

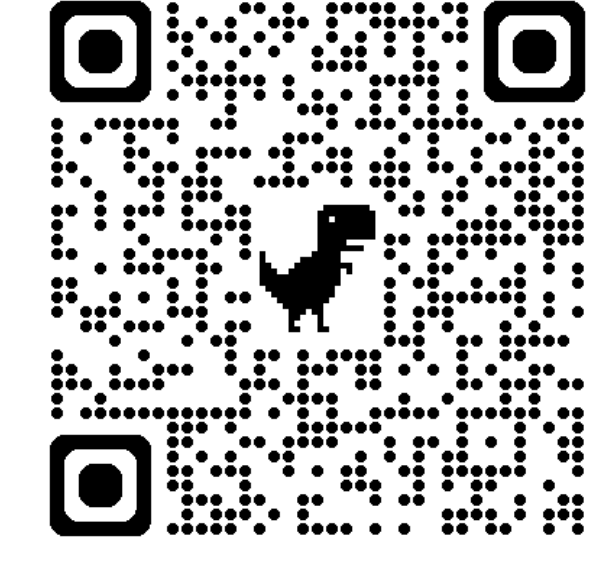
Periodic Skill Discovery

Jonghae Park¹ Daesol Cho² Jusuk Lee¹ Dongseok Shim¹ Inkyu Jang¹ H. Jin Kim¹

¹Seoul National University ²Georgia Institute of Technology



SEOUL
NATIONAL
UNIVERSITY

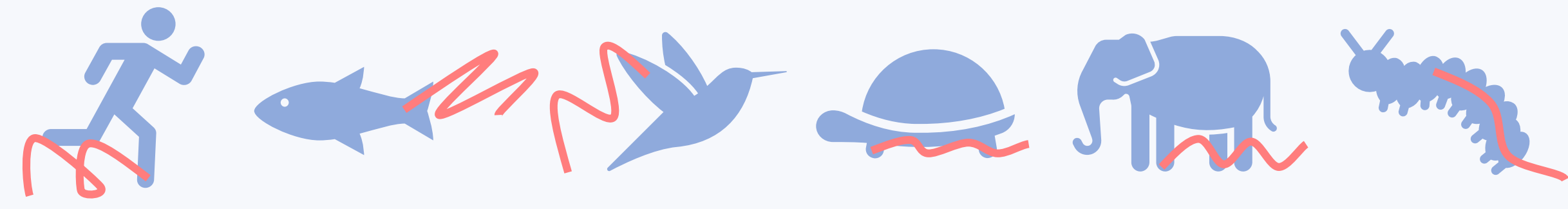


Project Page
(Demos & Code)

Why Periodic Skills Matter?

A **fundamental observation in nature** is that nearly all forms of locomotion are inherently periodic.

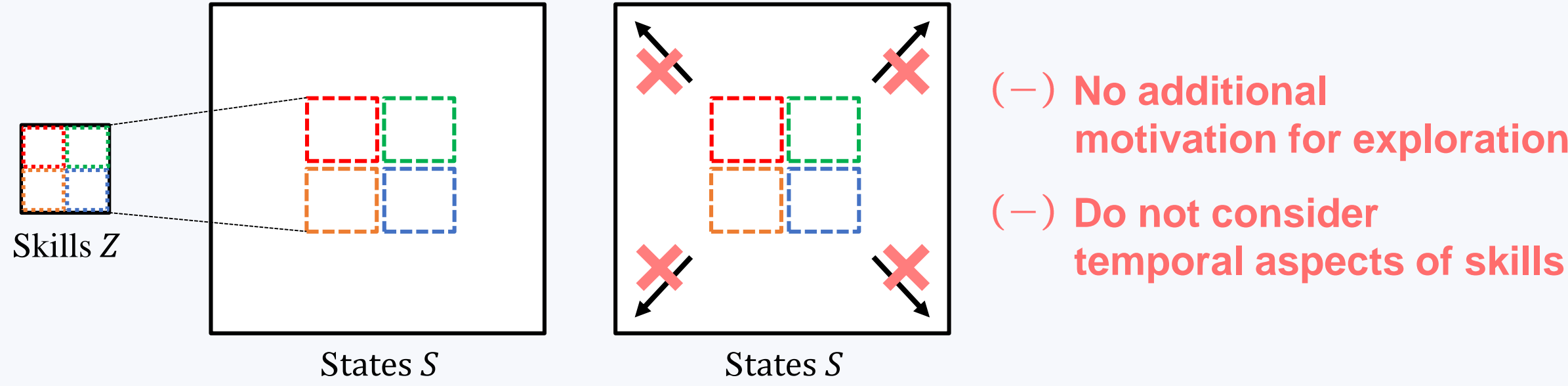
However, existing unsupervised skill discovery methods have rarely addressed **the role of periodicity**.



Prior works often fail to learn *multi-timescale* behaviors

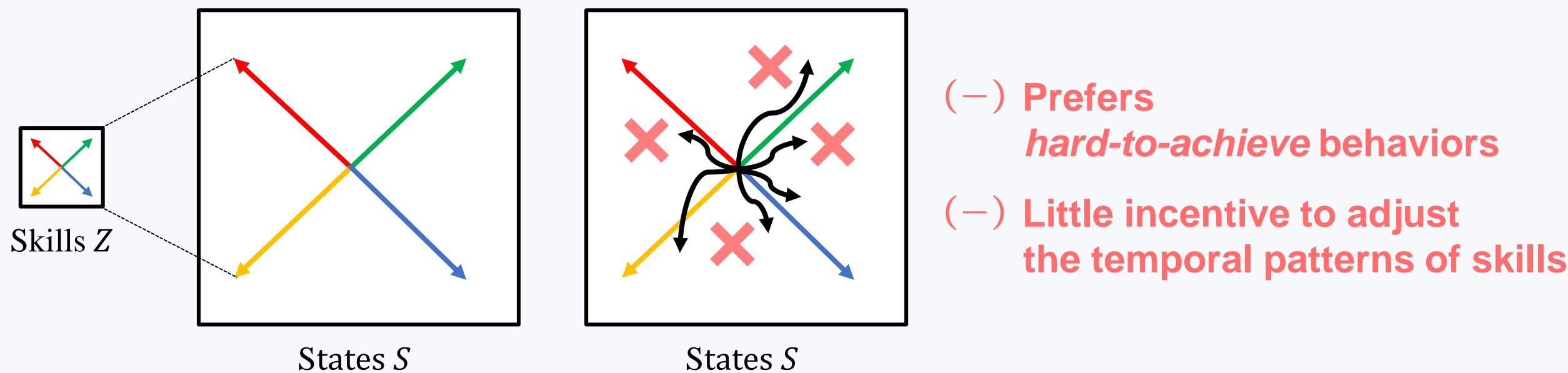
(1) Mutual Information (MI) - based skill discovery

$$I(S; Z) = -H(Z|S) + H(Z) = \mathbb{E}_{z,\tau}[\log p(z|s)] - \mathbb{E}_z[\log p(z)] \\ \geq \mathbb{E}_{z,\tau}[\log q_\theta(z|s)] + (\text{constant}) \simeq \mathbb{E}_{z,\tau} \left[-\frac{1}{2\sigma^2} \|z - \mu_\theta(s)\|_2^2 \right] + (\text{constant})$$



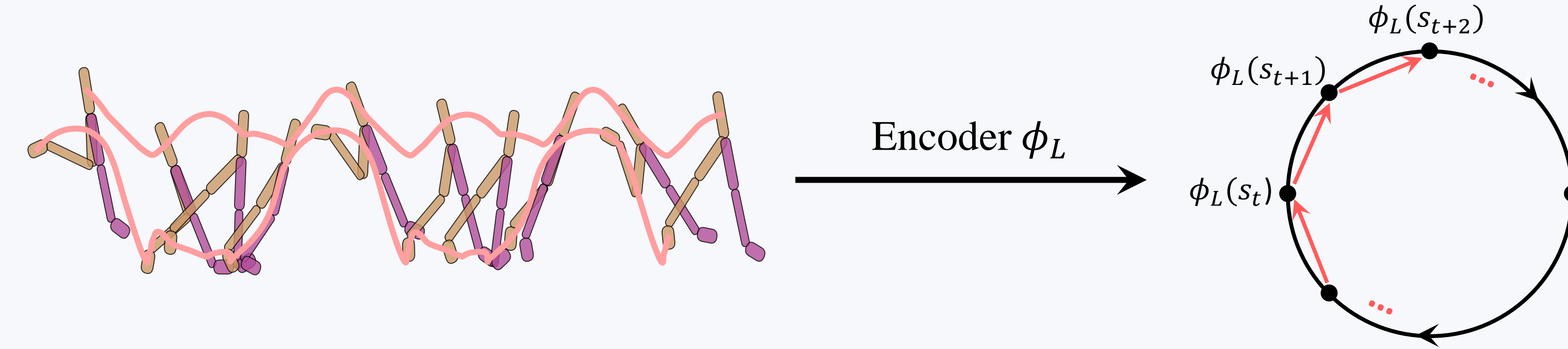
(2) Distance - maximizing skill discovery

$$\mathcal{J}_{\text{DSD}} := \mathbb{E}_{(z,\tau) \sim \mathcal{D}} \left[(\phi(s_{t+1}) - \phi(s_t))^\top z \right] \quad \text{s.t.} \quad \|\phi(x) - \phi(y)\| \leq d(x, y) \quad \forall x, y \in \mathcal{D}$$



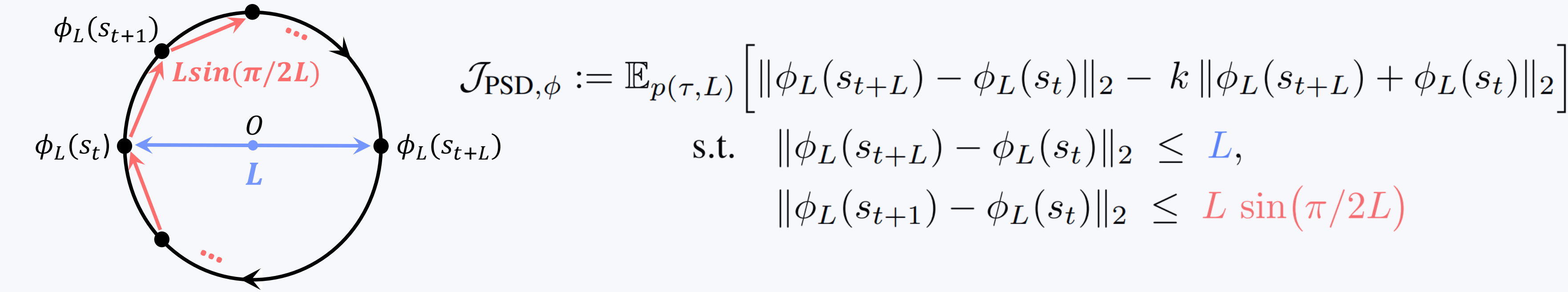
Intuition

- We introduce **Periodic Skill Discovery (PSD)**
- Key intuition: Construct a **circular latent space** for periodic behavior.



Representations for Periodicity

- Construct a **regular 2L-gon** inscribed in a circle of **diameter L**.



Single-step Intrinsic Reward for the policy $\pi(a|s, L)$.

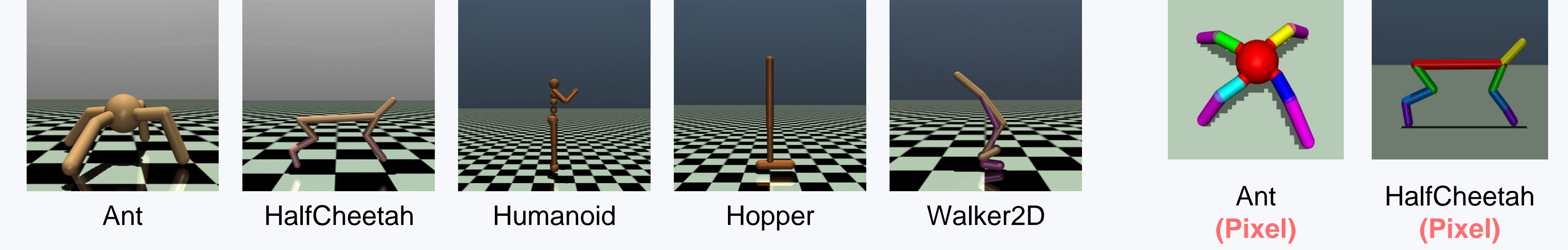
- Define a single-step intrinsic reward encouraging **2L-periodicity**.

$$\Delta := \|\phi_L(s_{t+1}) - \phi_L(s_t)\|_2 - L \sin(\pi/2L) \quad r_{\text{PSD}}(s_t, s_{t+1}, L) := \exp(-\kappa \Delta^2)$$

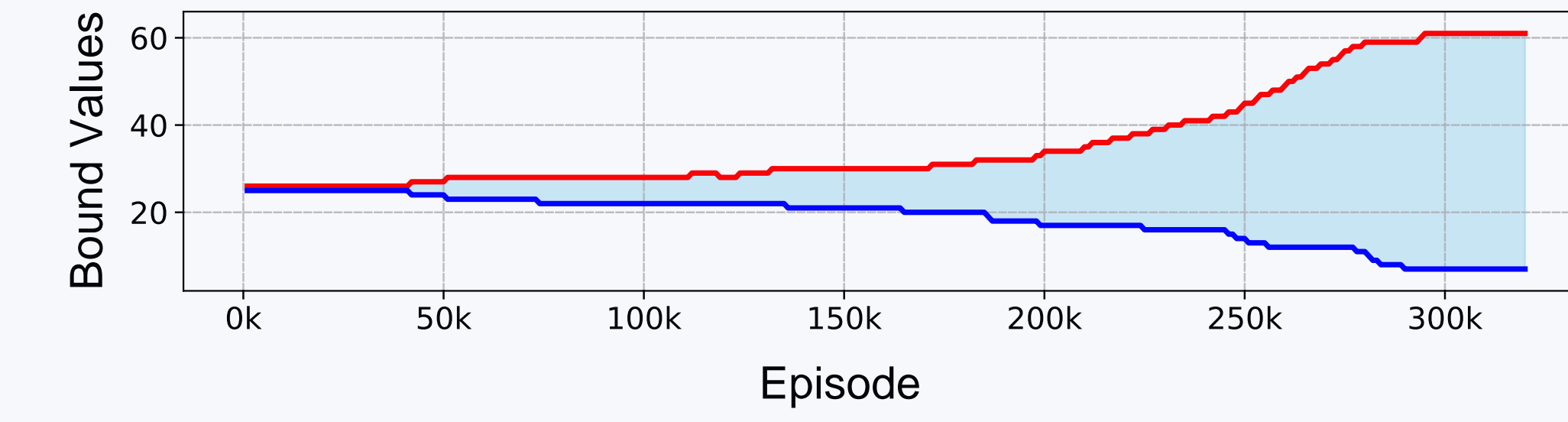
Adaptive Sampling Method

- To enable the agent to discover a **maximally diverse** range of periods without **any prior knowledge** of its inherent period ranges, we introduce an adaptive sampling method that dynamically adjusts the sampling range during training.
- Key idea: Evaluate the policy's **performance on the boundary of the current sampling range**, using $\sum_{t=0}^{T-1} r_{\text{PSD}}(L_{\text{bound}})$ as the evaluation criterion.

Benchmark Environments

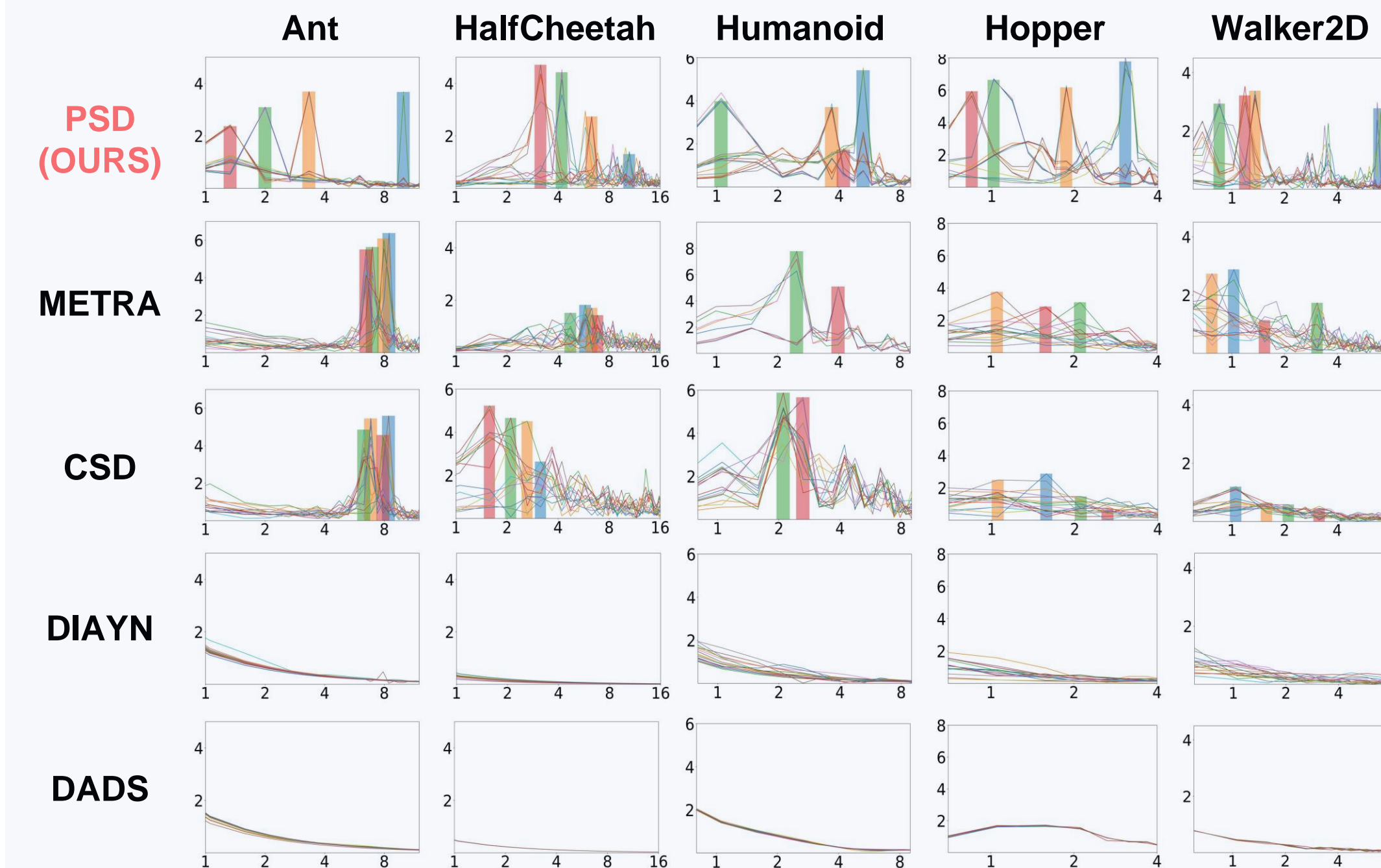


Learned Sampling Bounds Over Time



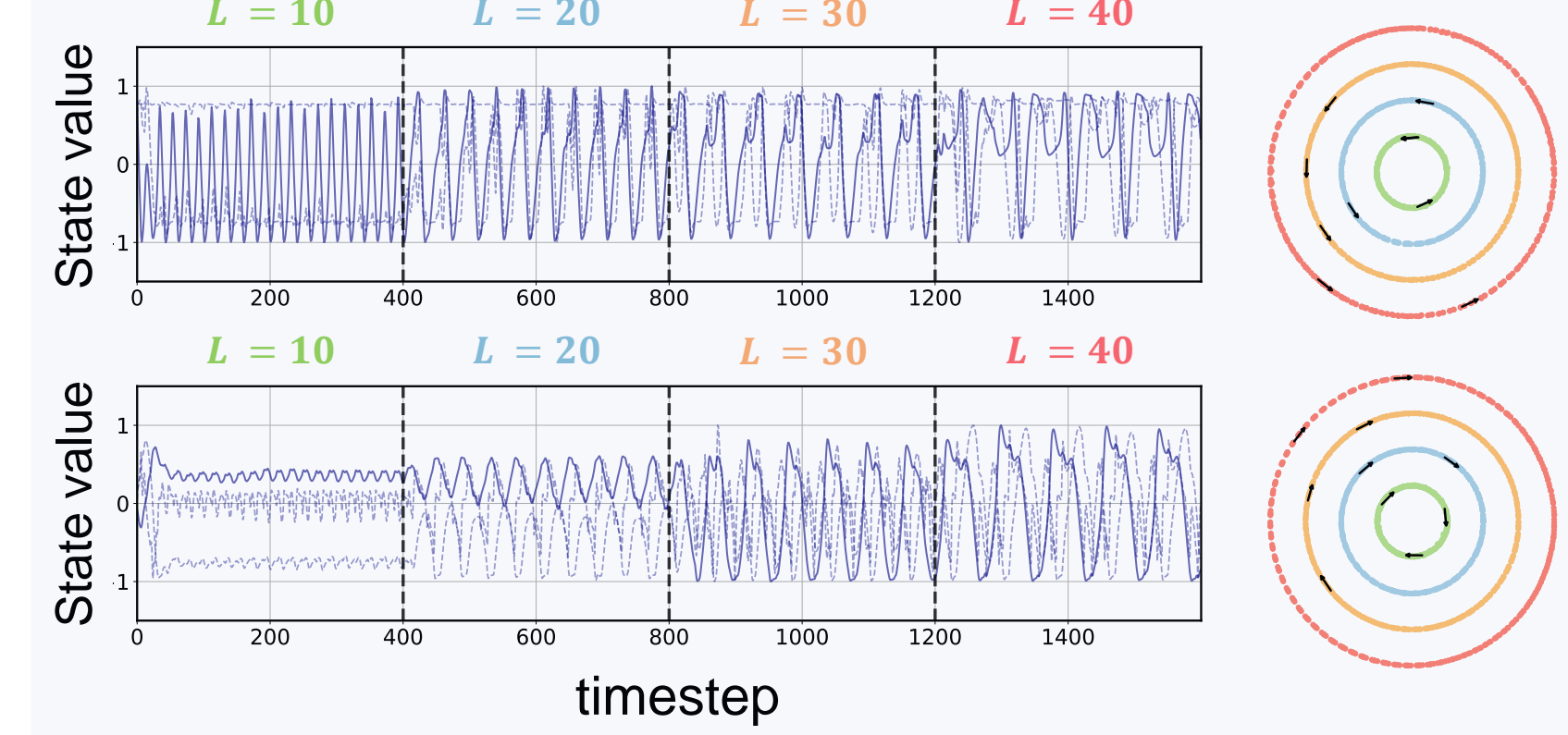
As training progresses, **increasingly challenging periods are proposed** to the agent, enabling the discovery of a wider range of periodic behaviors.

Frequency-Domain Skill Comparison



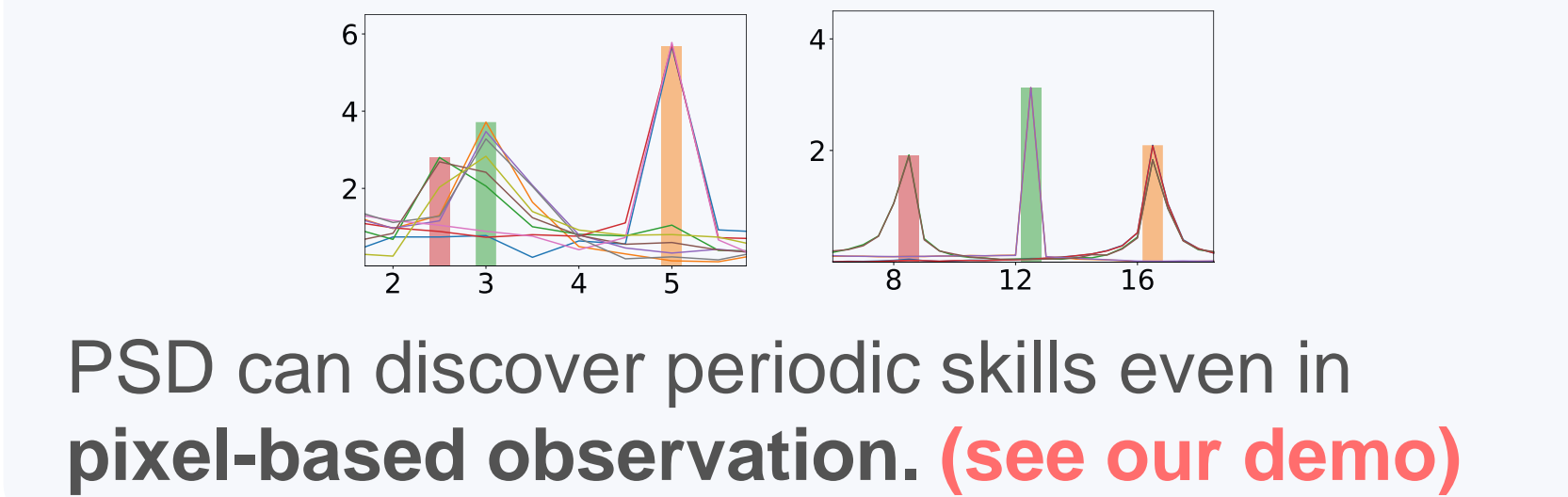
We apply a **Fourier transform** to the skill trajectories (frequency on the x-axis, amplitude on the y-axis). **PSD consistently discovers a wider range of frequencies than the baselines.**

State Trajectories & Latents (PCA)



For varying values of L, PSD successfully **constructs a circular latent space** whose diameter is L, and **learns behaviors with the desired period of 2L**.

PSD with Pixel-based observation



PSD combined with METRA (Park et al., 2023)

