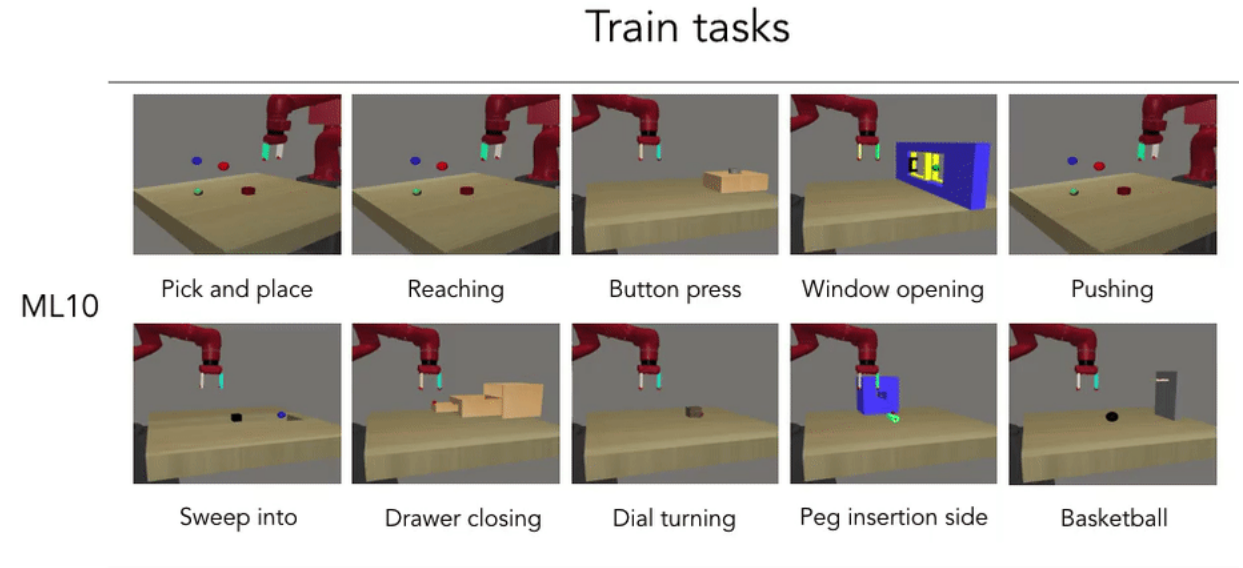


Unsupervised Curricula for Visual Meta-Reinforcement Learning

Allan Jabri, Kyle Hsu, Ben Eysenbach,
Abhishek Gupta, Sergey Levine, Chelsea Finn

NeurIPS 2019

From Specialist to Generalist



Source: Meta-World
meta-world.github.io

Multi-task Reinforcement Learning

Contextual Policies

$$\pi(a|o, z)$$

Task description is given

e.g. a goal


more general

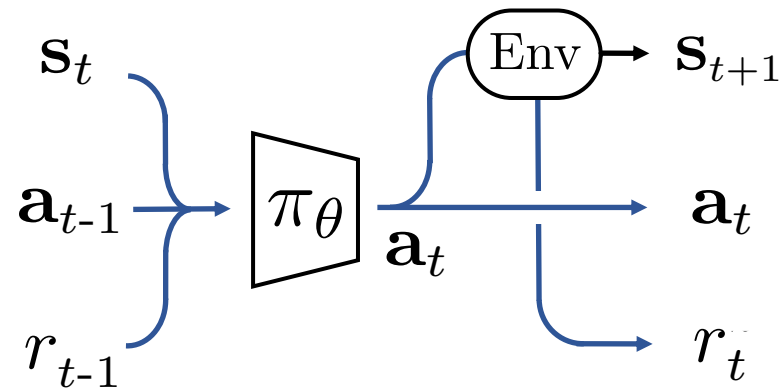
Meta-learning for RL

$$\pi(a|o, \mathcal{D}_{\mathcal{T}})$$

Task inferred from data

collected by policy

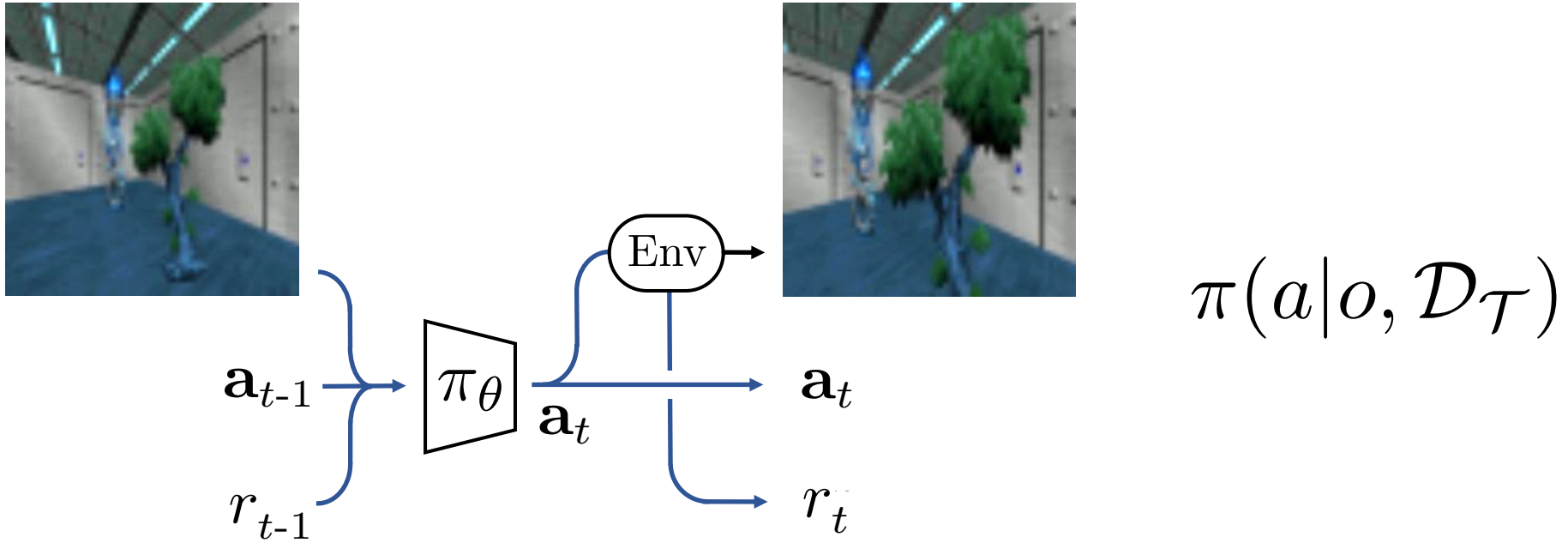
Meta-Reinforcement-Learning



$$\pi(a|o, \mathcal{D}_{\mathcal{T}})$$

Recurrent policy learns to **infer task** by collecting the right data

Visual Meta-Reinforcement-Learning

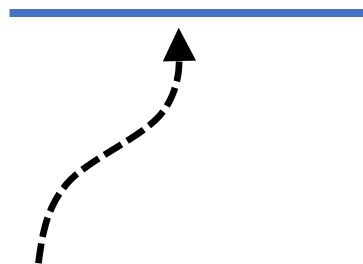


Search for and associate **stimulus** and **reward**.

The Task Distribution

$$\arg \max_{\theta} \sum_{i=1}^n \mathbb{E}_{\pi_{\theta}(\mathcal{D}_{\mathcal{M}_i})} [R(\tau)]$$

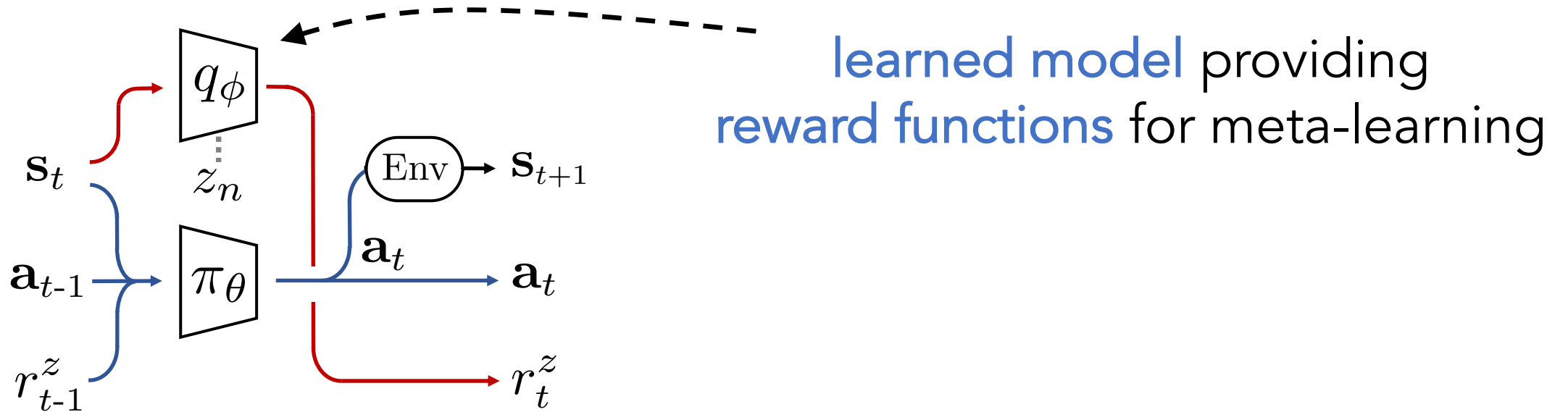
where $\mathcal{M}_i \sim p(\mathcal{M})$



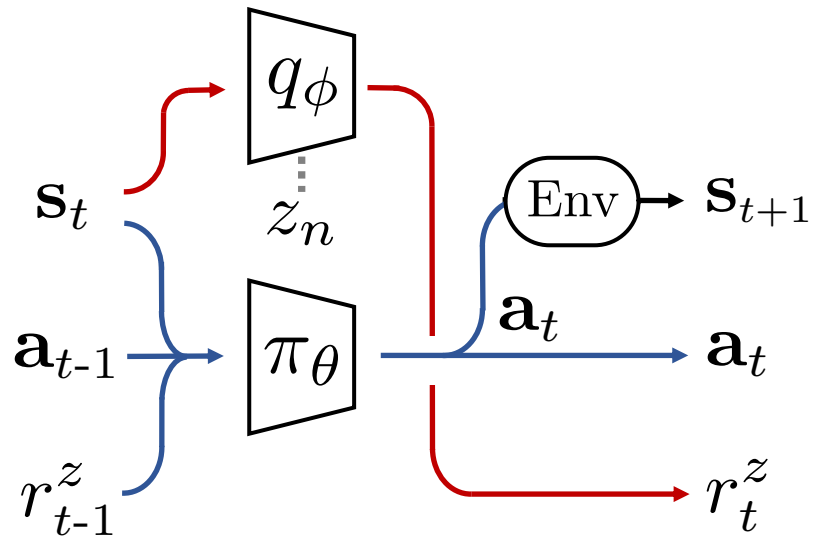
Meta-training tasks give rise to
task inference and execution strategies

Can we learn **useful** meta-RL strategies
with tasks formed **without supervision**?

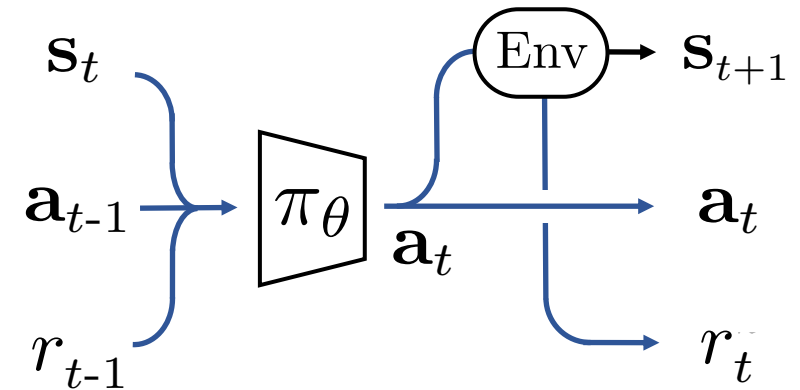
"Meta-Pre-training"



"Meta-Pre-training"



Unsupervised Pre-training



Transfer to Test Tasks

Task Acquisition

Unsupervised discovery of tasks



Meta-learning

Learn to learn to solve tasks

Task Acquisition

Unsupervised discovery of tasks



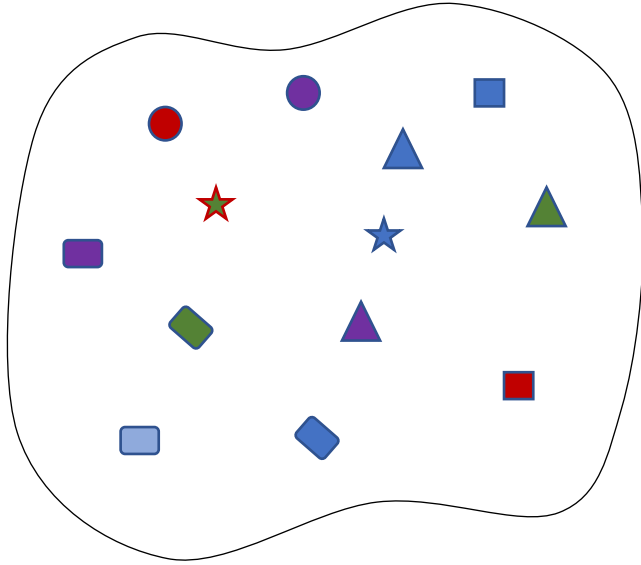
Meta-learning

Learn to learn to solve tasks

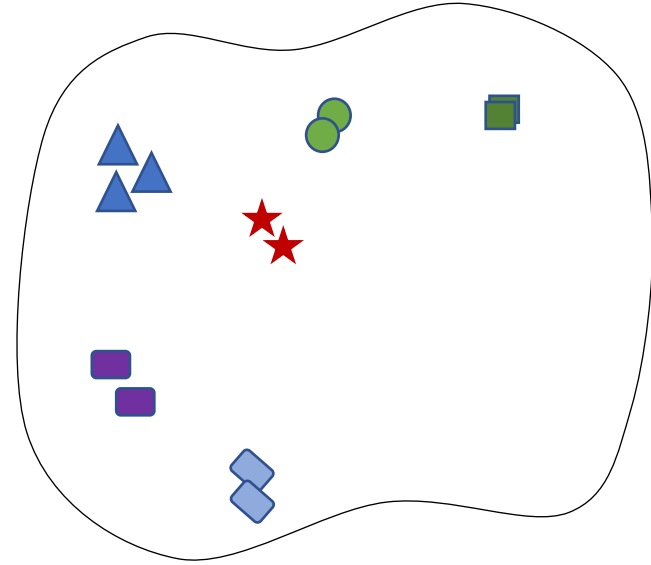


Should co-adapt

Criteria for Task Distribution

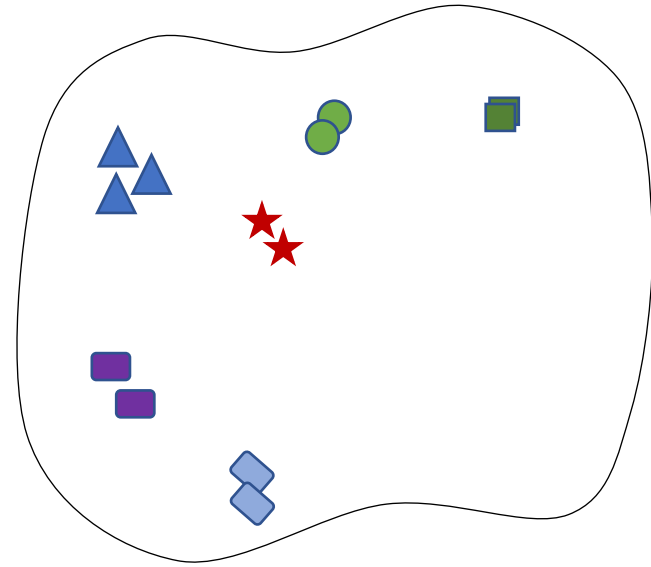
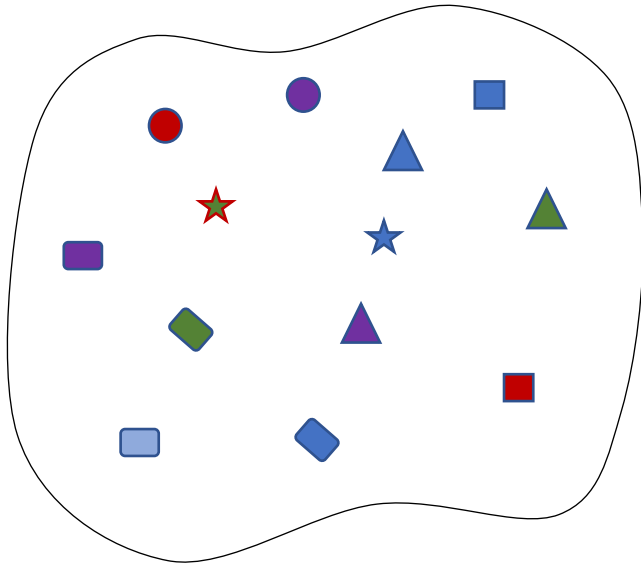


Diversity



Structure

Criteria for Task Distribution



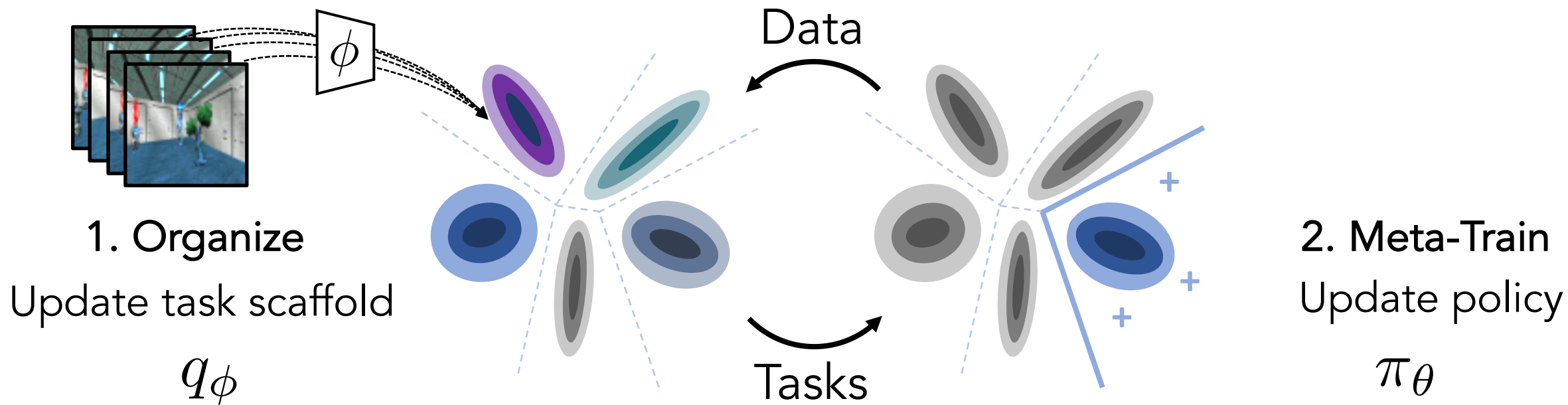
Diversity $H(\boldsymbol{\tau}) - H(\boldsymbol{\tau}|\mathbf{z})$ Structure

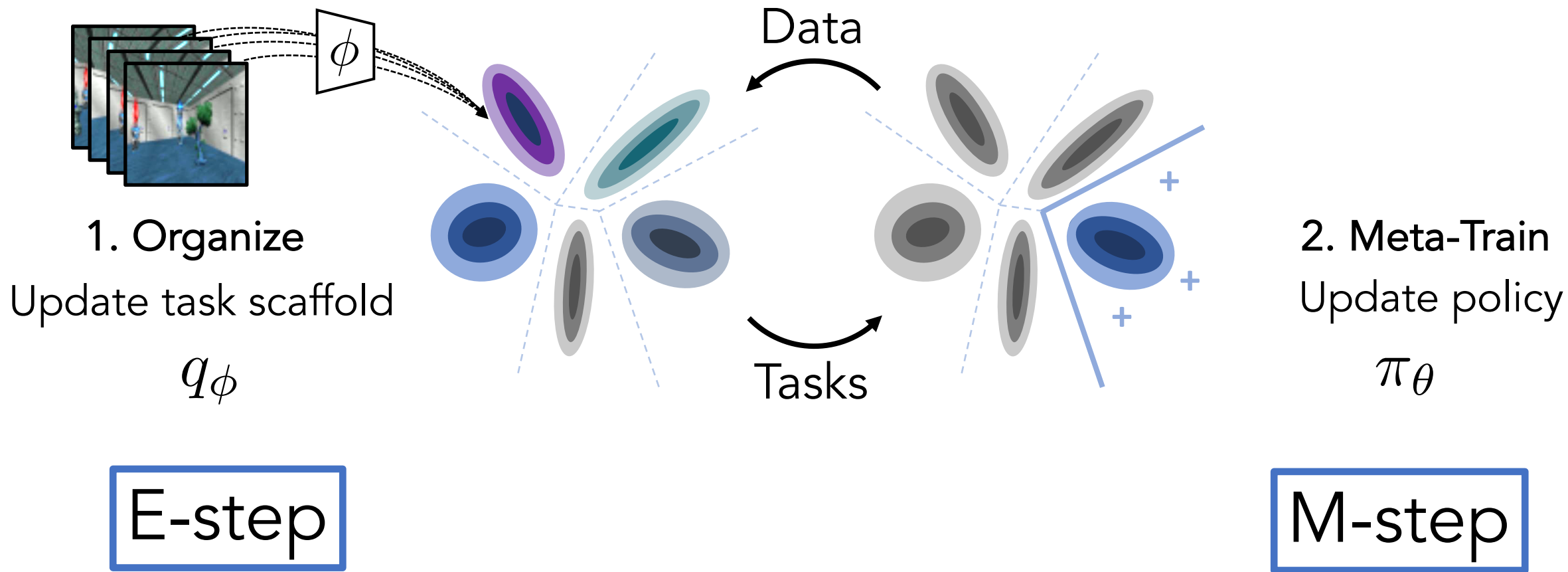
$$= I(\boldsymbol{\tau}; \mathbf{z})$$

Formulation

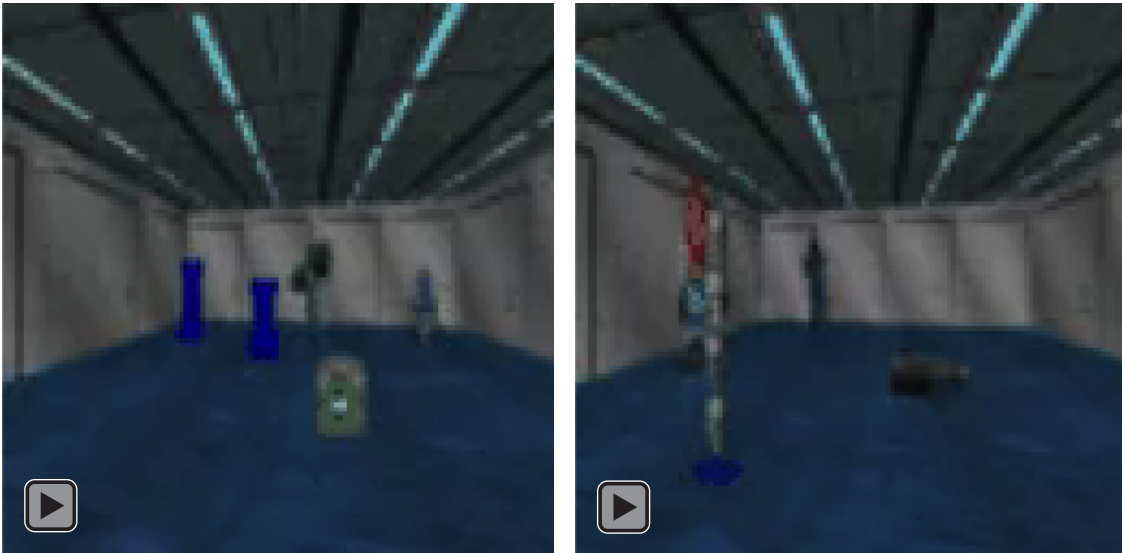
$$\max_{\theta, \phi} I(\boldsymbol{\tau}; \mathbf{z})$$

Policy	π_{θ}	$\boldsymbol{\tau}$	Post-update trajectories
Task scaffold	q_{ϕ}	\mathbf{z}	Task latent variable

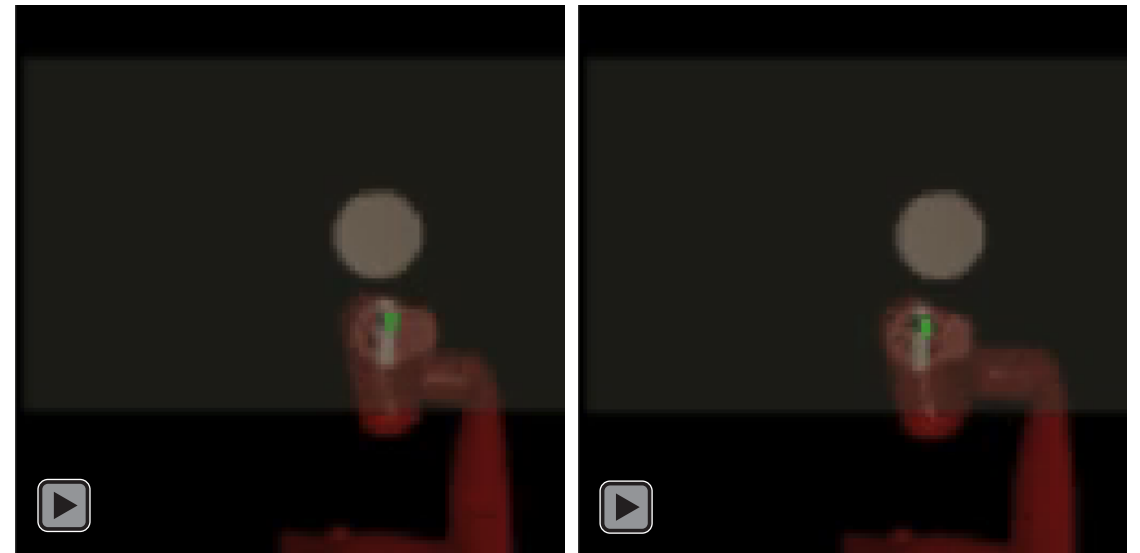




Experimental Setting

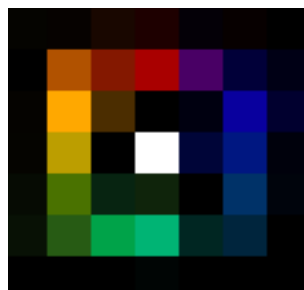
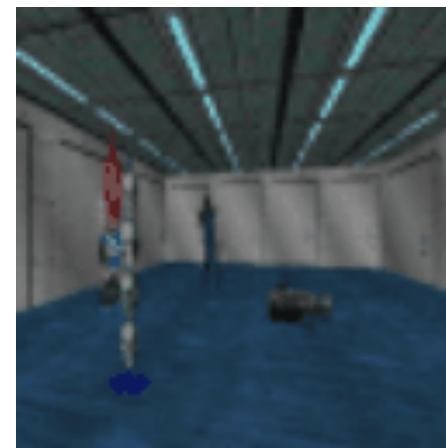
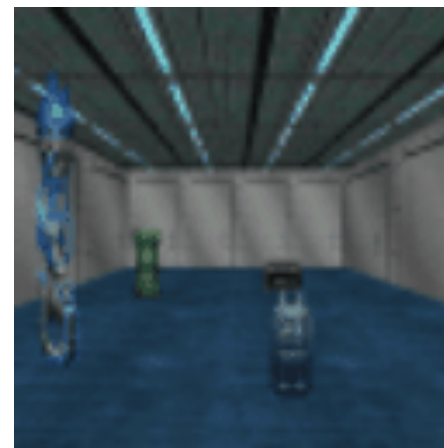
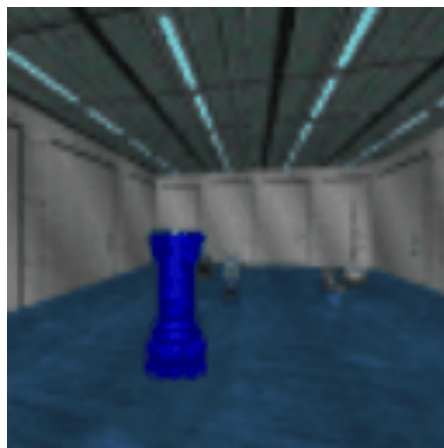
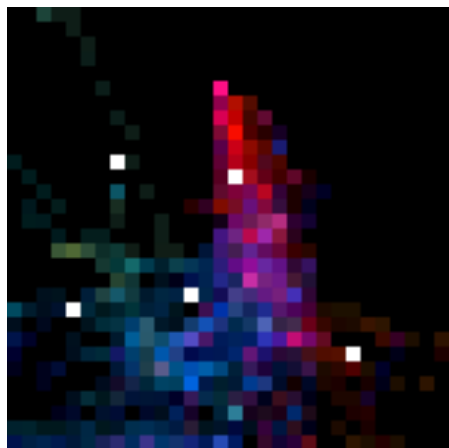


Visual Navigation
in VizDoom



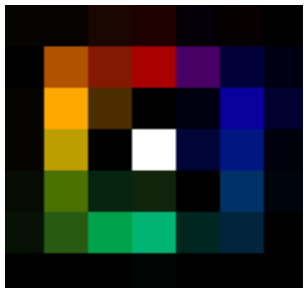
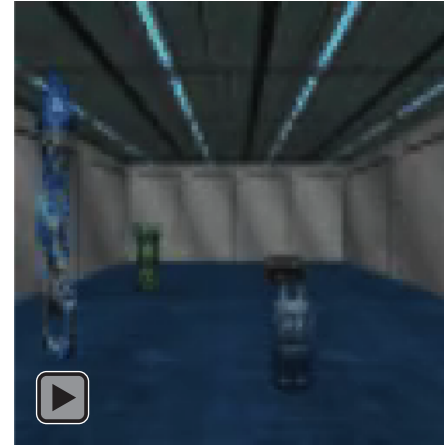
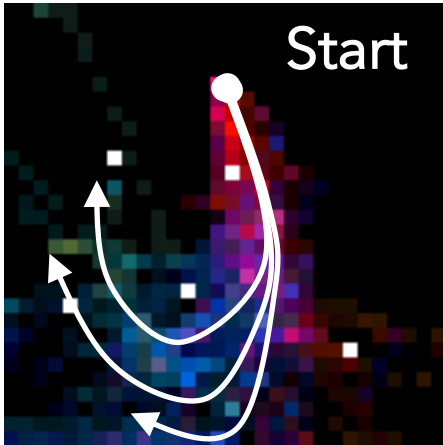
Object Pushing
with Sawyer in MuJoCo

What kind of tasks are discovered?



Direction encoded as color

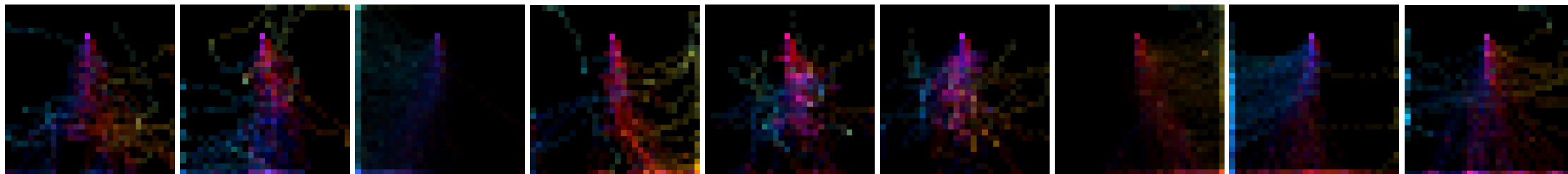
What kind of tasks are discovered?



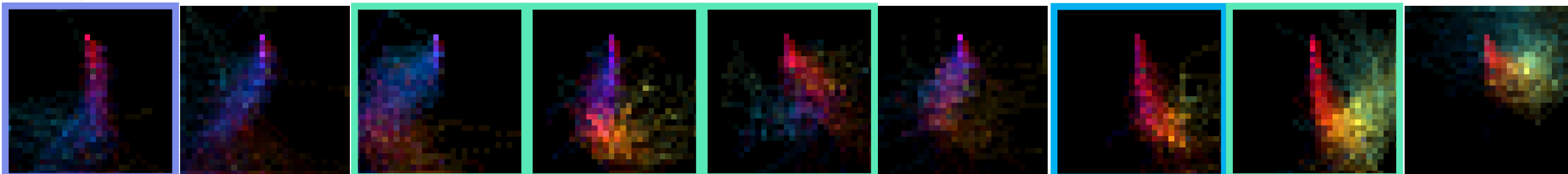
Direction encoded as color

What kind of tasks are discovered?

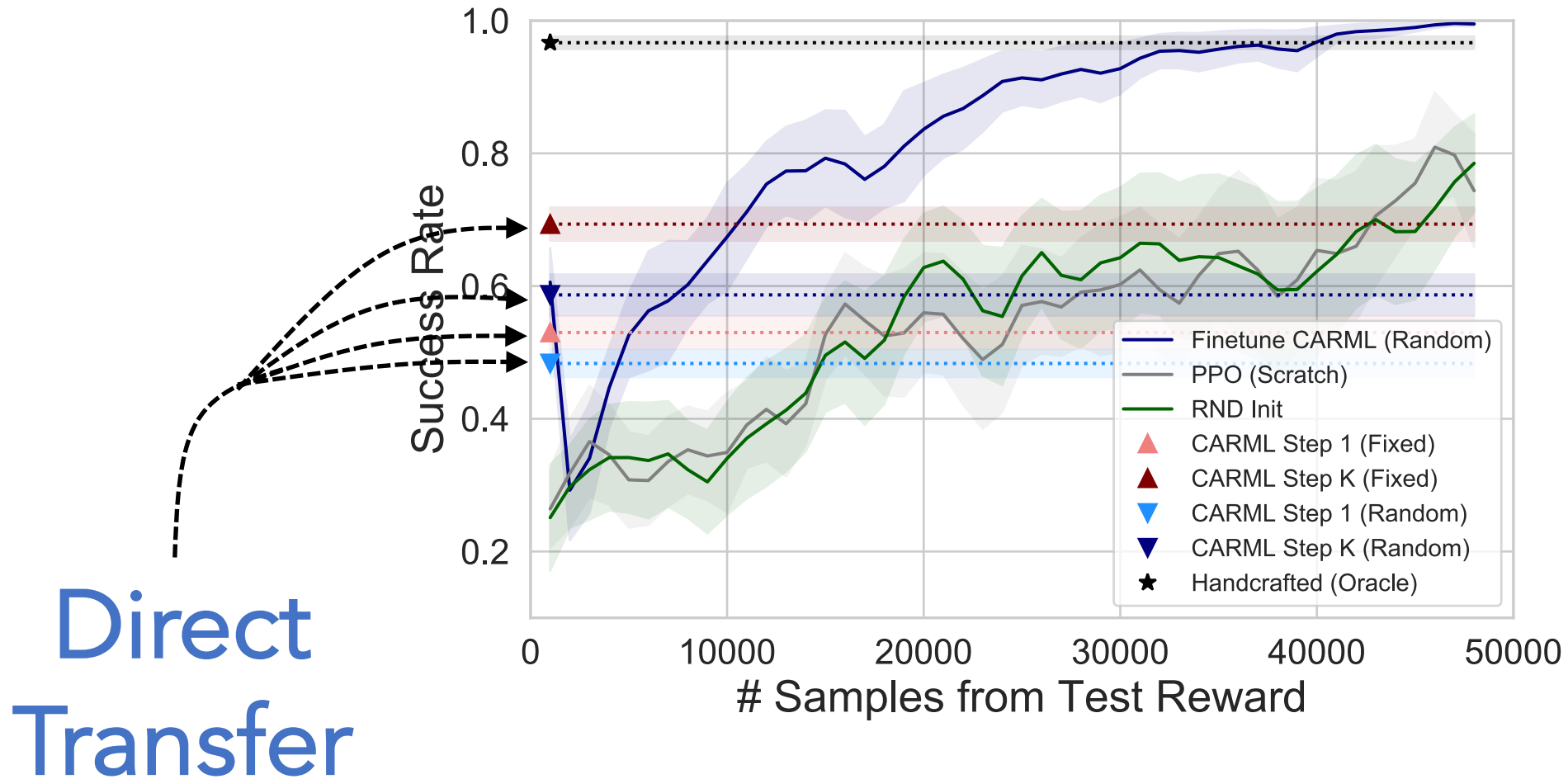
Step 1



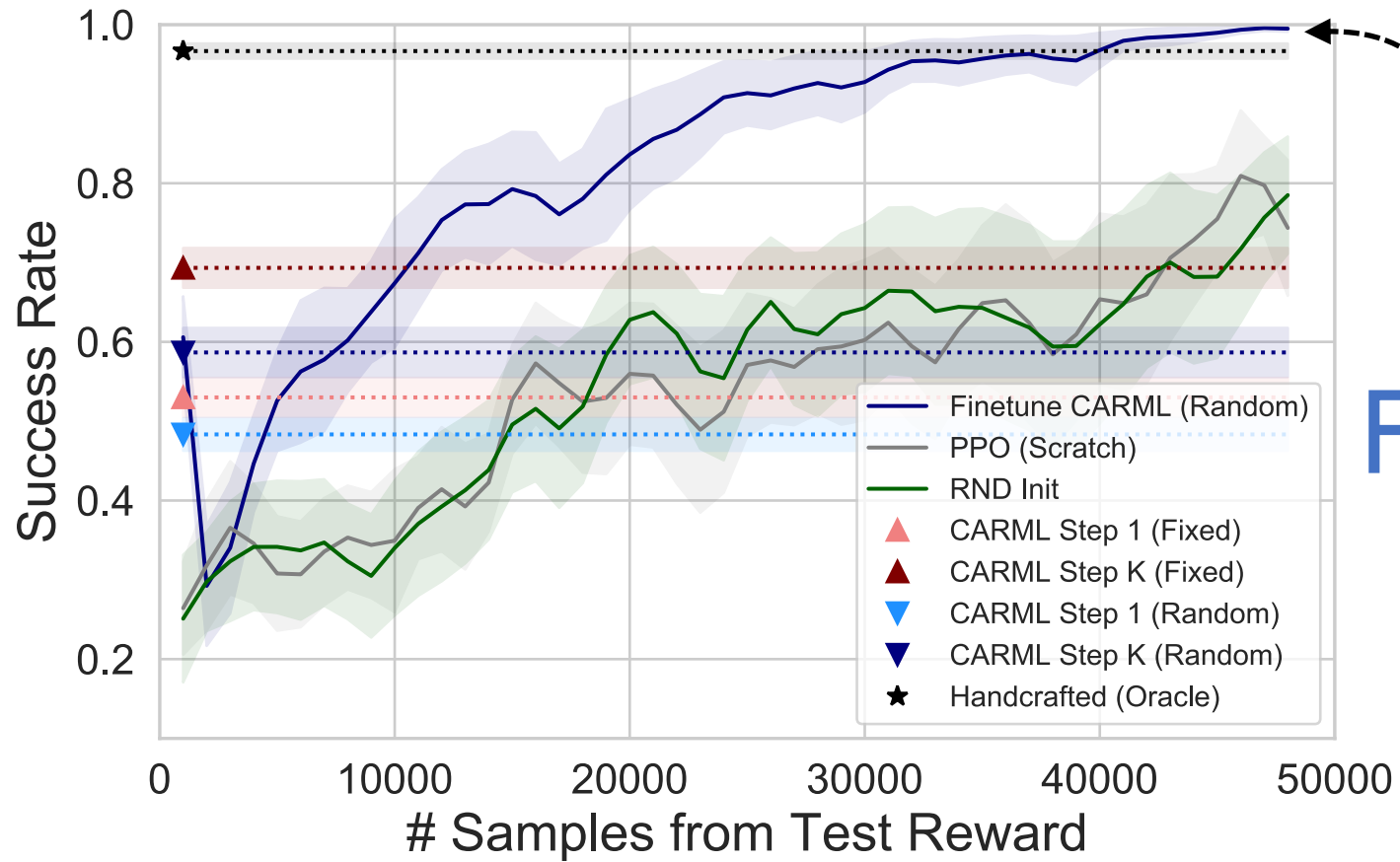
Step 5



Transfer to Test Tasks – VizDoom

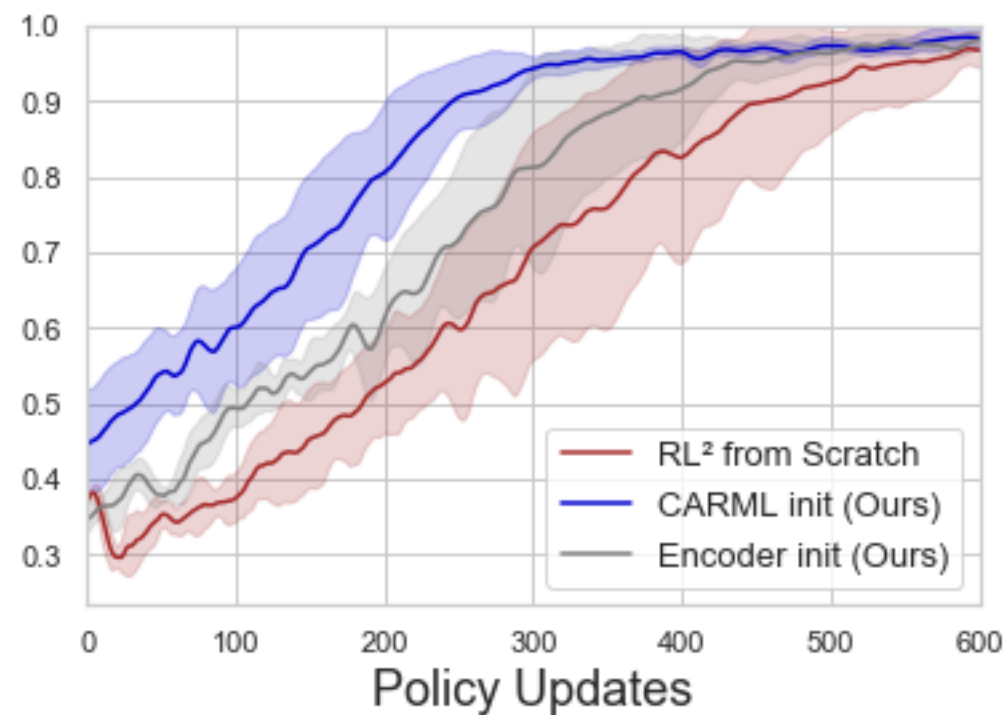
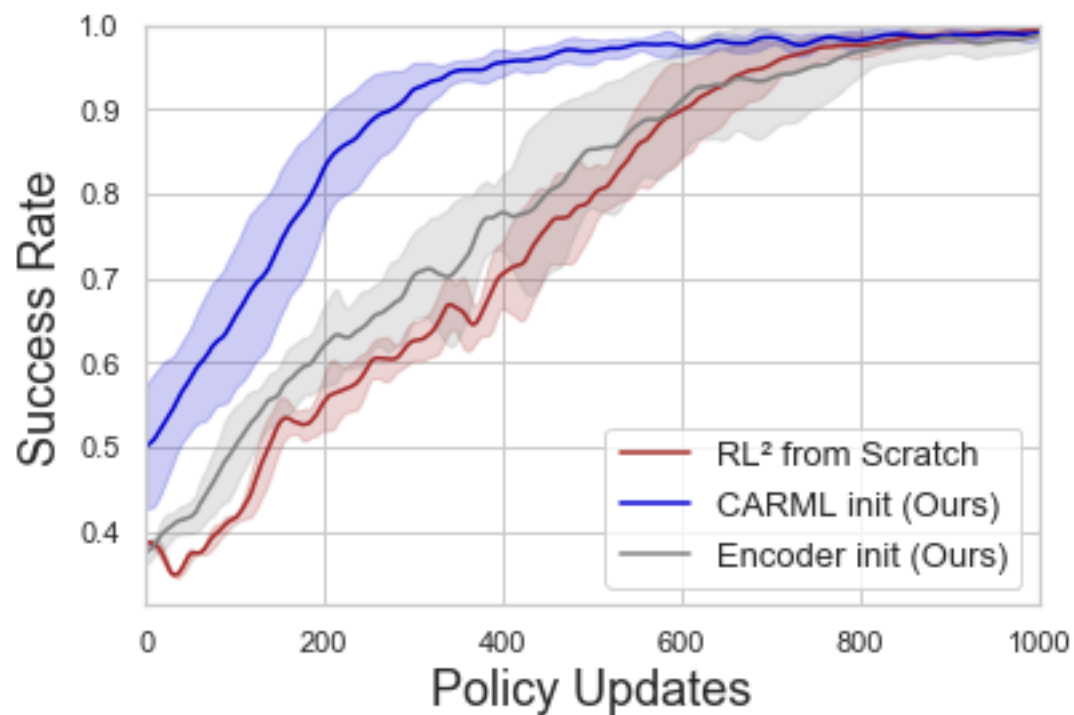


Transfer to Test Tasks – VizDoom



**Faster
Finetuning**

Faster Supervised Meta-RL



Thank You



Kyle Hsu



Ben Eysenbach



Abhishek Gupta



Sergey Levine



Chelsea Finn

Poster #35, East Exhibition Hall B + C

<https://sites.google.com/view/carm1>