

# Proximal Curriculum Learning [Tzannetos et al. 2023]

- ▶ Idea from educational psychology: choose tasks that are not too hard and not too easy [Vygotsky and Cole 1978]
- ▶ use *probability of success*  $PoS_{\theta_t}(s)$  as measure of difficulty
- ▶ control difficulty by adjusting start state
- ▶ get starting state by targeting  $PoS_{\theta_t}(s) = 0.5$

- ▶  $\mathbb{P}(s_t^0) \propto \exp(\beta \cdot PoS_{\theta_t}(s) \cdot (PoS^*(s) - PoS_{\theta_t}(s)))$
- ▶  $PoS^* \approx 1$ : optimal policy reaches goal
- ▶  $PoS_{\theta_t}(s)$ : approximated by rollouts or  $V_{\theta_t}(s)$
- ▶  $\beta$ : adjust sampling uniformity

- ▶  $\mathbb{P}(s_t^0) \propto \exp(\beta \cdot PoS_{\theta_t}(s) \cdot (PoS^*(s) - PoS_{\theta_t}(s)))$
- ▶  $PoS^* \approx 1$ : optimal policy reaches goal
- ▶  $PoS_{\theta_t}(s)$ : approximated by rollouts or  $V_{\theta_t}(s)$
- ▶  $\beta$ : adjust sampling uniformity

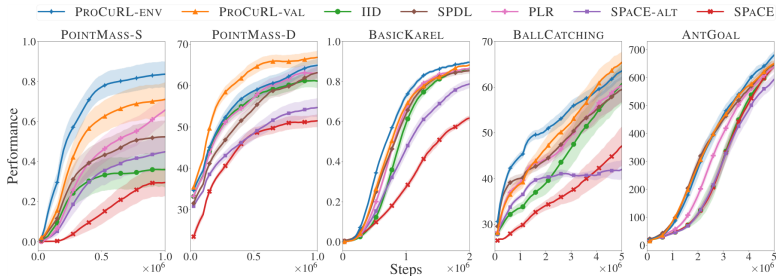


Figure 2: Performance comparison of RL agents trained using different curriculum strategies described in Section 4.2. The performance is measured as the mean reward ( $\pm 1$  standard error) on the training pool of tasks. The results are averaged over 20 random seeds for POINTMASS-S and POINTMASS-D, 10 random seeds for BASICKAREL and BALLCATCHING, and 5 random seeds for ANTGOAL. The plots are smoothed across 5 evaluation snapshots happening at over 25000 training steps.