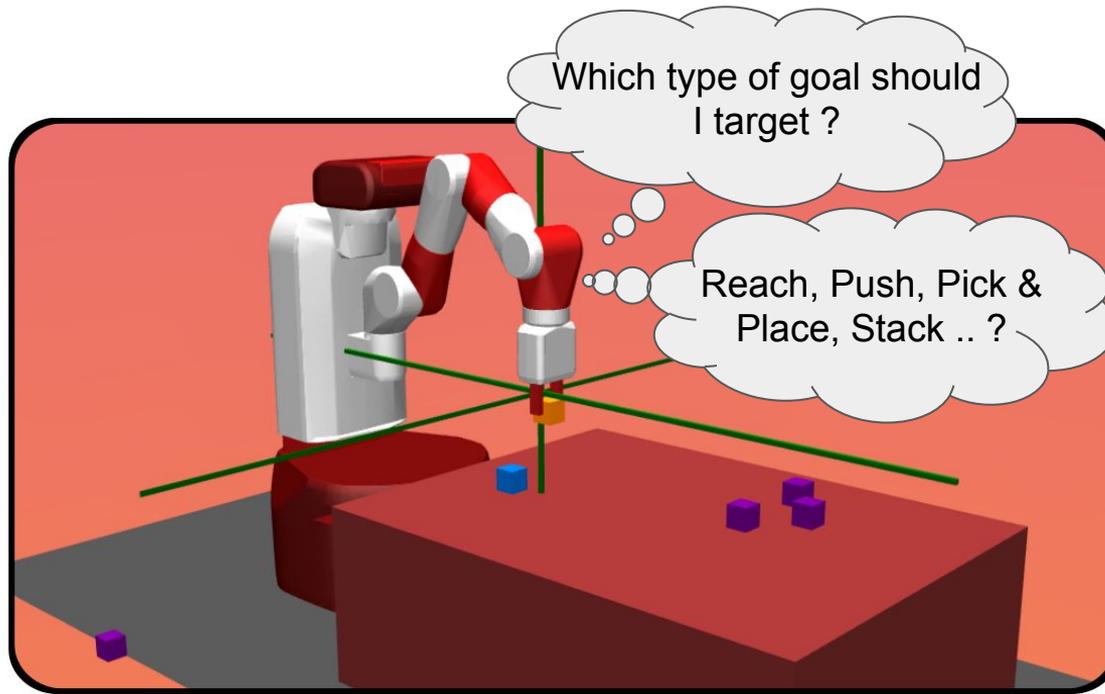# CURIOUS: Intrinsically Motivated Modular Multi-Goal RL

Cédric Colas
Phd student @ Flowers team, INRIA

Co-authors:
Pierre Fournier, Olivier Sigaud, Mohamed Chetouani, Pierre-Yves Oudeyer

# Problem: Intrinsically Motivated Modular Multi-Goal RL

Modular Multi-Goal Fetch Arm environment

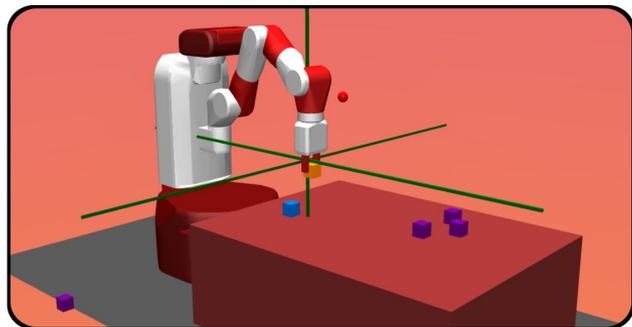# Problem: Intrinsically Motivated Modular Multi-Goal RL

Modular Multi-Goal Fetch Arm environment

Problem: Intrinsically Motivated Modular Multi-Goal RL
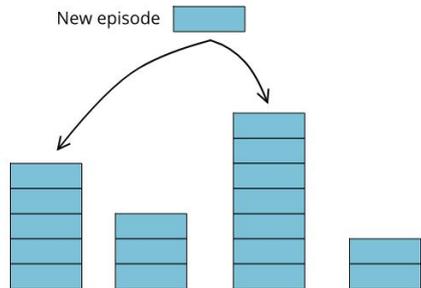

Modular Multi-Goal Fetch Arm environment

Controllable objects (learnable goals)

Distracting objects (unlearnable goals)
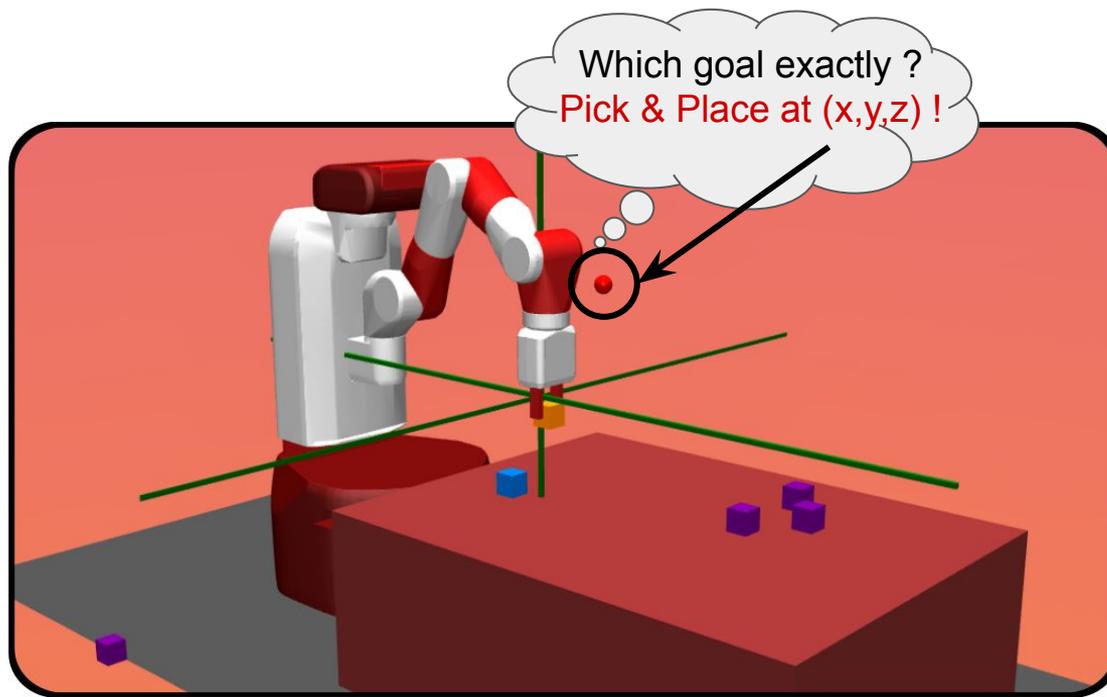
Curious: Intrinsically Motivated Modular Multi-Goal RL
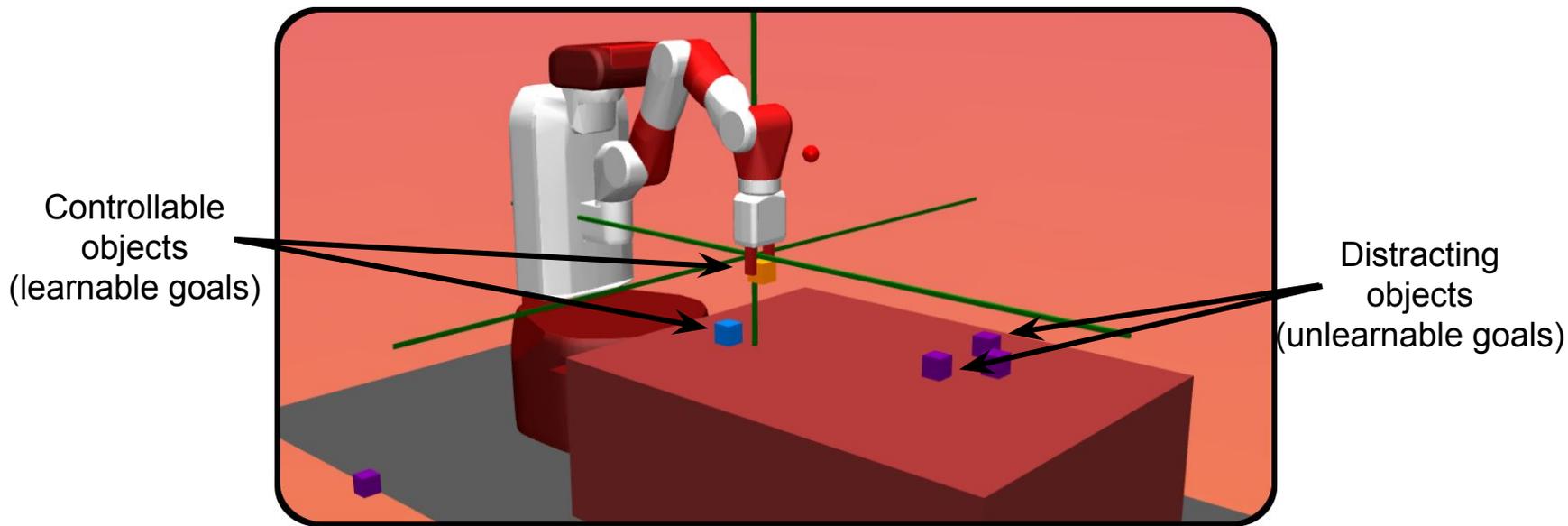
The Curious Algorithm

External world

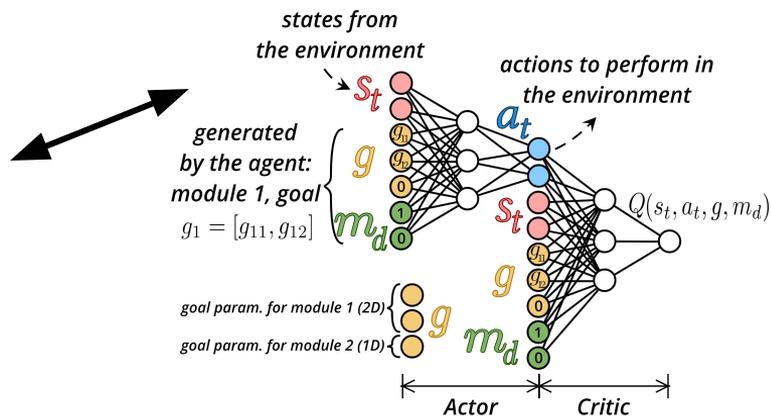**Modular goal encoding for UVFA:**[1]

e.g. of modular goals:

Move gripper to (x,y,z)
Pick & Place cube2 at (x,y,z)
Push cube1 at (x,y)

Sampling of modules and goals using **absolute learning progress**[2]
(using Bandit algorithm)

states from the environment

generated by the agent: module 1, goal
$g_1 = [g_{11}, g_{12}]$

actions to perform in the environment

$Q(s_t, a_t, g, m_d)$

goal param. for module 1 (2D)
goal param. for module 2 (1D)

Actor   *Critic*

New episode

Modular replay buffer:
with hindsight learning[3, 4]
(module and goal substitutions)

Replay buffers per task

1: UVFA, Schaul et al., 2015
2: IMGEP, Forestier, 2017
3: HER, Andrychowicz et al., 2017
4: Unicorn, Mankowitz et al., 2018

Curious: Intrinsically Motivated Modular Multi-Goal RL
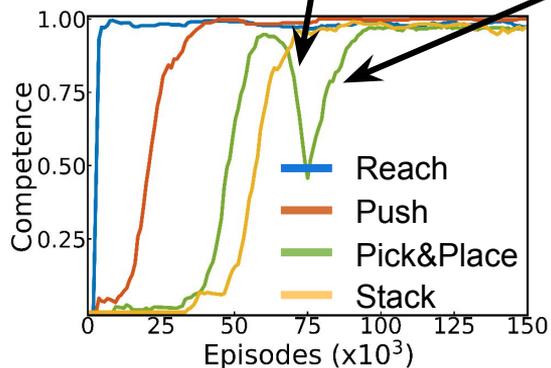
Modular goal encoding vs Multi-Goal Module Experts
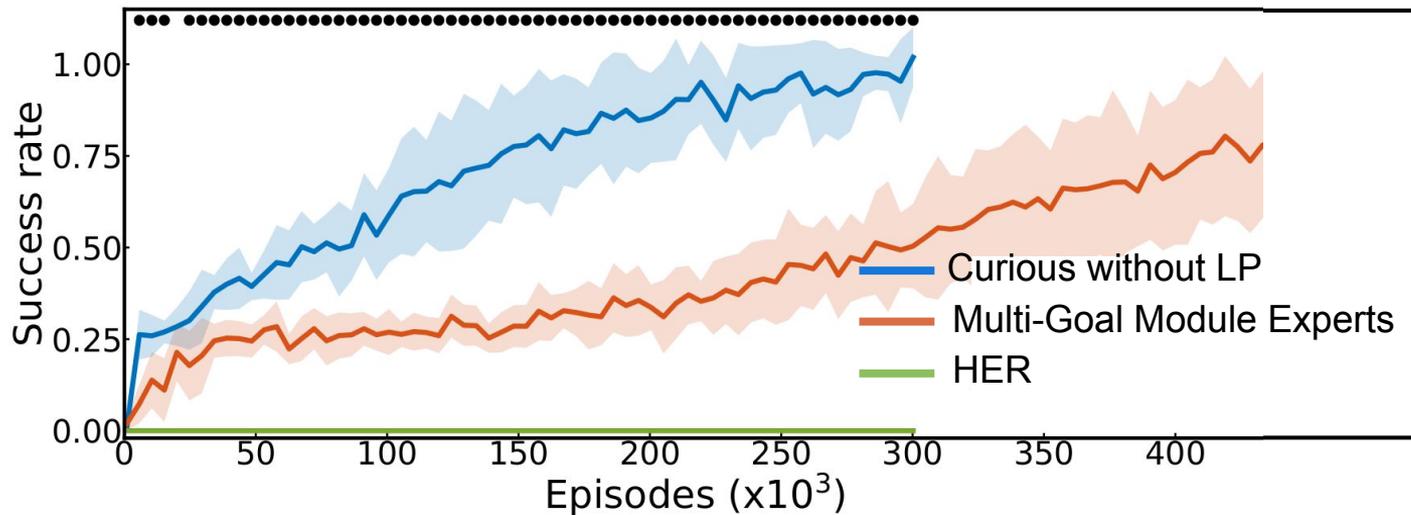
Impact of the policy and value function architecture. Average success rates over the set of tasks (mean +/- std, 10 seeds).

Curious: Intrinsically Motivated Modular Multi-Goal RL

Automatic Curriculum with Absolute Learning Progress

Forgetting due to interferences among modules/goals

Mitigated thanks to fast LP-based refocus
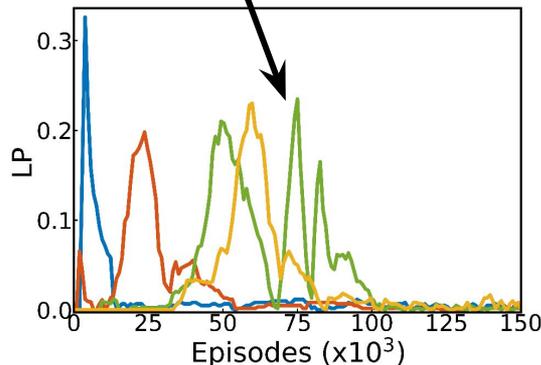
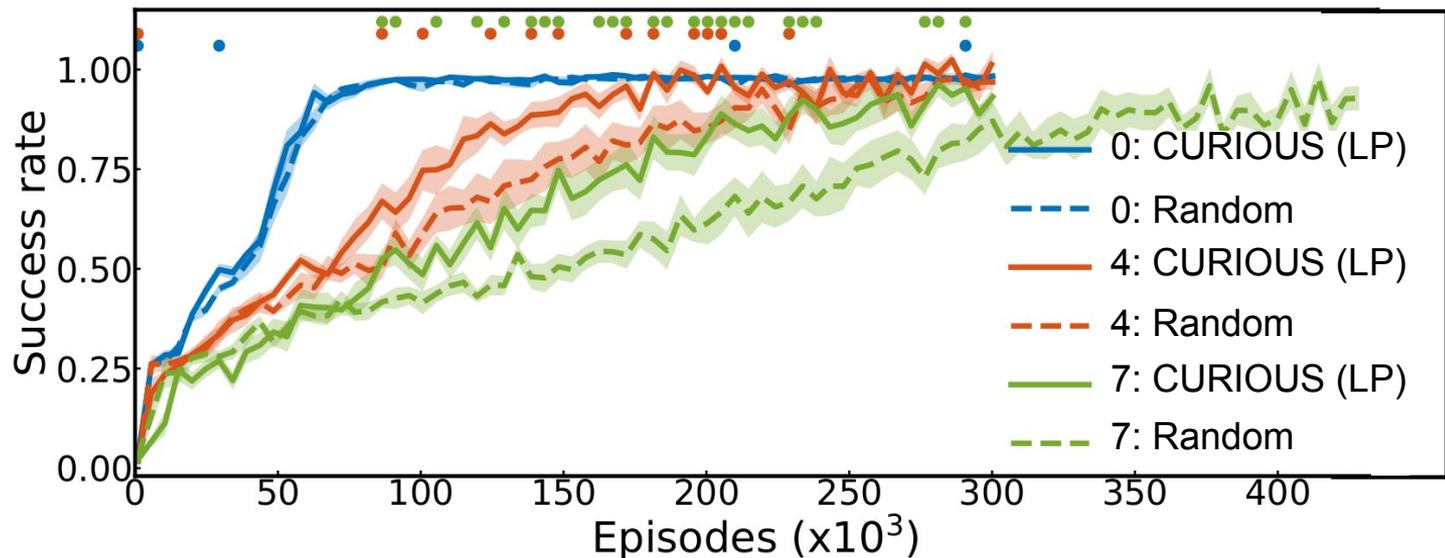Using a bandit for module selection and replay

Competence

Absolute Learning Progress
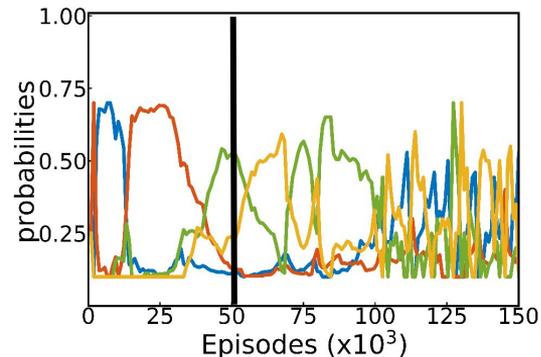
Selection Probabilities

Curious: Intrinsically Motivated Modular Multi-Goal RL
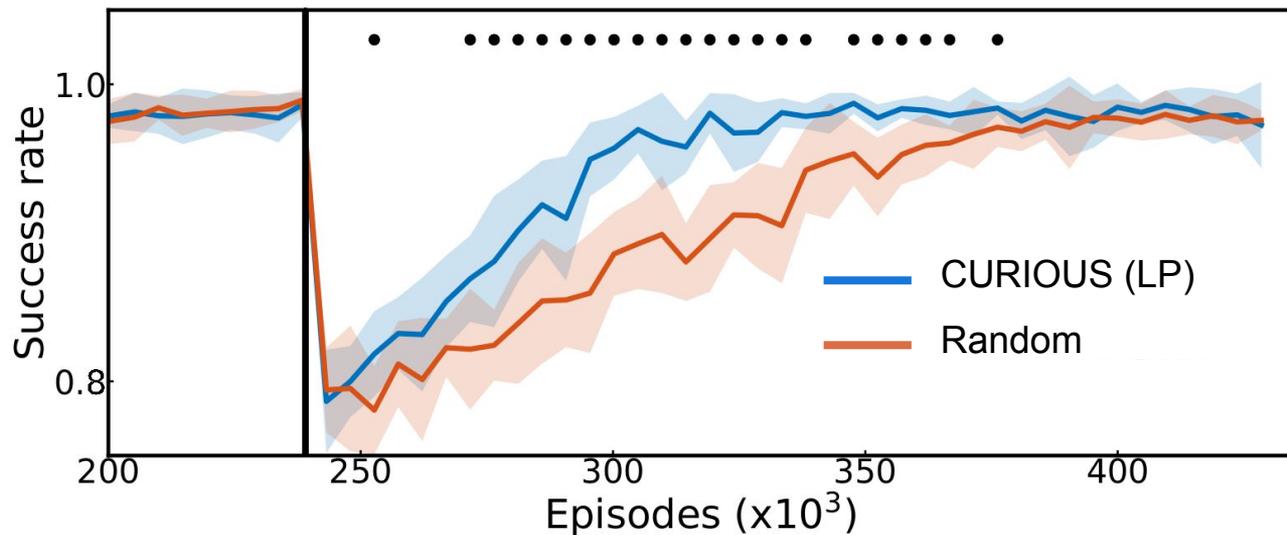
Resilience to Distracting Goals

Resilience to distracting goals: 0, 4 or 7 distracting modules. CURIOUS (intrinsically motivated) and Random (random module). Mean +/- sem, 10 seeds.

Curious: Intrinsically Motivated Modular Multi-Goal RL

Resilience to sensory failure: Recovery following a sensory failure.
Mean +/- std, 10 seeds.

CURIOUS recovers 95 % of its original performance twice as fast as Random.

Curious: Intrinsically Motivated Modular Multi-Goal RL
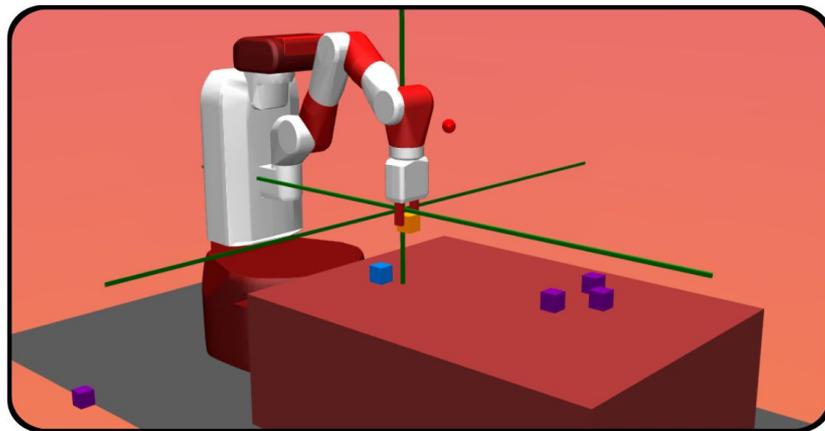
Curious: Intrinsically Motivated Modular Multi-Goal RL

Poster: Pacific Ballroom #42