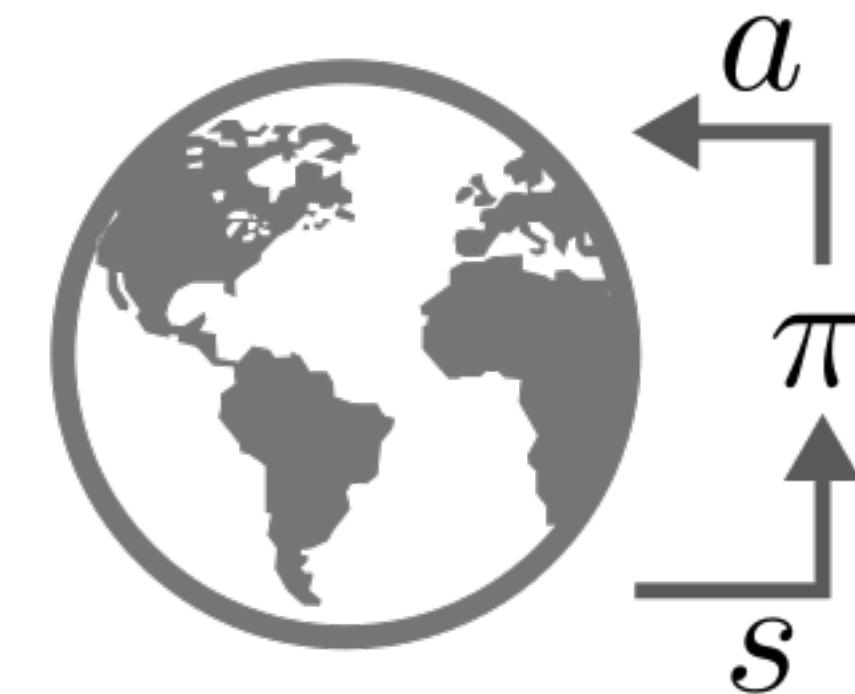


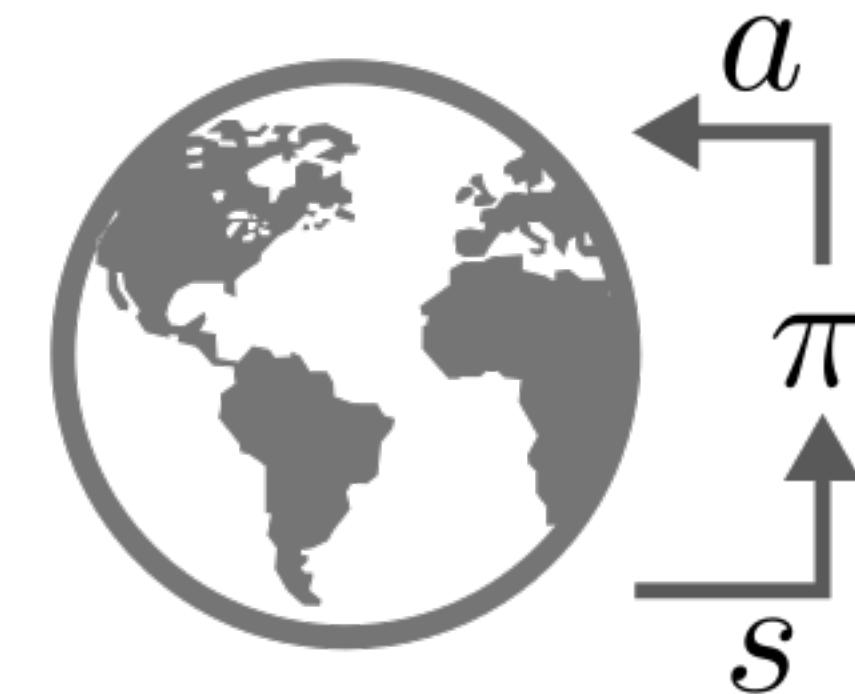
# Offline Meta-Reinforcement Learning with Online Self-Supervision

Vitchyr H. Pong, Ashvin Nair, Laura Smith, Catherine Huang, Sergey Levine  
UC Berkeley

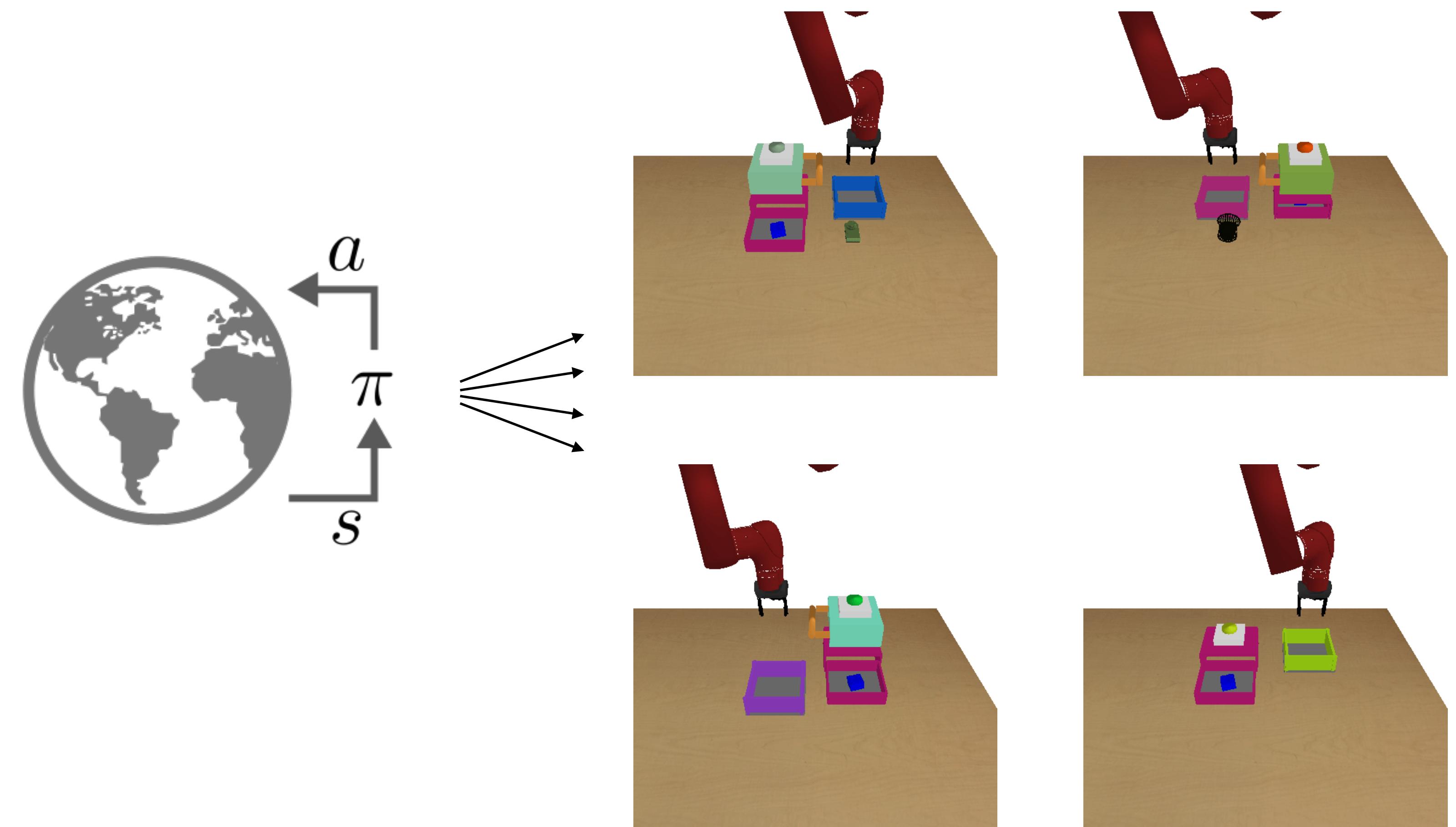
# Meta-Reinforcement Learning



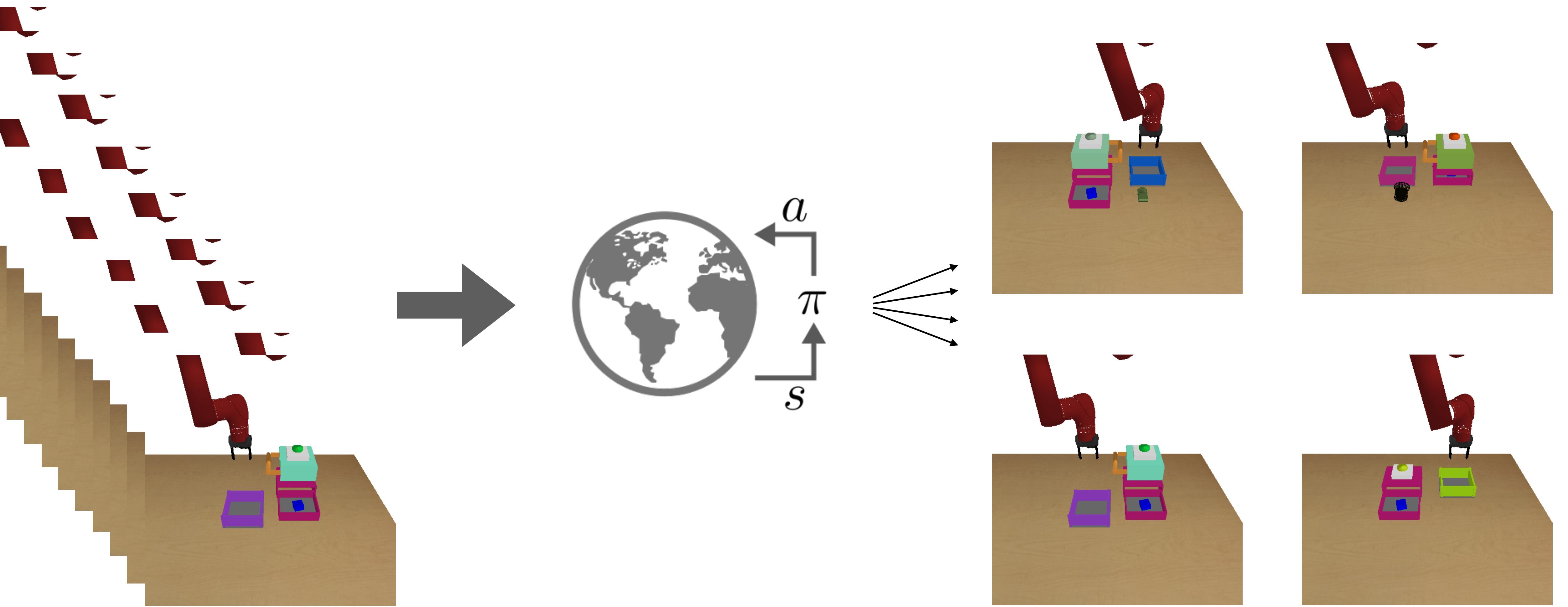
# Meta-Reinforcement Learning



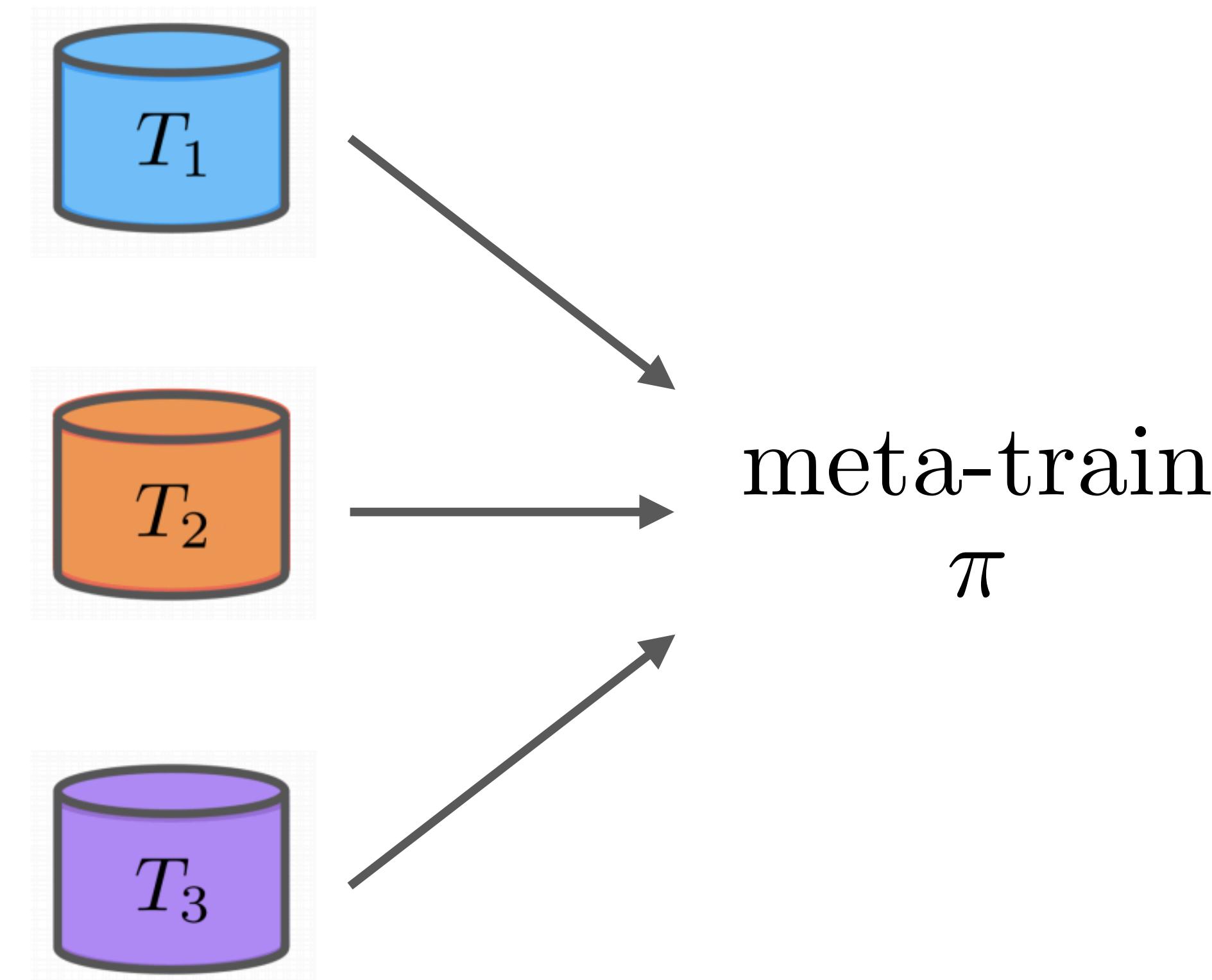
# Meta-Reinforcement Learning



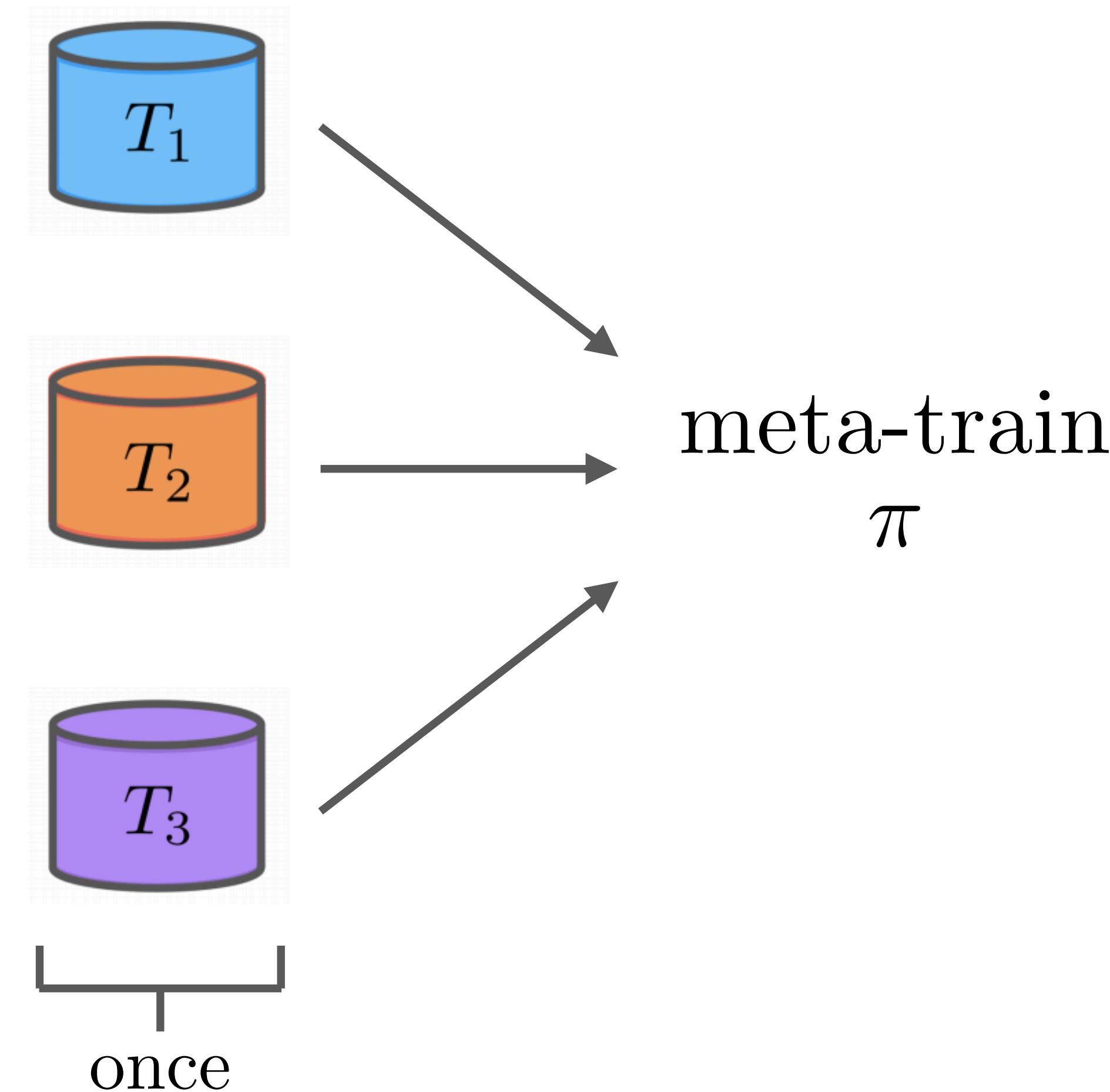
# Meta-Reinforcement Learning



# Offline Meta-Reinforcement Learning



# Offline Meta-Reinforcement Learning



# Challenge with Offline Meta-Reinforcement Learning

# Challenge with Offline Meta-Reinforcement Learning

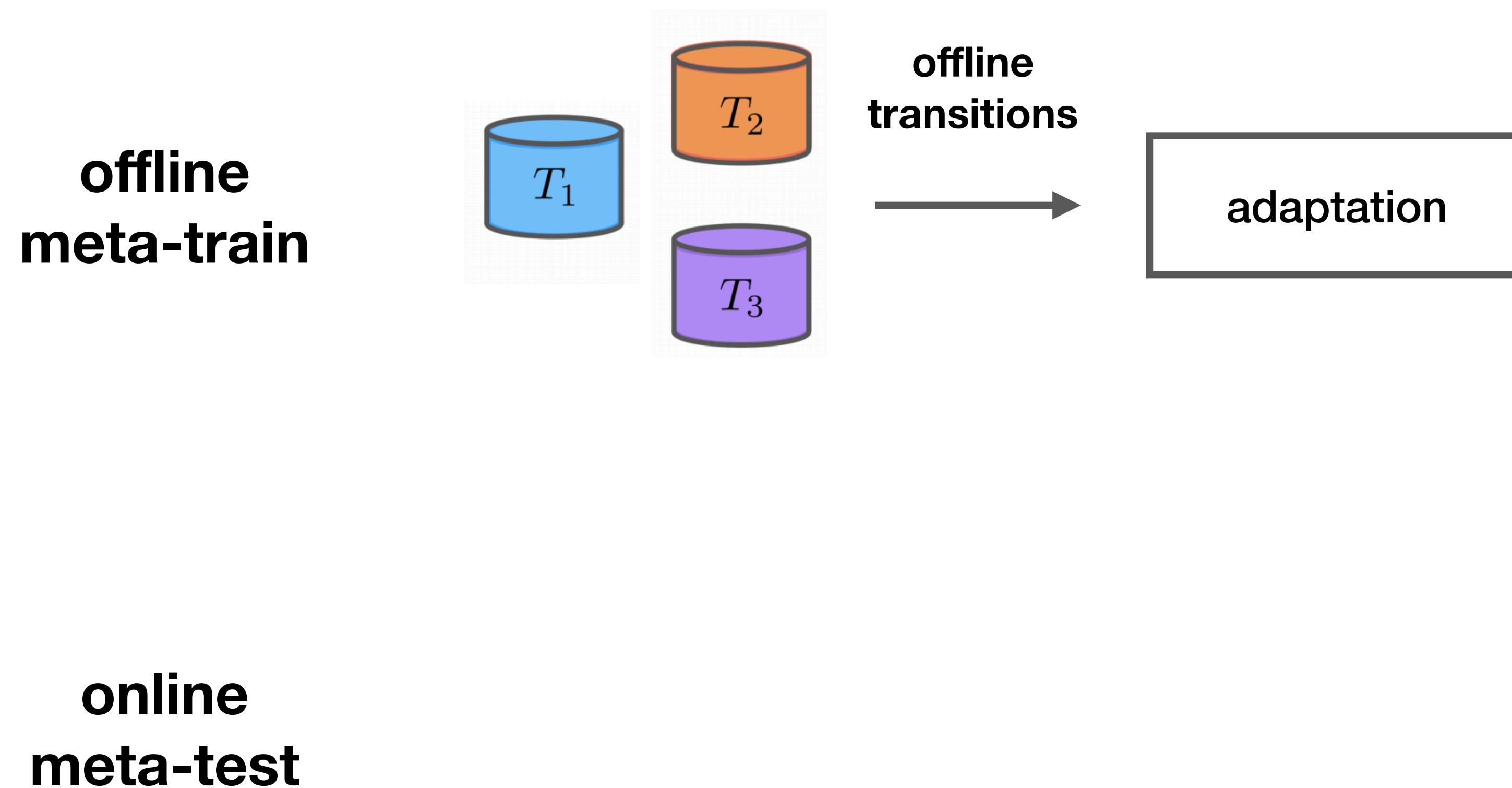
**offline  
meta-train**

# Challenge with Offline Meta-Reinforcement Learning

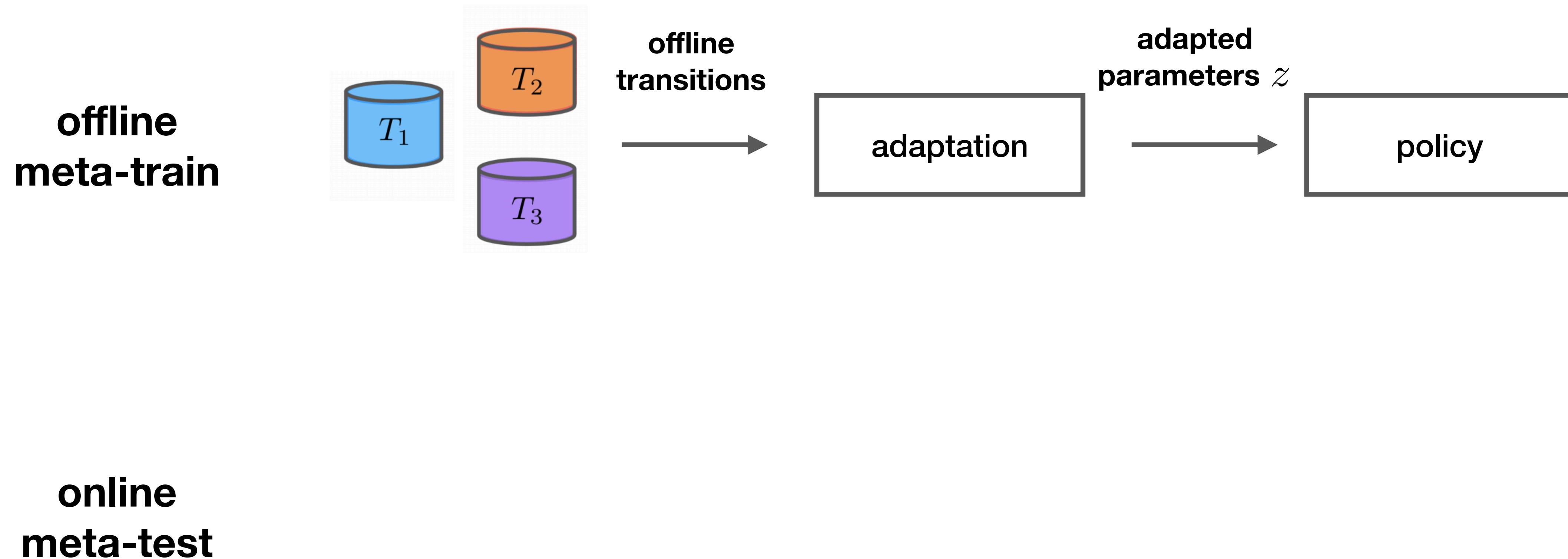
**offline**  
**meta-train**

**online**  
**meta-test**

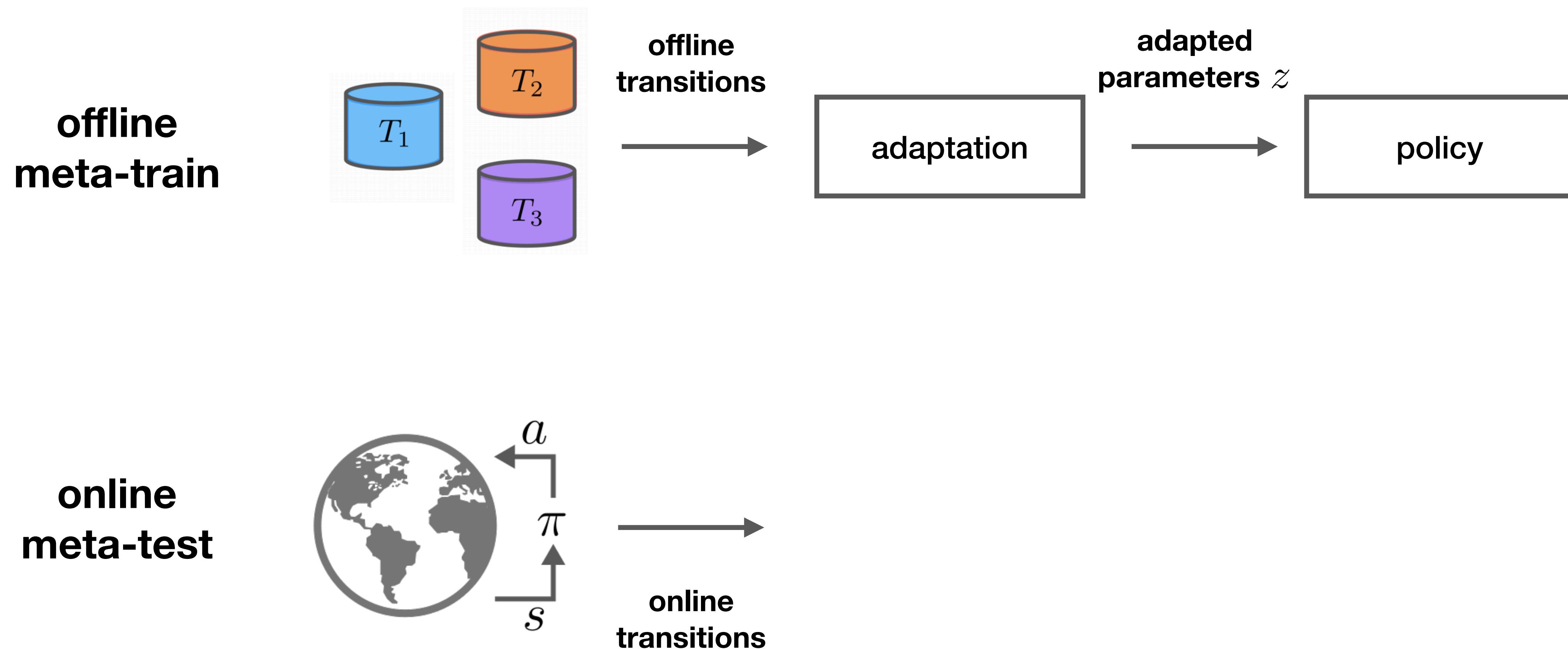
# Challenge with Offline Meta-Reinforcement Learning



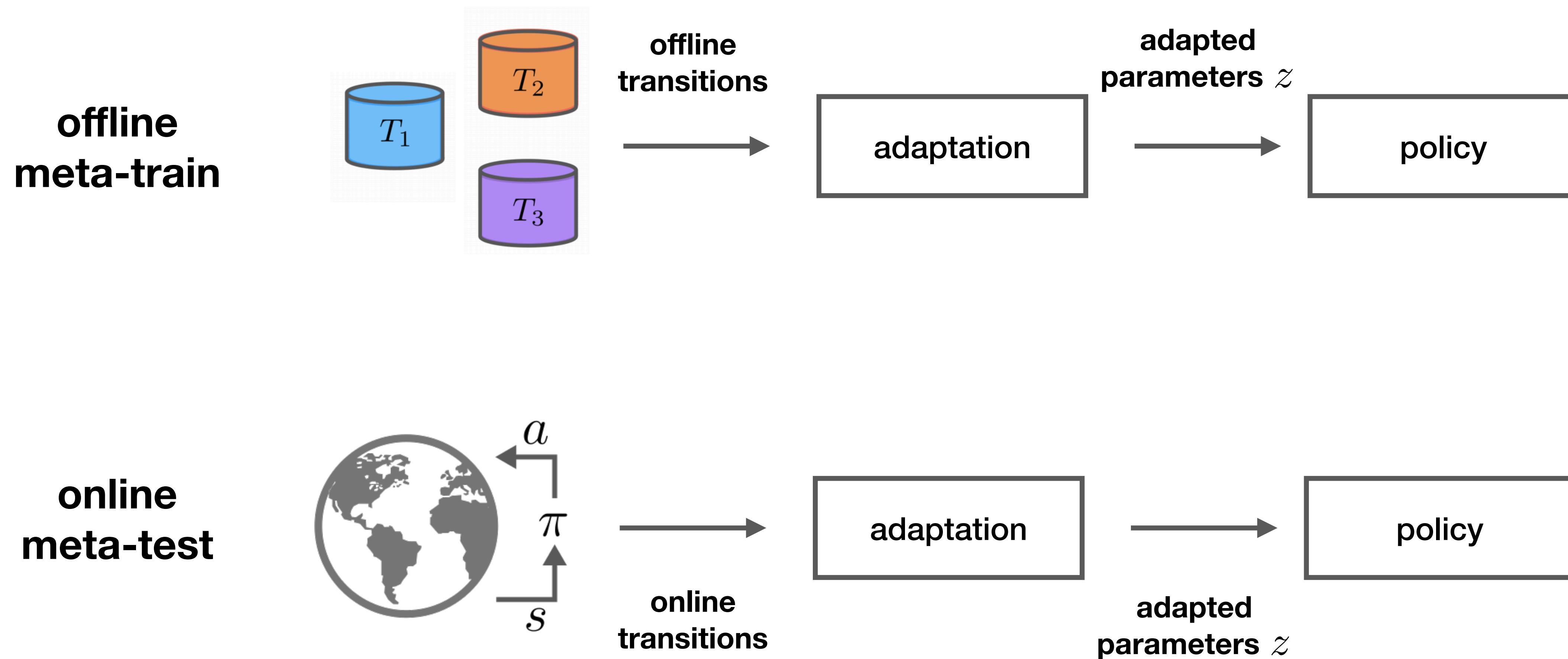
# Challenge with Offline Meta-Reinforcement Learning



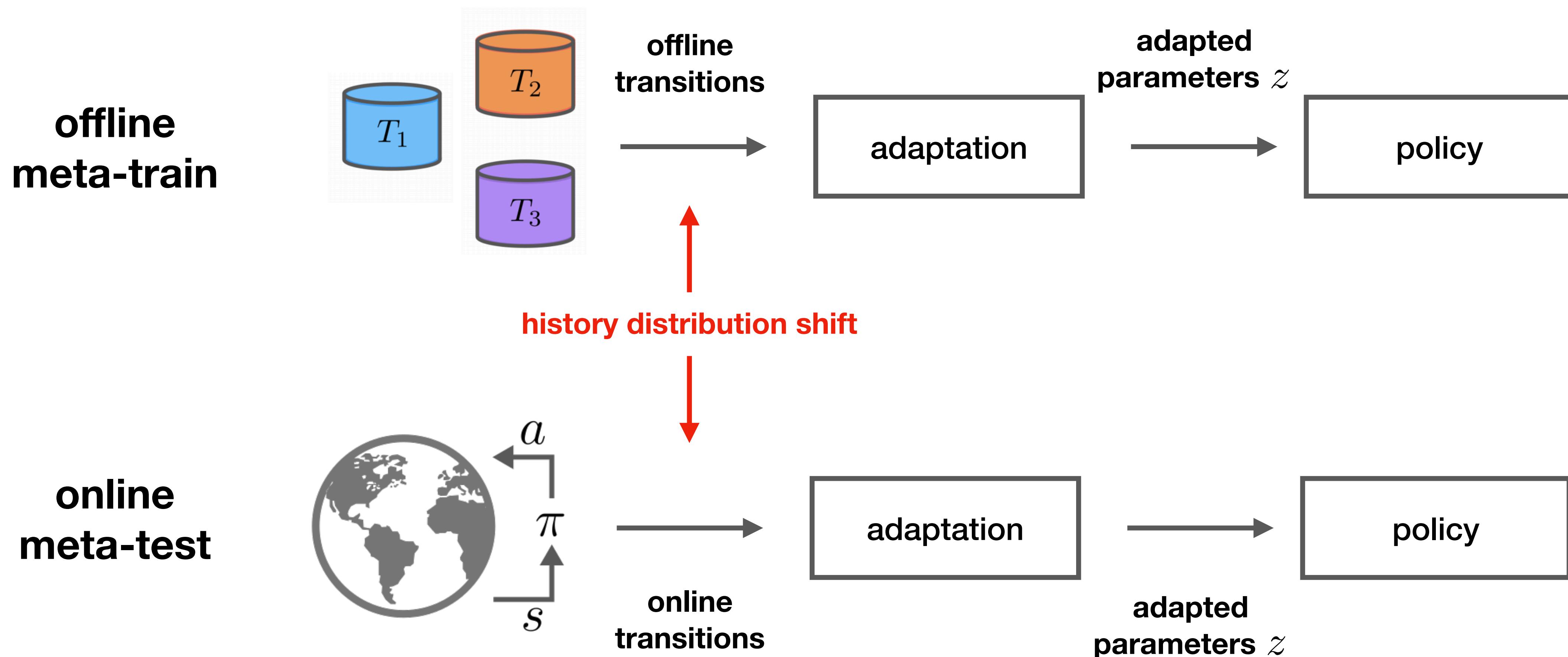
# Challenge with Offline Meta-Reinforcement Learning



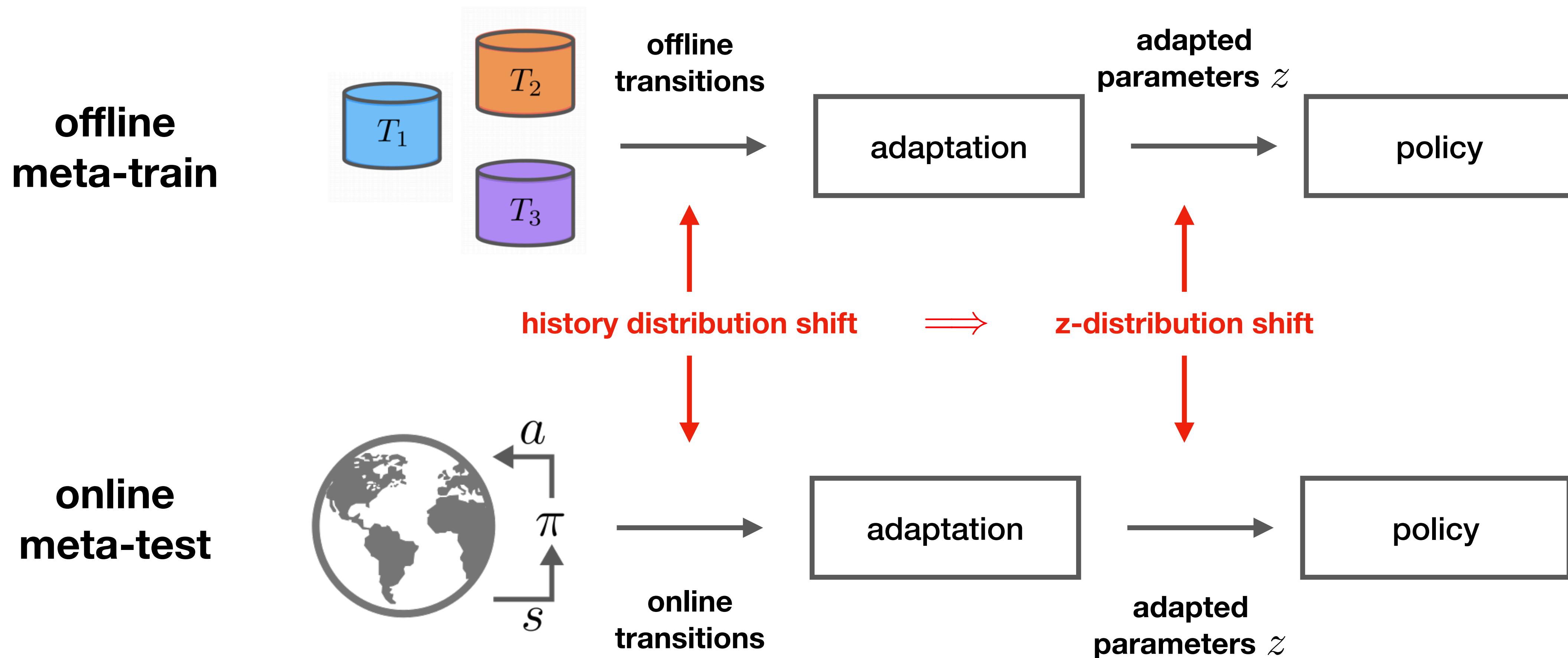
# Challenge with Offline Meta-Reinforcement Learning



# Challenge with Offline Meta-Reinforcement Learning



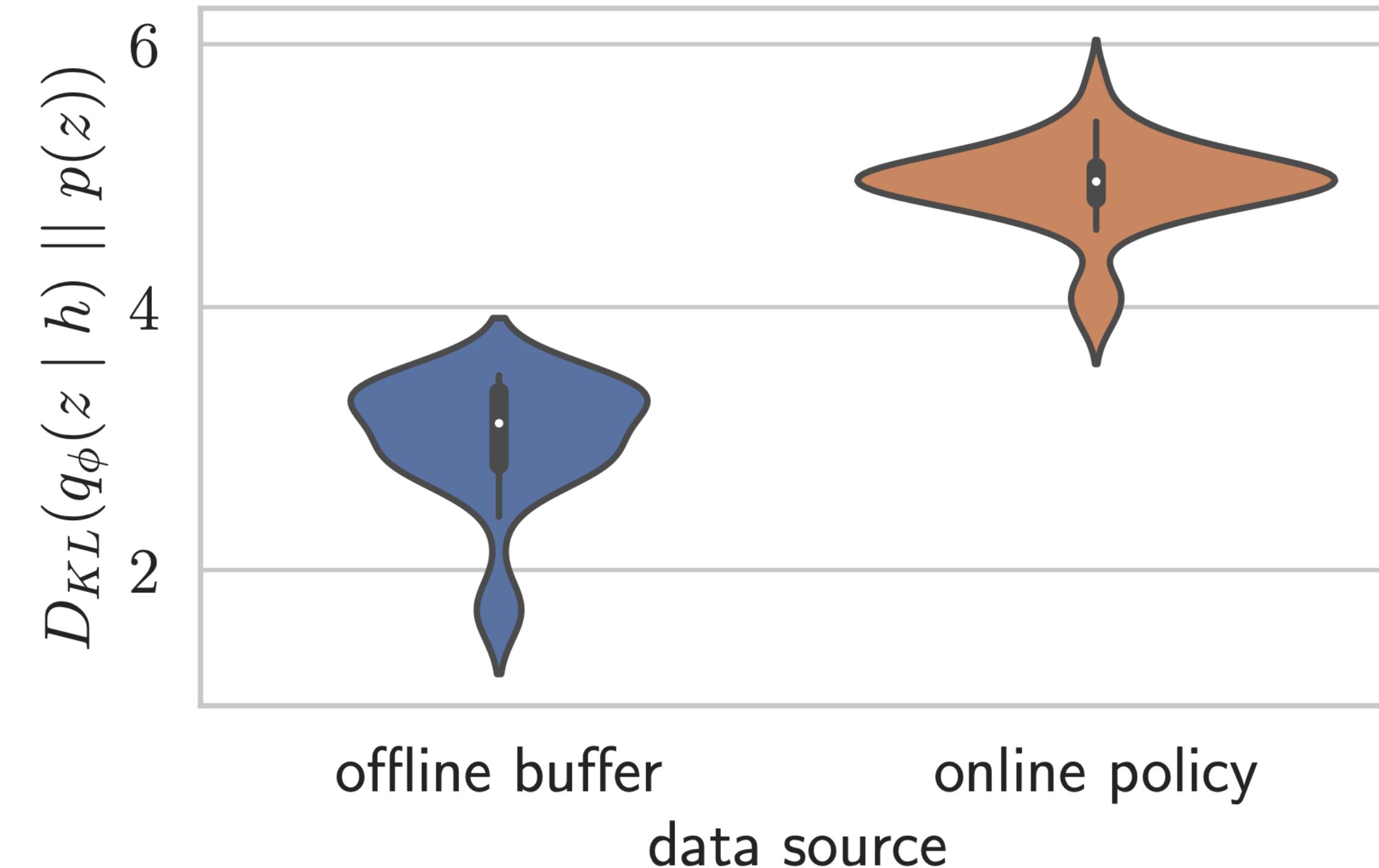
# Challenge with Offline Meta-Reinforcement Learning



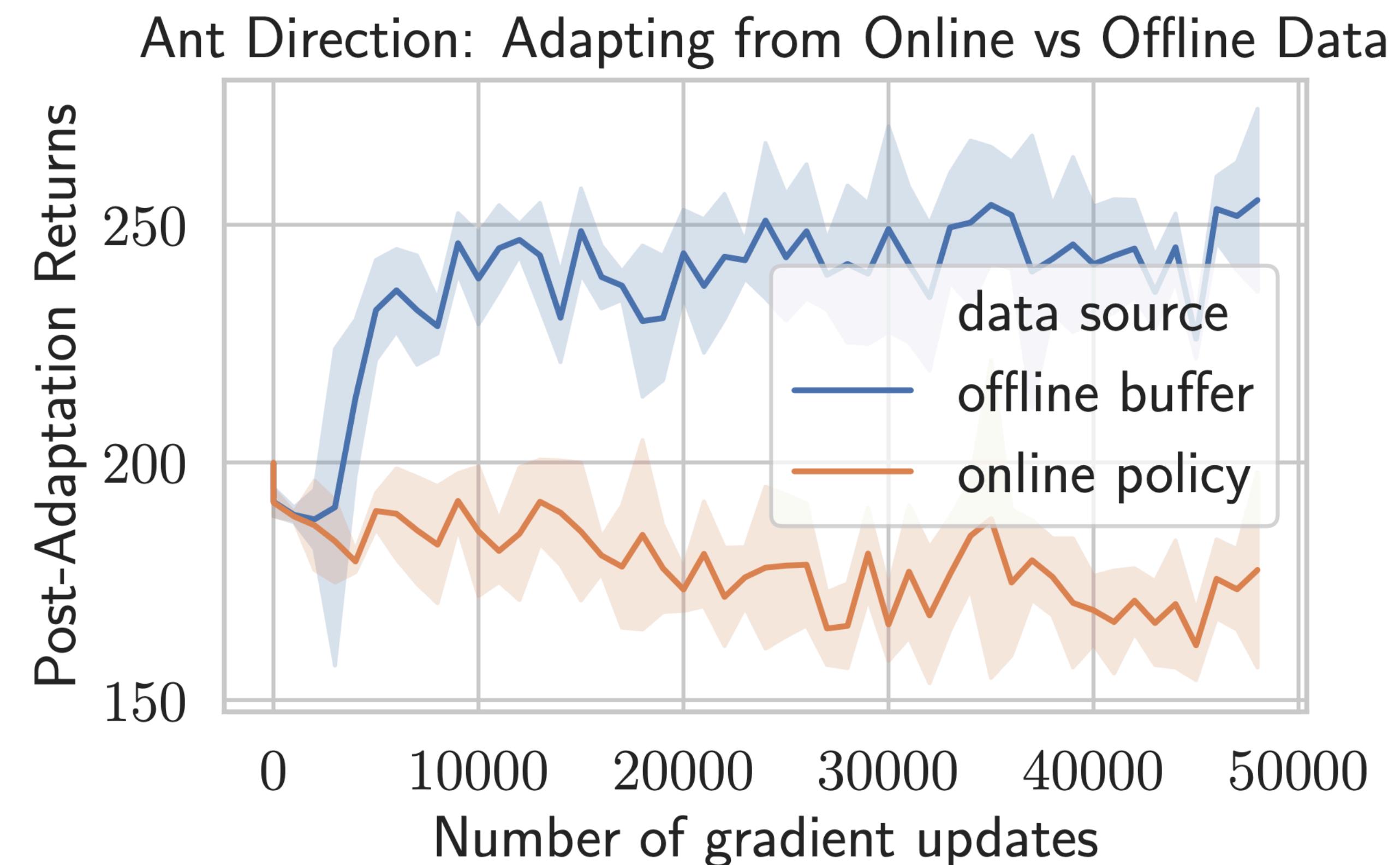
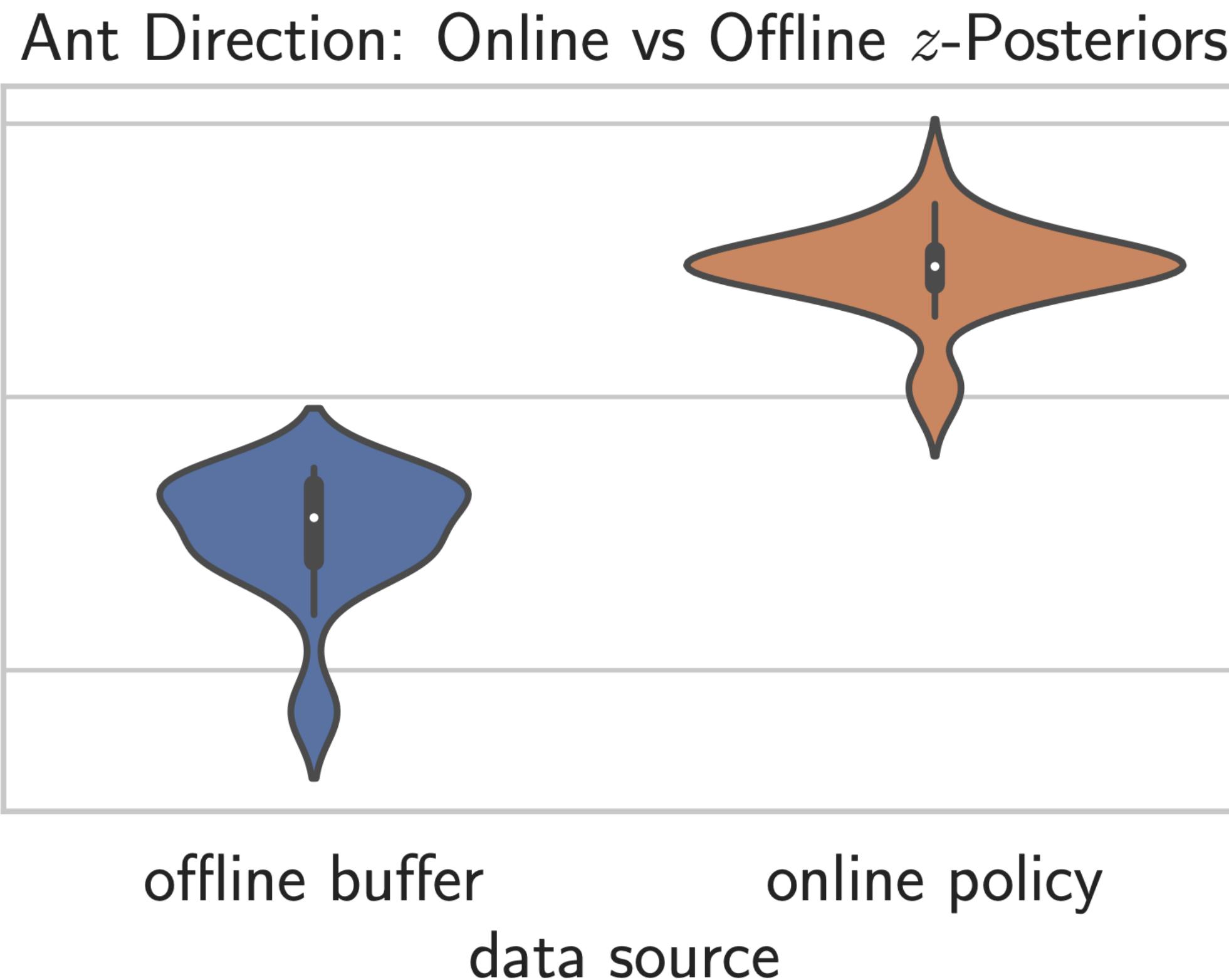
# **Evidence for Z-Space Distribution Shift**

# Evidence for Z-Space Distribution Shift

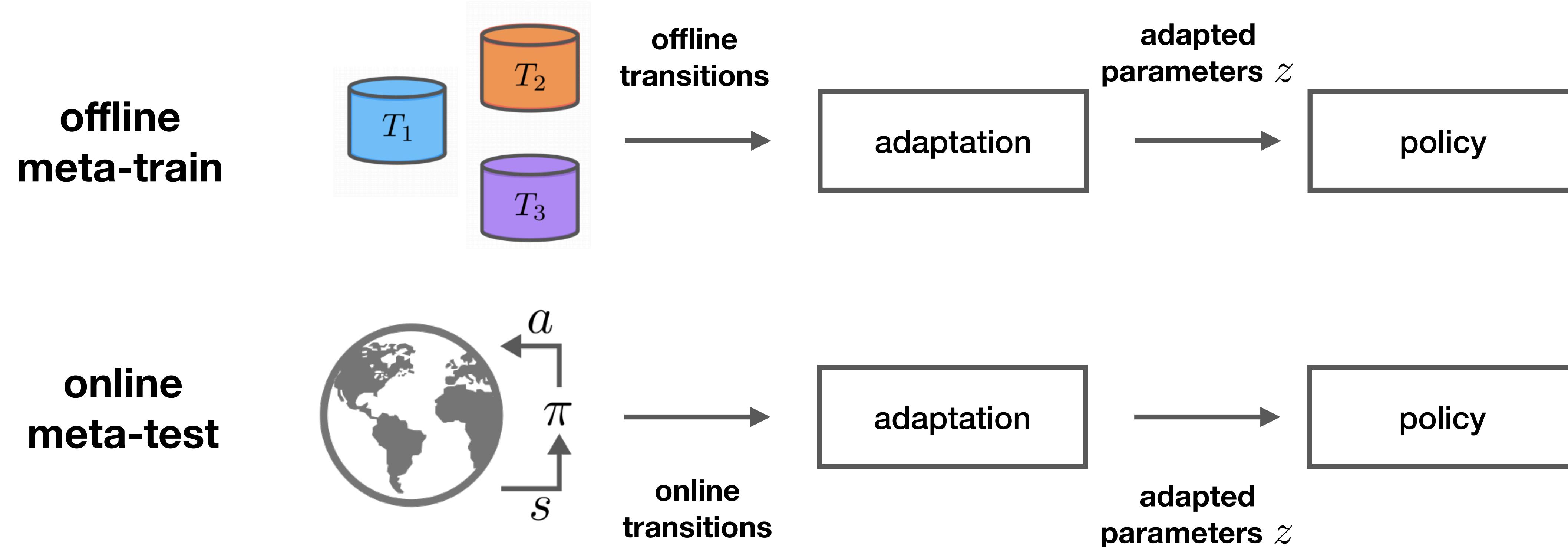
Ant Direction: Online vs Offline  $z$ -Posteriors



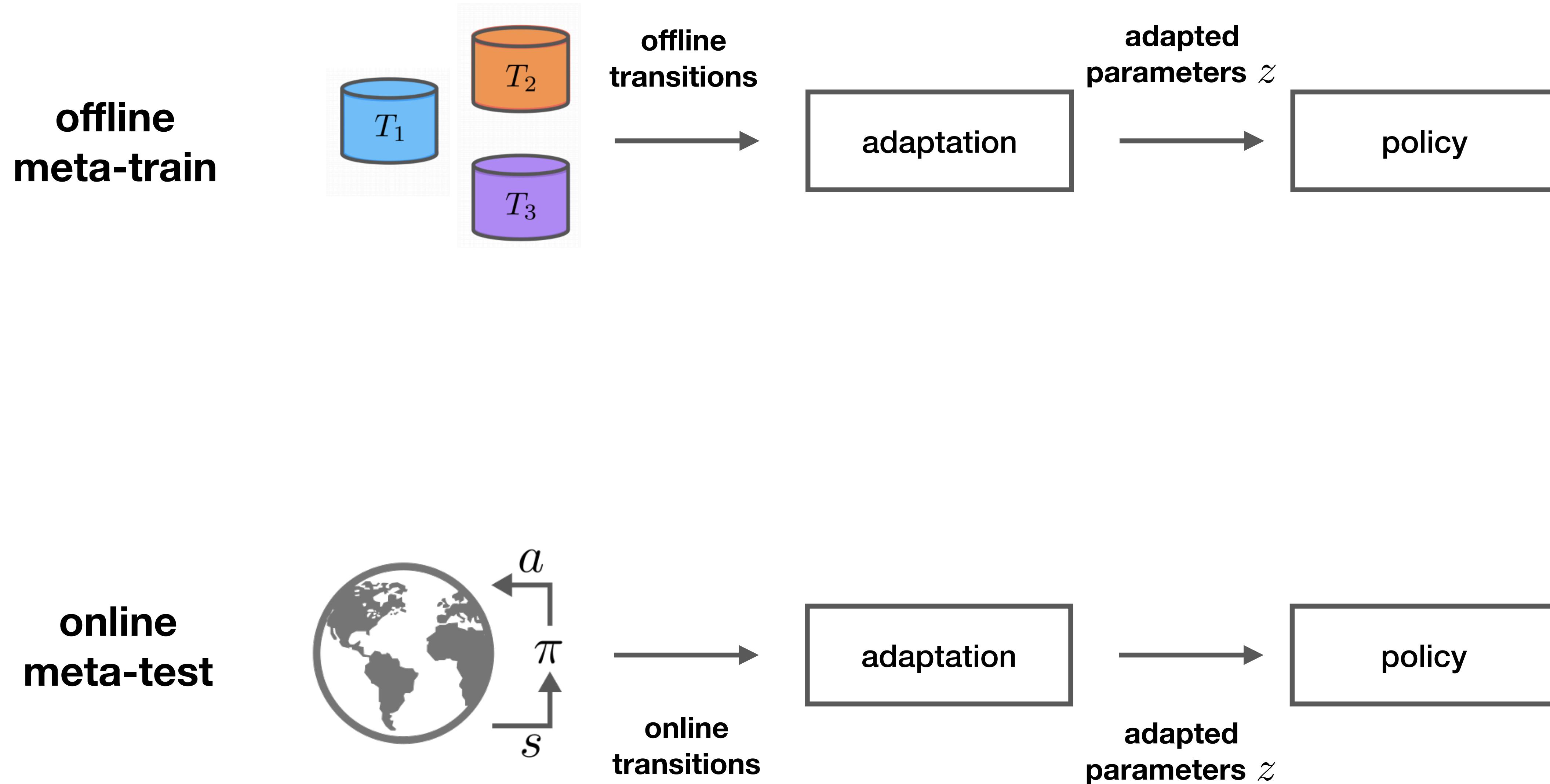
# Evidence for Z-Space Distribution Shift



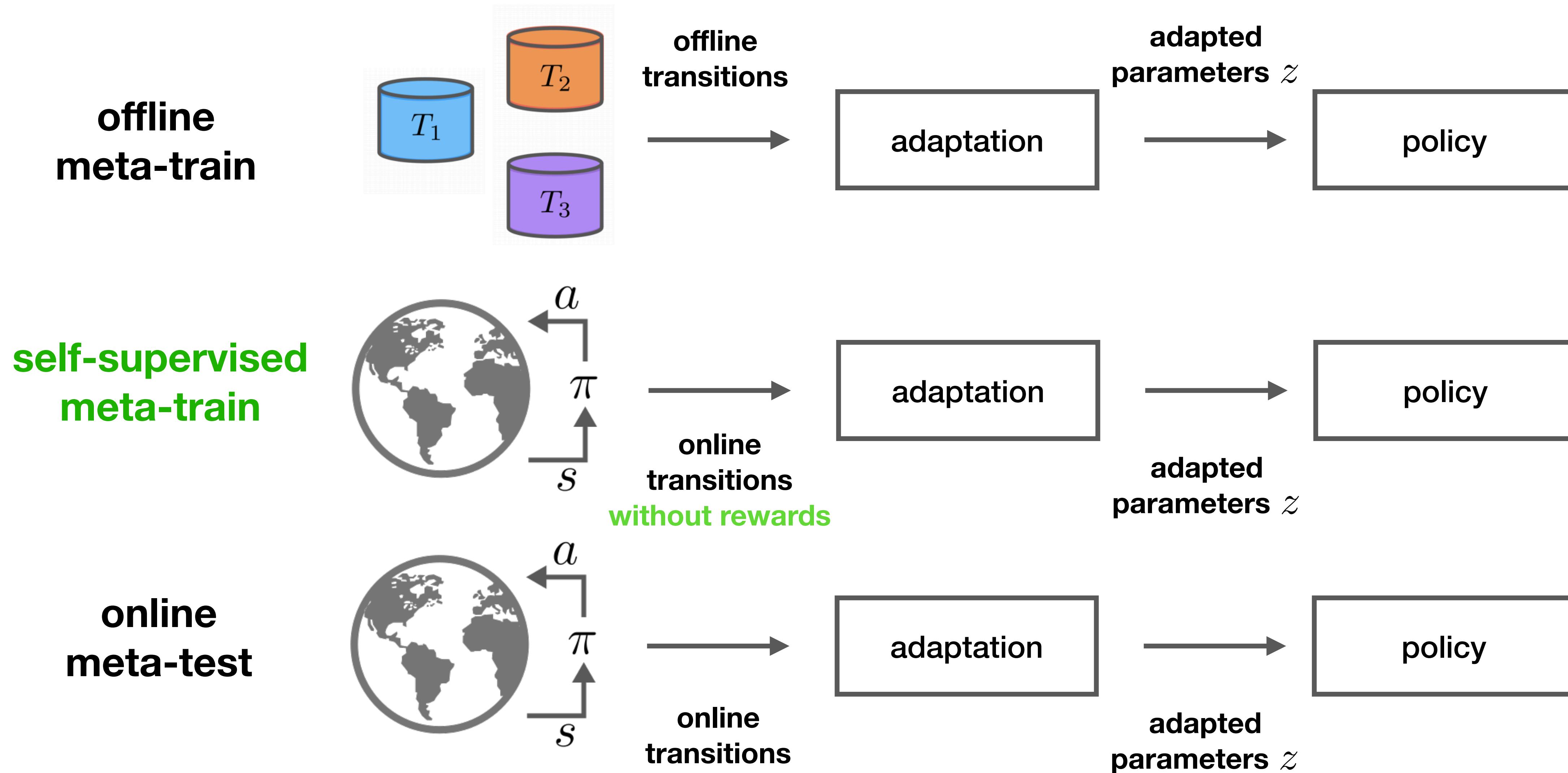
# Semi-Supervised Meta-Actor Critic (SMAC)



# Semi-Supervised Meta-Actor Critic (SMAC)



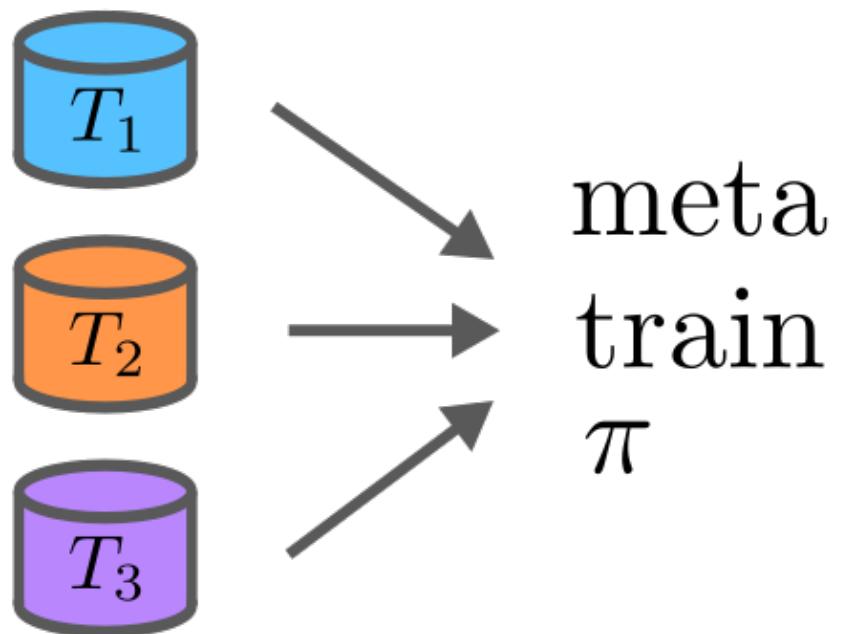
# Semi-Supervised Meta-Actor Critic (SMAC)



# **SMAC Overview**

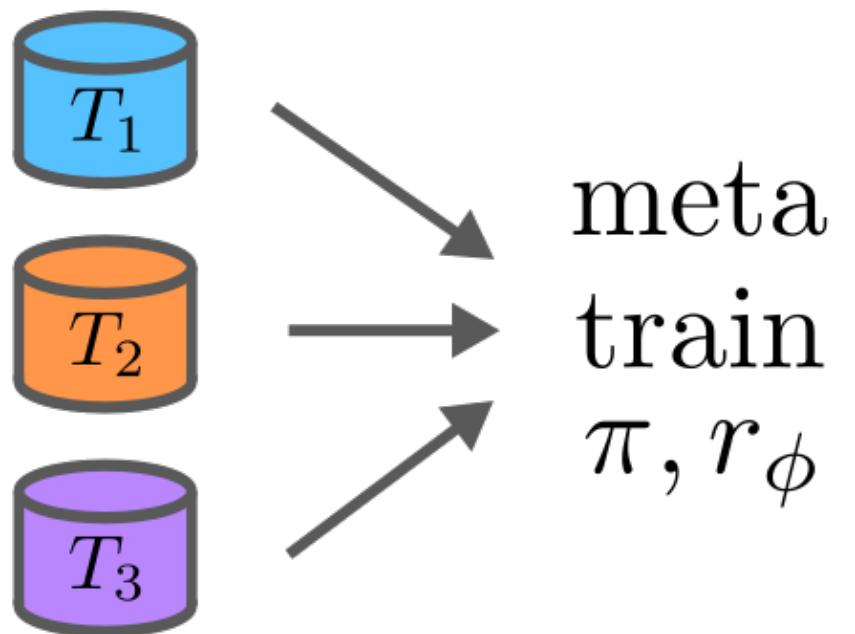
# SMAC Overview

## 1. offline meta-RL



# SMAC Overview

## 1. offline meta-RL

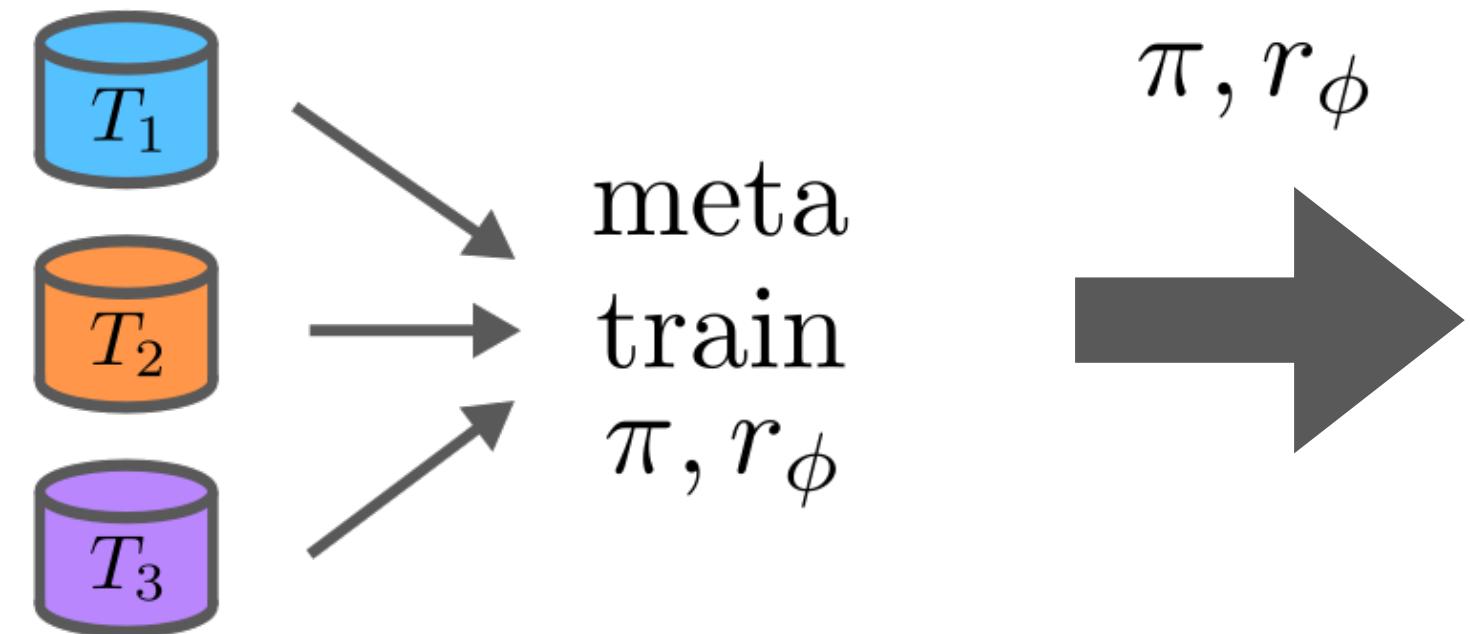


reward decoder     $r \sim r_\phi(\cdot \mid s, a, z)$

# SMAC Overview

1. offline meta-RL

2. self-supervised online training

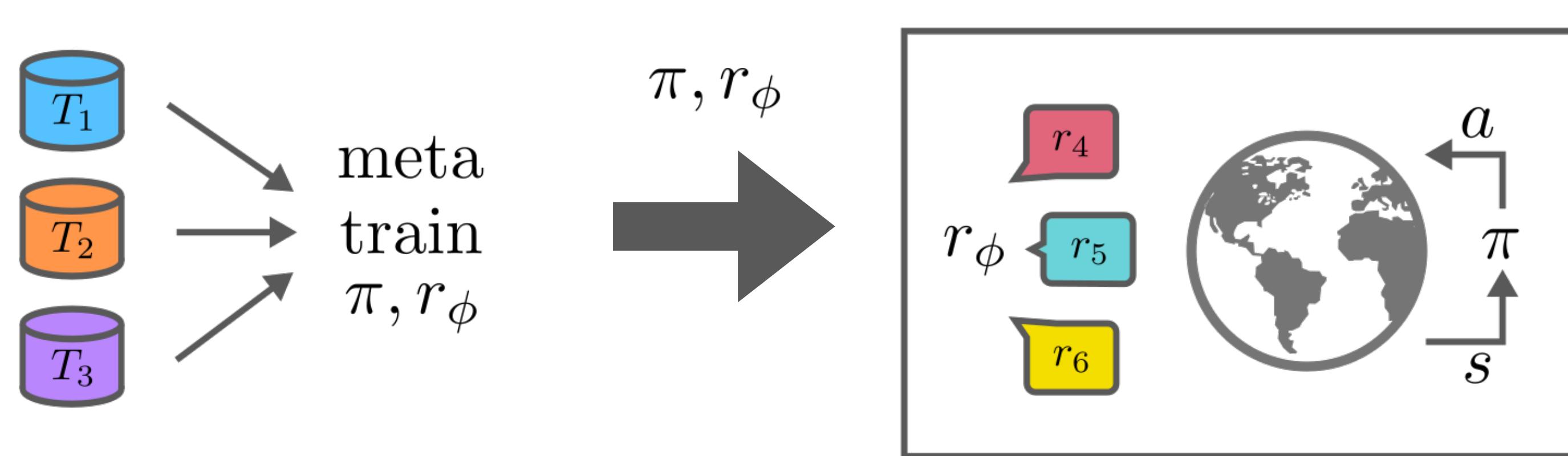


reward decoder     $r \sim r_\phi(\cdot \mid s, a, z)$

# SMAC Overview

1. offline meta-RL

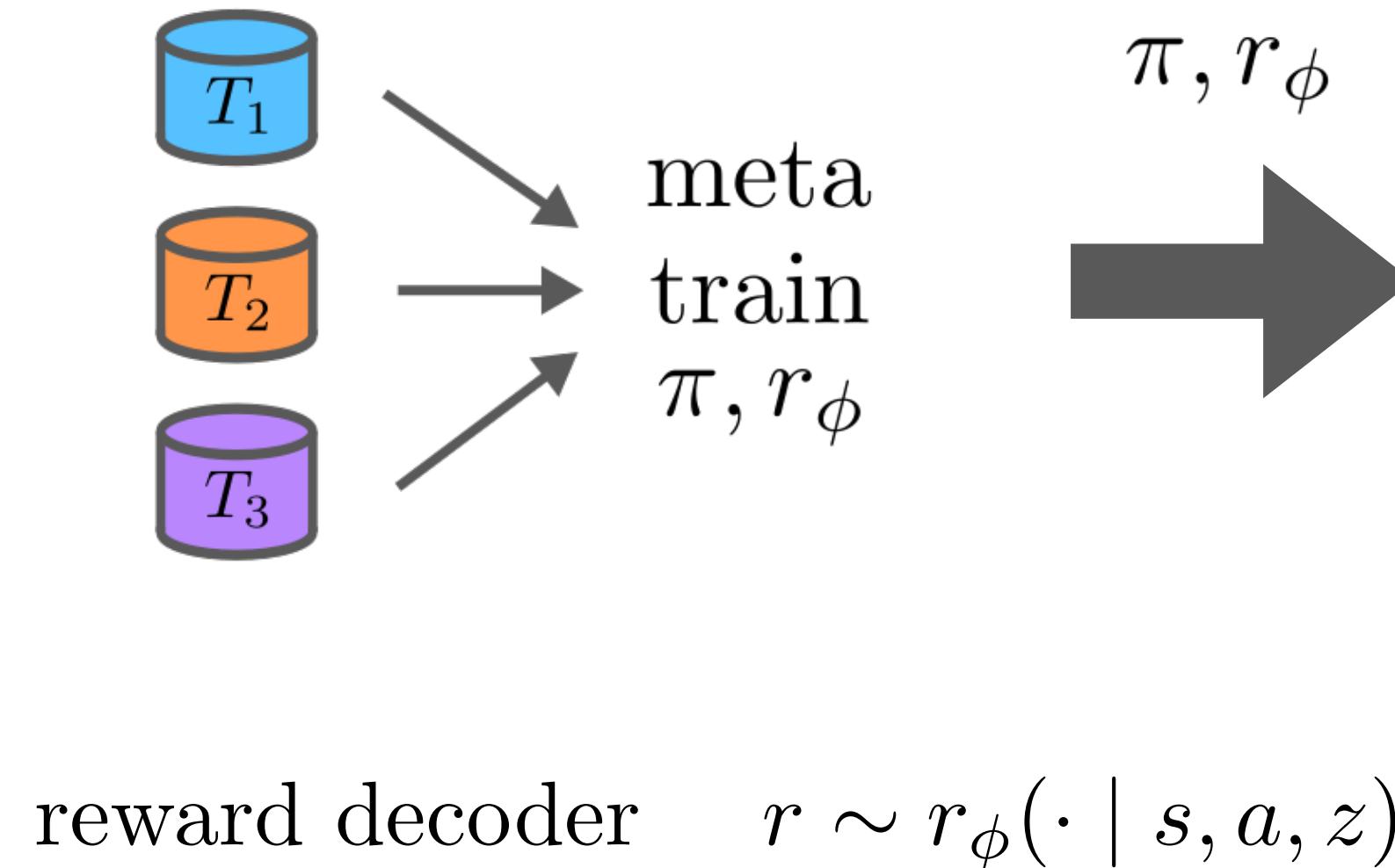
2. self-supervised online training



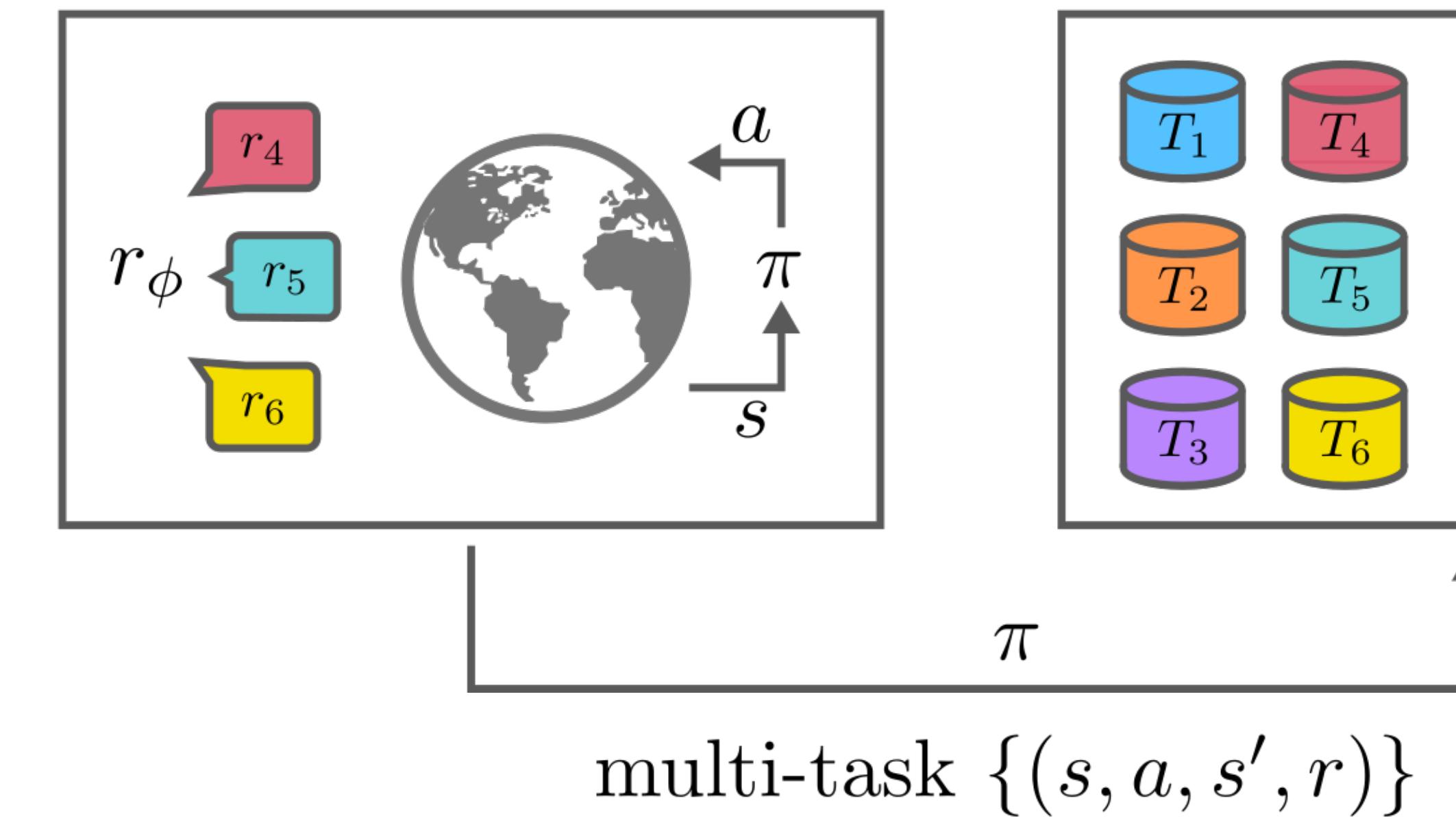
reward decoder     $r \sim r_\phi(\cdot \mid s, a, z)$

# SMAC Overview

1. offline meta-RL

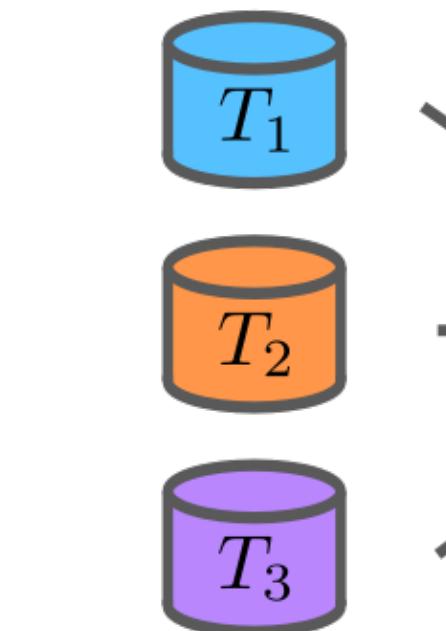


2. self-supervised online training



# SMAC Overview

1. offline meta-RL

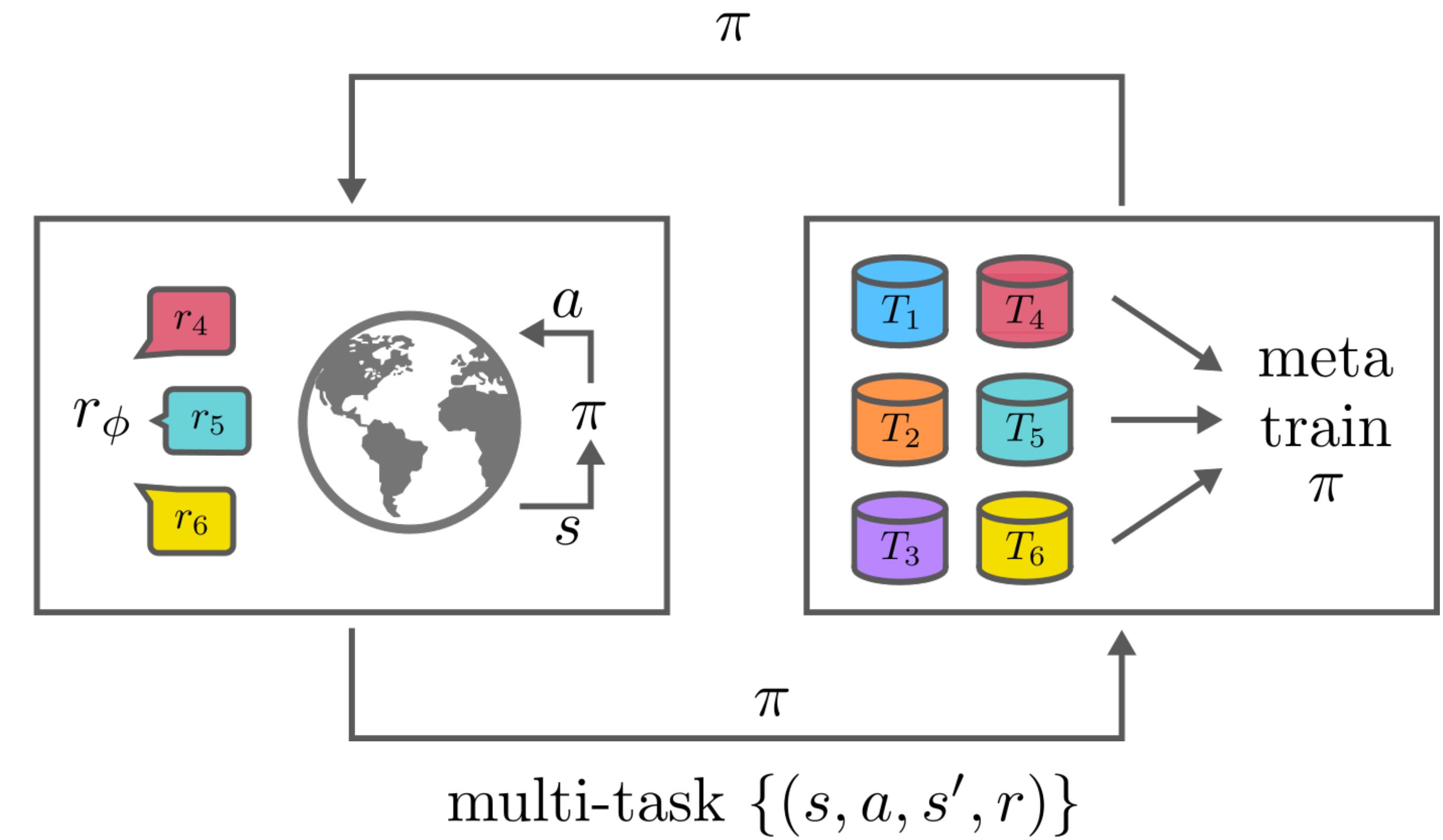


$\pi, r_\phi$

reward decoder

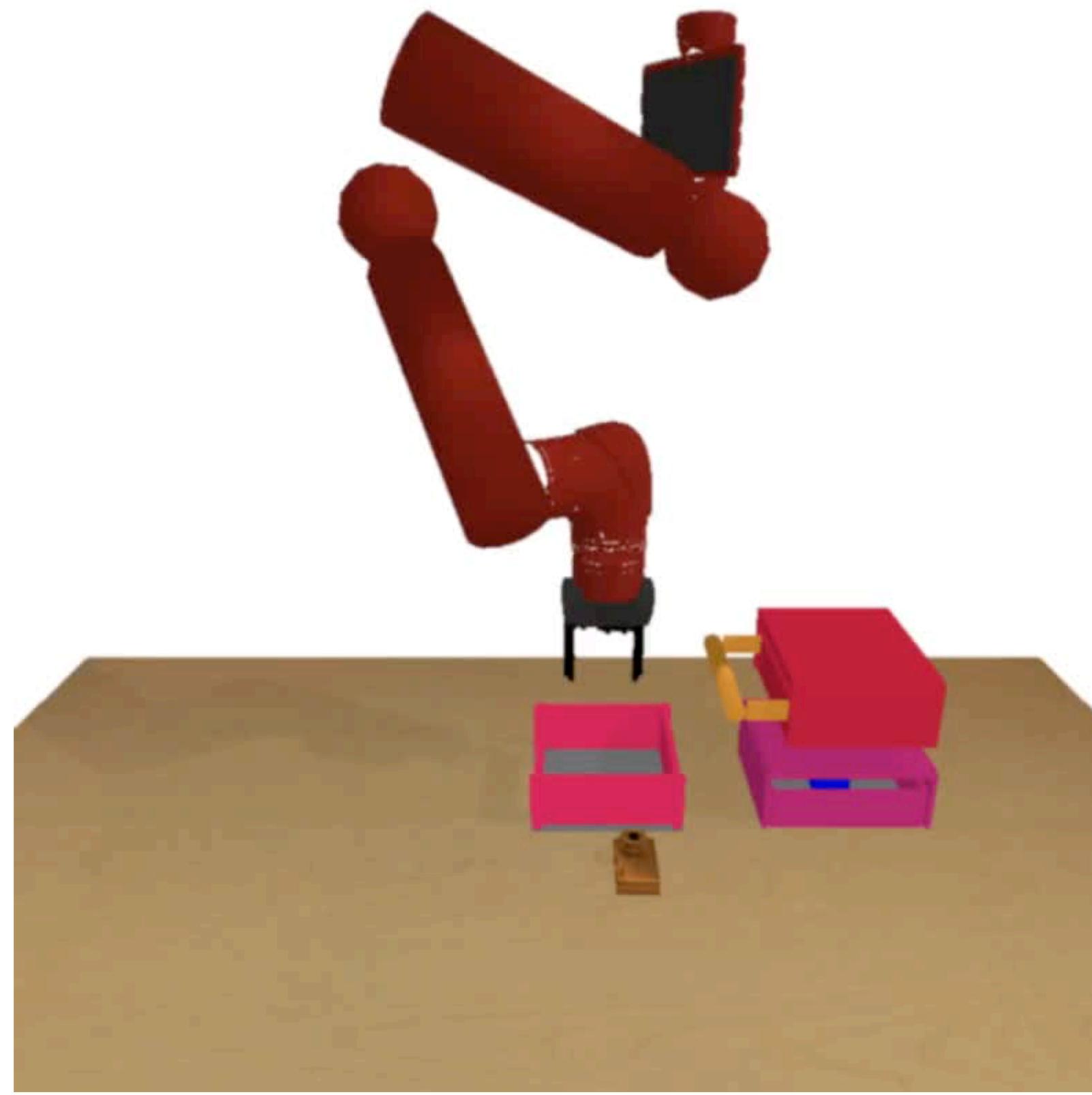
$$r \sim r_\phi(\cdot \mid s, a, z)$$

2. self-supervised online training



# SMAC meta-train improves skills

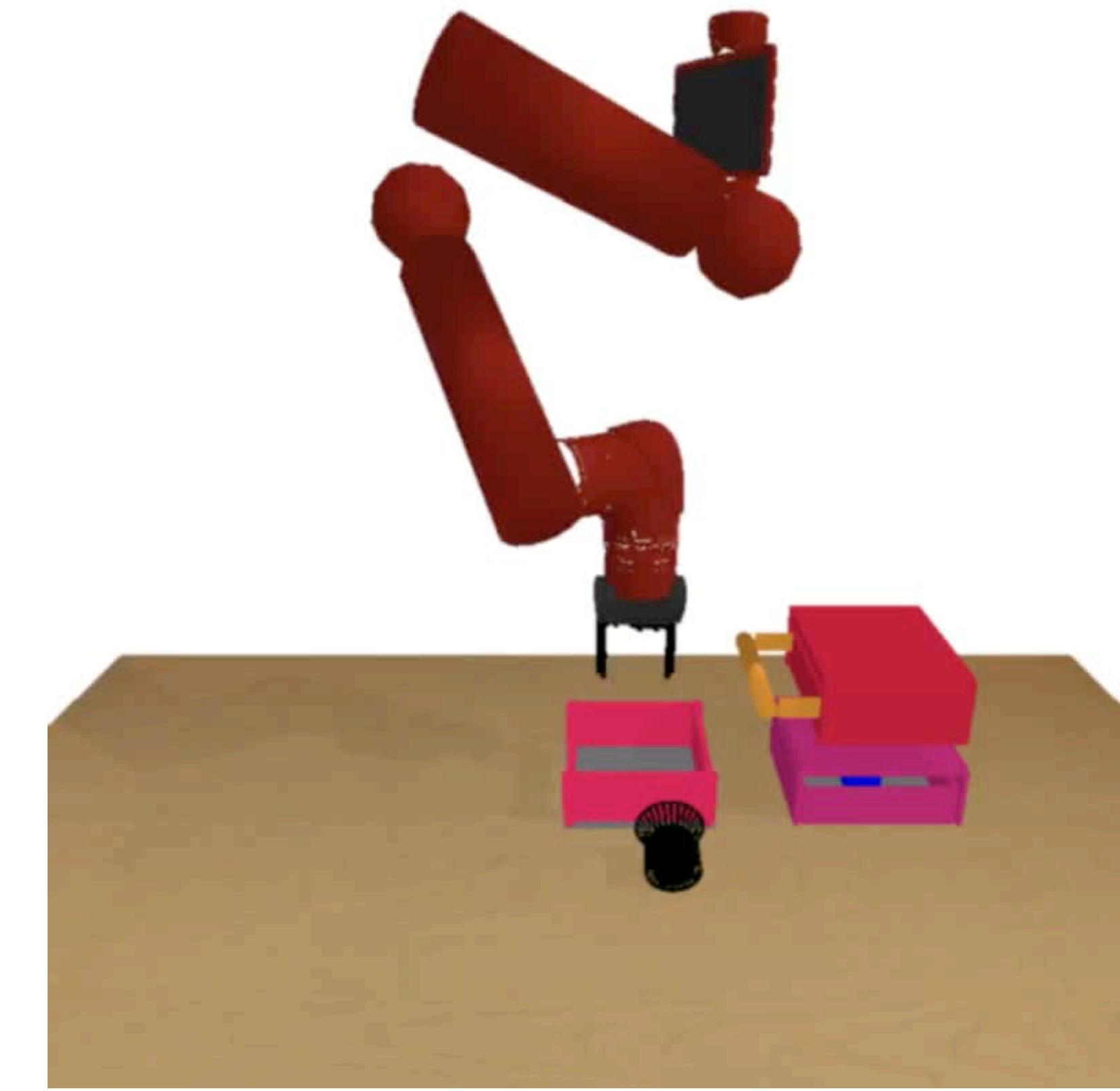
Iteration 0 (after offline training)



Meta test episode 1  
Meta test episode 2



Iteration 25 (after 150K steps of online training)

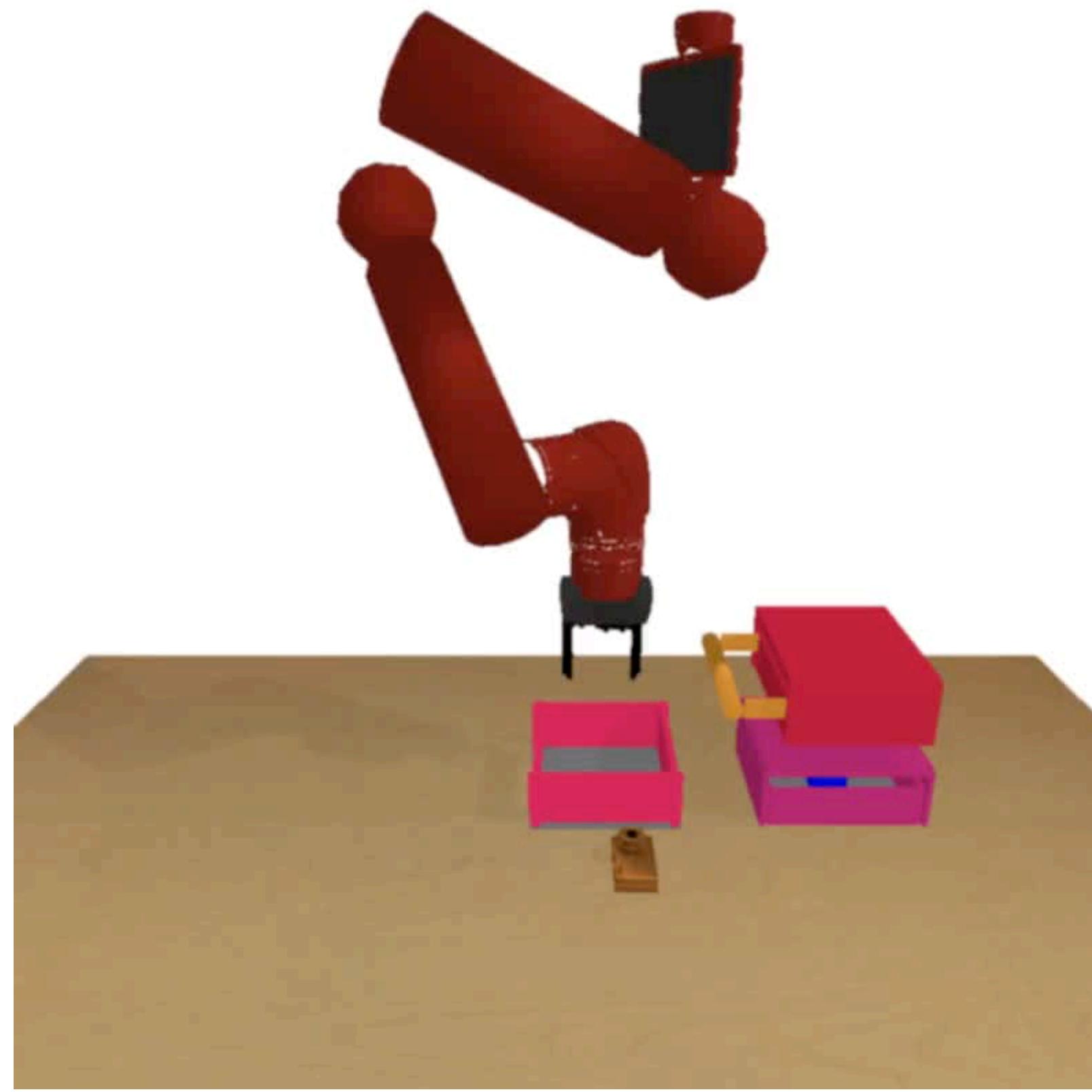


Meta test episode 1  
Meta test episode 2



# SMAC meta-train improves skills

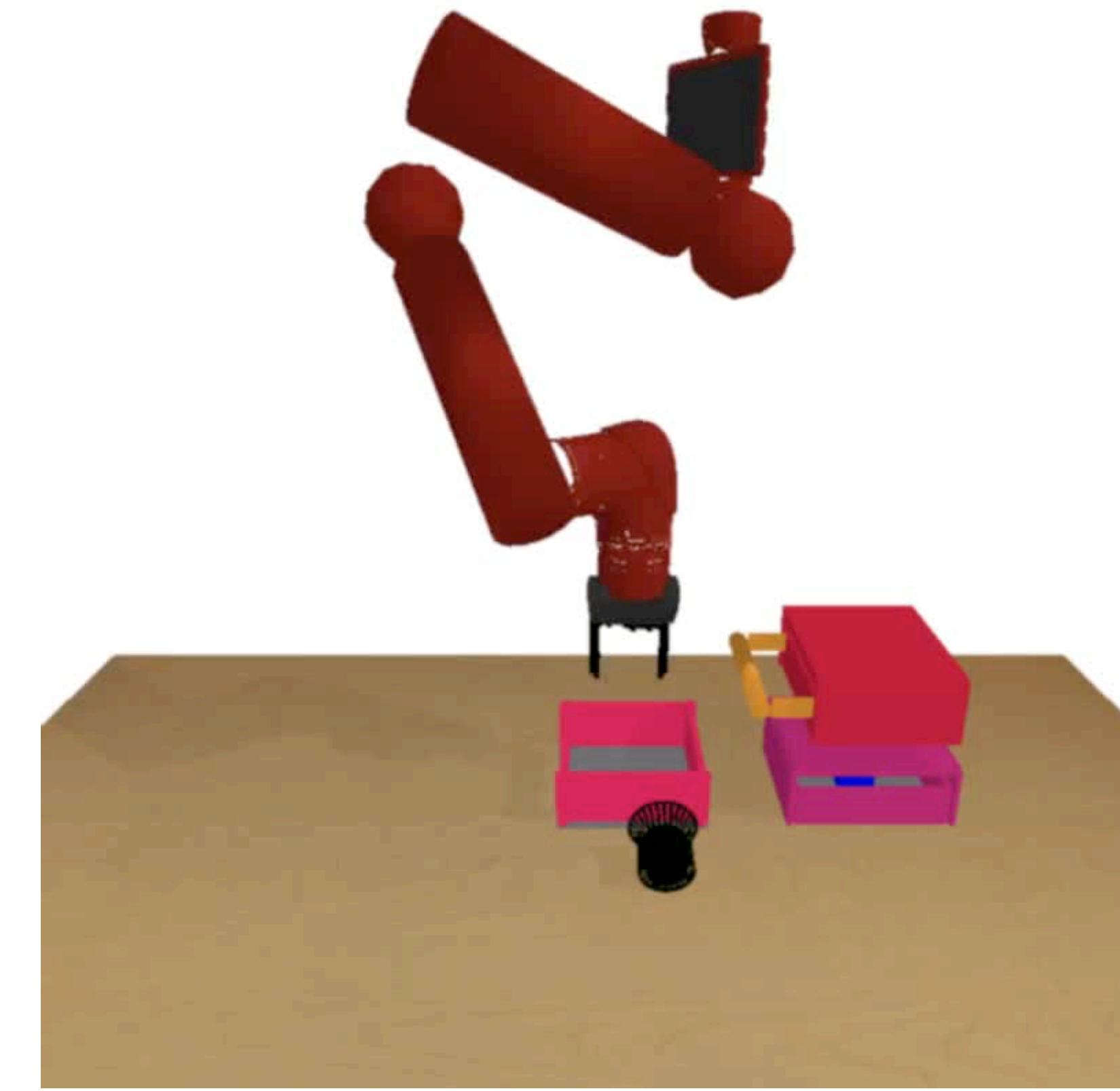
Iteration 0 (after offline training)



Meta test episode 1  
Meta test episode 2



Iteration 25 (after 150K steps of online training)

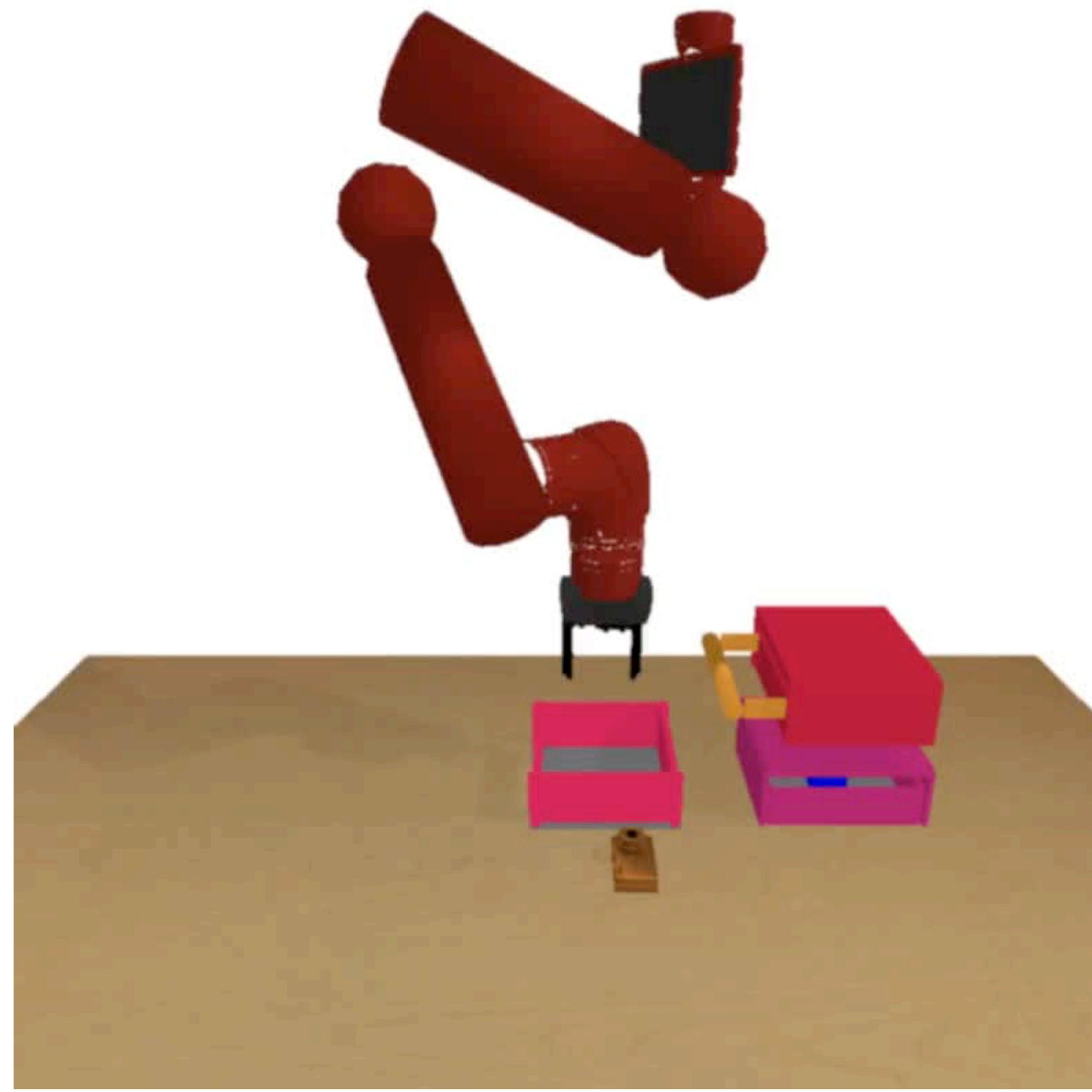


Meta test episode 1  
Meta test episode 2



# SMAC meta-train improves skills

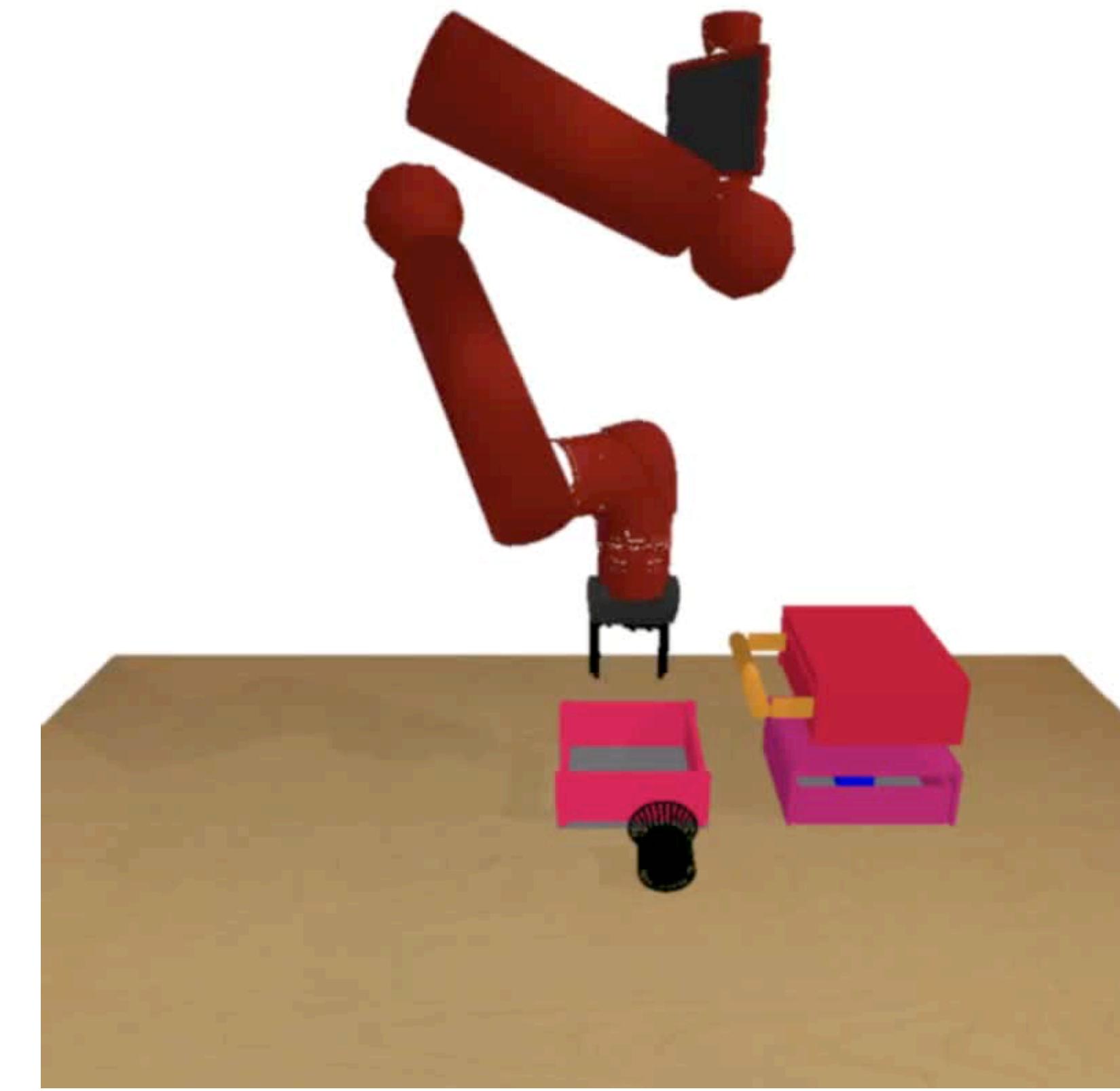
Iteration 0 (after offline training)



Meta test episode 1  
Meta test episode 2



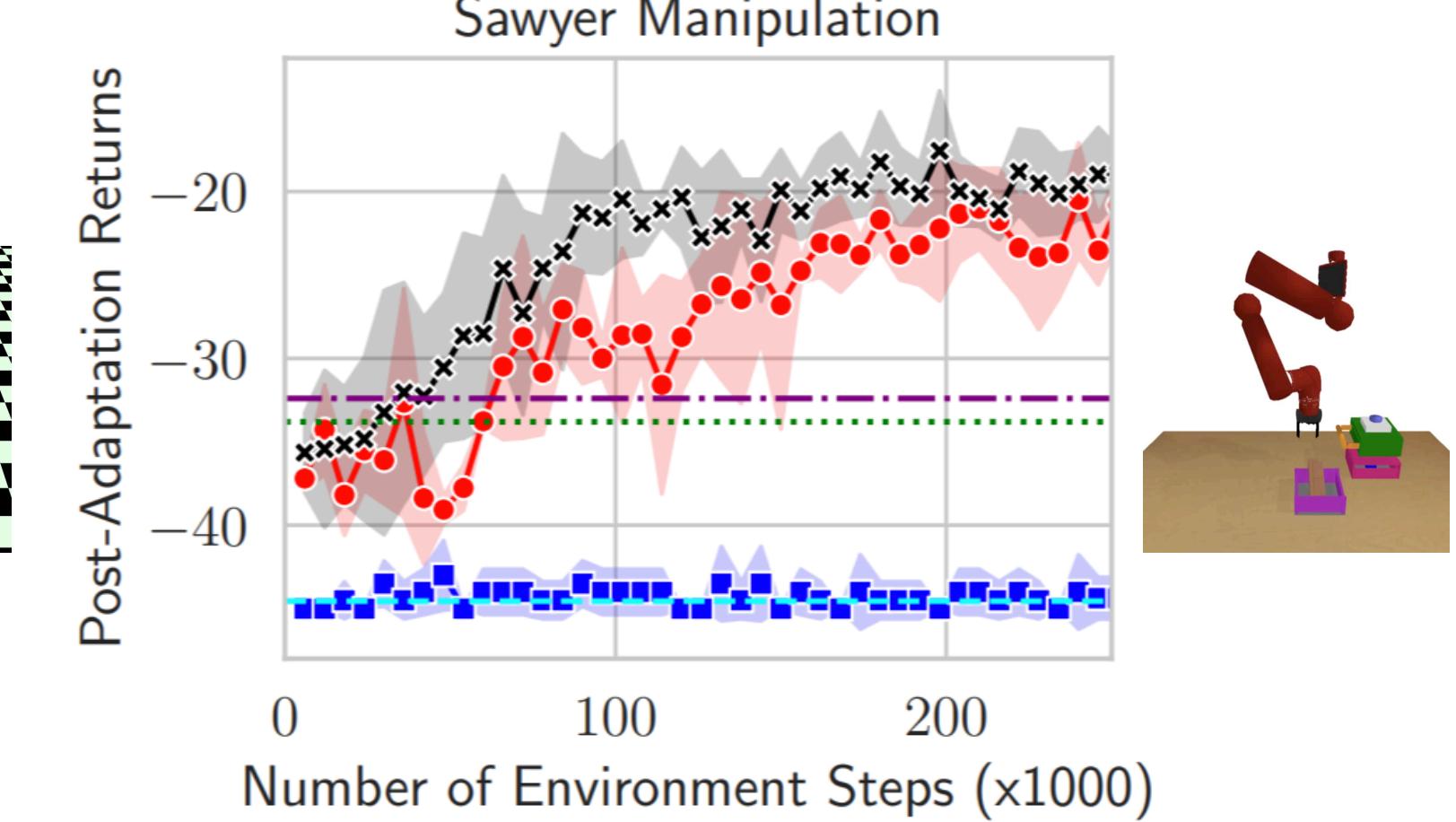
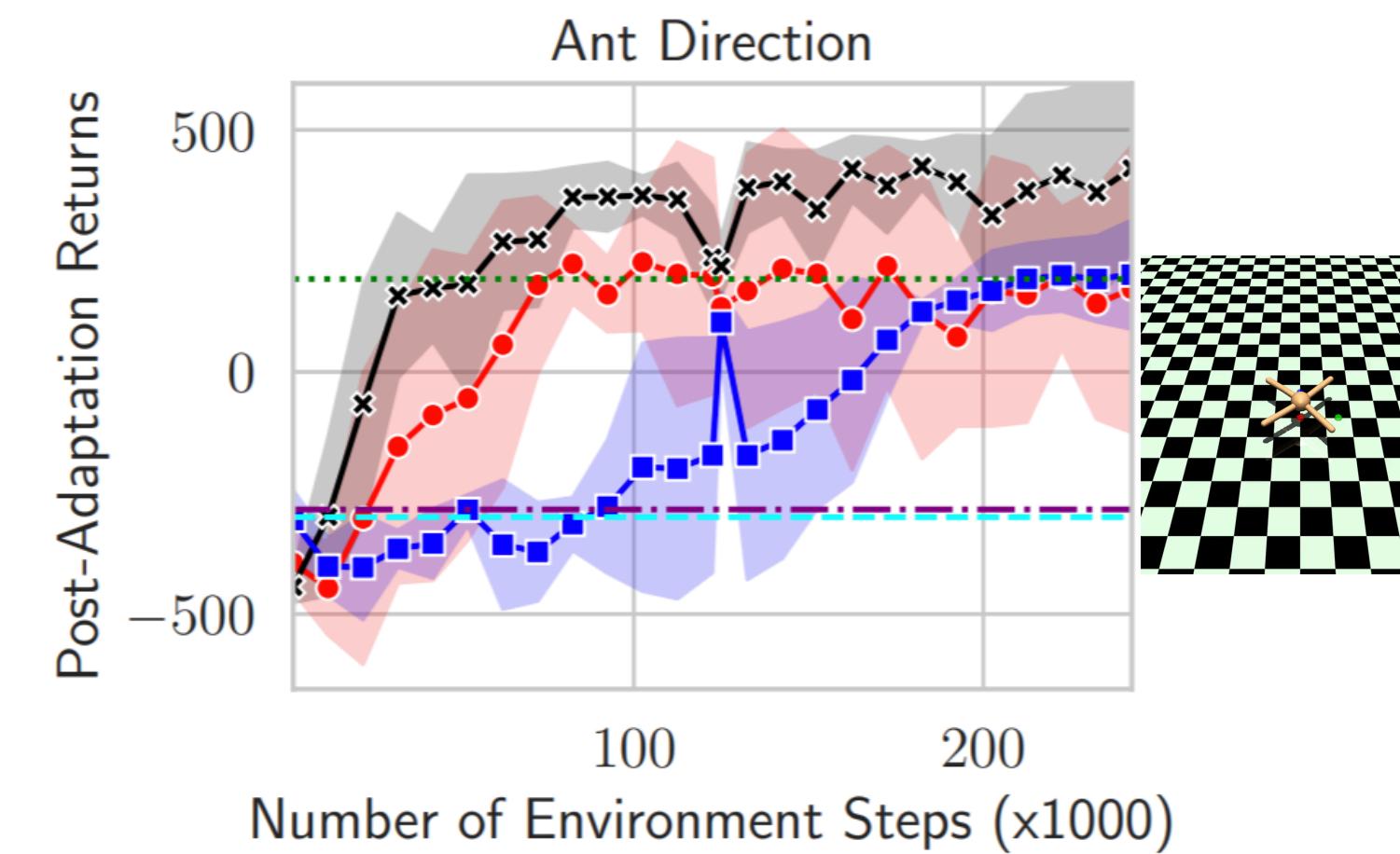
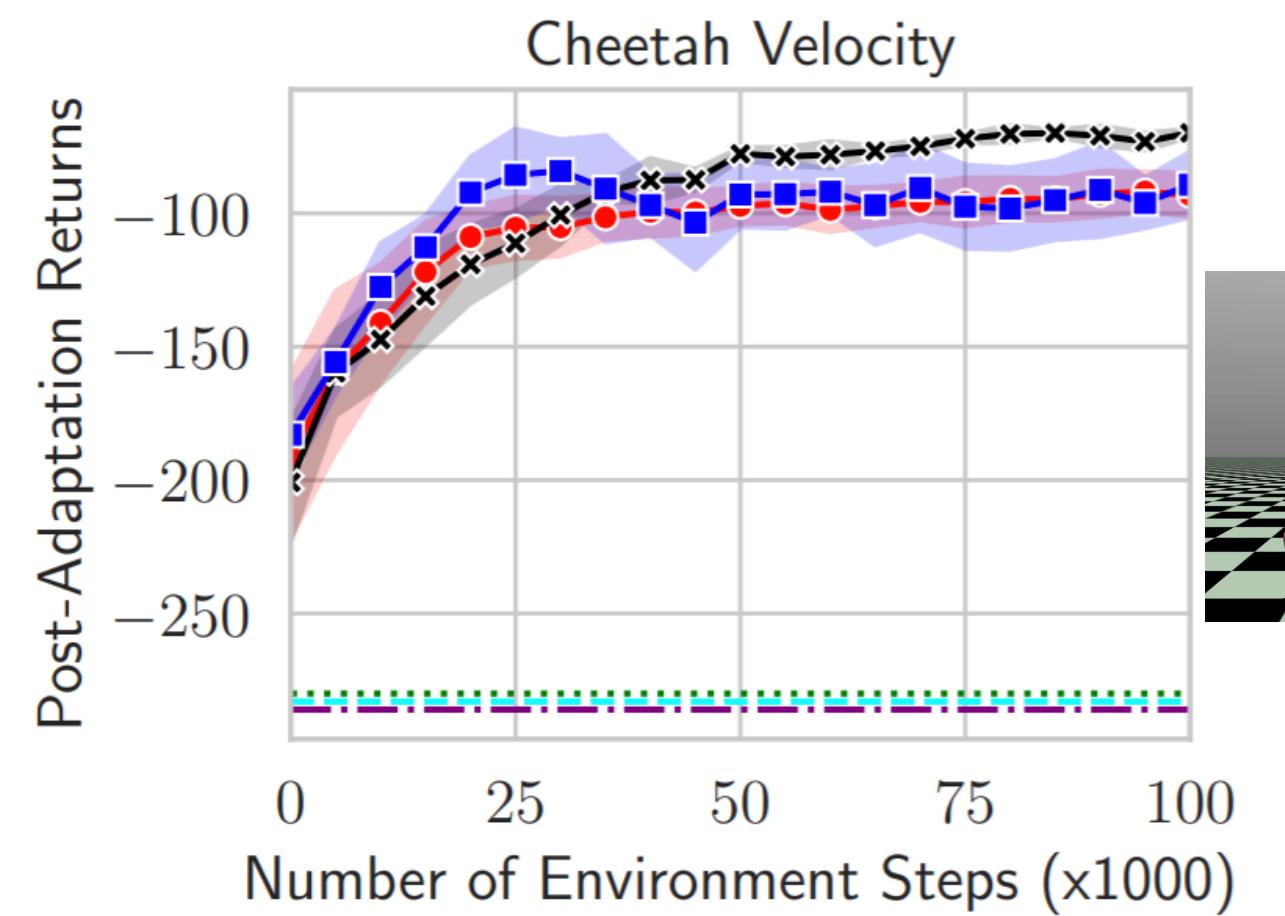
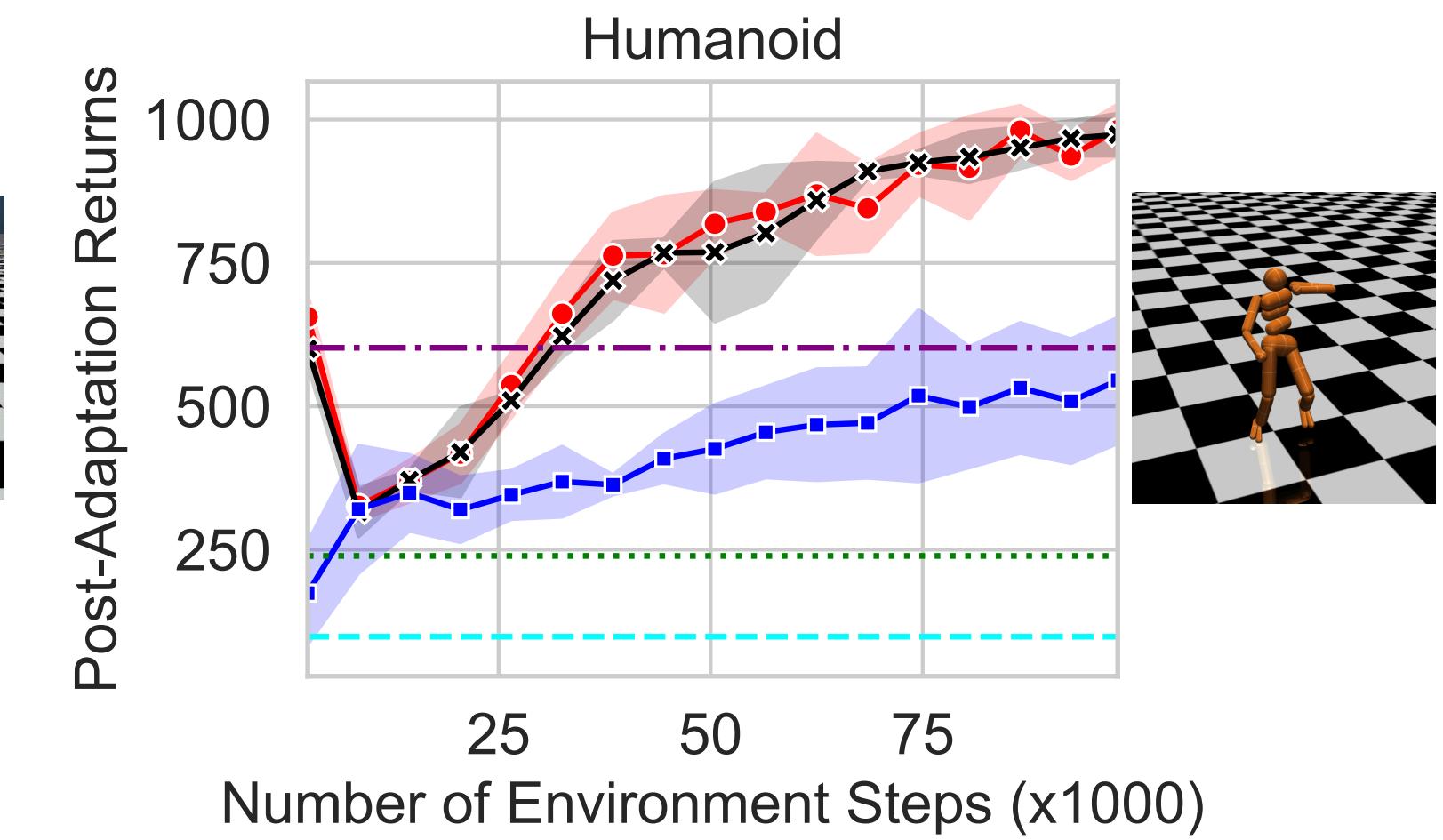
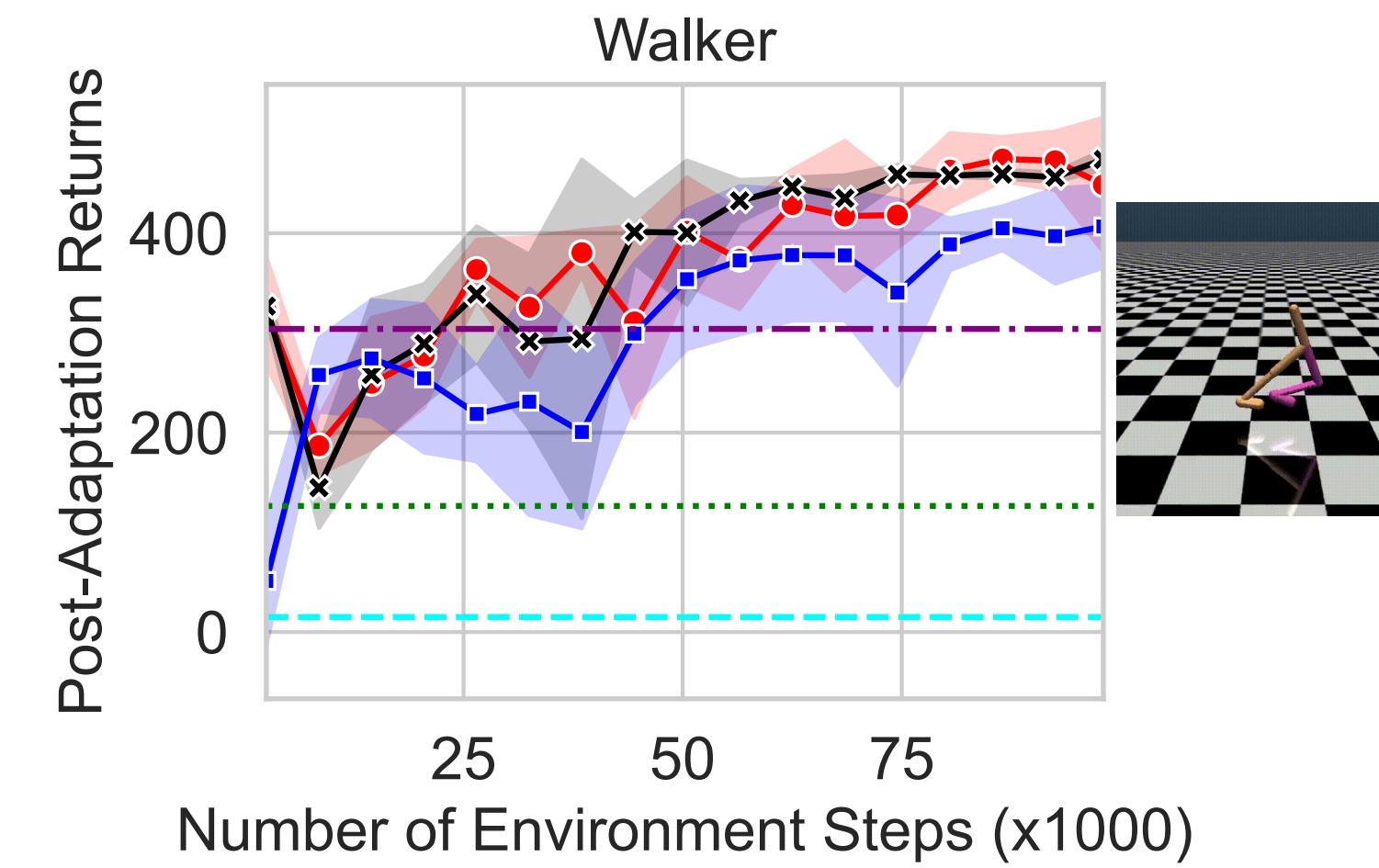
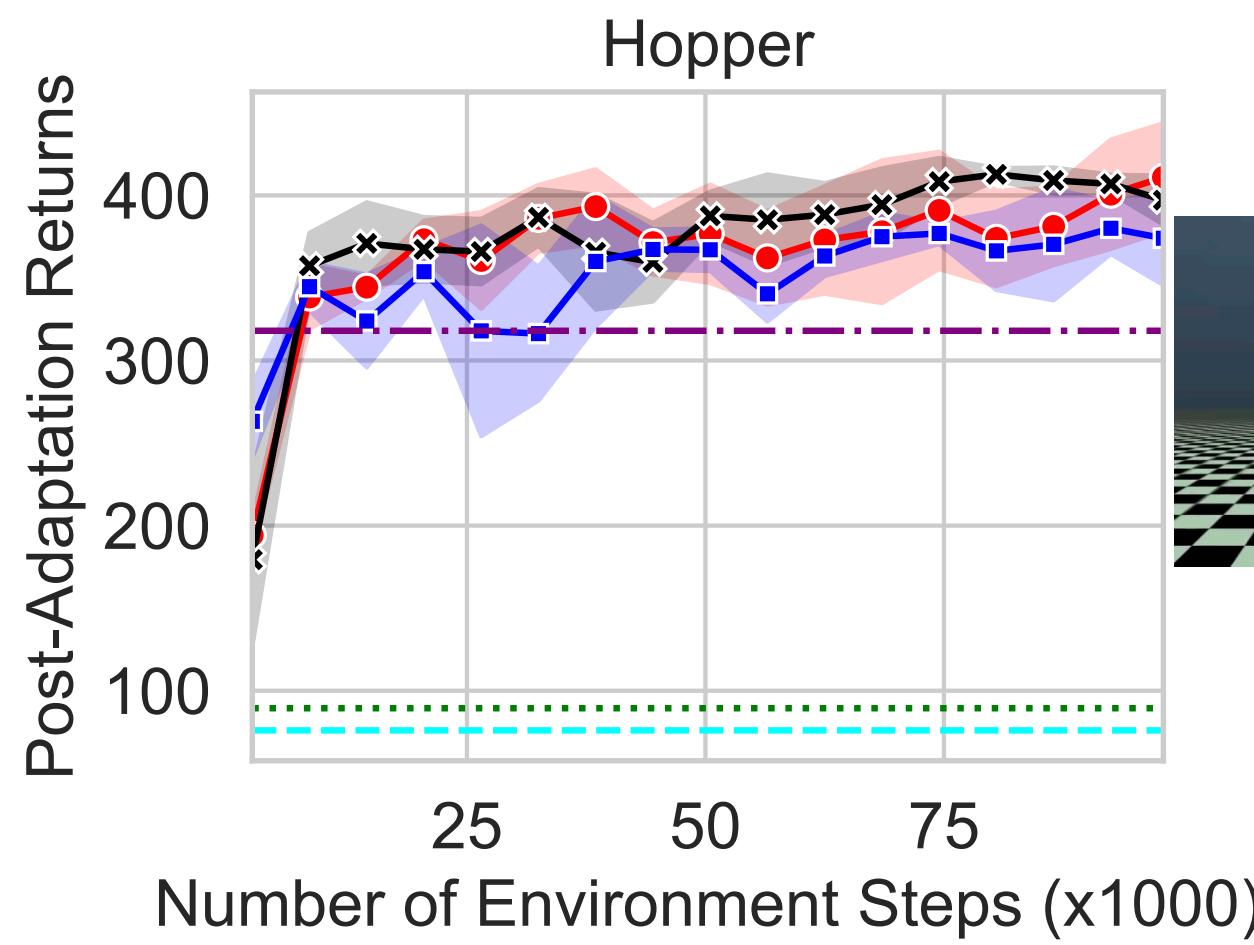
Iteration 25 (after 150K steps of online training)



Meta test episode 1  
Meta test episode 2

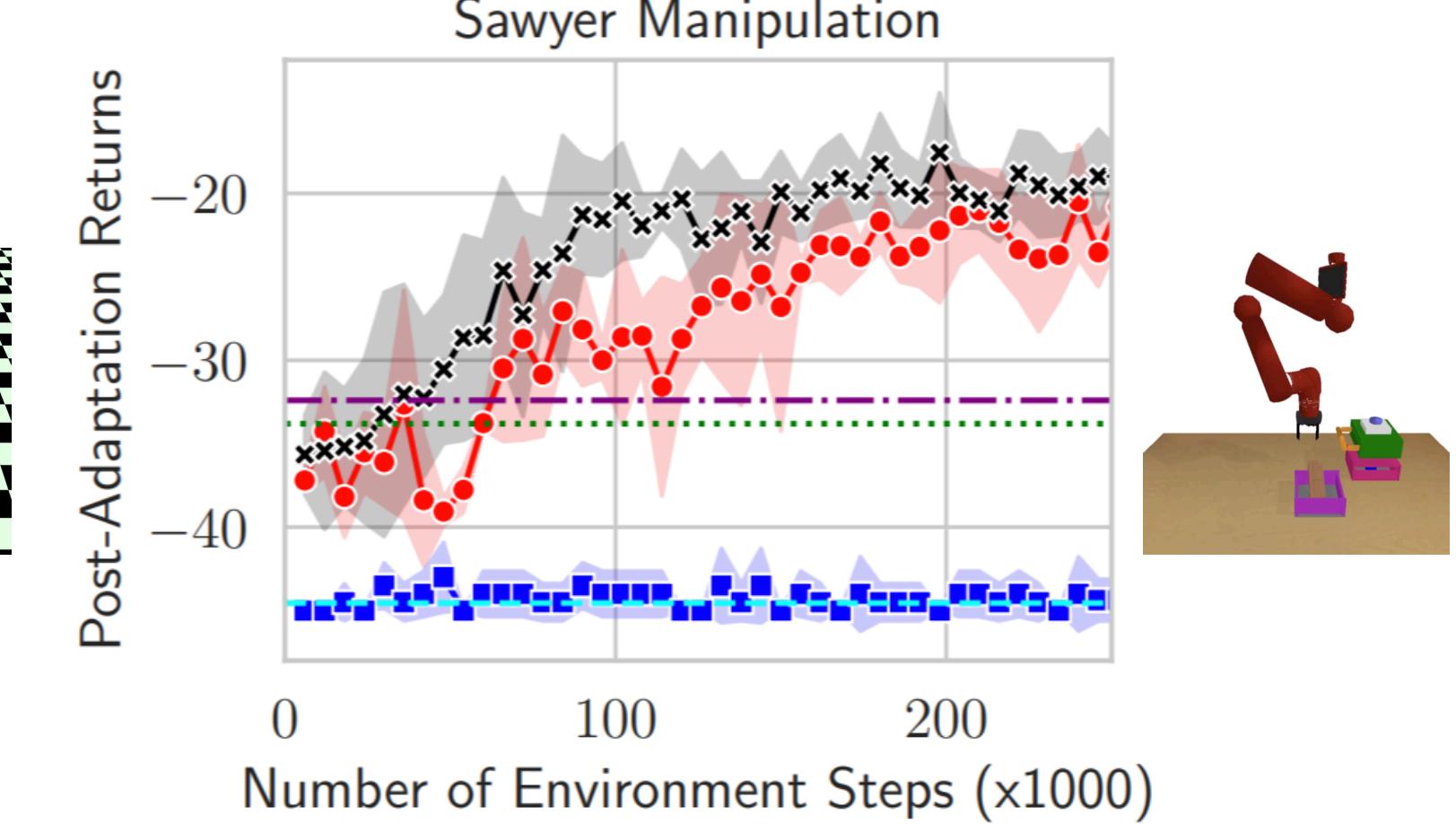
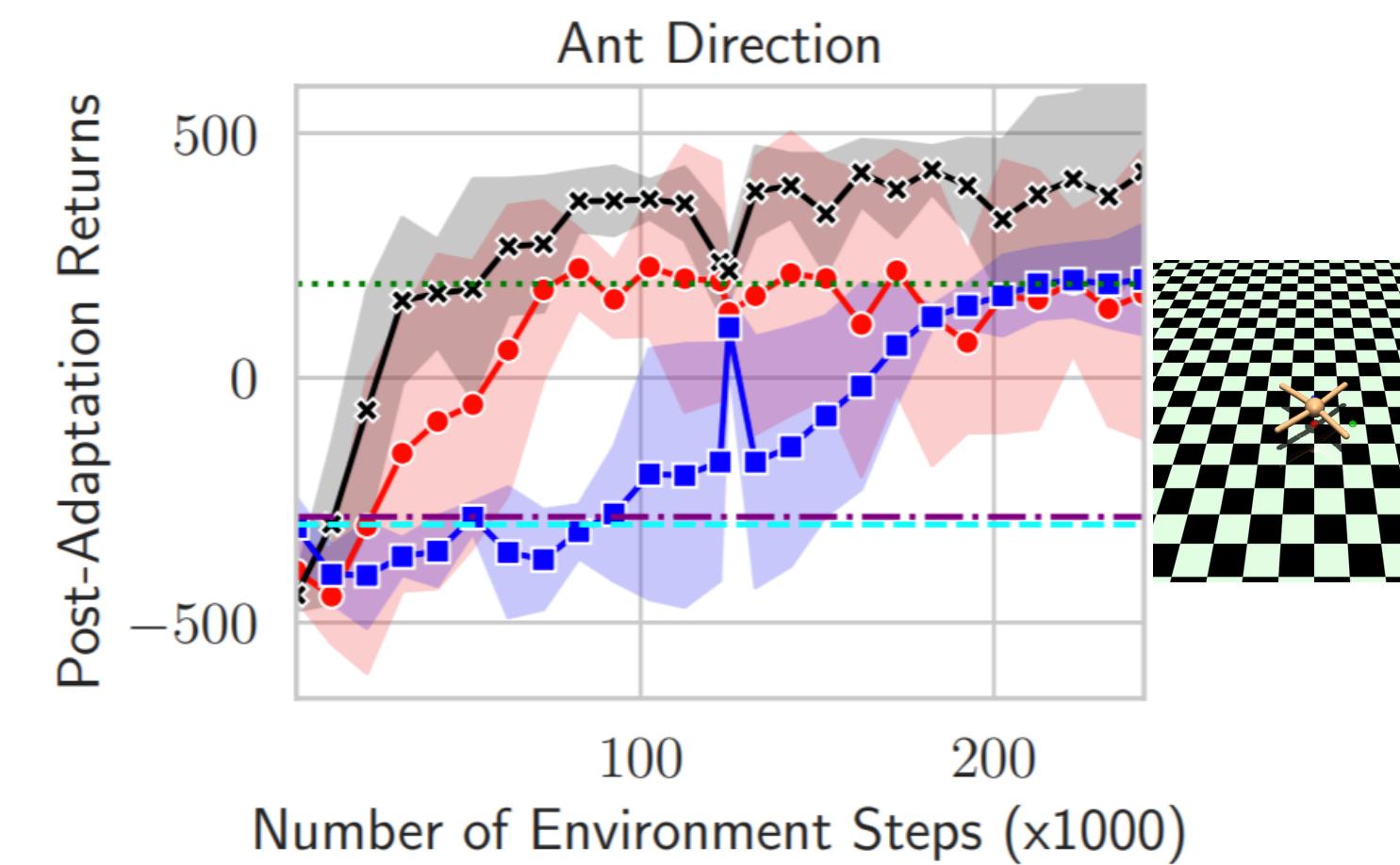
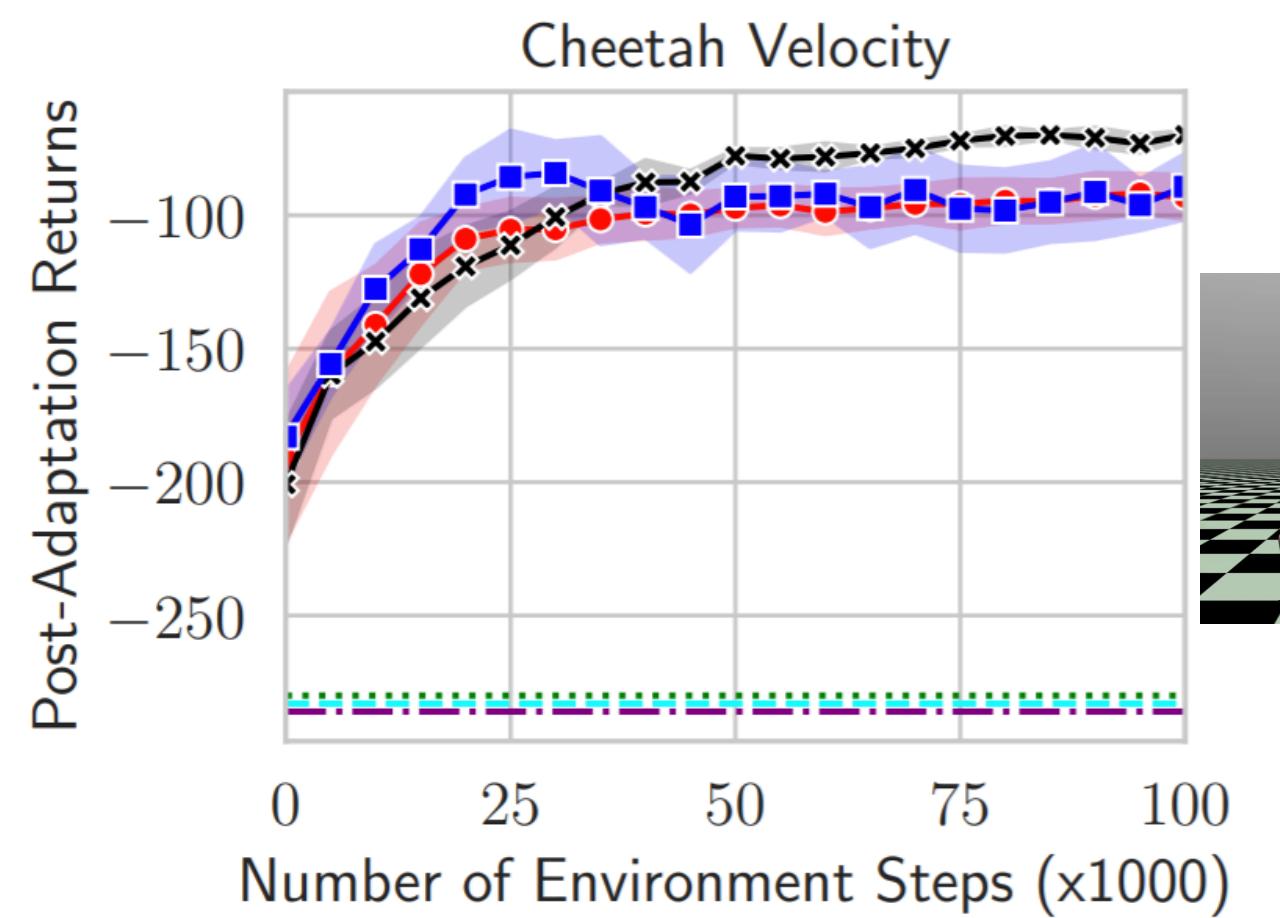
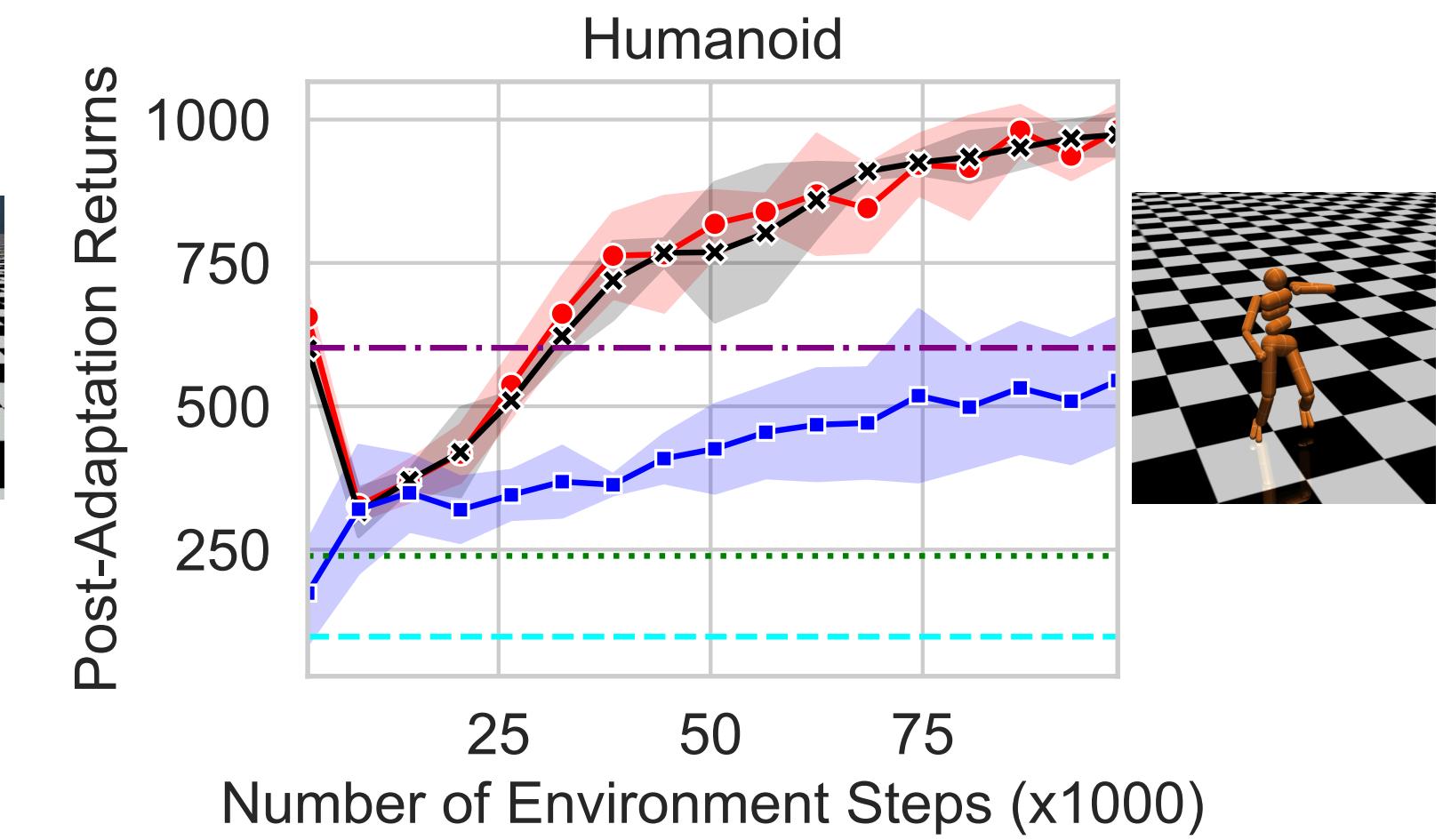
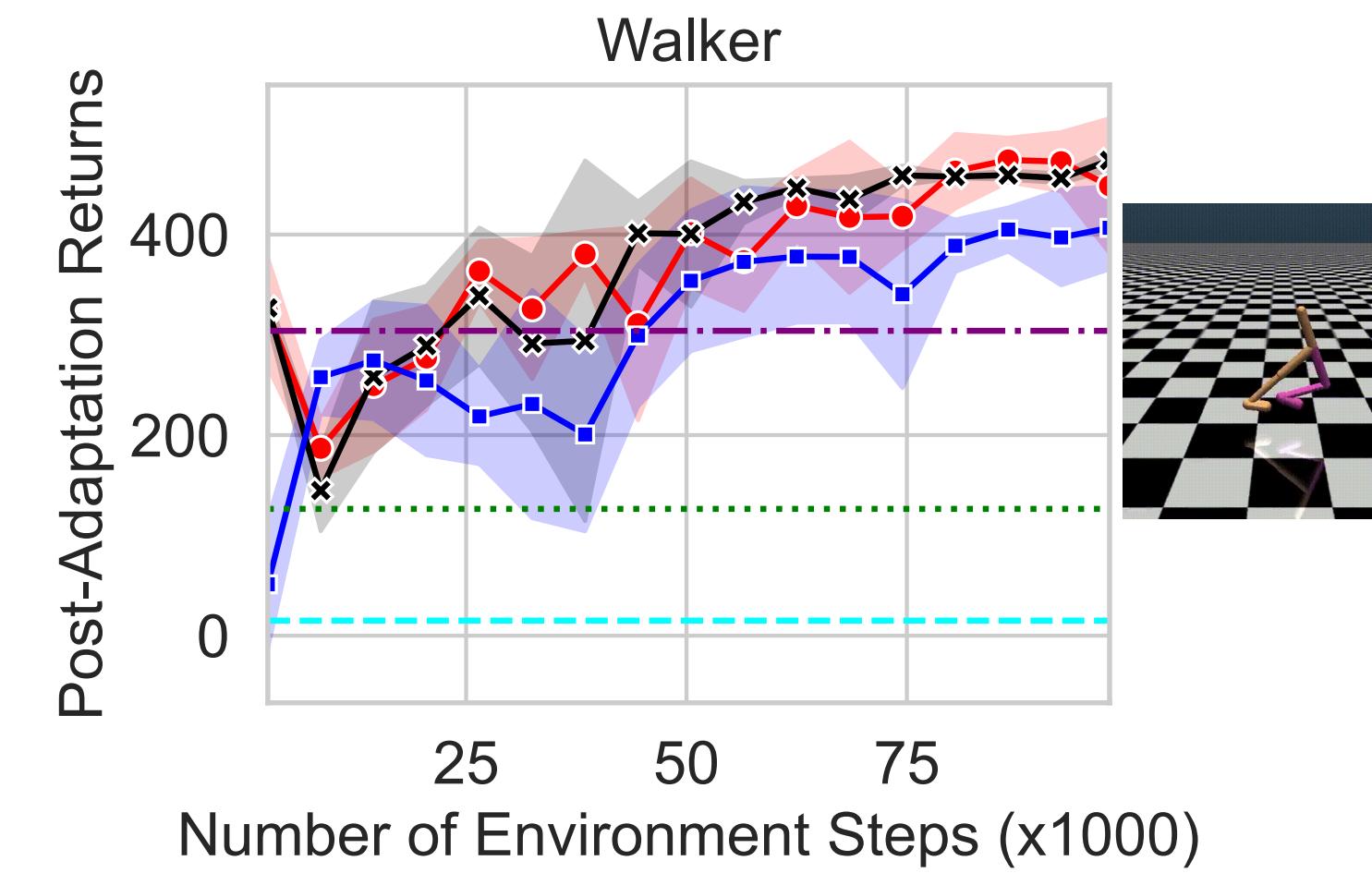
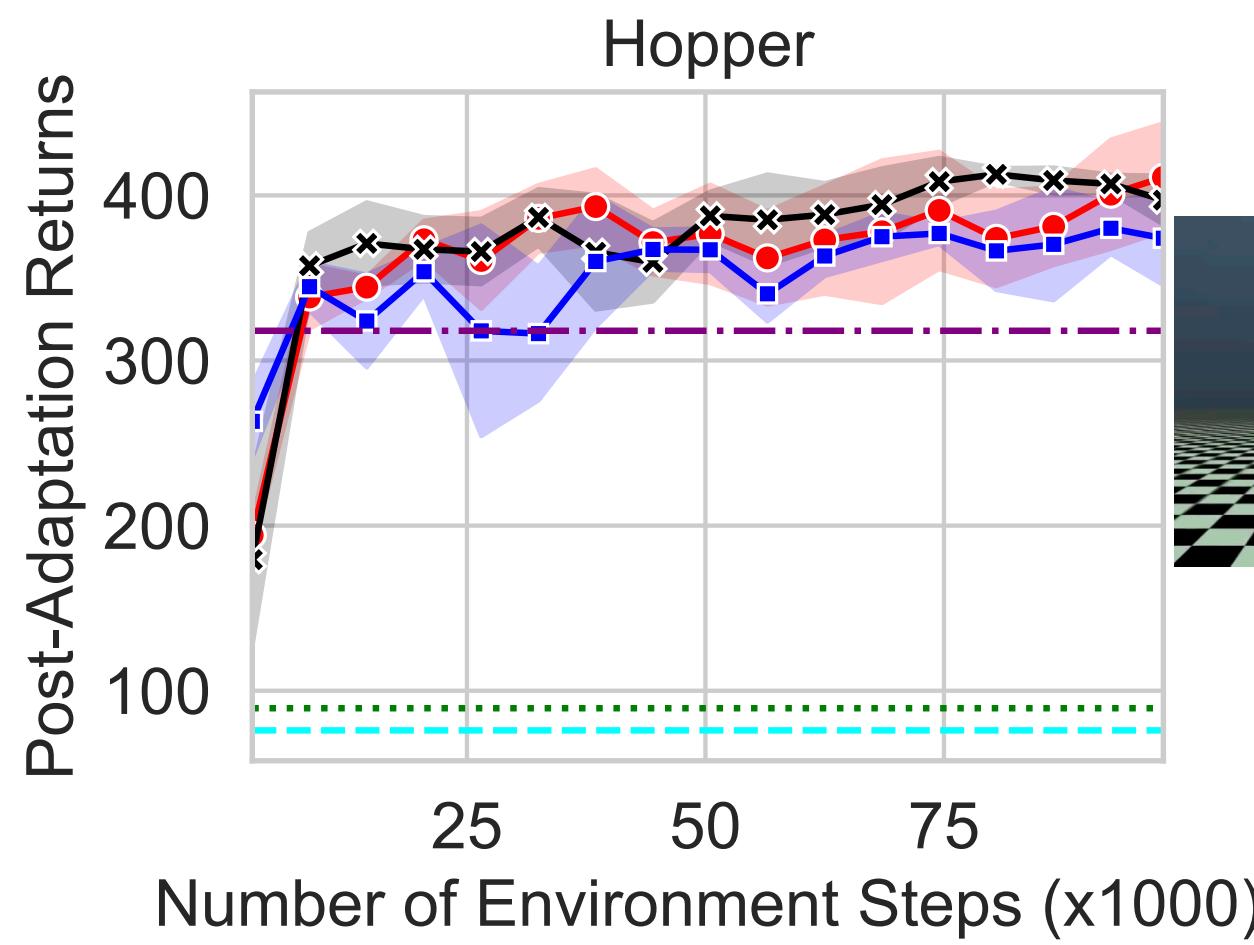


# SMAC performs well across many environments



—●— SMAC (Ours)   —◆— Online Oracle   —■— SMAC (actor ablation)   ······ MACAW   -··- BOReL   -··- meta behavior cloning

# SMAC performs well across many environments

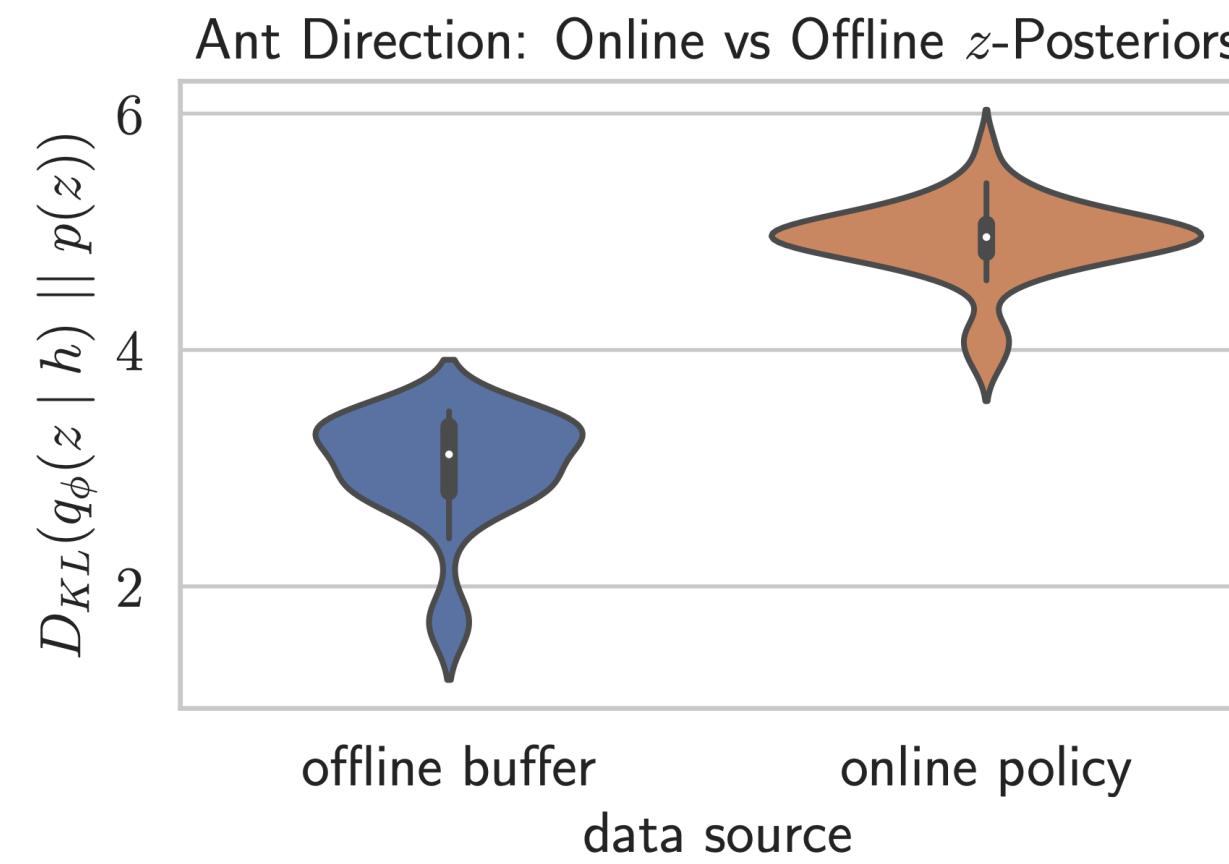


—●— SMAC (Ours)   —◆— Online Oracle   —■— SMAC (actor ablation)   ······ MACAW   ······ BOReL   ······ meta behavior cloning

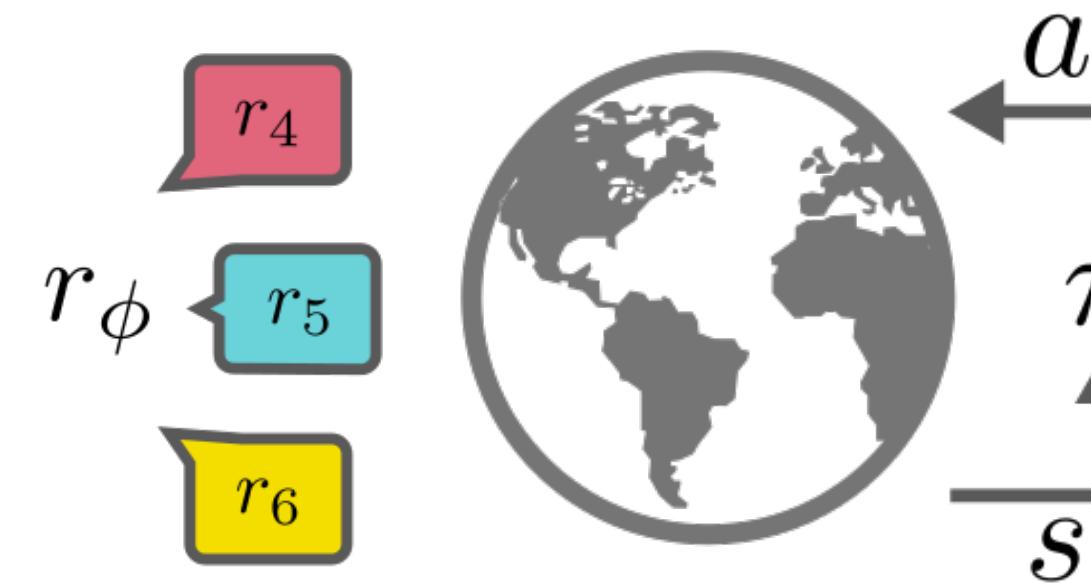
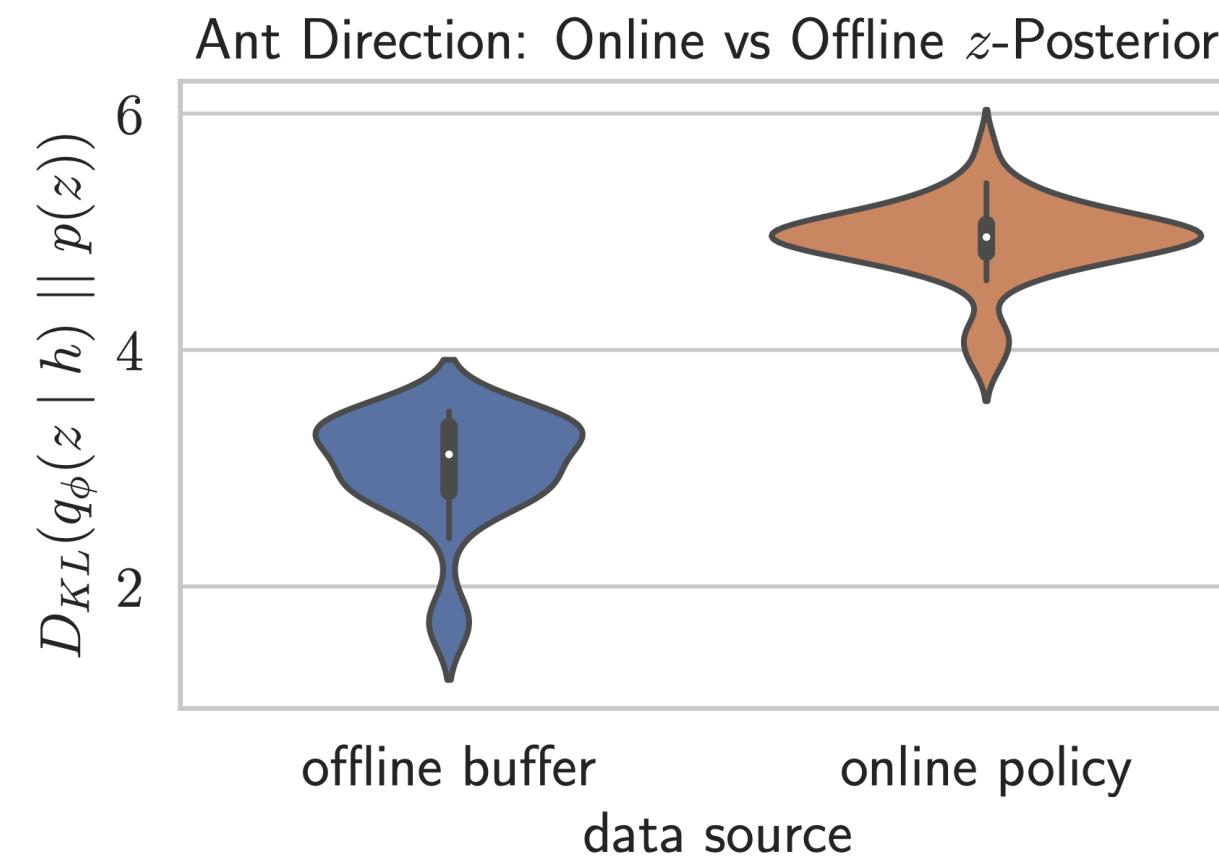
# Conclusion

Offline Meta-Reinforcement Learning with Online Self-Supervision.  
Vitchyr H. Pong, Ashvin Nair, Laura Smith, Catherine Huang, Sergey Levine.

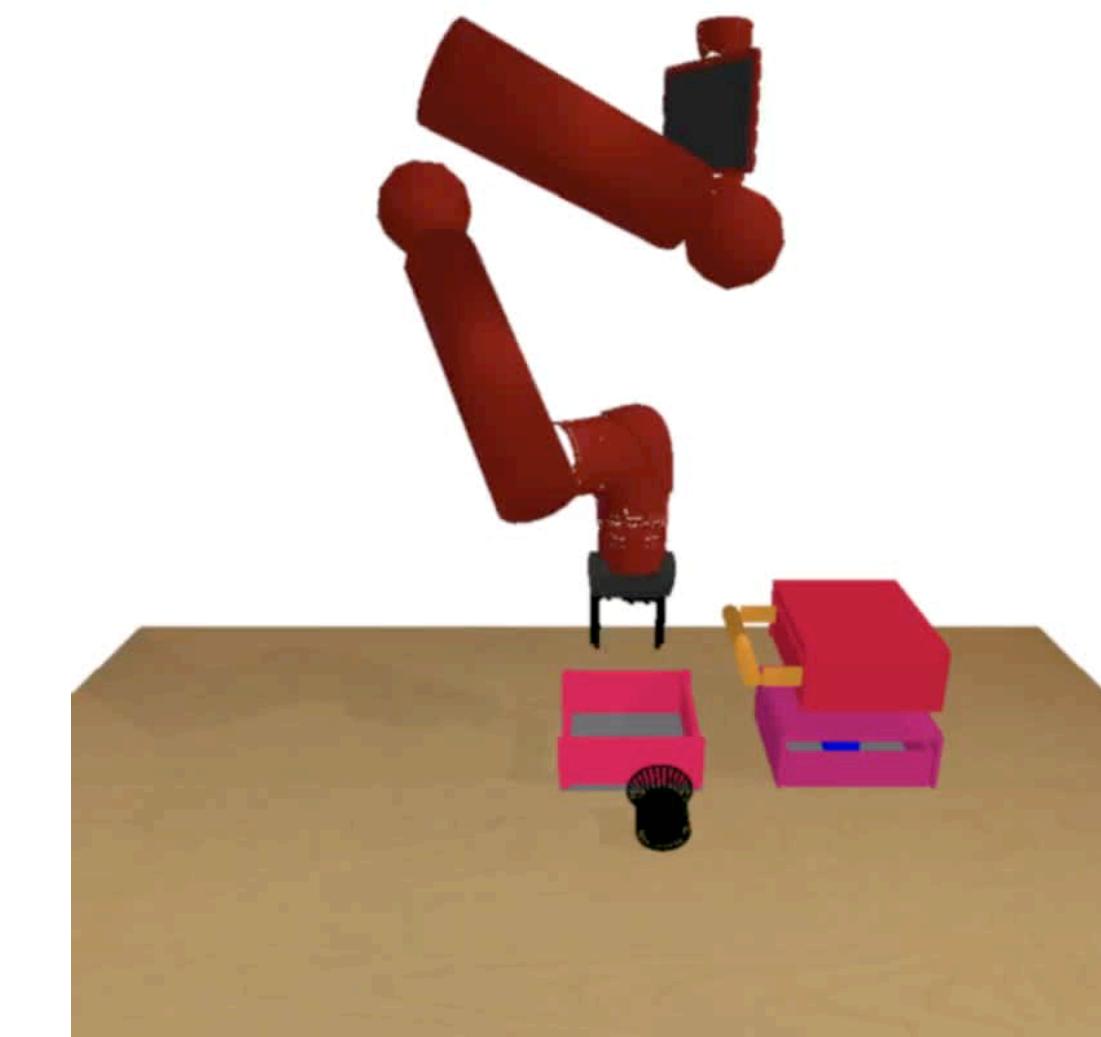
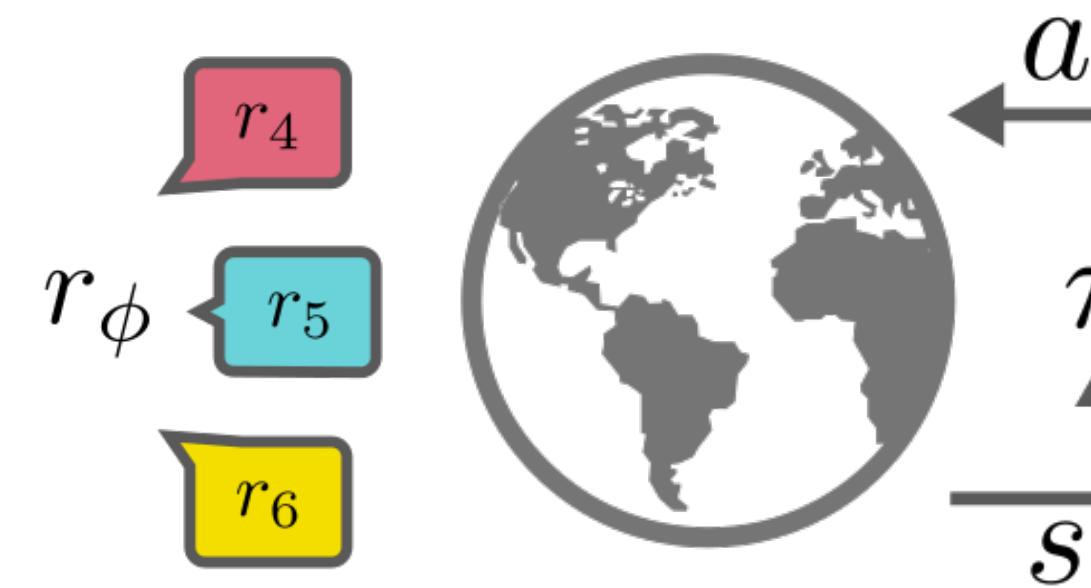
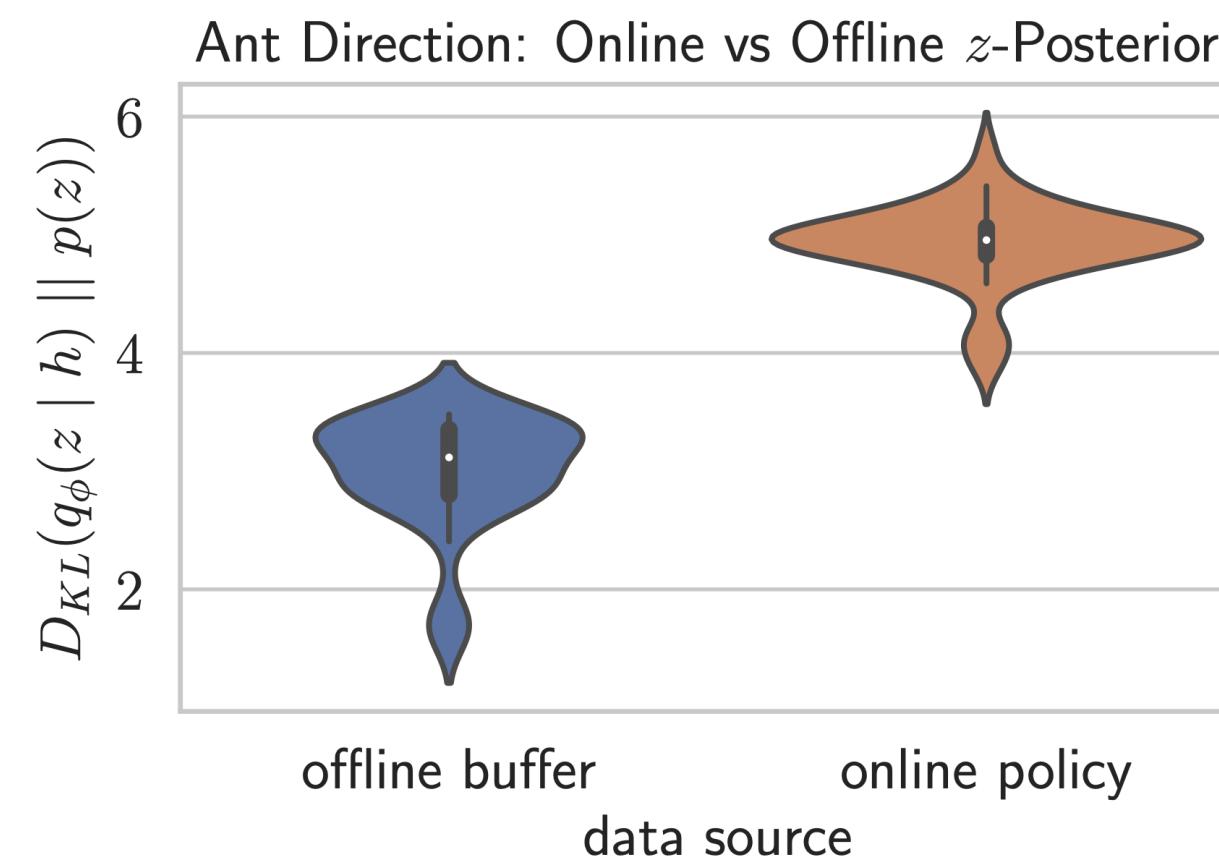
# Conclusion



# Conclusion



# Conclusion



Offline Meta-Reinforcement Learning with Online Self-Supervision.  
Vitchyr H. Pong, Ashvin Nair, Laura Smith, Catherine Huang, Sergey Levine.