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

[Published at AAAI'25]

Pierre Vandenhove

Joint work with Marius Belly, Nathanaël Fijalkow, Hugo Gimbert, Florian Horn, Guillermo A. Pérez

INFORTECH Day 2025 – May 23, 2025

UMONS Université de Mons

LABORATOIRE BORDELAIS DE RECHERCHE EN INFORMATIQUE LaBRI

IRIF INSTITUT DE RECHERCHE EN INFORMATIQUE FONDAMENTALE

University of Antwerp

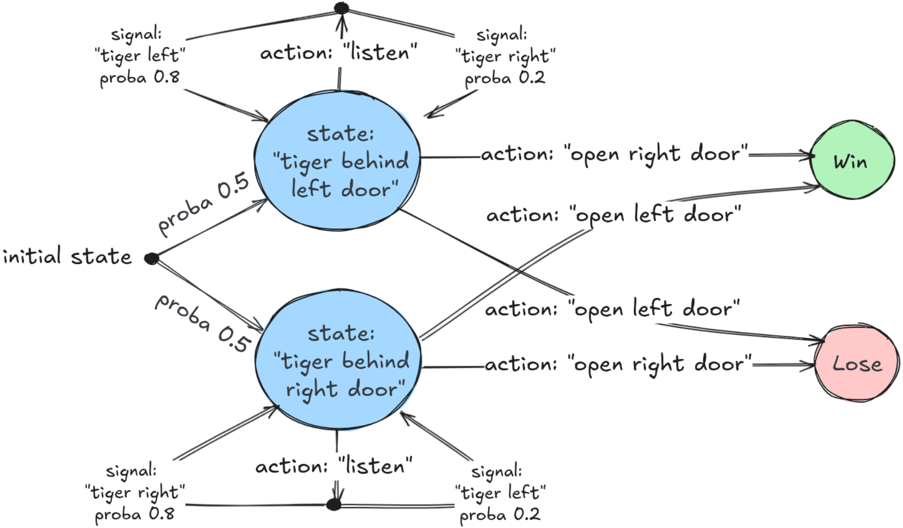
Outline

- A lot to unpack in the title!
- No technical detail here; mainly **examples** and **motivations**.
- The main object we consider is a **POMDP**: a

  partially observable Markov decision process.
- Before defining POMDPs formally, let us look at the well-known "tiger" example. . . [1]
  - A person is in front of **two closed doors**.
  - A tiger is behind **one** of the doors.
  - The person **has to open** the non-tiger door to win.
  - The person can **listen** to get some **imperfect** information about the tiger's location.

[1] Kaelbling, Littman, and Cassandra, "Planning and acting in partially observable stochastic domains", 1998.

Tiger POMDP



Maximal probability of reaching Win? No strategy wins with probability $1\ldots$

However, **for every** $\epsilon > 0$, there is a strategy that wins with probability $\geq 1 - \epsilon$.

Why is this *a POMDP*?

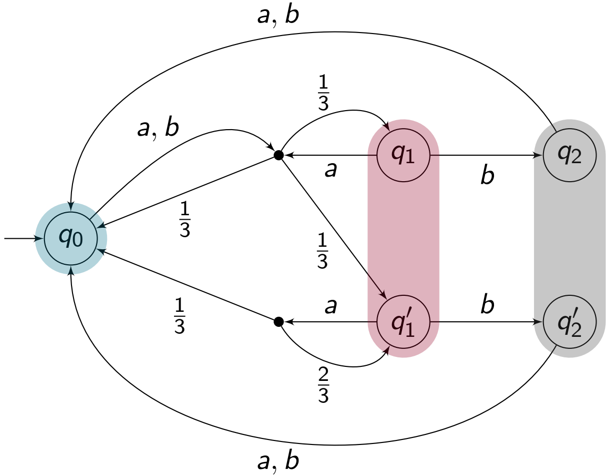
Features of a **partially observable Markov decision processes**:

- **nondeterminism** (multiple actions to choose from),
- **stochasticity** (probabilistic transitions),
- **uncertainty** about the actual state (the current state is not always known).

To make them interesting, we also need an **objective**.

- For the tiger: **maximize the probability** to not get eaten, i.e., **to reach** state "Win".

Partially observable Markov decision processes, more formally



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

Two approaches to analyse POMDPs

POMDPs are studied from two complementary points of view:

1. **Reinforcement learning**:
   - Learns a *good* strategy by **interacting** with the environment.
   - Often a **model-free**, online approach.
   - Can handle **big state spaces**, but **few guarantees** about the output.
2. **Model checking**:
   - Given a **static description** of the POMDP, computes the "best" strategy.
   - Often a **model-based**, offline approach.
   - Based on **computability/complexity** theory.
   - Can not always handle big state spaces (many problems need $\geq$ exponential time), but **mathematical guarantees**.

Here, **model-checking** approach: what is computable about POMDPs?

What is decidable about POMDPs?

Decidability in POMDPs with **reachability objectives**[2,3,4,5]

- Given a POMDP and a threshold $t \in (0,1)$, is there a strategy that reaches the target with probability $\geq t$? **Undecidable** ☹
- Given a POMDP, is it true that for all $\epsilon > 0$, there is a strategy that reaches the target with probability $\geq 1 - \epsilon$? **Undecidable** ☹
- Given a POMDP, is there an algorithm that **approximates** the supremum probability of reaching the target? **No** ☹
- Given a POMDP, is there a strategy that reaches the target with probability 1? **EXPTIME-complete**! ☺

Summary: **quantitative** problems are **all** undecidable in POMDPs.
Qualitative problems (e.g., existence of an **almost-sure strategy**): it depends!

[2] Madani, Hanks, and Condon, "On the undecidability of probabilistic planning and related stochastic optimization problems", 2003.
[3] Gimbert and Oualhadj, "Probabilistic Automata on Finite Words: Decidable and Undecidable Problems", 2010.
[4] Baier, Größer, and Bertrand, "Probabilistic ω-automata", 2012.
[5] Chatterjee, Chmelik, and Tracol, "What is decidable about partially observable Markov decision processes with ω-regular objectives", 2016.

Natural objectives

For almost-sure strategies, are there other decidable objectives?

- Common **objectives**:
  - **Reachability**: a good state is eventually visited,
  - **Safety**: a bad state is never 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**.
- Parity objectives encompass the crucial class of ω-**regular objectives** (hence the title!).

Decidability of **almost-sure** strategies in POMDPs[6,7]

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

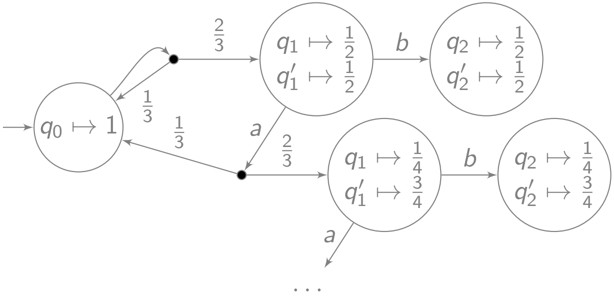
---

[6] Baier, Größer, and Bertrand, "Probabilistic ω-automata", 2012.

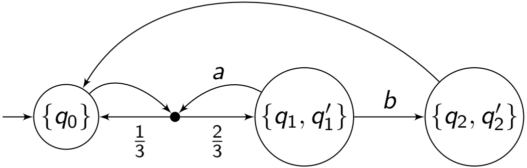
[7] Chatterjee, Chmelik, and Tracol, "What is decidable about partially observable Markov decision processes with ω-regular objectives", 2016.

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...

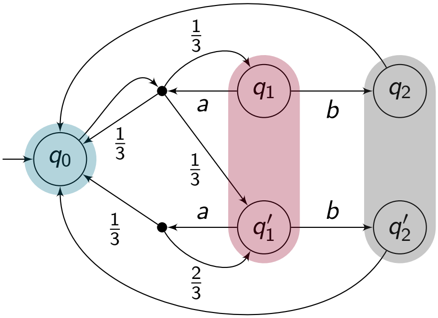
1. Weak Revelations

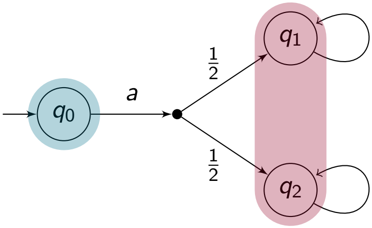
by restricting the information loss!

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.

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

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...

Undecidability

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

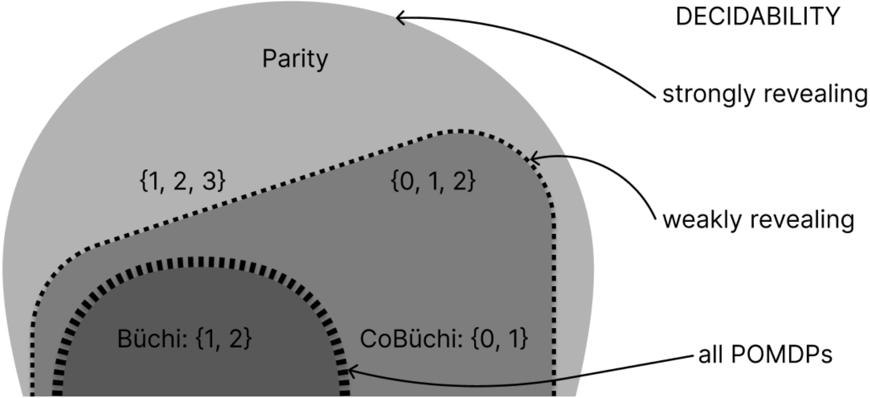
Looking for **more** decidable classes...

2. Strong Revelations

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

No time here: see the paper ☺

Summary



Decidable subclasses for almost-sure *parity* POMDPs w.r.t. **revelation** mechanisms.

Decidability frontier when we move to two-player **games**: **games with partial observation** remain **undecidable** no matter the revelation mechanism.

Final comments

Paper link:

- Implementation available at `https://github.com/gaperez64/pomdps-reveal`; currently pimping up the experiments for a journal version.
- **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!