Revelations: A Decidable Class of POMDPs with Omega-Regular Objectives

Marius Belly, Nathanaël Fijalkow, Hugo Gimbert, Florian Horn, Guillermo Alberto Pérez, Pierre Vandenhove

Revelations: A Decidable Class of POMDPs with Omega-Regular Objectives

Marius Belly[1], **Nathanaël Fijalkow**[1], **Hugo Gimbert**[1],
Florian Horn[2], **Guillermo A. Pérez**[3], **Pierre Vandenhove**[1*]

[1]CNRS, LaBRI, Université de Bordeaux, France
[2]CNRS, IRIF, Université de Paris, France
[3]University of Antwerp – Flanders Make, Antwerp, Belgium

Abstract

Partially observable Markov decision processes (POMDPs) form a prominent model for uncertainty in sequential decision making. We are interested in constructing algorithms with theoretical guarantees to determine whether the agent has a strategy ensuring a given specification with probability 1. This well-studied problem is known to be undecidable already for very simple omega-regular objectives, because of the difficulty of reasoning on uncertain events. We introduce a revelation mechanism which restricts information loss by requiring that almost surely the agent has eventually full information of the current state. Our main technical results are to construct exact algorithms for two classes of POMDPs called *weakly* and *strongly revealing*. Importantly, the decidable cases reduce to the analysis of a finite belief-support Markov decision process. This yields a conceptually simple and exact algorithm for a large class of POMDPs.

Code — https://github.com/gaperez64/pomdps-reveal

1 Introduction

Partially observable Markov decision processes (POMDPs) form a prominent model for uncertainty in sequential decision making. They were defined in the 1960s (Åström 1965) for operations research and introduced in artificial intelligence by the seminal paper of Kaelbling, Littman, and Cassandra (1998). We consider POMDPs from a model-based point of view common in planning and in formal methods. Our goal is to construct exact (as opposed to approximate) algorithms that take as an input a complete description of the POMDP and construct a strategy ensuring a given specification. A long line of work has established that most formulations of this problem are undecidable. For instance, even in the extreme case where the agent has no information and the goal is to reach a target state with arbitrarily high probability, complex convergence phenomena occur, implying strong undecidability results (Madani, Hanks, and Condon 2003; Gimbert and Oualhadj 2010; Fijalkow 2017).

In this work, we are interested in constructing *almost-sure strategies*, meaning strategies ensuring their specifications with probability 1. We consider the class of omega-regular objectives (all expressible as *parity objectives*), which is a robust class including properties expressible in Linear Temporal Logic (Pnueli 1977; Giacomo and Vardi 2013). Determining whether there exists an almost-sure strategy against the subclass of CoBüchi objectives (requiring to avoid a target from some point onwards) is undecidable (Chatterjee, Chmelik, and Tracol 2016; Bertrand, Genest, and Gimbert 2017). There is a vast body of work towards approximate and practical solutions: for instance, using interpolation in the belief space (Lovejoy 1991), approximation of the value function (Hauskrecht 2000), or Monte Carlo tree search approaches (Silver and Veness 2010). This is orthogonal to the current paper since we focus on exact algorithms.

Our starting point is a simple approach to construct almost-sure strategies: from the POMDP, we build a Markov decision process (MDP) whose states are *supports of the beliefs* of the POMDP. In other words, we store information about which states we can be in, but abstract away the probabilities. The *belief-support MDP* serves as a finite abstraction of the POMDP; one could expect that there exists an almost-sure strategy in the POMDP if and only if there exists one in the corresponding belief-support MDP. Unfortunately, this abstraction is neither sound nor complete; we present a simple counterexample in Figure 1.

The fundamental question we ask in this paper is whether **there are natural sufficient conditions which make the belief-support abstraction correct**. Conceptually, the failure of this abstraction is due to information loss and its accumulation over time.

We introduce a **revelation mechanism** which restricts information loss by requiring that, almost surely, the agent has eventually full information of the current state. Intuitively, by forbidding information loss from accumulating for an unbounded amount of time, the revelation mechanism removes the convergence issues leading to undecidability. Practically, we conjecture that revelation is a commonly occurring phenomenon in partial observability; a canonical example is systems with a small probability of resetting infinitely often, and where this reset is observable. We leave to future work to investigate this question further. Other approaches to restrict information loss have been proposed; we refer to the related works (Section 6) for an additional discussion.

Our contributions.

- We study two properties of POMDPs based on the reve-

*Authors are listed in alphabetical order.

Figure 1: We consider the POMDP on the LHS: there is a single signal s, so no information is ever given about the exact state we are in (a behavior the revelation mechanisms forbid!). Yet, almost surely, we reach q_1. The *priorities* indicated on states constitute a parity condition inducing the objective "eventually never visiting q_1", which clearly cannot be ensured almost surely. We represent the *belief-support MDP* on the RHS: the two states are $\{q_0\}$ and $\{q_0, q_1\}$, and only the state $\{q_0, q_1\}$ is visited infinitely often. To assign priorities to the states of this MDP, there are two natural candidates: "maximal priority semantics" and "minimal priority semantics", meaning that we assign either the maximal or minimal priority from the states in the belief support. In this figure, we use the maximal priority semantics: the priority of $\{q_0, q_1\}$ is thus 2, so the belief-support MDP is winning. This means that the analysis of the belief-support MDP is not sound in general. By tweaking the priorities in this example, one can show that both priority semantics are neither sound nor complete.

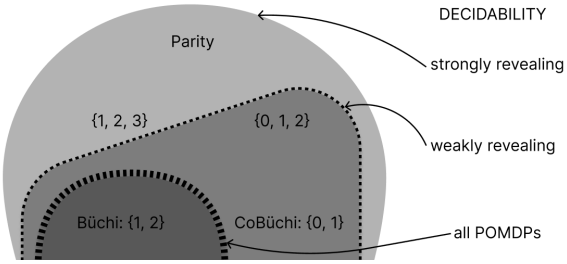

Figure 2: Summary of our results: decidable subclasses of the *parity* objective depending on the revelation mechanism.

lation mechanism, called *weak* and *strong revelations*.

- We obtain decidability (and undecidability) results for both classes. Importantly, the decidable cases reduce to the analysis of the finite belief-support MDP. A summary of our contributions for POMDPs is provided in Figure 2. We also briefly consider the class of *two-player games of partial information*, to show that our revealing mechanisms do not suffice for decidability on this larger class.
- We provide a simple implementation of the algorithm as a proof of concept. We provide a comparison between our algorithm and off-the-shelf deep reinforcement learning (DRL) trained via an observation wrapper. As we will show in the paper, the MDP induced by the belief supports carries sufficient information to play in revealing POMDPs; hence, we used a wrapper implementing a subset construction on the fly to generate the current belief support, and focused on algorithms intended for MDPs. Spending moderate effort on reward engineering and hyperparameter tuning, we have been unable to match the performance of our algorithm using such DRL approaches (see Figure 3).

This yields a conceptually simple and exact algorithm for a large class of POMDPs. The importance of our results can be appreciated by the following remark: instead of a subclass of POMDPs, the revelation mechanism can be seen as new semantics for *all* POMDPs. In that sense, we obtain decidability results for an *optimistic* semantics of POMDPs which, to the best of our knowledge, has not been done before. We refer to Section 5.3 for more details on this point of view.

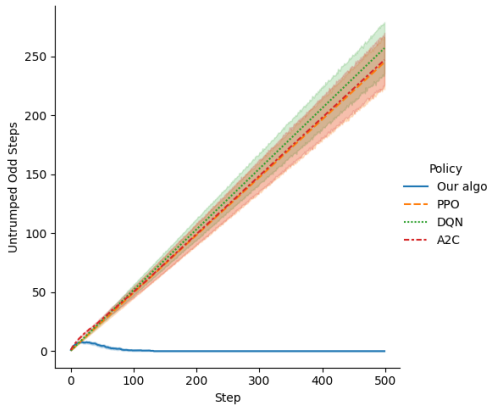

Figure 3: Omega-regular specifications have a natural interpretation in terms of *bad* events that must all be *trumped* by future *good* events. Along a simulation of the POMDP, one can keep track of the number of steps from the last bad event that has not yet been trumped (i.e., lower is better). Here, we depict this value, per step (from 1 to 500) over 500 simulations of a revealing version of the classical tiger POMDP (Cassandra, Kaelbling, and Littman 1994). A2C, DQN, and PPO are (`MlpPolicy`) strategies obtained from the *stable-baselines* library (Raffin et al. 2021), trained (for a total of 10k time steps) with default parameter values using a simple reward scheme: a good event yields a reward of 100; a bad one, -1. In the simulations, the trained models are queried for deterministic action predictions. The example used will be discussed in Section 5, Example 2.

Outline. We define our stochastic models in Section 2. We then introduce the belief-support MDP in Section 3. Sections 4 and 5 are devoted to our revealing mechanisms and to our (un)decidability results. We conclude with additional related works and perspectives in Sections 6 and 7. Due to a lack of space, most proofs are deferred to the appendix.

2 Preliminaries

A *(discrete) probability distribution* on a finite set X is a function $d\colon X \to [0,1]$ such that $\sum_{x\in X} d(x) = 1$. The set of all probability distributions on X is denoted $\mathcal{D}(X)$.

The *support* $\mathsf{supp}(d)$ of a probability distribution d is the set $\{x \in X \mid d(x) > 0\}$. We let $|X|$ denote the number of elements in a set X.

2.1 POMDPs

A *partially observable Markov decision process* (POMDP) is a tuple $\mathcal{P} = (Q, \mathsf{Act}, \mathsf{Sig}, \delta, q_0)$ such that Q is a finite set of *states*, Act is a finite set of *actions*, Sig is a finite state of *signals*, $\delta\colon Q \times \mathsf{Act} \to \mathcal{D}(\mathsf{Sig} \times Q)$ is the *transition function*, and $q_0 \in Q$ is an *initial state*.

A *play* of a POMDP $\mathcal{P} = (Q, \mathsf{Act}, \mathsf{Sig}, \delta, q_0)$ is an infinite sequence $\pi = q_0 a_1 s_1 q_1 a_2 s_2 \ldots \in (Q \cdot \mathsf{Act} \cdot \mathsf{Sig})^\omega$ such that, for all $i \geq 0$, $\delta(q_i, a_{i+1})(s_{i+1}, q_{i+1}) > 0$. A *history* h of a POMDP is a finite prefix of a play ending in a state (it is an element of $(Q \cdot \mathsf{Act} \cdot \mathsf{Sig})^* \cdot Q$). If $h = q_0 a_1 s_1 q_1 \ldots a_n s_n q_n$, we write $\mathsf{last}(h)$ for q_n. In practice, states are not fully observable; we define an *observable history* as the projection of a history to the subsequence in $(\mathsf{Act} \cdot \mathsf{Sig})^*$. We write $\mathsf{obs}(h)$ for the observable history derived from a history h, i.e., the same sequence with the states removed.

For q a state of a POMDP $\mathcal{P} = (Q, \mathsf{Act}, \mathsf{Sig}, \delta, q_0)$, we define $\mathcal{P}^q$ to be the POMDP $(Q, \mathsf{Act}, \mathsf{Sig}, \delta, q)$ with only a change of initial state. We let $\beta_\mathcal{P} = \min\{\delta(q,a)(s,q') \mid q, q' \in Q,\ a \in \mathsf{Act},\ s \in \mathsf{Sig}, \text{ and } \delta(q,a)(s,q') > 0\}$ denote the least non-zero probability occurring in $\mathcal{P}$.

A *Markov decision process* (MDP) is a tuple $\mathcal{M} = (Q, \mathsf{Act}, \delta, q_0)$ where $\delta\colon Q \times \mathsf{Act} \to \mathcal{D}(Q)$. Formally, an MDP $\mathcal{M} = (Q, \mathsf{Act}, \delta, q_0)$ can be seen as a POMDP $\mathcal{P} = (Q, \mathsf{Act}, \mathsf{Sig}, \delta, q_0)$ such that $\mathsf{Sig} = \{s_q \mid q \in Q\}$ and for all $q, q', q'' \in Q$ and $a \in \mathsf{Act}$, $\delta(q,a)(s_{q''}, q') > 0$ if and only if $q' = q''$. In practice, it means that the last signal always uniquely determines the current state. MDPs have "complete observation", whereas POMDPs have "partial observation". For a POMDP $\mathcal{P} = (Q, \mathsf{Act}, \mathsf{Sig}, \delta, q_0)$, we define the *underlying MDP of* $\mathcal{P}$ to be the MDP $(Q, \mathsf{Act}, \delta', q_0)$ with $\delta'(q,a)(q') = \sum_{s \in \mathsf{Sig}} \delta(q,a)(s,q')$.

Remark 1. *The observable information in POMDPs is here provided through* signals *that appear along transitions. This contrasts with state-based* observations *that partition the state space, which are also frequently used to model POMDPs. Both models are polynomially equivalent: a POMDP with observations can be transformed into an equivalent POMDP with signals on the same state space, while the converse requires an increase of the state space linear in* $|\mathsf{Sig}|$. *Both choices are convenient, but using signals make the definition of* strongly revealing *(Definition 2) more natural, which is why we opted for this convention.*

Strategies. Let $\mathcal{P} = (Q, \mathsf{Act}, \mathsf{Sig}, \delta, q_0)$ be a POMDP. An *(observation-based) strategy in* $\mathcal{P}$ is a function that makes decisions based on the current observable history, i.e., it is a function $\sigma\colon (\mathsf{Act} \cdot \mathsf{Sig})^* \to \mathcal{D}(\mathsf{Act})$. We can define strategies in MDPs similarly (i.e., assuming that Sig gives the information of the current state), but we assume for convenience that a strategy is a function $\sigma\colon (\mathsf{Act} \cdot Q)^* \to \mathcal{D}(\mathsf{Act})$ in this case. An observable history $a_1 s_1 \ldots a_n s_n$ is *consistent with a strategy* σ if for all $1 \leq i < n$, $\sigma(a_1 s_1 \ldots a_i s_i)(a_{i+1}) > 0$.

A strategy σ is *pure* if for all observable histories $h \in (\mathsf{Act} \cdot \mathsf{Sig})^*$, $\sigma(h)$ is a Dirac distribution; in other words, if σ is a function $(\mathsf{Act} \cdot \mathsf{Sig})^* \to \mathsf{Act}$. We let $\Sigma(\mathcal{P})$ denote the set of strategies in POMDP $\mathcal{P}$ and $\Sigma_\mathsf{P}(\mathcal{P})$ denote the set of pure strategies in $\mathcal{P}$.

For an MDP $\mathcal{M}$, a strategy σ in $\mathcal{M}$ is *memoryless* if its decisions are only based on the current state: i.e., if for all histories h_1, h_2, $\mathsf{last}(h_1) = \mathsf{last}(h_2)$ implies $\sigma(h_1) = \sigma(h_2)$. We only define the memoryless notion for MDPs.

Probability measure induced by a strategy. Let $\mathcal{P} = (Q, \mathsf{Act}, \mathsf{Sig}, \delta, q_0)$ be a POMDP. For a history h of $\mathcal{P}$, we define $\mathsf{Cyl}(h)$ (the *cylinder of* h) to be the set of all plays starting with h, i.e., $h(\mathsf{Act} \cdot \mathsf{Sig} \cdot Q)^\omega$. Given a strategy σ, we can define a probability measure $\mathbb{P}^\mathcal{P}_\sigma[\cdot]$ on infinite plays. This function is naturally defined over cylinders by induction. We define $\mathbb{P}^\mathcal{P}_\sigma[\mathsf{Cyl}(q_0)] = 1$, and $\mathbb{P}^\mathcal{P}_\sigma[\mathsf{Cyl}(q)] = 0$ for $q \in Q$, $q \neq q_0$. For a history $h = h' a s q$, we define $\mathbb{P}^\mathcal{P}_\sigma[\mathsf{Cyl}(h)] = \mathbb{P}^\mathcal{P}_\sigma[\mathsf{Cyl}(h')] \cdot \sigma(\mathsf{obs}(h'))(a) \cdot \delta(\mathsf{last}(h'), a)(s, q)$. By Ionescu-Tulcea extension theorem (Klenke 2007), this function can be uniquely extended to a probability distribution $\mathbb{P}^\mathcal{P}_\sigma[\cdot]$ over the Borel sets of infinite plays induced by all cylinders.

We also use this probability distribution to measure sets of infinite sequences in Q^ω, by associating a set $W \subseteq Q^\omega$ with the set $\bigcup_{q_0 q_1 \ldots \in W} q_0 \mathsf{ActSig} q_1 \mathsf{ActSig} q_2 \ldots \subseteq (Q \times \mathsf{Act} \times \mathsf{Sig})^\omega$. Similarly, we use this probability distribution to measure events based only on signals, by associating a set $S \subseteq \mathsf{Sig}^\omega$ with the set $\bigcup_{s_1 s_2 \ldots \in S} Q \mathsf{Act} s_1 Q \mathsf{Act} s_2 Q \ldots \subseteq (Q \times \mathsf{Act} \times \mathsf{Sig})^\omega$.

Objectives. Let $\mathcal{P} = (Q, \mathsf{Act}, \mathsf{Sig}, \delta, q_0)$ be a POMDP. An *objective* $W \subseteq Q^\omega$ is a measurable set of infinite sequences of states. Note that observing an infinite sequence of signals (but not the states) may not always be sufficient to determine whether a play satisfies an objective.

Given a set $F \subseteq Q$, the *reachability objective* $\mathsf{Reach}(F) = \{q_0 q_1 \ldots \in Q^\omega \mid \exists i \geq 0, q_i \in F\}$ is the set of plays that visit a state in F at least once. For $k \in \mathbb{N}$, we write $\mathsf{Reach}^{\leq k}(F) = \{q_0 q_1 \ldots \in Q^\omega \mid \exists i, 0 \leq i \leq k, q_i \in F\}$ for the set of plays that reach F in at most k steps. Given a set $F \subseteq Q$, the *safety objective* $\mathsf{Safety}(F)$ is the set of plays that never visit any state in F.

Given a *priority function* $p\colon Q \to \{0, \ldots, d\}$ (where $d \in \mathbb{N}$), the *parity objective* $\mathsf{Parity}(p) = \{q_0 q_1 \ldots \in Q^\omega \mid \limsup_{i \geq 0} p(q_i) \text{ is even}\}$ is the set of infinite plays whose highest priority seen infinitely often is even. A *Büchi objective* is a parity objective $\mathsf{Parity}(p)$ such that $p\colon Q \to \{1, 2\}$, and a *CoBüchi objective* is a parity objective $\mathsf{Parity}(p)$ such that $p\colon Q \to \{0, 1\}$. For $Q' \subseteq Q$, we write $\mathsf{Büchi}(Q')$ for the set of infinite plays that visit Q' infinitely often. It is equal to $\mathsf{Parity}(p)$ for the priority function p such that $p(q) = 2$ if $q \in Q'$, and $p(q) = 1$ otherwise.

For an objective W, a strategy σ is *almost sure* if $\mathbb{P}^\mathcal{P}_\sigma[W] = 1$, and is *positively winning* if $\mathbb{P}^\mathcal{P}_\sigma[W] > 0$. We say that an objective W has *value* 1 in a POMDP $\mathcal{P}$ if $\sup_{\sigma \in \Sigma(\mathcal{P})} \mathbb{P}^\mathcal{P}_\sigma[W] = 1$.

2.2 Beliefs and belief supports

Let $\mathcal{P} = (Q, \mathsf{Act}, \mathsf{Sig}, \delta, q_0)$ be a POMDP. A *belief* $\mathfrak{b} \in \mathcal{D}(Q)$ is a probability distribution on Q. A *belief support* $b \in 2^Q \setminus \{\emptyset\}$ is the support of a belief. For brevity, we write

$2^Q_\emptyset$ for $2^Q \setminus \{\emptyset\}$. At every step, beliefs and belief supports can be updated when playing an action and observing a signal. We show how to do so for belief supports: we define a function $\mathcal{B}\colon 2^Q_\emptyset \times \mathsf{Act} \times \mathsf{Sig} \to 2^Q_\emptyset$ that updates the belief support. For $b \in 2^Q_\emptyset$, $a \in \mathsf{Act}$, $s \in \mathsf{Sig}$, we define $\mathcal{B}(b,a,s) = \{q' \in Q \mid \exists q \in b, \delta(q,a)(s,q') > 0\}$. We extend this function in a natural way to a function $\mathcal{B}^*\colon 2^Q_\emptyset \times (\mathsf{Act} \cdot \mathsf{Sig})^* \to 2^Q_\emptyset$. Objectives $\mathsf{Reach}(B)$ and $\mathsf{Büchi}(B)$ can be naturally extended to sets of belief supports $B \subseteq 2^Q_\emptyset$ (see Appendix C).

Beliefs carry more information than belief supports, as they contain the exact probability of being in a particular state, while belief supports only contain the qualitative information of the possible current states. Observe that when the belief support is a singleton (i.e., $b = \{q\}$ for some $q \in Q$), knowing the precise belief does not yield more information than knowing the belief support, as all the probability mass is in one of the states. Our "revealing" restrictions on POMDPs defined later will exploit this fact.

3 The belief-support MDP

For a POMDP $\mathcal{P} = (Q, \mathsf{Act}, \mathsf{Sig}, \delta, q_0)$, the *belief-support MDP of* $\mathcal{P}$ is the MDP $\mathcal{P}_\mathcal{B} = (2^Q_\emptyset, \mathsf{Act}, \delta_\mathcal{B}, \{q_0\})$ where for $b, b' \in 2^Q_\emptyset$ and $a \in \mathsf{Act}$, $\delta_\mathcal{B}(b,a)(b') > 0$ if and only if there is $s \in \mathsf{Sig}$ such that $\mathcal{B}(b,a,s) = b'$. We assume the distribution to be uniform over successors with positive probability.

We can show that for multiple simple objectives, the POMDP and its belief-support MDP behave in a similar way. For example, sets of belief supports that can be reached with a positive probability are the same in the POMDP and its belief-support MDP (Appendix C, Lemma 2); if a set of belief supports is reachable almost surely in the POMDP, it is also the case in the belief-support MDP (Appendix C, Lemma 3).

There is a natural way to lift a strategy in the belief-support MDP to a strategy in the POMDP. We define a notation to go from a sequence of signals to the induced sequence of belief supports. Let $h = a_1 s_1 \dots a_n s_n \in (\mathsf{Act} \cdot \mathsf{Sig})^*$ be a possible observable history in $\mathcal{P}$. For $1 \leq i \leq n$, let $b_i = \mathcal{B}^*(\{q_0\}, a_1 s_1 \dots a_i s_i)$ be the belief support after i steps. We define B_h to be the history $a_1 b_1 \dots a_n b_n$ of $\mathcal{P}_\mathcal{B}$. Let $\sigma_\mathcal{B} \in \Sigma(\mathcal{P}_\mathcal{B})$ be a strategy in the belief-support MDP of a POMDP $\mathcal{P}$. We define a strategy $\widehat{\sigma_\mathcal{B}}$ in $\mathcal{P}$ derived from the strategy $\sigma_\mathcal{B}$: for $h \in (\mathsf{Act} \cdot \mathsf{Sig})^*$, we fix $\widehat{\sigma_\mathcal{B}}(h) = \sigma_\mathcal{B}(B_h)$.

4 Weakly revealing POMDPs

We define here our first *revealing* property for POMDPs, which requires that, infinitely often and almost surely, the current state can be deduced by looking at the previous sequence of signals. Formally, we write $B^\mathcal{P}_{\mathsf{sing}} = \{\{q\} \mid q \in Q\}$ for the set of singleton belief supports of a POMDP $\mathcal{P} = (Q, \mathsf{Act}, \mathsf{Sig}, \delta, q_0)$. An observable history $h \in (\mathsf{Act} \cdot \mathsf{Sig})^*$ such that $\mathcal{B}^*(\{q_0\}, h) \in B^\mathcal{P}_{\mathsf{sing}}$ is called a *revelation*.

Definition 1 (Weakly revealing)**.** *A POMDP $\mathcal{P}$ is* weakly revealing *if, for all strategies $\sigma \in \Sigma(\mathcal{P})$, we have $\mathbb{P}^\mathcal{P}_\sigma\big[\mathsf{Büchi}(B^\mathcal{P}_{\mathsf{sing}})\big] = 1$; i.e., infinitely many revelations occur almost surely for all strategies.*

In particular, POMDPs that "reset" infinitely often, and whose reset can be observed with a dedicated signal, are weakly revealing. We will use one such example in Figure 4.

One can give probabilistic bounds on the occurrence of a revelation for a weakly revealing POMDP (see Lemma 5 in Appendix D with $F = B^\mathcal{P}_{\mathsf{sing}}$): starting from any reachable belief, a revelation occurs within $2^{|Q|} - 1$ steps with probability at least $\beta_\mathcal{P}^{2^{|Q|}-1}$. The bound is asymptotically tight: there is a weakly revealing POMDP with $n+2$ states, 1 action, and n signals where we need at least $2^n - 1$ steps before observing a revelation with positive probability. Details are provided in Example 4, Appendix E.

4.1 Soundness of the belief-support MDP

In this section, we show that, for *weakly revealing* POMDPs, the existence of an almost-sure strategy in the belief-support MDP (with an adequate priority function) implies the existence of an almost-sure strategy in the POMDP.

For the priority function of the belief-support MDP, we consider the "maximal priority" semantics. Formally, let $\mathcal{P} = (Q, \mathsf{Act}, \mathsf{Sig}, \delta, q_0)$ be a POMDP, and $\mathcal{P}_\mathcal{B}$ be its belief-support MDP. Let $p\colon Q \to \{0, \dots, n\}$ be a priority function on $\mathcal{P}$, inducing the objective $\mathsf{Parity}(p)$. We extend this function to the belief-support MDP: for $b \in 2^Q_\emptyset$, we define

$$p_\mathcal{B}(b) = \max\{p(q) \mid q \in b\}.$$

Without any assumption, the belief-support MDP may be unsound, already for Büchi objectives; there may be an almost-sure strategy in the belief-support MDP, but not in the POMDP. An example illustrating this was given in Figure 1. Surprisingly, it is sound for CoBüchi objectives without any assumption (see Lemma 6 in Appendix E).

Under the weakly revealing semantics, almost-sure strategies of the belief-support MDP carry over to the POMDP for all parity objectives. In other words, the analysis of the belief-support MDP is sound. We recall that pure memoryless strategies suffice to reach the optimal value for parity objectives in MDPs (Chatterjee and Henzinger 2012).

Proposition 1. *Let $\mathcal{P} = (Q, \mathsf{Act}, \mathsf{Sig}, \delta, q_0)$ be a weakly revealing POMDP with priority function p, and let $\mathcal{P}_\mathcal{B}$ be its belief-support MDP with priority function $p_\mathcal{B}$. Assume there is an almost-sure strategy $\sigma_\mathcal{B}$ for $\mathsf{Parity}(p_\mathcal{B})$ in $\mathcal{P}_\mathcal{B}$; by (Chatterjee and Henzinger 2012), we may assume $\sigma_\mathcal{B}$ to be pure and memoryless. Then, $\widehat{\sigma_\mathcal{B}}$ is an almost-sure strategy for $\mathsf{Parity}(p)$ in $\mathcal{P}$.*

The complete proof is in Appendix E.

4.2 Decidability of parity for priorities 0, 1, and 2

We show that the existence of an almost-sure strategy in a weakly revealing POMDP implies the existence of an almost-sure strategy in its belief-support MDP when priorities are in $\{0,1,2\}$. This provides a converse to Proposition 1 when priorities are restricted to $\{0,1,2\}$. We will see that it is not the case for priorities in $\{1,2,3\}$ in the next section; this result is therefore optimal w.r.t. the priority used. We emphasize that parity objectives with priorities $\{0,1,2\}$ encompass both Büchi and CoBüchi objectives. This result

is false without the weakly revealing assumption; see the simple POMDP in Figure 1. The proof is in Appendix E.

Proposition 2. *Let $\mathcal{P} = (Q, \mathsf{Act}, \mathsf{Sig}, \delta, q_0)$ be a weakly revealing POMDP with priority function p with values in $\{0,1,2\}$. Let $\mathcal{P}_{\mathcal{B}}$ be its belief-support MDP with priority function $p_{\mathcal{B}}$. If there is an almost-sure strategy for $\mathsf{Parity}(p)$ in $\mathcal{P}$, then there is an almost-sure strategy for $\mathsf{Parity}(p_{\mathcal{B}})$ in $\mathcal{P}_{\mathcal{B}}$.*

From the above, we deduce a complexity upper bound; a matching lower bound is proved in the appendix.

Theorem 1. *The existence of an almost-sure strategy for parity objectives with priorities in $\{0,1,2\}$ in weakly revealing POMDPs is EXPTIME-complete.*

Proof. The EXPTIME algorithm is a consequence of the results from this section: by Proposition 1 (soundness of the belief-support MDP) and Proposition 2 (completeness), we reduce the problem to the existence of an almost-sure strategy for a parity objective with priorities in $\{0,1,2\}$ in an MDP of size exponential in $|Q|$. The existence of an almost-sure strategy for parity objectives is decidable in polynomial time in MDPs (Baier and Katoen 2008, Theorem 10.127). Proposition 1 also constructs an almost-sure strategy in $\mathcal{P}$.

The EXPTIME-hardness is proved in Proposition 4 (Appendix F), already for CoBüchi objectives (i.e., with priorities in $\{0,1\}$) and for the more restricted class of strongly revealing POMDPs. □

Remark 2. *The algorithm also gives an upper bound on the size of the strategies for parity objectives with priorities in $\{0,1,2\}$ in weakly revealing POMDPs. As we reduce to the analysis of an exponential-size MDP and that memoryless strategies suffice for parity objectives in MDPs, given Proposition 1, it means that a strategy of exponential size suffices in the POMDP. We can also prove an exponential lower bound: see Example 4 (Appendix E).*

The weakly revealing property is decidable in 2-EXPTIME. To see it, extend the POMDP with the information of the current belief support (this creates an exponential POMDP with state space $Q \times 2^Q_{\emptyset}$). This extended POMDP has no positively winning strategy for $\mathsf{Safety}(Q \times B^{\mathcal{P}}_{\mathsf{sing}})$ if and only if it is weakly revealing. As the existence of a positively winning strategy for safety objectives is EXPTIME-complete (Chatterjee, Chmelik, and Tracol 2016), the complexity of this algorithm is doubly exponential. We only show below that deciding whether a POMDP is weakly revealing is EXPTIME-hard, thereby not completely settling its complexity. We discuss in Section 7 future works related to this question.

Lemma 1. *Deciding whether a POMDP is weakly revealing is EXPTIME-hard.*

4.3 Undecidability of parity for priorities 1, 2, and 3

The previous section suggests that analyzing the belief-support MDP is a sound and complete approach for weakly revealing POMDPs with parity objectives with priorities in $\{0,1,2\}$. One may wonder whether it is complete for any priority function. Unfortunately, this fails to hold in general, already for priority functions taking values in $\{1,2,3\}$. We discuss one such example below.

Example 1. *Consider the POMDP $\mathcal{P}$ in Figure 4. This POMDP is weakly revealing, as state q_0 is visited infinitely often for any strategy and is revealed through signal s_0. The only choice in this POMDP is in states q_1 and q_1': whether to play a and move to q_0 or $\{q_1, q_1'\}$, or to play c and go to q_2 or q_3. Observe that when the game starts in q_0, the only reachable belief supports are $\{q_0\}$, $\{q_1, q_1'\}$, and $\{q_2, q_3\}$, which all have a maximal odd priority. Hence, the belief-support MDP with priority function $p_{\mathcal{B}}$ trivially has no almost-sure (and even positively) winning strategy. However, we show that there is an almost-sure strategy in $\mathcal{P}$.*

The only way to win in this POMDP is to visit q_2 infinitely often while visiting q_3 only finitely often. To do so, observe that when a is played multiple times in a row and only receives signal s_1, the probability to be in q_1' becomes arbitrarily close to 1. Formally, if σ_a is the strategy that only plays a, we have that for $n > 0$,

$$\mathbb{P}^{\mathcal{P}}_{\sigma_a}[Q^n q_1' \mid (s_1)^n] = 1 - \mathbb{P}^{\mathcal{P}}_{\sigma_a}[q_0 (q_1)^n \mid (s_1)^n] = 1 - \frac{1}{2^n}.$$

For $n > 0$, let σ_n be the strategy that plays only a until s_1 has been seen n times in a row*, and when that is the case, plays c. Let us divide a play in this POMDP into rounds 1, 2,...; every time we go back to q_0 after visiting q_2 or q_3, we move to the next round. Consider the strategy that plays σ_n in round n. This strategy ensures that infinitely many rounds happen, because at each round n, it will eventually succeed in seeing n occurrences of s_1 in a row. At each round n, c is eventually played with probability 1. By the above equation, q_3 is seen with probability $\frac{1}{2^n}$ and q_2 is seen with probability $1 - \frac{1}{2^n}$. State q_2 is clearly seen infinitely often almost surely. However, the probability that q_3 is never seen anymore after round n is equal to $\prod_{i=n}^{\infty}(1 - \frac{1}{2^i})$, which is positive and increases as n grows to ∞. We deduce that the probability that q_3 is seen at most finitely often is 1.*

Generalizing the above example, we show that if we allow p to take values in $\{1,2,3\}$, the existence of almost-sure strategies in weakly revealing POMDPs is undecidable. We provide here a proof sketch; a full proof is in Appendix E.

Theorem 2. *The existence of an almost-sure strategy in weakly revealing POMDPs with a parity objective with priorities in $\{1,2,3\}$ is undecidable. The same holds for the existence of a positively winning strategy.*

Our proof uses a reduction from the value-1 problem in *probabilistic automata*. A *probabilistic automaton* (Rabin 1963) is a tuple $\mathcal{A} = (Q, \mathsf{Act}, \delta, q_0)$. One can define their semantics through POMDPs: they behave like POMDPs in which we assume that the signals bring no information (Sig is a singleton). No useful information is provided by the signals along a play (beyond the number of steps played); pure strategies therefore correspond to words on alphabet Act.

Intuitively, the proof expands on the POMDP in Figure 4 by replacing states q_1, q_1' by a copy of a probabilistic automaton $\mathcal{A}$: the transition from q_0 goes to the initial state of

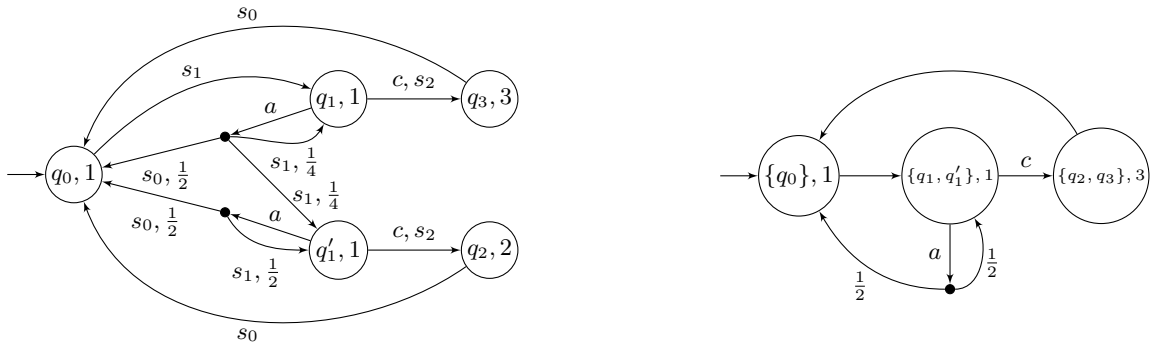

Figure 4: The POMDP $\mathcal{P}$ from Example 1 (depicted on the left) with an almost-sure strategy, but whose belief-support MDP (depicted on the right) has no winning strategy. Notation q, k inside a circle depicts a state q with priority k. Transitions from states involving a bullet $\bullet$ indicate a probabilistic transition. In POMDPs, we always write the signals along transitions. Actions are omitted when they all induce the same transition from a given state, and probabilities equal to 1 are omitted.

$\mathcal{A}$, and playing c goes to q_2 if the current state is a final state of $\mathcal{A}$, and to q_3 otherwise. We keep a positive probability to go back to q_0 at any point to make it weakly revealing. The idea of playing n times a in a row in the example is replaced by a (possible) sequence of words that have a probability arbitrarily close to 1 to reach a final state. One can show that there is an almost-sure strategy in this POMDP if and only if $\mathcal{A}$ has value 1 w.r.t. its final states.

5 Strongly revealing POMDPs

In this section, we introduce *strongly revealing POMDPs*, a stronger property entailing that infinitely many revelations occur in a POMDP almost surely. We show that the existence of almost-sure strategies is decidable for strongly revealing POMDPs with arbitrary parity objectives.

We define a notion of *revealing signals*: for q a state of a POMDP $\mathcal{P} = (Q, \mathsf{Act}, \mathsf{Sig}, \delta, q_0)$, we define $\mathsf{Revealing}(q) = \{s \in \mathsf{Sig} \mid \forall r, r' \in Q, r' \neq q \implies \delta(r, a)(s, r') = 0\}$ to be the set of signals that indicate surely that the next state is q. For convenience, we define $\mathsf{Succ}(q, a) = \{q' \in Q \mid \exists s \in \mathsf{Sig}, \delta(q, a)(s, q') > 0\}$ and $\mathsf{Succ}(q, a, s) = \{q' \in Q \mid \delta(q, a)(s, q') > 0\}$.

Definition 2. *POMDP $\mathcal{P} = (Q, \mathsf{Act}, \mathsf{Sig}, \delta, q_0)$ is* strongly revealing *if any transition between two states for a given action in the underlying MDP of $\mathcal{P}$ can also happen with a revealing signal. Formally, $\mathcal{P}$ is strongly revealing if for all $q, q' \in Q$ and $a \in \mathsf{Act}$, if $q' \in \mathsf{Succ}(q, a)$, then there is $s \in \mathsf{Revealing}(q')$ such that $q' \in \mathsf{Succ}(q, a, s)$.*

Under this definition, the set of belief supports $B^{\mathcal{P}}_{\mathsf{sing}}$ is visited infinitely often from the initial state for any given strategy, so a strongly revealing POMDP is in particular weakly revealing. Observe that the weakly revealing POMDP from Figure 4 is not strongly revealing: for instance, $q_1' \in \mathsf{Succ}(q_1, a)$, but there is no revealing signal that could for sure reveal q_1' after q_1. The strongly revealing property can be decided in polynomial time in the size of a POMDP by simply analyzing every transition.

Example 2. *We give an example of a strongly revealing POMDP inspired from the* tiger *of (Cassandra, Kaelbling, and Littman 1994), depicted in Figure 5. This example is the one used in Figure 3 in the introduction; the code to generate this example in our tool is also provided in Appendix A.*

In the tiger environment, an agent has to open the left or the right door, with action a_{L} or a_{R}, respectively. One of them has a (deadly) tiger behind it. Fortunately, the agent can choose to wait and listen (action $a_?$) to help its decision. Listening results in a signal that is biased towards the reality, i.e., the signal can be s_{L} or s_{R} and the former is more likely if the tiger really is on the left, and vice versa.

We present our version of the tiger environment in which listening guarantees one will eventually discern behind which door there is a tiger. This is achieved by adding new revealing signals $a_{\mathsf{L}!}$ or $a_{\mathsf{R}!}$ which, importantly, can only be obtained when the tiger is on the left or on the right, respectively. To keep things interesting, these signals can only be obtained with a small probability (yet, them being there already ensures that the POMDP is strongly revealing*). We also add signals for death ($s_\bot$) and victory ($s_\top$), which are missing from the original tiger environment.*

The addition of the signals $s_{\mathsf{L}!}$ and $s_{\mathsf{R}!}$ makes the objective easier to satisfy and therefore changes the semantics of the POMDP; indeed, it is possible to gather more information than in the original tiger environment. However, revealing the deadlock states $\top$ and $\bot$ through signals $s_\top$ and $s_\bot$ is necessary to make the POMDP strongly revealing but does not make the objective easier to satisfy, as no strategy can exit these states anyway.

5.1 Decidability of parity with strong revelations

The soundness of the analysis of the belief-support MDP for strongly revealing POMDPs follows from Proposition 1; it remains to show completeness (proof in Appendix F).

Proposition 3. *Let $\mathcal{P} = (Q, \mathsf{Act}, \mathsf{Sig}, \delta, q_0)$ be a strongly revealing POMDP with a priority function p, and let $\mathcal{P}_{\mathcal{B}}$ be its belief-support MDP with priority function $p_{\mathcal{B}}$. If there is an almost-sure strategy for $\mathsf{Parity}(p)$ in $\mathcal{P}$, then there is an almost-sure strategy for $\mathsf{Parity}(p_{\mathcal{B}})$ in $\mathcal{P}_{\mathcal{B}}$.*

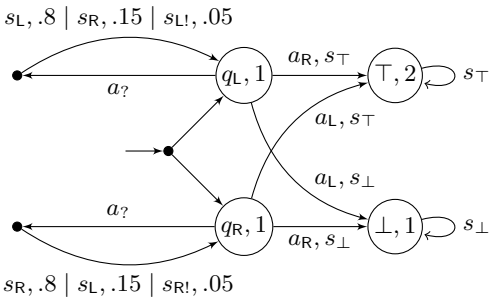

Figure 5: Strongly revealing *tiger* (Example 2).

We also show a complexity lower bound. The lower bound holds for CoBüchi in strongly revealing POMDPs; as strongly revealing POMDPs are a subclass of weakly ones, the hardness follows for weakly revealing POMDPs.

Proposition 4. *The following problem is EXPTIME-hard: given a strongly revealing POMDP with a CoBüchi objective, decide whether there is an almost-sure strategy.*

We obtain as before the decidability of the problem by reducing to the analysis of the belief-support MDP. The proof is similar to the one of Theorem 1, simply replacing the use of Proposition 2 by Proposition 3.

Theorem 3. *The existence of an almost-sure strategy for parity objectives in strongly revealing POMDPs is EXPTIME-complete.*

5.2 Undecidability of CoBüchi games with strong revelations

We study here whether the revealing semantics helps in *zero-sum games* of partial information with revealing semantics. In general, such games with CoBüchi objectives are undecidable (they encompass POMDPs) while Büchi games are decidable for almost-sure strategies (Bertrand, Genest, and Gimbert 2017). We show a negative result: the existence of an almost-sure strategy in CoBüchi *games* with partial information is undecidable, even when satisfying a natural extension of the strongly revealing property.

Two-player games of partial information are tuples $\mathcal{G} = (Q, \mathsf{Act}_1, \mathsf{Act}_2, \mathsf{Sig}, \delta, q_0)$, where Q is a finite set of states, Act_1 and Act_2 are finite sets of actions, Sig is a finite set of signals, $\delta \colon Q \times \mathsf{Act}_1 \times \mathsf{Act}_2 \to \mathcal{D}(\mathsf{Sig} \times Q)$ is the transition function, and $q_0 \in Q$ is an initial state. At each round, two players Player 1 and Player 2 respectively choose an action in Act_1 and Act_2. Histories and plays are defined as for POMDPs. For $i \in \{1, 2\}$, a strategy of Player i is a function $\sigma_i \colon (\mathsf{Act}_i \times \mathsf{Sig})^* \to \mathcal{D}(\mathsf{Act}_i)$. Given two strategies σ_1 and σ_2 for the two players, we can define as for POMDPs a probability measure $\mathbb{P}^{\mathcal{P}}_{\sigma_1,\sigma_2}[\cdot]$ on plays. A strategy σ_1 of Player 1 is *almost sure* for an objective W if for all strategies σ_2 of Player 2, $\mathbb{P}^{\mathcal{P}}_{\sigma_1,\sigma_2}[W] = 1$.

A game $\mathcal{G}$ is *strongly revealing* if for any possible transition between two states with a pair of actions, there is a chance of a signal that reveals the second state; the definition is the same as for POMDPs, assuming $\mathsf{Act} = \mathsf{Act}_1 \times \mathsf{Act}_2$.

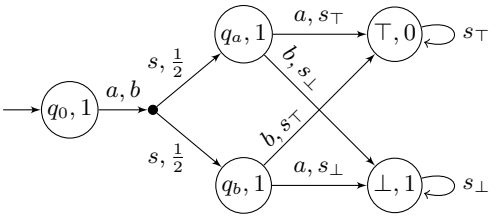

Figure 6: POMDP $\mathcal{P}$ with a CoBüchi objective such that there is an almost-sure strategy in the underlying MDP, but not in the strongly revealing $\mathcal{P}_{\mathsf{sr}}$ (Example 3).

Theorem 4. *The existence of an almost-sure strategy in strongly revealing CoBüchi games is undecidable.*

5.3 Optimistic semantics for POMDPs

In our revealing definitions, we adopted the point of view of considering *subclasses* of POMDPs. A limitation of this point of view is that our results say nothing about POMDPs which are not strongly (nor weakly) revealing. We argue that another fruitful formulation of our results concerns the class of *all* POMDPs, by defining alternative, *revealing* semantics.

Consider a POMDP $\mathcal{P}$. Let us define the extended POMDP $\mathcal{P}_{\mathsf{sr}}$ such that, at each transition, there is a small probability of revealing which state we reach after firing this action, using additional signals s_q, one for each state q of $\mathcal{P}$.

Theorem 5. *For any POMDP $\mathcal{P}$, $\mathcal{P}_{\mathsf{sr}}$ is strongly revealing. Moreover, if there is no almost-sure strategy ensuring an omega-regular objective in $\mathcal{P}_{\mathsf{sr}}$ (which is decidable by Theorem 3), then there is no almost-sure strategy ensuring the same objective in $\mathcal{P}$.*

The contrapositive is easily proved: any almost-sure strategy of $\mathcal{P}$ can be lifted to an almost-sure strategy of $\mathcal{P}_{\mathsf{sr}}$. This property justifies the term "optimistic semantics". Note that the converse implication cannot hold (as POMDPs with omega-regular objectives are undecidable).

We compare our approach with a well-known, even more optimistic semantics: revealing the exact state at each transition, which corresponds to working on the underlying MDP. The following example shows that studying the strongly revealing semantics is finer than studying the underlying MDP (i.e., it proves the non-existence of almost-sure strategies for more POMDPs).

Example 3. *Consider the POMDP $\mathcal{P}$ with a CoBüchi objective depicted in Figure 6. In this POMDP, the only way to win is to get to state $\top$; for this, it is necessary to play a from q_a or b from q_b. However, this is only possible by knowing the exact state (q_a or q_b) after one move. Therefore, there is an almost-sure strategy in the underlying MDP, but not in the strongly revealing $\mathcal{P}_{\mathsf{sr}}$.*

The fact that our approach is finer than considering the underlying MDP could already be guessed from the computational complexity of the problems: solving MDPs with parity objectives is in P (Baier and Katoen 2008), while solving strongly revealing POMDPs with parity objectives is EXPTIME-complete (Theorem 3). If presented with a POMDP with a parity objective beyond a Büchi one, we

therefore suggest to try to solve the underlying MDP in polynomial time; if there is no almost-sure strategy, then this is also the case for the original POMDP; otherwise, try to solve the strongly revealing POMDP in exponential time; if there is no almost-sure strategy, then this is also the case for the original POMDP.

6 Related works

We discuss additional references where a restriction is set to stochastic systems to make them decidable.

The closest idea to our revelations that we know of is in (Berwanger and Mathew 2017), defining a class of partial-information multi-player games with *sure* (not just almost-sure) revelations; from any point in the game, a "revelation" occurs surely within a bounded number of steps. This is a yet stronger kind of revelation mechanism under which even parity *games* are decidable.

The *decisiveness property* (Abdulla, Ben Henda, and Mayr 2007; Bertrand et al. 2020) is a useful property to decide reachability property in infinite stochastic systems (without decision-making). Decisiveness is implied by the existence of a *finite attractor*; there is such an attractor in weakly revealing POMDPs once we fix a finite-memory strategy (as in Proposition 1).

Another path to decidability and strong guarantees is to restrict strategies, such as studying "memoryless" (Vlassis, Littman, and Barber 2012) or finite-memory (Chatterjee, Chmelik, and Tracol 2016; Andriushchenko et al. 2022) strategies in POMDPs. As the strategies we obtain in our paper are memoryless strategies on the belief-support MDP, this is in the same spirit as our work. The sufficiency of belief-support-based strategies in POMDPs, which was known for almost-sure reachability (Baier, Größer, and Bertrand 2012), was also exploited to craft efficient algorithms in (Junges, Jansen, and Seshia 2021); such an approach could speed up our algorithms.

An interesting class of decidable POMDPs with parity objectives is given by the *multi-environment MDPs* (Raskin and Sankur 2014), which are POMDPs consisting of multiple copies of the same MDP, where only the transition function changes. The only partial observation comes from not knowing in which copy we play; this is therefore also a restriction on the information loss. This class is incomparable to our classes of revealing POMDPs.

In a quantitative setting, the idea of having actions with some cost (time or energy) that reveal the current state or decrease the uncertainty has appeared multiple times in the literature. Such an idea already appeared in 2011 (Bertrand and Genest 2011) for POMDPs with quantitative reachability objectives. Recently, *active-measuring POMDPs*, with a similar mechanism, have been considered in the online planning community (Bellinger et al. 2021; Krale, Simão, and Jansen 2023). Despite a different setting (online planning vs. model-checking), it carries an intuition similar to our work: precise states can be known, which helps find good strategies.

Also in online planning, the article (Liu et al. 2022) considers a subclass of POMDPs restricting information loss that make *reinforcement learning* sample efficient.

7 Perspectives

We presented classes of POMDPs for which many natural objectives become decidable, and showed that these lied close to undecidability frontiers (priorities $\{0, 1, 2\}$ vs. $\{1, 2, 3\}$, POMDPs vs. games).

Due to their intrinsic undecidability, POMDPs are not often studied through the prism of exact algorithms. We believe there is a lot to gain by understanding more closely (i) the *structural properties* of POMDPs that make them decidable for classes of objectives (such as weak/strong revelations), and (ii) the conditions that make *simple strategies* (such as belief-support-based strategies) sufficient. Our article is a new step towards these goals.

On a more specific note, an interesting step for (i) could involve framing the exact complexity of the existence of strategies for simple objectives involving beliefs and belief supports. This step includes settling precisely the complexity of deciding whether a POMDP is weakly revealing.

Acknowledgments

Pierre Vandenhove was funded by ANR project G4S (ANR-21-CE48-0010-01). This work was sparked by discussions with Guillaume Vigeral and Bruno Ziliotto, following a talk on a related model (Vigeral and Ziliotto 2022).

References

Abdulla, P. A.; Ben Henda, N.; and Mayr, R. 2007. Decisive Markov Chains. *Log. Methods Comput. Sci.*, 3(4).

Andriushchenko, R.; Ceska, M.; Junges, S.; and Katoen, J. 2022. Inductive synthesis of finite-state controllers for POMDPs. In Cussens, J.; and Zhang, K., eds., *Uncertainty in Artificial Intelligence, Proceedings of the Thirty-Eighth Conference on Uncertainty in Artificial Intelligence, UAI 2022, 1-5 August 2022, Eindhoven, The Netherlands*, volume 180 of *Proceedings of Machine Learning Research*, 85–95. PMLR.

Åström, K. J. 1965. Optimal Control of Markov Processes with Incomplete State Information I. *Journal of Mathematical Analysis and Applications*, 10: 174–205.

Baier, C.; Bertrand, N.; and Größer, M. 2008. On Decision Problems for Probabilistic Büchi Automata. In Amadio, R. M., ed., *Foundations of Software Science and Computational Structures, FoSSaCS*, volume 4962 of *Lecture Notes in Computer Science*, 287–301. Springer.

Baier, C.; Größer, M.; and Bertrand, N. 2012. Probabilistic ω-automata. *J. ACM*, 59(1): 1:1–1:52.

Baier, C.; and Katoen, J. 2008. *Principles of model checking*. MIT Press. ISBN 978-0-262-02649-9.

Bellinger, C.; Coles, R.; Crowley, M.; and Tamblyn, I. 2021. Active Measure Reinforcement Learning for Observation Cost Minimization. In Antonie, L.; and Zadeh, P. M., eds., *Canadian Conference on Artificial Intelligence*. Canadian Artificial Intelligence Association.

Bertrand, N.; Bouyer, P.; Brihaye, T.; and Fournier, P. 2020. Taming denumerable Markov decision processes with decisiveness. *CoRR*, abs/2008.10426.

Bertrand, N.; and Genest, B. 2011. Minimal Disclosure in Partially Observable Markov Decision Processes. In Chakraborty, S.; and Kumar, A., eds., *IARCS Annual Conference on Foundations of Software Technology and Theoretical Computer Science, FSTTCS 2011, December 12–14, 2011, Mumbai, India*, volume 13 of *LIPIcs*, 411–422. Schloss Dagstuhl – Leibniz-Zentrum für Informatik.

Bertrand, N.; Genest, B.; and Gimbert, H. 2017. Qualitative Determinacy and Decidability of Stochastic Games with Signals. *J. ACM*, 64(5): 33:1–33:48.

Berwanger, D.; and Mathew, A. B. 2017. Infinite games with finite knowledge gaps. *Information and Computation*, 254: 217–237.

Bloem, R.; Chatterjee, K.; and Jobstmann, B. 2018. Graph Games and Reactive Synthesis. In Clarke, E. M.; Henzinger, T. A.; Veith, H.; and Bloem, R., eds., *Handbook of Model Checking*, 921–962. Springer.

Cassandra, A. R.; Kaelbling, L. P.; and Littman, M. L. 1994. Acting Optimally in Partially Observable Stochastic Domains. In Hayes-Roth, B.; and Korf, R. E., eds., *Proceedings of the 12th National Conference on Artificial Intelligence, Seattle, WA, USA, July 31 - August 4, 1994, Volume 2*, 1023–1028. AAAI Press / The MIT Press.

Chatterjee, K.; Chmelik, M.; and Tracol, M. 2016. What is decidable about partially observable Markov decision processes with ω-regular objectives. *Journal of Computer and System Sciences*, 82(5): 878–911.

Chatterjee, K.; Doyen, L.; Gimbert, H.; and Henzinger, T. A. 2010. Randomness for Free. In Hlinený, P.; and Kucera, A., eds., *Mathematical Foundations of Computer Science 2010, 35th International Symposium, MFCS 2010, Brno, Czech Republic, August 23–27, 2010. Proceedings*, volume 6281 of *Lecture Notes in Computer Science*, 246–257. Springer.

Chatterjee, K.; Doyen, L.; and Henzinger, T. A. 2013. A survey of partial-observation stochastic parity games. *Formal Methods Syst. Des.*, 43(2): 268–284.

Chatterjee, K.; and Henzinger, T. A. 2012. A survey of stochastic ω-regular games. *Journal of Computer and System Sciences*, 78(2): 394–413.

de Alfaro, L. 1997. *Formal verification of probabilistic systems*. Ph.D. thesis, Stanford University, USA.

Fijalkow, N. 2017. Undecidability results for probabilistic automata. *ACM SIGLOG News*, 4(4): 10–17.

Giacomo, G. D.; and Vardi, M. Y. 2013. Linear Temporal Logic and Linear Dynamic Logic on Finite Traces. In Rossi, F., ed., *International Joint Conference on Artificial Intelligence*, IJCAI'13, 854–860. IJCAI/AAAI.

Gimbert, H.; and Oualhadj, Y. 2010. Probabilistic Automata on Finite Words: Decidable and Undecidable Problems. In Abramsky, S.; Gavoille, C.; Kirchner, C.; auf der Heide, F. M.; and Spirakis, P. G., eds., *International Colloquium on Automata, Languages and Programming, ICALP*, volume 6199 of *Lecture Notes in Computer Science*, 527–538. Springer.

Hauskrecht, M. 2000. Value-Function Approximations for Partially Observable Markov Decision Processes. *Journal of Artificial Intelligence Research*, 13: 33–94.

Junges, S.; Jansen, N.; and Seshia, S. A. 2021. Enforcing Almost-Sure Reachability in POMDPs. In Silva, A.; and Leino, K. R. M., eds., *Computer Aided Verification – 33rd International Conference, CAV 2021, Virtual Event, July 20-23, 2021, Proceedings, Part II*, volume 12760 of *Lecture Notes in Computer Science*, 602–625. Springer.

Kaelbling, L. P.; Littman, M. L.; and Cassandra, A. R. 1998. Planning and acting in partially observable stochastic domains. *Artificial Intelligence*, 101(1): 99–134.

Klenke, A. 2007. *Probability Theory: A Comprehensive Course*. Springer.

Krale, M.; Simão, T. D.; and Jansen, N. 2023. Act-Then-Measure: Reinforcement Learning for Partially Observable Environments with Active Measuring. In Koenig, S.; Stern, R.; and Vallati, M., eds., *Proceedings of the Thirty-Third International Conference on Automated Planning and Scheduling, Prague, Czech Republic, July 8–13, 2023*, 212–220. AAAI Press.

Liu, Q.; Chung, A.; Szepesvári, C.; and Jin, C. 2022. When Is Partially Observable Reinforcement Learning Not Scary? In Loh, P.; and Raginsky, M., eds., *Conference on Learning Theory, 2–5 July 2022, London, UK*, volume 178 of *Proceedings of Machine Learning Research*, 5175–5220. PMLR.

Lovejoy, W. S. 1991. Computationally Feasible Bounds for Partially Observed Markov Decision Processes. *Operations Research*, 39(1): 162–175.

Madani, O.; Hanks, S.; and Condon, A. 2003. On the undecidability of probabilistic planning and related stochastic optimization problems. *Artificial Intelligence*, 147(1-2): 5–34.

Pnueli, A. 1977. The temporal logic of programs. In *Symposium on Foundations of Computer Science*, SFCS'77.

Rabin, M. O. 1963. Probabilistic Automata. *Inf. Control.*, 6(3): 230–245.

Raffin, A.; Hill, A.; Gleave, A.; Kanervisto, A.; Ernestus, M.; and Dormann, N. 2021. Stable-Baselines3: Reliable Reinforcement Learning Implementations. *Journal of Machine Learning Research*, 22(268): 1–8.

Raskin, J.; and Sankur, O. 2014. Multiple-Environment Markov Decision Processes. In Raman, V.; and Suresh, S. P., eds., *34th International Conference on Foundation of Software Technology and Theoretical Computer Science, FSTTCS 2014, December 15–17, 2014, New Delhi, India*, volume 29 of *LIPIcs*, 531–543. Schloss Dagstuhl – Leibniz-Zentrum für Informatik.

Silver, D.; and Veness, J. 2010. Monte-Carlo Planning in Large POMDPs. In *Conference on Neural Information Processing Systems, NIPS*, 2164–2172. Curran Associates, Inc.

Vigeral, G.; and Ziliotto, B. 2022. Zero-sum stochastic games with intermittent observation of the state. Communication at the "Current Trends in Graph and Stochastic Games" workshop (GAMENET'22).

Vlassis, N.; Littman, M. L.; and Barber, D. 2012. On the Computational Complexity of Stochastic Controller Optimization in POMDPs. *ACM Trans. Comput. Theory*, 4(4): 12:1–12:8.

A Revealing Tiger Environment

We provide the code to generate the *strongly revealing tiger* POMDP, used in Figure 3 in the introduction, and described in Example 2 and Figure 5.

For concreteness, we provide the environment encoded in the *Cassandra* format for POMDPs. States `tiger-left`, `tiger-right`, `dead`, and `done` correspond respectively to q_{L}, q_{R}, $q_{\bot}$, and $q_{\top}$. Actions `listen`, `open-left`, and `open-right` correspond respectively to $a_{?}$, a_{L}, and a_{R}. Observations `maybe-left` and `maybe-right` correspond to signals s_{L} and s_{R}, `defo-left` and `defo-right` correspond to $s_{\mathsf{L}!}$ and $s_{\mathsf{R}!}$, and `dead-obs` and `done-obs` correspond to $s_{\bot}$ and $s_{\top}$. As a basis for comparison, the original (discounted) tiger environment is given here: https://www.pomdp.org/examples/tiger.aaai.POMDP.

```
states: tiger-left tiger-right dead done
actions: listen open-left open-right
observations: maybe-left maybe-right defo-left defo-right dead-obs done-obs
start include: tiger-left tiger-right

T:listen
identity

T:open-left
0.00 0.00 1.00 0.00
0.00 0.00 0.00 1.00
0.00 0.00 1.00 0.00
0.00 0.00 0.00 1.00

T:open-right
0.00 0.00 0.00 1.00
0.00 0.00 1.00 0.00
0.00 0.00 1.00 0.00
0.00 0.00 0.00 1.00

O:listen
0.80 0.15 0.05 0.00 0.00 0.00
0.15 0.80 0.00 0.05 0.00 0.00
0.00 0.00 0.00 0.00 1.00 0.00
0.00 0.00 0.00 0.00 0.00 1.00

O:open-left
0.00 0.00 0.00 0.00 1.00 0.00
0.00 0.00 0.00 0.00 0.00 1.00
0.00 0.00 0.00 0.00 1.00 0.00
0.00 0.00 0.00 0.00 0.00 1.00

O:open-right
0.00 0.00 0.00 0.00 0.00 1.00
0.00 0.00 0.00 0.00 1.00 0.00
0.00 0.00 0.00 0.00 1.00 0.00
0.00 0.00 0.00 0.00 0.00 1.00
```

B Additional preliminaries: End components in MDPs

In proofs, we will use extensively the notion of *end components* for various MDPs derived from POMDPs. Let $\mathcal{M} = (Q, \mathsf{Act}, \delta, q_0)$ be an MDP. For a pair $U = (R, A)$ with $R \subseteq Q$ and $A\colon R \to 2^{\mathsf{Act}}$, we define the graph *induced by* (R, A) as the graph (R, E), where E is the set of edges $(q, q') \in R \times R$ for which there is $a \in A(q)$ such that $\delta(q, a)(q') > 0$. An *end component* of $\mathcal{M}$ is a tuple $U = (R, A)$, with $R \subseteq Q$ and $A\colon R \to 2^{\mathsf{Act}}$, such that for all $q \in R$ and $a \in A(q)$, $\mathsf{supp}(\delta(q, a)) \subseteq R$, and such that the graph induced by (R, A) is strongly connected.

For $\pi \in (\mathsf{Act} \cdot Q)^\omega$ a play of $\mathcal{M}$, we write $\mathsf{inf}(\pi)$ for the pair (R, A) such that R is the set of states that π visits infinitely often and such that for $q \in R$, $A(q)$ is the set of actions played infinitely often when in q through π. The main result we need about end components is the following.

Theorem 6 (Fundamental theorem of end components (de Alfaro 1997))**.** *Let $\mathcal{M} = (Q, \mathsf{Act}, \delta, q_0)$ be an MDP.*

- *If $U = (R, A)$ is an end component of $\mathcal{M}$ and $q \in R$, there is a strategy σ such that $\mathbb{P}_\sigma^{\mathcal{M}^q}[\mathsf{inf}(\pi) = U] = 1$.*
- *For all strategies σ, $\mathbb{P}_\sigma^{\mathcal{M}}[\mathsf{inf}(\pi)$ is an end component$] = 1$.*

C Properties of the belief-support MDP (appendix to Section 3)

In this section, we provide full statements and proofs for claims in Section 3.

Observe that there is a syntactic difference in what strategies of a POMDP and strategies in the belief-support MDP can observe: strategies in the belief-support MDP can see the belief supports, but do not know through which signal they were generated (there may be multiple such signals). Strategies in a POMDP are therefore syntactically more general: they can observe the precise sequence of signals, and thereby compute the belief supports visited. Despite this difference between the power of strategies in both models, both models behave similarly for multiple simple objectives.

First, there is a correspondence between the sequences of belief supports reached positively in both models.

Lemma 2. *Let $\mathcal{P} = (Q, \mathsf{Act}, \mathsf{Sig}, \delta, q_0)$ be a POMDP and $\mathcal{P}_\mathcal{B}$ be its belief-support MDP.*

- *For all strategies $\sigma \in \Sigma(\mathcal{P})$, there is a strategy $\sigma_\mathcal{B} \in \Sigma(\mathcal{P}_\mathcal{B})$ such that, for all sequences of belief supports $B \in (2^Q_\emptyset)^*$,*
$$\mathbb{P}_\sigma^{\mathcal{P}}[\mathsf{Cyl}(B)] > 0 \iff \mathbb{P}_{\sigma_\mathcal{B}}^{\mathcal{P}_\mathcal{B}}[\mathsf{Cyl}(B)] > 0.$$
- *For all strategies $\sigma_\mathcal{B} \in \Sigma(\mathcal{P}_\mathcal{B})$, for all sequences of belief supports $B \in (2^Q_\emptyset)^*$,*
$$\mathbb{P}_{\widehat{\sigma_\mathcal{B}}}^{\mathcal{P}}[\mathsf{Cyl}(B)] > 0 \iff \mathbb{P}_{\sigma_\mathcal{B}}^{\mathcal{P}_\mathcal{B}}[\mathsf{Cyl}(B)] > 0.$$

Proof. We write $\mathcal{P}_\mathcal{B} = (2^Q_\emptyset, \mathsf{Act}, \delta_\mathcal{B}, \{q_0\})$.

We show the first claim. Let $\sigma \in \Sigma(\mathcal{P})$. We want to define a strategy $\sigma_\mathcal{B} \in \Sigma(\mathcal{P}_\mathcal{B})$ that imitates what σ does, but such a strategy cannot observe the signals. Let $h_\mathcal{B} = a_1 b_1 \dots a_n b_n$ be a history of $\mathcal{P}_\mathcal{B}$. We define $S_{h_\mathcal{B}}$ as the set $\{h \in (\mathsf{Act} \cdot \mathsf{Sig})^* \mid h \text{ is consistent with } \sigma \text{ and } B_h = h_\mathcal{B}\}$.

We define $\sigma_\mathcal{B}$ inductively on the length of histories, and show at the same time that $S_{h_\mathcal{B}}$ is non-empty for all histories $h_\mathcal{B}$ consistent with $\sigma_\mathcal{B}$. After 0 step, there is a single possible empty history $h_\mathcal{B}$ consistent with $\sigma_\mathcal{B}$. The set $S_{h_\mathcal{B}}$ contains only the empty history and is indeed non-empty.

Now, let $h_\mathcal{B} = a_1 b_1 \dots a_n b_n$ be a history of length n consistent with $\sigma_\mathcal{B}$. By induction hypothesis, $S_{h_\mathcal{B}}$ is non-empty. We use this fact to define $\sigma_\mathcal{B}(h_\mathcal{B})$. Let $A_{h_\mathcal{B}} = \{a \in \mathsf{Act} \mid \exists h \in S_{h_\mathcal{B}}, \sigma(h)(a) > 0\}$, which is also non-empty. We define $\sigma_\mathcal{B}(h_\mathcal{B})$ to randomize over all actions in $A_{h_\mathcal{B}}$. Let $a_{n+1} \in A_{h_\mathcal{B}}$ and b_{n+1} be such that $\delta_\mathcal{B}(b_n, a_{n+1})(b_{n+1}) > 0$. By induction hypothesis, there is $h \in S_{h_\mathcal{B}}$ of length n consistent with σ such that $\sigma(h)(a_{n+1}) > 0$. Hence, by construction of the belief-support MDP, $S_{h a_{n+1} b_{n+1}}$ is non-empty.

Let $B = b_1 \dots b_n \in (2^Q_\emptyset)^*$ be a sequence of belief supports. We now show that

$$\mathbb{P}_\sigma^{\mathcal{P}}[\mathsf{Cyl}(B)] > 0 \iff \mathbb{P}_{\sigma_\mathcal{B}}^{\mathcal{P}_\mathcal{B}}[\mathsf{Cyl}(B)] > 0.$$

If $\mathbb{P}_{\sigma_\mathcal{B}}^{\mathcal{P}_\mathcal{B}}[\mathsf{Cyl}(B)] > 0$, then there is $h_\mathcal{B}$ consistent with $\sigma_\mathcal{B}$ such that $h_\mathcal{B} = a_1 b_1 \dots a_n b_n$ for some $a_1, \dots, a_n \in \mathsf{Act}$. As $S_{h_\mathcal{B}}$ is non-empty, we also have $\mathbb{P}_\sigma^{\mathcal{P}}[\mathsf{Cyl}(B)] > 0$. If $\mathbb{P}_\sigma^{\mathcal{P}}[\mathsf{Cyl}(B)] > 0$, then there is $h \in S_{h_\mathcal{B}}$ consistent with σ such that $B_h = a_1 b_1 \dots a_n b_n$ for some $a_1, \dots, a_n \in \mathsf{Act}$. One can show by induction on the length of h that B_h is also consistent with $\sigma_\mathcal{B}$, so $\mathbb{P}_{\sigma_\mathcal{B}}^{\mathcal{P}_\mathcal{B}}[\mathsf{Cyl}(B)] > 0$.

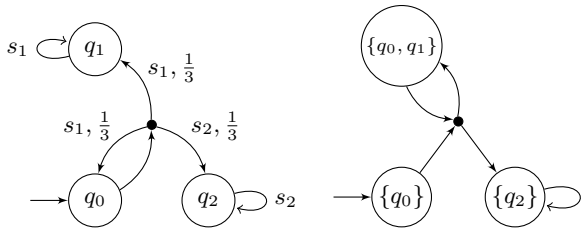

Figure 7: POMDP with a single action (left) and its belief-support MDP (right) such that the belief support $\{q_2\}$ is reached almost surely in the belief-support MDP, but not in the POMDP.

We now show the second claim. Let $\sigma_{\mathcal{B}} \in \Sigma(\mathcal{P}_{\mathcal{B}})$; consider $\widehat{\sigma_{\mathcal{B}}} \in \Sigma(\mathcal{P})$ (which was defined in Section 3). Clearly, the tree of belief supports induced by both strategies contains the same nodes (although they are attained with possibly different probabilities). Again, this can be shown with a similar (and easier) induction on the length of histories. Hence, both strategies see the same cylinders of belief supports with positive probability. □

Second, sets of belief supports reached almost surely in the POMDP can also be reached almost surely in the belief-support MDP.

Lemma 3. *Let $\mathcal{P} = (Q, \mathsf{Act}, \mathsf{Sig}, \delta, q_0)$ be a POMDP and $\mathcal{P}_{\mathcal{B}}$ be its belief-support MDP. Let $B \subseteq 2^Q_{\emptyset}$ be a set of belief supports.*

- *If there is $\sigma \in \Sigma(\mathcal{P})$ such that $\mathbb{P}^{\mathcal{P}}_{\sigma}[\mathsf{Reach}(B)] = 1$, then there is $\sigma_{\mathcal{B}} \in \Sigma(\mathcal{P}_{\mathcal{B}})$ such that $\mathbb{P}^{\mathcal{P}_{\mathcal{B}}}_{\sigma_{\mathcal{B}}}[\mathsf{Reach}(B)] = 1$.*
- *If for all $\sigma \in \Sigma(\mathcal{P})$ it holds that $\mathbb{P}^{\mathcal{P}}_{\sigma}[\mathsf{Reach}(B)] = 1$, then for all $\sigma_{\mathcal{B}} \in \Sigma(\mathcal{P}_{\mathcal{B}})$ we have $\mathbb{P}^{\mathcal{P}_{\mathcal{B}}}_{\sigma_{\mathcal{B}}}[\mathsf{Reach}(B)] = 1$.*

The converses of both claims of Lemma 3 are false in general; see Figure 7. The proof uses the notion of *end components*, which is discussed in Appendix B.

Proof. We first prove the first item. Let $\sigma \in \Sigma(\mathcal{P})$ such that $\mathbb{P}^{\mathcal{P}}_{\sigma}[\mathsf{Reach}(B)] = 1$. Without loss of generality, we modify σ and $\mathcal{P}$ such that σ moves to a sink state $\perp$ with a new action $a_{\perp}$ and a new signal $s_{\perp}$ whenever a belief support in B is seen. Let $B_{\sigma} = \{b \in 2^Q_{\emptyset} \mid b \neq \{\perp\} \text{ and } \mathbb{P}^{\mathcal{P}}_{\sigma}[\mathsf{Reach}(b)] > 0\}$. Any belief support $b \in B_{\sigma}$ is such that there is an almost-sure strategy σ^b from b for $\mathsf{Reach}(B)$ that only visits belief supports in B_{σ} (simply take σ^b to be a continuation of σ after b is visited). We build a strategy $\sigma_{\mathcal{B}} \in \Sigma(\mathcal{P}_{\mathcal{B}})$ such that $\mathbb{P}^{\mathcal{P}_{\mathcal{B}}}_{\sigma_{\mathcal{B}}}[\mathsf{Reach}(B)] = 1$.

By Lemma 2, for all $b \in B_{\sigma}$, there is a strategy $\sigma^b_{\mathcal{B}} \in \Sigma(\mathcal{P}_{\mathcal{B}})$ that reaches with a positive probability exactly the same belief supports as σ^b. In particular, $\mathbb{P}^{\mathcal{P}^b_{\mathcal{B}}}_{\sigma^b_{\mathcal{B}}}[\mathsf{Reach}(b')] > 0$ implies that $b' \in B_{\sigma}$. Moreover, $\mathbb{P}^{\mathcal{P}^b_{\mathcal{B}}}_{\sigma^b_{\mathcal{B}}}[\mathsf{Reach}(B)] > 0$. By the continuity of probabilities and the finiteness of B_{σ}, there is a uniform $k \in \mathbb{N}$ such that $\mathbb{P}^{\mathcal{P}^b_{\mathcal{B}}}_{\sigma^b_{\mathcal{B}}}\left[\mathsf{Reach}^{\leq k}(B)\right] > 0$ for all b. Moreover, as B_{σ} is finite, there is $\alpha > 0$ such that $\mathbb{P}^{\mathcal{P}^b_{\mathcal{B}}}_{\sigma^b_{\mathcal{B}}}\left[\mathsf{Reach}^{\leq k}(B)\right] \geq \alpha$ for all $b \in B_{\sigma}$.

We define the strategy $\sigma_{\mathcal{B}}$ as follows: we start by following $\sigma^{\{q_0\}}_{\mathcal{B}}$ for k steps, which gives a probability $\geq \alpha$ to reach B. If B is not reached within k steps, we know that the current belief support b is still in B_{σ}. We then move on to the strategy $\sigma^b_{\mathcal{B}}$ for k steps and iterate this process. Strategy $\sigma_{\mathcal{B}}$ has a lower-bounded probability to reach B infinitely many times, and therefore reaches B almost surely.

We now show the contrapositive of the second item. Assume that there exists $\sigma_{\mathcal{B}} \in \Sigma(\mathcal{P}_{\mathcal{B}})$ such that $\mathbb{P}^{\mathcal{P}_{\mathcal{B}}}_{\sigma_{\mathcal{B}}}[\mathsf{Reach}(B)] < 1$. By Theorem 6, it implies that there is an end component $U = (R, A)$ of $\mathcal{P}_{\mathcal{B}}$ such that $R \cap B = \emptyset$ and $\mathbb{P}^{\mathcal{P}_{\mathcal{B}}}_{\sigma_{\mathcal{B}}}[\pi \notin \mathsf{Reach}(B) \text{ and } \mathsf{inf}(\pi) = U] > 0$. By Lemma 2, the strategy $\widehat{\sigma_{\mathcal{B}}}$ therefore also has a non-zero probability to reach a belief support in U without reaching B and to then never visit a belief outside of U. Hence, $\mathbb{P}^{\mathcal{P}}_{\widehat{\sigma_{\mathcal{B}}}}[\mathsf{Reach}(B)] < 1$. □

D Probabilistic bounds on reachability objectives

We prove technical lemmas about reachability properties that we will use multiple times in the paper: roughly, if a set of states (or a set of belief supports) is reached with a positive probability by *all* strategies, we can upper bound the number of steps before reaching it and lower bound the probability to reach it.

We will use the result that pure strategies suffice in POMDPs for all objectives, both for almost-sure strategies and for attaining values, which we recall from (Chatterjee et al. 2010).

Theorem 7 (Chatterjee et al. (2010, Theorem 5))**.** *Let $\mathcal{P} = (Q, \mathsf{Act}, \mathsf{Sig}, \delta, q_0)$ be a POMDP and $W \subseteq Q^\omega$ be an objective. We have that $\sup_{\sigma\in\Sigma(\mathcal{P})} \mathbb{P}^{\mathcal{P}}_\sigma[W] = \sup_{\sigma\in\Sigma_{\mathsf{P}}(\mathcal{P})} \mathbb{P}^{\mathcal{P}}_\sigma[W]$. Moreover, if there exists $\sigma \in \Sigma(\mathcal{P})$ such that $\mathbb{P}^{\mathcal{P}}_\sigma[W] = \sup_{\sigma\in\Sigma(\mathcal{P})} \mathbb{P}^{\mathcal{P}}_\sigma[W]$, then there exists $\sigma' \in \Sigma_{\mathsf{P}}(\mathcal{P})$ such that $\mathbb{P}^{\mathcal{P}}_{\sigma'}[W] = \sup_{\sigma\in\Sigma(\mathcal{P})} \mathbb{P}^{\mathcal{P}}_\sigma[W]$.*

Before proving Lemma 5 about POMDPs, we prove a similar (and easier) result about MDPs.

Lemma 4. *Let $\mathcal{M} = (Q, \mathsf{Act}, \delta, q_0)$ be an MDP and $F \subseteq Q$. Assume that for all $\sigma \in \Sigma(\mathcal{M})$, $\mathbb{P}^{\mathcal{M}}_\sigma[\mathsf{Reach}(F)] > 0$. Then, for all $\sigma \in \Sigma(\mathcal{M})$, $\mathbb{P}^{\mathcal{M}}_\sigma\left[\mathsf{Reach}^{\leq|Q|}(F)\right] \geq \beta^{|Q|}_{\mathcal{M}}$.*

Proof. We consider a deterministic variant of the MDP $\mathcal{M}$ where an antagonistic opponent resolves the stochastic transitions, by choosing among any transition that had a non-zero probability in the original MDP. Formally, we consider the two-player deterministic game $\mathcal{G} = (Q, Q \times \mathsf{Act}, E, q_0)$ where the set of edges $E = \{(q, (q, a)) \mid q \in Q, a \in \mathsf{Act}\} \cup \{((q, a), q') \mid \delta(q, a)(q') > 0\}$: Player 1 chooses the action in states of Q, and Player 2 chooses the transitions in states of $Q \times \mathsf{Act}$.

Player 1 has no winning strategy for $\mathsf{Safety}(F)$ in this deterministic zero-sum game. We show the contrapositive: assume that Player 1 has a winning strategy for $\mathsf{Safety}(F)$ in $\mathcal{G}$. Then, this strategy would be surely winning in $\mathcal{G}$, and thus with probability 1, in $\mathcal{M}$ for $\mathsf{Safety}(F)$. So there exists a strategy $\sigma \in \Sigma(\mathcal{M})$ such that $\mathbb{P}^{\mathcal{M}}_\sigma[\mathsf{Reach}(F)] = 0$. This contradicts the hypothesis.

Therefore, Player 1 has no winning strategy for $\mathsf{Safety}(F)$ in $\mathcal{G}$. As reachability games are determined, this means that Player 2 has a strategy that ensures $\mathsf{Reach}(F)$. Classical results on reachability games (using the *attractor decomposition* (Bloem, Chatterjee, and Jobstmann 2018)) imply that Player 2 even has such a strategy ensuring that no state is visited twice along the path before reaching a state in F. Hence, Player 2 has a strategy ensuring that Player 1 plays at most $|Q|$ times before visiting F.

This means that, in the original MDP, no matter the strategy, there is a path from q_0 to F of length at most $|Q|$. For pure strategies, there are at most $|Q|$ probabilistic transitions along this path, so such a path has probability at least $\beta^{|Q|}_{\mathcal{M}}$. As pure strategies suffice for reachability objectives in MDPs, this result extends to all strategies. □

We can show such a result for POMDPs, both for sets of states and sets of belief supports. To make sense of this statement, we extend reachability objectives to deal with (sets of) belief supports. For $B \subseteq 2^Q_\emptyset$, let $S_B \subseteq \mathsf{Sig}^*$ be the set of observable histories $a_1 s_1 \dots a_n s_n$ such that $\mathcal{B}^*(\{q_0\}, a_1 s_1 \dots a_n s_n) \in B$. We define $\mathsf{Reach}(B)$ (resp. $\mathsf{Büchi}(B)$) to be the set of plays inducing a belief support in B at least once (resp. infinitely often). For $b \in 2^Q_\emptyset$, we write $\mathsf{Reach}(b)$ and $\mathsf{Büchi}(b)$ for $\mathsf{Reach}(\{b\})$ and $\mathsf{Büchi}(\{b\})$. Note that the proof of the following Lemma 5 uses properties of the belief-support MDP $\mathcal{P}_{\mathcal{B}}$ proved in Appendix C.

Lemma 5. *Let $\mathcal{P} = (Q, \mathsf{Act}, \mathsf{Sig}, \delta, q_0)$ be a POMDP.*

- *Let $B \subseteq 2^Q_\emptyset$ be a set of belief supports. If for all $\sigma \in \Sigma(\mathcal{P})$, $\mathbb{P}^{\mathcal{P}}_\sigma[\mathsf{Reach}(B)] > 0$, then for all $\sigma \in \Sigma(\mathcal{P})$, $\mathbb{P}^{\mathcal{P}}_\sigma\left[\mathsf{Reach}^{\leq 2^{|Q|}-1}(B)\right] \geq \beta^{2^{|Q|}-1}_{\mathcal{P}}$.*
- *Let $F \subseteq Q$ be a set of states. If for all $\sigma \in \Sigma(\mathcal{P})$, $\mathbb{P}^{\mathcal{P}}_\sigma[\mathsf{Reach}(F)] > 0$, then for all $\sigma \in \Sigma(\mathcal{P})$, $\mathbb{P}^{\mathcal{P}}_\sigma\left[\mathsf{Reach}^{\leq 2^{|Q|}-1}(F)\right] \geq \beta^{2^{|Q|}-1}_{\mathcal{P}}$.*

Proof. We first prove the claim about belief supports. Let $B \subseteq 2^Q_\emptyset$. Assume that for all $\sigma \in \Sigma(\mathcal{P})$, $\mathbb{P}^{\mathcal{P}}_\sigma[\mathsf{Reach}(B)] > 0$. Then, by Lemma 2, for all $\sigma_{\mathcal{B}} \in \Sigma(\mathcal{P}_{\mathcal{B}})$, $\mathbb{P}^{\mathcal{P}_{\mathcal{B}}}_{\sigma_{\mathcal{B}}}[\mathsf{Reach}(B)] > 0$. As we now work with an MDP with $|2^Q_\emptyset| = 2^{|Q|} - 1$ states, we can use Lemma 4 and obtain that for all $\sigma_{\mathcal{B}} \in \Sigma(\mathcal{P}_{\mathcal{B}})$, $\mathbb{P}^{\mathcal{P}_{\mathcal{B}}}_{\sigma_{\mathcal{B}}}\left[\mathsf{Reach}^{\leq 2^{|Q|}-1}(B)\right] > 0$. Going back to $\mathcal{P}$ with Lemma 2, we have that for all $\sigma \in \Sigma(\mathcal{P})$, $\mathbb{P}^{\mathcal{P}}_\sigma\left[\mathsf{Reach}^{\leq 2^{|Q|}-1}(B)\right] > 0$. Hence, for all strategies, there is a path of length at most $2^{|Q|} - 1$ reaching B. For pure strategies, such a path has probability at least $\beta^{2^{|Q|}-1}_{\mathcal{P}}$ to happen, as there are at most $2^{|Q|} - 1$ random transitions along this path. Hence, for all *pure* strategies $\sigma \in \Sigma_{\mathsf{P}}(\mathcal{P})$,

$\mathbb{P}^{\mathcal{P}}_{\sigma}\left[\mathsf{Reach}^{\leq 2^{|Q|}-1}(B)\right] \geq \beta_{\mathcal{P}}^{2^{|Q|}-1}$. By Theorem 7, we deduce that for all (even non-pure) strategies $\sigma \in \Sigma(\mathcal{P})$, $\mathbb{P}^{\mathcal{P}}_{\sigma}\left[\mathsf{Reach}^{\leq 2^{|Q|}-1}(B)\right] \geq \beta_{\mathcal{P}}^{2^{|Q|}-1}$.

For the second claim, let $F \subseteq Q$ be a set of states. Assume that for all $\sigma \in \Sigma(\mathcal{P})$, $\mathbb{P}^{\mathcal{P}}_{\sigma}[\mathsf{Reach}(F)] > 0$. Let $B_F = \{b \in 2^Q_{\emptyset} \mid b \cap F \neq \emptyset\}$ be the set of belief supports containing a state in F. As the event "visiting F" implies the event "visiting a belief support in B_F", we also have that for all $\sigma \in \Sigma(\mathcal{P})$, $\mathbb{P}^{\mathcal{P}}_{\sigma}[\mathsf{Reach}(B_F)] > 0$. By the first property, for all strategies $\sigma \in \Sigma(\mathcal{P})$, $\mathbb{P}^{\mathcal{P}}_{\sigma}\left[\mathsf{Reach}^{\leq 2^{|Q|}-1}(B_F)\right] > 0$. This means once again that for all strategies, there is a path of length at most $2^{|Q|}-1$ that reaches B_F, so that reaches F. We conclude in the same way as the first property: for all *pure* strategies $\sigma \in \Sigma_{\mathsf{P}}(\mathcal{P})$, $\mathbb{P}^{\mathcal{P}}_{\sigma}\left[\mathsf{Reach}^{\leq 2^{|Q|}-1}(F)\right] \geq \beta_{\mathcal{P}}^{2^{|Q|}-1}$; so, by Theorem 7, for all strategies $\sigma \in \Sigma(\mathcal{P})$, $\mathbb{P}^{\mathcal{P}}_{\sigma}\left[\mathsf{Reach}^{\leq 2^{|Q|}-1}(F)\right] \geq \beta_{\mathcal{P}}^{2^{|Q|}-1}$. □

As a corollary, we obtain that if a safety objective has value 1, then there is actually an almost-sure strategy for the safety objective. This contrasts with reachability objectives, for which there are POMDPs that have value 1 but no almost-sure strategy. For reachability, the computational complexity of the two problems are widely different: the existence of an almost-sure strategy is EXPTIME-complete (Baier, Bertrand, and Größer 2008), but deciding value 1 is undecidable (Gimbert and Oualhadj 2010).

Corollary 1. *Let $\mathcal{P} = (Q, \mathsf{Act}, \mathsf{Sig}, \delta, q_0)$ be a POMDP and $F \subseteq Q$ be a set of states. We have that*

$$\exists \sigma \in \Sigma(\mathcal{P}), \mathbb{P}^{\mathcal{P}}_{\sigma}[\mathsf{Safety}(F)] = 1 \iff \sup_{\sigma \in \Sigma(\mathcal{P})} \mathbb{P}^{\mathcal{P}}_{\sigma}[\mathsf{Safety}(F)] = 1.$$

Proof. The implication $\Longrightarrow$ is trivial; we focus on the other implication. We prove the contrapositive. Assume that there is no almost-sure strategy, i.e., that for all strategies σ, we have $\mathbb{P}^{\mathcal{P}}_{\sigma}[\mathsf{Safety}(F)] < 1$. In other words, this means that for all strategies $\sigma \in \Sigma(\mathcal{P})$, we have $\mathbb{P}^{\mathcal{P}}_{\sigma}[\mathsf{Reach}(F)] > 0$. By Lemma 5, there is thus $\alpha > 0$ such that for all strategies $\sigma \in \Sigma(\mathcal{P})$, we have $\mathbb{P}^{\mathcal{P}}_{\sigma}\left[\mathsf{Reach}^{\leq 2^{|Q|}-1}(F)\right] \geq \alpha$. As $\mathbb{P}^{\mathcal{P}}_{\sigma}[\mathsf{Reach}(F)] \geq \mathbb{P}^{\mathcal{P}}_{\sigma}\left[\mathsf{Reach}^{\leq 2^{|Q|}-1}(F)\right]$, we have that for all strategies $\sigma \in \Sigma(\mathcal{P})$, $\mathbb{P}^{\mathcal{P}}_{\sigma}[\mathsf{Reach}(F)] \geq \alpha$. Hence, for all strategies $\sigma \in \Sigma(\mathcal{P})$, $\mathbb{P}^{\mathcal{P}}_{\sigma}[\mathsf{Safety}(F)] \leq 1 - \alpha$, so $\sup_{\sigma \in \Sigma(\mathcal{P})} \mathbb{P}^{\mathcal{P}}_{\sigma}[\mathsf{Safety}(F)] < 1$. □

E Additional details for Section 4

This section is devoted to the proofs of statements about weakly revealing POMDPs from Section 4. We restate and prove the soundness of the analysis of the belief-support MDP.

Proposition 1. *Let $\mathcal{P} = (Q, \mathsf{Act}, \mathsf{Sig}, \delta, q_0)$ be a weakly revealing POMDP with priority function p, and let $\mathcal{P}_{\mathcal{B}}$ be its belief-support MDP with priority function $p_{\mathcal{B}}$. Assume there is an almost-sure strategy $\sigma_{\mathcal{B}}$ for $\mathsf{Parity}(p_{\mathcal{B}})$ in $\mathcal{P}_{\mathcal{B}}$; by (Chatterjee and Henzinger 2012), we may assume $\sigma_{\mathcal{B}}$ to be pure and memoryless. Then, $\widehat{\sigma_{\mathcal{B}}}$ is an almost-sure strategy for $\mathsf{Parity}(p)$ in $\mathcal{P}$.*

Proof. As $\sigma_{\mathcal{B}}$ is pure and memoryless, we assume it is a function $2^Q_{\emptyset} \to \mathsf{Act}$. Consider strategy $\widehat{\sigma_{\mathcal{B}}}$: it is a pure strategy in $\mathcal{P}$ with exponentially many memory states, as it simply looks at the current belief support and plays as $\sigma_{\mathcal{B}}$ would. By Lemma 2, $\widehat{\sigma_{\mathcal{B}}}$ and $\sigma_{\mathcal{B}}$ reach the same belief supports with a positive probability.

We first prove a correspondence between the sets of belief supports reached almost surely by $\widehat{\sigma_{\mathcal{B}}}$ and $\sigma_{\mathcal{B}}$, in the other direction than the one of Lemma 3. Using that $\mathcal{P}$ is weakly revealing and that the strategies only use finite memory, we show that for all $B \subseteq 2^Q_{\emptyset}$,

$$\mathbb{P}^{\mathcal{P}_{\mathcal{B}}}_{\sigma_{\mathcal{B}}}[\mathsf{Büchi}(B)] = 1 \Longrightarrow \mathbb{P}^{\mathcal{P}}_{\widehat{\sigma_{\mathcal{B}}}}[\mathsf{Büchi}(B)] = 1. \tag{1}$$

Let $B \subseteq 2^Q_{\emptyset}$ be such that $\mathbb{P}^{\mathcal{P}_{\mathcal{B}}}_{\sigma_{\mathcal{B}}}[\mathsf{Büchi}(B)] = 1$. Let $Q_\infty = \{q \in Q \mid \mathbb{P}^{\mathcal{P}}_{\widehat{\sigma_{\mathcal{B}}}}[\mathsf{Büchi}(\{q\})] > 0\}$. Due to $\mathcal{P}$ being weakly revealing, Q_∞ is non-empty and $\mathbb{P}^{\mathcal{P}}_{\widehat{\sigma_{\mathcal{B}}}}[\mathsf{Büchi}(Q_\infty)] = 1$. Every $\{q\}$ with $q \in Q_\infty$ is reached positively by $\sigma_{\mathcal{B}}$ (Lemma 2), so $\widehat{\sigma_{\mathcal{B}}}$ still reaches an element of B with non-zero probability from these singleton beliefs. Note that whenever a $\{q\}$ with $q \in Q_\infty$ is reached by $\widehat{\sigma_{\mathcal{B}}}$, the probability to reach B afterwards is independent from the past: this is due to belief support $\{q\}$ describing precisely the current belief (the probability to be in q is 1) and $\widehat{\sigma_{\mathcal{B}}}$

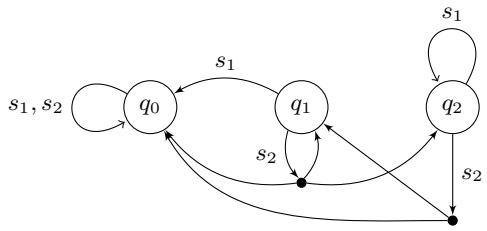

Figure 8: The POMDP from Example 4 (with $n = 2$): a partially observable Markov chain which is weakly revealing, and such that a revelation takes at best exponentially many steps from the initial state q_{init}. The initial state q_{init} is not represented, it simply moves to any other state in a non-observable way. Precise probabilities are omitted. One transition of each type appears: when the signal has a lower index than the state (e.g., s_1 from q_2), the probability mass stays in the state; when it matches the state (e.g., s_2 from q_2), the probability mass is spread over the smaller states; when it is greater (e.g., s_2 from q_1), the probability mass is spread over all states. The shortest revealing sequence from q_{init} is $\{q_{\mathsf{init}}\} \to \{q_0, q_1, q_2\} \xrightarrow{s_1} \{q_0, q_2\} \xrightarrow{s_2} \{q_0, q_1\} \xrightarrow{s_1} \{q_0\}$.

being memoryless w.r.t. belief supports. Hence, infinitely many visits to singleton belief supports $\{q\}$ with $q \in Q_\infty$ are done by $\widehat{\sigma_{\mathcal{B}}}$, and the probability to reach B afterwards is lower-bounded by a uniform value. We conclude that $\mathbb{P}^{\mathcal{P}}_{\sigma}[\mathsf{Büchi}(B)] = 1$.

We now consider the set of end components $\mathsf{EC}_{\sigma_{\mathcal{B}}} = \{U = (R_U, A_U) \mid \mathbb{P}^{\mathcal{P}_{\mathcal{B}}}_{\sigma_{\mathcal{B}}}[\inf(\pi) = U] > 0\}$ that strategy $\sigma_{\mathcal{B}}$ can exactly end up in in $\mathcal{P}_{\mathcal{B}}$. Set $\mathsf{EC}_{\sigma_{\mathcal{B}}}$ is non-empty by Theorem 6. Each such end component $U \in \mathsf{EC}_{\sigma_{\mathcal{B}}}$

- contains a belief singleton $\{q_U\}$ for some $q_U \in Q$: otherwise, using that $\mathcal{P}$ is weakly revealing, this contradicts Lemma 3;
- has an even maximal priority w.r.t. $p_{\mathcal{B}}$, due to $\sigma_{\mathcal{B}}$ being almost sure for $\mathsf{Parity}(p_{\mathcal{B}})$ and U being an end component of $\sigma_{\mathcal{B}}$ with positive probability.

The second property implies that each end component $U \in \mathsf{EC}_{\sigma_{\mathcal{B}}}$ is such that $\max\{p(q) \mid q \in b \in R_U\}$ is even. Let $q_U^{\max}$ be a state in some belief support of U achieving this maximal priority in U.

Let us consider the effect of the strategy $\widehat{\sigma_{\mathcal{B}}}$ in $\mathcal{P}$. First, observe that this strategy almost surely reaches a belief support $\{q\}$ for some $U \in \mathsf{EC}_{\sigma_{\mathcal{B}}}$ and $q \in R_U$, which follows from (1).

We fix an end component $U = (R_U, A_U) \in \mathsf{EC}_{\sigma_{\mathcal{B}}}$. Assume $\widehat{\sigma_{\mathcal{B}}}$ has reached a belief support $\{q\}$ with $q \in R_U$. Using again (1), we know that such singleton belief supports are almost surely visited infinitely often. Following strategy $\widehat{\sigma_{\mathcal{B}}}$, by definition of the belief-support MDP, there is a non-zero probability to reach $q_U^{\max}$ from every singleton belief support $\{q\}$ with $q \in R_U$. Due to $\widehat{\sigma_{\mathcal{B}}}$ being memoryless w.r.t. belief supports, this probability is the same after each visit to such a $\{q\}$. Hence, state $q_U^{\max}$ has infinitely often a lower-bounded probability to be visited. As these probabilities are independent, $q_U^{\max}$ is visited infinitely often. To conclude, observe that once a belief support in U is reached, $\widehat{\sigma_{\mathcal{B}}}$ only visits states in the belief supports of U and visits $q_U^{\max}$ infinitely often. The strategy $\widehat{\sigma_{\mathcal{B}}}$ is therefore almost sure for $\mathsf{Parity}(p)$ in $\mathcal{P}$. □

We now show an example illustrating that revelations may take exponentially many steps to occur with positive probability in weakly revealing POMDPs. More generally, this example also shows that bounds for belief supports in Lemma 5 are tight, even for weakly revealing POMDPs.

Example 4. *Let $\mathcal{P} = (Q, \mathsf{Act}, \mathsf{Sig}, \delta, q_{\mathsf{init}})$ be a POMDP with $Q = \{q_{\mathsf{init}}, q_0, \ldots, q_n\}$, $\mathsf{Act} = \{a\}$, and $\mathsf{Sig} = \{s_1, \ldots, s_n\}$. We depict the construction for $n = 2$ in Figure 8. As $|\mathsf{Act}| = 1$, it is actually a* partially observable Markov chain*. The transition function is defined as follows:*

- $\delta(q_{\mathsf{init}}, a)(s, q_i) > 0$ *for all* $s \in \mathsf{Sig}$, $i \in \{0, \ldots, n\}$,
- $\delta(q_i, a)(s_j, q_k) > 0$ *if and only if* $(i = k = 0)$ *or* $(i > j$ *and* $i = k)$ *or* $(i = j$ *and* $i > k)$ *or* $(i < j)$.

Here is how a play happens: first, the probability mass is spread from q_{init} to all the other states. The state q_0 is absorbing: there will always be some probability mass in state q_0, so a revelation can only happen in this state. Here is how the other states with index $i \in \{1, \ldots, n\}$ behave:

- *if a signal s_j with index $j < i$ is seen, then any probability mass already in q_i simply remains in q_i;*
- *if the signal s_i is seen, then any probability mass in q_i is spread out over the states q_k with $k < i$.*
- *if a signal s_j with index $j > i$ is seen, then any probability mass in q_i is spread out over all states in $\{q_0, \ldots, q_n\}$.*

To reason on this POMDP, we associate every belief support with a number that we read in binary on the belief support: we assume that a belief support $b \in 2^Q_\emptyset$ is associated with number $f(b) = \sum_{q_i \in b, i \geq 1} 2^{i-1}$ (in particular, $f(\{q_0\}) = 0$). The initial belief (after one step) is number $2^n - 1 = f(\{q_0, \dots, q_n\})$. Let us study how the number $f(b)$ evolves given the signals seen:

- *for $j = \min\{i \geq 1 \mid q_i \in b\}$, we have $f(\mathcal{B}(b, as_j)) = f(b) - 1$: indeed, state q_j in the belief support becomes $\{q_0, \dots, q_{j-1}\}$, while other states remain the same;*
- *for $j < \min\{i \geq 1 \mid q_i \in b\}$, we have $f(\mathcal{B}(b, as_j)) = f(b)$: the belief support remains exactly the same;*
- *for $j > \min\{i \geq 1 \mid q_i \in b\}$, we have $f(\mathcal{B}(b, as_j)) = 2^n - 1$, as a state q_i with $i < j$ is such that $\mathcal{B}(\{q_i\}, as_j) = \{q_0, \dots, q_n\}$.*

Hence, the shortest path towards a revelation from belief support $\{q_0, \dots, q_n\}$ has length $2^n - 1$.

As mentioned in Remark 2, one can show an exponential lower bound on the memory that strategies need to play almost surely in weakly revealing POMDPs. To do so, we slightly modify the POMDP $\mathcal{P}$ used above. We assume all states have priority 1, and add two sink states $q_\top$ and $q_\bot$ with priorities respectively 0 and 1. Add a second action c such that c goes to $q_\top$ from state q_0 and to $q_\bot$ from any other state. There is an almost-sure strategy for the CoBüchi objective: play action a until it is certain that the current state is q_0, and then play c. This strategy has an exponential size, and by the above analysis, any smaller strategy cannot ensure that $q_\top$ is reached almost surely.

We note that the soundness of the belief-support MDP holds for the CoBüchi objective for general POMDPs, without any revealing hypothesis.

Lemma 6. *Let $\mathcal{P} = (Q, \mathsf{Act}, \mathsf{Sig}, \delta, q_0)$ be a POMDP with a CoBüchi objective specified by priority function $p\colon Q \to \{0, 1\}$, and let $\mathcal{P}_\mathcal{B}$ be its belief-support MDP with priority function $p_\mathcal{B}$. If there is an almost-sure strategy for $\mathsf{Parity}(p_\mathcal{B})$ in $\mathcal{P}_\mathcal{B}$, then there is an almost-sure strategy for $\mathsf{Parity}(p)$ in $\mathcal{P}$.*

Proof. The proof carries out as the one of Proposition 1, also using strategy $\widehat{\sigma_\mathcal{B}}$. However, the analysis of end components is simpler: simply notice that no end component in $\mathsf{EC}_{\sigma_\mathcal{B}}$ can contain a belief support with priority 1. In particular, strategy $\widehat{\sigma_\mathcal{B}}$ eventually does not encounter any state with priority 1. □

We now restate and prove the completeness of the belief-support MDP for weakly revealing POMDPs with priorities restricted to $\{0, 1, 2\}$.

Proposition 2. *Let $\mathcal{P} = (Q, \mathsf{Act}, \mathsf{Sig}, \delta, q_0)$ be a weakly revealing POMDP with priority function p with values in $\{0, 1, 2\}$. Let $\mathcal{P}_\mathcal{B}$ be its belief-support MDP with priority function $p_\mathcal{B}$. If there is an almost-sure strategy for $\mathsf{Parity}(p)$ in $\mathcal{P}$, then there is an almost-sure strategy for $\mathsf{Parity}(p_\mathcal{B})$ in $\mathcal{P}_\mathcal{B}$.*

Before giving the proof, we compare the technique to the proof of soundness of belief-support MDPs for weakly revealing POMDPs (Proposition 1). Two elements make this proof less straightforward:

- we start from an almost-sure strategy in a POMDP, on which we cannot exploit memory bounds (as opposed to the case of MDPs, for which we know that pure memoryless strategies suffice for parity objectives); in general, infinite-memory strategies are required in POMDPs (Chatterjee, Doyen, and Henzinger 2013, Theorem 2);
- we cannot just play a "copy" of the winning strategy of the POMDP in the MDP. In general, a strategy may be almost sure in a POMDP with a CoBüchi objective, while visiting infinitely many belief supports containing a state with priority 1. Starting from an arbitrary winning strategy σ in the POMDP, we may need to modify it to win in the MDP.

Proof of Proposition 2. Let $\sigma \in \Sigma(\mathcal{P})$ be an almost-sure strategy for $\mathsf{Parity}(p)$ in $\mathcal{P}$ (we recall that p takes values in $\{0, 1, 2\}$). Let $Q_\infty = \{q \in Q \mid \mathbb{P}^\mathcal{P}_\sigma[\mathsf{Büchi}(\{q\})] > 0\}$. Due to $\mathcal{P}$ being weakly revealing, the set Q_∞ is non-empty. Moreover, $\mathbb{P}^\mathcal{P}_\sigma\left[\bigcup_{q \in Q_\infty} \mathsf{Büchi}(\{q\})\right] = 1$.

The strategy we build in $\mathcal{P}_\mathcal{B}$ first tries to reach a belief support $\{q\}$ for some $q \in Q_\infty$. By Lemma 3, there is a strategy that achieves this with probability 1. We then show that if we reach a state $\{q\}$ for some $q \in Q_\infty$, then we can continue with an almost-sure strategy in $\mathcal{P}_\mathcal{B}$ for $\mathsf{Parity}(p_\mathcal{B})$. We distinguish two cases.

Assume first there is a strategy from $q \in Q_\infty$ in $\mathcal{P}$ that satisfies $\mathsf{Safety}(p^{-1}(1))$ almost surely. In terms of belief supports, this is equivalent to satisfying $\mathsf{Safety}(B_1)$ almost surely, where $B_1 = \{b \in 2^Q_\emptyset \mid \exists q \in b, p(q) = 1\}$. By Lemma 2, there is therefore a strategy from $\{q\}$ in $\mathcal{P}_\mathcal{B}$ that avoids any belief in B_1 almost surely, which is winning almost surely.

We now assume that there is no almost-sure strategy for $\mathsf{Safety}(p^{-1}(1))$ from q. By Corollary 1, this implies that this safety objective does not have value 1. Therefore, a state with priority 1 has a lower-bounded probability to be visited after each visit to q: there is $\alpha > 0$ such that for all strategies $\sigma' \in \Sigma(\mathcal{P})$, $\mathbb{P}^{\mathcal{P}^q}_{\sigma'}\left[\mathsf{Reach}(p^{-1}(1))\right] \geq \alpha$. From this, we deduce that $\mathbb{P}^{\mathcal{P}}_{\sigma}\left[\mathsf{Büchi}(p^{-1}(1)) \mid \mathsf{Büchi}(\{q\})\right] = 1$. Since $\mathbb{P}^{\mathcal{P}}_{\sigma}[\mathsf{Büchi}(\{q\})] > 0$ and $\mathbb{P}^{\mathcal{P}}_{\sigma}[\mathsf{Parity}(p) \mid \mathsf{Büchi}(\{q\})] = 1$, we conclude that $\mathbb{P}^{\mathcal{P}}_{\sigma}\left[\mathsf{Büchi}(p^{-1}(2)) \mid \mathsf{Büchi}(\{q\})\right] = 1$. Deducing from this a property over belief supports, if we take $B_2 = \{b \in 2^Q_\emptyset \mid \exists q \in b, p(q) = 2\}$, we have $\mathbb{P}^{\mathcal{P}}_{\sigma}[\mathsf{Büchi}(B_2) \mid \mathsf{Büchi}(\{q\})] = 1$. We deduce that in $\mathcal{P}_{\mathcal{B}}$, there must be an end component containing both $\{q\}$ and a belief support in B_2. By Theorem 6, there is an almost-sure strategy from $\{q\}$ in $\mathcal{P}_{\mathcal{B}}$ for $\mathsf{Parity}(p_{\mathcal{B}})$, which ends the proof. □

Lemma 1. *Deciding whether a POMDP is weakly revealing is EXPTIME-hard.*

Proof. To show that the problem is EXPTIME-hard, we reduce from the existence of a positively winning strategy for safety objectives in POMDPs, which is EXPTIME-complete (Chatterjee, Chmelik, and Tracol 2016).

Let $\mathcal{P} = (Q, \mathsf{Act}, \mathsf{Sig}, \delta, q_0)$ be a POMDP with a safety objective $\mathsf{Safety}(F)$, where $F \subseteq Q$. The general idea of the proof is to build a POMDP $\mathcal{P}'$ in which a revelation (and actually, infinitely many revelations) occur with probability 1 for all strategies if and only if there is no positively winning strategy in $\mathcal{P}$ for $\mathsf{Safety}(F)$. We apply two transformations to $\mathcal{P}$.

- First, we make two parallel, indistinguishable copies of the state space. This ensures that no revelations can happen, as there is no mechanism to know in which copy we currently are. This prevents existing revelations in $\mathcal{P}$ to alter our reduction.
- Second, we add a single sink state $\bot$ to which we redirect all transitions outgoing from F. We make sure that this sink state is revealed through a dedicated signal whenever it is reached. This ensures that infinitely many revelations happen if and only if F is reached.

Formally, let $\mathcal{P}' = (Q', \mathsf{Act}, \mathsf{Sig}', \delta', q'_0)$, where

- $Q' = (Q \times \{1, 2\}) \uplus \{q'_0, \bot\}$ ($\uplus$ denotes a *disjoint* union),
- $\mathsf{Sig}' = \mathsf{Sig} \uplus \{s_0, s_\bot\}$,
- for all $a \in \mathsf{Act}$, $\delta'(q'_0, a)(s_0, (q_0, 1)) = \delta'(q'_0, a)(s_0, (q_0, 2)) = \frac{1}{2}$ and for $q \in F$, $i \in \{1, 2\}$, $\delta'((q, i), a)(s_\bot, \bot) = \delta'(\bot, a)(s_\bot, \bot) = 1$. All other transitions are copied from $\mathcal{P}$ are stay within their own copy of $\mathcal{P}$.

We show that $\mathcal{P}'$ is weakly revealing if and only if $\mathcal{P}$ has no positively winning strategy for $\mathsf{Safety}(F)$.

By construction, infinitely many revelations happen in $\mathcal{P}'$ if and only if F is reached. Therefore, $\mathcal{P}'$ is weakly revealing if and only if for all strategies $\sigma \in \Sigma(\mathcal{P}')$, $\mathbb{P}^{\mathcal{P}'}_{\sigma}[\mathsf{Reach}(F \times \{1, 2\})] = 1$. If we restrict our focus to histories that have not gone through F, there is a natural bijection between strategies of $\mathcal{P}'$ and of $\mathcal{P}$. Based on this, one can show that for all strategies $\sigma \in \Sigma(\mathcal{P}')$, $\mathbb{P}^{\mathcal{P}'}_{\sigma}[\mathsf{Reach}(F \times \{1, 2\})] = 1$ if and only if for all strategies $\sigma \in \Sigma(\mathcal{P})$, $\mathbb{P}^{\mathcal{P}}_{\sigma}[\mathsf{Reach}(F)] = 1$. This last property says exactly that there is no positively winning strategy for $\mathsf{Safety}(F)$ in $\mathcal{P}$, ending the proof. □

We now move on the proof of undecidability of parity objectives with priorities 1, 2, and 3 for weakly revealing POMDPs.

Theorem 2. *The existence of an almost-sure strategy in weakly revealing POMDPs with a parity objective with priorities in $\{1, 2, 3\}$ is undecidable. The same holds for the existence of a positively winning strategy.*

A *probabilistic automaton* (Rabin 1963) is a tuple $\mathcal{A} = (Q, \mathsf{Act}, \delta, q_0)$. One can define their semantics through POMDPs: they behave like POMDPs in which we assume that the signals bring no information (Sig is a singleton). No useful information is provided by the signals along a play (beyond the number of steps played); pure strategies therefore correspond to words on alphabet Act.

We define the *value-1 problem* for probabilistic automata, which has been shown to be undecidable (Gimbert and Oualhadj 2010; Fijalkow 2017). We will reduce from this problem to prove Theorem 2. Let $\mathcal{A} = (Q, \mathsf{Act}, \delta, q_0)$ be a probabilistic automaton and $F \subseteq Q$ be a set of "final" states. For $\sigma \in A^*$, we denote by $\mathfrak{b}^{\mathcal{A}}_\sigma$ the belief after playing σ. We write $\mathfrak{b}^{\mathcal{A}}_\sigma(F) = \sum_{q \in F} \mathfrak{b}^{\mathcal{A}}_\sigma(q)$. The value-1 problem is the following: given a probabilistic automaton $\mathcal{A} = (Q, \mathsf{Act}, \delta, q_0)$ and a set of states $F \subseteq Q$, do we have $\sup_{\sigma \in A^*} \mathfrak{b}^{\mathcal{A}}_\sigma(F) = 1$?

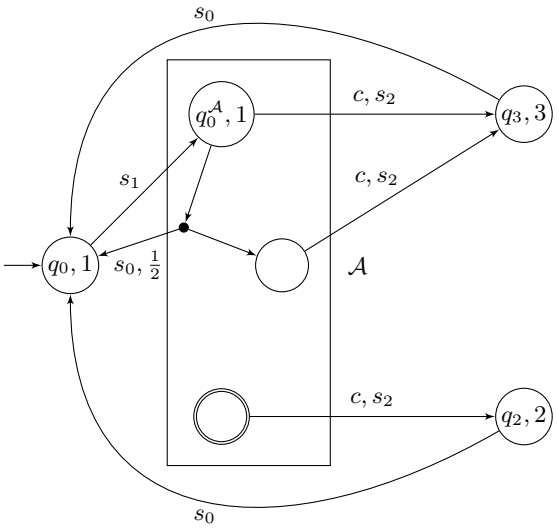

Figure 9: POMDP $\mathcal{P}^{\mathcal{A}}$ used in the proof of Theorem 2. The rectangle contains a copy of probabilistic automaton $\mathcal{A}$, with all transitions having probability $\frac{1}{2}$ to go back to q_0. When playing c from a state of $\mathcal{A}$, either q_3 is reached if the state is not in F, or q_2 is reached if the state is in F (represented by the double circle). This POMDP has an almost-sure strategy for the parity objective if and only if $\mathcal{A}$ has value 1 w.r.t. F.

Proof of Theorem 2. Let $\mathcal{A} = (Q^{\mathcal{A}}, \mathsf{Act}^{\mathcal{A}}, \delta^{\mathcal{A}}, q_0^{\mathcal{A}})$ be a probabilistic automaton, and $F \subseteq Q^{\mathcal{A}}$. To interpret $\mathcal{A}$ as a POMDP, we assume that all actions receive a single signal $s^{\mathcal{A}}$. We consider a generalization of the POMDP of Example 1: roughly, we replace the two states q_1 and q_1' by a copy of $\mathcal{A}$, where q_1' plays the role of states of F.

Formally, we consider the POMDP $\mathcal{P}^{\mathcal{A}} = (Q, \mathsf{Act}, \mathsf{Sig}, \delta, q_0)$ (depicted in Figure 9) such that

- $Q = \{q_0, q_2, q_3\} \uplus Q^{\mathcal{A}}$,
- $\mathsf{Act} = \{c\} \uplus \mathsf{Act}^{\mathcal{A}}$,
- $\mathsf{Sig} = \{s_0, s_1, s_2, s^{\mathcal{A}}\}$,
- for all $a \in \mathsf{Act}$, $\delta(q_0, a)(s_1, q_0^{\mathcal{A}}) = \delta(q_2, a)(s_0, q_0) = \delta(q_3, a)(s_0, q_0) = 1$,
- for all $q, q' \in Q^{\mathcal{A}}$, $a \in \mathsf{Act}^{\mathcal{A}}$, $\delta(q, a)(s^{\mathcal{A}}, q') = \frac{\delta_{\mathcal{A}}(q,a)(q')}{2}$, $\delta(q, a)(s_0, q_0) = \frac{1}{2}$,
- for $q \in Q^{\mathcal{A}} \setminus F$, $\delta(q, c)(s_2, q_3) = 1$, and for $q \in F$, $\delta(q, c)(s_2, q_2) = 1$.

POMDP $\mathcal{P}^{\mathcal{A}}$ is weakly revealing: all strategies visit q_0 infinitely often almost surely, and all visits to q_0 are revealed through signal s_0. We define a priority function p in $\mathcal{P}^{\mathcal{A}}$ with values in $\{1, 2, 3\}$: $p(q_2) = 2$, $p(q_3) = 3$, and $p(q) = 1$ for all $q \in Q^{\mathcal{A}} \cup \{q_0\}$.

The following claim suffices to conclude.

Claim 1. *If $\mathcal{A}$ does not have value 1 w.r.t. F, then there is no positively winning strategy in $\mathcal{P}^{\mathcal{A}}$ for $\mathsf{Parity}(p)$. If $\mathcal{A}$ has value 1 w.r.t. F, then there is an almost-sure strategy in $\mathcal{P}^{\mathcal{A}}$ for $\mathsf{Parity}(p)$.*

We prove the claim. We first assume that $\mathcal{A}$ does not have value 1 w.r.t. F. Let $\alpha > 0$ such that, for all $\sigma \in \mathsf{Act}^*$, $\mathfrak{b}_\sigma^{\mathcal{A}}(F) \leq 1 - \alpha$. To be an almost-sure strategy in $\mathcal{P}^{\mathcal{A}}$, a strategy needs to almost surely play c when the belief is a subset of $Q^{\mathcal{A}}$ infinitely often; otherwise, there is a positive probability to only see priority 1. Whenever c is played from a state in $Q^{\mathcal{A}}$, let us consider the actions played since the last visit to $q_0^{\mathcal{A}}$: it is a word $\sigma \in (\mathsf{Act}^{\mathcal{A}})^*$. Since $\mathfrak{b}_\sigma^{\mathcal{A}}(F) \leq 1 - \alpha$, we have that the probability to visit q_3 at the next step when playing c is $\geq \alpha$. Therefore, a strategy in $\mathcal{P}^{\mathcal{A}}$ that almost surely plays c infinitely often when the belief is a subset of $Q^{\mathcal{A}}$ will almost surely visit q_3 infinitely often, and almost surely lose since $p(q_3) = 3$.

Let us now assume that $\mathcal{A}$ has value 1 w.r.t. F. Therefore, for all $n \geq 1$, there is a word $\sigma_n' \in (\mathsf{Act}^{\mathcal{A}})^*$ such that $\mathfrak{b}_{\sigma_n'}^{\mathcal{A}}(F) \geq 1 - \frac{1}{2^n}$. The proof carries out as for Example 1. Let us divide a play in this POMDP into rounds 1, 2,...;

every time we go back to q_0 after visiting q_2 or q_3, we move to the next round. We define a strategy σ_n that tries to play σ'_n from $q_0^{\mathcal{A}}$ while staying in $Q^{\mathcal{A}}$: if it fails to do so (i.e., it sees signal s_0 which means that it is back to q_0), it goes back to $q_0^{\mathcal{A}}$ and retries. Whenever σ'_n could be fully played while staying in $Q^{\mathcal{A}}$, it plays c.

Consider the strategy σ that plays σ_n in round n; we show that σ is almost sure. This strategy ensures that infinitely many rounds happen, because at each round n, it will eventually succeed in playing σ'_n fully. At each round n, c is eventually played with probability 1. When c is played in round n, q_3 is visited with probability $\leq \frac{1}{2^n}$ and q_2 is visited with probability $\geq 1 - \frac{1}{2^n}$. State q_2 is clearly seen infinitely often almost surely, as the probability it is seen at each round is lower bounded by $\frac{1}{2}$. However, the probability that q_3 is never seen anymore after round n is greater than $\prod_{i=n}^{\infty}(1 - \frac{1}{2^i})$, which is positive and increases as n grows to ∞. We deduce that the probability that q_3 is seen at most finitely often is 1. □

Remark 3. *Although not our focus, a similar reduction shows that the value-*1 *problem for* reachability *objectives in weakly revealing POMDPs is undecidable. To see it, consider the POMDP $\mathcal{P}'$ which is like $\mathcal{P}^{\mathcal{A}}$ except that q_2 and q_3 are absorbing and revealing. POMDP $\mathcal{P}'$ is still weakly revealing. Consider the reachability objective* $\mathsf{Reach}(q_2)$. *We can similarly show that* $\sup_{\sigma \in \Sigma(\mathcal{P}')} \mathbb{P}_\sigma^{\mathcal{P}'}[\mathsf{Reach}(q_2)] = 1$ *if and only if $\mathcal{A}$ has value* 1 *w.r.t. F.*

F Additional details for Section 5

In this section, we give missing proofs of statements from Section 5. We start with the complexity lower bound on the existence of almost-sure strategies.

Proposition 4. *The following problem is* EXPTIME*-hard: given a strongly revealing POMDP with a CoBüchi objective, decide whether there is an almost-sure strategy.*

Proof. Our proof is by reduction from the existence of an almost-sure strategy for safety objectives in general POMDPs, which is EXPTIME-complete (Chatterjee, Chmelik, and Tracol 2016).

Let $\mathcal{P} = (Q, \mathsf{Act}, \mathsf{Sig}, \delta, q_0)$ be a POMDP along with an objective $\mathsf{Safety}(F)$ for some $F \subseteq Q$. We define a new POMDP $\mathcal{P}' = (Q, \mathsf{Act}, \mathsf{Sig}', \delta', q_0)$ on the same state and action space, in which we split each transition of $\mathcal{P}$ into three transitions: the original transition, a transition with a dedicated signal revealing the target state, and a transition resetting to the initial state, also with a signal indicating the reset. Formally, let

- $\mathsf{Sig}' = \mathsf{Sig} \uplus \{s_q \mid q \in Q\} \uplus \{s_{\mathsf{reset}}\}$,
- *copy of the original transition*: for all $q, q' \in Q$, $a \in \mathsf{Act}$, $s \in \mathsf{Sig}$, we define $\delta'(q,a)(s,q') = \frac{\delta(q,a)(s,q')}{3}$,
- *revelations over all possible states*: for all $q \in Q$ and $a \in \mathsf{Act}$, let $\mathsf{Succ}(q,a) = \{q' \mid \exists s \in \mathsf{Sig}, \delta(q,a)(s,q') > 0\}$ be the set of states that can be reached from q playing action a in one step; for $q' \in \mathsf{Succ}(q,a)$, we define $\delta'(q,a)(s_{q'},q') = \frac{1}{3 \cdot |\mathsf{Succ}(q,a)|}$,
- *reset*: for all $q \in Q$, $a \in \mathsf{Act}$, we define $\delta'(q,a)(s_{\mathsf{reset}}, q_0) = \frac{1}{3}$,

POMDP $\mathcal{P}'$ is strongly revealing: for every transition, there is a corresponding transition revealing the target state. The CoBüchi objective we define is the one induced by the priority function p such that $p(q) = 1$ if $q \in F$, and $p(q) = 0$ if $q \notin F$.

We show that there is an almost-sure strategy for the CoBüchi objective $\mathsf{Parity}(p)$ in $\mathcal{P}'$ if and only if there is an almost-sure strategy for $\mathsf{Safety}(F)$ in $\mathcal{P}$.

Assume first that there is an almost-sure strategy σ for $\mathsf{Safety}(F)$ in $\mathcal{P}$. The idea is to try to play σ in $\mathcal{P}'$ to avoid F, but there are two new events to take into account:

- whenever signal s_{reset} is seen, simply replay σ from the start;
- whenever a revealing signal $s_{q'}$ is seen on a transition $\delta(q,a)(s_{q'},q') > 0$, simply assume that the signal seen was actually a signal $s \in \mathsf{Sig}$ such that $\delta(q,a)(s,q') > 0$ (which exists by construction of $\mathcal{P}'$). This way, the assumed belief support is just an overapproximation of the actual singleton belief support. As σ wins almost surely for $\mathsf{Safety}(F)$ in $\mathcal{P}$, even the overapproximated belief support will never contain any state in F.

The strategy built is actually almost sure for $\mathsf{Safety}(F)$ in $\mathcal{P}'$, so it is in particular almost sure for $\mathsf{Parity}(p)$.

Assume now that there is no almost-sure strategy for $\mathsf{Safety}(F)$ in $\mathcal{P}$. Then, by Lemma 5, there is $n \in \mathbb{N}$ and $\alpha > 0$ such that for all strategies $\sigma \in \Sigma(\mathcal{P})$, $\mathbb{P}_\sigma^{\mathcal{P}}\left[\mathsf{Reach}^{\leq n}(F)\right] \geq \alpha$.

Consider now any strategy σ' on $\mathcal{P}'$. Almost surely, infinitely many resets (with signal s_{reset}) happen. Also almost surely, infinitely often, after each reset, the game lasts for more than n steps without any revelation or reset. Under this condition, the probability to visit F is at least α. Hence, there is almost surely and infinitely often a lower-bounded probability to visit F, so F is almost surely visited infinitely often. Hence, there is no almost-sure strategy for $\mathsf{Parity}(p)$ in $\mathcal{P}$. □

We now show that the analysis of the belief-support MDP is complete for strongly revealing POMDPs.

Proposition 3. *Let $\mathcal{P} = (Q, \mathsf{Act}, \mathsf{Sig}, \delta, q_0)$ be a strongly revealing POMDP with a priority function p, and let $\mathcal{P}_\mathcal{B}$ be its belief-support MDP with priority function $p_\mathcal{B}$. If there is an almost-sure strategy for $\mathsf{Parity}(p)$ in $\mathcal{P}$, then there is an almost-sure strategy for $\mathsf{Parity}(p_\mathcal{B})$ in $\mathcal{P}_\mathcal{B}$.*

Proof. Let $\sigma \in \Sigma(\mathcal{P})$ be an almost-sure strategy for $\mathsf{Parity}(p)$ in $\mathcal{P}$.

Let $\mathsf{EC}_\sigma^\mathcal{P} = \{U = (R, A) \mid \mathbb{P}_\sigma^\mathcal{P}[\inf(\pi) = U] > 0\}$ be the end components that σ can end up in, and visit exactly all their states and actions infinitely often, in the underlying MDP of $\mathcal{P}$. We build an almost-sure strategy in $\mathcal{P}_\mathcal{B}$. By Lemma 3, there is a strategy on $\mathcal{P}_\mathcal{B}$ that almost surely reaches a singleton belief support from some end component in $\mathsf{EC}_\sigma^\mathcal{P}$. We now define what the strategy does after such a singleton belief is reached.

Let $U = (R, A) \in \mathsf{EC}_\sigma^\mathcal{P}$. As it is a possible end component of σ, $p_{\mathsf{max}} = \max p(U)$ is even. Let $q_{\mathsf{max}}^U \in R$ such that $p(q_{\mathsf{max}}^U) = p_{\mathsf{max}}^U$. As an end component is strongly connected, from any state $q \in R$, there is a history from q to q_{max}^U. We denote it $h_q^U = q_0 \xrightarrow{a_1} q_1 \xrightarrow{a_2} \dots \xrightarrow{a_n} q_n$, where $q_0 = q$ and $q_n = q_{\mathsf{max}}^U$.

Whenever any belief support singleton $\{q\}$ for some $q \in R$ is reached, we play the following strategy on $\mathcal{P}_\mathcal{B}$:

- we try to achieve exactly the history h_q^U, hoping for a revelation after *every* action (this exploits that $\mathcal{P}$ is strongly revealing);
- if that fails, let $B_{\neg R} = \{b \in 2_\emptyset^\mathcal{P} \mid b \cap (Q \setminus R) \neq \emptyset\}$, i.e., all the belief supports that indicate that R may have been left. We show below that we can play an almost-sure strategy for $\mathsf{Safety}(B_{\neg R})$. We play this strategy until we fall back to a singleton belief support in U, which happens eventually due to $\mathcal{P}$ being strongly revealing.

This strategy is almost sure for $\mathsf{Parity}(p_\mathcal{B})$: eventually, it reaches a singleton belief support from some end component U in $\mathsf{EC}_\sigma^\mathcal{P}$. Then, it only sees beliefs whose states are all in U. And whenever a singleton belief support is reached, which happens infinitely often, we have a lower-bounded probability to reach q_{max}^U, so q_{max}^U is reached infinitely often almost surely. It remains to show that whenever we deviate from a revelation at every step of a history h_q^U, we have an almost-sure strategy for $\mathsf{Safety}(B_{\neg R})$.

Let $U = (R, A) \in \mathsf{EC}_\sigma^\mathcal{P}$, $q \in R$, and $h_q^U = q_0 \xrightarrow{a_1} q_1 \xrightarrow{a_2} \dots \xrightarrow{a_n} q_n$. Let $b \in 2_\emptyset^Q$ be such that $\delta_\mathcal{B}(\{q_i\}, a_{i+1})(b) > 0$ (i.e., b is a possible successor of $\{q_i\}$ along the path h_q). We show that there is an almost-sure strategy from b for $\mathsf{Safety}(B_{\neg R})$ in $\mathcal{P}_\mathcal{B}$.

Assume on the contrary that for all strategies $\sigma' \in \Sigma(\mathcal{P}_\mathcal{B}^b)$, we have $\mathbb{P}_{\sigma'}^{\mathcal{P}_\mathcal{B}^b}[\mathsf{Reach}(B_{\neg R})] > 0$. Then, by Lemma 4, for all strategies $\sigma' \in \Sigma(\mathcal{P}_\mathcal{B}^b)$, we have $\mathbb{P}_{\sigma'}^{\mathcal{P}_\mathcal{B}^b}\left[\mathsf{Reach}^{\leq 2^{|Q|}-1}(B_{\neg R})\right] > 0$. Hence, by Lemma 2, for all strategies $\sigma \in \Sigma(\mathcal{P}^b)$, $\mathbb{P}_\sigma^{\mathcal{P}^b}\left[\mathsf{Reach}^{\leq 2^{|Q|}-1}(B_{\neg R})\right] > 0$. This means that for all strategies, there is a history of length at most $2^{|Q|} - 1$ that exits R from b. For pure strategies, the probability to exit R is therefore at least $\beta_\mathcal{P}^{2^{|Q|}-1}$; by Theorem 7, this extends to all (not only pure) strategies. This means that every time $\{q_i\}$ is visited and a_{i+1} is played, there is a lower-bounded probability to exit R. This contradicts that U is an end component. □

To conclude the missing proofs, we discuss our undecidability result for strongly revealing CoBüchi games. The syntax for games was defined in Section 5.2.

Remark 4. *The model of (Bertrand, Genest, and Gimbert 2017) allows for distinct signals for both players, and is therefore slightly more general. Our undecidability proof works even when the signals given to both players are the same, which is why we opted for this restriction.*

Theorem 4. *The existence of an almost-sure strategy in strongly revealing CoBüchi games is undecidable.*

Proof. We reduce from the undecidable *value-*1 *problem for probabilistic automata*. This problem was already used in the reduction of Theorem 2; we refer to Appendix E for a definition and an introduction to the problem.

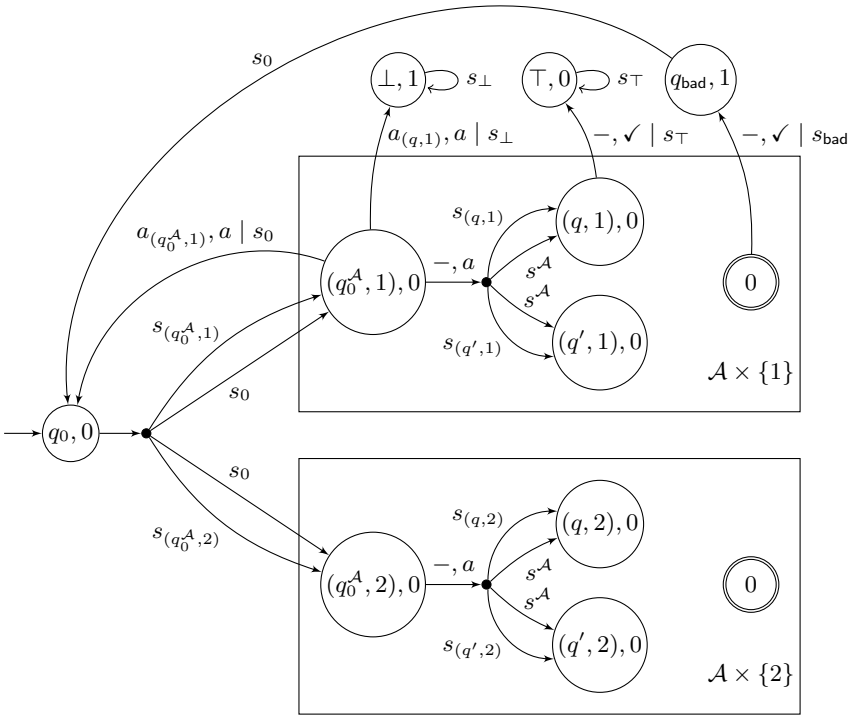

Figure 10: Game $\mathcal{G}^{\mathcal{A}}$ used in the proof of Theorem 4. States with a double circle correspond to states of F. Many transitions are not represented, but we illustrate at least one transition of each kind (randomization in the initial state, possible revelations for each state inside $\mathcal{A} \times \{1, 2\}$, right and wrong F-guesses, right and wrong state guesses for Player 1).

Let $\mathcal{A} = (Q^{\mathcal{A}}, \mathsf{Act}^{\mathcal{A}}, \delta^{\mathcal{A}}, q_0^{\mathcal{A}})$ be a probabilistic automaton and $F \subseteq Q^{\mathcal{A}}$ be a set of final states. We build a game $\mathcal{G}^{\mathcal{A}}$ with a CoBüchi objective such that Player 1 wins almost surely if and only if $\mathcal{A}$ does not have value 1 w.r.t. F.

Our construction is illustrated in Figure 10. We briefly give an intuition of the construction before a more formal definition. The game happens in a copy of $\mathcal{A}$, in which all states are given priority 0. Player 2 picks the letters in $\mathsf{Act}^{\mathcal{A}}$ to induce transitions in $\mathcal{A}$. To see priority 1, Player 2 needs to "F-guess", i.e., claim that the current state is in F by playing a special action $\checkmark$: if the guess is right, Player 2 goes to a special state q_{bad} that produces priority 1, and the game resets; if the guess is wrong, Player 2 immediately loses as the next state is a sink state $\top$ with priority 0.

We artificially make the game strongly revealing by adding a special signal which may reveal the state at each transition. This makes the game easier for Player 2, as it is easier for Player 2 to reach a state of F if revelations happen. To balance the game, we give to Player 1 the power to reset the game by guessing the current state; Player 1 has a special action for each state. If the guess is correct (which is easy after a revelation), the game resets to the initial state; if the guess is wrong, a sink state $\bot$ with priority 1 is reached, which immediately loses for Player 1. Player 1 may also choose not to guess anything by playing the waiting move $-$. To prevent revelations that may be occurring in $\mathcal{A}$ from helping Player 1, we make two copies of $\mathcal{A}$ and start the game randomly in one of the two copies, which are never distinguished except by the added revealing signals (the same trick was used in the proof of Lemma 1). This way, Player 1 can only guess the current state with certainty after one of the added revealing signals occurred. We also give to Player 2 the power to reset the game by correctly guessing the current state, which prevents a state that cannot reach F anymore to be attained and revealed, with no way for Player 2 to ever go back to F.

Formally, we define the game $\mathcal{G}^{\mathcal{A}} = (Q, \mathsf{Act}_1, \mathsf{Act}_2, \mathsf{Sig}, \delta, q_0)$ as follows:

- $Q = (Q^{\mathcal{A}} \times \{1, 2\}) \uplus \{q_0, q_{\mathsf{bad}}, \bot, \top\}$,
- $\mathsf{Act}_1 = \{a_{(q,i)} \mid (q, i) \in Q^{\mathcal{A}} \times \{1, 2\}\} \uplus \{-\}$,

- $\mathsf{Act}_2 = \{a_{(q,i)} \mid (q,i) \in Q^{\mathcal{A}} \times \{1,2\}\} \uplus \mathsf{Act}^{\mathcal{A}} \uplus \{\checkmark\}$,
- $\mathsf{Sig} = \{s^{\mathcal{A}}, s_0, s_{\mathsf{bad}}, s_\top, s_\bot\} \uplus \{s_{(q,i)} \mid (q,i) \in Q^{\mathcal{A}} \times \{1,2\}\}$,
- $\top$ and $\bot$ are sinks: for all $a \in \mathsf{Act}_1 \times \mathsf{Act}_2$, $\delta(\top, a)(s_\top, \top) = \delta(\bot, a)(s_\bot, \bot) = 1$,
- we give the following priorities to states: $p(q_0) = 0, p(q_{\mathsf{bad}}) = 1, p(\top) = 0, p(\bot) = 1$, and for $(q,i) \in Q \times \{1,2\}$, $p((q,i)) = 0$,
- after q_{bad}, we always move back to q_0: for $(a_1, a_2) \in \mathsf{Act}_1 \times \mathsf{Act}_2$, $\delta(q_{\mathsf{bad}}, (a_1, a_2))(s_0, q_0) = 1$,
- the initial state randomizes over the two copies of the initial state of $\mathcal{A}$ (there is already a positive probability of a revelation): for all $a \in \mathsf{Act}_1 \times \mathsf{Act}_2$, $\delta(q_0, a)(s_0, (q_0^{\mathcal{A}}, 1)) = \delta(q_0, a)(s_0, (q_0^{\mathcal{A}}, 2)) = \delta(q_0, a)(s_{(q_0^{\mathcal{A}},1)}, (q_0^{\mathcal{A}}, 1)) = \delta(q_0, a)(s_{(q_0^{\mathcal{A}},2)}, (q_0^{\mathcal{A}}, 2)) = \frac{1}{4}$,
- when Player 1 plays $-$, transitions in each copy of $\mathcal{A}$ behave like in $\mathcal{A}$, with a positive probability of a revelation at each transition: for $(q,i) \in Q^{\mathcal{A}} \times \{1,2\}$, $q' \in Q^{\mathcal{A}}$, and $a \in \mathsf{Act}_2$, $\delta((q,i), (-,a))(s^{\mathcal{A}}, (q',i)) = \delta((q,i), (-,a))(s_{(q',i)}, (q',i)) = \frac{\delta^{\mathcal{A}}(q,a)(q')}{2}$,
- at each round, players can guess the current state of $Q^{\mathcal{A}} \times \{1,2\}$ by playing the corresponding action: a wrong guess ends the game by leading to the sink state that is losing for the player guessing wrong, while a right guess simply resets the game. For a state $(q,i) \in Q^{\mathcal{A}} \times \{1,2\}$, we say that action $a_{(q,i)}$ is a *right guess*, while an action $a_{(q',j)}$ is a *wrong guess* if $q' \neq q$ or $j \neq i$. Formally, for all $(q,i) \in Q^{\mathcal{A}} \times \{1,2\}$, $a_1 \in \mathsf{Act}_1$, $a_2 \in \mathsf{Act}_2$, we define $\delta((q,i), (a_1, a_2))(s_\bot, \bot) = 1$ if a_1 is a wrong guess, and $\delta((q,i), (a_{(q',j)}, a_2))(s_\top, \top) = 1$ if a_2 (but not a_1) is a wrong guess. For all $(q,i) \in Q^{\mathcal{A}} \times \{1,2\}$, $a_1 \in \mathsf{Act}_1$, $a_2 \in \mathsf{Act}_2$, we define $\delta((q,i), (a_1, a_2))(s_0, q_0) = 1$ if a_1 or a_2 is a right guess (and none is a wrong guess).
- If there are no other guesses, Player 2 can "F-guess" whether the current state is in $F \times \{1,2\}$ with action $\checkmark$: if it is a right F-guess, then q_{bad} is reached, producing priority 1, and the game then resets. If it is a wrong F-guess, the next state is $\top$, an immediate win for Player 1. Formally, for all $(q,i) \in Q^{\mathcal{A}} \times \{1,2\}$, we define $\delta((q,i), (-,\checkmark))(s_{\mathsf{bad}}, q_{\mathsf{bad}}) = 1$ if $q \in F$, and $\delta((q,i), (-,\checkmark))(s_\top, q_\top) = 1$ if $q \notin F$.

Game $\mathcal{G}^{\mathcal{A}}$ is strongly revealing: every state has a dedicated signal, which is produced with non-zero probability on each incoming transition. To win, Player 1 needs to avoid ending up in $\bot$ and avoid visiting q_{bad} infinitely often.

We show the following claim, which suffices to conclude.

Claim 2. *Player 1 has an almost-sure strategy for the CoBüchi objective of the strongly revealing game $\mathcal{G}^{\mathcal{A}}$ if and only if $\mathcal{A}$ does not have value 1 w.r.t. F.*

We prove the claim. Assume first that $\mathcal{A}$ has value 1 w.r.t. F. We show that Player 2 has a positively winning strategy. As $\mathcal{A}$ has value 1 w.r.t. F, there is a sequence $\sigma_1, \sigma_2, \ldots$ of words in $(\mathsf{Act}^{\mathcal{A}})^*$ such that $\mathfrak{b}^{\mathcal{A}}_{\sigma_i}(F) \geq 1 - \frac{1}{2^i}$ (notation $\mathfrak{b}^{\mathcal{A}}_{\sigma_i}$ was defined in Appendix E). To define the strategy of Player 2, we split a play into rounds. The game starts at round 1. At round i, Player 2 tries to play word σ_i fully in a copy of $\mathcal{A}$ without any revelation (i.e., a signal $s_{(q,i)}$ for some $(q,i) \in Q^{\mathcal{A}} \times \{1,2\}$); if such a revelation $s_{(q,i)}$ happens, Player 2 plays the corresponding action $a_{(q,i)}$ to reset the game and tries to play σ_i again. Once σ_i can be played fully without a revelation (which happens eventually with probability 1), Player 2 plays $\checkmark$, which goes to q_{bad} with probability $\geq 1 - \frac{1}{2^i}$. The strategy then moves over to round $i+1$. Assuming Player 1 only plays $-$, this strategy wins for Player 2 with probability greater than the infinite product $\prod_{i \geq 1}(1 - \frac{1}{2^i})$, which is a positive number. If Player 1 tries a guess when in a copy of $\mathcal{A}$, either a revelation just happened, in which case Player 2 was resetting the game anyway, or the probability of a wrong guess for Player 1 is at least $\frac{1}{2}$, which immediately wins the game with positive probability for Player 2.

Assume now that $\mathcal{A}$ does not have value 1 w.r.t. F. This means that there is $\alpha > 0$ such that for all finite words $\sigma \in (\mathsf{Act}^{\mathcal{A}})^*$, $\mathfrak{b}^{\mathcal{A}}_{\sigma}(F) \leq 1 - \alpha$. We define a strategy of Player 1: play a right guess whenever a revelation happens in $Q^{\mathcal{A}} \times \{1,2\}$, and otherwise always play $-$. We show that this strategy is almost sure for Player 1. Observe that this strategy guarantees that $\bot$ is never reached and that the game resets infinitely often (unless the winning state $\top$ is reached due to a mistake of Player 2). To win, Player 2 needs to visit q_{bad} infinitely often, which requires to play $\checkmark$ infinitely often while in a state of $F \times \{1,2\}$. After every reset, two things can happen when the play moves into $Q^{\mathcal{A}} \times \{1,2\}$:

- either a revelation happens, in which case Player 1 immediately resets the game;
- or Player 2 plays some word $\sigma \in (\mathsf{Act}^{\mathcal{A}})^*$ without any revelation and then attempts to play $\checkmark$. This reaches $\top$ with probability $\geq \alpha$.

Either Player 2 plays ✓ only finitely often, in which case priority 1 is seen at most finitely often, or Player 2 attempts ✓ infinitely often from a non-revealed state of $Q^{\mathcal{A}} \times \{1, 2\}$, which eventually leads to $\top$ almost surely. In both cases, Player 1 wins almost surely. □