player 2 has perfect observation, where pure strategies for positive winning require only exponential memory. Surprisingly and perhaps counter-intuitively when player 1 has more information and player 2 has less information, the positive winning strategies for player 1 require much more memory (non-elementary as compared to exponential). With more information player 1 can win from more states, but the winning strategy is much harder to implement.
   b) We present a non-elementary upper bound for the memory needed by pure strategies for positive winning. We then show with an example that for almost-sure winning more memory may be required as compared to positive winning. Finally, we show how to combine pure strategies for positive winning in a recharging scheme to obtain a non-elementary upper bound for the memory required by pure strategies for almost-sure winning. Thus we establish non-elementary complete bounds for pure strategies both for positive and almost-sure winning.
3) *General (two-sided) case.* We show that in the general case finite memory strategies are sufficient both for positive and almost-sure winning. The result is obtained essentially by a simple generalization of König's Lemma [28]. A non-elementary lower bound for memory follows from the special case when player 1 has perfect observation and player 2 has partial observation.
4) *Randomized strategies with actions invisible.* For randomized strategies with actions invisible we give two reductions to establish connections with pure strategies. First, we show that the almost-sure problem for randomized strategies with actions invisible reduces in polynomial time to the almost-sure problem for pure strategies. The reduction requires to first establish that finite-memory randomized strategies are sufficient in two-sided games. Second, we show that the problem of almost-sure winning with pure strategies reduces in polynomial time to the problem of randomized strategies with actions invisible. For this reduction it is crucial that the actions are not visible.

   Our reductions have deep consequences. They unexpectedly imply that the problems of almost-sure winning with *pure* strategies or *randomized* strategies with actions invisible are polynomial-time *equivalent*. Moreover, it follows that even in one-sided games with player 1 partial and player 2 perfect, belief-based randomized strategies with actions invisible are not sufficient for almost-sure winning. This shows that the remark (without proof) of [16] that the results (such as existence of belief-based strategies) of randomized strategies with actions visible carry over to actions invisible is an oversight. However from our first reduction and our results for pure strategies it follows that there is an exponential upper bound on memory and the problem is EXPTIME-complete for one-sided games with player 1 partial and player 2 perfect. More importantly, our results exhibit a serious flaw[1] in the main result of [25] which showed that belief-based randomized strategies with actions invisible are sufficient for almost-sure winning in two-sided games, and concluded that enumerating over such strategies yields a 2EXPTIME algorithm for the problem. Our second reduction and lower bound for pure strategies show that the result is incorrect, and that the exponential (belief-based) upper bound is far off. Instead, the lower bound on memory for almost-sure winning with randomized strategies and actions invisible is non-elementary. Thus, contrary to the general belief, there is a sharp contrast for randomized strategies with or without actions visible: if actions are visible, then exponential memory is sufficient for almost-sure winning while if actions are not visible, then memory of non-elementary size is necessary in general.

The memory requirements are summarized in Table I and the results of this paper are shown in bold font. We explain

[1]This flaw was presented in [13] to the authors of [25] and acknowledged in August 2011.

how the other results of the table follow from results of the literature. For randomized strategies (with or without actions visible), if a positive winning strategy exists, then a memoryless strategy that plays all actions uniformly at random is also positive winning. Thus the memoryless result for positive winning strategies follows for all cases of randomized strategies. The belief-based bound for memory of almost-sure winning randomized strategies with actions visible follows from [16], [6]. The memoryless strategies results for almost-sure winning for one-sided games with player 1 perfect and player 2 partial are obtained as follows: when actions are visible, then belief-based strategies coincide with memoryless strategies as player 1 has perfect observation. If player 1 has perfect observation, then for memoryless strategies whether actions are visible or not is irrelevant and thus the memoryless result also follows for randomized strategies with actions invisible. Thus we obtain Table I. Proofs omitted due to lack of space are available in a technical report released in July 2011 [13].

II. Definitions

A *probability distribution* on a finite set S is a function $\kappa : S \to [0,1]$ such that $\sum_{s \in S} \kappa(s) = 1$. The *support* of κ is the set $\mathsf{Supp}(\kappa) = \{s \in S \mid \kappa(s) > 0\}$. We denote by $\mathcal{D}(S)$ the set of probability distributions on S. Given $s \in S$, the *Dirac distribution* on s assigns probability 1 to s.

Games. Given finite alphabets A_i of actions for player i ($i = 1, 2$), a *stochastic game* on A_1, A_2 is a tuple $G = \langle Q, q_0, \delta \rangle$ where Q is a finite set of states, $q_0 \in Q$ is the initial state, and $\delta : Q \times A_1 \times A_2 \to \mathcal{D}(Q)$ is a probabilistic transition function that, given a current state q and actions a, b for the players gives the transition probability $\delta(q, a, b)(q')$ to the next state q'. The game is called *deterministic* if $\delta(q, a, b)$ is a Dirac distribution for all $(q, a, b) \in Q \times A_1 \times A_2$. A state q is *absorbing* if $\delta(q, a, b)$ is the Dirac distribution on q for all $(a, b) \in A_1 \times A_2$. In some examples, we allow an initial distribution of states. This can be encoded in our game model by a probabilistic transition from the initial state.

A *player*-1 *state* is a state q where $\delta(q, a, b) = \delta(q, a, b')$ for all $a \in A_1$ and all $b, b' \in A_2$. We use the notation $\delta(q, a, -)$. *Player*-2 *states* are defined analogously. In figures, we use boxes to emphasize that a state is a player-2 state, and we represent probabilistic branches using diamonds (which are not real 'states', e.g., as in Fig. 1).

In a (two-sided) *partial-observation* game, the players have a partial or incomplete view of the states visited and of the actions played in the game. This view may be different for the two players and it is defined by equivalence relations $\approx_i$ on the states and on the actions ($i = 1, 2$). For player i, equivalent states (or actions) are indistinguishable. We denote by $\mathcal{O}_i \subseteq 2^Q$ ($i = 1, 2$) the $\approx_i$-equivalence classes of states which define two partitions of the state space Q, and we call them *observations* (for player i). These partitions uniquely define functions $\mathsf{obs}_i : Q \to \mathcal{O}_i$ such that $q \in \mathsf{obs}_i(q)$ for all $q \in Q$, that map each state q to its observation for player i.

In the case where all states and actions are equivalent (i.e., the relation $\approx_i$ is the set $(Q \times Q) \cup (A_1 \times A_1) \cup (A_2 \times A_2)$), we say that player i is *blind* and the actions are *invisible*. In this case, we have $\mathcal{O}_i = \{Q\}$ because all states have the same observation. Note that the case of perfect observation for player i corresponds to the case $\mathcal{O}_i = \{\{q_0\}, \{q_1\}, \ldots, \{q_n\}\}$ (given $Q = \{q_0, q_1, \ldots, q_n\}$), and $a \approx_i b$ iff $a = b$, for all actions a, b.

For $s \subseteq Q$, $a \in A_1$, and $b \in A_2$, let $\mathsf{Post}_{a,b}(s) = \bigcup_{q \in s} \mathsf{Supp}(\delta(q, a, b))$ denote the set of possible successors of q given action a and b, and let $\mathsf{Post}_{a,-}(s) = \bigcup_{b \in A_2} \mathsf{Post}_{a,b}(s)$.

Plays and observations. Initially, the game starts in the initial state q_0. In each round, player 1 chooses an action $a \in A_1$, player 2 (simultaneously and independently) chooses an action $b \in A_2$, and the successor of the current state q is chosen according to the probabilistic transition function $\delta(q, a, b)$. A *play* in G is an infinite sequence $\rho = q_0\, a_0 b_0\, q_1\, a_1 b_1\, q_2 \ldots$ such that q_0 is the initial state and $\delta(q_j, a_j, b_j)(q_{j+1}) > 0$ for all $j \geq 0$ (the actions a_j's and b_j's are the actions *associated* to the play). Its *length* is $|\rho| = \infty$. The length of a play prefix $\rho = q_0\, a_0 b_0\, q_1 \ldots q_k$ is $|\rho| = k$, and its last element is $\mathsf{Last}(\rho) = q_k$. A state $q \in Q$ is *reachable* if it occurs in some play. We denote by $\mathsf{Plays}(G)$ the set of plays in G, and by $\mathsf{Prefs}(G)$ the set of corresponding finite prefixes. For $i = 1, 2$, the *observation sequence* for player i of a play (prefix) ρ is the unique (in)finite sequence $\mathsf{obs}_i(\rho) = \gamma_0 \gamma_1 \ldots$ such that $\gamma_j = \mathsf{obs}_i(q_j)$ for all $0 \leq j \leq |\rho|$.

The games with *one-sided partial-observation* are the special case where either $\approx_1$ is equality and hence $\mathcal{O}_1 = \{\{q\} \mid q \in Q\}$ (player 1 has complete observation) or $\approx_2$ is equality and hence $\mathcal{O}_2 = \{\{q\} \mid q \in Q\}$ (player 2 has complete observation). The games with *perfect observation* are the special cases where $\approx_1$ and $\approx_2$ are equality, i.e., every state and action is visible to both players.

Strategies. A *pure strategy* in G for player 1 is a function $\sigma : \mathsf{Prefs}(G) \to A_1$. A *randomized strategy* in G for player 1 is a function $\sigma : \mathsf{Prefs}(G) \to \mathcal{D}(A_1)$. A (pure or randomized) strategy σ for player 1 is *observation-based* if for all prefixes $\rho = q_0\, a_0 b_0\, q_1 \ldots$ and $\rho' = q'_0\, a'_0 b'_0\, q'_1 \ldots$, if $a_j \approx_1 a'_j$ and $b_j \approx_1 b'_j$ for all $j \geq 0$, and $\mathsf{obs}_1(\rho) = \mathsf{obs}_1(\rho')$, then $\sigma(\rho) = \sigma(\rho')$. In the sequel, strategies are meant to be observation-based in partial-observation games. If for all actions a and b we have $a \approx_1 b$ iff $a = b$, and $a \approx_2 b$ iff $a = b$ (all actions are distinguishable), then the strategy is *action visible*, and if for all actions a and b we have $a \approx_1 b$ and $a \approx_2 b$ (all actions are indistinguishable), then the strategy is *action invisible*. We say that a play (prefix) $\rho = q_0\, a_0 b_0\, q_1 \ldots$ is *compatible* with a pure (resp., randomized) strategy σ if the associated action of player 1 in step j is $a_j = \sigma(q_0\, a_0 b_0 \ldots q_{j-1})$ (resp., $a_j \in \mathsf{Supp}(\sigma(q_0\, a_0 b_0 \ldots q_{j-1}))$) for all $0 \leq j \leq |\rho|$.

We omit analogous definitions of strategies for player 2. We denote by Σ_G, Σ_G^O, Σ_G^P, Π_G, Π_G^O, and Π_G^P the set of all player-1 strategies, the set of all observation-based player-1 strategies, the set of all pure player-1 strategies, the set of all player-2 strategies in G, the set of all observation-based

player-2 strategies, and the set of all pure player-2 strategies, respectively.

Remark 1. *The model of games with partial observation on both actions and states can be encoded in a model of games with actions invisible and observations on states only: when actions are invisible, we can use the state space to keep track of the last action played, and reveal information about the last action played using observations on the states [25]. Therefore, in the sequel we assume that the actions are invisible to the players with partial observation. A play is then viewed as a sequence of states only, and the definition of strategies is updated accordingly. Note that a player with perfect observation has actions and states visible (and the equivalence relation $\approx_i$ is equality).*

Remark 2. *The important special case of partial-observation Markov decision processes (POMDP) corresponds to the case where either all states in the game are player-*1 *states (player-1 POMDP) or all states are player-*2 *states (player-2 POMDP). For POMDP it is known that randomization is not necessary, and pure strategies are as powerful as randomized strategies [14].*

Finite-memory strategies. A player-1 strategy uses *finite-memory* if it can be encoded by a deterministic transducer $\langle \mathsf{Mem}, m_0, \alpha_u, \alpha_n \rangle$ where Mem is a finite set (the memory of the strategy), $m_0 \in \mathsf{Mem}$ is the initial memory value, $\alpha_u : \mathsf{Mem} \times \mathcal{O}_1 \to \mathsf{Mem}$ is an update function, and $\alpha_n : \mathsf{Mem} \times \mathcal{O}_1 \to \mathcal{D}(A_1)$ is a next-move function. The *size* of the strategy is the number $|\mathsf{Mem}|$ of memory values. If the current observation is o, and the current memory value is m, then the strategy chooses the next action according to the probability distribution $\alpha_n(m, o)$, and the memory is updated to $\alpha_u(m, o)$. Formally, $\langle \mathsf{Mem}, m_0, \alpha_u, \alpha_n \rangle$ defines the strategy σ such that $\sigma(\rho \cdot q) = \alpha_n(\hat{\alpha}_u(m_0, \mathsf{obs}_1(\rho)), \mathsf{obs}_1(q))$ for all $\rho \in Q^*$ and $q \in Q$, where $\hat{\alpha}_u$ extends α_u to sequences of observations as expected. This definition extends to infinite-memory strategies by dropping the assumption that the set Mem is finite. A strategy is *memoryless* if $|\mathsf{Mem}| = 1$.

Objectives and winning modes. An *objective* (for player 1) in G is a set $\varphi \subseteq \mathsf{Plays}(G)$ of plays. A play $\rho \in \mathsf{Plays}(G)$ *satisfies* the objective φ, denoted $\rho \models \varphi$, if $\rho \in \varphi$. Objectives are generally Borel measurable: a Borel objective is a Borel set in the Cantor topology [27]. Given strategies σ and π for the two players, the probabilities of a measurable objective φ is uniquely defined [44]. We denote by $\Pr_{q_0}^{\sigma,\pi}(\varphi)$ the probability that φ is satisfied by the play obtained from the starting state q_0 when the strategies σ and π are used.

We specifically consider the following well-known objectives. Given a set $\mathcal{T} \subseteq Q$ of target states, the *reachability objective* requires that the play visit the set $\mathcal{T}$: $\mathsf{Reach}(\mathcal{T}) = \{q_0 a_0 b_0 q_1 \ldots \in \mathsf{Plays}(G) \mid \exists i \geq 0 : q_i \in \mathcal{T}\}$, and the *Büchi objective* requires that the play visit the set $\mathcal{T}$ infinitely often, $\mathsf{Büchi}(\mathcal{T}) = \{q_0 a_0 b_0 q_1 \ldots \in \mathsf{Plays}(G) \mid \forall i \geq 0 \cdot \exists j \geq i : q_j \in \mathcal{T}\}$. Our solution for reachability objectives will also use the dual notion of *safety objectives* that require the play to stay within the set $\mathcal{T}$: $\mathsf{Safe}(\mathcal{T}) = \{q_0 a_0 b_0 q_1 \ldots \in \mathsf{Plays}(G) \mid \forall i \geq 0 : q_i \in \mathcal{T}\}$. In figures, the target states in $\mathcal{T}$ are double-lined and labeled by ☺.

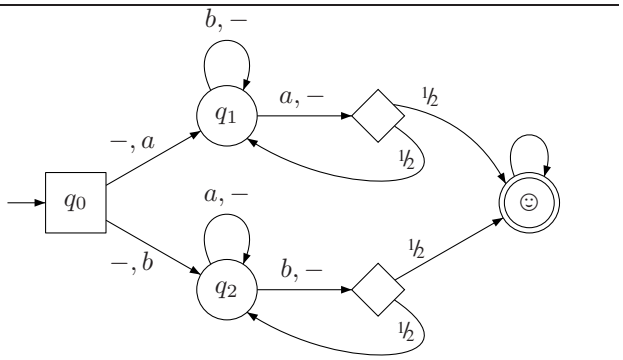

Fig. 1. Belief-based pure strategies are not sufficient for positive and almost-sure reachability.

Given a game structure G and a state q, an observation-based strategy σ for player 1 is *almost-sure winning* (resp. *positive winning*) for the objective φ from q if for all observation-based randomized strategies π for player 2, we have $\Pr_q^{\sigma,\pi}(\varphi) = 1$ (resp. $\Pr_q^{\sigma,\pi}(\varphi) > 0$). The strategy σ is *sure winning* if all plays compatible with σ satisfy φ. We also say that the state q is almost-sure (or positive, or sure) winning for player 1.

Positive and almost-sure winning problems. We are interested in the problems of deciding, given a game structure G, a state q, and an objective φ, whether there exists a {pure, randomized} strategy which is {almost-sure, positive} winning from q for the objective φ. For safety objectives almost-sure winning coincides with sure winning, however for reachability objectives they are different. The sure winning problem for the objectives we consider has been studied in [36], [16], [12]. The almost-sure winning problem for Büchi objectives can be easily reduced to the almost-sure winning problem for reachability objectives [3]. The positive winning problem for Büchi objectives is undecidable even for POMDPs [3]. Hence in this paper we mostly focus on reachability objectives.

Remark 3. *(Almost-sure Büchi to almost-sure reachability [3]). The reduction of almost-sure Büchi to almost-sure reachability is as follows: given a two-sided stochastic game with Büchi objective* $\mathsf{Büchi}(\mathcal{T})$*, we add a new absorbing state q_T, make q_T the target state for the reachability objective, and from every state $q \in \mathcal{T}$ we add positive probability transitions to q_T (details and correctness proof follow from [3, Lemma 13]).*

III. One-sided Games: Player 1 Partial and Player 2 Perfect

In Sections III and IV, we consider one-sided games with partial observation: one player has perfect observation, and the other player has partial observation. The player with perfect observation sees the states visited and the actions played in the game. We present the results for positive and almost-sure

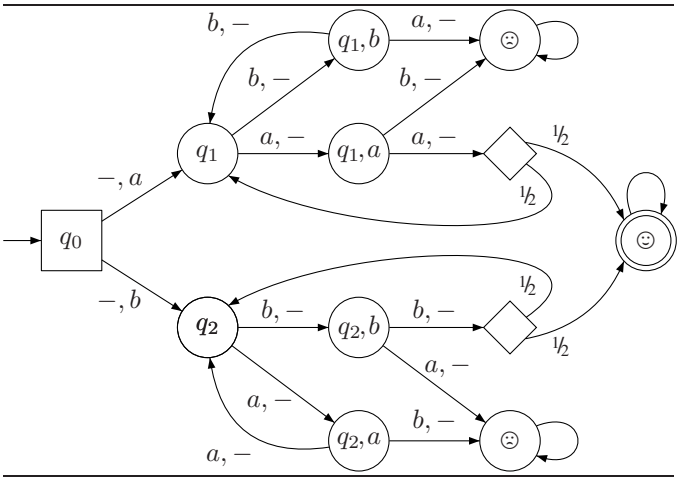

Fig. 2. Belief-based randomized action-invisible strategies are not sufficient for almost-sure reachability.

winning for reachability objectives along with examples that illustrate key elements of the problem such as the memory required for winning strategies.

Note that the case of player 1 partial and player 2 perfect is important in the context of controller synthesis as it is a conservative approximation of two-sided games for player 1 (if player 1 wins in the one-sided game, then he also wins in the two-sided game). In the following example we show that for pure strategies *belief-based* strategies are not sufficient for positive as well as almost-sure winning. A strategy is belief-based if its memory relies only on the subset construction, i.e., the strategy plays only depending on the set of possible current states of the game which is called *belief*.

Example 1. Belief is not sufficient for positive (as well as almost-sure) reachability. *Consider the game in Fig. 1 where player* 1 *is blind (all states have the same observation except the target state, and actions are invisible) and player* 2 *has perfect observation. Initially, player* 2 *chooses the state* q_1 *or* q_2 *(which player* 1 *does not see). The belief of player* 1 *is thus the set* $\{q_1, q_2\}$ *(see Fig. 3). We claim that the belief is not a sufficient information to win with a pure strategy for player* 1 *because the belief-based subset construction in Fig. 3 suggests that playing always the same action (say* a*) when the belief is* $\{q_1, q_2\}$ *is an almost-sure winning strategy. However, in the original game this is not even a positive winning strategy (the counter strategy of player* 2 *is to choose* q_2 *initially). A winning strategy for player* 1 *is to alternate between* a *and* b *when the belief is* $\{q_1, q_2\}$*, showing that remembering the belief is not sufficient.* ■

We present reductions of the almost-sure and positive winning problem for reachability objective to the problem of sure-winning in a game of perfect observation with Büchi objective, and reachability objective respectively. The two reductions are based on the same construction of a game where the state space $L = \{(s, o) \mid o \subseteq s \subseteq Q\}$ contains the subset construction s enriched with *obligation sets* $o \subseteq s$ which ensure that from all

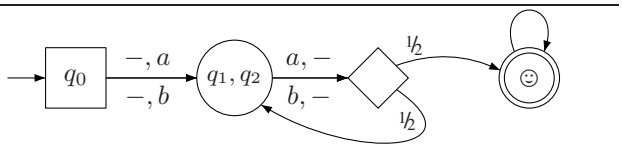

Fig. 3. A belief-based subset construction for the reachability game of Fig. 1.

states in s, the target set $\mathcal{T}$ is reached with positive probability. The Büchi (resp. reachability) objective is to visit the empty obligation set infinitely often (resp. at least once). Instead of a naive solution which would keep track of all successors of probabilistic choices, we use a rank-based argument on the obligation set to show correctness of the construction. The key argument considers arbitrary (possibly infinite-memory) almost-sure winning strategy σ, and proves the existence of a finite ranking in the infinite tree obtained from σ such that target states have rank 0, the rank is strictly decreasing for non-target states, and the root gets a finite rank.

The construction is as follows. Given $G = \langle Q, q_0, \delta \rangle$ over alphabets A_1, A_2 and observation set $\mathcal{O}_1$ for player 1, with reachability objective $\mathsf{Reach}(\mathcal{T})$, we construct the following (deterministic) game of perfect observation $H = \langle L, \ell_0, \delta_H \rangle$ over alphabets A_1', A_2' such that player 1 has a pure observation-based almost-sure (resp., positive) winning strategy in G from q_0 if and only if player 1 has a sure winning strategy in H from ℓ_0 for the objective $\mathsf{Büchi}(\alpha)$ (resp., $\mathsf{Reach}(\alpha)$) defined by $\alpha \subseteq L$ where:

- $L = \{(s, o) \mid o \subseteq s \subseteq Q\}$. Intuitively, s is the belief of player 1 and o is a set of obligation states that "owe" a visit to $\mathcal{T}$ with positive probability.
- $\ell_0 = (\{q_0\}, \{q_0\})$ if $q_0 \notin \mathcal{T}$, and $\ell_0 = (\emptyset, \emptyset)$ if $q_0 \in \mathcal{T}$;
- $A_1' = A_1 \times 2^Q$. In a pair $(a, u) \in A_1'$, we call a the action, and u the witness set;
- $A_2' = \mathcal{O}_1$. In the game H, player 2 simulate player 2's choice in game G, as well as resolves the probabilistic choices. This amounts to choosing a possible successor state, and revealing its observation;
- $\alpha = \{(s, \emptyset) \in L\}$;
- δ_H is defined as follows. First, the state $(\emptyset, \emptyset)$ is absorbing. Second, in every other state $(s, o) \in L$ the function δ_H ensures that (i) player 1 chooses a pair (a, u) such that $\mathsf{Supp}(\delta(q, a, b)) \cap u \neq \emptyset$ for all $q \in o$ and $b \in A_2$, and (ii) player 2 chooses an observation $\gamma \in \mathcal{O}_1$ such that $\mathsf{Post}_{a,-}(s) \cap \gamma \neq \emptyset$. If a player violates this, then a losing absorbing state is reached with probability 1. Assuming the above condition on (a, u) and γ is satisfied, define $\delta_H((s, o), (a, u), \gamma)$ as the Dirac distribution on the state (s', o') such that:
  - $s' = (\mathsf{Post}_{a,-}(s) \cap \gamma) \setminus \mathcal{T}$;
  - $o' = s'$ if $o = \emptyset$; and $o' = (\mathsf{Post}_{a,-}(o) \cap \gamma \cap u) \setminus \mathcal{T}$ if $o \neq \emptyset$.

Lemma 1. *Given a one-sided partial-observation stochastic game* G *with player 1 partial and player 2 perfect with a reachability objective for player 1, we can construct in time*

exponential in the size of the game and polynomial in the size of action sets a perfect-information deterministic game H with a Büchi objective (resp. reachability objective) such that player 1 has a pure almost-sure (resp. positive) winning strategy in G iff player 1 has a sure-winning strategy in H.

It follows from the construction in the proof of Lemma 1 that pure strategies with exponential memory are sufficient for positive (as well as almost-sure) winning, and the exponential lower bound follows from the special case of non-stochastic games [7]. Lemma 1 also gives EXPTIME upper bound for the problem since perfect-observation Büchi games can be solved in polynomial time [42]. The EXPTIME-hardness follows from the sure winning problem for non-stochastic games [37], where pure almost-sure (positive) winning strategies coincide with sure winning strategies. Theorem 1 summarizes the results, and note that by Remark 3 all the results of the theorem for almost-sure winning also hold for Büchi objectives.

Theorem 1. *Given one-sided partial-observation stochastic games with player 1 partial and player 2 perfect, the following assertions hold for reachability objectives for player 1:*

1) (Memory bound). *Belief-based pure strategies are not sufficient both for positive and almost-sure winning; exponential memory is necessary and sufficient both for positive (memory of size $\sum_{\gamma \in \mathcal{O}_1} 2^{|\gamma|}$ is sufficient) and almost-sure winning (memory of size $\sum_{\gamma \in \mathcal{O}_1} 3^{|\gamma|}$ is sufficient) for pure strategies, where $|\gamma|$ is the cardinality of γ.*
2) (Algorithm). *The problems of deciding the existence of a pure almost-sure and a pure positive winning strategy can be solved in time exponential in the state space of the game and polynomial in the size of the action sets.*
3) (Complexity). *The problems of deciding the existence of a pure almost-sure and a pure positive winning strategy are EXPTIME-complete.*

Symbolic algorithms. The exponential Büchi (or reachability) game constructed in the proof of Lemma 1 can be solved by computing classical fixpoint formulas [23]. However, it is not necessary to construct the exponential game structure explicitly. Instead, we can exploit the structure induced by the pre-order $\preceq$ defined by $(s, o) \preceq (s', o')$ if (i) $s \subseteq s'$, (ii) $o \subseteq o'$, and (iii) $o = \emptyset$ iff $o' = \emptyset$. Intuitively, if a state (s', o') is winning for player 1, then all states $(s, o) \preceq (s', o')$ are also winning because they correspond to a better belief and a looser obligation. Hence all sets computed by the fixpoint algorithm are downward-closed and thus they can be represented symbolically by the antichain of their maximal elements (see [16] for details related to antichain algorithms). This technique provides a symbolic algorithm without explicitly constructing the exponential game.

IV. One-sided Games: Player 1 Perfect and Player 2 Partial

Recall that we are interested in finding a pure winning strategy for player 1. We present the key ideas of the main three results for one-sided games with player 1 perfect and player 2 partial.

Lower bound on memory. We present a family of games where player 1 needs memory of non-elementary size to satisfy both almost-sure and positive reachability. The key idea is that player 1 needs to remember not only the possible current states of the game (belief of player 2), but also how many paths that player 2 cannot distinguish end up in each state. Then we show that player 1 needs to simulate a counter system where the operations on counters are increment and division by 2 (with round down) which requires to store non-elementary values of the counters in the worst case. The key challenge is to construct a polynomial-size game to simulate non-elementary counter values. We show how to use the partial observation of player 2 to achieve this. This establishes the surprising non-elementary lower bound. See [13, Theorem 2] for details.

Upper bound for positive reachability with almost-sure safety. We show a matching non-elementary upper bound for pure strategies to ensure positive reachability along with almost-sure safety. We obtain the solution for positive reachability as a special case and on the other hand it will be required for solving almost-sure reachability. The result is achieved in the following steps. First, we compute the set of states from which player 1 can satisfy the objective with a *randomized action-visible* strategy. Second, we show how pure strategies can simulate randomized strategies by using the stochasticity of the transition relation and the fact that player 2 cannot distinguish observationally-equivalent paths. This is the main novel idea behind this proof. Finally, we show that if the number of indistinguishable paths is non-elementary, then player 1 achieves the full power of randomized action-visible strategies and is winning using the computation of the first step. The crux of the final step is to analyze a new class of counter systems (with division by a constant and increment) and show that counters with non-elementary value suffice. See [13, Theorem 3] for details.

Upper bound for almost-sure reachability. We show an example of a game where memoryless positive winning strategies exist, but almost-sure winning strategies require memory [13, Example 4]. We then present a construction of a pure almost-sure winning strategy (when such a strategy exists) by repeatedly playing a strategy for positive reachability along with almost-sure safety in a *recharging* scheme. As a consequence we obtain a non-elementary upper bound on the memory size of almost-sure winning strategies. Let Q_B be the set of states such that if the belief of player 2 is a state in Q_B, then against all strategies of player 1, player 2 can ensure that with positive probability the target is not reached. Hence an almost-sure winning strategy must ensure almost-sure safety for the set $Q_G = Q \setminus Q_B$. From Q_G player 1 can ensure both positive reachability to the target as well as safety for the set Q_G. We show that repeatedly playing a strategy for positive reachability along with almost-sure safety is an almost-sure winning strategy for the reachability objective (details in [13, Theorem 4]). By Remark 3, the results of Theorem 2 and

Corollary 1 for almost-sure winning also hold for Büchi objectives.

Theorem 2. *In one-sided partial-observation stochastic games with player 1 perfect and player 2 partial, the following assertions hold:*

1) *Both pure almost-sure and pure positive winning strategies for reachability objectives for player 1 require memory of non-elementary size in general.*
2) *Non-elementary size memory is sufficient for pure strategies to ensure positive probability reachability along with almost-sure safety for player 1; and hence for pure positive winning strategies for reachability objectives for player 1 non-elementary memory bound is optimal.*
3) *Non-elementary size memory is sufficient for pure strategies to ensure almost-sure reachability for player 1; and hence for pure almost-sure winning strategies for reachability objectives for player 1 non-elementary memory bound is optimal.*

Corollary 1. *In one-sided partial-observation stochastic games with player 1 perfect and player 2 partial, the problem of deciding the existence of pure almost-sure and positive winning strategies for reachability objectives for player 1 can be solved in non-elementary time complexity.*

Discussion about the surprising non-elementary memory bound. We now discuss the surprising non-elementary memory bound for positive winning with reachability objectives for pure strategies in player-1 perfect player-2 partial stochastic games, comparing it with other related questions. We consider four related questions: two are related to stochasticity in transitions and strategies, and the other two are related to the information of the players (see also Fig. 4).

1) *Question 1.* If we consider player-1 perfect player-2 partial deterministic games with reachability objective, then for positive winning pure memoryless strategies are sufficient. This follows from the results of [36] because in deterministic games positive winning coincides with sure winning, and the results of [36] shows (see [16] for an explicit proof) that for sure winning the observation of player 2 is irrelevant. Hence the problem is same as sure winning in perfect-information deterministic games with reachability objective for which pure memoryless strategies exist.
2) *Question 2.* If we consider player-1 perfect player-2 partial stochastic games with reachability objective, but instead of pure strategies consider randomized strategies, then memoryless strategies are sufficient. It follows from [6] that if there is a randomized strategy to ensure reachability with positive probability, then the randomized memoryless strategy that plays all actions uniformly at random is also a positive winning strategy.
3) *Question 3.* If we consider perfect-information stochastic games (both players have perfect information) with reachability objective, then for positive winning pure memoryless strategies are sufficient. This follows from a more general result of [17] that in perfect-information stochastic games with reachability objective, pure memoryless optimal strategies exist.
4) *Question 4.* If we consider player-1 partial player-2 perfect stochastic games with reachability objective, then for positive winning exponential memory pure strategies are sufficient (by Theorem 1).

Observe that the question we study is a natural extension of the above questions: (1) adding stochasticity to the transition as compared to question 1; (2) restricting strategies to pure strategies as compared to randomized strategies of question 2; (3) player 2 is less informed as compared to question 3; and (4) player 1 is more informed and player 2 is less informed as compared to question 4. Our results show the natural variant of question 1 and question 2 obtained by adding stochasticity to transitions or removing stochasticity from strategies, and the variant of question 3 and question 4 by making player 1 most well informed lead to a sunrising memory bound for strategies (non-elementary complete memory bound, whereas for all the related questions memoryless or exponential-size memory strategies are sufficient). See also Fig. 4.

V. Two-sided Games

We show the existence of finite-memory pure strategies for positive and almost-sure winning in two-sided games.

Positive reachability with almost-sure safety. We show that to ensure positive reachability along with almost-sure safety, finite-memory strategies suffice. The proof is in two parts: (1) we show that if there is an infinite-memory strategy σ, then the strategy ensures positive reachability within a finite number N of steps and almost-sure safety (the result is shown by a simple extension of König's Lemma [28]), and (2) then a finite-memory strategy plays like σ for N steps and then switches to a strategy for almost-sure safety (and for almost-sure safety finite-memory strategies suffice [12]). See [13, Theorem 5] for details.

Almost-sure reachability. The proof to show that finite-memory strategies suffice for almost-sure winning is analogous to the proof of the previous section for player 1 perfect and player 2 partial, where an almost-sure winning strategy is constructed by repeatedly playing finite-memory strategies (of [13, Theorem 5]) for positive reachability along with almost-sure safety in a recharging scheme. See [13, Theorem 6] for details.

Theorem 3. *In two-sided partial-observation stochastic games finite memory is sufficient (and non-elementary memory is required in general) for pure strategies both for positive and almost-sure winning for reachability objectives for player 1.*

VI. Equivalence of Randomized Action-invisible and Pure Strategies

In this section, we show that for two-sided partial-observation games, the problem of almost-sure winning with randomized action-invisible strategies is inter-reducible with the problem of almost-sure winning with pure strategies. The reductions are polynomial in the number of states in the

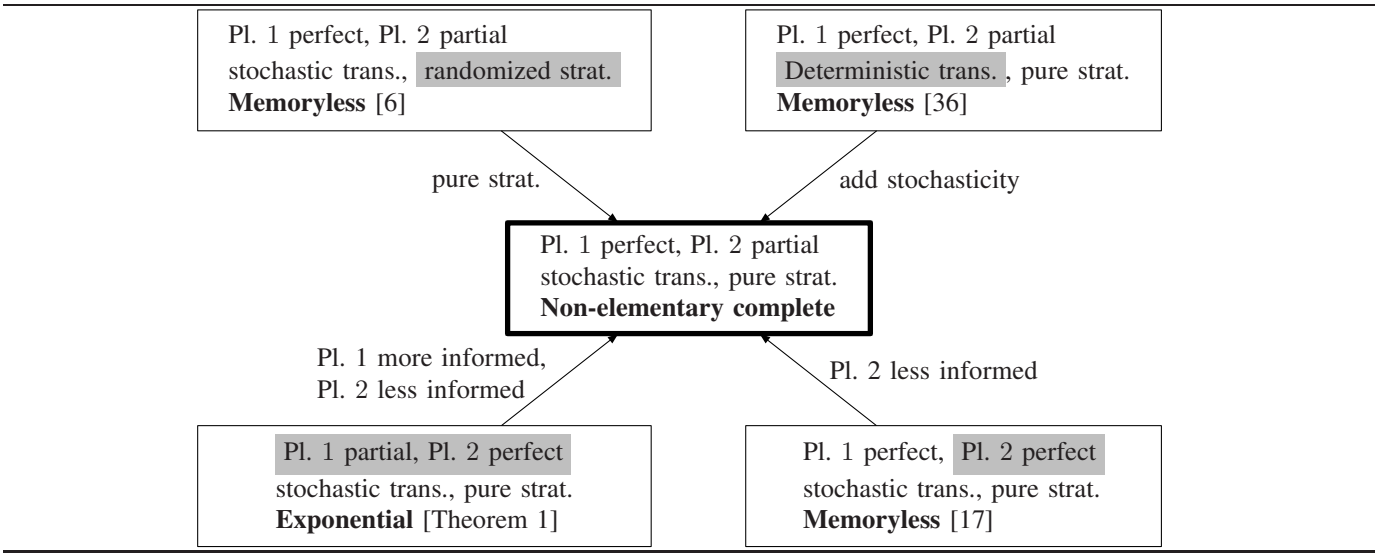

Fig. 4. The surprising non-elementary bound for memory of pure strategies in one-sided partial-observation stochastic games for player 1 perfect and player 2 partial for positive winning with reachability objectives (Theorem 2).

game (the reduction from randomized to pure strategies is exponential in the number of actions).

Reduction of randomized action-invisible strategies to pure strategies. We give a reduction for almost-sure winning for randomized action-invisible strategies to pure strategies. Given a stochastic game G we will construct another stochastic game H such that there is a randomized action-invisible almost-sure winning strategy in G iff there is a pure almost-sure winning strategy in H. The idea of the reduction is as follows: in the game H, player 1 can choose a non-empty subset $A \subseteq A_1$ of actions, and the probabilistic transition function in H is as follows: for $q \in Q$, $A \subseteq A_1$ and $b \in A_2$, we have $\delta_H(q, A, b)(q') = \frac{1}{|A|} \cdot \sum_{a \in A} \delta(q, a, b)(q')$. The observation mapping is same as in G. Further details in [13, Section 6.1] establish the following theorem and corollary.

Theorem 4. *Given a two-sided (resp. one-sided) partial-observation stochastic game G with a reachability objective we can construct in time polynomial in the size of the game and exponential in the size of the action sets a two-sided (resp. one-sided) partial-observation stochastic game H such that there exists a randomized action-invisible almost-sure winning strategy in G iff there exists a pure almost-sure winning strategy in H.*

For positive winning, randomized memoryless strategies are sufficient (both for action-visible and action-invisible) and the problem is PTIME-complete for one-sided and EXPTIME-complete for two-sided [6]. The above theorem along with Theorem 1 gives us the following corollary.

Corollary 2. *Given one-sided partial-observation stochastic games with player 1 partial and player 2 perfect, the following assertions hold for reachability objectives for player 1. (1) Exponential memory is sufficient for randomized action-invisible strategies for almost-sure winning. (2) The existence of a randomized action-invisible almost-sure winning strategy can be decided in time exponential in the state space of the game and exponential in the size of the action sets. (3) The problem of deciding the existence of a randomized action-invisible almost-sure winning strategy is EXPTIME-complete.*

Reduction of pure strategies to randomized action-invisible strategies. We present a reduction for almost-sure winning with pure strategies to randomized action-invisible strategies. Given a stochastic game G we construct another stochastic game H such that there exists a pure almost-sure winning strategy in G iff there exists a randomized almost-sure winning strategy in H. The idea of the reduction is to force player 1 to play a pure strategy in H. The game H simulates G and requires player 1 to repeat each action played (i.e., to play each action two times). Then, if player 1 uses randomization, he has to repeat the actions chosen randomly in the previous step. Since the actions are invisible, this can be achieved only if the support of the randomized actions is a singleton, i.e., the strategy is pure. Note that the reduction works for randomized strategies with actions invisible, and not when the actions are visible (details in [13, Section 6.2]).

Theorem 5. *Given a two-sided partial-observation stochastic game G with a reachability objective we can construct in time polynomial in the size of the game and size of the action sets a two-sided partial-observation stochastic game H such that there exists a pure almost-sure winning strategy in G iff there exists a randomized action-invisible almost-sure winning strategy in H.*

Belief-based strategies are not sufficient. We illustrate our reduction with the following example that shows belief-based (belief-only) randomized action-invisible strategies are not sufficient for almost-sure reachability in one-sided partial-observation games (player 1 partial and player 2 perfect),

showing that a remark (without proof) of [16, p.4] and the result and construction of [25, Theorem 1] are wrong.

Example 2. *We illustrate the reduction on the example of Fig. 1. The result of the reduction is given in Fig. 2. Remember that Example 1 showed that belief-based pure strategies are not sufficient for almost-sure winning. We show that belief-based randomized strategies are not sufficient for almost-sure winning in the game of Fig. 2. First, in* $\{q_1, q_2\}$ *player* 1 *has to play pure since he has to be able to repeat the same action to avoid reaching a sink state* ☹ *with positive probability. Now, the argument is the same as in Example 1: playing always the same action (either* a *or* b*) in* $\{q_1, q_2\}$ *is not even positive winning as player* 2 *can choose either* q_2 *or* q_1*.* ■

Note that our reduction preserves the structure and memory of almost-sure winning strategies, hence the non-elementary lower bound given in Theorem 3 for pure strategies also holds for randomized action-invisible strategies.

Corollary 3. *For one-sided partial-observation stochastic games, with player 1 partial and player 2 perfect, belief-based randomized action-invisible strategies are not sufficient for almost-sure winning for reachability objectives. For two-sided partial-observation stochastic games, memory of non-elementary size is necessary in general for almost-sure winning for randomized action-invisible strategies for reachability objectives.*

Acknowledgement. This work was partially supported by FWF Grant No P 23499-N23, FWF NFN Grant No S11407-N23 (RiSE), ERC Start grant (279307: Graph Games), and Microsoft faculty fellows award.

REFERENCES

[1] M. Abadi, L. Lamport, and P. Wolper. Realizable and unrealizable specifications of reactive systems. In *Proc. of ICALP*, LNCS 372, pages 1–17. Springer, 1989.
[2] R. Alur, T. A. Henzinger, and O. Kupferman. Alternating-time temporal logic. *Journal of the ACM*, 49:672–713, 2002.
[3] C. Baier, N. Bertrand, and M. Größer. On decision problems for probabilistic Büchi automata. In *Proc. of FoSSaCS*, LNCS 4962, pages 287–301. Springer, 2008.
[4] C. Baier, N. Bertrand, and M. Größer. The effect of tossing coins in omega-automata. In *Proc. of CONCUR*, LNCS 5710, pages 15–29. Springer, 2009.
[5] C. Baier and M. Größer. Recognizing omega-regular languages with probabilistic automata. In *Proc. of LICS*, pages 137–146, 2005.
[6] N. Bertrand, B. Genest, and H. Gimbert. Qualitative determinacy and decidability of stochastic games with signals. In *Proc. of LICS*, pages 319–328, 2009.
[7] D. Berwanger and L. Doyen. On the power of imperfect information. In *Proc. of FSTTCS*, Dagstuhl Seminar Proceedings 08004. IBFI, 2008.
[8] P. Cerný, K. Chatterjee, T. A. Henzinger, A. Radhakrishna, and R. Singh. Quantitative synthesis for concurrent programs. In *Proc. of CAV*, LNCS 6806, pages 243–259. Springer, 2011.
[9] R. Chadha, A. P. Sistla, and M. Viswanathan. On the expressiveness and complexity of randomization in finite state monitors. *Journal of the ACM*, 56:1–44, 2009.
[10] R. Chadha, A. P. Sistla, and M. Viswanathan. Power of randomization in automata on infinite strings. In *Proc. of CONCUR*, LNCS 5710, pages 229–243. Springer, 2009.
[11] R. Chadha, A. P. Sistla, and M. Viswanathan. Model checking concurrent programs with nondeterminism and randomization. In *Proc. of FSTTCS*, volume 8 of *LIPIcs*, pages 364–375, 2010.
[12] K. Chatterjee and L. Doyen. The complexity of partial-observation parity games. In *Proc. of LPAR*, LNCS 6397, pages 1–14. Springer, 2010.
[13] K. Chatterjee and L. Doyen. Partial-observation stochastic games: How to win when belief fails. *CoRR*, abs/1107.2141, July 2011.
[14] K. Chatterjee, L. Doyen, H. Gimbert, and T. A. Henzinger. Randomness for free. In *Proc. of MFCS*, LNCS 6281, pages 246–257. Springer, 2010.
[15] K. Chatterjee, L. Doyen, and T. A. Henzinger. Qualitative analysis of partially-observable Markov decision processes. In *Proc. of MFCS*, LNCS 6281, pages 258–269. Springer, 2010.
[16] K. Chatterjee, L. Doyen, T. A. Henzinger, and J.-F. Raskin. Algorithms for omega-regular games of incomplete information. *Logical Methods in Computer Science*, 3(3:4), 2007.
[17] A. Condon. The complexity of stochastic games. *Information and Computation*, 96(2):203–224, 1992.
[18] L. de Alfaro and T. A. Henzinger. Interface automata. In *Proc. of FSE*, pages 109–120. ACM Press, 2001.
[19] L. de Alfaro, T. A. Henzinger, and O. Kupferman. Concurrent reachability games. *Theor. Comput. Sci.*, 386(3):188–217, 2007.
[20] M. De Wulf, L. Doyen, and J.-F. Raskin. A lattice theory for solving games of imperfect information. In *Proc. of HSCC*, LNCS 3927, pages 153–168. Springer, 2006.
[21] D. L. Dill. *Trace Theory for Automatic Hierarchical Verification of Speed-independent Circuits*. The MIT Press, 1989.
[22] L. Doyen and J.-F. Raskin. Antichains algorithms for finite automata. In *Proc. of TACAS*, LNCS 6015, pages 2–22. Springer, 2010.
[23] E. A. Emerson and C. Jutla. Tree automata, mu-calculus and determinacy. In *Proc. of FOCS*, pages 368–377, 1991.
[24] H. Gimbert and Y. Oualhadj. Probabilistic automata on finite words: Decidable and undecidable problems. In *Proc. of ICALP (2)*, LNCS 6199, pages 527–538. Springer, 2010.
[25] V. Gripon and O. Serre. Qualitative concurrent stochastic games with imperfect information. In *Proc. of ICALP (2)*, LNCS 5556, pages 200–211. Springer, 2009.
[26] T. A. Henzinger and P. W. Kopke. Discrete-time control for rectangular hybrid automata. *Theor. Comp. Science*, 221:369–392, 1999.
[27] A. Kechris. *Classical Descriptive Set Theory*. Springer, 1995.
[28] D. König. *Theorie der endlichen und unendlichen Graphen*. Akademische Verlagsgesellschaft, Leipzig, 1936.
[29] H. Kress-Gazit, G. E. Fainekos, and G. J. Pappas. Temporal-logic-based reactive mission and motion planning. *IEEE Transactions on Robotics*, 25(6):1370–1381, 2009.
[30] O. Kupferman and M. Y. Vardi. Synthesis with incomplete informatio. In *Advances in Temporal Logic*, pages 109–127. Kluwer Academic Publishers, 2000.
[31] C. H. Papadimitriou and J. N. Tsitsiklis. The complexity of Markov decision processes. *Math. of Op. Research*, 12:441–450, 1987.
[32] A. Paz. *Introduction to probabilistic automata*. Academic Press, 1971.
[33] A. Pnueli and R. Rosner. On the synthesis of a reactive module. In *Proc. of POPL*, pages 179–190. ACM Press, 1989.
[34] M. O. Rabin. Probabilistic automata. *Inf. & Cont.*, 6:230–245, 1963.
[35] P. J. Ramadge and W. M. Wonham. Supervisory control of a class of discrete-event processes. *SIAM Journal of Control and Optimization*, 25(1):206–230, 1987.
[36] J. H. Reif. Universal games of incomplete information. In *Proc. of STOC*, pages 288–308. ACM, 1979.
[37] J. H. Reif. The complexity of two-player games of incomplete information. *JCSS*, 29:274–301, 1984.
[38] J. H. Reif and G. L. Peterson. A dynamic logic of multiprocessing with incomplete information. In *Proc. of POPL*, pages 193–202. ACM, 1980.
[39] D. Rosenberg, E. Solan, and N. Vieille. Stochastic games with a single controller and incomplete information. *SIAM J. Control and Optimization*, 43(1):86–110, 2004.
[40] L. S. Shapley. Stochastic games. *Proc. Nat. Acad. Sci. USA*, 39:1095–1100, 1953.
[41] S. Sorin. *A first course in zero-sum repeated games*. Springer, 2002.
[42] W. Thomas. Languages, automata, and logic. In *Handbook of Formal Languages*, volume 3, Beyond Words, chapter 7, pages 389–455. Springer, 1997.
[43] M. Tracol, C. Baier, and M. Größer. Recurrence and transience for probabilistic automata. In *Proc. of FSTTCS*, volume 4 of *LIPIcs*, pages 395–406, 2009.
[44] M. Y. Vardi. Automatic verification of probabilistic concurrent finite-state systems. In *Proc. of FOCS*, pages 327–338, 1985.
[45] M. T. Vechev, E. Yahav, and G. Yorsh. Inferring synchronization under limited observability. In *Proc. of TACAS*, LNCS 5505, pages 139–154. Springer, 2009.

Full Version

Partial-Observation Stochastic Games: How to Win when Belief Fails

Krishnendu Chatterjee (IST Austria) Laurent Doyen (LSV, ENS Cachan & CNRS, France)

Abstract

In two-player finite-state stochastic games of partial observation on graphs, in every state of the graph, the players simultaneously choose an action, and their joint actions determine a probability distribution over the successor states. The game is played for infinitely many rounds and thus the players construct an infinite path in the graph. We consider reachability objectives where the first player tries to ensure a target state to be visited almost-surely (i.e., with probability 1*) or positively (i.e., with positive probability), no matter the strategy of the second player.*

We classify such games according to the information and to the power of randomization available to the players. On the basis of information, the game can be one-sided *with either* (a) *player* 1*, or* (b) *player* 2 *having partial observation (and the other player has perfect observation), or* two-sided *with* (c) *both players having partial observation. On the basis of randomization,* (a) *the players may not be allowed to use randomization (pure strategies), or* (b) *they may choose a probability distribution over actions but the actual random choice is external and not visible to the player (actions invisible), or* (c) *they may use full randomization.*

Our main results for pure strategies are as follows: (1) For one-sided games with player 2 perfect observation we show that (in contrast to full randomized strategies) belief-based *(subset-construction based) strategies are not sufficient, and we present an exponential upper bound on memory both for almost-sure and positive winning strategies; we show that the problem of deciding the existence of almost-sure and positive winning strategies for player 1 is EXPTIME-complete and present symbolic algorithms that avoid the explicit exponential construction. (2) For one-sided games with player 1 perfect observation we show that non-elementary memory is both necessary and sufficient for both almost-sure and positive winning strategies. (3) We show that for the general (two-sided) case finite-memory strategies are sufficient for both positive and almost-sure winning, and at least non-elementary memory is required. We establish the equivalence of the almost-sure winning problems for pure strategies and for randomized strategies with actions invisible. Our equivalence result exhibit serious flaws in previous results in the literature: we show a non-elementary memory lower bound for almost-sure winning whereas an exponential upper bound was previously claimed.*

Keywords: Partial-observation games; Reachability objectives; Positive and Almost-sure winning.

1 Introduction

Games on graphs. Two-player games on graphs play a central role in several important problems in computer science, such as controller synthesis [35, 37], verification of open systems [2], realizability and compatibility checking [1, 22, 19], and many others. Most results about two-player games on graphs make the hypothesis of *perfect observation* (i.e., both players have perfect or complete observation about the state of the game). This assumption is often not realistic in practice. For example in the context of hybrid systems, the controller acquires information about the state of a plant using digital sensors with finite precision, which gives imperfect information about the state of the plant [21, 28]. Similarly, in a concurrent system where the players represent individual processes, each process has only access to the public variables of the other processes, not to their private variables [39, 2]. Such problems are better modeled in the more general framework of *partial-observation* games [38, 39, 40, 17, 7] and have been studied in the context of verification and synthesis [32, 23] (also see [3] for pushdown partial-observation games).

Partial-observation stochastic games and subclasses. In two-player partial-observation stochastic games on graphs with a finite state space, in every round, both players independently and simultaneously choose actions which along with the current state give a probability distribution over the successor states in the game. In a general setting, the players may not be able to distinguish certain states which are observationally equivalent for them (e.g., if they differ only by the value of private variables). The state space is partitioned into *observations* defined as equivalence classes and the players do not see the actual state of the game, but only an observation (which is typically different for the two players). The model of partial-observation games we consider is the same as the model of stochastic games with signals [7] and is a standard model in game theory [41, 43]. It subsumes other classical game models such as concurrent games [42, 20], probabilistic automata [36, 9, 34], and partial-observation Markov decision processes (POMDPs) [33] (see also the recent decidability and complexity results for probabilistic automata [4, 5, 6, 11, 12, 13, 26] and for POMDPs [16, 4, 45]).

The special case of *perfect observation* for a player corresponds to every observation for this player being a singleton. Depending on which player has perfect observation, we consider the following *one-sided* subclasses of the general two-sided partial-observation stochastic games: (1) *player* 1 *partial and player* 2 *perfect* where player 2 has perfect observation, and player 1 has partial observation; and (2) *player* 1 *perfect and player* 2 *partial* where player 1 has perfect observation, and player 2 has partial observation. The case where the two players have perfect observation corresponds to the well-known perfect-information (perfect-observation) stochastic games [42, 18, 20].

Note that in a given game G, if player 1 wins in the setting of player 1 partial and player 2 perfect, then player 1 wins in the game G as well. Analogously, if player 1 cannot win in the setting of player 1 perfect and player 2 partial, then player 1 does not win in the game G either. In this sense, the one-sided games are conservative over- and under-approximations of two-sided games. In the context of applications in verification and synthesis, the conservative approximation is that the adversary is all powerful, and hence player-1 partial and player-2 perfect games provide the important worst-case analysis of partial-observation games.

Objectives and qualitative problems. In this work we consider partial-observation stochastic games with *reachability* objectives where the goal of player 1 is to reach a set of target states, and games with *Büchi* objective where the goal for player 1 is to visit some target state infinitely often. The study of partial-observation games is considerably more complicated than games of perfect observation. For example, in contrast to perfect-observation games, strategies in partial-observation games require both randomization and memory for reachability objectives; and the *quantitative* problem of deciding whether there exists a strategy for player 1 to ensure that the target is reached with probability at least $\frac{1}{2}$ can be decided in NP $\cap$ coNP for perfect-observation stochastic games [18], whereas the problem is undecidable even for partial-observation stochastic games with only one player [34]. Since the quantitative problem is undecidable we consider the following *qualitative* problems: the *almost-sure* (resp. *positive*)

problem asks whether there exists a strategy for player 1 to ensure that the target set is reached with probability 1 (resp. positive probability). The qualitative problems for Büchi objectives are defined similarly where the goal is to visit the target set infinitely often with probability 1 (resp. positive probability) for the almost-sure (resp. positive) problem. For Büchi objectives, the positive problem is undecidable [4], and the almost-sure problem is polynomially equivalent to the almost-sure problem for reachability objective [4]. Therefore, we discuss reachability objectives, and the results for Büchi objective follow.

Classes of strategies. In general, randomized strategies are necessary to win with probability 1 in a partial-observation game with reachability objective [17]. However, there exist two types of randomized strategies where either (i) actions are visible, the player can observe the action he played [17, 7], or (ii) actions are invisible, the player may choose a probability distribution over actions, but the source of randomization is external and the actual choice of the action is invisible to the player [27]. The second model is more general since the qualitative problems of randomized strategies with actions visible can be reduced in polynomial time to randomized strategies with actions invisible, by modeling the visibility of actions using the observations on states.

With actions visible, the almost-sure (resp. positive) problem was shown to be EXPTIME-complete (resp. PTIME-complete) for one-sided games with player 1 partial and player 2 perfect [17], and 2EXPTIME-complete (resp. EXPTIME-complete) in the two-sided case [7]. For the positive problem memoryless randomized strategies exist, and for the almost-sure problem *belief-based* strategies exist (strategies based on subset construction that consider the possible current states of the game).

It was remarked (without any proof) in [17, p.4] that these results easily extend to randomized strategies with actions invisible for one-sided games with player 1 partial and player 2 perfect. It was claimed in [27] (Theorems 1 & 2) that the almost-sure problem is 2EXPTIME-complete for randomized strategies with actions invisible for two-sided games, and that belief-based strategies are sufficient for player 1. Thus it is believed that the two qualitative problems with actions visible or actions invisible are essentially equivalent.

In this paper, we consider the class of *pure* strategies, which do not use randomization at all. Pure strategies arise naturally in the synthesis of controllers and processes that do not have access to any source of randomization, such as synchronizers for lock placement in concurrent programs [10], and controllers for robot planning [31]. Moreover we will establish deep connections between the qualitative problems for pure strategies and for randomized strategies with actions invisible, which on one hand exhibit major flaws in previous results of the literature (the remark without proof of [17] and the main results of [27]), and on the other hand show that the solution for almost-sure winning randomized strategies with actions invisible (which is the most general case) can be surprisingly obtained by solving the problem for pure strategies.

Contributions. The contributions of the paper are summarized below.

1. *Player* 1 *partial and player* 2 *perfect.* We show that both for almost-sure and positive winning, belief-based pure strategies are not sufficient. This implies that the classical approaches relying on the belief-based subset construction cannot work for solving the qualitative problems for pure strategies. However, we present an optimal exponential upper bound on the memory needed by pure strategies (the exponential lower bound follows from the special case of non-stochastic games [8]). By a reduction to a perfect-observation game of exponential size, we show that both the almost-sure and positive problems are EXPTIME-complete for one-sided games with perfect-observation for player 2. In contrast to the previous proofs of EXPTIME upper bound that rely either on subset constructions or enumeration of belief-based strategies, our correctness proof relies on a novel rank-based argument that works uniformly both for positive and almost-sure winning. The structure of this construction also provides symbolic antichain-based algorithms (see [24] for a survey of the antichain approach) for solving the qualitative problems that avoids the explicit exponential construction. Thus for the important special case of player 1 partial and player 2 perfect we establish optimal memory bound, complexity bound, and present symbolic algorithmic solutions for the qualitative problems.
2. *Player* 1 *perfect and player* 2 *partial.*

(a) We show a very surprising result that both for positive and almost-sure winning, pure strategies for player 1 require memory of non-elementary size (i.e., a tower of exponentials). This is in sharp contrast with (i) the case of randomized strategies (with or without actions visible) where memoryless strategies are sufficient for positive winning, and with (ii) the previous case where player 1 has partial observation and player 2 has perfect observation, where pure strategies for positive winning require only exponential memory. Surprisingly and perhaps counter-intuitively when player 1 has more information and player 2 has less information, the positive winning strategies for player 1 require much more memory (non-elementary as compared to exponential). With more information player 1 can win from more states, but the winning strategy is much harder to implement.

(b) We present a non-elementary upper bound for the memory needed by pure strategies for positive winning. We then show with an example that for almost-sure winning more memory may be required as compared to positive winning. Finally, we show how to combine pure strategies for positive winning in a recharging scheme to obtain a non-elementary upper bound for the memory required by pure strategies for almost-sure winning. Thus we establish non-elementary complete bounds for pure strategies both for positive and almost-sure winning.

3. *General (two-sided) case.* We show that in the general case finite memory strategies are sufficient both for positive and almost-sure winning. The result is obtained essentially by a simple generalization of König's Lemma [30]. The non-elementary lower bound for memory follows from the special case when player 1 has perfect observation and player 2 has partial observation.

4. *Randomized strategies with actions invisible.* For randomized strategies with actions invisible we present two reductions to establish connections with pure strategies. First, we show that the almost-sure problem for randomized strategies with actions invisible can be reduced in polynomial time to the almost-sure problem for pure strategies. The reduction requires to first establish that finite-memory randomized strategies are sufficient in two-sided games. Second, we show that the problem of almost-sure winning with pure strategies can be reduced in polynomial time to the problem of randomized strategies with actions invisible. For this reduction it is crucial that the actions are not visible.

   Our reductions have deep consequences. They unexpectedly imply that the problems of almost-sure winning with *pure* strategies or *randomized* strategies with actions invisible are polynomial-time *equivalent*. Moreover, it follows that even in one-sided games with player 1 partial and player 2 perfect, belief-based randomized strategies with actions invisible are not sufficient for almost-sure winning. This shows that the remark (without proof) of [17] that the results (such as existence of belief-based strategies) of randomized strategies with actions visible carry over to actions invisible is an oversight. However from our first reduction and our results for pure strategies it follows that there is an exponential upper bound on memory and the problem is EXPTIME-complete for one-sided games with player 1 partial and player 2 perfect. More importantly, our results exhibit a serious flaw in the main result of [27] which showed that belief-based randomized strategies with actions invisible are sufficient for almost-sure winning in two-sided games, and concluded that enumerating over such strategies yields a 2EXPTIME algorithm for the problem. Our second reduction and lower bound for pure strategies show that the result is incorrect, and that the exponential (belief-based) upper bound is far off. Instead, the lower bound on memory for almost-sure winning with randomized strategies and actions invisible is non-elementary. Thus, contrary to the general belief, there is a sharp contrast for randomized strategies with or without actions visible: if actions are visible, then exponential memory is sufficient for almost-sure winning while if actions are not visible, then memory of non-elementary size is necessary in general.

The memory requirements are summarized in Table 1 and the results of this paper are shown in bold font. We explain how the other results of the table follow from results of the literature. For randomized strategies (with or without actions visible), if a positive winning strategy exists, then a memoryless strategy that plays all actions

uniformly at random is also positive winning. Thus the memoryless result for positive winning strategies follows for all cases of randomized strategies. The belief-based bound for memory of almost-sure winning randomized strategies with actions visible follows from [17, 7]. The memoryless strategies results for almost-sure winning for one-sided games with player 1 perfect and player 2 partial are obtained as follows: when actions are visible, then belief-based strategies coincide with memoryless strategies as player 1 has perfect observation. If player 1 has perfect observation, then for memoryless strategies whether actions are visible or not is irrelevant and thus the memoryless result also follows for randomized strategies with actions invisible. Thus along with our results we obtain Table 1.

	one-sided player 2 perfect		one-sided player 1 perfect		two-sided	
	Positive	Almost-sure	Positive	Almost-sure	Positive	Almost-sure
Randomized (actions visible)	Memoryless	Exponential (belief-based)	Memoryless	Memoryless	Memoryless	Exponential (belief-based)
Randomized (actions invisible)	Memoryless	**Exponential (belief is not sufficient)**	Memoryless	Memoryless	Memoryless	**Non-elem. low. bound** **Finite upp. bound**
Pure	**Exponential (belief is not sufficient)**	**Exponential (belief is not sufficient)**	**Non-elem. complete**	**Non-elem. complete**	**Non-elem. low. bound** **Finite upp. bound**	**Non-elem. low. bound** **Finite upp. bound**

Table 1. Memory requirement for player 1 and reachability objective.

2 Definitions

A *probability distribution* on a finite set S is a function $\kappa : S \to [0, 1]$ such that $\sum_{s \in S} \kappa(s) = 1$. The *support* of κ is the set $\mathsf{Supp}(\kappa) = \{s \in S \mid \kappa(s) > 0\}$. We denote by $\mathcal{D}(S)$ the set of probability distributions on S. Given $s \in S$, the *Dirac distribution* on s assigns probability 1 to s.

Games. Given finite alphabets A_i of actions for player i ($i = 1, 2$), a *stochastic game* on A_1, A_2 is a tuple $G = \langle Q, q_0, \delta \rangle$ where Q is a finite set of states, $q_0 \in Q$ is the initial state, and $\delta : Q \times A_1 \times A_2 \to \mathcal{D}(Q)$ is a probabilistic transition function that, given a current state q and actions a, b for the players gives the transition probability $\delta(q, a, b)(q')$ to the next state q'. The game is called *deterministic* if $\delta(q, a, b)$ is a Dirac distribution for all $(q, a, b) \in Q \times A_1 \times A_2$. A state q is *absorbing* if $\delta(q, a, b)$ is the Dirac distribution on q for all $(a, b) \in A_1 \times A_2$. In some examples, we allow an initial distribution of states. This can be encoded in our game model by a probabilistic transition from the initial state.

A *player-*1 *state* is a state q where $\delta(q, a, b) = \delta(q, a, b')$ for all $a \in A_1$ and all $b, b' \in A_2$. We use the notation $\delta(q, a, -)$. *Player-*2 *states* are defined analogously. In figures, we use boxes to emphasize that a state is a player-2 state, and we represent probabilistic branches using diamonds (which are not real 'states', e.g., as in Figure 1).

In a (two-sided) *partial-observation* game, the players have a partial or incomplete view of the states visited and of the actions played in the game. This view may be different for the two players and it is defined by equivalence relations $\approx_i$ on the states and on the actions. For player i, equivalent states (or actions) are indistinguishable. We denote by $\mathcal{O}_i \subseteq 2^Q$ ($i = 1, 2$) the equivalence classes of $\approx_i$ which define two partitions of the state space Q, and we call them *observations* (for player i). These partitions uniquely define functions $\mathsf{obs}_i : Q \to \mathcal{O}_i$ ($i = 1, 2$) such that $q \in \mathsf{obs}_i(q)$ for all $q \in Q$, that map each state q to its observation for player i.

In the case where all states and actions are equivalent (i.e., the relation $\approx_i$ is the set $(Q \times Q) \cup (A_1 \times A_1) \cup (A_2 \times A_2)$), we say that player i is *blind* and the actions are *invisible*. In this case, we have $\mathcal{O}_i = \{Q\}$ because

all states have the same observation. Note that the case of perfect observation for player i corresponds to the case $\mathcal{O}_i = \{\{q_0\}, \{q_1\}, \ldots, \{q_n\}\}$ (given $Q = \{q_0, q_1, \ldots, q_n\}$), and $a \approx_i b$ iff $a = b$, for all actions a, b.

For $s \subseteq Q$, $a \in A_1$, and $b \in A_2$, let $\mathsf{Post}_{a,b}(s) = \bigcup_{q \in s} \mathsf{Supp}(\delta(q, a, b))$ denote the set of possible successors of q given action a and b, and let $\mathsf{Post}_{a,-}(s) = \bigcup_{b \in A_2} \mathsf{Post}_{a,b}(s)$.

Plays and observations. Initially, the game starts in the initial state q_0. In each round, player 1 chooses an action $a \in A_1$, player 2 (simultaneously and independently) chooses an action $b \in A_2$, and the successor of the current state q is chosen according to the probabilistic transition function $\delta(q, a, b)$. A *play* in G is an infinite sequence $\rho = q_0 a_0 b_0 q_1 a_1 b_1 q_2 \ldots$ such that q_0 is the initial state and $\delta(q_j, a_j, b_j)(q_{j+1}) > 0$ for all $j \geq 0$ (the actions a_j's and b_j's are the actions *associated* to the play). Its *length* is $|\rho| = \infty$. The length of a play prefix $\rho = q_0 a_0 b_0 q_1 \ldots q_k$ is $|\rho| = k$, and its last element is $\mathsf{Last}(\rho) = q_k$. A state $q \in Q$ is *reachable* if it occurs in some play. We denote by $\mathsf{Plays}(G)$ the set of plays in G, and by $\mathsf{Prefs}(G)$ the set of corresponding finite prefixes. The *observation sequence* for player i ($i = 1, 2$) of a play (prefix) ρ is the unique (in)finite sequence $\mathsf{obs}_i(\rho) = \gamma_0 \gamma_1 \ldots$ such that $q_j \in \gamma_j \in \mathcal{O}_i$ for all $0 \leq j \leq |\rho|$.

The games with *one-sided partial-observation* are the special case where either $\approx_1$ is equality and hence $\mathcal{O}_1 = \{\{q\} \mid q \in Q\}$ (player 1 has complete observation) or $\approx_2$ is equality and hence $\mathcal{O}_2 = \{\{q\} \mid q \in Q\}$ (player 2 has complete observation). The games with *perfect observation* are the special cases where $\approx_1$ and $\approx_2$ are equality, i.e., every state and action is visible to both players.

Strategies. A *pure strategy* in G for player 1 is a function $\sigma : \mathsf{Prefs}(G) \to A_1$. A *randomized strategy* in G for player 1 is a function $\sigma : \mathsf{Prefs}(G) \to \mathcal{D}(A_1)$. A (pure or randomized) strategy σ for player 1 is *observation-based* if for all prefixes $\rho = q_0 a_0 b_0 q_1 \ldots$ and $\rho' = q'_0 a'_0 b'_0 q'_1 \ldots$, if $a_j \approx_1 a'_j$ and $b_j \approx_1 b'_j$ for all $j \geq 0$, and $\mathsf{obs}_1(\rho) = \mathsf{obs}_1(\rho')$, then $\sigma(\rho) = \sigma(\rho')$. It is assumed that strategies are observation-based in partial-observation games. If for all actions a and b we have $a \approx_1 b$ and $a \approx_2 b$ iff $a = b$ (all actions are distinguishable), then the strategy is *action visible*, and if for all actions a and b we have $a \approx_1 b$ and $a \approx_2 b$ (all actions are indistinguishable), then the strategy is *action invisible*. We say that a play (prefix) $\rho = q_0 a_0 b_0 q_1 \ldots$ is *compatible* with a pure (resp., randomized) strategy σ if the associated action of player 1 in step j is $a_j = \sigma(q_0 a_0 b_0 \ldots q_{j-1})$ (resp., $a_j \in \mathsf{Supp}(\sigma(q_0 a_0 b_0 \ldots q_{j-1}))$) for all $0 \leq j \leq |\rho|$.

We omit analogous definitions of strategies for player 2. We denote by Σ_G, Σ_G^O, Σ_G^P, Π_G, Π_G^O, and Π_G^P the set of all player-1 strategies, the set of all observation-based player-1 strategies, the set of all pure player-1 strategies, the set of all player-2 strategies in G, the set of all observation-based player-2 strategies, and the set of all pure player-2 strategies, respectively.

Remark 1. *The model of games with partial observation on both actions and states can be encoded in a model of games with actions invisible and observations on states only: when actions are invisible, we can use the state space to keep track of the last action played, and reveal information about the last action played using observations on the states [27]. Therefore, in the sequel we assume that the actions are invisible to the players with partial observation. A play is then viewed as a sequence of states only, and the definition of strategies is updated accordingly. Note that a player with perfect observation has actions and states visible (and the equivalence relation $\approx_i$ is equality).*

Remark 2. *The important special case of partial-observation Markov decision processes (POMDP) corresponds to the case where either all states in the game are player-*1 *states (player-1 POMDP) or all states are player-*2 *states (player-2 POMDP). For POMDP it is known that randomization is not necessary, and pure strategies are as powerful as randomized strategies [15].*

Finite-memory strategies. A player-1 strategy uses *finite-memory* if it can be encoded by a deterministic transducer $\langle \mathsf{Mem}, m_0, \alpha_u, \alpha_n \rangle$ where Mem is a finite set (the memory of the strategy), $m_0 \in \mathsf{Mem}$ is the initial memory value, $\alpha_u : \mathsf{Mem} \times \mathcal{O}_1 \to \mathsf{Mem}$ is an update function, and $\alpha_n : \mathsf{Mem} \times \mathcal{O}_1 \to \mathcal{D}(A_1)$ is a next-move function. The *size* of the strategy is the number $|\mathsf{Mem}|$ of memory values. If the current observation is o, and the current

memory value is m, then the strategy chooses the next action according to the probability distribution $\alpha_n(m, o)$, and the memory is updated to $\alpha_u(m, o)$. Formally, $\langle \mathsf{Mem}, m_0, \alpha_u, \alpha_n \rangle$ defines the strategy σ such that $\sigma(\rho \cdot q) = \alpha_n(\hat{\alpha}_u(m_0, \mathsf{obs}_1(\rho)), \mathsf{obs}_1(q))$ for all $\rho \in Q^*$ and $q \in Q$, where $\hat{\alpha}_u$ extends α_u to sequences of observations as expected. This definition extends to infinite-memory strategies by dropping the assumption that the set Mem is finite. A strategy is *memoryless* if $|\mathsf{Mem}| = 1$. For a strategy σ, we denote by G_σ the player-2 POMDP obtained as the synchronous product of G with the transducer defining σ.

Objectives and winning modes. An *objective* (for player 1) in G is a set $\phi \subseteq \mathsf{Plays}(G)$ of plays. A play $\rho \in \mathsf{Plays}(G)$ *satisfies* the objective ϕ, denoted $\rho \models \phi$, if $\rho \in \phi$. Objectives are generally Borel measurable: a Borel objective is a Borel set in the Cantor topology [29]. Given strategies σ and π for the two players, the probabilities of a measurable objective ϕ is uniquely defined [46]. We denote by $\Pr_{q_0}^{\sigma,\pi}(\phi)$ the probability that ϕ is satisfied by the play obtained from the starting state q_0 when the strategies σ and π are used.

We specifically consider the following objectives. Given a set $\mathcal{T} \subseteq Q$ of target states, the *reachability objective* requires that the play visit the set $\mathcal{T}$: $\mathsf{Reach}(\mathcal{T}) = \{q_0 a_0 b_0 q_1 \ldots \in \mathsf{Plays}(G) \mid \exists i \geq 0 : q_i \in \mathcal{T}\}$, and the *Büchi objective* requires that the play visit the set $\mathcal{T}$ infinitely often, $\mathsf{Büchi}(\mathcal{T}) = \{q_0 a_0 b_0 q_1 \ldots \in \mathsf{Plays}(G) \mid \forall i \geq 0 \cdot \exists j \geq i : q_j \in \mathcal{T}\}$. Our solution for reachability objectives will also use the dual notion of *safety objectives* that require the play to stay within the set $\mathcal{T}$: $\mathsf{Safe}(\mathcal{T}) = \{q_0 a_0 b_0 q_1 \ldots \in \mathsf{Plays}(G) \mid \forall i \geq 0 : q_i \in \mathcal{T}\}$. In figures, the target states in $\mathcal{T}$ are double-lined and labeled by ☺.

Given a game structure G and a state q, an observation-based strategy σ for player 1 is *almost-sure winning* (resp. *positive winning*) for the objective ϕ from q if for all observation-based randomized strategies π for player 2, we have $\Pr_q^{\sigma,\pi}(\phi) = 1$ (resp. $\Pr_q^{\sigma,\pi}(\phi) > 0$). The strategy σ is *sure winning* if all plays compatible with σ satisfy ϕ. We also say that the state q is almost-sure (or positive, or sure) winning for player 1.

Positive and almost-sure winning problems. We are interested in the problems of deciding, given a game structure G, a state q, and an objective ϕ, whether there exists a $\{$pure, randomized$\}$ strategy which is $\{$almost-sure, positive$\}$ winning from q for the objective ϕ. For safety objectives almost-sure winning coincides with sure winning, however for reachability objectives they are different. The sure winning problem for the objectives we consider has been studied in [38, 17, 14]. The almost-sure winning problem for Büchi objectives can be easily reduced to the almost-sure winning problem for reachability objectives [4]. The positive winning problem for Büchi objectives is undecidable even for POMDPs [4]. Hence in this paper we only focus on reachability objectives. In all our analysis, the counter strategies of player 2 can be restricted to pure strategies, because once a strategy for player 1 is fixed, then we obtain a POMDP for player 2 in which pure strategies are as powerful as randomized strategies [15].

Remark 3. *(Almost-sure Büchi to almost-sure reachability [4]). The reduction of almost-sure Büchi to almost-sure reachability is as follows: given a two-sided stochastic game with Büchi objective* $\mathsf{Büchi}(\mathcal{T})$*, we add a new absorbing state* q_T*, make* q_T *the target state for the reachability objective, and from every state* $q \in \mathcal{T}$ *we add positive probability transitions to* q_T *(details and correctness proof follow from [4, Lemma 13]). The key idea of the correctness of the reduction is as follows. If in the original game, Büchi states are visited infinitely often almost-surely, then the new target state is reached almost-surely (due to positive transition probability from the original Büchi states to the new target state). Conversely, if in the original game, Büchi states are visited infinitely often with probability less than 1, then since the only way to reach the new target state in the reduced game is through the Büchi states, it follows that the target state is reached with probability less than 1. This holds for any pair of strategies, and establishes the reduction.*

3 One-sided Games: Player 1 Partial and Player 2 Perfect

In Sections 3 and 4, we consider one-sided games with partial observation: one player has perfect observation, and the other player has partial observation. The player with perfect observation sees the states visited and the

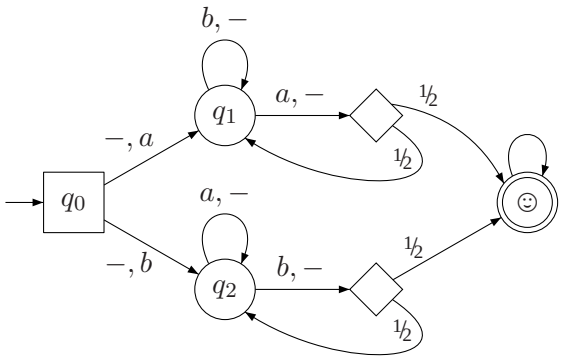

Figure 1. Belief-only is not enough for positive (as well as almost-sure) reachability. A one-sided reachability game with reachability objective in which player 1 is blind and player 2 has perfect observation. If we consider pure strategies, then player 1 has a positive (as well as almost-sure) winning strategy, but there is no belief-based memoryless positive winning strategy.

actions played in the game. We present the results for positive and almost-sure winning for reachability objectives along with examples that illustrate key elements of the problem such as the memory required for winning strategies.

Note that the case of player 1 partial and player 2 perfect is important in the context of controller synthesis as it is a conservative approximation of two-sided games for player 1 (if player 1 wins in the one-sided game, then he also wins in the two-sided game). In the following example we show that for pure strategies *belief-based* strategies are not sufficient for positive as well as almost-sure winning. A strategy is belief-based if its memory relies only on the subset construction, i.e., the strategy plays only depending on the set of possible current states of the game which is called *belief*.

Example 1. Belief-only is not enough for positive (as well as almost-sure) reachability. *Consider the game in Figure 1 where player 1 is blind (all states have the same observation except the target state, and actions are invisible) and player 2 has perfect observation. Initially, player 2 chooses the state q_1 or q_2 (which player 1 does not see). The belief of player 1 is thus the set $\{q_1, q_2\}$ (see Figure 2). We claim that the belief is not a sufficient information to win with a pure strategy for player 1 because the belief-based subset construction in Figure 2 suggests that playing always the same action (say a) when the belief is $\{q_1, q_2\}$ is an almost-sure winning strategy. However, in the original game this is not even a positive winning strategy (the counter strategy of player 2 is to choose q_2 initially). A winning strategy for player 1 is to alternate between a and b when the belief is $\{q_1, q_2\}$, showing that remembering the belief is not sufficient.* ■

We present reductions of the almost-sure and positive winning problem for reachability objective to the problem of sure-winning in a game of perfect observation with Büchi objective, and reachability objective respectively. The two reductions are based on the same construction of a game where the state space $L = \{(s, o) \mid o \subseteq s \subseteq Q\}$ contains the subset construction s enriched with *obligation sets* $o \subseteq s$ which ensure that from all states in s, the target set $\mathcal{T}$ is reached with positive probability.

Lemma 1. *Given a one-sided partial-observation stochastic game G with player 1 partial and player 2 perfect with a reachability objective for player 1, we can construct in time exponential in the size of the game and polynomial in the size of action sets a perfect-information deterministic game H with a Büchi objective (resp. reachability objective) such that player 1 has a pure almost-sure (resp. positive) winning strategy in G iff player 1*

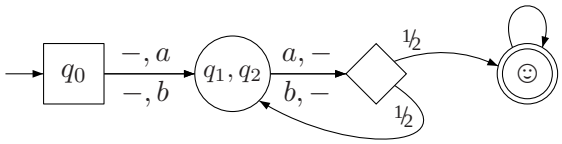

Figure 2. The belief-based subset construction for the reachability game of Figure 1. Player 1 has a pure strategy for positive (as well as almost-sure) winning in the subset construction. However, belief-based memoryless pure strategies are not sufficient in the original game.

has a sure-winning strategy in H.

Proof. We present the construction and the proof in details for almost-sure reachability. The construction is the same for positive reachability, and the argument is described succinctly afterwards.

Construction. Given $G = \langle Q, q_0, \delta \rangle$ over alphabets A_1, A_2 and observation set $\mathcal{O}_1$ for player 1, with reachability objective $\mathsf{Reach}(\mathcal{T})$, we construct the following (deterministic) game of perfect observation $H = \langle L, \ell_0, \delta_H \rangle$ over alphabets A_1', A_2' with Büchi objective $\mathsf{Büchi}(\alpha)$ defined by $\alpha \subseteq L$ where:

- $L = \{(s, o) \mid o \subseteq s \subseteq Q\}$. Intuitively, s is the belief of player 1 and o is a set of obligation states that "owe" a visit to $\mathcal{T}$ with positive probability.
- $\ell_0 = (\{q_0\}, \{q_0\})$ if $q_0 \notin \mathcal{T}$, and $\ell_0 = (\emptyset, \emptyset)$ if $q_0 \in \mathcal{T}$;
- $A_1' = A_1 \times 2^Q$. In a pair $(a, u) \in A_1'$, we call a the action, and u the witness set;
- $A_2' = \mathcal{O}_1$. In the game H, player 2 simulate player 2's choice in game G, as well as resolves the probabilistic choices. This amounts to choosing a possible successor state, and revealing its observation;
- $\alpha = \{(s, \emptyset) \in L\}$;
- δ_H is defined as follows. First, the state $(\emptyset, \emptyset)$ is absorbing. Second, in every other state $(s, o) \in L$ the function δ_H ensures that (i) player 1 chooses a pair (a, u) such that $\mathsf{Supp}(\delta(q, a, b)) \cap u \neq \emptyset$ for all $q \in o$ and $b \in A_2$, and (ii) player 2 chooses an observation $\gamma \in \mathcal{O}_1$ such that $\mathsf{Post}_{a,-}(s) \cap \gamma \neq \emptyset$. If a player violates this, then a losing absorbing state is reached with probability 1. Assuming the above condition on (a, u) and γ is satisfied, define $\delta_H((s, o), (a, u), \gamma)$ as the Dirac distribution on the state (s', o') such that:
  - $s' = (\mathsf{Post}_{a,-}(s) \cap \gamma) \setminus \mathcal{T}$;
  - $o' = s'$ if $o = \emptyset$; and $o' = (\mathsf{Post}_{a,-}(o) \cap \gamma \cap u) \setminus \mathcal{T}$ if $o \neq \emptyset$.

Note that for every reachable state (s, o) in H, there exists a unique observation $\gamma \in \mathcal{O}_1$ such that $s \subseteq \gamma$ (which we denote by $\mathsf{obs}_1(s)$).

We show the following property of this construction. Player 1 has a pure observation-based almost-sure winning strategy in G for the objective $\mathsf{Reach}(\mathcal{T})$ if and only if player 1 has a sure winning strategy in H for the objective $\mathsf{Büchi}(\alpha)$.

Mapping of plays. Given a play prefix $\rho_H = (s_0, o_0)(s_1, o_1) \dots (s_k, o_k)$ in H with associated actions for player 1 of the form $(a_i, \cdot)$ in step i $(0 \leq i < k)$, and a play prefix $\rho_G = q_0 q_1 \dots q_k$ in G with associated actions a_i' $(0 \leq i < k)$ for player 1, we say that ρ_G is *matching* ρ_H if $a_i = a_i'$ for all $0 \leq i < k$, and $q_i \in \mathsf{obs}_1(s_i)$ for all $0 \leq i \leq k$.

By induction on the length of ρ_H, we show that (i) for each $q_k \in s_k$ there exists a matching play ρ_G (which visits no $\mathcal{T}$-state) such that $\mathsf{Last}(\rho_G) = q_k$, and (ii) for all play prefixes ρ_G matching ρ_H, if ρ_G does not visit any $\mathcal{T}$-state, then $\mathsf{Last}(\rho_G) \in s_k$.

For $|\rho_H| = 0$ (i.e., $\rho_H = (s_0, o_0)$ where $(s_0, o_0) = \ell_0$) it is easy to see that $\rho_G = q_0$ is a matching play with $q_0 \notin \mathcal{T}$ if and only if $s_0 = o_0 = \{q_0\}$. For the induction step, assume that we have constructed matching plays for all play prefixes of length $k-1$, and let $\rho_H = (s_0, o_0)(s_1, o_1)\dots(s_k, o_k)$ be a play prefix of length k in H with associated actions of the form $(a_i, \cdot)$ in step i ($0 \leq i < k$). To prove (i), pick $q_k \in s_k$. By definition of δ_H, we have $q_k \in \mathsf{Post}_{a_{k-1},-}(s_{k-1})$, hence there exists $b \in A_2$ and $q_{k-1} \in s_{k-1}$ such that $q_k \in \mathsf{Supp}(\delta(q_{k-1}, a_{k-1}, b))$. By induction hypothesis, there exists a play prefix ρ_G in G matching $(s_0, o_0)\dots(s_{k-1}, o_{k-1})$ and with $\mathsf{Last}(\rho_G) = q_{k-1}$, which we can extend to $\rho_G.q_k$ to obtain a play prefix matching ρ_H. To prove (ii), it is easy to see that every play prefix matching ρ_H is an extension of play prefix matching $(s_0, o_0)\dots(s_{k-1}, o_{k-1})$ with a non $\mathcal{T}$-state q_k in $\gamma_k = \mathsf{obs}_1(s_k)$ and in $\mathsf{Post}_{a_{k-1},-}(s_{k-1})$, therefore $q_k \in (\mathsf{Post}_{a_{k-1},-}(s_{k-1}) \cap \gamma_k) \setminus \mathcal{T} = s_k$.

Mapping of strategies, from G to H (ranking argument). First, assume that player 1 has a pure observation-based almost-sure winning strategy σ in G for the objective $\mathsf{Reach}(\mathcal{T})$. We construct an infinite-state MDP $G_\sigma = \langle Q^+, \rho_0, \delta_\sigma \rangle$ where:

- Q^+ is the set of nonempty finite sequences of states;
- $\rho_0 = q_0 \in Q$;
- $\delta_\sigma : Q^+ \times A_2 \to \mathcal{D}(Q^+)$ is defined as follows: for each $\rho \in Q^+$ and $b \in A_2$, if $\mathsf{Last}(\rho) \notin \mathcal{T}$ then $\delta_\sigma(\rho, b)$ assigns probability $\delta(\mathsf{Last}(\rho), \sigma(\rho), b)(q')$ to each $\rho' = \rho q' \in Q^+$, and probability 0 to all other $\rho' \in Q^+$; if $\mathsf{Last}(\rho) \in \mathcal{T}$, then ρ is an absorbing state;

We define a ranking of the reachable states of G_σ. Assign rank 0 to all $\rho \in Q^+$ such that $\mathsf{Last}(\rho) \in \mathcal{T}$. For $i = 1, 2, \dots$ assign rank i to all non-ranked ρ such that for all player 2 actions $b \in A_2$, there exists $\rho' \in \mathsf{Supp}(\delta_\sigma(\rho, b))$ with a rank (and thus with a rank smaller than i). We claim that all reachable states of G_σ get a rank. By contradiction, assume that a reachable state $\hat{\rho} = q_0 q_1 \dots q_k$ is not ranked (note that $q_i \notin \mathcal{T}$ for each $0 \leq i \leq k$). Fix a strategy π for player 2 as follows. Since $\hat{\rho}$ is reachable in G_σ, there exist actions $b_0, \dots, b_{k-1}$ such that $q_{i+1} \in \mathsf{Supp}(\delta_\sigma(q_0 \dots q_i, b_i))$ for all $0 \leq i < k$. Then, define $\pi(q_0 \dots q_i) = b_i$. This ensures that $\mathsf{Last}(\hat{\rho})$ is reached with positive probability in G under strategies σ and π. From $\hat{\rho}$, the strategy π continues playing as follows. If the current state ρ is not ranked (which is the case of $\hat{\rho}$), then choose an action b such that all states in $\mathsf{Supp}(\delta_\sigma(\rho, b))$ are not ranked. The fact that ρ is not ranked ensures that such an action b exists. Now, under σ and π all paths from $\mathsf{Last}(\hat{\rho})$ in G avoid $\mathcal{T}$-sates. Hence the set $\mathcal{T}$ is not reached almost-surely, in contradiction with the fact that σ is almost-sure winning. Hence all states in G_σ get a rank. We denote by $\mathsf{Rank}(\rho)$ the rank of a reachable state ρ in G_σ.

From the strategy σ and the ranking in G_σ, we construct a strategy σ' in the game H as follows. Given a play $\rho_H = (s_0, o_0)(s_1, o_1)\dots(s_k, o_k)$ in H (with $s_k \neq \emptyset$), define $\sigma'(\rho_H) = (a, u)$ where $a = \sigma(\rho_G)$ for a play prefix ρ_G matching ρ_H and $u = \{q \in \mathsf{Supp}(\delta(\mathsf{Last}(\rho_G), a, b)) \mid b \in A_2, \rho_G \text{ is matching } \rho_H \text{ with } \mathsf{Last}(\rho_G) \in o_k \text{ and } \mathsf{Rank}(\rho_G.q) < \mathsf{Rank}(\rho_G)\}$ is a witness set which selects successor states of o_k with decreased rank along each branch of the MDP G_σ.

Note that all matching play prefixes ρ_G have the same observation sequence. Therefore, the action $a = \sigma(\rho_G)$ is unique and well-defined since σ is an observation-based strategy. Note also that the pair (a, u) is an allowed choice for player 1 by definition of the ranking, and that for each $q \in o_k$, all matching play prefixes ρ_G with $\mathsf{Last}(\rho_G) = q$ have the same rank in G_σ. Therefore we abuse notation and write $\mathsf{Rank}(q)$ for $\mathsf{Rank}(\rho_G)$, assuming that the set o_k to which q belongs is clear from the context. Let $\mathsf{MaxRank}(o_k) = \max_{q \in o_k} \mathsf{Rank}(q)$. If $o_k \neq \emptyset$, then $\mathsf{MaxRank}(o_{k+1}) < \mathsf{MaxRank}(o_k)$ since $o_{k+1} \subseteq u$ (by definition of δ_H).

Correctness of the mapping. We show that σ' is sure winning for $\mathsf{Büchi}(\alpha)$ in H. Fix an arbitrary strategy π' for player 2 in H and consider an arbitrary play $\rho_H = (s_0, o_0)(s_1, o_1)\dots$ compatible with σ' and π'. By the properties of the witness set played by σ', for each pair (s_i, o_i) with $o_i \neq \emptyset$, an α-pair $(\cdot, \emptyset)$ is reached within at

most $\mathsf{MaxRank}(o_i)$ steps. And by the properties of the mapping of plays and strategies, if $o_i = \emptyset$ then $o_{i+1} = s_{i+1}$ contains only states from which σ is almost-sure winning for $\mathsf{Reach}(\mathcal{T})$ in G and therefore have a finite rank, showing that $\mathsf{MaxRank}(o_{i+1})$ is defined and finite. This shows that an α-pair is visited infinitely often in ρ_H and σ' is sure winning for $\mathsf{Büchi}(\alpha)$.

Mapping of strategies, from H to G. Given a strategy σ' in H, we construct a pure observation-based strategy σ in G.

We define $\sigma(\rho_G)$ by induction on the length of ρ_G. In fact, we need to define $\sigma(\rho_G)$ only for play prefixes ρ_G which are compatible with the choices of σ for play prefixes of length smaller than $|\rho_G|$ (the choice of σ for other play prefixes can be fixed arbitrarily). For all such ρ_G, our construction is such that there exists a play prefix $\rho_H = \theta(\rho_G)$ compatible with σ' such that ρ_G is matching ρ_H, and if $\sigma(\rho_G) = a$ and $\sigma'(\rho_H) = (a', \cdot)$, then $a = a'$ $(\star)$.

We define σ and $\theta(\cdot)$ as follows. For $|\rho_G| = 0$ (i.e., $\rho_G = q_0$), let $\rho_H = \theta(\rho_G) = (s_0, o_0)$ where $s_0 = o_0 = \{q_0\}$ if $q_0 \notin \mathcal{T}$, and $s_0 = o_0 = \emptyset$ if $q_0 \in \mathcal{T}$, and let $\sigma(\rho_G) = a$ if $\sigma'(\rho_H) = (a, \cdot)$. Note that property $(\star)$ holds. For the induction step, let $k \geq 1$ and assume that from every play prefix ρ_G of length smaller than k, we have defined $\sigma(\rho_G)$ and $\theta(\rho_G)$ satisfying $(\star)$. Let $\rho_G = q_0 q_1 \dots q_k$ be a play prefix in G of length k. Let $\rho_H = \theta(q_0 q_1 \dots q_{k-1})$ and $\gamma_k = \mathsf{obs}_1(q_k)$, and let (s_k, o_k) be the (unique) successor state in the Dirac distribution $\delta_H(\mathsf{Last}(\rho_H), \sigma'(\rho_H), \gamma_k)$. Note that $q_k \in s_k$. Define $\theta(\rho_G) = \rho_H.(s_k, o_k)$ and $\sigma(\rho_G) = a$ if $\sigma'(\rho_H.(s_k, o_k)) = (a, \cdot)$. Therefore, the property $(\star)$ holds.

Note that the strategy σ is observation-based because if $\mathsf{obs}_1(\rho_G) = \mathsf{obs}_1(\rho'_G)$, then $\theta(\rho_G) = \theta(\rho'_G)$.

Correctness of the mapping. If player 1 has a sure winning strategy σ' in H for the objective $\mathsf{Büchi}(\alpha)$, then we can assume that σ' is memoryless (since in perfect-observation deterministic games with Büchi objectives memoryless strategies are sufficient for sure winning [25, 44]), and we show that the strategy σ defined above is almost-sure winning in G for the objective $\mathsf{Reach}(\mathcal{T})$.

Since σ' is memoryless and sure winning for $\mathsf{Büchi}(\alpha)$, in every play compatible with σ' there are at most $n = |L| \leq 3^{|Q|}$ steps between two consecutive visits to an α-state.

The properties of matching plays entail that if a play prefix ρ_G compatible with σ has no visit to $\mathcal{T}$-states, and $(s, o) = \mathsf{Last}(\theta(\rho_G))$, then $\mathsf{Last}(\rho_G) \in s$. Moreover if $s = o$, then under strategy σ for player 1 and arbitrary strategy π for player 2, there is a way to fix the probabilistic choices such that all plays extension of ρ_G visit a $\mathcal{T}$-state. To see this, consider the probabilistic choices given at each step by the witness component u of the action $(\cdot, u)$ played by σ'. By the definition of the mapping of plays and of the transition function in H, it can be shown that if $(s_i, o_i)(s_{i+1}, o_{i+1}) \dots (s_k, o_k)$ is a play fragment of $\theta(\rho_G)$ (hence compatible with σ') where $s_i = o_i$ and $o_j \neq \emptyset$ for all $i \leq j < k$, then the "owe" set o_k is the set of all states that can be reached in G from states s_i along a path which is compatible with both the action played by the strategy σ' (and σ) and the probabilistic choices fixed by σ', and visits no $\mathcal{T}$-states. Since the "owe" set gets empty within at most n steps regardless of the strategy of player 2, all paths compatible with the probabilistic choices must visit an $\mathcal{T}$-state. This shows that under any player 2 strategy, within n steps, a $\mathcal{T}$-state is visited with probability at least r^n where $r > 0$ is the smallest non-zero probability occurring in G. Therefore, the probability of not having visited a $\mathcal{T}$-state after $z \cdot n$ steps is at most $(1 - r^n)^z$ which vanishes for $z \to \infty$ since $r^n > 0$. Hence, against arbitrary strategy of player 2, the strategy σ ensures the objective $\mathsf{Reach}(\mathcal{T})$ with probability 1.

Memory bound. Since H is a perfect-information game, pure memoryless sure winning strategies exist in H for Büchi objectives [25, 44]. Consider a pure memoryless sure winning strategy in H, and the strategy ensures that if a state (s, o) visited in the play, then it satisfies that $o \subseteq s \subseteq \gamma$ for some $\gamma \in \mathcal{O}_1$ (i.e., the first component is a subset of some observation). The number of distinct states (s, o) such that $o \subseteq s \subseteq \gamma$ for some $\gamma \in \mathcal{O}_1$ is bounded by $\sum_{\gamma \in \mathcal{O}_1} 3^{|\gamma|}$, where $|\gamma|$ is the cardinality of γ (i.e., the number of different states in the observation γ). It follows

that memory of size $\sum_{\gamma \in \mathcal{O}_1} 3^{|\gamma|}$ suffices for almost-sure winning for pure strategies for reachability objectives in one-sided games with player 1 partial and player 2 perfect.

Argument for positive reachability. The proof for positive reachability follows the same line as for almost-sure reachability, with the following differences. The construction of the game of perfect information H is now interpreted as a reachability game with objective $\mathsf{Reach}(\alpha)$. The mapping of plays is the same as above. In the mapping of strategies from G to H, we use the same ranking construction, but we only claim that the initial state gets a rank. The argument is that if the initial state would get no rank, then player 2 would have a strategy to ensure that all paths avoid the target states, in contradiction with the fact that player 1 has fixed a positive winning strategy. The rest of the proof is analogous to the case of almost-sure reachability.

Memory bound. We first observe that if the objective is $\mathsf{Reach}(\alpha)$, then all states in α can be converted to absorbing states. Hence it follows that if the objective in H is the reachability objective, then the obligation component does not need to be recharged when it becomes empty (in contrast to the case when the objective in H is the Büchi objective). Hence a sure winning strategy in H for the objective $\mathsf{Reach}(\alpha)$ can only depend on the obligation component (i.e., for a state (s, o), the sure winning strategy only depends on o). We also remark that if the game G is a non-stochastic game, then the obligation component coincides with belief. As before, if a state (s, o) is reachable, then $o \subseteq s \subseteq \gamma$ for some $\gamma \in \mathcal{O}_1$. Since H is a perfect-information game, pure memoryless sure winning strategies exist in H for reachability objectives [25, 44]. Hence it follows that memory of size $\sum_{\gamma \in \mathcal{O}_1} 2^{|\gamma|}$ suffices for positive winning for pure strategies for reachability objectives in one-sided games with player 1 partial and player 2 perfect. □

It follows from the construction in the proof of Lemma 1 that pure strategies with exponential memory are sufficient for positive (as well as almost-sure) winning, and the exponential lower bound follows from the special case of non-stochastic games [8]. Lemma 1 also gives EXPTIME upper bound for the problem since perfect-observation Büchi games can be solved in polynomial time [44]. The EXPTIME-hardness follows from the sure winning problem for non-stochastic games [39], where pure almost-sure (positive) winning strategies coincide with sure winning strategies. We have the following theorem summarizing the results.

Theorem 1. *For one-sided partial-observation stochastic games with player 1 partial and player 2 perfect, the following assertions hold for reachability objectives for player 1:*

1. (Memory complexity). *Belief-based pure strategies are not sufficient both for positive and almost-sure winning; exponential memory is necessary and sufficient both for positive and almost-sure winning for pure strategies. Memory of size $\sum_{\gamma \in \mathcal{O}_1} 2^{|\gamma|}$ for positive, and $\sum_{\gamma \in \mathcal{O}_1} 3^{|\gamma|}$ for almost-sure winning is sufficient.*
2. (Algorithm). *The problems of deciding the existence of a pure almost-sure and a pure positive winning strategy can be solved in time exponential in the state space of the game and polynomial in the size of the action sets.*
3. (Complexity). *The problems of deciding the existence of a pure almost-sure and a pure positive winning strategy are EXPTIME-complete.*

From Theorem 1 and Remark 3 we obtain the following corollary.

Corollary 1. *The problem of deciding the existence of a pure almost-sure winning strategy for one-sided partial-observation stochastic games with player 1 partial and player 2 perfect, and Büchi objective for player 1 is EXPTIME-complete, and memory of size $\sum_{\gamma \in \mathcal{O}_1} 3^{|\gamma|}$ is sufficient for pure almost-sure winning strategies.*

Also note that we have $\sum_{\gamma \in \mathcal{O}_1} 2^{|\gamma|} \leq \prod_{\gamma \in \mathcal{O}_1} 2^{|\gamma|} = 2^n$ and $\sum_{\gamma \in \mathcal{O}_1} 3^{|\gamma|} \leq \prod_{\gamma \in \mathcal{O}_1} 3^{|\gamma|} = 3^n$, where n is the number of states in the one-sided game.

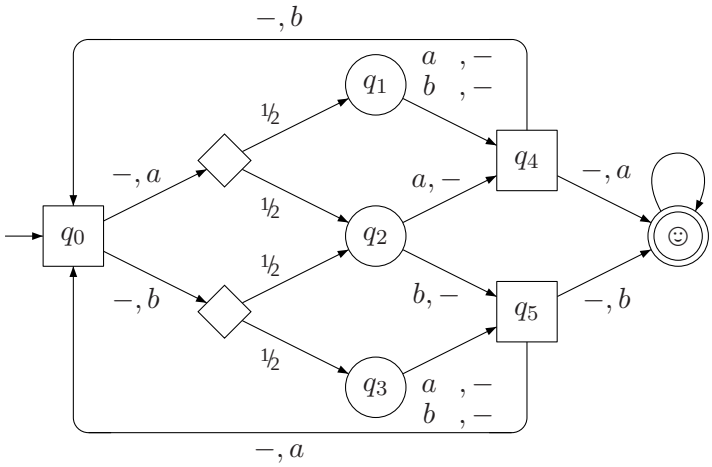

Figure 3. Remembering the belief of player 2 is necessary. A one-sided reachability game where player 1 (round states) has perfect observation, player 2 (square states) is blind. Player 1 has a pure almost-sure winning strategy that depends on the belief of player 2 (in q_2), but no pure memoryless strategy is almost-sure winning.

Symbolic algorithms. The exponential Büchi (or reachability) game constructed in the proof of Theorem 1 can be solved by computing classical fixpoint formulas [25]. However, it is not necessary to construct the exponential game structure explicitly. Instead, we can exploit the structure induced by the pre-order $\preceq$ defined by $(s, o) \preceq (s', o')$ if (i) $s \subseteq s'$, (ii) $o \subseteq o'$, and (iii) $o = \emptyset$ iff $o' = \emptyset$. Intuitively, if a state (s', o') is winning for player 1, then all states $(s, o) \preceq (s', o')$ are also winning because they correspond to a better belief and a looser obligation. Hence all sets computed by the fixpoint algorithm are downward-closed and thus they can be represented symbolically by the antichain of their maximal elements (see [17] for details related to antichain algorithms). This technique provides a symbolic algorithm without explicitly constructing the exponential game.

4 One-sided Games: Player 1 Perfect and Player 2 Partial

Recall that we are interested in finding a pure winning strategy for player 1. Therefore, when we construct counter-strategies for player 2, we always assume that player 1 has already fixed a pure strategy. This is important for the way the belief of player 2 is updated. Although player 2 does not have perfect information about the actions played by player 1, the belief of player 2 can be updated according to the precise actions of player 1 because the response and the counter-strategy of player 2 is designed after player 1 has fixed a strategy.

4.1 Lower bound on memory

We present the following examples to illustrate two properties of the problem.

Example 2. Remembering the belief of player 2 is necessary. *We present an example of a game where player 1 has perfect observation but needs to remember the belief of player 2 to ensure positive or almost-sure reachability. The game is shown in Figure 3. The target is $\mathcal{T} = \{q_{\text{☺}}\}$. Player 2 is blind. If player 2 chooses a in the initial state q_0, then his belief will be $\{q_1, q_2\}$, and if he plays b, then his belief will be $\{q_2, q_3\}$. In q_2, the choice of player 1 depends on the belief of player 2. If the belief is $\{q_1, q_2\}$, then playing a in q_2 is not a good choice because the belief of player 2 would be $\{q_4\}$ and player 2 could surely avoid $q_{\text{☺}}$ by further playing b. For symmetrical reasons, if the belief of player 2 is $\{q_2, q_3\}$ in q_2, then playing b is not a good choice for player 1. Therefore, there is no*

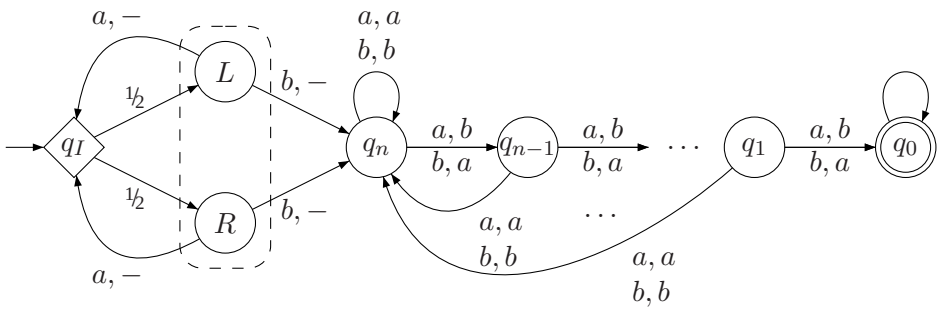

Figure 4. A one-sided reachability game L_n with reachability objective in which player 1 is has perfect observation and player 2 is blind. Player 1 needs exponential memory to win positive reachability.

positively winning memoryless *strategy for player* 1*. However, we show that there exists an almost-sure winning* belief-based *strategy for player* 1 *as follows: in* q_2*, play* b *if the belief of player* 2 *is* $\{q_1, q_2\}$*, and play* a *if the belief of player* 2 *is* $\{q_2, q_3\}$*. Note that player* 1 *has perfect observation and thus can observe the actions of player* 2*. This ensures the next belief of player* 2 *to be* $\{q_3, q_4\}$ *and therefore no matter the next action of player* 2*, the state* $q_{\text{☺}}$ *is reached with probability* $\frac{1}{2}$*. Repeating this strategy ensures to reach* $q_{\text{☺}}$ *with probability* 1*.* ■

Example 3. Memory of non-elementary size may be necessary for positive and almost-sure reachability. *We show that player* 1 *may need memory of non-elementary size to win positively (as well as almost-surely) in a reachability game. We present a family of one-sided games* G_n *where player* 1 *has perfect observation, and player* 2 *has partial observation both about the state of the game, and the actions played by player* 1*. We explain the example step by step. The key idea of the example is that the winning strategy of player* 1 *in game* G_n *will need to simulate a counter systems (with* n *integer-valued counters) where the operations on counters are* increment *and* division by 2 *(with round down), and to reach strictly positive counter values.*

Counters. *First, we use a simple example to show that counters appear naturally in the analysis of the game under pure strategies.*

Consider the family of games $(L_n)_{n\in\mathbb{N}}$ *shown in Figure 4, where the reachability objective is* $\mathsf{Reach}(\{q_0\})$*. In the first part, the states* L *and* R *are indistinguishable for player* 2*. Consider the strategy of player* 1 *that plays* b *in* L *and* R*. Then, the state* q_n *is reached by two play prefixes* $\rho_{up} = q_I L q_n$ *and* $\rho_{dw} = q_I R q_n$ *that player* 2 *cannot distinguish. Therefore, player* 2 *has to play the same action in both play prefixes, while perfectly-informed player* 1 *can play different actions. In particular, if player* 1 *plays* a *in* ρ_{up} *and* b *in* ρ_{dw}*, then no matter the action chosen by player* 2 *the state* q_{n-1} *is reached with positive probability. However, because only one play prefix reaches* q_{n-1}*, this strategy of player* 1 *cannot ensure to reach* q_{n-2} *with positive probability.*

Player 1 *can ensure to reach* q_{n-2} *(and* q_0*) with positive probability with the following exponential-memory strategy. For the first* $n-1$ *visits to either* L *or* R*, play* a*, and on the* n*th visit, play* b*. This strategy produces* 2^n *different play prefixes from* q_I *to* q_n*, each with probability* $\frac{1}{2^n}$*. Considering the mapping* $L \mapsto a$*,* $R \mapsto b$*, each such play prefix* ρ *is mapped to a sequence* w_ρ *of length* n *over* $\{a, b\}$ *(for example, the play prefix* $q_I L q_I R q_I L q_n$ *is mapped to* aba*). The strategy of player* 1 *is to play the sequence* w_ρ *in the next* n *steps after* ρ*. This strategy ensures that for all* $0 \le i \le n$*, there are* 2^i *play prefixes which reach* q_i *with positive probability, all being indistinguishable for player* 2*. The argument is an induction on* i*. The claim is true for* $i = n$*, and if it holds for* $i = k$*, then no matter the action chosen by player* 2 *in* q_k*, the state* q_{k-1} *is reached with positive probability by half of the* 2^k *play prefixes, i.e.* 2^{k-1} *play prefixes. This establishes the claim. As a consequence, one play prefix reaches* q_0 *with positive probability. This strategy requires exponential memory, and an inductive argument shows that this memory is necessary because player* 1 *needs to have at least* 2 *play prefixes that are indistinguishable for player* 2 *in state* q_1*, and at least* 2^i *play prefixes in* q_i *for all* $0 \le i \le n$*.*

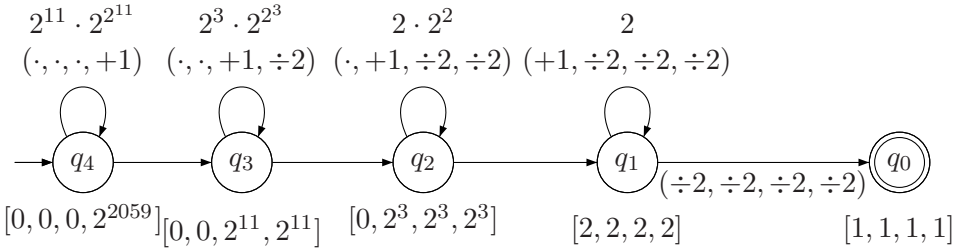

Figure 5. A family $(C_n)_{n\in\mathbb{N}}$ of counter systems with n counters and $n+1$ states where the shortest execution to reach $(q_0, k_1, \ldots, k_n)$ with positive counters (i.e., $k_i > 0$ for all $1 \leq i \leq n$) from $(q_n, 0, \ldots, 0)$ is of non-elementary length. The numbers above the self-loops show the number of times each self-loop is taken along the shortest execution.

Non-elementary counters. *Now, we present a family C_n of counter systems where the shortest execution is of non-elementary length (specifically, the shortest length is greater than a tower $2^{2^{\cdot^{\cdot^{2}}}}$ of exponentials of height n). The counter system C_4 (for $n = 4$) is shown in Figure 5. The operations on counters can be* increment *($+1$),* division by 2 *($\div 2$), and* idle *($\cdot$). In general, C_n has n counters $c_1, \ldots, c_n$ and $n+1$ states $q_0, \ldots, q_n$. In state q_i of C_n ($0 \leq i \leq n$), the counter c_i can be incremented and at the same time all the counters c_j for $j > i$ are divided by 2. From q_n, to reach q_0 with strictly positive counters (i.e., all counters have value at least 1), we show that it is necessary to execute the self-loop on state q_n a non-elementary number of times. In Figure 5, the numbers above the self-loops show the number of times they need to be executed. When leaving q_1, the counters need to have value at least 2 in order to survive the transition to q_0 which divides all counters by 2. Since the first counter can be incremented only in state q_1, the self-loop in q_1 has to be executed 2 times. Hence, when leaving q_2, the other counters need to have value at least $2 \cdot 2^2 = 2^3$ in order to survive the self-loops in q_1. Therefore, the self-loop in q_2 is executed 2^3 times. And so on. In general, if the self-loop on state q_i is executed k times (in order to get $c_i = k$), then the counters $c_{i+1}, \ldots, c_n$ need to have value $k \cdot 2^k$ when entering q_i (in order to guarantee a value at least k of these counters). In q_n, the last counter c_n needs to have value $f^n(1)$ where f^n is the nth iterate of the function $f : \mathbb{N} \to \mathbb{N} : x \mapsto x \cdot 2^x$. This value is greater than a tower of exponentials of height n.*

Gadgets for increment and division. *In Figure 6, we show the gadgets that are used to simulate operations on counters. The gadgets are game graphs where the player-1 actions a, b are indistinguishable for player 2 (but player 2 can observe and distinguish the action $\#$). The actions a, b are used by player 1 to simulate the operations on the counters. The $\#$ is used to simulate the transitions from state q_i to q_{i-1} of the counter system of Figure 5. All states of the gadgets have the same observation for player 2. Recall that player 1 has perfect observation.*

The idle gadget is straightforward. The actions a, b have no effect. In the other gadgets, the value of the counters is represented by the number of paths that are indistinguishable for player 2, and that end up in the entry state of the gadget (for the value of the counter before the operation) or in the exit state (for the value of the counter after the operation).

Consider the division gadget div_2. *If player 2 plays an action that matches the choice of player 1, then the game leaves the gadget and the transition will go to the initial state of the game we construct (which is shown on Figure 8). Otherwise, the action of player 2 does not match the action of player 1 and the play reaches the exit state of the gadget. Let k be the number of indistinguishable[1] paths in the entry state of the gadget. By playing a after k_1 such paths and b after k_2 paths (where $k_1 + k_2 = k$), player 1 ensures that $\min\{k_1, k_2\}$ indistinguishable paths reach the exit state of the gadget (because in the worst case, player 2 can choose his action to match the action of player 1 over $\max\{k_1, k_2\}$ paths). Hence, player 1 can ensure that $\lfloor \frac{k}{2} \rfloor$ indistinguishable paths get to*

[1] In the rest of this section, the word *indistinguishable* means *indistinguishable for player 2*.

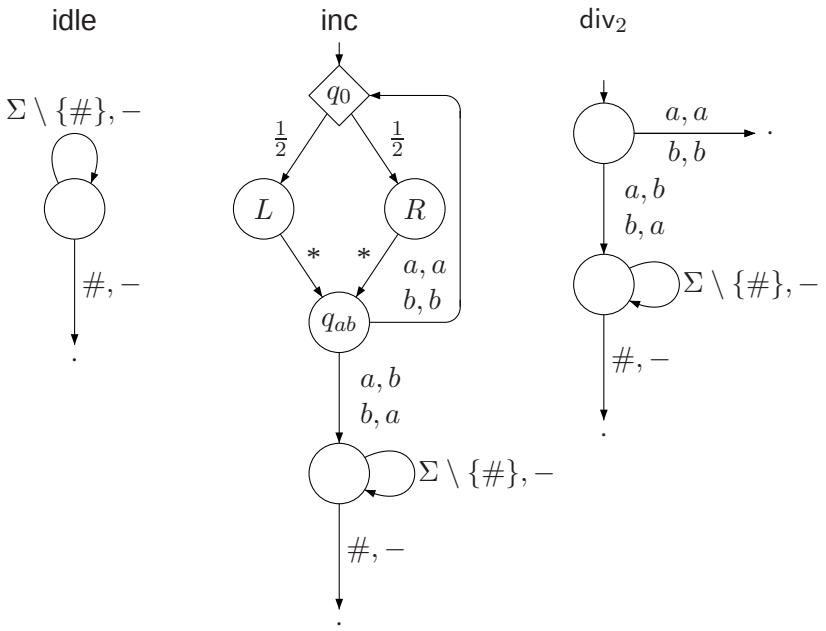

Figure 6. Gadgets to simulate idle**,** increment**, and** division by 2**.**

the exit state. In the game of Figure 8, the entry and exit state of division gadgets are merged. The argument still holds.

Consider the increment gadget inc *on Figure 6. We use this gadget with the assumption that the entry state is not reached by more than one indistinguishable path. This will be the case in the game of Figure 8. Player* 1 *can achieve* k *indistinguishable paths in the exit state as follows. In state* q_{ab}*, play action* a *if the last visited state is* q_L*, and play action* b *if the last visited state is* q_R*. No matter the choice of player* 2*, one path will reach the exit state, and the other path will get to the entry state. Repeating this scenario* k *times gives* k *paths in the exit state. We show that there is essentially no faster way to obtain* k *paths in the exit state. Indeed, if player* 1 *chooses the same action (say* a*) after the two paths ending up in* q_{ab}*, then against the action* b *from player* 2*, two paths reach the exit state, and no state get to the entry state. Then, player* 1 *can no longer increment the number of paths. Therefore, to get* k *paths in the exit state, the fastest way is to increment one by one up to* $k-2$*, and then get* 2 *more paths as a last step. Note that it is not of the interest of player* 2 *to match the action of player* 1 *if player* 1 *plays the same action, because this would double the number of paths.*

Structure of the game. *The game* G_n *which requires memory of non-elementary size is sketched in Figure 8 for* $n = 3$*. Its abstract structure is shown in Figure 7, corresponding to the structure of the counter system in Figure* 5*. The alphabet of player* 1 *is* $\{a, b, \#\}$*. For the sake of clarity, some transitions are not depicted in Figure 8. It is assumed that for player* 1*, playing an action from a state where this action has no transition depicted leads to the initial state of the game. For example, playing* $\#$ *in state* q_4 *goes to the initial state, and from the target state* $q_{\text{☺}}$*, all transitions go to the initial state.*

Figure 8 shows the initial state q_I *of the game from which a uniform probabilistic transition branches to the three states* q_7, r_7, s_7*. The idea of this game is that player* 1 *needs to ensure that the states* q_1, r_1, s_1 *are reached with positive probability, so as to ensure that no matter the action (*a*,* b*, or* c*) chosen by player* 2*, the state* $q_{\text{☺}}$ *is reached with positive probability. From* q_1, r_1, s_1*, the other actions of player* 2 *(i.e.,* b *and* c *from* q_1*,* a *and* c *from* r_1*, etc.) lead to the initial state. Player* 2 *can observe the initial state. All the other states are indistinguishable.*

*Intuitively, each "*line*" of states (*q*'s,* r*'s, and* s*'s) simulate one counter. Synchronization of the operations on*

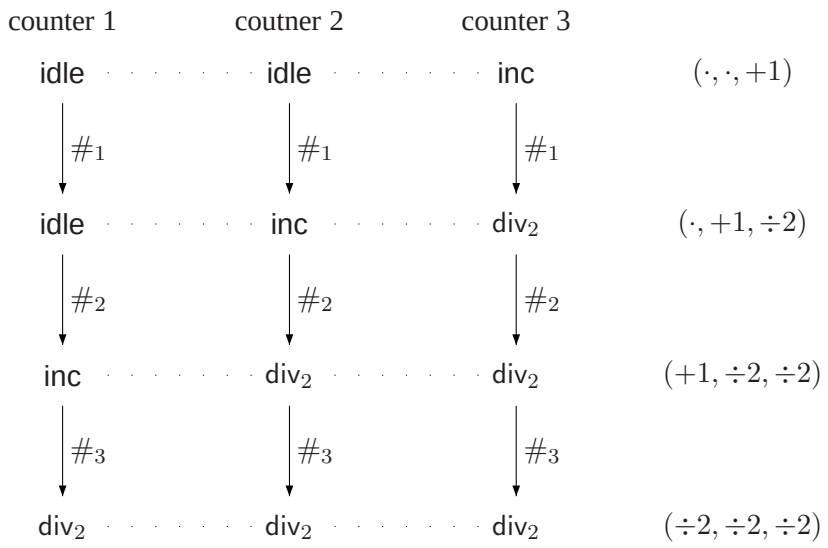

Figure 7. Abstract view of the game in Figure 8 as a 3-counter system.

the three counters is ensured by the special (and visible to player 2) symbol $\#$. Intuitively, since $\#$ is visible to player 2, player 1 must play $\#$ at the same "time" in the three lines of states (i.e., after the same number of steps in each line). Otherwise, player 2 may eliminate one line of states from his belief. For example, say after k steps, the game could be in state q_7, r_7, or some state s_j ($5 \leq j \leq 7$), and if player 1 plays $\#_1$ in the states q_7 and r_7, but plays a different action from s_j, then player 2 observing $\#_1$ after k steps can safely update his belief to $\{q_6, r_6\}$, and thus avoid to play c when one of the states q_1, r_1 is reached. In Figure 8, the dotted lines and the subscripts on $\#$ emphasize the layered structure of the game, corresponding to the structure of Figure 7.

From all the above, it follows that player 1 needs memory of size non-elementary in order to ensure indistinguishable paths ending up in each of the states q_1, r_1, s_1, and win with positive probability. Since all other paths are going back to the initial state, this strategy can be repeated over and over again to achieve almost-sure reachability as well. ■

Theorem 2. *In one-sided partial-observation stochastic games with player 1 perfect and player 2 partial, both pure almost-sure and pure positive winning strategies for reachability objectives for player 1 require memory of non-elementary size in general.*

4.2 Upper bound for positive reachability with almost-sure safety

We present the solution of one-sided games with a conjunction of positive reachability and almost-sure safety objectives, in which player 1 has perfect observation and player 2 has partial observation. This will be useful in Section 4.3 to solve almost-sure reachability, and using a trivial safety objective (safety for the whole state space) it also gives the solution for positive reachability.

Let $G = \langle Q, q_0, \delta_G \rangle$ be a game over alphabets A_1, A_2 and observation set $\mathcal{O}_2$ for player 2, with reachability objective $\mathsf{Reach}(\mathcal{T})$ (where $\mathcal{T} \subseteq Q$) and safety objective $\mathsf{Safe}(Q_G)$ (where $Q_G \subseteq Q$ represents a set of good states) for player 1. We assume that the states in $\mathcal{T}$ are absorbing and that $\mathcal{T} \subseteq Q_G$. This assumption is satisfied by the games we consider in Section 4.3, as well as by the case of a trivial safety objective ($Q_G = Q$). The goal of player 1 is to ensure positive probability to reach $\mathcal{T}$ and almost-sure safety for the set Q_G.

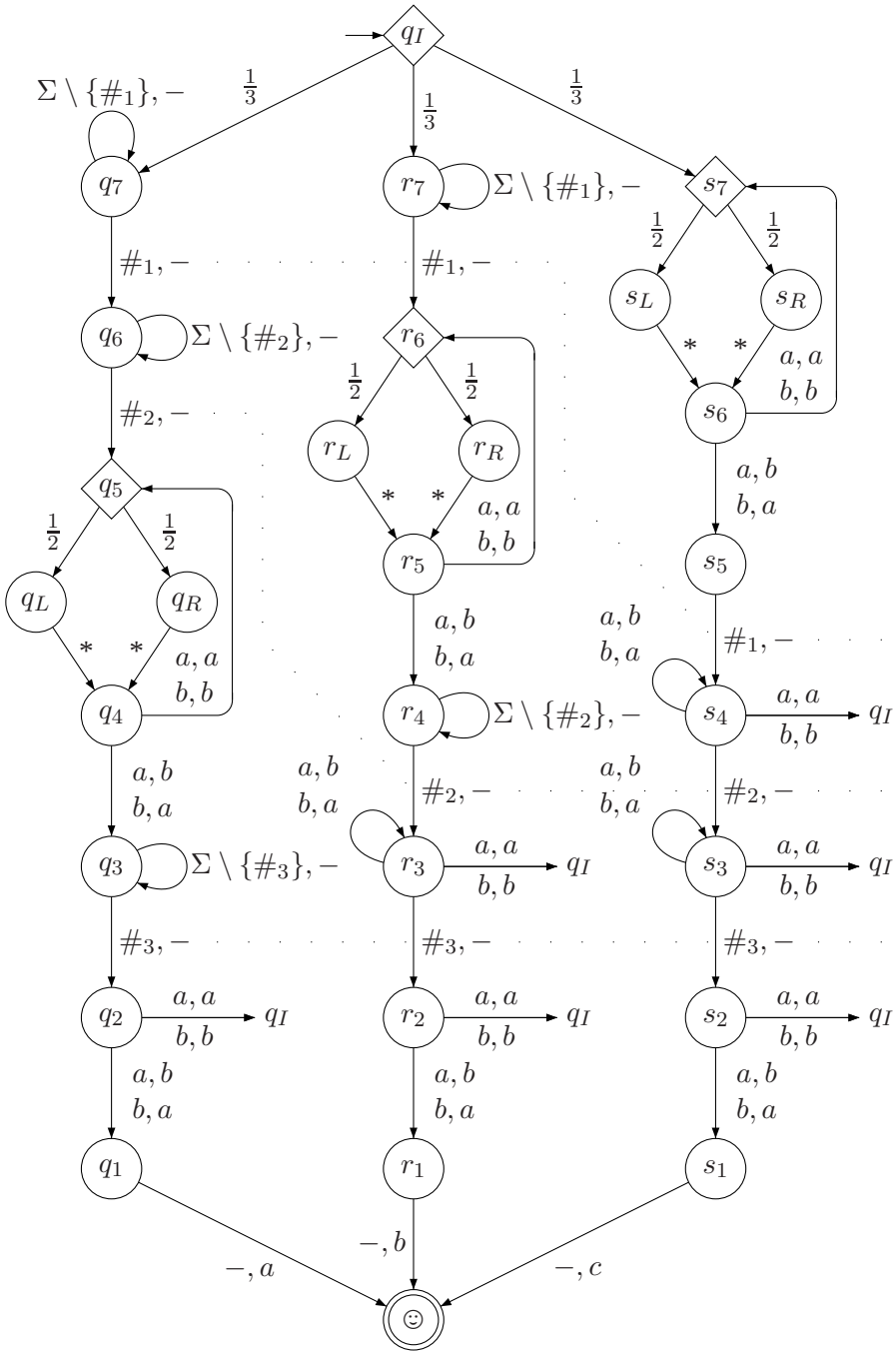

Figure 8. Memory of non-elementary size may be necessary for positive and almost-sure reachability. A family of one-sided reachability games in which player 1 is has perfect observation. Player 1 needs memory of non-elementary size to win positive reachability (as well as almost-sure reachability).

Before presenting the algorithm for solving these games in pure strategies, we consider the case of randomized strategies. After, we use the results of randomized strategies to solve the case of pure strategies.

Step 1 - Winning with randomized strategies. First, we show that with randomized strategies, memoryless strategies are sufficient. It suffices to play uniformly at random the set of safe actions. In a state q, an action $a \in A_1$ is *safe* if $\mathsf{Post}_G(q,a,b) \subseteq \mathsf{Win}_{safe}$ for all $b \in A_2$, where Win_{safe} is the set of states that are sure winning[2] for player 1 in G for the safety objective $\mathsf{Safe}(Q_G)$. This strategy ensures that the set $Q \setminus Q_G$ of bad states is never reached, and from the positive winning region of player 1 for $\mathsf{Reach}(\mathcal{T})$ it ensures that the set $\mathcal{T}$ is reached with positive probability. Therefore, computing the set Z of states that are winning for player 1 in randomized strategies can be done by fixing the uniformly randomized safe strategy for player 1, and checking that player 2 does not almost-surely win the safety objective $\mathsf{Safe}(Q \setminus \mathcal{T})$, which requires the analysis of a POMDP for almost-sure safety and can be done in exponential time using a simple subset construction [16, Theorem 2].

Note that $\mathcal{T} \subseteq Z$ and that from all states in Z, player 1 can ensure that $\mathcal{T}$ is reached with positive probability within at most $2^{|Q|}$ steps, while from any state $q \notin Z$, player 1 cannot win positively with a randomized strategy, and therefore also not with a pure strategy.

Step 2 - Pure strategies to simulate randomized strategies. Second, we show that pure strategies can in some cases simulate the behavior of randomized strategies. As we have seen in the gadget inc of Figure 6, if there are two play prefixes ending up in the same state and that are indistinguishable for player 2 (e.g., $q_0 L q_{ab}$ and $q_0 R q_{ab}$ in the example), then player 1 can simulate a random choice of action over support $\{a,b\}$ by playing a after $q_0 L q_{ab}$, and playing b after $q_0 R q_{ab}$. No matter the choice of player 2, one of the plays will reach q_0 and the other will reach the exit state of the gadget. Intuitively, this corresponds to a uniform probabilistic choice of the actions a and b: the state q_0 and the exit state are reached with probability $\frac{1}{2}$.

In general, if there are $|A_1|$ indistinguishable play prefixes ending up in the same state q, then player 1 can simulate a random choice of actions over A_1 from q. However, the number of indistinguishable play prefixes in a successor state q' may have decreased by a factor $|A_1|$ (there may be just one play reaching q'). Hence, in order to simulate a randomized strategy during k steps, player 1 needs to have $|A_1|^k$ indistinguishable play prefixes. Since $2^{|Q|}$ steps are sufficient for a randomized strategy to achieve the reachability objective, an upper bound on the number of play prefixes that are needed to simulate a randomized strategy using a pure strategy is $\mathsf{Num} = |A_1|^{2^{|Q|}}$. More precisely, if the belief of player 2 is $B \subseteq Z$ and in each state $q \in B$ there are at least Num indistinguishable play prefixes, then player 1 wins with a pure strategy that essentially simulates a winning randomized strategy (which exists since $q \in Z$) for 2^n steps.

Step 3 - Counting abstraction for pure strategies. We present a construction of a game of perfect observation H such that player 1 wins in H if and only if player 1 wins in G. The objective in H is a conjunction of positive reachability and almost-sure safety objectives, for which pure memoryless winning strategies exist: for every state we restrict the set of actions to safe actions, and then we solve positive reachability on a perfect-observation game. The result follows since for perfect-observation games pure memoryless positive winning strategies exist for reachability objectives [18].

State space. The idea of this construction is to keep track of the belief set $B \subseteq Q$ of player 2, and for each state $q \in B$, of the number of indistinguishable play prefixes that end up in q. For $k \in \mathbb{N}$, we denote by $[k]$ the set $\{0,1,\ldots,k\}$. A state of H is a *counting function* $f : Q \to [\mathsf{K}_*] \cup \{\omega\}$ where $\mathsf{K}_* \in \mathbb{N}$ is of order $|A_1|^{|A_1|^{\cdot^{\cdot^{\cdot^{|A_1|^{2^{O(n)}}}}}}}$ where the number of nested exponentials is in $O(n)$ (where $n = |Q|$).

[2]Note that for safety objectives, the notion of sure winning and almost-sure winning coincide, and pure strategies are sufficient.

As we have seen in the example of Figure 8, it may be necessary to keep track of a non-elementary number of play prefixes. We show that the bound K_* is sufficient, and that we can substitute larger numbers by the special symbol ω to obtain a *finite* counting abstraction. The belief associated with a counting function f is the set $\mathsf{Supp}(f) = \{q \in Q \mid f(q) \neq 0\}$, and the states q such that $f(q) = \omega$ are called ω*-states*.

Action alphabet. In H, an action of player 1 is a function $\hat{a} : Q \times [\mathsf{K}_*] \to A_1$ that assigns to each copy of a state in the current belief (of player 2), the action played by player 1 after the corresponding play prefix in G. We denote by $\mathsf{Supp}(\hat{a}(q,\cdot)) = \{\hat{a}(q,i) \mid i \in [\mathsf{K}_*]\}$ the set of actions played by $\hat{a}$ in $q \in Q$.

The action set of player 2 in the game H is the same as in G.

Transitions. Let $\mathbf{1}(a, A)$ be 1 if $a \in A$, and 0 if $a \notin A$. We denote this function by $\mathbf{1}(a \in A)$. Given f and $\hat{a}$ as above, given an action $b \in A_2$ and an observation $\gamma \in \mathcal{O}_2$, let $f' = \mathsf{Succ}(f, \hat{a}, b, \gamma)$ be the function such that $f'(q') = 0$ for all $q' \notin \gamma$, and such that for all $q' \in \gamma$:

$$f'(q') = \begin{cases} \omega & \text{if } \exists a \in \mathsf{Supp}(\hat{a}(q,\cdot)) \cdot \exists q \in Q : f(q) = \omega \wedge q' \in \mathsf{Post}_G(q,a,b) \\ x & \text{otherwise} \end{cases}$$

$$\text{where } x = \textstyle\sum_{q \in \mathsf{Supp}(f)} \sum_{i=0}^{f(q)-1} \mathbf{1}(q' \in \mathsf{Post}_G(q, \hat{a}(q,i), b)).$$

Note that if the current state q is an ω-state, then only the support $\mathsf{Supp}(\hat{a}(q,\cdot))$ of the function $\hat{a}$ matters.

Now $f' = \mathsf{Succ}(f, \hat{a}, b, \gamma)$ may not be a counting function because it may assign values greater than K_* to some states. We show that beyond certain bounds, it is not necessary to remember the exact value of the counters and we can replace such large values by ω. Intuitively, the ω value can be interpreted as "very large and definitely positive value". This abstraction needs to be done carefully in order to obtain the desired upper bound (namely, K_*). When a counter $f(q)$ has value ω, the successors of q have value ω according to $\mathsf{Succ}(\cdot)$, which is faithful if the exact value of the counter $f(q)$ is large enough. In fact, large enough means that the counter has value at least $|A_1|$ as this allows player 1 to play each action at least once. Hence the abstraction remains faithful during K steps if the counters with value greater than $|A_1|^{\mathsf{K}}$ are set to ω. We know that if all counters have value greater than $\mathsf{K}_1 = |A_1|^{2^n}$, then player 1 wins by simulating a randomized strategy. Therefore, when all counters but one have already value ω, we set the last counter to ω if it has value greater than K_1. Since this can take at most K_1 steps, the other counters with value ω need to have value at least $\mathsf{K}_2 = \mathsf{K}_1 \cdot |A_1|^{\mathsf{K}_1}$.

Therefore, when all counters but two have already value ω, whenever a counter gets value greater than K_2 we set it to ω. This can take at most $(\mathsf{K}_2)^2$ steps and the other counters with value ω need to have value at least $\mathsf{K}_3 = \mathsf{K}_2 \cdot |A_1|^{(\mathsf{K}_2)^2}$. In general, when all counters but k have value ω, we set a counter to ω if it has value at least $\mathsf{K}_{k+1} = \mathsf{K}_k \cdot |A_1|^{(\mathsf{K}_k)^k}$. It can be shown by induction that K_k is of order $|A_1|^{|A_1|^{\cdot^{\cdot^{\cdot^{|A_1|^{2^{O(n)}}}}}}}$ where the tower of exponential is of height k, and thus we do not need to store counter values greater than K_*. We define the abstraction mapping $f' = \mathsf{Abs}(f)$ for $f : Q \to \mathbb{N}$ as follows:

> Let $k = |\{q \mid f(q) = \omega\}|$ be the number of counters with value ω in f. If there is a state $\hat{q}$ with finite value $f(\hat{q})$ greater than K_{n-k}, then $f'(\hat{q}) = \omega$ and f' agrees with f on all states except $\hat{q}$ (i.e., $f'(q) = f(q)$ for all $q \neq \hat{q}$). Otherwise, $f' = f$.

Actually, we define $\mathsf{Abs}(f)$ as the nth iterate of the above procedure. Given f, $\hat{a}$, and b, let $\delta_H(f, \hat{a}, b)$ be the uniform distribution over the set of counting functions f' such that there exists an observation $\gamma \in \mathcal{O}_2$ such that $f' = \mathsf{Abs}(\mathsf{Succ}(f, \hat{a}, b, \gamma))$ and $\mathsf{Supp}(f') \neq \emptyset$.

Note that the operators $\mathsf{Succ}(\cdot)$ and $\mathsf{Abs}(\cdot)$ are *monotone*, that is $f \leq f'$ implies $\mathsf{Abs}(f) \leq \mathsf{Abs}(f')$ as well as $\mathsf{Succ}(f, \hat{a}, b, \gamma) \leq \mathsf{Succ}(f', \hat{a}, b, \gamma)$ for all $\hat{a}, b, \gamma$ (where $\leq$ is the componentwise order).

Objective. Given $\mathcal{T} \subseteq Q$ and $Q_G \subseteq Q$ defining the reachability and safety objectives in G, the objective in the game H is a conjunction of positive reachability and almost-sure safety objectives, defined by $\mathsf{Reach}(\mathcal{T}_H)$ where[3] $\mathcal{T}_H = \{f \mid \mathsf{Supp}(f) \subseteq Z \wedge \forall q \in \mathsf{Supp}(f) : f(q) = \omega\} \cup \{f \mid \mathsf{Supp}(f) \cap \mathcal{T} \neq \emptyset\}$ and by $\mathsf{Safe}(\mathsf{Good}_H)$ where $\mathsf{Good}_H = \{f \mid \mathsf{Supp}(f) \subseteq Q_G\}$.

Step 4 - Correctness argument. First, assume that there exists a pure winning strategy σ for player 1 in G, and we show how to construct a winning strategy σ^H in H. As we play the game in G using σ, we keep track of the exact number of indistinguishable play prefixes ending up in each state. This allows to define the action $\hat{a}$ to play in H by collecting the actions played by σ in all the indistinguishable play prefixes. Note that by monotonicity, the counting abstractions in the corresponding play prefix of H are at least as big (assuming $\omega > k$ for all $k \in \mathbb{N}$), and thus the action $\hat{a}$ is well-defined. Since σ is winning, $\mathcal{T}$ is reached with positive probability in G, and the set $Q \setminus Q_G$ is never hit, and therefore a counting function $f \in \mathcal{T}_H$ (such that $\mathsf{Supp}(f) \cap \mathcal{T} \neq \emptyset$) is reached with positive probability in H, and all plays remain safe in the set Good_H.

Second, assume that there exists a winning strategy σ^H for player 1 in H, and we show how to construct a pure winning strategy σ in G. We can assume that σ^H is pure memoryless. Fix an arbitrary strategy π for player 2 and consider the unfolding tree of the game H when σ^H and π are fixed (we get a tree and not just a path because the game is stochastic). In this tree, there is a shortest path to reach $\mathcal{T}_H$ and this path has no loop since strategy σ^H is memoryless. we show that the length of this path can be bounded, and that the bounds used in the counting abstraction with ω's are faithful, showing that the strategy σ^H can be simulated in G (in particular, we need to show that there are sufficiently many indistinguishable play prefixes in G to simulate the action 'functions' $\hat{a}$ played by σ^H). More precisely, the bounds $\mathsf{K}_1, \mathsf{K}_2, \ldots$ have been chosen in such a way that counters with value ω keep a positive value until all counters get value ω. For example, when all counters but k have value ω, it takes at most $(K_k)^k$ steps to get one more counter with value ω by the argument given in Step 3. Therefore, along the shortest path to $\mathcal{T}_H$, either we reach a counting function f with $f(q) = \omega$ for all $q \in \mathsf{Supp}(f)$, or a counting function f with $\mathsf{Supp}(f) \cap \mathcal{T} \neq \emptyset$. In the first case, we can simulate σ^H in G to this point, and then win by simulating a winning randomized strategy, and in the second case the reachability objective $\mathsf{Reach}(\mathcal{T})$ is achieved in G with positive probability. Since the strategy σ^H ensures that the support of the counting functions never hit the set $Q \setminus Q_G$, player 1 wins in G for the positive reachability and almost-sure safety objectives.

Theorem 3. *In one-sided partial-observation stochastic games with player 1 perfect and player 2 partial, non-elementary size memory is sufficient for pure strategies to ensure positive probability reachability along with almost-sure safety for player 1; and hence for pure positive winning strategies for reachability objectives for player 1 non-elementary memory bound is optimal.*

4.3 Upper bound for almost-sure reachability

In this section we present the algorithm to solve the almost-sure reachability problem. We start with an example to illustrate that in general strategies for almost-sure winning may be more complicated than positive winning for reachability objectives.

Example 4. Almost-sure winning strategy may require more memory than positive winning strategies. *The example of Figure 9 illustrates a key insight in the algorithmic solution of almost-sure reachability games where player 1 has perfect observation and player 2 has partial observation (he is blind in this case). For player 1, playing a in q_1 and in q_2 is a positive winning strategy to reach $q_{\text{☺}}$. This is because from $\{q_1, q_2\}$, the belief of player 2 becomes $\{q_3, q_4\}$ and no matter the action chosen by player 2, the state $q_{\text{☺}}$ is reached with positive probability from either q_3 or q_4.*

[3] Recall that Z is the set of states that are winning in G for player 1 in randomized strategies.

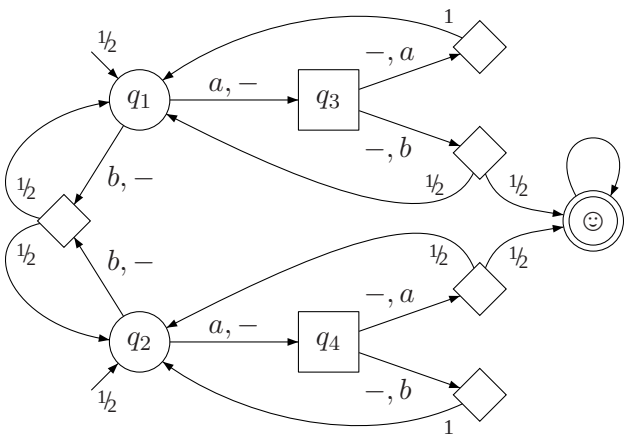

Figure 9. Almost-sure winning strategy may require more memory than positive winning strategies. A one-sided reachability game where player 1 (round states) has perfect observation, player 2 (square states) is blind. Player 1 has a pure almost-sure winning strategy, but no pure belief-based memoryless strategy is almost-sure winning. However, player 1 has a pure belief-based memoryless strategy that is positive winning.

However, always playing a when the belief of player 2 is $\{q_1, q_2\}$ is not almost-sure winning because if player 2 chooses always the same action (say a) in $\{q_3, q_4\}$, then with probability $\frac{1}{2}$ the state $q_{\text{☺}}$ is not reached. Intuitively, this happens because player 2 can guess that the initial state is, say q_1, and be right with positive probability (here $\frac{1}{2}$). To be almost-surely winning, player 1 needs to alternate actions a and b when the belief is $\{q_1, q_2\}$. The action b corresponds to the restart phase *of the strategy, i.e. even assuming that player 2's belief would be, say $\{q_1\}$, the action b ensures that $q_{\text{☺}}$ is reached with positive probability by make the belief to be $\{q_1, q_2\}$.* ■

Notation. We will consider $\mathcal{T}$ as the set of target states and without loss of generality assume that all target states are absorbing. In this section the belief of player 2 represents the set of states that can be with positive probability. Given strategies σ and π for player 1 and player 2, respectively, a state q and a set $K \subseteq Q$ we denote by $\Pr^{\sigma,\pi}_{q,K}(\cdot)$ the probability measure over sets of paths when the players play the strategies, the initial state is q and the initial belief for player 2 is K.

In rest of this section we omit the subscript G (such as we write Π^O instead of Π^O_G) as the game is clear from the context.

Bad states. Let $\overline{\mathcal{T}} = Q \setminus \mathcal{T}$. Let

$$Q_B = \{\, q \in Q \mid \forall \sigma \in \Sigma^P \cdot \exists \pi \in \Pi^O :\ \Pr^{\sigma,\pi}_{q,\{q\}}(\mathsf{Safe}(\overline{\mathcal{T}})) > 0 \,\}$$

be the set of states q such that given the initial belief of player 2 is the singleton $\{q\}$, for all pure strategies for player 1 there is a counter observation-based strategy for player 2 to ensure that $\mathsf{Safe}(\overline{\mathcal{T}})$ is satisfied with positive probability. We will consider Q_B as the set of *bad* states.

Property of an almost-sure winning strategy. Consider a pure almost-sure winning strategy for player 1 that ensures against all observation-based strategies of player 2 that $\mathcal{T}$ is reached with probability 1. Then we claim that the belief of player 2 must never intersect with Q_B: otherwise if the belief intersects with Q_B, let q be the state in Q_B that is reached with positive probability. Then player 2 simply assumes that the current state is q, updates the belief to $\{q\}$, and the guess is correct with positive probability. Given the belief is $\{q\}$, since $q \in Q_B$, it follows that against all player-1 pure strategies there is an observation-based strategy for player 2 to ensure with positive probability that $\mathcal{T}$ is not reached. This contradicts that the strategy for player 1 is almost-sure winning.

Transformation. We transform the game by changing all states in Q_B as absorbing. Let $Q_G = Q \setminus Q_B$. By definition we have

$$Q_G = \{\, q \in Q \mid \exists \sigma \in \Sigma^P \cdot \forall \pi \in \Pi^O : \Pr^{\sigma,\pi}_{q,\{q\}}(\mathsf{Reach}(\mathcal{T})) = 1 \,\}.$$

By the argument above that for a pure almost-sure winning strategy the belief must never intersect with Q_B we have

$$\begin{aligned} Q_G &= \{\, q \in Q \mid \exists \sigma \in \Sigma^P \cdot \forall \pi \in \Pi^O : \Pr^{\sigma,\pi}_{q,\{q\}}(\mathsf{Reach}(\mathcal{T})) = 1 \\ &\qquad \text{and } \Pr^{\sigma,\pi}_{q,\{q\}}(\mathsf{Safe}(Q \setminus Q_B)) = 1 \,\}. \end{aligned}$$

Let

$$\begin{aligned} Q^p_G &= \{\, q \in Q \mid \exists \sigma \in \Sigma^P \cdot \forall \pi \in \Pi^O : \Pr^{\sigma,\pi}_{q,\{q\}}(\mathsf{Reach}(\mathcal{T})) > 0 \\ &\qquad \text{and } \Pr^{\sigma,\pi}_{q,\{q\}}(\mathsf{Safe}(Q \setminus Q_B)) = 1 \,\}. \end{aligned}$$

We now show that $Q^p_G = Q_G$. The inclusion $Q_G \subseteq Q^p_G$ is trivial, and we now show the other inclusion $Q^p_G \subseteq Q_G$. Observe that in Q^p_G we have the property of positive reachability and almost-sure safety and we will use strategies for positive reachability and almost-sure safety to construct an almost-sure winning strategy. We consider Q_B as the set of unsafe states (i.e., Q_G is the safe set), and $\mathcal{T}$ as the target and invoke the results of the Section 4.2: for all $q \in Q^p_G$ there is a pure finite-memory strategy σ_q of memory at most B (where B is non-elementary) to ensure that from q, within $N = 2^{O(B)}$ steps, $\mathcal{T}$ is reached with probability at least some positive constant $\eta_q > 0$, even when the initial belief for player 2 is $\{q\}$. Let $\eta = \min_{q \in Q^p_G} \eta_q$. A pure finite-memory almost-sure winning strategy is described below. The strategy plays in two-phases: (1) the *Restart* phase; and (1) the *Play* phase. We define them as follows:

1. *Restart phase.* Let the current state be q, assume that the belief for player 2 is $\{q\}$ and goto the Play phase with strategy σ_q that ensures that Q_G is never left and $\mathcal{T}$ is reached within N steps with probability at least $\eta > 0$.
2. *Play phase.* Let σ be the strategy defined in the Restart phase, then play σ for N steps and go back to the Restart phase.

The strategy is almost-sure winning as for all states in Q^p_G and for all histories, in every N steps the probability to reach $\mathcal{T}$ is at least $\eta > 0$, and Q_G (and hence Q^p_G) is never left. Thus probability to reach $\mathcal{T}$ in $N \cdot \ell$ steps, for $\ell \in \mathbb{N}$, is at least $1-(1-\eta)^\ell$ and this is 1 as $\ell \to \infty$. Thus the desired result follows and we obtain the almost-sure winning strategy.

Memory bound and algorithm. The memory upper bound for the almost-sure winning strategy constructed is as follows: $|Q| \cdot B + \log N$, we require $|Q|$ strategies of Section 4.2 of memory size B and a counter to count up to $N = 2^{O(B)}$ steps. We now present an algorithm for almost-sure reachability that works in time $2^{|Q|} \times O(\text{POSREACHSURESAFE})$, where POSREACHSURESAFE denote the complexity to solve the positive reachability along with almost-sure safety problem. The algorithm enumerates all subset $Q' \subseteq Q$ and then verify that forall $q \in Q'$ player 1 can ensure to reach $\mathcal{T}$ with positive probability staying safe in Q' with probability 1. In other words the algorithm enumerates all subsets $Q' \subseteq Q$ to obtain the set Q_G. The enumeration is exponential and the verification requires solving the positive reachability with almost-sure safety problem.

Theorem 4. *In one-sided partial-observation stochastic games with player 1 perfect and player 2 partial, non-elementary size memory is sufficient for pure strategies to ensure almost-sure reachability for player 1; and hence for pure almost-sure winning strategies for reachability objectives for player 1 non-elementary memory bound is optimal.*

Corollary 2. *In one-sided partial-observation stochastic games with player 1 perfect and player 2 partial, the problem of deciding the existence of pure almost-sure and positive winning strategies for reachability objectives for player 1 can be solved in non-elementary time complexity.*

From the previous results and Remark 3 we obtain the following corollary.

Corollary 3. *The problem of deciding the existence of a pure almost-sure winning strategy for one-sided partial-observation stochastic games with player 1 perfect and player 2 partial, and Büchi objective for player 1, can be solved in non-elementary time complexity, and non-elementary memory is necessary and sufficient for pure almost-sure winning strategies.*

Discussion about the surprising non-elementary memory bound. We now discuss the surprising non-elementary memory bound for positive winning with reachability objectives for pure strategies in player-1 perfect player-2 partial stochastic games, comparing it with other related questions. We consider four related questions: two are related to stochasticity in transitions and strategies, and the other two are related to the information of the players.

1. *Question 1.* If we consider player-1 perfect player-2 partial deterministic games with reachability objective, then for positive winning pure memoryless strategies are sufficient. This follows from the results of [38] because in deterministic games positive winning coincides with sure winning, and the results of [38] shows (see [17] for an explicit proof) that for sure winning the observation of player 2 is irrelevant. Hence the problem is same as sure winning in perfect-information deterministic games with reachability objective for which pure memoryless strategies exist.
2. *Question 2.* If we consider player-1 perfect player-2 partial stochastic games with reachability objective, but instead of pure strategies consider randomized strategies, then memoryless strategies are sufficient. It follows from [7] that if there is a randomized strategy to ensure reachability with positive probability, then the randomized memoryless strategy that plays all actions uniformly at random is also a positive winning strategy.
3. *Question 3.* If we consider perfect-information stochastic games (both players have perfect information) with reachability objective, then for positive winning pure memoryless strategies are sufficient. This follows from a more general result of [18] that in perfect-information stochastic games with reachability objective, pure memoryless optimal strategies exist.
4. *Question 4.* If we consider player-1 partial player-2 perfect stochastic games with reachability objective, then for positive winning exponential memory pure strategies are sufficient (by Theorem 1).

Observe that the question we study is a very natural extension of the above questions: (1) adding stochasticity to the transition as compared to question 1; (2) restricting strategies to pure strategies as compared to randomized strategies of question 2; (3) player 2 is less informed as compared to question 3; and (4) player 1 is more informed and player 2 is less informed as compared to question 4. Our results show the natural variant of question 1 and question 2 obtained by adding stochasticity to transitions or removing stochasticity from strategies; and the variant of question 3 and question 4 by making player 1 most well informed lead to a sunrising memory bound for strategies (non-elementary complete memory bound, whereas for all the related questions memoryless or exponential-size memory strategies are sufficient). Also see Figure 10 for a pictorial illustration.

5 Finite-memory Strategies for Two-sided Games

In this section we show the existence of finite-memory pure strategies for positive and almost-sure winning in two-sided games.

5.1 Positive reachability with almost-sure safety

Let $\mathcal{T}$ be the set of target states for reachability (such that all the target states are absorbing) and Q_G be the set of good states for safety with $\mathcal{T} \subseteq Q_G$. Our goal is to show that for pure strategies to ensure positive probability reachability to $\mathcal{T}$ and almost-sure safety for Q_G, finite-memory strategies suffice. Note that with Q_G as the whole state space we obtain the result for positive reachability as a special case.

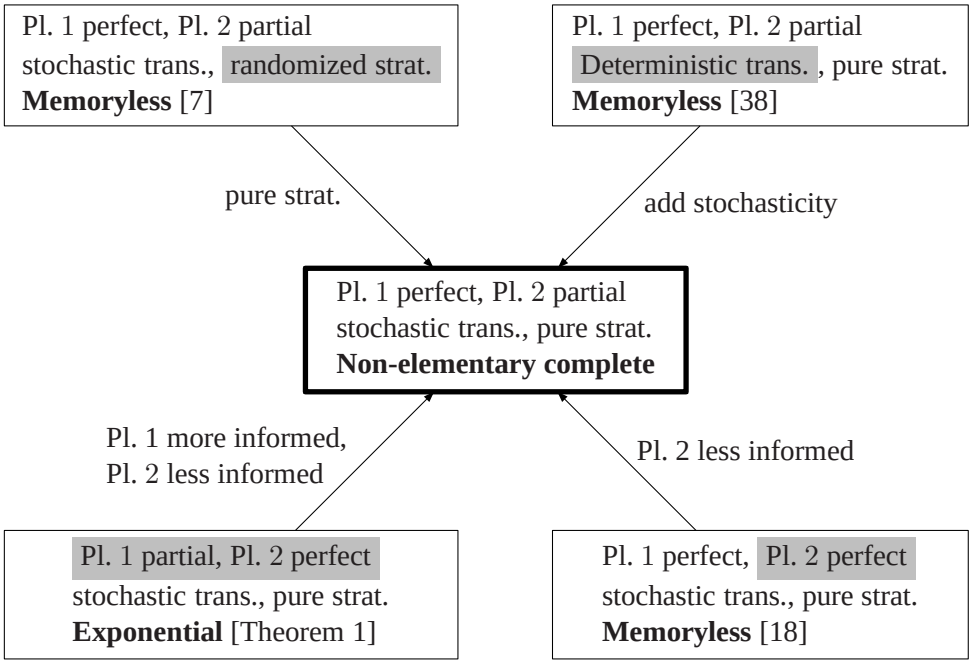

Figure 10. The surprising non-elementary bound for memory of pure strategies in one-sided partial-observation stochastic games for player 1 perfect and player 2 partial for positive winning with reachability objectives (Theorem 4).

Lemma 2. *For all games G, for all $q \in Q$, if there exists a pure strategy $\sigma \in \Sigma^O \cap \Sigma^P$ such that for all strategies $\pi \in \Pi^O$ of player 2 we have*

$$\Pr_q^{\sigma,\pi}(\mathsf{Reach}(\mathcal{T})) > 0 \quad \textit{and} \quad \Pr_q^{\sigma,\pi}(\mathsf{Safe}(Q_G)) = 1;$$

then there exists a finite-memory pure strategy $\sigma^f \in \Sigma^O \cap \Sigma^P$ such that for all strategies $\pi \in \Pi^O$ of player 2 we have

$$\Pr_q^{\sigma^f,\pi}(\mathsf{Reach}(\mathcal{T})) > 0 \quad \textit{and} \quad \Pr_q^{\sigma^f,\pi}(\mathsf{Safe}(Q_G)) = 1.$$

We prove the result with the following two claims. We fix a (possibly infinite memory) strategy $\sigma \in \Sigma^O \cap \Sigma^P$ such that for all strategies $\pi \in \Pi^O$ of player 2 we have

$$\Pr_q^{\sigma,\pi}(\mathsf{Reach}(\mathcal{T})) > 0 \quad \text{and} \quad \Pr_q^{\sigma,\pi}(\mathsf{Safe}(Q_G)) = 1.$$

Claim 1. If there exists $N \in \mathbb{N}$ such that for all strategies $\pi \in \Pi^O$ of player 2 we have

$$\Pr_q^{\sigma,\pi}(\mathsf{Reach}^{\leq N}(\mathcal{T})) > 0 \quad \text{and} \quad \Pr_q^{\sigma,\pi}(\mathsf{Safe}(Q_G)) = 1$$

where $\mathsf{Reach}^{\leq N}$ denotes reachability within first N-steps; then there exists a finite-memory pure strategy $\sigma^f \in \Sigma^O \cap \Sigma^P$ such that for all strategies $\pi \in \Pi^O$ of player 2 we have

$$\Pr_q^{\sigma^f,\pi}(\mathsf{Reach}(\mathcal{T})) > 0 \quad \text{and} \quad \Pr_q^{\sigma^f,\pi}(\mathsf{Safe}(Q_G)) = 1.$$

Proof. The finite-memory strategy σ^f is as follows: play like the strategy σ for the first N-steps, and then switch to a strategy to ensure $\mathsf{Safe}(Q_G)$ with probability 1. The strategy ensure positive probability reachability to $\mathcal{T}$ as for the first N-steps it plays like σ and σ already ensures positive reachability within N-steps. Moreover, since σ ensures $\mathsf{Safe}(Q_G)$ with probability 1, it must also ensure $\mathsf{Safe}(Q_G)$ for the first N-steps, and since σ^f after the first N-steps only plays a strategy for almost-sure safety, it follows that σ^f guarantees $\mathsf{Safe}(Q_G)$ with probability 1. The strategy σ^f is a finite-memory strategy since it needs to play like σ for the first N-steps (which requires finite-memory) and then it switches to an almost-sure safety strategy for which exponential size memory is sufficient (for safety objective almost-sure winning coincides with sure winning and then belief-based strategies are sufficient; see [14] for details). □

Claim 2. There exists $N \in \mathbb{N}$ such that for all strategies $\pi \in \Pi^O$ of player 2 we have

$$\Pr_q^{\sigma,\pi}(\mathsf{Reach}^{\leq N}(\mathcal{T})) > 0 \quad \text{and} \quad \Pr_q^{\sigma,\pi}(\mathsf{Safe}(Q_G)) = 1$$

where $\mathsf{Reach}^{\leq N}$ denotes reachability within first N-steps.

Proof. The proof is by contradiction. Towards contradiction, assume that for all $n \in \mathbb{N}$, there exists a strategy $\pi_n \in \Pi^O$ such that either $\Pr_q^{\sigma,\pi_n}(\mathsf{Reach}^{\leq n}(\mathcal{T})) = 0$ or $\Pr_q^{\sigma,\pi_n}(\mathsf{Safe}(Q_G)) < 1$.

If for some $n \geq 0$ we have $\Pr_q^{\sigma,\pi_n}(\mathsf{Safe}(Q_G)) < 1$, then we get a contradiction with the fact that $\Pr_q^{\sigma,\pi}(\mathsf{Safe}(Q_G)) = 1$ for all $\pi \in \Pi^O$. Hence $\Pr_q^{\sigma,\pi_n}(\mathsf{Safe}(Q_G)) = 1$ for all $n \in \mathbb{N}$, and therefore $\Pr_q^{\sigma,\pi_n}(\mathsf{Reach}^{\leq n}(\mathcal{T})) = 0$ for all $n \in \mathbb{N}$. Equivalently, all play prefixes of length at most n and compatible with σ and π_n avoid to hit $\mathcal{T}$, and thus $\Pr_q^{\sigma,\pi_n}(\mathsf{Safe}^{\leq n}(Q \setminus \mathcal{T})) = 1$ for all $n \in \mathbb{N}$. Note that we can assume that each strategy π_n is pure because once the strategy σ of player 1 is fixed we get a POMDP for player 2, and for POMDPs pure strategies are as powerful as randomized strategies [15] (in [15] the result was shown for finite POMDPs with finite action set, but the proof is based on induction on the action set and also works for countably infinite POMDPs).

Using a simple extension of König's Lemma [30], we construct a strategy $\pi' \in \Pi^O$ such that $\Pr_q^{\sigma,\pi'}(\mathsf{Safe}(Q \setminus \mathcal{T})) = 1$. The construction is as follows. In the initial state q, there is an action $b_0 \in A_2$ which is played by infinitely many strategies π_n. We define $\pi'(q) = b_0$ and let P_0 be the set $\{\pi_n \mid \pi_n(q) = b_0\}$. Note that P_0 is an infinite set. We complete the construction as follows. Having defined $\pi'(\rho)$ for all play prefixes ρ of length at most k, and given the infinite set P_k, we define $\pi'(\rho')$ for all play prefixes ρ' of length $k+1$ and the infinite set P_{k+1} as follows. Consider the tuple $b_{\pi_n} \in A_2^m$ of actions played by the strategy $\pi_n \in P_k$ after the m prefixes ρ' of length $k+1$. Clearly, there exists an infinite subset P_{k+1} of P_k in which all strategies play the same tuple b_{k+1}. We define $\pi(\rho')$ using the tuple b_{k+1}. This construction ensures that no play prefix of length $k+1$ compatible with σ and π' hit the set $\mathcal{T}$, since π' agrees with some strategy π_n for arbitrarily large n. Repeating this inductive argument yields a strategy π' such that $\Pr_q^{\sigma,\pi'}(\mathsf{Safe}(Q \setminus \mathcal{T})) = 1$, in contradiction with the fact that $\Pr_q^{\sigma,\pi}(\mathsf{Reach}(\mathcal{T})) > 0$ for all $\pi \in \Pi^O$. Hence, the desired result follows. □

The above two claims establish Lemma 2 and gives the following result.

Theorem 5. *In two-sided partial-observation stochastic games finite memory is sufficient for pure strategies to ensure positive probability reachability along with almost-sure safety for player 1; and hence for pure positive winning strategies for reachability objectives finite memory is sufficient and non-elementary memory is required in general for player 1.*

5.2 Almost-sure reachability

We now show that for pure strategies for almost-sure reachability, finite-memory strategies suffice. The proof is a straight forward extension of the results of Section 4.3, and for finite-memory strategies for positive reachability with almost-sure safety we use the result of the previous subsection.

Notation. We will consider $\mathcal{T}$ as the set of target states and without loss of generality assume that all target states are absorbing. In this section the belief of player 2 represents the set of states that can be with positive probability. Given strategies σ and π for player 1 and player 2, respectively, a state q and a set $K \subseteq Q$ we denote by $\Pr^{\sigma,\pi}_{q,K}(\cdot)$ the probability distribution when the players play the strategies, the initial state is q and the initial belief for player 2 is K.

In rest of this section we omit subscript G (such as we write Π^O instead of Π^O_G) as the game is clear from the context.

Bad beliefs. Let $\overline{\mathcal{T}} = Q \setminus \mathcal{T}$. Let

$$Q_B = \{\, \mathcal{B} \in 2^Q \mid \forall \sigma \in \Sigma^O \cap \Sigma^P \cdot \exists \pi \in \Pi^O \cdot \exists q \in \mathcal{B} : \Pr^{\sigma,\pi}_{q,\{q\}}(\mathsf{Safe}(\overline{\mathcal{T}})) > 0 \,\}$$

be the set of beliefs $\mathcal{B}$ such that for all pure strategies for player 1 there is a counter strategy for player 2 with a state $q \in \mathcal{B}$ to ensure that given the initial belief of player 2 is the singleton $\{q\}$, $\mathsf{Safe}(\overline{\mathcal{T}})$ is satisfied with positive probability. We will consider Q_B as the set of *bad* beliefs.

Property of an almost-sure winning strategy. Consider a pure almost-sure winning strategy for player 1 that ensures against all strategies of player 2 that $\mathcal{T}$ is reached with probability 1. Then we claim that the belief of player 2 must never intersect with Q_B: otherwise if the belief intersects with Q_B, let $\mathcal{B}$ be the belief in Q_B that is reached with positive probability. Then there exists $q \in \mathcal{B}$ such that player 2 can simply assume that the current state is q, update the belief to $\{q\}$, and the guess is correct with positive probability, and then player 2 can ensure that against all player-1 pure strategies there is a strategy for player 2 to ensure with positive probability that $\mathcal{T}$ is not reached. This contradicts that the strategy for player 1 is almost-sure winning. Let $Q_G = 2^Q \setminus Q_B$. By definition we have

$$Q_G = \{\, \mathcal{B} \in 2^Q \mid \exists \sigma \in \Sigma^O \cap \Sigma^P \cdot \forall \pi \in \Pi^O \cdot \forall q \in \mathcal{B} : \Pr^{\sigma,\pi}_{q,\{q\}}(\mathsf{Reach}(\mathcal{T})) = 1 \,\}.$$

By the argument above that for a pure almost-sure winning strategy the belief must never intersect with Q_B we have

$$\begin{aligned} Q_G \;=\; \{\, \mathcal{B} \in 2^Q \mid \exists \sigma \in \Sigma^O \cap \Sigma^P \cdot \forall \pi \in \Pi^O \cdot \forall q \in \mathcal{B} : \Pr^{\sigma,\pi}_{q,\{q\}}(\mathsf{Reach}(\mathcal{T})) = 1 \\ \text{and } \Pr^{\sigma,\pi}_{q,\{q\}}(\mathsf{Safe}(2^Q \setminus Q_B)) = 1 \,\}. \end{aligned}$$

Let

$$\begin{aligned} Q^p_G \;=\; \{\, \mathcal{B} \in 2^Q \mid \exists \sigma \in \Sigma^O \cap \Sigma^P \cdot \forall \pi \in \Pi^O \cdot \forall q \in \mathcal{B} : \Pr^{\sigma,\pi}_{q,\{q\}}(\mathsf{Reach}(\mathcal{T})) > 0 \\ \text{and } \Pr^{\sigma,\pi}_{q,\{q\}}(\mathsf{Safe}(2^Q \setminus Q_B)) = 1 \,\}. \end{aligned}$$

We now show that $Q^p_G = Q_G$. The inclusion $Q_G \subseteq Q^p_G$ is trivial, and we now show the other inclusion $Q^p_G \subseteq Q_G$. Observe that in Q^p_G we have the property of positive reachability and almost-sure safety and we will use strategies for positive reachability and almost-sure safety to construct a witness finite-memory almost-sure winning strategy. Note that here we have safety for a set of beliefs (instead of set of states, and it is straight forward to verify that the argument of the previous subsection holds when the safe set is a set of beliefs). We consider Q_B as the set of unsafe beliefs (i.e., Q_G is the safe set), and $\mathcal{T}$ as the target and invoke the results of the previous subsection: for all $\mathcal{B} \in Q^p_G$ there is a pure finite-memory strategy $\sigma_{\mathcal{B}}$ of to ensure that from all states $q \in \mathcal{B}$, within N steps (for some finite $N \in \mathbb{N}$), $\mathcal{T}$ is reached with probability at least some positive constant $\eta_{\mathcal{B}} > 0$, even when the initial belief for player 2 is $\{q\}$. Let $\eta = \min_{\mathcal{B} \in Q^p_G} \eta_{\mathcal{B}}$. A pure finite-memory almost-sure winning strategy is described below. The strategy plays in two-phases: (1) the *Restart* phase; and (1) the *Play* phase. We define them as follows:

1. *Restart phase.* Let the current belief be $\mathcal{B}$, the belief for player 2 is any perfect belief $\{q\}$, for $q \in \mathcal{B}$; and goto the Play phase with strategy $\sigma_{\mathcal{B}}$ that ensures that Q_G is never left and $\mathcal{T}$ is reached within N steps with probability at least $\eta > 0$.
2. *Play phase.* Let σ be the strategy defined in the Restart phase, then play σ for N steps and go back to the Restart phase.

The strategy is almost-sure winning as for all states in Q_G^p and for all histories, in every N steps the probability to reach $\mathcal{T}$ is at least $\eta > 0$, and Q_G (and hence Q_G^p) is never left. Thus probability to reach $\mathcal{T}$ in $N \cdot \ell$ steps, for $\ell \in \mathbb{N}$, is at least $1 - (1-\eta)^\ell$ and this is 1 as $\ell \to \infty$. Thus the desired result follows and we obtain the required finite-memory almost-sure winning strategy.

Memory bound and algorithm. The memory upper bound for the almost-sure winning strategy constructed is as follows: $|2^Q| \cdot B + \log N$, we require $|2^Q|$ strategies of the previous subsection of memory size B and a counter to count up to N steps; where B is the memory required for strategies to ensure positive reachability with almost-sure safety objectives.

Theorem 6. *In two-sided partial-observation stochastic games, finite memory is sufficient (and non-elementary memory is required in general) for pure strategies for almost-sure winning for reachability and Büchi objectives for player 1.*

6 Equivalence of Randomized Action-invisible Strategies and Pure Strategies

In this section, we show that for two-sided partial-observation games, the problem of almost-sure winning with randomized action-invisible strategies is inter-reducible with the problem of almost-sure winning with pure strategies. The reductions are polynomial in the number of states in the game (the reduction from randomized to pure strategies is exponential in the number of actions).

It follows from the reduction of pure to randomized action-invisible strategies that the memory lower bounds for pure strategies transfer to randomized strategies, and in particular belief-based memoryless strategies are not sufficient, showing that a remark (without proof) of [17, p.4] and the result and construction of [27, Theorem 1] are wrong.

6.1 Reduction of randomized action-invisible strategies to pure strategies

We give a reduction for almost-sure winning for randomized action-invisible strategies to pure strategies. Given a stochastic game G we will construct another stochastic game H such that there is a randomized action-invisible almost-sure winning strategy in G iff there is a pure almost-sure winning strategy in H. We first show in Lemma 3 the correctness of the reduction for finite-memory randomized action-invisible strategies, and then show in Lemma 4 that finite memory is sufficient in two-sided partial-observation games for randomized action-invisible strategies.

Construction. Given a stochastic game $G = \langle Q, q_0, \delta \rangle$ over action sets A_1 and A_2, and observations $\mathcal{O}_1$ and $\mathcal{O}_2$ (along with the corresponding observation mappings obs_1 and obs_2), we construct a game $H = \langle Q, q_0, \delta_H \rangle$ over action sets $2^{A_1} \setminus \{\emptyset\}$ and A_2 and observations $\mathcal{O}_1$ and $\mathcal{O}_2$. The transition function δ_H is defined as follows:

- for all $q \in Q$ and $A \in 2^{A_1} \setminus \{\emptyset\}$ and $b \in A_2$ we have $\delta_H(q, A, b)(q') = \frac{1}{|A|} \cdot \sum_{a \in A} \delta(q, a, b)(q')$, i.e., in a state in Q player 1 selects a non-empty subset $A \subseteq A_1$ of actions and the transition function δ_H simulates the transition function δ along with the uniform distribution over the set A of actions.

The observation mappings obs_i^H in H, for $i \in \{1, 2\}$ are as follows: $\mathsf{obs}_i^H(q) = \mathsf{obs}_i(q)$, where obs_i is the observation mapping in G.

Lemma 3. *The following assertions hold for reachability objectives:*

1. *If there is a pure almost-sure winning strategy in H, then there is a randomized action-invisible almost-sure winning strategy in G.*
2. *If there is a finite-memory randomized action-invisible almost-sure winning strategy in G, then there is a pure almost-sure winning strategy in H.*

Proof. We present both parts of the proof below.

1. Let σ_H be a pure almost-sure winning strategy in H. We construct a randomized action-invisible almost-sure winning strategy σ_G in G. The strategy σ_G is as constructed as follows. Let $\rho_G = q_0 q_1 \dots q_k$ be a play prefix in G, and we consider the same play prefix $\rho_H = q_0 q_1 \dots q_k$ in H, and let $A_k = \sigma_H(\rho_H)$. The strategy $\sigma_G(\rho_G)$ plays all actions in A_k uniformly at random. Since σ_H is an almost-sure winning strategy it follows σ_G is also almost-sure winning. Also observe that if σ_H is observation-based, then so is σ_G.
2. Let σ_G be a finite-memory randomized action-invisible almost-sure winning strategy in G. If the strategy σ_G is fixed in G we obtain a finite POMDP, and by the results of [16] it follows that in an POMDP the precise transition probabilities do not affect almost-sure winning. Hence if σ_G is almost-sure winning, then the uniform version σ_G^u of the strategy σ_G that always plays the same support of the probability distribution as σ_G but plays all actions in the support uniformly at random is also almost-sure winning. Given σ_G^u we construct a pure almost-sure winning strategy σ_H in H. Given a play prefix $\rho_H = q_0 q_1 \dots q_k$ in H, consider the same play prefix $\rho_G = q_0 q_1 \dots q_k$ in G. Let $A_k = \mathsf{Supp}(\sigma_G^u(\rho_G))$, then $\sigma_H(\rho_H)$ plays the action $A_k \in (2^{A_1} \setminus \{\emptyset\})$. Since σ_G^u is almost-sure winning it follows that σ_H is almost-sure winning. Observe that if σ_G is observation-based, then so is σ_G^u, and then so is σ_H.

The desired result follows. □

Lemma 4. *For reachability objectives, if there exists a randomized action-invisible almost-sure winning strategy in G, then there exists also a finite-memory randomized action-invisible almost-sure winning strategy in G.*

Proof. Let $\mathcal{W} = \{ \mathcal{B} \mid \mathcal{B} \in 2^Q$ is the belief of player 1 such that $\exists \sigma \in \Sigma^O \cdot \forall \pi \in \Pi^O \cdot \forall q \in \mathcal{B} : \Pr_q^{\sigma,\pi}(\mathsf{Reach}(\mathcal{T})) = 1 \}$ denote the set of belief sets $\mathcal{B}$ for player 1 such that player 1 has a (possibly infinite-memory) randomized action-invisible almost-sure winning strategy from all starting states in $\mathcal{B}$. It follows that the almost-sure winning strategy must ensure that the set $\mathcal{W}$ is never left: this is because from the complement set of $\mathcal{W}$ against all randomized action-invisible strategies for player 1 there is a counter strategy for player 2 to ensure that with positive probability the target is not reached. Moreover for all $\mathcal{B} \in \mathcal{W}$ the almost-sure winning strategy also ensures that $\mathcal{T}$ is reached with positive probability. Hence we have again the problem of positive reachability with almost-sure safety. We simply repeat the proof for the pure strategy case, treating sets of actions (that is the support of the randomized strategy) as actions (for pure strategy) and played uniformly at random (as in the reduction from G to H), and thus obtain a witness finite-memory strategy σ_G to ensure positive reachability and almost-sure safety. Repeating the strategy σ_G with play phase and repeat phase (as in the case of pure strategies) we obtain the desired finite-memory almost-sure winning strategy. □

The following theorem follows from the previous two lemmas.

Theorem 7. *Given a two-sided (resp. one-sided) partial-observation stochastic game G with a reachability objective we can construct in time polynomial in the size of the game and exponential in the size of the action sets a two-sided (resp. one-sided) partial-observation stochastic game H such that there exists a randomized action-invisible almost-sure winning strategy in G iff there exists a pure almost-sure winning strategy in H.*

For positive winning, randomized memoryless strategies are sufficient (both for action-visible and action-invisible) and the problem is PTIME-complete for one-sided and EXPTIME-complete for two-sided [7]. The above theorem along with Theorem 1 gives us the following corollary for almost-sure winning for randomized action-invisible strategies.

Corollary 4. *Given one-sided partial-observation stochastic games with player 1 partial and player 2 perfect, the following assertions hold for reachability objectives for player 1:*

1. (Memory complexity). *Exponential memory (of size $\sum_{\gamma \in \mathcal{O}_1} 3^{|\gamma|}$) is sufficient for randomized action-invisible strategies for almost-sure winning.*

2. (Algorithm). *The existence of a randomized action-invisible almost-sure winning strategy can be decided in time exponential in the state space of the game and exponential in the size of the action sets.*

3. (Complexity). *The problem of deciding the existence of a randomized action-invisible almost-sure winning strategy is EXPTIME-complete.*

Corollary 5. *The problem of deciding the existence of a pure almost-sure winning strategy for one-sided partial-observation stochastic games with player 1 partial and player 2 perfect, and Büchi objective for player 1 is EXPTIME-complete, and memory of size* $\sum_{\gamma \in \mathcal{O}_1} 3^{|\gamma|}$ *is sufficient for pure winning strategies.*

6.2 Reduction of pure strategies to randomized action-invisible strategies

We present a reduction for almost-sure winning for pure strategies to randomized action-invisible strategies. Given a stochastic game G we construct another stochastic game H such that there exists a pure almost-sure winning strategy in G iff there exists a randomized almost-sure winning strategy in H.

The idea of the reduction is to force player 1 to play a pure strategy in H. The game H simulates G and requires player 1 to repeat each actions played (i.e. to play each action two times). Then, if player 1 uses randomization, he has to repeat the actions chosen randomly in the previous step. Since the actions are invisible, this can be achieved only if the support of the randomized actions is a singleton, i.e., the strategy is pure. Note that the reduction works for randomized strategies with actions invisible, and not when the actions are visible.

Construction. Given a stochastic game $G = \langle Q, q_0, \delta_G \rangle$ over action sets A_1 and A_2, and observations $\mathcal{O}_1$ and $\mathcal{O}_2$ (along with the corresponding observation mappings obs_1 and obs_2), we construct a game $H = \langle Q \cup (Q \times A_1) \cup \{\mathsf{sink}\}, q_0, \delta_H \rangle$ over the same action sets A_1 and A_2 and observations $\mathcal{O}_1$ and $\mathcal{O}_2$. The transition function δ_H is defined as follows:

- for all $q \in Q$ and $a \in A_1$ and $b \in A_2$ we have $\delta_H(q, a, b)((q, a)) = 1$, i.e., in a state q for action a of player 1, irrespective of the choice of player 2, the game stores player 1's action with probability 1;
- for all $(q, a) \in Q \times A_1$, for all $b \in A_2$ we have $\delta_H((q, a), a, b) = \delta_G(q, a, b)$, i.e. if player 1 repeats the action played in the previous step, then the probabilistic transition function is the same as in G; and for all $a' \in A_1 \setminus \{a\}$, we have $\delta_H((q, a), a', b)(\mathsf{sink}) = 1$, i.e. if player 1 does not repeat the same action, then the sink state is reached.
- for all $a \in A_1$ and $b \in A_2$, we have $\delta_H(\mathsf{sink}, a, b)(\mathsf{sink}) = 1$.

The observation mappings obs_i^H in H ($i \in \{1, 2\}$) are as follows: $\mathsf{obs}_i^H(q) = \mathsf{obs}_i^H((q, a)) = \mathsf{obs}_i(q)$, where obs_i is the observation mapping in G. Note that H is of size polynomial in the size of G.

Lemma 5. *Let $\mathcal{T} \subseteq Q$ be a set of target states. There exists a pure almost-sure winning strategy in G for $\mathsf{Reach}(\mathcal{T})$ if and only if there exists a randomized action-invisible almost-sure winning strategy in H for objective $\mathsf{Reach}(\mathcal{T})$.*

Proof. We present both directions of the proof below.

1. Let σ_H be a randomized action-invisible almost-sure winning strategy in H. We show that we can assume wlog that σ_H is actually a pure strategy. To see this, assume that under strategy σ_H there is a prefix $\rho_H = q_0(q_0, a_0)q_1(q_1, a_1) \dots q_k$ in H compatible with σ_H from which σ_H plays a randomized action with support $A \subseteq A_1$ and $|A| > 1$. Then, with positive probability the states (q_k, a_k) and (q_k, a'_k) are reached where $a_k, a'_k \in A$ and $a_k \neq a'_k$. No matter the action(s) played by σ_H in the next step, the state sink is reached with positive probability in the next step, either from (q_k, a_k) or from (q_k, a'_k). This contradicts that σ_H is almost-sure winning. Therefore, we can assume that σ_H is a pure strategy that repeats each action two times. We construct a pure almost-sure winning strategy in G by removing these repetitions.

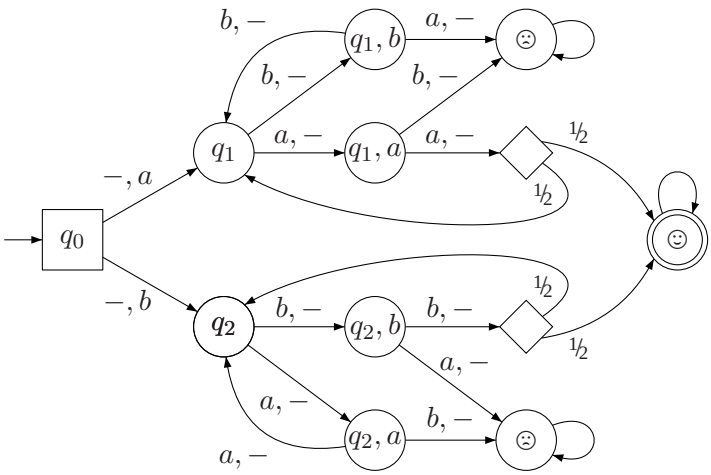

Figure 11. Belief-based strategies are not sufficient. The game graph obtained by the reduction of pure to randomized strategies on the game of Figure 1 (for almost-sure reachability objective). Player 1 is blind and player 2 has perfect observation. There exists an almost-sure winning randomized strategy (with invisible actions), but there is no *belief-based memoryless* almost-sure winning randomized strategy.

2. Let σ_G be a pure almost-sure winning strategy in G. Consider the strategy σ_H in H that always repeats two times the actions played by σ_G. The strategy σ_H is observation-based and almost-sure winning since H simulates G when actions are repeated twice.

The desired result follows. □

Theorem 8. *Given a two-sided partial-observation stochastic game G with a reachability objective we can construct in time polynomial in the size of the game and size of the action sets a two-sided partial-observation stochastic game H such that there exists a pure almost-sure winning strategy in G iff there exists a randomized action-invisible almost-sure winning strategy in H.*

Belief-based strategies are not sufficient. We illustrate our reduction with the following example that shows belief-based (belief-only) randomized action-invisible strategies are not sufficient for almost-sure reachability in one-sided partial-observation games (player 1 partial and player 2 perfect), showing that a remark (without proof) of [17, p.4] and the result and construction of [27, Theorem 1] are wrong.

Example 5. *We illustrate the reduction of on the example of Figure 1. The result of the reduction is given in Figure 11. Remember that Example 1 showed that belief-based pure strategies are not sufficient for almost-sure winning. We show that belief-based randomized strategies are not sufficient for almost-sure winning in the game of Figure 11. First, in $\{q_1, q_2\}$ player 1 has to play pure since he has to be able to repeat the same action to avoid reaching a sink state ☹ with positive probability. Now, the argument is the same as in Example 1: playing always the same action (either a or b) in $\{q_1, q_2\}$ is not even positive winning as player 2 can choose the state in this set (either q_2 or q_1).* ■

Note that our reduction preserves the structure and memory of almost-sure winning strategies, hence the non-elementary lower bound given in Theorem 2 for pure strategies also transfers to randomized action-invisible strategies by the same reduction.

Corollary 6. *For one-sided partial-observation stochastic games, with player 1 partial and player 2 perfect, belief-based randomized action-invisible strategies are not sufficient for almost-sure winning for reachability objectives. For two-sided partial-observation stochastic games, memory of non-elementary size is necessary in general for almost-sure winning for randomized action-invisible strategies for reachability and Büchi objectives.*

References

[1] M. Abadi, L. Lamport, and P. Wolper. Realizable and unrealizable specifications of reactive systems. In *Proc. of ICALP*, LNCS 372, pages 1–17. Springer, 1989.

[2] R. Alur, T. A. Henzinger, and O. Kupferman. Alternating-time temporal logic. *Journal of the ACM*, 49:672–713, 2002.

[3] B. Aminof, A. Murano, and M. Y. Vardi. Pushdown module checking with imperfect information. In *Proc. of CONCUR*, LNCS 4703, pages 460–475. Springer, 2007.

[4] C. Baier, N. Bertrand, and M. Größer. On decision problems for probabilistic Büchi automata. In *Proc. of FoSSaCS*, LNCS 4962, pages 287–301. Springer, 2008.

[5] C. Baier, N. Bertrand, and M. Größer. The effect of tossing coins in omega-automata. In *Proc. of CONCUR*, LNCS 5710, pages 15–29. Springer, 2009.

[6] C. Baier and M. Größer. Recognizing omega-regular languages with probabilistic automata. In *Proc. of LICS*, pages 137–146, 2005.

[7] N. Bertrand, B. Genest, and H. Gimbert. Qualitative determinacy and decidability of stochastic games with signals. In *Proc. of LICS*, pages 319–328, 2009.

[8] D. Berwanger and L. Doyen. On the power of imperfect information. In *Proc. of FSTTCS*, Dagstuhl Seminar Proceedings 08004. IBFI, 2008.

[9] R. G. Bukharaev. Probabilistic automata. *Journal of Mathematical Sciences*, 13:359–386, 1980.

[10] P. Cerný, K. Chatterjee, T. A. Henzinger, A. Radhakrishna, and R. Singh. Quantitative synthesis for concurrent programs. In *Proc. of CAV*, LNCS 6806, pages 243–259. Springer, 2011.

[11] R. Chadha, A. P. Sistla, and M. Viswanathan. On the expressiveness and complexity of randomization in finite state monitors. *Journal of the ACM*, 56:1–44, 2009.

[12] R. Chadha, A. P. Sistla, and M. Viswanathan. Power of randomization in automata on infinite strings. In *Proc. of CONCUR*, LNCS 5710, pages 229–243. Springer, 2009.

[13] R. Chadha, A. P. Sistla, and M. Viswanathan. Model checking concurrent programs with nondeterminism and randomization. In *Proc. of FSTTCS*, volume 8 of *LIPIcs*, pages 364–375, 2010.

[14] K. Chatterjee and L. Doyen. The complexity of partial-observation parity games. In *Proc. of LPAR*, LNCS 6397, pages 1–14. Springer, 2010.

[15] K. Chatterjee, L. Doyen, H. Gimbert, and T. A. Henzinger. Randomness for free. In *Proc. of MFCS*, LNCS 6281, pages 246–257. Springer, 2010.

[16] K. Chatterjee, L. Doyen, and T. A. Henzinger. Qualitative analysis of partially-observable Markov decision processes. In *Proc. of MFCS*, LNCS 6281, pages 258–269. Springer, 2010.

[17] K. Chatterjee, L. Doyen, T. A. Henzinger, and J.-F. Raskin. Algorithms for omega-regular games of incomplete information. *Logical Methods in Computer Science*, 3(3:4), 2007.

[18] A. Condon. The complexity of stochastic games. *Information and Computation*, 96(2):203–224, 1992.

[19] L. de Alfaro and T. A. Henzinger. Interface automata. In *Proc. of FSE*, pages 109–120. ACM Press, 2001.

[20] L. de Alfaro, T. A. Henzinger, and O. Kupferman. Concurrent reachability games. *Theor. Comput. Sci.*, 386(3):188–217, 2007.

[21] M. De Wulf, L. Doyen, and J.-F. Raskin. A lattice theory for solving games of imperfect information. In *Proc. of HSCC*, LNCS 3927, pages 153–168. Springer, 2006.

[22] D. L. Dill. *Trace Theory for Automatic Hierarchical Verification of Speed-independent Circuits*. The MIT Press, 1989.

[23] R. Dimitrova and B. Finkbeiner. Abstraction refinement for games with incomplete information. In *Proc. of FSTTCS*, volume 2 of *LIPIcs*, pages 175–186, 2008.

[24] L. Doyen and J.-F. Raskin. Antichains algorithms for finite automata. In *Proc. of TACAS*, LNCS 6015, pages 2–22. Springer, 2010.

[25] E. A. Emerson and C. Jutla. Tree automata, mu-calculus and determinacy. In *Proc. of FOCS*, pages 368–377, 1991.

[26] H. Gimbert and Y. Oualhadj. Probabilistic automata on finite words: Decidable and undecidable problems. In *Proc. of ICALP (2)*, LNCS 6199, pages 527–538. Springer, 2010.

[27] V. Gripon and O. Serre. Qualitative concurrent stochastic games with imperfect information. In *Proc. of ICALP (2)*, LNCS 5556, pages 200–211. Springer, 2009.

[28] T. A. Henzinger and P.W. Kopke. Discrete-time control for rectangular hybrid automata. *Theor. Comp. Science*, 221:369–392, 1999.

[29] A. Kechris. *Classical Descriptive Set Theory*. Springer, 1995.

[30] D. König. *Theorie der endlichen und unendlichen Graphen*. Akademische Verlagsgesellschaft, Leipzig, 1936.

[31] H. Kress-Gazit, G. E. Fainekos, and G. J. Pappas. Temporal-logic-based reactive mission and motion planning. *IEEE Transactions on Robotics*, 25(6):1370–1381, 2009.

[32] O. Kupferman and M. Y. Vardi. Synthesis with incomplete informatio. In *Advances in Temporal Logic*, pages 109–127. Kluwer Academic Publishers, 2000.

[33] C. H. Papadimitriou and J. N. Tsitsiklis. The complexity of Markov decision processes. *Math. of Op. Research*, 12:441–450, 1987.

[34] A. Paz. *Introduction to probabilistic automata*. Academic Press, Inc. Orlando, FL, USA, 1971.

[35] A. Pnueli and R. Rosner. On the synthesis of a reactive module. In *Proc. of POPL*, pages 179–190. ACM Press, 1989.

[36] M. O. Rabin. Probabilistic automata. *Information and Control*, 6:230–245, 1963.

[37] P. J. Ramadge and W. M. Wonham. Supervisory control of a class of discrete-event processes. *SIAM Journal of Control and Optimization*, 25(1):206–230, 1987.

[38] J. H. Reif. Universal games of incomplete information. In *Proc. of STOC*, pages 288–308. ACM, 1979.

[39] J. H. Reif. The complexity of two-player games of incomplete information. *JCSS*, 29:274–301, 1984.

[40] J. H. Reif and G. L. Peterson. A dynamic logic of multiprocessing with incomplete information. In *Proc. of POPL*, pages 193–202. ACM, 1980.

[41] D. Rosenberg, E. Solan, and N. Vieille. Stochastic games with imperfect monitoring (discussion paper). Technical Report 1376, Northwestern University, Center for Mathematical Studies in Economics and Management Science, July, 2003.

[42] L. S. Shapley. Stochastic games. *Proc. Nat. Acad. Sci. USA*, 39:1095–1100, 1953.

[43] S. Sorin. *A first course in zero-sum repeated games*. Springer, 2002.

[44] W. Thomas. Languages, automata, and logic. In *Handbook of Formal Languages*, volume 3, Beyond Words, chapter 7, pages 389–455. Springer, 1997.

[45] M. Tracol, C. Baier, and M. Größer. Recurrence and transience for probabilistic automata. In *Proc. of FSTTCS*, volume 4 of *LIPIcs*, pages 395–406, 2009.

[46] M. Y. Vardi. Automatic verification of probabilistic concurrent finite-state systems. In *Proc. of FOCS*, pages 327–338, 1985.