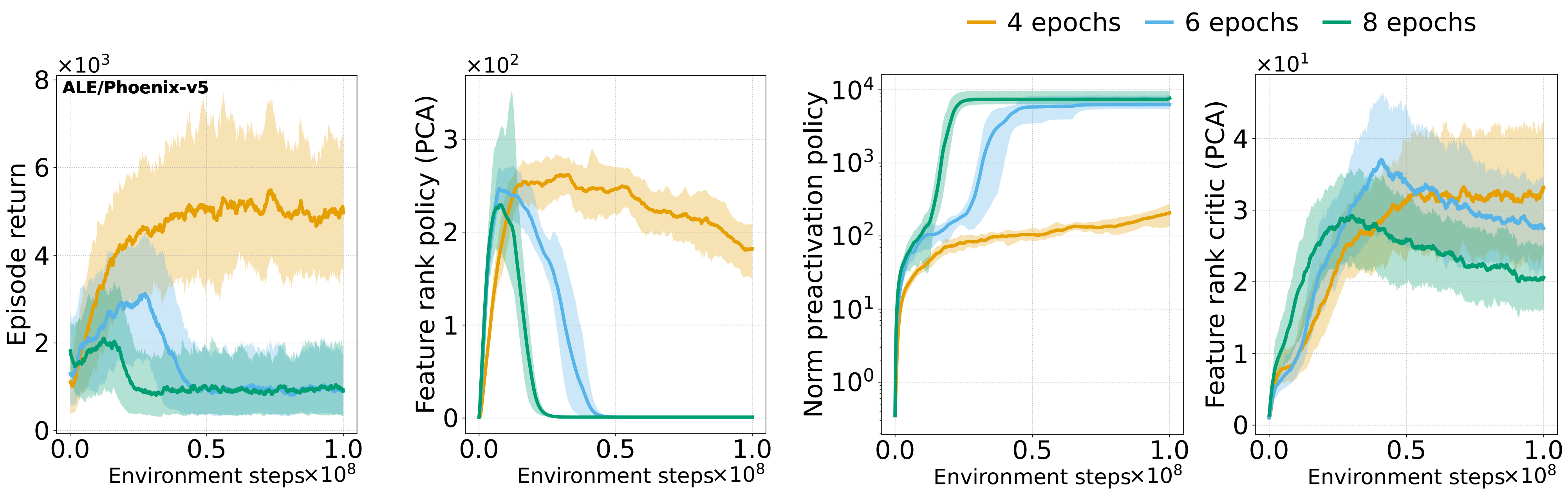
PPO suffers from a **deteriorating representation** that **breaks** its **trust region.**

No Representation, No Trust: Connecting Representation, Collapse, and Trust Issues in PPO

Skander Moalla, Andrea Miele, Razvan Pascanu, Caglar Gulcehre

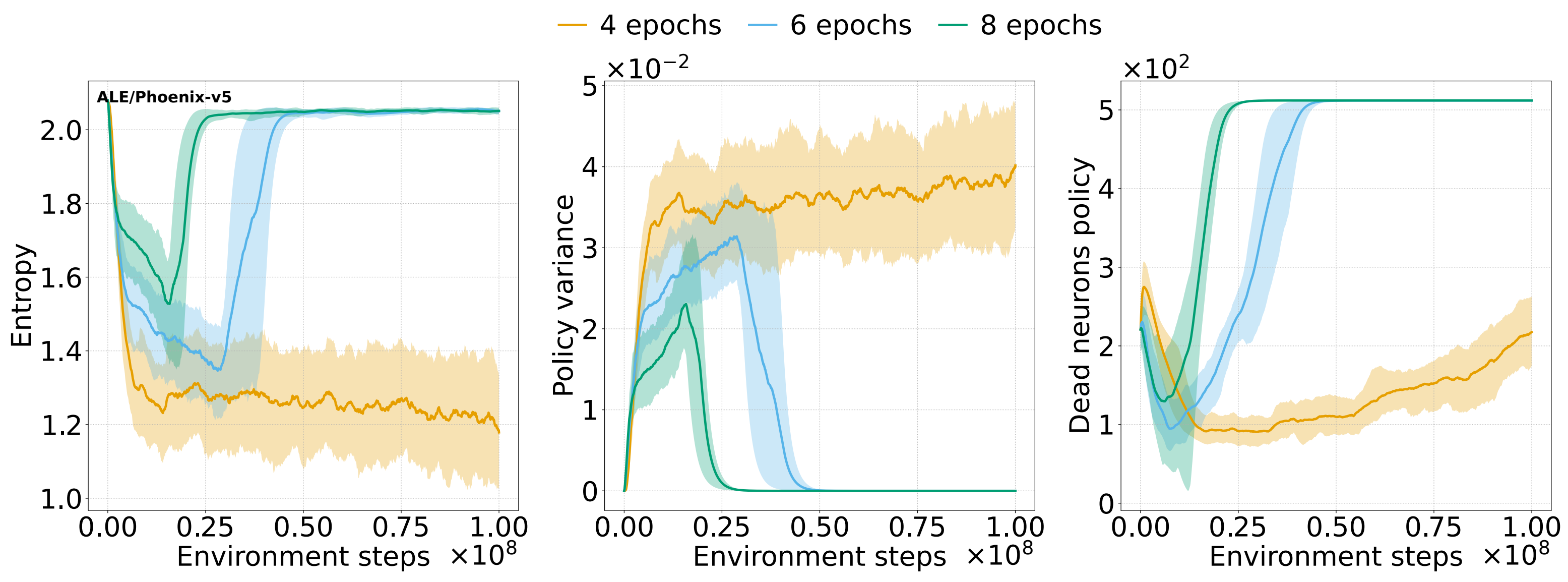
EPFL Google DeepMind

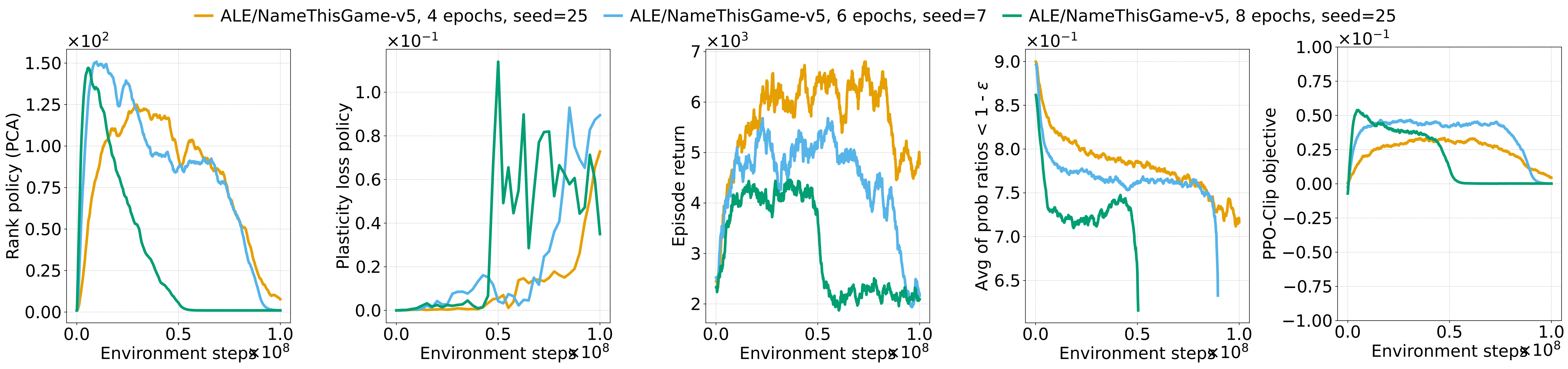
Fully reproducible and replicable



PPO agents on Atari and MuJoCo suffer from **representation collapse causing performance collapse**. The collapse is **faster with stronger non-stationarity,** achieved with more epochs.

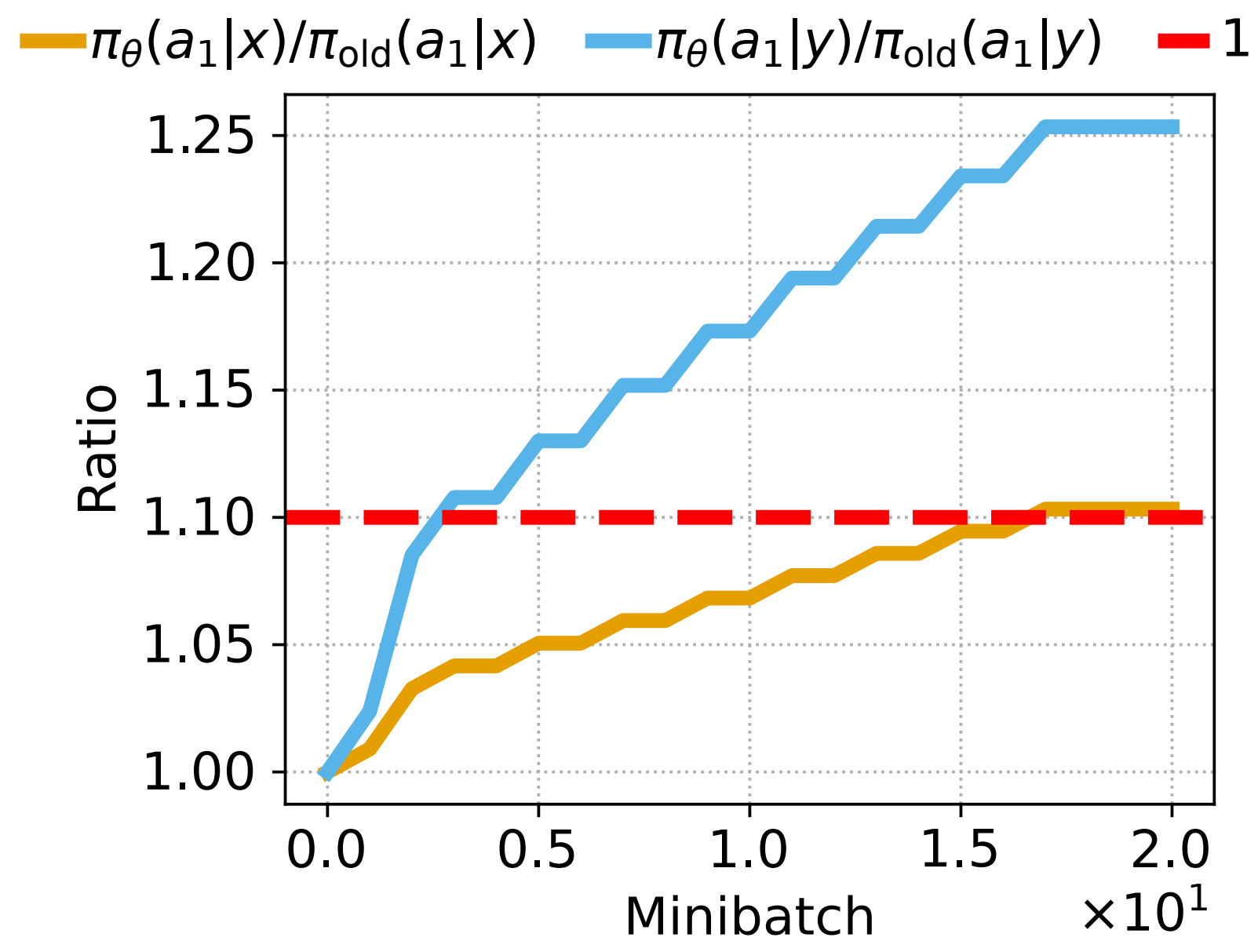
The collapsed policy has **high entropy, but zero variance across states**. All neurons are dead and the **model doesn't distinguish between states**, leading to trivial performance.

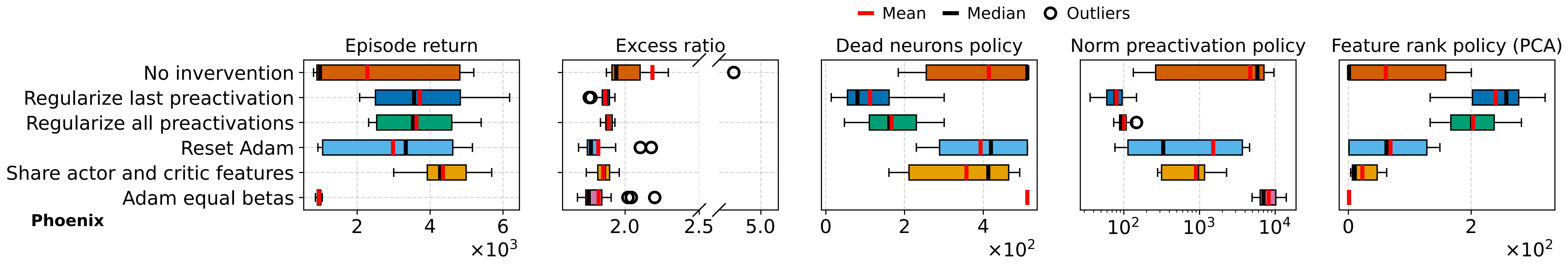




The **trust region cannot prevent this catastrophic change** as **it also breaks down with a poor representation.** The **policy's plasticity also becomes so poor** that the **agent cannot recover** by optimizing the surrogate objective.

We show that it's **not possible to maintain the trust region with a collapsed representation.**





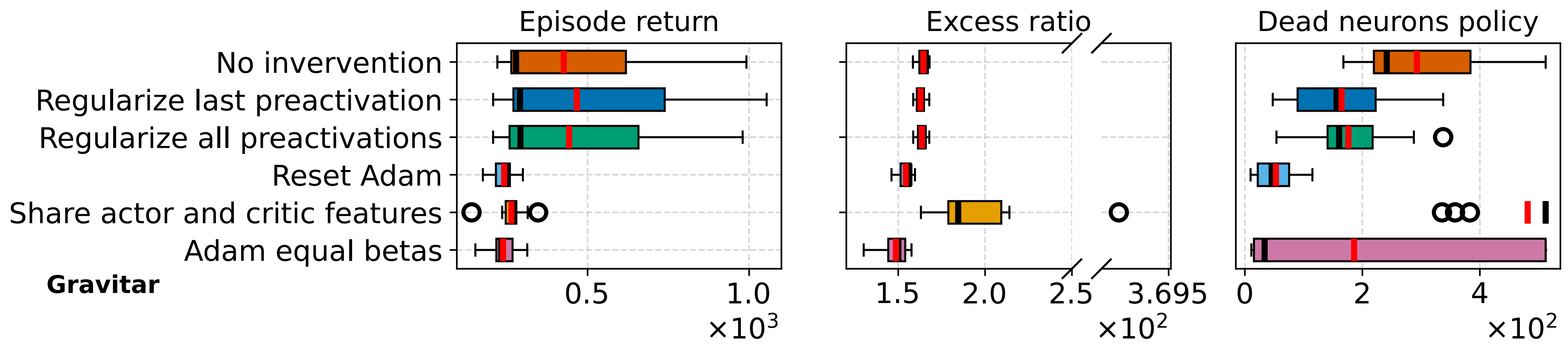
Regularizing representations and non-stationarity results in a better trust region and **mitigates performance collapse.**

Proximal Feature Optimization (PFO)

A simple auxiliary loss to motivate controlling the representation

$$L^{PFO}_{\pi_{old}}(\theta) = \mathbb{E}_{\pi_{old}}\left[\sum_{t=0}^{t_{\max}-1}\left(\phi_\theta(S_t) - \phi_{old}(S_t)\right)^2\right]$$

Bonus: when should you share actor-critic features? Probably not when rewards are sparse as it drives the critic to rank collapse, and the actor with it.


Episode return
Excess ratio
Dead neurons policy
No invervention
Regularize last preactivation
Regularize all preactivations
Reset Adam
Share actor and critic features
Adam equal betas
Gravitar
×10³
×10²
×10²