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

Pierre Vandenhove

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

LaBRI, Université de Bordeaux

UMONS Formal Methods Reading Group – October 7, 2024

# Outline

**Partially observable Markov decision processes** (POMDPs):

- stochastic,
- nondeterministic,
- **uncertainty** about the actual state.

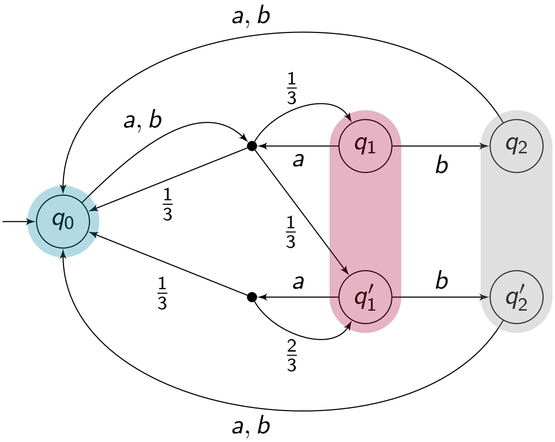
## Goal

Strategy synthesis for **parity objectives** ($\rightsquigarrow$ $\omega$-regular objectives).
Undecidable in general; **decidable subclasses**?

## Means

**Two subclasses** with probabilistic guarantees about sometimes **knowing the actual state**; restrictions about **information loss**.

# Partially observable MDPs

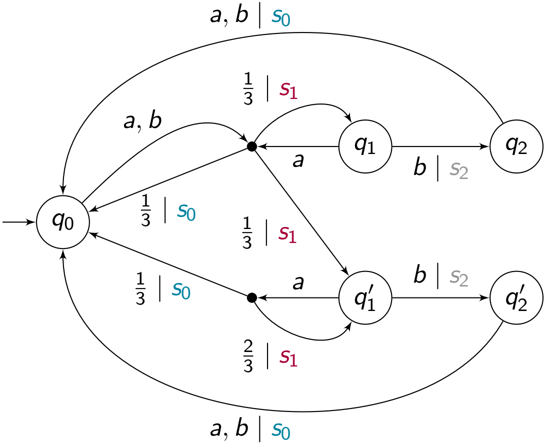



**States** $Q$, **actions** Act, **observations** Obs.

Strategies are functions $(\mathrm{Act} \times \mathrm{Obs})^* \to \mathcal{D}(\mathrm{Act})$.

# Same model, but **signals** instead of observations

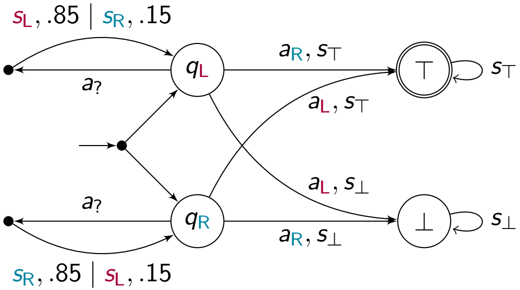
For convenience, "transition-based" **signals** Sig instead of "state-based" observations.




$\rightsquigarrow$ Equivalent models (increase linear in $|\mathrm{Sig}|$ when going from signals to observations).

# Tiger example[1]

- Tiger behind one of two doors: the L door or the R door.
- You can *listen* ($a_?$) or *open a door* ($a_L$ or $a_R$).




- Probability to **reach** $\top$ can be arbitrarily close to 1 (the POMDP has *value* 1), but no **almost-sure strategy**.

[1]Cassandra, Kaelbling, and Littman, "Acting Optimally in Partially Observable Stochastic Domains", 1994.

# Objective

- Function $p\colon Q \to \{0,\dots,d\}$ assigning **priorities** to **states**.
- **Parity objective**: the **maximal** priority seen infinitely often is **even**.
- Common subclasses:
  - **Büchi**: $p\colon Q \to \{1,2\}$: something good (2) occurs infinitely often,
  - **coBüchi**: $p\colon Q \to \{0,1\}$: something bad (1) occurs finitely often.
- **Almost-sure** strategies; "qualitative".

**Theorem**[2,3]

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

Undecidability already for **probabilistic automata** ($|\mathrm{Sig}| = 1$).
Quantitative problems (e.g., value-1 problem) are undecidable for reachability objectives.[4]

---

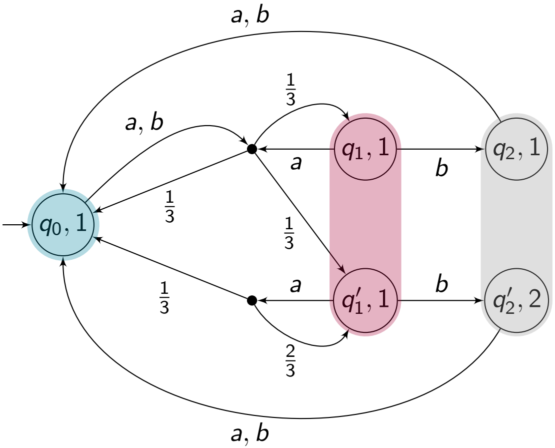
[2] Baier, Größer, and Bertrand, "Probabilistic $\omega$-automata", 2012.

[3] Chatterjee, Chmelik, and Tracol, "What is decidable about partially observable Markov decision processes with $\omega$-regular objectives", 2016.

[4] Gimbert and Oualhadj, "Probabilistic Automata on Finite Words: Decidable and Undecidable Problems", 2010.

# Example

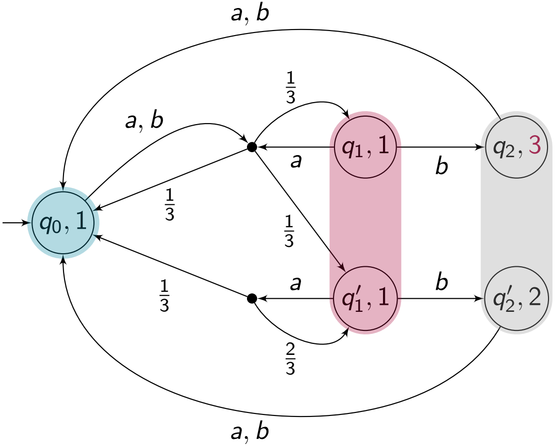
Added priorities 1, 2 to the previous POMDP.




**Almost-sure strategy**? Yes! Move to $q_2/q'_2$ infinitely often.

# Example

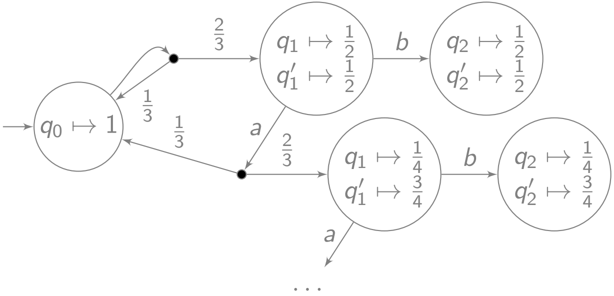
Added priorities 1, 2, 3 to the previous POMDP. **Changed the priority of $q_2$ to 3.**



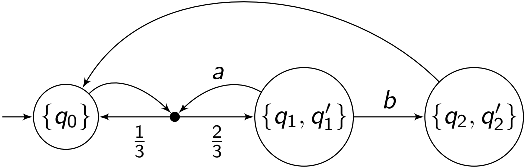

**Almost-sure strategy**? Yes! Move to $q_2/q_2'$ when *increasingly high probability* to be in $q_1'$.

# Belief (support) MDP

POMDPs induce **infinite belief MDPs**:



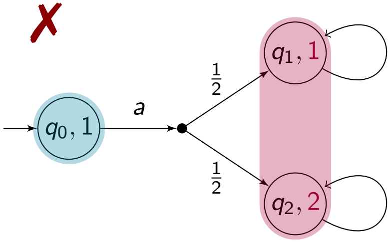

**Finite**: only keep belief **supports**:



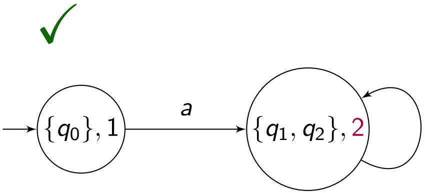

When does the analysis of the belief **support** MDP suffice?

# Non-soundness of the belief support MDP

**No** almost-sure strategy in the POMDP, but OK in the belief support MDP.




POMDP

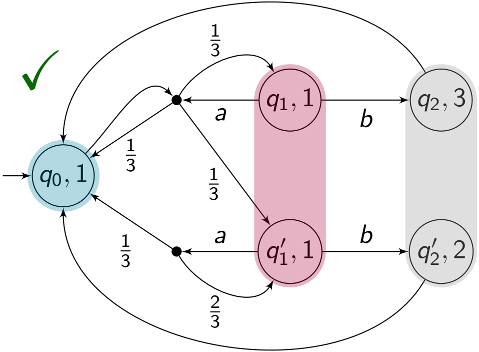



Belief support MDP

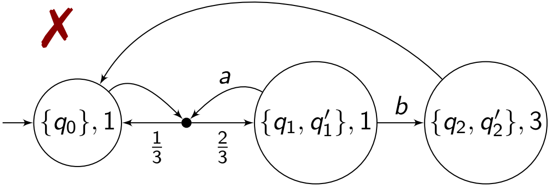
(Technical detail: how to lift the priority function? Take the **max**.)

# Incompleteness of the belief support MDP

Almost-sure strategy in the POMDP, **not** in the belief support MDP.




POMDP



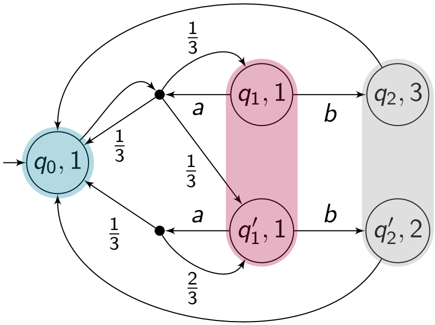

Belief support MDP

# First revealing property

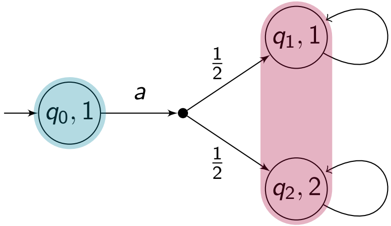
# First revealing property

## Property 1

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




Weakly revealing




Not weakly revealing

# First revealing property

## Property 1

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

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

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

Includes POMDPs that *reset* to the initial state with probability 1.

# Weakly revealing POMDPs

"Weakly revealing" is a semantic property:

## Deciding the property

Deciding whether a POMDP is **weakly revealing** is EXPTIME-hard and in 2-EXPTIME (**update**: actually EXPTIME-complete, WIP).

Let $\mathcal{P}$ be a **weakly revealing** POMDP with a parity objective.

## Soundness for **parity**

Almost-sure winning strategy in the **belief support MDP** of $\mathcal{P} \implies$ also in **POMDP** $\mathcal{P}$.

**Proof**: similar ideas to *decisiveness* (the "singletons" belief supports are a finite attractor).

## Completeness for priorities $\{0, 1, 2\}$

Almost-sure winning strategy in **POMDP** $\mathcal{P} \implies$ also in the **belief support MDP** of $\mathcal{P}$.

Analysing the belief support MDP is **sound** and **complete** for parity $\{0, 1, 2\}$.

# Decidability of weakly revealing POMDPs

### Decidability

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

**Algorithm**: solve the **belief support MDP** $\rightsquigarrow$ in EXPTIME.
**EXPTIME-hardness**: already for coBüchi; reduction from almost-sure safety in POMDPs.

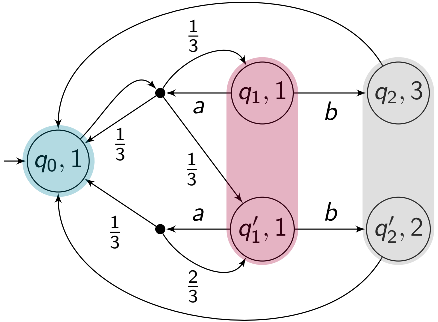
Compared to general POMDPs:
$\rightsquigarrow$ makes **coBüchi decidable**,
$\rightsquigarrow$ gives a (conceptually) **simpler algorithm** for Büchi (state space is $2^Q$, instead of $Q \times 2^Q$ in general[5]).

Exponential strategies ($2^Q \to \text{Act}$) suffice; this bound is tight.

[5] Baier, Größer, and Bertrand, "Probabilistic $\omega$-automata", 2012.

# Parity still not decidable

Belief support MDP is "incomplete" for this weakly revealing POMDP with priorities 1, 2, 3:




## Undecidability

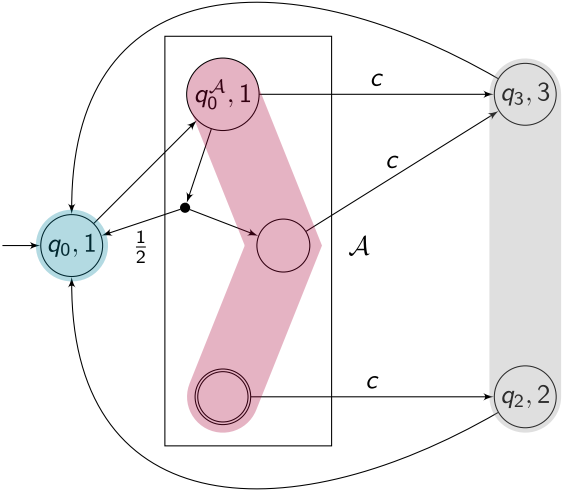
Almost-sure **parity** $\{1, 2, 3\}$ is **undecidable** for **weakly revealing** POMDPs.

Reduction from the value-1 problem for probabilistic automata.[6]

[6] Gimbert and Oualhadj, "Probabilistic Automata on Finite Words: Decidable and Undecidable Problems", 2010.

# Proof sketch

- We take a prob. automaton $\mathcal{A} = (Q^{\mathcal{A}}, \mathsf{Act}^{\mathcal{A}}, \delta^{\mathcal{A}}, q_0^{\mathcal{A}})$ with an accepting set $F$.
- We replace $\{q_1, q_1'\}$ by a copy of $\mathcal{A}$.
- We add a non-zero probability to go back to the initial state from every transition ($\rightsquigarrow$ weakly revealing).
- We add a new action $c$ that reaches $q_2$ if and only if we are in an accepting state of $\mathcal{A}$.
- $\mathcal{A}$ has value 1 $\iff$ there is an almost-sure strategy in the parity-$\{1, 2, 3\}$ POMDP.

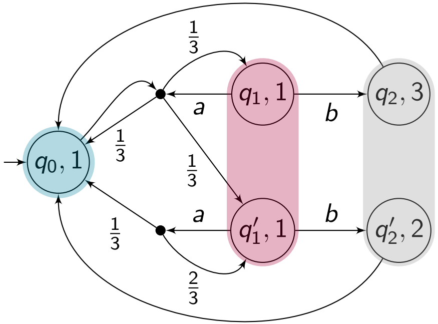


# Second revealing property

# Second revealing property

## Property 2

A POMDP is **strongly revealing** if for every transition $q \xrightarrow{a} q'$,
there is a **non-zero probability to see a signal that uniquely identifies $q'$**.

- Syntactic property.
- Strongly revealing $\Longrightarrow$ weakly revealing.
- 

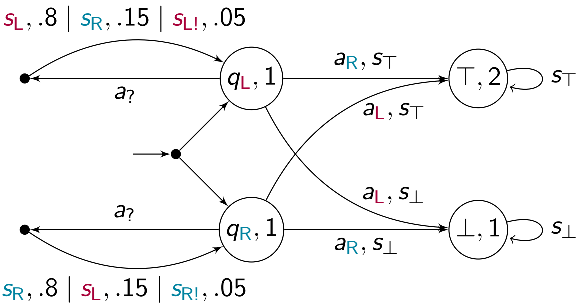
  Not strongly revealing: $q_1 \xrightarrow{a} q_1'$ is a possible transition, but nothing can reveal $q_1'$ with certainty.

# Second revealing property

## Property 2

A POMDP is **strongly revealing** if for every transition $q \xrightarrow{a} q'$,
there is a **non-zero probability to see a signal that uniquely identifies** $q'$.

- Syntactic property.
- Strongly revealing $\Longrightarrow$ weakly revealing.
- **Strongly revealing** variant of the Tiger example ("revealing signals" $s_{L!}$ and $s_{R!}$):



# Strongly revealing: results

### Completeness for **parity**

Almost-sure winning strategy in **strongly revealing POMDP** $\mathcal{P} \implies$ also in the **belief support MDP** of $\mathcal{P}$.

Soundness for full parity follows already from weakly revealing POMDPs.

### Theorem

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

Already EXPTIME-hard for coBüchi.

# Optimistic semantic

Another way to see the strongly revealing property:

### Optimistic semantic

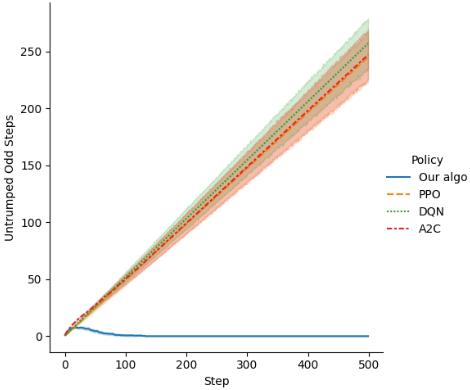
From a POMDP $\mathcal{P}$, one can define a related **strongly revealing** POMDP $\mathcal{P}_{\text{opt}}$ by adding a small probability of a revelation along all transitions.

### Proposition

If there is no almost-sure strategy in $\mathcal{P}_{\text{opt}}$, then this is also the case in $\mathcal{P}$.

# Empirical evaluation

- Classical algorithms (PPO, DQN, A2C)[7] for reinforcement learning in POMDPs do not solve the revealing tiger well.
- Not a completely fair comparison (e.g., model-based vs. model-free, parity vs. rewards), but indicates that more **structural observations** could be useful.




[7] Raffin et al., "Stable-Baselines3: Reliable Reinforcement Learning Implementations", 2021.

# Games with partial observation

The strongly revealing property seems very strong, but decidability frontier when we move to games:

**Theorem**

Almost-sure **coBüchi** for **strongly revealing** **games with partial information** is undecidable.

More complex reduction from the value-1 problem for probabilistic automata.

# Summary for POMDPs

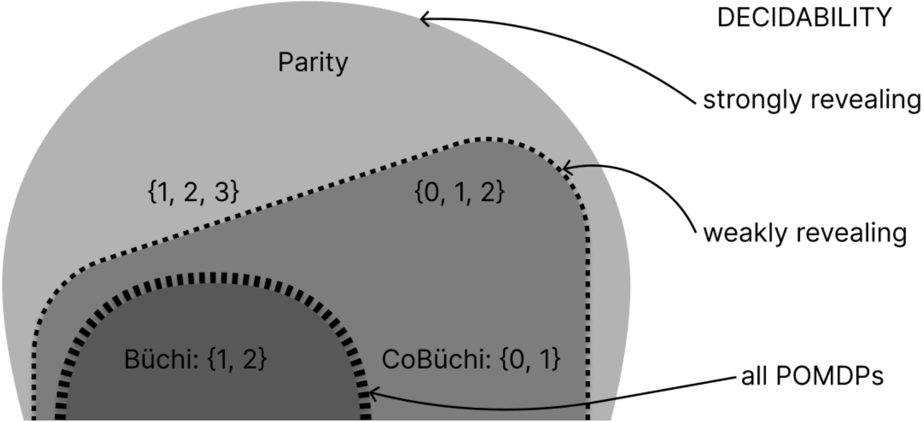



Figure: Decidable subclasses of the *parity* objective depending on the revelation mechanism.

# Related works

- Same philosophy: models with **sure** revelations (not just almost sure).[8]
  $\rightsquigarrow$ even **games** are decidable!
  To appreciate the difference, very different bounds on the frequency of revelations:

**Weakly revealing**:
for all $\sigma \in \Sigma(\mathcal{P})$,
$$\mathbb{P}_\sigma^{\mathcal{P}}\left[\mathsf{Reach}^{\leq 2^{|Q|}-1}(\mathsf{Revelations})\right] \geq \beta_{\mathcal{P}}^{2^{|Q|}-1}.$$

**Sure revelations**:
for all $\sigma \in \Sigma(\mathcal{P})$,
$$\mathbb{P}_\sigma^{\mathcal{P}}\left[\mathsf{Reach}^{\leq |Q|}(\mathsf{Revelations})\right] = 1.$$

- We study strategies $2^Q \to \mathsf{Act}$ and give sufficient conditions for their sufficiency.
  Similar studies exist for (less general) "memoryless" strategies $\mathsf{Obs} \to \mathsf{Act}$.[9]
- *Active-measuring POMDPs*: a cost may be paid to acquire additional information about the next state.[10]
- *Multi-environment MDPs*: multiple MDPs on the same state space with different transition functions.[11]

---

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

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

[10] 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.

[11] Raskin and Sankur, "Multiple-Environment Markov Decision Processes", 2014.

# Future works

**Open problems**:

- Larger class where the **belief support MDP** is sound and complete?
- Larger **decidable classes** for coBüchi/parity?
- **More general models** that the revealing mechanisms make decidable?

## Thanks!