Generalization to New Actions in Reinforcement Learning

Ayush Jain*, Andrew Szot*, Joseph J. Lim

USC University of Southern California

Cooking Tools

Cutting

Mixing

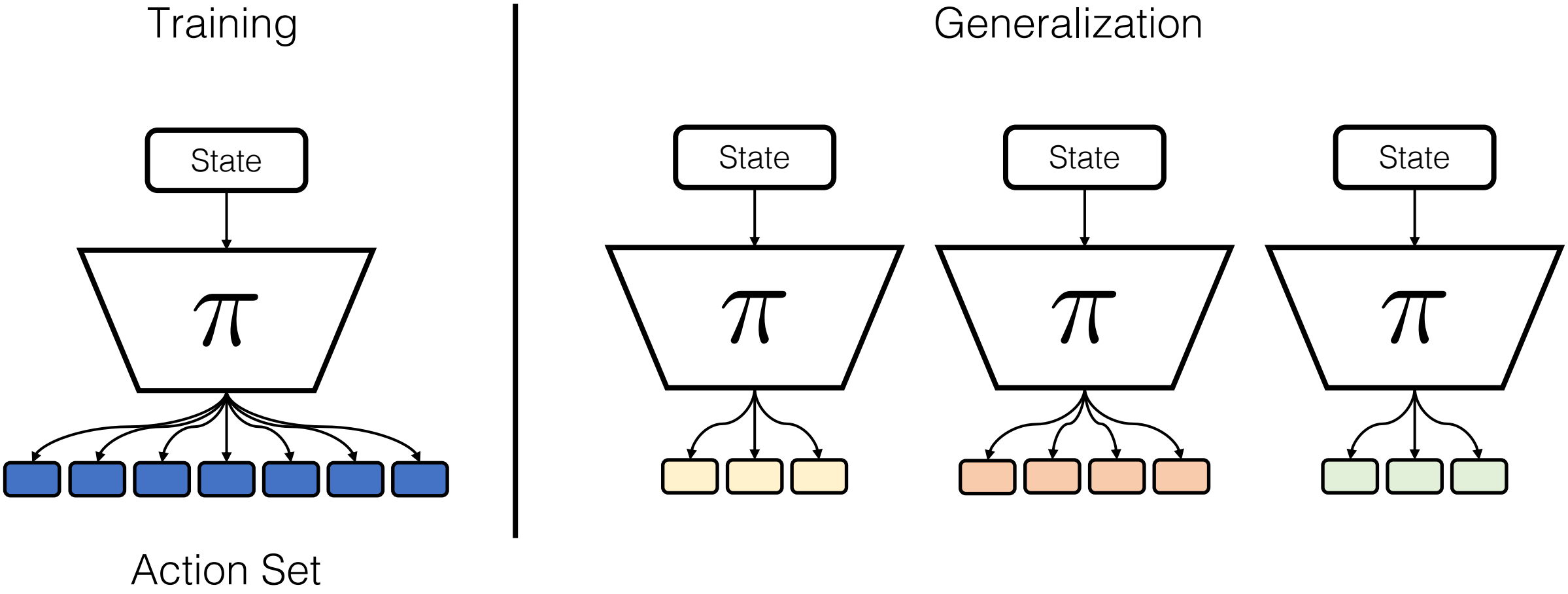
How to decide between new tools?

New Cooking Tools

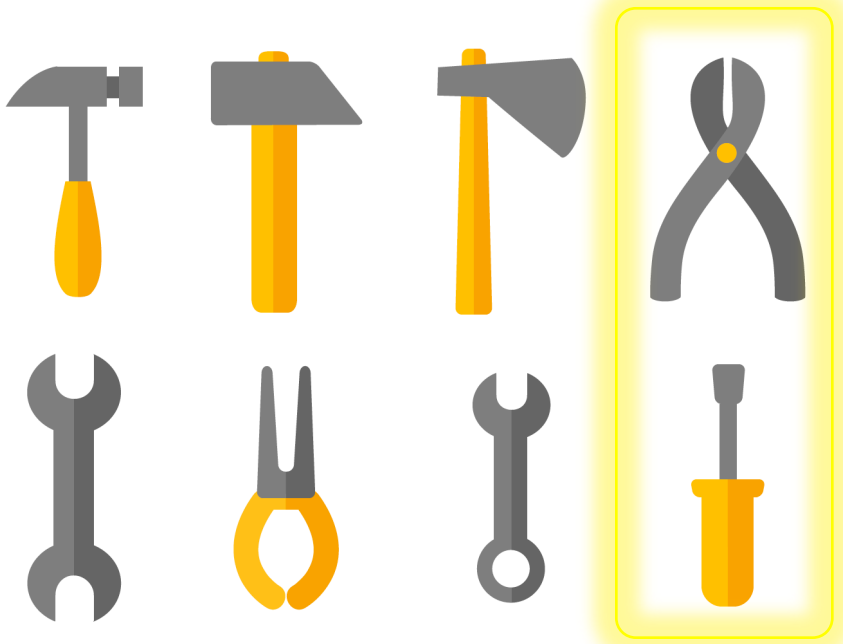
Cutting

Mixing

Generalization to New Actions

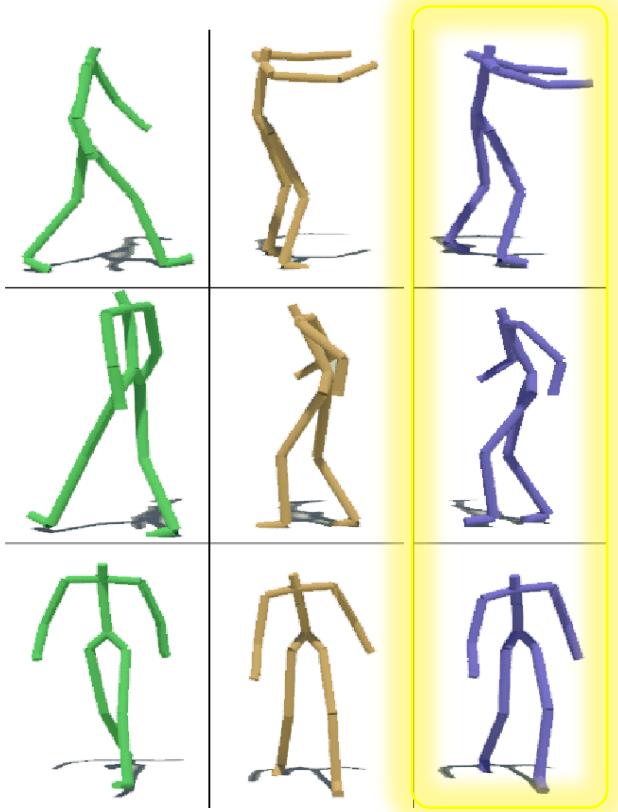


Using new New Actions



Tool Improvisation

New Recommendations



Acquired Skill Set

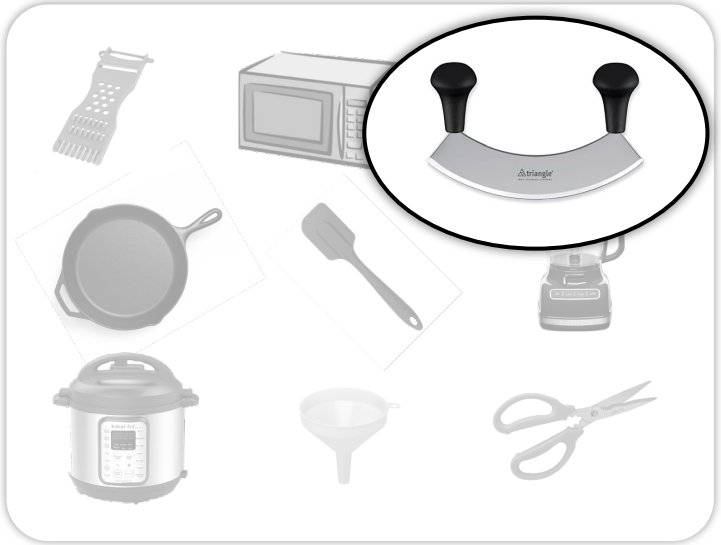
Approach Intuition

1. Observation
2. Inference
3. Decision-making

Actions are characterized by their behaviors

Approach Intuition

1. Observation
2. Inference
3. Decision-making

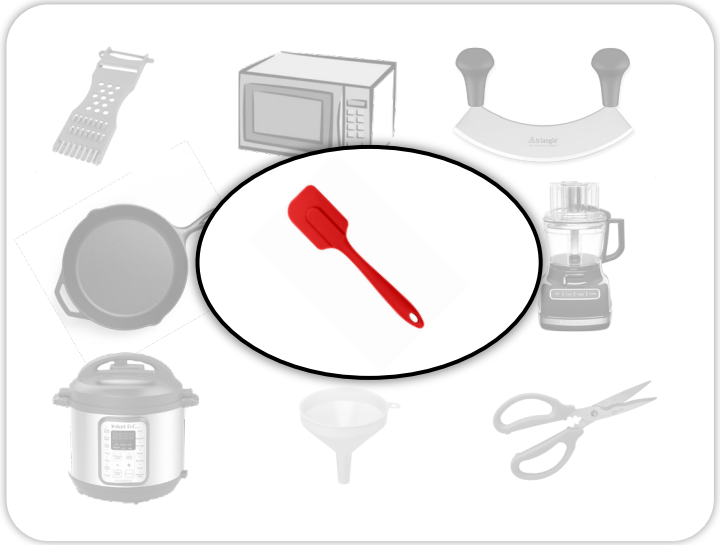


Cutting

Actions are characterized by their behaviors

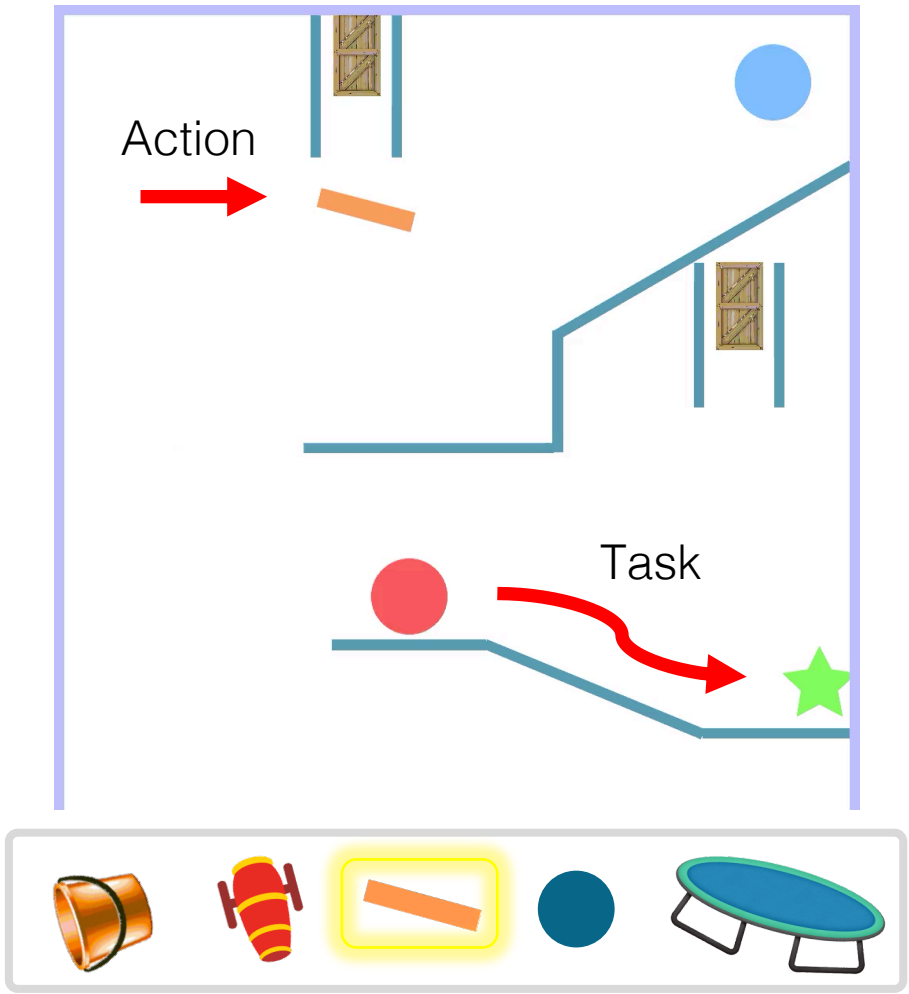
Approach Intuition

1. Observation
2. Inference
3. Decision-making

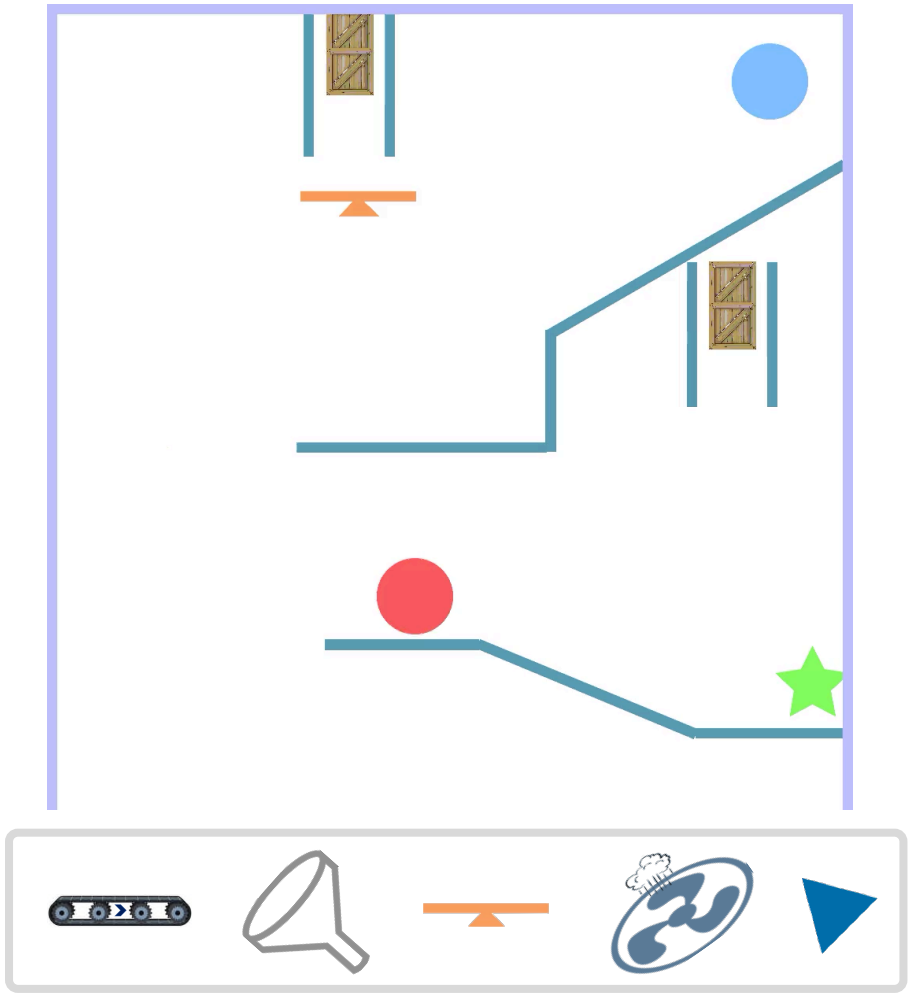


Actions are characterized by their behaviors

Results: CREATE

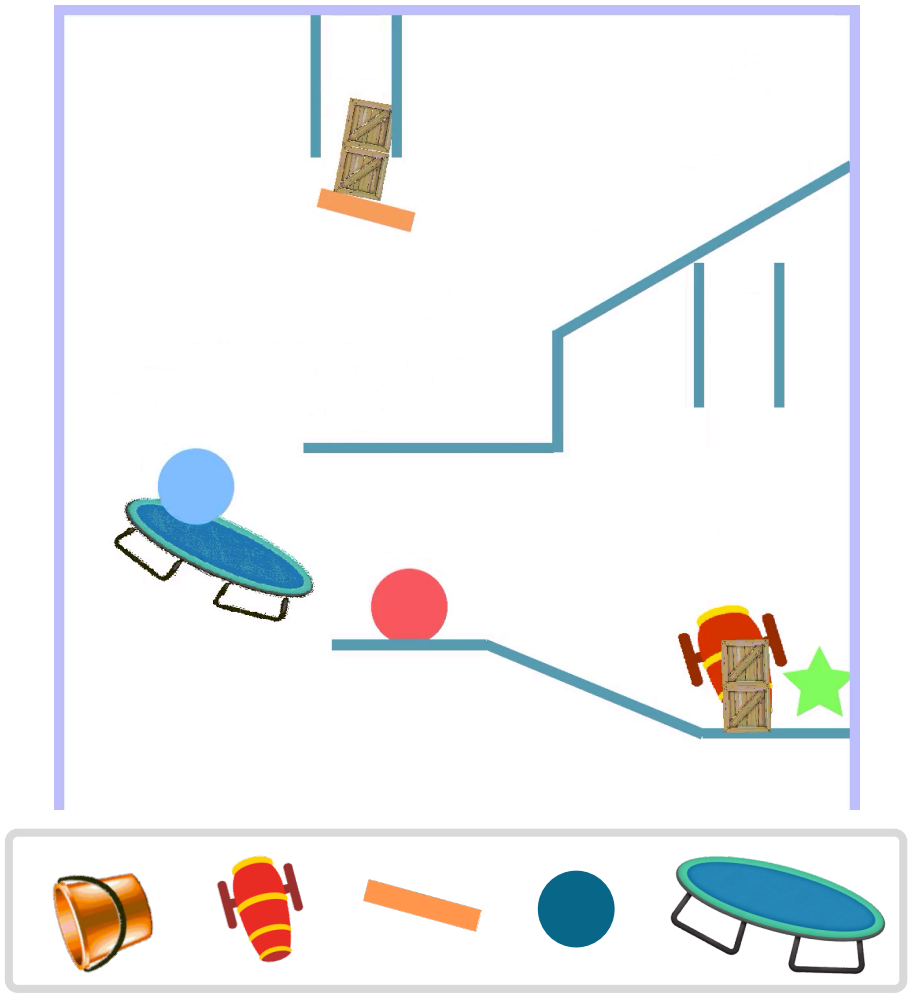


Training actions

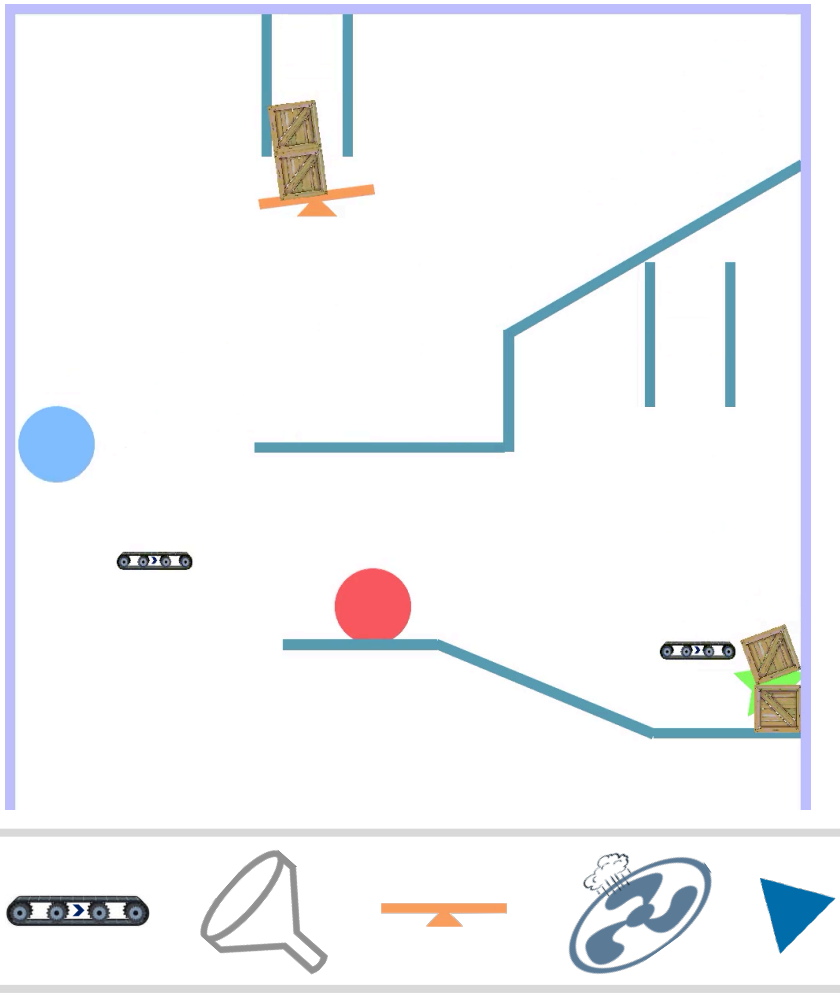


Generalization

Results: CREATE

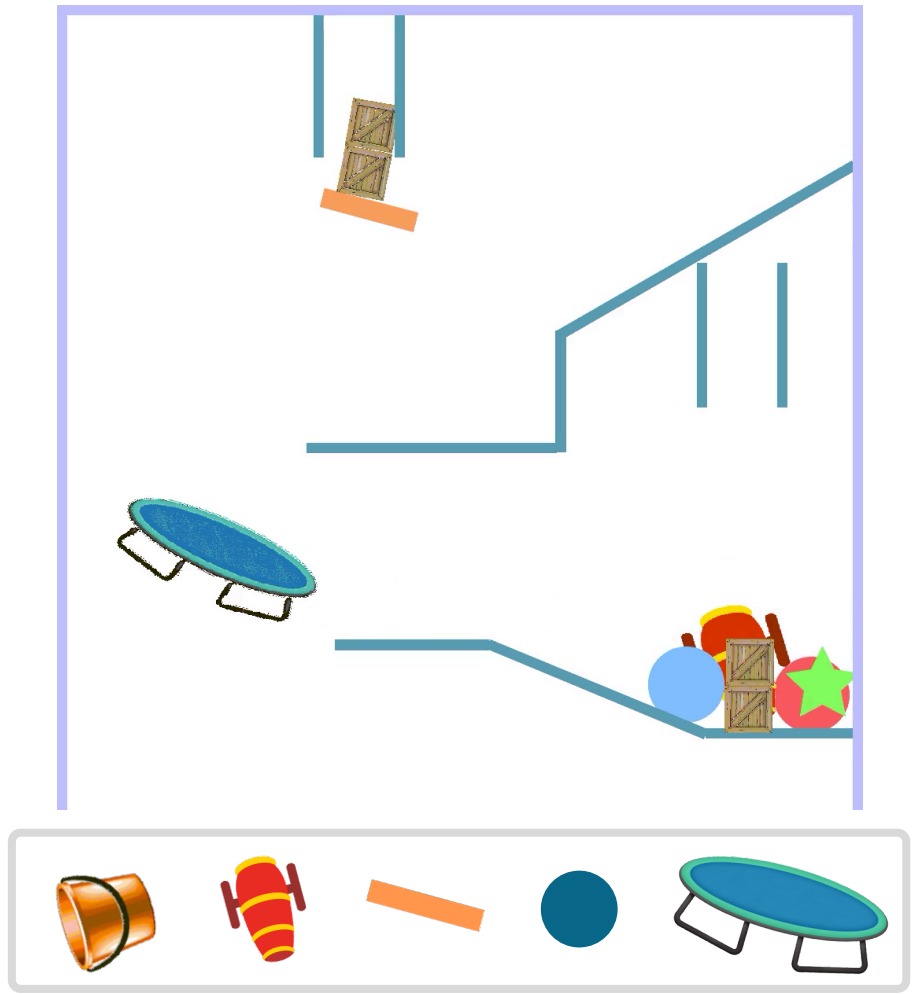


Training actions

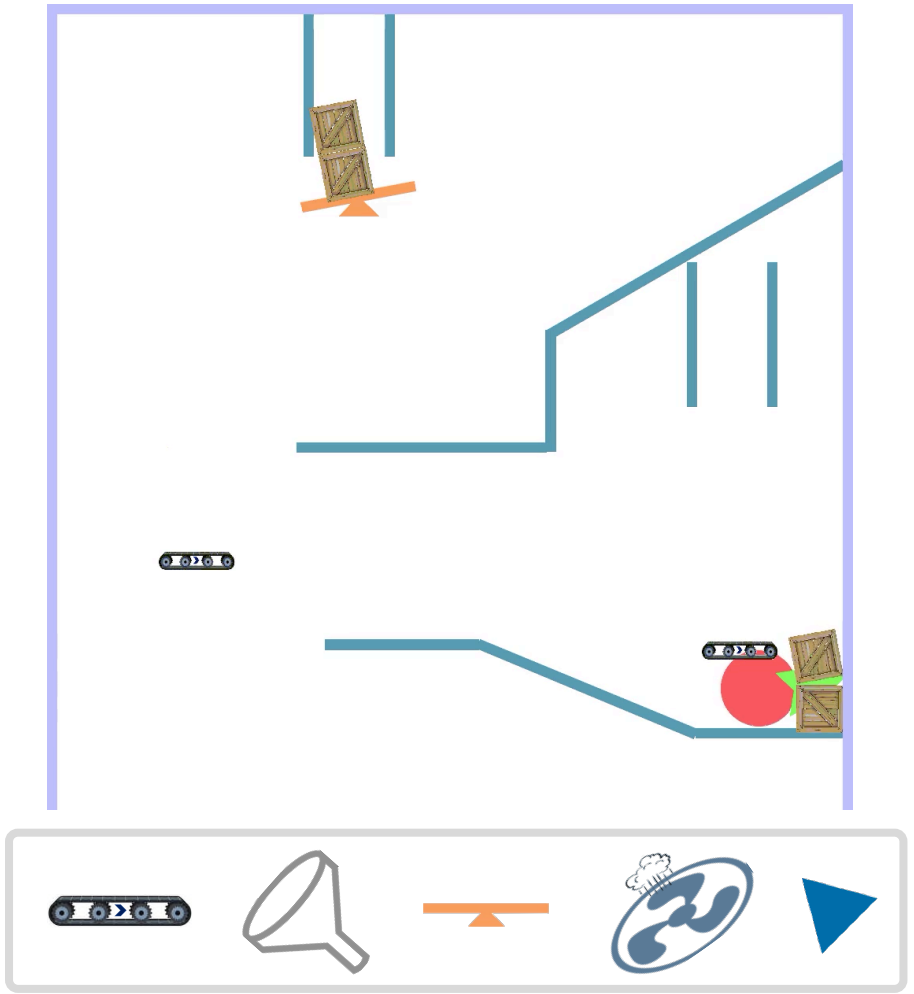


Generalization

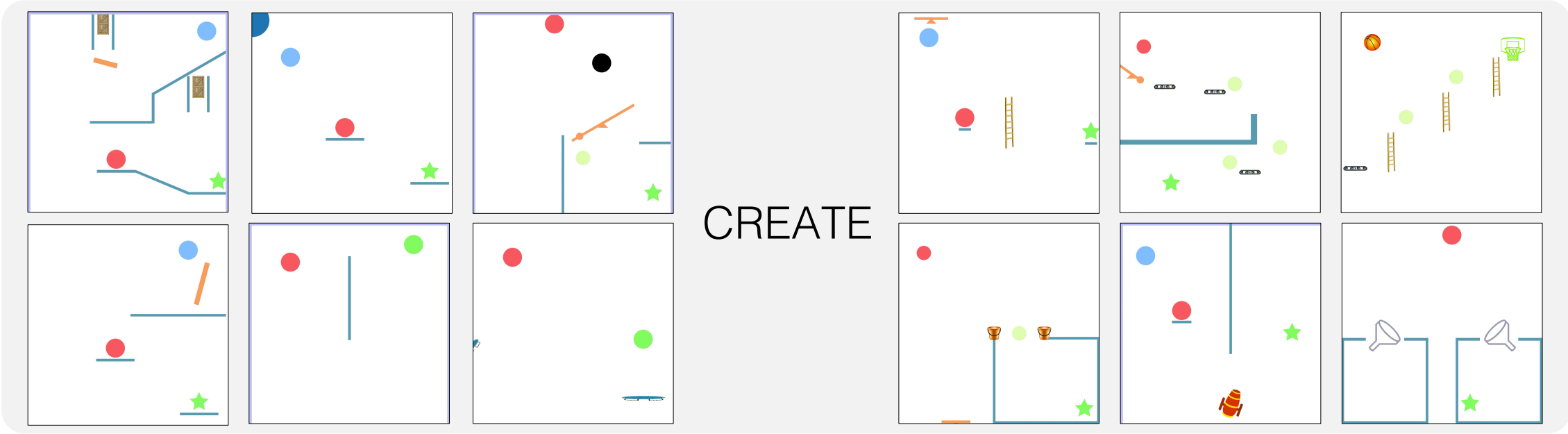
Results: CREATE



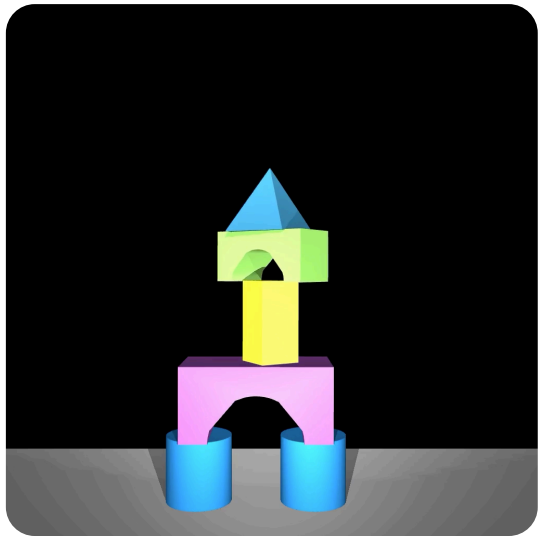
Training actions



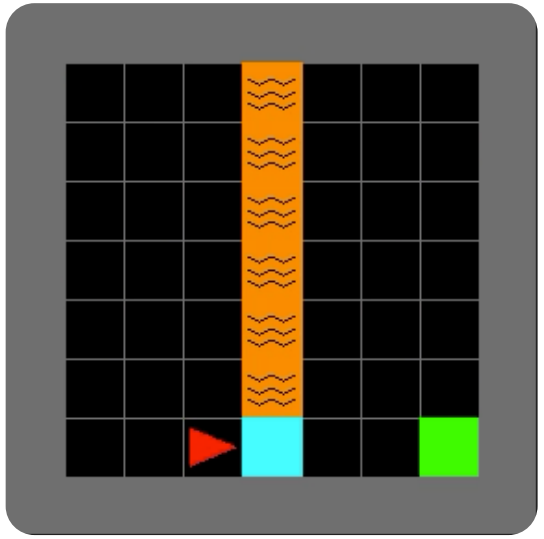
Generalization



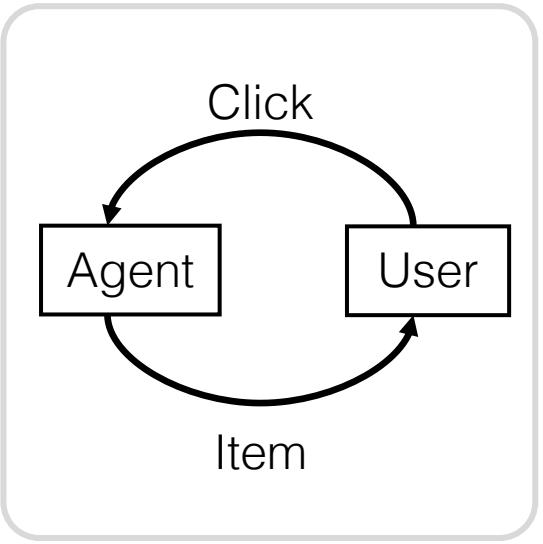
clvrai.com/create



Shape Stacking



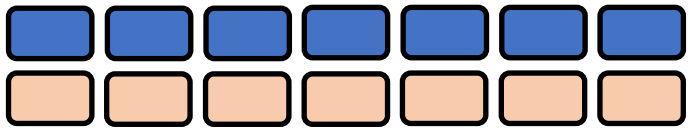
2D Grid



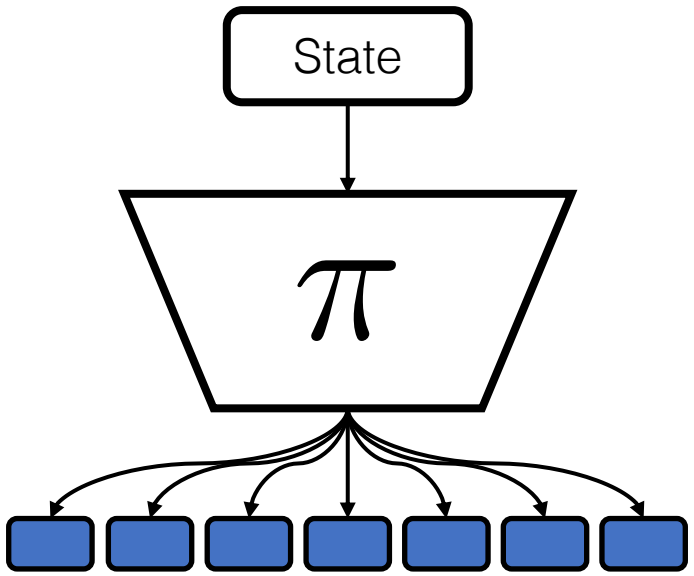
Recommender

Problem Formulation

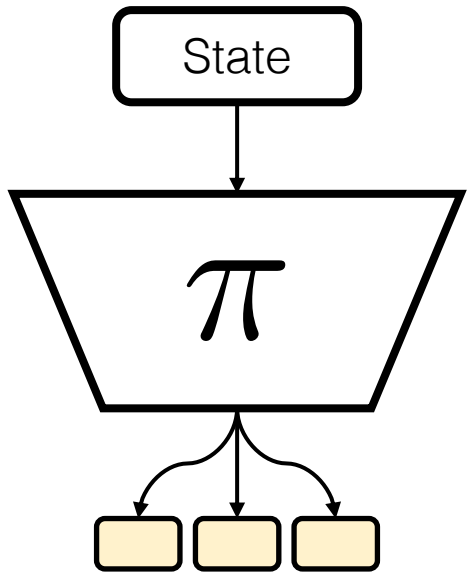
Problem Formulation



Available Actions

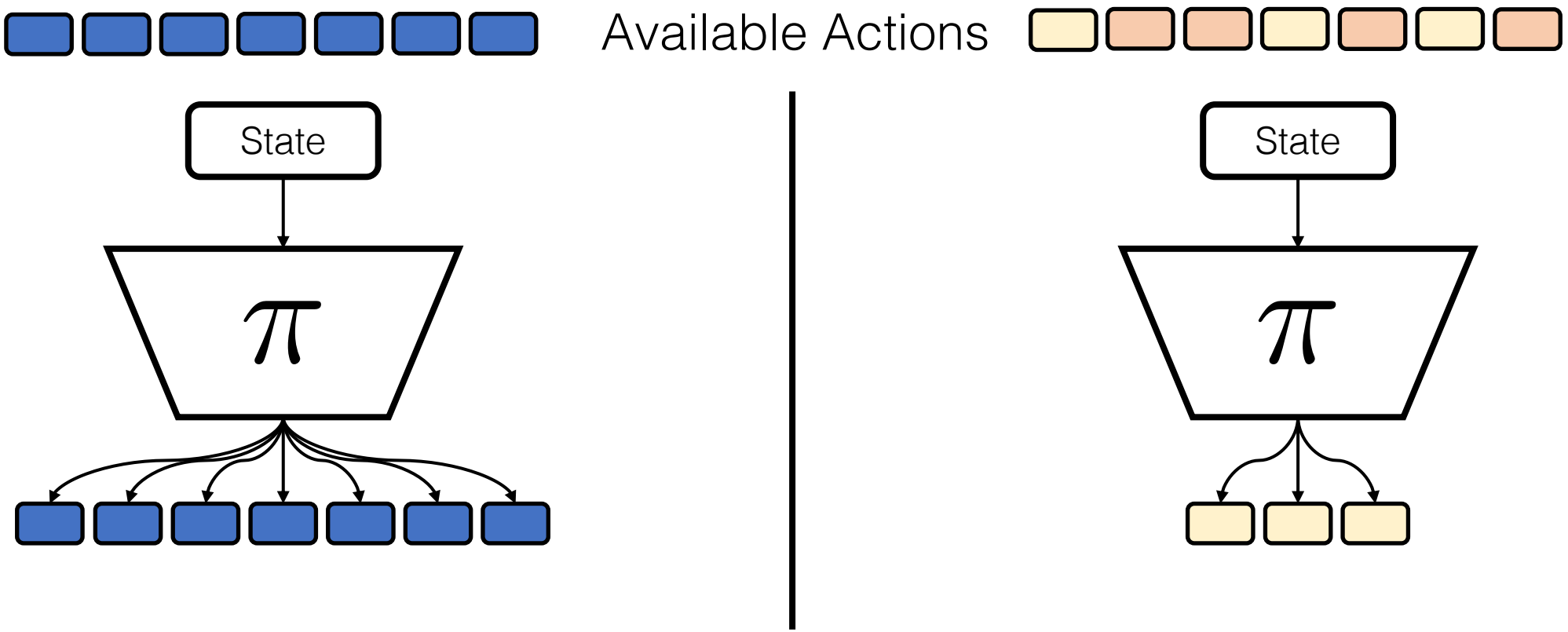


Training

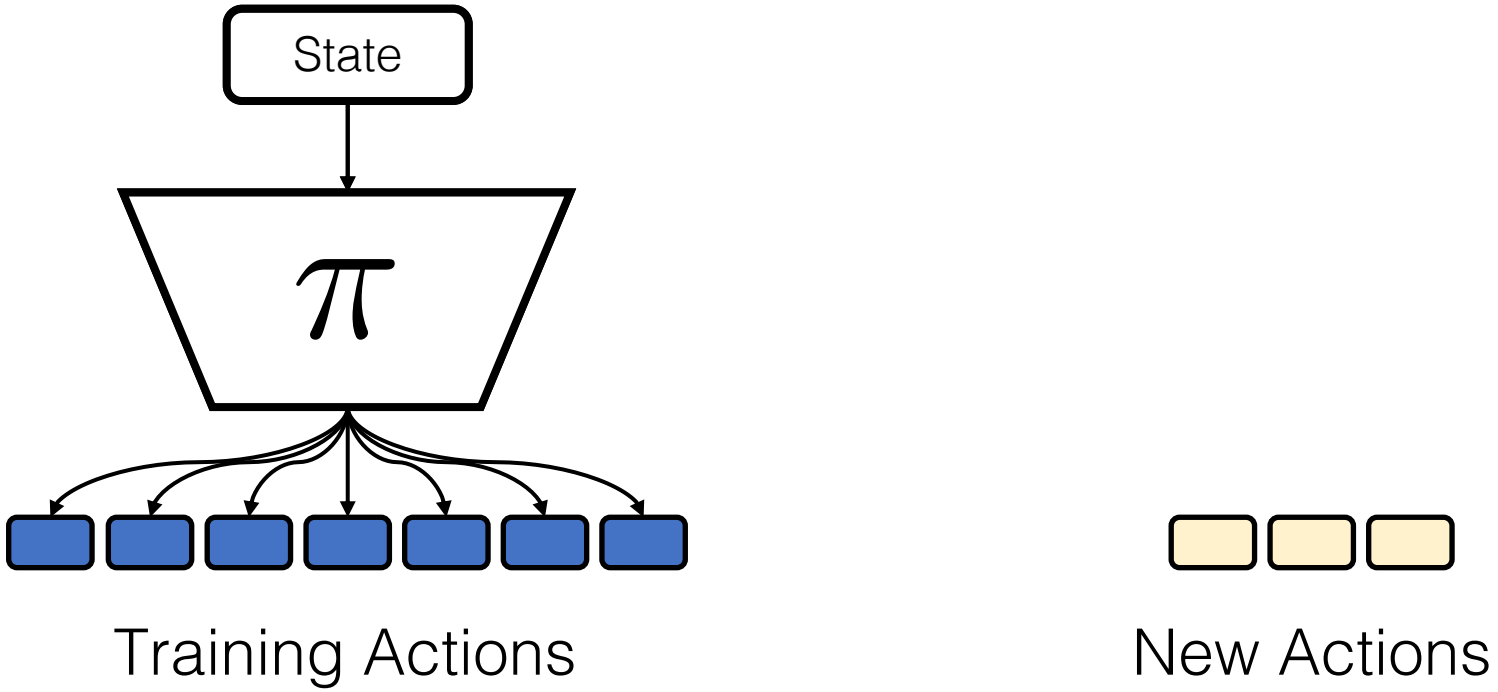


Evaluation

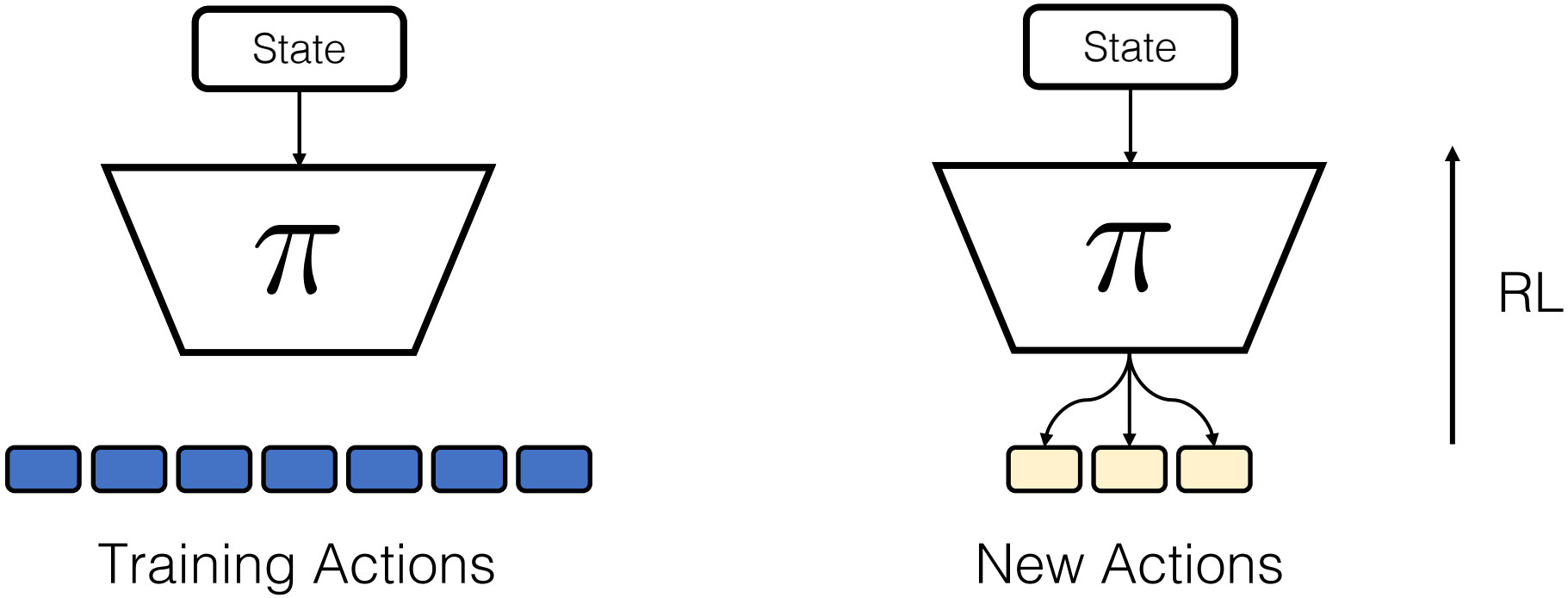
Problem Formulation



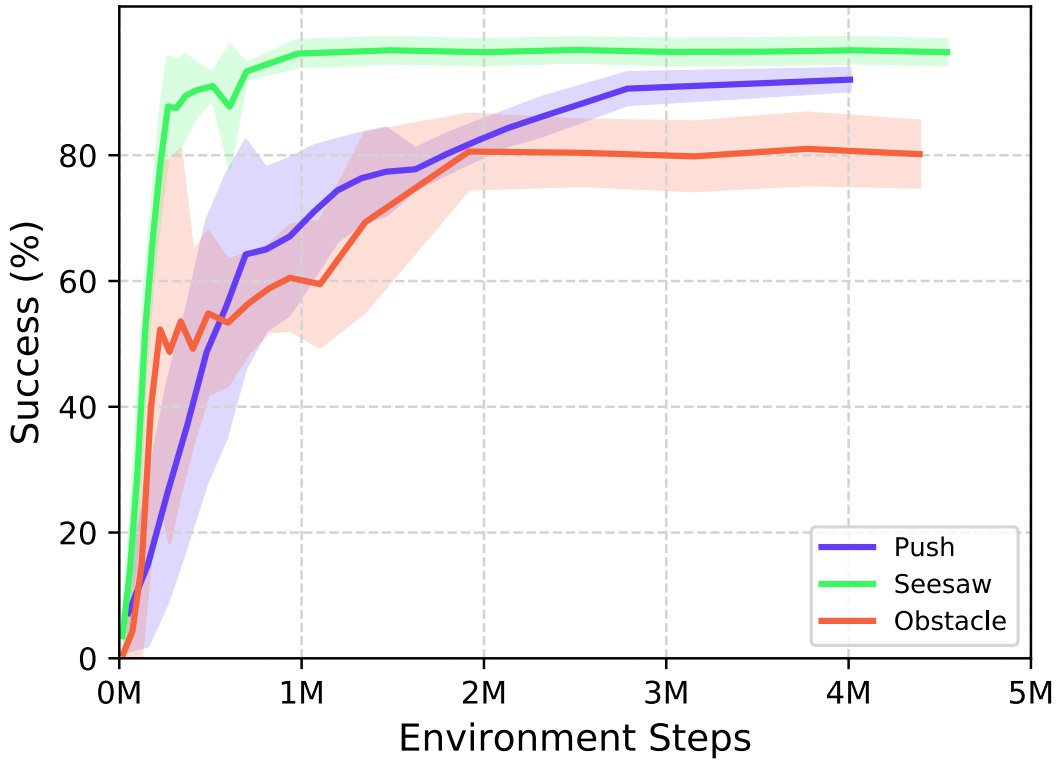
Fine-tuning on New Action Set



Fine-tuning on New Action Set

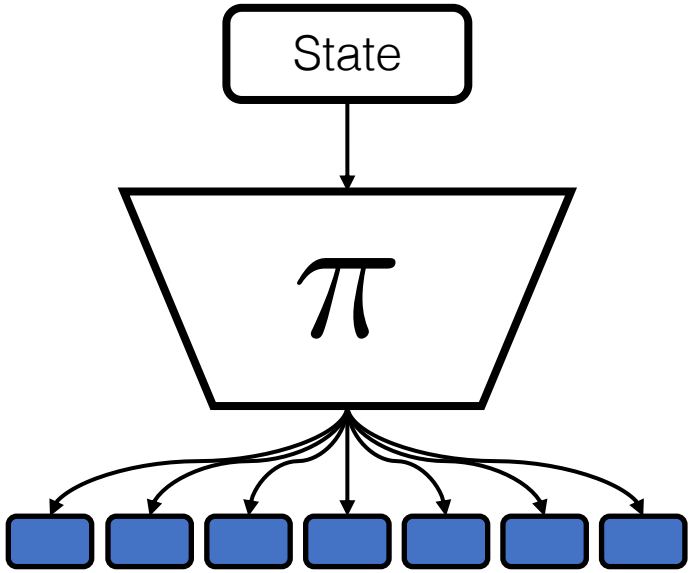


Fine-tuning is expensive!

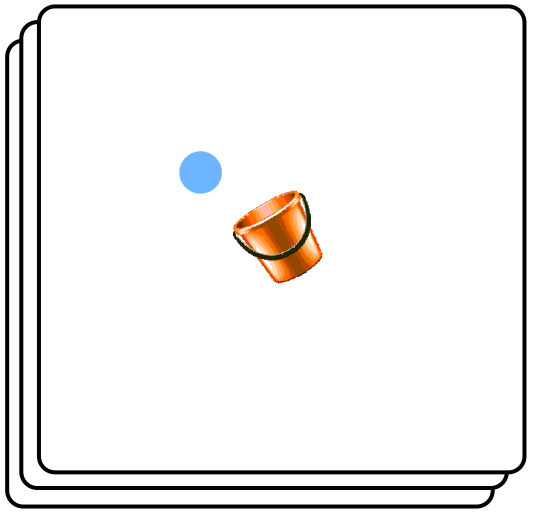


Zero-Shot Generalization to New Action Sets is Important

Actions Characterized by Observations

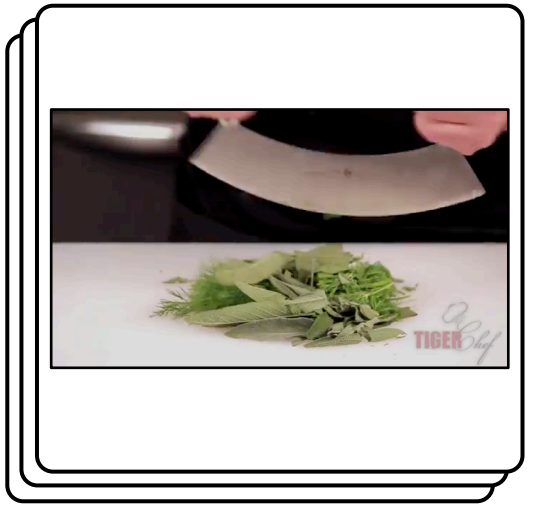


Actions

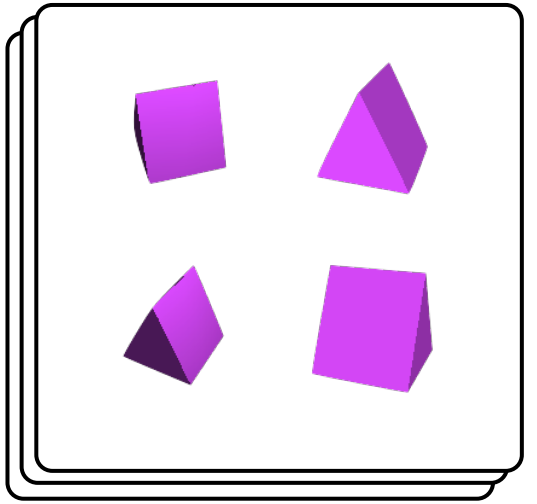


Action Observations

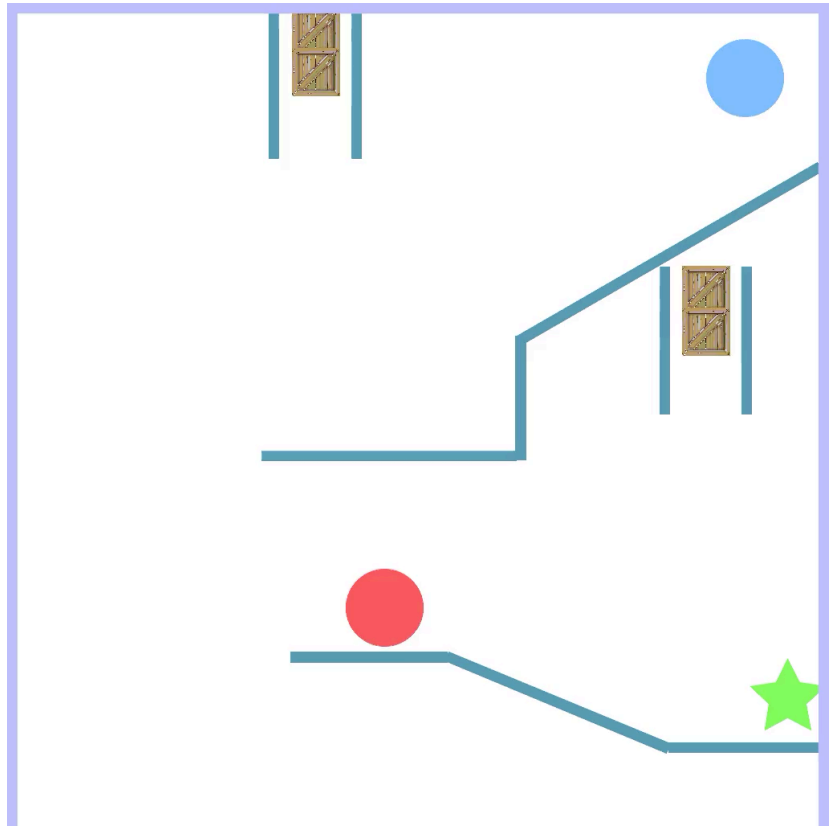
Tools



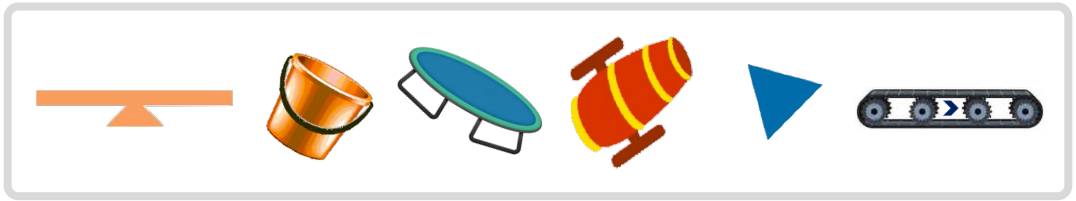
Shapes



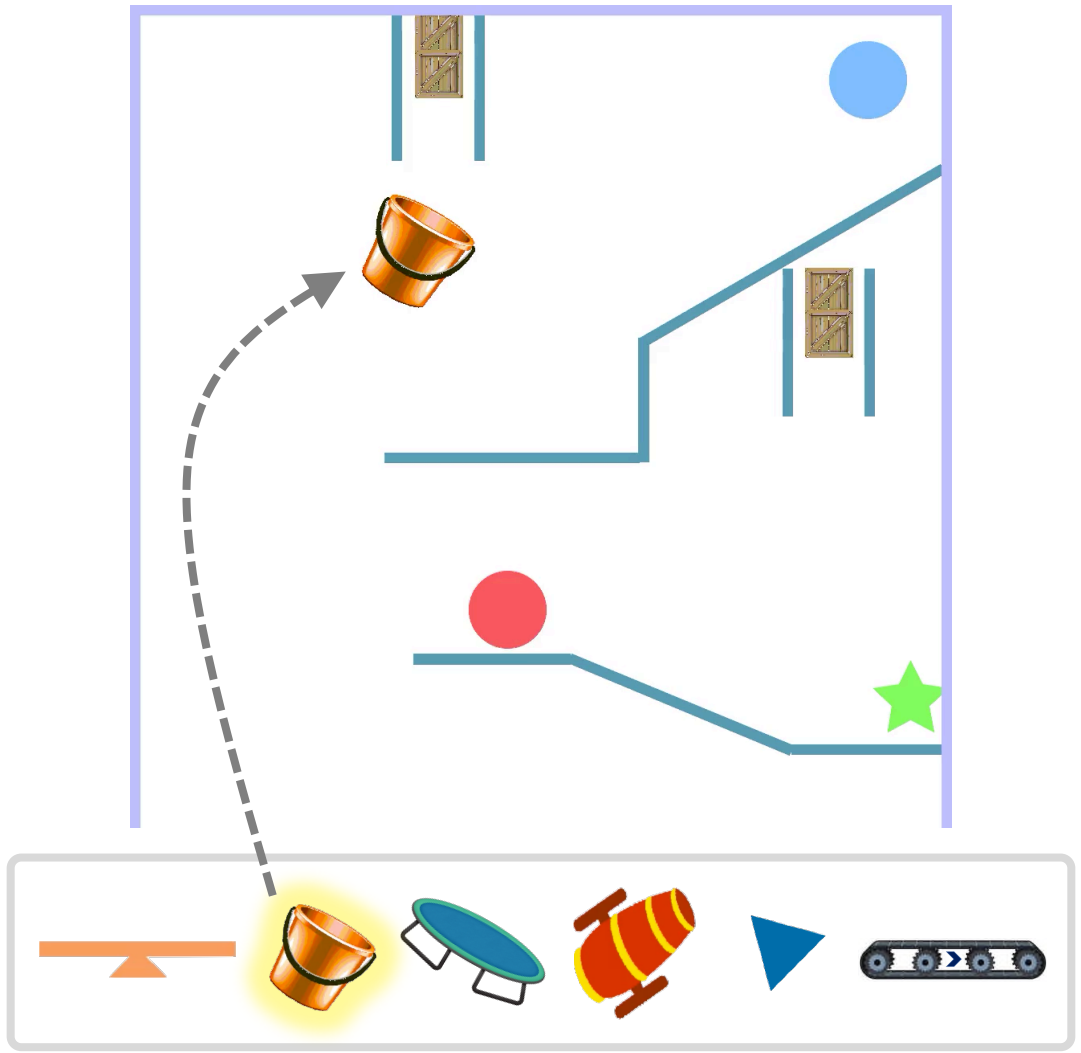
Chain REAction Tool Environment (CREATE)



Available Tools

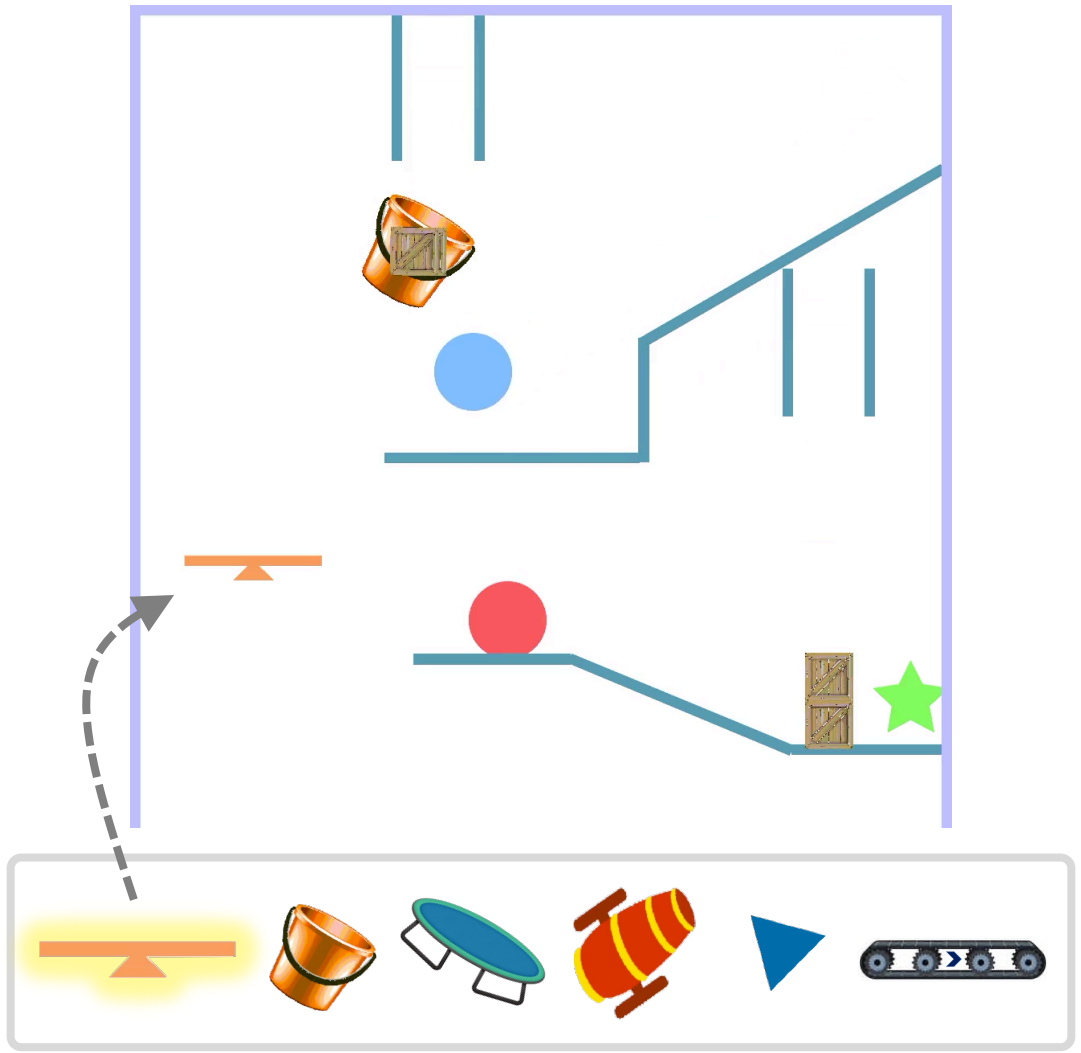


Select and Place Tools

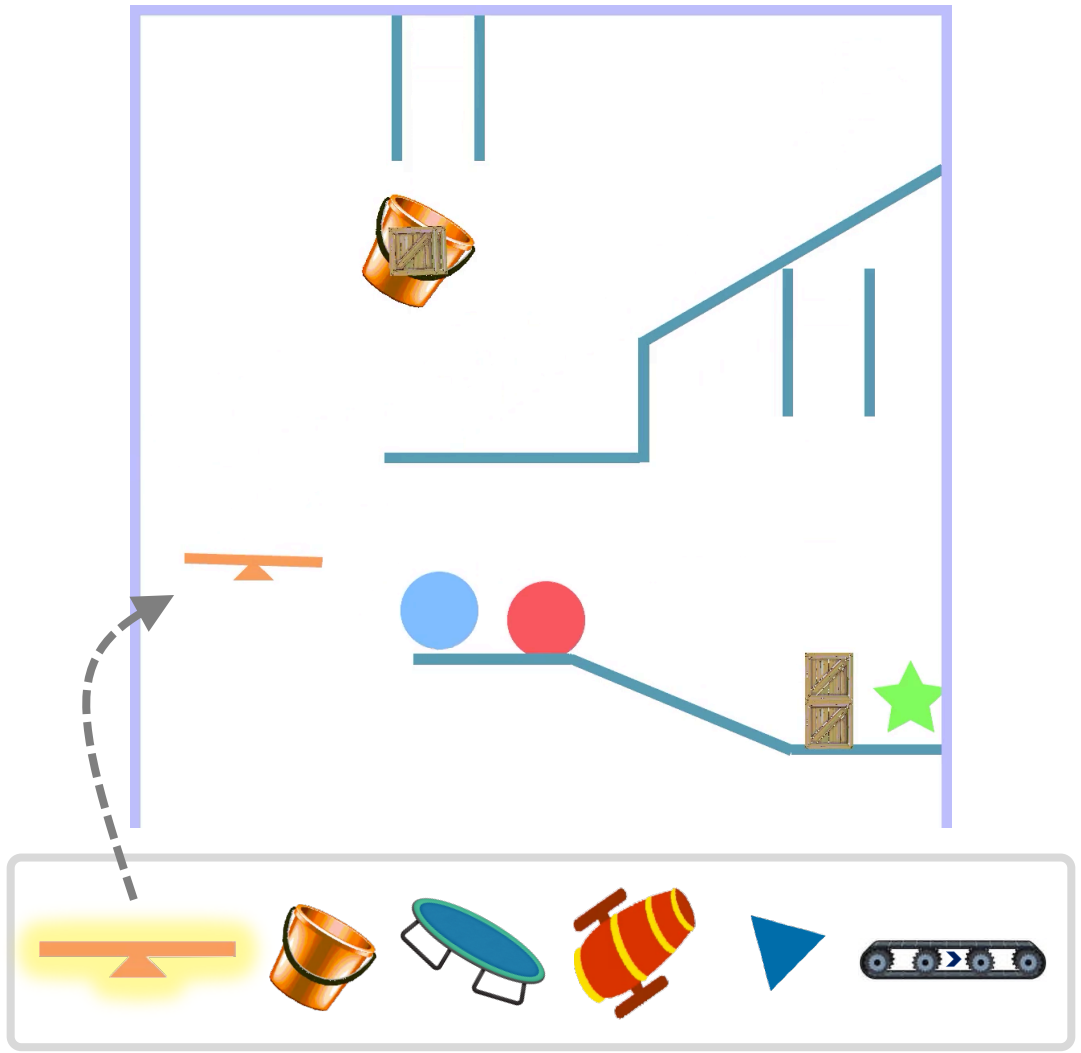


Select and Place Tools

Sequential Decision-making

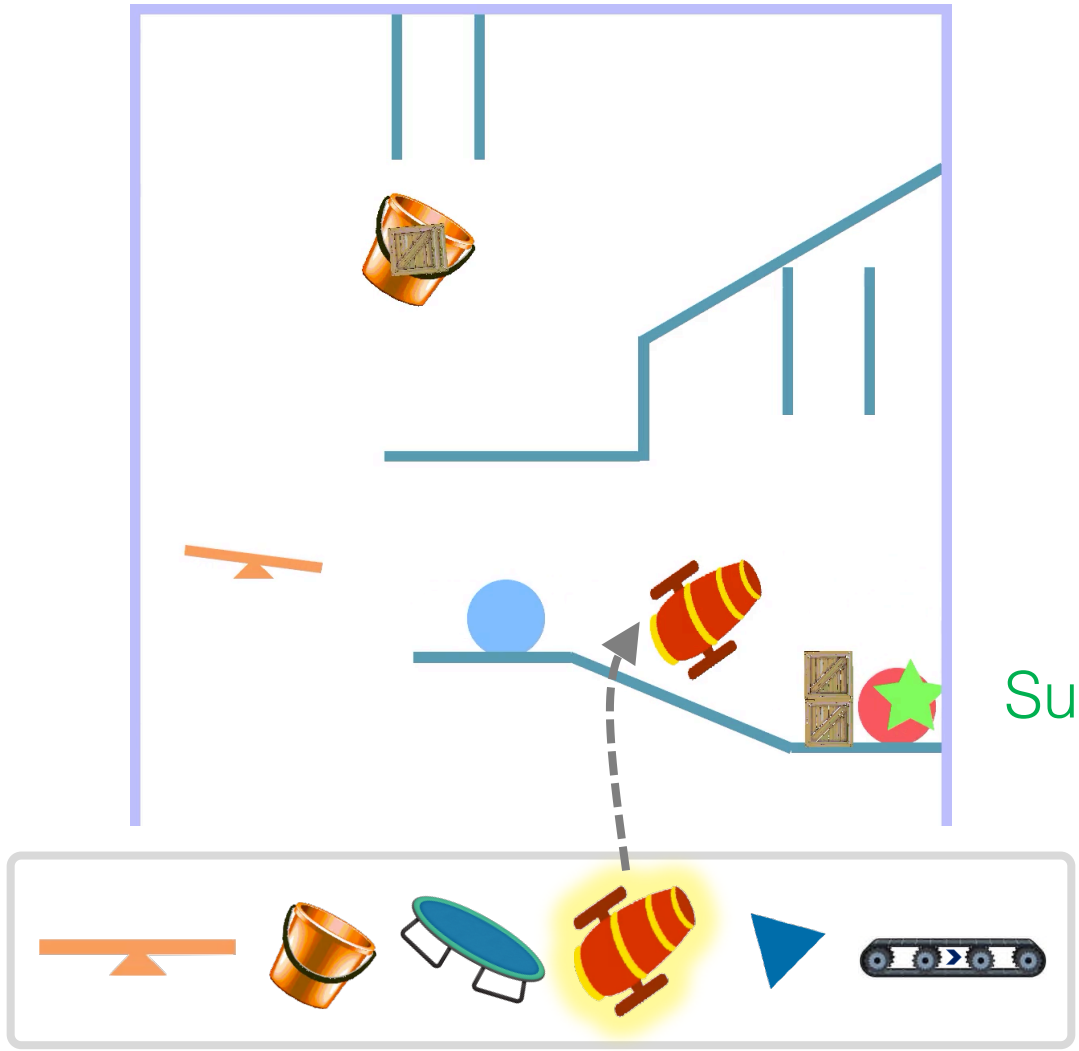


Sequential Decision-making



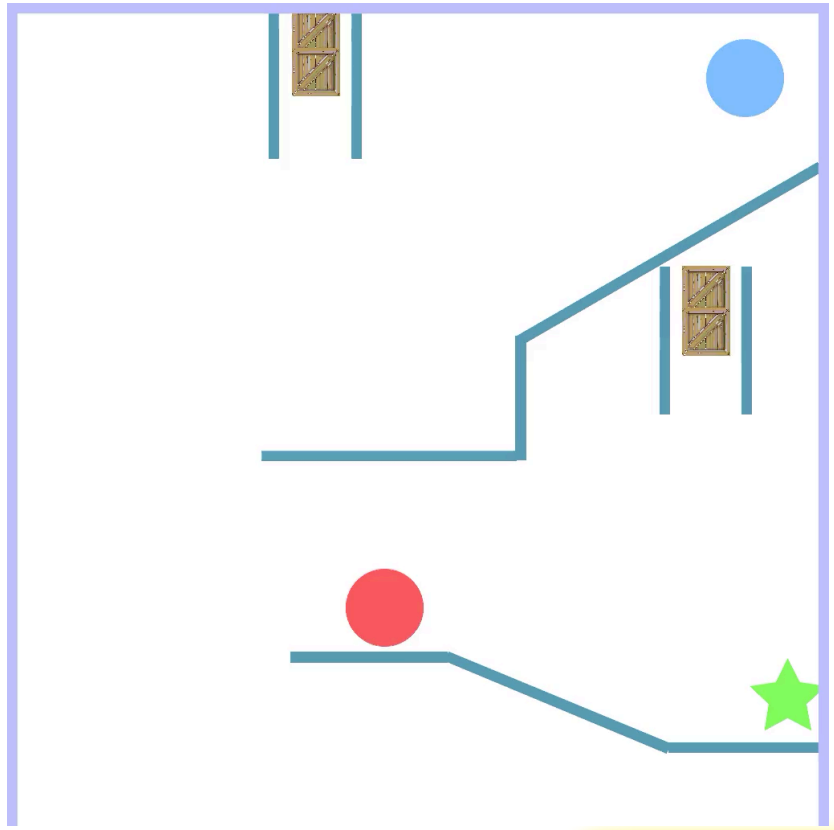
Environment Reward

Environment Reward

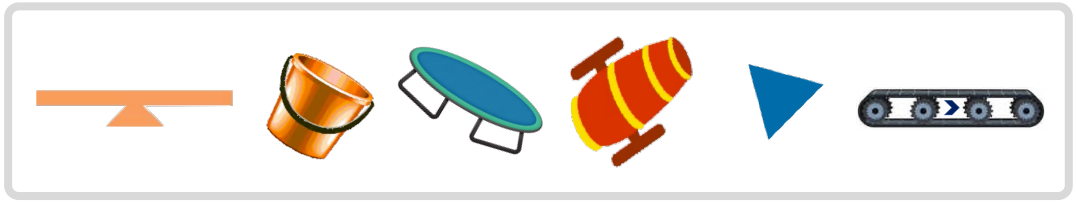


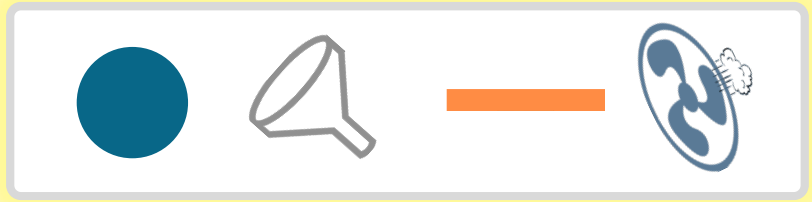
Success!

How to solve the task with new tools?

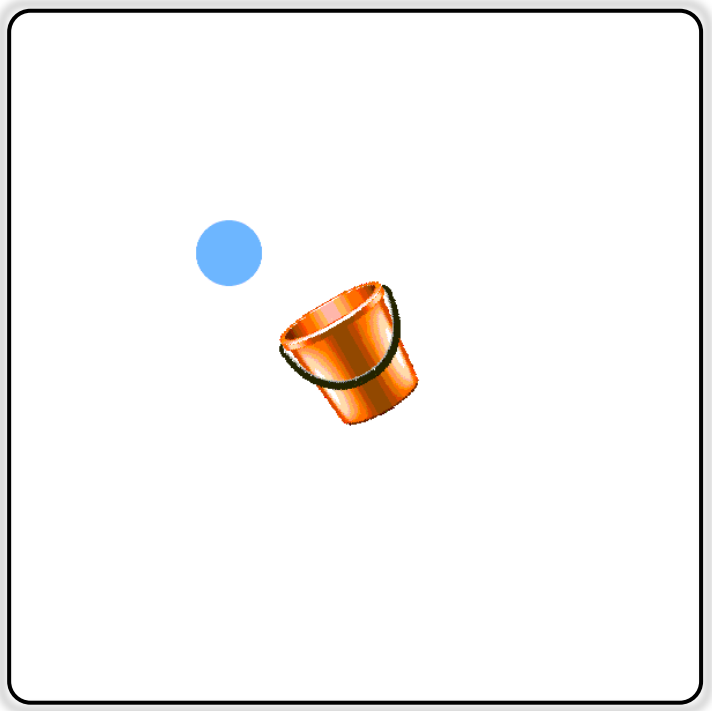


New Tools

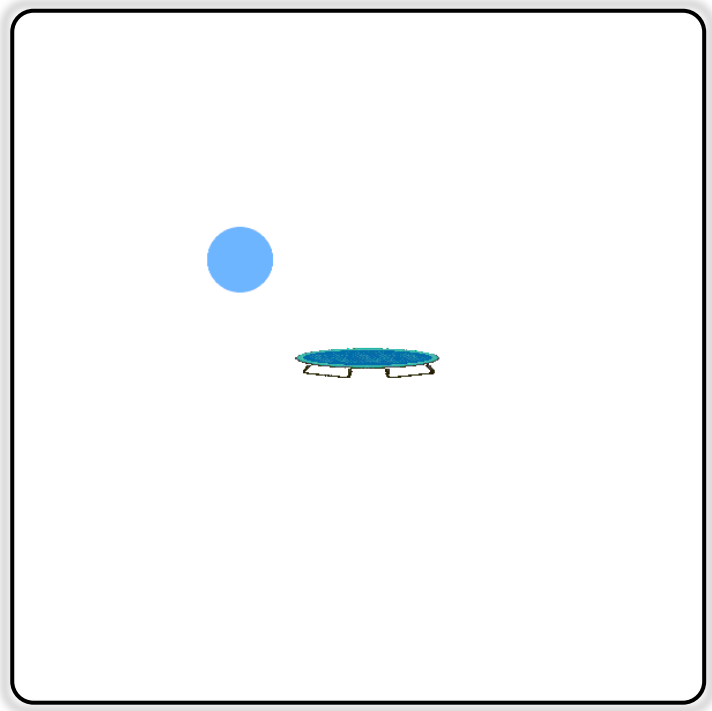




Action Observations



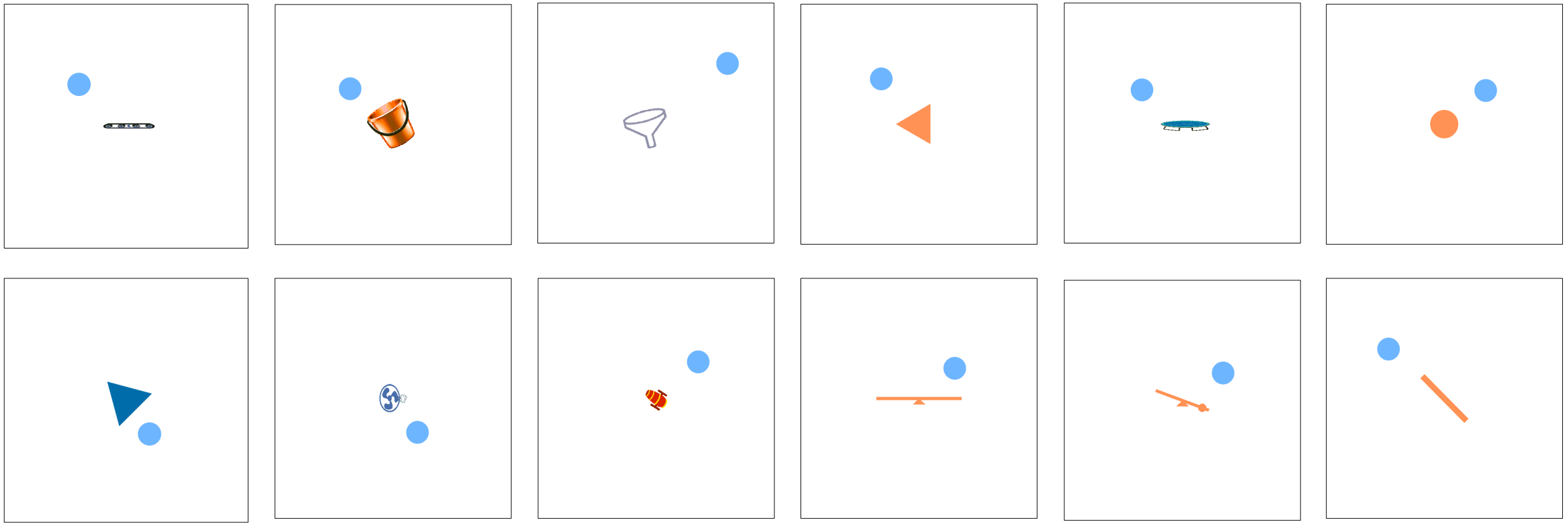
Bucket



Trampoline

Diverse behaviors of tools

CREATE Tools

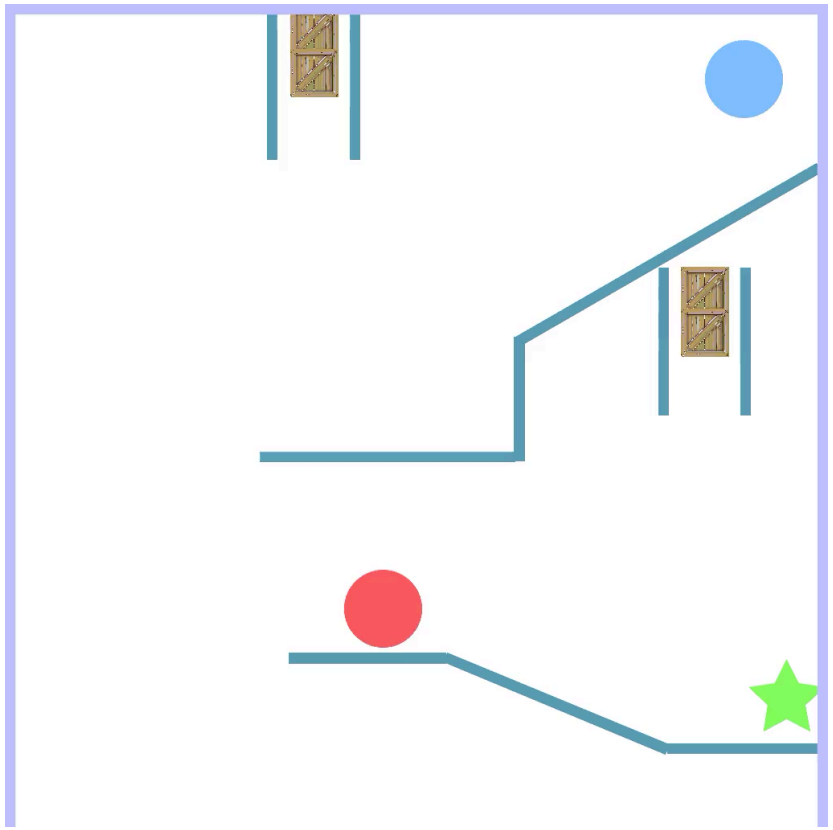


More tools generated by varying the parameters of these tool types

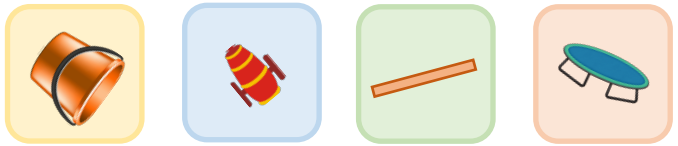
Approach

Approach

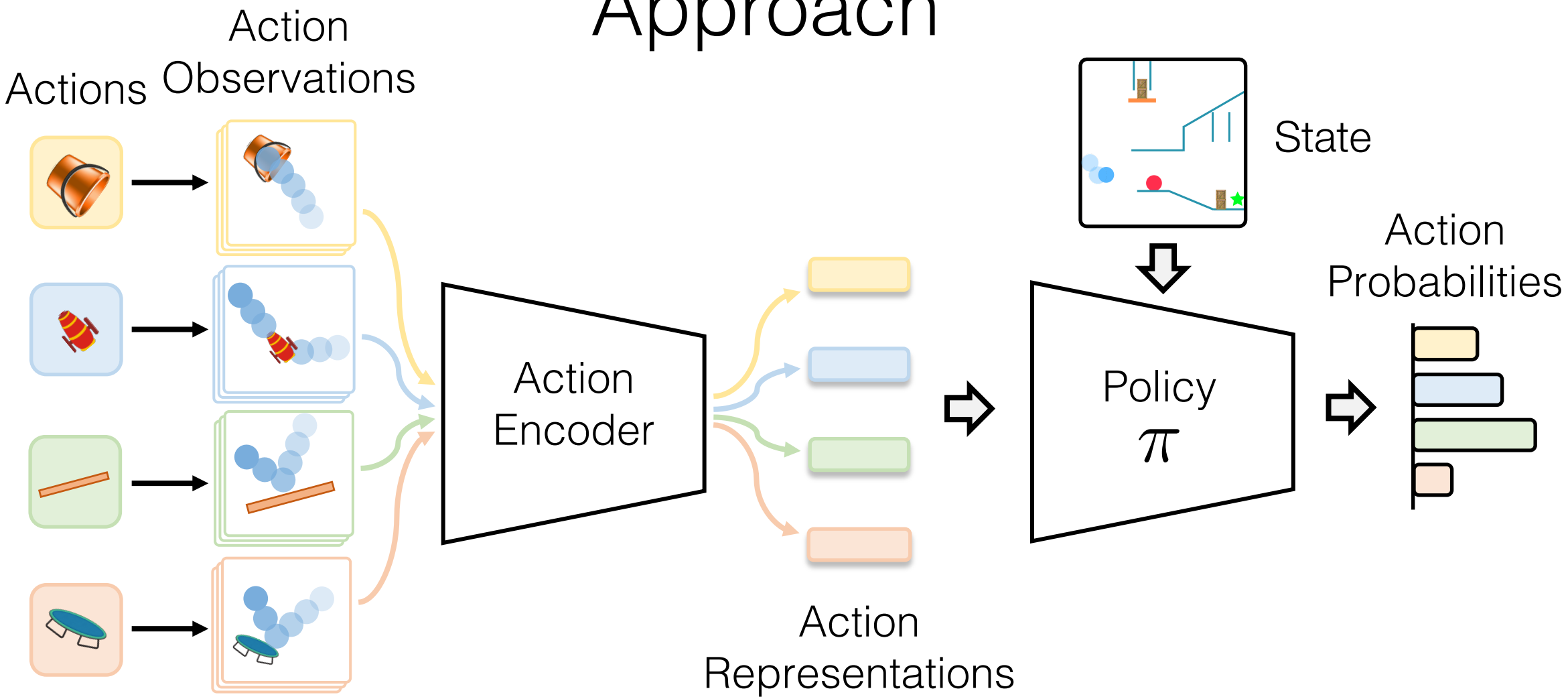
Task



Action Space



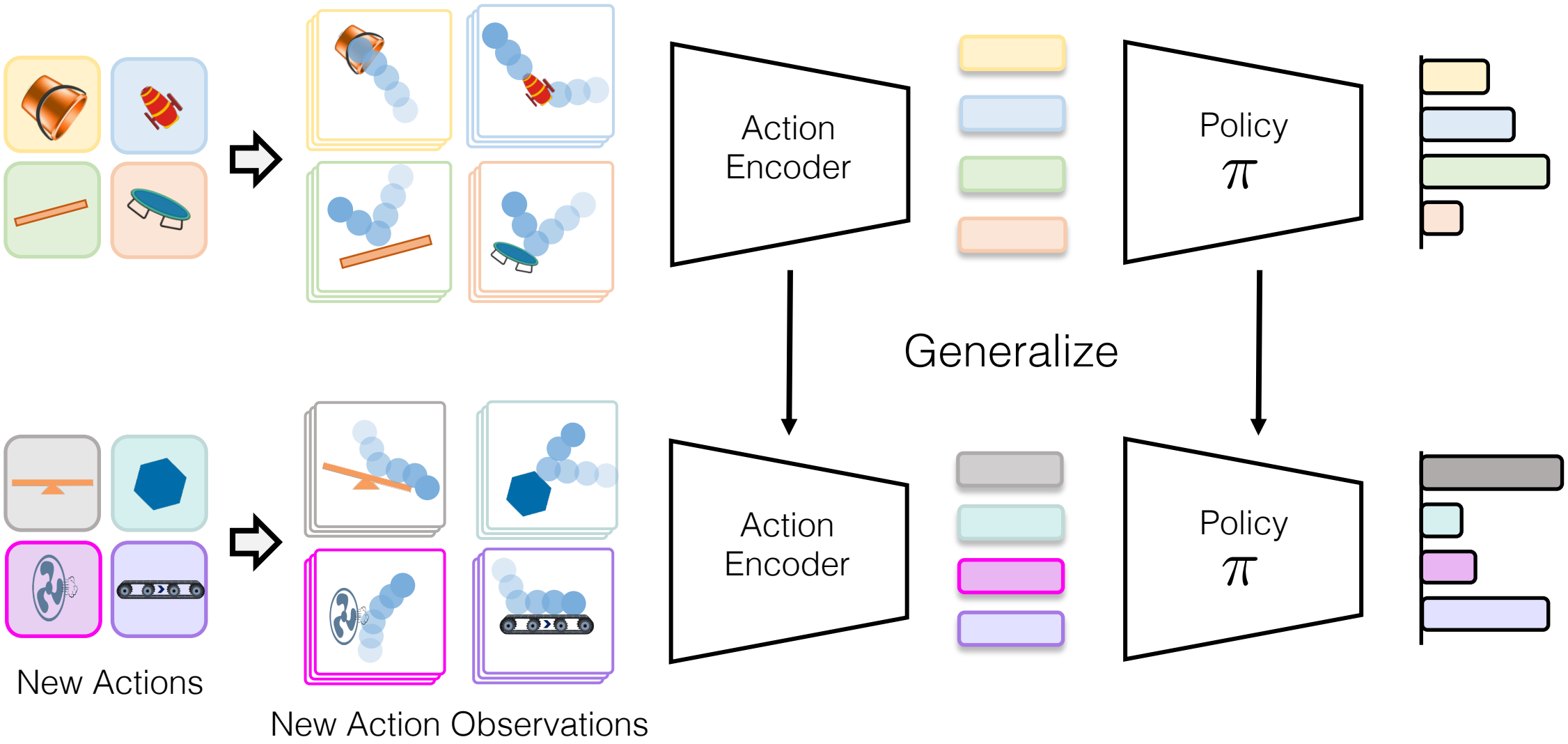
Approach



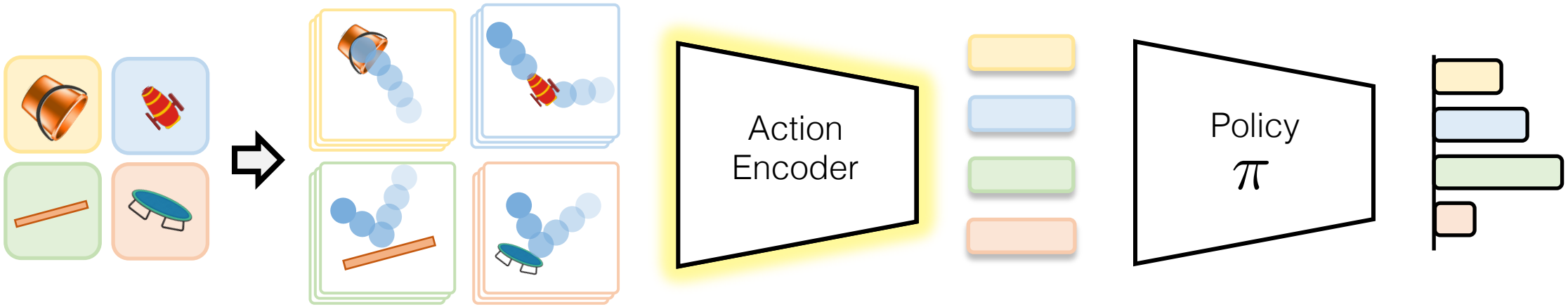
(1) Learn Action Representations

(2) Learn Generalizable Policy

Same Pipeline for New Actions



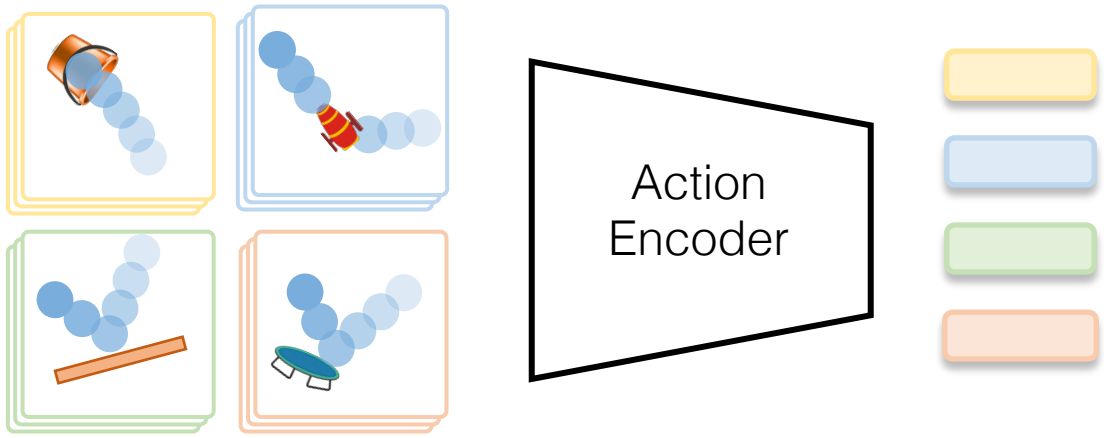
Training Procedure



Unsupervised Representation Learning

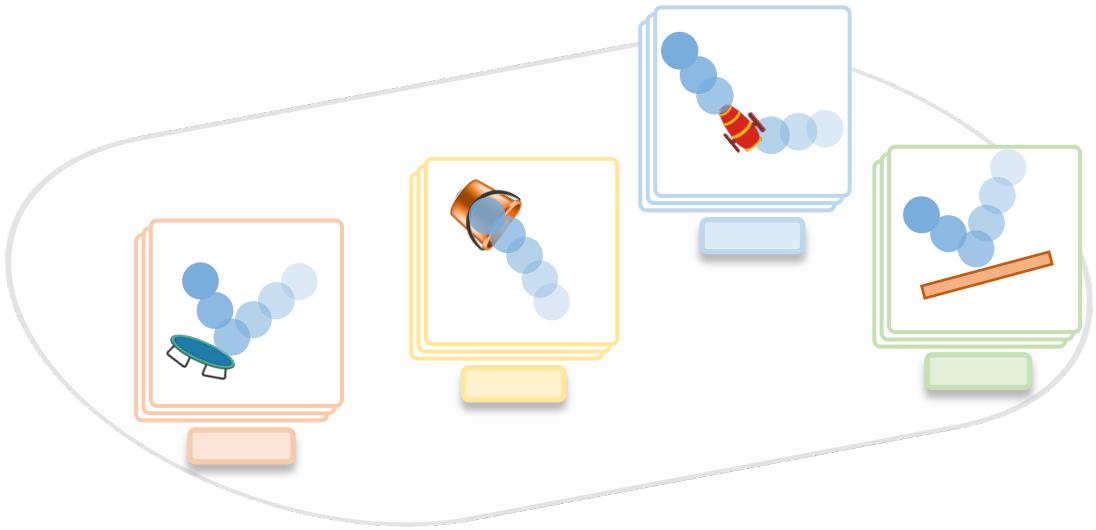
Reinforcement Learning

Action Encoder



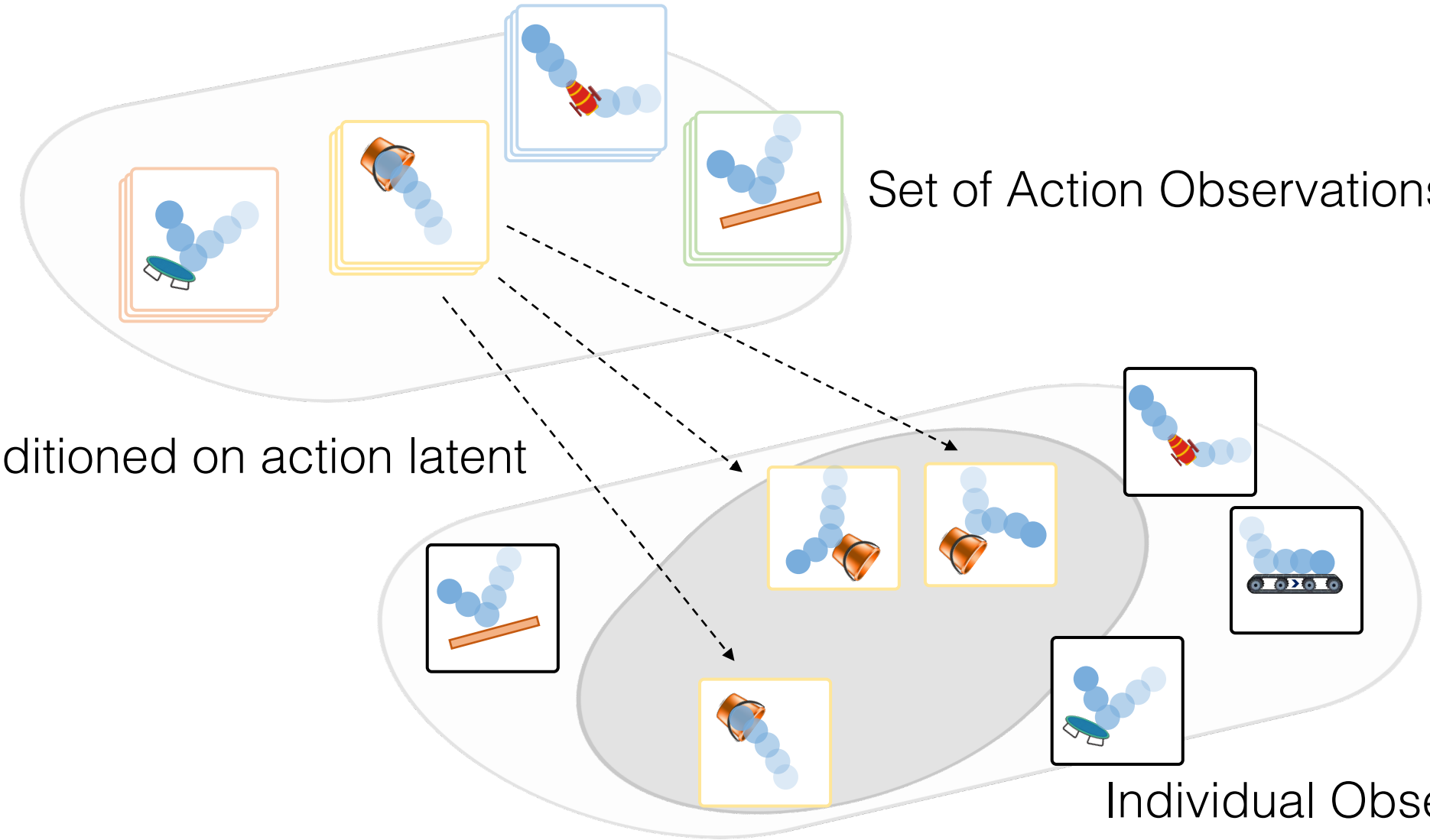
Hierarchical Variational Auto-encoder (HVAE) architecture (Edwards & Storkey, 2017)

Hierarchical Latent Spaces

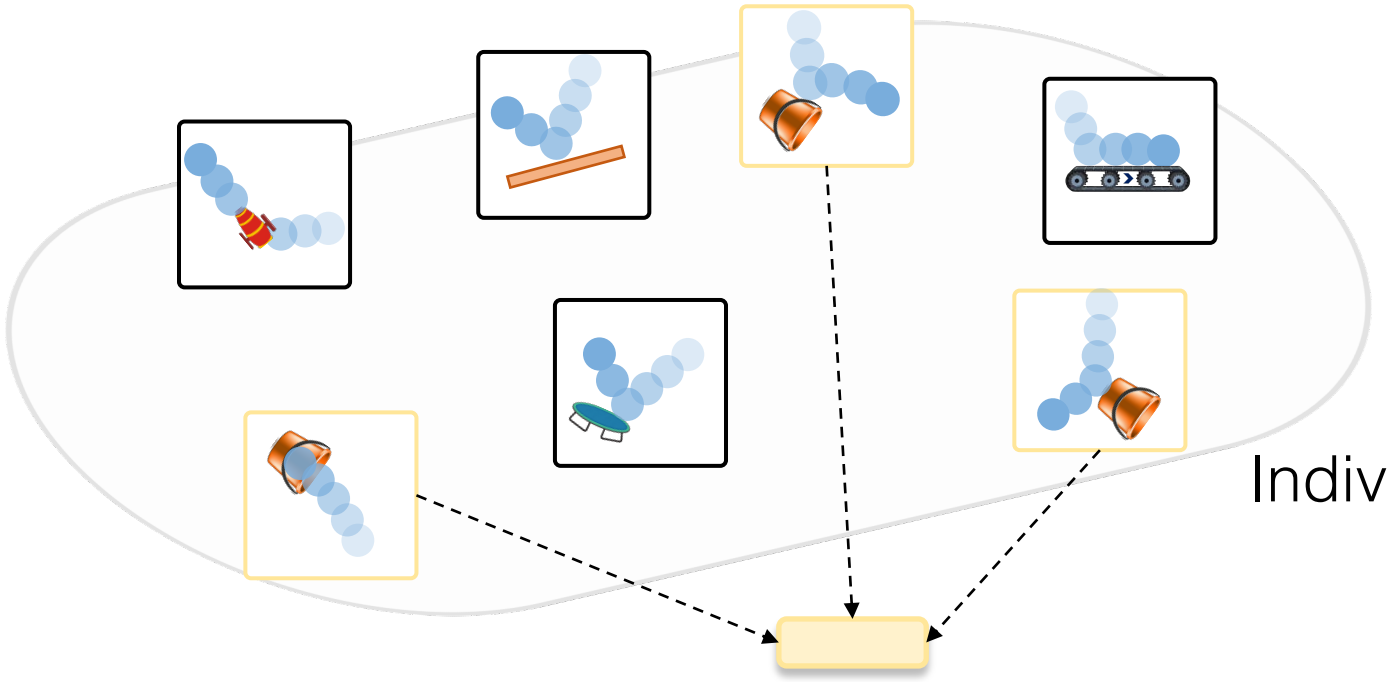


Set of Action Observations

Hierarchical Latent Spaces



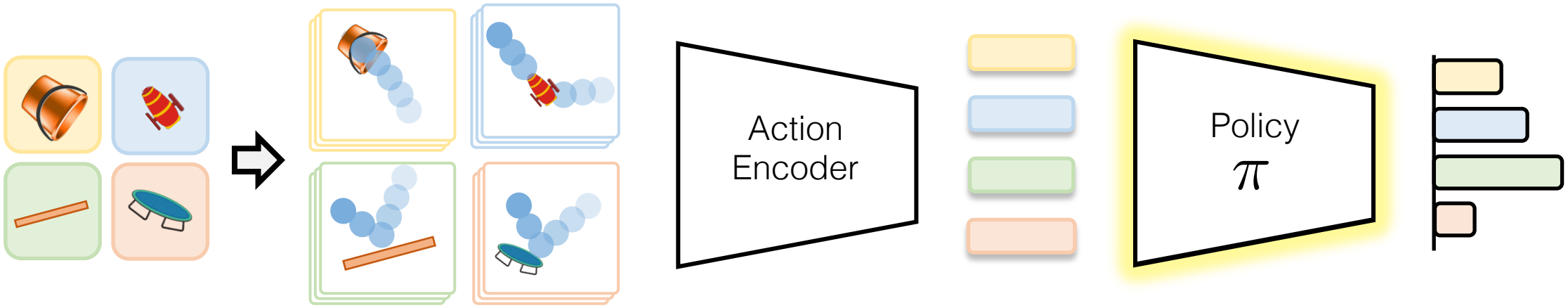
Flat VAE without Hierarchy?



Individual Observations

Averaging observation latents

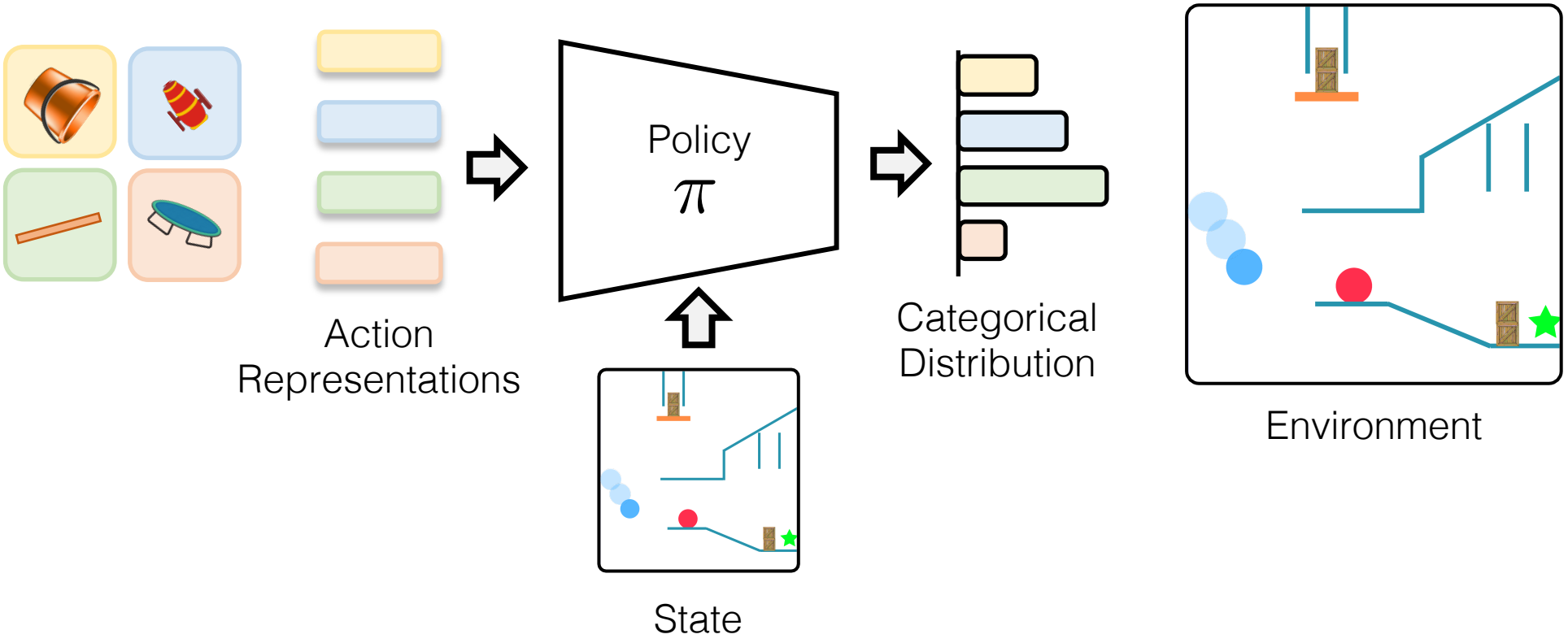
Policy Architecture



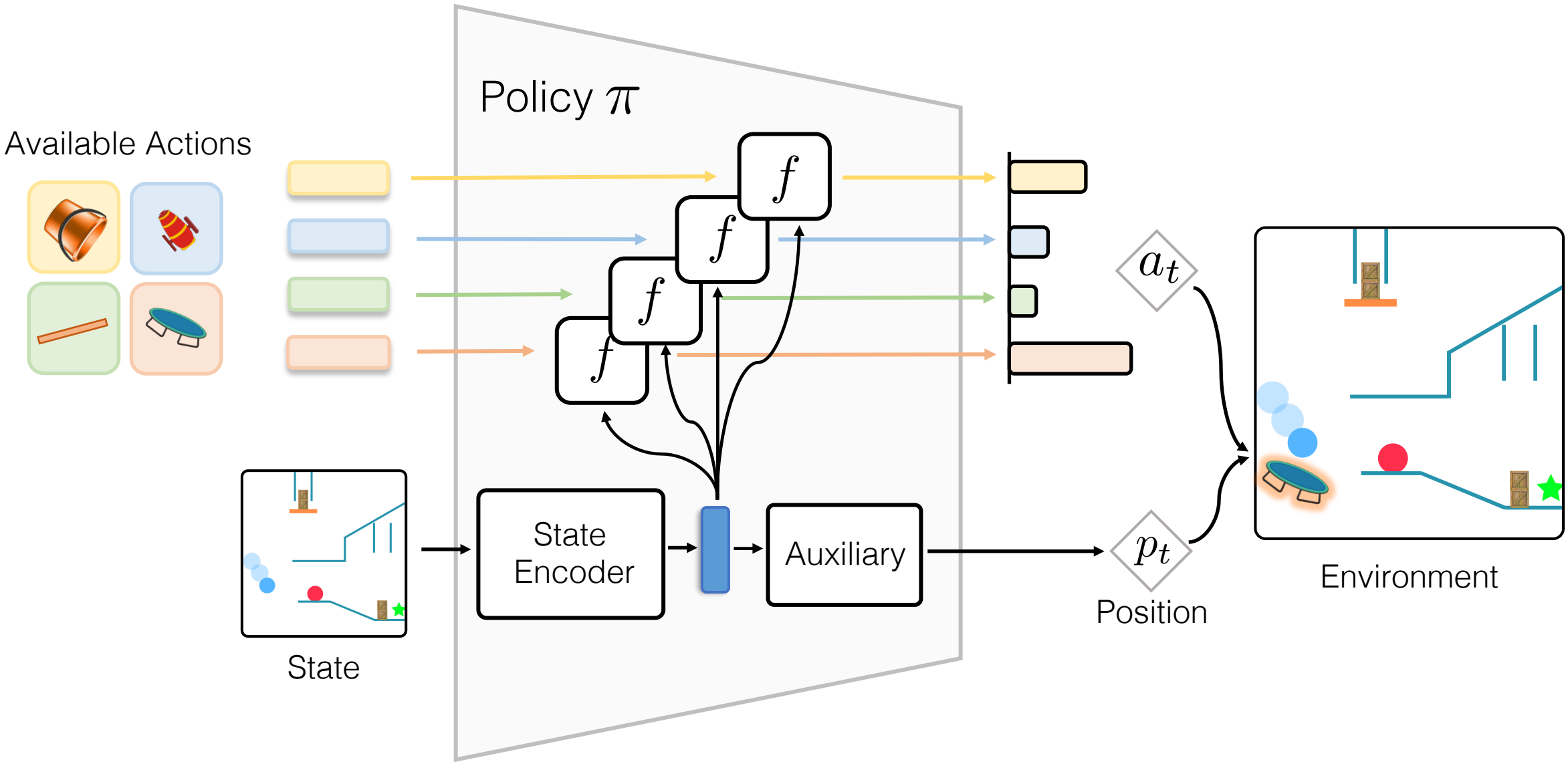
Unsupervised Representation Learning

Reinforcement Learning

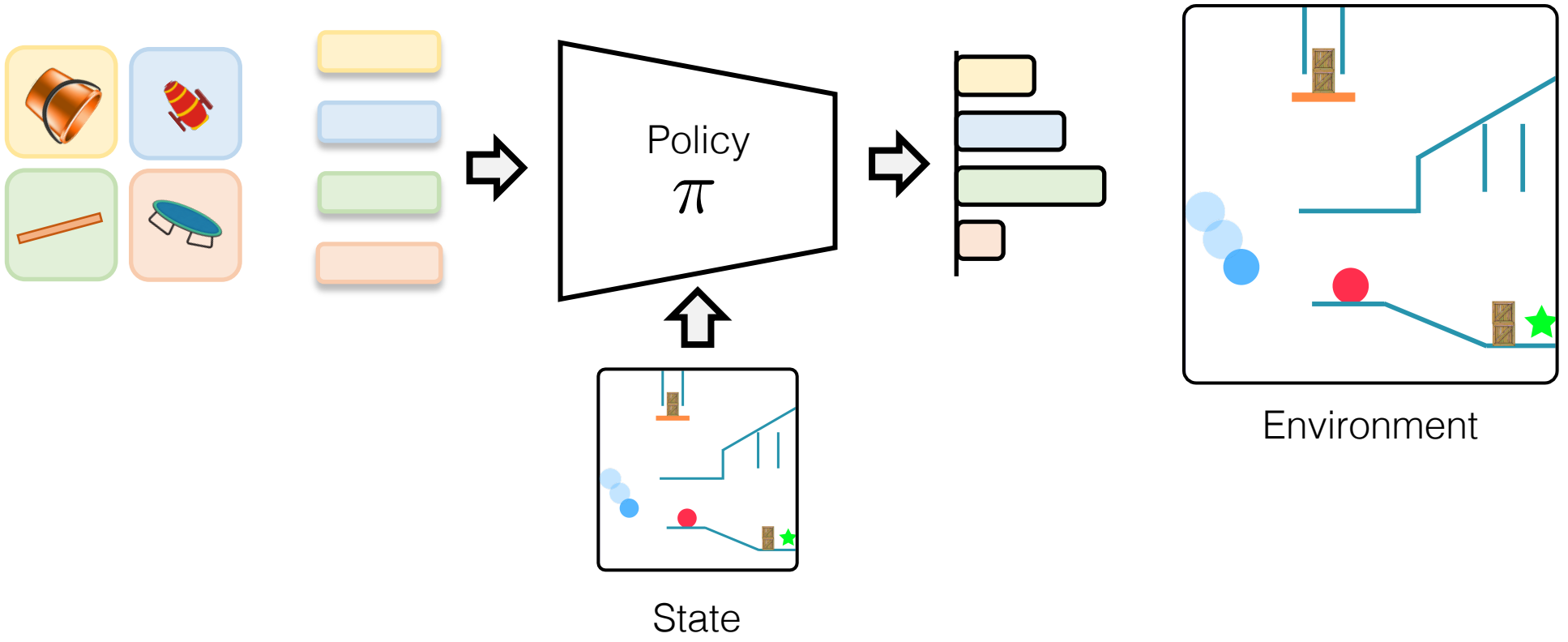
Policy Architecture



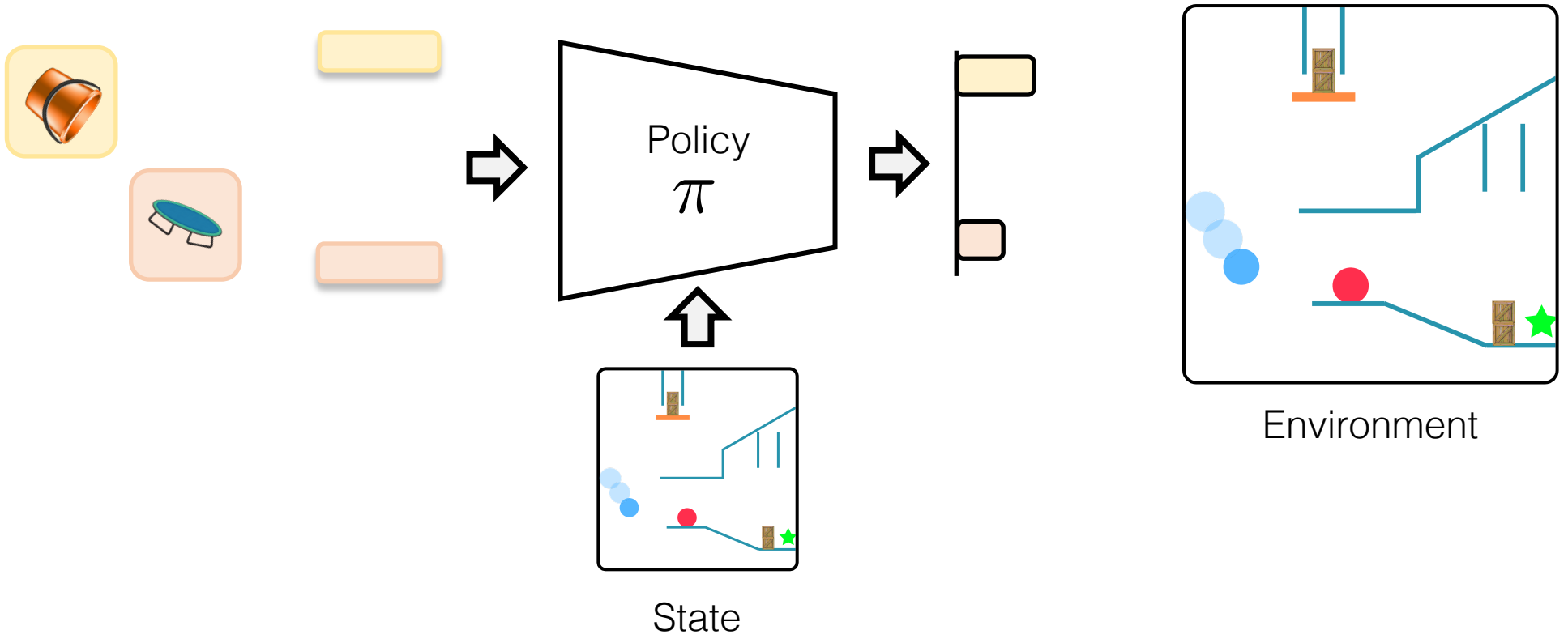
Policy Architecture



Avoiding Overfitting in RL

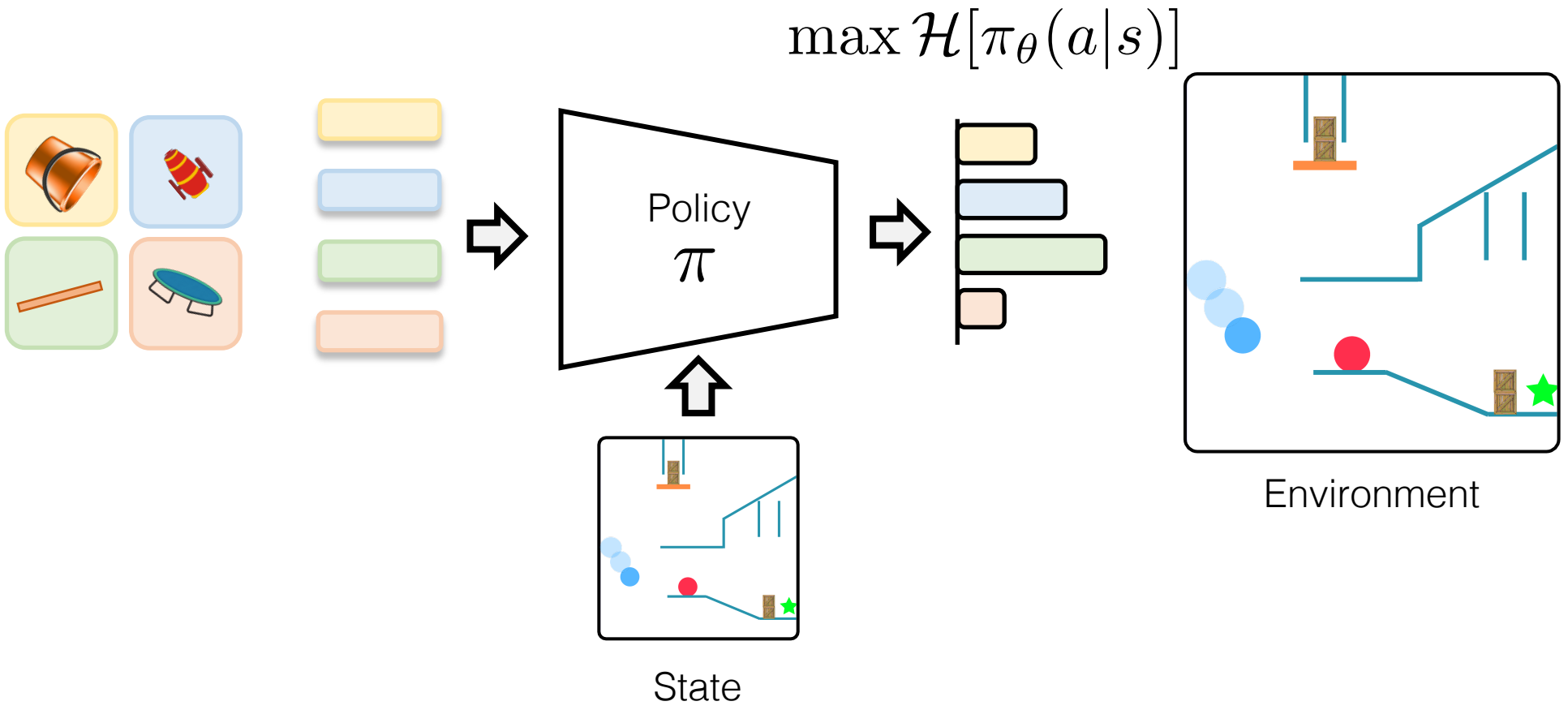


Action Dropout



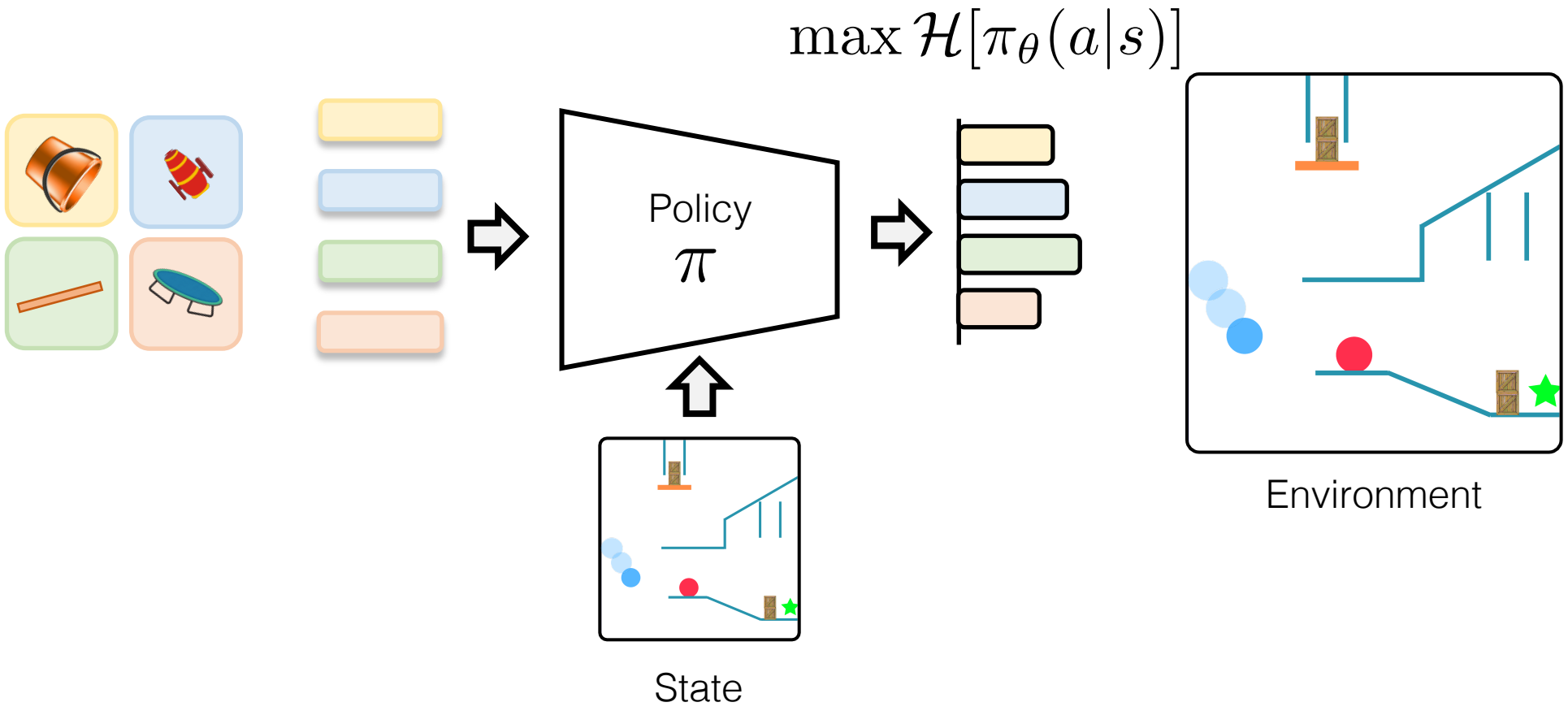
Avoid overfitting to certain actions

Maximum Entropy RL



Encourage diverse actions

Maximum Entropy RL

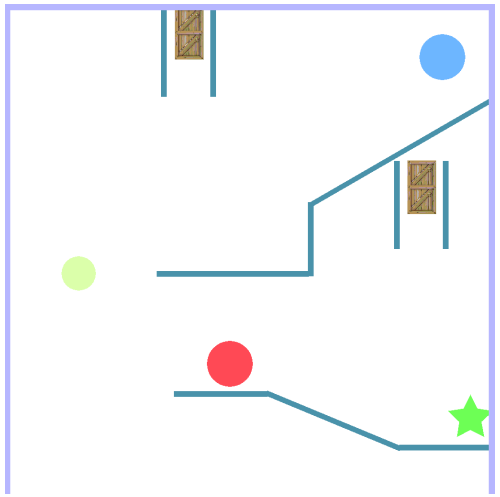


Encourage diverse actions

Environments

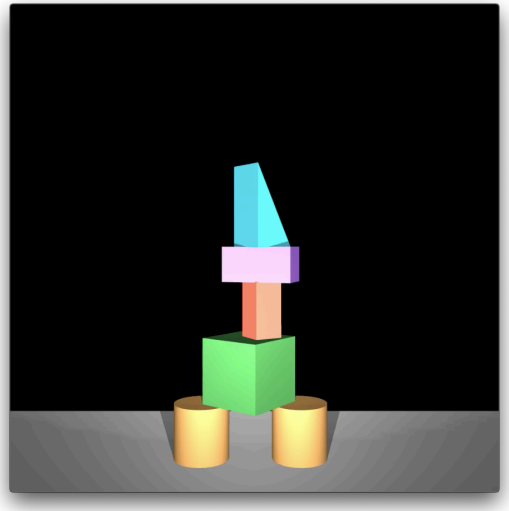
Environments

CREATE



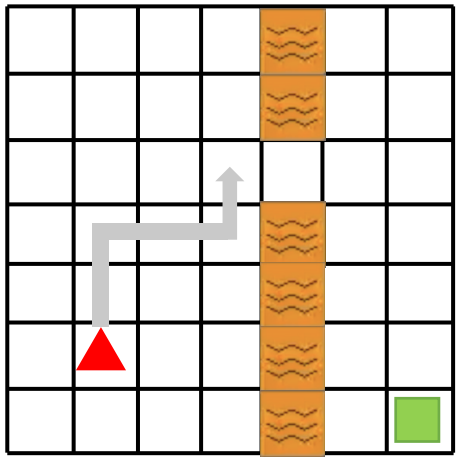
Unseen Tools

Shape Stacking



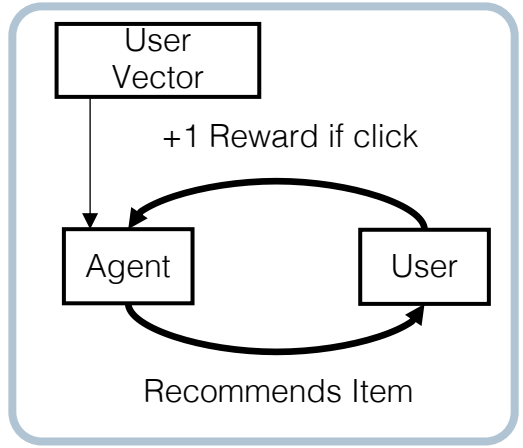
Unseen 3D Shapes

2D Navigation



Unseen Skills

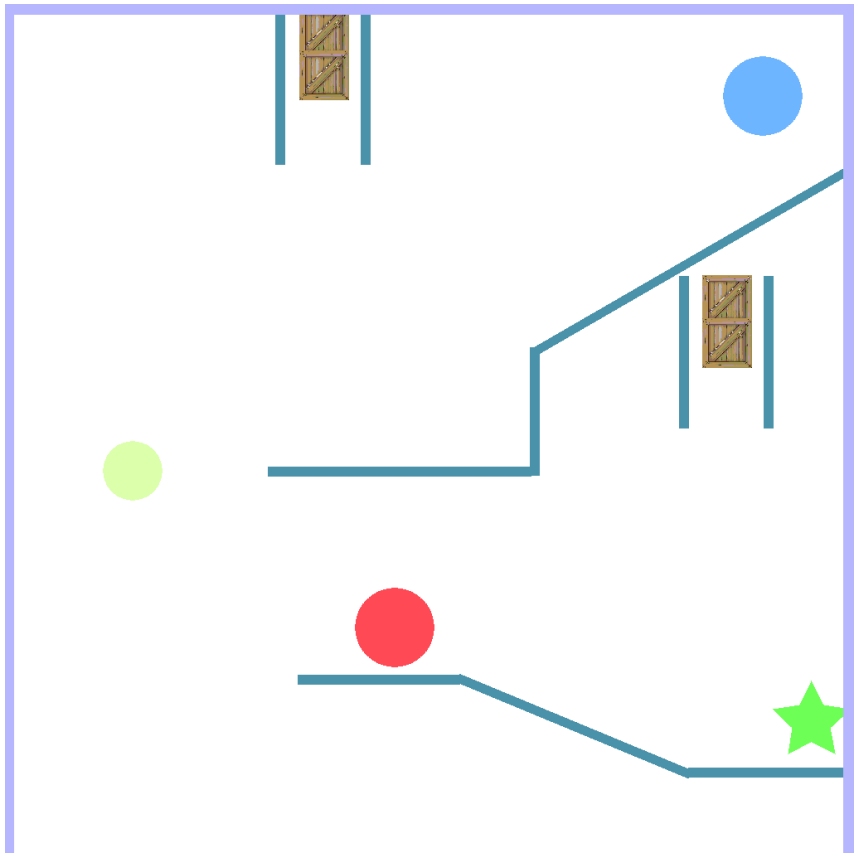
Recommender System



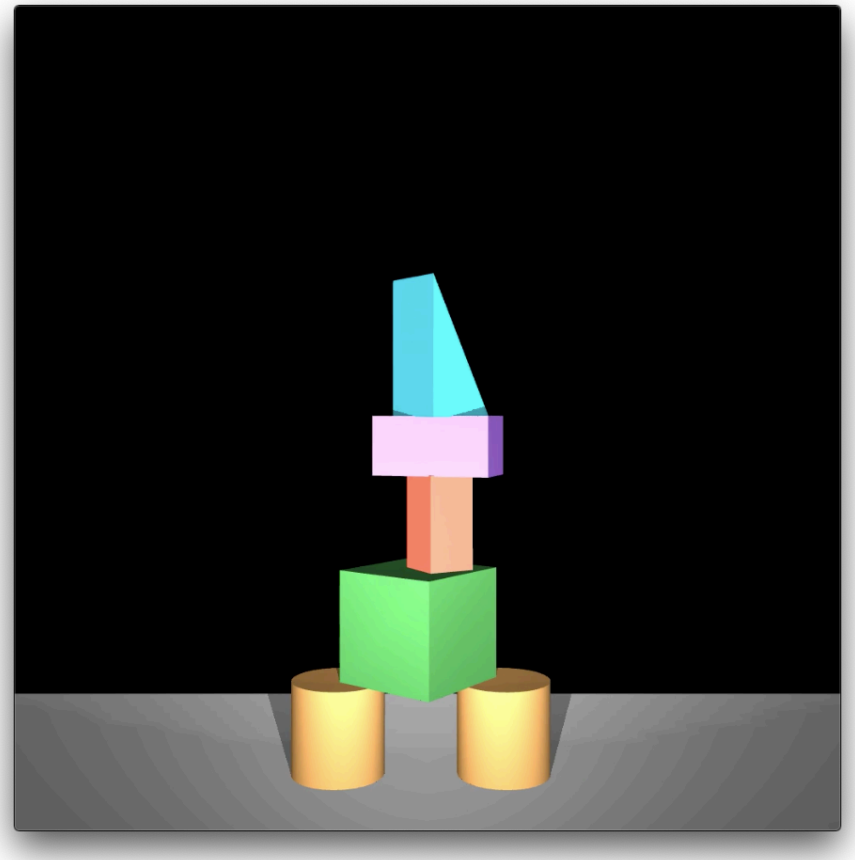
Unseen Products

Environments

CREATE

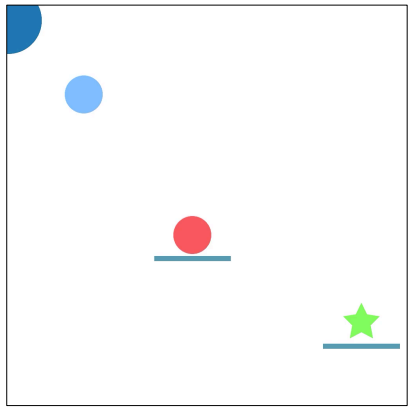


Shape Stacking

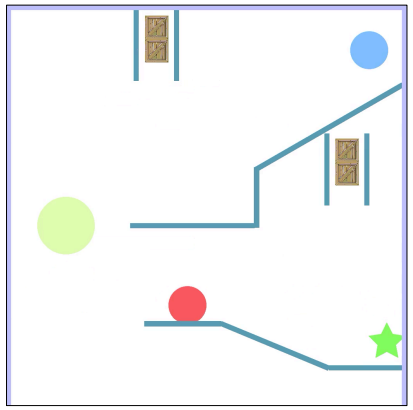


12 CREATE Tasks

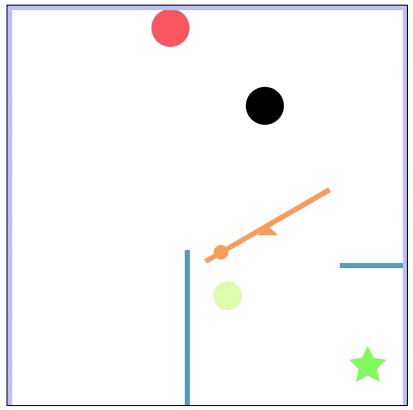
Push



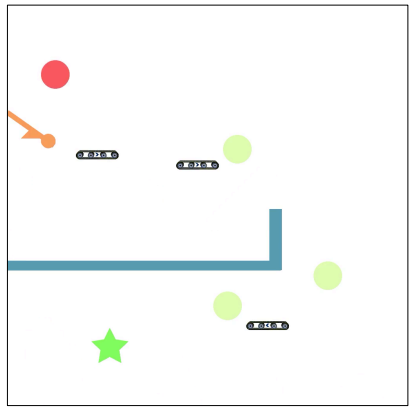
Obstacle



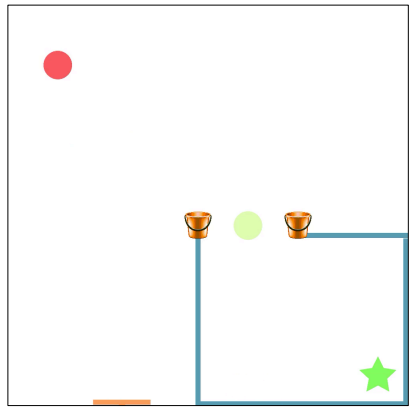
See-Saw



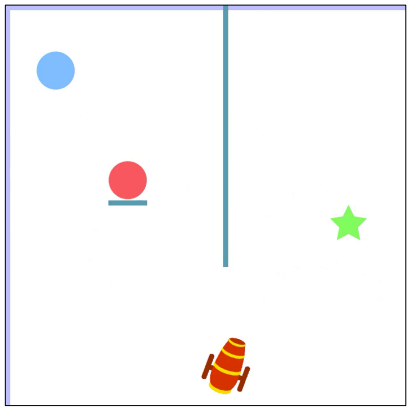
Belt



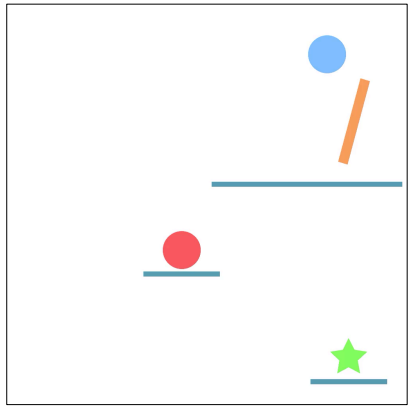
Bucket



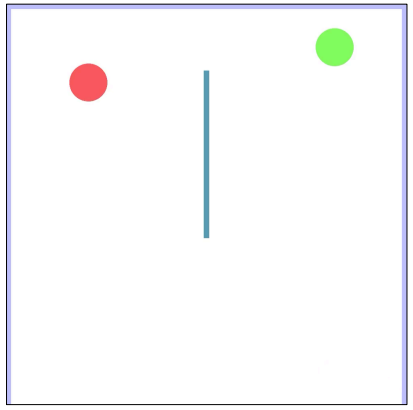
Cannon



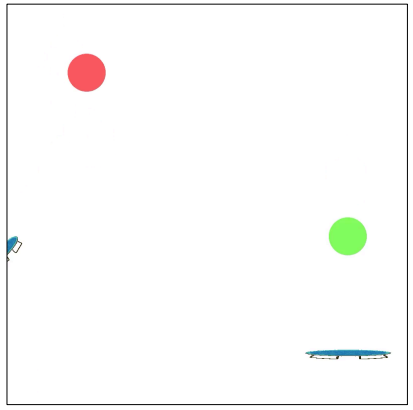
Navigate



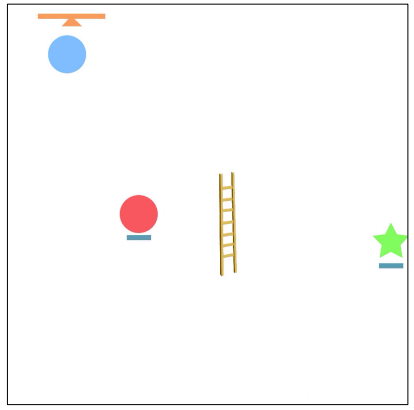
Collide



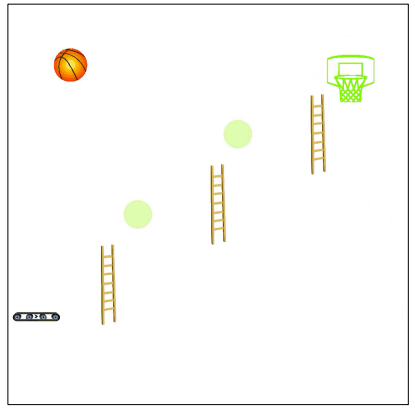
Moving



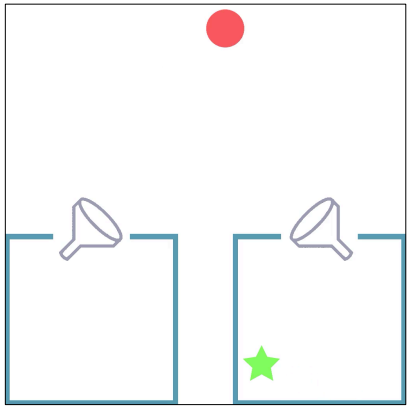
Ladder



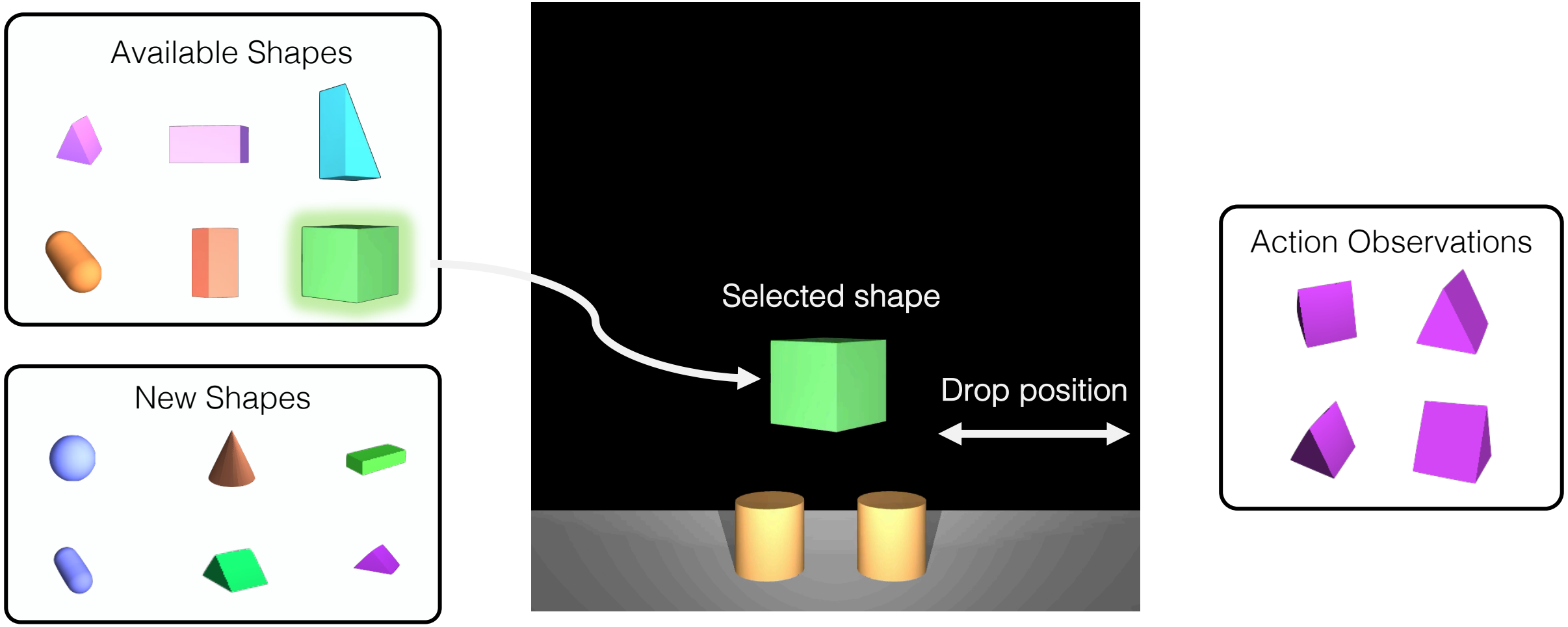
Basket



Funnel

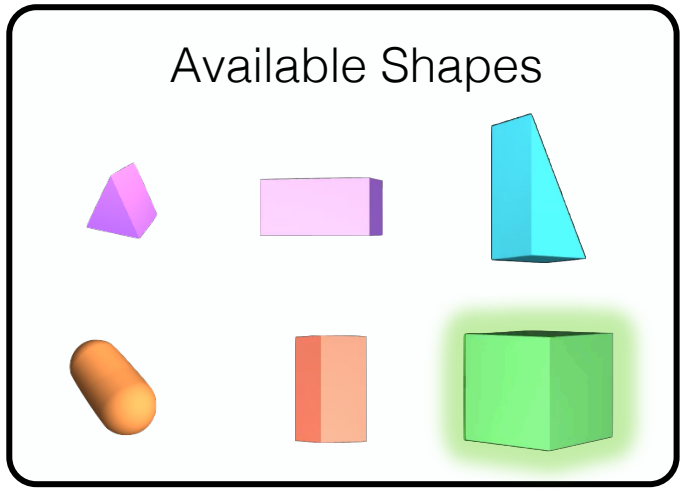


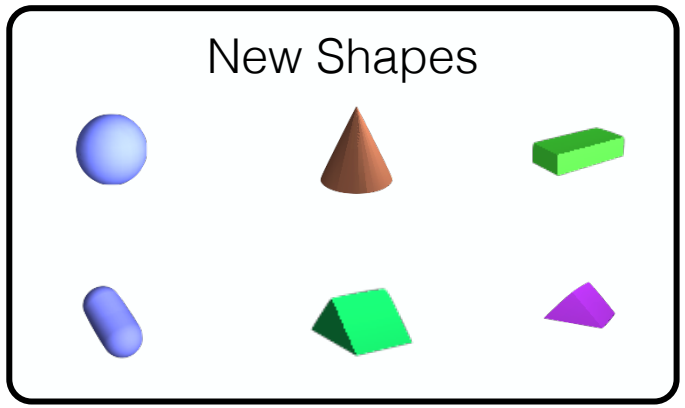
Shape Stacking

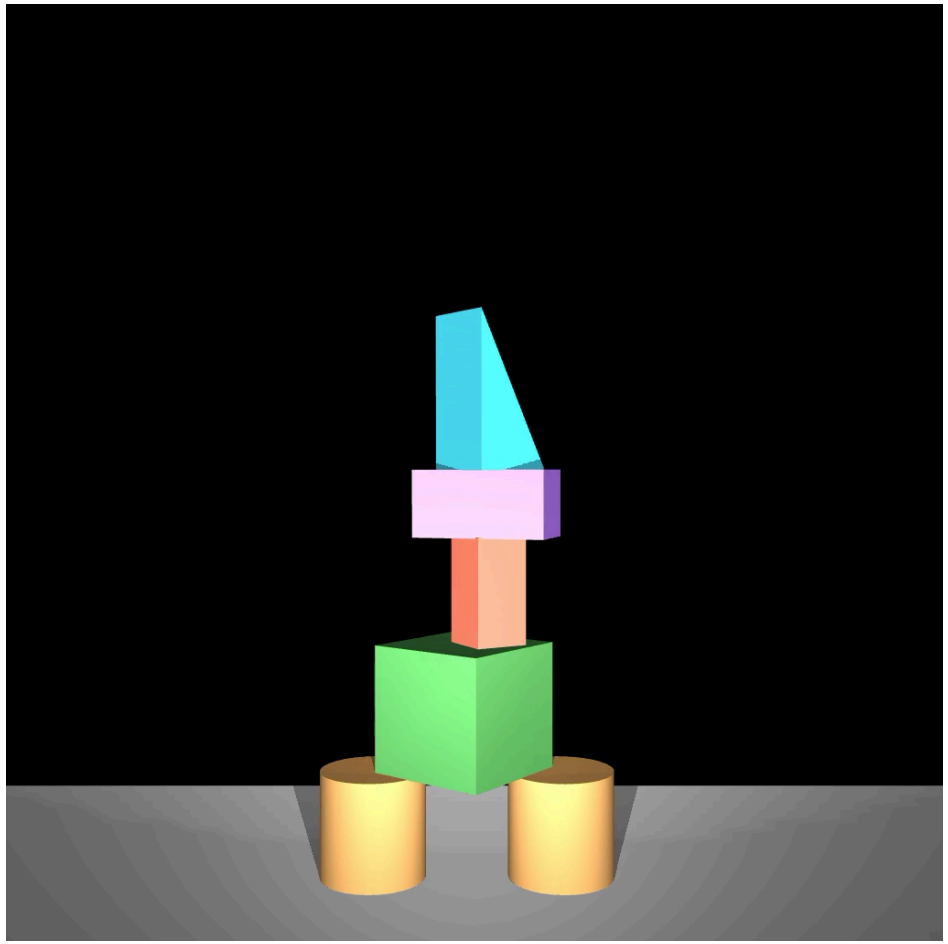


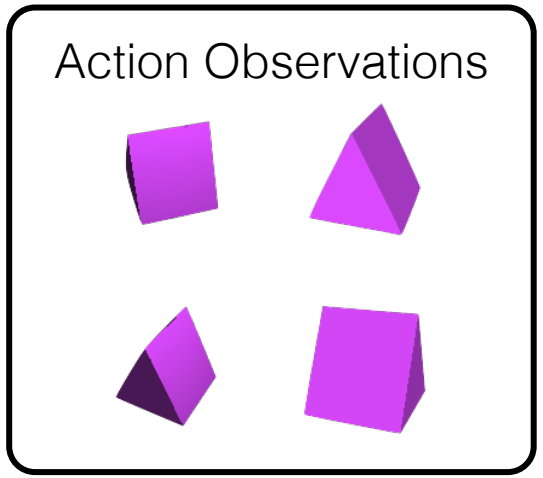
Task: Stack a stable tower

Shape Stacking





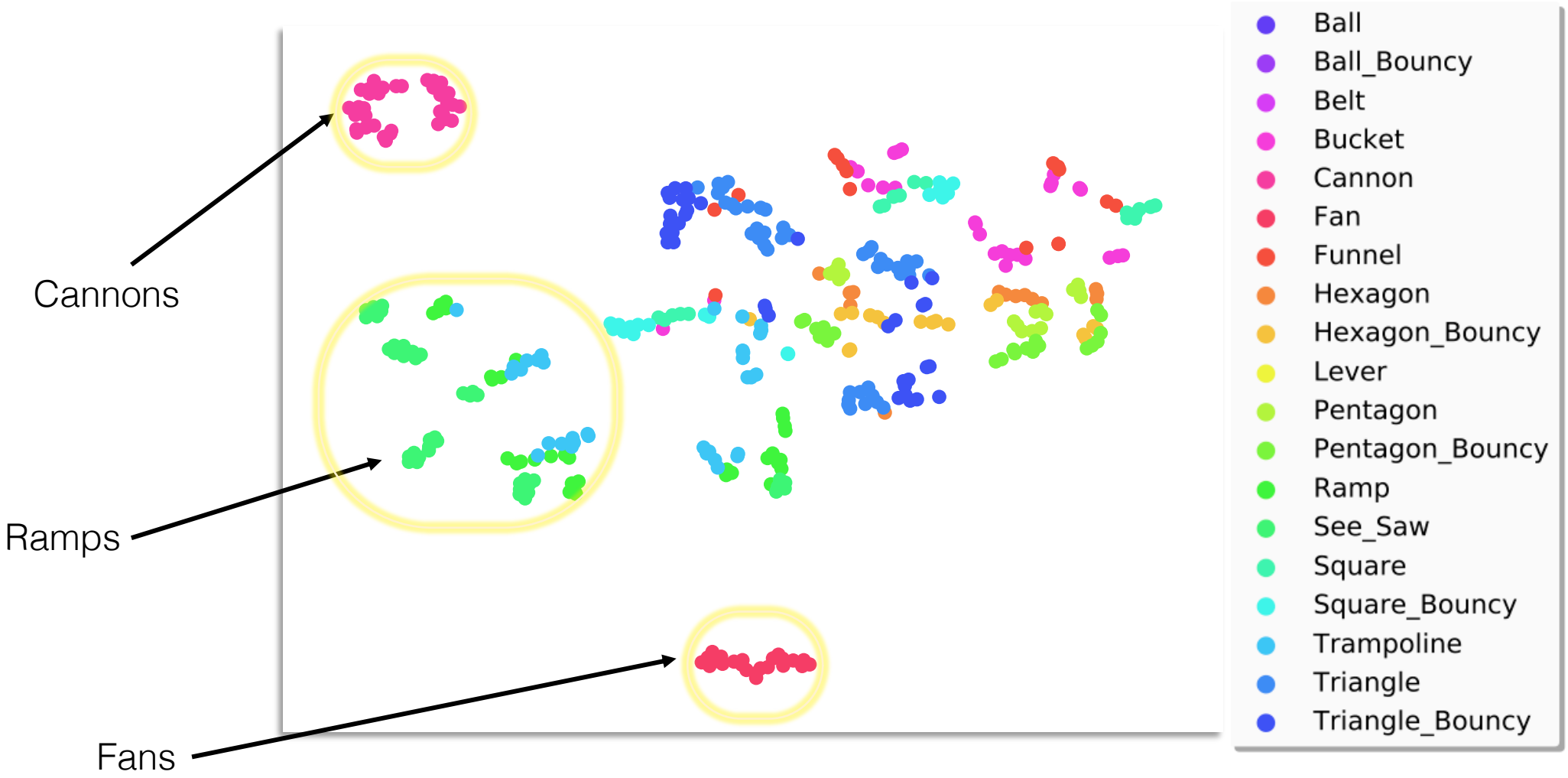




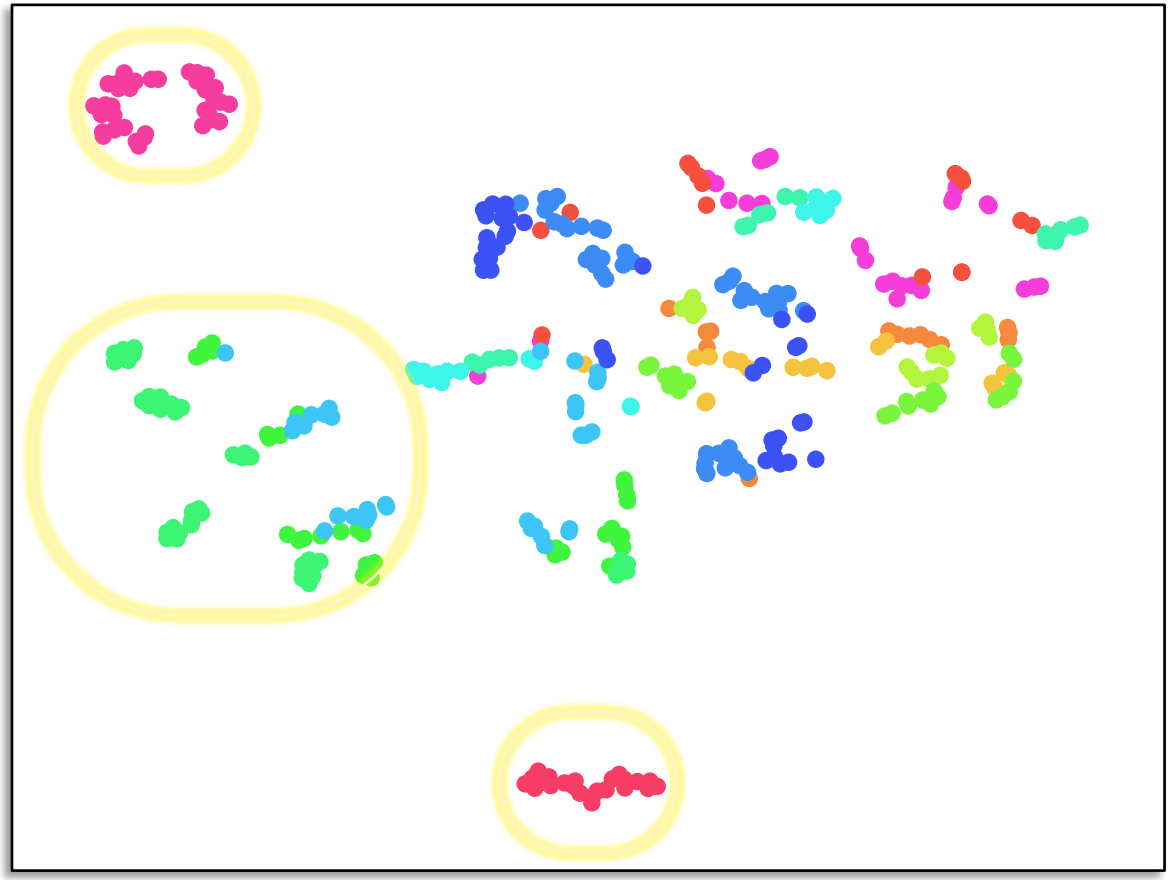
Task: Stack a stable tower

Results

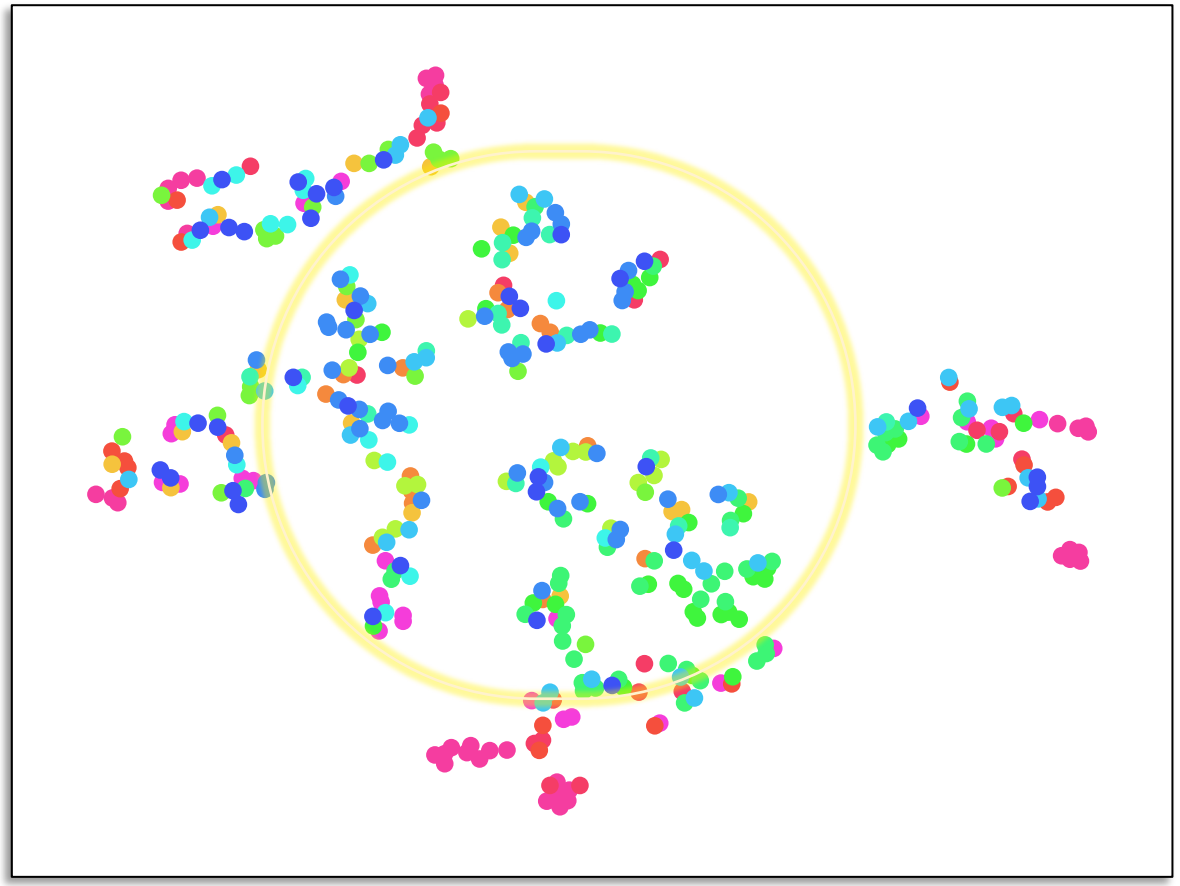
t-SNE Visualization of Representations



Hierarchy extracts semantic information

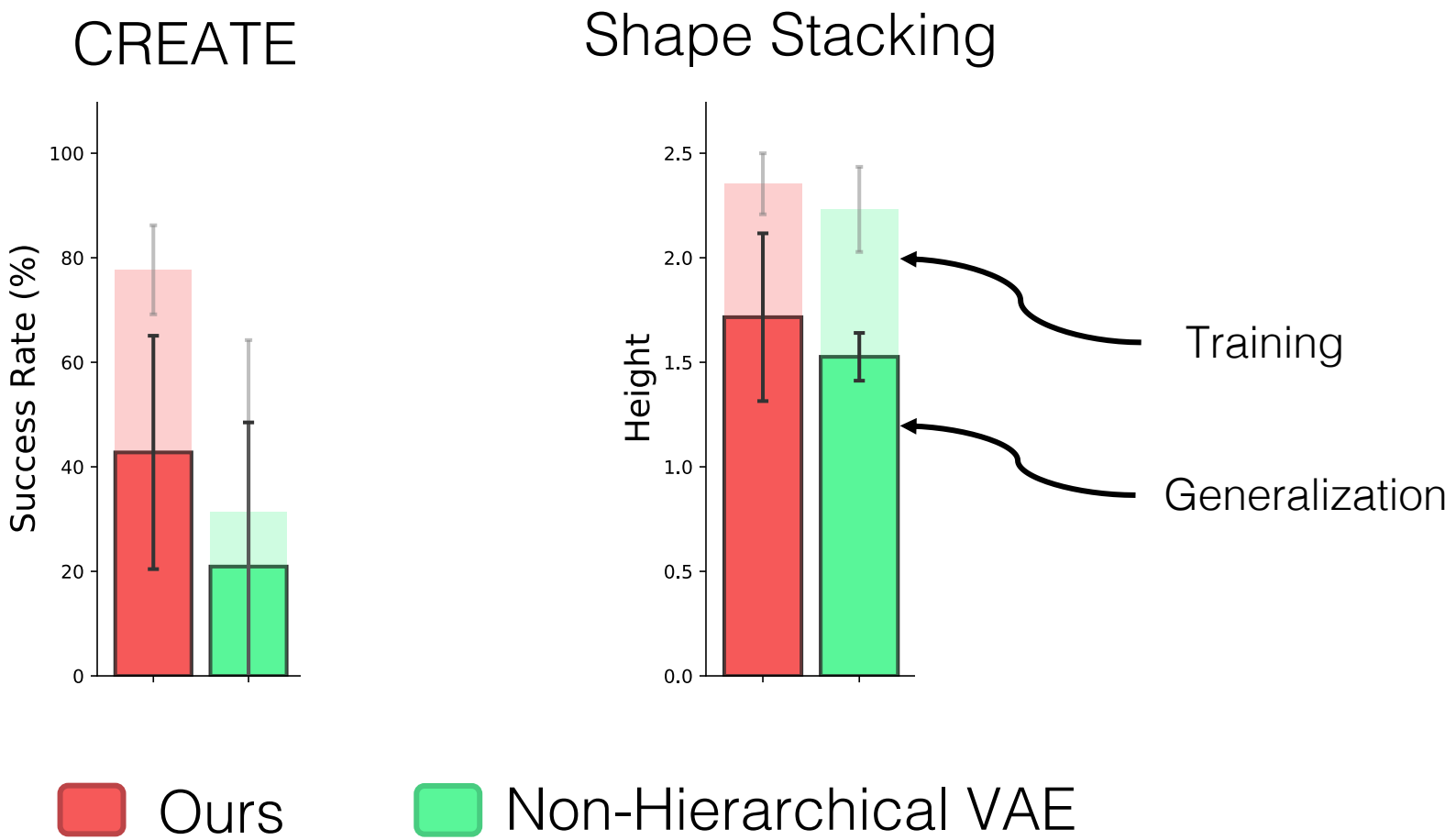


Hierarchical VAE

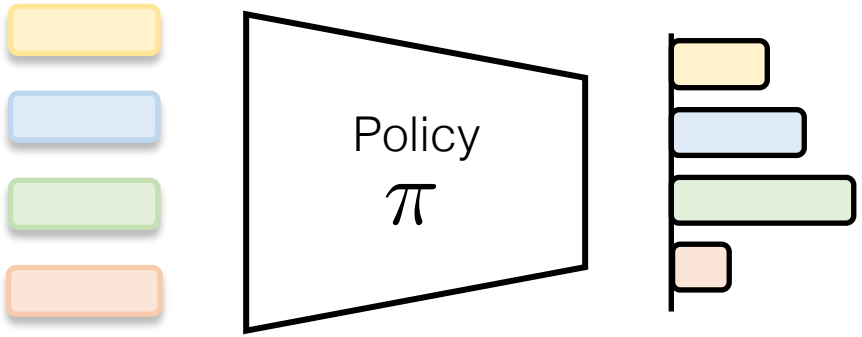


Flat VAE

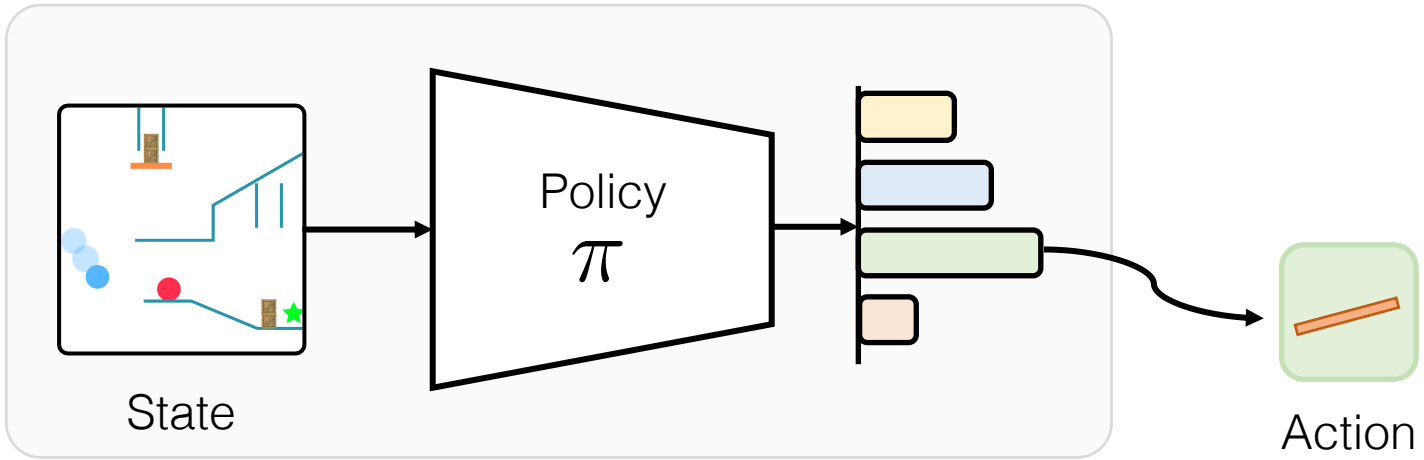
Hierarchy helps policy learning



Policy Architecture Baselines

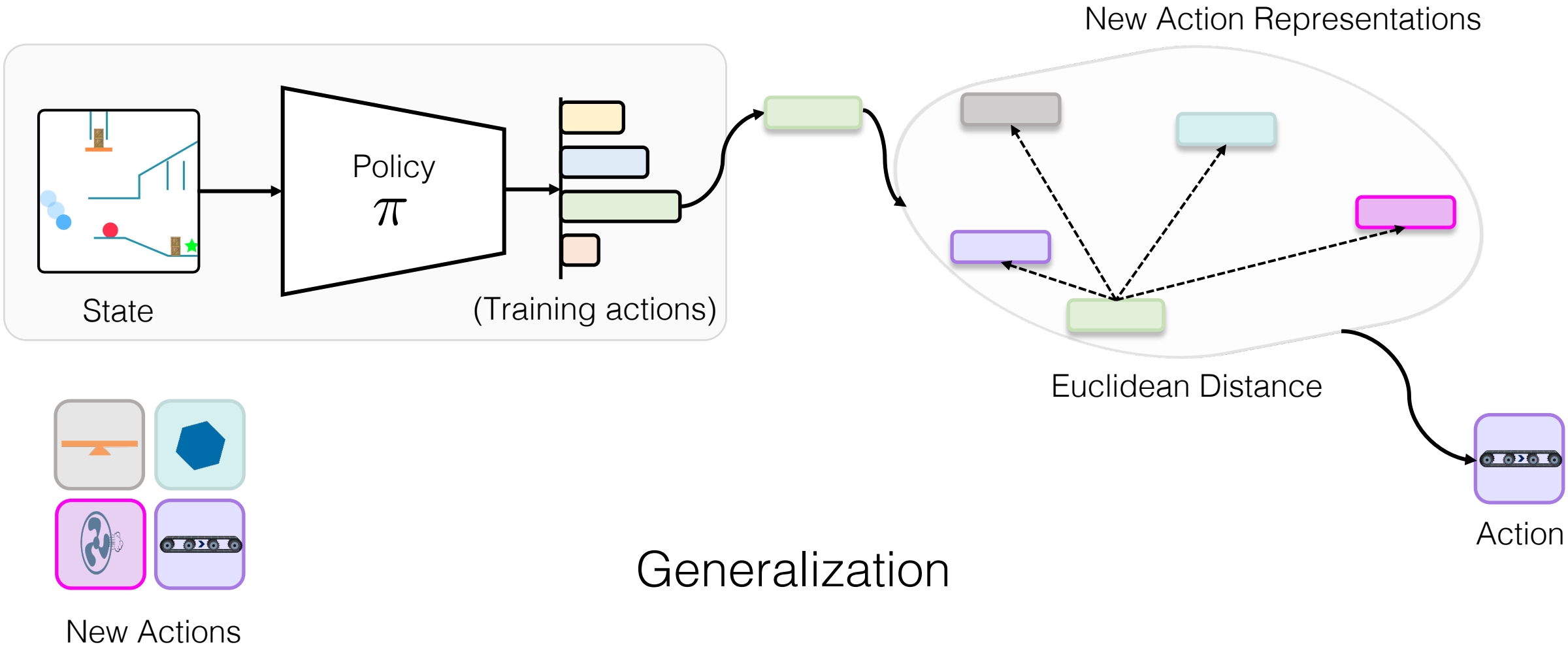


Nearest Neighbor Policy Baseline

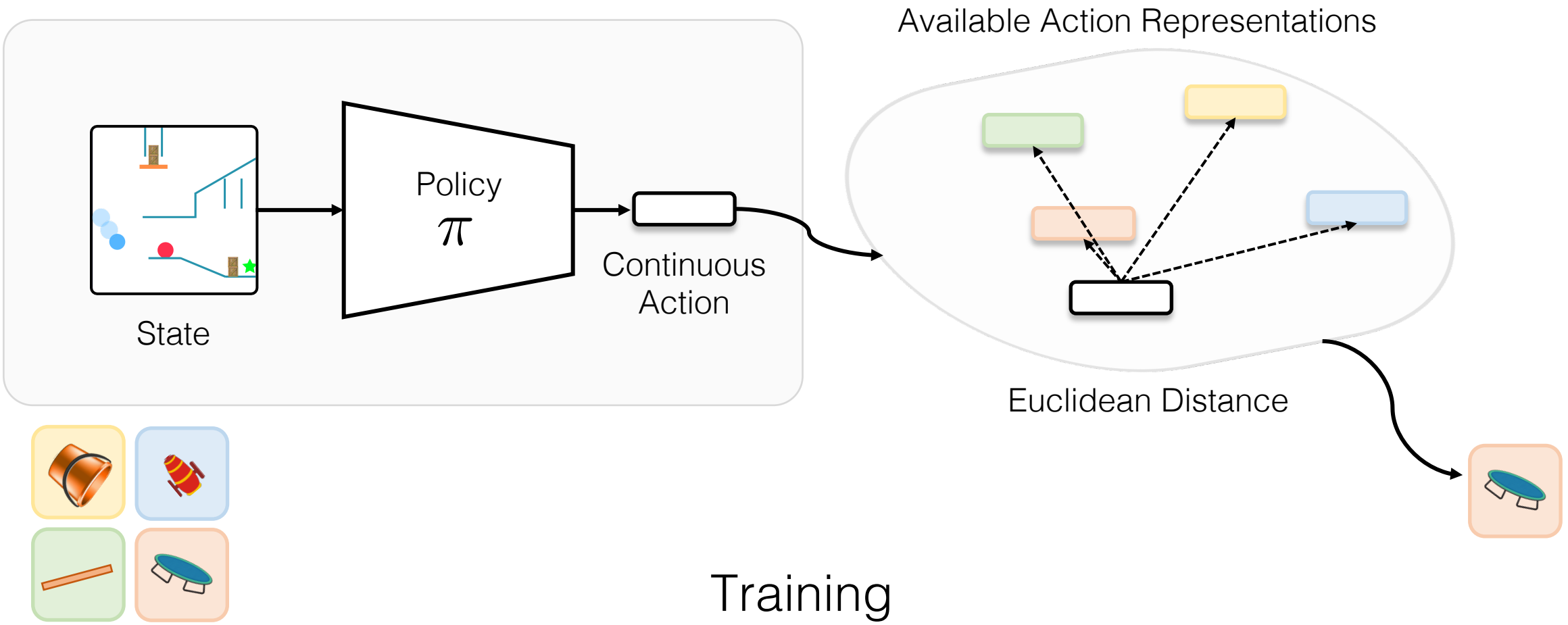


Training

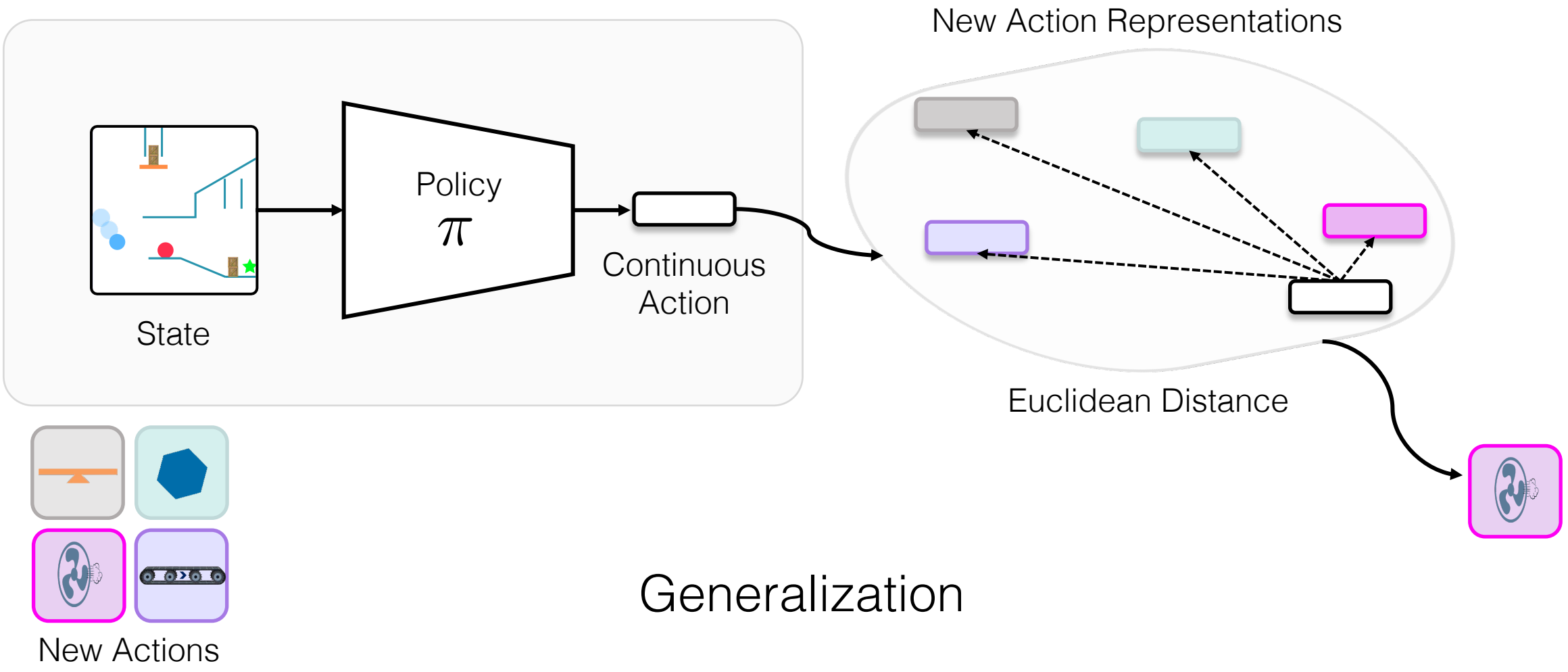
Nearest Neighbor Policy Baseline



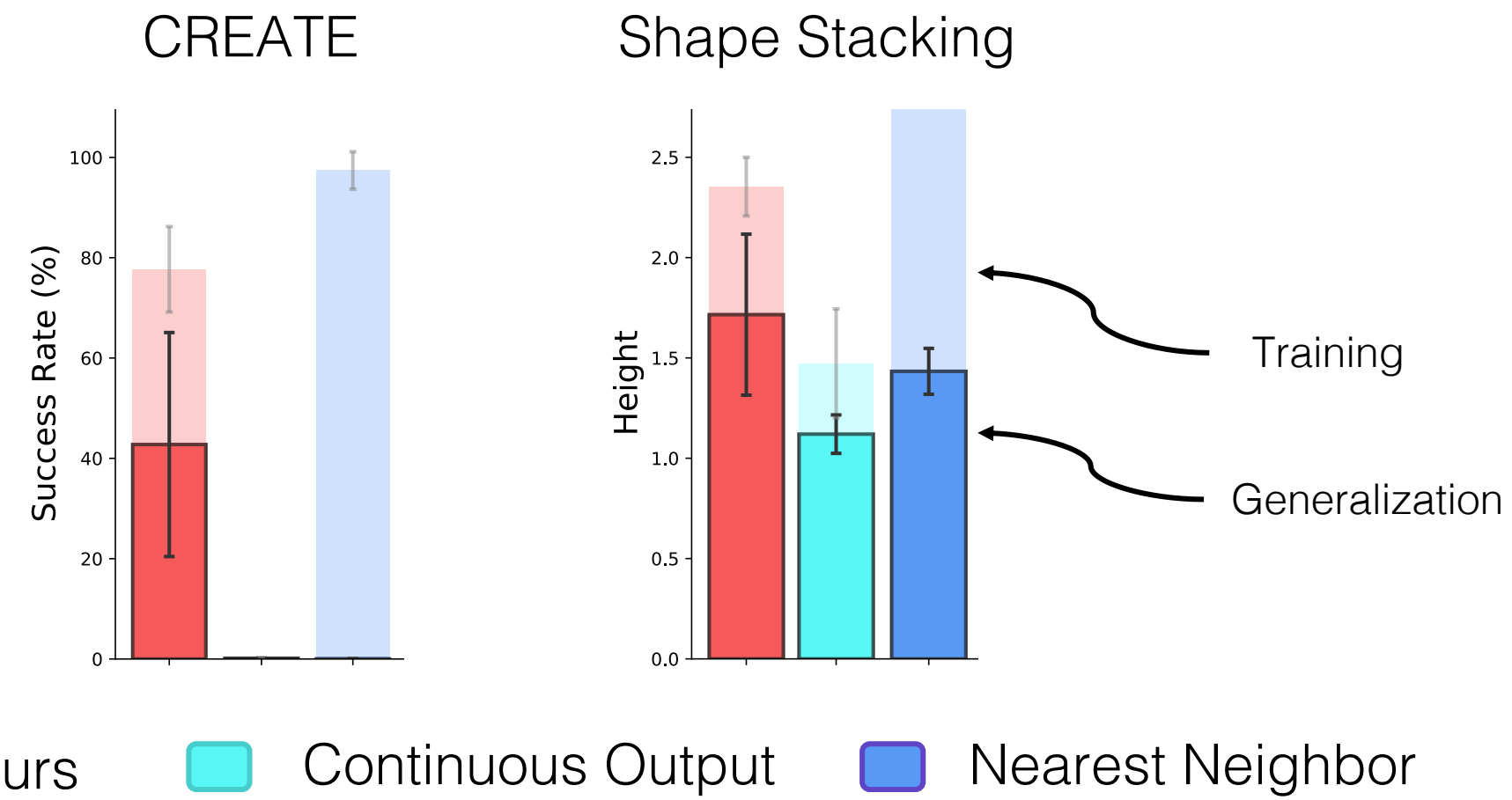
Continuous Output Policy Baseline



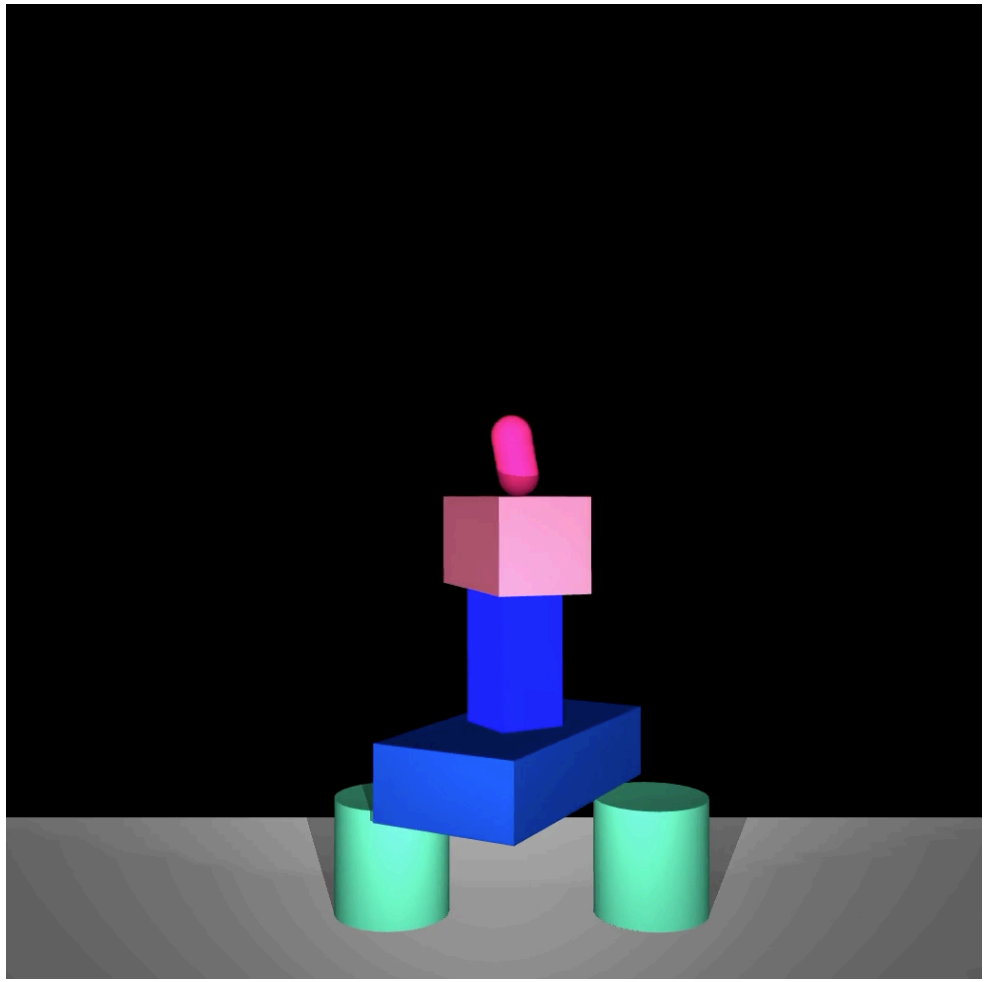
Continuous Output Policy Baseline



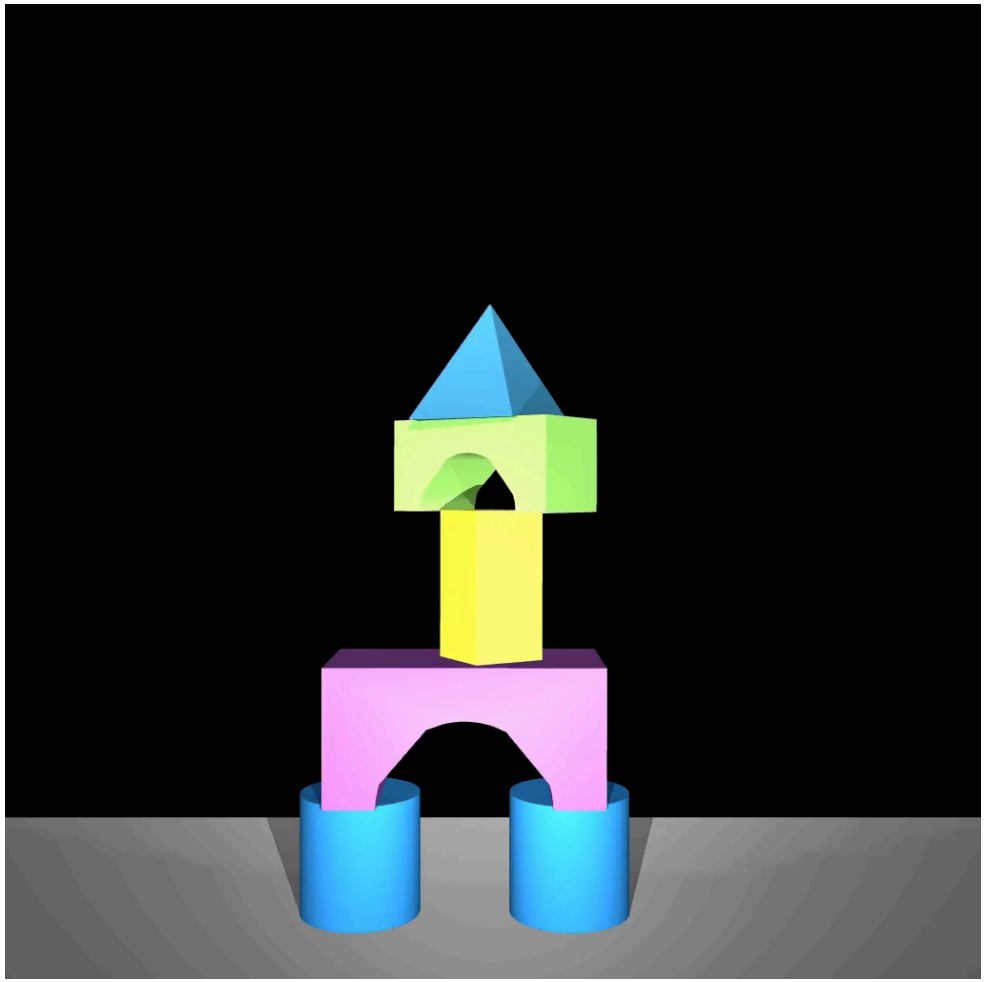
Learning task utility helps generalization



Qualitative Results: Shape Stacking

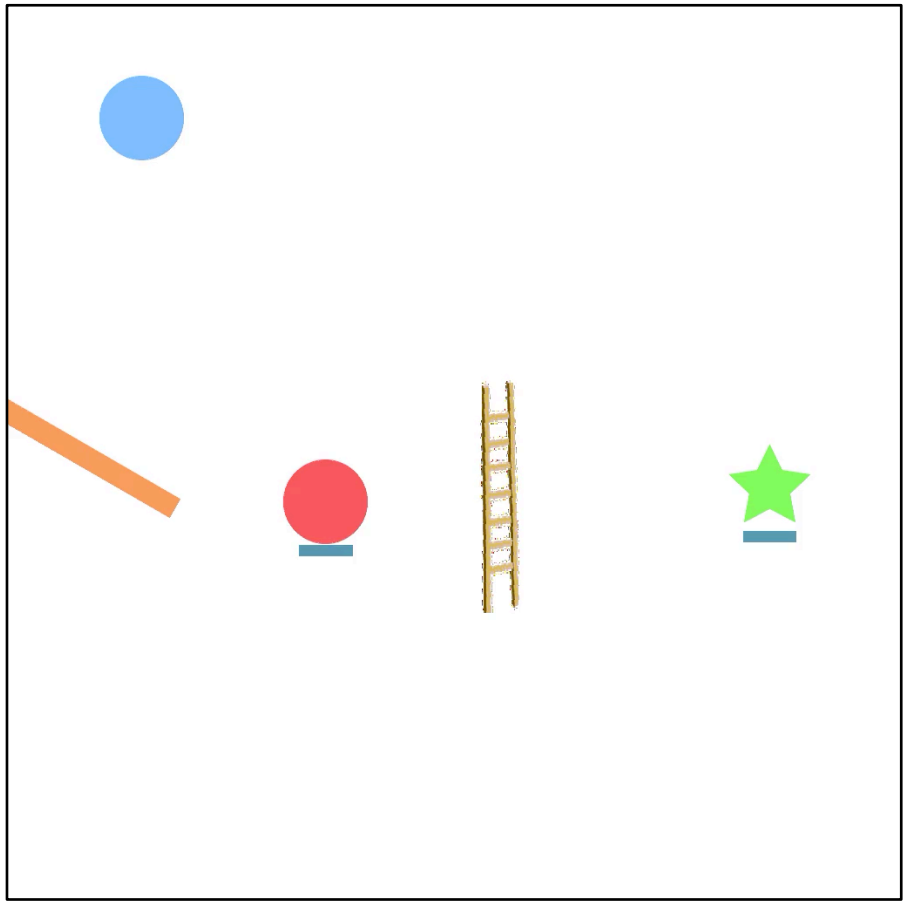


Training

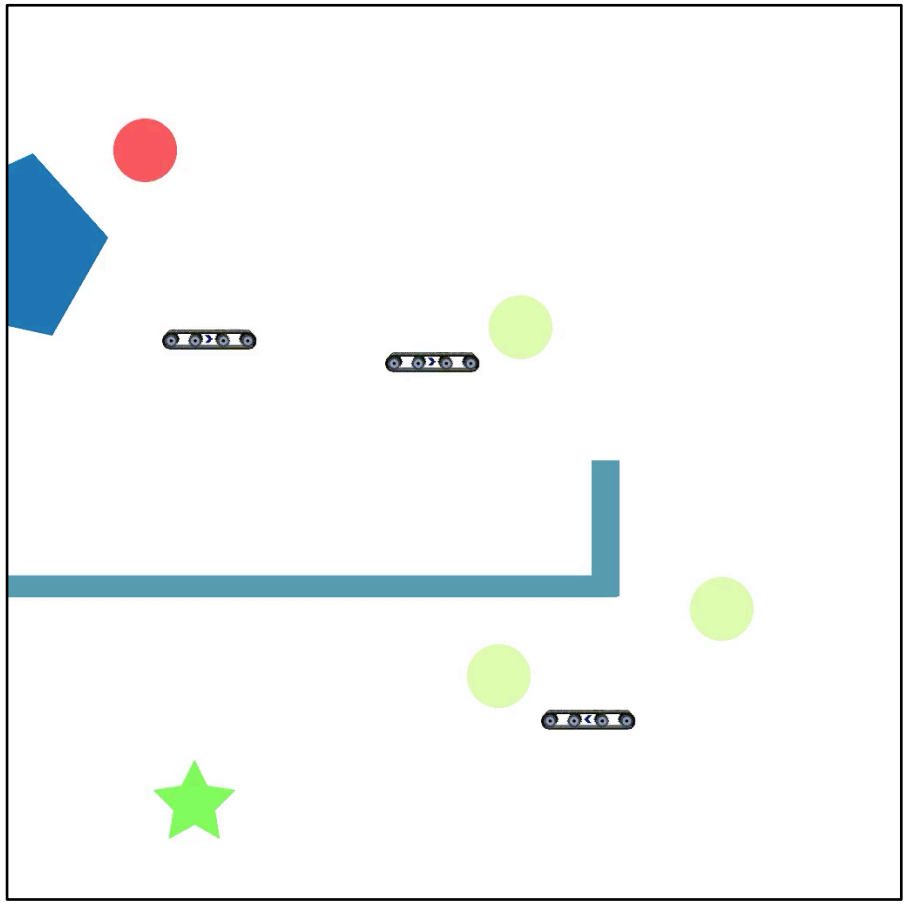


Generalization

Qualitative Results: CREATE

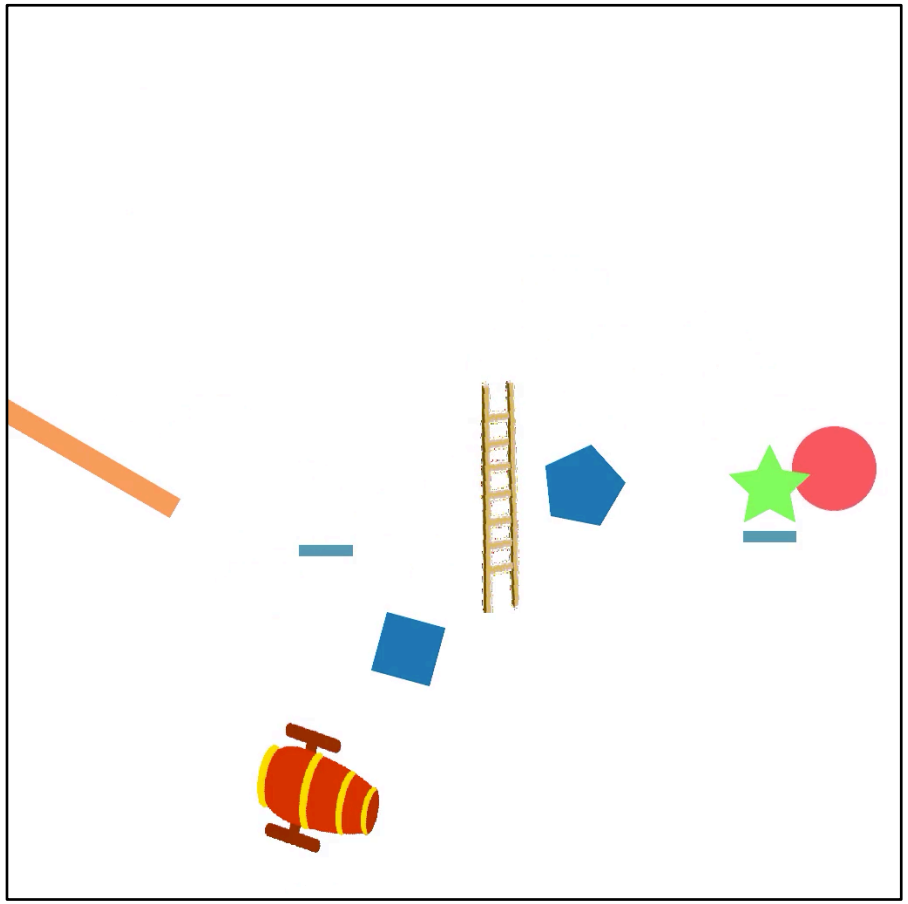


Ladder

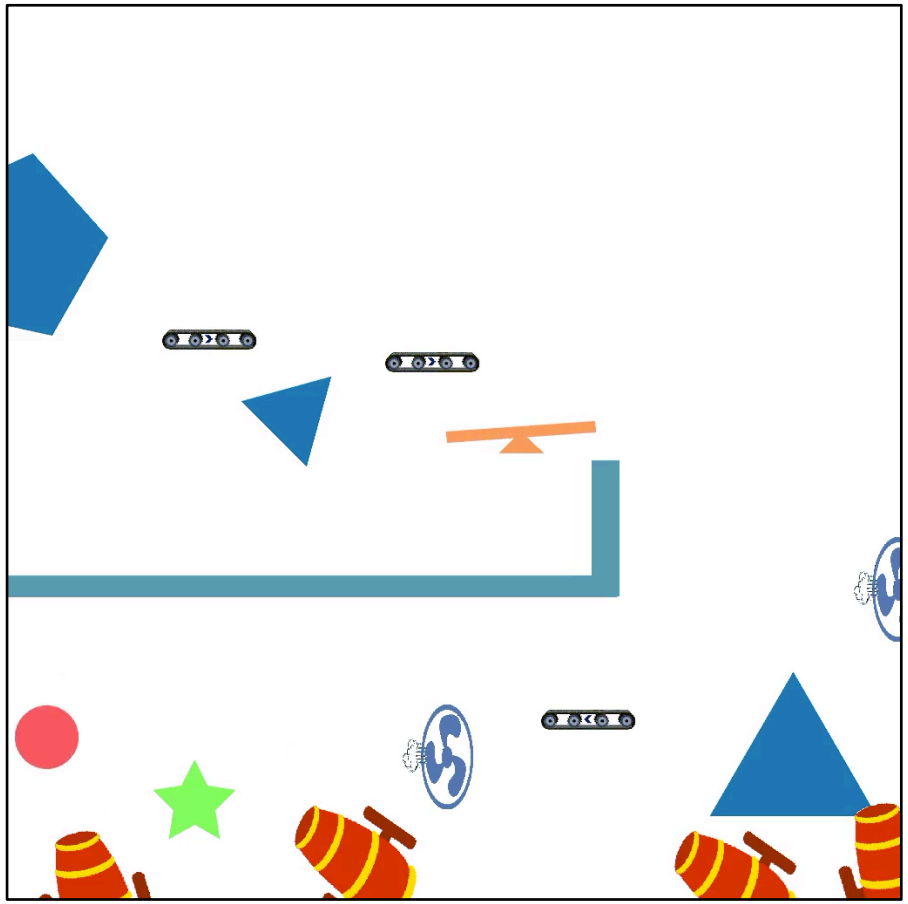


Belt

Qualitative Results: CREATE



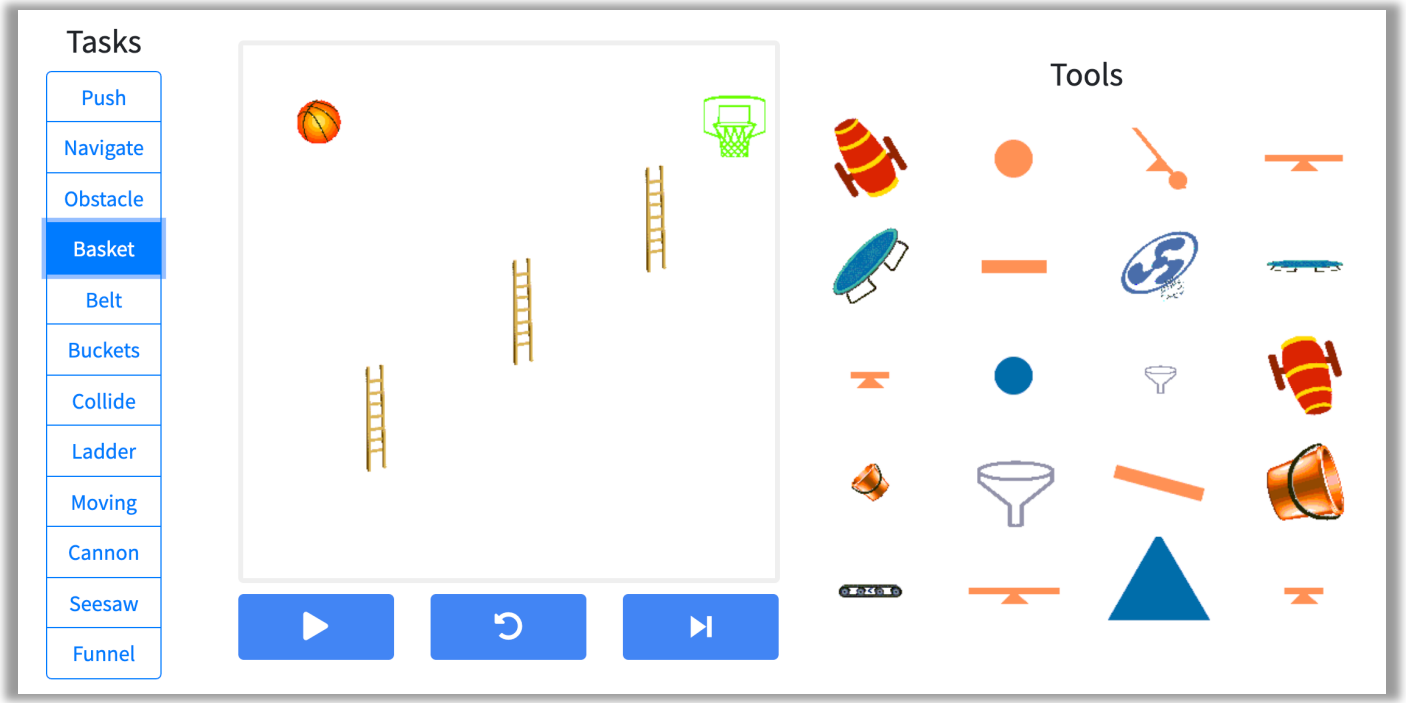
Ladder



Belt

Takeaways

Solving tasks using new action choices without retraining!



Explore CREATE, Results & Code at clvrai.com/create