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
[2]CNRS, IRIF, Universite de Paris, France
[3]University of Antwerp – Flanders Make, Antwerp, Belgium

AAAI 2025 – March 1, 2025

LABORATOIRE BORDELAIS DE RECHERCHE EN INFORMATIQUE LaBRI

INSTITUT DE RECHERCHE EN INFORMATIQUE FONDAMENTALE

University of Antwerp

Outline

Partially observable Markov decision processes (POMDPs):

- nondeterminism,
- stochasticity,
- **uncertainty** about the actual state.

Offline approach: complete description of the POMDP as an input.

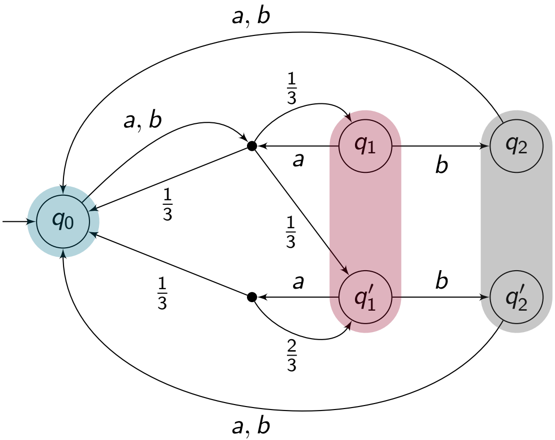
Goal

Strategy synthesis for ω-*regular objectives* (e.g., reachability, safety, *Büchi*...).
Undecidable in general; **decidable subclasses**?

Means

Two subclasses with probabilistic guarantees about sometimes **knowing the actual state**.
Natural algorithm that applies to this class.

Partially observable MDPs



States Q, **initial state** q_0, **actions** Act, **observations** Obs.
Strategies are functions $(\mathrm{Act} \times \mathrm{Obs})^* \to \mathcal{D}(\mathrm{Act})$.

Objective

- Common **objectives**:
  - **Reachability**: a good state is eventually visited,
  - **Büchi**: $p\colon Q \to \{1, 2\}$; good states (2) are visited infinitely often,
  - **coBüchi**: $p\colon Q \to \{0, 1\}$; bad states (1) are visited finitely often.
- More generally: function $p\colon Q \to \{0, \ldots, d\}$ assigning **priorities** to **states**.
- **Parity objective**: the **maximal** priority seen infinitely often is **even**.
- **Question**: does there exist an **almost-sure** strategy?

Decidability in POMDPs[1,2]

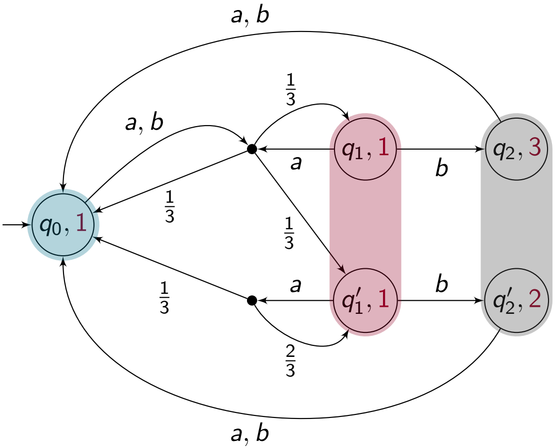
- Almost-sure **reachability**, **safety**, and **Büchi** are **EXPTIME-complete**.
- Almost-sure **coBüchi** (and therefore **parity**) are **undecidable**.

[1] Baier, Größer, and Bertrand, “Probabilistic ω-automata”, 2012.

[2] Chatterjee, Chmelik, and Tracol, “What is decidable about partially observable Markov decision processes with ω-regular objectives”, 2016.

Example of a difficult POMDP

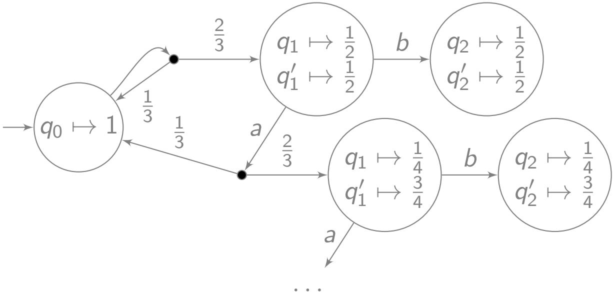
Added priorities 1, 2, 3 to the previous POMDP.



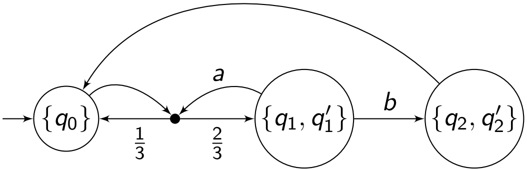
Almost-sure strategy? Yes! Move to q_2/q'_2 when *increasingly high probability* to be in q'_1.

Belief (support) MDP

POMDPs induce **infinite** **belief MDPs**:



Finite: only keep belief **supports**:



When does the analysis of the belief **support** MDP suffice?

In general, neither sound nor complete...

Looking for decidable classes...

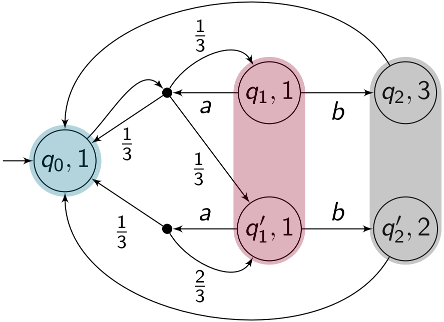
1. Weak Revelations

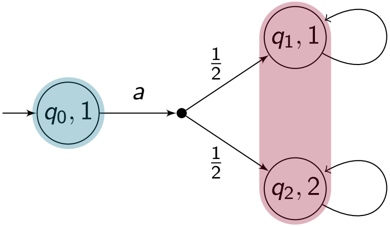
by restricting the information loss!

Weak revelations

Weak revelations

A POMDP is **weakly revealing** if for all strategies,
almost surely, the **current state can be known** infinitely often.





Not weakly revealing

Weakly revealing: q_0 is visited infinitely often

Weak revelations

Weak revelations

A POMDP is **weakly revealing** if for all strategies,
almost surely, the **current state can be known** infinitely often.

When a *revealing history* happens, the finite belief **support** MDP contains **as much information** as the infinite belief MDP.

$$\{q_0\} \qquad \approx \qquad q_0 \mapsto 1$$

Weak revelations: results

"Weakly revealing" is a semantic property, but is **decidable**.

Priorities $\{0, 1, 2\}$ (encompassing Büchi and coBüchi)

There exists an almost-sure strategy...
in a **weakly revealing POMDP** $\mathcal{P} \iff$ in the **belief support MDP** of $\mathcal{P}$.

Decidability

Almost-sure **parity** $\{0, 1, 2\}$ for **weakly revealing** POMDPs is EXPTIME-complete.

Algorithm: solve the **belief support MDP** $\rightsquigarrow$ in EXPTIME.

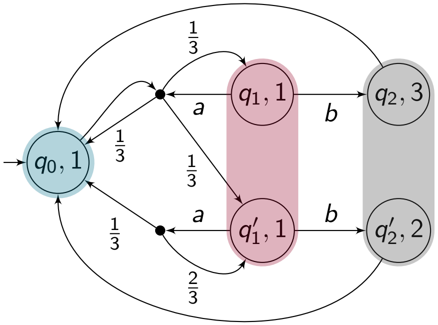
Why restrict to parity $\{0, 1, 2\}$? Unfortunately...

Full parity remains undecidable

Undecidability

Almost-sure **parity** $\{1, 2, 3\}$ is **undecidable** for **weakly revealing** POMDPs.

Belief support MDP does not help for this **weakly revealing** POMDP with priorities 1, 2, 3.



Looking for **more** decidable classes...

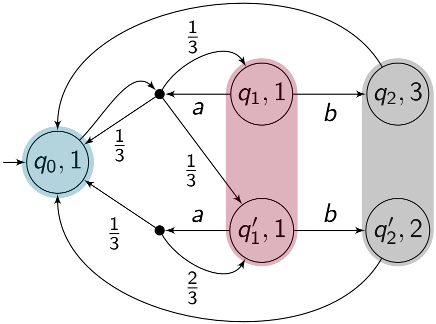
2. Strong Revelations

by restricting the information loss **even more**!

Strong revelations

Strong revelations

A POMDP is **strongly revealing** if for every transition $q \xrightarrow{a} q'$, there is a non-zero probability of **revealing** q'.



Not strongly revealing: $q_1 \xrightarrow{a} q_1'$ is a possible transition, but nothing can reveal q_1' with certainty.

Strong revelations: results

Full parity

There exists an almost-sure strategy...

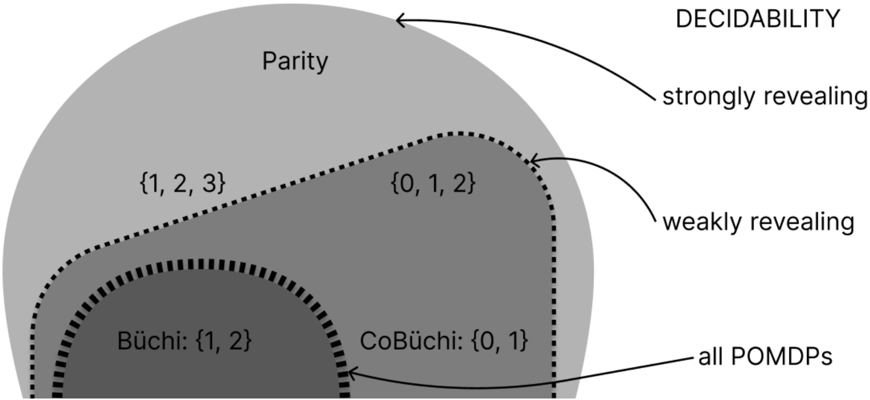
in a **strongly revealing POMDP** $\mathcal{P}$ $\iff$ in the **belief support MDP** of $\mathcal{P}$.

Theorem

Almost-sure **parity** for **strongly revealing** POMDPs is EXPTIME-complete.

Algorithm: solve the **belief support MDP** $\rightsquigarrow$ in EXPTIME (again!).

Summary



Decidable subclasses for *parity* POMDPs depending on the **revelation** mechanism.

Decidability frontier when we move to **games**: **games with partial observation** remain **undecidable** for coBüchi under **strong revelations**.

Final comments

Paper link:

- A few works with similar approaches.[3,4,5]
- Implementation available at `https://github.com/gaperez64/pomdps-reveal`.
- **Take-home message**: While POMDPs are undecidable in general, they are not hopeless: there exist **natural and expressive decidable subclasses**.
- **Future directions**:
  - **more general** decidable classes,
  - **more expressive** objectives (e.g., quantitative reachability),
  - other **algorithms** than solving the belief support MDP?

Thanks!

[3] Berwanger and Mathew, "Infinite games with finite knowledge gaps", 2017.

[4] Vlassis, Littman, and Barber, "On the Computational Complexity of Stochastic Controller Optimization in POMDPs", 2012.

[5] Bellinger et al., "Active Measure Reinforcement Learning for Observation Cost Minimization", 2021; Krale, Simão, and Jansen, "Act-Then-Measure: Reinforcement Learning for Partially Observable Environments with Active Measuring", 2023.