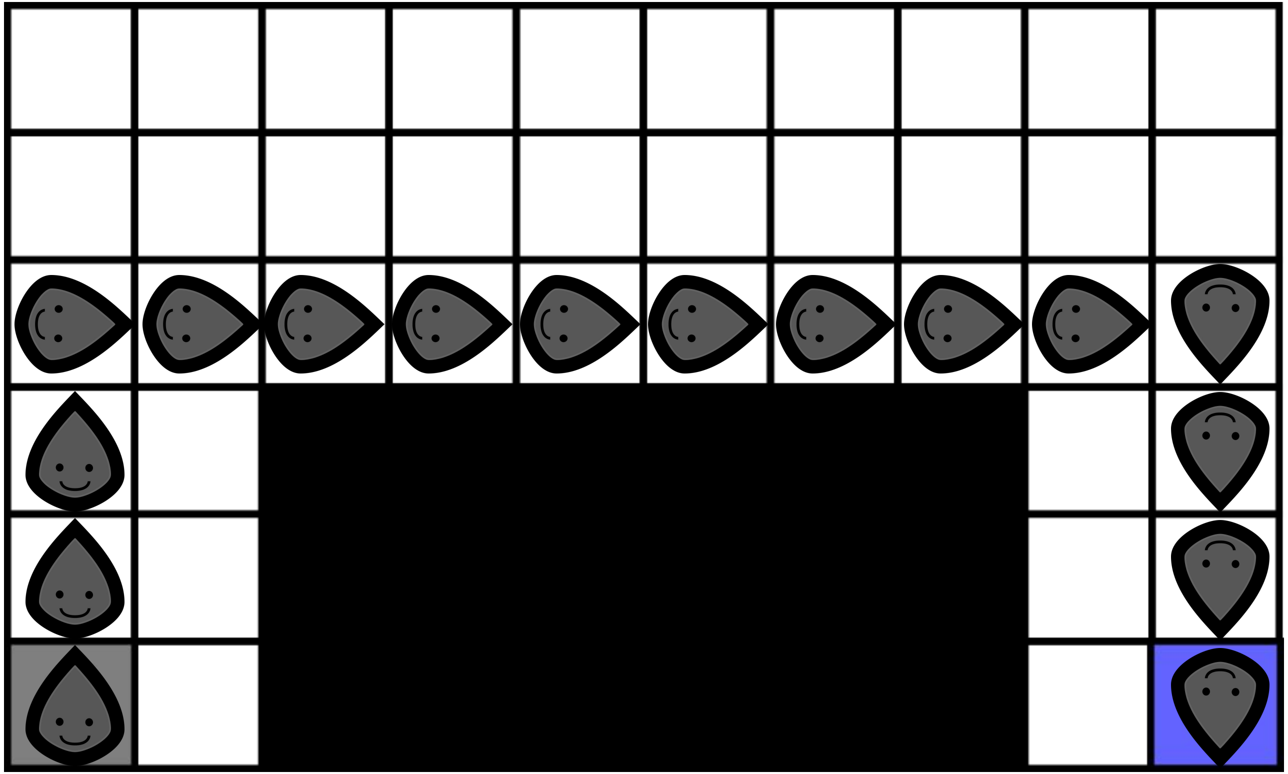
Towards TempoRL: Learning When to Act

André Biedenkapp, Raghu Rajan,
Frank Hutter & Marius Lindauer

In a Nutshell

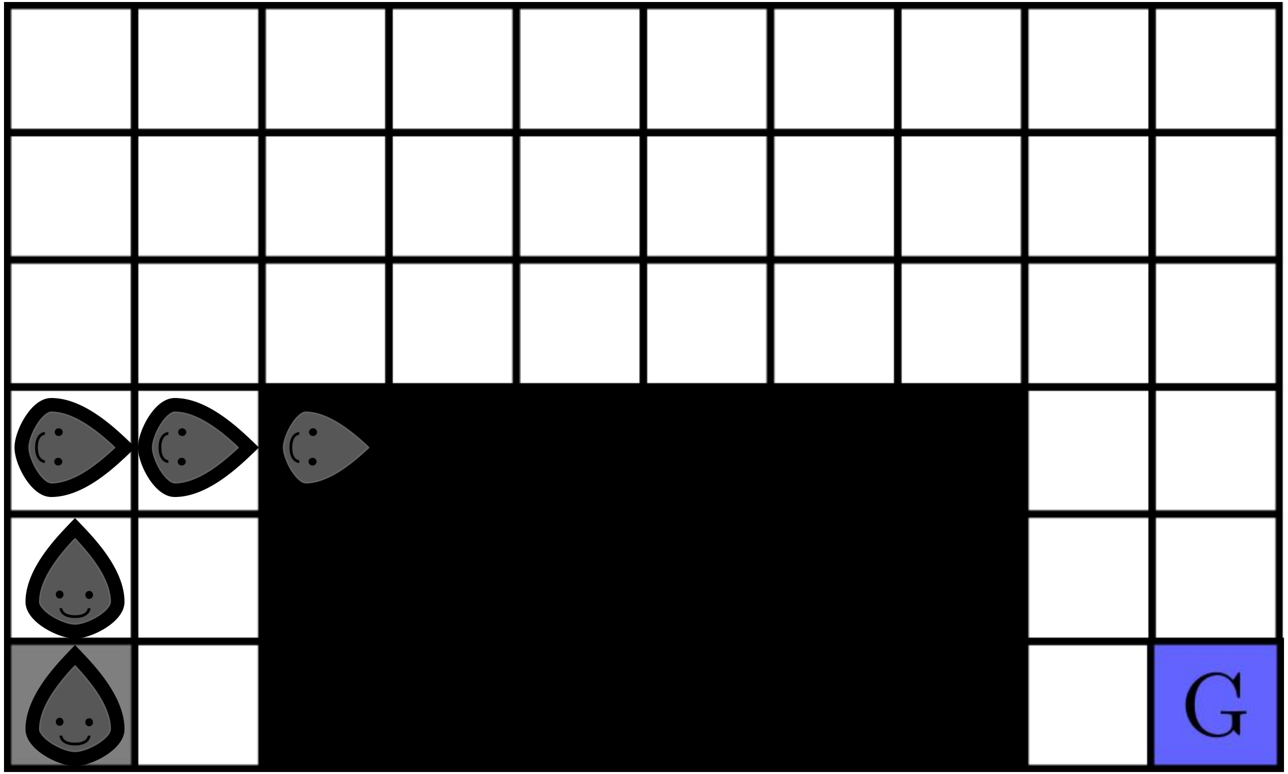
1. We propose a proactive way of doing RL
2. We introduce skip-connections into MDPs
    - through action repetition
    - allows for faster propagation of rewards
3. We propose a novel algorithm using skip-connections
    - learn *what* action to take & *when* to make new decisions
    - condition *when* on *what*
4. We evaluate our approach with tabular Q-learning on small grid worlds

Motivation



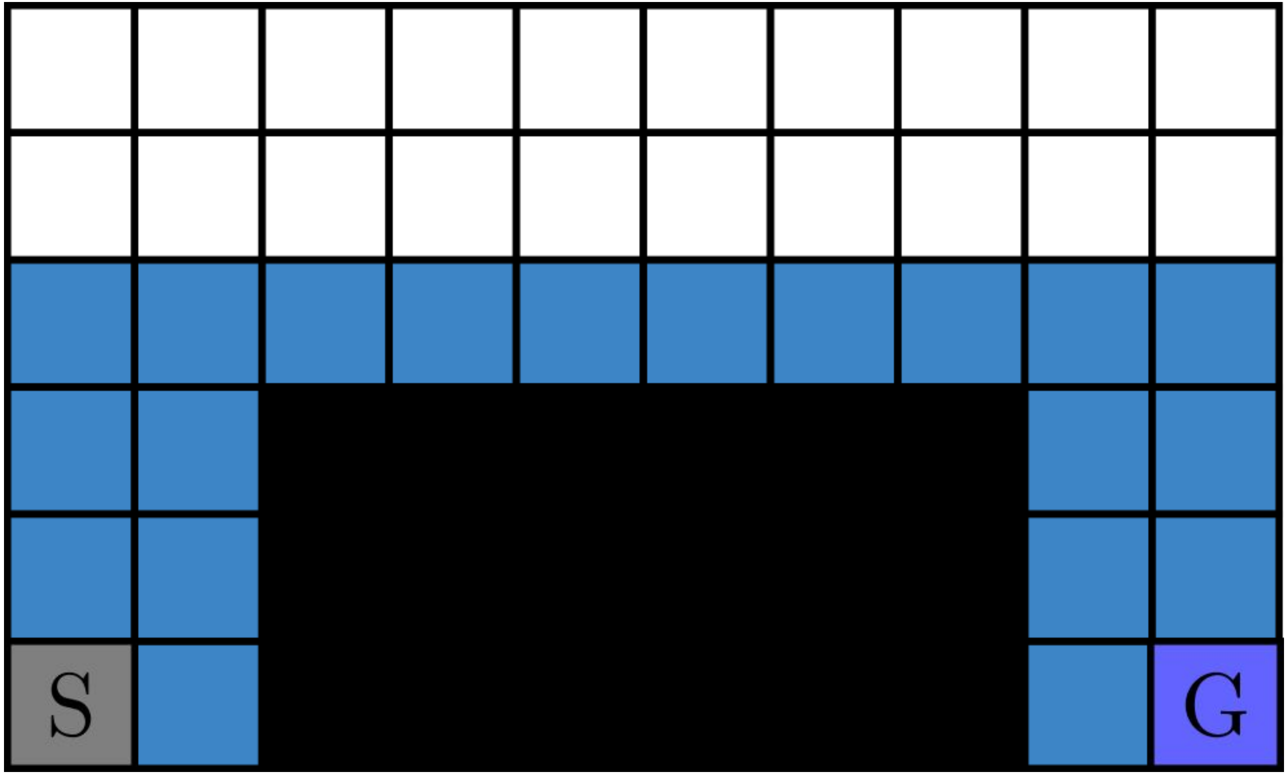
r = 0

Motivation

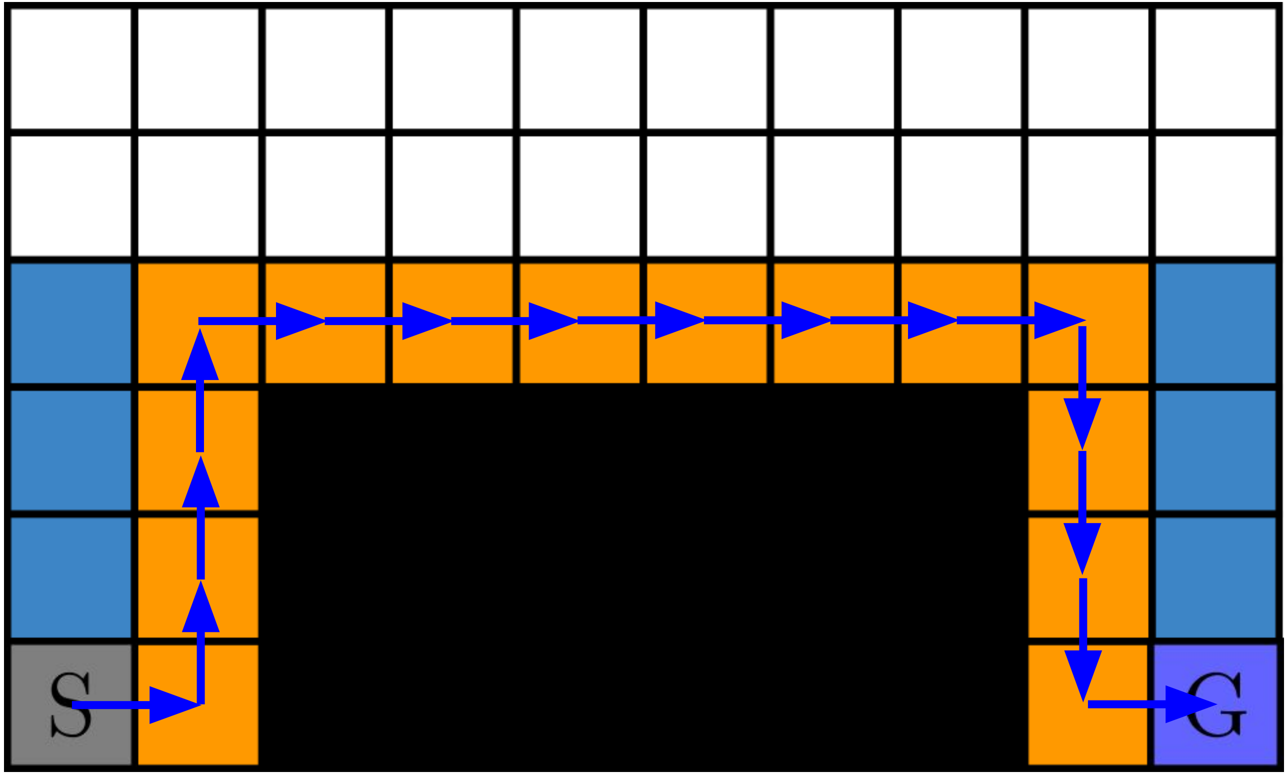


r = -1

Optimal Policies

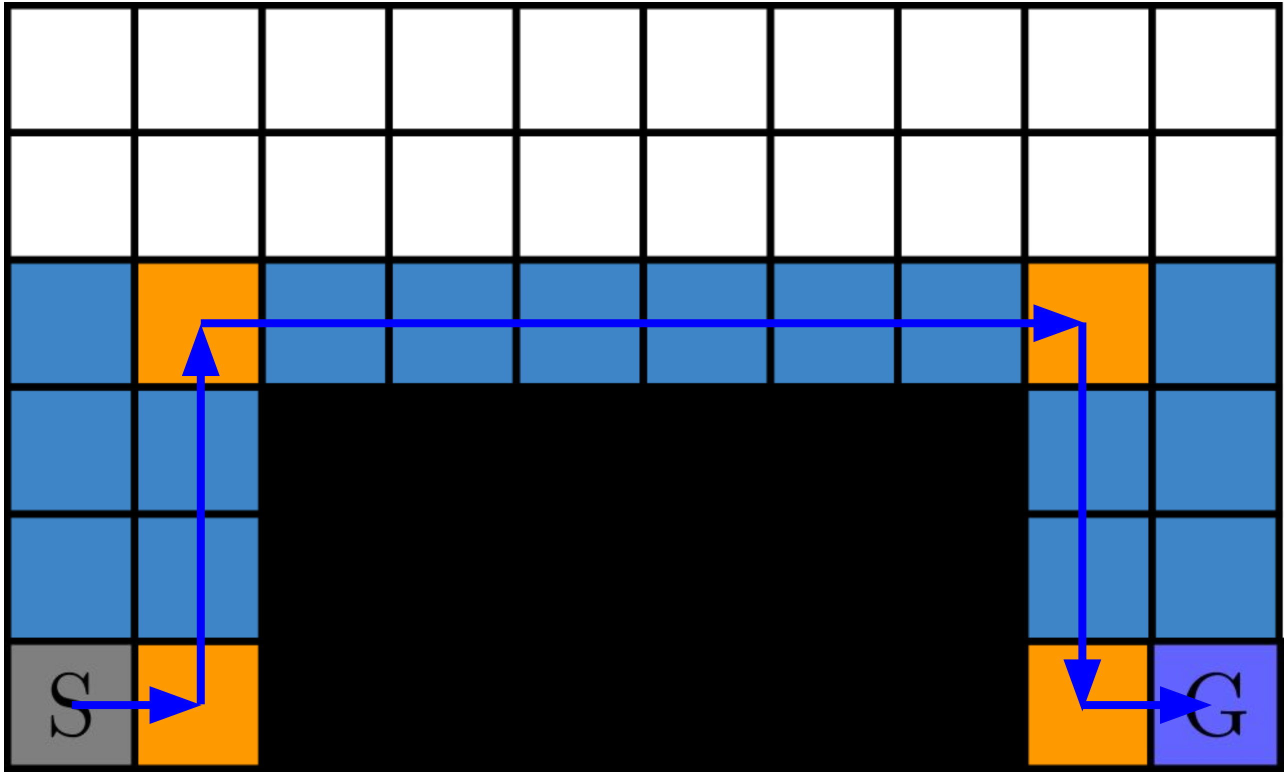


Optimal Policies: When do we need to act?



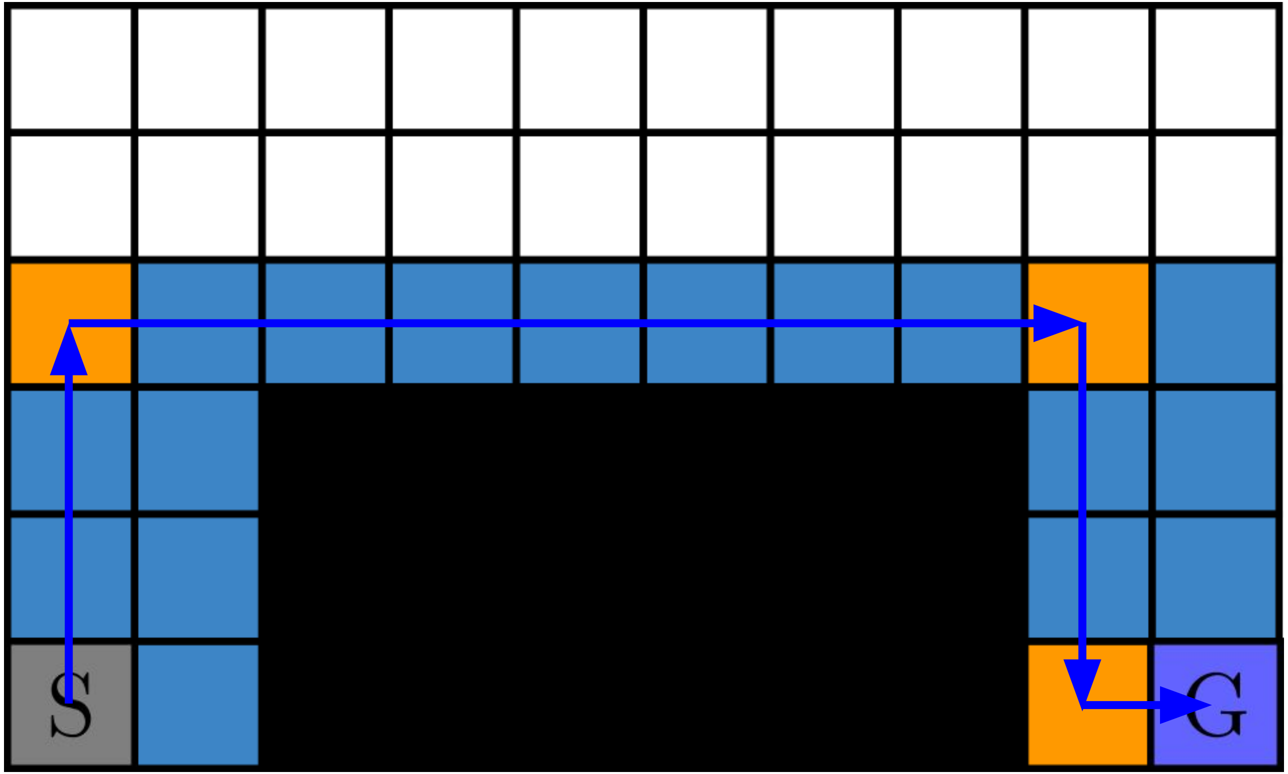
Steps: 16
Decisions: 16

Optimal Policies: When do we need to act?



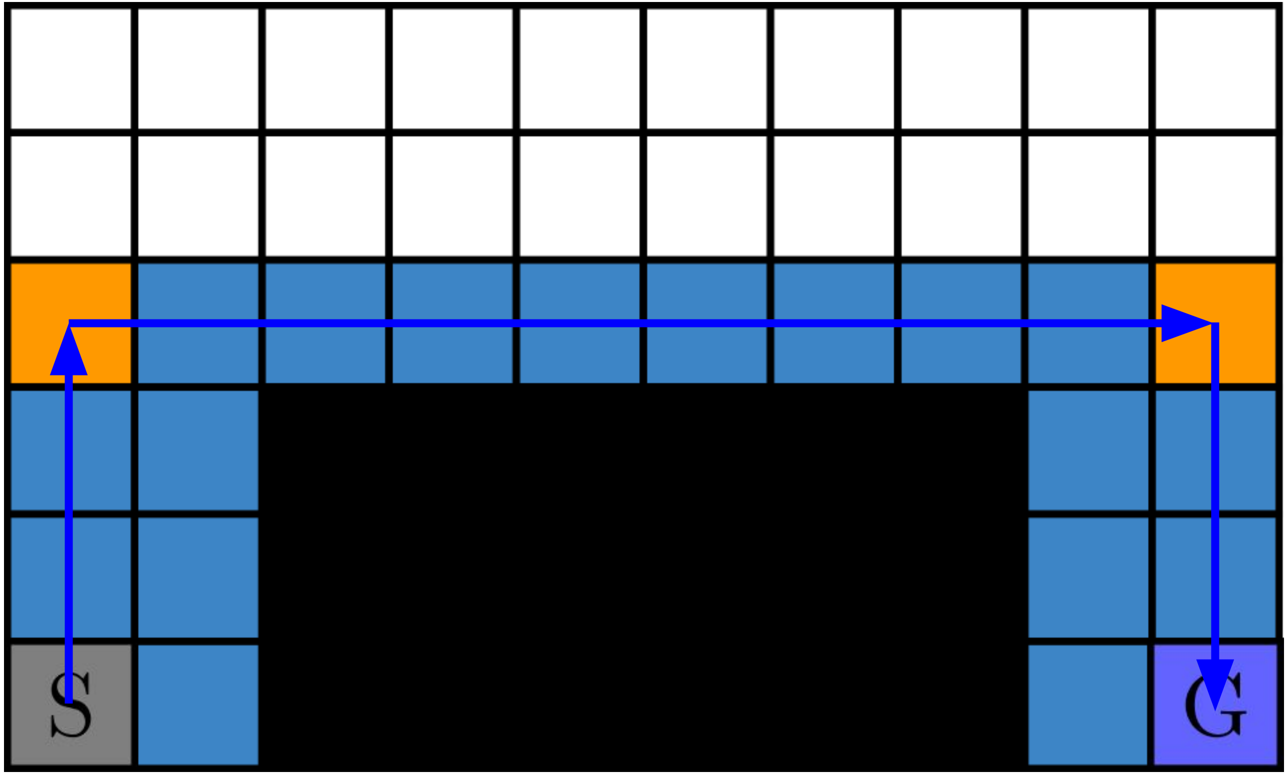
Steps: 16
Decisions: 5

Optimal Policies: When do we need to act?



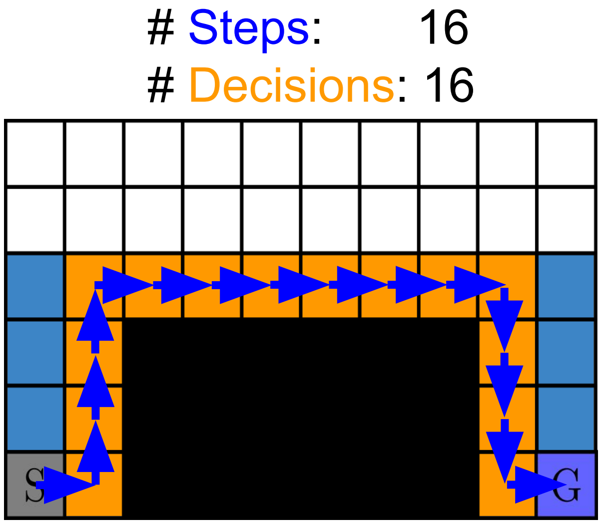
Steps: 16
Decisions: 4

Optimal Policies: When do we need to act?

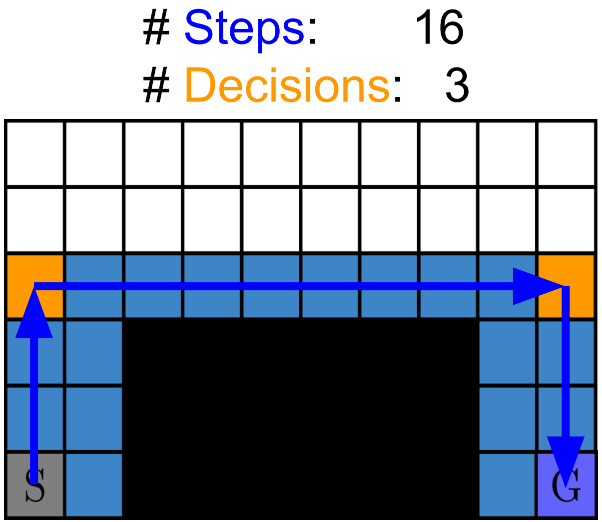


\# Steps: 16
\# Decisions: 3

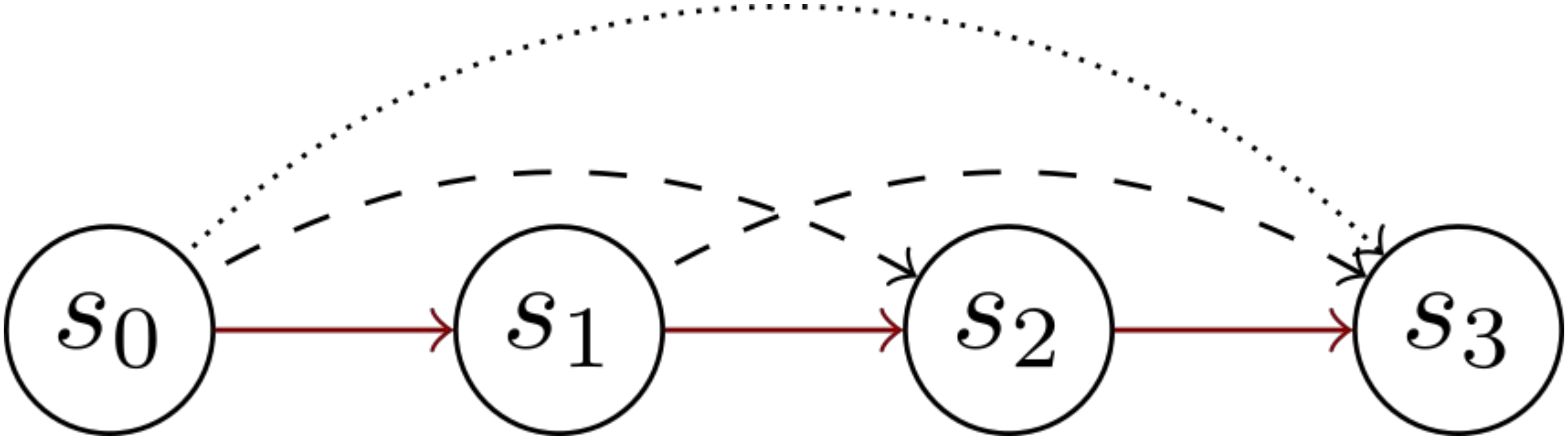
Proactive Decision Making



~80% fewer
Decision points



Skip MDPs



Flat Hierarchy

1. Use standard Q-learning to determine the behaviour

   $Q^{\pi}(s_t, a) \longrightarrow a$

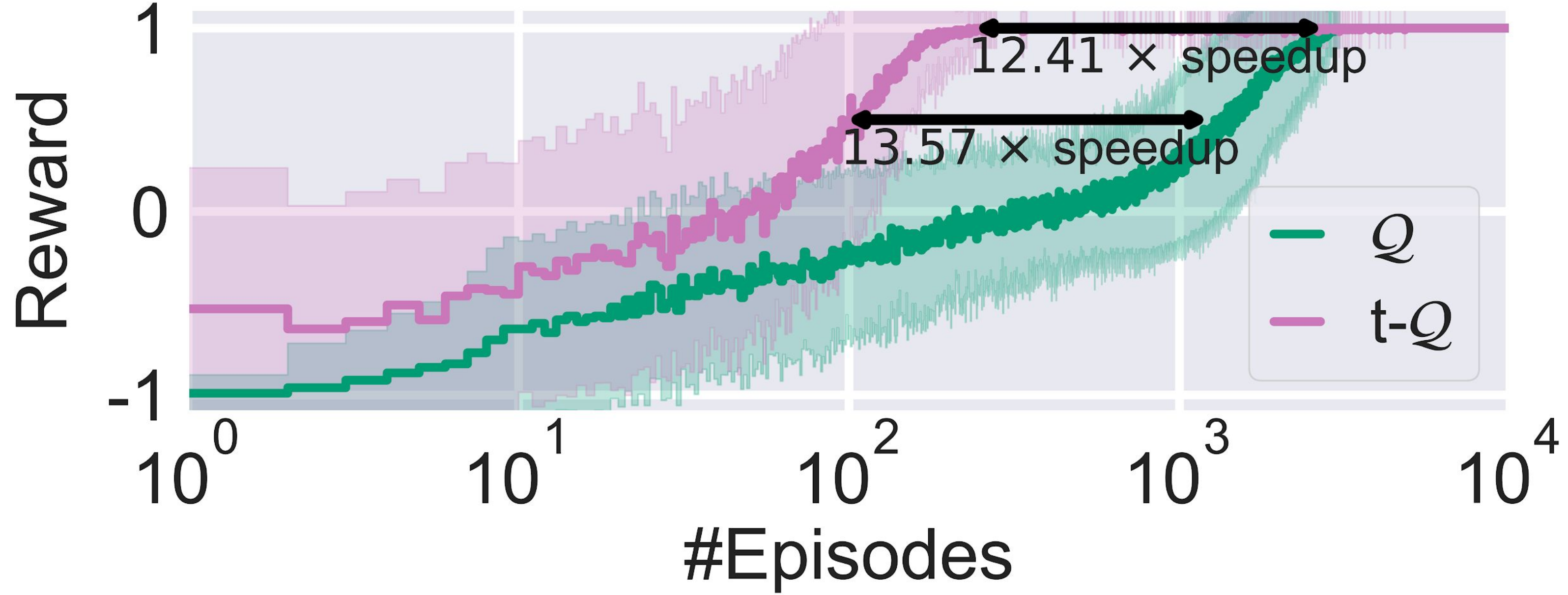
2. Condition skips on the chosen action.

   $Q^{\pi_j}(s_t, j|a) \longrightarrow j$

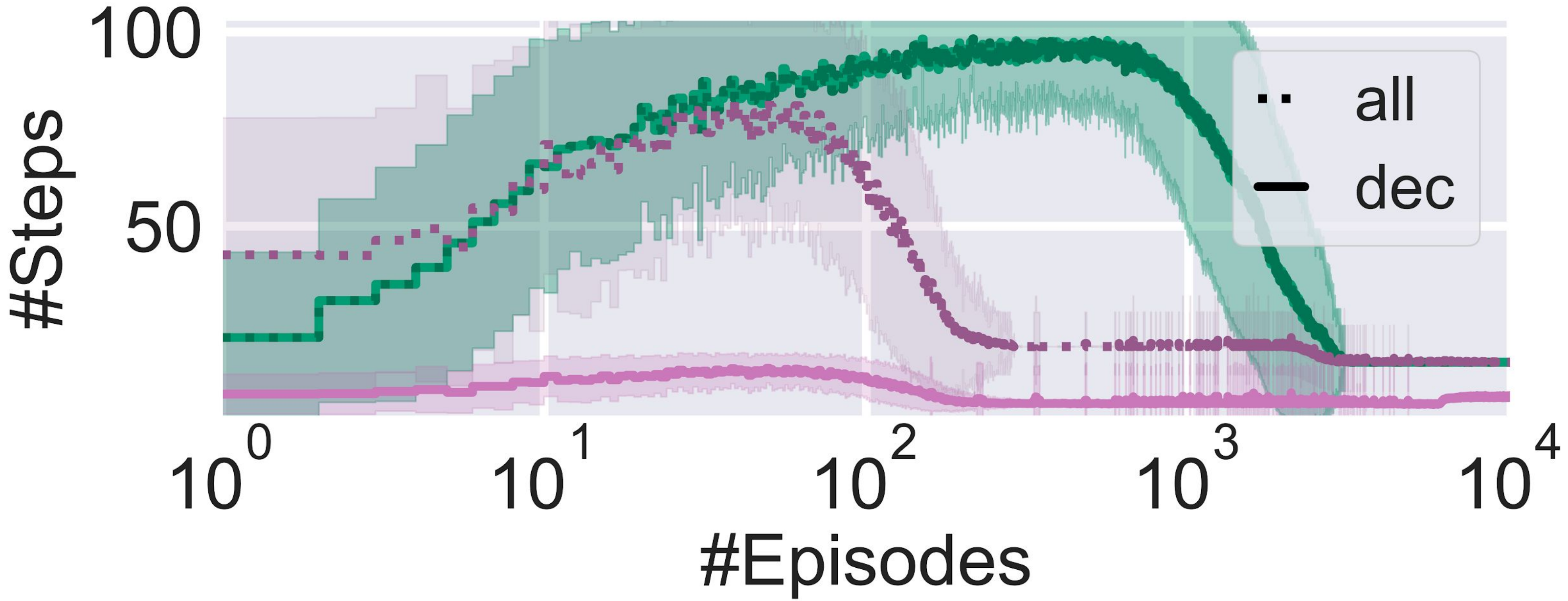
3. Play action a for the next j steps

The action Q-function
The skip Q-function can be learned using n-step updates

Experimental Evaluation



Experimental Evaluation



Wrap-Up

Code & Data available:

https://github.com/automl/TabularTempoRL

Future work:

- Use deep function approximation
- Different exploration mechanisms for skip and behaviour policies

AutoML.org
Freiburg-Hannover