Bath Reinforcement Learning Laboratory

Joshua Evans & Daniel Beechey

Joshua Evans

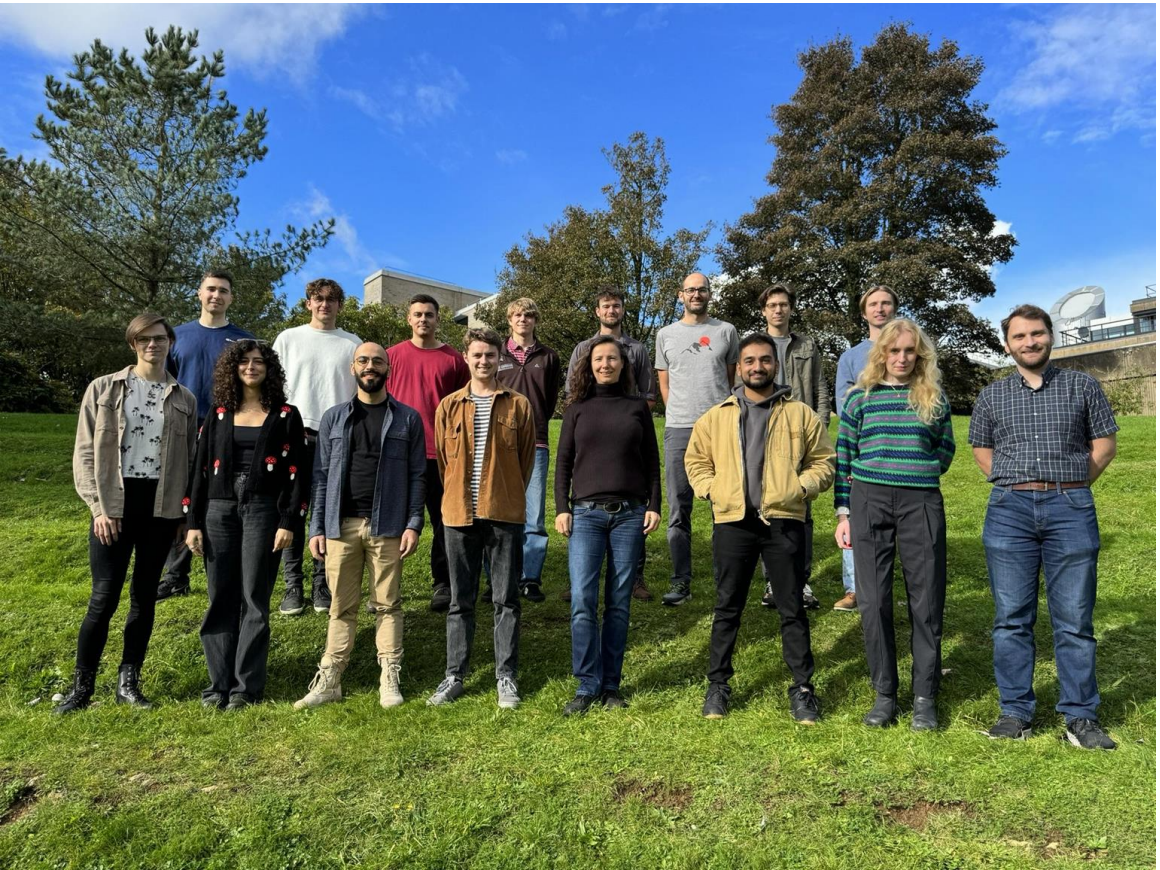
- Bath graduate
  - MComp Computer Science
- Fifth-year Computer Science PhD student and AI Lecturer (CM30359, CM50270).

Daniel Beechey

- Bath graduate
  - BSc Mathematics, MSc Data Science
- Second-year PhD student at the Centre of Doctoral Training in Accountable, Responsible and Transparent AI (ART-AI).

Members of the Bath Reinforcement Learning Laboratory

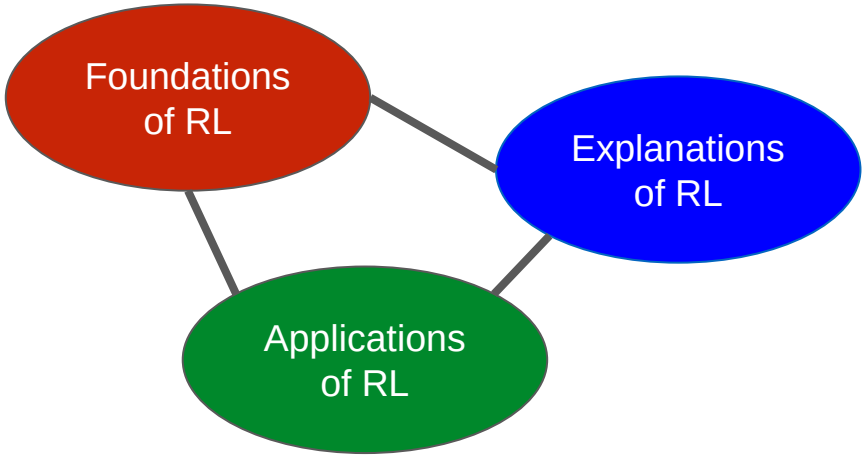
Bath Reinforcement Learning Laboratory



- Lead by Prof Özgür Şimşek
- 21 PhD students
- Supervise undergraduate and postgraduate dissertations

Bath Reinforcement Learning Laboratory

Creating Multi-level Skill Hierarchies in Reinforcement Learning. Evans & Şimşek, 2023, NeurIPS.



Explaining Reinforcement Learning with Shapley Values. Beechey, Smith, & Şimşek, 2023, ICML.

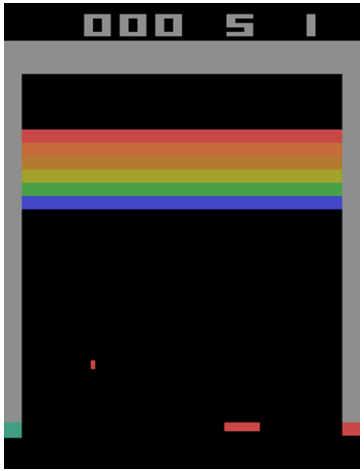
- Resource Constrained Station-Keeping for Latex Balloons, Saunders et al., 2023, IROS
- Designing Printed Circuit Boards
- High-volume, high-variation manufacturing of injectable medicines
- Patient scheduling for the NHS

What can reinforcement learning do?

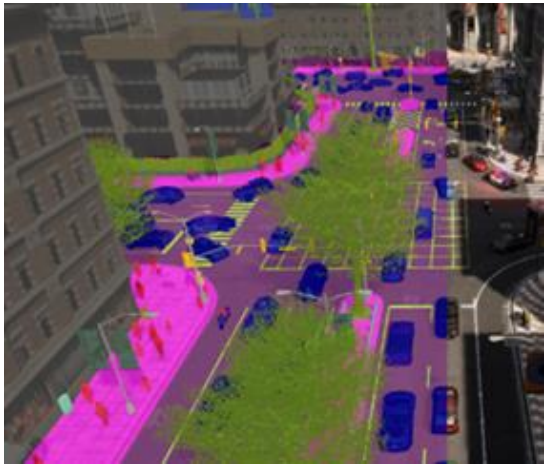
TD-Gammon (Tesauro, 1992)

AlphaGo (Silver et al., 2016)

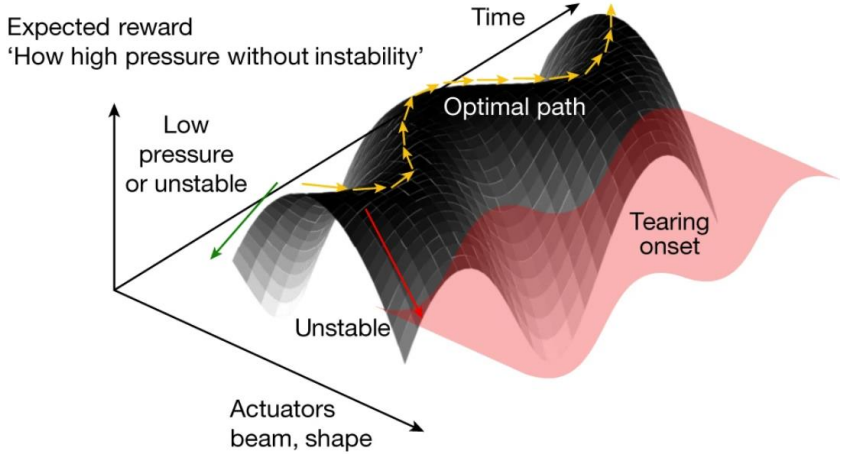
RoboCup (Riedmiller & Gabel, 2007)



Atari (Minh et al. 2015)



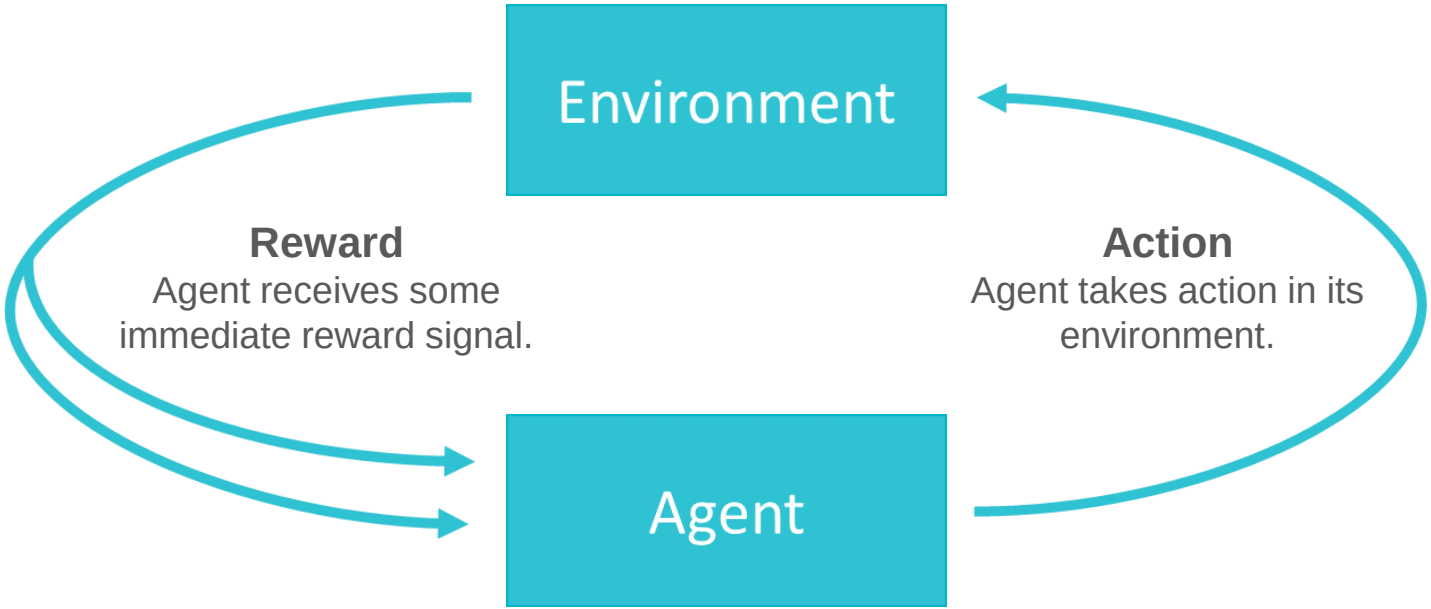
Autonomous Driving (ML4AD@NeurIPS)



Nuclear Fusion Reactor Control (Seo et al. 2024)

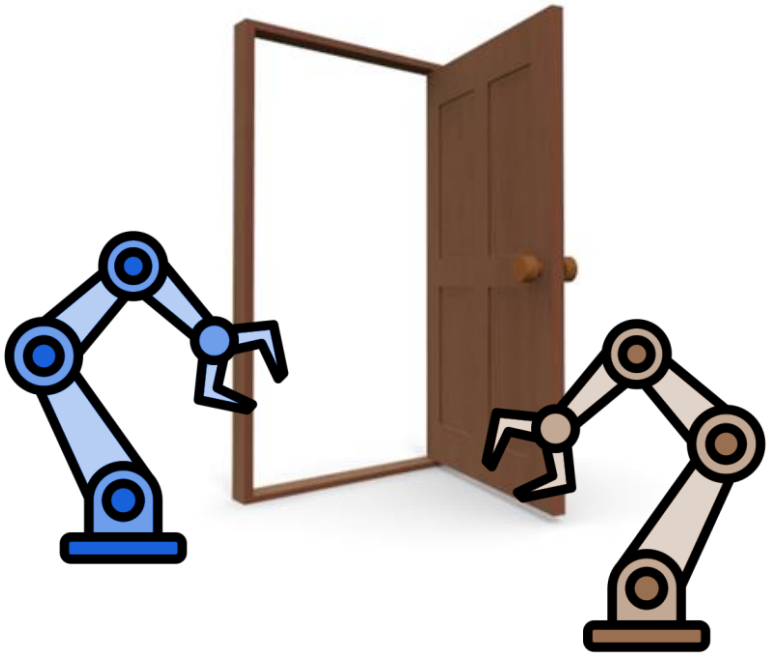
What is reinforcement learning?

What do all of these problems have in common?



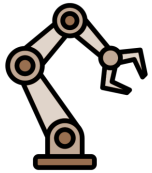
RL is **learning how to act** through **trial and error** interaction with the world.

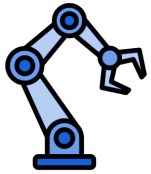
How to **map states to actions** in order to **maximise long term reward**.



Skill Discovery

How should autonomous agents develop useful behaviour hierarchies on their own as they interact with the world?





Creating Multi-Level Skill Hierarchies in Reinforcement Learning

Joshua B. Evans
Department of Computer Science
University of Bath
Bath, United Kingdom
jbe25@bath.ac.uk

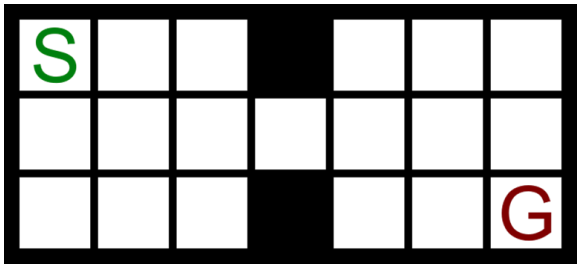
Özgür Şimşek
Department of Computer Science
University of Bath
Bath, United Kingdom
o.simsek@bath.ac.uk

Evans, J.B. and Şimşek, Ö., 2023. Creating Multi-Level Skill Hierarchies in Reinforcement Learning.
In *Advances in Neural Information Processing Systems, 36 (NeurIPS '23)*.

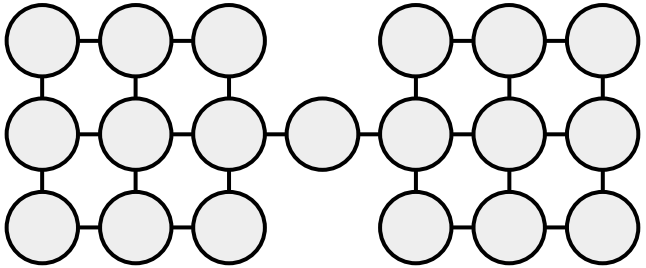
State Transition Graphs

We can represent the interaction between an agent and its environment as a **state transition graph**.

- Each node represents a state.
- Each edge represents a transition between nodes via primitive actions.



Two-Room Gridworld



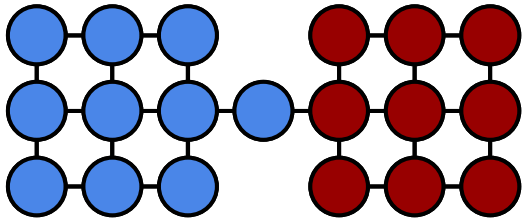
State Transition Graph

Edge directions omitted for clarity.

Graph-Based Skill Discovery

We can use tools from graph theory to analyse this graph and define skills.

- Use partitioning methods to identify meaningful regions of the state space.
- Specifically, find a partition that maximises **modularity**.



Graph Partitioning

↓

Skills for navigating between rooms.

Partitions with high modularity have relatively dense connections within their clusters, and relatively sparse connections between them.

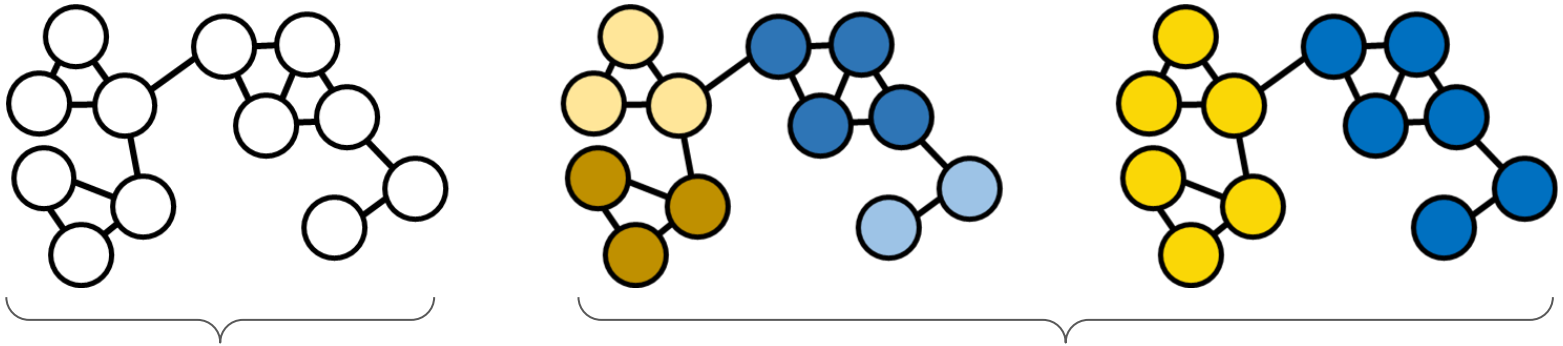
↓

Regions of the environment that are easy to move within, but difficult to move between.

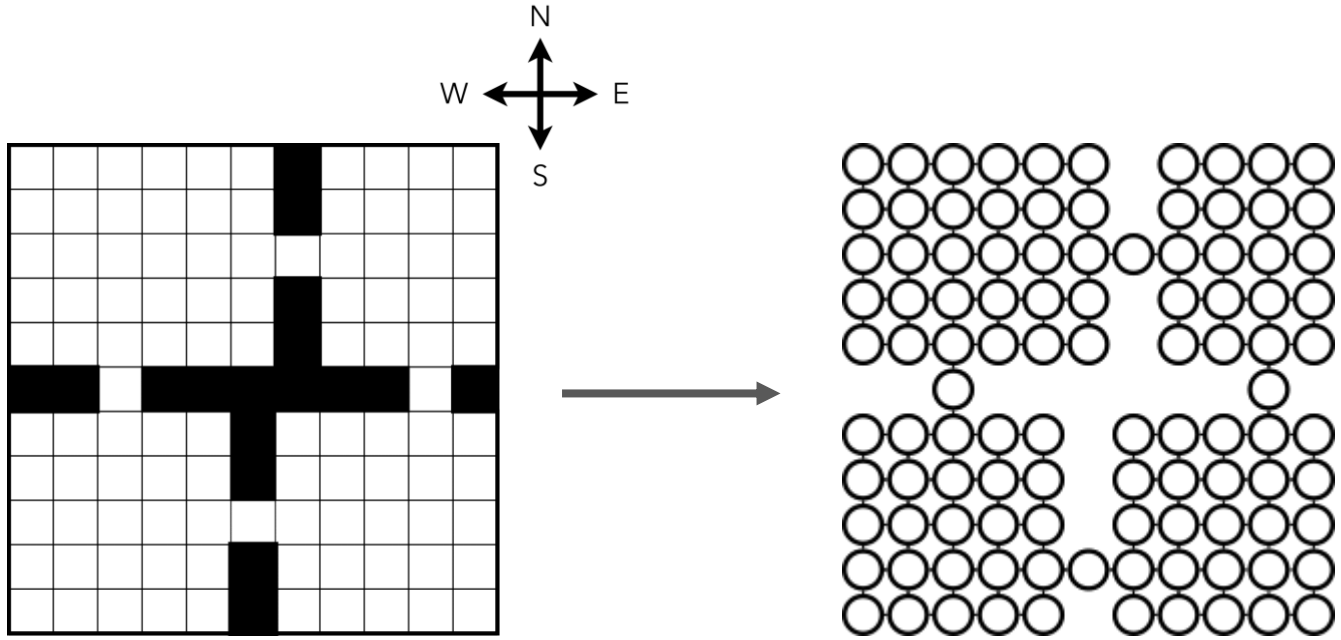
Finding a partition that maximises modularity is NP-complete.

Approximations exist - we use the **Louvain algorithm**.

- Linear time complexity in the number of graph edges.
- Agglomerative **hierarchical** graph clustering algorithm.
- Outputs a sequence of partitions with a useful hierarchical structure.



Results

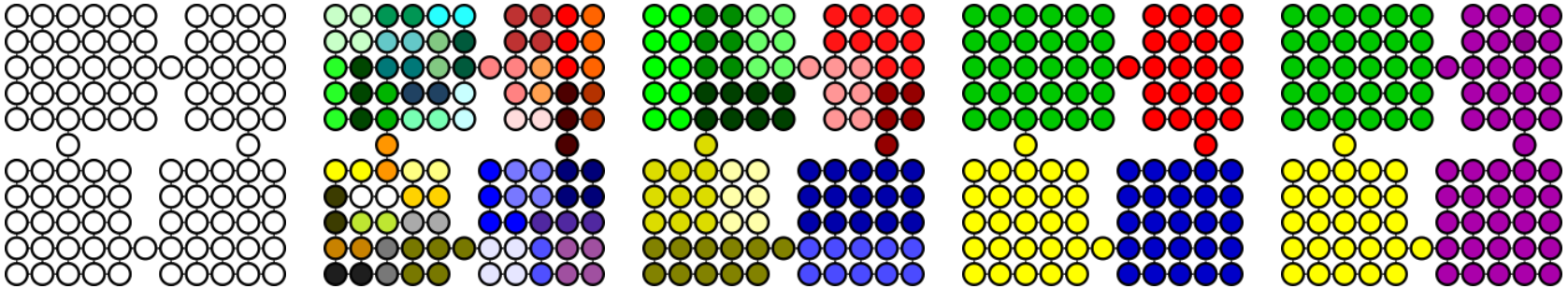


Results

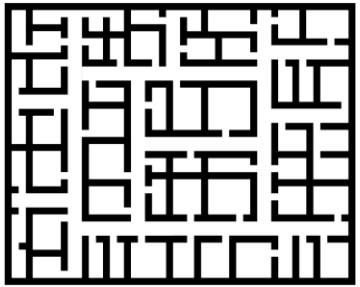
Contains skills that operate at varying timescales!

The skills naturally form a multi-level hierarchy!

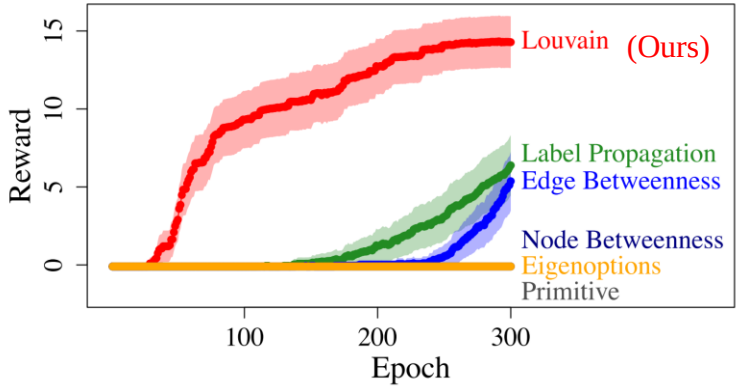
Entire hierarchy is generate entirely automatically!



Level 1 Level 2 Level 3 Level 4



Two-Floor “Office”
(2537 States)

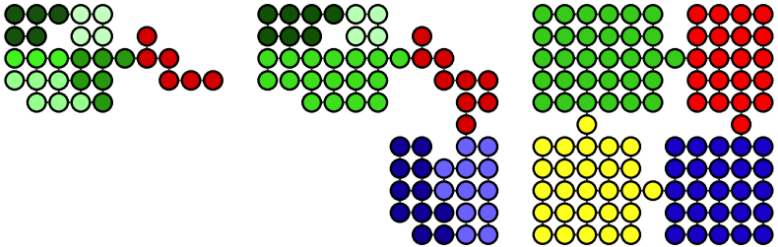


Scaling Up

Large discrete domains? Continuous domains? Function approximation?

Incremental Learning

Open-ended learning of Louvain skills?



How to Explain Reinforcement Learning

Thomas Smith

tmss20@bath.ac.uk

Özgür Şimşek

os435@bath.ac.uk

Beechey, D., Smith, T.M. and Şimşek, Ö., 2023, July. Explaining reinforcement learning with shapley values. In *International Conference on Machine Learning* (pp. 2003-2014). PMLR.

Understanding AI Behaviour

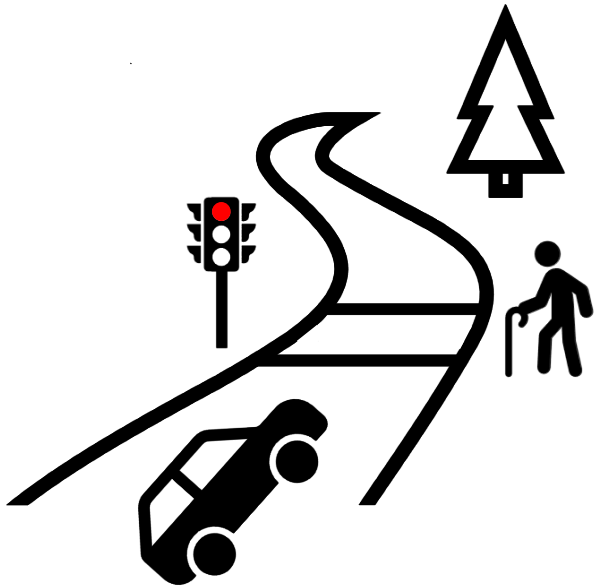
The behaviour of RL agents is influenced by certain features of their observations.

Contribution: A mathematical framework for explaining agent behaviour using the influence of features.

How to Calculate the Influence of Features?

Compute the influence of features by observing the behaviour change caused by their removal.

Features are correlated, removing one feature does not properly capture their influence.



Shapley Values

In a **cooperative game**, players work together to produce a measurable outcome.

How to assign the contribution of a player to the outcome of a game?

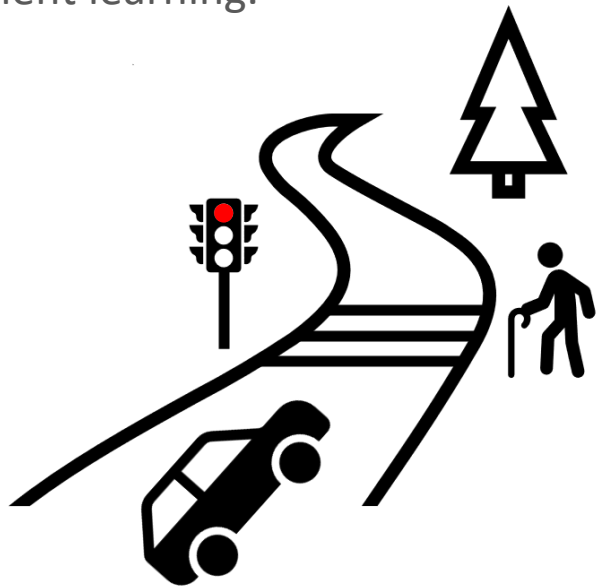
Shapley values: the **unique solution** to the **contribution assignment problem** satisfying the four axioms of **efficiency, symmetry, nullity and linearity**.

Shapley Values for Explaining Reinforcement Learning (SVERL)

A collection of cooperative games played by features of an agent's observation whose outcomes are different behavioural qualities of reinforcement learning.

Explaining Policy

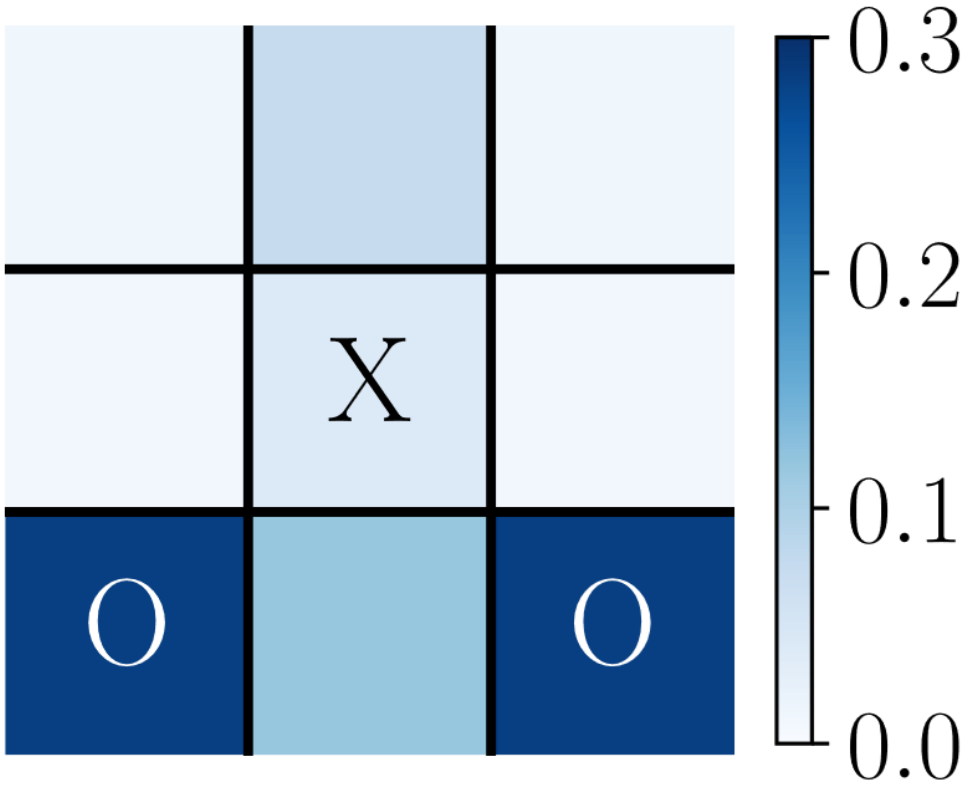
The contribution of features to the probability of selecting action a in state s.



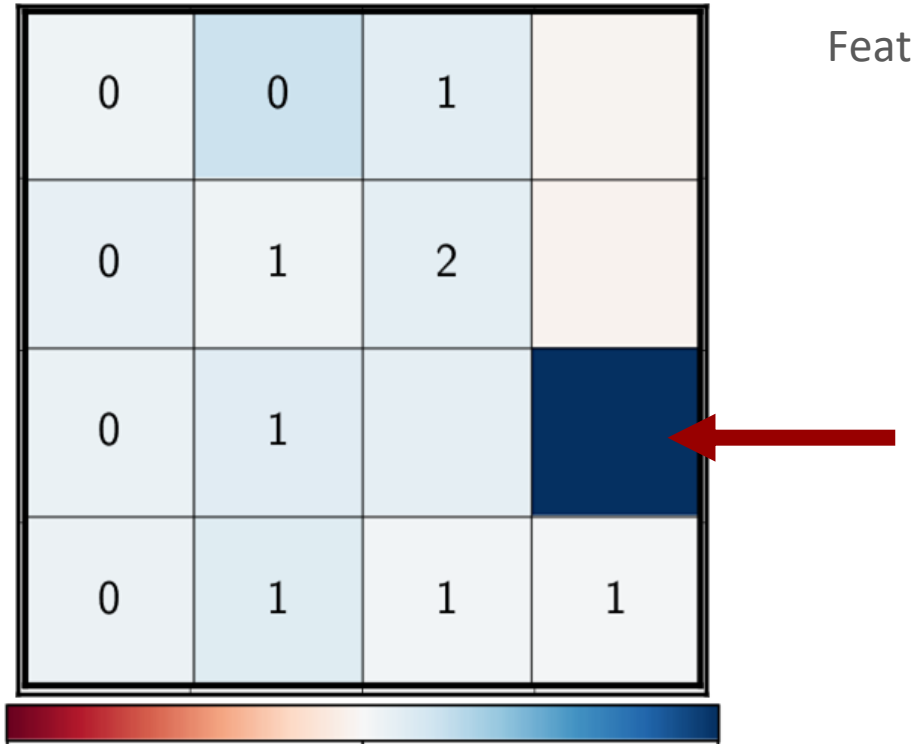
Explaining Tic Tac Toe

Agent plays as X.

Features are the grid squares.



Explaining Minesweeper



Features are the 16 grid squares.

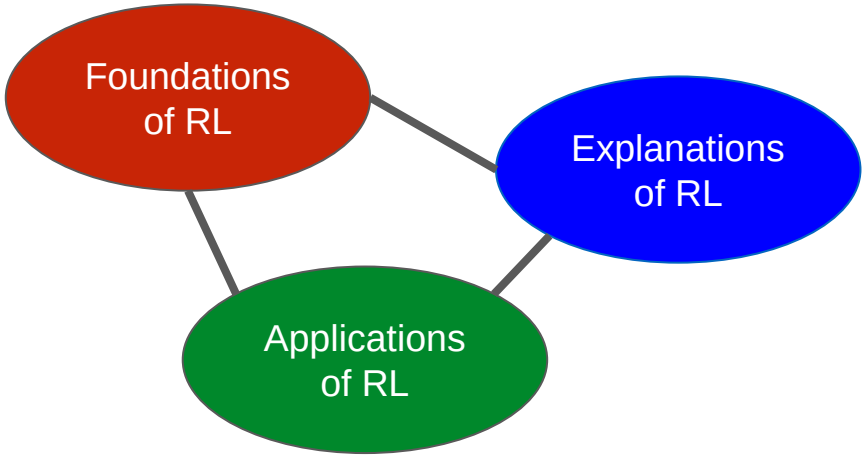
Future Work

- SVERL: The complete framework.
- How to use SVERL in large and complicated domains.
- Real-world applications of SVERL.

Beechey, D., Smith, T.M. and Şimşek, Ö., 2023, July. Explaining reinforcement learning with shapley values. In *International Conference on Machine Learning* (pp. 2003-2014). PMLR.

Bath Reinforcement Learning Laboratory

Creating Multi-level Skill Hierarchies in Reinforcement Learning.
Evans & Şimşek, 2023, NeurIPS.



Explaining Reinforcement Learning with Shapley Values.
Beechey, Smith, & Şimşek, 2023, ICML.

- **Resource Constrained Station-Keeping for Latex Balloons**, Saunders et al., 2023, IROS.
- Designing Printed Circuit Boards.
- High-volume, high-variation manufacturing of injectable medicines.
- Patient scheduling for the NHS.