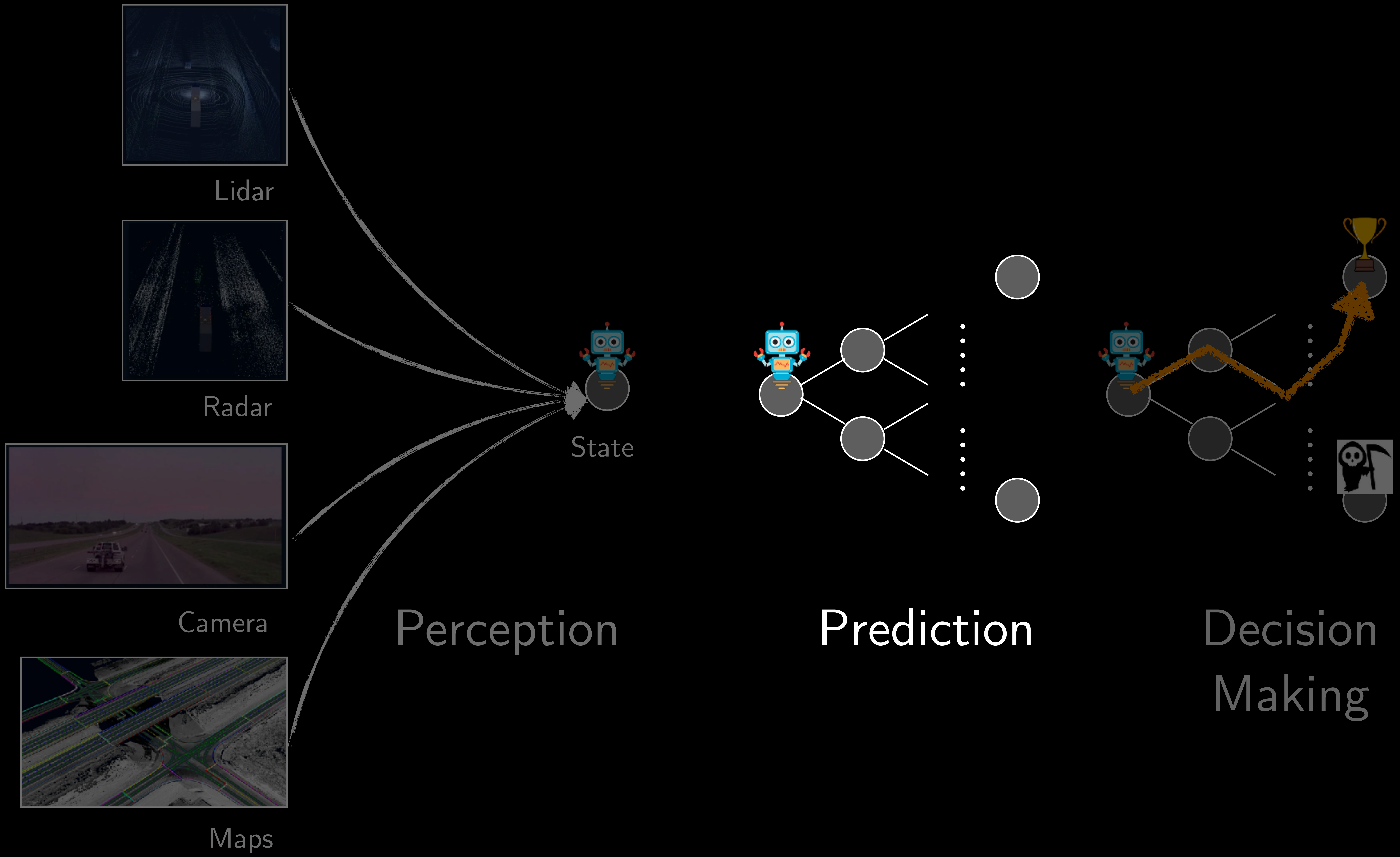
Generative World Models: The Dreamer Models

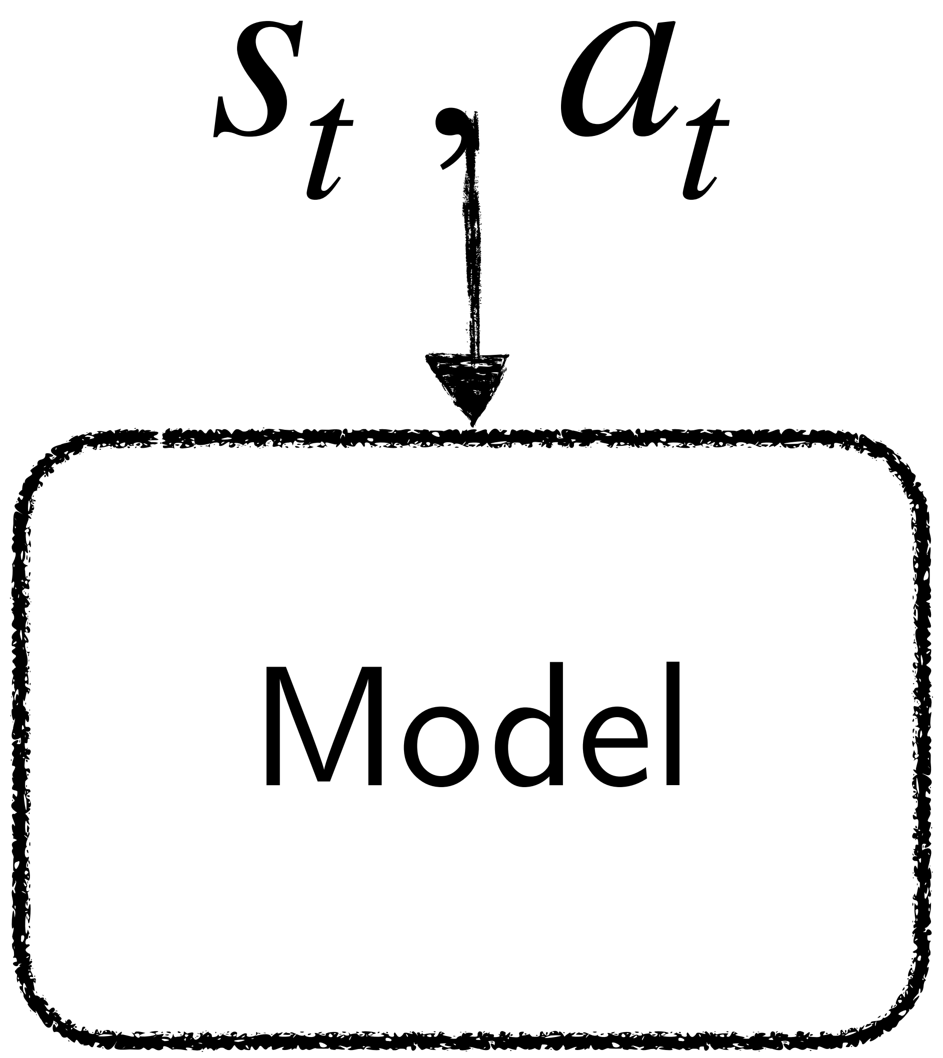
Sanjiban Choudhury

Cornell Bowers CIS Computer Science


Lidar
Radar
Camera
Maps
State
Perception
Prediction
Decision
Making

Models.

What is a model?



What is a model?

s_t , a_t

$\downarrow$

Model

$\downarrow$

s_{t+1}

What is a model?

$$P_\theta(s_{t+1} \mid s_t, a_t)$$

Why Model?

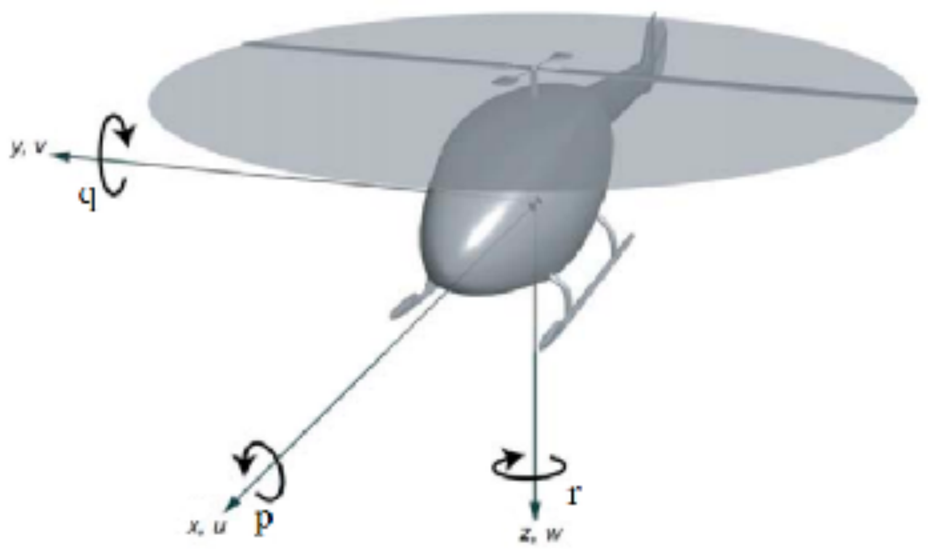
Models are *necessary*

Robots can't just try out random actions in the world!

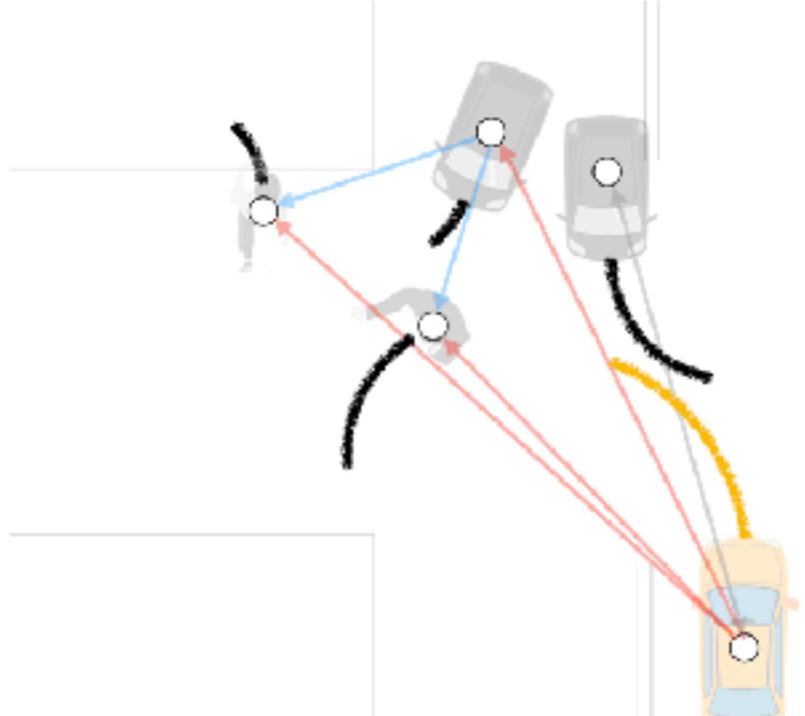
Learning Models.

(Early work in Model Based RL by Pieter Abeel et al. 2010
https://people.eecs.berkeley.edu/~pabbeel/autonomous_helicopter.html)

Models: From Simple to Complex



Physics Models

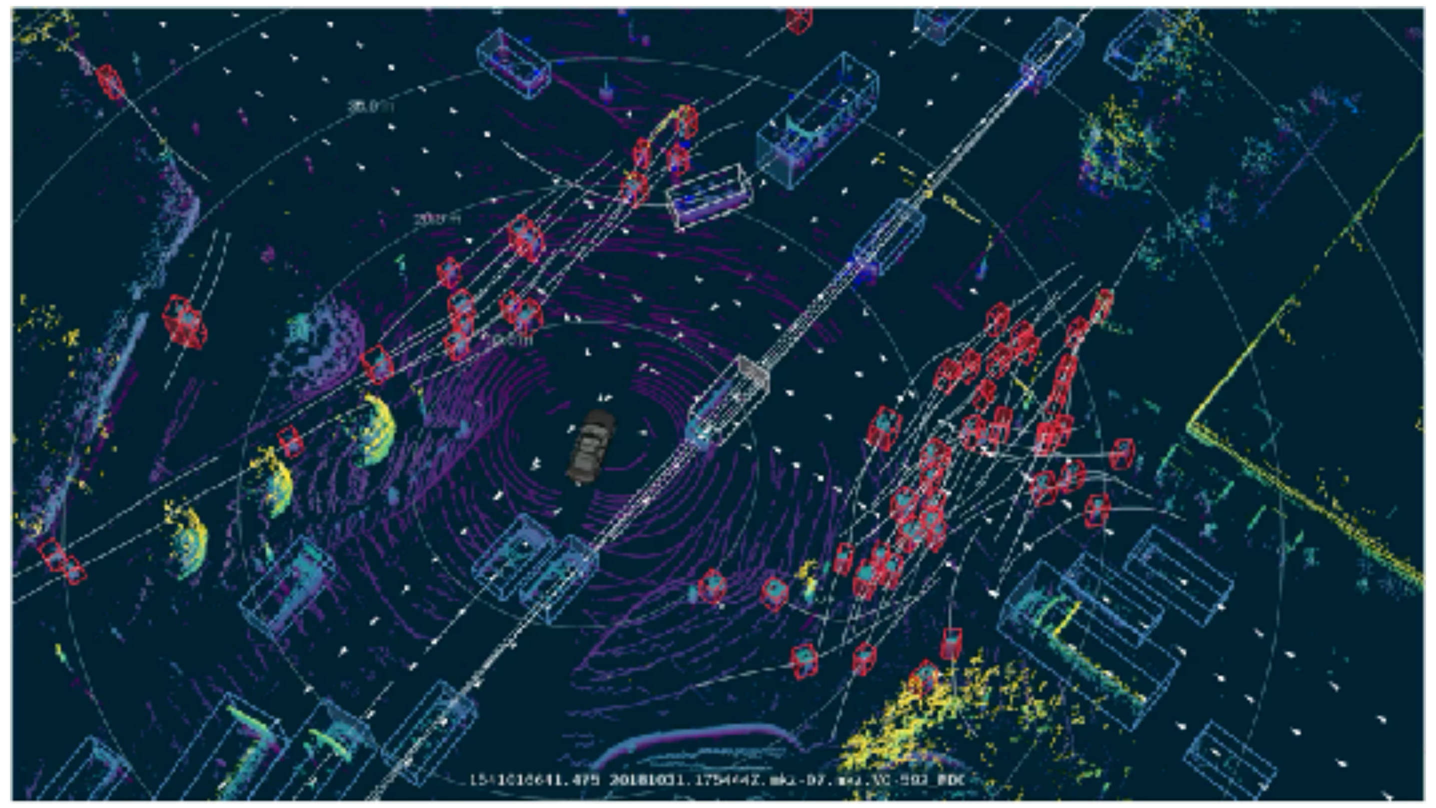


Human Models

Open World Models

Simple

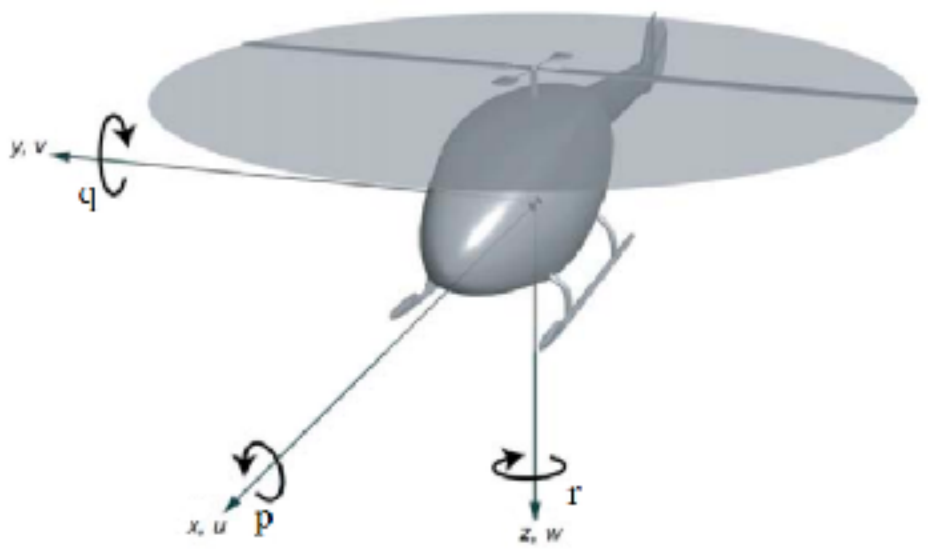
Complex



Models: From Simple to Complex

Simple Complex

Models: From Simple to Complex



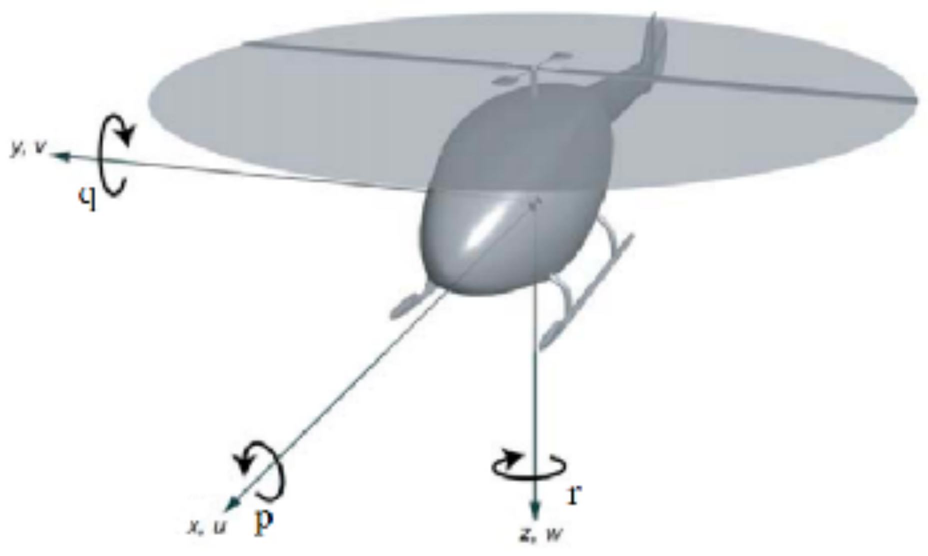
Physics Models

Simple

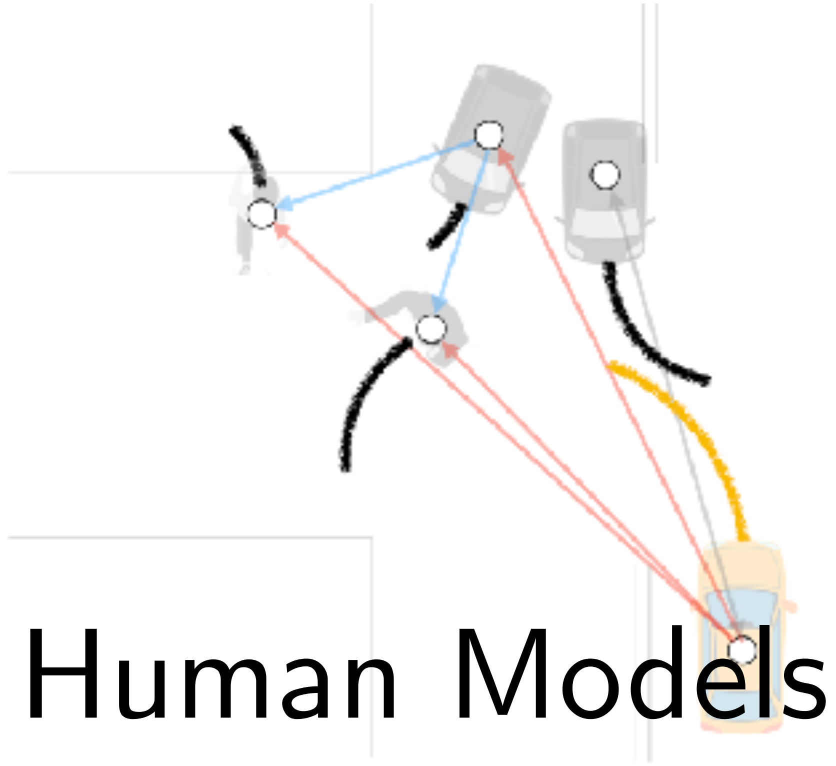
Know state

Strong prior
on dynamics

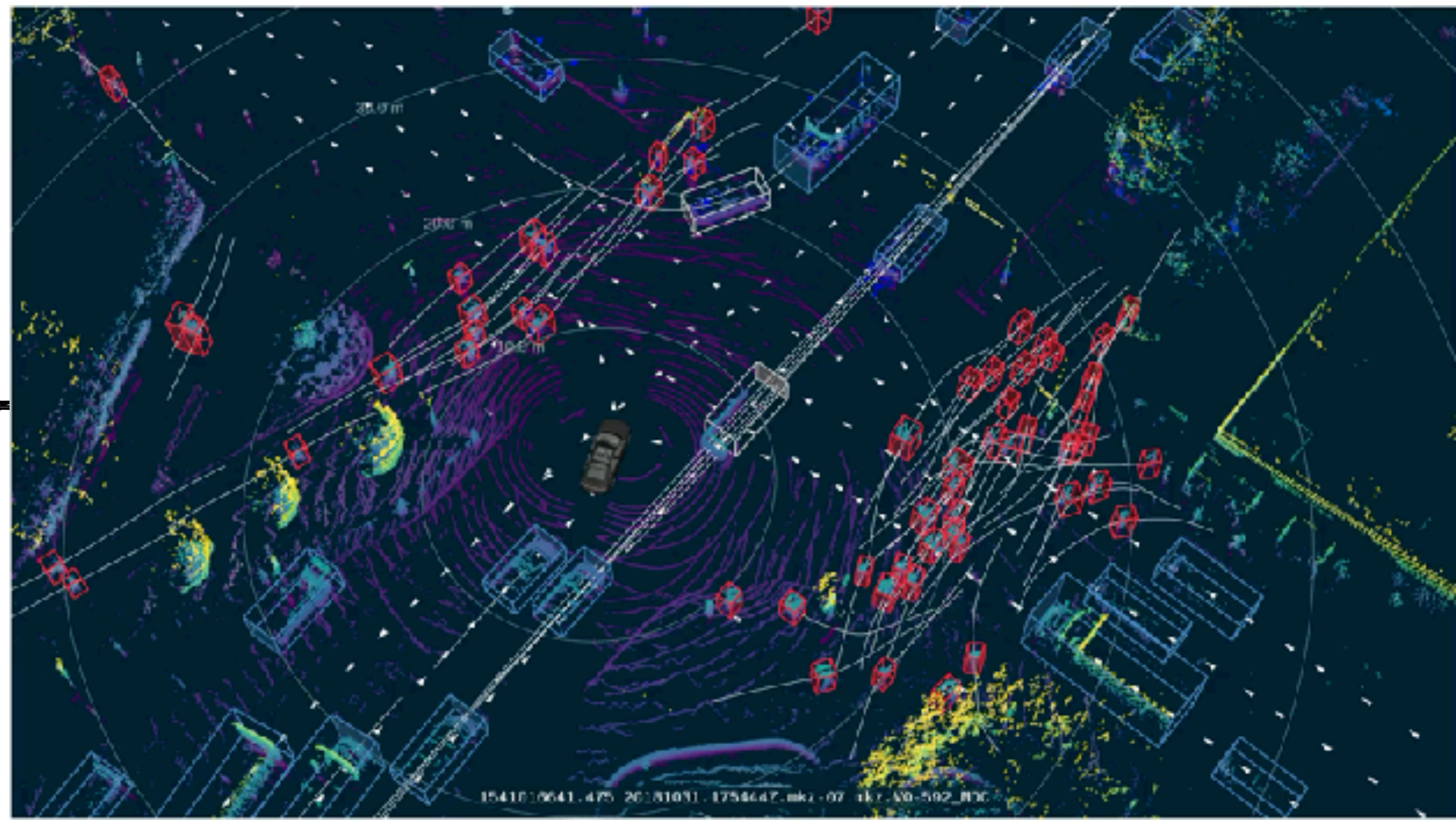
Models: From Simple to Complex



Physics Models



Human Models



Simple

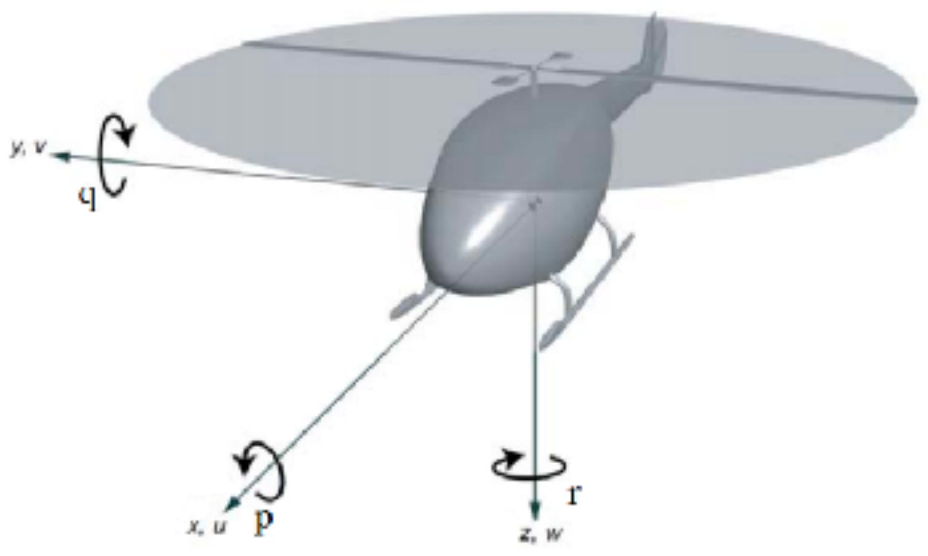
Know state

Strong prior
on dynamics

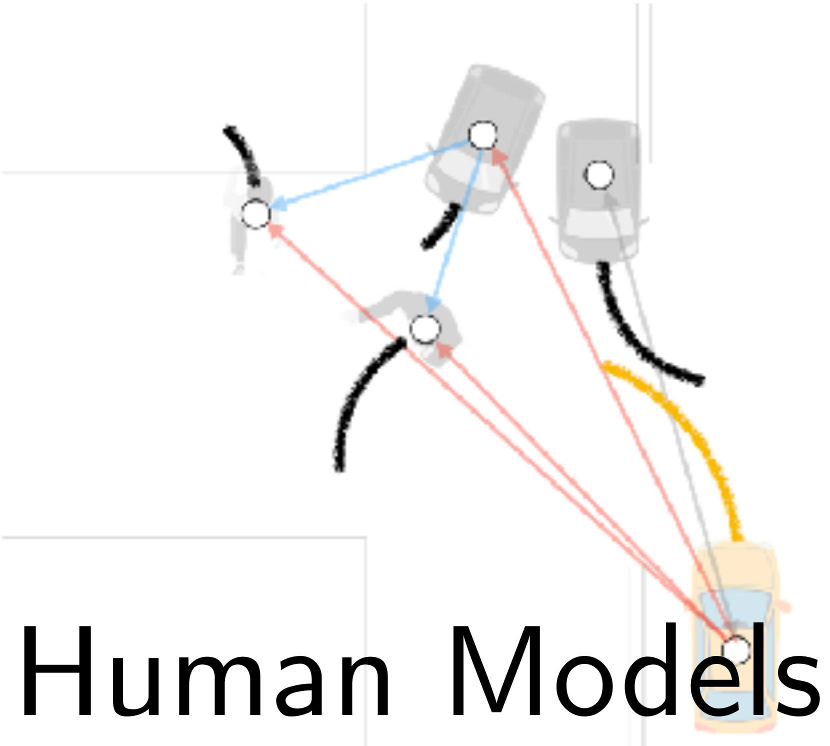
Know state

Unknown
dynamics

Models: From Simple to Complex



Physics Models



Human Models

Open World Models

Simple → Complex

Physics Models	Human Models	Open World Models
Know state	Know state	Unknown state
Strong prior on dynamics	Unknown dynamics	Unknown dynamics

Activity!

Modelling Tamago Sushi

Think-Pair-Share!

Think (30 sec): How would you model making tamago sushi?

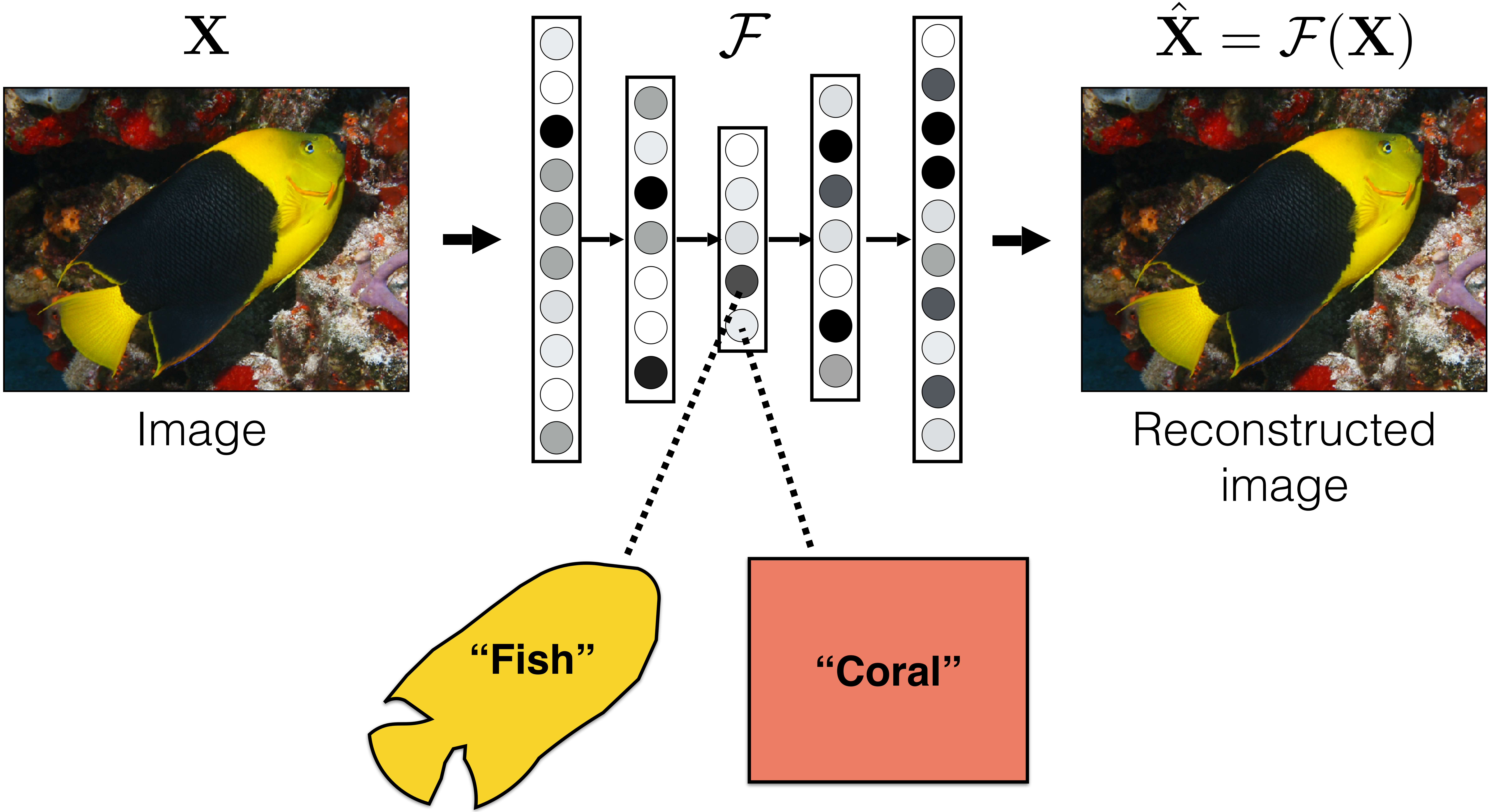
Pair: Find a partner

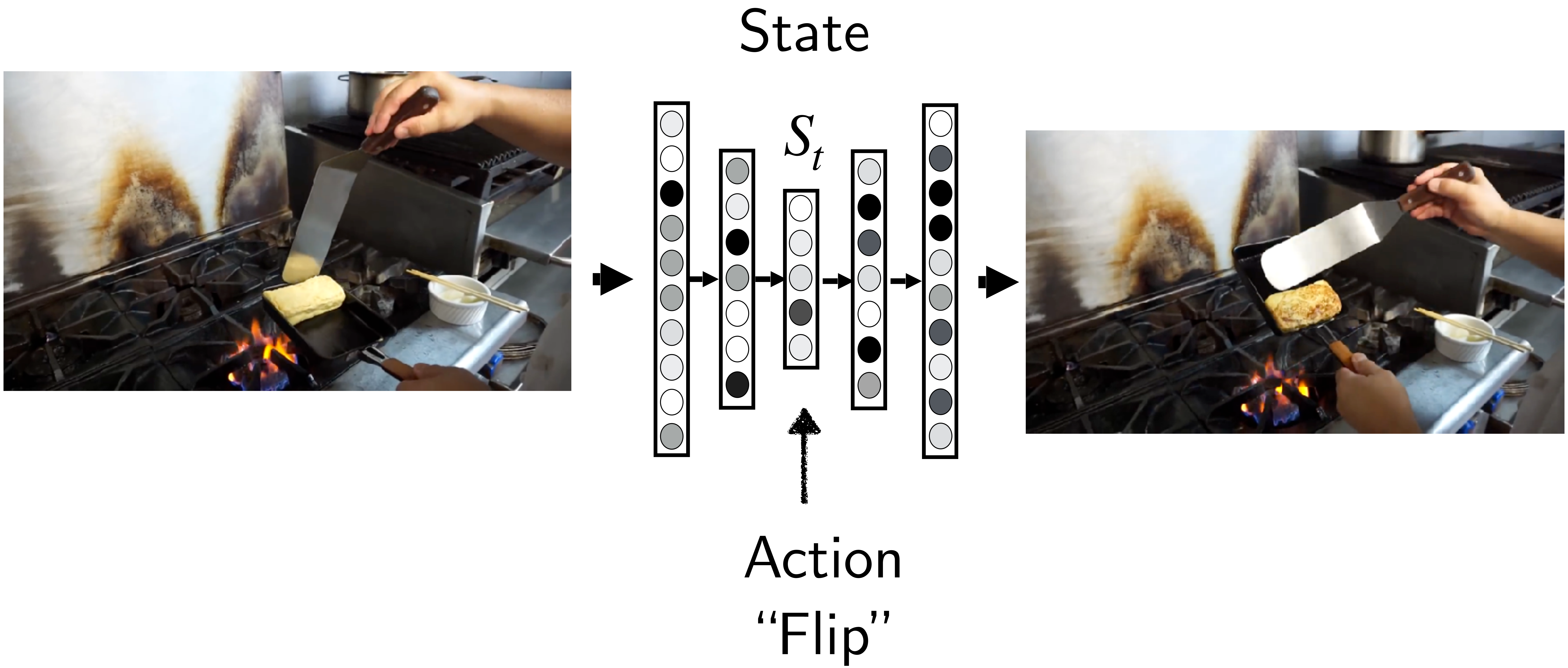
Share (45 sec): Partners exchange ideas

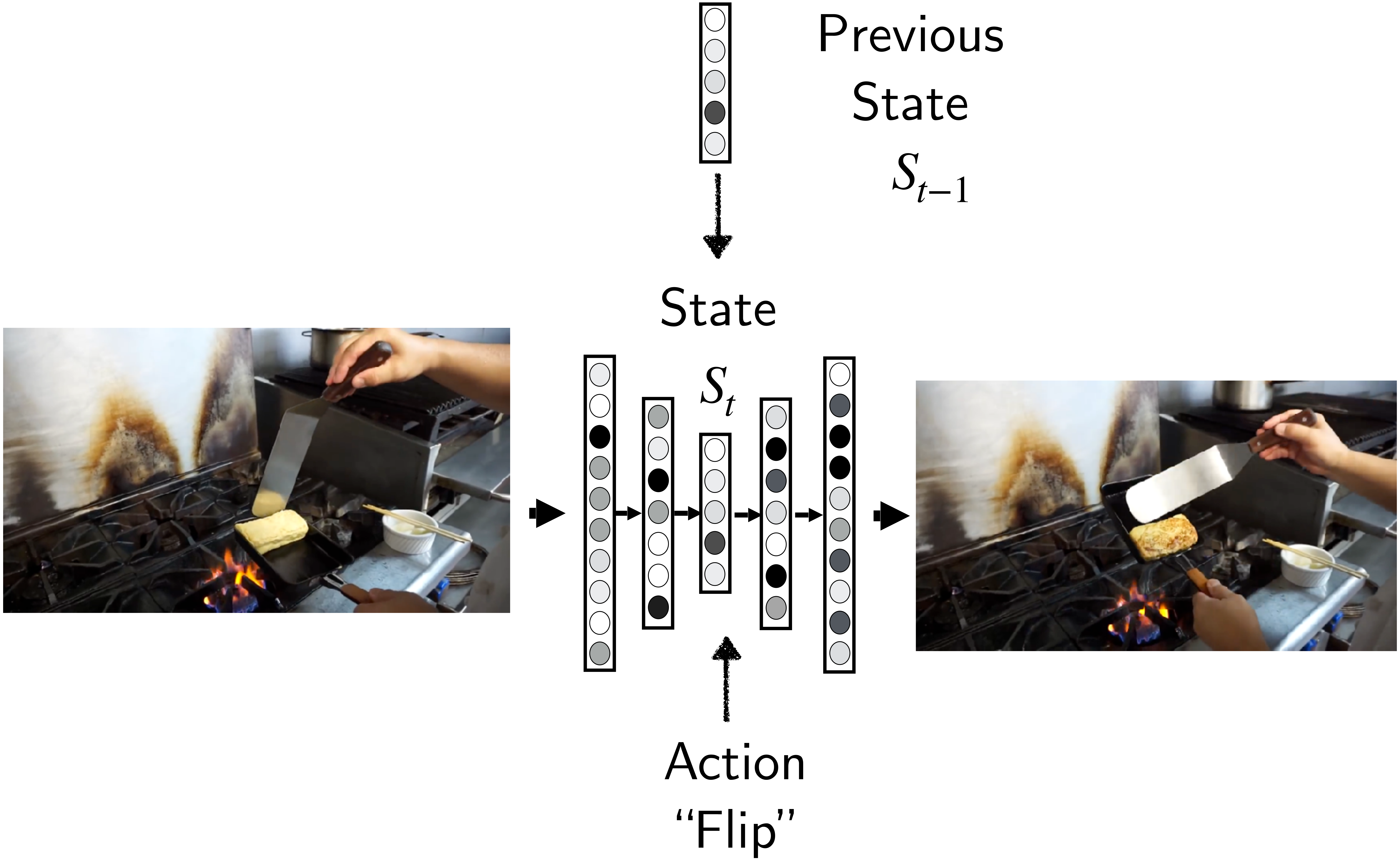
Challenges with learning complex models

Challenge 1: High-dimensional observations

Challenge 2: Planning with Complex Dynamics

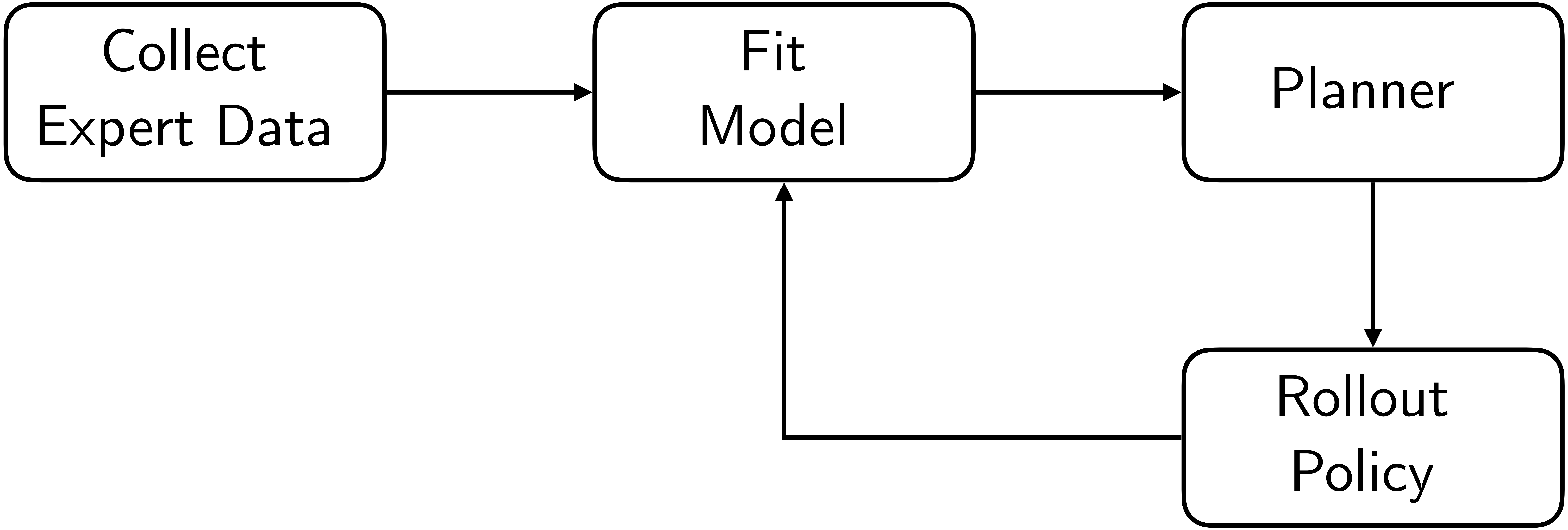

$\mathbf{X}$
Image
$\mathcal{F}$
$\hat{\mathbf{X}} = \mathcal{F}(\mathbf{X})$
Reconstructed image
"Fish"
"Coral"


State
S_t
Action
“Flip”


Previous
State
S_{t-1}
State
S_t
Action
“Flip”

Recall from previous lecture!

(Ross & Bagnell, 2012)



What if we don't have expert data?

The DREAMER Algorithm

Dream to Control: Learning Behaviors by Latent Imagination

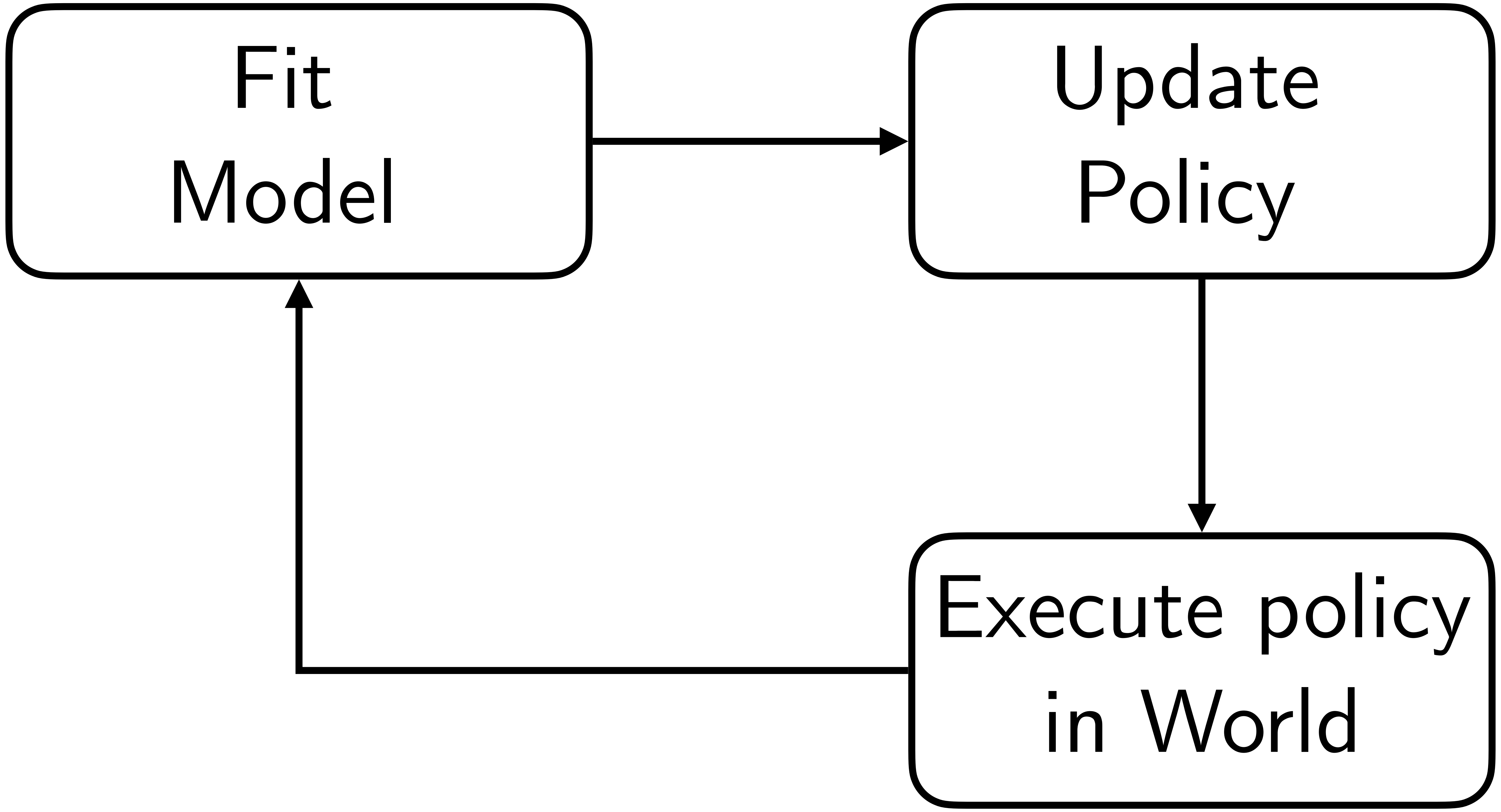
Danijar Hafner *
University of Toronto
Google Brain

Timothy Lillicrap
DeepMind

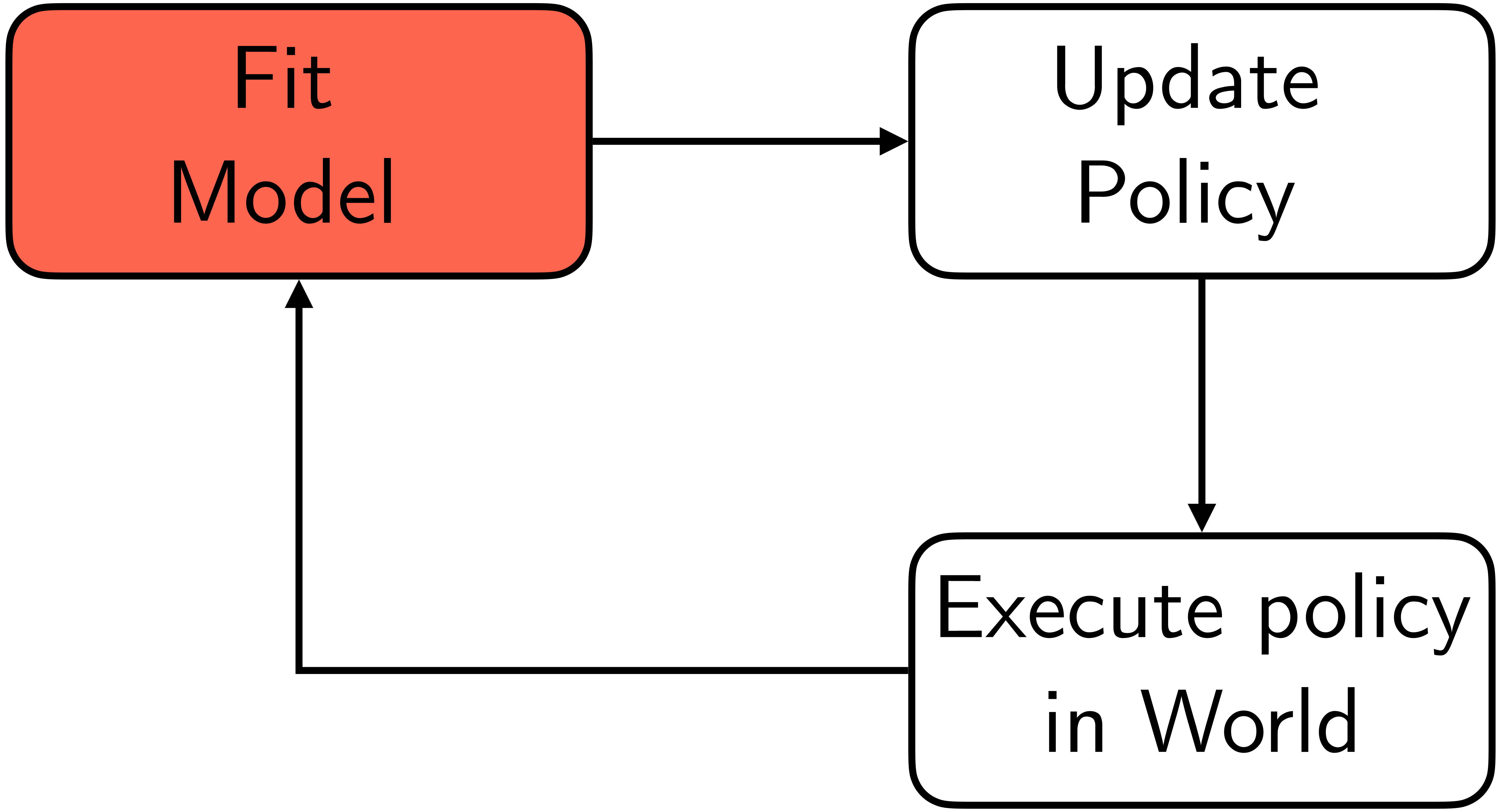
Jimmy Ba
University of Toronto

Mohammad Norouzi
Google Brain

DREAMER



DREAMER



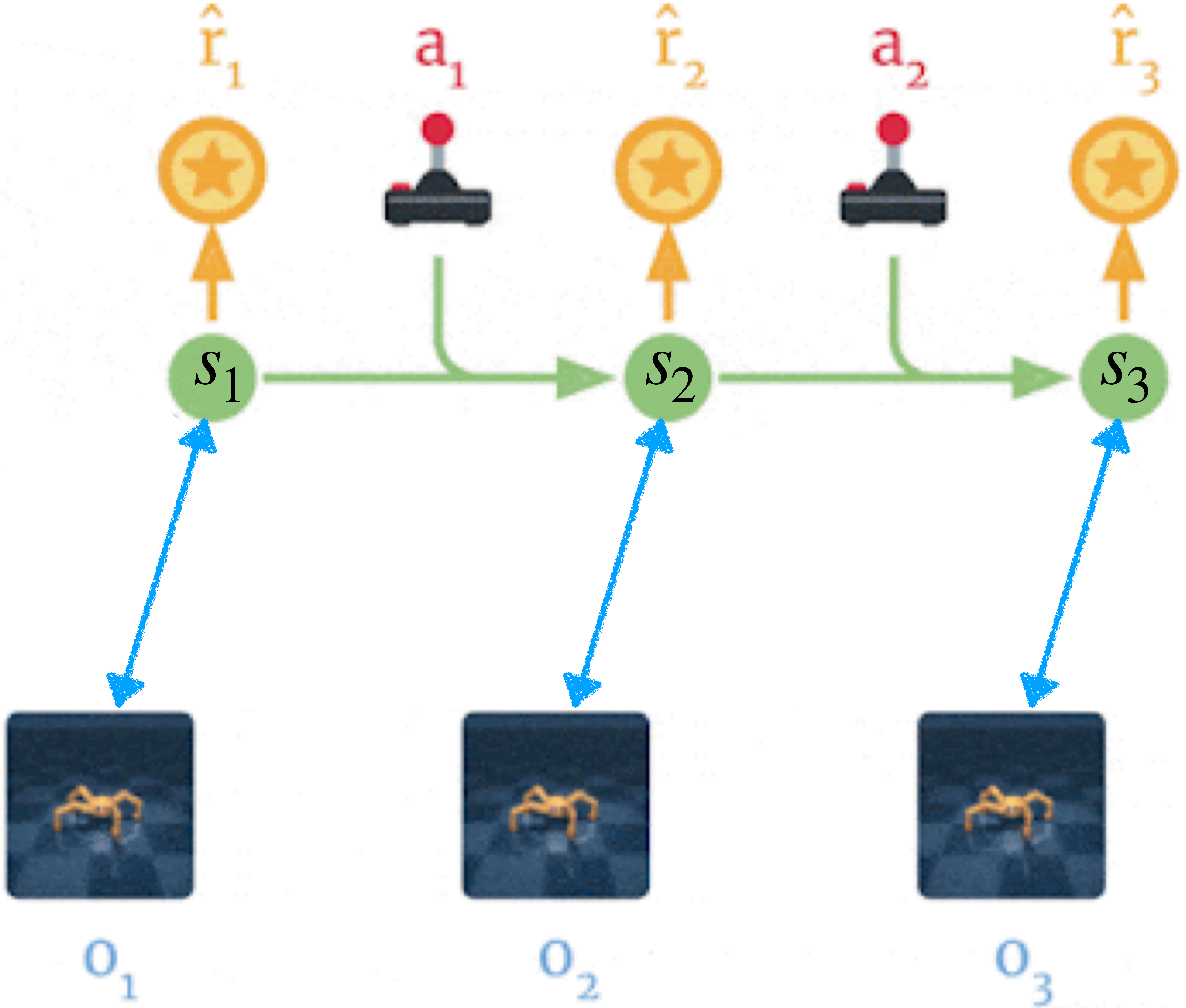
Goal: Fit a Model

Given:
Observations, rewards, actions

Goal: Fit a Model

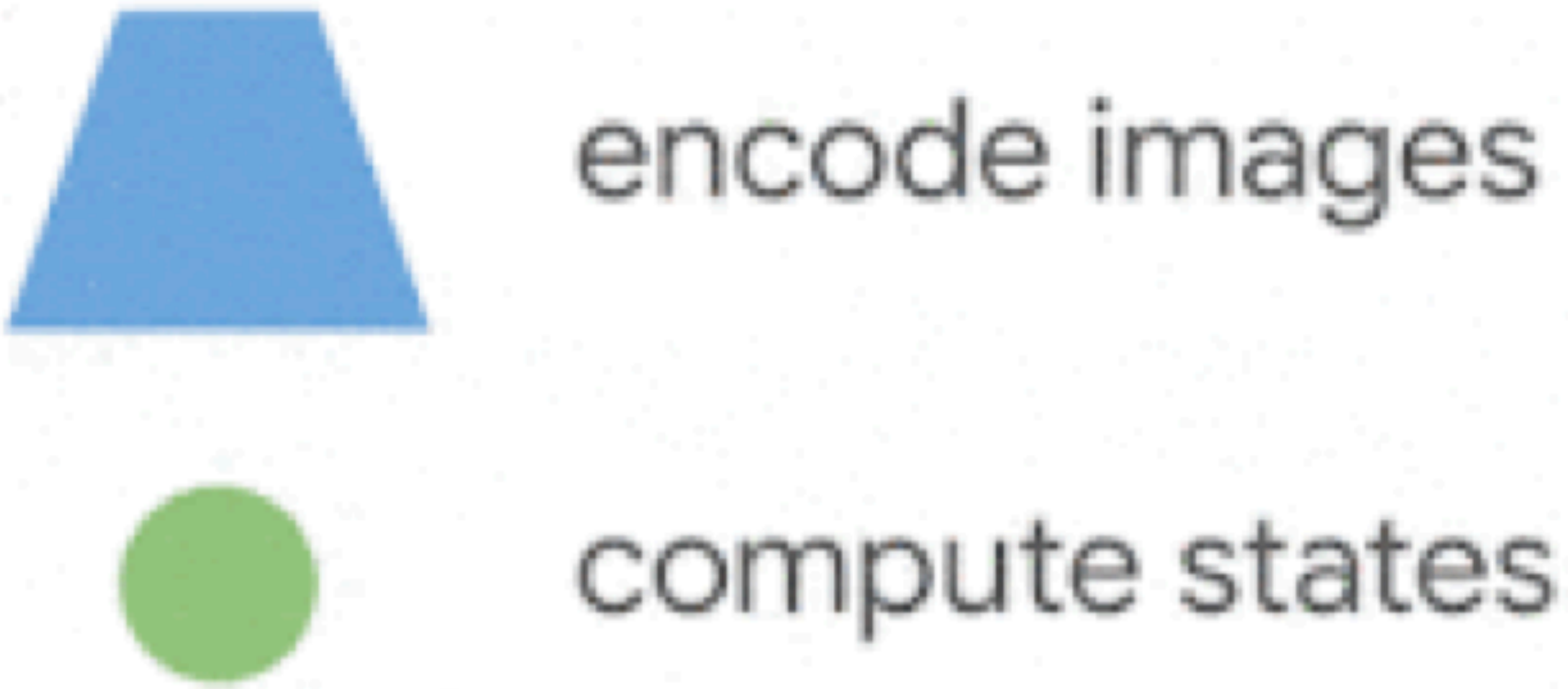
Given:
Observations, rewards,
actions

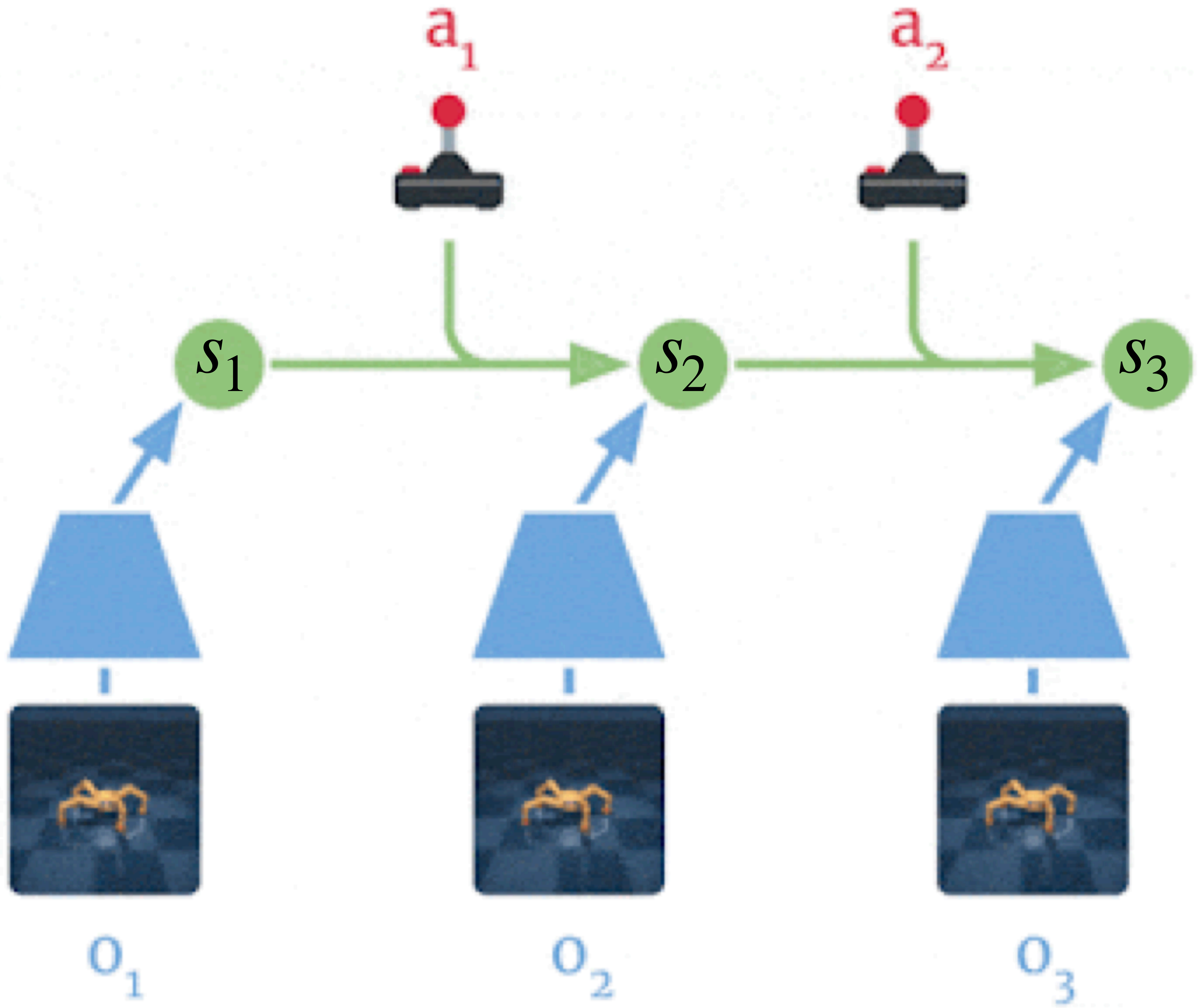
Predict:
States,
Dynamics Function,
Reward Function



Actions

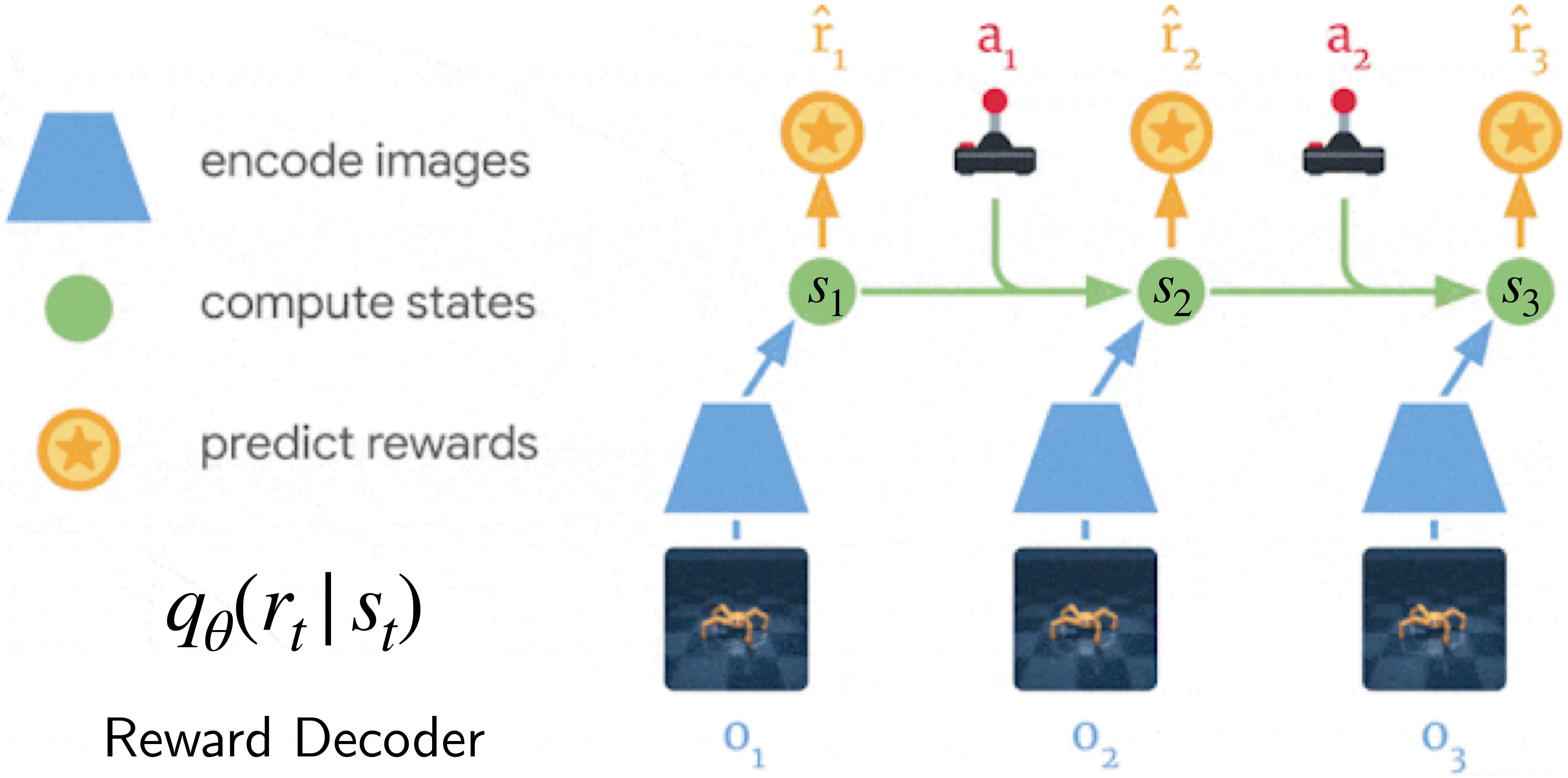
Observations





$$p_\theta(s_t \mid o_t, s_{t-1}, a_{t-1})$$

State Encoder

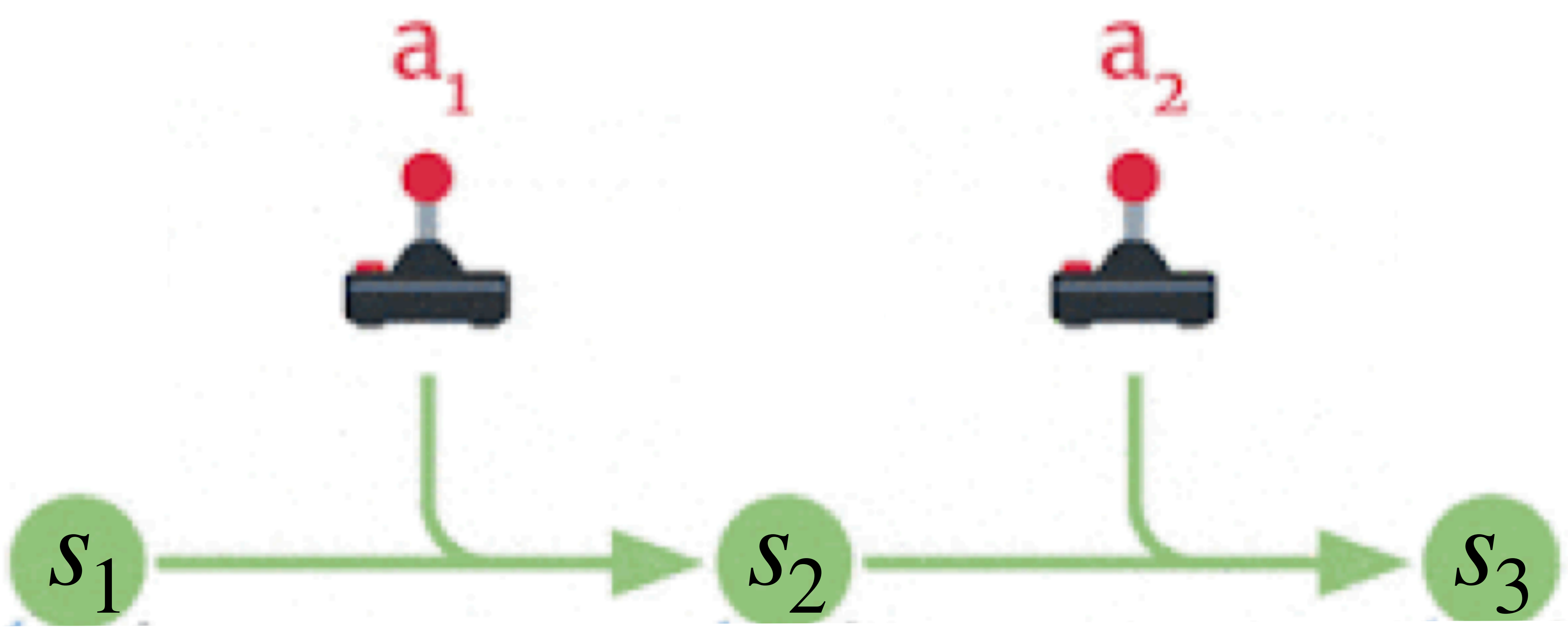


$$q_\theta(r_t \mid s_t)$$

Reward Decoder

$$q_\theta(o_t \mid s_t)$$

Observation Decoder

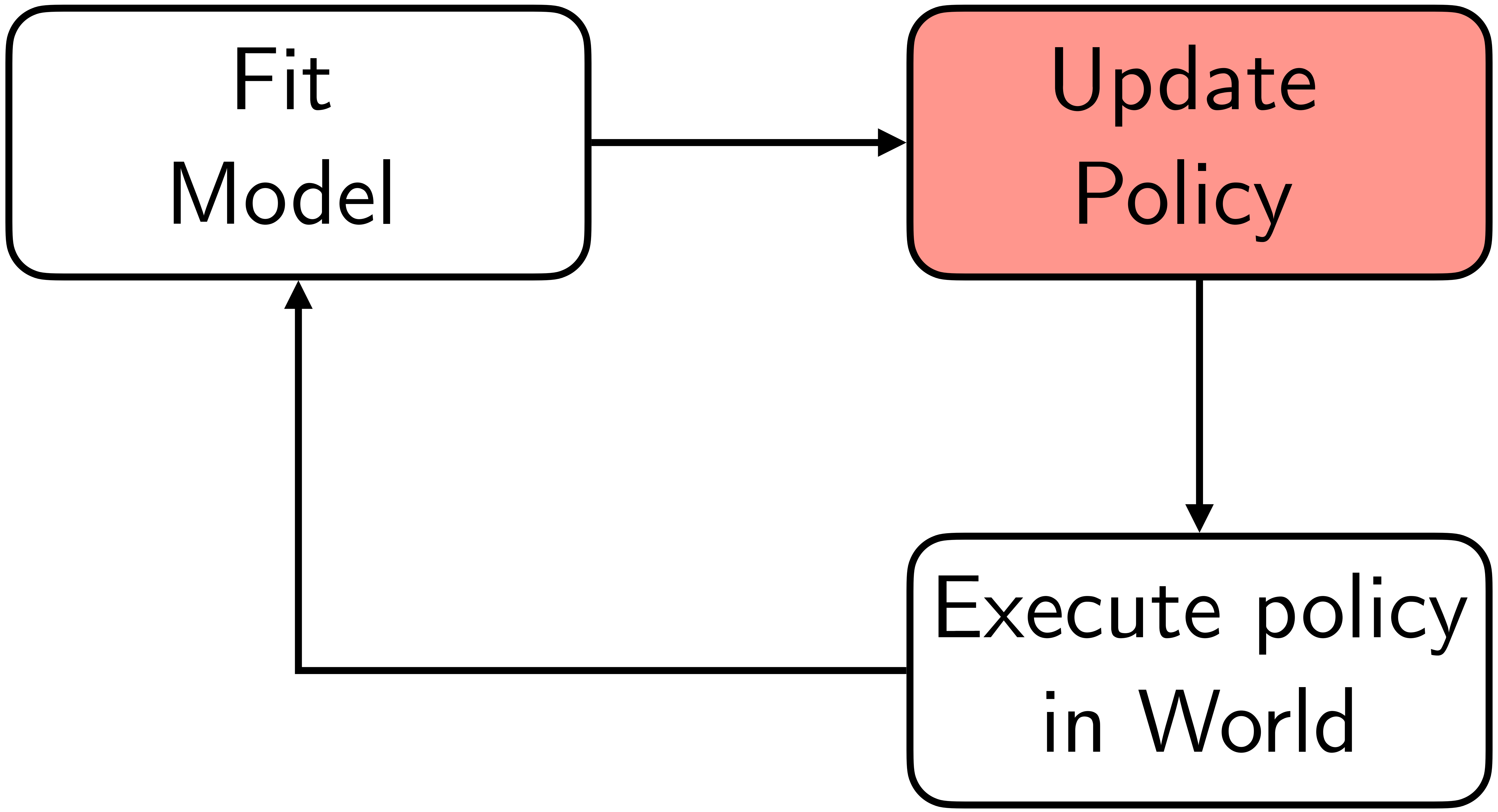


$$q_\theta(s_t \mid s_{t-1}, a_{t-1})$$

Dynamics
Function

Results: Learning World Model

DREAMER



Goal: Learn a Policy using Actor-Critic

$\pi_\phi(a_t \mid s_t)$

Actor

$V_\psi(s_t)$

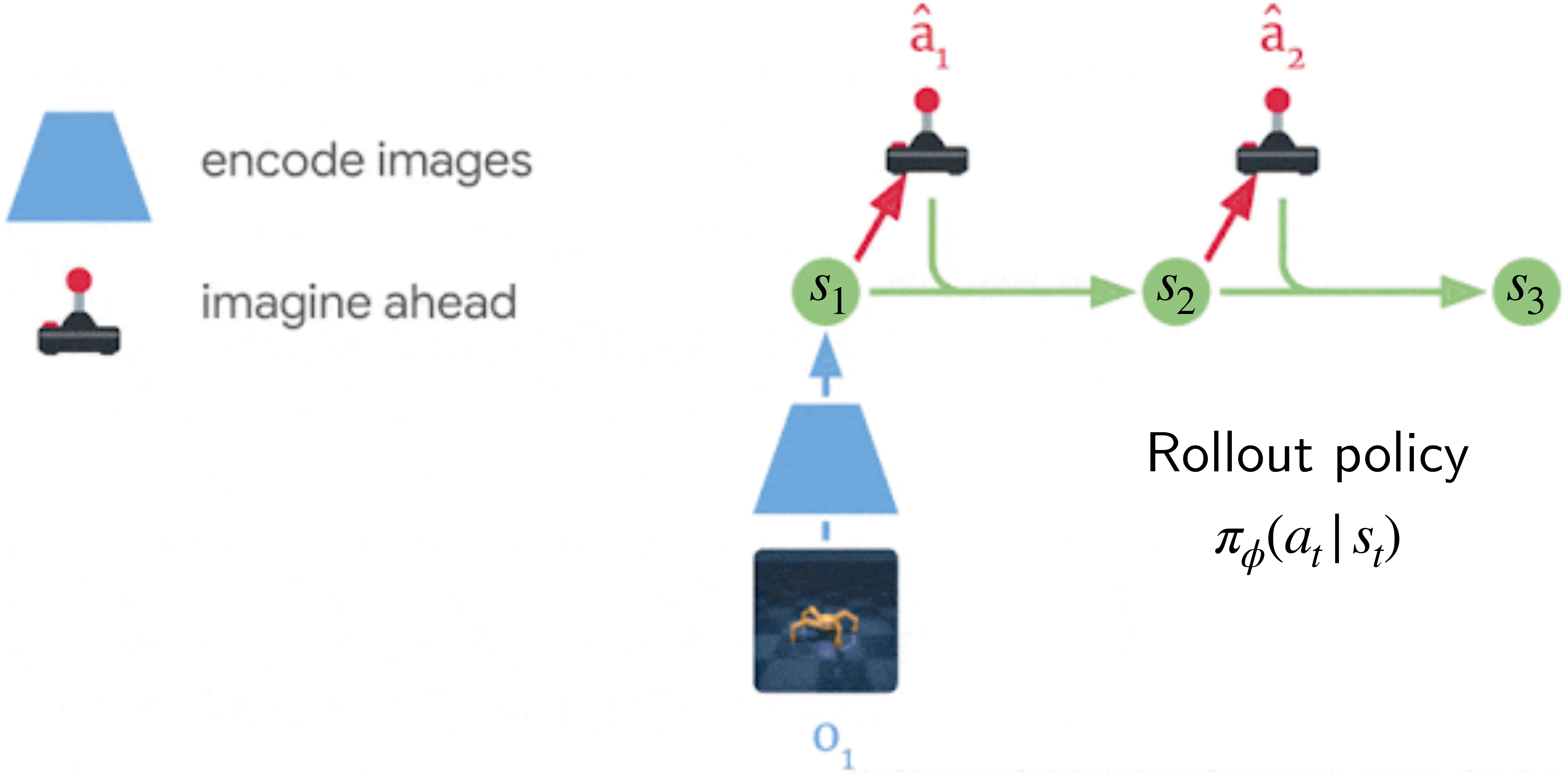
Critic

From rollouts in the model

$q_\theta(s_t \mid s_{t-1}, a_{t-1})$

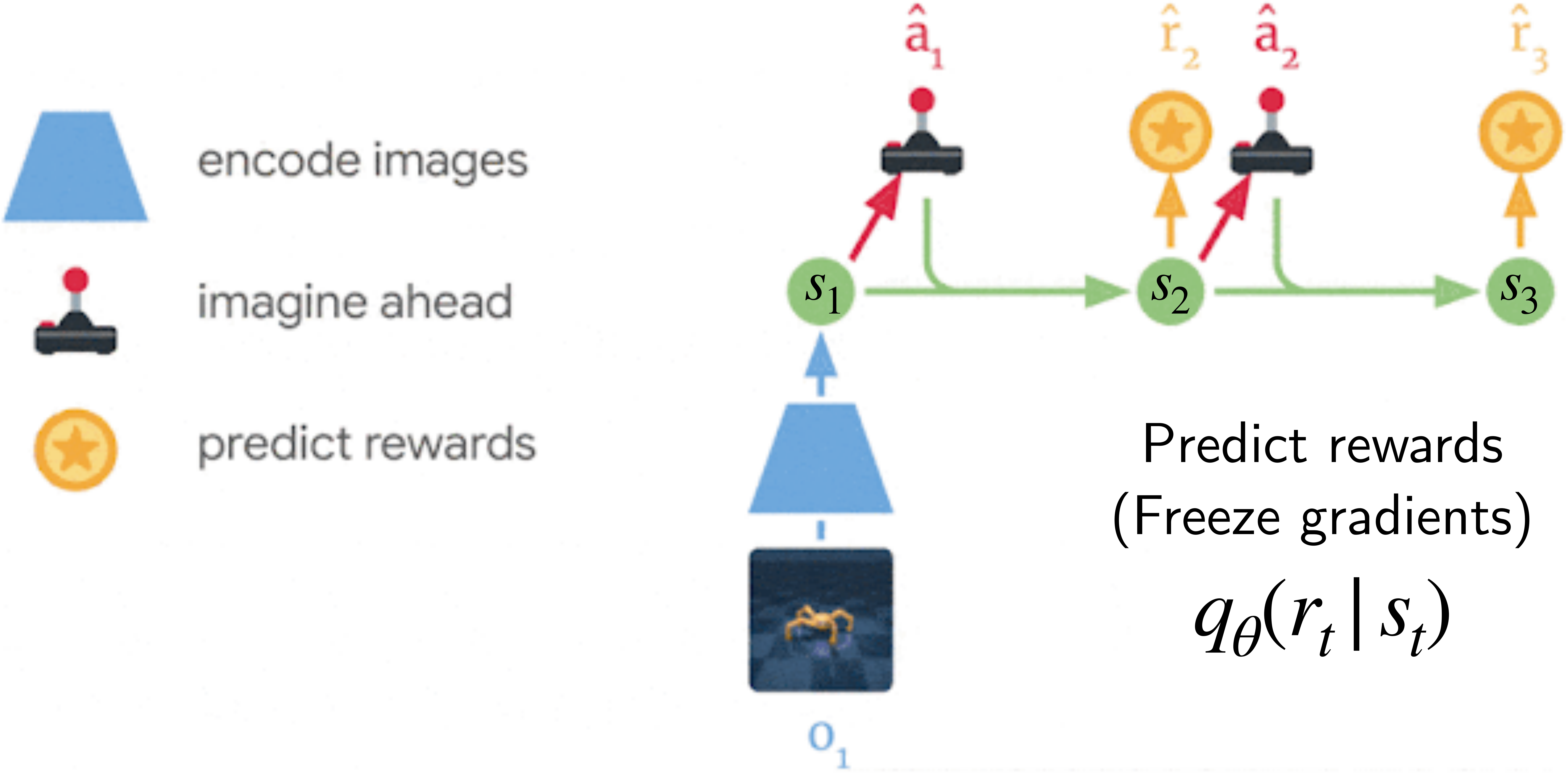
o_1

encode images



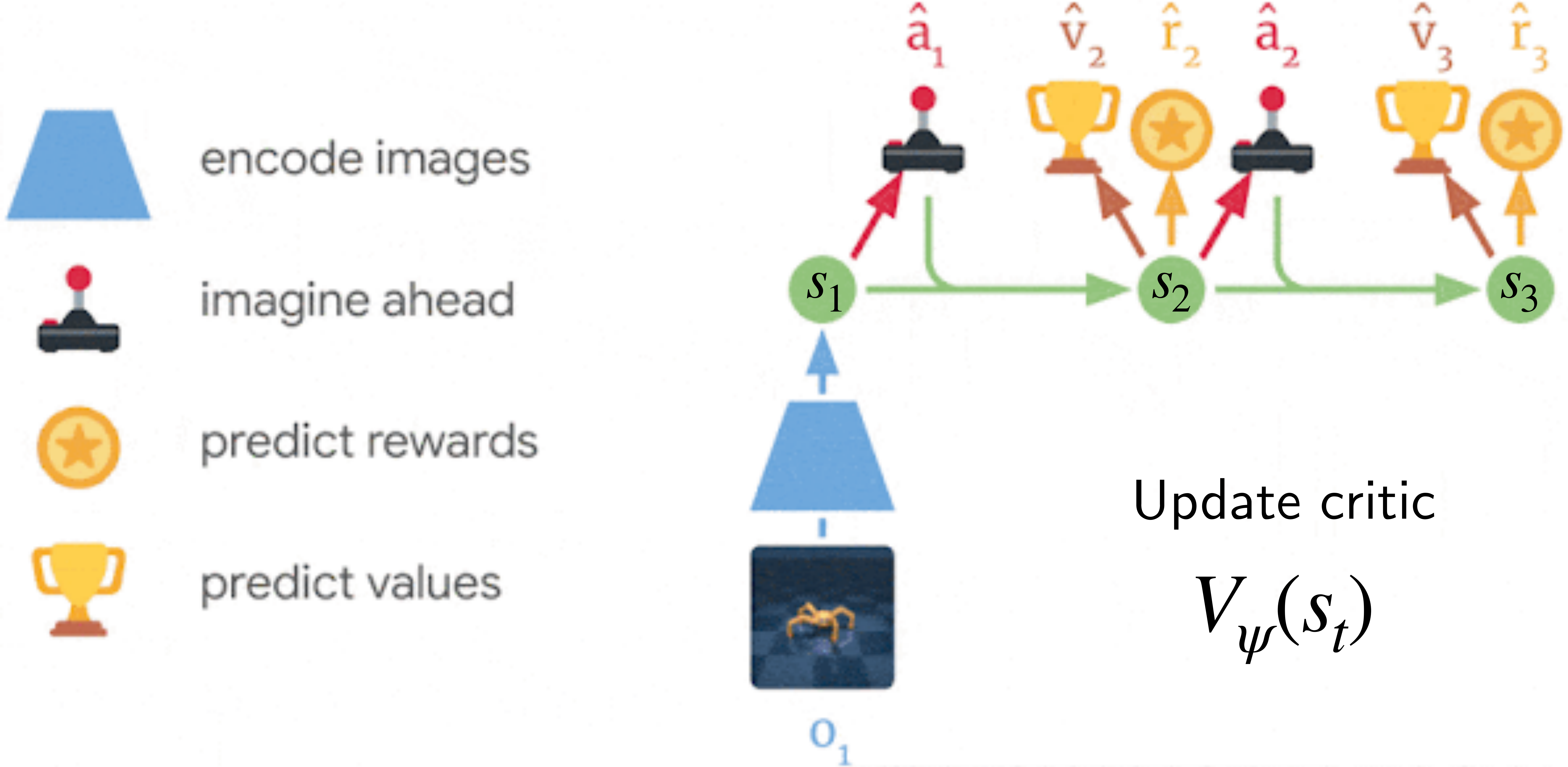
Rollout policy

$\pi_\phi(a_t \mid s_t)$



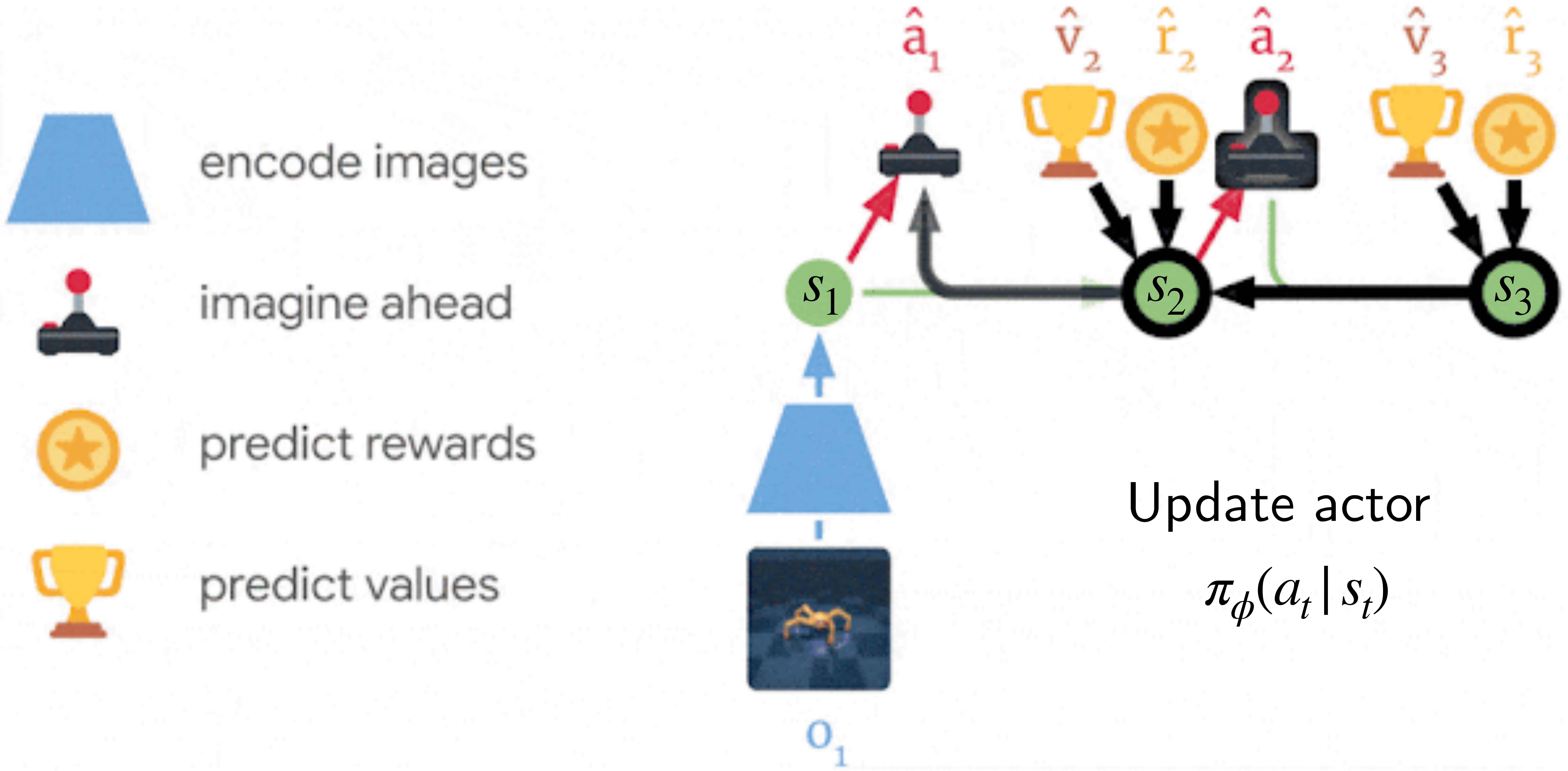
Predict rewards
(Freeze gradients)

$$q_\theta(r_t \mid s_t)$$



Update critic

$$V_\psi(s_t)$$

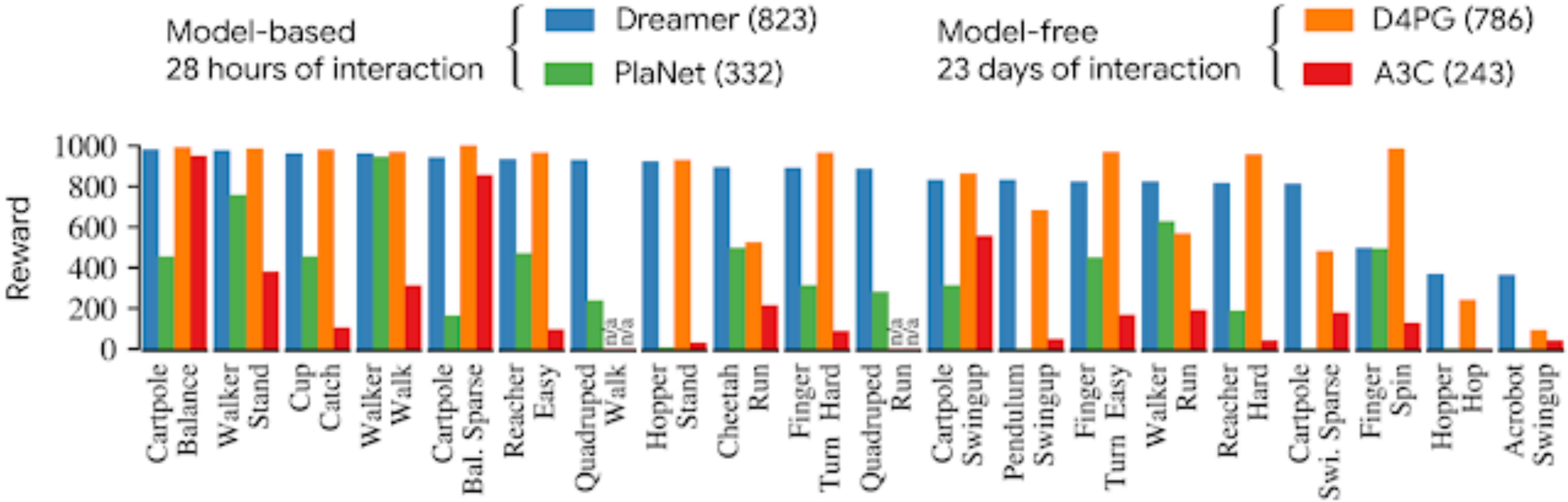


Update actor

$$\pi_\phi(a_t \mid s_t)$$

DREAMER: Results

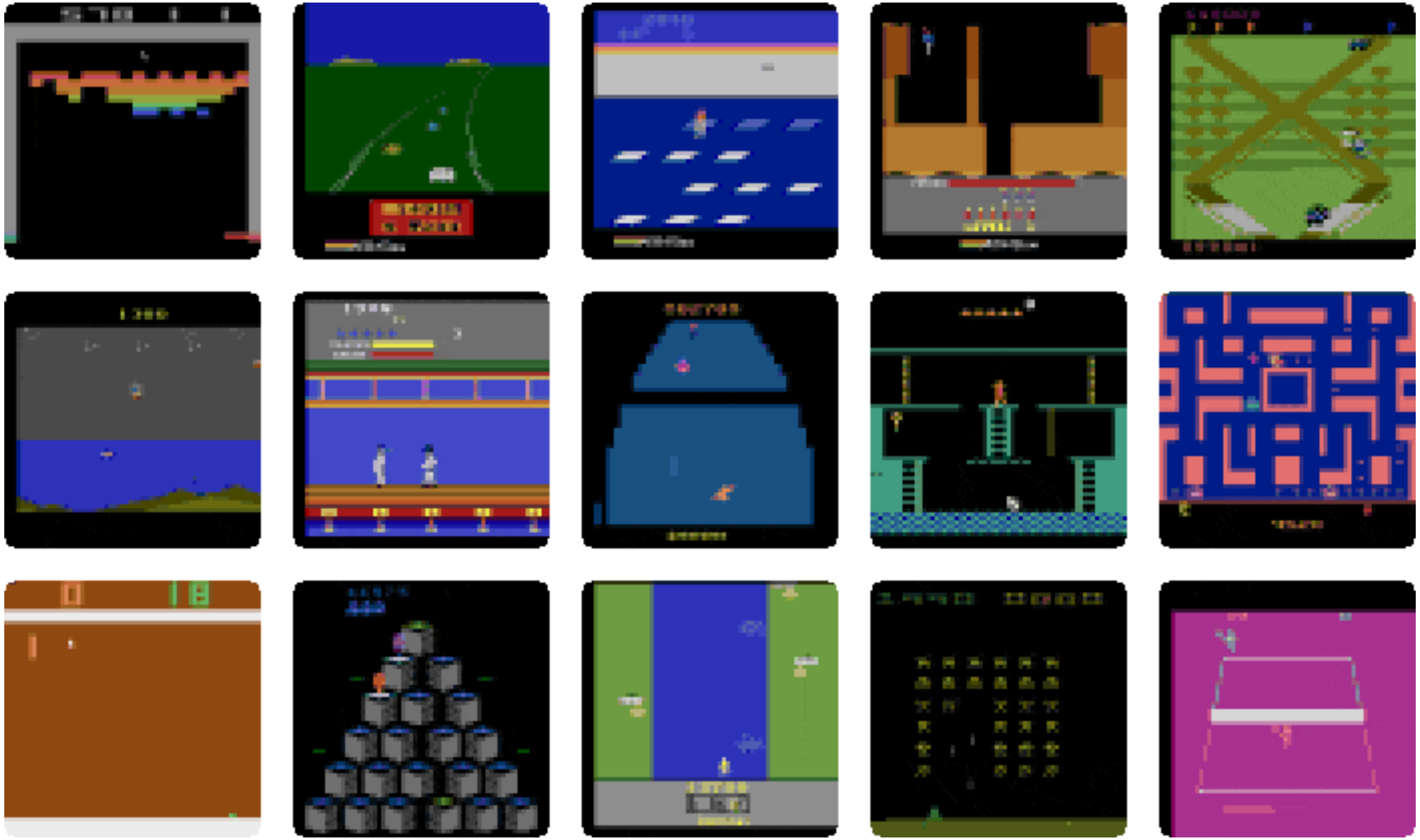
Sparse Cartpole | Acrobot Swingup | Hopper Hop | Walker Run | Quadruped Run | Boxing | Freeway | Frostbite | Collect Objects | Watermaze



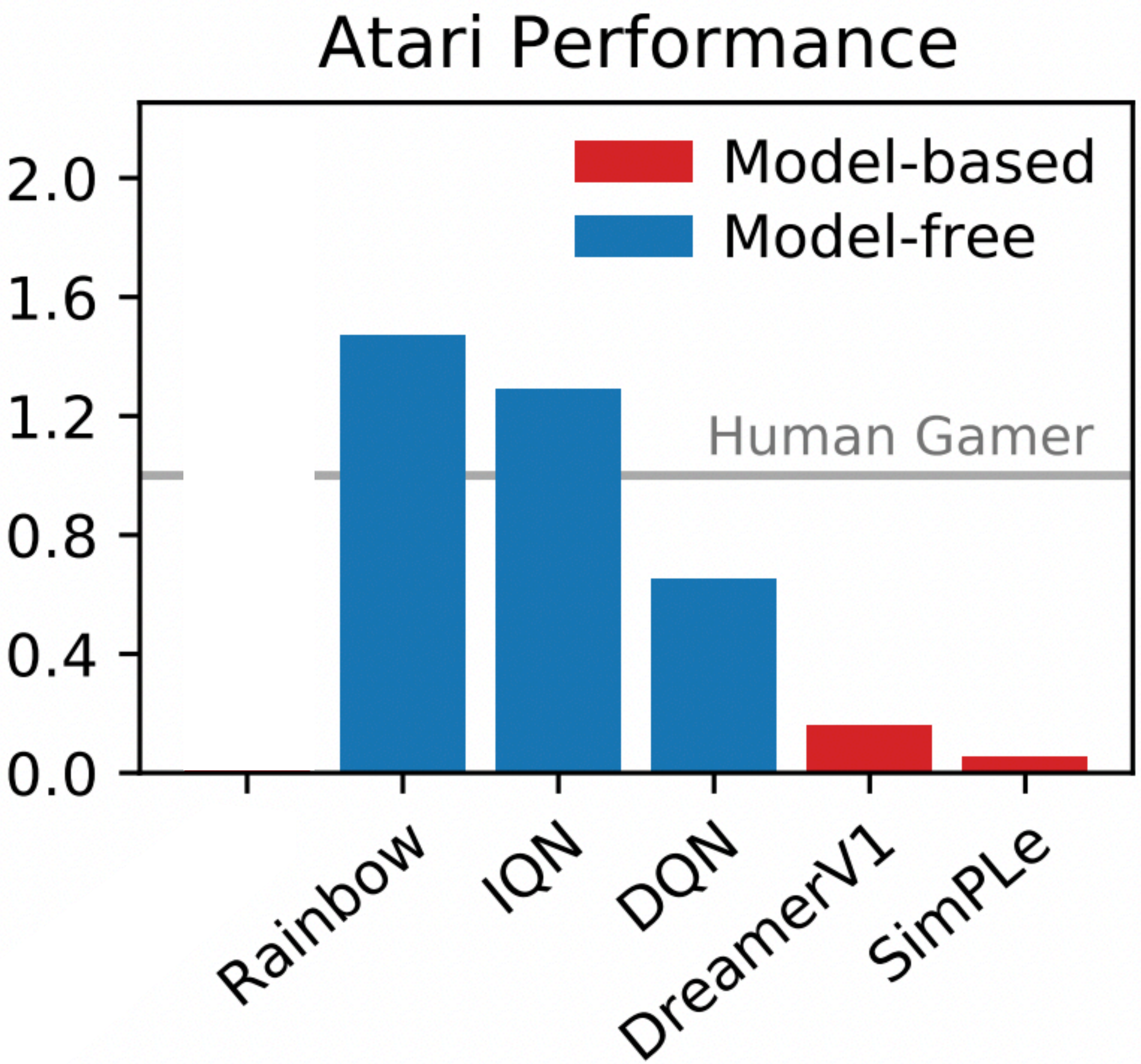
Are we done?

Mastering Atari with Discrete World Models

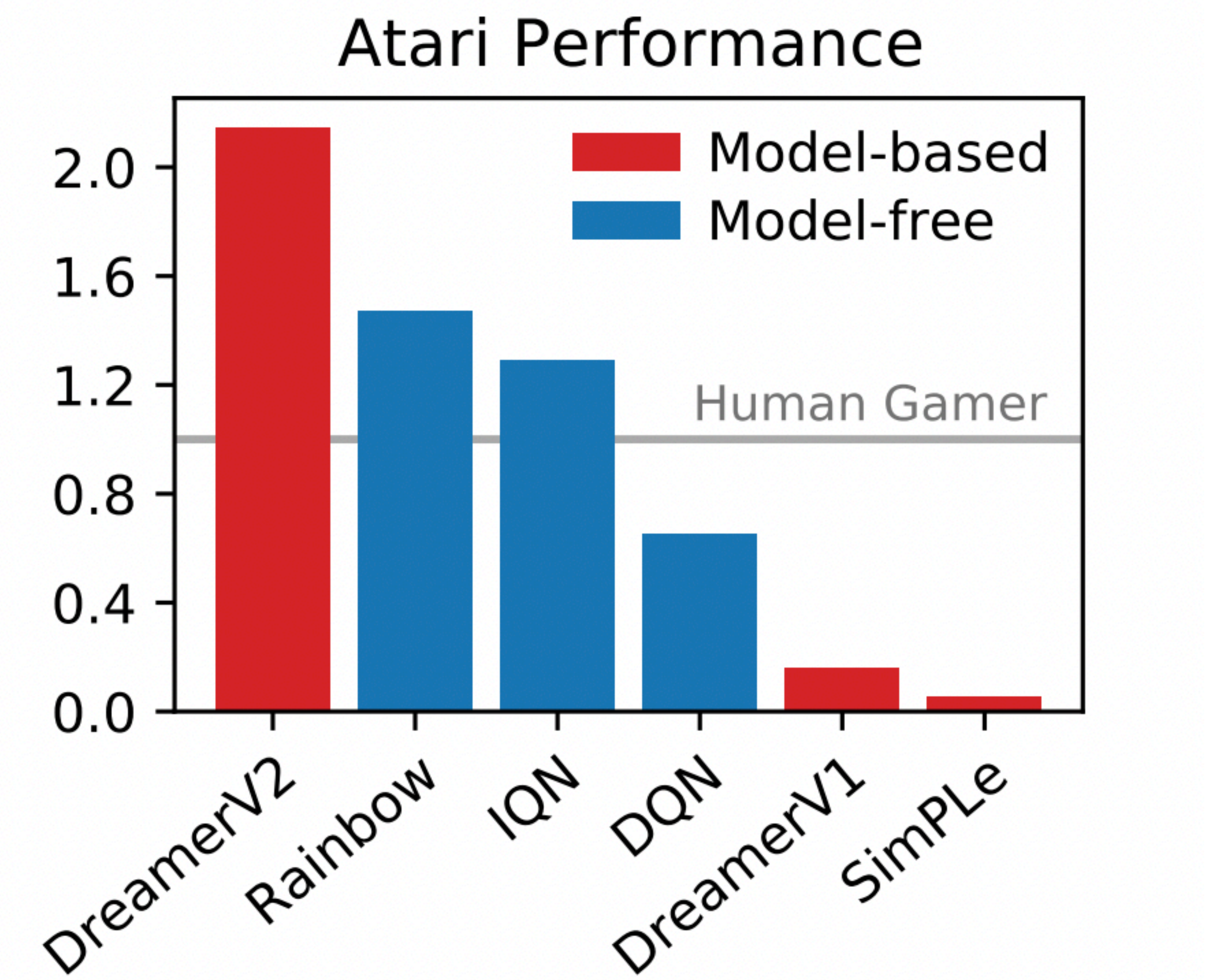
Danijar Hafner * Google Research
Timothy Lillicrap DeepMind
Mohammad Norouzi Google Research
Jimmy Ba University of Toronto



Atari was hard for Model Based RL

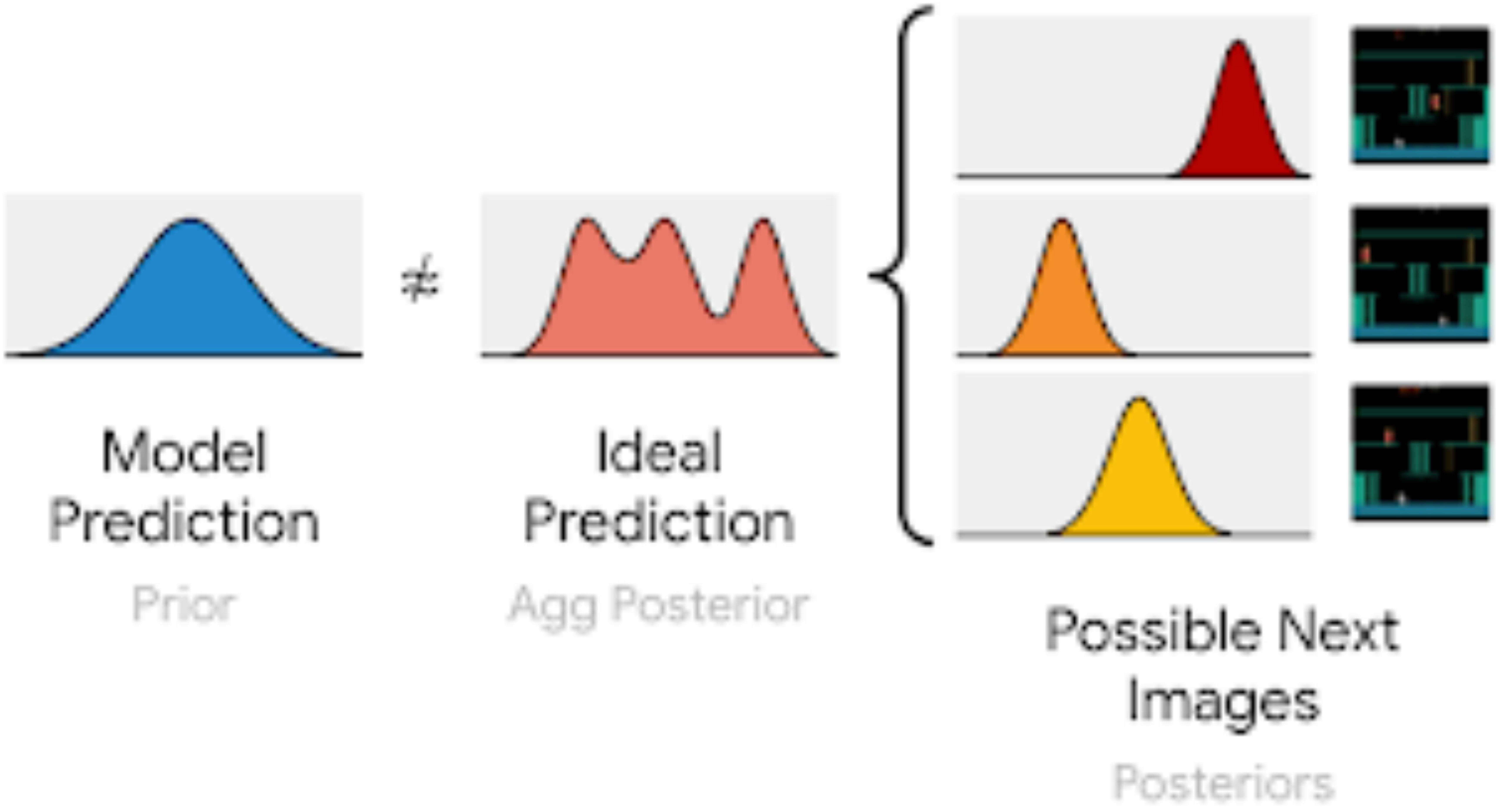


DreamerV2 beats all model free!

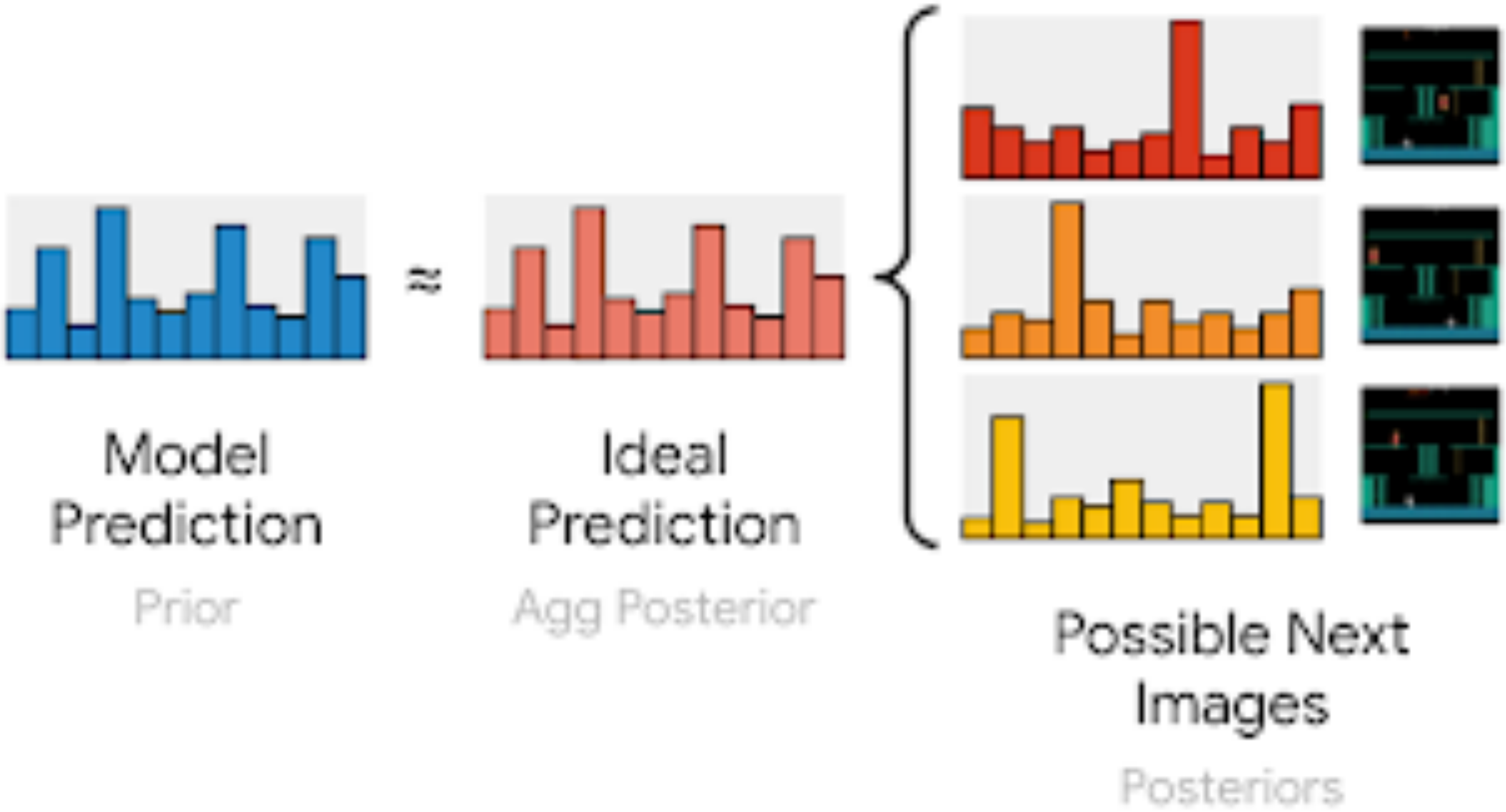


Problem: Dreamer V1 predicts a single mode of dynamics

Dreamer V1 predicts single mode dynamics

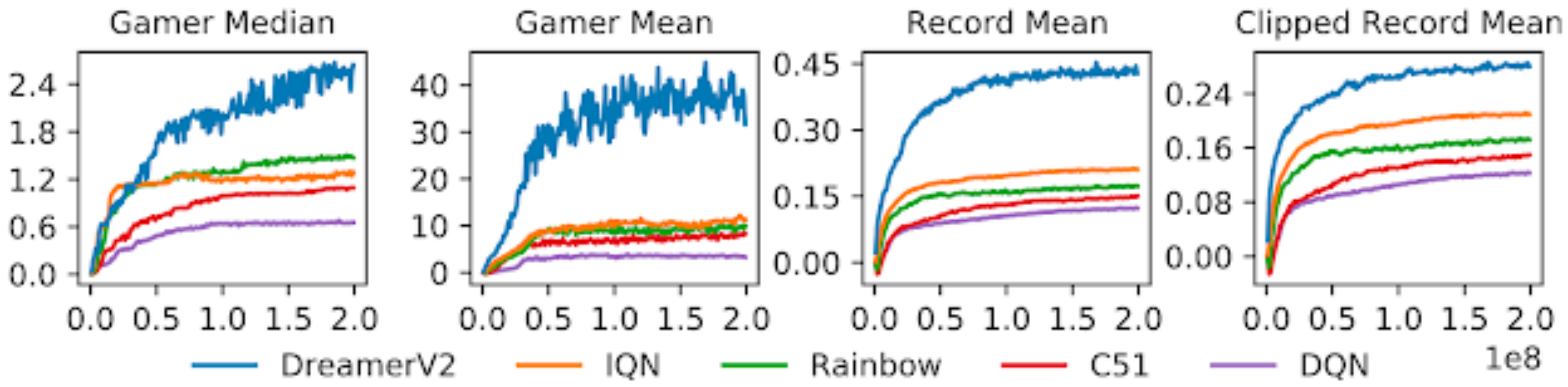


Idea: Predict multiple discrete modes!




$\hat{x}_1$
$\hat{x}_2$
$\hat{x}_3$
a_1
a_2
Sparse Representation
h_1
h_2
h_3
$\hat{z}_1$
z_1
$\hat{z}_2$
z_2
$\hat{z}_3$
z_3
min KL
min KL
min KL
32 categoricals
32 classes each
x_1
x_2
x_3


True
Model


Gamer Median
Gamer Mean
Record Mean
Clipped Record Mean
2.4
1.8
1.2
0.6
0.0
40
30
20
10
0
0.45
0.30
0.15
0.00
0.24
0.16
0.08
0.00
0.0 0.5 1.0 1.5 2.0
1e8
DreamerV2
IQN
Rainbow
C51
DQN

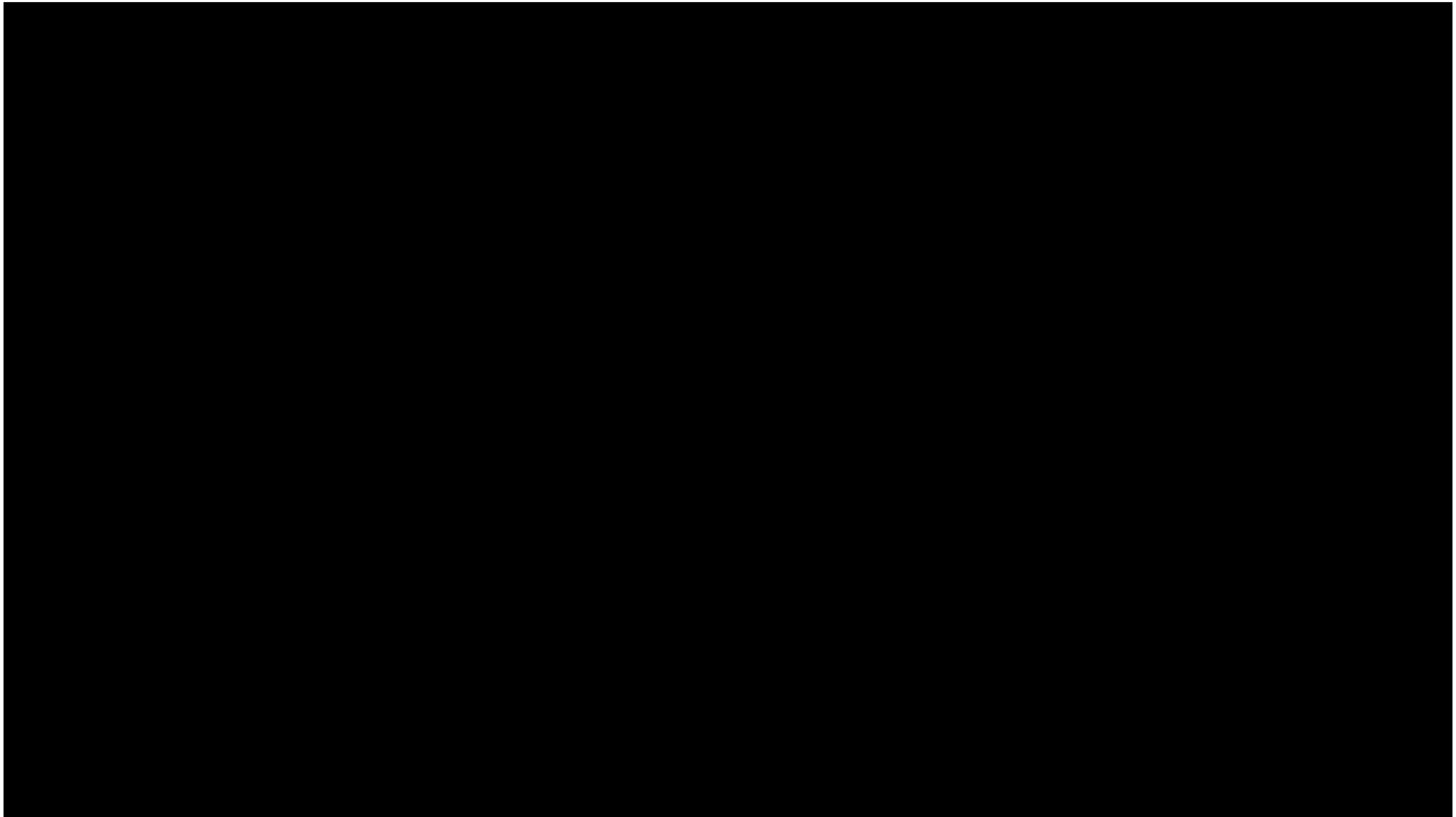
Are we done?

Mastering Diverse Domains through World Models

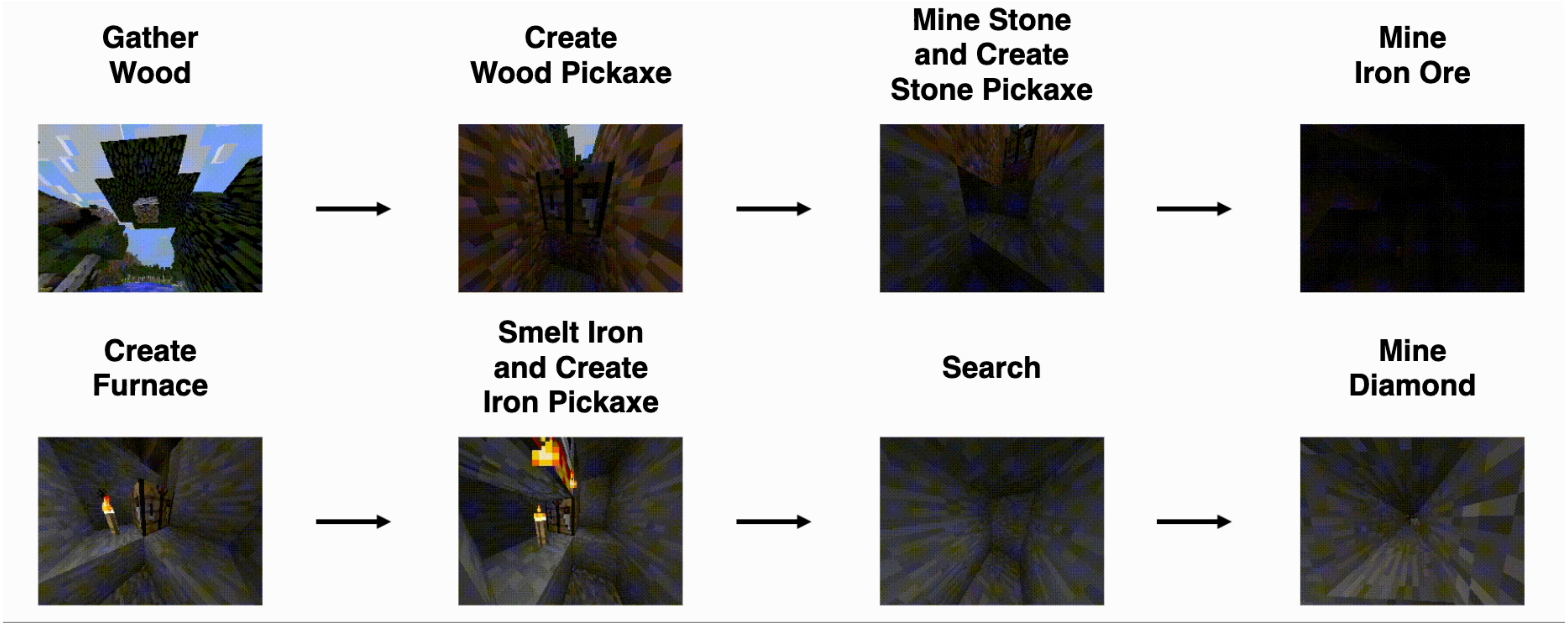
Danijar Hafner[1,2] Jurgis Pasukonis[1] Jimmy Ba[2] Timothy Lillicrap[1]

[1]DeepMind [2]University of Toronto

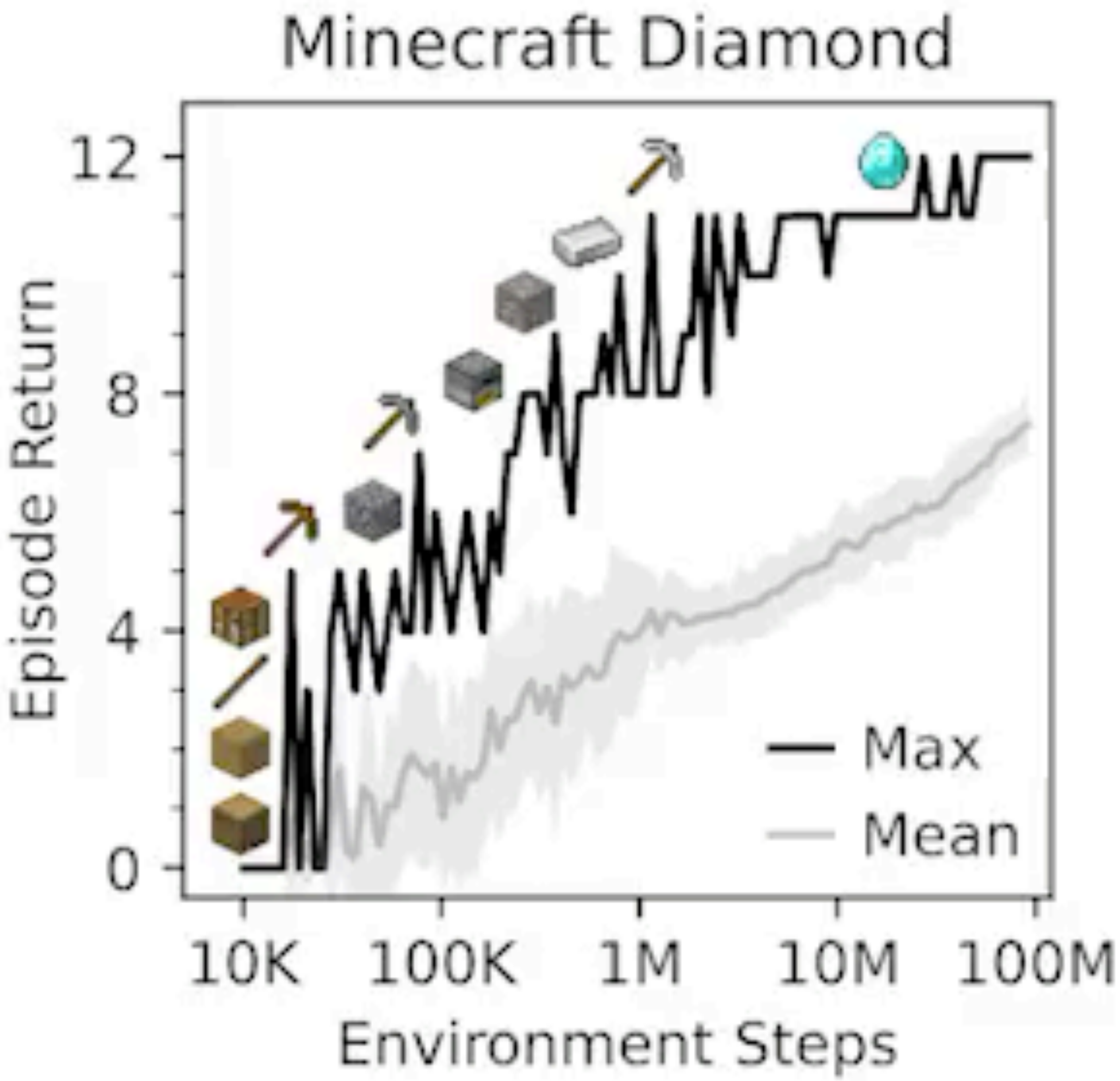
MineRL Diamond Challenge

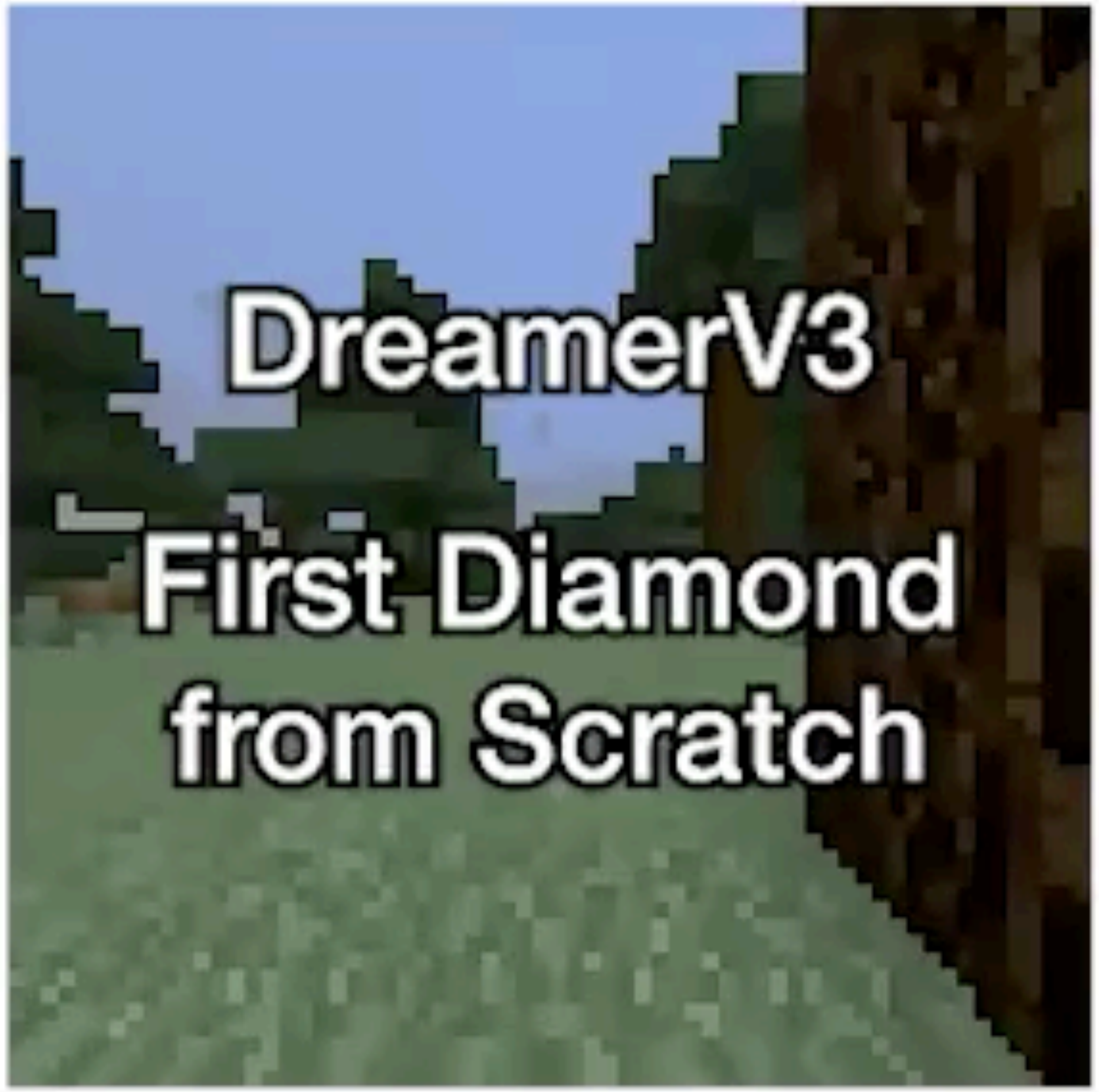


MineRL Diamond Challenge



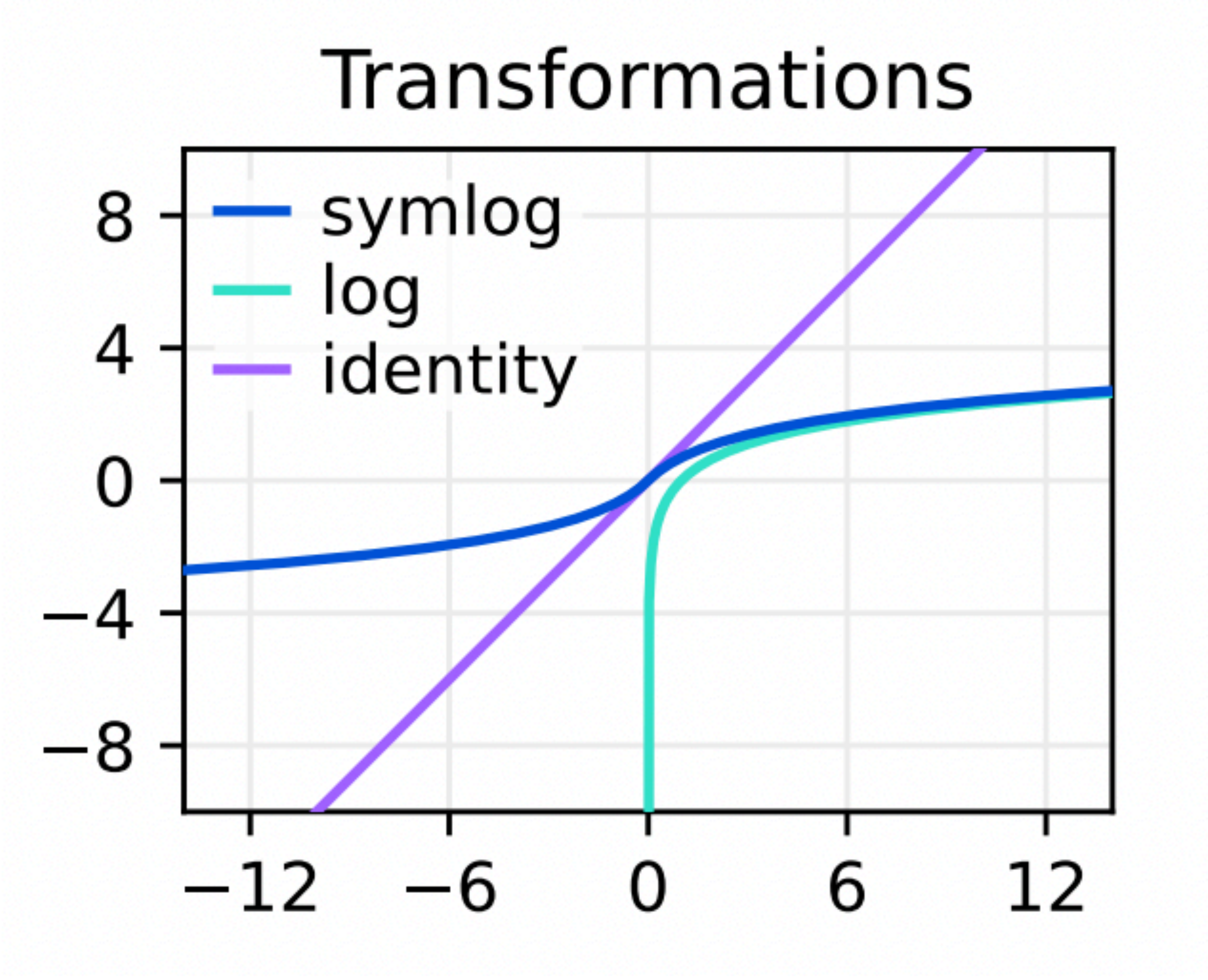
DreamerV3 solved this task!



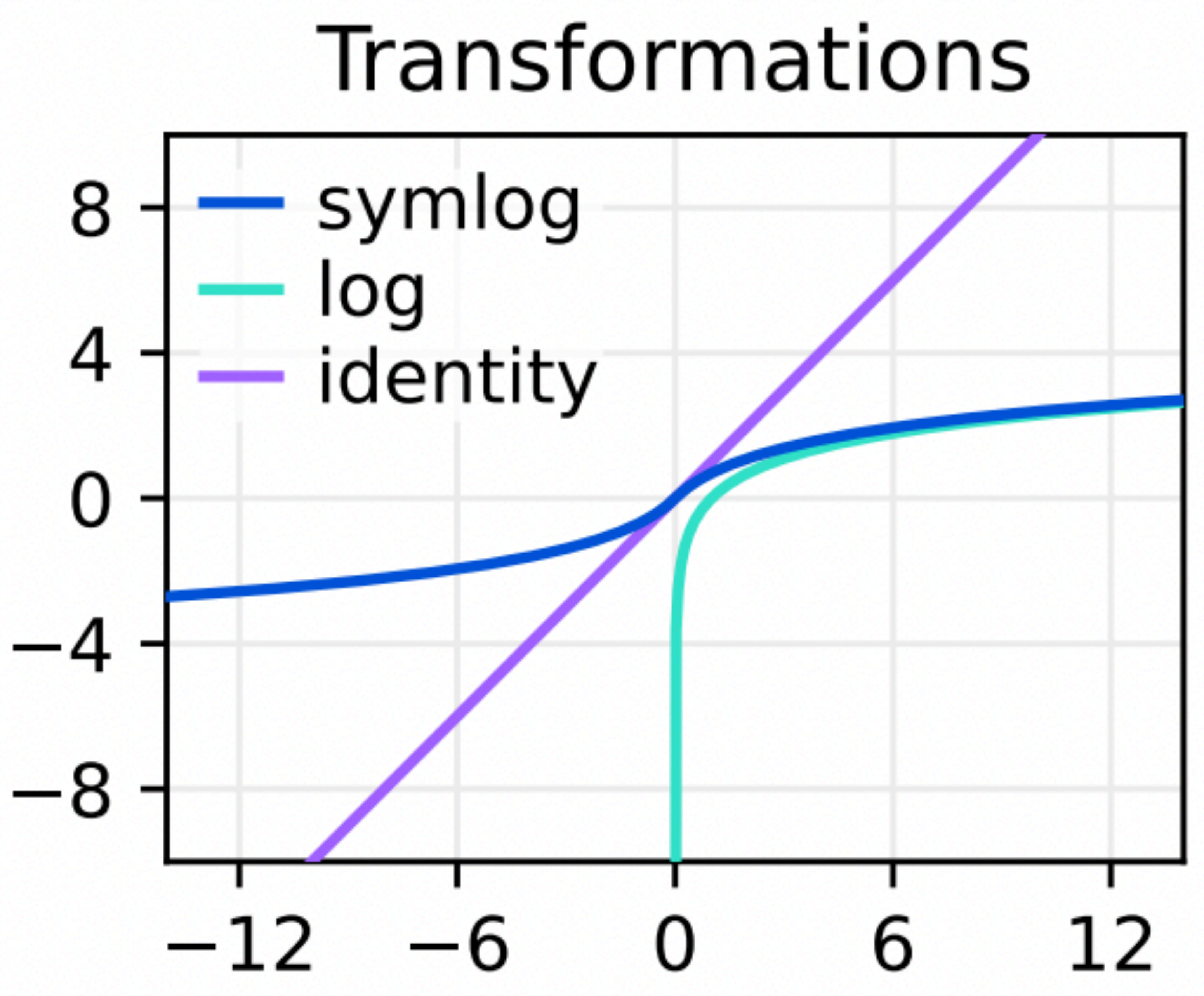


Problem: Scale of rewards, values vary wildly across domains

Solution: “Squash” predictions with symlog function



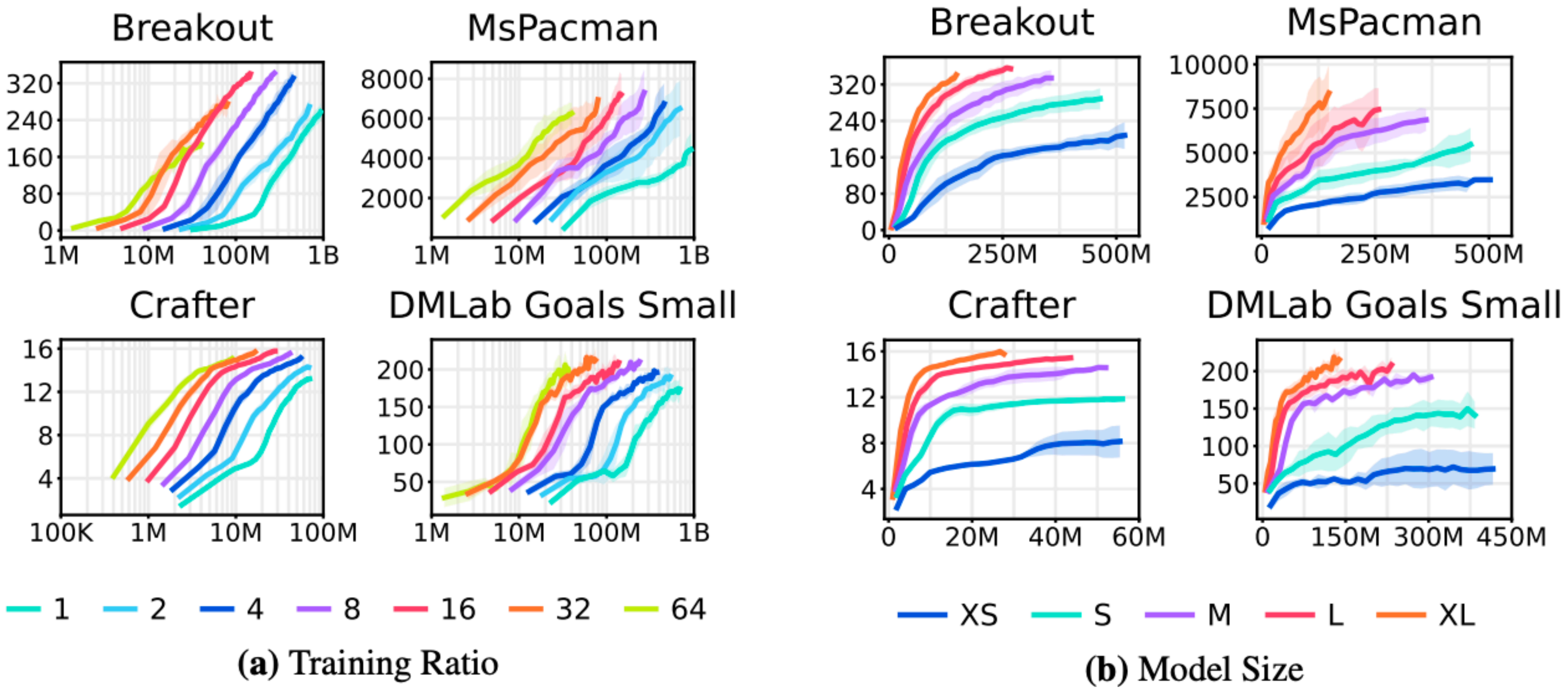
Solution: "Squash" predictions with symlog function



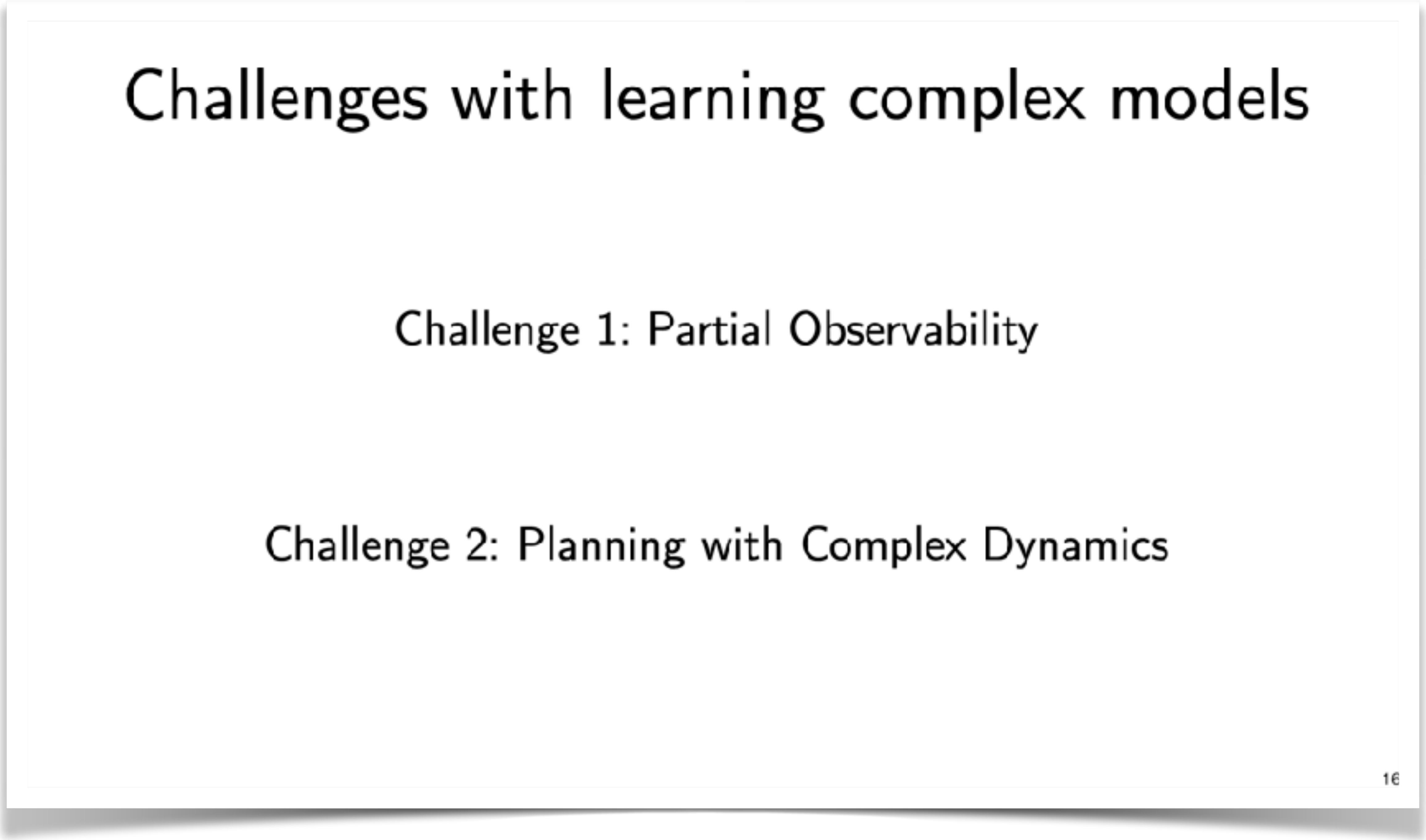
$$\mathcal{L}(\theta) \doteq \frac{1}{2}\Big(f(x,\theta) - \operatorname{symlog}(y)\Big)^2 \qquad \hat{y} \doteq \operatorname{symexp}\big(f(x,\theta)\big)$$

$$\operatorname{symlog}(x) \doteq \operatorname{sign}(x)\ln\big(|x|+1\big) \qquad \operatorname{symexp}(x) \doteq \operatorname{sign}(x)\big(\exp(|x|)-1\big)$$

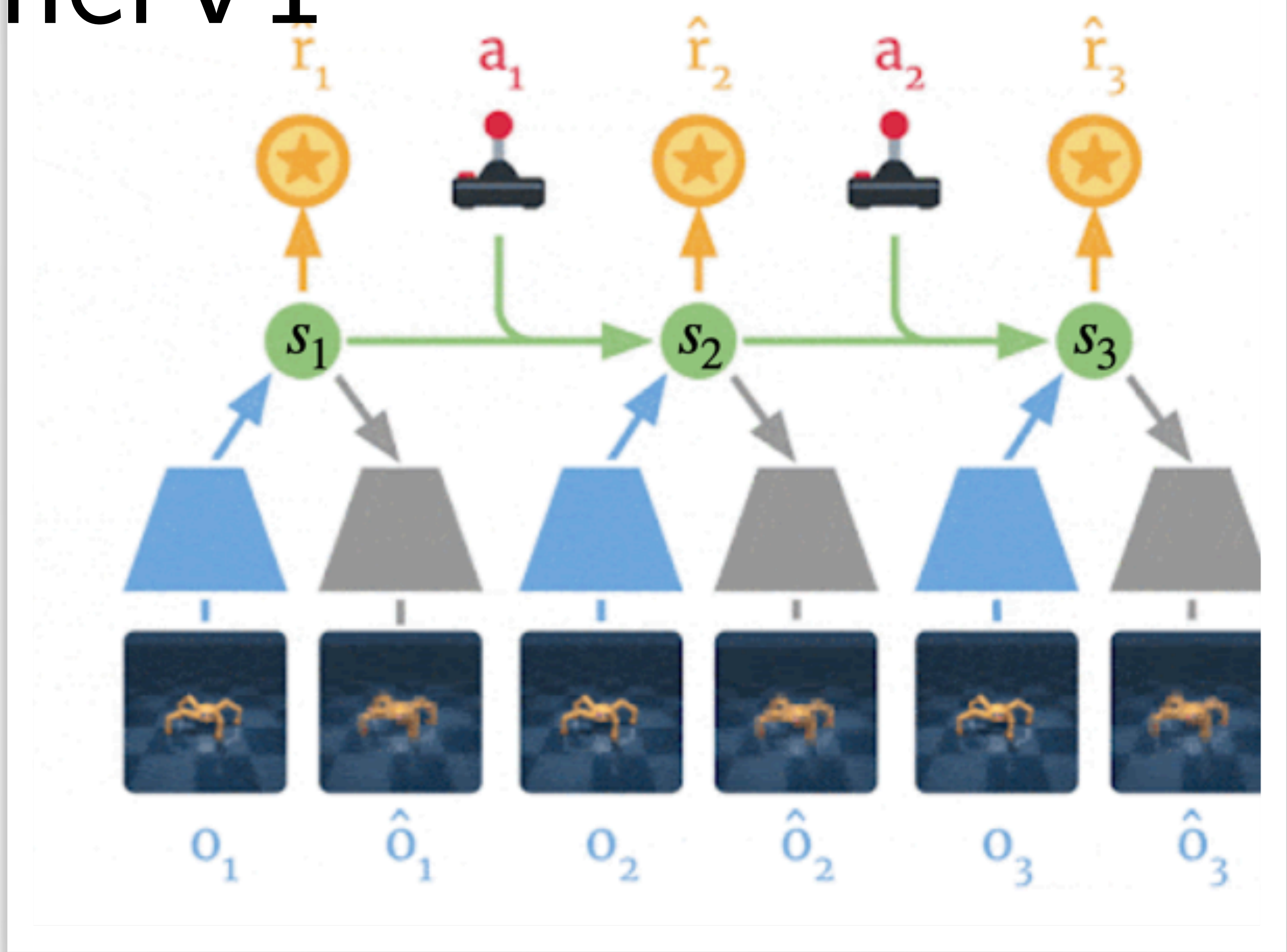
DreamerV3 scales really well!



tl;dr



DreamerV1



Extensions (V2, V3)