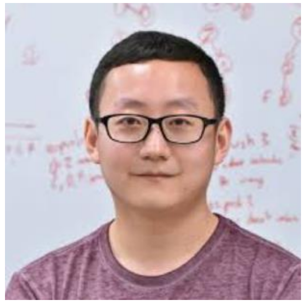
Discovering Generalizable Spatial Goal Representations via Graph-based Active Reward Learning

Aviv Netanyahu*



Tianmin Shu*

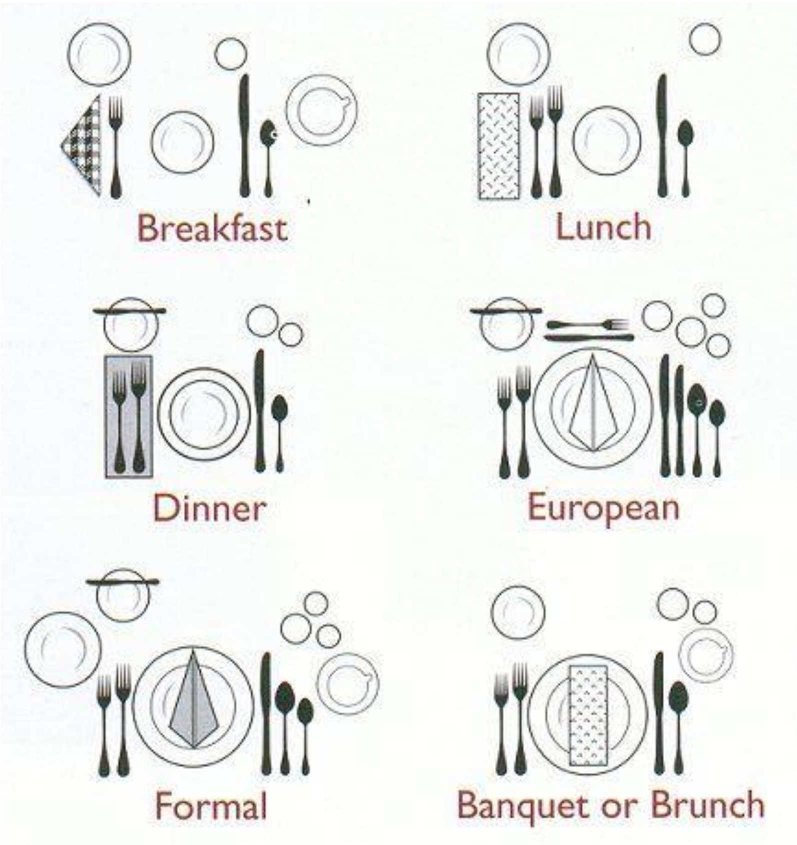
Josh Tenenbaum

Pulkit Agrawal

Massachusetts Institute of Technology

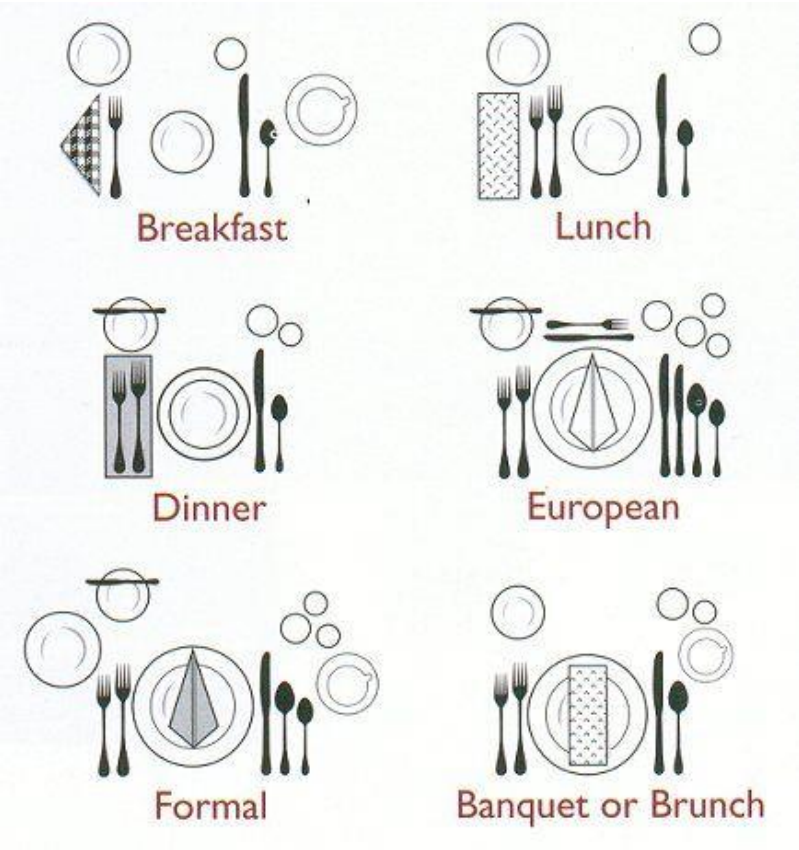
(* equal contribution)

AI assistants need to adapt to *any* task



AI assistants need to adapt to *any* task

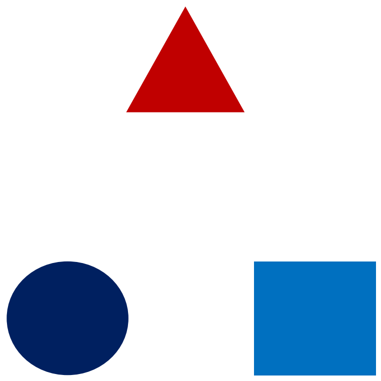
Key: **Learn the goal specification for any new task**



This work

Spatial goals for object rearrangement tasks

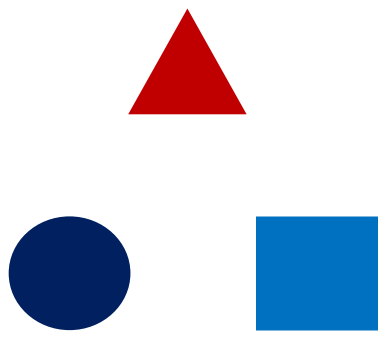
A Expert Demonstration



This work

Spatial goals for object rearrangement tasks

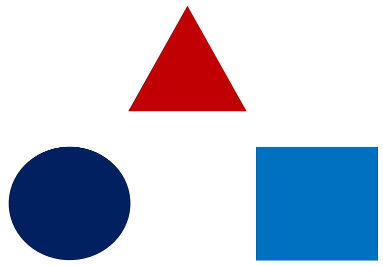
A Expert Demonstration



This work

Spatial goals for object rearrangement tasks

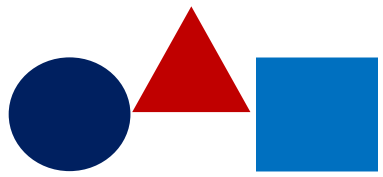
A Expert Demonstration



This work

Spatial goals for object rearrangement tasks

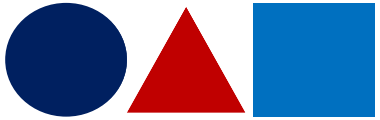
A Expert Demonstration



This work

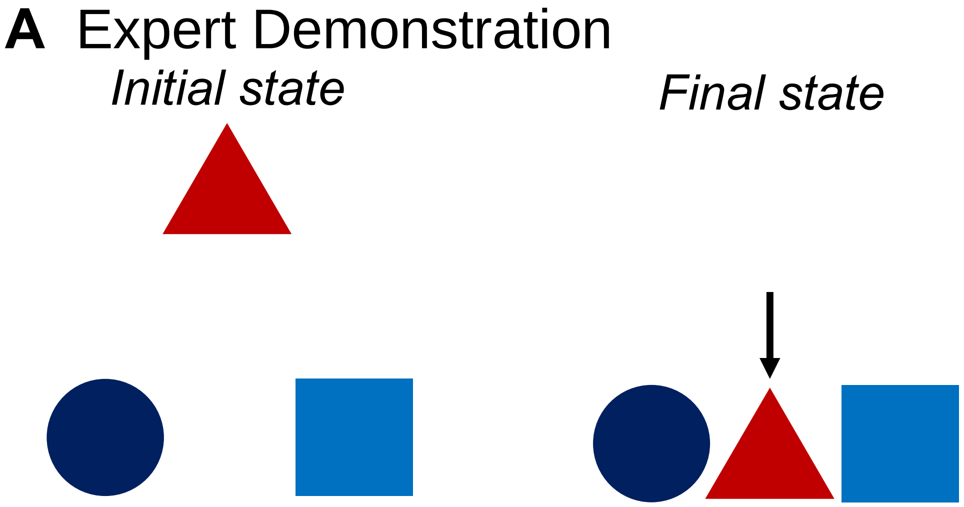
Spatial goals for object rearrangement tasks

A Expert Demonstration



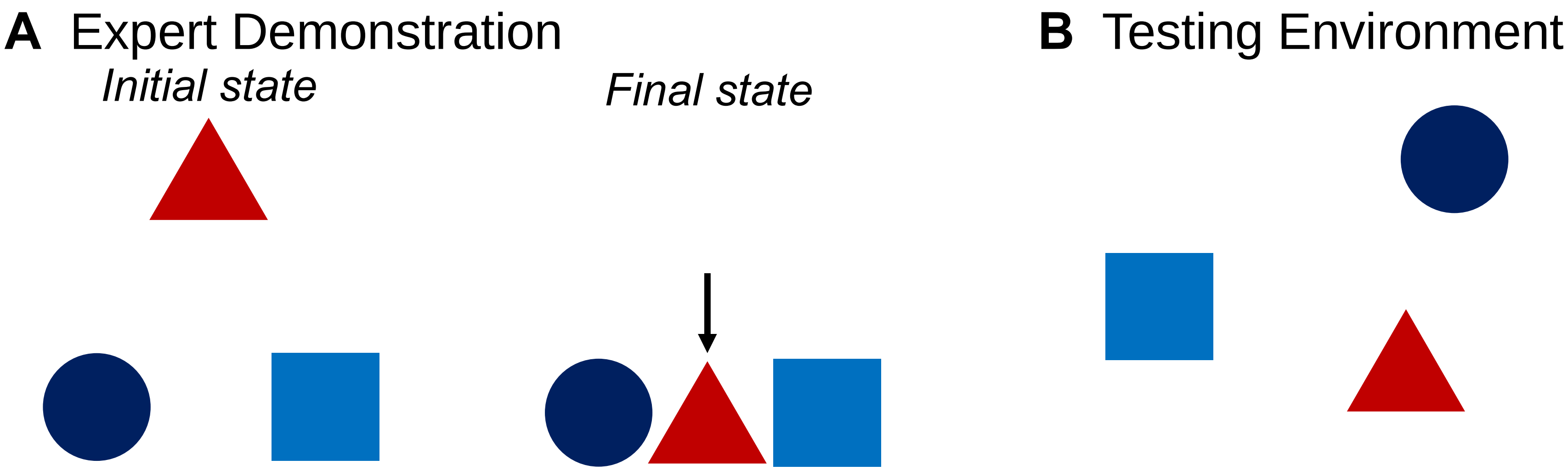
This work

Spatial goals for object rearrangement tasks



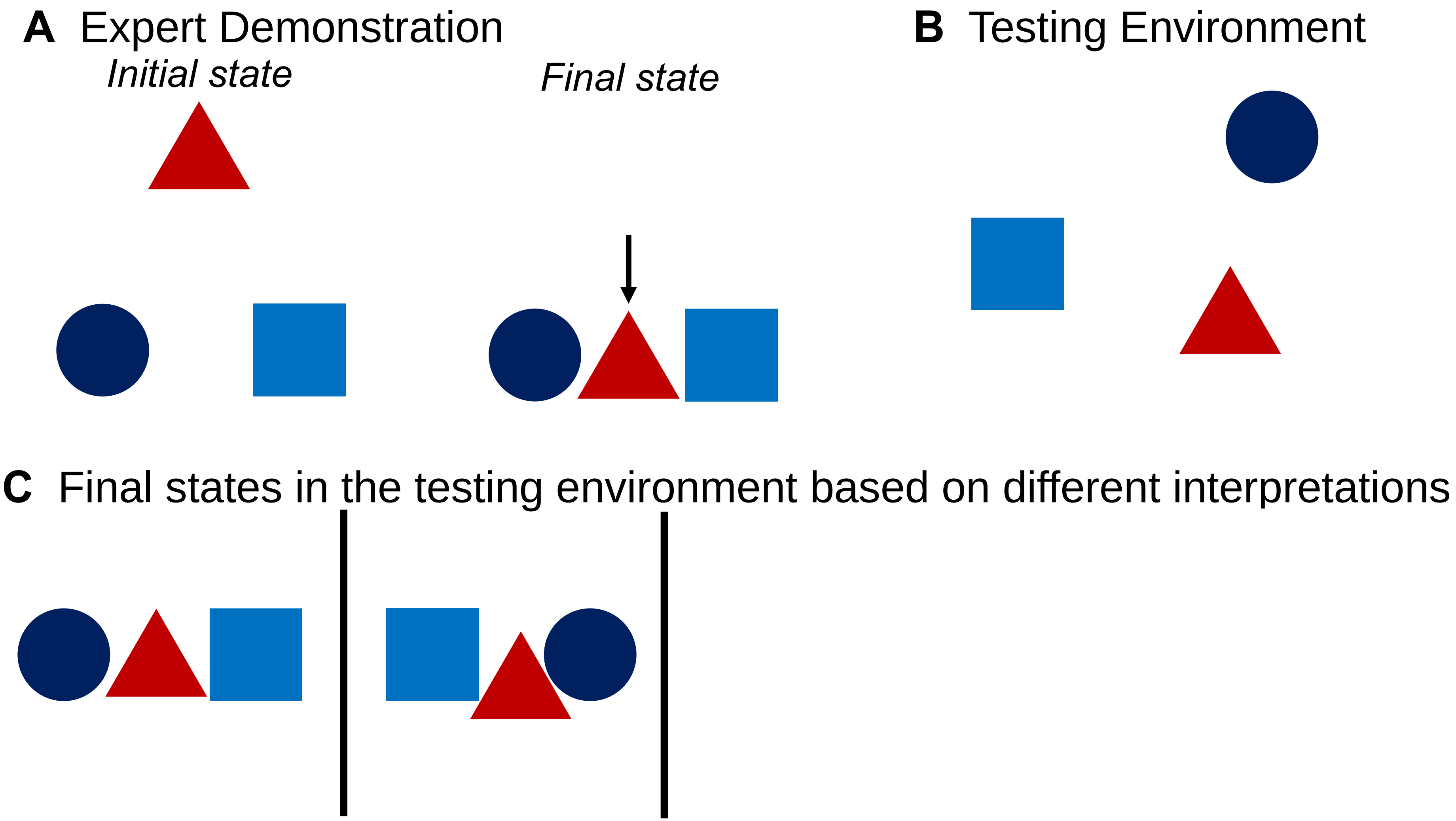
This work

Spatial goals for object rearrangement tasks



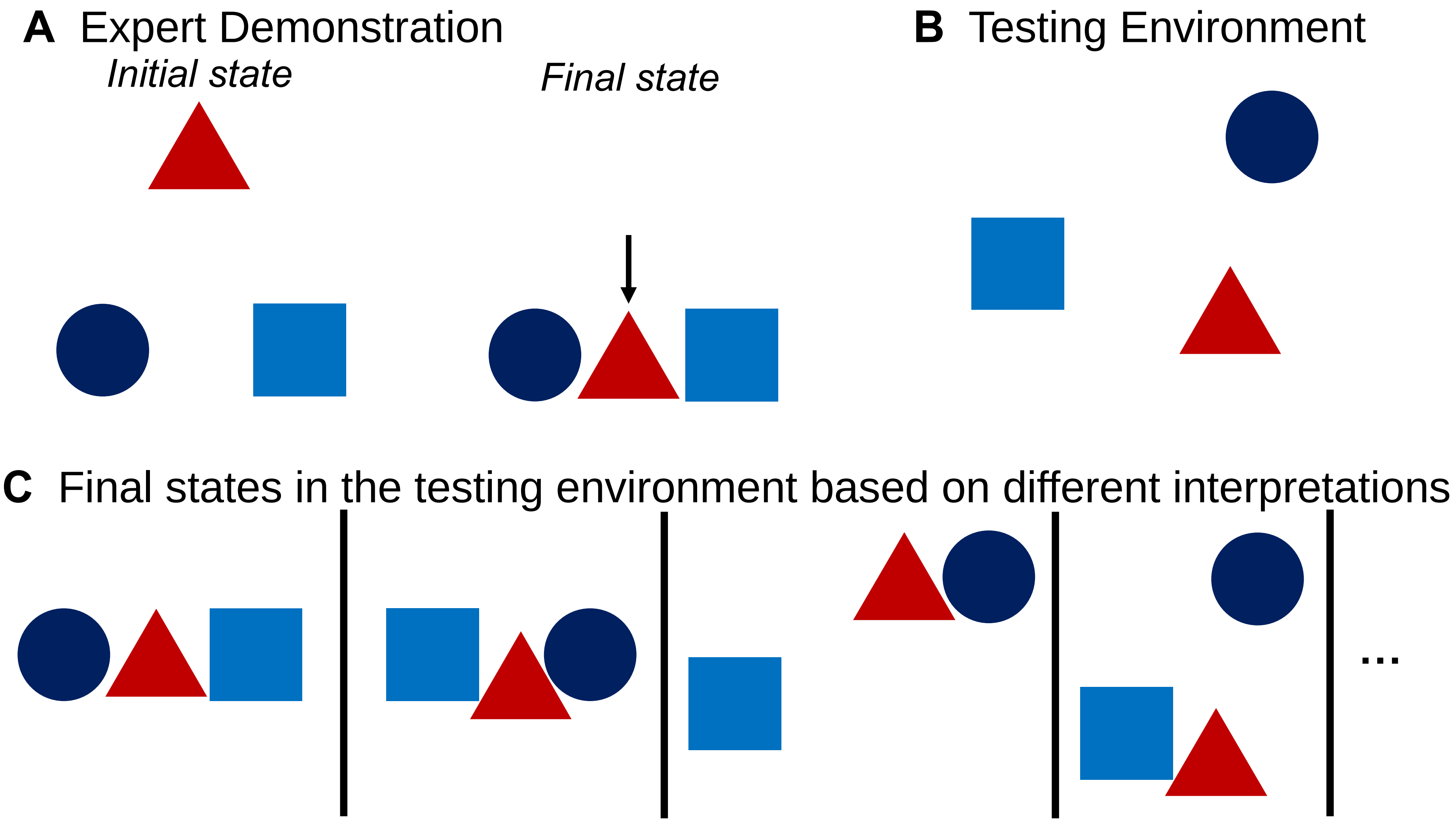
This work

Spatial goals for object rearrangement tasks

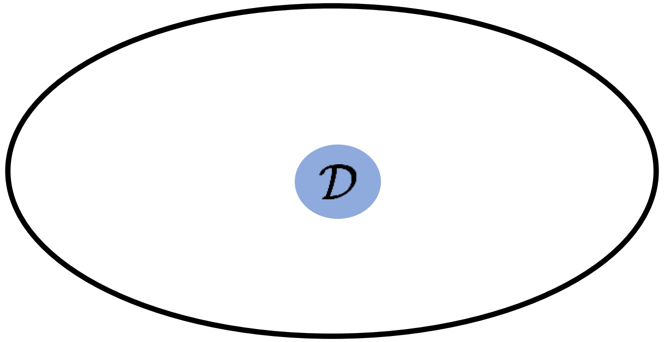


This work

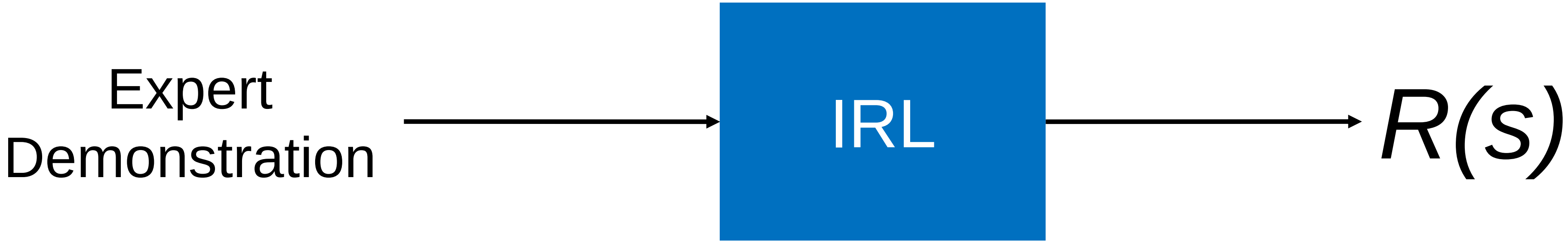
Spatial goals for object rearrangement tasks



Prior work



Inverse Reinforcement Learning (IRL)

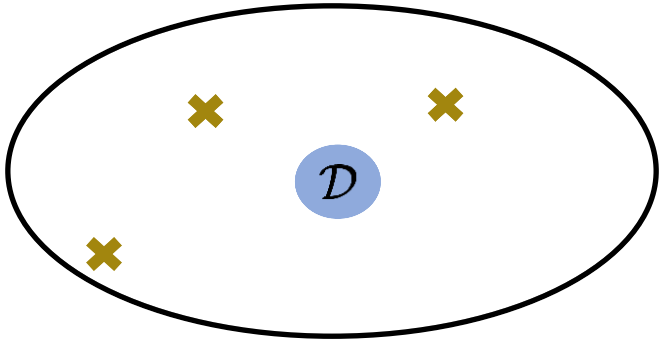


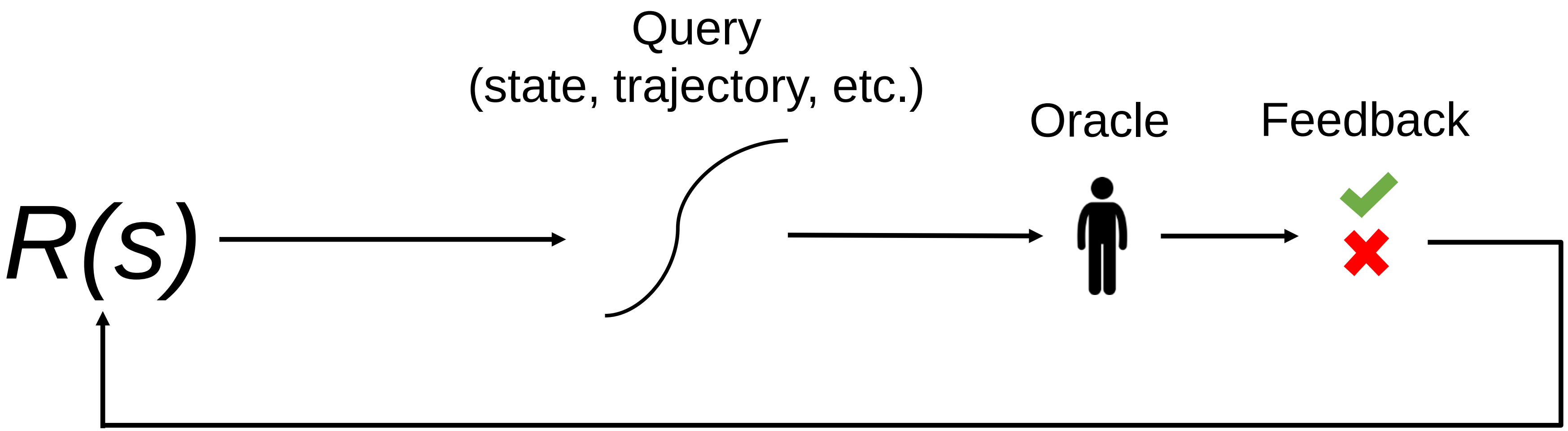
Learn *R(s)* from demos
Approximates the rewards for states **in the demo**

Fu et al. (2018)

Prior work

Active reward learning

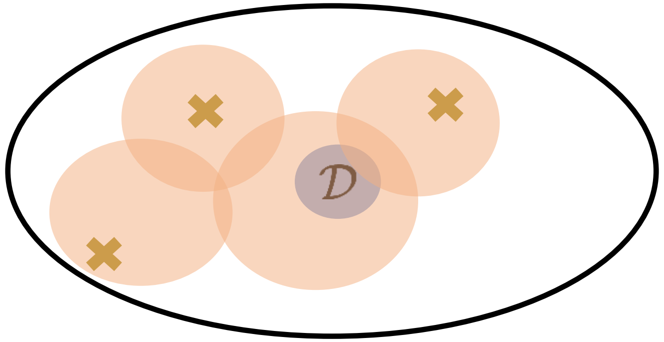




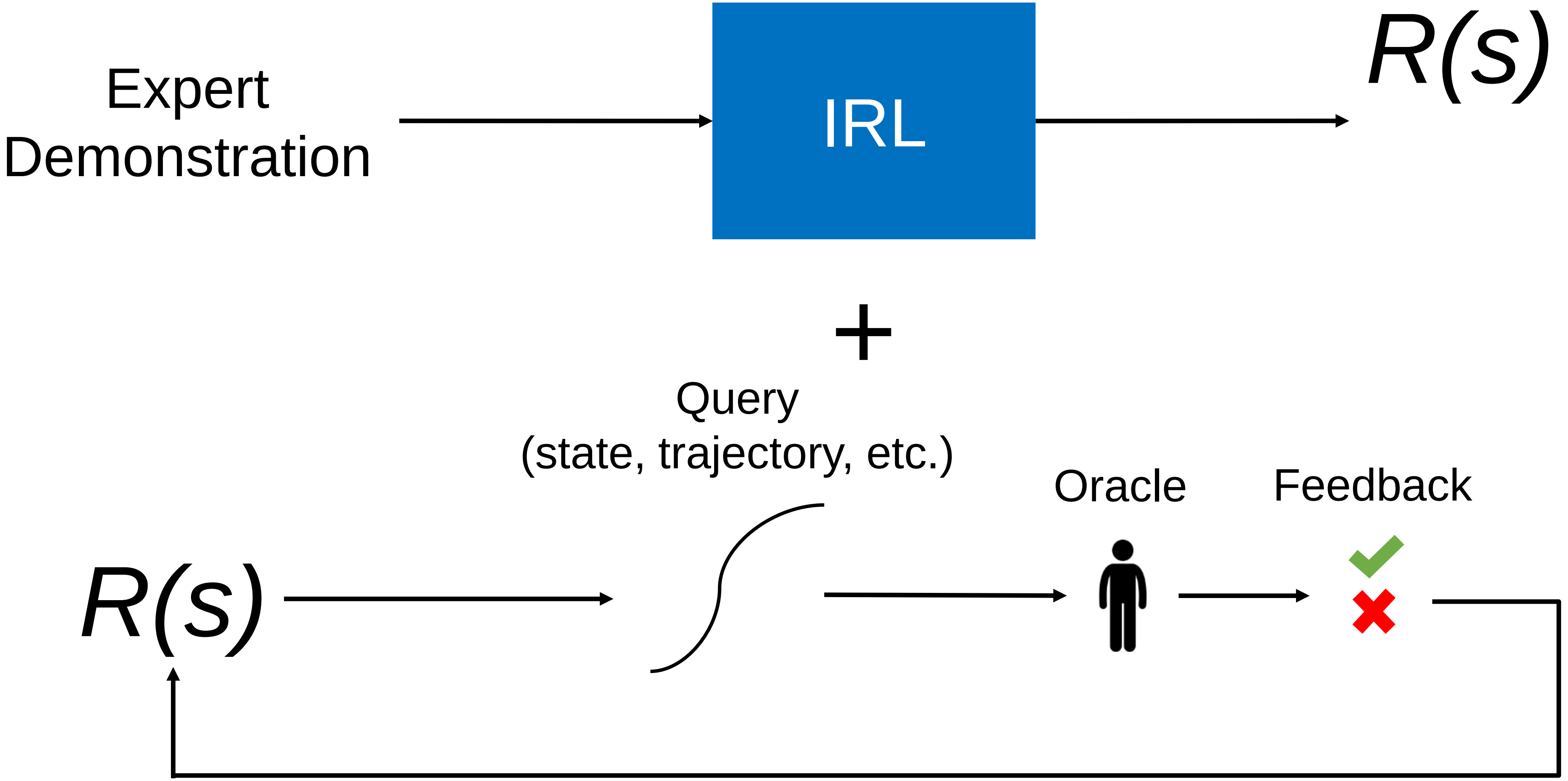
Collect additional training data from oracle feedback
Each query provides **one data point**

Reddy et al. (2020)

Our approach

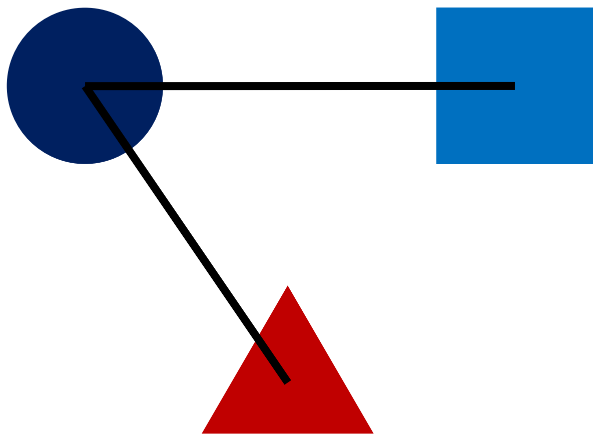


Combine IRL and active reward learning



Graph-based rewards for spatial goals

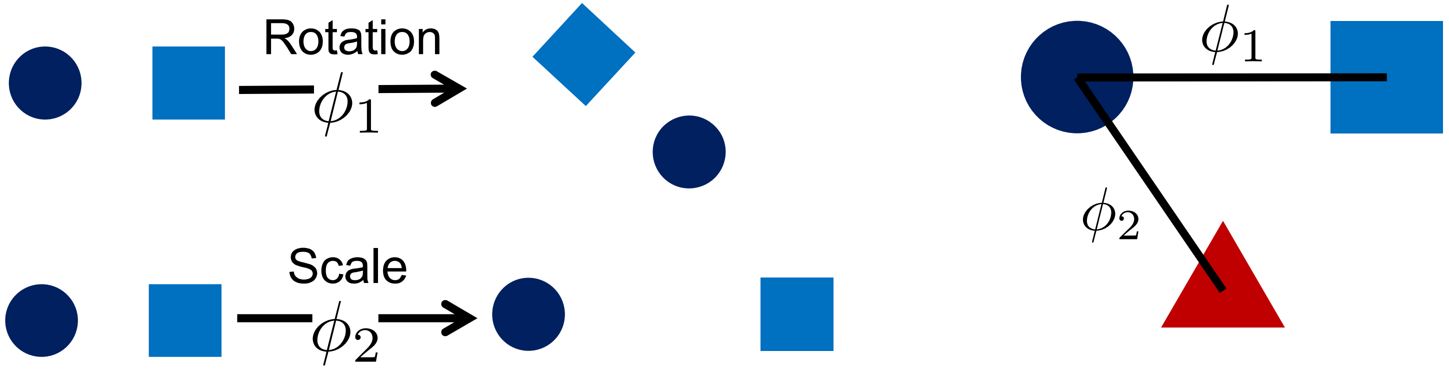
Graph structure: ***which*** relations?



$$R(\bullet, \blacksquare, \blacktriangle, G) = r(\bullet, \blacksquare) + r(\bullet, \blacktriangle)$$

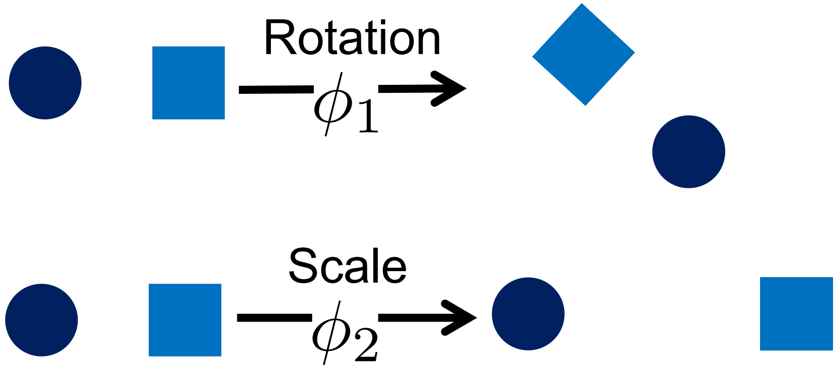
Equivalence mappings for data augmentation

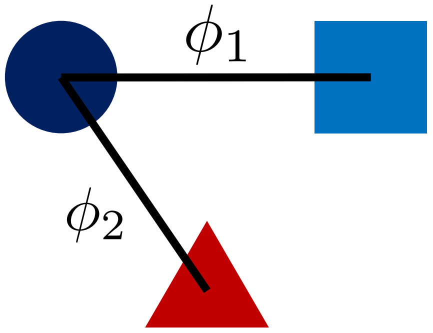
Equivalence mappings: *what* relations?



Equivalence mappings for data augmentation

Equivalence mappings: *what* relations?



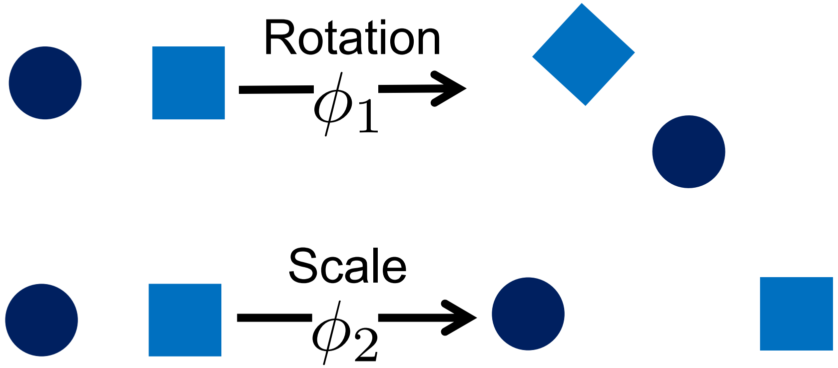


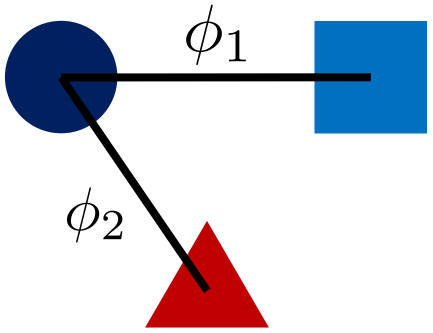
Edge equivalence → same reward for transformed edge

$$r(\bullet, \blacksquare) = r(\phi_k(\bullet, \blacksquare))$$

Equivalence mappings for data augmentation

Equivalence mappings: ***what*** relations?





Edge equivalence → same reward for transformed edge

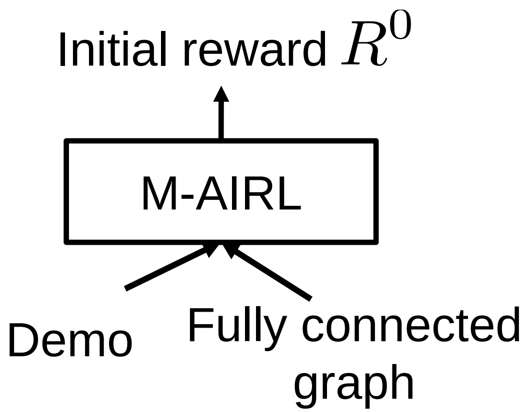
$$r(\bullet, \blacksquare) = r(\phi_k(\bullet, \blacksquare))$$

This enables data augmentation

GEM: Graph-based Equivalence Mappings

Reward function initialization

via **model-based** adversarial **inverse RL** (M-AIRL), an extension of Fu et al. (2018)

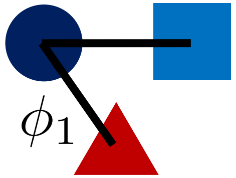


GEM: Graph-based Equivalence Mappings

Active reward refinement

Discover the **graph** and **equivalence mappings** for data augmentation

Current state, graph, and mapping assignment

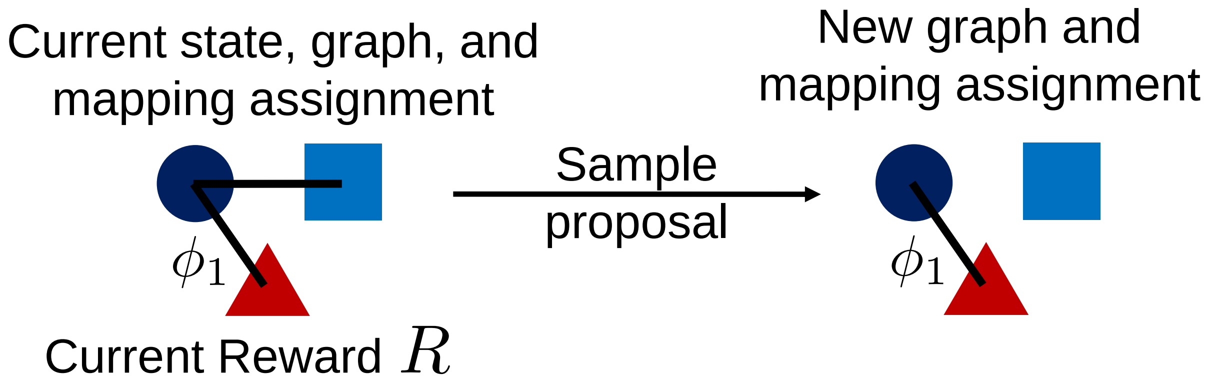


Current Reward R

GEM: Graph-based Equivalence Mappings

Active reward refinement

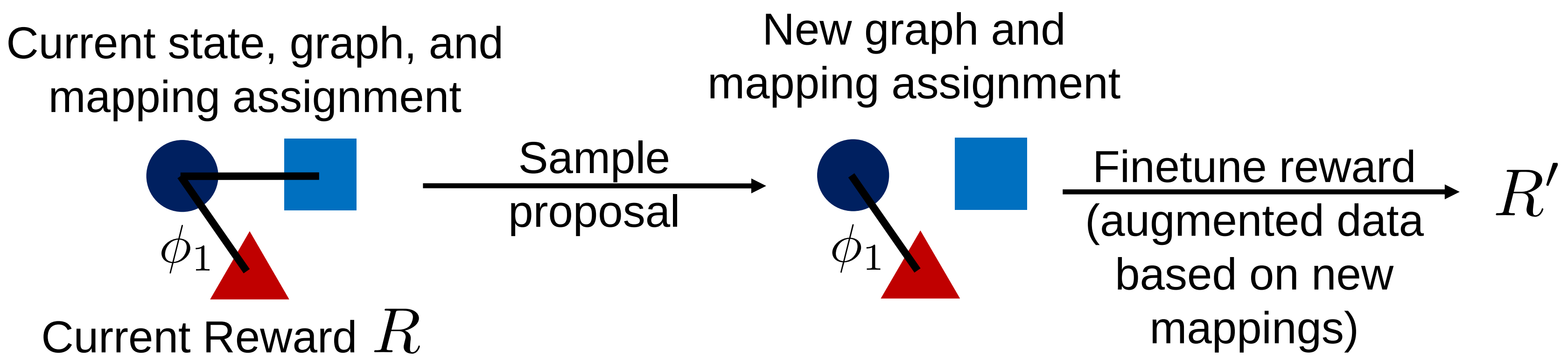
Discover the **graph** and **equivalence mappings** for data augmentation



GEM: Graph-based Equivalence Mappings

Active reward refinement

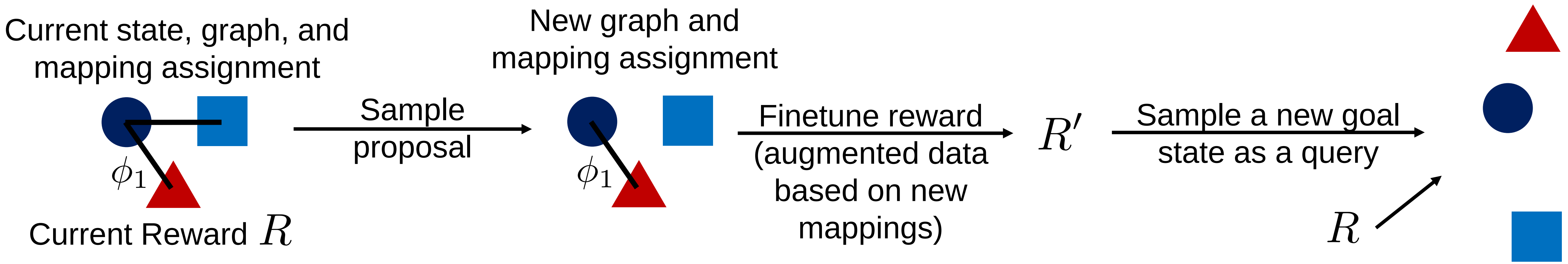
Discover the **graph** and **equivalence mappings** for data augmentation



GEM: Graph-based Equivalence Mappings

Active reward refinement

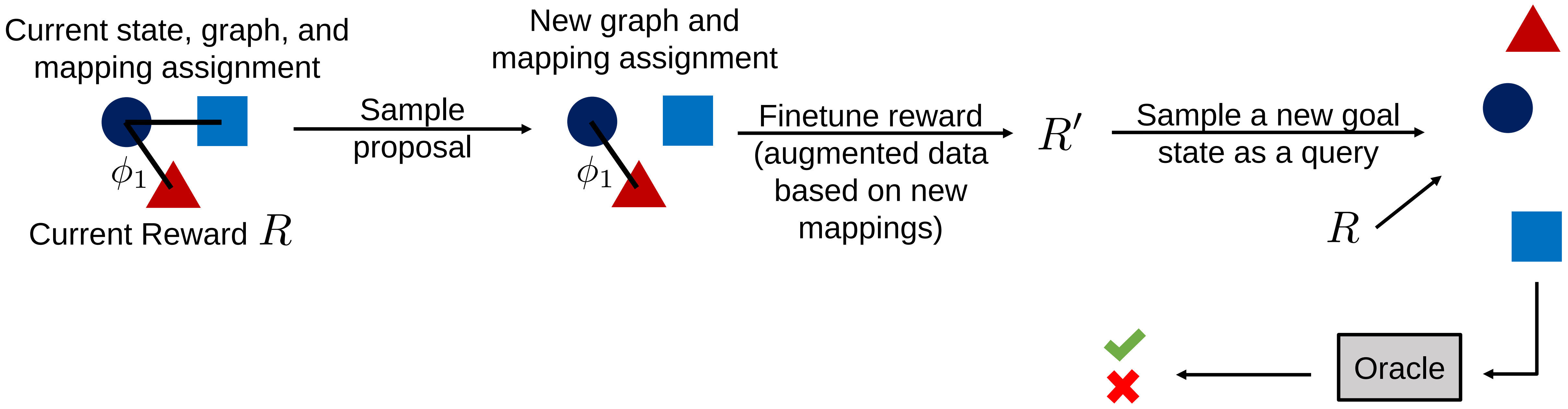
Discover the **graph** and **equivalence mappings** for data augmentation



GEM: Graph-based Equivalence Mappings

Active reward refinement

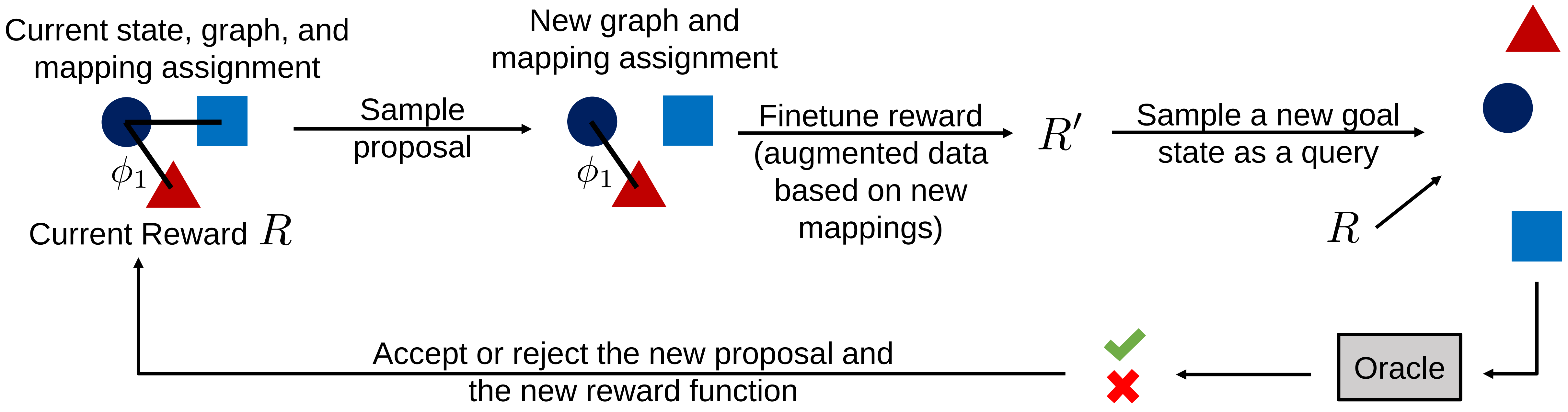
Discover the **graph** and **equivalence mappings** for data augmentation



GEM: Graph-based Equivalence Mappings

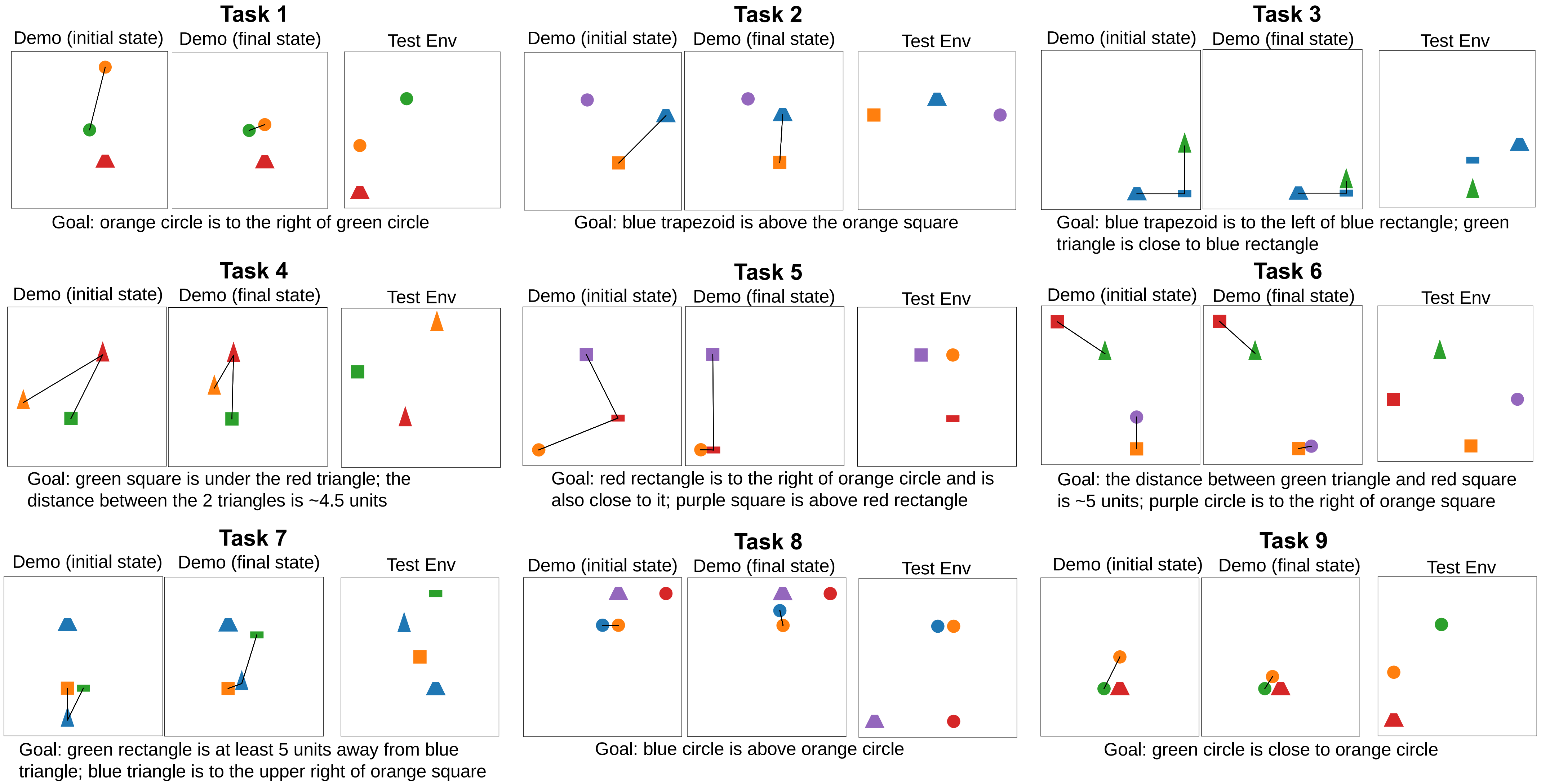
Active reward refinement

Discover the **graph** and **equivalence mappings** for data augmentation



Watch&Move tasks

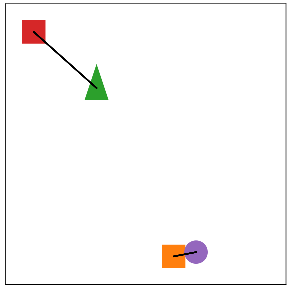
Object rearrangement tasks in a physics simulation from a single demonstration
Evaluation in an unseen environment



Results with a **simulated** oracle

Task 6

Demo (final state)

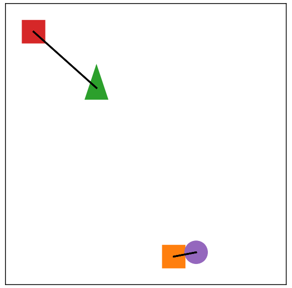


Goal: the distance between green triangle and red square is ~5, and purple circle is to the right of orange square

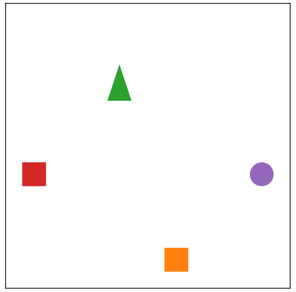
Results with a **simulated** oracle

Task 6

Demo (final state)



Test Env



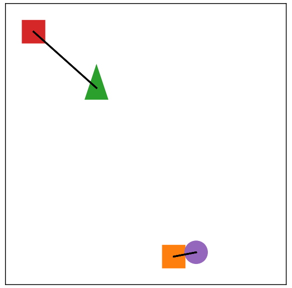
Goal: the distance between green triangle and red square is ~5, and purple circle is to the right of orange square

Results with a **simulated** oracle

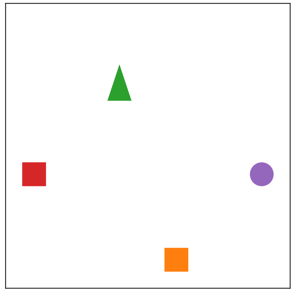
$$R_{\text{eval}} = \underbrace{\mathbf{1}(s^T \text{ satisfies the goal})}_{\text{Task completion}} - \underbrace{0.02\, d(s^0, s^T)}_{\text{Cost (displacement)}}$$

Task 6

Demo (final state)



Test Env



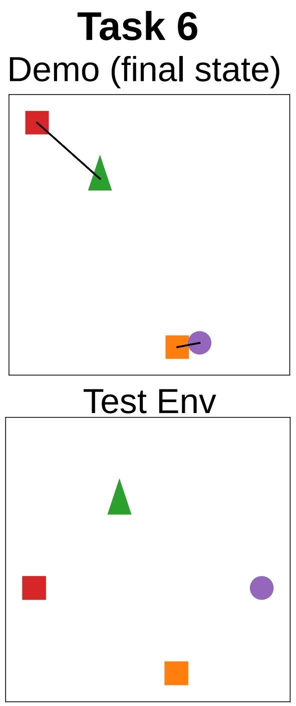
Goal: the distance between green triangle and red square is ~5, and purple circle is to the right of orange square

R eval

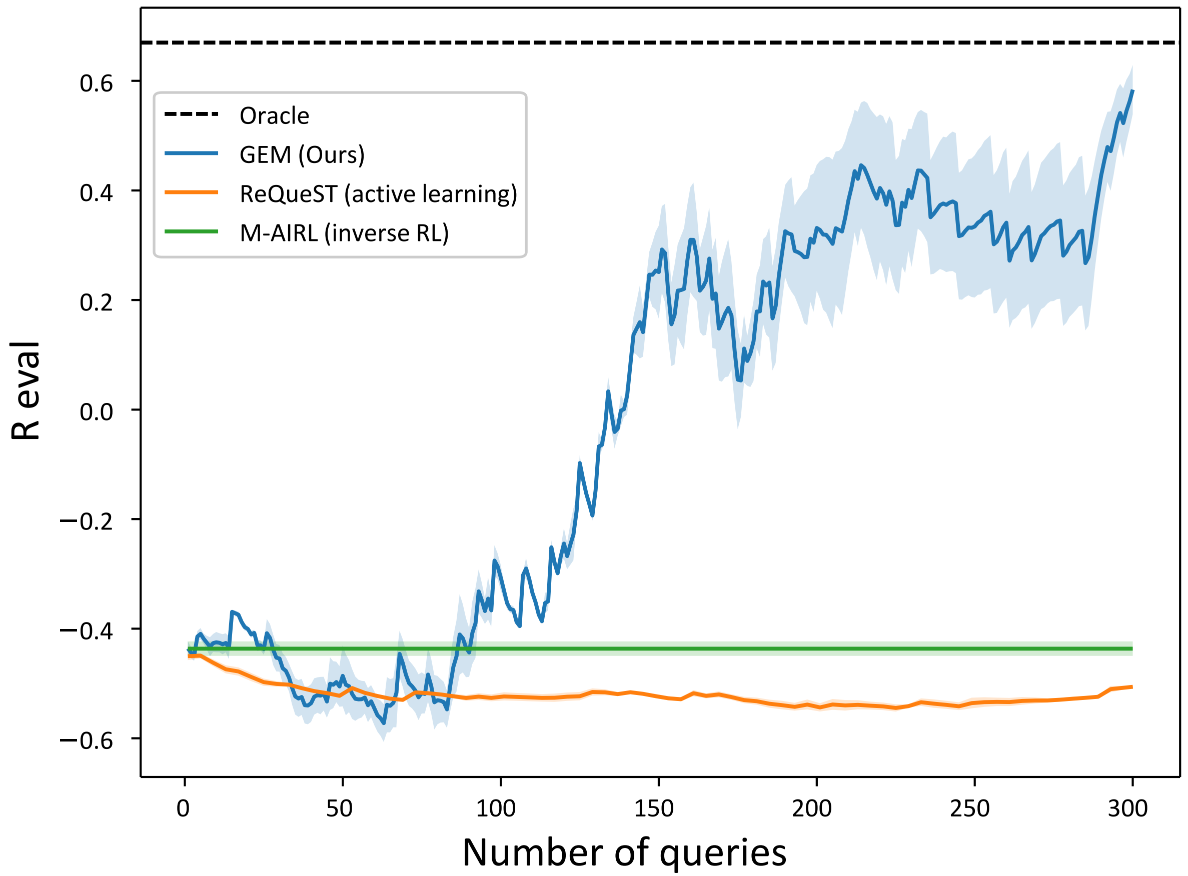
Number of queries

Results with a **simulated** oracle

$$R_{\text{eval}} = \underbrace{\mathbf{1}(s^T \text{ satisfies the goal})}_{\text{Task completion}} - \underbrace{0.02\, d(s^0, s^T)}_{\text{Cost (displacement)}}$$

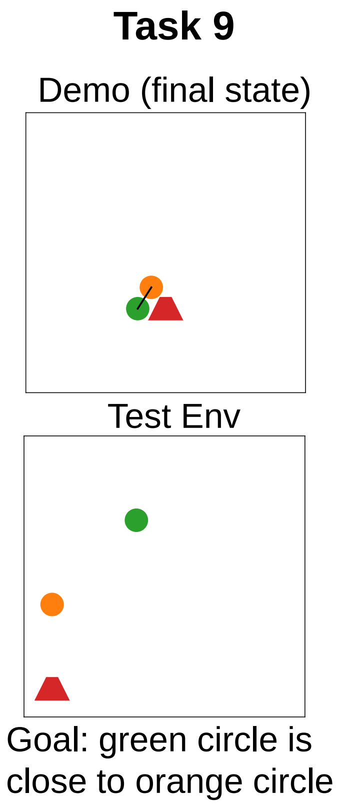


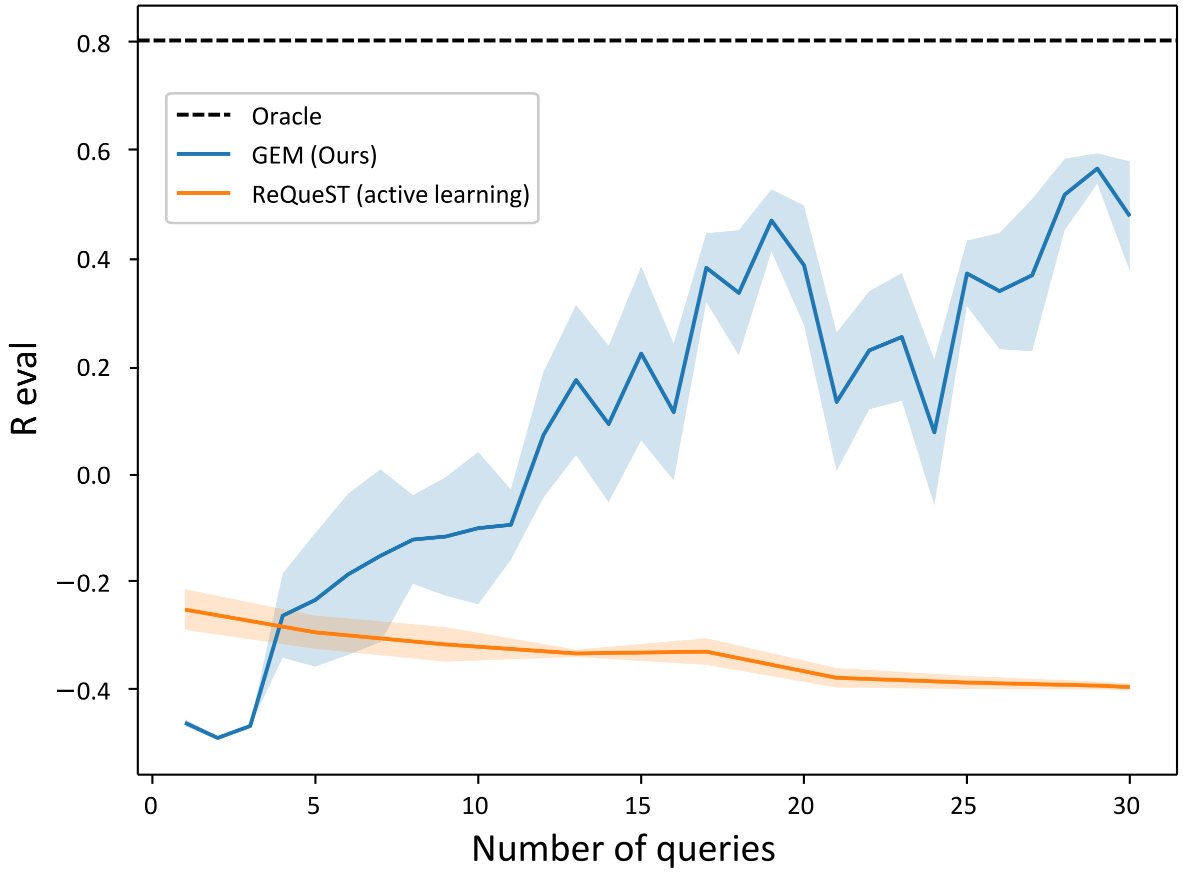
Goal: the distance between green triangle and red square is ~5, and purple circle is to the right of orange square



Results with a **human** oracle

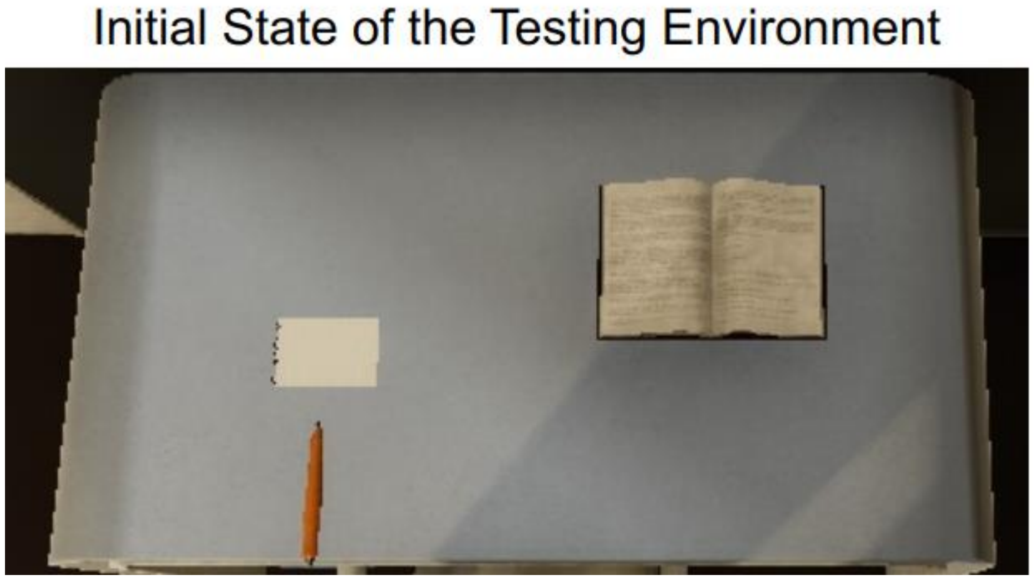
$$R_{\text{eval}} = \underbrace{\mathbf{1}(s^T \text{ satisfies the goal})}_{\text{Task completion}} - \underbrace{0.02\, d(s^0, s^T)}_{\text{Cost (displacement)}}$$

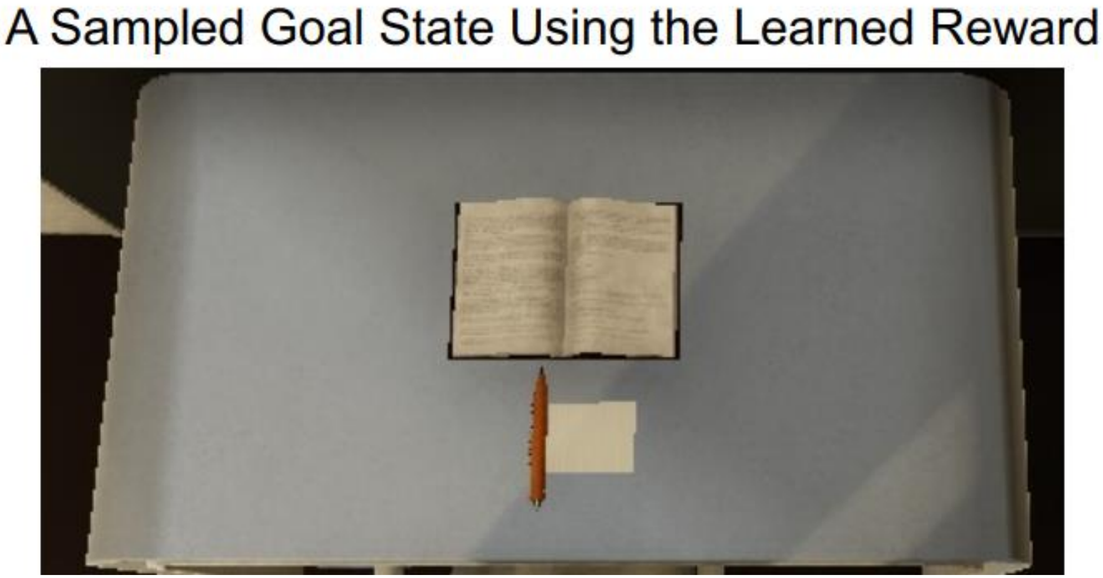




Results on a **real-life** task

Virtual home simulator





Puig et al. (2021)

Summary

A graph-based active reward learning algorithm, GEM, for learning spatial goals from a single demonstration and oracle feedback

Summary

- A graph-based active reward learning algorithm, GEM, for learning spatial goals from a single demonstration and oracle feedback
- GEM achieves much greater success in learning generalizable spatial goal specification compared to SOTA IRL and active reward learning baselines

Summary

A graph-based active reward learning algorithm, GEM, for learning spatial goals from a single demonstration and oracle feedback

GEM achieves much greater success in learning generalizable spatial goal specification compared to SOTA IRL and active reward learning baselines

GEM can also learn from human users

Summary

A graph-based active reward learning algorithm, GEM, for learning spatial goals from a single demonstration and oracle feedback

GEM achieves much greater success in learning generalizable spatial goal specification compared to SOTA IRL and active reward learning baselines

GEM can also learn from human users

Project website: https://www.tshu.io/GEM/

Please visit our poster today for more details!