# DISCOVER: Automated Curricula for Sparse-Reward RL

Leander Diaz-Bone [*,1], Marco Bagatella [*,1,2], Jonas Hübotter [*,1], Andreas Krause [1]

[1] ETH Zürich, [2] Max Planck Institute for Intelligent Systems

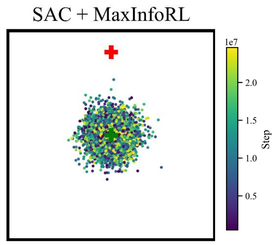
ETH zürich

Learning & Adaptive Systems

## Background

- Exploration in long-horizon, sparse-reward RL is challenging.
- Standard RL (+ curiosity) does not explore effectively.

**How can an agent explore towards an unseen target?**




## Method

- We propose **DISCOVER**, which selects exploratory goals trading off **achievability**, **novelty** & **relevance** to the target:

$$g_t = \arg\max_{g \in \mathcal{G}_{\text{ach}}} \alpha_t \Big[ \underbrace{V(s_0, g)}_{\textbf{Achievability}} + \beta_t \underbrace{\sigma(s_0, g)}_{\textbf{Novelty}} \Big] + (1-\alpha_t) \Big[ \underbrace{V(g, g^\star) + \beta_t \sigma(g, g^\star)}_{\textbf{Relevance}} \Big]$$

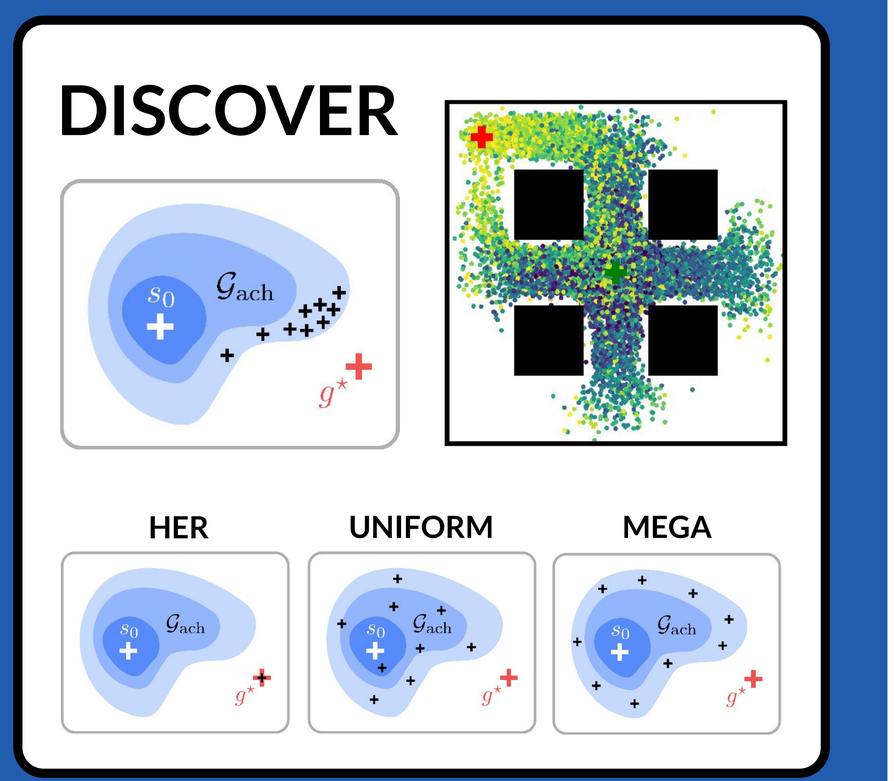
- All components are estimated using an ensemble of goal-conditioned value functions, learning how goals relate.
- We automatically adapt α to match 50% goal achievement rate.
- We connect DISCOVER to principled exploration in bandits, bounding the target task achieval time:

$$N \leq \widetilde{O}\left(\frac{Dd^2}{\kappa^3}\right)$$

$N$: Episodes until $g^\star$ becomes reachable
$D = V^*(s_0, g^\star)$: Optimal distance between $s_0$ and $g^\star$
$\kappa$: Expansion rate
$d$: Feature dimensionality

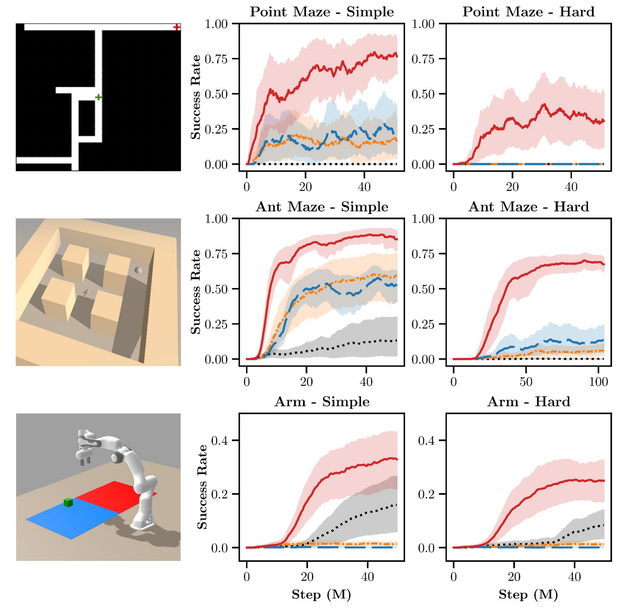
## Long-horizon, sparse-reward RL tasks are challenging.

## Only agents that ***direct*** exploration can solve them.




## Results

- DISCOVER outperforms baselines on complex sparse-reward control tasks.

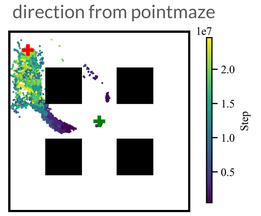



- DISCOVER effectively directs exploration.

Number of Steps (M) required to reach 10% success rate in d-dimensional pointmazes

| Dimension | 2 | 3 | 4 | 5 | 6 |
|---|---|---|---|---|---|
| HER | ∞ | ∞ | ∞ | ∞ | ∞ |
| MEGA | 4.8 | ∞ | ∞ | ∞ | ∞ |
| Ach. + Nov. | 5.2 | ∞ | ∞ | ∞ | ∞ |
| DISCOVER | **2.9** | **3.1** | **7.4** | **5.4** | **18.7** |

DISCOVER + pre-trained direction from pointmaze




Components of DISCOVER after 8M steps (antmaze-hard)