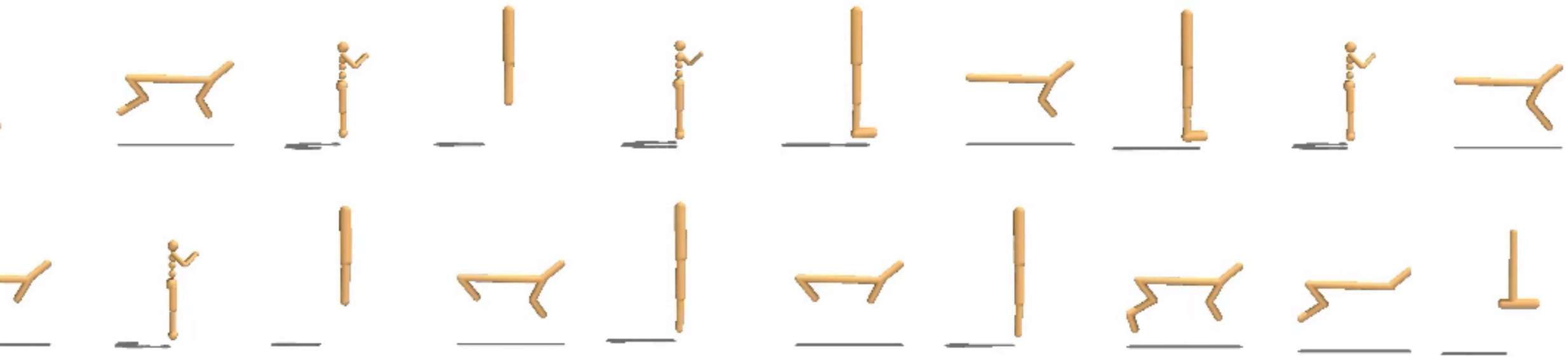
One Policy to Control Them All:

Shared Modular Policies for Agent-Agnostic Control

Wenlong Huang
UC Berkeley

Igor Mordatch
Google

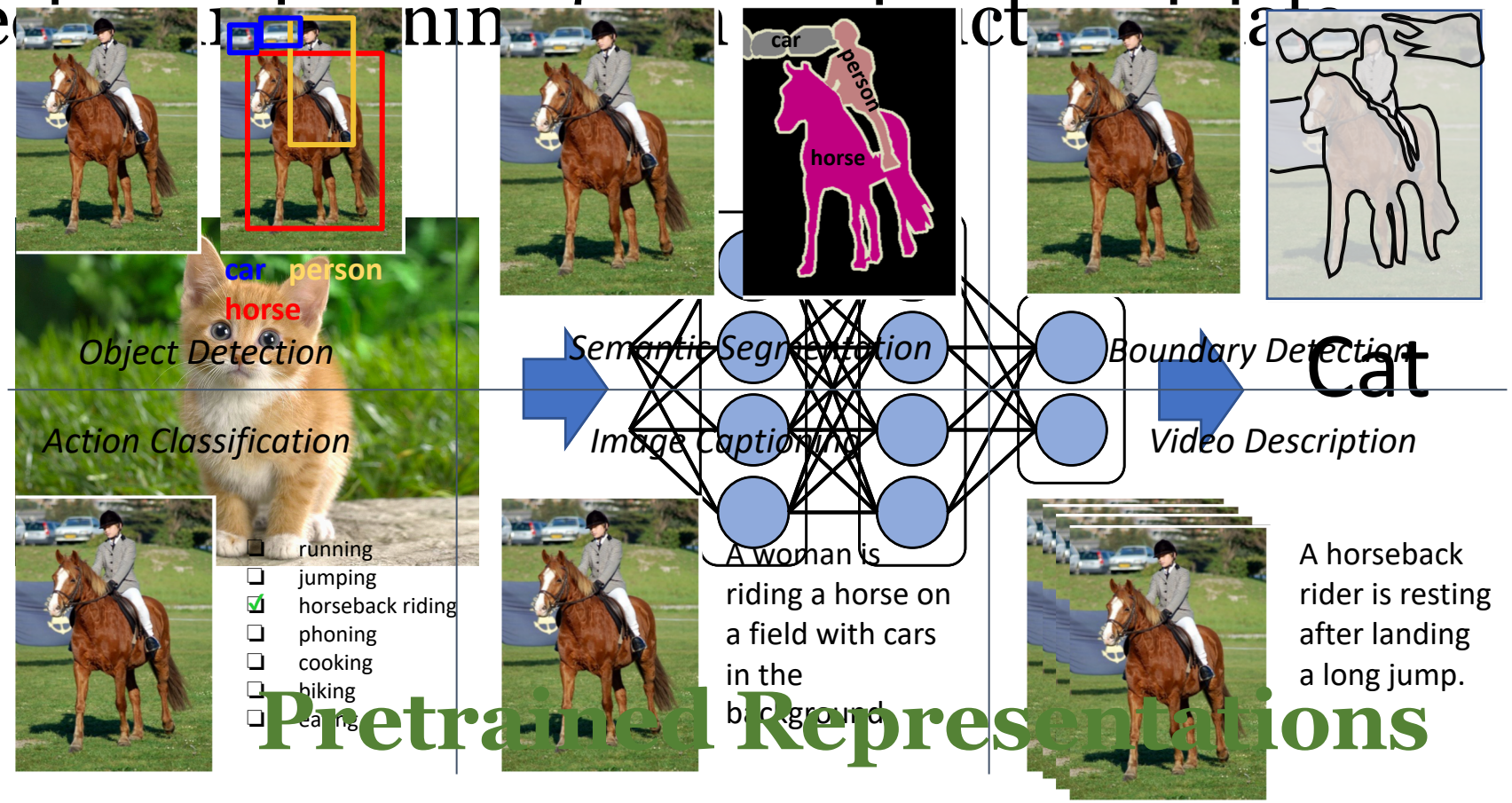
Deepak Pathak
FAIR, CMU



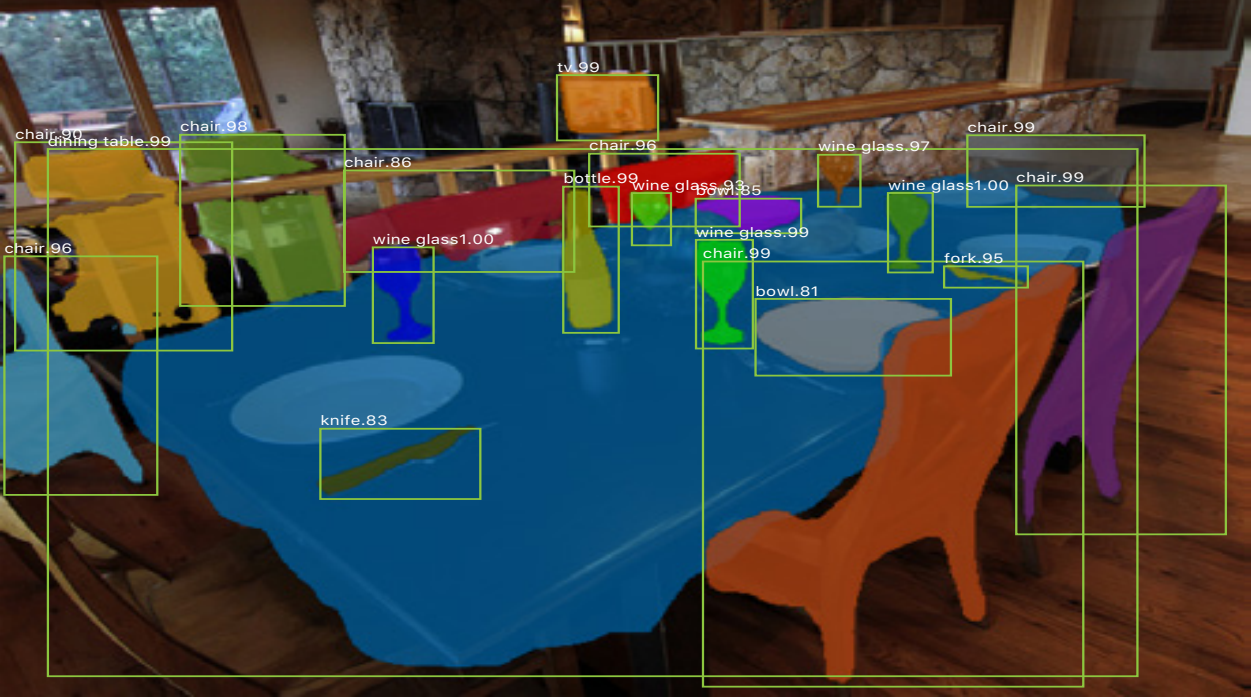
Deep Learning Success

Input	Target
Image	Task

- Effe… …nin… …ct… …a…



[Slide adapted from Jeff Donahue]

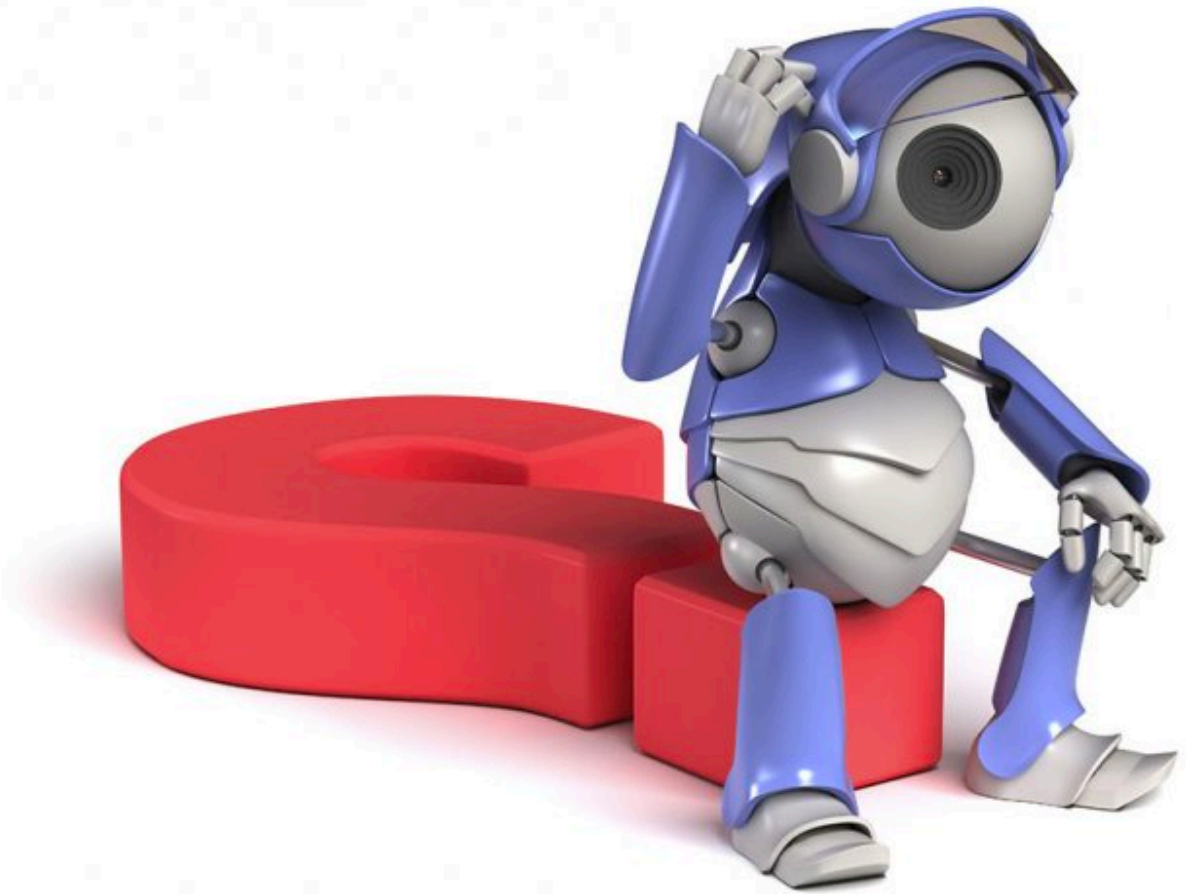


For any given problem:

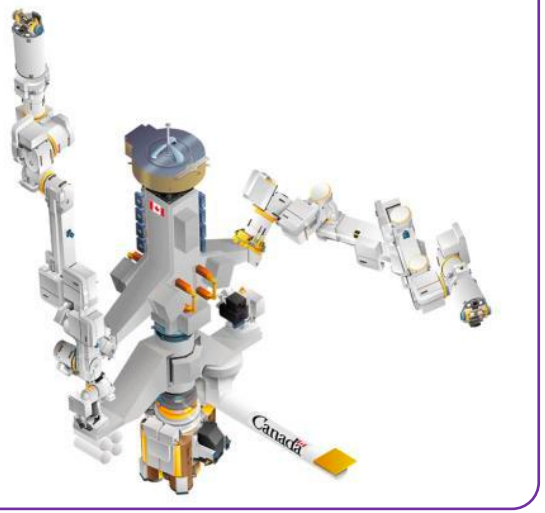
1. Download a pretrained model (ResNet, BERT etc.)
2. Label some new data for task
3. Finetune the last layer

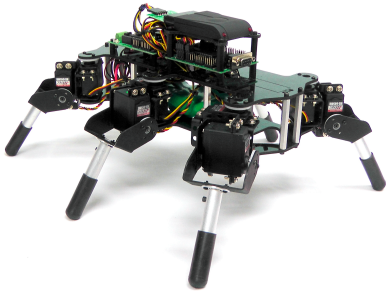
4. Declare victory!

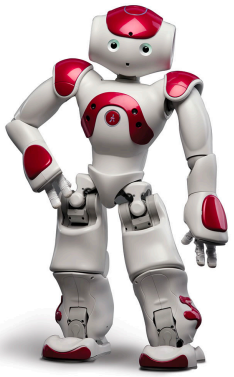
Can we translate this ***pretraining*** success to robotics?

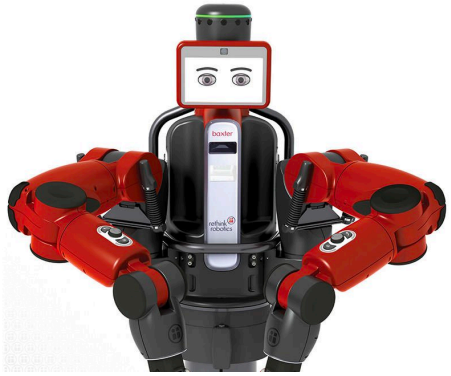


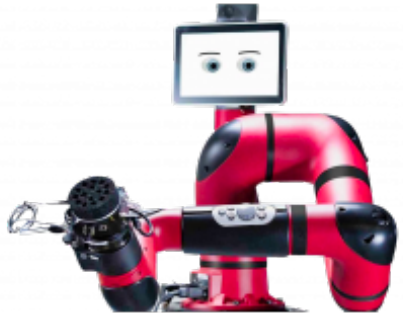
Challenging because every robot is different!





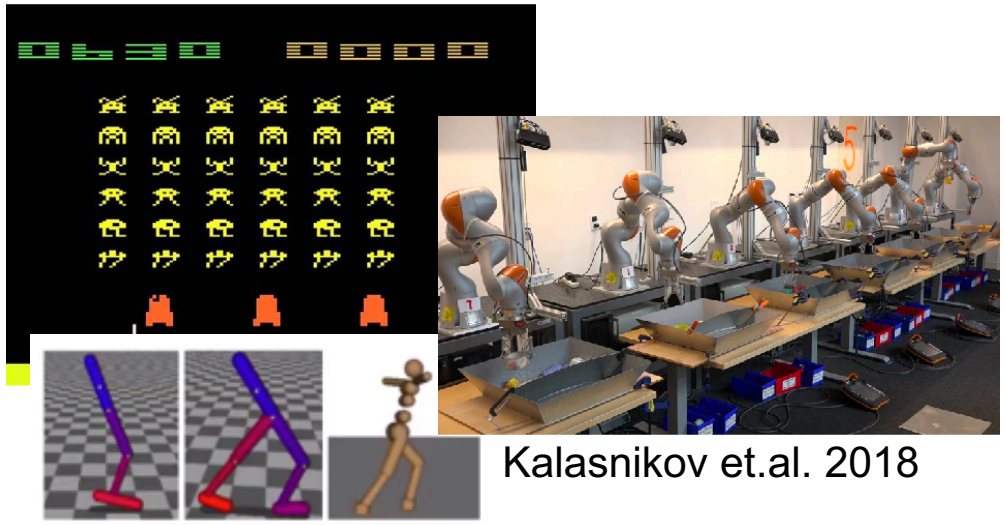






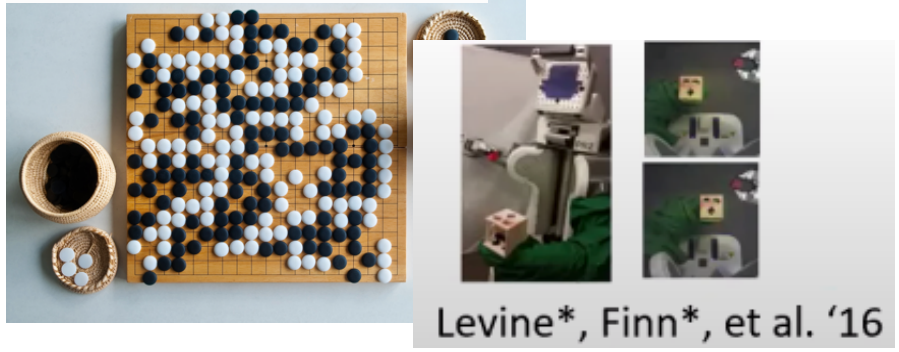
.... etc.

Deep Reinforcement Learning for Robot Control



Kalasnikov et.al. 2018

Schulman et al. ’14 & ‘15



Levine*, Finn*, et al. ‘16

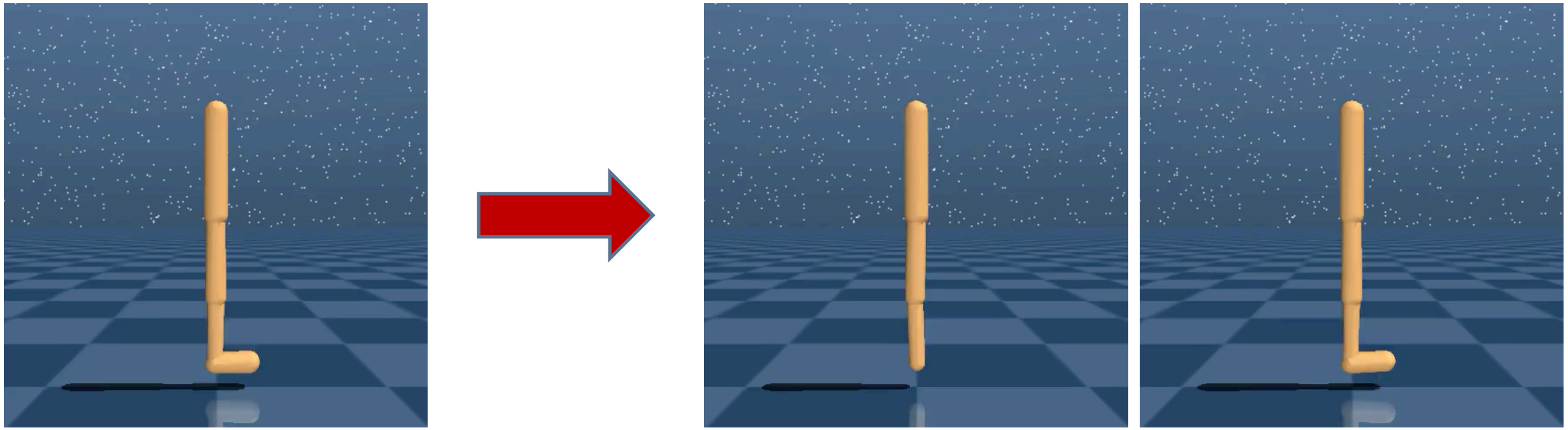
A separate policy is trained for each robotics setup.

How well does it generalize?

Reinforcement Learning for Robot Control

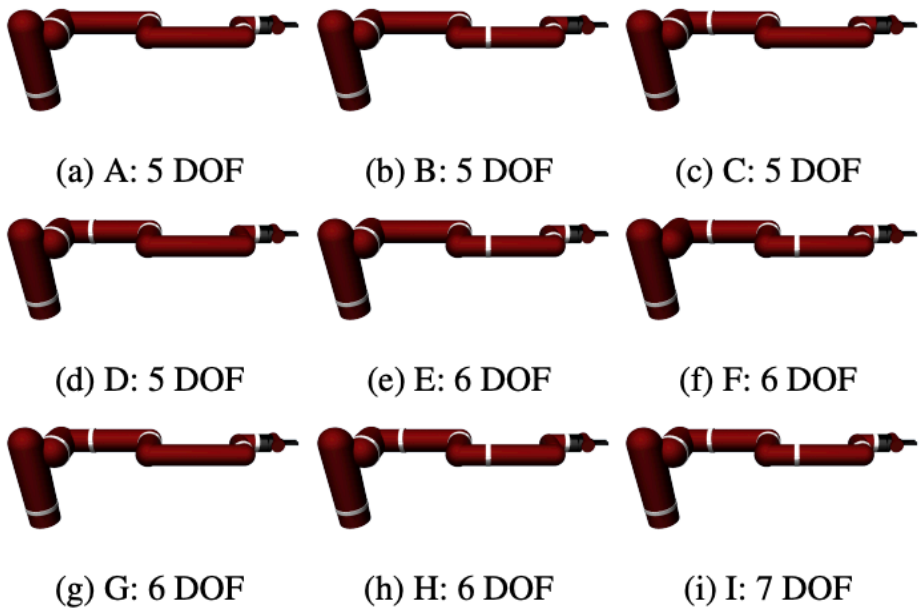
- Lots of tuning needed even for one robotic agent
- Yet, it doesn't generalize!

What about training on multiple robots?

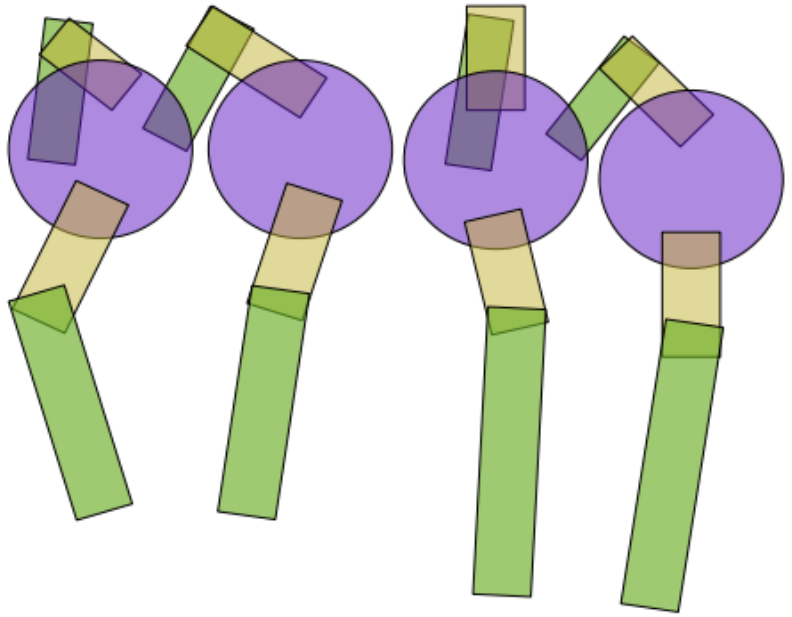


Prior Attempts

Same topology!

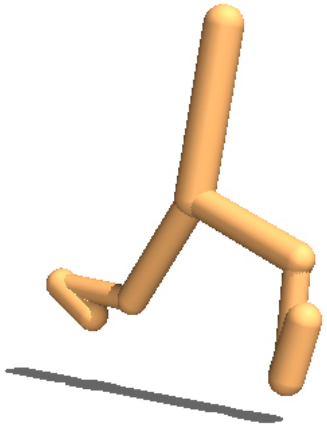


[Chen et.al. NeurIPS 2018]

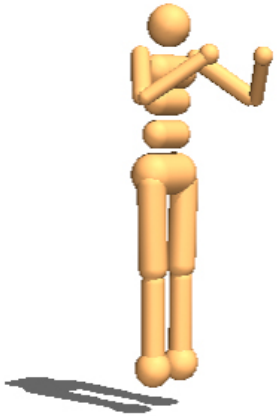


[Wang et.al. ICLR 2018]

How to train same policy for these different shapes?



Same Policy?

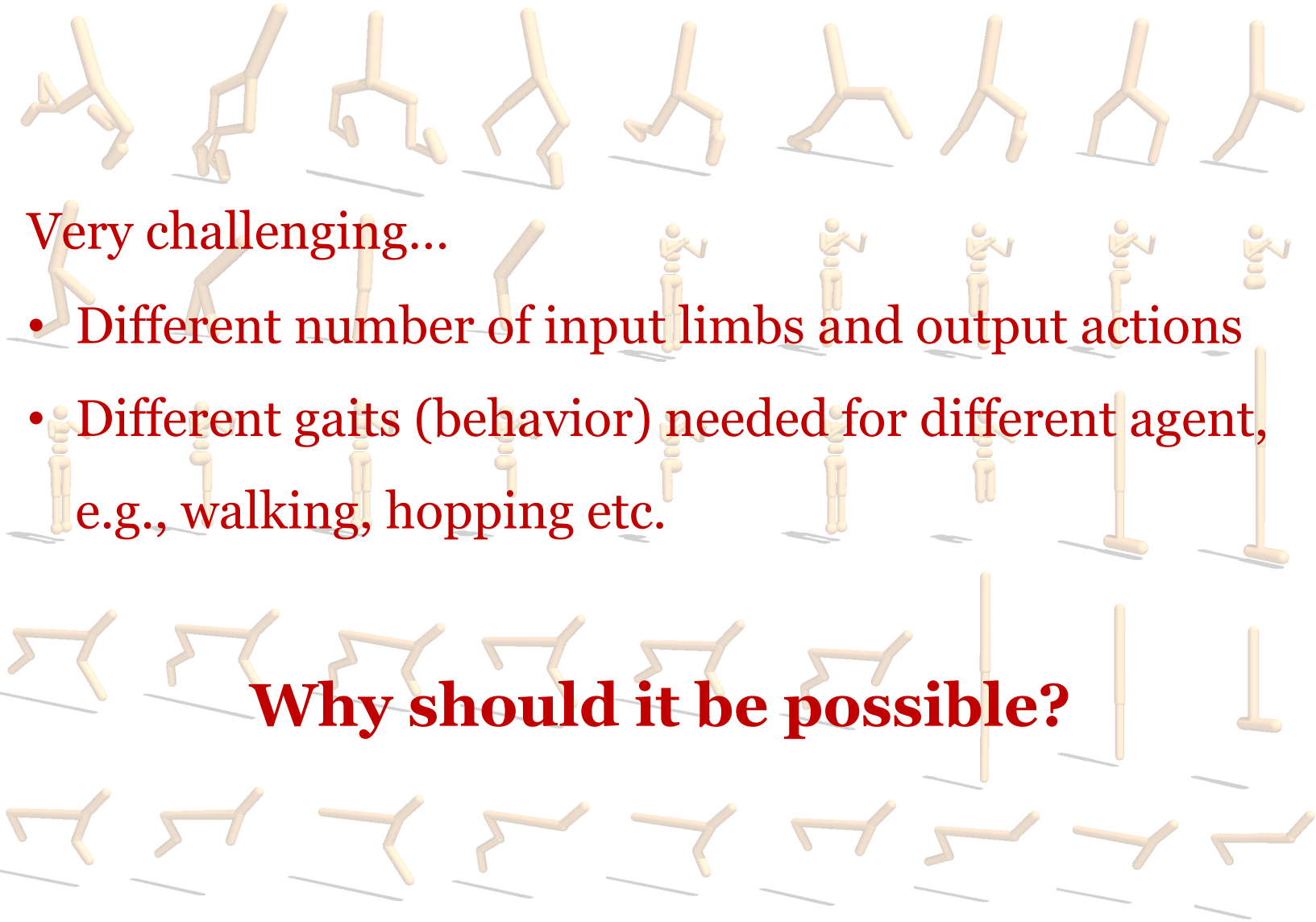


Walker

Humanoid

This work

- One policy for several agents
- Zero-shot generalize to new agents!

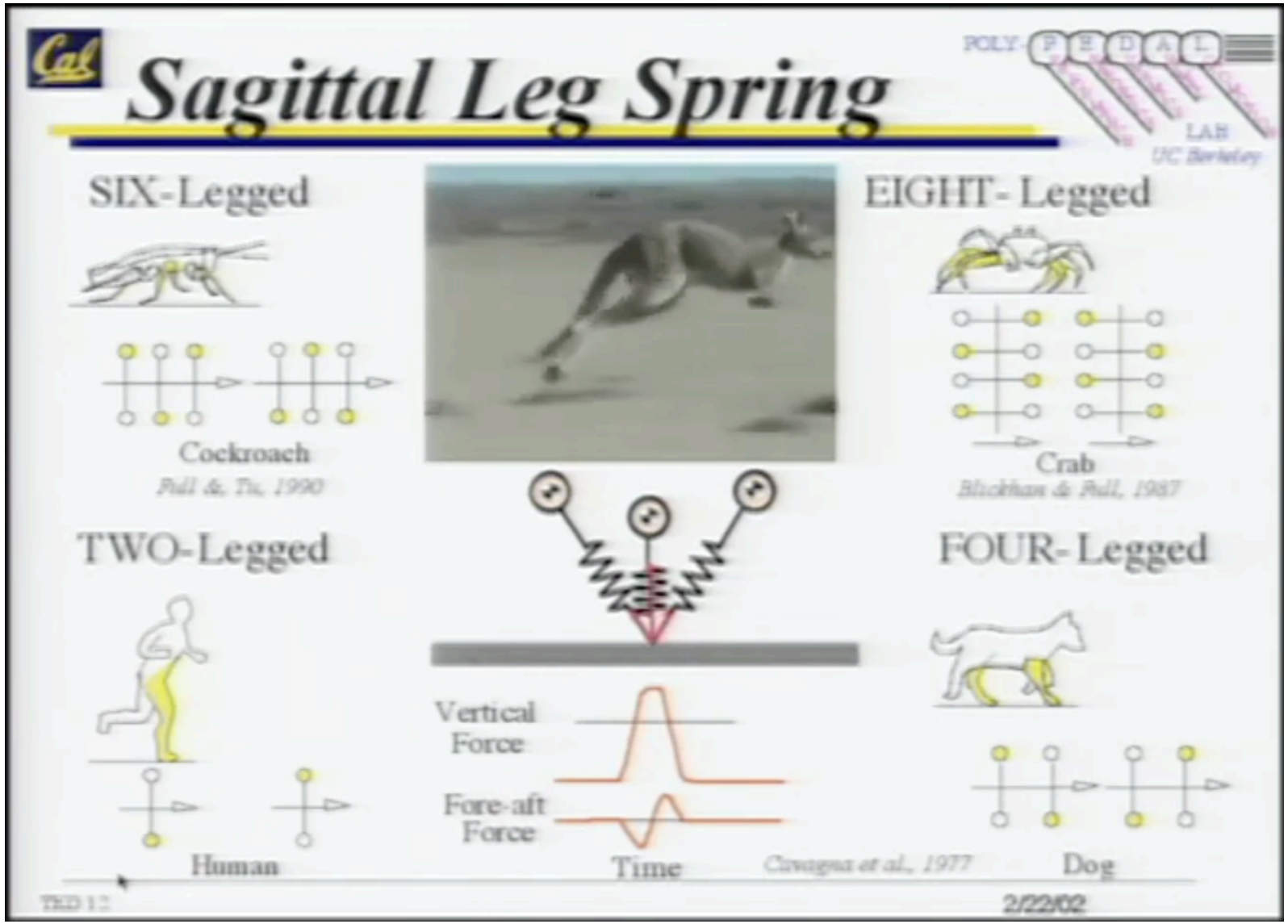


Very challenging…

- Different number of input limbs and output actions
- Different gaits (behavior) needed for different agent, e.g., walking, hopping etc.

Why should it be possible?

Shared Locomotion patterns in Biology



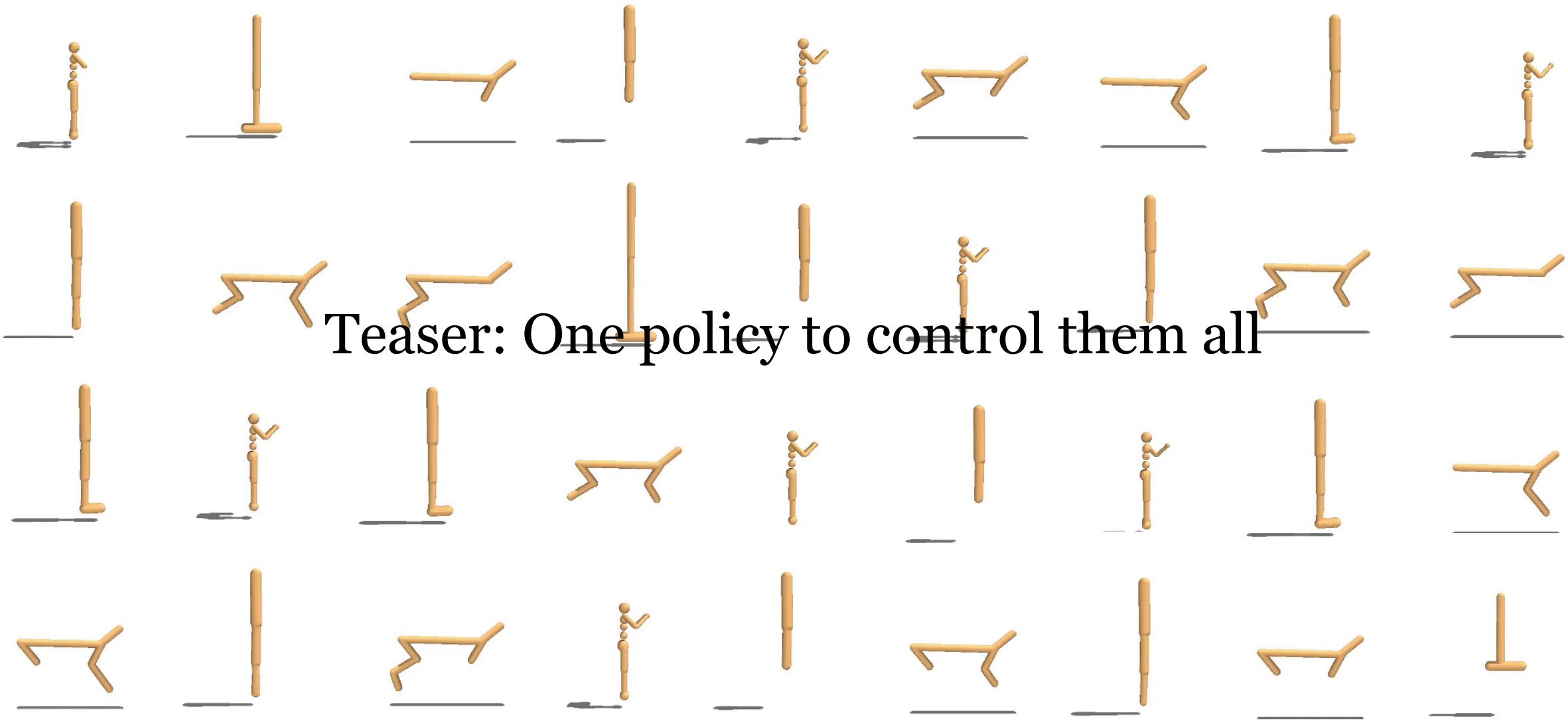
[slide from Robert Full]

Precocial Animals -- horses, giraffes, zebras etc.

- “foals can walk as fast as their parents within hours”

1-2 hrs after birth

4 days after birth

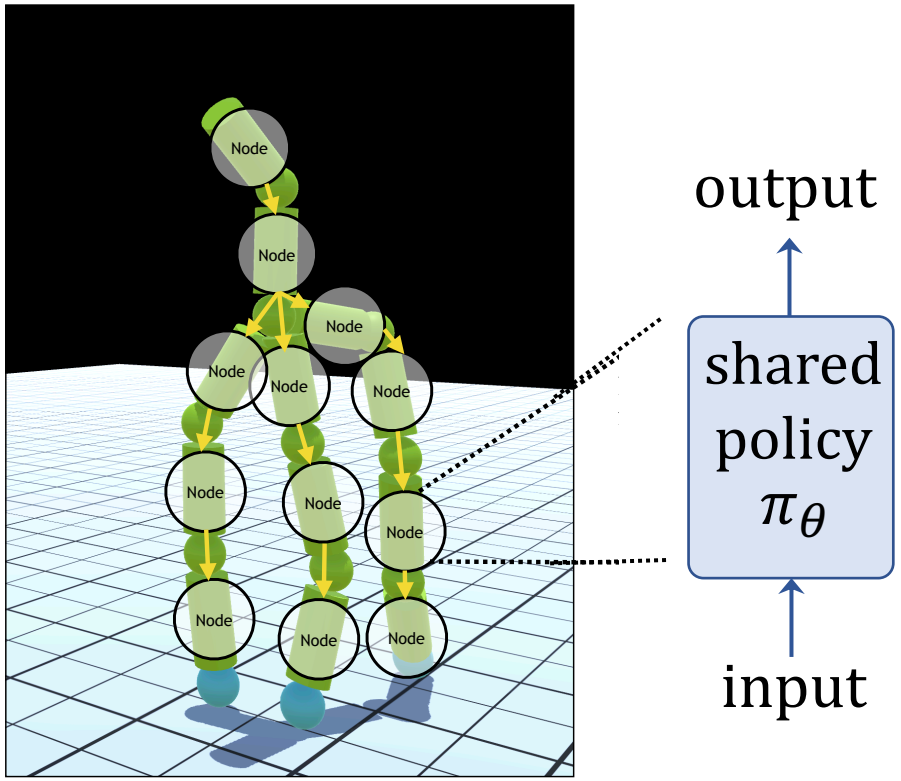


Teaser: One policy to control them all

How do we train a single policy for all?

INSPIRATION

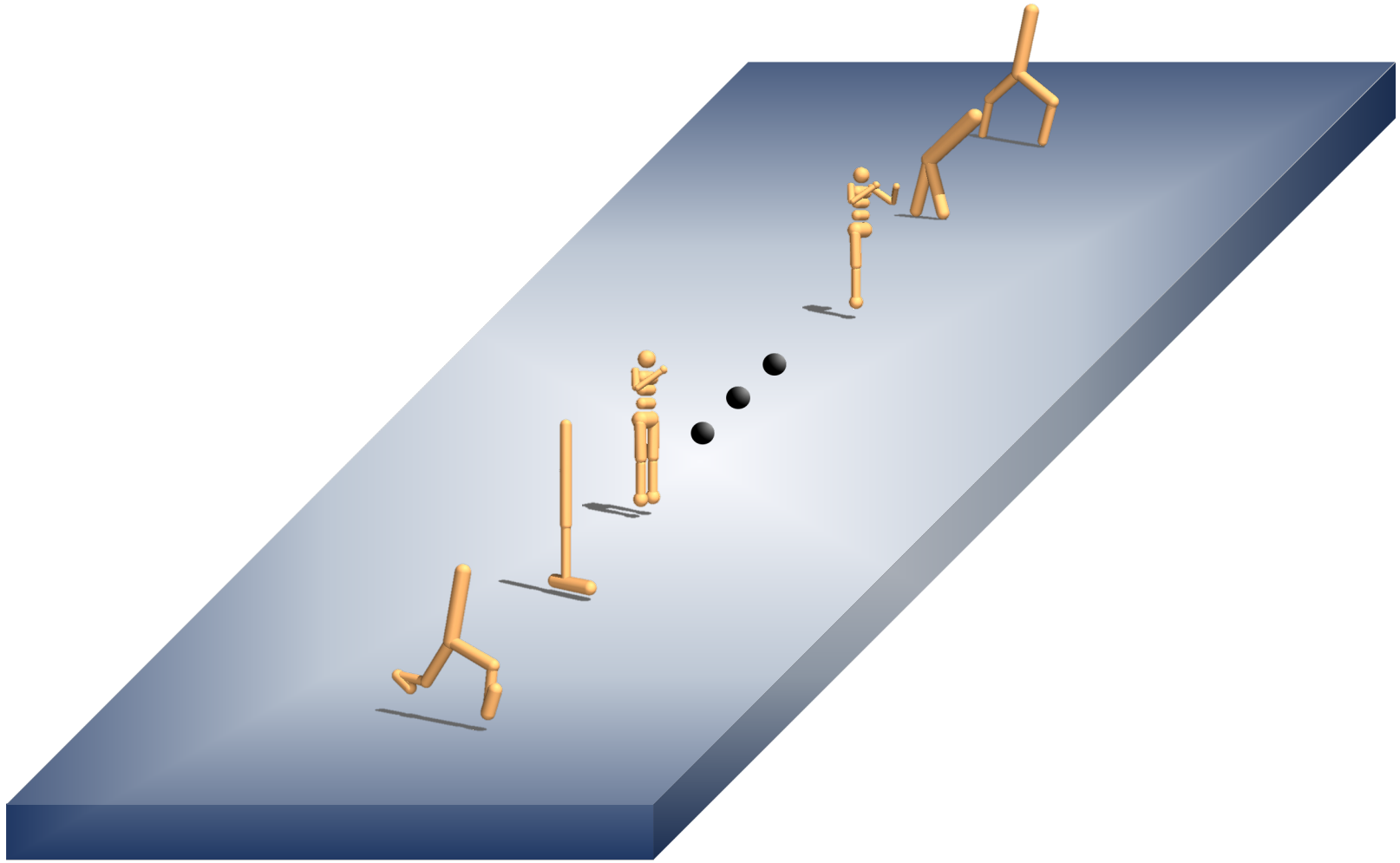
Train the controller for each motor/limb and share it across all the motors/limbs.

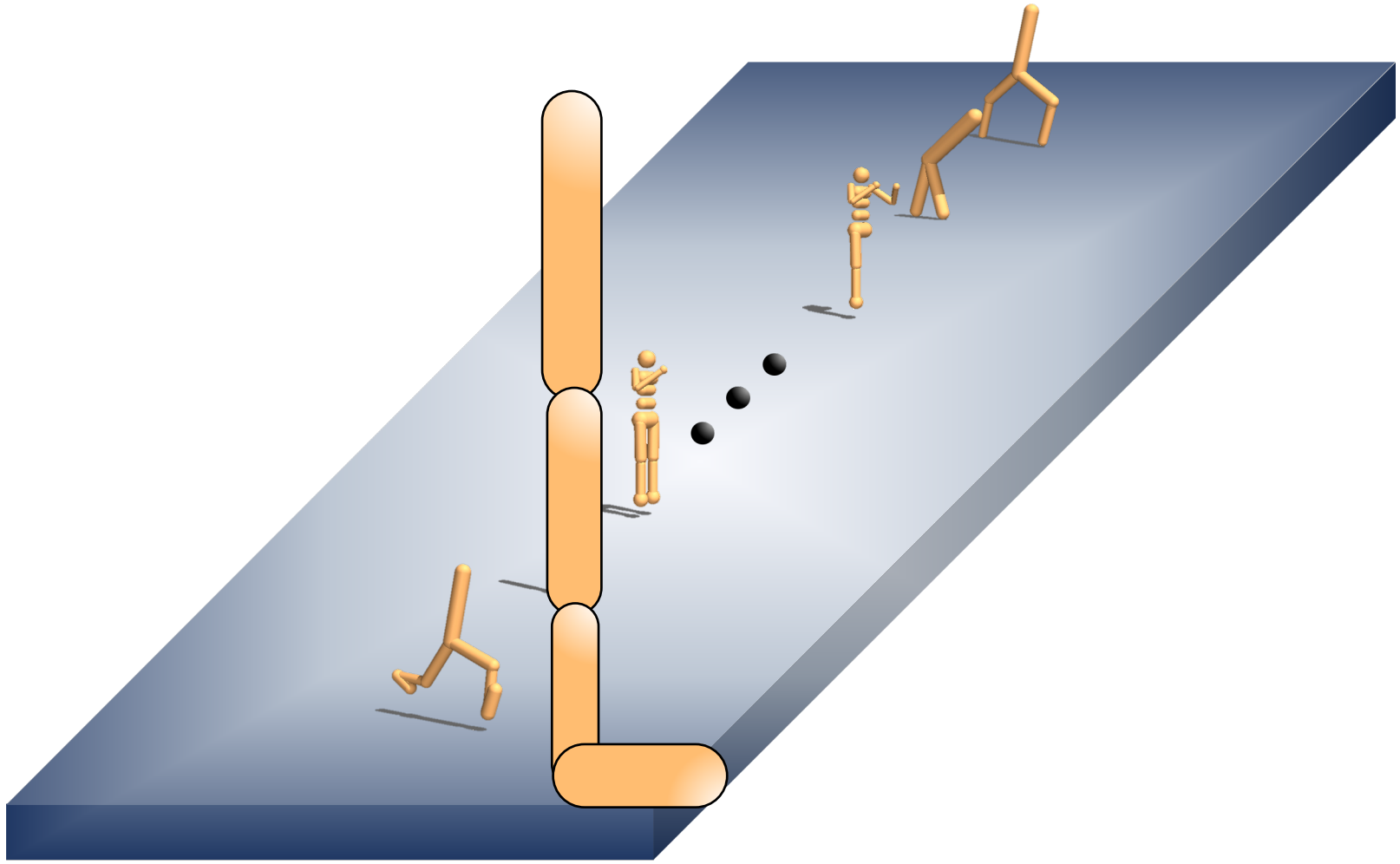


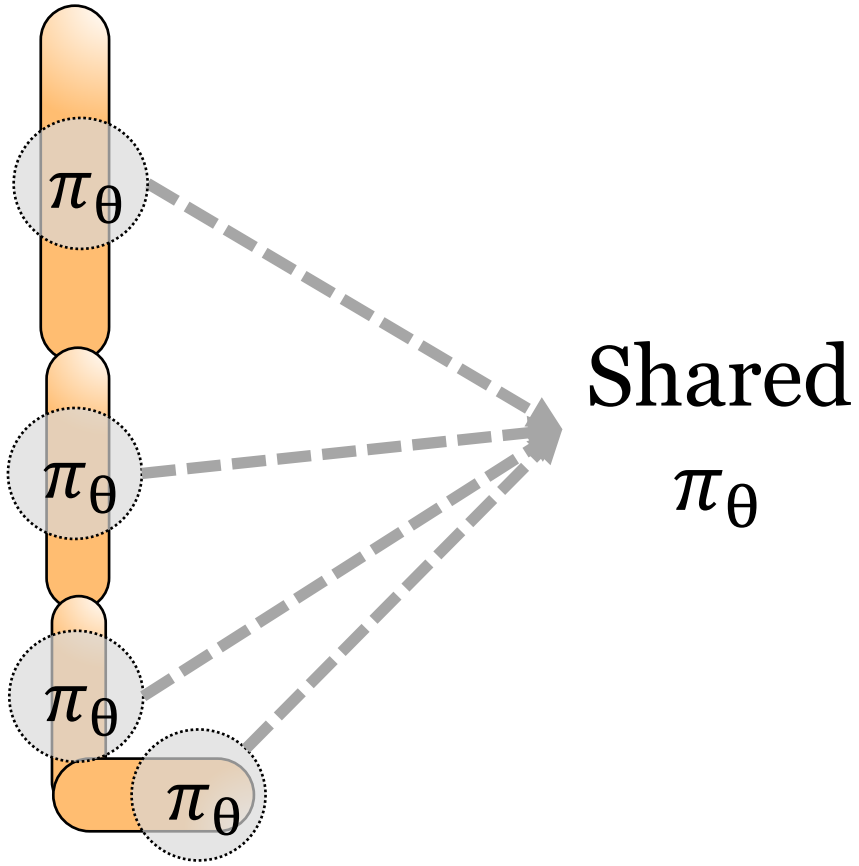
Pathak*, Lu*, Darrell, Isola, Efros. NeurIPS 2019.

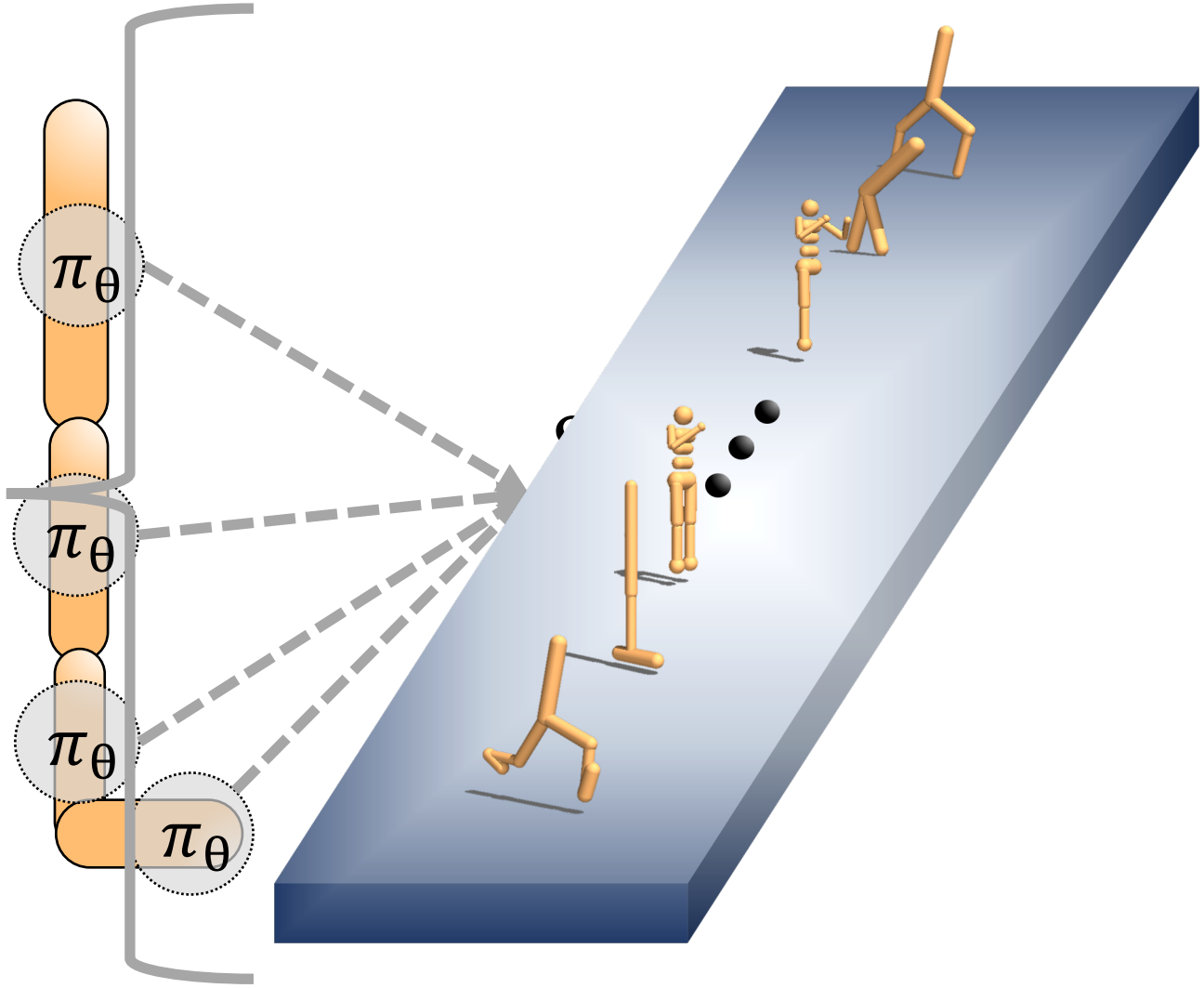
Modularity at the level of software and hardware!

[slides adapted from Pathak et al., NeurIPS 2019]



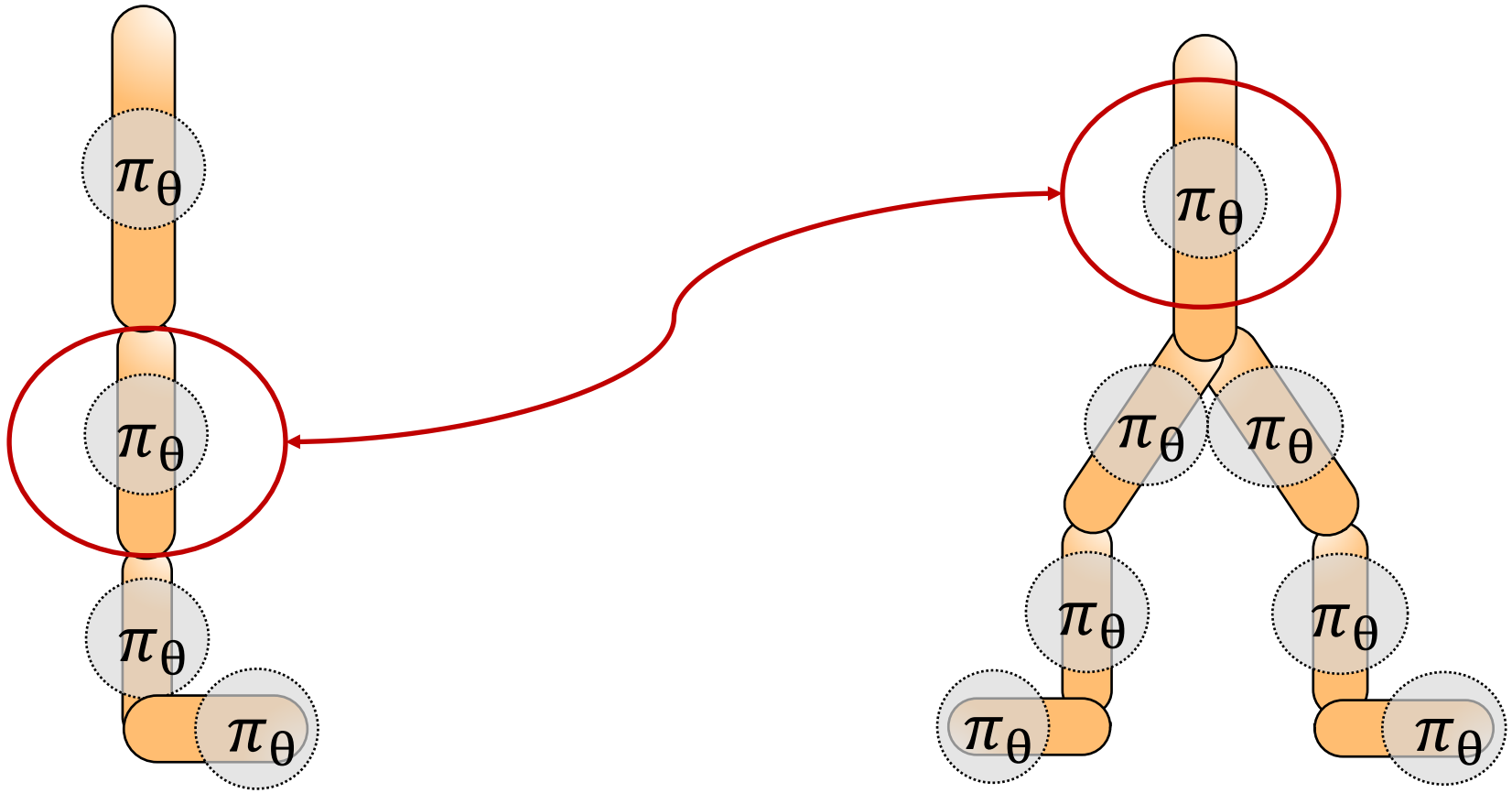



π_θ
π_θ
π_θ
π_θ
Shared
π_θ


π_θ
π_θ
π_θ
π_θ

How to train the same policy for different agents?

Hopper

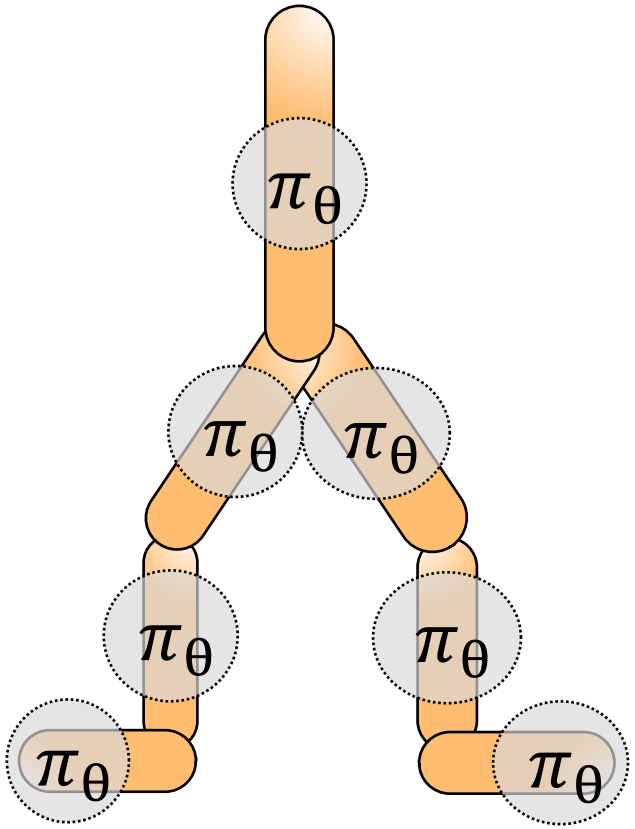


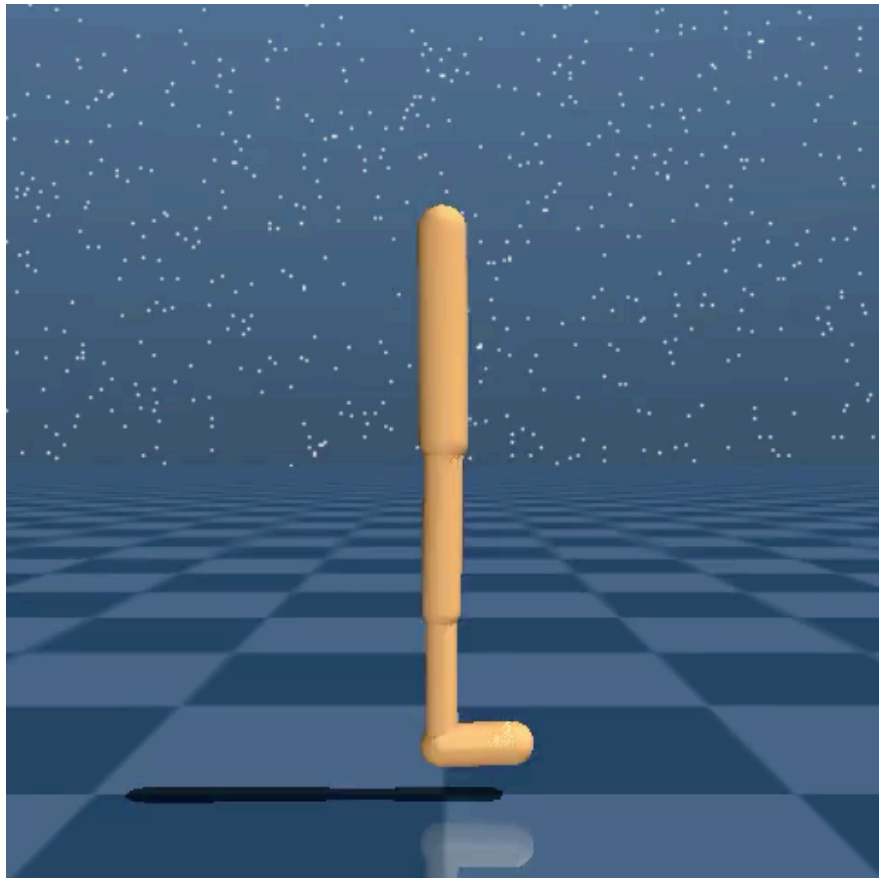
Walker

Same network is applied at each limb/motor.

[slides adapted from Pathak et al., NeurIPS 2019]

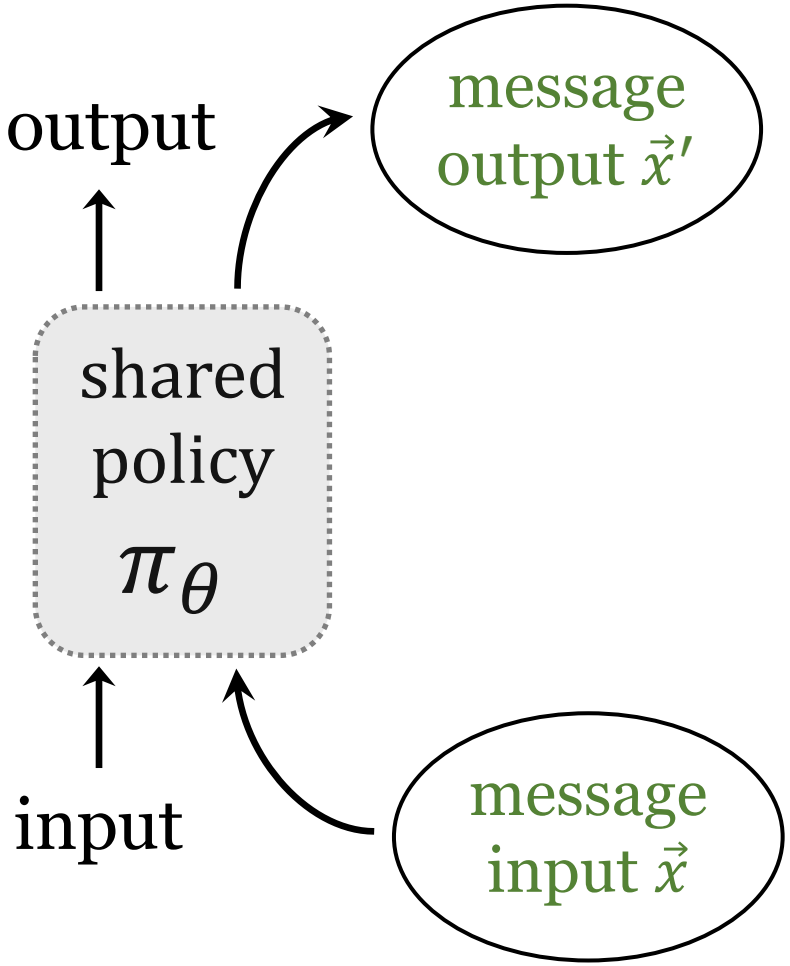
Does it just work like that?





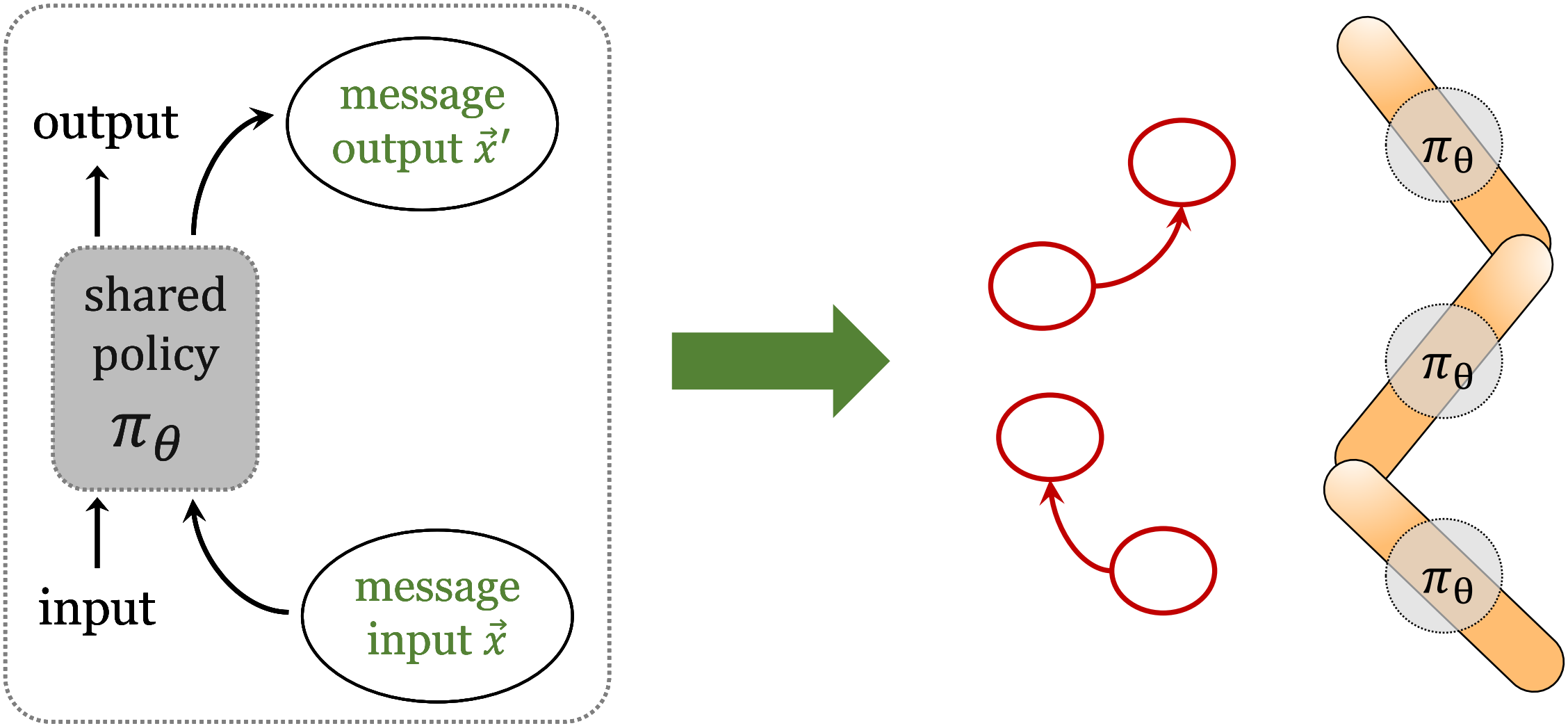
No global coordination!

Network as reusable and communicating LEGO Blocks



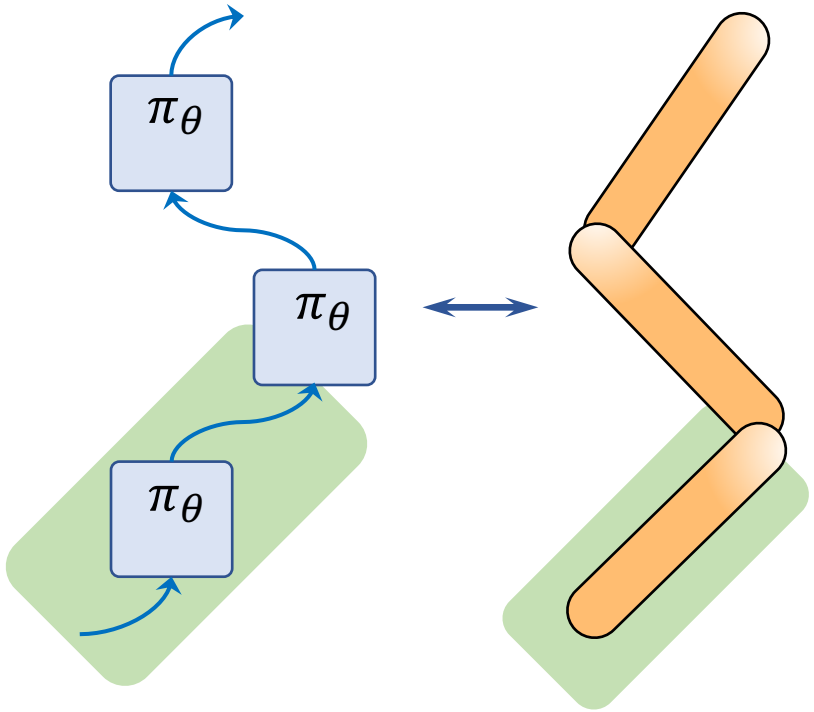
[slides adapted from Pathak et al., NeurIPS 2019]

Network as reusable and communicating LEGO Blocks



[slides adapted from Pathak et al., NeurIPS 2019]

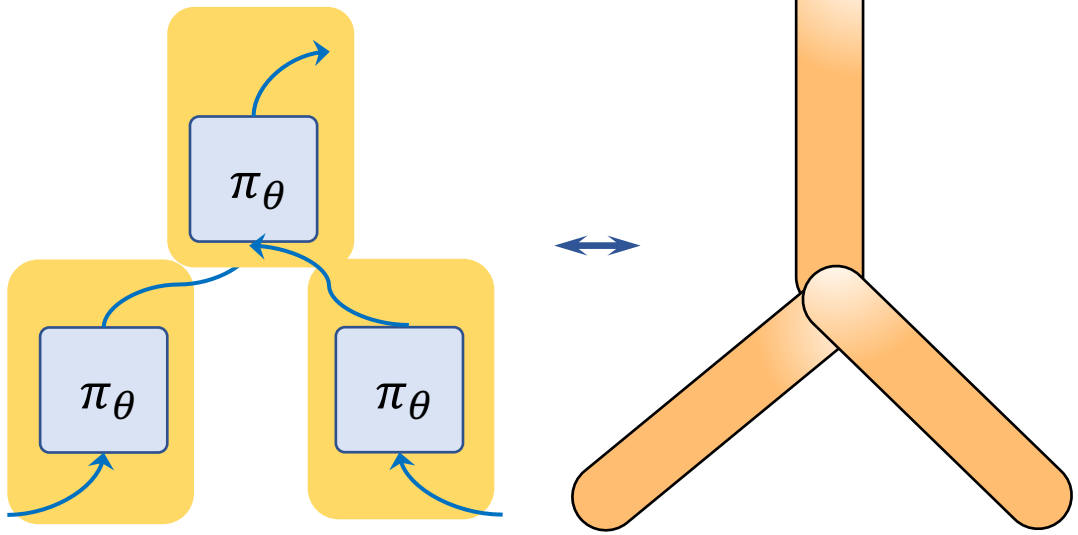
Network as reusable and communicating LEGO Blocks



copy

and

paste

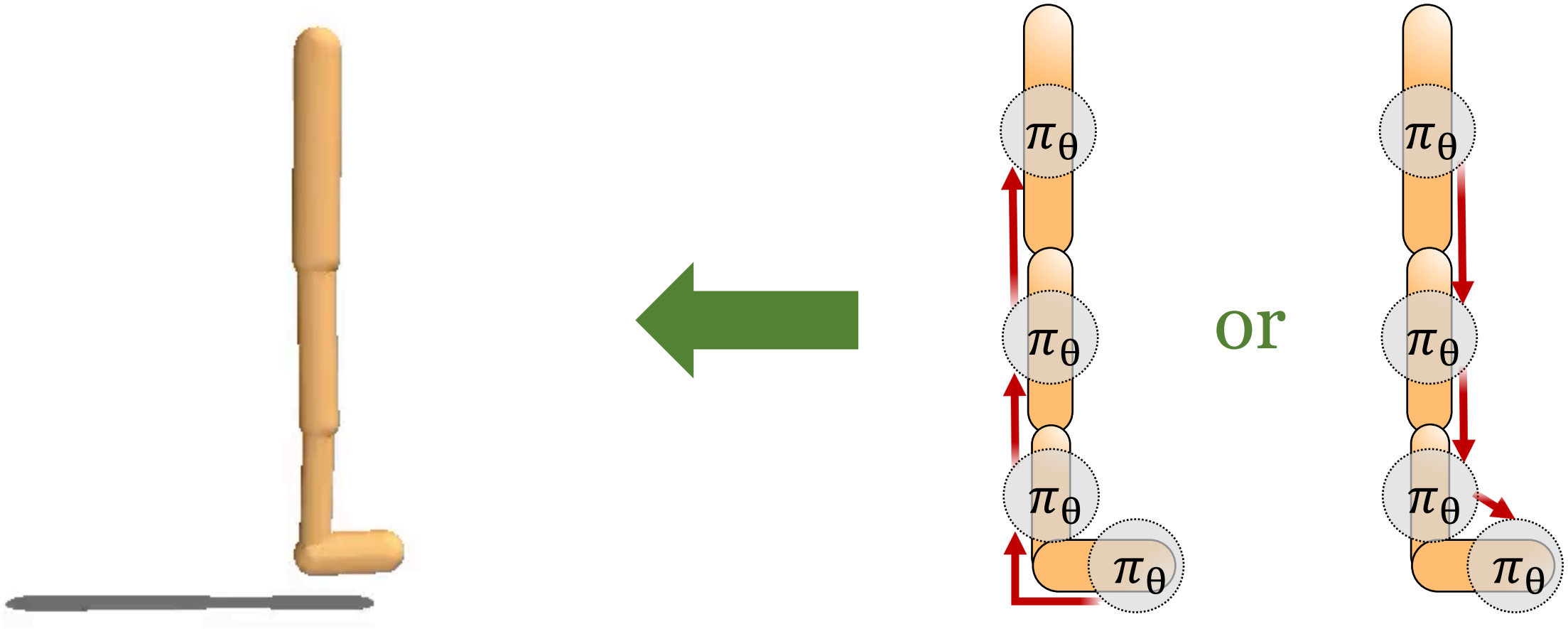


Trained Robot

New Robot at Test Time

[slides adapted from Pathak et al., NeurIPS 2019]

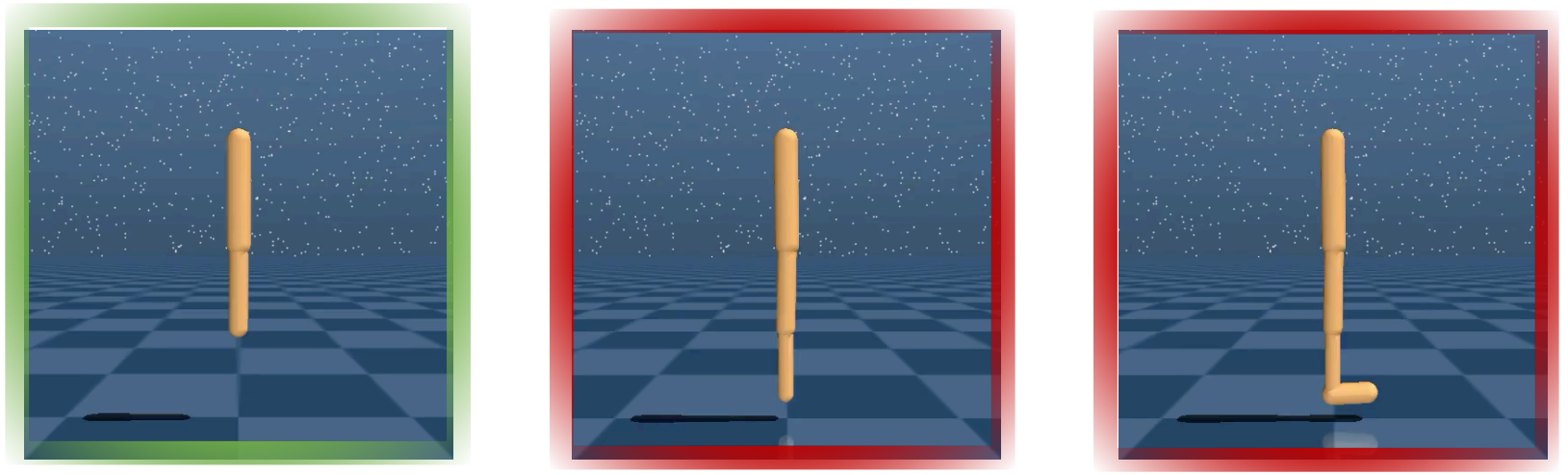
Message Passing Scheme



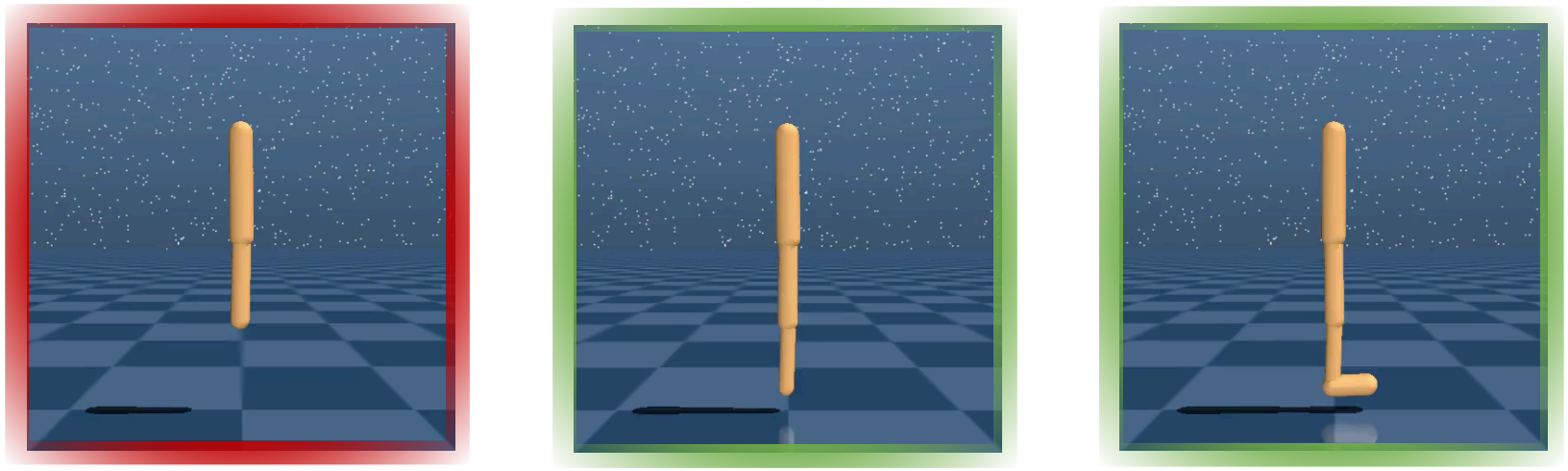
One-way: Leaves to root or root to leaves?

Limitation of One-Way Message Passing

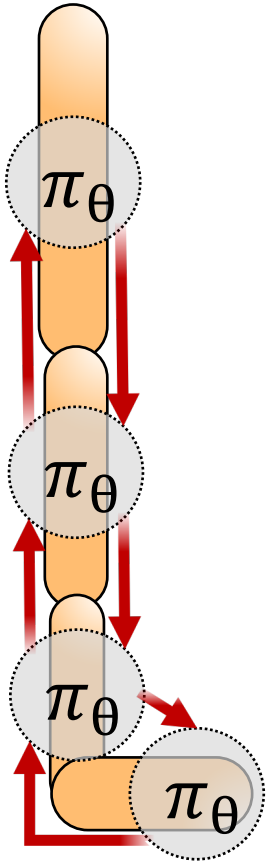
Hopping Gait



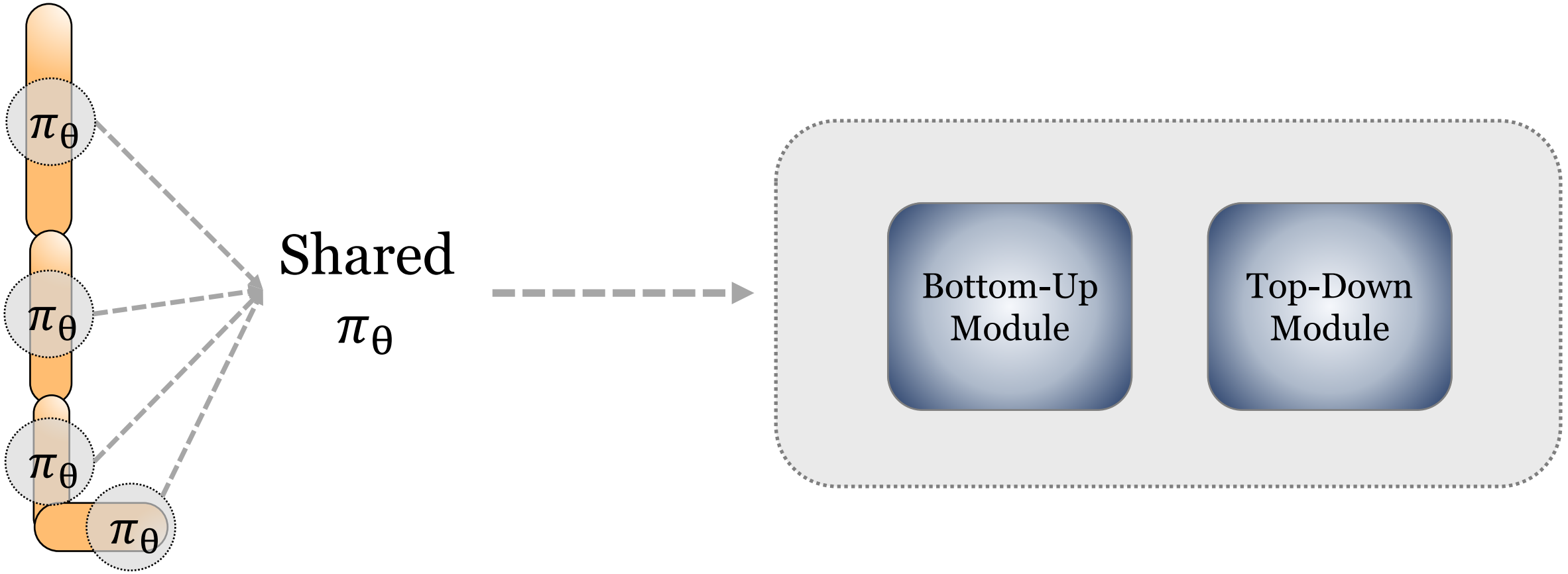
Walking Gait



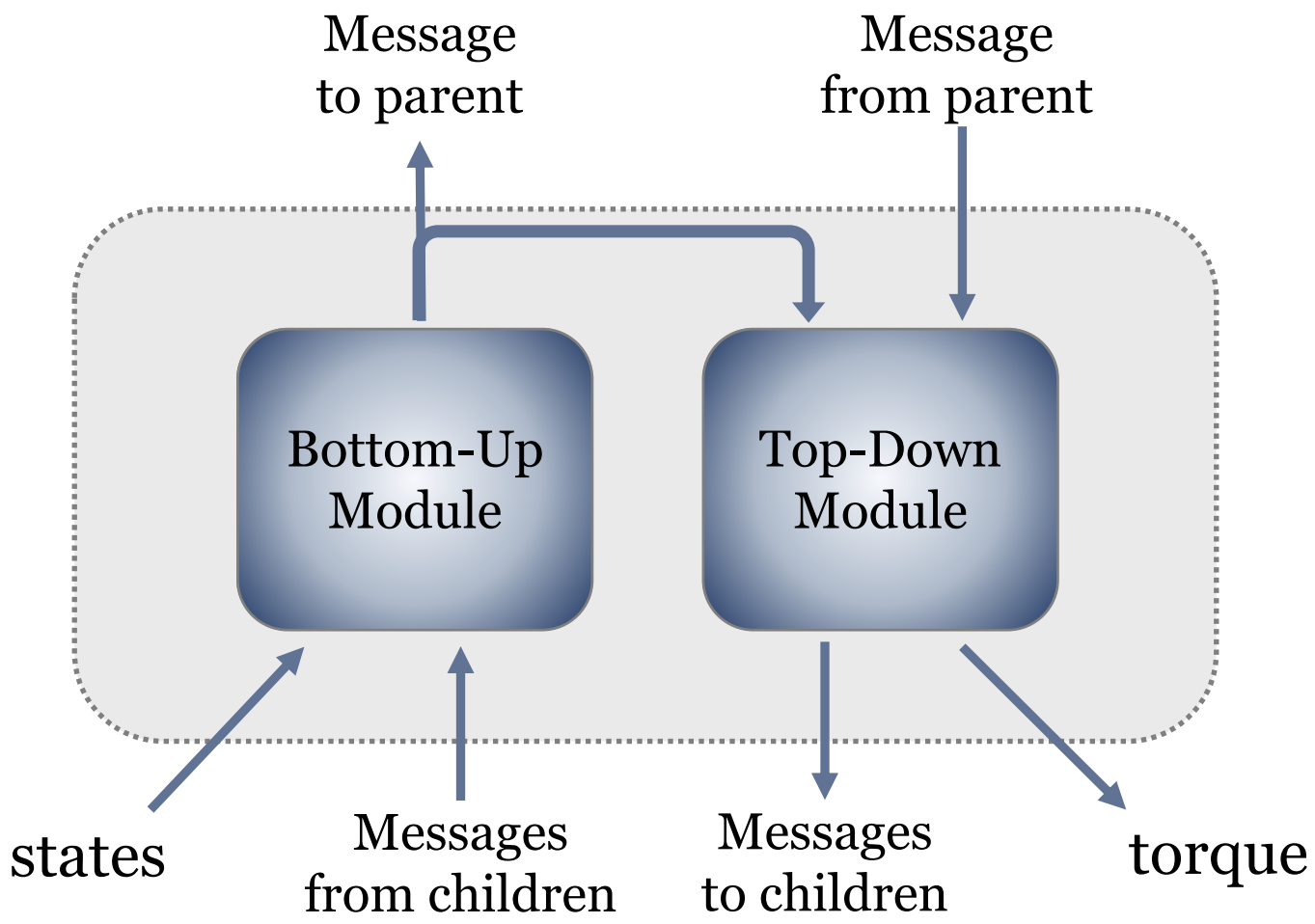
Centralized Control via ***Both-way*** Message Passing



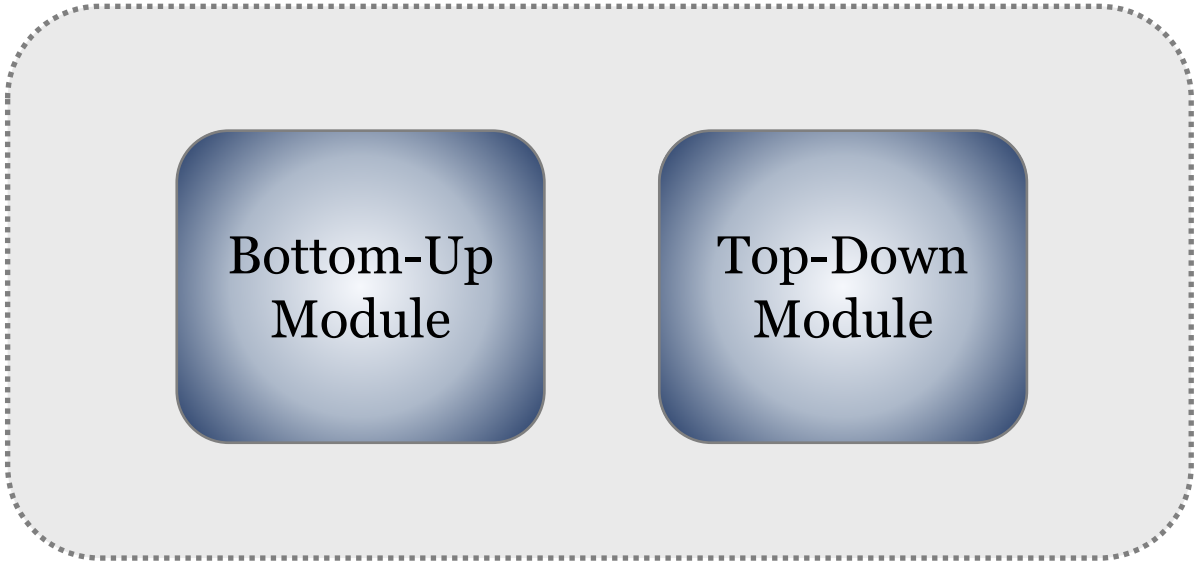
Shared Modular Policies with ***Both-way*** Message Passing

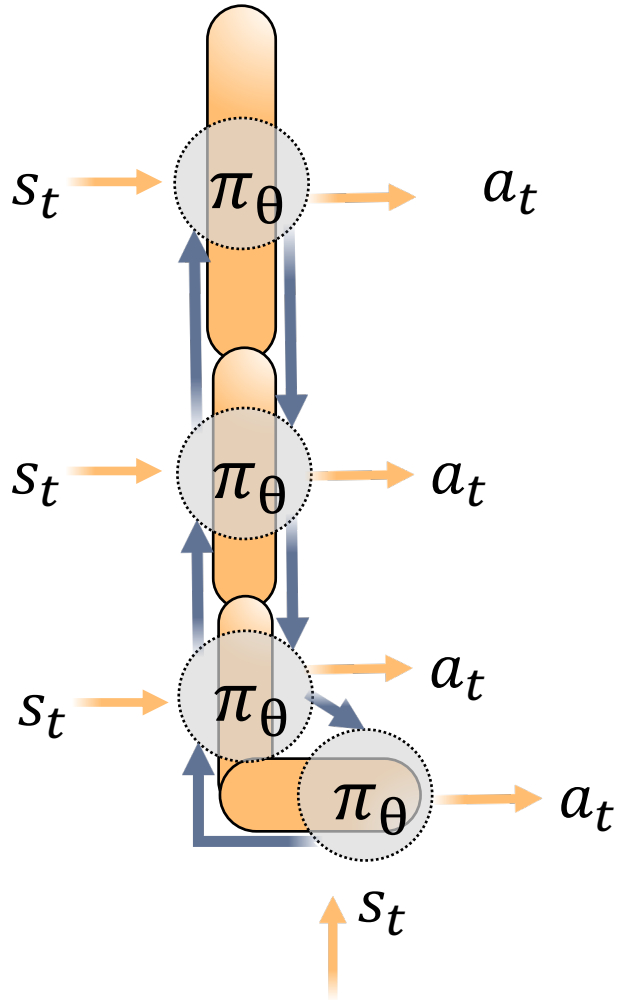


Shared Modular Policies with ***Both-way*** Message Passing



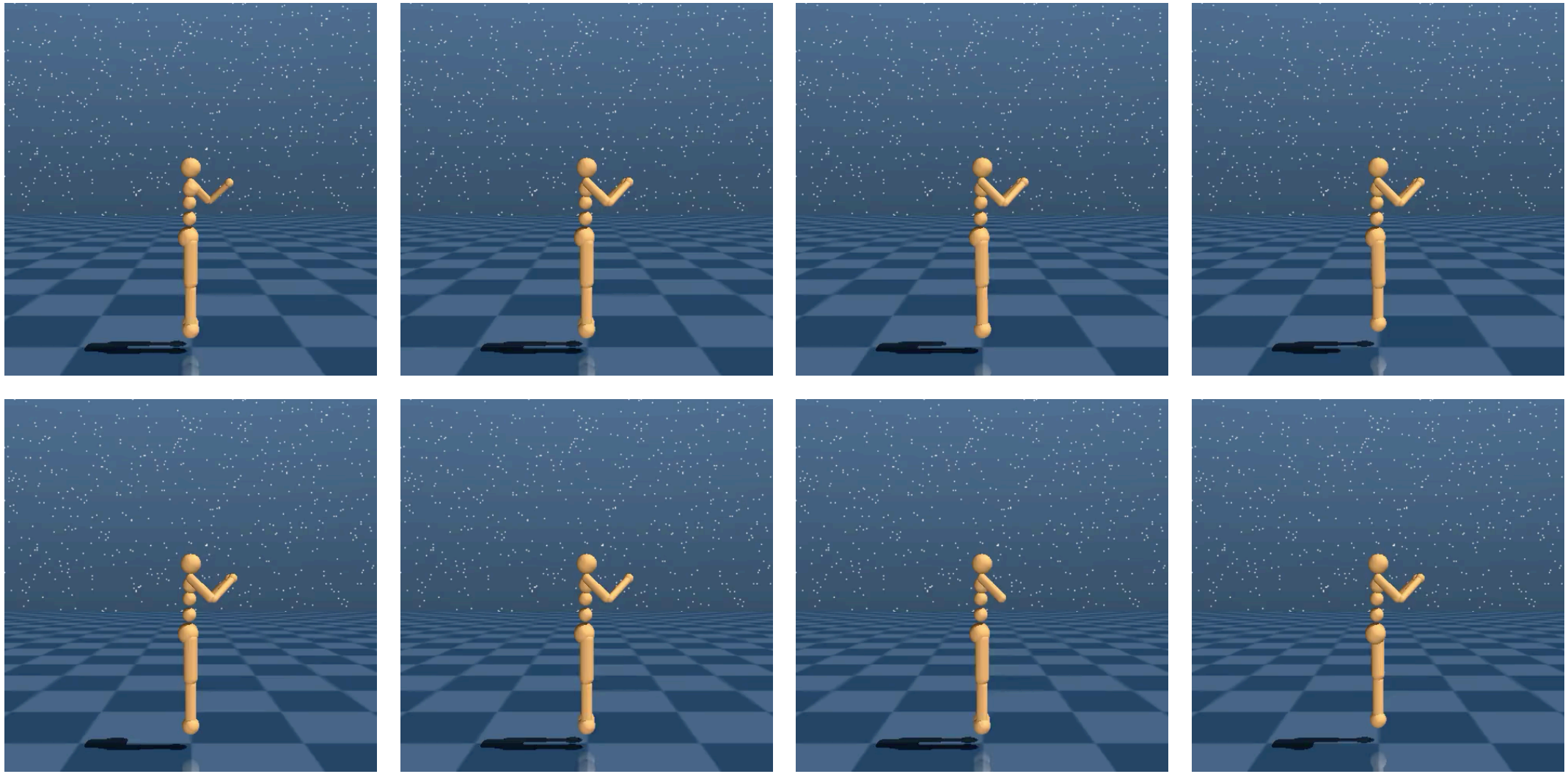
Example of Shared Modular Policies



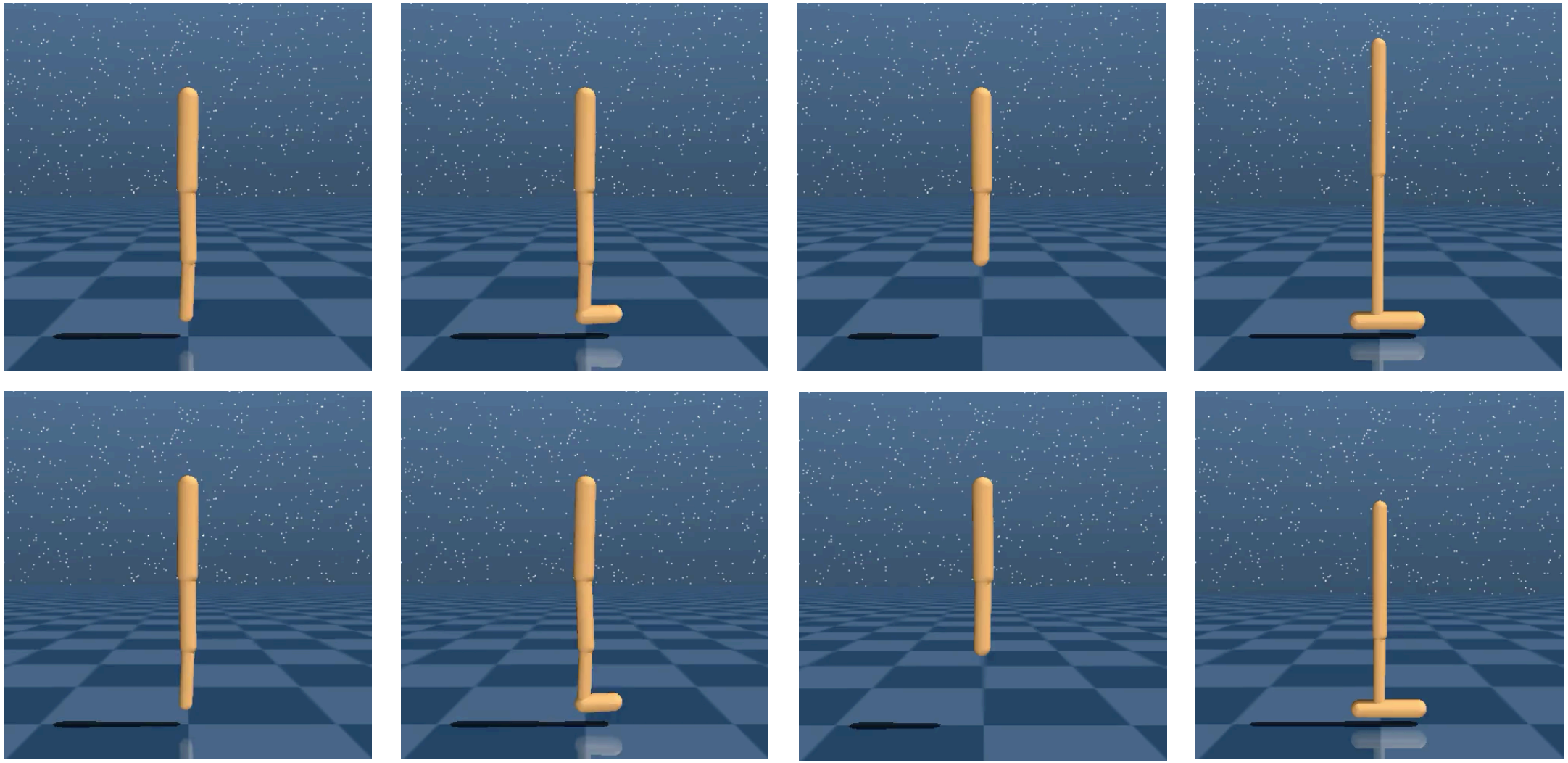


Results

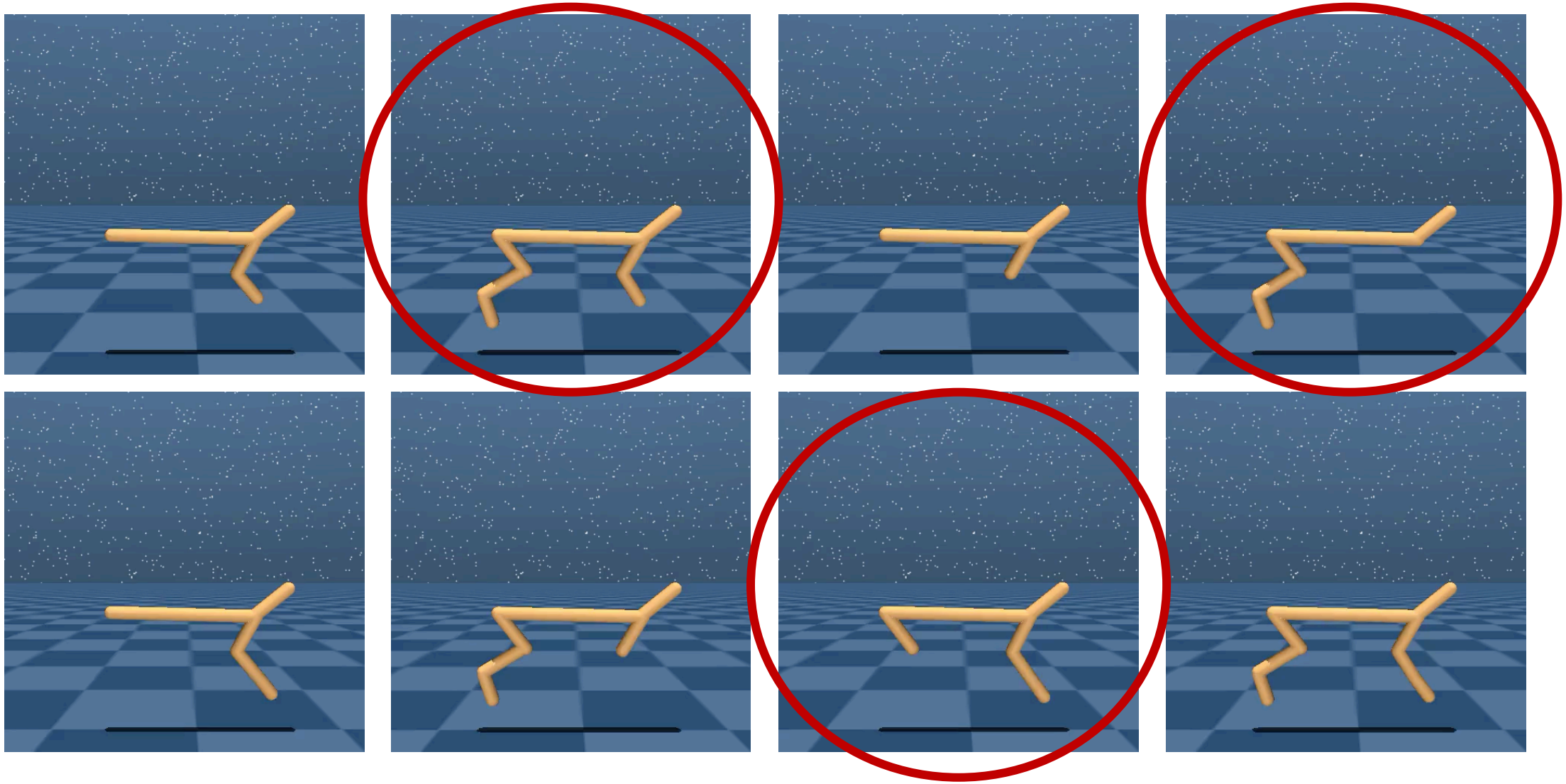
One single controller for different variants of 2D humanoid



One single controller for different variants of walker and hopper

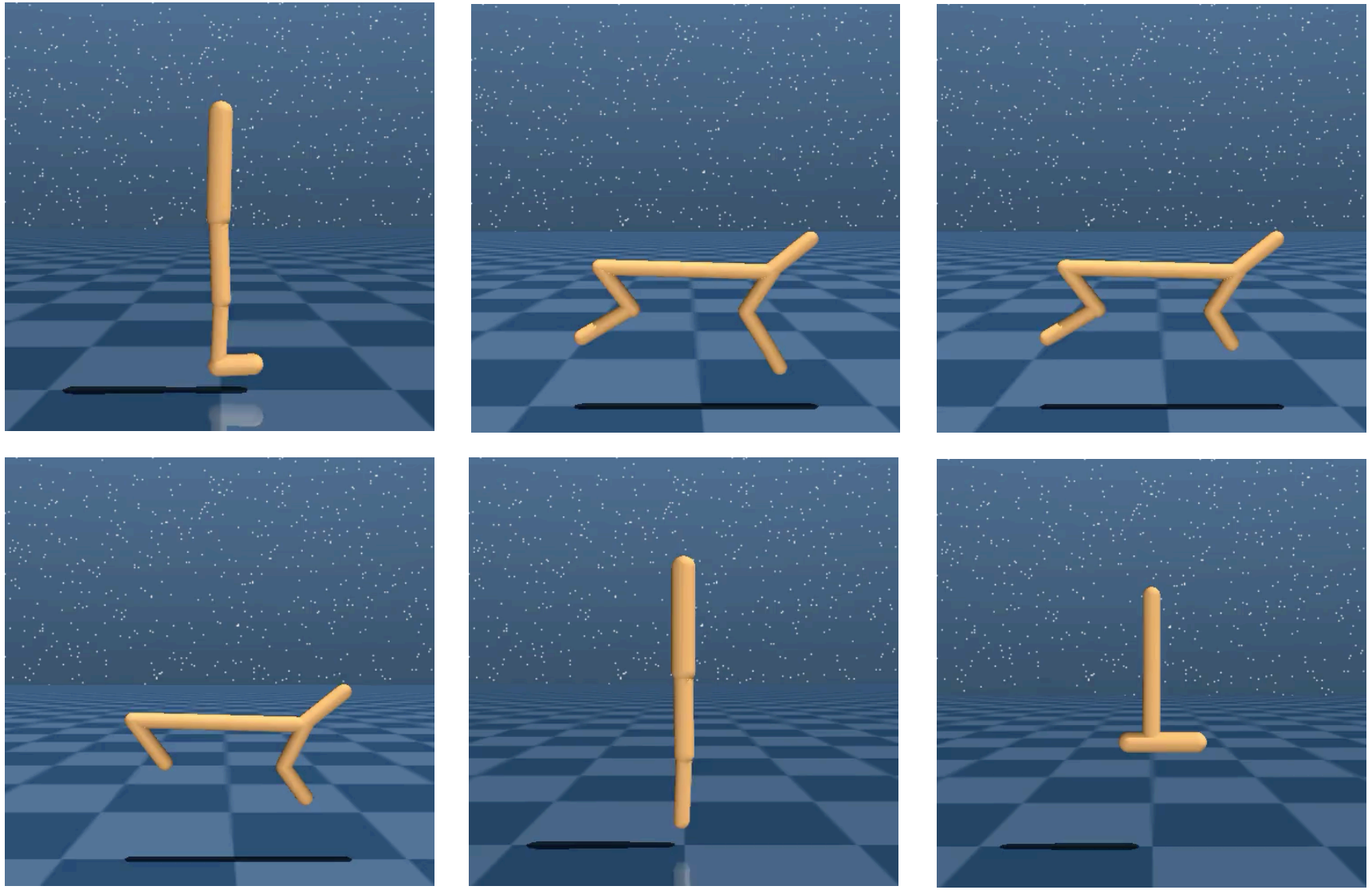


One single controller for different variants of cheetah



Zero-Shot Generalization to Unseen Agents!

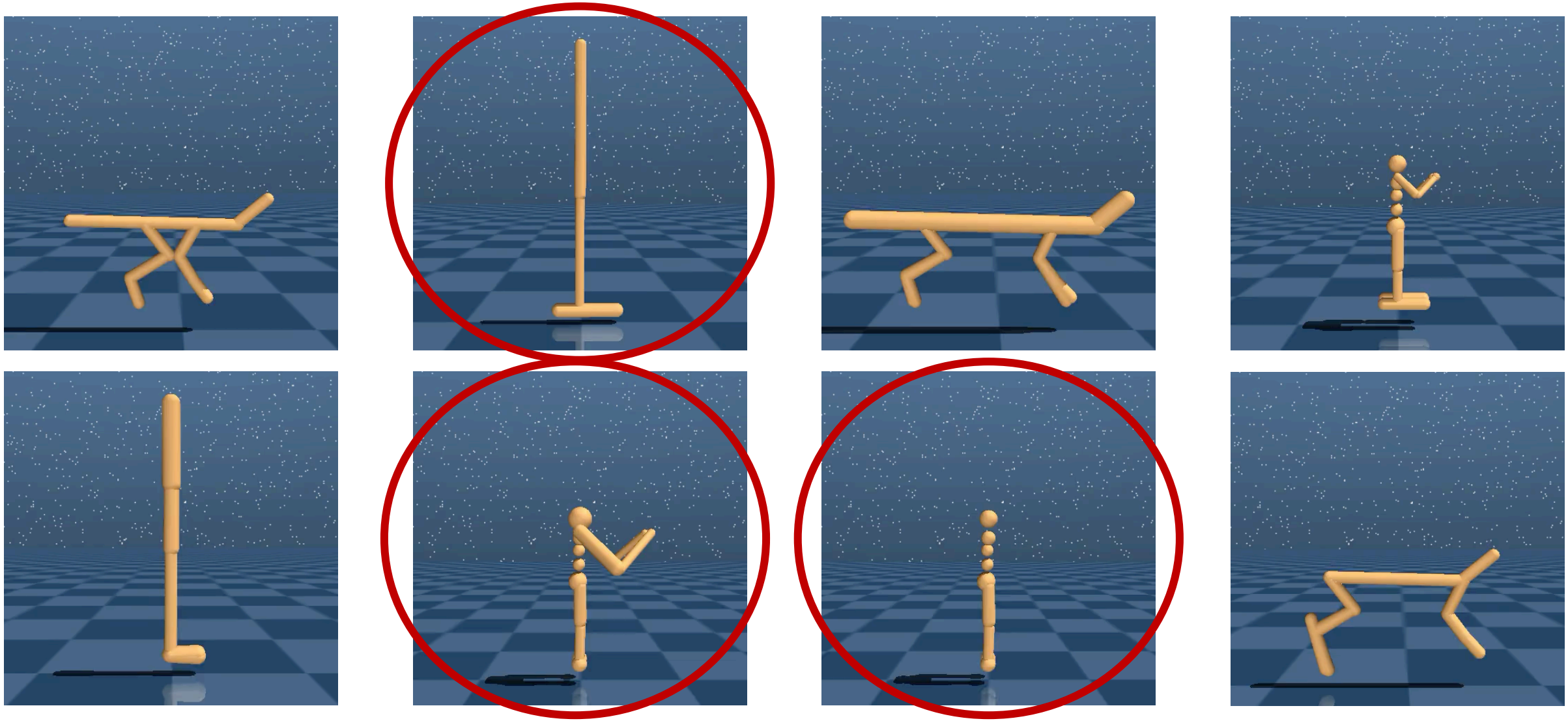
Zero-Shot Generalization: held-out, similar distribution



Zero-shot testing on **out of distribution**

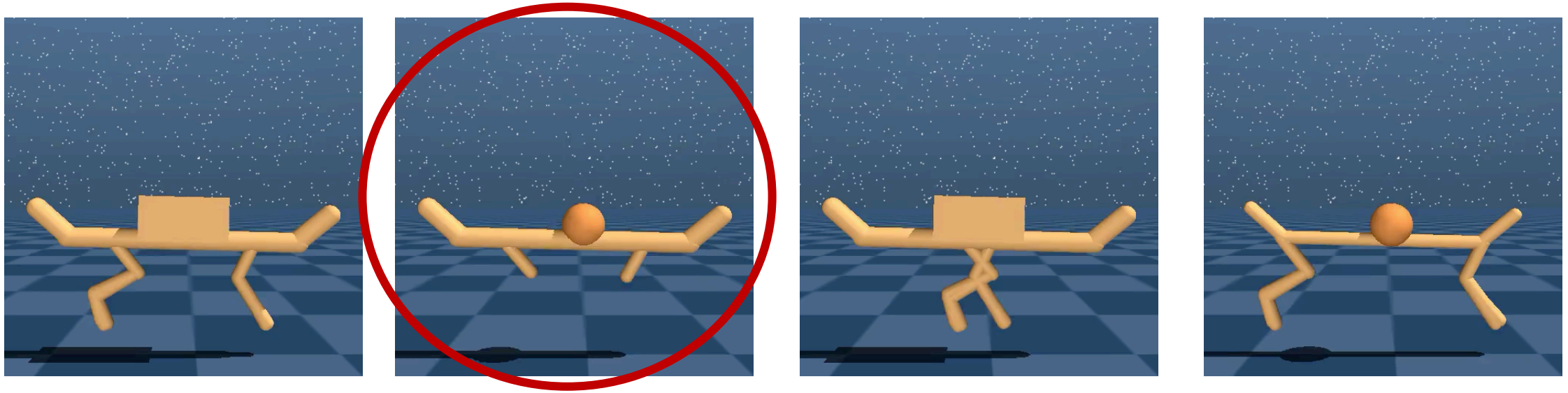
Zero-Shot Generalization – very different agents!

Doesn't excel, but still tries to maintain balance and not fall!



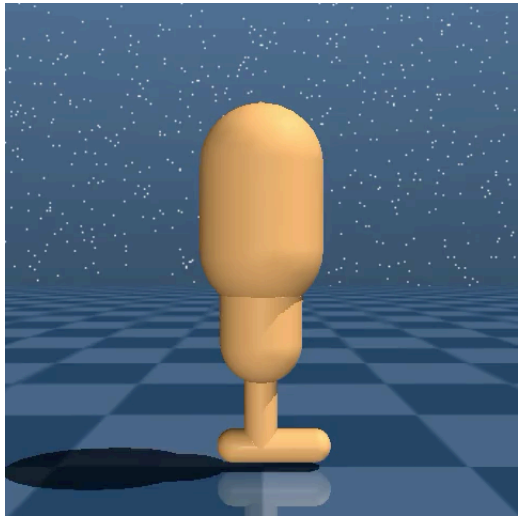
Zero-Shot Generalization – add objects on new agents

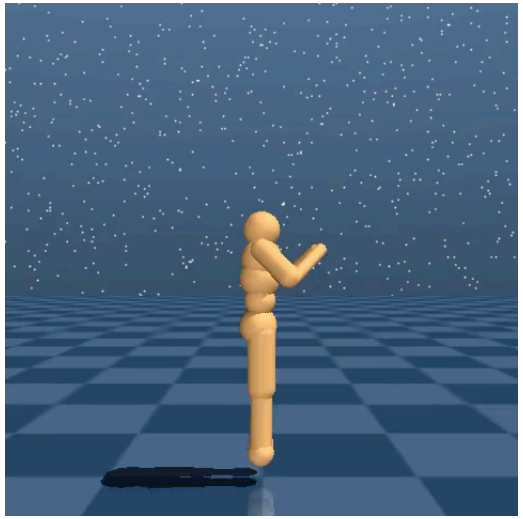
No objects at training time!

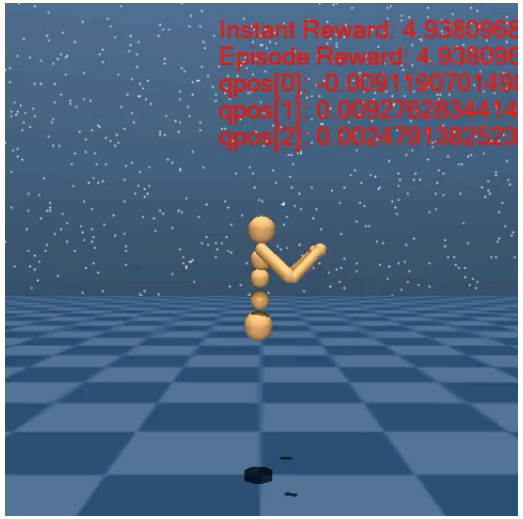


Doesn't excel in moving forward, but still tries to maintain balance and posture

Zero-Shot Generalization: Failures…



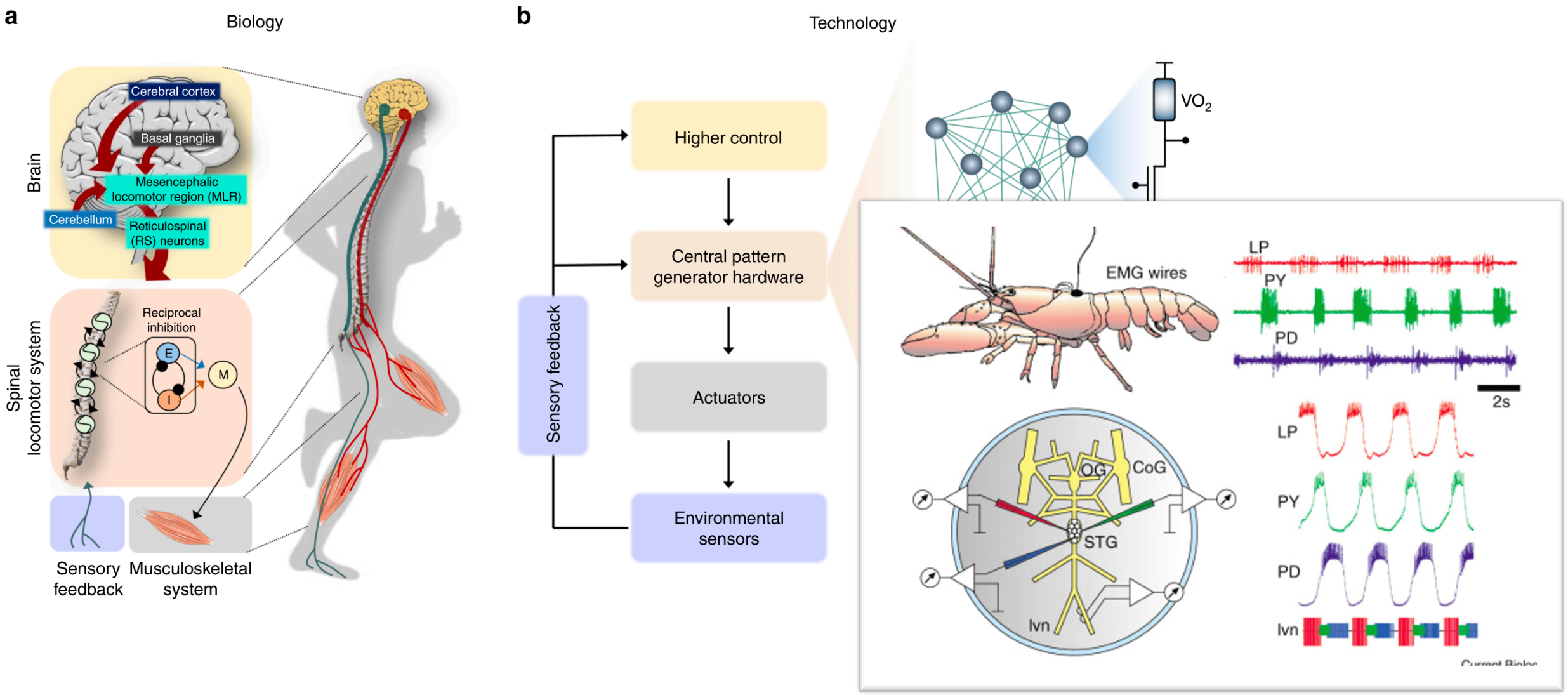




Hard because muscle strength is not enough to carry the weight of these limbs.

Does message coordination really emerge?

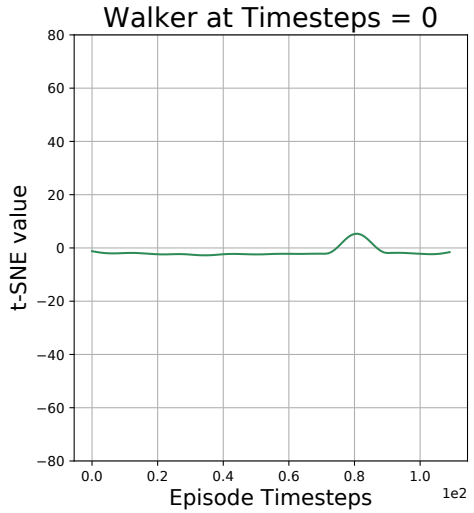
Biological Central Pattern Generators

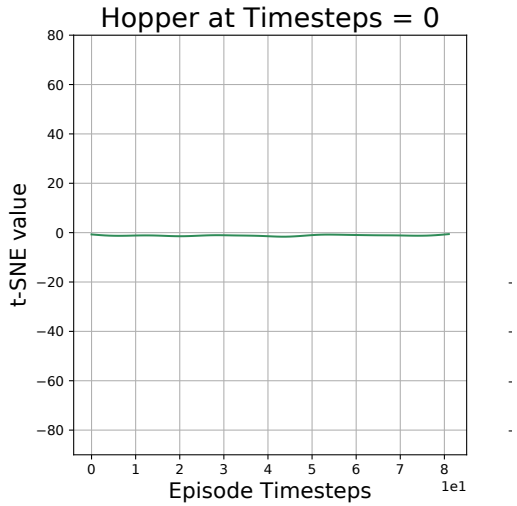


Does message coordination really emerge?

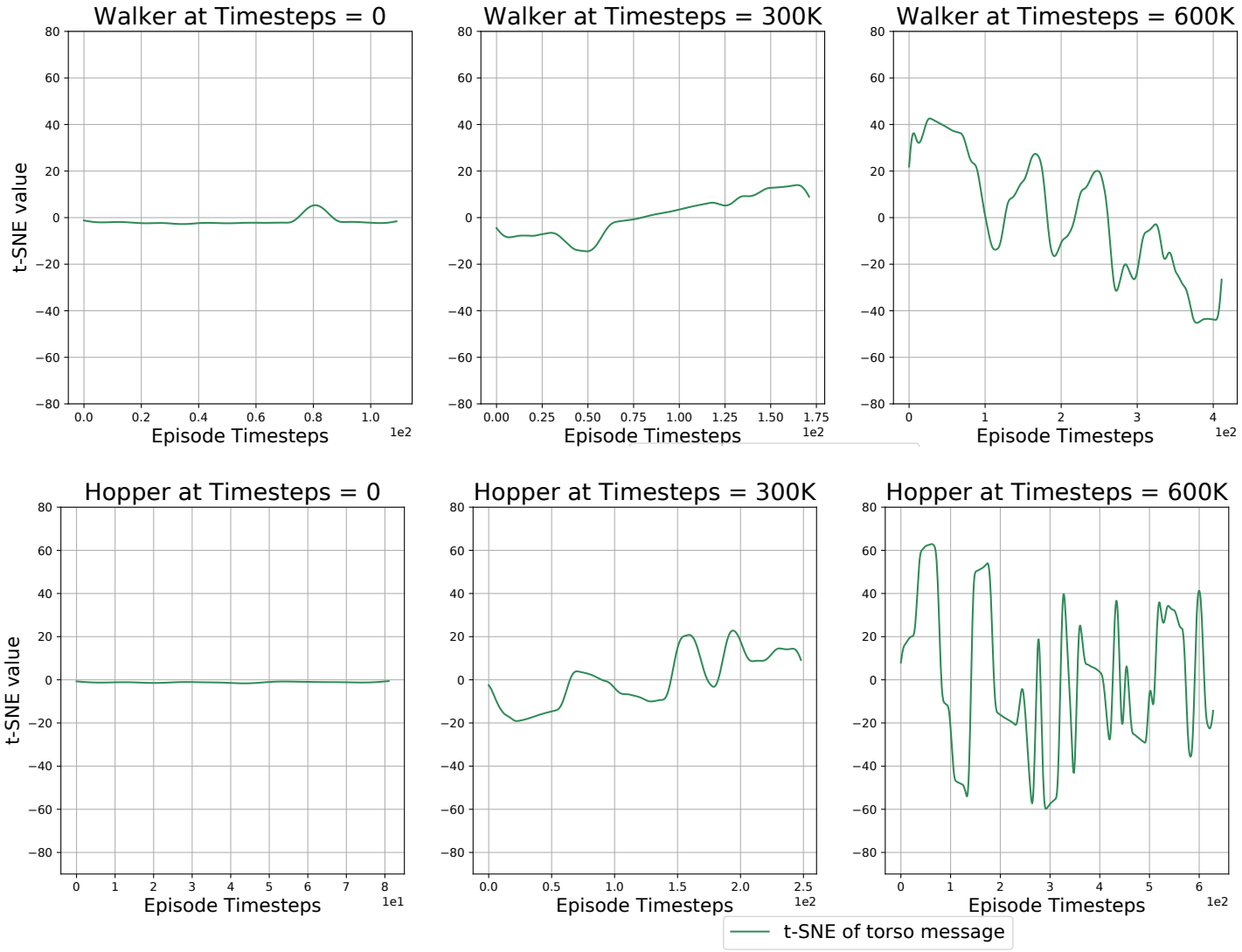
Visualization of out-going message from root node

Does message coordination really emerge?

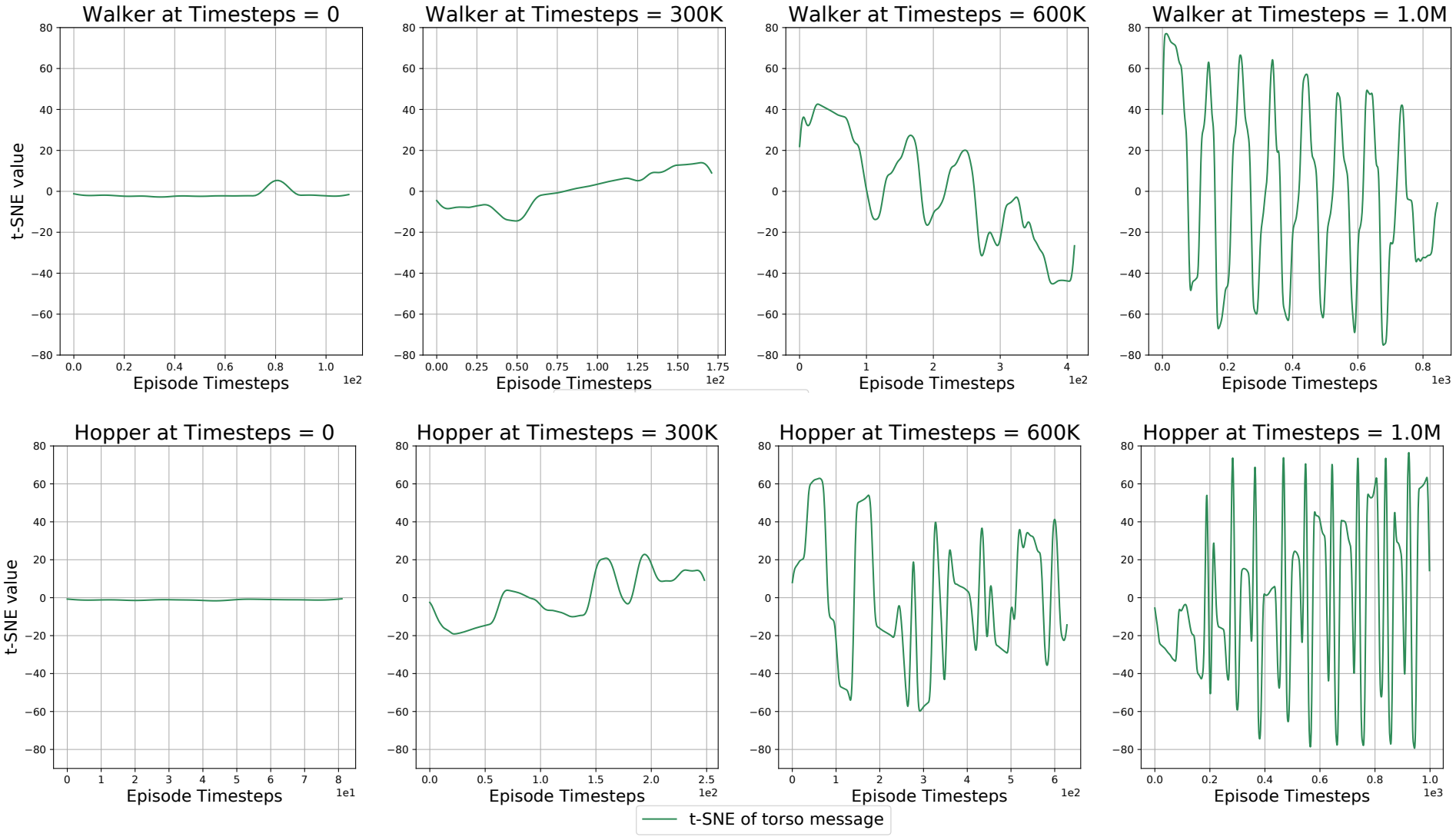




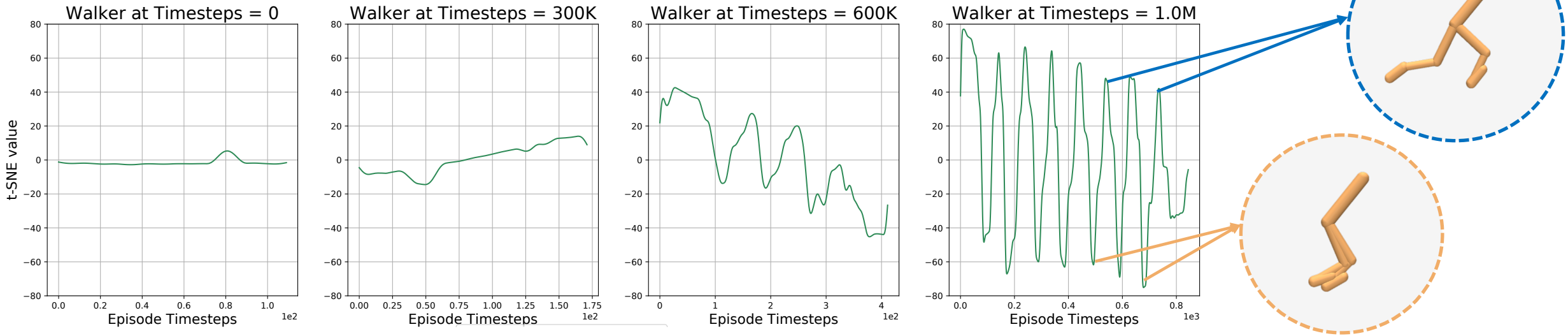
Does message coordination really emerge?

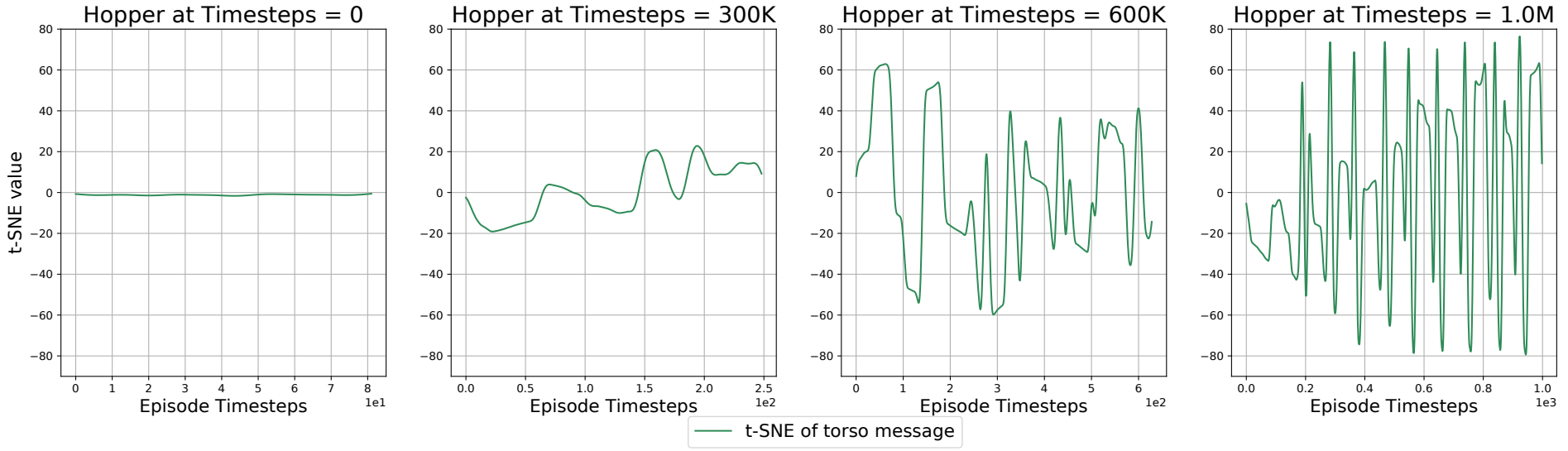


Does message coordination really emerge?

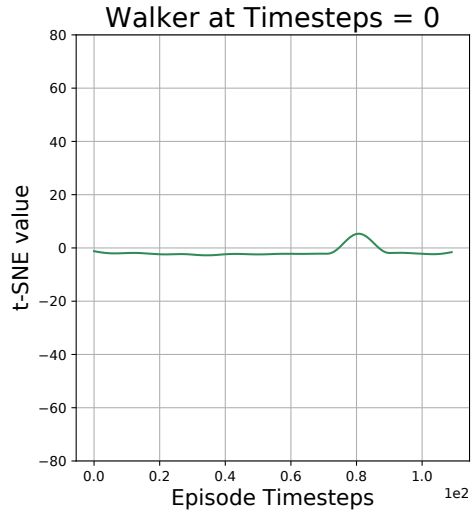


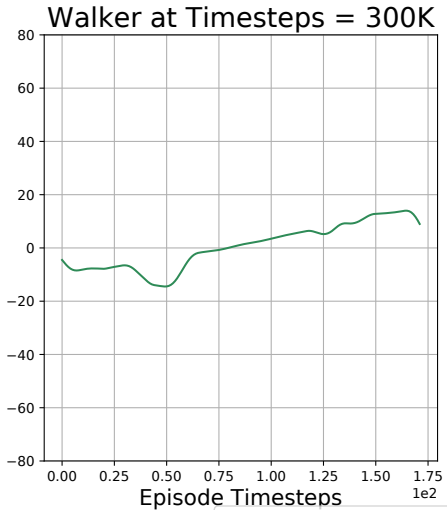
Does message coordination really emerge?

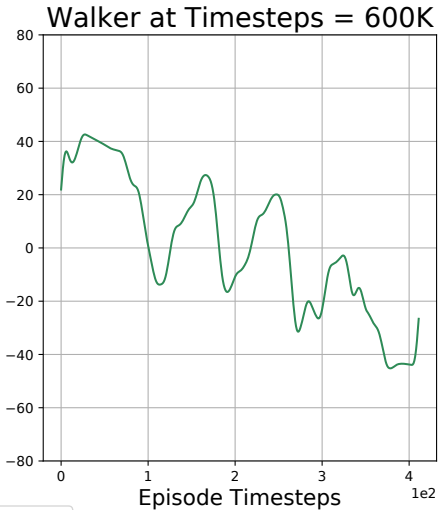


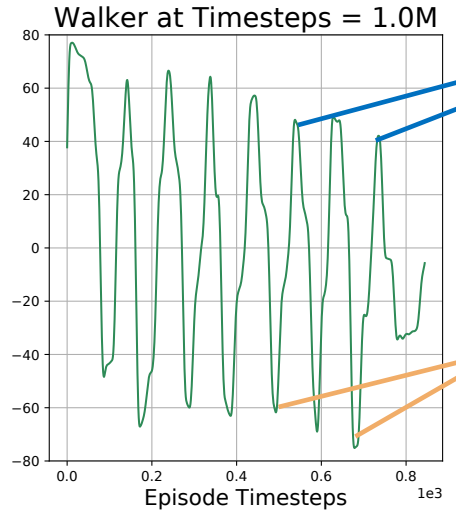


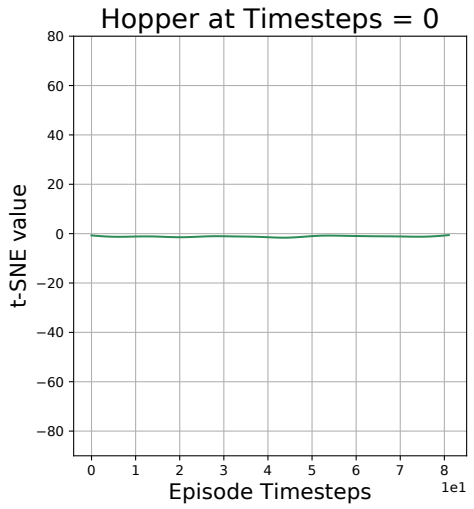
Does message coordination really emerge?

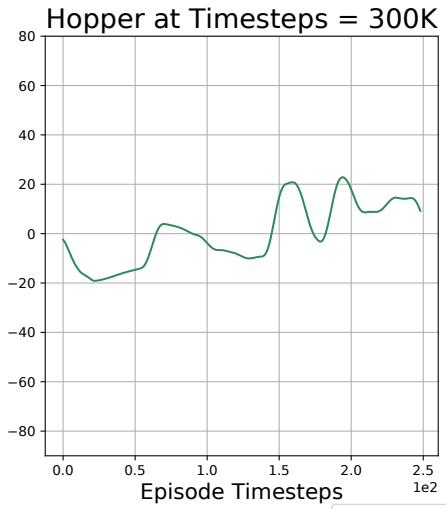


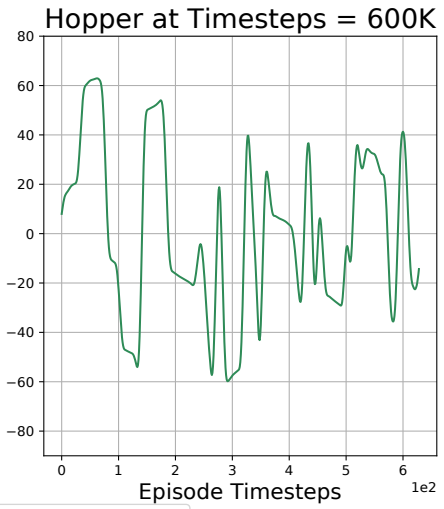


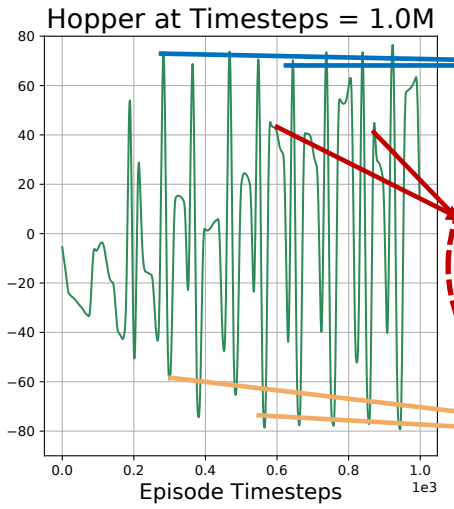


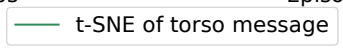


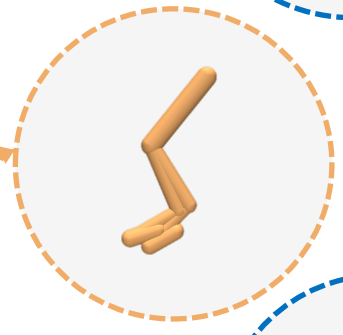


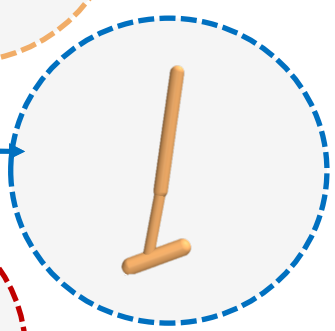


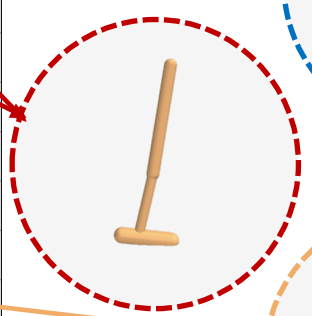


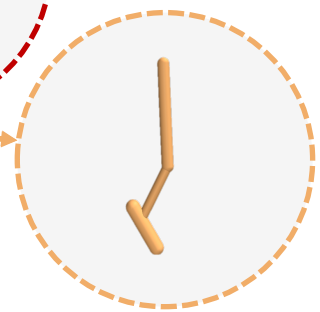










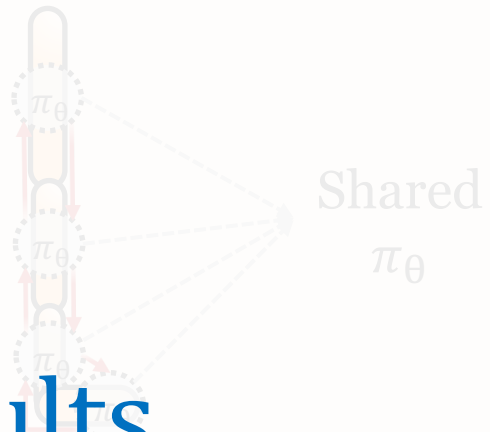


Summary: Shared Modular Policies

1. Training Single Policy on Diverse Robots

- Have **modularity at the level of actuators** instead

- Emergent centralized control by **both-way** message passing



Super-early but promising results…

2. Toward General Pretrained Controllers

- State-of-the-art training information **may not** transfer **robots by a single policy**

- **Zero-shot** generalization to unseen robot shapes

…still a long way to go!

Thank you!