Prioritized Level Replay

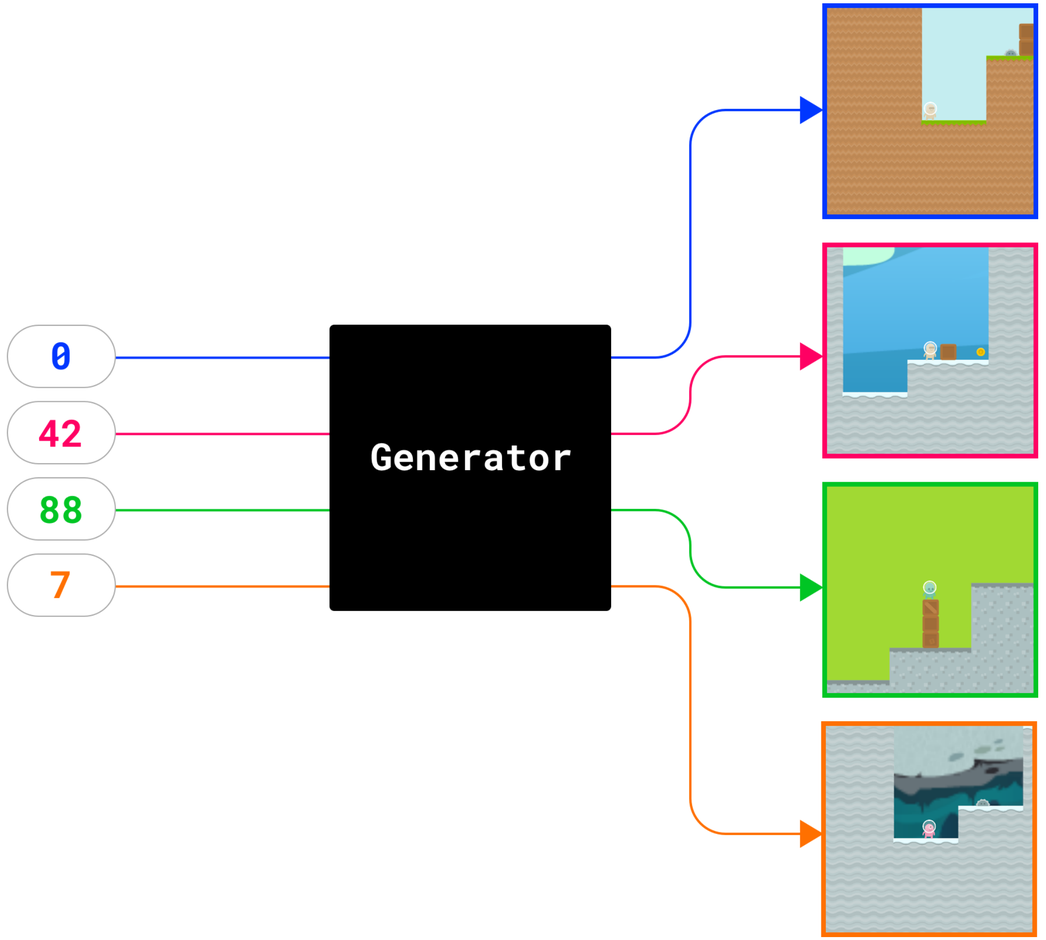
Minqi Jiang, Edward Grefenstette, Tim Rocktäschel
{msj,egrefen,rockt}@fb.com



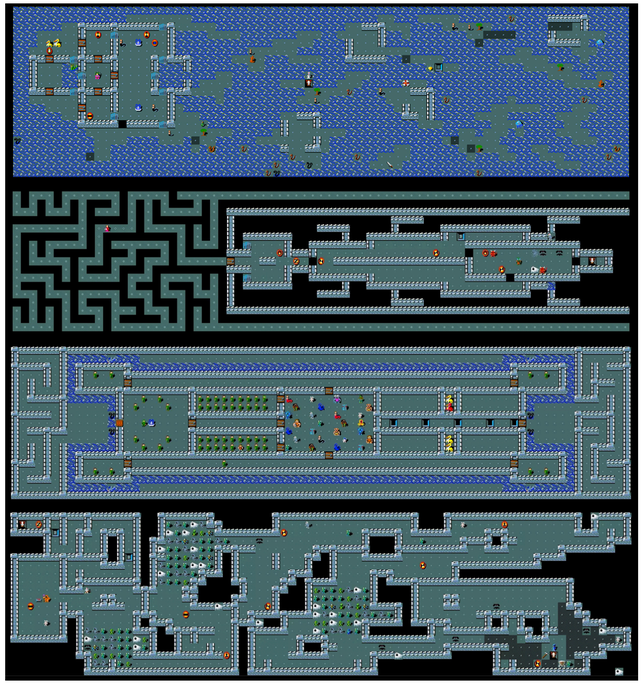
UCL
FACEBOOK AI

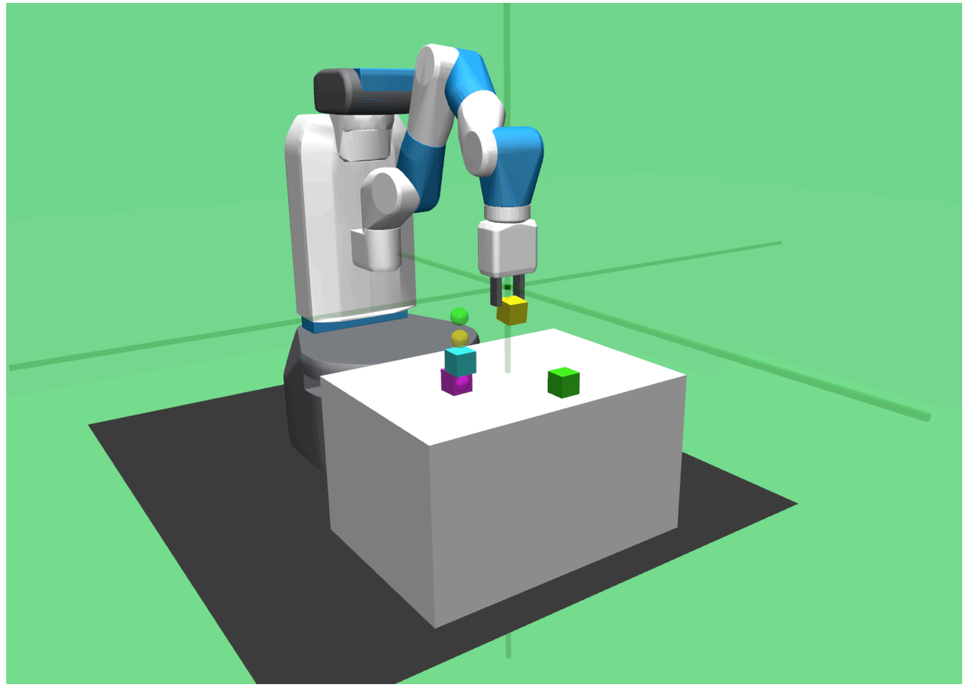
Procedural Content Generation (PCG)

- Environment varies across episodes
- We assume a **blackbox** seeded simulator
- Can test generalization from N training levels



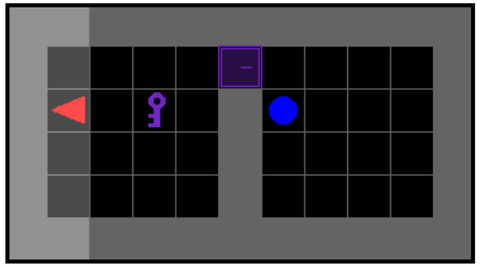
The Generality of PCG



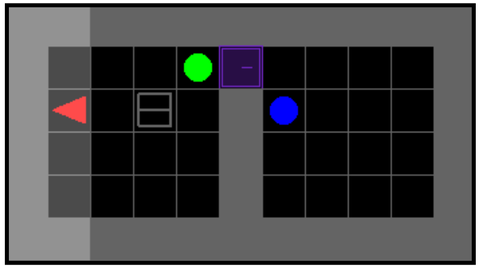


Preliminary motivation

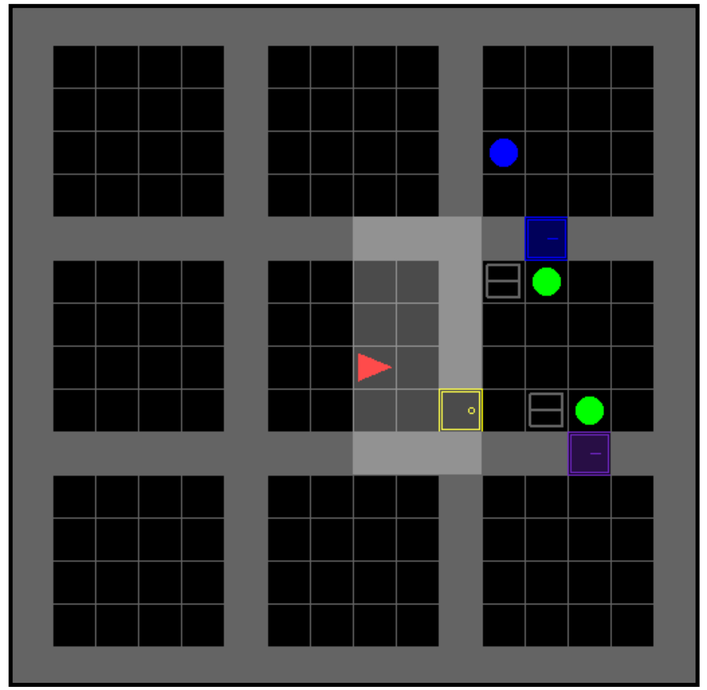
Difficulty: 0

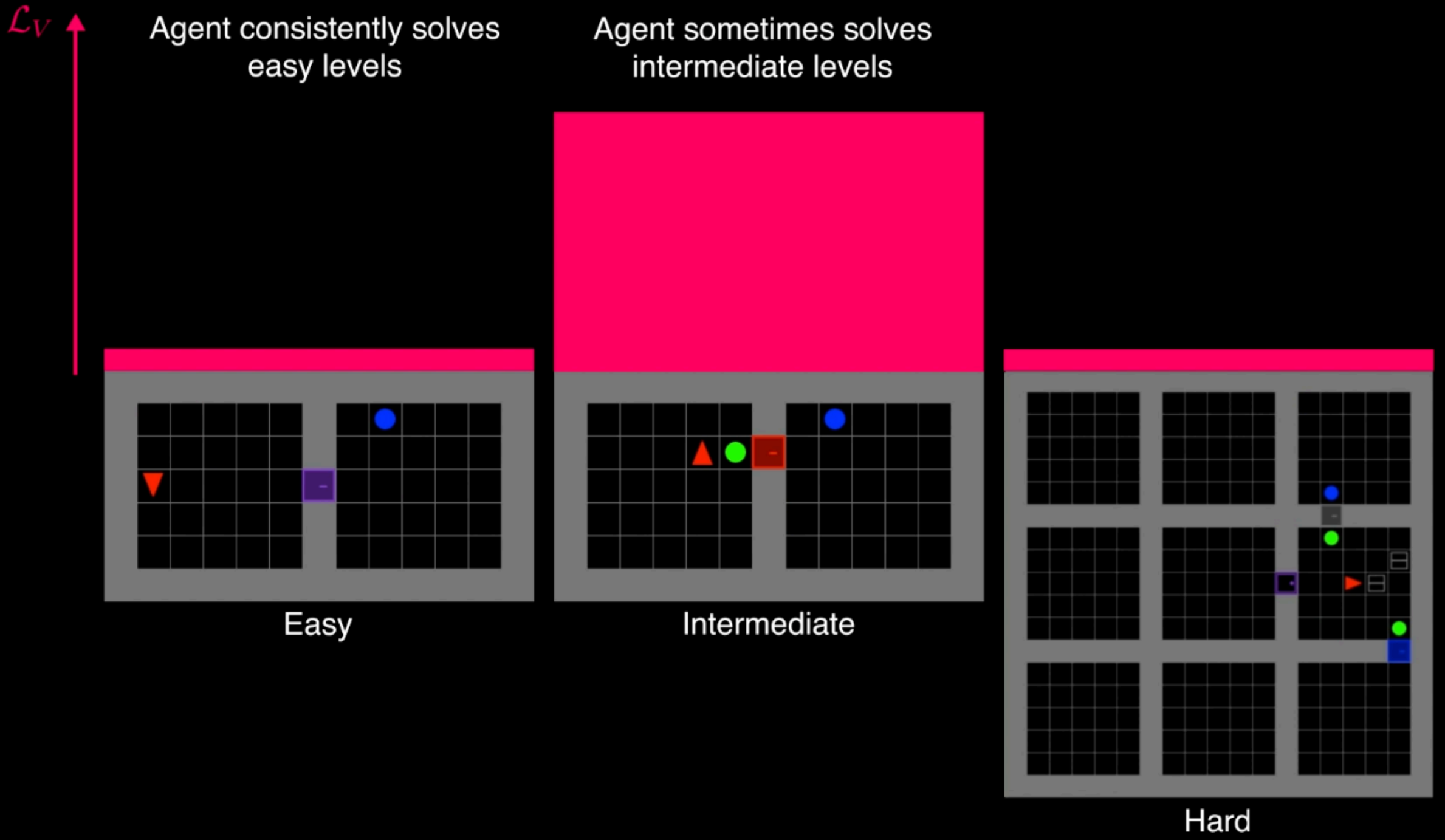


Difficulty: 3

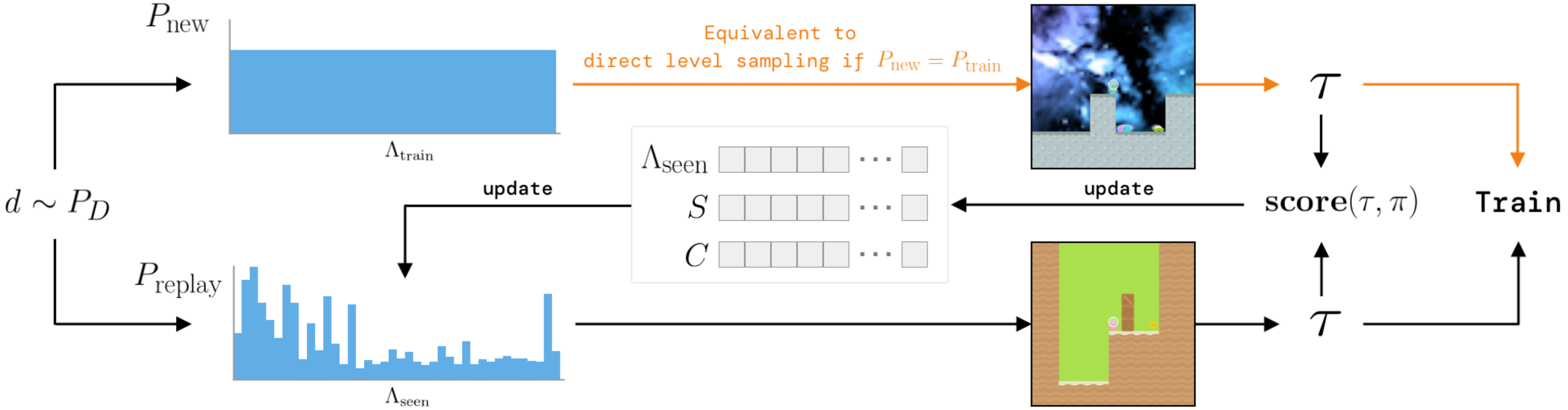


Difficulty: 6




$\mathcal{L}_V$
Agent consistently solves easy levels
Agent sometimes solves intermediate levels
Easy
Intermediate
Hard

PLR Overview



Level scores → sampling distribution

$$P_S(l_i|\Lambda_{\text{seen}}, S) = \frac{h(S_i)^{1/\beta}}{\sum_j h(S_j)^{1/\beta}}$$

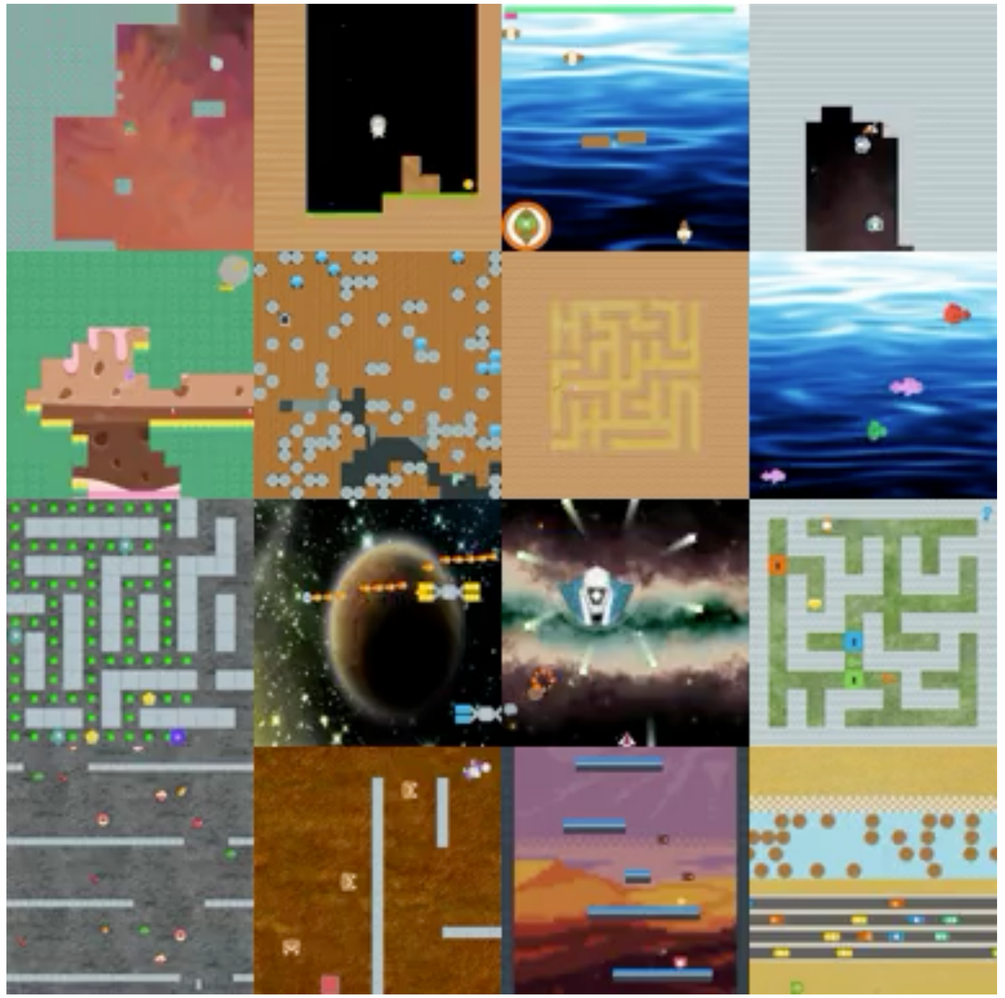
Rank prioritization: *h* is 1/rank(S_i)
Proportional prioritization: *h* is identity

$$P_C(l_i|\Lambda_{\text{seen}}, C, c) = \frac{c - C_i}{\sum_{C_j \in C} c - C_j}$$

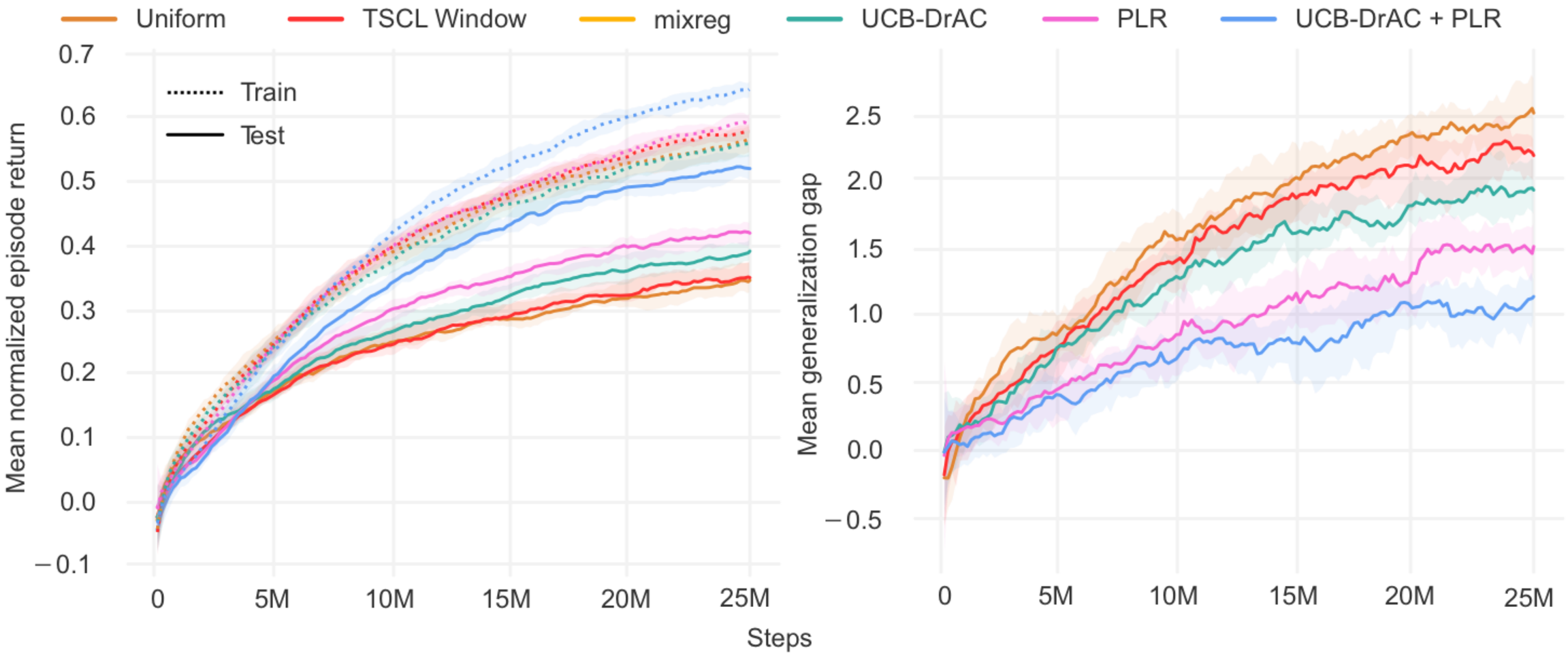
$$P_{\text{replay}}(l_i) = (1-\rho) \cdot P_S(l_i|\Lambda_{\text{seen}}, S) + \rho \cdot P_C(l_i|\Lambda_{\text{seen}}, C, c)$$

OpenAI Procgen Benchmark

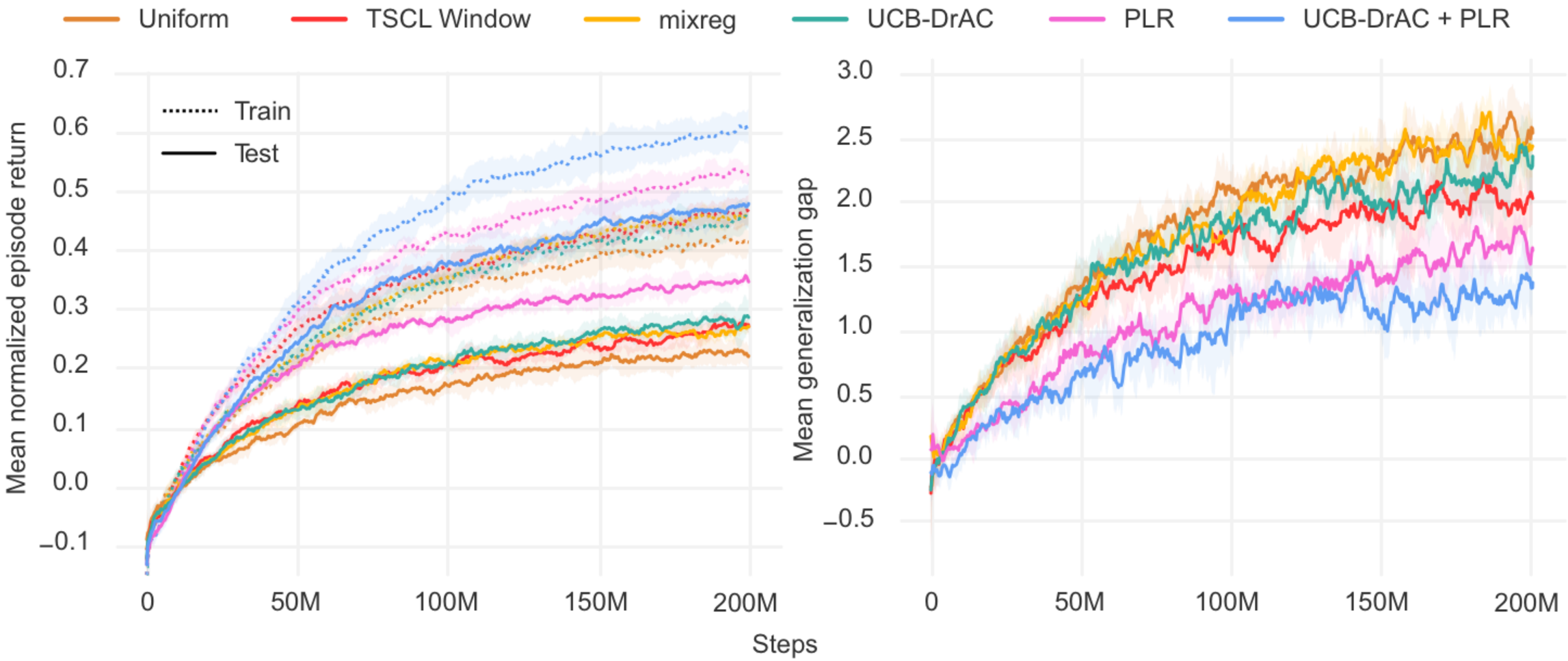
- 16 diverse PCG environments
- Pixel observations
- Generalization from limited training seeds



Results on Procgen easy (+28%, +76%)



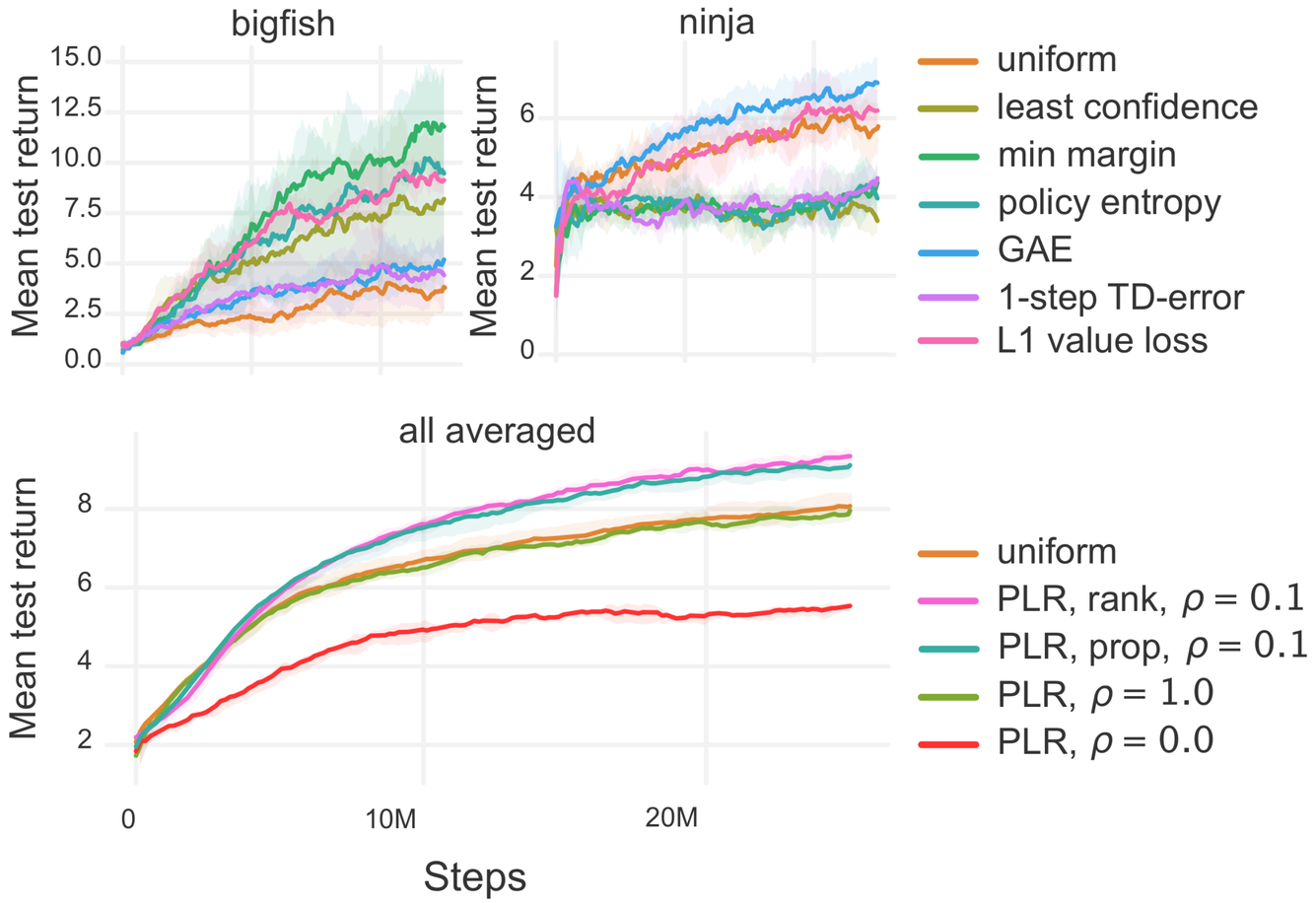
Results on Procgen hard (+35%, +83%)



UCL FACEBOOK AI

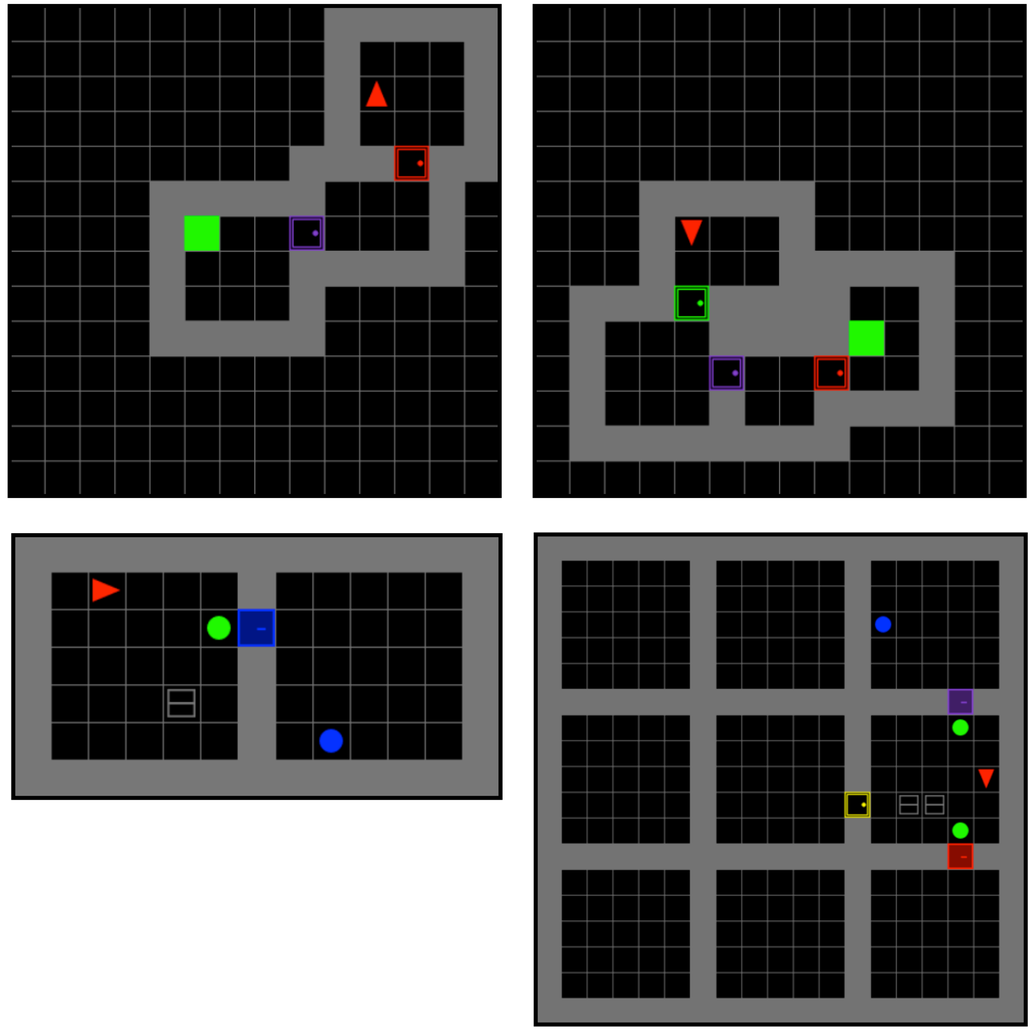
Comparing design choices

Scoring function	$\mathbf{score}(\tau, \pi)$
Policy entropy	$\frac{1}{T}\sum_{t=0}^{T}\sum_{a}\pi(a, s_t)\log\pi(a, s_t)$
Policy min-margin	$\frac{1}{T}\sum_{t=0}^{T}(\max_a \pi(a, s_t) - \max_{a\neq\max_a \pi(a,s_t)} \pi(a, s_t))$
Policy least-confidence	$\frac{1}{T}\sum_{t=0}^{T}(1 - \max_a \pi(a, s_t))$
1-step TD error	$\frac{1}{T}\sum_{t=0}^{T}\lvert\delta_t\rvert$
GAE	$\frac{1}{T}\sum_{t=0}^{T}\sum_{k=t}^{T}(\gamma\lambda)^{k-t}\delta_k$
GAE magnitude (L1 value loss)	$\frac{1}{T}\sum_{t=0}^{T}\left\lvert\sum_{k=t}^{T}(\gamma\lambda)^{k-t}\delta_k\right\rvert$

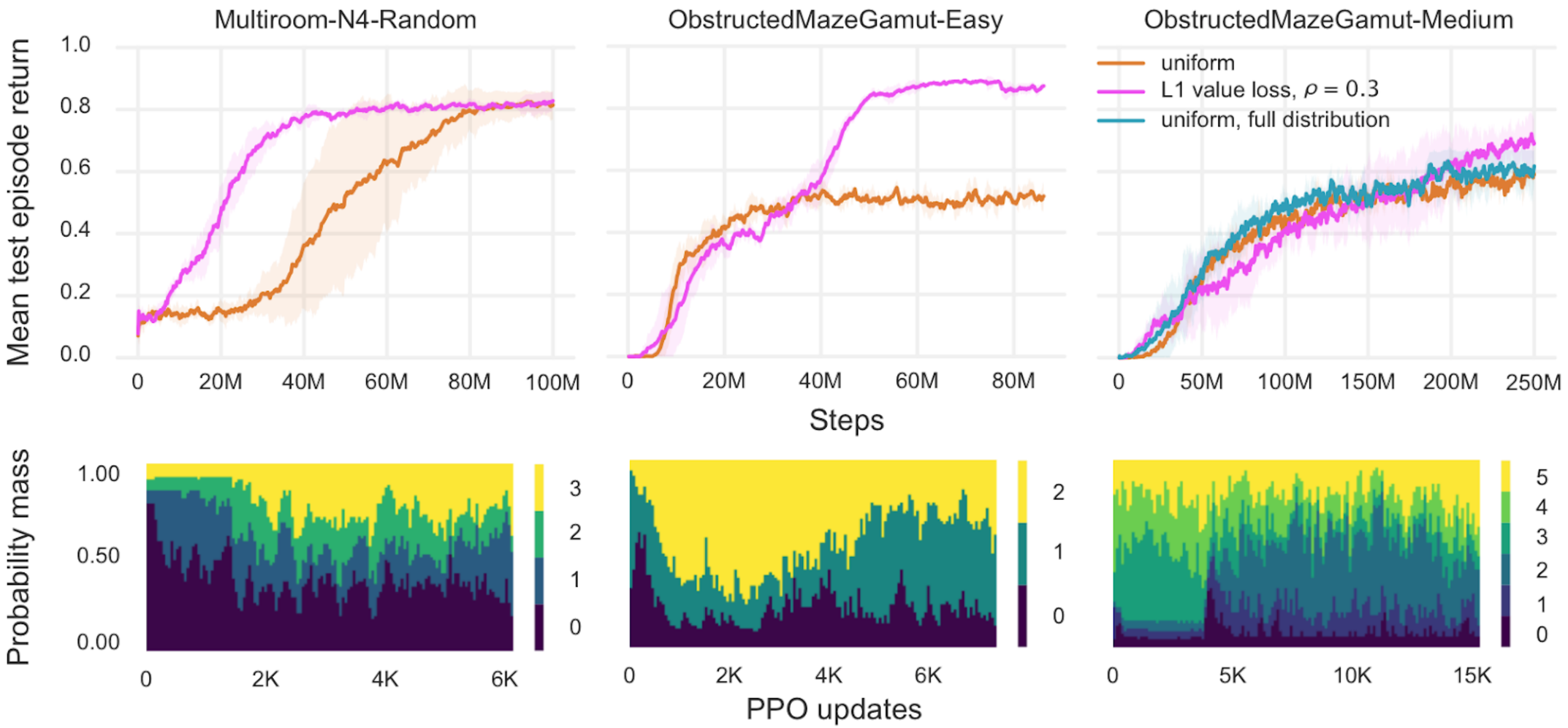


MiniGrid

- Hard environments with discrete difficulty settings.
- Generalization from limited training seeds of each difficulty.



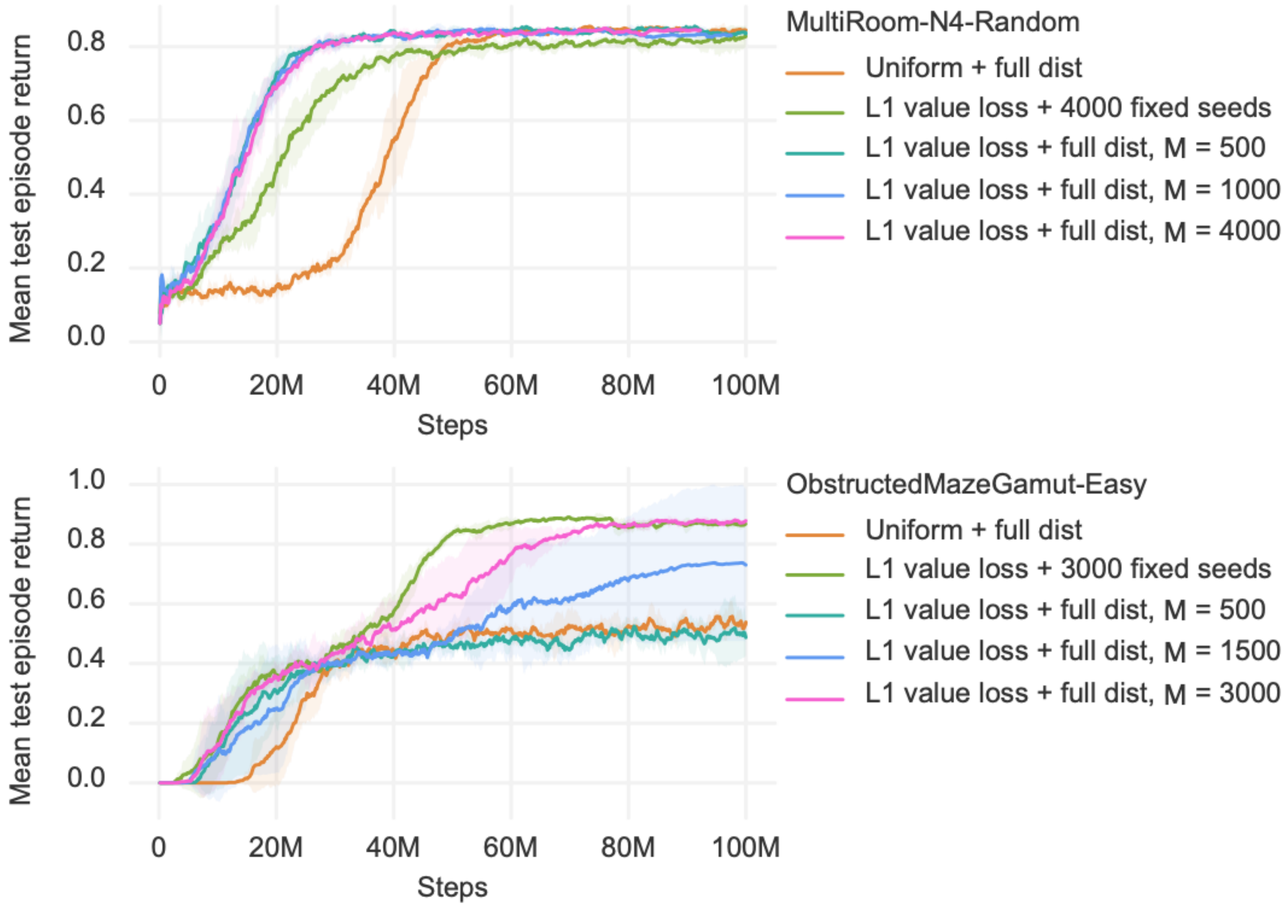
Results on MiniGrid



Training on the full level distribution

- Track top *M* levels
- Replace $l_{\min} = \arg\min_{l} P_{\text{replay}}(l)$

  if new level has higher learning potential



Thanks for listening!

PLR is open source
https://github.com/facebookresearch/level-replay

We look forward to any questions and feedback at our poster session

UCL FACEBOOK AI