Task Generalisation in Multi-Agent Reinforcement Learning

AAMAS 2022, Doctoral Consortium

Lukas Schäfer

PhD student, University of Edinburgh

May 9, 2022

1. Multi-Agent Reinforcement Learning

Multi-Agent Reinforcement Learning (MARL)

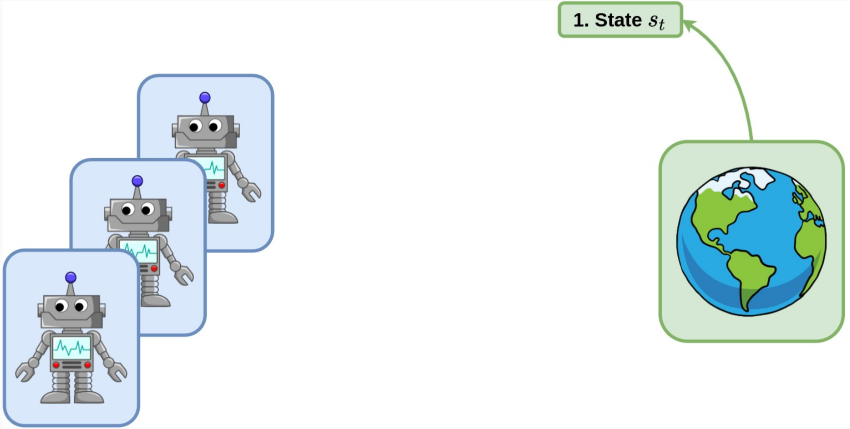


Figure 1: Multi-agent reinforcement learning loop

Multi-Agent Reinforcement Learning (MARL)

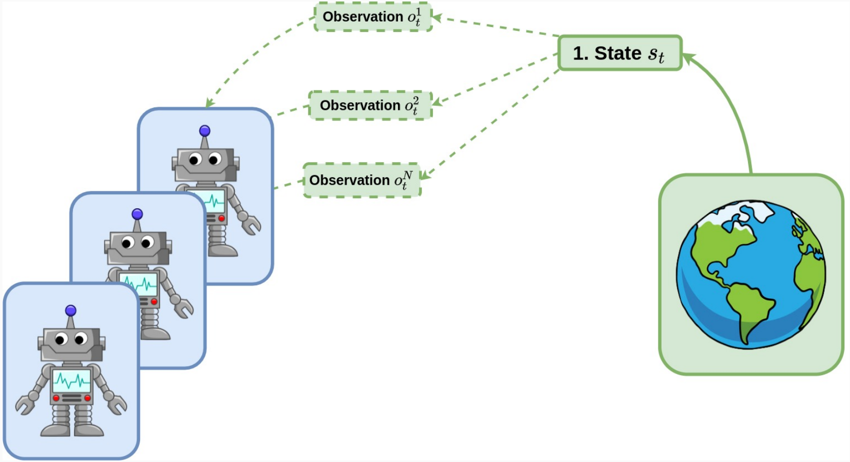


Figure 1: Multi-agent reinforcement learning loop

Multi-Agent Reinforcement Learning (MARL)

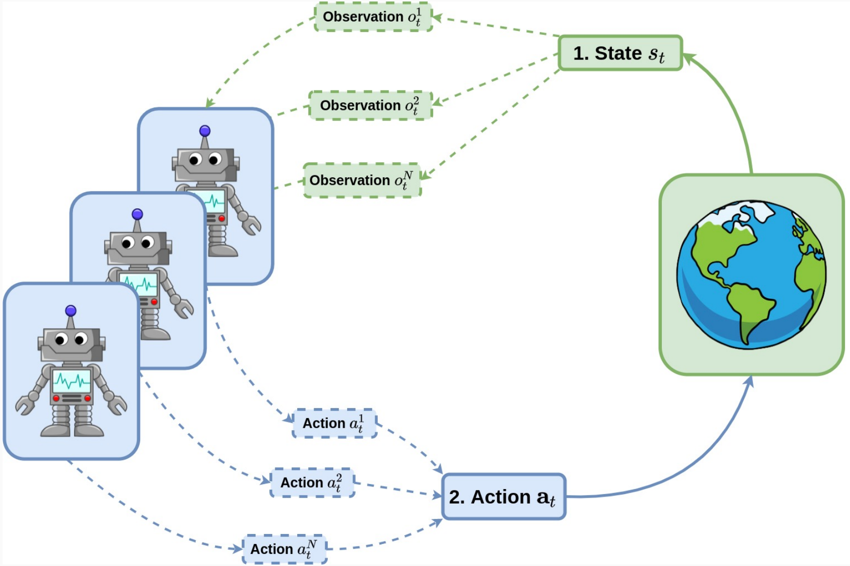


Figure 1: Multi-agent reinforcement learning loop

Multi-Agent Reinforcement Learning (MARL)

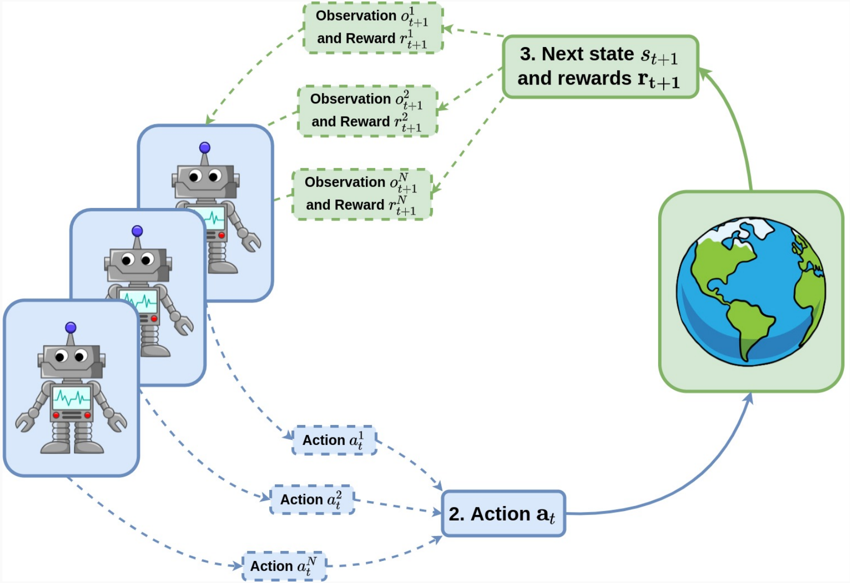


Figure 1: Multi-agent reinforcement learning loop

2. Generalisation in MARL

Motivation

- Learned behaviour typically highly task-specific
- Can be desirable, but often limiting applicability in real-world tasks

Motivation

- Learned behaviour typically highly task-specific
- Can be desirable, but often limiting applicability in real-world tasks
- Tasks require robustness and generalisation capabilities to varying circumstances

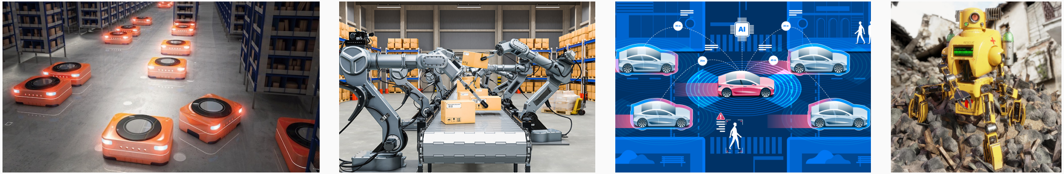


Figure 2: Applications: distributed robotic logistics, autonomous vehicles and rescue robots.

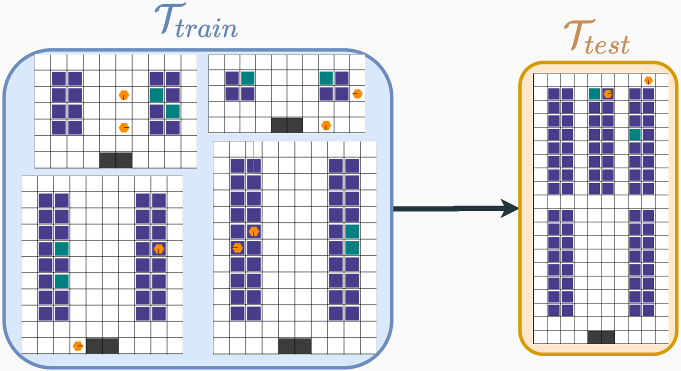
What Does Generalisation Mean?

- (MA)RL lacks unified view on generalisation
- Train joint policy π in a set of training tasks and generalise to testing tasks
- But what is the relationship between tasks in $\mathcal{T}_{train}$ and $\mathcal{T}_{test}$?
  $\rightarrow$ need assumptions on task similarity

Task Generalisation in MARL

Challenge task: Multi-robot warehouse navigation [1]

- Agents need to navigate a warehouse to collect and deliver shelves
- Generalise to different layouts of warehouses



[1]Environment available at `https://github.com/uoe-agents/robotic-warehouse`

3. Preliminary Experiments

Generalisation Experiments

- **Goal:** identify the limitations of existing approaches

Generalisation Experiments

- **Goal:** identify the limitations of existing approaches
- Train agents using independent synchronous Advantage Actor-Critic (IA2C)
- Train in tasks of similar layout but varying height of blocks of shelves
- Evaluate based on zero-shot generalisation after 50M timesteps of training

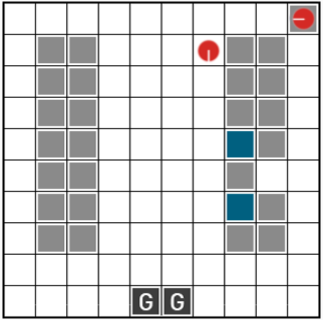
Generalisation Experiments

- **Goal:** identify the limitations of existing approaches
- Train agents using independent synchronous Advantage Actor-Critic (IA2C)
- Train in tasks of similar layout but varying height of blocks of shelves
- Evaluate based on zero-shot generalisation after 50M timesteps of training
- Investigate the impact on generalisation of
    1. Observation encoding
    2. Domain randomisation (train in set of tasks)
    3. Neural network architectures

Generalisation Experiments - Observation Encoding

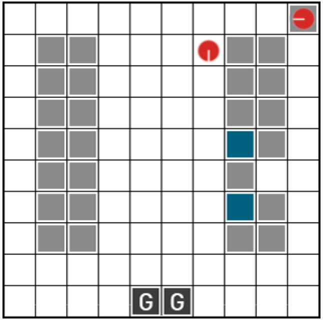
Default observations

- Absolute x- and y-coordinate of agent
- 3×3 grid centered on agent including
  - Agents: load, direction, on "highway"
  - Shelves: requested



Default observations

- Absolute x- and y-coordinate of agent
- 3×3 grid centered on agent including
  - Agents: load, direction, on "highway"
  - Shelves: requested



Generalisation Experiments - Observation Encoding

Default observations

- Absolute x- and y-coordinate of agent
- 3×3 grid centered on agent including
  - Agents: load, direction, on "highway"
  - Shelves: requested

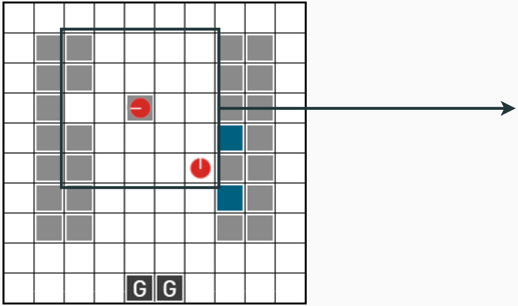
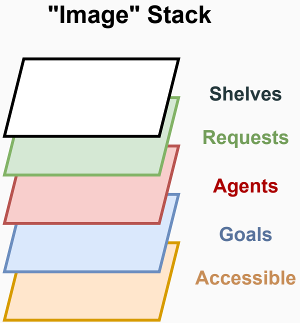
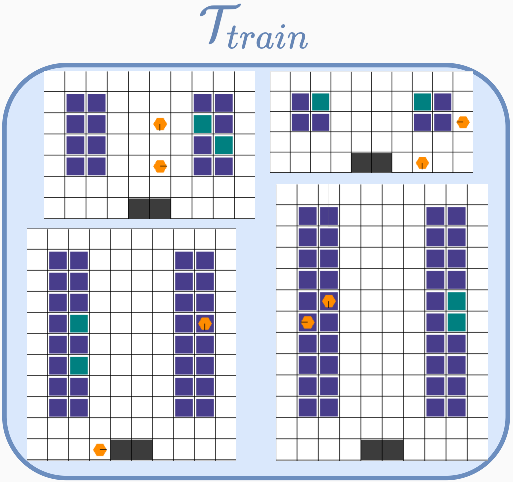


Image observations

- Stack of binary information
- All information is relative



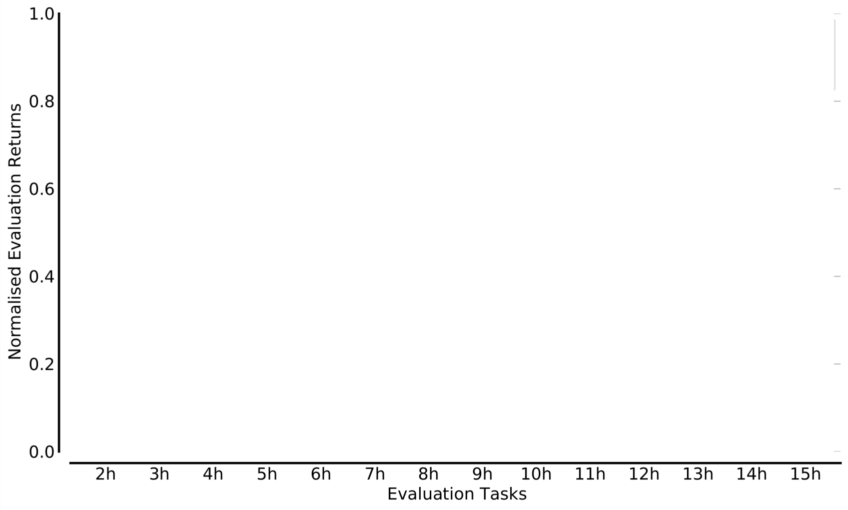
Generalisation Experiments - Domain Randomisation



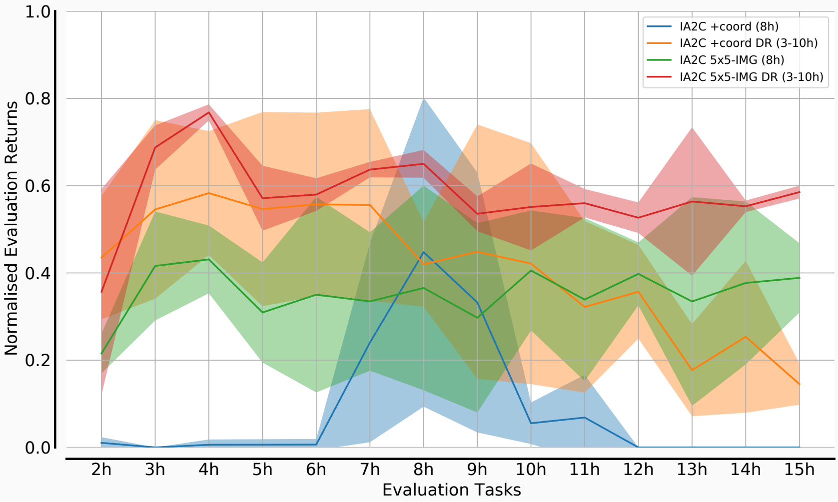
No DR: train in single task (column height of 8 - bottom left)

DR: train in set of tasks with column height 3 – 10

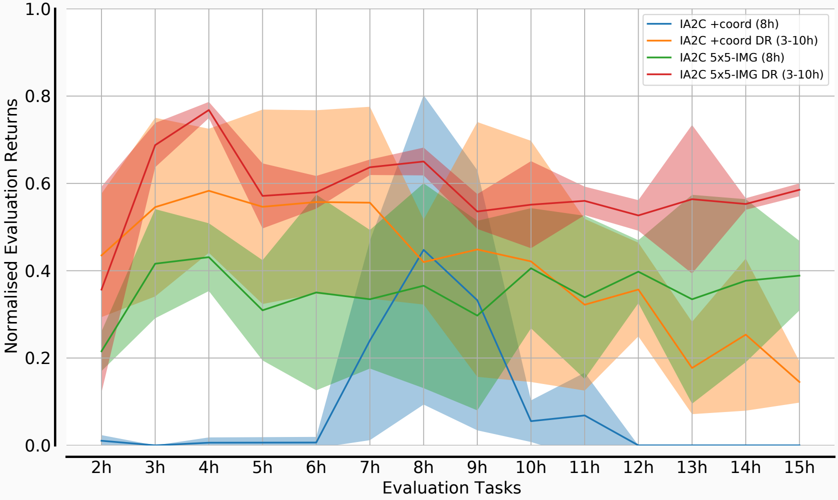
Generalisation Experiments - Observations and DR Results



Generalisation Experiments - Observations and DR Results



Generalisation Experiments - Observations and DR Results



- Observations with coordinates only generalise with DR unlike image observations
- DR improves generalisation in all cases

Generalisation Experiments - Network Architecture Results

- **Recurrent networks**
  - Commonly applied in partially observable tasks
  - Improve performance of agents (but not generalisation specific)

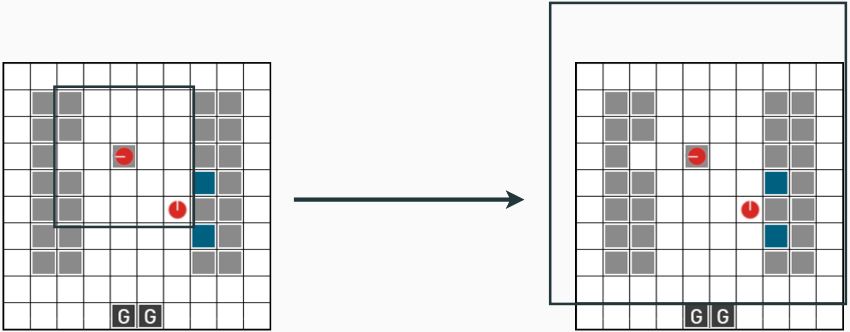
Generalisation Experiments - Network Architecture Results

- **Recurrent networks**
  - Commonly applied in partially observable tasks
  - Improve performance of agents (but not generalisation specific)
- **Convolution neural networks**
  - CNNs did not make significant difference by themselves

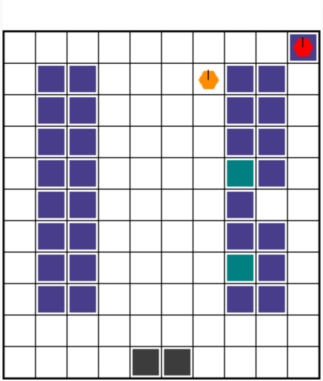
Generalisation Experiments - Network Architecture Results

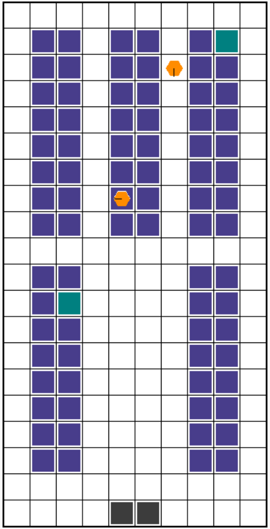
- **Recurrent networks**
  - Commonly applied in partially observable tasks
  - Improve performance of agents (but not generalisation specific)

- **Convolution neural networks**
  - CNNs did not make significant difference by themselves
  - But CNNs allowed to train on larger image observations

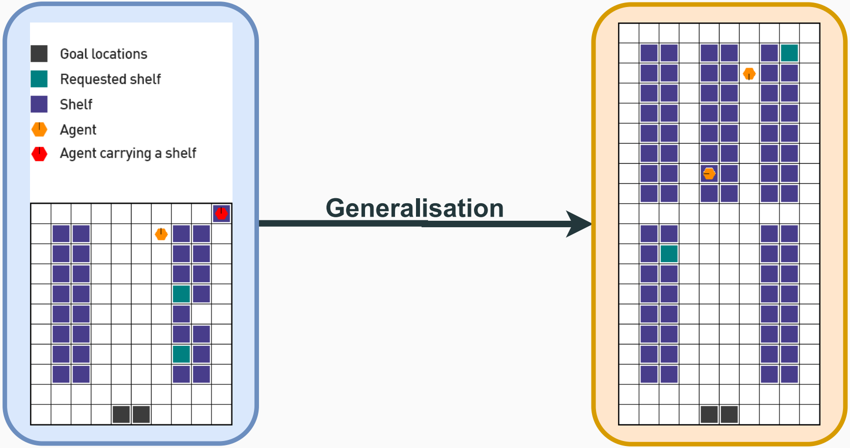


Generalisation Experiments - Failure Case





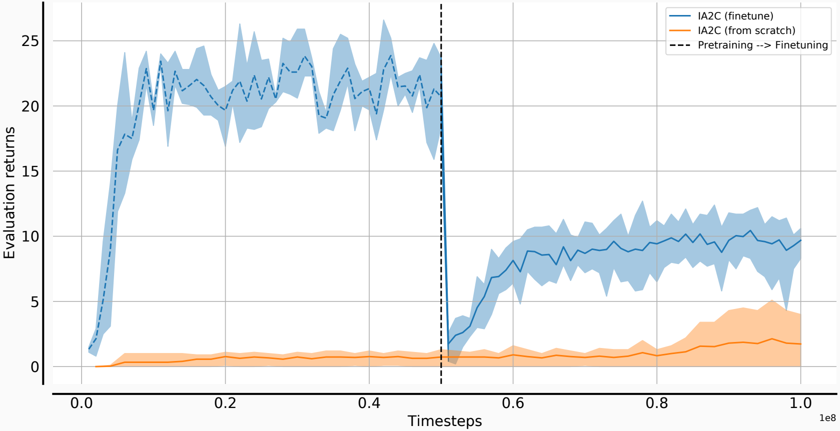
How To Bridge This Gap?



Promising directions:

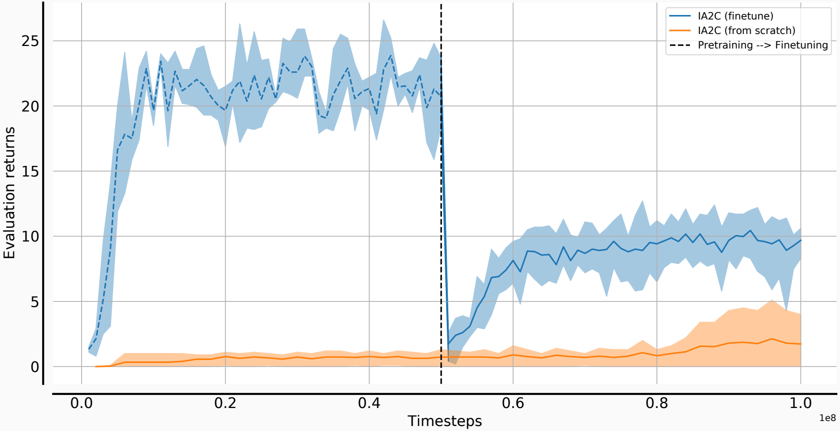
1. Reason over **high-level** information (relational/ neurosymbolic and hierarchical RL)
2. Allow for limited **finetuning** in testing tasks

Generalisation Finetuning Result



→ Opportunity for **curriculum learning**

Generalisation Finetuning Result



- Representations are valuable and generalise with limited finetuning!
- Finetuned agents outperform agents trained in harder task from scratch

$\rightarrow$ Opportunity for **curriculum learning**

4. Conclusion

Conclusion

- We demonstrated the challenge of generalisation in MARL
- Existing approaches are sensitive to task-specific details in observations
  $\rightarrow$ Zero-shot generalisation quickly reaches its limits
- Finetuning experiments demonstrate representations can generalise with limited training in new tasks
- Future directions
  1. Few-shot generalisation with finetung in testing tasks (e.g. meta RL)
  2. Condition policy on high-level information (hierarchical and relational RL, neurosymbolic models)

Feel free to reach out to me!

https://www.lukaschaefer.com/
l.schaefer@ed.ac.uk
@LukasSchaefer96