# An Introduction to Probabilistic Models for Reactive Systems: Markov Decision Processes and Stochastic Games

**James C. A. Main**

UMONS – Université de Mons, Belgium


1st Meeting of the Probability FNRS Contact Group – October 17, 2025

# Overview

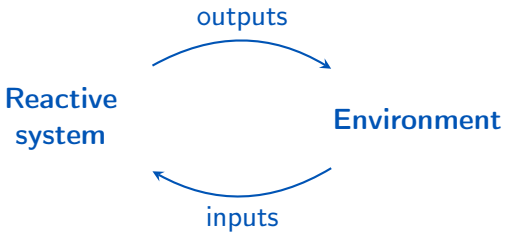
We focus on **Markov decision processes (MDPs)**, an extension of **Markov chains**.

**Structure of the talk**:

- Motivations of studying **(sequential) decision making**;
- MDPs and **strategies**;
- **Optimal reachability probabilities** in MDPs;
- MDPs with **multiple objectives**;
- Generalisation of MDPs.

# Formal verification

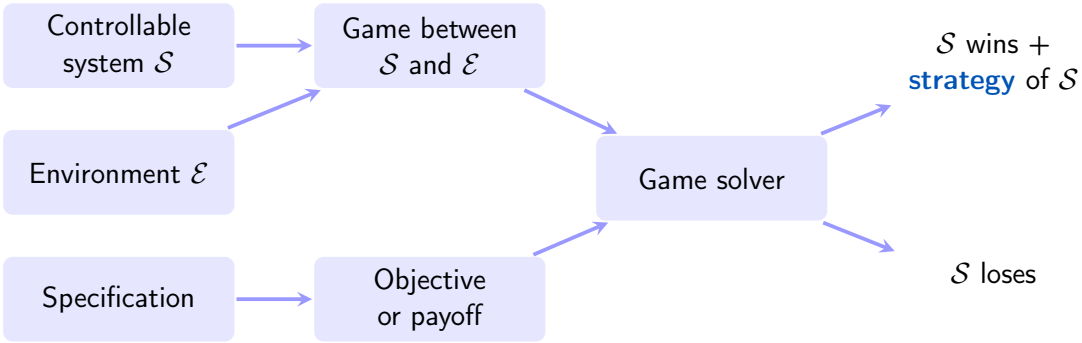
We are interested in analysing **reactive systems**.




**Model checking**: **formally verifying** that (a model of) a system **behaves well**.

# Reactive synthesis

What if we want to **design a well-performing controller** a reactive system?



# Table of contents

# Markov decision processes

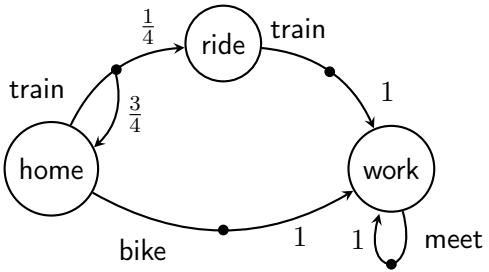
A **Markov decision process** is a tuple $\mathcal{M} = (S, A, \delta)$, where

- a **finite** set of **states** $S$,
- a **finite** set of **actions** $A$ and
- a **randomised transition** function $\delta \colon S \times A \to \mathcal{D}(S)$.

A **play** of an MDP is an **infinite path** along transitions.

**Ex.** home train home bike $(\text{work meet})^{\omega}$

**MDP example:** commuting to work



# Strategies

A **strategy** is a function $\sigma \colon (SA)^* S \to \mathcal{D}(A)$ describing the **possibly randomised** choices to be made depending on the **past history**.

Given a **strategy** $\sigma$ and an **initial state** $s$, we obtain a **distribution over plays of the MDP**.

Formally, for all histories $h = s_0 a_0 s_1 \dots a_{r-1} s_r$, we define

$$\mathbb{P}_s^\sigma(\mathsf{Cyl}\,(h)) = \prod_{\ell=0}^{r-1} \sigma(s_0 \dots s_\ell)(a_\ell) \cdot \delta(s_\ell, a_\ell)(s_{\ell+1}),$$

where $\mathsf{Cyl}\,(h)$ is the set of plays extending $h$.

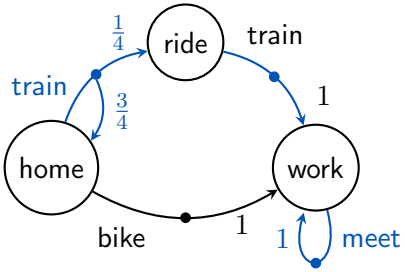
The **expectation** with respect to $\mathbb{P}_s^\sigma$ is denoted $\mathbb{E}_s^\sigma$.

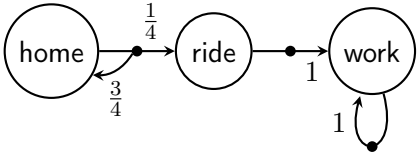
# Memoryless strategies

A strategy is **memoryless** if it makes decisions based on the **current state**.

A memoryless strategy can be seen as a function $\sigma : S \to A$.

**Markov chains** induced by **memoryless strategies** can be seen as Markov chains over $S$.



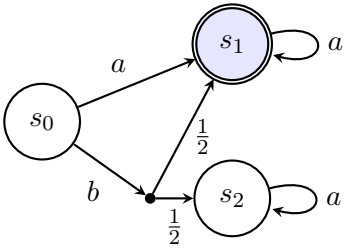



# Objectives

We formalise **qualitative** specifications via **objectives**.

> **Objective**: measurable subset of Plays($\mathcal{M}$).

**Reachability objective**: given $T \subseteq S$, Reach($T$) is the set of plays visiting $T$.




> A strategy $\sigma$ is **optimal** for an objective $\Omega$ from $s_0 \in S$ if $\mathbb{P}^{\sigma}_{s_0}(\Omega) = \sup_{\tau} \mathbb{P}^{\tau}_{s_0}(\Omega)$.

# Payoffs

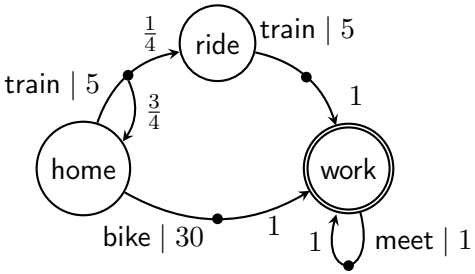
We formalise **quantitative** specifications via **payoff functions**.

**Payoff**: measurable function $f\colon \mathsf{Plays}(\mathcal{M}) \to \bar{\mathbb{R}}$.

**Shortest-path function**: given $T \subseteq S$ and $w\colon S \times A \to [0, +\infty[$, we let $\mathsf{SPath}_w^T(\pi)$ be

$$\begin{cases} +\infty & \text{if } \pi \notin \mathsf{Reach}(T) \\ \sum_{\ell=0}^{r-1} w(s_\ell, a_\ell) & \text{else, where } r = \min\{\ell \mid s_\ell \in T\} \end{cases}$$

for all plays $\pi = s_0 a_0 s_1 \ldots \in \mathsf{Plays}(\mathcal{M})$.




A strategy $\sigma$ is **optimal** for a payoff $f$ from $s_0 \in S$ if $\mathbb{E}_{s_0}^{\sigma}(f) = \sup_\tau \mathbb{E}_{s_0}^{\tau}(f)$.

# Algorithms and strategy complexity

What kind of **problems** do we study?

> **Reactive synthesis problem (for MDPs).** Design algorithms to **automatically construct** well-performing strategies for a **given class of payoffs**.

Furthermore, we are interested in **simple strategies**.

- When do **pure memoryless strategies** suffice to play optimally?
- **How much memory** do we need to play optimally?
- What kind of **randomisation** do we need for a given specification?

# Table of contents

# Reachability

**Why focus on reachability?**

- Reachability is a **central specification** in formal methods.
- The analysis of several objectives either **relies on reachability** or **uses analogous techniques**.

How can we **characterise** the **supremum probability** of a reachability objective from an initial state?

Do **optimal strategies** always exist?

- If they do, how **complex** are they?
- How can they be **computed** if they exist?

# Values in reachability

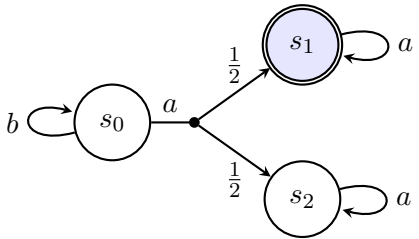
Let $T \subseteq S$ be a target. For all $s \in S$, let $\mathrm{val}(s) = \sup_\sigma \mathbb{P}_s^\sigma(\mathrm{Reach}(T))$ denote the **value** of $s$.

**Theorem**. The vector $(\mathrm{val}(s))_{s \in S}$ is the **least solution** of the equation system

$$x_s = \begin{cases} 1 & \text{if } s \in T, \\ 0 & \text{if } T \text{ is not reachable from } s, \\ \max_{a \in A(s)} \sum_{s' \in S} \delta(s,a)(s') \cdot x_{s'} & \text{otherwise.} \end{cases}$$

This system may not have a **unique solution**.

$\rightarrow$ The value vector is $(\frac{1}{2}, 1, 0)$ is a solution, but **so is** $(1,1,0)$.



# Establishing the value equations

## Least fixed point formulation

We want to show that $(\mathsf{val}(s))_{s \in S}$ is the **least fixed point** of the operator $\mathcal{F}\colon [0,1]^S \to [0,1]^S$ defined, for all $\mathbf{v} = (v_j)_{j \in [\![1,d]\!]} \in [0,1]^S$, by

$$(\mathcal{F}(\mathbf{v}))_s = \begin{cases} 1 & \text{if } s \in T \\ \max_{a \in A(s)} \sum_{s' \in S} \delta(s,a)(s') \cdot v_s & \text{else.} \end{cases}$$

$\mathcal{F}$ has several nice properties:

- for all $\mathbf{v}, \mathbf{w} \in [0,1]^S$, if $\mathbf{v} \leq \mathbf{w}$, then $\mathcal{F}(\mathbf{v}) \leq \mathcal{F}(\mathbf{w})$;
- $\mathcal{F}$ is continuous.

**Lemma**. The **least fixed point** of $\mathcal{F}$ is given by $\lim_{k \to \infty} \mathcal{F}^k(\mathbf{0})$.

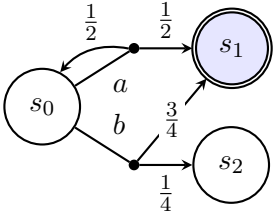
# Establishing the value equations

## Least fixed point formulation

What does the sequence $(\mathcal{F}^k(\mathbf{0}))_{k\in\mathbb{N}}$ represent?

**Example**: focus on $s_0$.

- $\mathcal{F}^1(\mathbf{0})_{s_0} = 0 = \max_\sigma \mathbb{P}^\sigma_{s_0}(\text{Reach}^{\leq 0}(T))$;
- $\mathcal{F}^2(\mathbf{0})_{s_0} = \frac{3}{4} = \max_\sigma \mathbb{P}^\sigma_{s_0}(\text{Reach}^{\leq 1}(T))$;
- $\mathcal{F}^3(\mathbf{0})_{s_0} = \frac{7}{8} = \max_\sigma \mathbb{P}^\sigma_{s_0}(\text{Reach}^{\leq 2}(T))$.




The sequence is related to **step-bounded reachability**. For all $k \in \mathbb{N}$, we let

- $\text{Reach}^{\leq k}(T) = \{s_0 a_0 s_1 \ldots \in \text{Plays}(\mathcal{M}) \mid \exists \ell \leq k,\ s_\ell \in T\}$;
- $\text{val}^{\leq k}(s) = \sup_\sigma \mathbb{P}^\sigma_s(\text{Reach}^{\leq k}(T))$ for all $s \in S$.

# Step-bounded reachability

**Lemma**. For all $k \in \mathbb{N}$ and $s \in S$, $\mathsf{val}^{\leq k}(s) = \mathcal{F}^{k+1}(\mathbf{0})_s$.

**Proof by induction**. Case $k = 0$ is direct. Assume the lemma holds for $k \in \mathbb{N}$.

For all strategies $\sigma$ and all states $s \in S$, we have

$$\begin{aligned}
\mathbb{P}_s^\sigma(\mathsf{Reach}^{\leq k+1}(T)) &= \sum_{a \in A(s)} \sigma(s)(a) \cdot \sum_{s' \in S} \delta(s,a)(s') \cdot \mathbb{P}_s^\sigma(\mathsf{Reach}^{\leq k+1}(T) \mid sas') \\
&\leq \sum_{a \in A(s)} \sigma(s)(a) \cdot \sum_{s' \in S} \delta(s,a)(s') \cdot \mathsf{val}^{\leq k}(s') \\
&\leq \mathcal{F}^{k+2}(\mathbf{0})_s.
\end{aligned}$$

It follows that $\mathsf{val}^{\leq k+1}(s) \leq \mathcal{F}^{k+2}(\mathbf{0})_s$.

# Step-bounded reachability

**Lemma**. For all $k \in \mathbb{N}$ and $s \in S$, $\mathsf{val}^{\leq k}(s) = \mathcal{F}^{k+1}(\mathbf{0})_s$.

**Proof by induction**. Case $k = 0$ is direct. Assume the lemma holds for $k \in \mathbb{N}$.

To reach $T$ within $k+1$ steps with probability at least $\mathcal{F}^{k+2}(\mathbf{0})_s$ from $s \in S$, it suffices to select an action in the set

$$\operatorname*{argmax}_{a \in A(s)} \sum_{s' \in S} \delta(s,a)(s') \cdot \mathcal{F}^{k+1}(\mathbf{0})_{s'}$$

in $s$ when $k+1$ steps remain and follow an **optimal strategy** for $k$ steps after.

This implies $(\mathsf{val}^{\leq k+1}(s))_{s \in S} \geq \mathcal{F}((\mathsf{val}^{\leq k}(t))_{t \in S}) = \mathcal{F}^{k+2}(\mathbf{0})$. □

# Putting it all together

**Theorem**. The vector $(\mathsf{val}(s))_{s \in S}$ is the **least solution** of the equation system

$$x_s = \begin{cases} 1 & \text{if } s \in T, \\ 0 & \text{if } T \text{ is not reachable from } s, \\ \max_{a \in A(s)} \sum_{s' \in S} \delta(s,a)(s') \cdot x_{s'} & \text{otherwise.} \end{cases}$$

**Proof**. It suffices to show that for all $s \in S$, $\mathsf{val}(s) = \lim_{k \to \infty} \mathsf{val}^{\leq k}(s)$.

On the one hand, for all strategies $\sigma$ and $s \in S$, we have

$$\mathbb{P}_s^\sigma(\mathsf{Reach}(T)) = \lim_{k \to \infty} \mathbb{P}_s^\sigma(\mathsf{Reach}^{\leq k}(T)) \leq \lim_{k \to \infty} \mathsf{val}^{\leq k}(s).$$

# Putting it all together

**Theorem**. The vector $(\mathsf{val}(s))_{s \in S}$ is the **least solution** of the equation system

$$x_s = \begin{cases} 1 & \text{if } s \in T, \\ 0 & \text{if } T \text{ is not reachable from } s, \\ \max_{a \in A(s)} \sum_{s' \in S} \delta(s,a)(s') \cdot x_{s'} & \text{otherwise.} \end{cases}$$

**Proof**. It suffices to show that for all $s \in S$, $\mathsf{val}(s) = \lim_{k \to \infty} \mathsf{val}^{\leq k}(s)$.

For all strategies $\sigma$, and all $s \in S$ and $k \in \mathbb{N}$, we have

$$\mathbb{P}_s^\sigma(\mathsf{Reach}^{\leq k}(T)) \leq \mathbb{P}_s^\sigma(\mathsf{Reach}(T)) \leq \mathsf{val}(s).$$

It follows that $\mathsf{val}(s) \geq \lim_{k \to \infty} \mathsf{val}^{\leq k}(s)$. □

# Constructing optimal strategies

**Theorem**. There exist **pure memoryless uniformly optimal strategies** for reachability in MDPs.

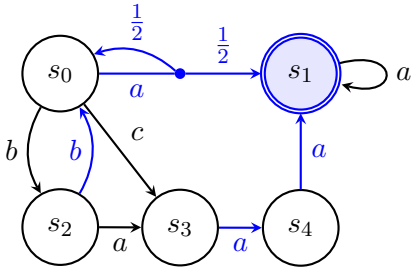
We construct strategies using the equation, for $s \in S \setminus T$,

$$\mathsf{val}(s) = \max_{a \in A(s)} \sum_{s' \in S} \delta(s,a)(s') \cdot \mathsf{val}(s').$$

For each state $s \in S$, we select an action such that

- the action **witnesses the maximum above**;
- if $\mathsf{val}(s) > 0$, there is a **path from $s$ to $T$** in the induced Markov chain.

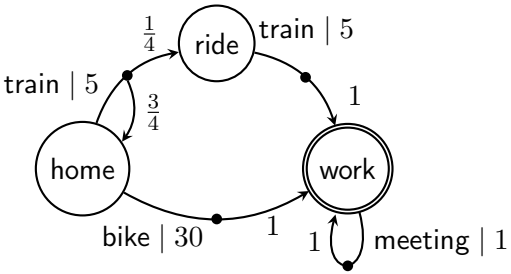
Any such strategy is **uniformly optimal**.

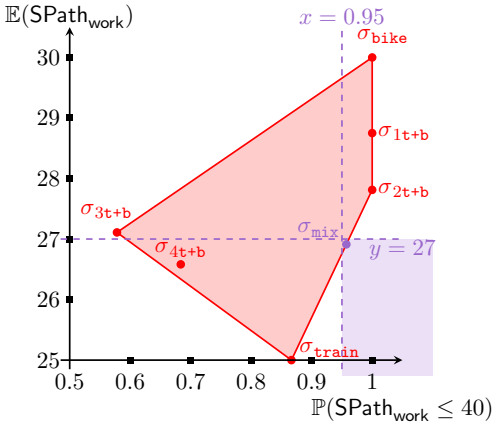


# Table of contents

# MDPs with multiple objectives

Some applications require **optimising multiple goals**.



# Multi-objective MDPs

We study MDPs with **multi-dimensional payoff functions** $\bar{f}$: Plays$(\mathcal{M}) \to \bar{\mathbb{R}}^d$.

Achieving some payoff vectors may require both **memory** and **randomisation**.

We focus on **randomisation** in this framework.

- What is the **purpose of randomisation** in this framework?
- What is the relationship between payoffs of **pure** and **randomised** strategies?

For a **multi-dimensional payoff** $\bar{f} = (f_1, \ldots, f_d)$ and $s \in S$, we study:

- $\mathsf{Pay}_s(\bar{f}) = \{\mathbb{E}_s^\sigma(\bar{f}) \mid \sigma \text{ strategy}\}$;
- $\mathsf{Pay}_s^{\mathsf{pure}}(\bar{f}) = \{\mathbb{E}_s^\sigma(\bar{f}) \mid \sigma \text{ pure strategy}\}$.

# Restrictions on payoffs

The expectation of a payoff from a state under a strategy is **not always well-defined**.

**Which payoffs $f$ do we consider?**

- ▷ A payoff $f$ is **good** if it has a **well-defined expectation** under all strategies from all initial states.
- ▷ A payoff $f$ is **universally integrable** if its expectation is finite under all strategies from all initial states.

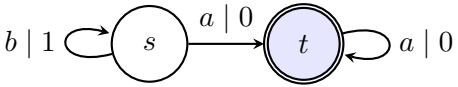
# The structure of payoff sets

## Example

In our initial example, we had

$$\mathsf{Pay}_{\mathsf{home}}(\bar{f}) = \mathsf{conv}(\mathsf{Pay}^{\mathsf{pure}}_{\mathsf{home}}(\bar{f}))$$

**When does this generalise?**




For all $\pi = s_0 a_0 s_1 \ldots \in \mathsf{Plays}(\mathcal{M})$, let

$$f(\pi) = \begin{cases} \sum_{\ell \in \mathbb{N}} w(s_\ell, a_\ell) & \text{if } \pi \in \mathsf{Reach}(t) \\ 0 & \text{otherwise.} \end{cases}$$

We have

- $\mathsf{Pay}^{\mathsf{pure}}_s(f) = [0, +\infty[$;
- $\mathsf{Pay}_s(f) = [0, +\infty]$.

**The property does not apply to all good payoffs.**

# The structure of payoff sets

## Results

**Theorem (M., Randour 2025).** If $\bar{f}$ is **universally integrable**, then for all $s \in S$, $\mathsf{Pay}_s(\bar{f}) = \mathsf{conv}(\mathsf{Pay}_s^{\mathsf{pure}}(\bar{f}))$.

For **good payoffs**, we can obtain a weaker result.

**Theorem (M., Randour 2025).** If $\bar{f}$ is **good**, then for all $s \in S$, $\mathsf{cl}(\mathsf{Pay}_s(\bar{f})) = \mathsf{cl}(\mathsf{conv}(\mathsf{Pay}_s^{\mathsf{pure}}(\bar{f})))$.

These results imply that **limited randomisation** is sufficient in multi-objective MDPs.

The **role of randomisation** in multi-objective MDPs is mainly to **balance the different objectives**.

# Table of contents

# Multiple sources of non-determinism

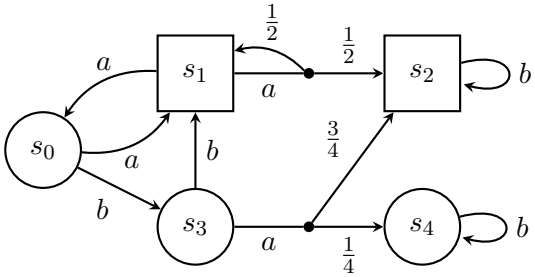
MDPs model the interaction of a decision maker or **player** with a **fully stochastic environment**.

If the environment is not fully stochastic, we can model it as **another player**.

A **two-player stochastic game** is a tuple $\mathcal{G} = (S_1, S_2, A, \delta)$ where

- $S_1$ and $S_2$ partition the **finite state space** $S$;
- $A$ and $\delta$ are defined as in MDPs.

**Strategies** of $\mathcal{P}_i$ are functions $\sigma_i \colon (SA)^* S_i \to \mathcal{D}(A)$.



# Values in two-player zero-sum games

We consider **zero-sum games**: the two players have **opposite goals**.

Given a payoff $f$:

- the **supremum expected value** that $\mathcal{P}_1$ can ensure from $s \in S$ is $\sup_{\sigma_1} \inf_{\sigma_2} \mathbb{E}_s^{\sigma_1,\sigma_2}(f)$;
- the **infimum expected value** that $\mathcal{P}_2$ can ensure from $s$ is $\inf_{\sigma_2} \sup_{\sigma_1} \mathbb{E}_s^{\sigma_1,\sigma_2}(f)$.

If $\sup_{\sigma_1} \inf_{\sigma_2} \mathbb{E}_s^{\sigma_1,\sigma_2}(f) = \inf_{\sigma_2} \sup_{\sigma_1} \mathbb{E}_s^{\sigma_1,\sigma_2}(f)$, we call this number the **value of** $s$.

**When is the value well-defined?**

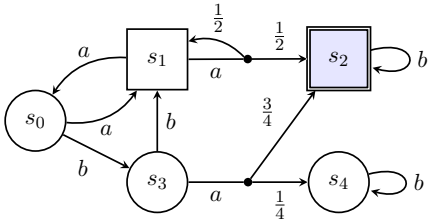
**Theorem (determinacy).** If $f$ is bounded, then the value is **well-defined** in all states.

# Two-player reachability games

**Theorem**. The value vector in a zero-sum reachability game with target $T$ is the **least solution** of the equation system given by, for $s \in S$,

$$x_s = \begin{cases} 1 & \text{if } s \in T \\ \max_{a \in A(s)} \sum_{s' \in S} \delta(s,a)(s') \cdot x_{s'} & \text{if } s \in S_1 \setminus T \\ \min_{a \in A(s)} \sum_{s' \in S} \delta(s,a)(s') \cdot x_{s'} & \text{if } s \in S_2 \setminus T. \end{cases}$$

- Both players have **uniformly optimal pure memoryless** strategies.
- We can prove this by adapting our arguments for MDPs.



# Table of contents

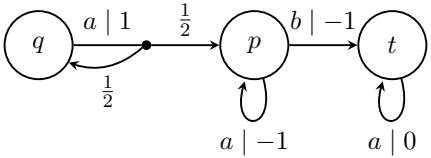
# One-counter MDPs

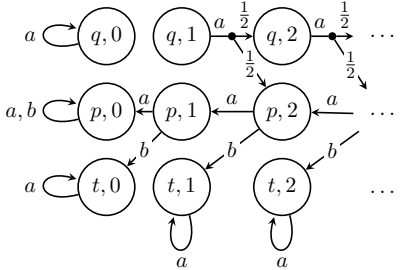
**One-counter MDP (OC-MDP) $\mathcal{Q}$:**

- **Finite** MDP $(Q, A, \delta)$.
- **Weight function** $w \colon Q \times A \to \{-1, 0, 1\}$.

**MDP $\mathcal{M}^{\leq\infty}(\mathcal{Q})$ induced by $\mathcal{Q}$:**

- **Countable** MDP over $S = Q \times \mathbb{N}$.
- State transitions via $\delta$.
- Counter updates via $w$.



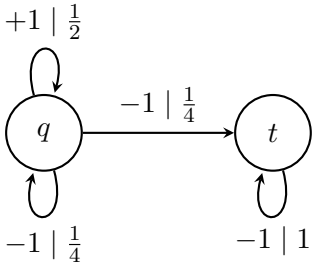



# Analysing OC-MDPs

The **selective termination objective** is defined as Reach$(T \times \{0\})$ for some $T \subseteq Q$.

- Specific algorithms are required due to the **infinite state space**.
- Relevant probabilities may be **irrational**.

In **(Ajdarów, M., Novotný, Randour 2025)**, we provide algorithms to verify **concisely represented counter-based strategies** in OC-MDPs.




The probability of terminating in $t$ from $(q, 1)$ in this one-counter Markov chain is $\frac{\sqrt{2}}{2}$.

# Conclusion

We have introduced **MDPs**, some **extensions** and **results**.

There exist many more extensions:

- there can be **more than two players**;
- **time** can be included (as in continuous-time Markov chains or like timed automata)
- players can be made to act **simultaneously**;
- the players can be **imperfectly informed**;
- ...

Thank you for your attention!

# References – books

**Reference on model checking including MDPs:**

Christel Baier and Joost-Pieter Katoen. *Principles of model checking*. MIT Press, 2008. ISBN: 978-0-262-02649-9.

**Reference on games on graphs, including MDPs and stochastic games**

Nathanaël Fijalkow (ed.). *Games on Graphs: From Logic and Automata to Algorithms*. Cambridge University Press, 2026.

# References (II) – cited papers

Michal Ajdarów, James C. A. Main, Petr Novotný and Mickael Randour. *Taming Infinity one Chunk at a Time: Concisely Represented Strategies in One-Counter MDPs*. In: Proceedings of the 52th International Colloquium on Automata, Languages, and Programming, ICALP 2025, July 8–11, 2025, Aarhus, Denmark. Ed. by Keren Censor-Hillel et al. Vol. 334. LIPIcs. Schloss Dagstuhl – Leibniz-Zentrum für Informatik, 2025.

James C. A. Main and Mickael Randour. *Mixing Any Cocktail with Limited Ingredients: On the Structure of Payoff Sets in Multi-Objective POMDPs and its Impact on Randomised Strategies*. In: CoRR abs/2502.18296 (2025). DOI: 10.48550/arXiv.2502.18296.