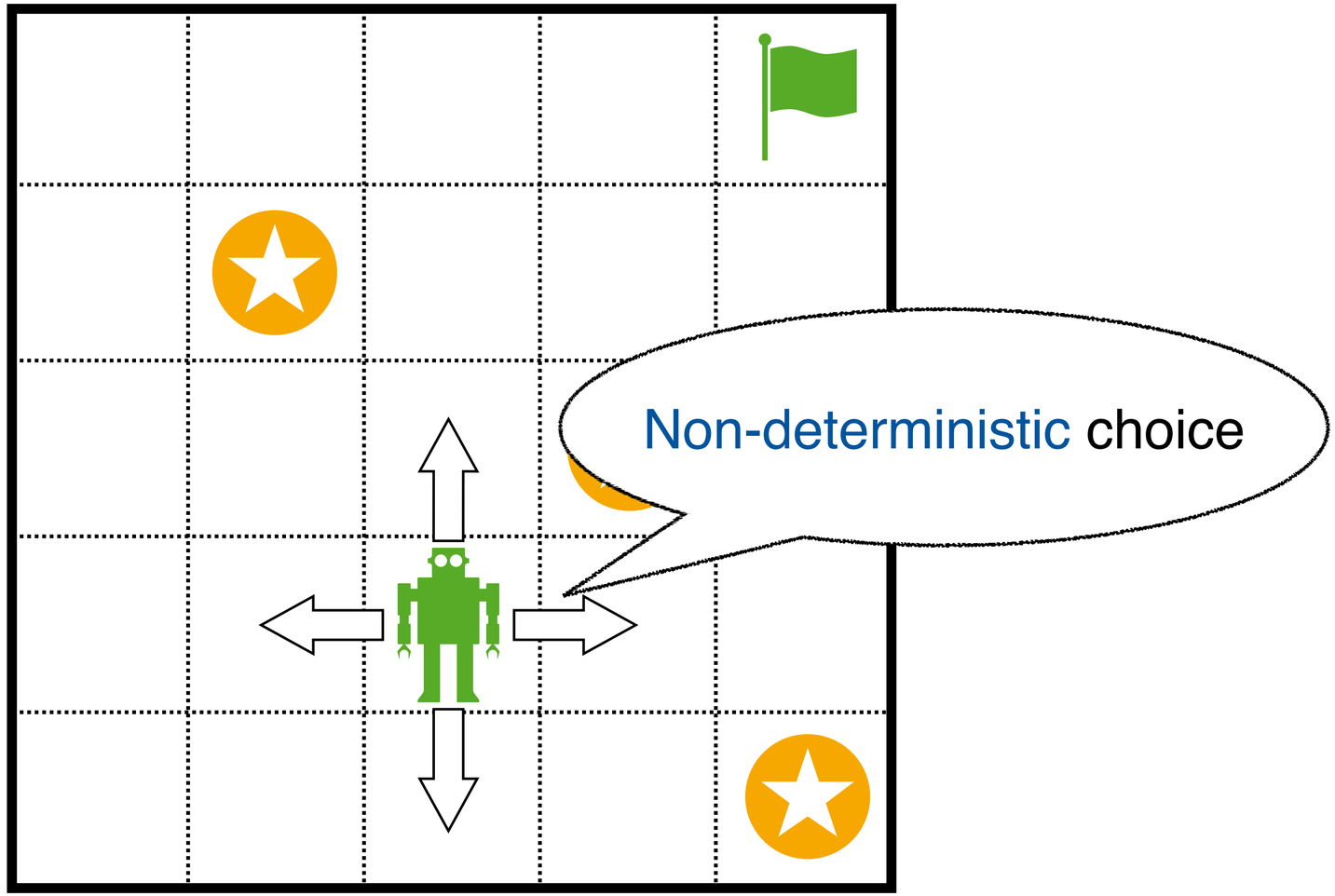
Advances in SAYNT
Symbiotic Policy Synthesis in POMDPs

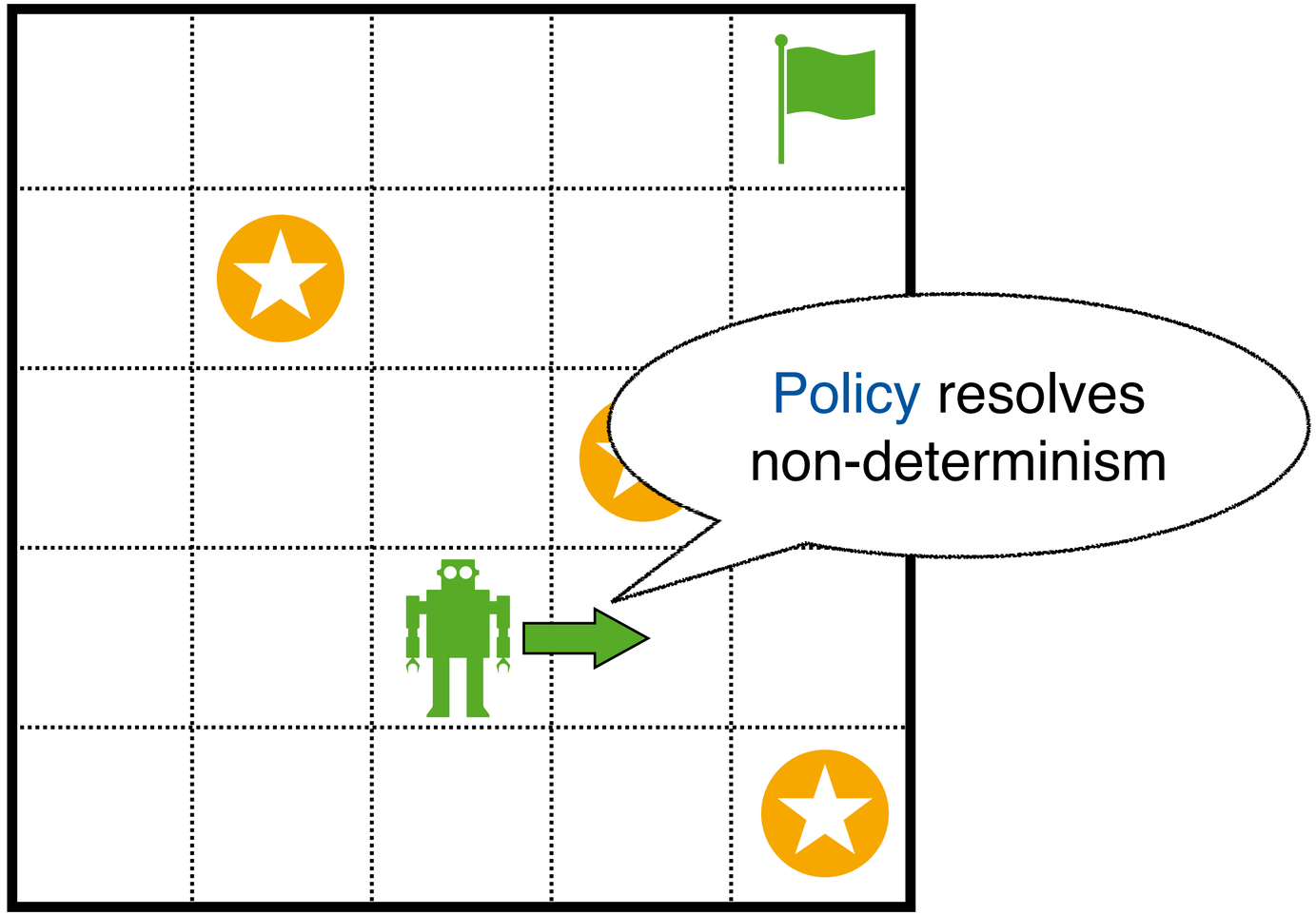
Roman Andriushchenko[1], **Alexander Bork[2]**, Milan Češka[1], Sebastian Junges[3], Joost-Pieter Katoen[2], Fillip Macák[1]

[1] Brno University of Technology, CZ
[2] RWTH Aachen University, DE
[3] Radboud University Nijmegen, NL

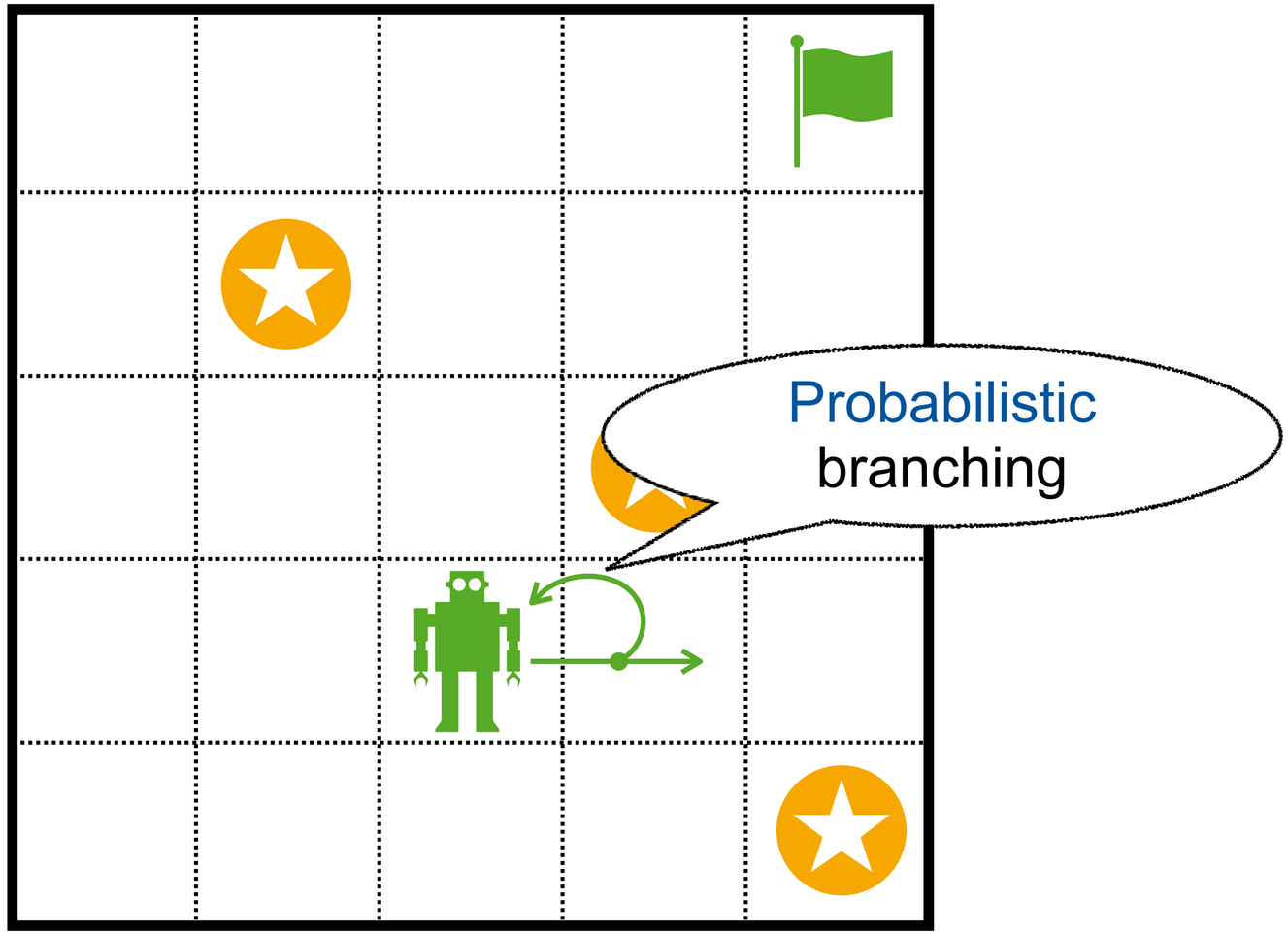
Motivation



Motivation

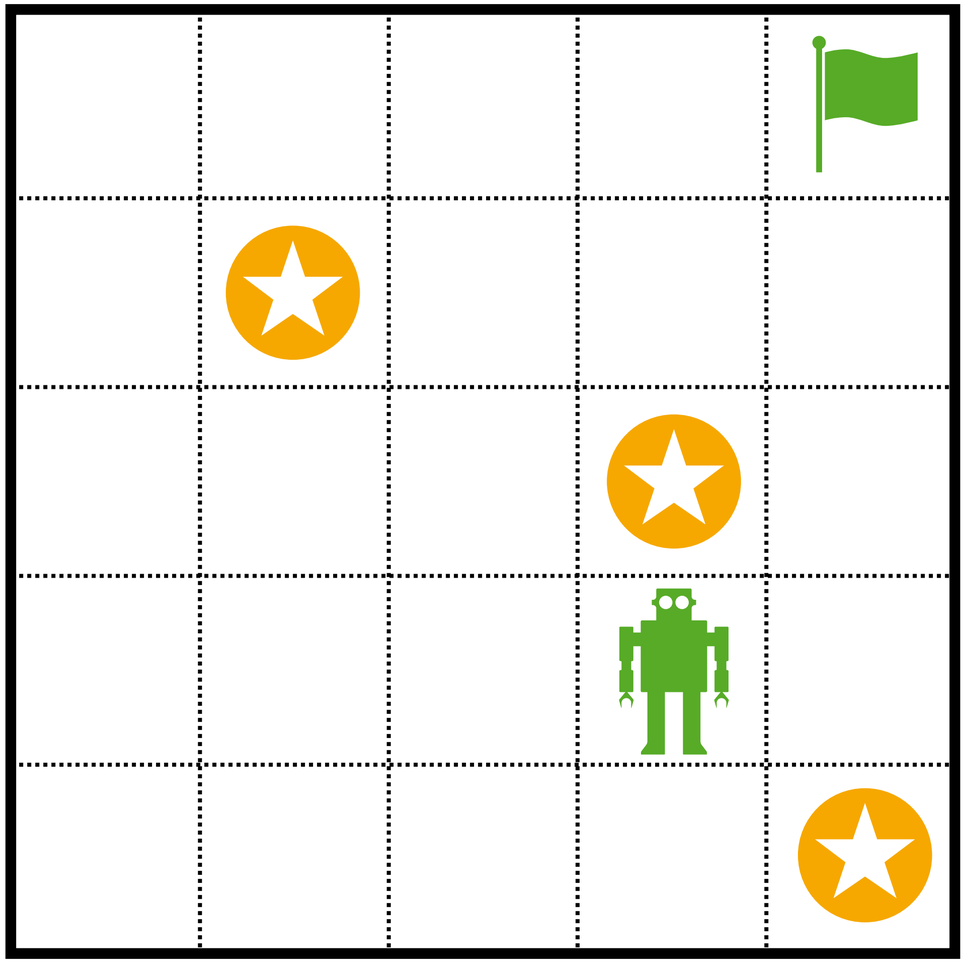


Motivation



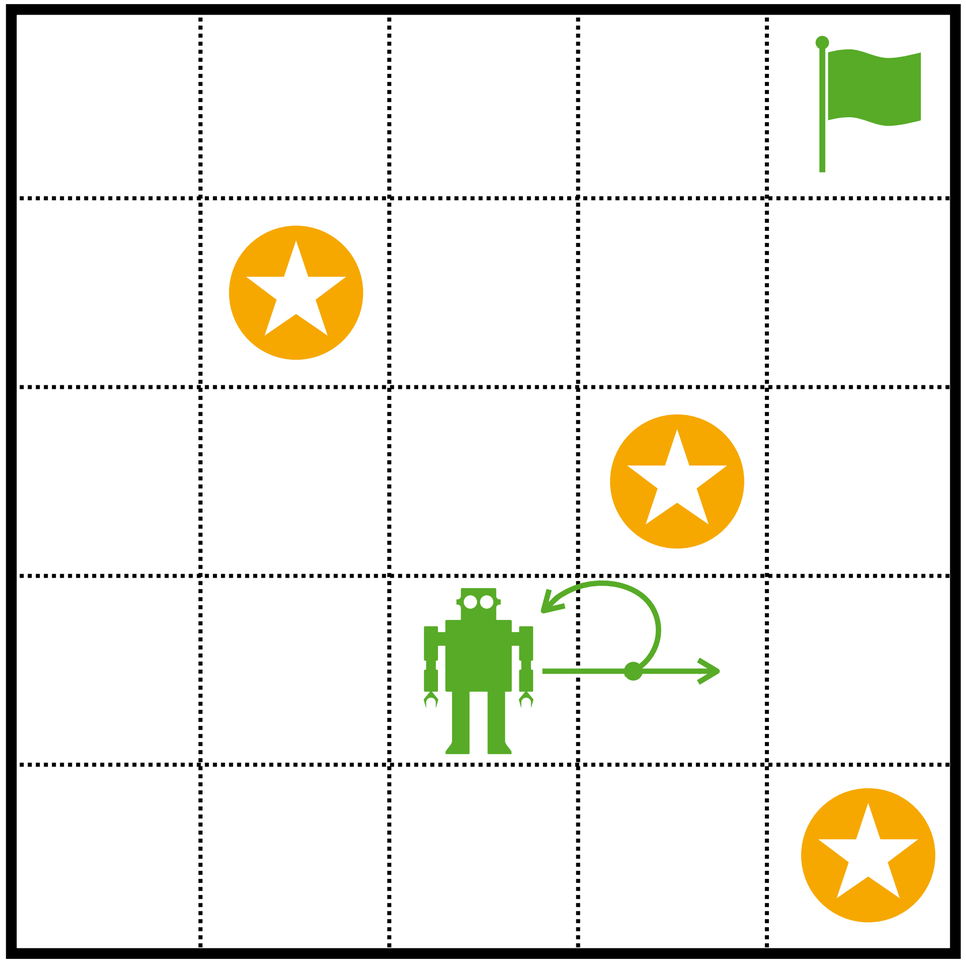
Motivation

Possible state after action



Motivation — MDP

MDPs are a pivotal model for decision making under uncertainty



Markov Decision Process (MDP)

- Non-deterministic choice
- Probabilistic branching

Rewards

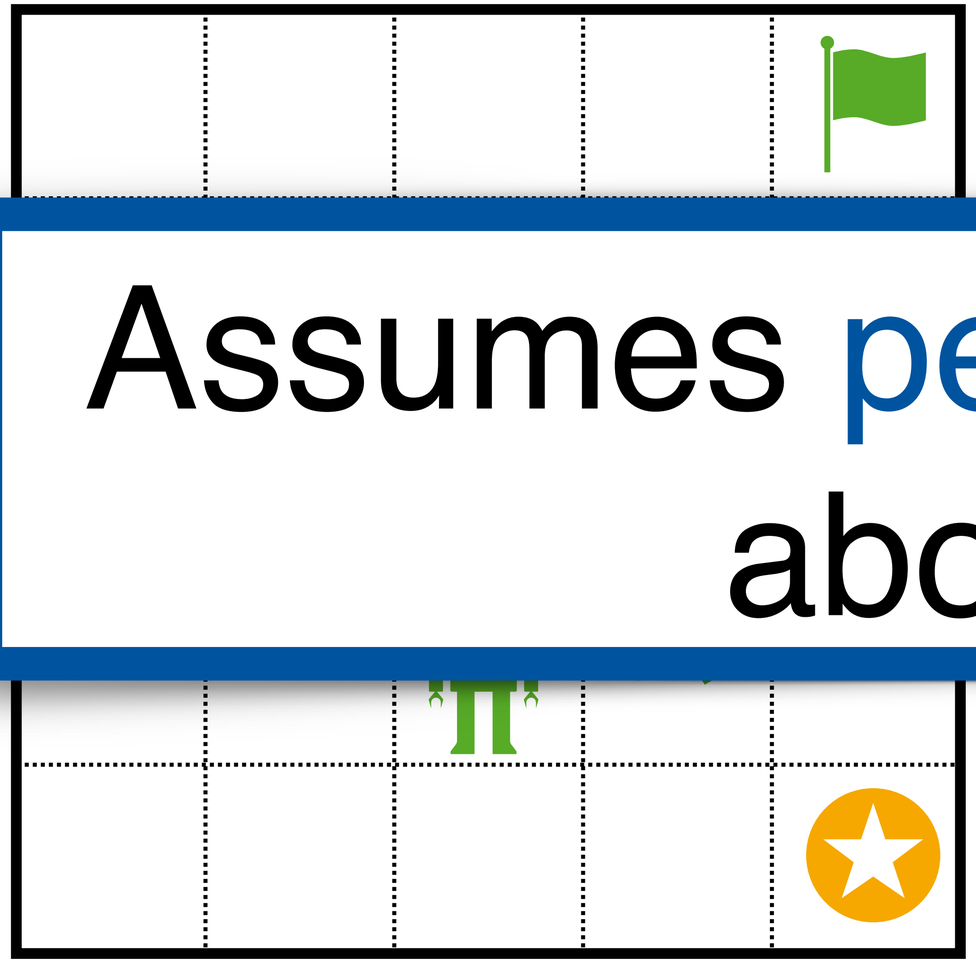
- Used to model steps, costs, …
- Collected when taking a transition

Policy

- Resolves non-determinism
- Maximising/minimising reachability objective:
  Only state-dependent, no memory necessary

Motivation — MDP

MDPs are a pivotal model for decision making under uncertainty

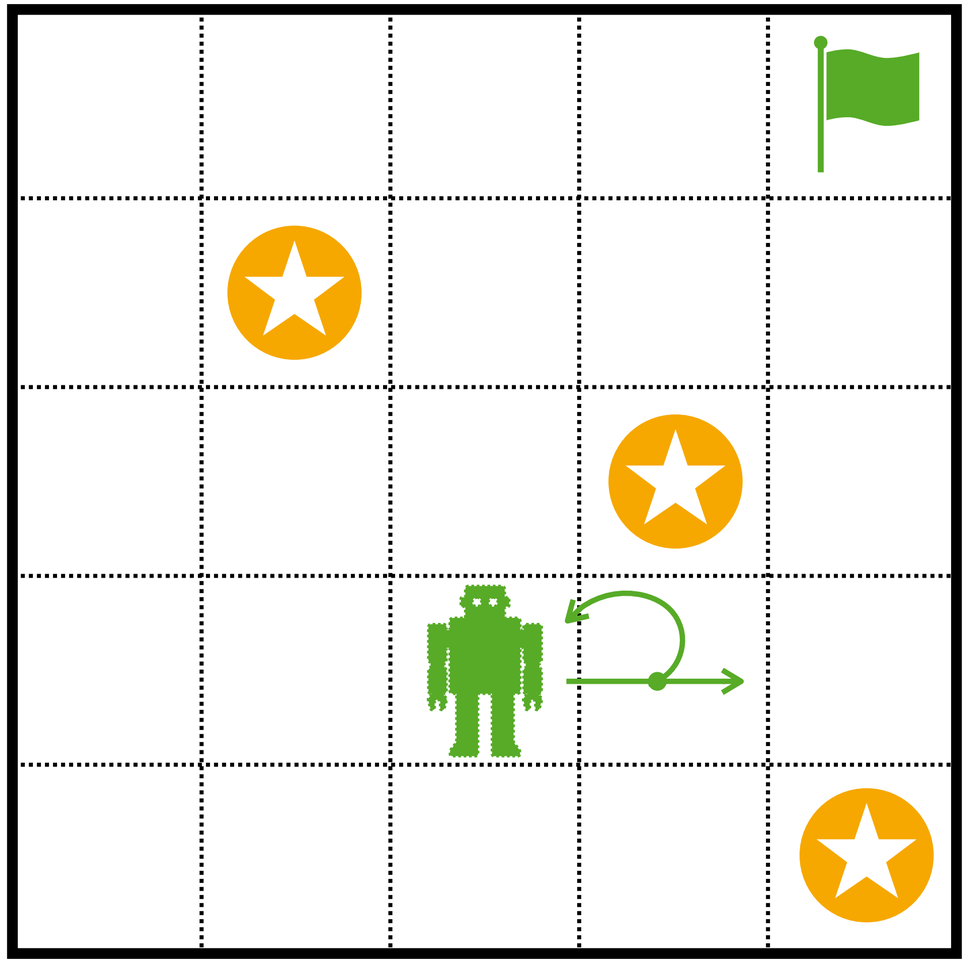


Markov Decision Process (MDP)

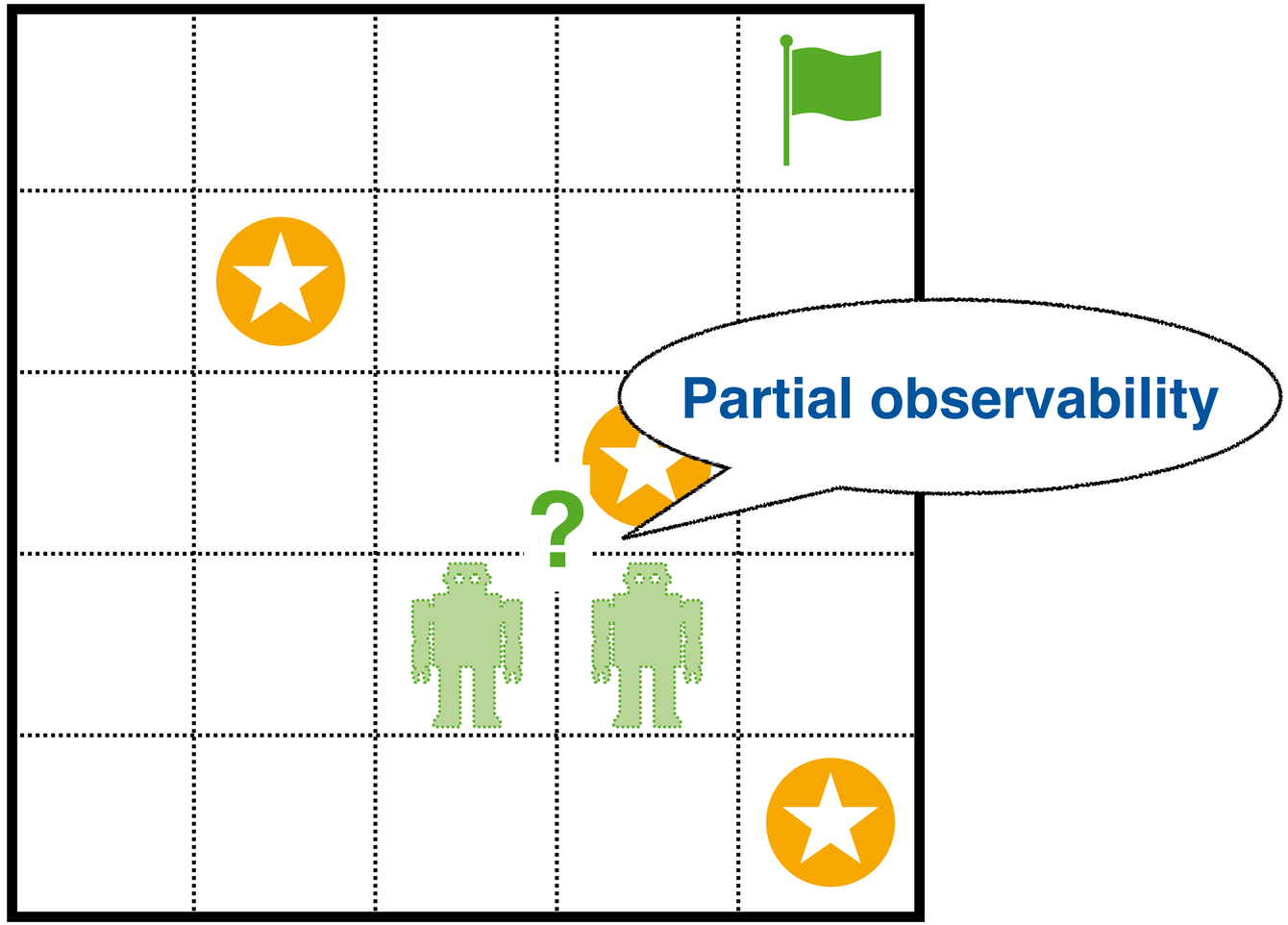
- Non-deterministic choice
- Maximising/minimising reachability objective:
  Only state-dependent, no memory necessary

Assumes perfect information about state!

Motivation

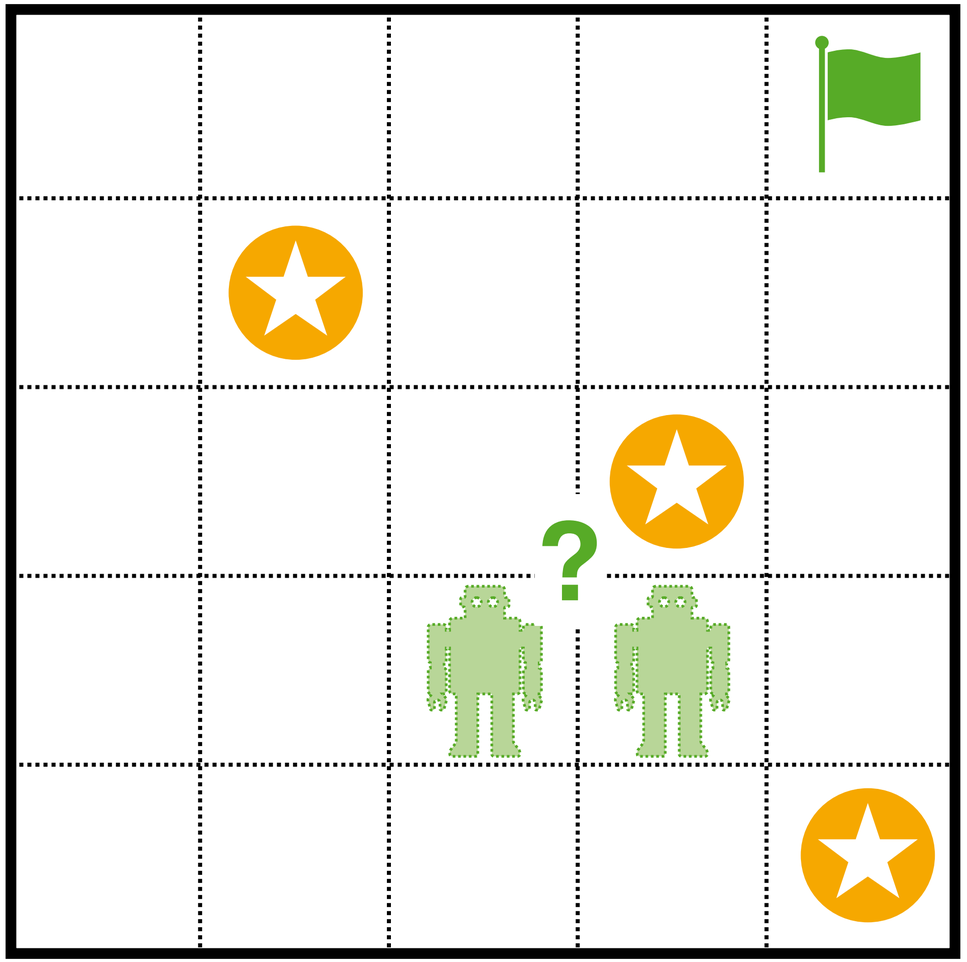


Motivation



Motivation — POMDP

POMDPs play an important role for planning in AI



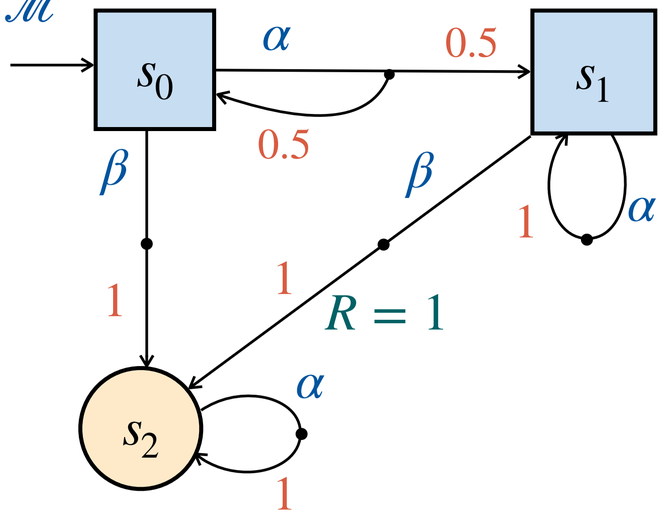
Partially Observable MDP (POMDP)

- Extension by observation labels

Observation-based policy

- Decisions using observable state information
- **Memory is crucial!**

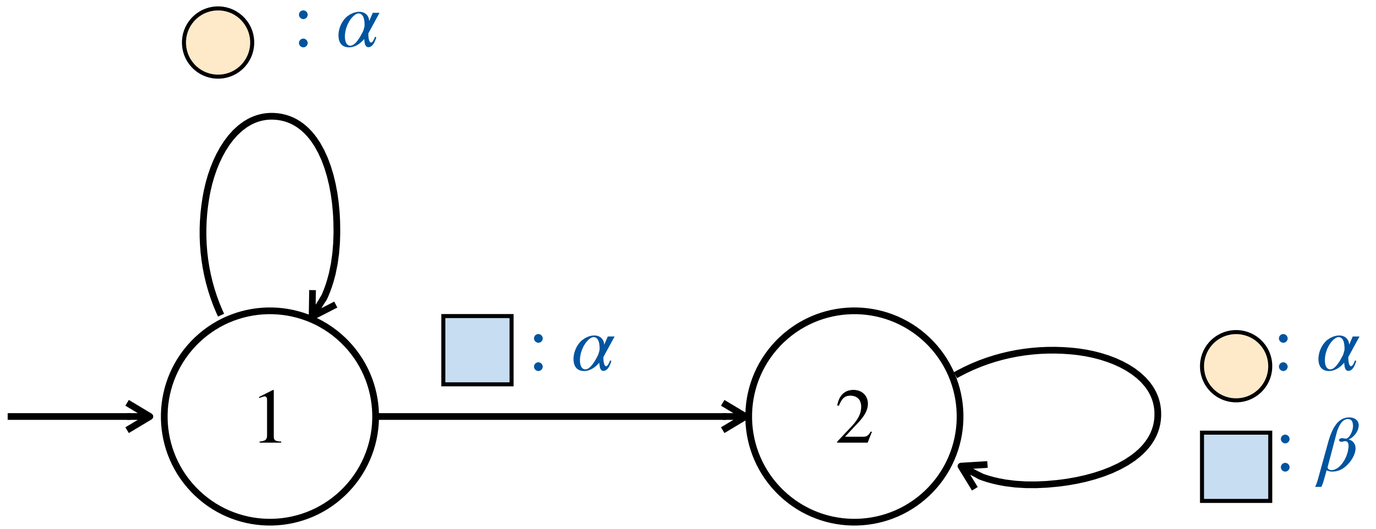
POMDP $\mathscr{M}$



Policy Synthesis

- Goal: find policy that maximises expected total reward
  - reward collected along all paths until goal state is reached
- Undiscounted and infinite time horizon

- Optimal policy might not exist
  ➔ Synthesise good policies (Finite State Controllers)



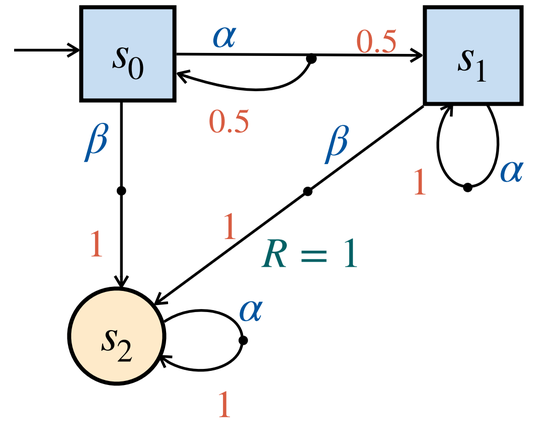
Symbiotic Policy Synthesis

- **PAYNT: Inductive Policy Synthesis**
  - Synthesise FSC directly
  - Use induced MC for value approximation

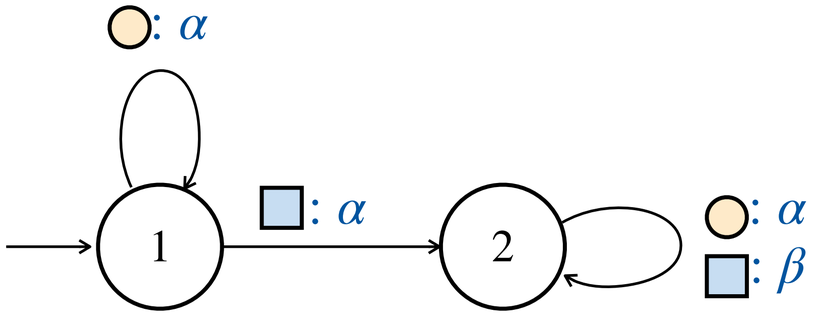
- **STORM: Belief Exploration**
  - Construct belief model
  - Compute policy using model checking
  - Obtain controller from computed policy

- **SAYNT: Symbiotic Approach**

POMDP $\mathscr{M}$



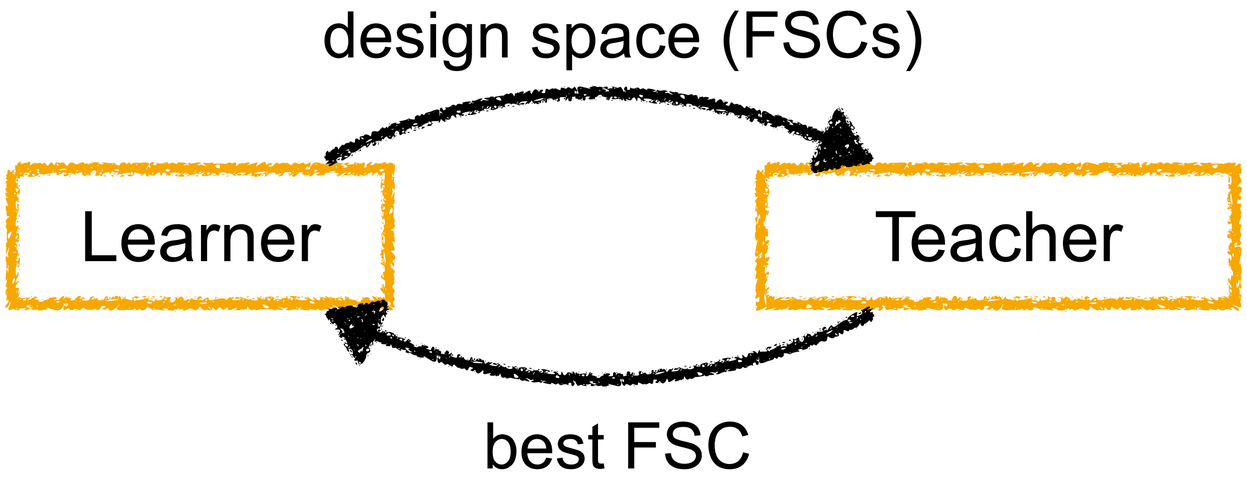
FSC F



Inductive Synthesis for POMDPs — Outer Loop

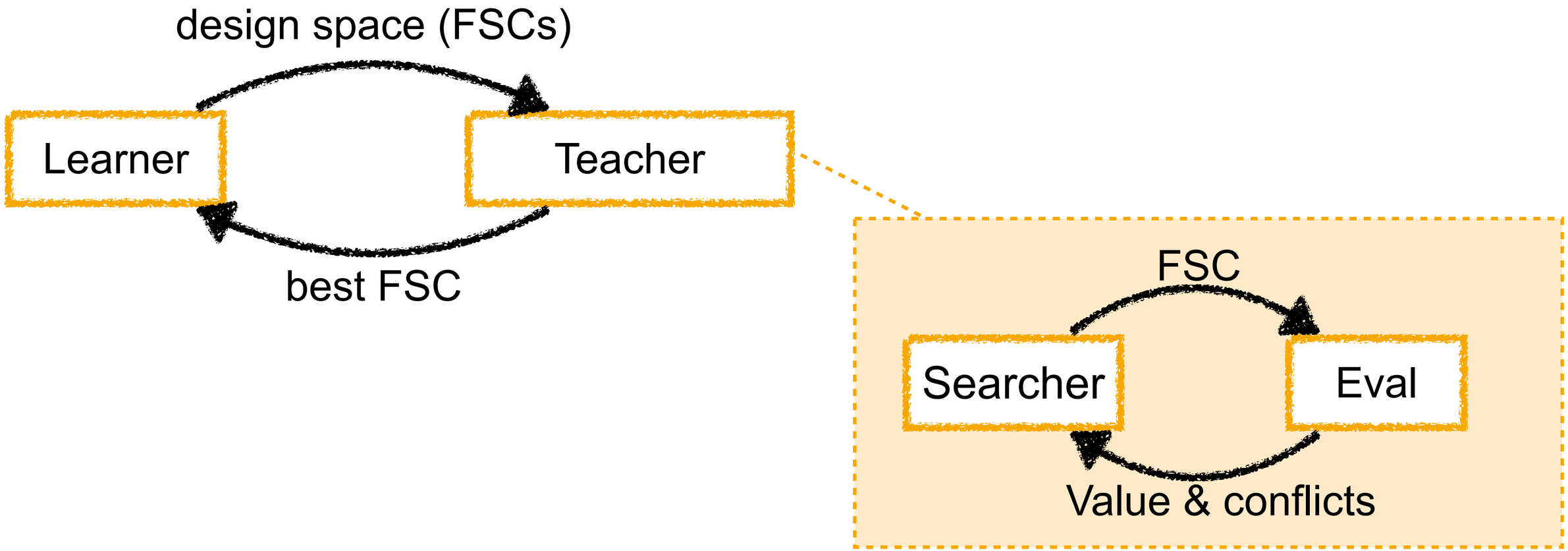
[Andriushchenko et al. 2022]

- Goal: learn deterministic FSC
- Limiting factor: design space size
- Access to oracle can improve design space



Inductive Synthesis for POMDPs

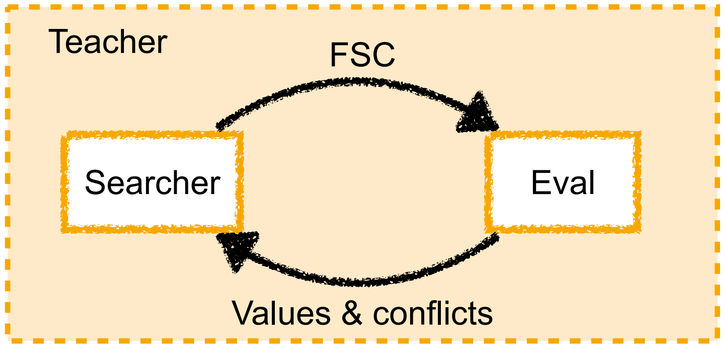
[Andriushchenko et al. 2022]

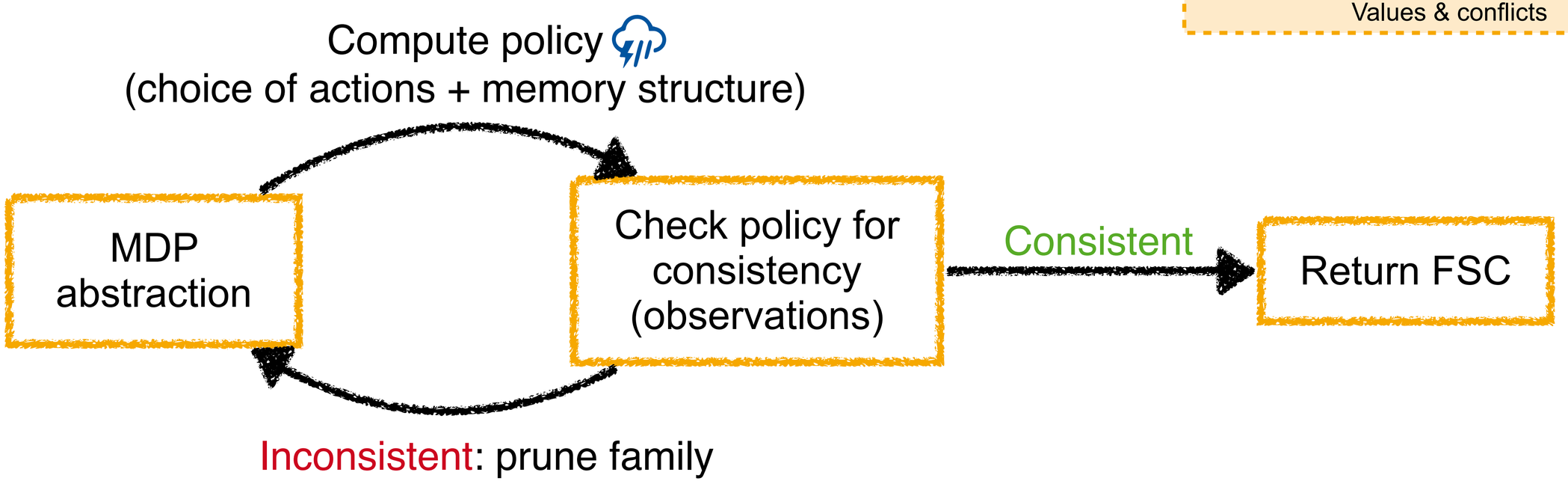


Inductive Synthesis for POMDPs — Inner Loop

[Andriushchenko et al. 2022]

- Teacher gets family of k-FSC
  - FSC parameterised in action-choice and memory transitions
- MDP abstraction of family of induced MCs





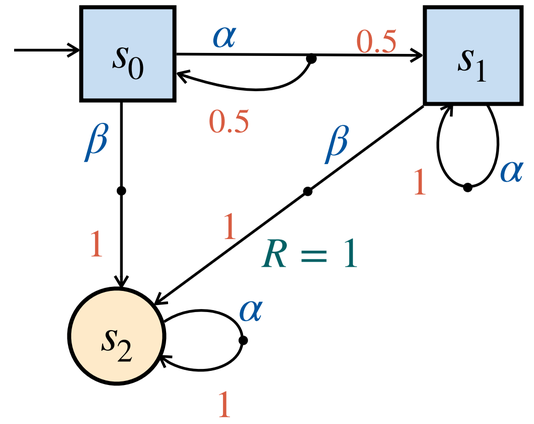
Symbiotic Policy Synthesis

- **PAYNT: Inductive Policy Synthesis**
  - Synthesise a finite state controller
  - Use induced MC for value approximation

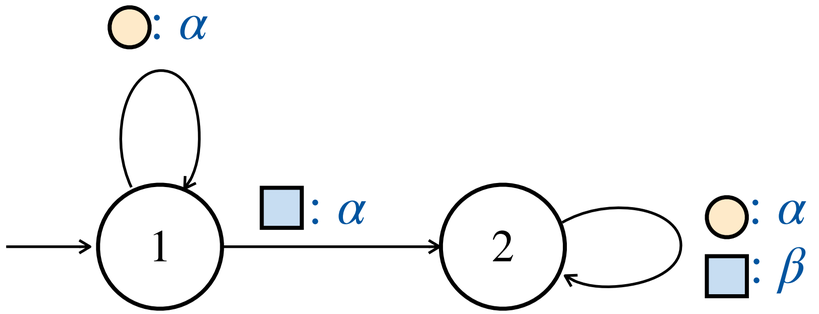
- **STORM: Belief Exploration**
  - Construct belief model
  - Compute policy using model checking
  - Obtain controller from computed policy

- **SAYNT: Symbiotic Approach**

POMDP $\mathscr{M}$



FSC F



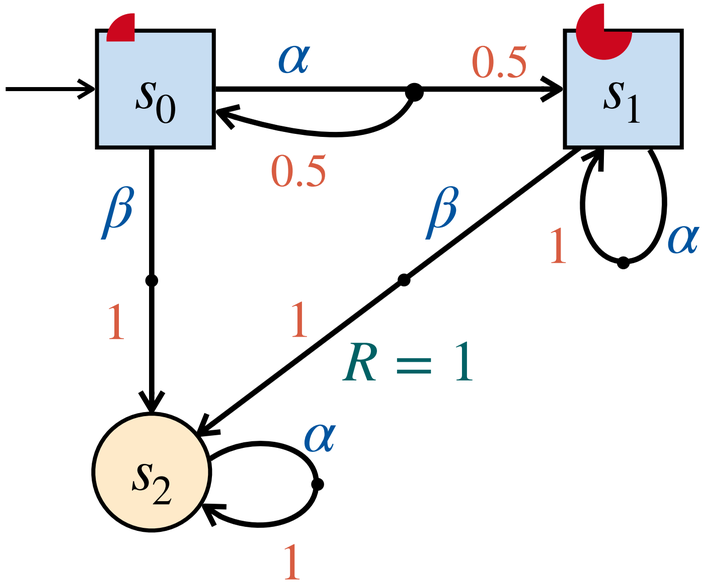
POMDP Semantics — Belief MDP

[Åström, 1965]

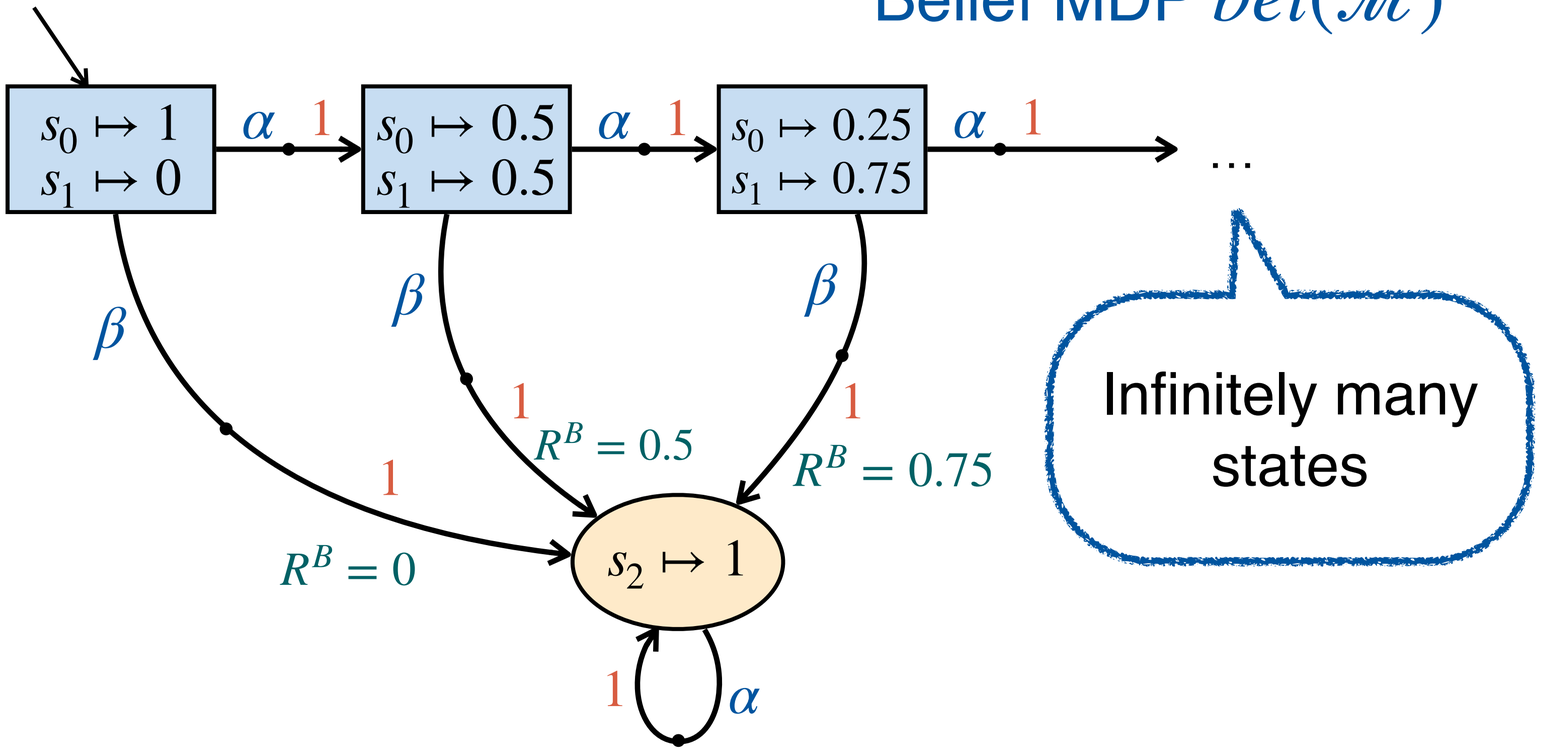
Belief

- Distribution over POMDP states
- Describes likelihood to be in state given observation history

POMDP $\mathscr{M}$



Belief MDP $bel(\mathscr{M})$



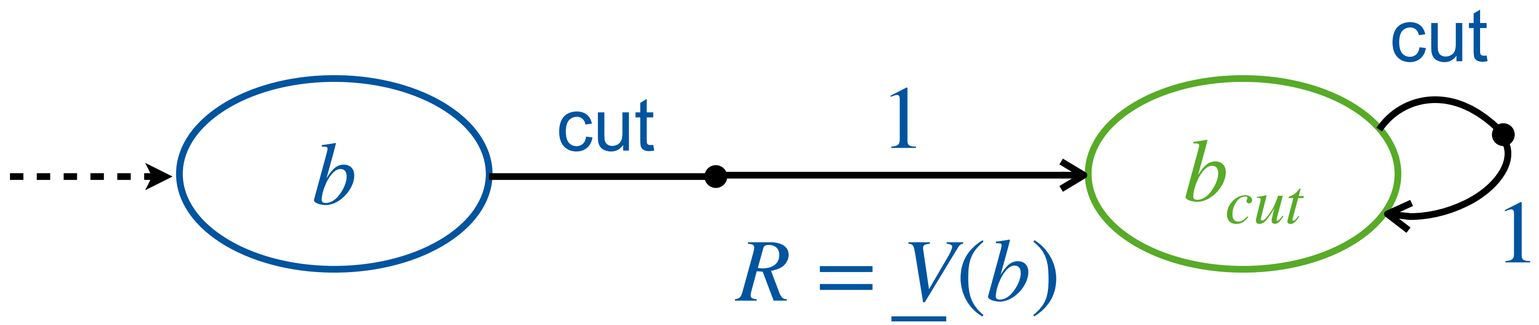
Belief Exploration with Cut-Offs

[B., Katoen, Quatmann, 2022]

- Obtain finite MDP for model checking
- Explore part of belief space, approximate values (Cut-Offs)

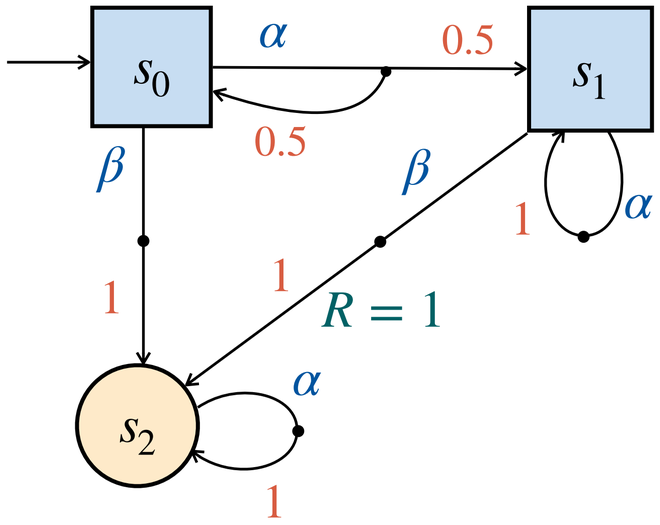
- Approximation: based on some policy for POMDP

- Weight values by belief distribution, add goal transition + approx. reward

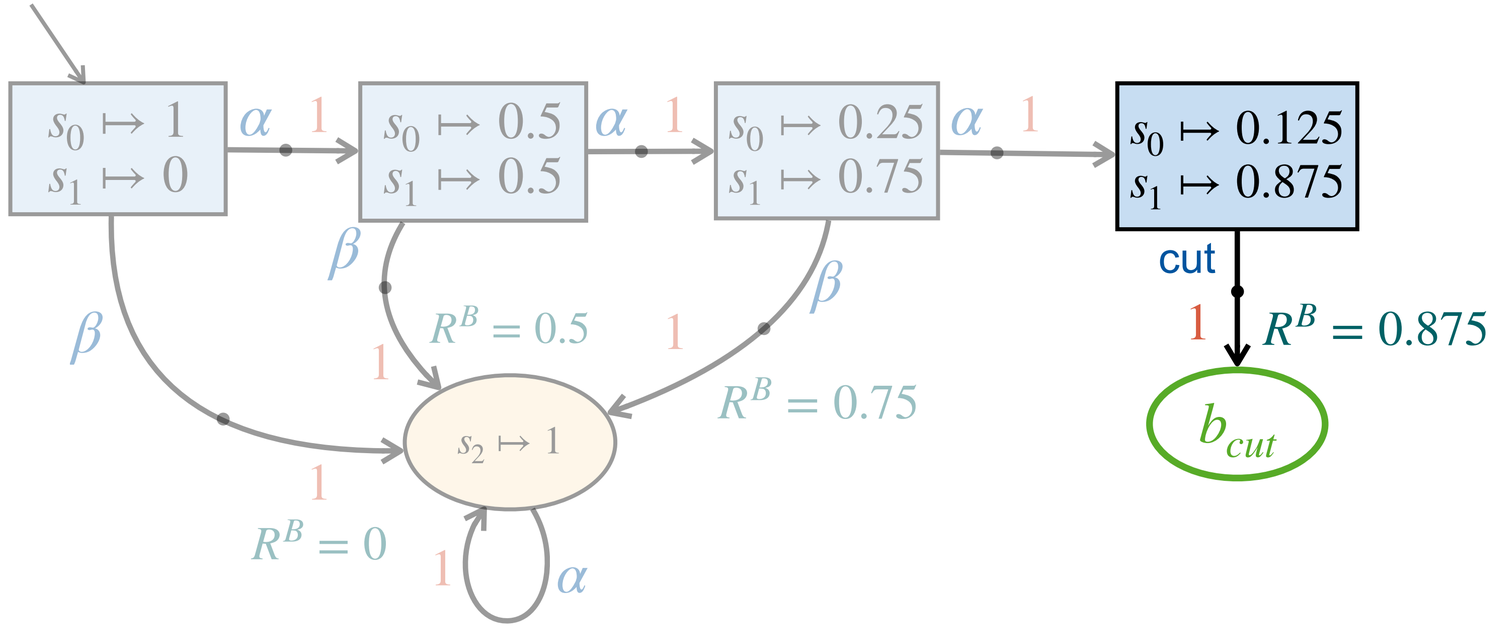


Belief Exploration — Example

POMDP $\mathscr{M}$



Cut-Off MDP



$$\sigma_{cut}(s) = \begin{cases} \alpha, & \text{if } O(s) = \bigcirc \\ \beta, & \text{otherwise} \end{cases}$$

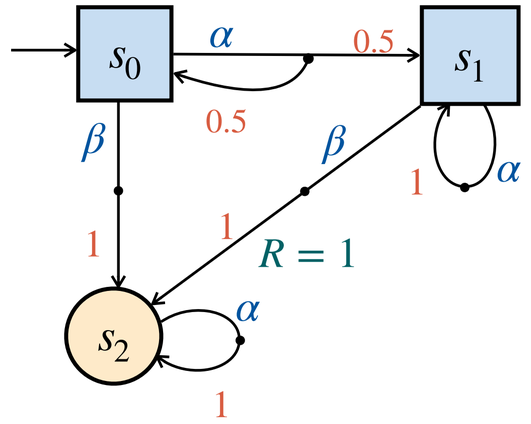
Symbiotic Policy Synthesis

- **PAYNT: Inductive Policy Synthesis**
  - Synthesise finite state controller
  - Use induced MC for value approximation

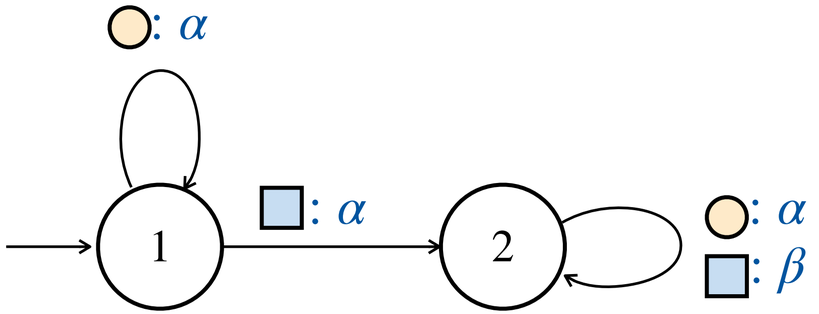
- **STORM: Belief Exploration**
  - Construct belief model
  - Compute policy using model checking
  - Obtain controller from computed policy

- **SAYNT: Symbiotic Approach**

POMDP $\mathscr{M}$

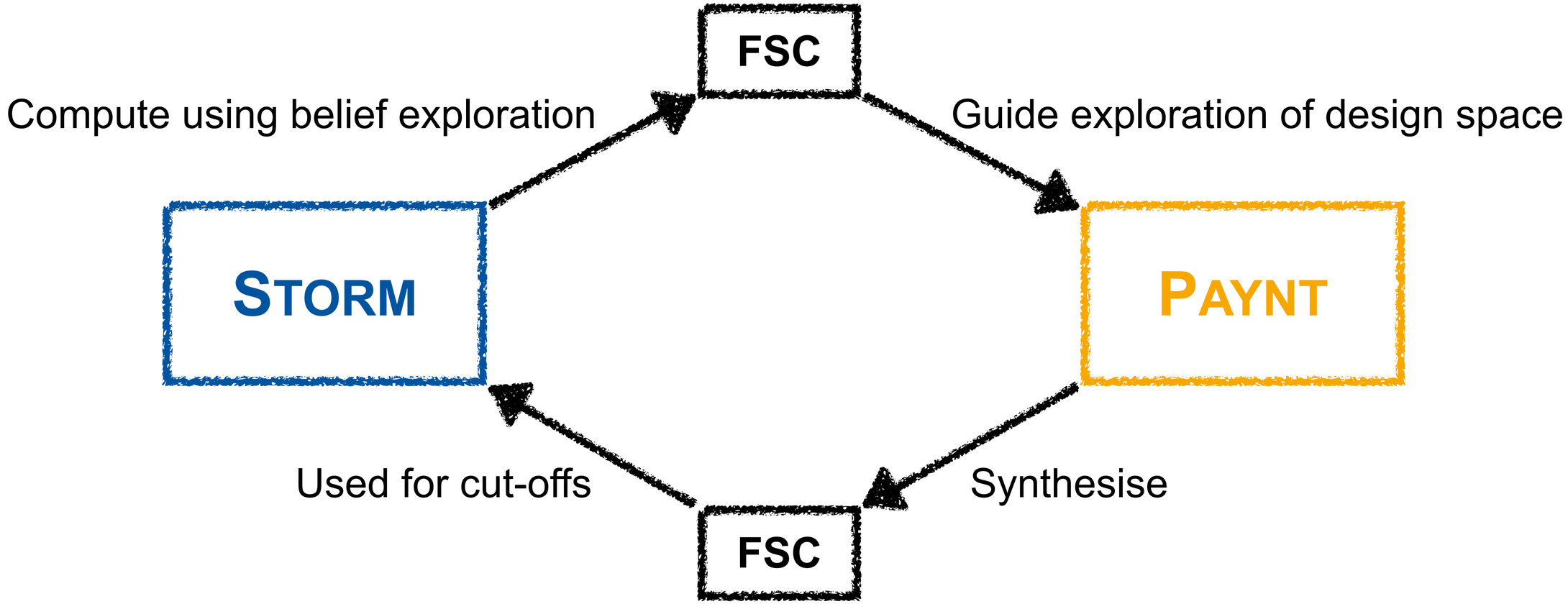


FSC F



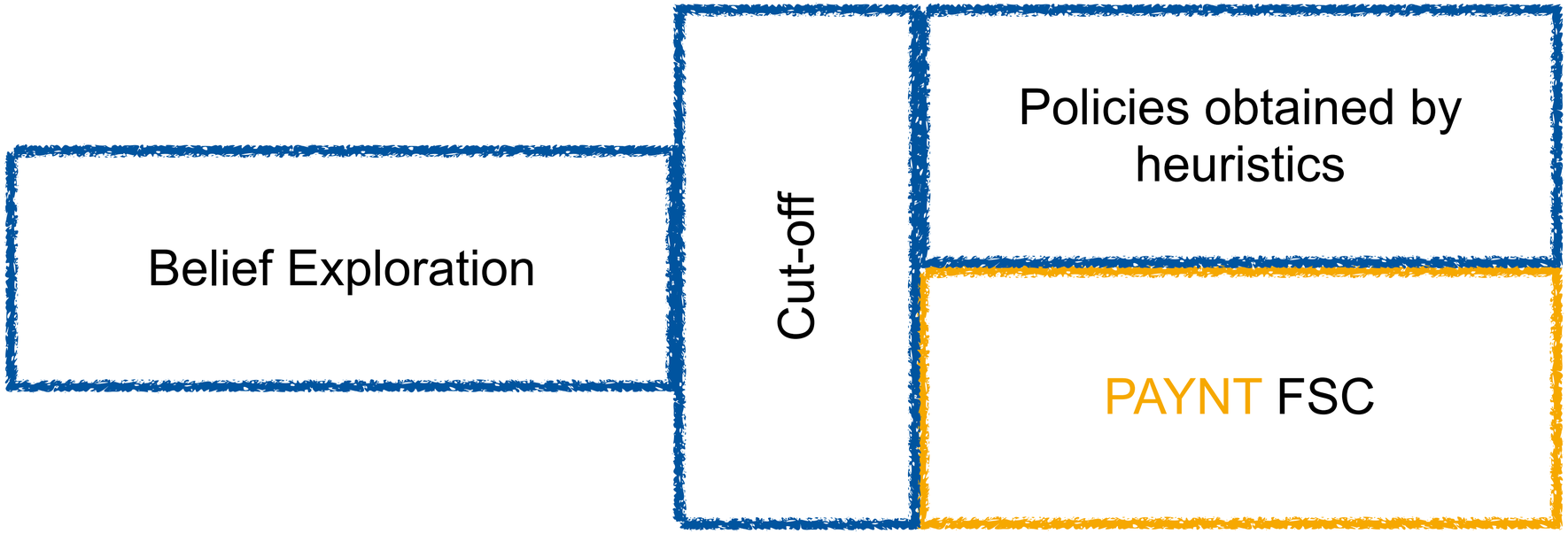
Symbiotic Approach — Overview

[Andriushchenko et al. 2023]



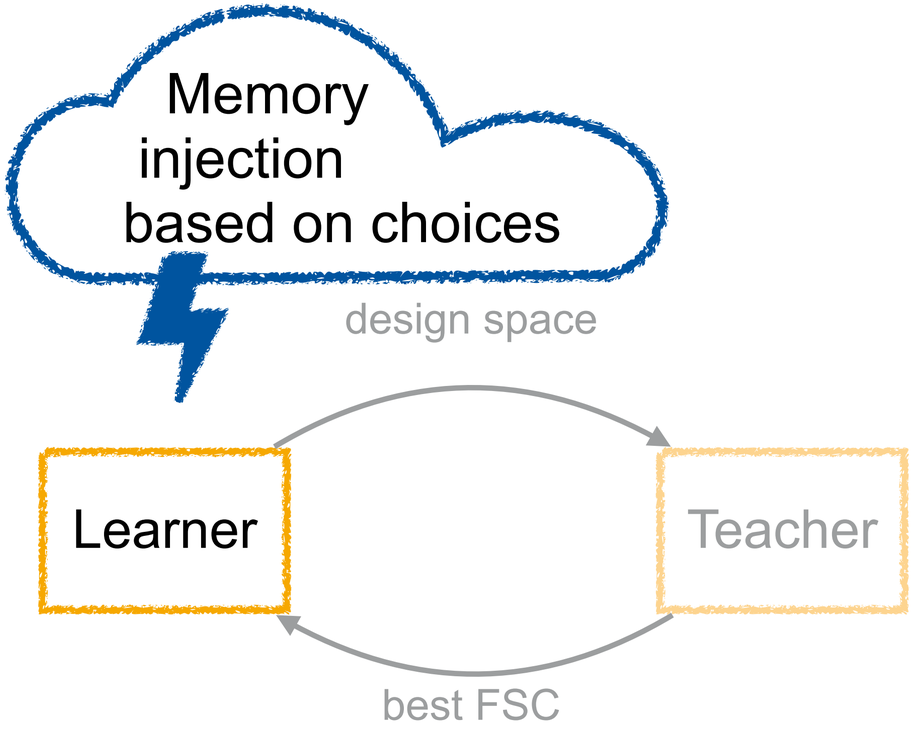
PAYNT → STORM

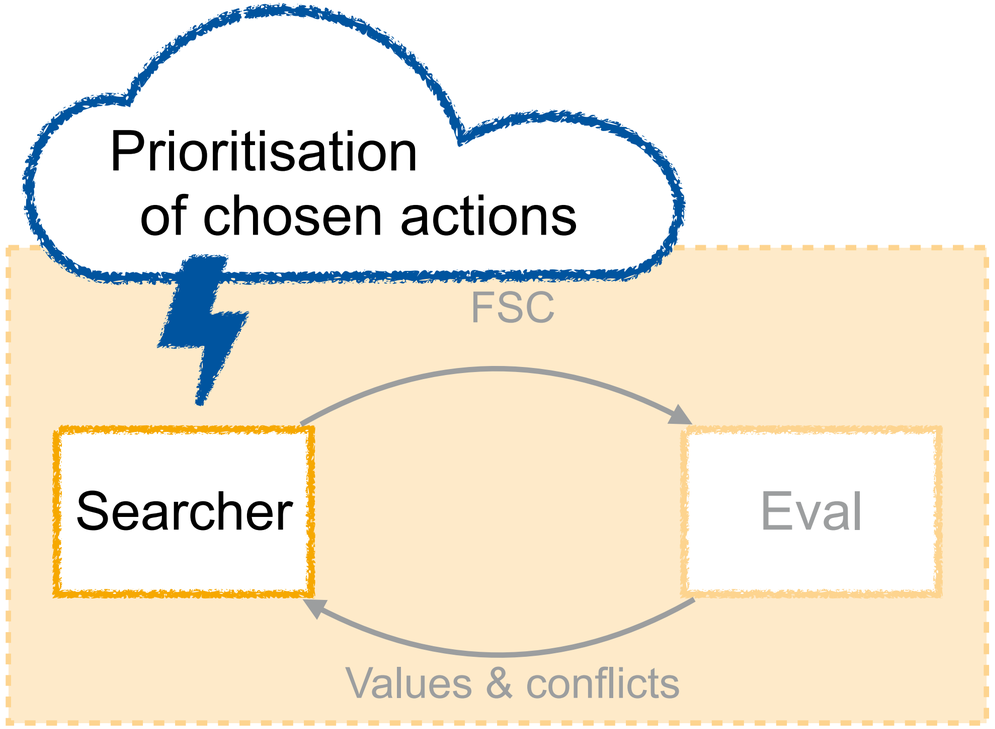
- Use FSC synthesis for cut-off values
- FSC induces state values
  - Convex combination with belief
  - Maximisation over memory nodes in induced MC



STORM → PAYNT

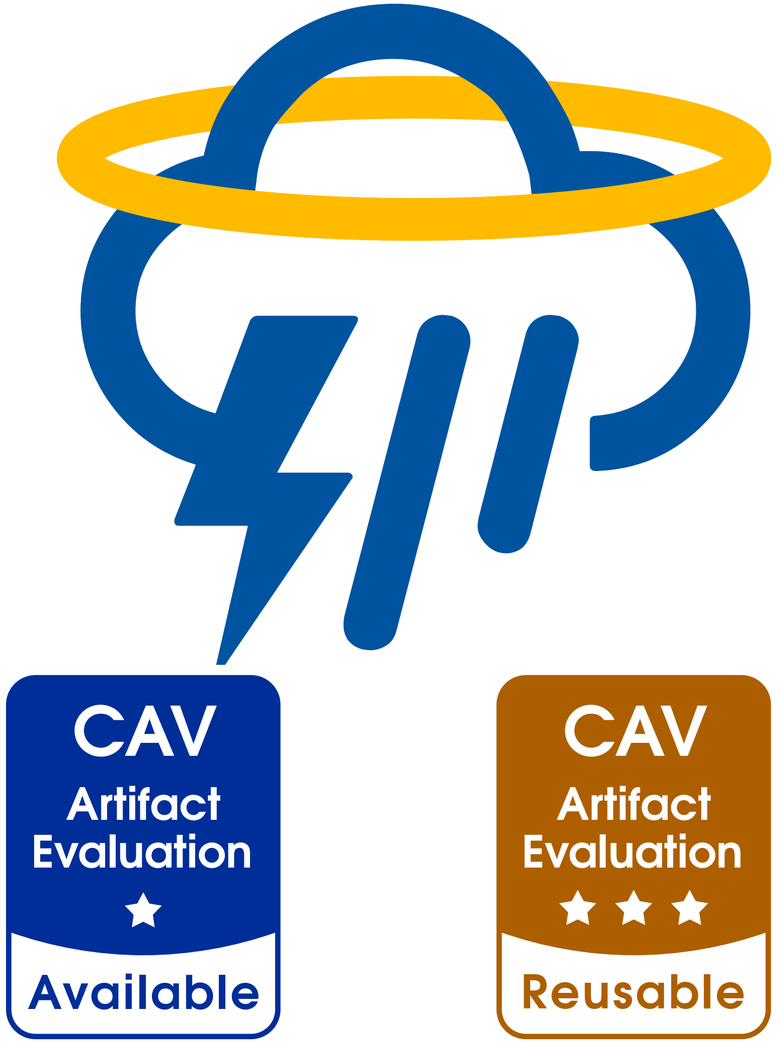
- STORM provides policy on cut-off MDP = FSC





Implementation

- Integrated in **STORM**/**PAYNT**
- Minimisation/Maximisation
  - Reachability Probabilities
  - Expected Total Rewards
- Part of main releases



Results — CAV ’23 (Excerpt)

Intel i5-12600KF @4.9GHz CPU / 64GB RAM
Timeout: 15min

Benchmark (*States/Act./Obs.*)	Over-Approx.	PAYNT	STORM	SAYNT
Refuel 20 - max (6834/24k/66)	≤ 0.99	0.02 922s	0.15 468s	**0.24** 386s
Drone 8-2 - max (13k/32k/3195)	≤ 0.99	0.9 260s	0.68 98s	**0.96** **247s**
Netw 3-8-20 - min (17k/30k/2205)	≥ 4.31	11.04 638s	10.27 238s	**10** 742s
Lanes+ - min (2741/5289/11)	≥ 4805	8223 118s	18870 376s	**4805** 173s
4x5x2 95 - max (79/310/7)	≤ 3.26	0.94 305s	2.08 3s	2.08 71s

Better result, sometimes faster

No improvement

max: larger is better
min: smaller is better

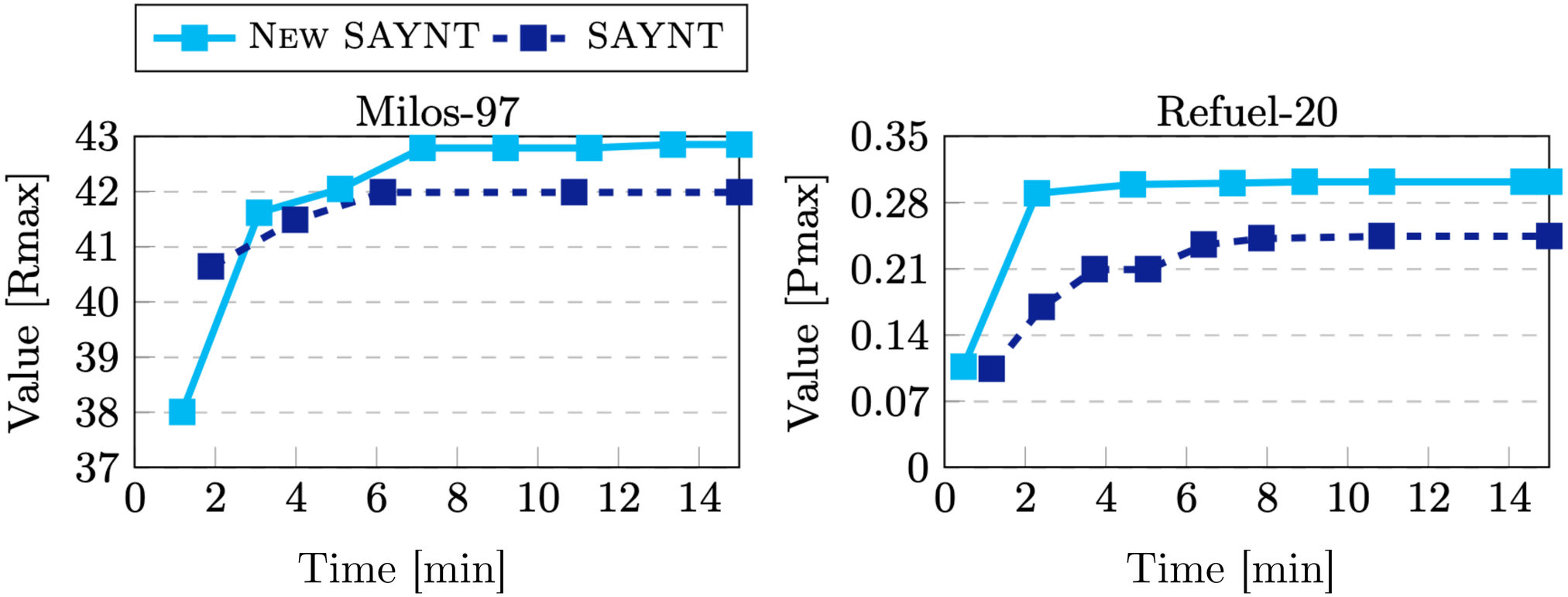
Lots of additional results in the paper

Advances — Focused FSC Synthesis

- **Motivation**:
  - PAYNT - optimise FSC for initial state
  - good in initial state ≠ good for all beliefs

- Seed synthesis in cut-off beliefs
- Use over-approximation as guide
- Prioritise large gaps

Focused FSC Synthesis — Prelim. Results



BUT: **STORM**'s over-approximations are costly for small benefit

➔ better over-approximations?

Advances — Discounting

- Discounted reward: standard in AI applications
- Solvers available (**SARSOP** [Kurniawati, Hsu, Lee 2008], …)

- Added support in **STORM**
  - Modify MDP model checking engine
  - enables discounting in PAYNT and STORM POMDP

No results to report yet, stay tuned!

Conclusion

Policy Synthesis in POMDPs

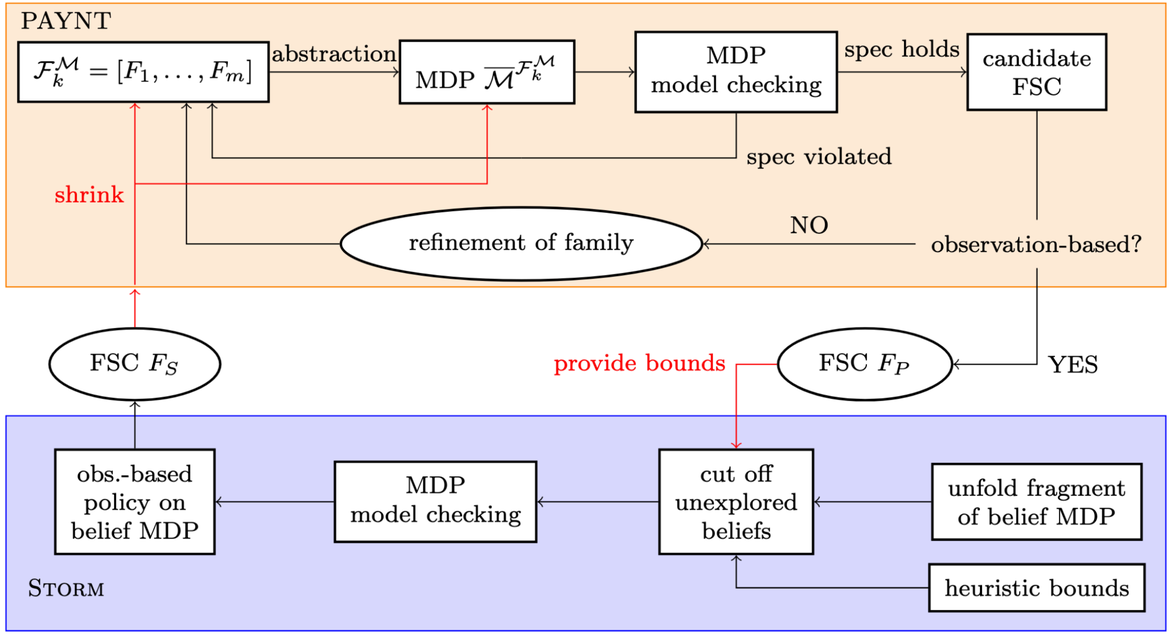
- Difficult problem, practically relevant
- Approximation necessary

Our Approach SAYNT

- Inductive synthesis + belief exploration
- Experiments show potential of symbiosis

Current Developments

- Multiple FSCs
- Integration of over-approximations
- Discounting



Scan for CAV '23 Paper