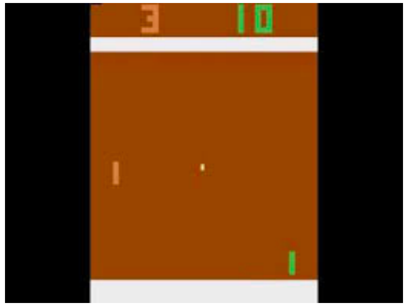
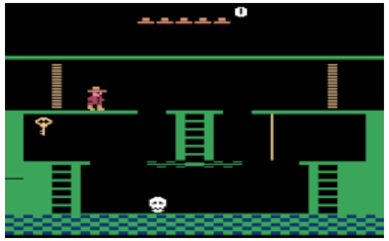
Unsupervised Curricula for Visual Meta-Reinforcement Learning

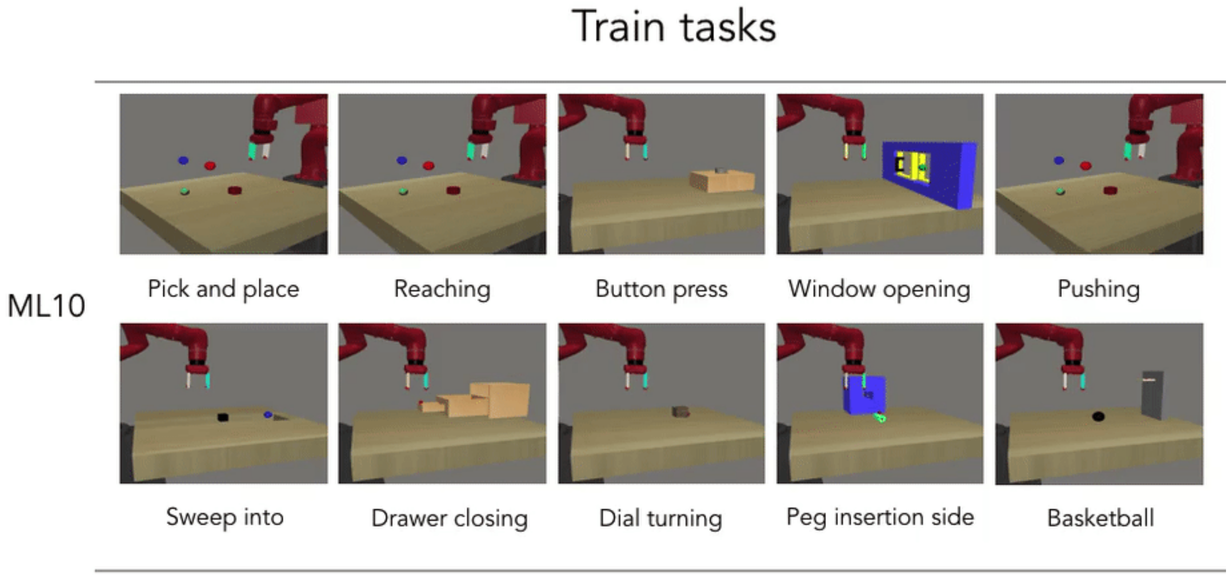
Allan Jabri, Kyle Hsu, Ben Eysenbach,
Abhishek Gupta, Sergey Levine, Chelsea Finn

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

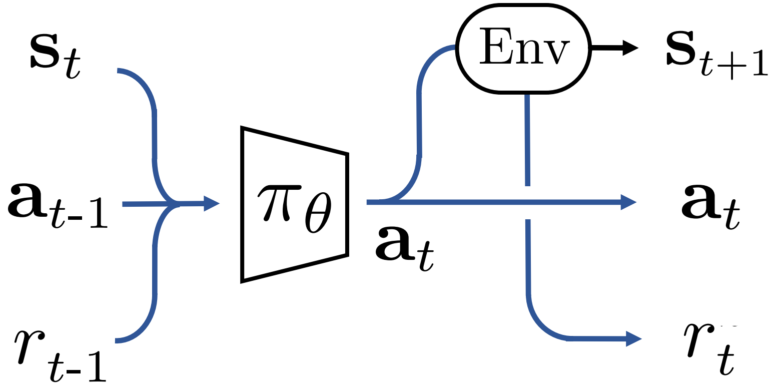
Meta-learning for RL

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

Task inferred from data

collected by policy

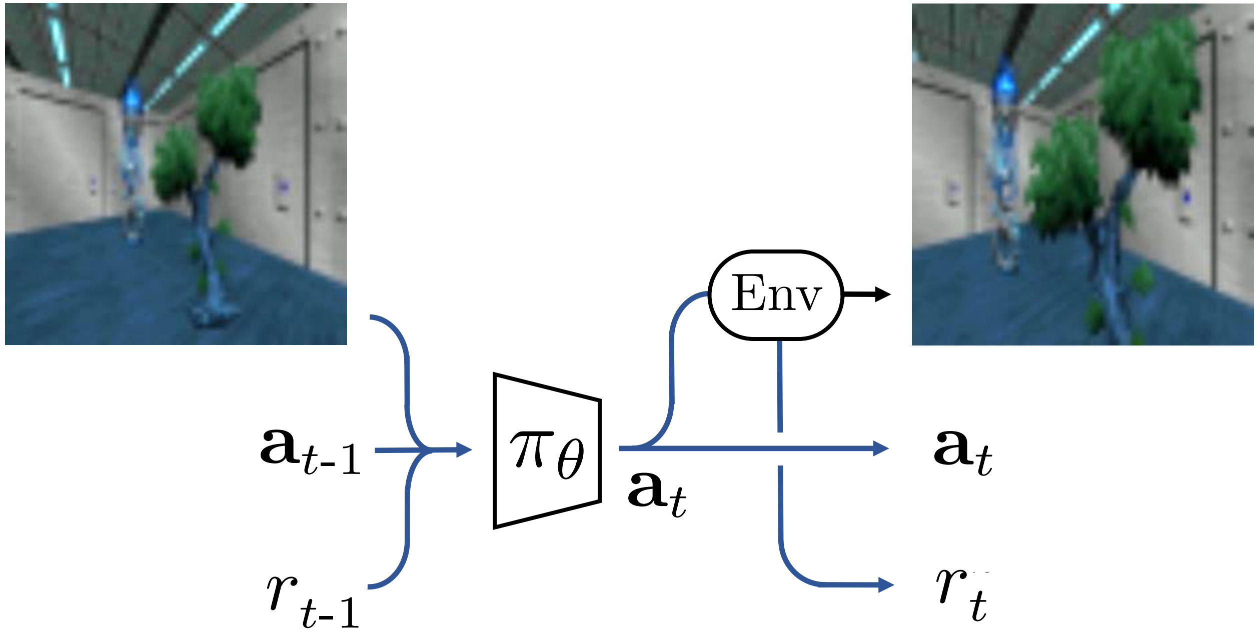
Meta-Reinforcement-Learning



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

Recurrent policy learns to infer task by collecting the right data

Visual Meta-Reinforcement-Learning



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

Search for and associate stimulus and reward.

The Task Distribution

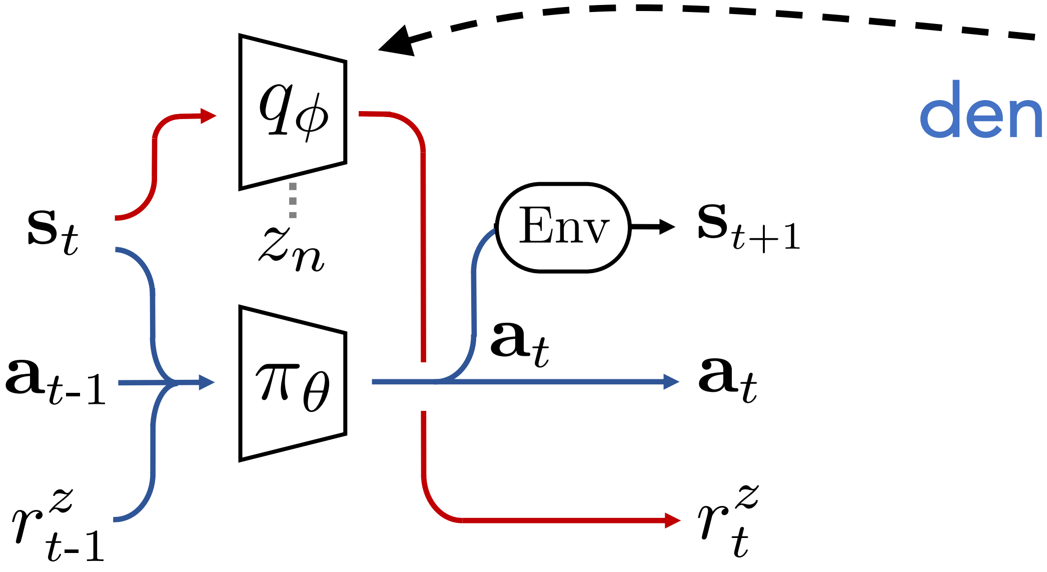
$$\arg\max_{\theta} \sum_{i=1}^{n} \mathbb{E}_{\pi_\theta(\mathcal{D}_{\mathcal{M}_i})}[R(\tau)]$$

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

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

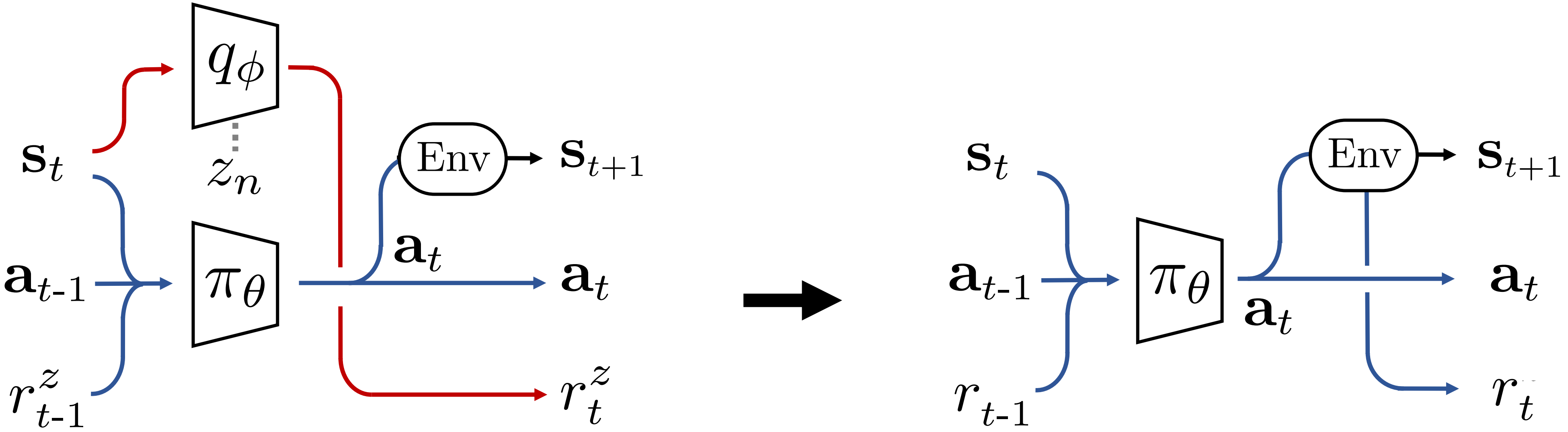
Can we learn useful meta-RL strategies with tasks formed without supervision?

“Meta-Pre-training”



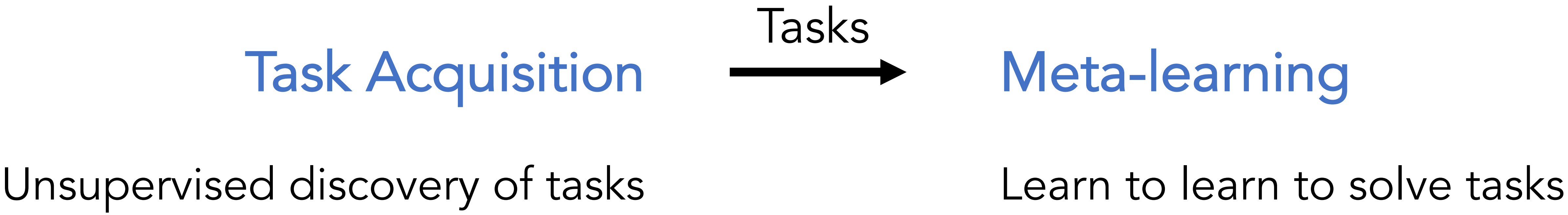
density model of trajectories providing reward functions for meta-RL

"Meta-Pre-training"



Unsupervised Pre-training

Transfer to Test Tasks



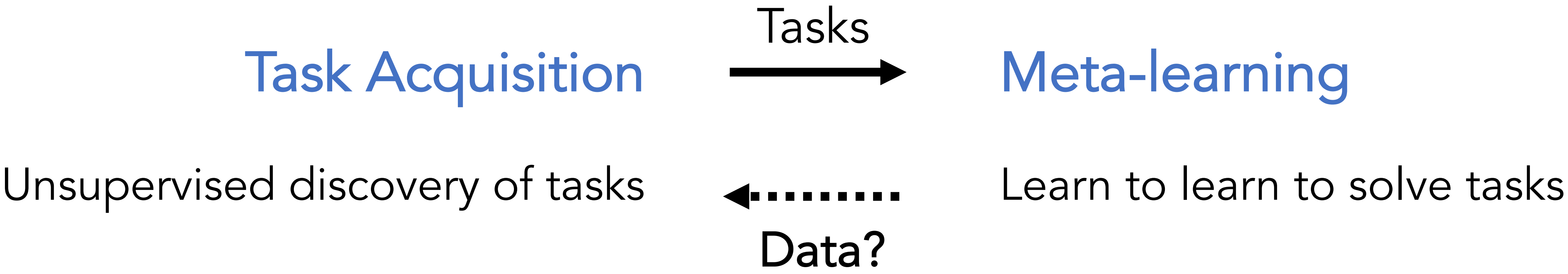
Task Acquisition

Unsupervised discovery of tasks

Tasks

Meta-learning

Learn to learn to solve tasks



Task Acquisition

Tasks

Meta-learning

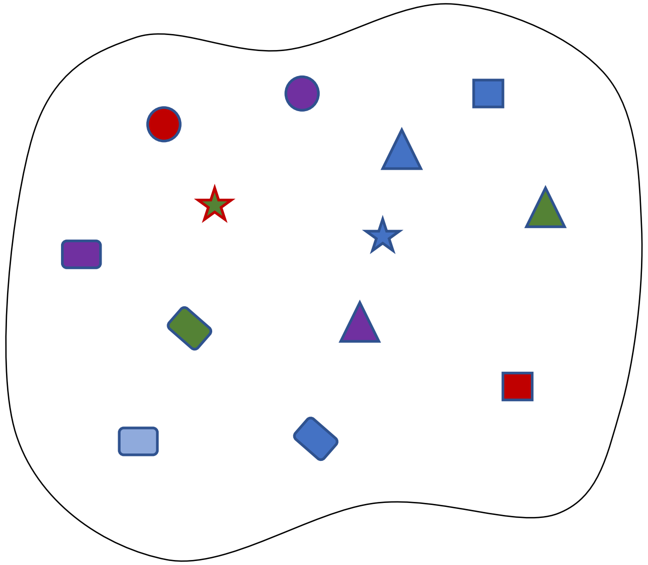
Unsupervised discovery of tasks

Learn to learn to solve tasks

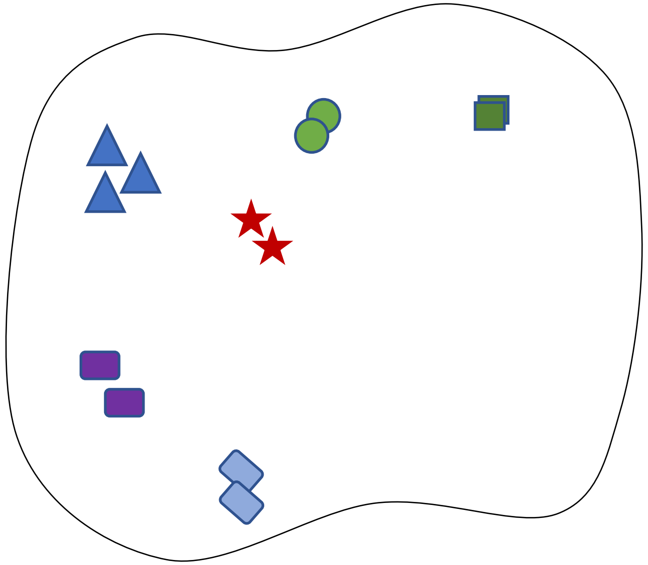
Data?

Should co-adapt

Criteria for Task Distribution

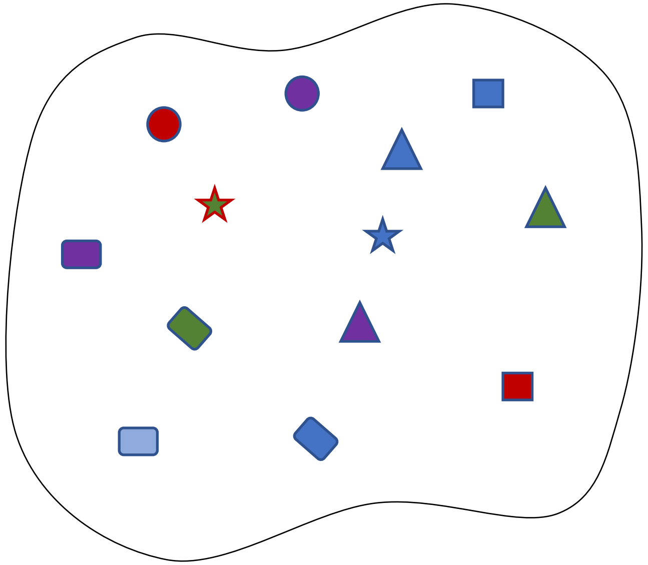


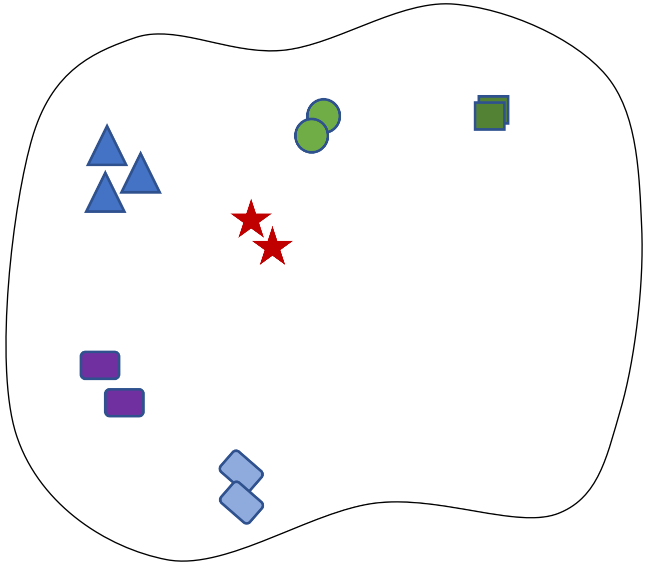
Diversity



Structure

Criteria for Task Distribution





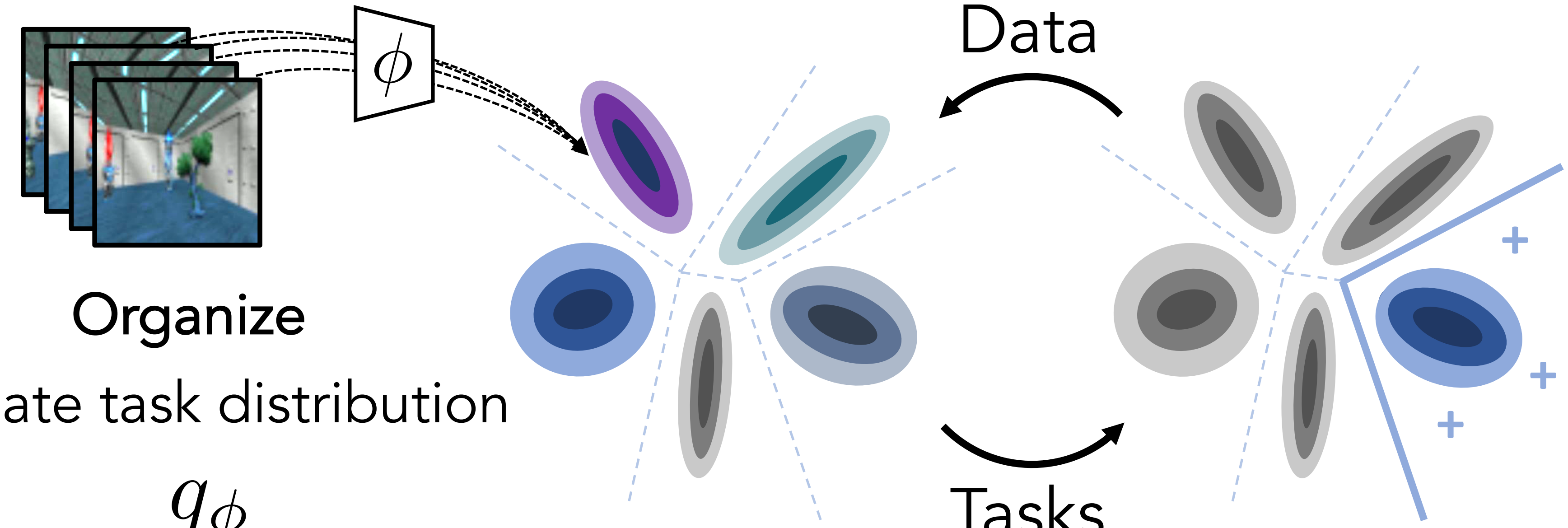
Diversity $H(\boldsymbol{\tau})$ $-H(\boldsymbol{\tau}|\mathbf{z})$ Structure

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

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



Data

Tasks

Organize

Update task distribution

q_ϕ

Meta-Train

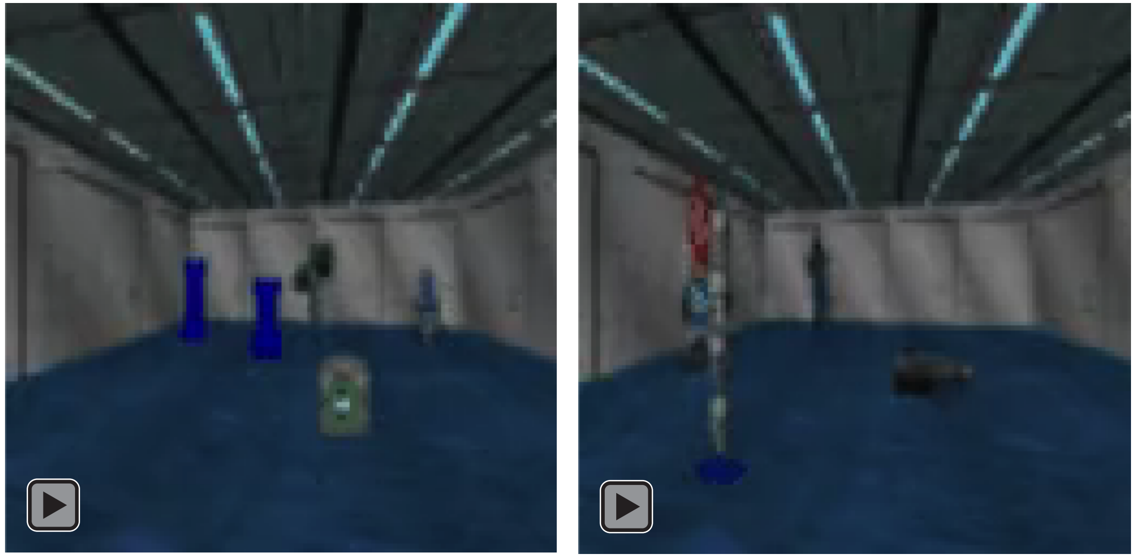
Update policy

π_θ

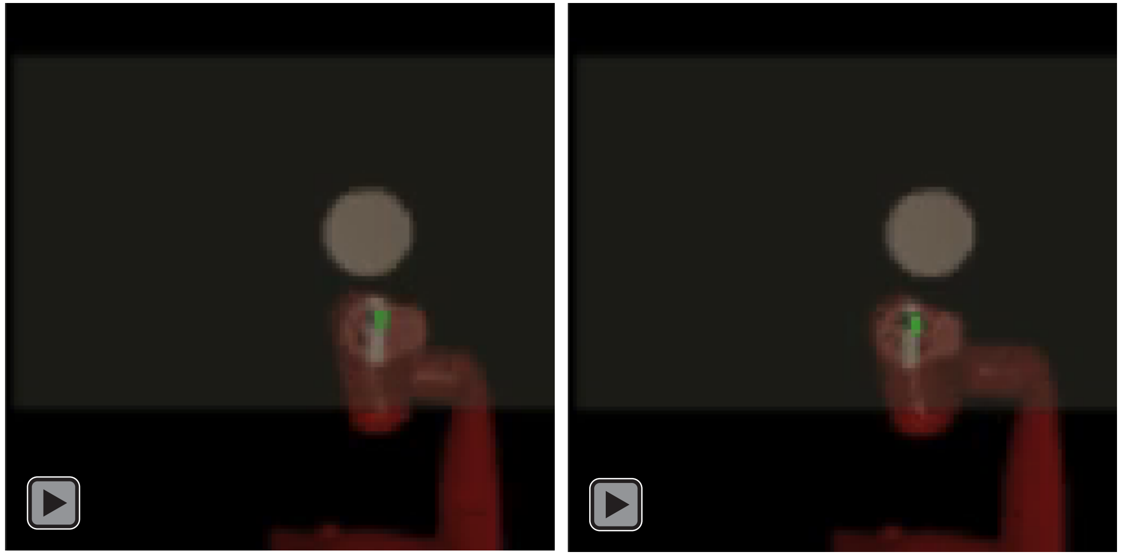
E-step

M-step

Experimental Setting

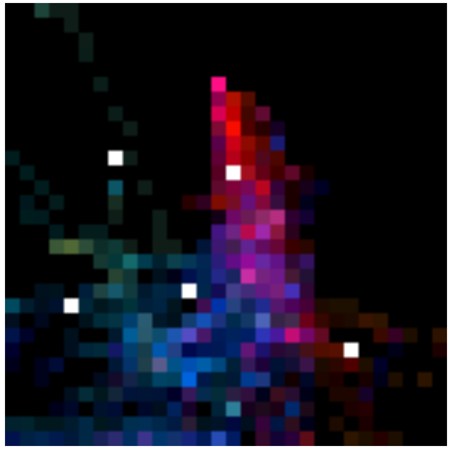


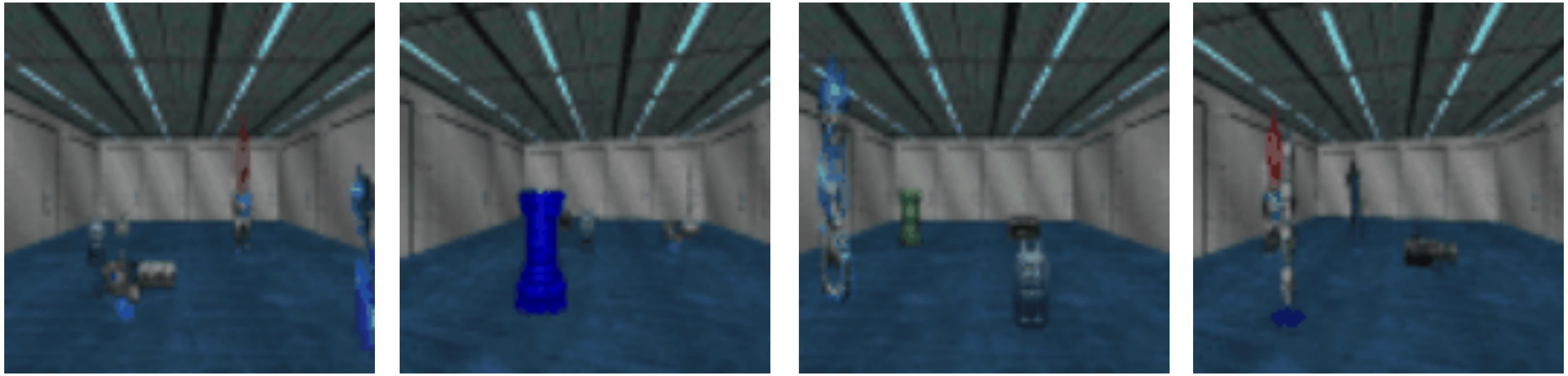
Visual Navigation
in VizDoom

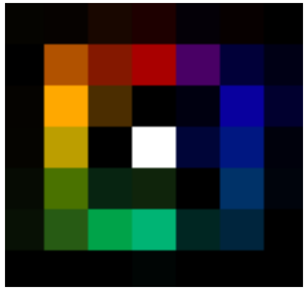


Object Pushing
with Sawyer in MuJoCo

What kind of tasks are discovered?

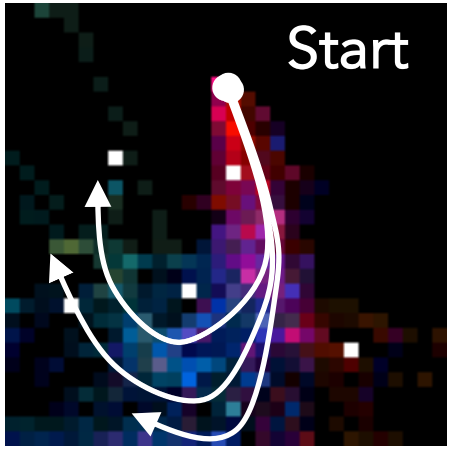


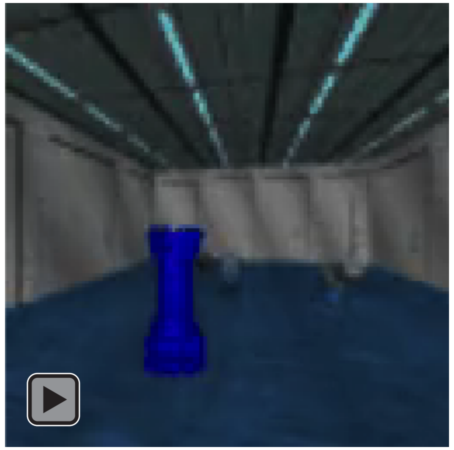


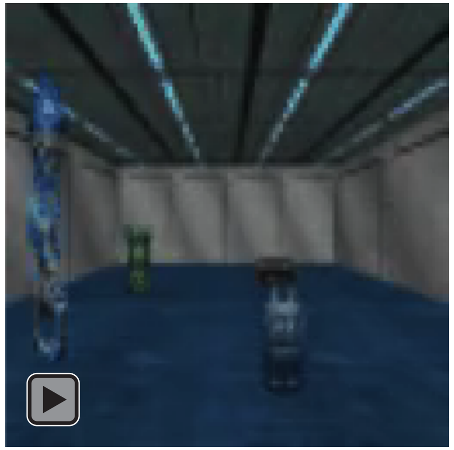


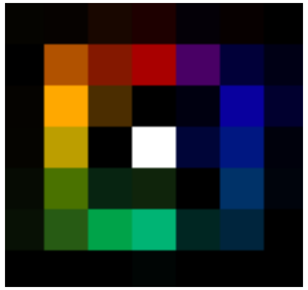
Direction encoded as color

What kind of tasks are discovered?





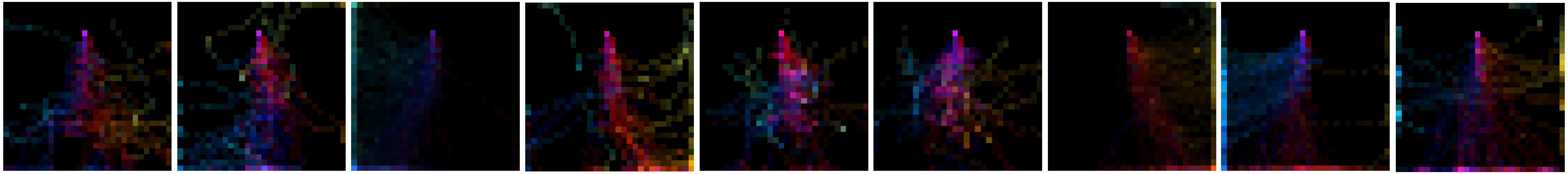




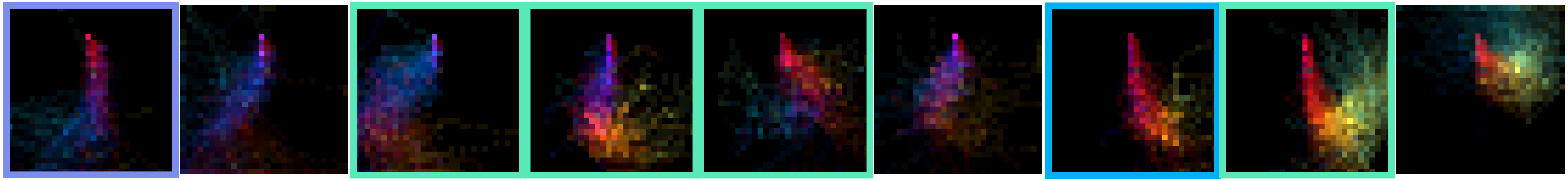
Direction encoded as color

What kind of tasks are discovered?

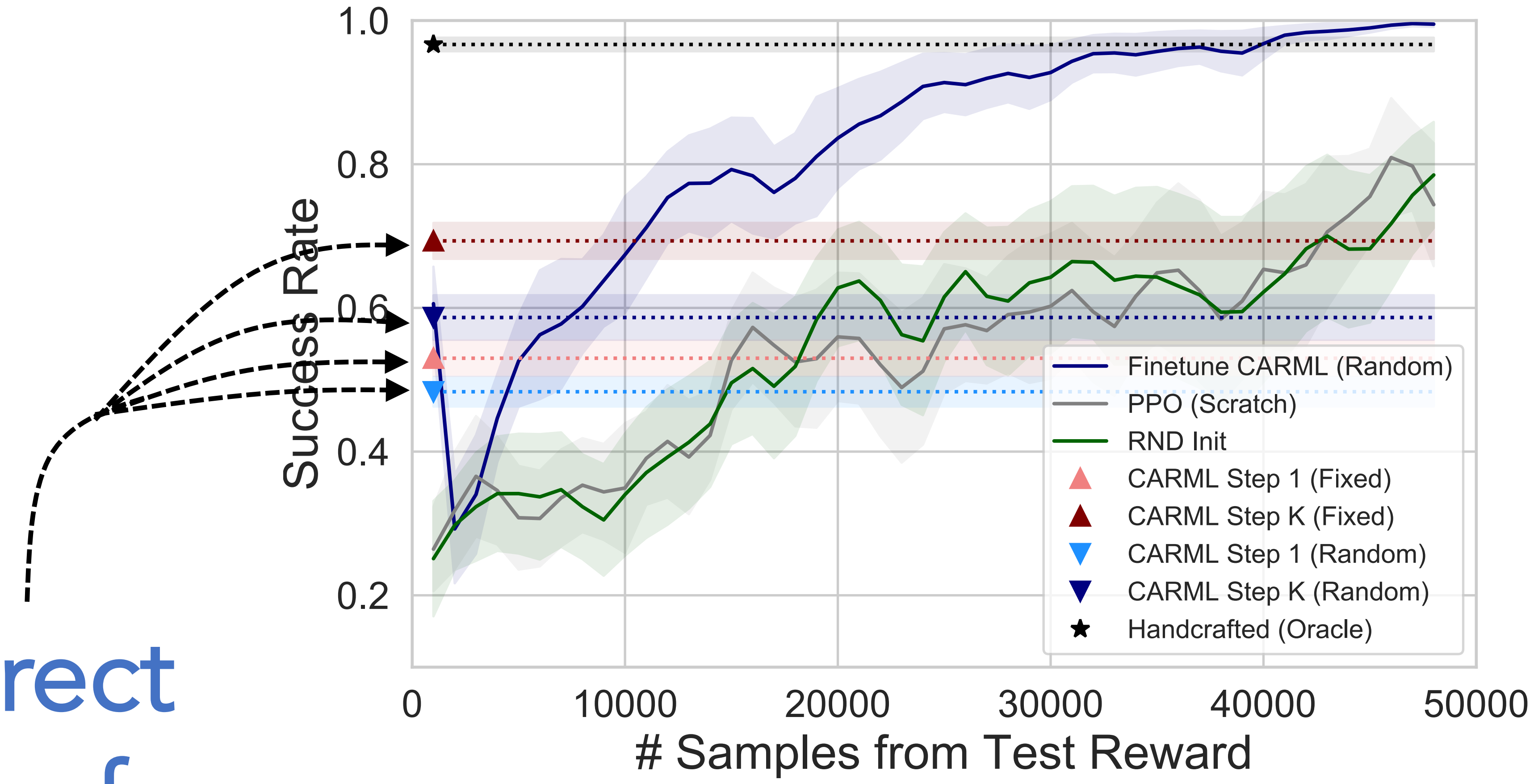
Step 1



Step 5

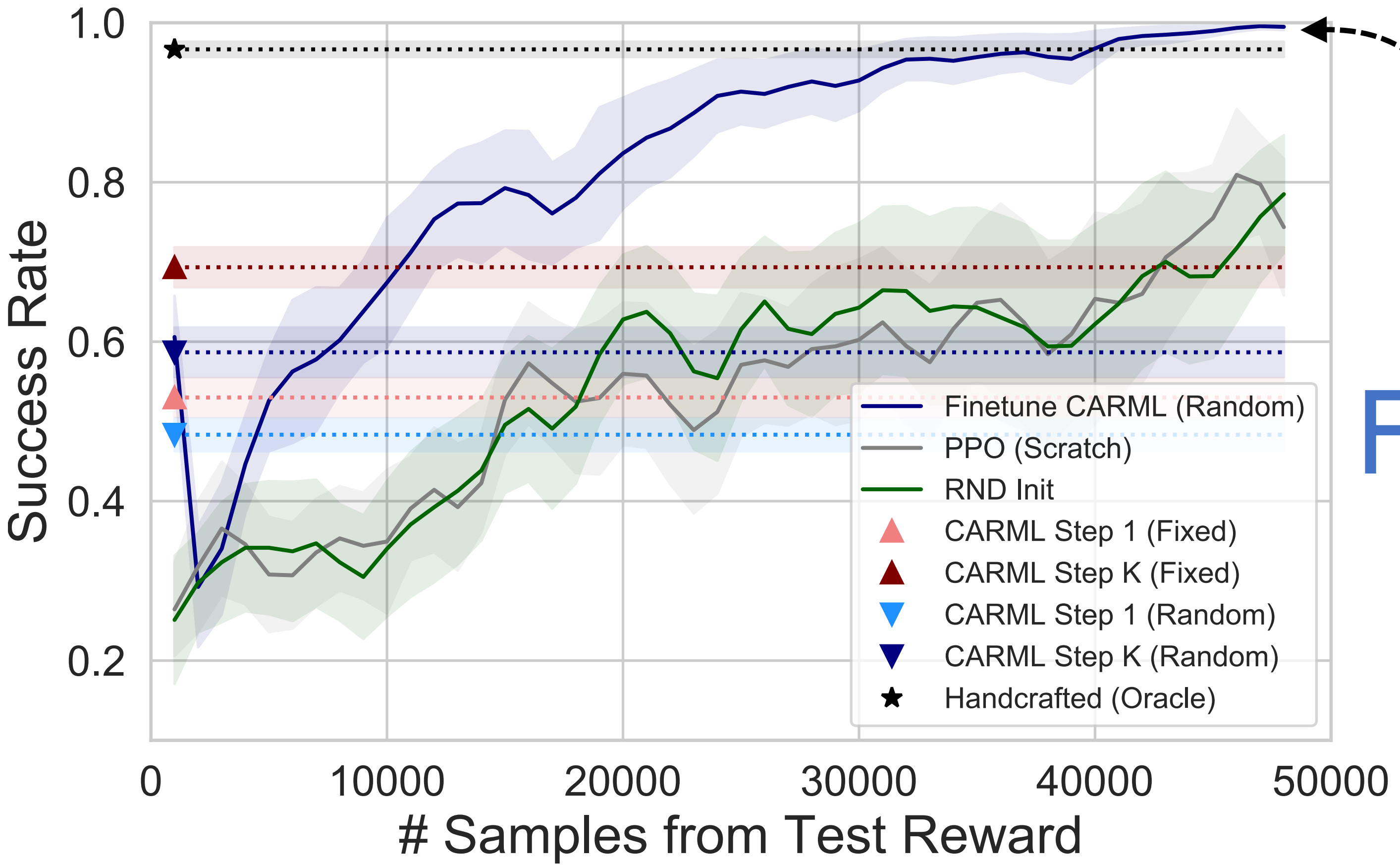


Transfer to Test Tasks – VizDoom



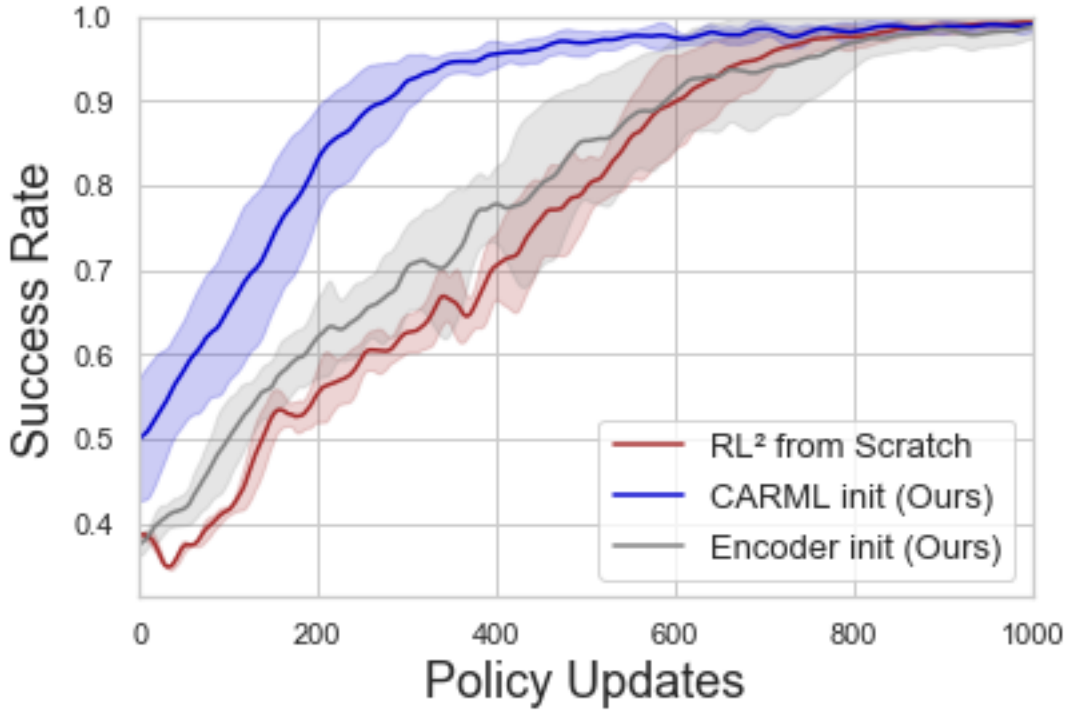
Direct Transfer

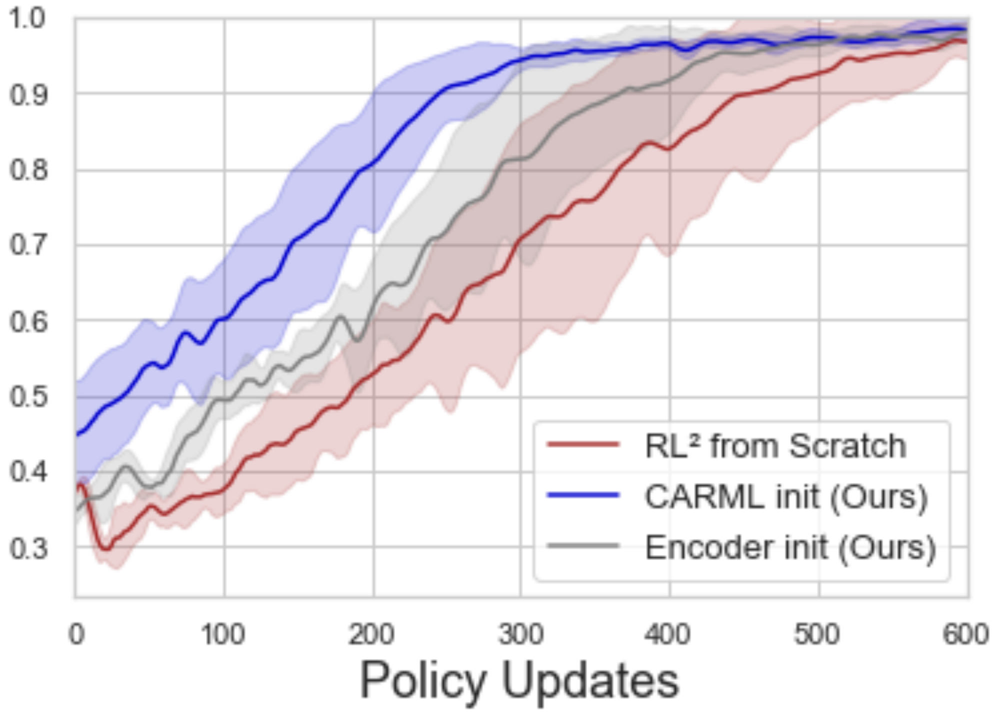
Transfer to Test Tasks – VizDoom



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/carml