A

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

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

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 having 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 having 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 of the literature: we show a non-elementary memory lower bound for almost-sure winning whereas an exponential upper bound was previously claimed.

Categories and Subject Descriptors: F.4.1 [**MATHEMATICAL LOGIC AND FORMAL LANGUAGES**]; F.1.2 [**COMPUTATION BY ABSTRACT DEVICES**]: Modes of Computation

General Terms: Verification, Theory, Algorithms

Additional Key Words and Phrases: Partial-observation games, Stochastic games, Reachability and Büchi objectives, Positive and Almost-sure winning, Strategy complexity, Memory bounds.

ACM Reference Format:
Krishnendu Chatterjee and Laurent Doyen, 2013. Partial-Observation Stochastic Games: How to Win when Belief Fails. ACM Trans. Comput. Logic V, N, Article A (January YYYY), 45 pages.
DOI = 10.1145/0000000.0000000 http://doi.acm.org/10.1145/0000000.0000000

A preliminary version of this paper appeared in the Proceedings of the 27th Annual Symposium on Logic in Computer Science (LICS), IEEE Computer Society Press, 2012.
This work was partially supported by Austrian Science Fund (FWF) Grant No P 23499-N23, FWF NFN Grant No S11407-N23 (RiSE), ERC Start grant (279307: Graph Games), and Microsoft faculty fellows award.

DOI 10.1145/0000000.0000000 http://doi.acm.org/10.1145/0000000.0000000

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 [Pnueli and Rosner 1989; Ramadge and Wonham 1987], verification of open systems [Alur et al. 2002], realizability and compatibility checking [Abadi et al. 1989; Dill 1989; de Alfaro and Henzinger 2001], 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 [De Wulf et al. 2006; Henzinger and Kopke 1999]. 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 [Reif 1984; Alur et al. 2002]. Such problems are better modeled in the more general framework of *partial-observation* games [Reif 1979; 1984; Reif and Peterson 1980; Chatterjee et al. 2007; Bertrand et al. 2009] and have been studied in the context of verification and synthesis [Kupferman and Vardi 2000; Dimitrova and Finkbeiner 2008].

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 that 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 [Bertrand et al. 2009] and is a standard model in game theory [Rosenberg et al. 2003; Sorin 2002]. It subsumes other classical game models such as concurrent games [Shapley 1953; de Alfaro et al. 2007], probabilistic automata [Rabin 1963; Bukharaev 1980; Paz 1971], and partial-observation Markov decision processes (POMDPs) [Papadimitriou and Tsitsiklis 1987] (see also the recent decidability and complexity results for probabilistic automata [Baier et al. 2008; 2009; Baier and Größer 2005; Chadha et al. 2009a; 2009b; 2010; Gimbert and Oualhadj 2010; Chatterjee and Tracol 2012; Fijalkow et al. 2012], for POMDPs [Chatterjee et al. 2010b; Baier et al. 2008; Tracol et al. 2009; Chatterjee et al. 2013a], and stochastic games [Chatterjee et al. 2013b; Nain and Vardi 2013; Chatterjee et al. 2013c]).

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 [Shapley 1953; Condon 1992; de Alfaro et al. 2007].

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* objectives 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 PSPACE for perfect-observation stochastic games [Etessami and Yannakakis 2006], (in NP $\cap$ coNP for turn-based perfect-observation stochastic games [Condon 1992] where players make moves in turns), whereas the problem is undecidable even for partial-observation stochastic games with only one player [Paz 1971]. 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, and the almost-sure problem is polynomially equivalent to the almost-sure problem for reachability objective [Baier et al. 2008]. 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 [Chatterjee et al. 2007]. However, there exist two types of randomized strategies where either (i) actions are visible, the player can observe the action he played [Chatterjee et al. 2007; Bertrand et al. 2009], 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 [Gripon and Serre 2009]. 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 [Chatterjee et al. 2007], and 2EXPTIME-complete (resp., EXPTIME-complete) in the two-sided case [Bertrand et al. 2009]. 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) [Chatterjee et al. 2007; Bertrand et al. 2009].

It was remarked (without any proof) in [Chatterjee et al. 2007, 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 [Gripon and Serre 2009] (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. In graph games, the adversary is all powerful, and a randomized strategy against all-powerful strategies requires perfect randomization. Since the practical sources of randomization are pseudo-random, they are effectively pure strategies against the adversary in graph games. Hence the pure strategies question has been considered in many synthesis problems in applications related to programming languages as well as planning, such as synchronizers for lock placement in concurrent programs [Cerný et al. 2011], and controllers for robot planning [Kress-Gazit et al. 2009]. 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 [Chatterjee et al. 2007] and the main results of [Gripon and Serre 2009]), 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.

Relevance of two-sided games. While for controller synthesis problems the player 1 partial and player 2 perfect model is relevant, the two-sided model (partial information for both players) is also of considerable interest. In protocol synthesis with multiple processes, where each process has private variables invisible to other processes, the two-sided model is more appropriate. For example, synthesis of fair-exchange protocols (a security protocol for exchange of digital signature) under partial information requires the two-sided model [Jamroga et al. 2012]. Moreover, for synthesis of lock synchronization in concurrent programs the two-sided model is more appropriate; for example, in concurrent programs accessing a shared list, the lock synchronization and the scheduling policies do not depend on the variables in the list [Cerný et al. 2011], and therefore both players have partial observation. Finally, the two-sided partial observation stochastic games with action-invisible strategies is the most general form of two-player partial-observation stochastic games, and has been extensively studied in the stochastic game theory community [Aumann 1995; Rosenberg et al. 2003; Sorin 2002; Renault 2012]. We will show that our results for pure strategies imply memory bounds for two-sided games with action-invisible strategies.

Contributions. In summary, our contributions are: (1) For player 1 partial and player 2 perfect, we show that (in contrast to all previous results in the literature) belief-based pure strategies are not sufficient, and present optimal (exponential) memory and tight (EXPTIME) complexity bound for the qualitative problems. (2) For player 1 perfect and player 2 partial we show a counter-intuitive result that the optimal memory bound for pure strategies is non-elementary (in sharp contrast to the previous exponential bound). (3) For two-sided games we show finite memory is sufficient for pure strategies. (4) For memory bounds, we establish an unexpected equivalence of pure and randomized action-invisible strategies for almost-sure winning, that implies a non-elementary lower bound whereas previously an exponential upper bound was claimed. We now present the details.

(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. 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. Our construction also provides symbolic algorithms for solving the qualitative problems that avoids the explicit exponential construction. Thus for the important case of player 1 partial and player 2 perfect we establish optimal memory and complexity bounds, and present symbolic algorithms 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 in 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.

(3) *General (two-sided) case.* We show that in the general (two-sided) case finite memory strategies are sufficient both for positive and almost-sure winning.

(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 [Chatterjee et al. 2007] 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 [Gripon and Serre 2009] 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 intuition, there is a sharp contrast for randomized strategies with or without actions visible.

The memory requirements are summarized in Table I 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 [Chatterjee et al. 2007; Bertrand et al. 2009]. 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 I.

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

	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**

Related works. We have already mentioned the most relevant related works about the various (partial-observation) games considered in this paper. The seminal works of [Reif 1979; 1984] introduced partial-observation non-stochastic games in computer science, and used subset construction as the main technique to analyze such games for winning with certainty (which is same as winning with pure strategies in non-stochastic games). It was shown in [Chatterjee et al. 2007] that belief-based randomized strategies with actions visible are sufficient for almost-sure winning in partial-observation games with player 1 partial and player 2 perfect for reachability and Büchi objectives. The work of [Bertrand et al. 2009] extended the results of [Chatterjee et al. 2007] showing that even for two-sided games belief-based randomized strategies with actions visible are sufficient. The work of [Gripon and Serre 2009] considered two-sided games under randomized strategies with actions invisible. In the study of partial-observation games, the previous results led to the following intuitions for reachability and Büchi objectives: (1) belief-based strategies are sufficient; (2) the memory requirement for randomized strategies with actions visible and actions invisible is the same; and (3) the memory requirement for two-sided games is the same as for games with player 1 partial and player 2 perfect. Our results show that none of the previous intuitions is correct.

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 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. 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 $\mathcal{O}_i = \{\{q_0\},\{q_1\},\dots,\{q_n\}\}$ (given $Q = \{q_0, q_1, \dots, q_n\}$), 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 the states in s given actions 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 $\rho = q_0 a_0 b_0 q_1 a_1 b_1 q_2 \dots$ 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 $\rho = q_0 a_0 b_0 q_1 \dots q_k$ 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 (in)finite sequence $\mathsf{obs}_i(\rho) = \mathsf{obs}_i(q_0)\mathsf{obs}_i(q_1)\dots$, and given a play $\rho' = q'_0 a'_0 b'_0 q'_1 a'_1 b'_1 q'_2 \dots$ we write $\rho' \approx_i \rho$ if $q_j \approx_i q'_j$, $a_j \approx_i a'_j$, and $b_j \approx_i b'_j$ for all $j \geq 0$.

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 $a \approx_1 b$ iff $a \approx_2 b$ iff $a = b$ (all actions are distinguishable by both players), 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, the set of all observation-based player-2 strategies, and the set of all pure player-2 strategies in G, respectively.

Beliefs. The belief of a player is the set of states in which the game can possibly be according to all the informations available to the player, namely their own strategy and the observation of the play prefix. Formally, the belief of player 1 after the play prefix ρ when strategy σ is fixed is the set $B(\rho, \sigma) = \{\mathsf{Last}(\rho') \mid \rho' \approx_1 \rho \text{ and } \rho' \text{ is compatible with } \sigma\}$.

Given a randomized strategy σ fixed by player 1, the belief is computed inductively along a play both in the case of actions visible and actions invisible, however, the computation is different in the two cases. Initially, the belief of player 1 is $B_0 = \{q_0\}$, and it is updated as follows:

— (*Actions visible*). Given a play prefix ρ, if the belief of player 1 in ρ is B_i, then define the belief of player 1 in the play prefix $\rho\, a_i b_i\, q_{i+1}$ by

$$B_{i+1} = \mathsf{Post}_{a_i, b_i}(B_i) \cap \mathsf{obs}_1(q_{i+1}). \tag{1}$$

— (*Actions invisible*). Given a play prefix ρ, let $A = \mathsf{Supp}(\sigma(\rho))$ be the support of the distribution on actions played by player 1 in ρ. If the belief of player 1 in ρ is B_i, then define the belief of player 1 in the play prefix $\rho a_i b_i q_{i+1}$ (where $a_i \in A$) by

$$B_{i+1} = \left(\bigcup_{a \in A} \mathsf{Post}_{a,-}(B_i) \right) \cap \mathsf{obs}_1(q_{i+1}). \tag{2}$$

It is easy to show by induction on the length of play prefixes that $B(\rho, \sigma) = B_{|\rho|}$ (for example, see [Chatterjee et al. 2007, Lemma 2.1] for actions visible). Note that in the case of actions invisible, the update of the belief depends on the support of the distribution on actions played by the strategy but not on the precise action chosen since the strategy does not see the actions, whereas in the case of actions visible, the belief update depends on the precise actions chosen.

Remark 2.1. 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 [Gripon and Serre 2009]. 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).

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 $\sigma(\rho \cdot q) = \alpha_n(\widehat{\alpha}_u(m_0, \mathsf{obs}_1(\rho)), \mathsf{obs}_1(q))$ for all $\rho \in Q^*$ and $q \in Q$, where $\widehat{\alpha}_u$ 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 $|\mathsf{Mem}| = 1$. A *belief-based* strategy for player 1 is a strategy that uses the set of beliefs as memory (say $\mathsf{Mem} = 2^Q$), and updates its memory according to Eq. (2) where the set A is given by the next-move function.

Objectives. 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 [Kechris 1995]. Given strategies σ and π for the two players, the probabilities of a measurable objective φ is uniquely defined [Vardi 1985]. We denote by $\mathrm{Pr}_{q_0}^{\sigma,\pi}(\varphi)$ 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 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 objective* that requires 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 ☺.

Winning modes: positive, almost-sure, and sure winning problems. 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 $\mathrm{Pr}_q^{\sigma,\pi}(\varphi) = 1$ (resp., $\mathrm{Pr}_q^{\sigma,\pi}(\varphi) > 0$). 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. 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.

Remark 2.2. The important special case of partial-observation Markov decision processes (POMDPs) corresponds to the case where either all states in the game are player-1 states (player-1 POMDPs) or all states are player-2 states (player-2 POMDPs). 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. In all our analysis, the counter strategies of player 2 can be restricted to pure strategies, because once a strategy for player 1 is

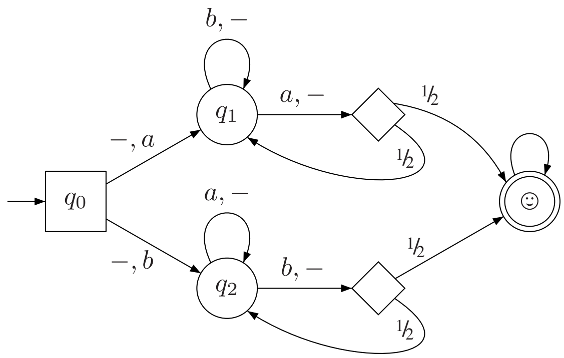

Fig. 1. **Belief-only is not enough for positive (as well as almost-sure) reachability.** A one-sided reachability game with reachability objective in which player 1 is blind and player 2 has perfect observation. If we consider pure strategies, then player 1 has a positive (as well as almost-sure) winning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

fixed, then we obtain a POMDP for player 2 in which pure strategies are as powerful as randomized strategies [Chatterjee et al. 2010a].

Remark 2.3. (Almost-sure Büchi to almost-sure reachability [Baier et al. 2008]). A reduction of almost-sure Büchi to almost-sure reachability was presented in [Baier et al. 2008] for POMDPs. The idea of the reduction also works for two-sided games and 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 [Baier et al. 2008, 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.

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 [Reif 1979; Chatterjee et al. 2007; Chatterjee and Doyen 2010]. The almost-sure winning problem for Büchi objectives reduces to the almost-sure winning problem for reachability objectives (Remark 2.3). The positive winning problem for Büchi objectives is undecidable even for POMDPs [Baier et al. 2008]. Hence in this paper we only focus on reachability objectives.

3. ONE-SIDED GAMES: PLAYER 1 PARTIAL AND PLAYER 2 PERFECT

In this section, we consider one-sided partial-observation games with player 1 partial and player 2 perfect (player 2 sees the states visited in the game, and thus the actions played as well). 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.

In the following example we show that *belief-based* strategies are not sufficient for positive as well as almost-sure winning for pure strategies. It is known that for reachability objectives, belief-based strategies are sufficient in both one-sided and two-sided games, for sure-winning with pure and randomized strategies [Reif 1979; Chatterjee et al. 2007], for almost-sure winning with action-visible randomized strategies [Chatterjee et al. 2007; Bertrand et al. 2009], and randomized memoryless strategies are sufficient for positive winning [Bertrand et al. 2009]. A simple self-contained proof of the almost-sure winning case with action-visible randomized strategies is presented in the appendix. We now consider the case of pure strategies for almost-sure and positive winning.

Example 3.1. **Belief-only is not enough 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 $\{q_1, q_2\}$.We claim that no belief-based pure strategy is positive (nor almost-sure) winning for player 1. Consider the belief-based pure strategy that plays action a when the belief is $\{q_1, q_2\}$. Then a counter strategy for player 2 is to choose q_2 initially. This ensures that the game remains in q_2, while the belief of player 1 remains $\{q_1, q_2\}$ and thus player 1 keeps playing a. A symmetric argument shows that playing b when the belief is $\{q_1, q_2\}$ is also not even positive winning. However, there exists an almost-sure winning strategy for player 1 which is to alternate between playing a and b when the belief is $\{q_1, q_2\}$, showing that belief-based strategies are 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 3.2. *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-observation 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.*

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 a one-sided game $G = \langle Q, q_0, \delta \rangle$ 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 (1) 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 (2) 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 (1), 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 (2), 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 $\widehat{\rho} = q_0 q_1 \dots q_k$ 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 $\widehat{\rho}$ is reachable in G_σ, there exist actions $b_0, \ldots, b_{k-1}$ such that $q_{i+1} \in \mathsf{Supp}(\delta_\sigma(q_0 \ldots q_i, b_i))$ for all $0 \leq i < k$. Then, define $\pi(q_0 \ldots q_i) = b_i$. This ensures that $\mathsf{Last}(\widehat{\rho})$ is reached with positive probability in G under strategies σ and π. From $\widehat{\rho}$, the strategy π continues playing as follows. If the current state ρ is not ranked (which is the case of $\widehat{\rho}$), 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 $\mathsf{Last}(\widehat{\rho})$ 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 $\rho_H = (s_0, o_0)(s_1, o_1) \ldots (s_k, o_k)$ 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 $u = \{q \in \mathsf{Supp}(\delta(\mathsf{Last}(\rho_G), a, b)) \mid b \in A_2, \rho_G$ is matching ρ_H with $\mathsf{Last}(\rho_G) \in o_k$ 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 $\rho_H = (s_0, o_0)(s_1, o_1) \ldots$ 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 $\rho_G = q_0 q_1 \ldots q_k$ be a play prefix in G of length k. Let $\rho_H = \theta(q_0 q_1 \ldots q_{k-1})$ 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 [Emerson and Jutla 1991; Thomas 1997]), 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 a $\mathcal{T}$-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 [Emerson and Jutla 1991; Thomas 1997]. 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 [Emerson and Jutla 1991; Thomas 1997]. 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. $\square$

It follows from the construction in the proof of Lemma 3.2 that pure strategies with exponential memory are sufficient for positive (as well as almost-sure) winning, and the exponential lower bound follows from the problem of shortest witness for non-universality of non-deterministic finite automata. Lemma 3.2 also gives EXPTIME upper bound for the problem since perfect-observation Büchi games can be solved in polynomial time [Thomas 1997]. The EXPTIME-hardness follows from the sure winning problem for non-stochastic games [Reif 1984], where pure almost-sure (positive) winning strategies coincide with sure winning strategies. We have the following theorem summarizing the results.

THEOREM 3.3. *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 3.3 and Remark 2.3 we obtain the following corollary.

COROLLARY 3.4. *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 $\sum_{\gamma \in \mathcal{O}_1} 2^{|\gamma|} \leq \prod_{\gamma \in \mathcal{O}_1} 2^{|\gamma|} = 2^{|Q|}$ and $\sum_{\gamma \in \mathcal{O}_1} 3^{|\gamma|} \leq \prod_{\gamma \in \mathcal{O}_1} 3^{|\gamma|} = 3^{|Q|}$, where $|Q|$ is the number of states in the one-sided game.

Symbolic algorithms. The exponential Büchi (or reachability) game constructed in the proof of Theorem 3.3 can be solved by computing classical fixpoint formulas [Emerson and Jutla 1991]. 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 [Chatterjee et al. 2007] for details related to antichain algorithms and [Doyen and Raskin 2010] for a survey of the antichain approach). This technique provides a symbolic algorithm without explicitly constructing the exponential game.

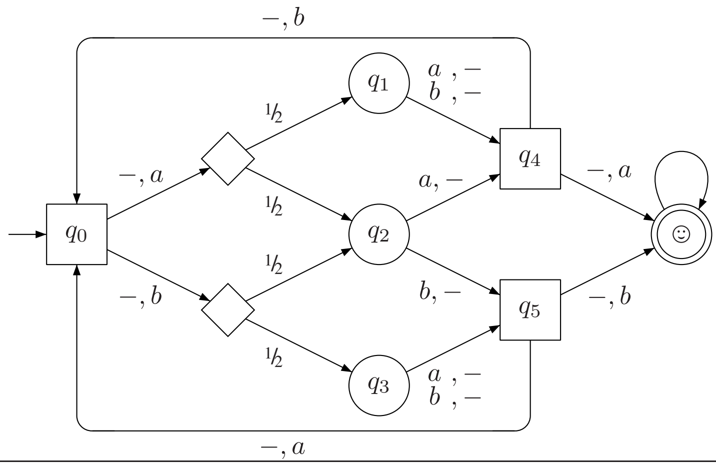

Fig. 2. **Remembering the belief of player 2 is necessary.** A one-sided reachability game where player 1 (round states) has perfect observation, player 2 (square states) is blind. Player 1 has a pure almost-sure winning strategy that depends on the belief of player 2 (in q_2), but no pure memoryless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

4. ONE-SIDED GAMES: PLAYER 1 PERFECT AND PLAYER 2 PARTIAL

In this section, we consider one-sided partial-observation games with 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 chosen after player 1 fixed a strategy.

The results of this section show that the optimal memory bound for pure strategies is non-elementary, in sharp and unexpected contrast to the exponential bound of Section 3, and thus the intuition that the memory requirement for two-sided games is the same as for games with player 1 partial and player 2 perfect is incorrect.

4.1. Lower bound on memory

We start with an example to illustrate that player 1 needs to remember the belief of player 2 both for positive and almost-sure winning.

Example 4.1. **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 Fig. 2. The target is $\mathcal{T} = \{☺\}$. 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 become $\{q_4\}$ and player 2 could surely avoid ☺ 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 positively winning *memoryless* strategy for player 1. However, we show that there exists an almost-sure winning strategy for player 1 that uses the belief of player 2 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

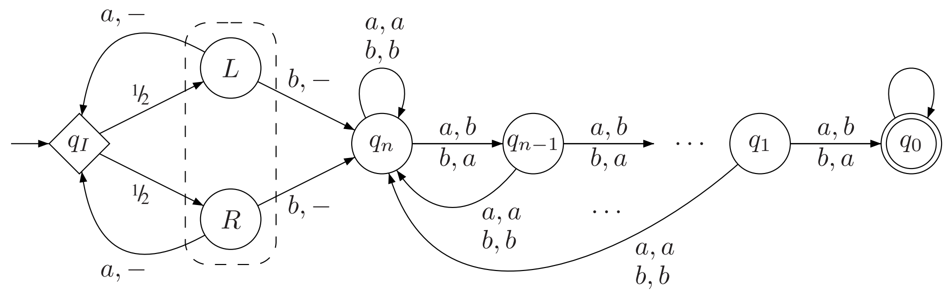

Fig. 3. A one-sided reachability game L_n with reachability objective in which player 1 is has perfect observation and player 2 is blind. Player 1 needs exponential memory to win positive reachability.

can observe the actions of player 2 (and thus compute the current belief of player 2). This ensures the next belief of player 2 to be $\{q_4, q_5\}$ and therefore no matter the next action of player 2, the state ☺ is reached with probability $\frac{1}{2}$. Repeating this strategy ensures to reach ☺ with probability 1. ■

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 will establish the surprising non-elementary lower bound.

Example 4.2. **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 system (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 Fig. 3, 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. Since by Remark 2.2, pure strategies are sufficient for player 2, consider an arbitrary pure strategy for player 2. It follows that the same action is played by player 2 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.

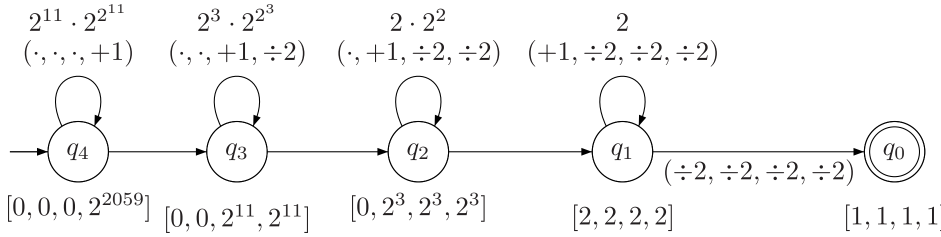

Fig. 4. A family $(C_n)_{n \in \mathbb{N}}$ of counter systems with n counters and $n+1$ states where the shortest execution to reach $(q_0, k_1, \dots, k_n)$ with positive counters (i.e., $k_i > 0$ for all $1 \leq i \leq n$) from $(q_n, 0, \dots, 0)$ is of non-elementary length. The numbers above the self-loops show the number of times each self-loop is taken along the shortest execution.

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 $\rho = q_I L q_I R q_I L q_n$ is mapped to $w_\rho = aba$). 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 $0 \leq i \leq n$, there are 2^i play prefixes that 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 $0 \leq i \leq n$.

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 Fig. 4. The operations on counters can be *increment* $(+1)$, *division by* 2 $(\div 2)$, and *idle* $(\cdot)$. In general, C_n has n counters $c_1, \dots, c_n$ and $n+1$ states $q_0, \dots, q_n$. 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 Fig. 4, 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 $c_{i+1}, \dots, c_n$ need to have value $k \cdot 2^k$ when entering q_i (in order to guarantee a value at least k of these counters after k successive division by 2). In q_n, the last counter c_n needs to have value $f^n(1)$ where f^n is the n-th iterate of the function $f : \mathbb{N} \to \mathbb{N}$ such that $f(x) = x \cdot 2^x$. This value is greater than a tower of exponentials of height n.

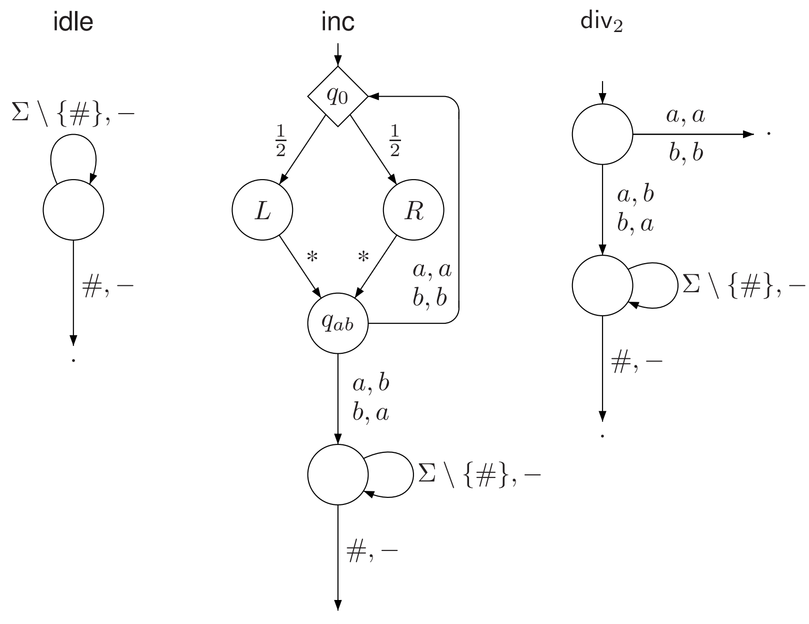

Fig. 5. Gadgets to simulate idle, increment, and division by 2.

Gadgets for increment and division. In Fig. 5, 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 Fig. 4. 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 Fig. 7). 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 the exit state. In the game of Fig. 7, the entry and exit state of some division gadgets are merged. The argument still holds.

Consider the increment gadget inc on Fig. 5. We use this gadget with the assumption that the entry state is not reached by more than one indistinguishable path. This will

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

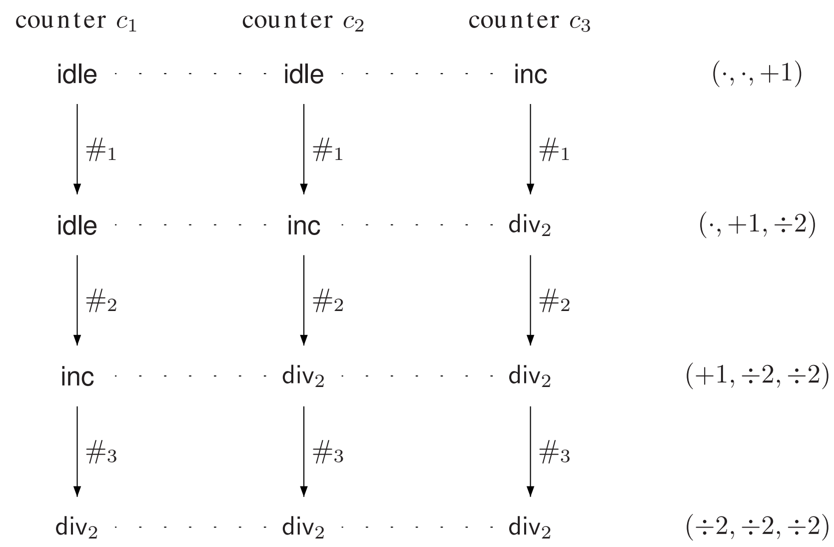

Fig. 6. Abstract view of the game in Fig. 7 as a 3-counter system.

be the case in the game of Fig. 7. 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 L, and play action b if the last visited state is 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. Note that the self-loop in the exit state is used to prolong the paths in the exit state, and make the k paths of the same length (hence indistinguishable by player 2).

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 Fig. 7 for $n = 3$. Its abstract structure is shown in Fig. 6, corresponding to the structure of the counter system in Fig. 4. The alphabet of player 1 is $\{a, b, \#\}$. For the sake of clarity, some transitions are not depicted in Fig. 7. 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 ☺, all transitions go to the initial state.

Fig. 7 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 by indistinguishable paths, so as to ensure that no matter the action (a, b, or c) chosen by player 2, the state ☺ 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 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 Fig. 7, the dotted lines and the subscripts on # emphasize the layered structure of the game, corresponding to the structure of Fig. 6.

To ensure that there are indistinguishable paths ending up in each of the states q_1, r_1, s_1, player 1 needs to generate paths in each "line", and this can be done only using the increment gadget in each line. Concretely, for the first counter c_1 the value must be at least $x_1 = 2$ when leaving the increment gadget (before playing $\#_3$ and entering q_2) in order to survive the transition leading to q_1. Since in the second line, the division gadget is executed in r_3 at the same time (thus after $\#_2$, and before $\#_3$ is played), the value of the second counter must be at least $x_2 = x_1 \cdot 2^{x_1} = 2 \cdot 2^2$ when leaving the increment gadget, corresponding to the value shown in the second state of Fig. 4. In the third line, the value of the counter must be at least $x_3 = x_2 \cdot 2^{x_2}$ when leaving the increment gadget, and so on (in general with more counters). We now present the three steps of the proof.

— *Positive winning strategy exists.* An inductive argument shows that if player 1 simulates the counter system of Fig. 4, then a positive counter value is reached simultaneously in the last states of each line. This ensures that no matter the choice of player 2 in the last step of each line, the target state is reached at least from one of the last states of each line. This ensure positive winning for player 1.
— *Lower bound*. The analysis of the family C_n of counter systems (illustrated in Fig. 4) showed that the number of increments needed in the n-th line (i.e., the n-th counter) is greater than a tower of exponentials of height n (by inductive argument on n). It follows that player 1 needs memory of size non-elementary in order to ensure indistinguishable paths ending up in each of the states $q_1, r_1, \ldots$, and win with positive probability. Note that if indistinguishable paths do not end up in each of the last step of each line (say for example, that player 2 knows that q_1 is not reached in Fig. 7), then player 2 can play an action to prevent that the target is reached (action a in Fig. 7).
— *Bound for almost-sure winning.* In the item above we already argued that in the example even for positive winning non-elementary memory is required. To claim the lower bound for almost-sure winning we need to establish the existence of a pure almost-sure winning strategy. Since all paths that do not reach the target state go back to the initial state, the pure positive winning strategy can be repeated over and over again to achieve almost-sure reachability as well.

The desired result follows. ∎

THEOREM 4.3. *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 algorithmic solution of one-sided games with a conjunction of positive reachability and almost-sure safety objectives, in which player 1 has perfect observa-

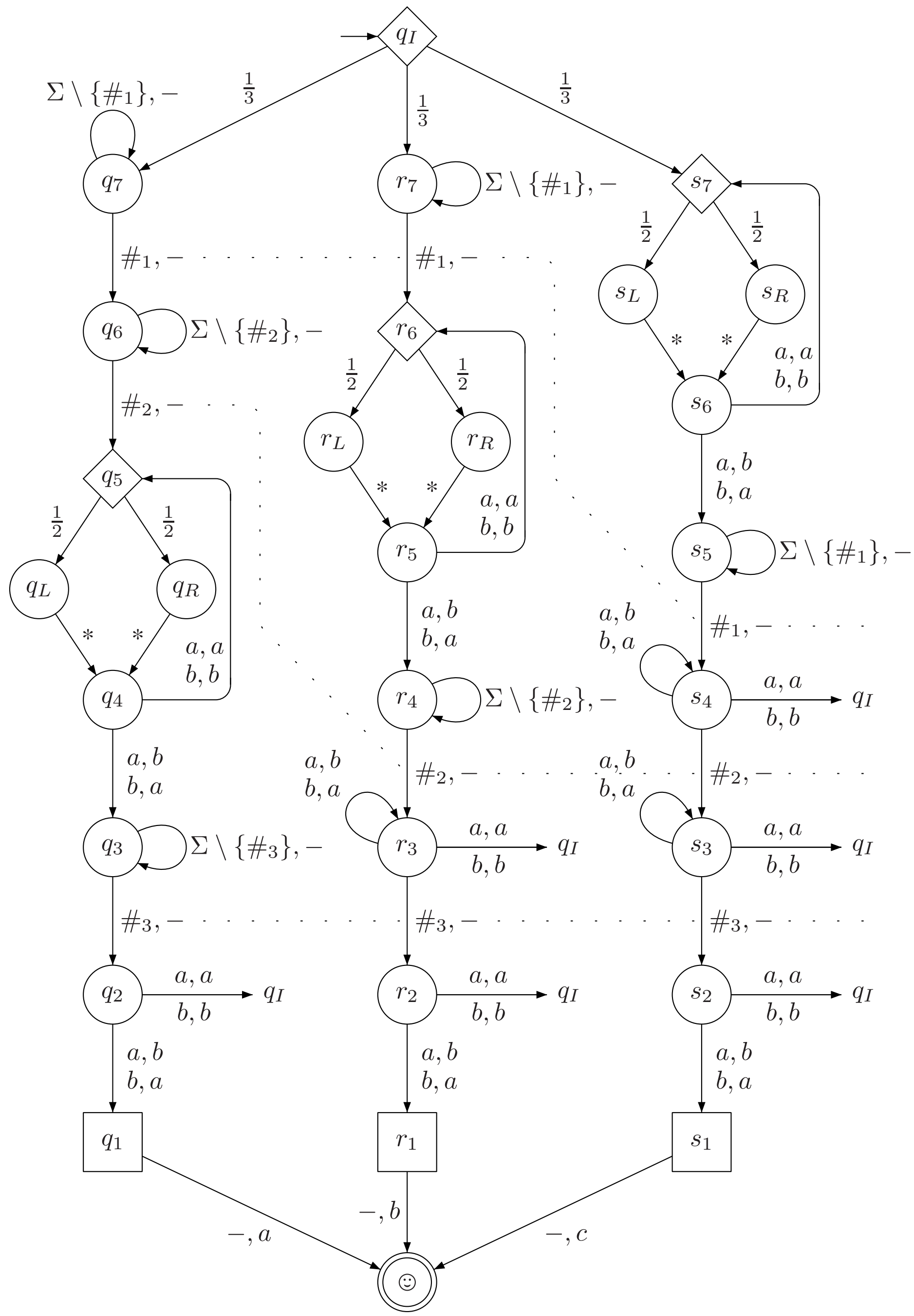

Fig. 7. **Memory of non-elementary size may be necessary for positive and almost-sure reachability.** A family of one-sided reachability games in which player 1 is has perfect observation. Player 1 needs memory of non-elementary size to win positive reachability (as well as almost-sure reachability).

tion 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 algorithmic 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.

We first describe the key steps of the proof. 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. Thus we first consider the case of randomized strategies. After that, 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 $\mathsf{Safe}(Q_G)$. Playing all safe actions uniformly at random ensures that the set $Q \backslash Q_G$ 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 with a randomized strategy 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 [Chatterjee et al. 2010b, 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 (because after fixing the positive winning strategy for player 1, there must be an acyclic path from all states in Z to the target in the subset-construction-based POMDP for player 2), 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 Fig. 5, 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 $\frac{1}{2}$.

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

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 with positive probability, an upper bound on the number of play prefixes that are needed to simulate a randomized strategy using a pure strategy is $\mathsf{K}_1 = |A_1|^{2^{|Q|}}$. More precisely, if the belief of player 2 is $B \subseteq Z$ and in each state $q \in B$ there are at least K_1 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 $2^{|Q|}$ 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 [Condon 1992].

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 $q \in B$, the number of indistinguishable play prefixes that end up in q. For $k \in \mathbb{N}$, we denote by $[k]$ the set $\{0, 1, \dots, k\}$. A state of H is a *counting function* $f : Q \to [\mathsf{K}_*] \cup \{\omega\}$ where $\mathsf{K}_* \in \mathbb{N}$, to be chosen later, is of order $|A_1|^{|A_1|^{\cdot^{\cdot^{|A_1|^{2^{O(|Q|)}}}}}}$ where the number of nested exponentials is of order $O(|Q|)$. We often view a counting function as a set of $|Q|$ *counters*.

As we have seen in the example of Fig. 7, 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 $\widehat{a} : Q \times [\mathsf{K}_*] \to A_1$ 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 (according to an arbitrary ordering on the play prefixes). We denote by $\mathsf{Supp}(\widehat{a}(q, \cdot)) = \{\widehat{a}(q, i) \mid i \in [\mathsf{K}_*]\}$ the set of actions played by $\widehat{a}$ 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 $\widehat{a}$ as above, given an action $b \in A_2$ and an observation $\gamma \in \mathcal{O}_2$, let $f' = \mathsf{Succ}(f, \widehat{a}, b, \gamma)$ be the function such that $f'(q') = 0$ for all $q' \notin \gamma$, and such that for all $q' \in \gamma$:

$$f'(q') = \begin{cases} \omega & \text{if } \exists a \in \mathsf{Supp}(\widehat{a}(q, \cdot)) \cdot \exists q \in Q : \\ & f(q) = \omega \wedge q' \in \mathsf{Post}_G(q, a, b); \\ \displaystyle\sum_{q \in \mathsf{Supp}(f)} \sum_{i=0}^{f(q)-1} \mathbf{1}(q' \in \mathsf{Post}_G(q, \widehat{a}(q, i), b)) & \text{otherwise.} \end{cases}$$

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

Now $f' = \mathsf{Succ}(f, \widehat{a}, b, \gamma)$ 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 a counters $f(q)$ and we can replace such large values by ω. Intuitively, the ω value can be interpreted as "large enough to be positive for sufficiently long". This abstraction needs to be done carefully in order to obtain the desired upper bound (namely, K_*). As explained before, the counter values represent the number of indistinguishable paths. A player-1 strategy by playing different actions in different indistinguishable paths can simulate a randomized strategy. Intuitively, a counter value is large enough if it allows to play all actions in the support of a randomized strategy. Thus a counter value is large enough for a single step if the counter has value at least $|A_1|$, as this allows player 1 to play every action at least once among the indistinguishable paths. A counter value is large enough for K steps if the value is at least $|A_1|^{\mathsf{K}}$ since this allows to simulate a randomized strategy for K steps. We now present an inductive argument to formalize the notions of "large enough" and "sufficiently long".

— We first consider the base case when all counters can be set to ω. We know that if all counters have value greater than $\mathsf{K}_1 = |A_1|^{2^{|Q|}}$, then player 1 wins by simulating a randomized strategy for sufficiently long which is $2^{|Q|}$ steps. 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. In other words, when all counters are set to ω, then sufficiently long is $2^{|Q|}$ steps.

— Now we consider the inductive case. First we consider when all counters but one are set to ω. By the base case when the value of the last counter exceeds K_1 then it is set to ω. Since reaching the value upto K_1 can take at most K_1 steps, the other counters with value ω need to have value at least $\mathsf{K}_2 = \mathsf{K}_1 \cdot |A_1|^{\mathsf{K}_1}$ to remain large enough (i.e., at least K_1) for sufficiently long (i.e., at least K_1 steps). 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 $\mathsf{K}_{k+1} = \mathsf{K}_k \cdot |A_1|^{(\mathsf{K}_k)^k}$. In other words, when all counters but k have value ω, then sufficiently long is K_k steps.

Taking $\mathsf{K}_* = \mathsf{K}_{|Q|}$, it follows by induction that K_* is of order $|A_1|^{|A_1|^{\cdot^{\cdot^{|A_1|^{2^{O(|Q|)}}}}}}$ where the tower of exponential is of height $|Q|$, 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 the $|Q|$-th iterate of the following procedure:

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

Given f, $\widehat{a}$, and b, let $\delta_H(f, \widehat{a}, b)$ be the uniform distribution over the set of counting functions f' such that there exists an observation $\gamma \in \mathcal{O}_2$ such that $f' = \mathsf{Abs}(\mathsf{Succ}(f, \widehat{a}, b, \gamma))$ 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 $\mathsf{Succ}(f, \widehat{a}, b, \gamma) \leq \mathsf{Succ}(f', \widehat{a}, b, \gamma)$ for all $\widehat{a}, b, \gamma$ (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] $\mathcal{T}_H = \{f \mid \exists q \in Z : f(q) = \omega\} \cup \{f \mid \mathsf{Supp}(f) \cap \mathcal{T} \neq \emptyset\}$ 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 can keep track along the play of the exact number of indistinguishable play prefixes ending up in each state. This allows to define the action $\widehat{a}$ to play in H by collecting the actions played by σ in all the play prefixes that are indistinguishable with the current play prefix. Note that by monotonicity, the counting abstractions along the corresponding play prefix of H are at least as big (assuming $\omega > k$ for all $k \in \mathbb{N}$), and thus the action $\widehat{a}$ is well-defined. Since σ is winning in G, regardless of the actions played by player 2, the set $\mathcal{T}$ is reached with positive probability, and the set $Q \backslash Q_G$ is never hit. It follows that in H, the strategy σ^H against the same sequence of actions of player 2 produces sequences of counting abstractions in Good_H (thus with support in Q_G), and that visit with positive probability a counting function $f \in \mathcal{T}_H$ (such that $\mathsf{Supp}(f) \cap \mathcal{T} \neq \emptyset$), thus satisfying both sure safety for $\mathsf{Safe}(\mathsf{Good}_H)$, and positive reachability for $\mathsf{Reach}(\mathcal{T}_H)$.

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. Since H is a game of perfect information, 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 in H 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 ω-abstraction ensure 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' $\widehat{a}$ played by σ^H). More precisely, the bounds $\mathsf{K}_1, \mathsf{K}_2, \dots$ 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 some $q \in Z$, or a counting function f with $\mathsf{Supp}(f) \cap \mathcal{T} \neq \emptyset$. In the first case, since $q \in Z$ is a winning state for player 1 with randomized strategies in G, we can simulate σ^H in G until q is reached (which happens with positive probability), and then from q win positively by simulating a randomized strategy for at most $2^{|Q|}$ steps; 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 4.4. *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 positive probability reachability along with almost-sure safety for player 1.*

From Theorem 4.3 and Theorem 4.4 we have the following result.

COROLLARY 4.5. *In one-sided partial-observation stochastic games with player 1 perfect and player 2 partial, for pure positive winning strategies for reachability objectives for player 1 non-elementary memory size is sufficient, and necessary in general.*

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

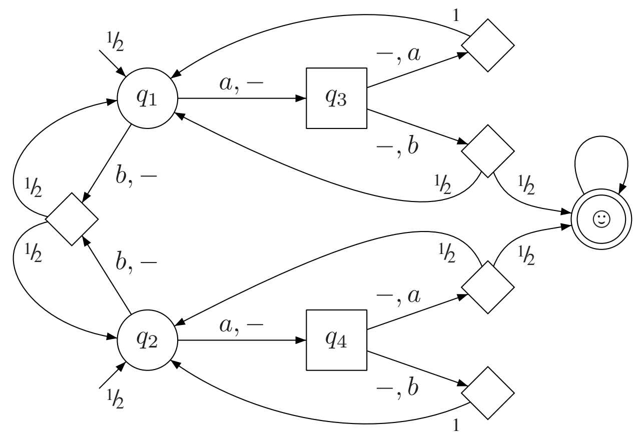

Fig. 8. **Almost-sure winning strategy may require more memory than positive winning strategies.** A one-sided reachability game where player 1 (round states) has perfect observation, player 2 (square states) is blind. Player 1 has a pure almost-sure winning strategy, but no pure belief-based memoryless strategy is almost-sure winning. However, player 1 has a pure belief-based memoryless strategy that is positive winning.

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 4.6. **Almost-sure winning strategy may require more memory than positive winning strategies.** The example of Fig. 8 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 ☺. 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 ☺ is reached with positive probability from either q_3 or q_4.

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 ☺ 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 ☺ is reached with positive probability by making the belief to be $\{q_1, q_2\}$. ■

We now 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. We first present an overview of the argument. 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 will show that repeatedly playing a strategy for positive reachability along with almost-sure safety is an almost-sure winning strategy for the reachability objective.

Notation. We 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 current belief of player 2, given a prefix and the strategies, represents the set of states in which the game can be currently with positive probability given the prefix and the strategies. Given strategies σ and π for player 1 and player 2, respectively, a state q and a set $K \subseteq Q$ we denote by $\Pr_{q,K}^{\sigma,\pi}(\cdot)$ 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_G^O) as the game is clear from the context.

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

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

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 observation-based counter 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

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

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

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

Let

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

We now show that $Q_G^p = Q_G$. The inclusion $Q_G \subseteq Q_G^p$ is trivial, and we now show the other inclusion $Q_G^p \subseteq Q_G$. Observe that in Q_G^p 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 $q \in Q_G^p$ 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 $\eta = \min_{q \in Q_G^p} \eta_q$. 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_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 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 $O(2^{|Q|} \times \text{PosReachSureSafe})$, 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 verifies that for all $q \in Q'$ 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 4.7. *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.*

From Theorem 4.3 and Theorem 4.7 we have the following results.

COROLLARY 4.8. *In one-sided partial-observation stochastic games with player 1 perfect and player 2 partial, for reachability objectives non-elementary memory size is sufficient, and necessary in general for pure almost-sure winning strategies for player 1.*

COROLLARY 4.9. *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 2.3 we obtain the following corollary.

COROLLARY 4.10. *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.*

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. Finite-memory for almost-sure winning is proved by constructing a winning strategy that ensures (i) that the almost-sure winning set is never left (this is a safety condition), and (ii) within a fixed number N

of steps, the target is reached with at least a constant probability η. We show that these two conditions can be ensured by a finite-memory strategy and the key arguments are as follows: we use the fact that belief-based strategies are sufficient for the safety condition, and then by a simple extension of König's Lemma [König 1936] establish that positive reachability implies that the target is reached after a finite number of steps. The result for almost-sure winning follows by combining these strategies in a recharging scheme analogous to the approach in Section 4.3.

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.

LEMMA 5.1. *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*

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

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

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

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

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

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

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

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

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

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 [Chatterjee and Doyen 2010] for details). □

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

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

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 $\mathrm{Pr}_q^{\sigma,\pi_n}(\mathsf{Reach}^{\leq n}(\mathcal{T})) = 0$ or $\mathrm{Pr}_q^{\sigma,\pi_n}(\mathsf{Safe}(Q_G)) < 1$.

If for some $n \geq 0$ we have $\mathrm{Pr}_q^{\sigma,\pi_n}(\mathsf{Safe}(Q_G)) < 1$, then we get a contradiction with the fact that $\mathrm{Pr}_q^{\sigma,\pi}(\mathsf{Safe}(Q_G)) = 1$ for all $\pi \in \Pi^O$. Hence $\mathrm{Pr}_q^{\sigma,\pi_n}(\mathsf{Safe}(Q_G)) = 1$ for all $n \in \mathbb{N}$, and therefore $\mathrm{Pr}_q^{\sigma,\pi_n}(\mathsf{Reach}^{\leq n}(\mathcal{T})) = 0$ 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 $\mathrm{Pr}_q^{\sigma,\pi_n}(\mathsf{Safe}^{\leq n}(Q \setminus \mathcal{T})) = 1$ 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 [Chatterjee et al. 2010a] (in [Chatterjee et al. 2010a] 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 [König 1936], we construct a strategy $\pi' \in \Pi^O$ such that $\mathrm{Pr}_q^{\sigma,\pi'}(\mathsf{Safe}(Q \setminus \mathcal{T})) = 1$. 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 $\mathrm{Pr}_q^{\sigma,\pi'}(\mathsf{Safe}(Q \setminus \mathcal{T})) = 1$, in contradiction with the fact that $\mathrm{Pr}_q^{\sigma,\pi}(\mathsf{Reach}(\mathcal{T})) > 0$ for all $\pi \in \Pi^O$. Hence, the desired result follows. $\square$

The above two claims establish Lemma 5.1 and gives the following result.

THEOREM 5.2. *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. The only differences are that since player 1 has partial-observation we need to consider beliefs for player 1, and for finite-memory strategies for positive reachability with almost-sure safety we use the result of the previous subsection.

Notation. We 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 $\mathrm{Pr}_{q,K}^{\sigma,\pi}(\cdot)$ 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

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

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

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

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

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

Let

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

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. A non-elementary lower bound on memory follows from Corollary 4.8.

THEOREM 5.3. *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).

This equivalence of randomized action-invisible with pure strategies is counter-intuitive, and the intuition that randomized strategies with actions visible and actions invisible are equivalent fails, since it follows from the reduction of pure to randomized action-invisible strategies that the non-elementary lower bounds obtained in Section 4 for pure strategies transfer to randomized strategies with actions invisible, whereas belief-based strategies are sufficient with actions visible.

Remark 6.1. We highlight two important aspects of the equivalence: (1) the equivalence holds for action-invisible strategies, but not in the special case of action-visible strategies; (2) the equivalence holds for almost-sure winning, but not positive winning. One reason of the first aspect fact is that when actions are visible, the belief update (according to Eq. (1), p. 8)) is based on the actual action played, even though the player was randomizing over a set of actions. The reason for the second aspect is for positive winning memoryless strategies that plays all actions uniformly at random are sufficient both for action-visible as well as action-invisible strategies.

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 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 6.2 the correctness of the reduction for finite-memory randomized action-invisible strategies, and then show in Lemma 6.3 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 6.2. *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 [Chatterjee et al. 2010b] it follows that in a 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 6.3. *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 6.4. *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 [Bertrand et al. 2009]. The above theorem along with Theorem 3.3 gives us the following corollary for almost-sure winning for randomized action-invisible strategies.

COROLLARY 6.5. *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 6.6. *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 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 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. The reduction would not work with actions visible because player 1 could easily repeat every action (since the strategy can observe the action) without using a pure strategy (thus use the power of randomization).

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 6.7. *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.
(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 6.8. *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 [Chatterjee et al. 2007, p.4] and the result and construction of [Gripon and Serre 2009, Theorem 1] are wrong.

Example 6.9. We illustrate the reduction of on the example of Fig. 1. The result of the reduction is given in Fig. 9. Remember that Example 3.1 showed that belief-based

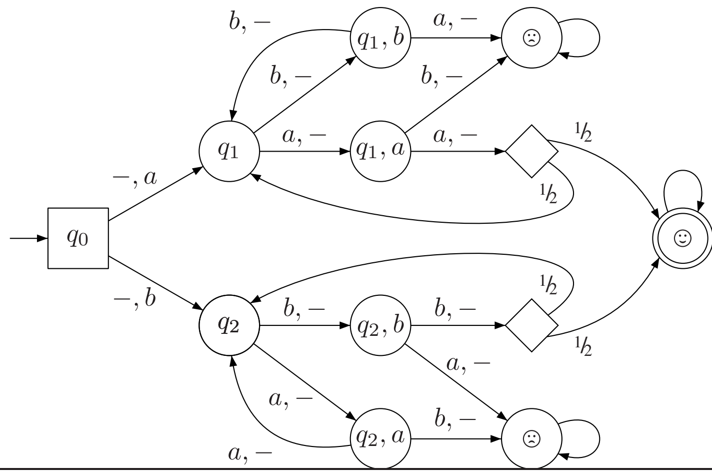

Fig. 9. **Belief-based strategies are not sufficient.** The game graph obtained by the reduction of pure to randomized strategies on the game of Fig. 1 (for almost-sure reachability objective). Player 1 is blind and player 2 has perfect observation. There exists an almost-sure winning randomized strategy (with invisible actions), but there is no *belief-based memoryless* almost-sure winning randomized strategy.

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. 9. 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 3.1: 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 4.3 for pure strategies also transfers to randomized action-invisible strategies by the same reduction.

COROLLARY 6.10. *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.*

7. DISCUSSION ON THE NON-ELEMENTARY MEMORY BOUND

In this section we discuss our results on the non-elementary memory bound. We start with one-sided games.

Discussion about the surprising non-elementary memory bound. We first 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. 10).

(1) *Question 1.* If we consider player-1 perfect player-2 partial deterministic games with reachability objective, then for positive winning pure memoryless strategies

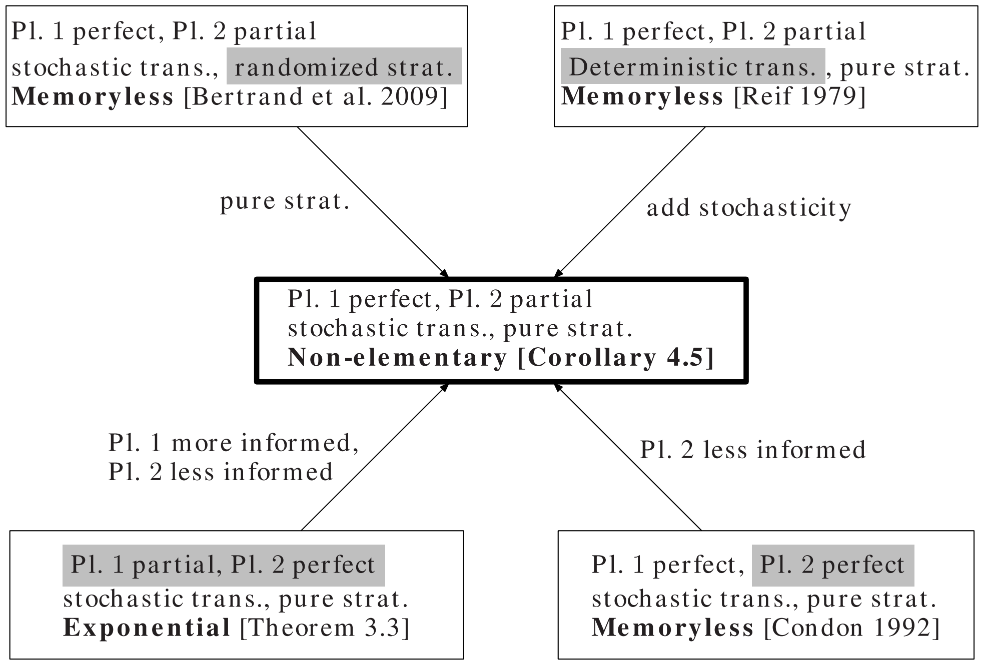

Fig. 10. The surprising non-elementary bound for memory of pure strategies in one-sided partial-observation stochastic games for player 1 perfect and player 2 partial for positive winning with reachability objectives (Theorem 4.7).

are sufficient. This follows from the results of [Reif 1979] because in deterministic games positive winning coincides with sure winning, and the results of [Reif 1979] shows (see [Chatterjee et al. 2007] 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 [Bertrand et al. 2009] 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 [Condon 1992] that in perfect-information stochastic games with reachability objective, pure memoryless optimal strategies exist (note that for pure strategies, perfect-information stochastic games coincide with turn-based perfect-information stochastic games).

(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 3.3).

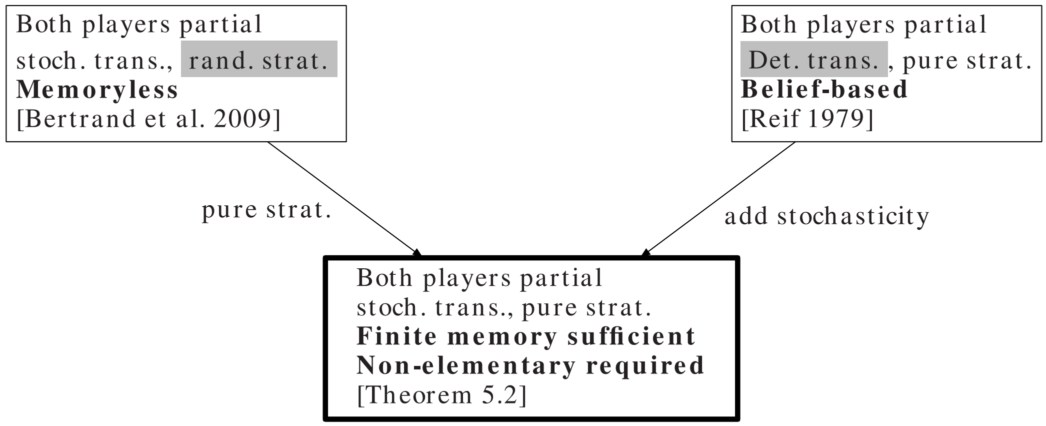

Fig. 11. The memory bounds in two-sided partial-observation stochastic games for positive winning with reachability objectives.

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 surprising memory bound for strategies (non-elementary memory bound is sufficient and necessary in general, whereas for all the related questions memoryless or exponential-size memory strategies are sufficient). See also Fig. 10 for a pictorial illustration.

Discussion for two-sided games. We now discuss the memory requirements for positive winning strategies for two-sided games. For two-sided deterministic games with pure strategies, positive winning coincide with sure winning, and hence it follows from the results of [Reif 1979] that belief-based strategies are sufficient. For two-sided games (deterministic or stochastic) with randomized strategies (action-visible or action-invisible) for positive winning memoryless strategies are sufficient where the randomized memoryless strategy plays all actions uniformly at random [Bertrand et al. 2009]. For two-sided stochastic games with pure strategies, our results show that for positive winning finite-memory strategies are sufficient and non-elementary memory is required in general. Fig. 11 presents a pictorial illustration.

8. CONCLUSION

In this work we considered the classical model of strategies with deterministic update of the memory. We briefly discuss the case when the memory-update function is stochastic and refer to them as stochastic-memory strategies. In a stochastic-memory *randomized* strategy, both the memory update and the next-move function of the strategy are stochastic. In a stochastic-memory *pure* strategy the memory update of the strategy is stochastic, but the next-move function is deterministic (i.e., non-stochastic). A stochastic-memory randomized strategy (with action visible) can be converted to a stochastic-memory pure strategy by simply encoding in an enlarged memory (product with the action space) which action to play next and use the stochastic update of memory states to simulate the randomization of the next-move function. Thus in case of stochastic-memory randomized strategy (both with action visible and action invisible), pure and randomized strategies are all equivalent. Translating a stochastic-memory strategy to a deterministic-memory strategy in general requires infinite memory (even in perfect-information games) [de Alfaro et al. 2007]. Partial-observation games with stochastic-memory strategies have been considered in [Bertrand et al. 2009]. An interesting direction of future work is to obtain computational complexity bounds for the almost-sure and positive problems in one-sided partial-observation stochastic games with player 1 perfect and player 2 partial (see Section 4), and to establish the decidability and complexity of the same problems for two-sided partial-observation stochastic games (Section 6). Our work also makes an unexpected connection between partial-observation stochastic games and counter systems. While counter systems and games have been extensively studied for Petri nets and Vector Addition Systems with States (VASS) [Rackoff 1978; Rosier and Yen 1986; Brázdil et al. 2010; Finkel and Goubault-Larrecq 2012], they have been studied in the context where the operations on the counters are increments and decrements. We show that the analysis of partial-observation stochastic games requires to consider counter systems with operations of increment and division by 2. The analysis of games over counters with the new class of operations gives rise to an interesting research direction for games over counters.

REFERENCES

ABADI, M., LAMPORT, L., AND WOLPER, P. 1989. Realizable and unrealizable specifications of reactive systems. In *Proc. of ICALP*. LNCS 372. Springer, 1–17.

ALUR, R., HENZINGER, T. A., AND KUPFERMAN, O. 2002. Alternating-time temporal logic. *Journal of the ACM 49*, 672–713.

AUMANN, R. 1995. *Repeated Games with Incomplete Information*. MIT Press.

BAIER, C., BERTRAND, N., AND GRÖSSER, M. 2008. On decision problems for probabilistic Büchi automata. In *Proc. of FoSSaCS*. LNCS 4962. Springer, 287–301.

BAIER, C., BERTRAND, N., AND GRÖSSER, M. 2009. The effect of tossing coins in omega-automata. In *Proc. of CONCUR*. LNCS 5710. Springer, 15–29.

BAIER, C. AND GRÖSSER, M. 2005. Recognizing omega-regular languages with probabilistic automata. In *Proc. of LICS*. 137–146.

BERTRAND, N., GENEST, B., AND GIMBERT, H. 2009. Qualitative determinacy and decidability of stochastic games with signals. In *Proc. of LICS*. 319–328.

BERWANGER, D. AND DOYEN, L. 2008. On the power of imperfect information. In *Proc. of FSTTCS*. Dagstuhl Seminar Proceedings 08004. IBFI.

BRÁZDIL, T., JANCAR, P., AND KUCERA, A. 2010. Reachability games on extended vector addition systems with states. In *Proc. of ICALP*. LNCS 6199. Springer, 478–489.

BUKHARAEV, R. G. 1980. Probabilistic automata. *Journal of Mathematical Sciences 13*, 359–386.

CERNÝ, P., CHATTERJEE, K., HENZINGER, T. A., RADHAKRISHNA, A., AND SINGH, R. 2011. Quantitative synthesis for concurrent programs. In *Proc. of CAV*. LNCS 6806. Springer, 243–259.

CHADHA, R., SISTLA, A. P., AND VISWANATHAN, M. 2009a. On the expressiveness and complexity of randomization in finite state monitors. *Journal of the ACM 56*, 1–44.

CHADHA, R., SISTLA, A. P., AND VISWANATHAN, M. 2009b. Power of randomization in automata on infinite strings. In *Proc. of CONCUR*. LNCS 5710. Springer, 229–243.

CHADHA, R., SISTLA, A. P., AND VISWANATHAN, M. 2010. Model checking concurrent programs with nondeterminism and randomization. In *Proc. of FSTTCS*. LIPIcs Series, vol. 8. 364–375.

CHATTERJEE, K., CHMELIK, M., AND TRACOL, M. 2013a. What is decidable about partially observable Markov decision processes with omega-regular objectives. In *Proc. of CSL*.

CHATTERJEE, K. AND DOYEN, L. 2010. The complexity of partial-observation parity games. In *Proc. of LPAR*. LNCS 6397. Springer, 1–14.

CHATTERJEE, K., DOYEN, L., GIMBERT, H., AND HENZINGER, T. A. 2010a. Randomness for free. In *Proc. of MFCS*. LNCS 6281. Springer, 246–257.

CHATTERJEE, K., DOYEN, L., AND HENZINGER, T. A. 2010b. Qualitative analysis of partially-observable Markov decision processes. In *Proc. of MFCS*. LNCS 6281. Springer, 258–269.

CHATTERJEE, K., DOYEN, L., AND HENZINGER, T. A. 2013b. A survey of partial-observation stochastic parity games. *Formal Methods in System Design 43,* 2, 268–284.

CHATTERJEE, K., DOYEN, L., HENZINGER, T. A., AND RASKIN, J.-F. 2007. Algorithms for omega-regular games of incomplete information. *Logical Methods in Computer Science 3,* 3:4.

CHATTERJEE, K., DOYEN, L., NAIN, S., AND VARDI, M. Y. 2013c. The complexity of partial-observation stochastic parity games with finite-memory strategies. Tech. rep., IST Austria. Technical report: IST Austria IST-2013-141 (September 2013).

CHATTERJEE, K. AND TRACOL, M. 2012. Decidable problems for probabilistic automata on infinite words. In *LICS*. 185–194.

CONDON, A. 1992. The complexity of stochastic games. *Information and Computation 96(2)*, 203–224.

DE ALFARO, L. AND HENZINGER, T. A. 2001. Interface automata. In *Proc. of FSE*. ACM Press, 109–120.

DE ALFARO, L., HENZINGER, T. A., AND KUPFERMAN, O. 2007. Concurrent reachability games. *Theor. Comput. Sci. 386,* 3, 188–217.

DE WULF, M., DOYEN, L., AND RASKIN, J.-F. 2006. A lattice theory for solving games of imperfect information. In *Proc. of HSCC*. LNCS 3927. Springer, 153–168.

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

DIMITROVA, R. AND FINKBEINER, B. 2008. Abstraction refinement for games with incomplete information. In *Proc. of FSTTCS*. LIPIcs Series, vol. 2. 175–186.

DOYEN, L. AND RASKIN, J.-F. 2010. Antichains algorithms for finite automata. In *Proc. of TACAS*. LNCS 6015. Springer, 2–22.

EMERSON, E. A. AND JUTLA, C. 1991. Tree automata, mu-calculus and determinacy. In *Proc. of FOCS*. 368–377.

ETESSAMI, K. AND YANNAKAKIS, M. 2006. Recursive concurrent stochastic games. In *ICALP 06: Automata, Languages, and Programming*. Springer.

FIJALKOW, N., GIMBERT, H., AND OUALHADJ, Y. 2012. Deciding the value 1 problem for probabilistic leak-tight automata. In *LICS*. 295–304.

FINKEL, A. AND GOUBAULT-LARRECQ, J. 2012. Forward analysis for WSTS, part II: Complete WSTS. *Logical Methods in Computer Science 8,* 3.

GIMBERT, H. AND OUALHADJ, Y. 2010. Probabilistic automata on finite words: Decidable and undecidable problems. In *Proc. of ICALP (2)*. LNCS 6199. Springer, 527–538.

GRIPON, V. AND SERRE, O. 2009. Qualitative concurrent stochastic games with imperfect information. In *Proc. of ICALP (2)*. LNCS 5556. Springer, 200–211.

HENZINGER, T. A. AND KOPKE, P. 1999. Discrete-time control for rectangular hybrid automata. *Theor. Comp. Science 221*, 369–392.

JAMROGA, W., MAUW, S., AND MELISSEN, M. 2012. Fairness in non-repudiation protocols. In *Proc. of STM*. LNCS 7170. Springer, 122–139.

KECHRIS, A. 1995. *Classical Descriptive Set Theory*. Springer.

KÖNIG, D. 1936. *Theorie der endlichen und unendlichen Graphen*. Akademische Verlagsgesellschaft, Leipzig.

KRESS-GAZIT, H., FAINEKOS, G. E., AND PAPPAS, G. J. 2009. Temporal-logic-based reactive mission and motion planning. *IEEE Transactions on Robotics 25,* 6, 1370–1381.

KUPFERMAN, O. AND VARDI, M. Y. 2000. Synthesis with incomplete informatio. In *Advances in Temporal Logic*. Kluwer Academic Publishers, 109–127.

NAIN, S. AND VARDI, M. Y. 2013. Solving partial-information stochastic parity games. In *LICS*. 341–348.

PAPADIMITRIOU, C. H. AND TSITSIKLIS, J. N. 1987. The complexity of Markov decision processes. *Math. of Op. Research 12*, 441–450.

PAZ, A. 1971. *Introduction to probabilistic automata*. Academic Press, Inc. Orlando, FL, USA.

PNUELI, A. AND ROSNER, R. 1989. On the synthesis of a reactive module. In *Proc. of POPL*. ACM Press, 179–190.

RABIN, M. O. 1963. Probabilistic automata. *Information and Control 6*, 230–245.

RACKOFF, C. 1978. The covering and boundedness problems for vector addition systems. *Theor. Comput. Sci. 6*, 223–231.

RAMADGE, P. J. AND WONHAM, W. M. 1987. Supervisory control of a class of discrete-event processes. *SIAM Journal of Control and Optimization 25,* 1, 206–230.

REIF, J. H. 1979. Universal games of incomplete information. In *Proc. of STOC*. ACM, 288–308.

REIF, J. H. 1984. The complexity of two-player games of incomplete information. *JCSS 29*, 274–301.

REIF, J. H. AND PETERSON, G. L. 1980. A dynamic logic of multiprocessing with incomplete information. In *Proc. of POPL*. ACM, 193–202.

RENAULT, J. 2012. The value of repeated games with an informed controller. *Math. Oper. Res. 37,* 1, 154–179.

ROSENBERG, D., SOLAN, E., AND VIEILLE, N. 2003. The maxmin value of stochastic games with imperfect monitoring. *Int. J. Game Theory 32,* 1, 133–150.

ROSIER, L. E. AND YEN, H.-C. 1986. A multiparameter analysis of the boundedness problem for vector addition systems. *J. Comput. Syst. Sci. 32,* 1, 105–135.

SHAPLEY, L. S. 1953. Stochastic games. *Proc. Nat. Acad. Sci. USA 39*, 1095–1100.

SORIN, S. 2002. *A first course in zero-sum repeated games*. Springer.

THOMAS, W. 1997. Languages, automata, and logic. In *Handbook of Formal Languages*. Vol. 3, Beyond Words. Springer, Chapter 7, 389–455.

TRACOL, M., BAIER, C., AND GRÖSSER, M. 2009. Recurrence and transience for probabilistic automata. In *Proc. of FSTTCS*. LIPIcs Series, vol. 4. 395–406.

VARDI, M. Y. 1985. Automatic verification of probabilistic concurrent finite-state systems. In *Proc. of FOCS*. 327–338.

Appendix

In this section we show that for two-sided partial-observation stochastic games with reachability objectives and randomized action-visible strategies, belief-based randomized strategies are sufficient for almost-sure winning[4]. By Remark 2.3 the result also holds for Büchi objectives. The result was shown in [Chatterjee et al. 2007, Theorem 4.12] for one-sided games with player 1 partial and player 2 perfect, and it was already mentioned in [Bertrand et al. 2009, Theorem 2] that the result can be extended to two-sided games (the proof of [Bertrand et al. 2009, Theorem 2] mentioned that the same technique of [Chatterjee et al. 2007] achieves the result). We present the relevant details of the result, and also the proof we present is simpler as compared to [Chatterjee et al. 2007]. After the relevant details we will present a comparison with the results of [Chatterjee et al. 2007].

Notation. We consider $\mathcal{T}$ as the set of target states and without loss of generality assume that all target states are absorbing. Let Q_W be the set of almost-sure winning beliefs, i.e., the set $\mathcal{B}$ of beliefs such that there is an observation-based randomized action-visible strategy for player 1 to ensure that $\mathcal{T}$ is reached almost-surely against all observation-based strategies of player 2 from the starting belief $\mathcal{B}$. Let $\overline{Q_W}$ be the complement set of beliefs, and it follows by definition that for all beliefs in $\overline{Q_W}$, for all observation-based randomized action-visible strategies for player 1, there exists a counter observation-based strategy for player 2 to ensure that with positive probability $\mathcal{T}$ is not reached.

Allowable actions and the winning strategy. For every $\mathcal{B} \in Q_W$, we say an action a is *allowable* in $\mathcal{B}$ if for every observation $\gamma \in \mathcal{O}_1$, if $(\mathsf{Post}_{a,-}(\mathcal{B}) \cap \gamma) \setminus \mathcal{T} \neq \emptyset$, then $(\mathsf{Post}_{a,-}(\mathcal{B}) \cap \gamma) \setminus \mathcal{T} \in Q_W$ (we follow the terminology of allowable actions introduced in [Chatterjee et al. 2007]). In other words, an action is allowable for a belief $\mathcal{B}$ in Q_W, if the action ensures that the next belief remains within the set Q_W of almost-sure winning beliefs. Thus it follows that if player 1 plays an action that is not allowable, then player 2 can ensure to reach a belief in $\overline{Q_W}$ with positive probability by playing all actions uniformly at random. We consider the belief-based randomized action-visible strategy σ^* for player 1 that for a belief $\mathcal{B} \in Q_W$ plays all allowable actions of $\mathcal{B}$ uniformly at random. First observe that playing strategy σ^* ensures that the belief set is always in Q_W, and to ensure that the next belief is in Q_W it is crucial that the strategy is an action-visible randomized strategy. Our goal is to show that σ^* is an almost-sure winning strategy for all beliefs in Q_W. Once the strategy σ^* is fixed we obtain a player-2 POMDP G_{σ^*}, and let $X_W \subseteq Q_W$ be the set of almost-sure winning beliefs in the POMDP G_{σ^*} for player 1. Let $Y_W = Q_W \setminus X_W$, and we assume towards contradiction that Y_W is non-empty. Hence in the POMDP, from Y_W player 2 can ensure positive safety (i.e., ensure with positive probability that $\mathcal{T}$ is not reached, and thus ensure positive safety for the complement set of $\mathcal{T}$). From the results of [Chatterjee et al. 2010b, Lemma 1] and from the fact that for safety objectives almost-sure and sure winning coincide, it follows that in a POMDP if the set of beliefs from where positve safety can be ensured is non-empty, then the set of beliefs from where sure safety can be ensured is also non-empty. Hence let $Z_W \subseteq Y_W$ be the set of beliefs for player 1 such that player 2 can ensure sure safety in G_{σ^*}. Since we assume Y_W is non-empty we also have Z_W is non-empty. Also it is known that for sure safety, a POMDP is equivalent to a game where all probabilistic choices are resolved by the opponent. Thus we have the following property in Z_W: there is an observation-based strategy π^* for player 2 such that against all observation-based randomized action-visible strate-

[4]The randomized action-visible strategies constructed have deterministic memory update function as defined in Section 2, and not stochastic memory update function as described in Section 8.

gies for player 1 that only plays allowable actions with positive probability (and all actions that are not allowable are played with probability 0), the set of beliefs visited is always in Z_W with certainty. We refer to the above property as sure winning property for player 2 in Z_W. We are now ready to show the contraditcion. Consider a belief in Z_W and an observation-based randomized action-visible strategy σ for player 1 (not necessarily belief-based), and we show that the strategy is not almost-sure winning and thus arrive at a contradiction that $Z_W \subseteq Q_W$ and Q_W is the set of almost-sure winning beliefs. We have two cases:

(1) If there exists a history that is possible with positive probability under σ such that the strategy σ plays an action that is not allowable with positive probability, then the counter strategy for player 2 plays all actions uniformly at random till an action that is not allowable is played and a belief in the set $\overline{Q}_W$ is reached with positive probability. Moreover, player 2 guesses that a belief in $\overline{Q}_W$ is reached in step i with probability $\frac{1}{2^{i+1}}$ and the guess is correct with positive probability for some i. Once a belief in the set $\overline{Q}_W$ is reached with positive probability, then by definition against all observation-based randomized action-visible strategies for player 1, there is a strategy for player 2 to ensure that with positive probability $\mathcal{T}$ is not reached. This ensures that σ is not an almost-sure winning strategy from Z_W.

(2) Otherwise for all possible histories the strategy σ plays only allowable actions, and then the strategy π^* ensures that σ is not almost-sure winning by the sure winning property for player 2 in Z_W.

The above case analysis establishes the contradiction and gives us the following result.

THEOREM 8.1. *In two-sided partial-observation games, for reachability and Büchi objectives, if there exists an observation-based randomized action-visible almost-sure winning strategy, then there exists a belief-based randomized action-visible almost-sure winning strategy.*

Comparison with related results. In [Chatterjee et al. 2007] the result was established for one-sided games with player 1 partial and player 2 perfect, and it was mentioned in [Bertrand et al. 2009] that the same proof can be extended to two-sided games. The proof of [Chatterjee et al. 2007] was more involved as along with proving the result for belief-based strategy, it also provided an EXPTIME upper bound, and the proof required an involved construction of an exponential-size game and new notions of strategies (called equivalence-preserving strategies). As shown above, the proof for belief-based strategies can be obtained for the general two-sided games with a simpler argument that does not construct the involved exponential-size game. However the simpler argument does not give an EXPTIME upper bound (even in the one-sided case), and the naive complexity bound that can be obtained from the sufficiency of belief-based strategies for almost-sure winning is 2EXPTIME (enumerating over the set of double-exponentially many belief-based strategies, and verifying the exponential-size POMDP obtained after fixing the strategy in double exponential time using the results of [Chatterjee et al. 2010b]). However, for two-sided games the 2EXPTIME complexity bound is optimal as a matching 2EXPTIME lower bound for almost-sure reachability in two-sided games with randomized action-visible strategies was established in [Bertrand et al. 2009]. The above proof only shows that if there is an almost-sure winning strategy, then there is a belief-based one, and does not establish that in the complement set there is a spoiling strategy for the opponent player that prevents almost-sure winning against all strategies (i.e., the above proof does not establish the important property of determinacy). The fundamental determinacy property for randomized action-visible strategies for almost-sure reachability for two-sided games was established in [Bertrand et al. 2009]. For the lower bound and other fundamental results related to qualitative determinacy of two-sided games we refer the reader to [Bertrand et al. 2009].