# ERGODIC CONTROL IN ROBOT LEARNING

## OR, WHERE DO TARGET DISTRIBUTIONS COME FROM?

**ICRA 2024**

Todd Murphey
Mechanical Engineering
Physical Therapy and Human Movement Sciences
Northwestern University
@todd_murphey https://murpheylab.github.io/

Northwestern | CENTER FOR ROBOTICS AND BIOSYSTEMS

# Why Does Learning Need Ergodic Control?

**What do robots need to learn?**

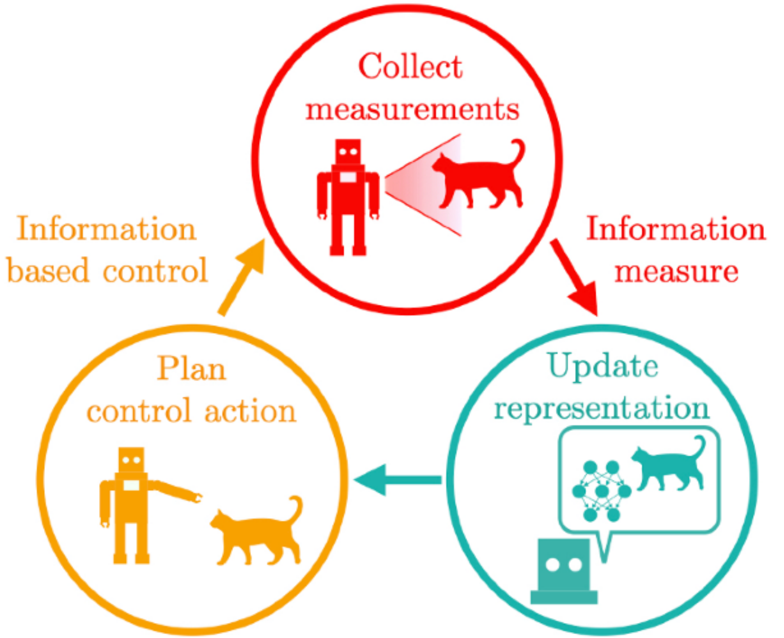
- environment (e.g., active SLAM)
- dynamical models (e.g., robot bodies moving)
- interactions (e.g., social navigation)
- perception models (e.g., ANN models of novel objects with novel sensors)
- task learning (e.g., embodied single-shot RL)

**What advantages do robots have?**

- Robots *do not need to learn passively*

**Why does ergodicity matter?**

- ergodic control automates data collection




Robotic Active Learning

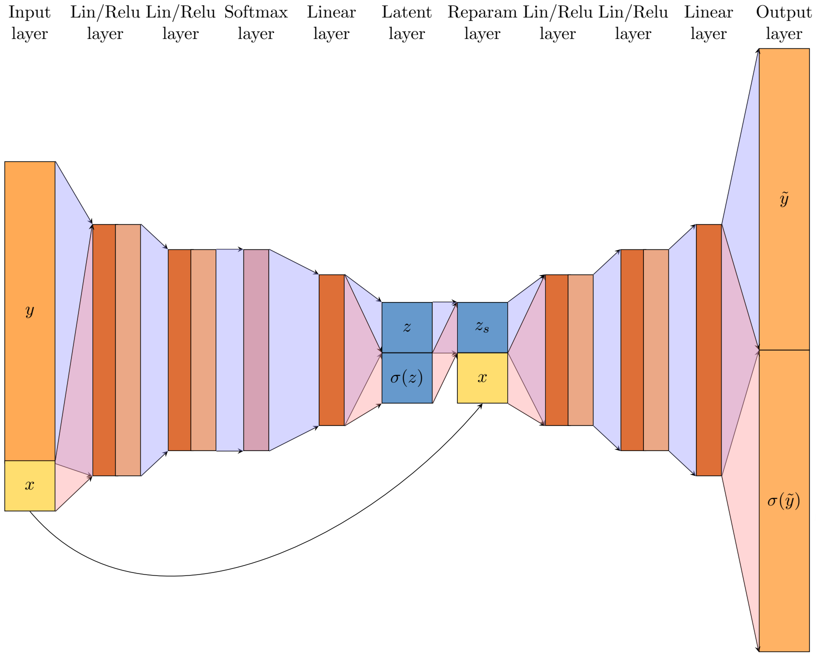


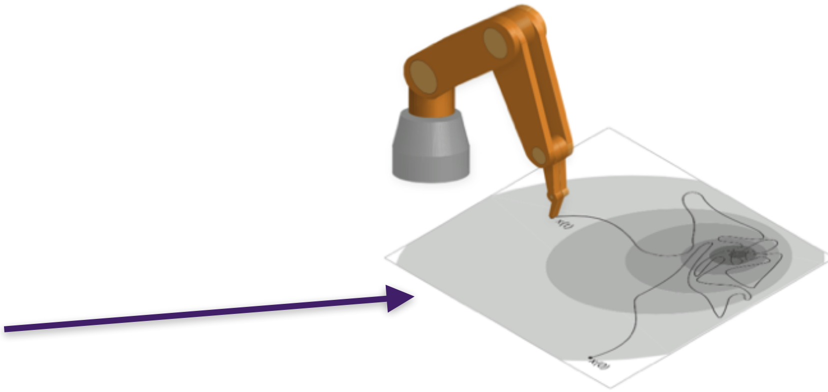
A. Taylor, T. Berrueta, and T. D. Murphey, "Active learning in robotics: A review of control principles," *Mechatronics*, vol. 77, pg. 102756, 2021.

# Robots Do Not Always Have Datasets

1. Non-traditional sensors: MRI / AFM / electrosense / tactile
2. Human-Machine Interfaces, decision support
3. Manufacturing with micro-/nano-scale non-heterogenous physics
4. Data for physical systems is almost always both ***scarce*** and ***sparse***
5. Many physical interactions cannot be simulated
6. Austere environments will not be simulated, even with access to cloud
7. Compute needs to be synchronous for active data collect

# Ergodic Control for Perception Models



Conditional Variational Autoencoders (CVAEs) are unsupervised learning methods that generate a latent representation conditioned on a set of parameters; use whatever learning model you like best, but one that predicts entropy as a function of state.



A. Prabhakar and T. D. Murphey, "Mechanical intelligence for learning embodied sensor-object relationships," *Nature Communications*, 2022.

# Ergodic Control for Perception Models

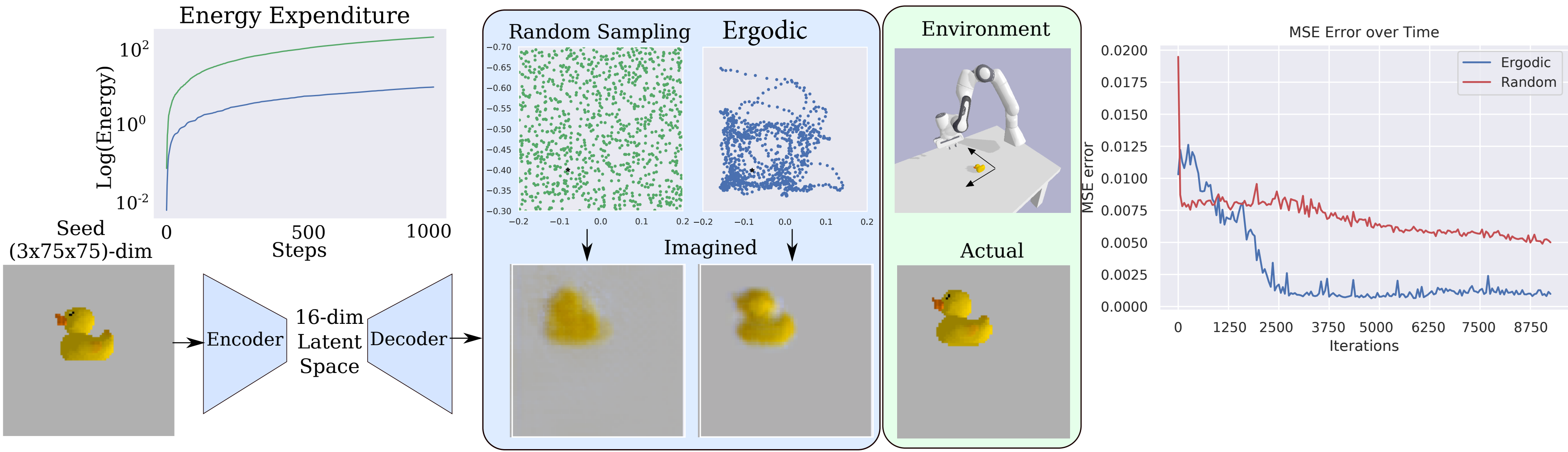
This simulated robot has no model of its sensor or the object—it creates a perception model through automated data collection.


The robot's motion is ergodic with respect to the spatially distributed entropy



A. Prabhakar and T. D. Murphey, "Mechanical intelligence for learning embodied sensor-object relationships," *Nature Communications*, 2022.

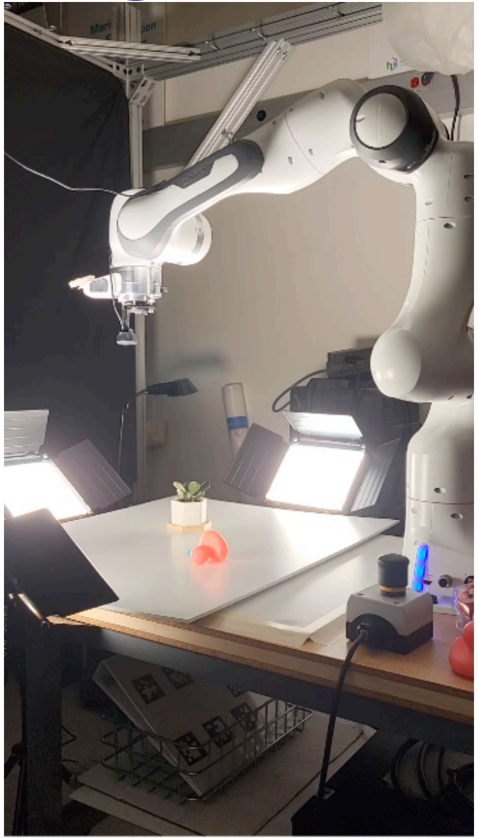
# Ergodic Control for Perception Models

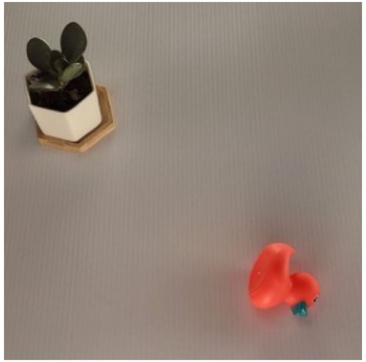


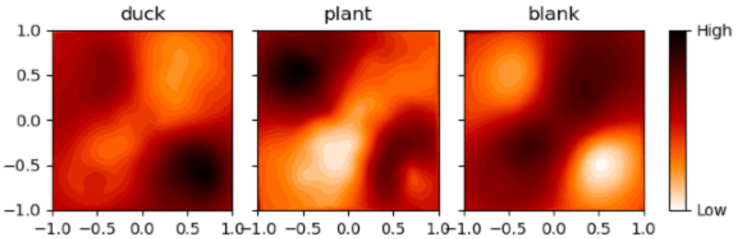



A. Prabhakar and T. D. Murphey, "Mechanical intelligence for learning embodied sensor-object relationships," *Nature Communications*, 2022.

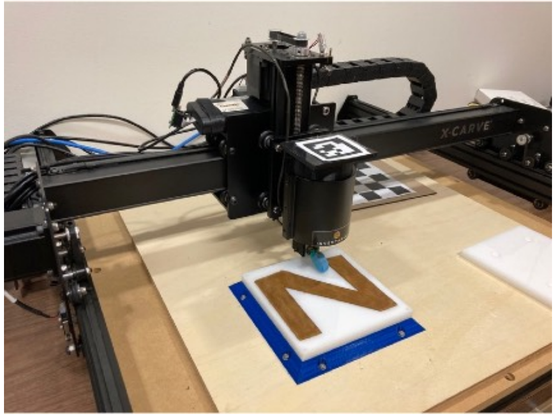
# Ergodic Control for Perception Models








*Real-time* / *Online* deep learning for perception involves compute and automated data collection



New sensors, such as touch sensing (above) or almost anything that isn't vision or audio, will need novel datasets. These datasets need to come from somewhere.

**The learning pipeline for all sensory modalities can be the *same*, using inference and control principles. This can be used to characterize novel materials using new sensor technologies.**



**See J. Ketchum's talk on Wednesday WeBT16-AX.8**

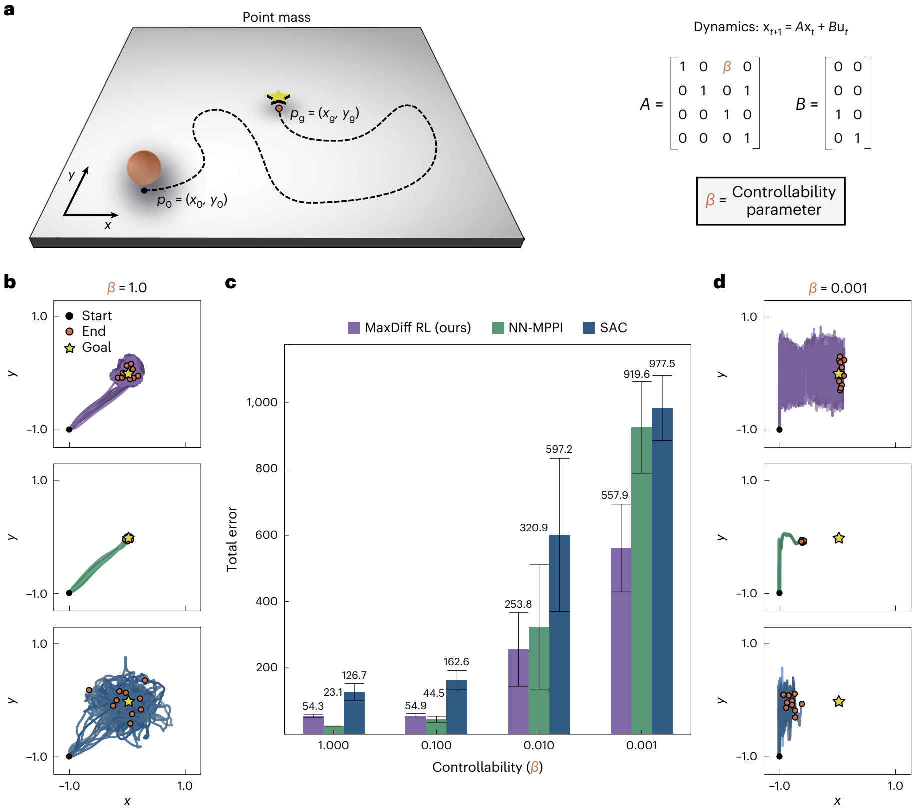
- Upshot: ergodic control provides a way to implement 'proportional response' to the predicted error of a neural network
- The robot can go and collect data while ensuring coverage and (spatial) independence of the data
- This connects to needs in reinforcement learning….

# Maximum Diffusion Reinforcement Learning

- RL typically variation in the *inputs* of a system to generate variation in the outputs (e.g., stochastic policies and MaxEnt strategies)
- The problem with this approach is that variation in the inputs may not create very much variation in the state
- For a continuous-time process, i.i.d. sampling is hard, we show there is an ergodic strategy, and the optimal strategy is diffusive

$$P_{\max}[x(t)] = \frac{1}{Z}\exp\left[-\frac{1}{2}\int_{t_0}^{t}\dot{x}(\tau)^T C^{-1}[x(\tau)]\dot{x}(\tau)\mathrm{d}\tau\right]$$

- Moreover, controllability plays a critical and explicit role

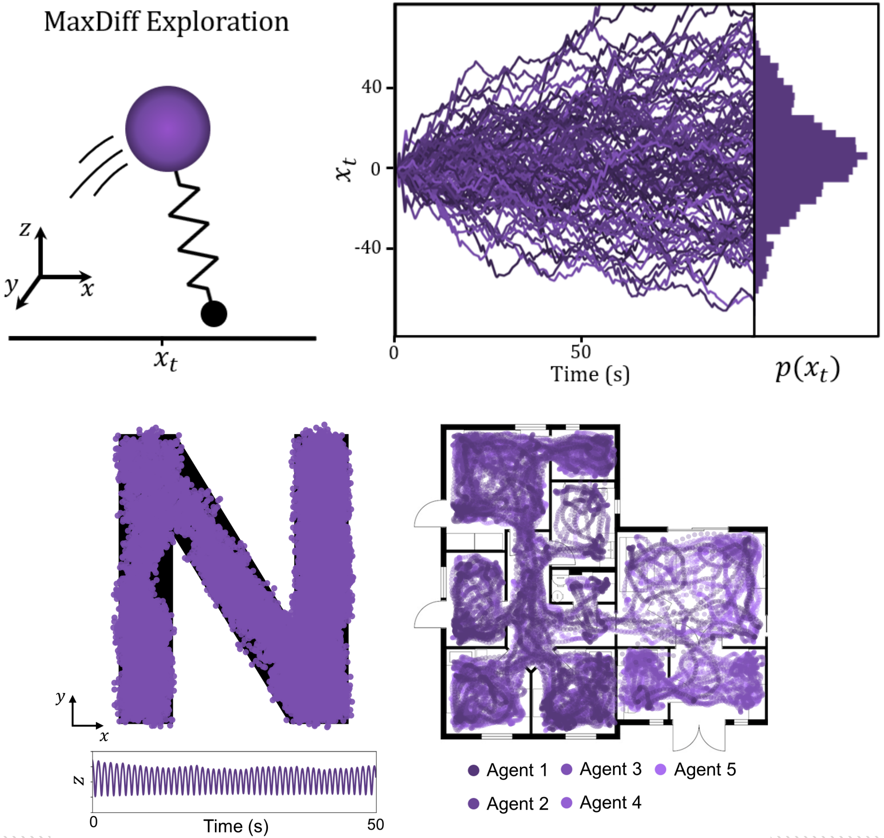





T. Berrueta, A. Pinosky, T. Murphey, "Maximum diffusion reinforcement learning," *Nature Machine Intelligence*, 2024.

# Maximum Diffusion Reinforcement Learning

- A strategy that maximizes the i.i.d. property does create an ergodic trajectory.
- This is model-based in the it uses an ANN model of the dynamics to synthesize the optimal ergodic path
- Moreover, the exploration has to take into account another ANN's estimate of the reward landscape

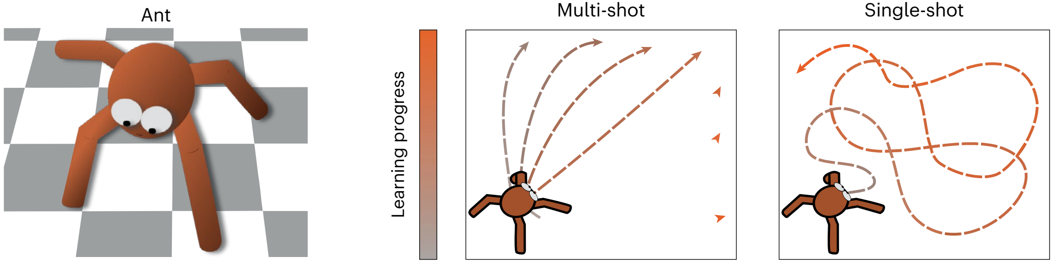


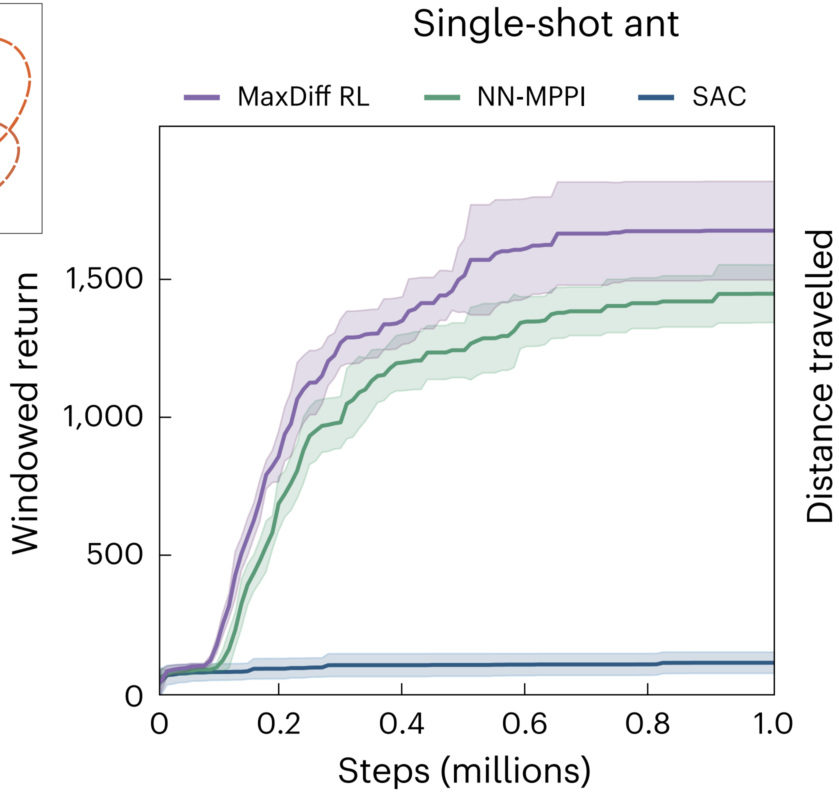




T. Berrueta, A. Pinosky, T. Murphey, "Maximum diffusion reinforcement learning," *Nature Machine Intelligence*, 2024.

# Maximum Diffusion Reinforcement Learning



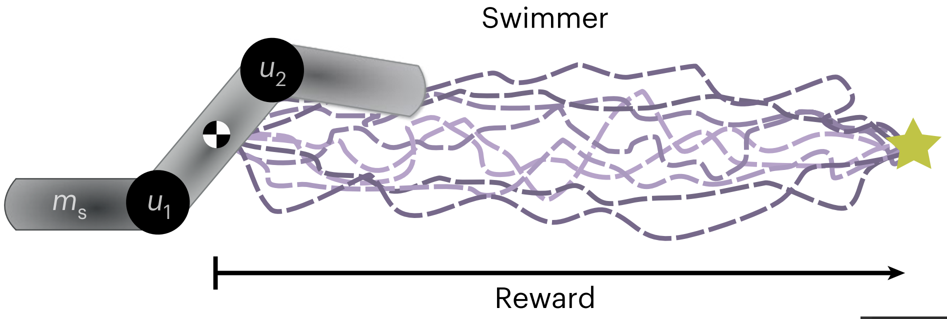




- Ergodic / i.i.d. guarantees enable stronger guarantees on single-shot performance in the form of seed invariance
- Benchmarking single-shot performance specifically highlights the quality of ergodicity-supported exploration



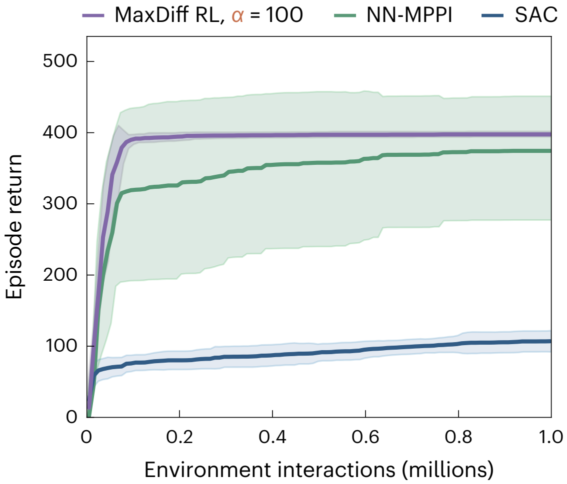
T. Berrueta, A. Pinosky, T. Murphey, "Maximum diffusion reinforcement learning," *Nature Machine Intelligence*, 2024.

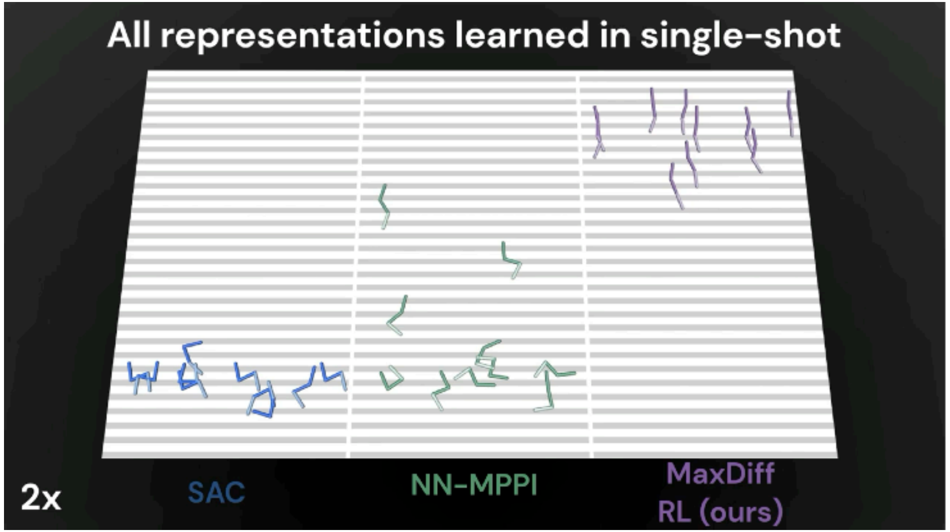
# Maximum Diffusion Reinforcement Learning




MaxDiff RL

$$\underset{\pi}{\operatorname{argmax}}\ E_{P_\pi}\Big[\sum_t \gamma^t \Big(\underbrace{r(x_t, u_t)}_{\text{Task exploitation}} + \underbrace{\alpha S[p(x_{t+1}|x_t, u_t)\pi(u_t|x_t)]}_{\text{Diffusive exploration}}\Big)\Big]$$



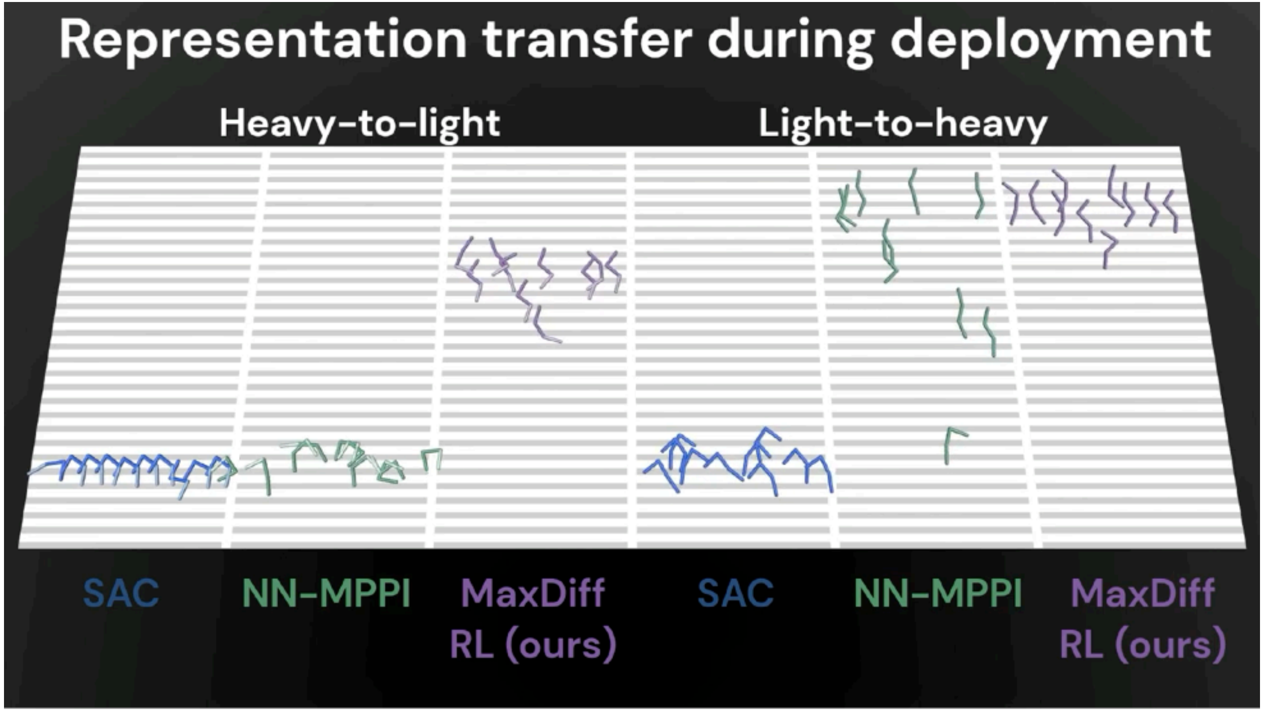




T. Berrueta, A. Pinosky, T. Murphey, “Maximum diffusion reinforcement learning,” *Nature Machine Intelligence*, 2024.

# Maximum Diffusion Reinforcement Learning

- Robustness (or "zero shot" learning) is also a benefit
- This is largely an empirical statement, not a mathematical one
- However, it should not be surprising that carefully data will protect the policy against sensitivity to small variations






T. Berrueta, A. Pinosky, T. Murphey, "Maximum diffusion reinforcement learning," *Nature Machine Intelligence*, 2024.

# Conclusions

- Machine learning is data sensitive
- Most ML techniques address this with more data rather than better data
- Robots can curate their own data during learning
- Ergodic approaches lead to superior learning outcomes
  - with mathematical single-shot properties
  - and empirical robustness properties

# Thanks!

## Amazing Students

- Lauren Miller
- Anastasia Mavrommati
- **Ian Abraham**
- **Ahalya Prabhakar**
- Katie Fitzsimons
- Jake Ketchum
- Giorgos Mamakoukas
- Ana Pervan
- **Thomas Berrueta**
- Millicent Schlafly
- Annalisa Taylor
- **Allison Pinosky**

## Amazing Collaborators

- Brenna Argall
- Alex Broad