

Creating Multi-Level Skill Hierarchies in Reinforcement Learning

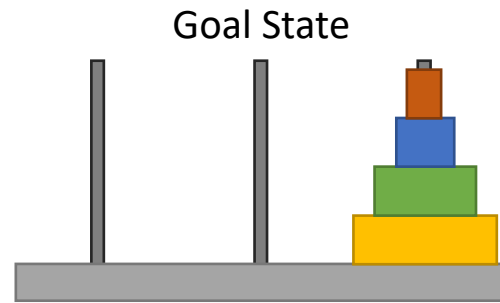
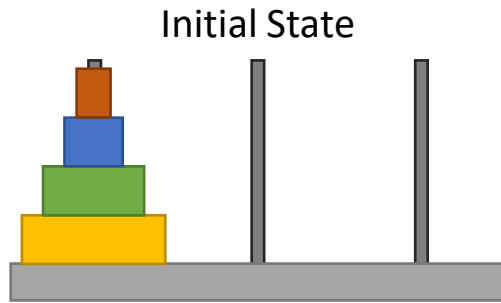
Joshua Evans and Özgür Şimşek

Bath Reinforcement Learning Laboratory

Department of Computer Science

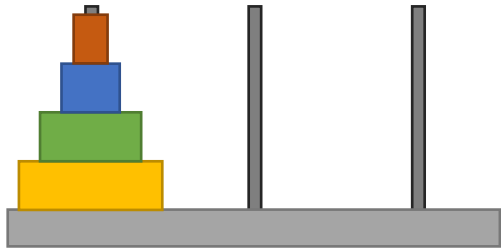


UNIVERSITY OF
BATH

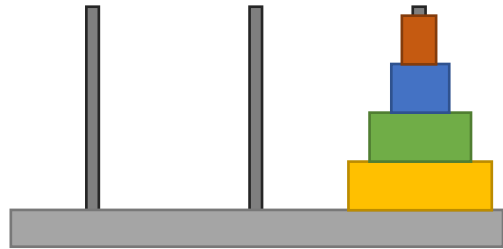


What **behaviours** would be useful?

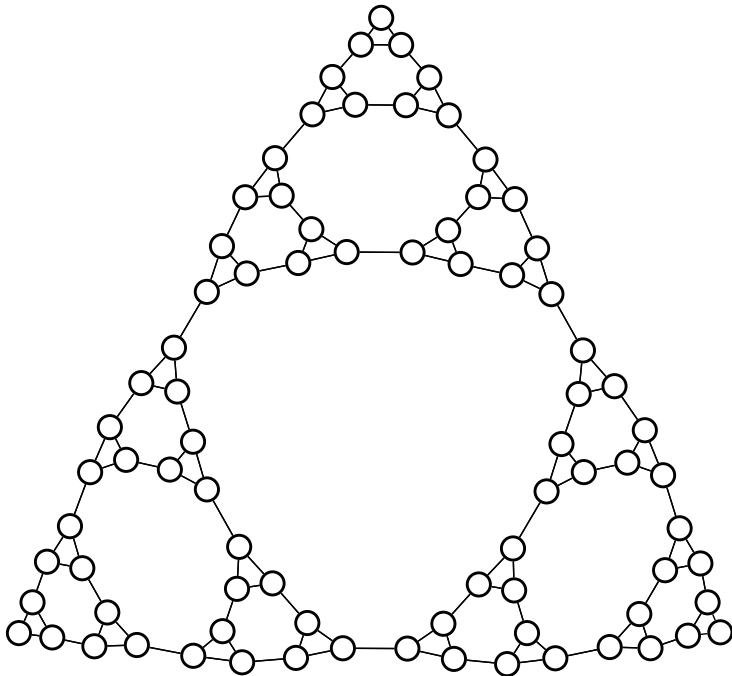
Initial State



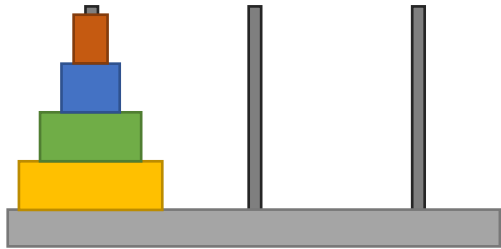
Goal State



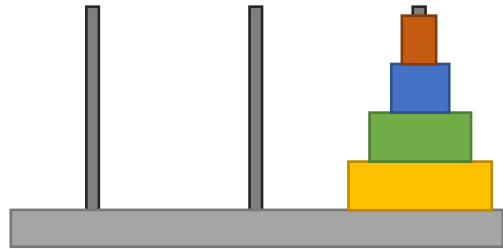
What **behaviours** would be useful?



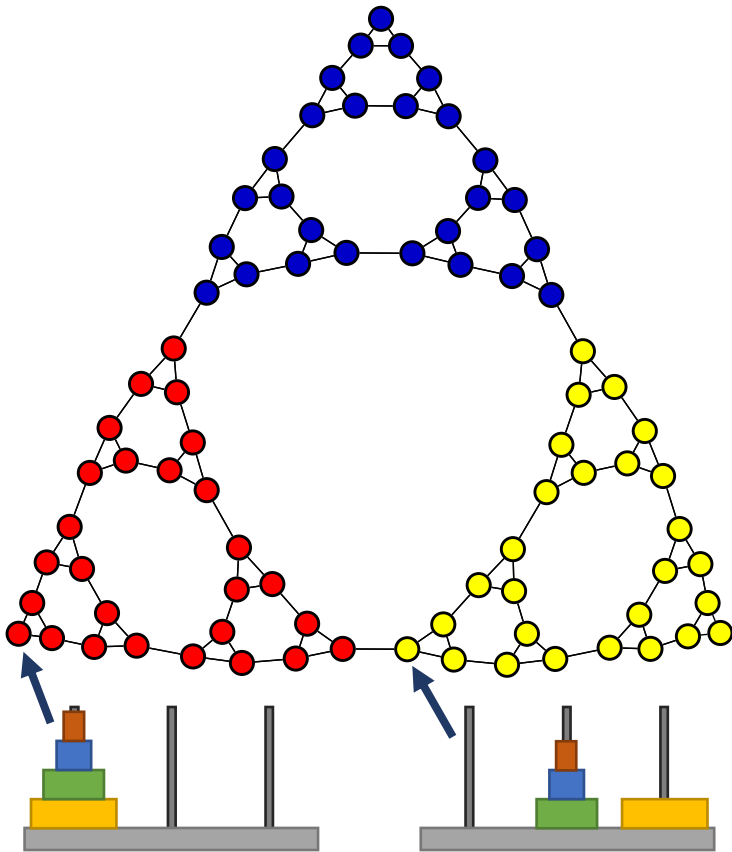
Initial State



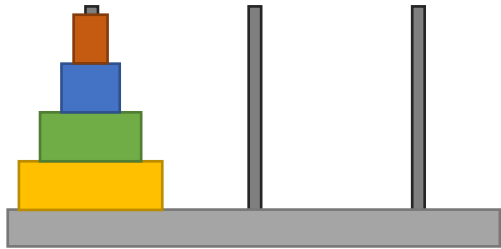
Goal State



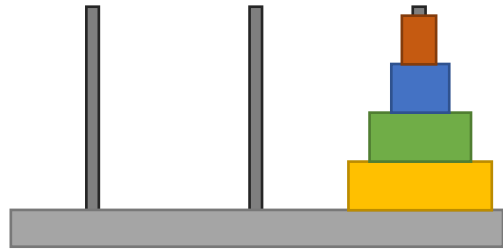
What **behaviours** would be useful?



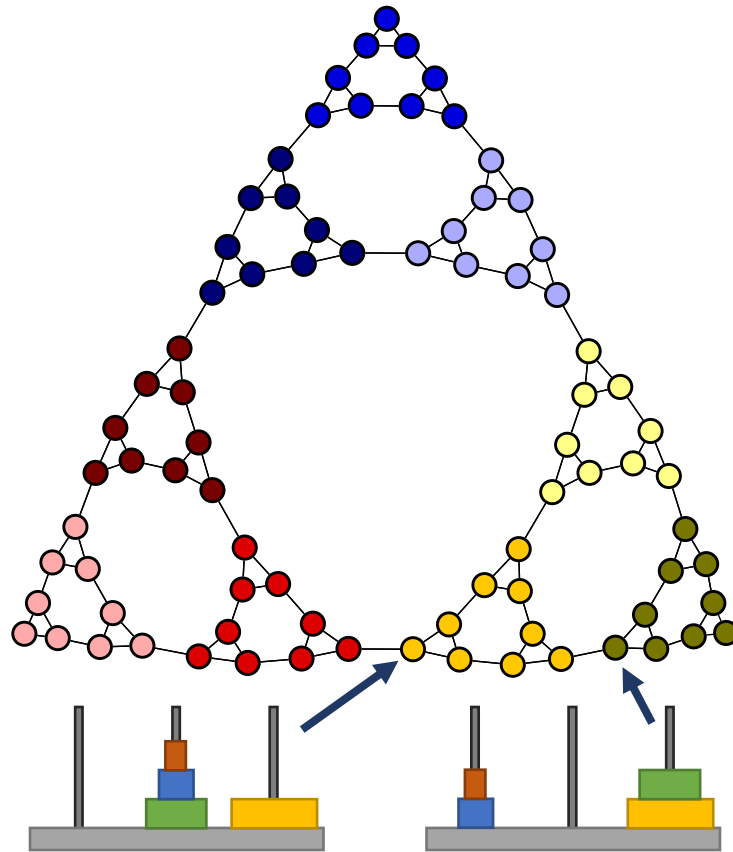
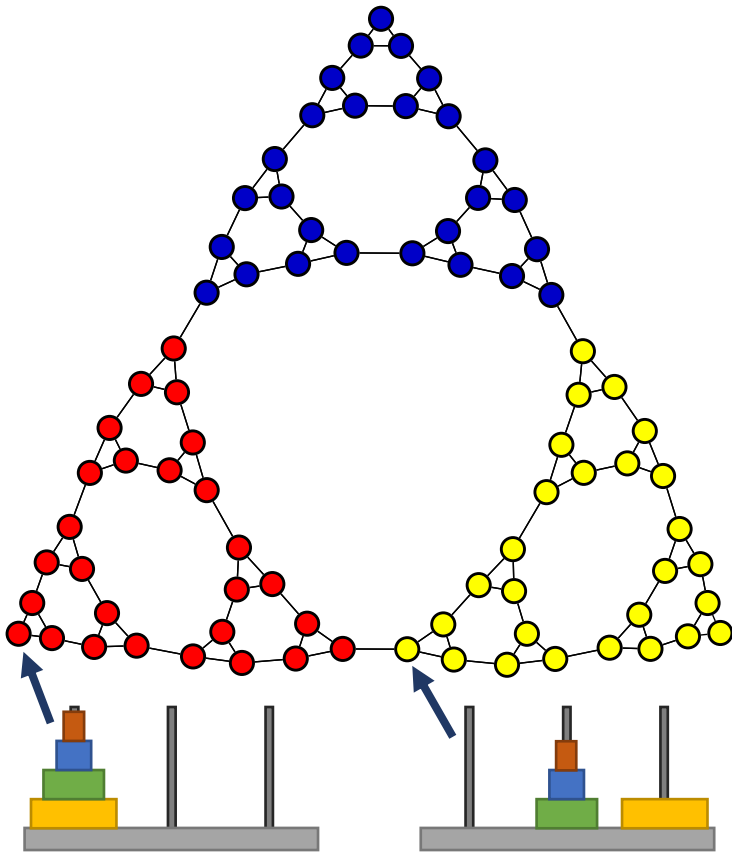
Initial State



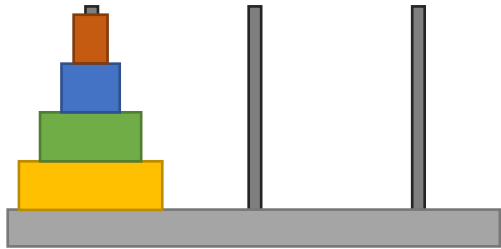
Goal State



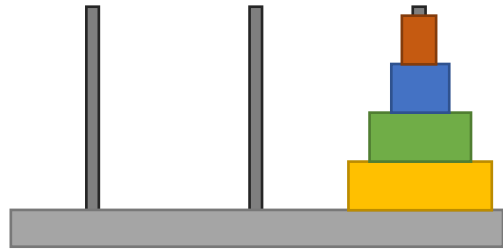
What **behaviours** would be useful?



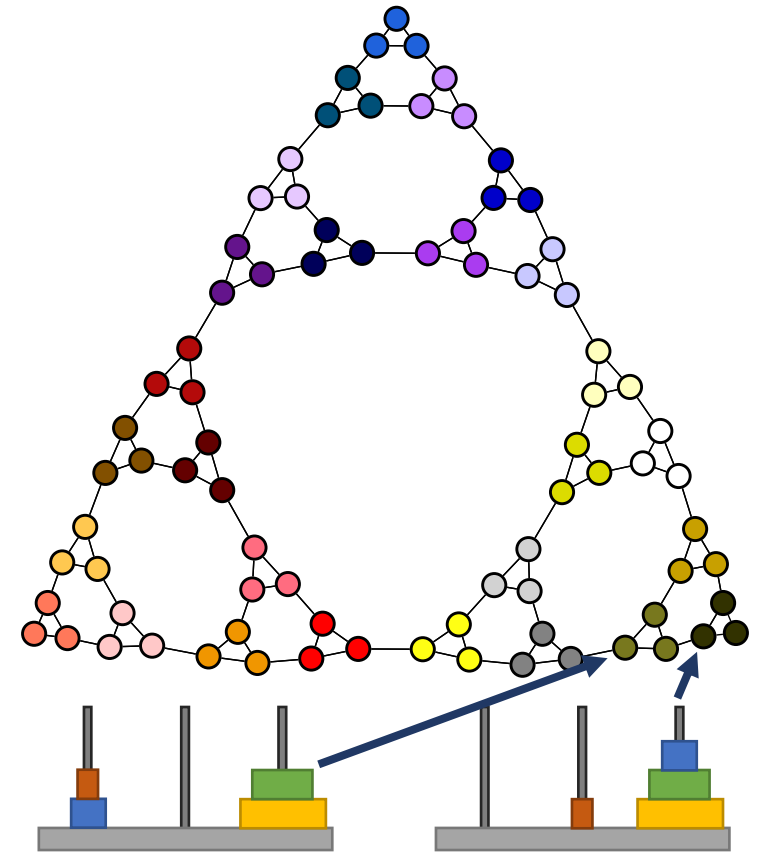
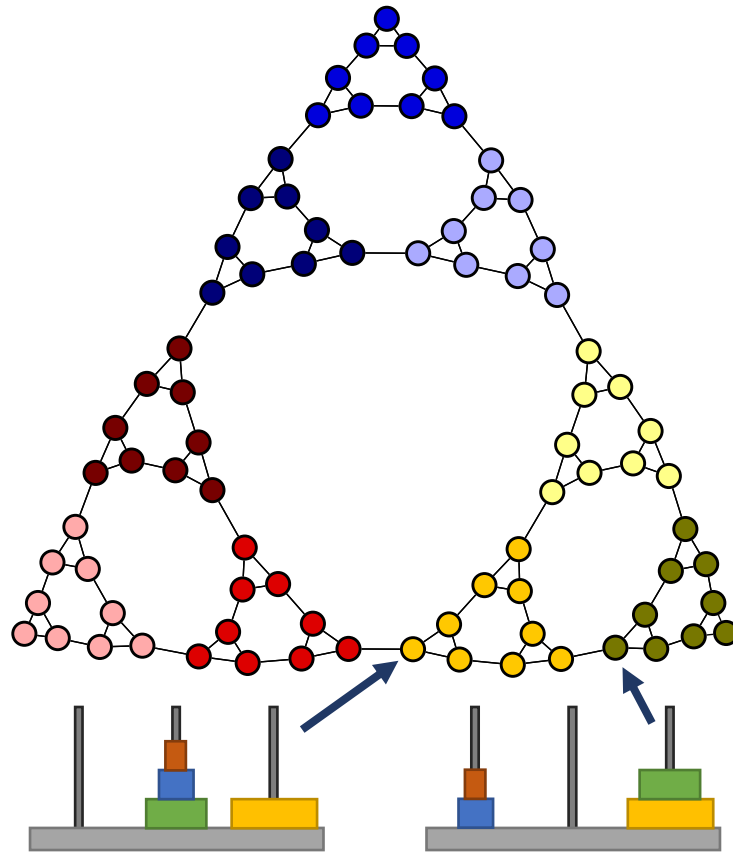
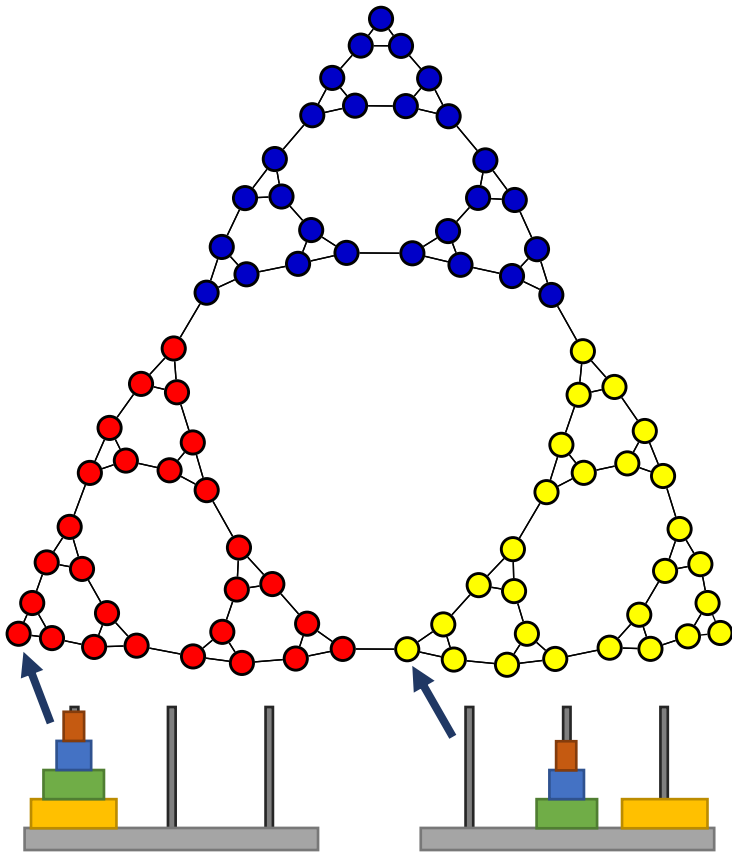
Initial State



Goal State



What **behaviours** would be useful?



Contribution

We propose a novel characterisation of a useful behaviour hierarchy.

[1] Newman, M.E. and Girvan, M., 2004. *Finding and evaluating community structure in networks*. Physical Review E.

[2] Blondel, V.D., Guillaume, J.L., Lambiotte, R. and Lefebvre, E., 2008. *Fast unfolding of communities in large networks*. Journal of Statistical Mechanics: Theory and Experiment.

Contribution

We propose a novel characterisation of a useful behaviour hierarchy.

- It uses the **graphical structure** of the agent-environment interaction.

[1] Newman, M.E. and Girvan, M., 2004. *Finding and evaluating community structure in networks*. Physical Review E.

[2] Blondel, V.D., Guillaume, J.L., Lambiotte, R. and Lefebvre, E., 2008. *Fast unfolding of communities in large networks*. Journal of Statistical Mechanics: Theory and Experiment.

Contribution

We propose a novel characterisation of a useful behaviour hierarchy.

- It uses the **graphical structure** of the agent-environment interaction.
- It finds partitions of this graph the **maximise modularity** [1].

[1] Newman, M.E. and Girvan, M., 2004. *Finding and evaluating community structure in networks*. Physical Review E.

[2] Blondel, V.D., Guillaume, J.L., Lambiotte, R. and Lefebvre, E., 2008. *Fast unfolding of communities in large networks*. Journal of Statistical Mechanics: Theory and Experiment.

Contribution

We propose a novel characterisation of a useful behaviour hierarchy.

- It uses the **graphical structure** of the agent-environment interaction.
- It finds partitions of this graph the **maximise modularity** [1].
- It does this at **multiple time scales**, leading to a multi-level hierarchy.

[1] Newman, M.E. and Girvan, M., 2004. *Finding and evaluating community structure in networks*. Physical Review E.

[2] Blondel, V.D., Guillaume, J.L., Lambiotte, R. and Lefebvre, E., 2008. *Fast unfolding of communities in large networks*. Journal of Statistical Mechanics: Theory and Experiment.

Contribution

We propose a novel characterisation of a useful behaviour hierarchy.

- It uses the **graphical structure** of the agent-environment interaction.
- It finds partitions of this graph the **maximise modularity** [1].
- It does this at **multiple time scales**, leading to a multi-level hierarchy.
- It uses the **Louvain algorithm** [2], a hierarchical graph partitioning algorithm, to expose the graph's structure at multiple levels of granularity.

[1] Newman, M.E. and Girvan, M., 2004. *Finding and evaluating community structure in networks*. Physical Review E.

[2] Blondel, V.D., Guillaume, J.L., Lambiotte, R. and Lefebvre, E., 2008. *Fast unfolding of communities in large networks*. Journal of Statistical Mechanics: Theory and Experiment.

Contribution

We propose a novel characterisation of a useful behaviour hierarchy.

- It uses the **graphical structure** of the agent-environment interaction.
- It finds partitions of this graph the **maximise modularity** [1].
- It does this at **multiple time scales**, leading to a multi-level hierarchy.
- It uses the **Louvain algorithm** [2], a hierarchical graph partitioning algorithm, to expose the graph's structure at multiple levels of granularity.

The result is a behaviour hierarchy with four desirable properties:

[1] Newman, M.E. and Girvan, M., 2004. *Finding and evaluating community structure in networks*. Physical Review E.

[2] Blondel, V.D., Guillaume, J.L., Lambiotte, R. and Lefebvre, E., 2008. *Fast unfolding of communities in large networks*. Journal of Statistical Mechanics: Theory and Experiment.

Contribution

We propose a novel characterisation of a useful behaviour hierarchy.

- It uses the **graphical structure** of the agent-environment interaction.
- It finds partitions of this graph the **maximise modularity** [1].
- It does this at **multiple time scales**, leading to a multi-level hierarchy.
- It uses the **Louvain algorithm** [2], a hierarchical graph partitioning algorithm, to expose the graph's structure at multiple levels of granularity.

The result is a behaviour hierarchy with four desirable properties:

1. It contains behaviours that operate over **multiple time scales**.

[1] Newman, M.E. and Girvan, M., 2004. *Finding and evaluating community structure in networks*. Physical Review E.

[2] Blondel, V.D., Guillaume, J.L., Lambiotte, R. and Lefebvre, E., 2008. *Fast unfolding of communities in large networks*. Journal of Statistical Mechanics: Theory and Experiment.

Contribution

We propose a novel characterisation of a useful behaviour hierarchy.

- It uses the **graphical structure** of the agent-environment interaction.
- It finds partitions of this graph the **maximise modularity** [1].
- It does this at **multiple time scales**, leading to a multi-level hierarchy.
- It uses the **Louvain algorithm** [2], a hierarchical graph partitioning algorithm, to expose the graph's structure at multiple levels of granularity.

The result is a behaviour hierarchy with four desirable properties:

1. It contains behaviours that operate over **multiple time scales**.
2. It contains behaviours that naturally form a **hierarchy**.

[1] Newman, M.E. and Girvan, M., 2004. *Finding and evaluating community structure in networks*. Physical Review E.

[2] Blondel, V.D., Guillaume, J.L., Lambiotte, R. and Lefebvre, E., 2008. *Fast unfolding of communities in large networks*. Journal of Statistical Mechanics: Theory and Experiment.

Contribution

We propose a novel characterisation of a useful behaviour hierarchy.

- It uses the **graphical structure** of the agent-environment interaction.
- It finds partitions of this graph the **maximise modularity** [1].
- It does this at **multiple time scales**, leading to a multi-level hierarchy.
- It uses the **Louvain algorithm** [2], a hierarchical graph partitioning algorithm, to expose the graph's structure at multiple levels of granularity.

The result is a behaviour hierarchy with four desirable properties:

1. It contains behaviours that operate over **multiple time scales**.
2. It contains behaviours that naturally form a **hierarchy**.
3. It contains behaviours that are **specified fully**.

[1] Newman, M.E. and Girvan, M., 2004. *Finding and evaluating community structure in networks*. Physical Review E.

[2] Blondel, V.D., Guillaume, J.L., Lambiotte, R. and Lefebvre, E., 2008. *Fast unfolding of communities in large networks*. Journal of Statistical Mechanics: Theory and Experiment.

Contribution

We propose a novel characterisation of a useful behaviour hierarchy.

- It uses the **graphical structure** of the agent-environment interaction.
- It finds partitions of this graph the **maximise modularity** [1].
- It does this at **multiple time scales**, leading to a multi-level hierarchy.
- It uses the **Louvain algorithm** [2], a hierarchical graph partitioning algorithm, to expose the graph's structure at multiple levels of granularity.

The result is a behaviour hierarchy with four desirable properties:

1. It contains behaviours that operate over **multiple time scales**.
2. It contains behaviours that naturally form a **hierarchy**.
3. It contains behaviours that are **specified fully**.
4. It is generated **automatically**, with no human input.

[1] Newman, M.E. and Girvan, M., 2004. *Finding and evaluating community structure in networks*. Physical Review E.

[2] Blondel, V.D., Guillaume, J.L., Lambiotte, R. and Lefebvre, E., 2008. *Fast unfolding of communities in large networks*. Journal of Statistical Mechanics: Theory and Experiment.

R				G
Y			B	

The taxi starts at a random square.

The passenger starts at R, G, B, or Y, with destination R, G, B, or Y.

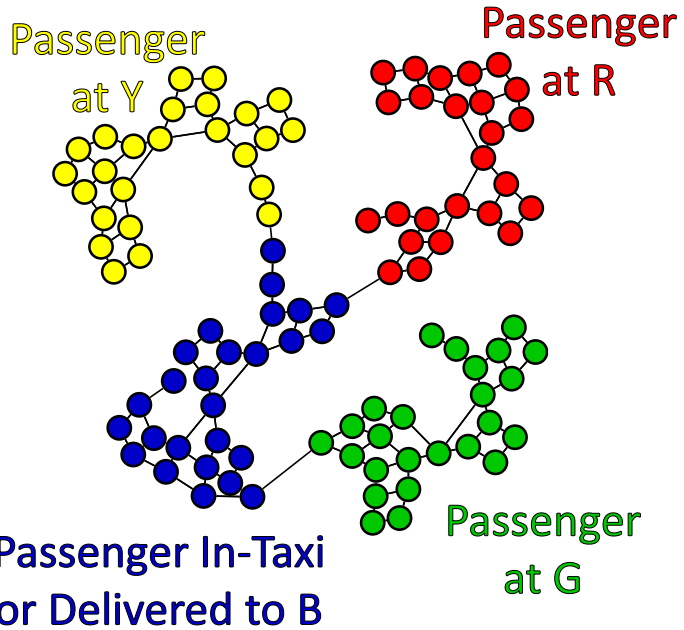
Actions: North, South, East, West, Pick Up Passenger, Put Down Passenger.

R				G
Y			B	

The taxi starts at a random square.

The passenger starts at R, G, B, or Y, with destination R, G, B, or Y.

Actions: North, South, East, West, Pick Up Passenger, Put Down Passenger.

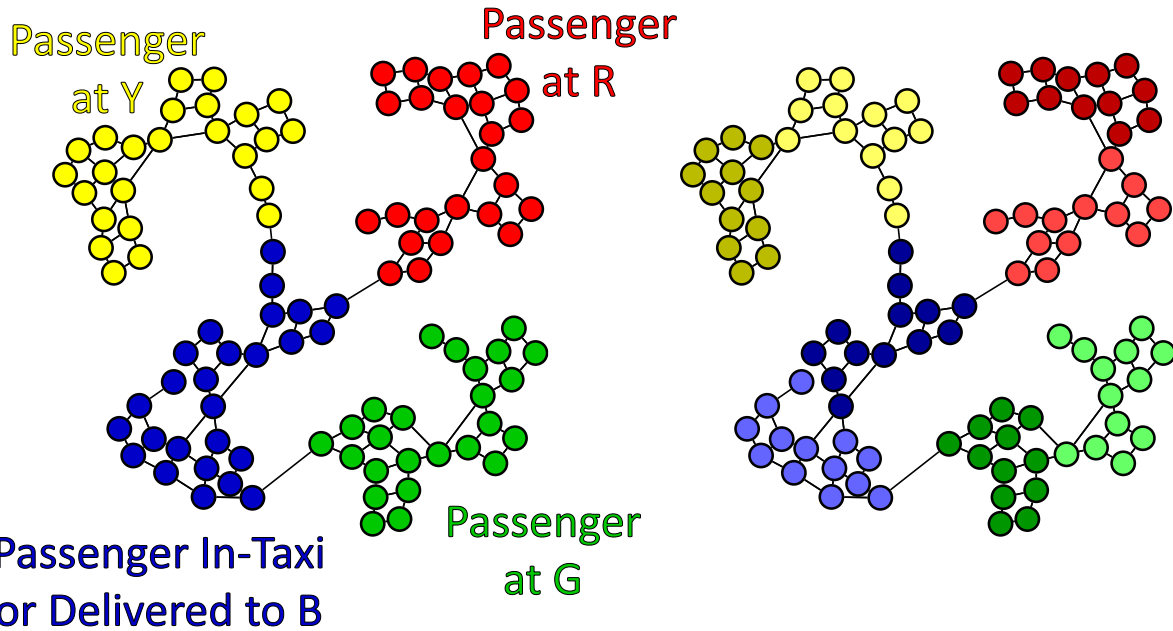


R				G
Y			B	

The taxi starts at a random square.

The passenger starts at R, G, B, or Y, with destination R, G, B, or Y.

Actions: North, South, East, West, Pick Up Passenger, Put Down Passenger.

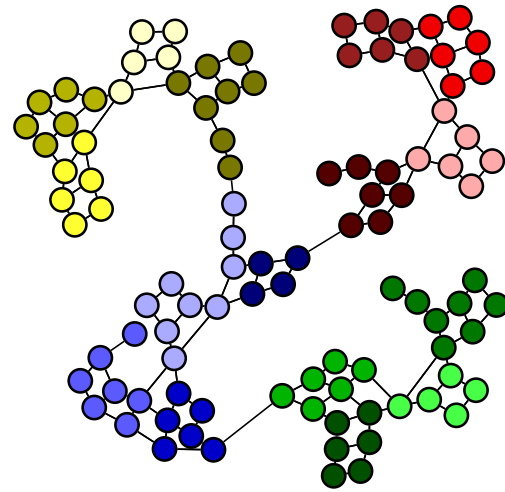
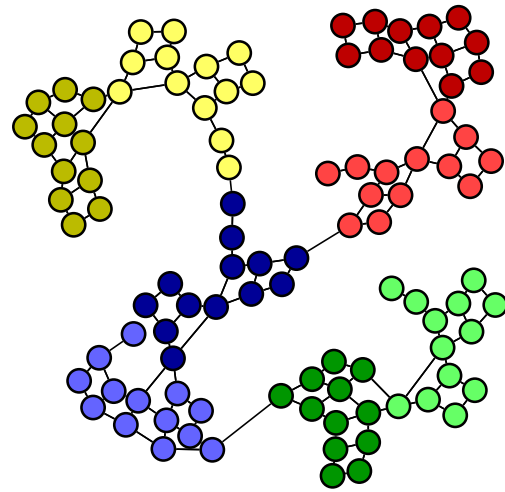
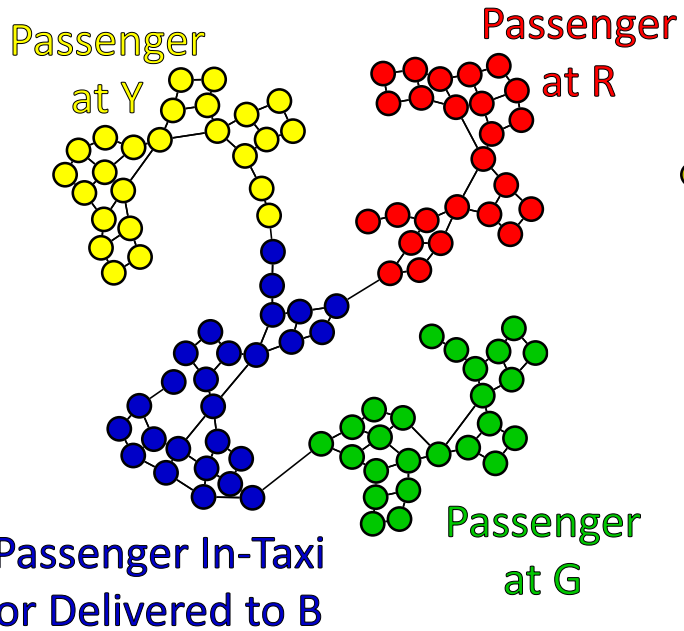


R				G
Y			B	

The taxi starts at a random square.

The passenger starts at R, G, B, or Y, with destination R, G, B, or Y.

Actions: North, South, East, West, Pick Up Passenger, Put Down Passenger.

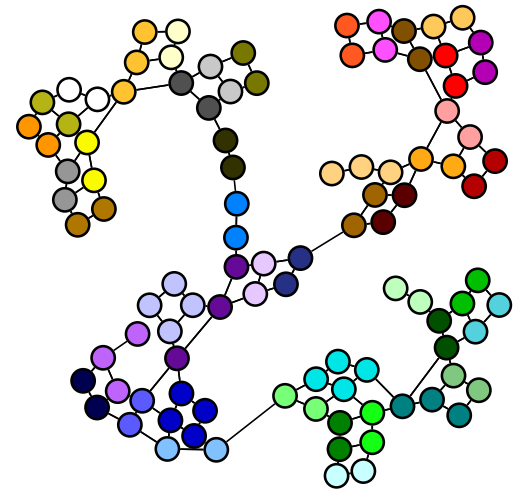
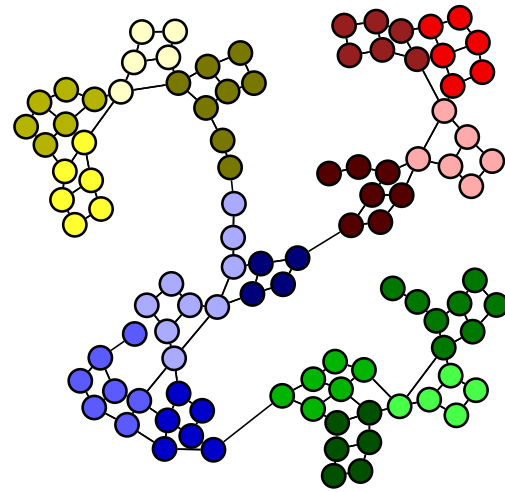
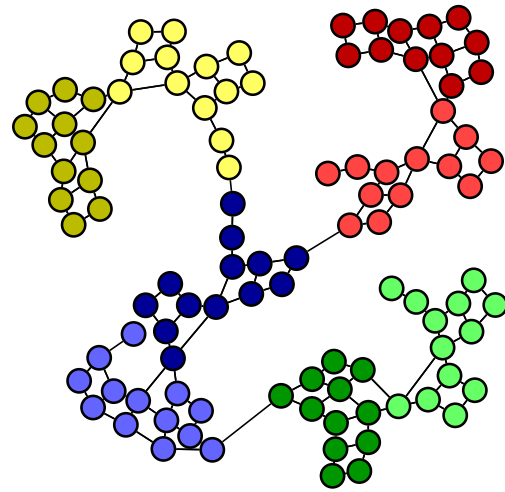
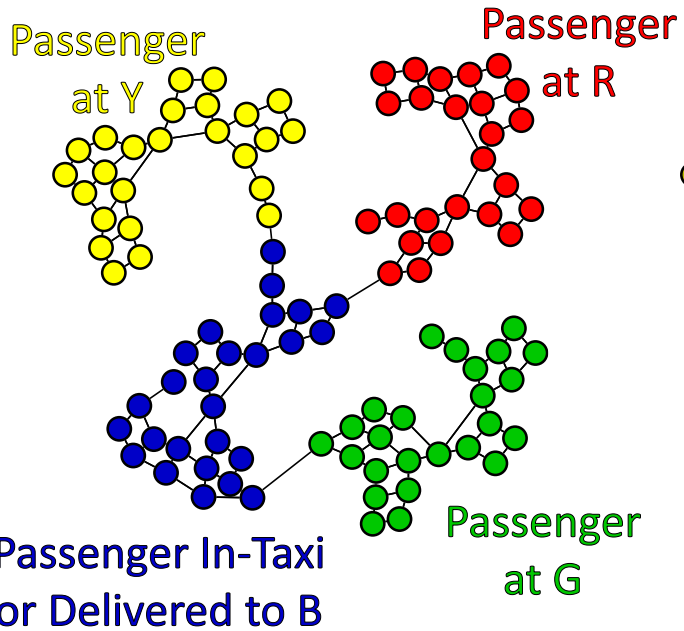


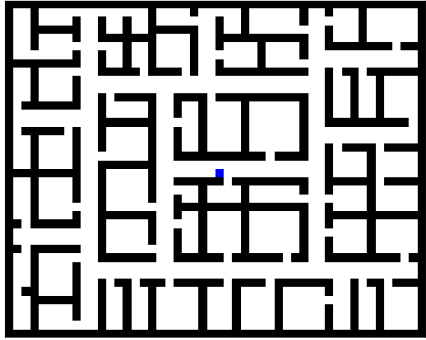
R				G
Y			B	

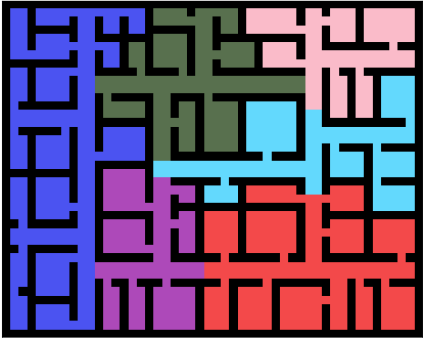
The taxi starts at a random square.

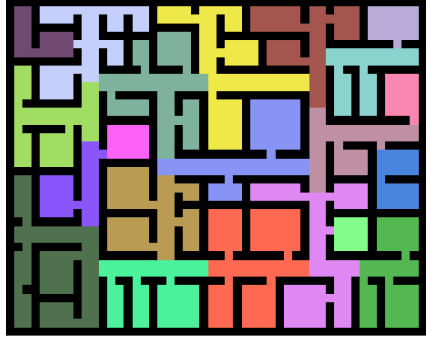
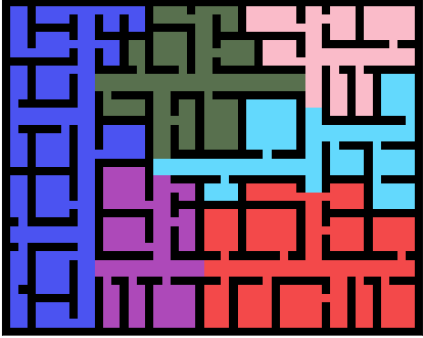
The passenger starts at R, G, B, or Y, with destination R, G, B, or Y.

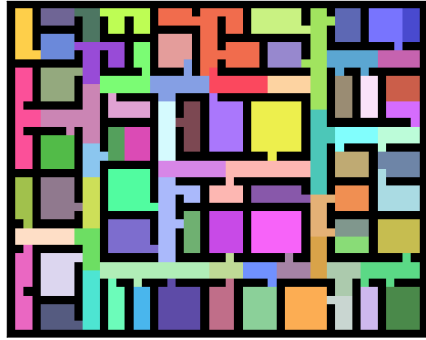
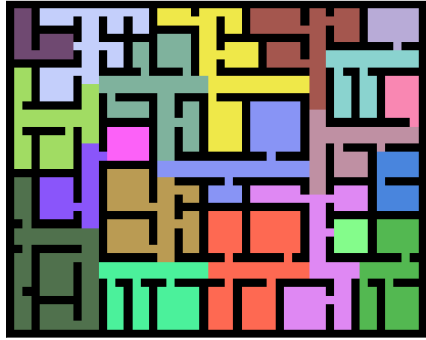
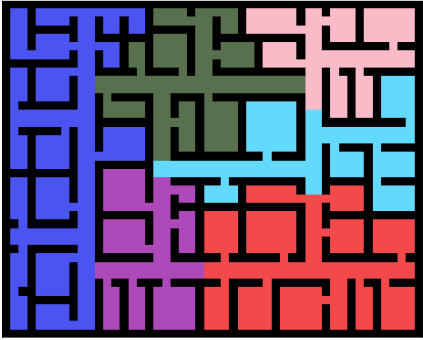
Actions: North, South, East, West, Pick Up Passenger, Put Down Passenger.

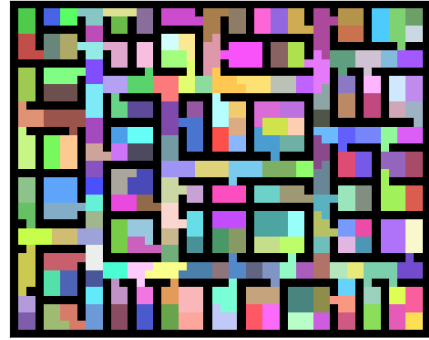
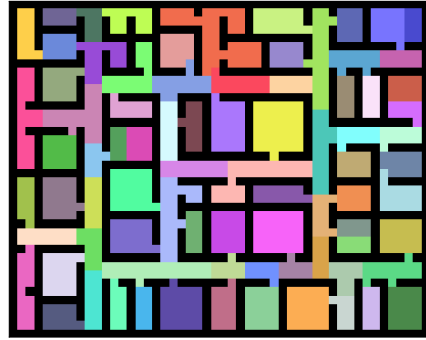
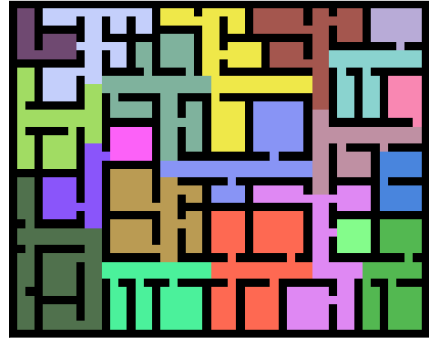
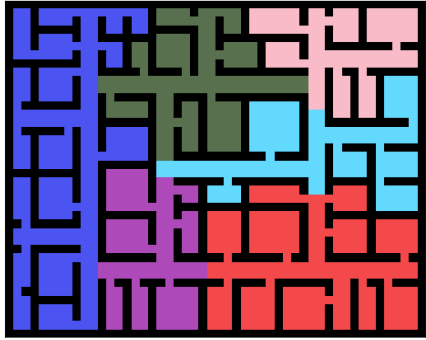


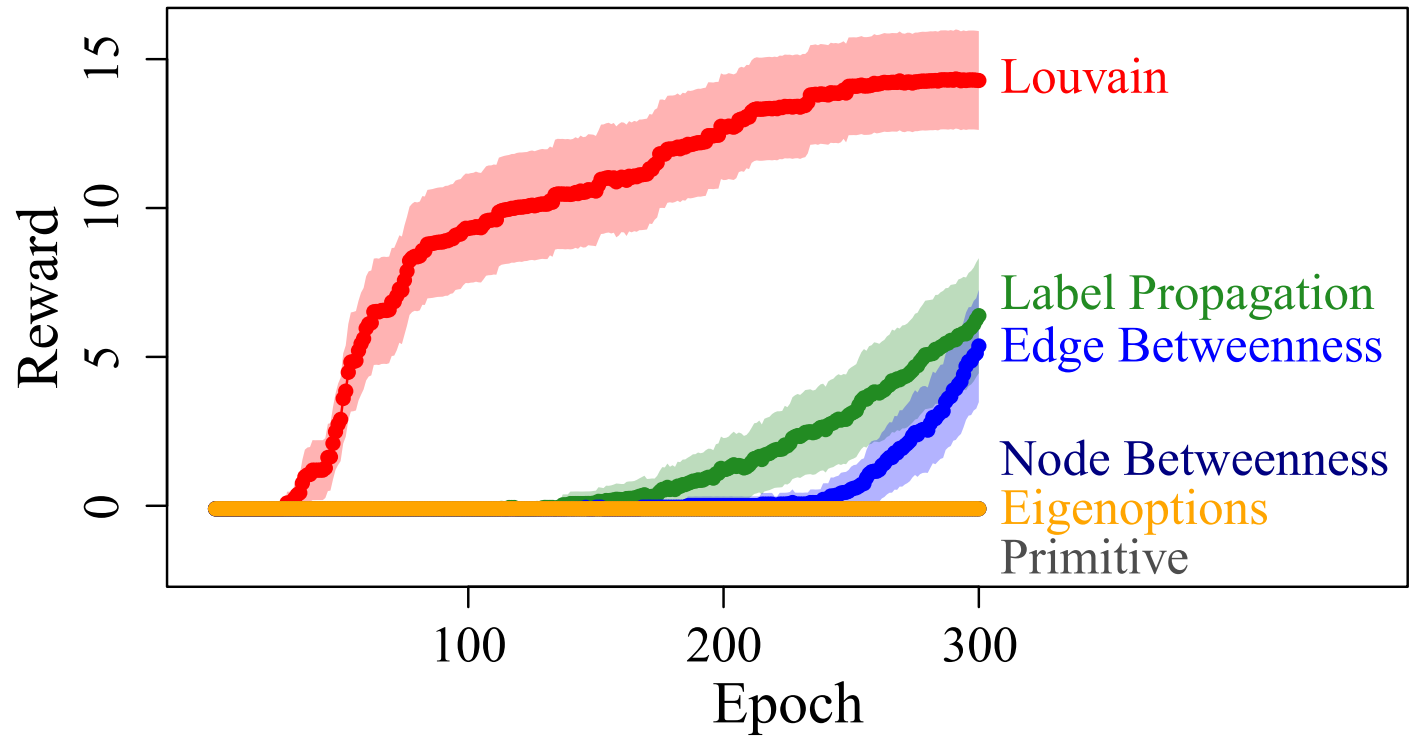
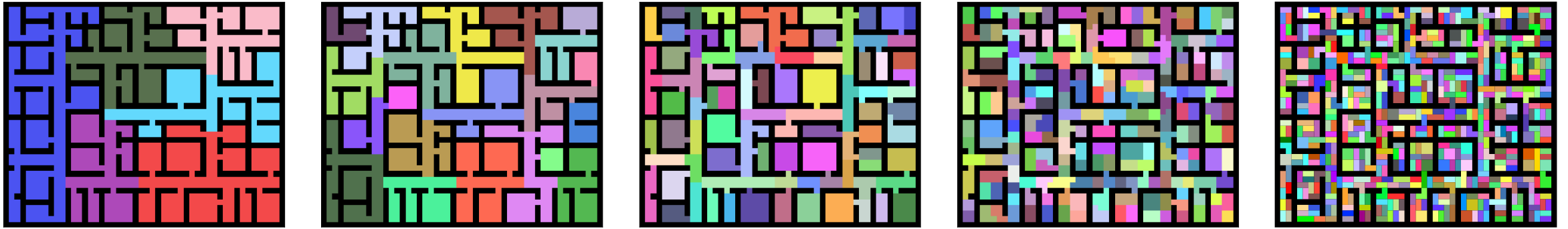


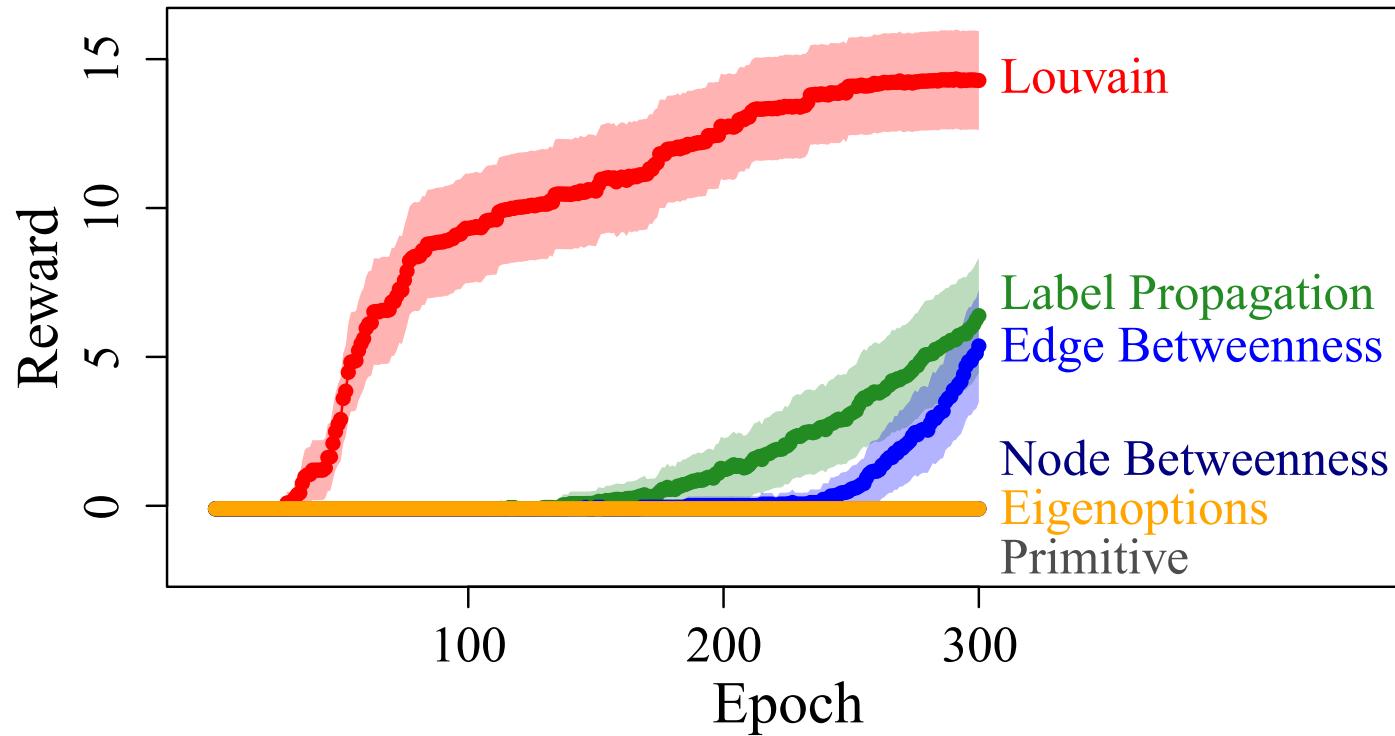
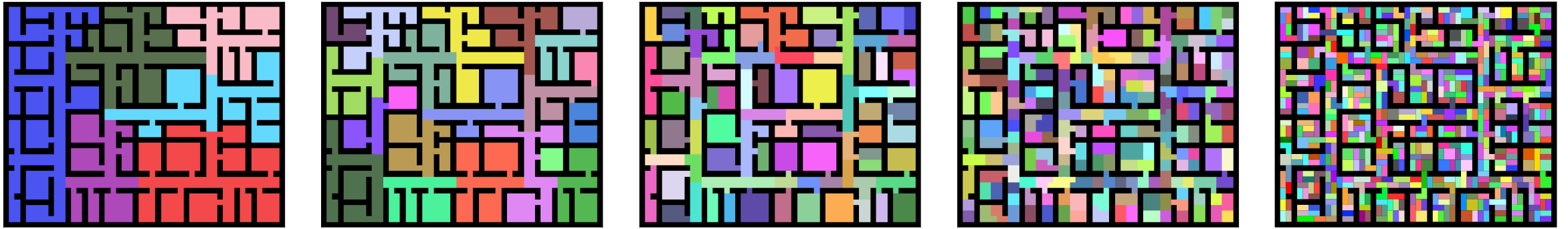








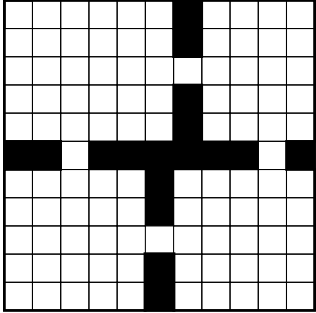




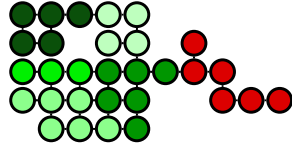
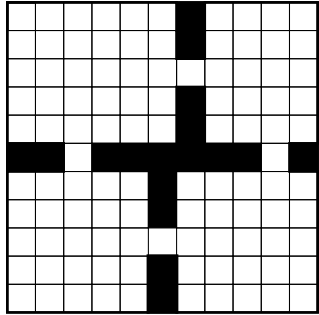
- The proposed behaviour hierarchy:
- **closely matches human intuition, and**
 - **improves learning performance.**

Incremental Skill Discovery

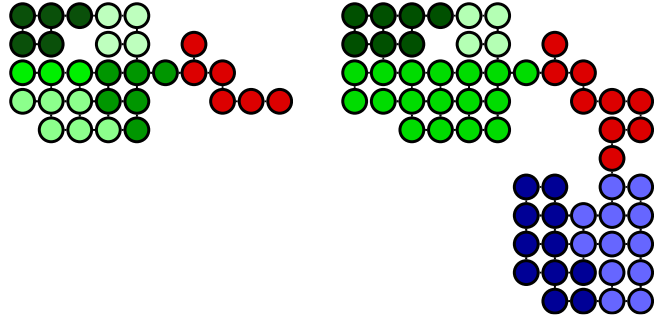
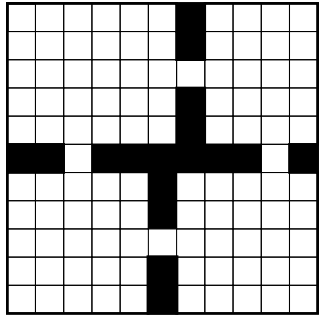
Incremental Skill Discovery



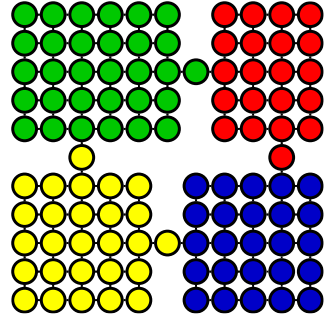
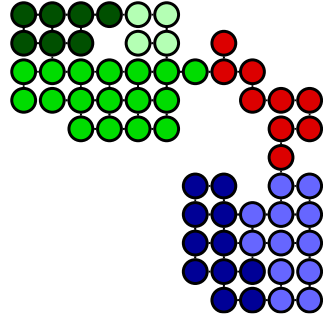
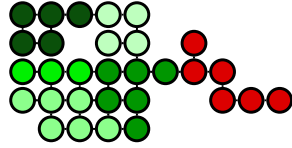
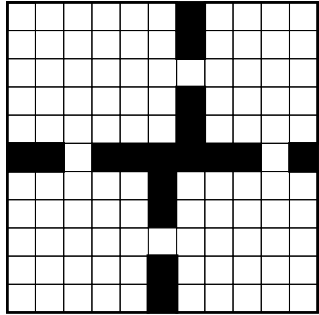
Incremental Skill Discovery



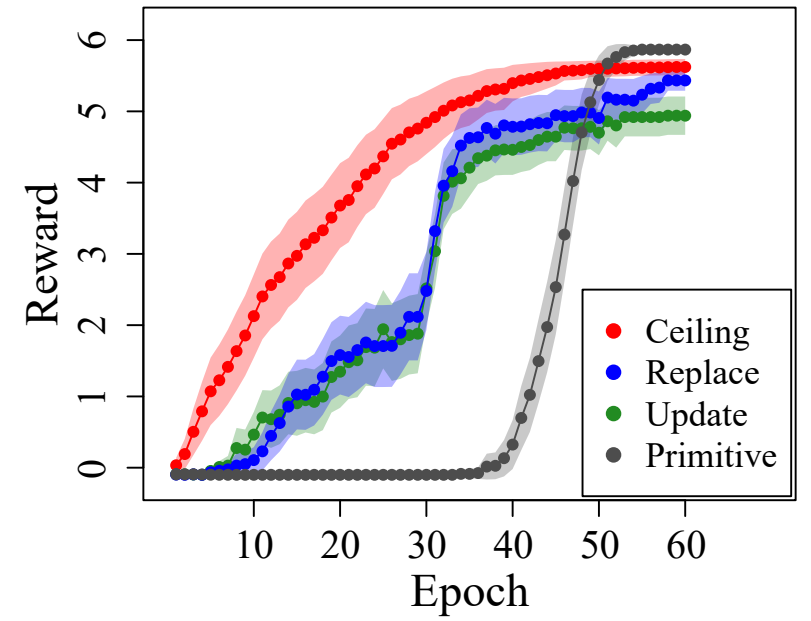
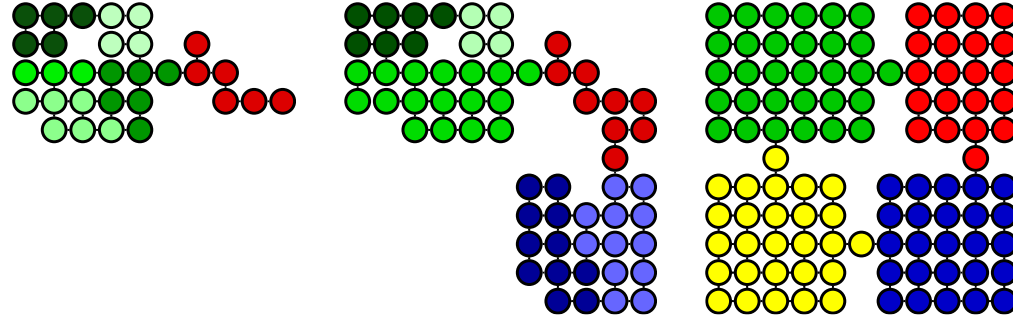
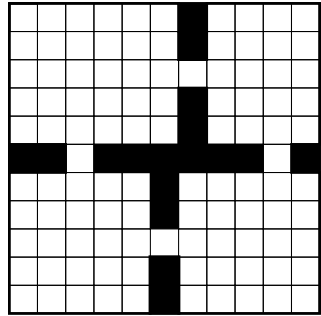
Incremental Skill Discovery



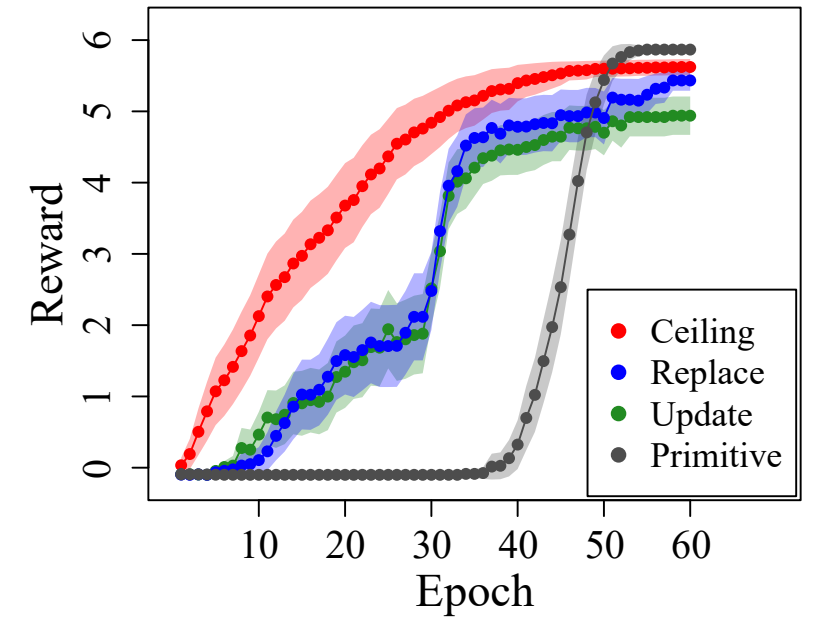
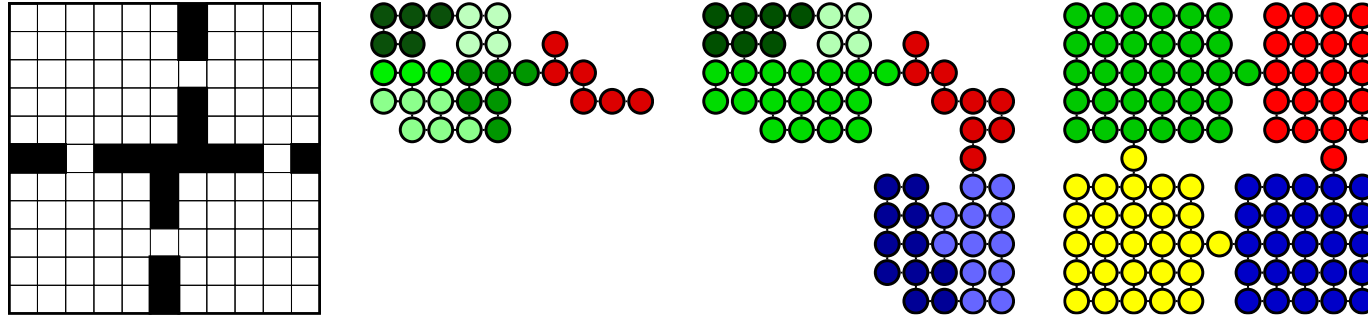
Incremental Skill Discovery



Incremental Skill Discovery

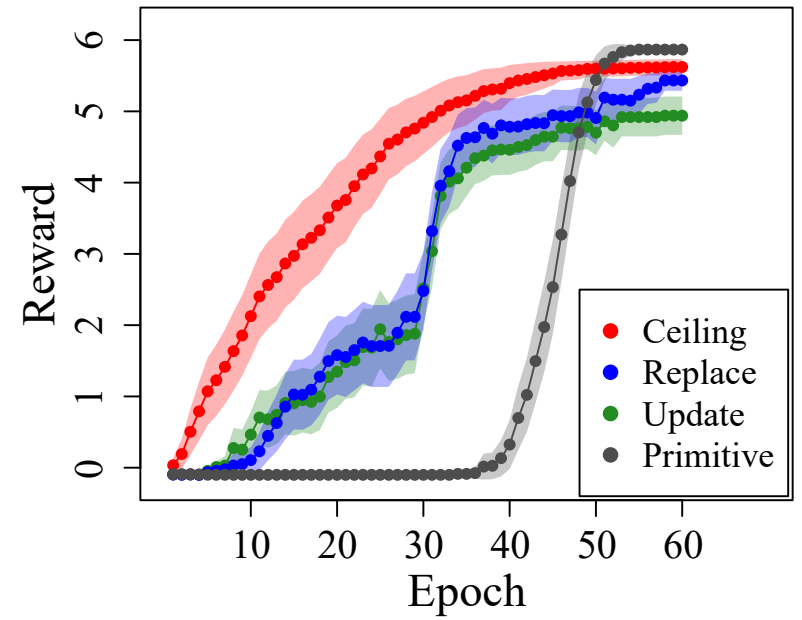
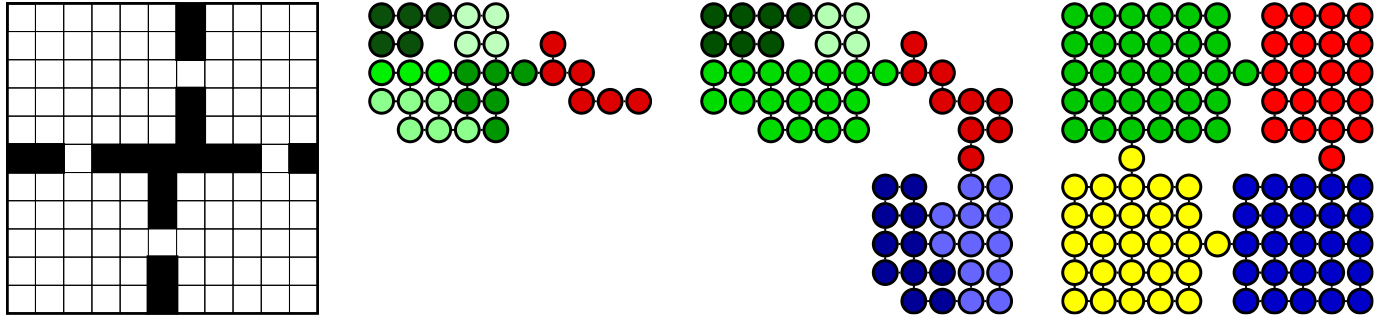


Incremental Skill Discovery

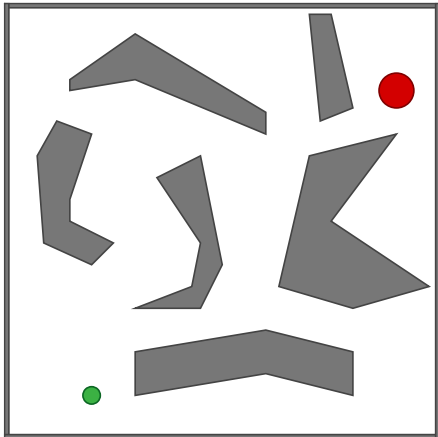


Continuous State-Spaces

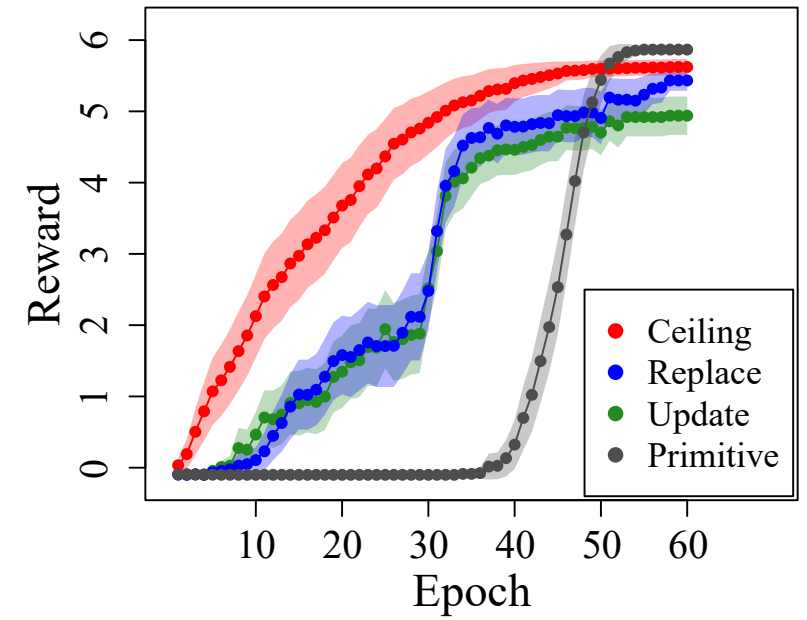
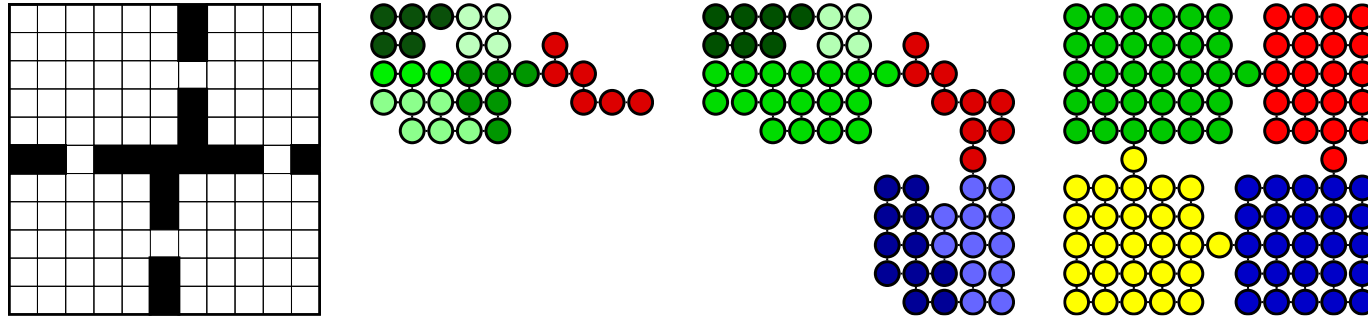
Incremental Skill Discovery



Continuous State-Spaces



Incremental Skill Discovery



Continuous State-Spaces

