Events as Triggers for Behavioral Diversity in Multi-Agent Reinforcement Learning

Agents must adopt diverse behaviors and switch them at the right moment, yet current MARL binds fixed behaviors to fixed agent identities. We argue the missing ingredient is **events**.

Hannes Büchi, Manon Flageat*, Eduardo Sebastián*, Amanda Prorok University of Cambridge · *Equal co-supervision

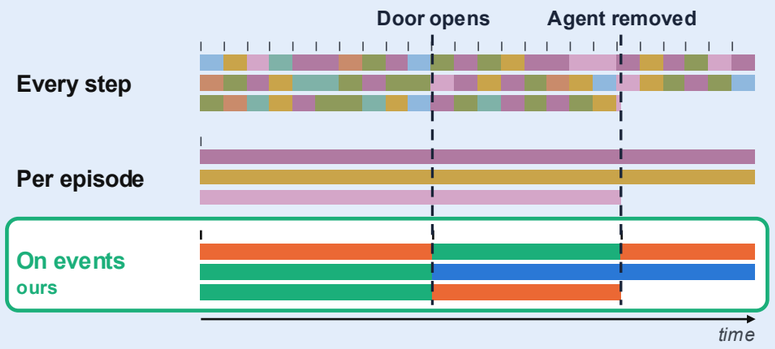
UNIVERSITY OF CAMBRIDGE

NEURAL INFORMATION PROCESSING SYSTEMS

Project page

1 When should an agent switch behavior?

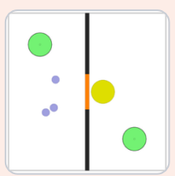
Three agents, two events. Color = behavior; tick = (re)assignment.



Our view: agents draw behaviors from a shared manifold, and an **event** decides when they switch.

Events are state changes that alter the task: a door opens, a teammate is removed, a target is detected.

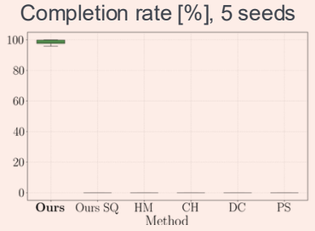
2 Spotlight: Pressure Plate



Needs roles switched in sequence:

1. One agent holds plate 1; the others cross.
2. They hold plate 2; the first agent follows.
3. All agents move to the goal.

Completion rate [%], 5 seeds



Only ours solves it: 98.4% median completion.

All baselines, and ours with a single query at the start (SQ), stay at ≈0%.

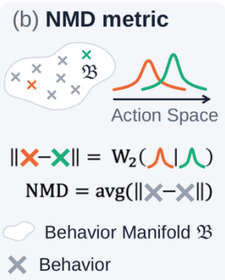
Unseen event: remove the agent that first reaches plate 2, and the team still completes **81.2%**.

3 Framework

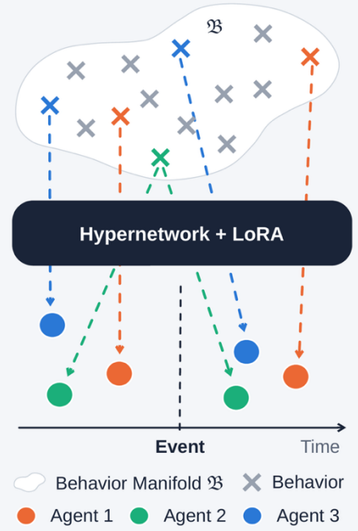
(a) **Events**



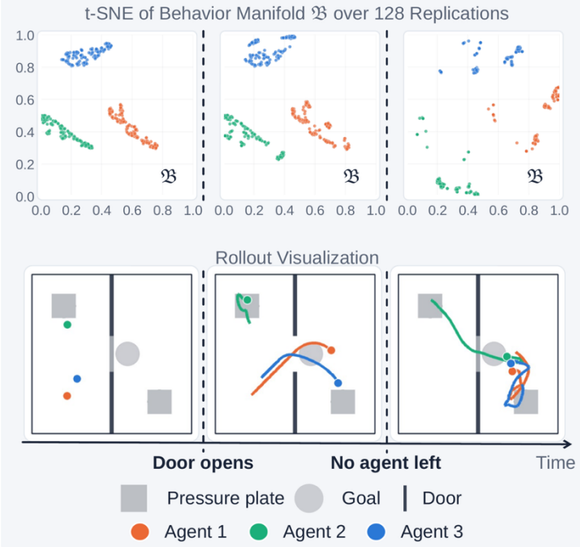
(b) **NMD metric**



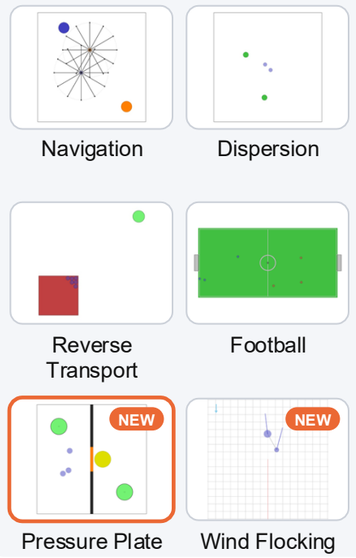
(c) **Proposed Framework**



(d) **Pressure Plate Task Visualization**



Six tasks in VMAS



4 Method

Neural Manifold Diversity (NMD) measures diversity on behaviors, not on agents:

$$\widehat{\mathrm{NMD}}\left(\{\pi_m\}_{m=1}^{B}\right) = \frac{2}{B(B-1)\,|\mathcal{O}_{\mathrm{ep}}|} \sum_{m<n} \sum_{o\in\mathcal{O}_{\mathrm{ep}}} W_2\left(\pi_m(o), \pi_n(o)\right)$$

Event-driven hypernetwork + LoRA: queried at the start and at each event, it emits a low-rank adapter for a shared policy:

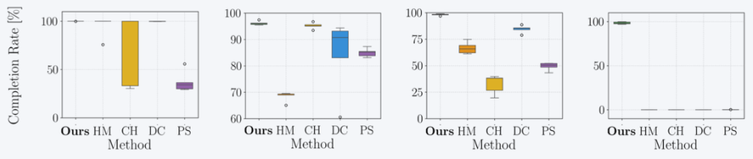
$$z_m(o_t) = W_{\mathrm{shared}}\,\phi(o_t) + \alpha\, D_m C_m\, \phi(o_t)$$

Diversity control by projection: rescaling the adapters hits the target NMD exactly, and reward gradients that would change NMD are projected out (Thm. 3):

$$\nabla_{u_m} R = \alpha\, P_{u_m}\left(\nabla_{z_m} R\right)^{\top}, \qquad P_{u_m} = I - \frac{u_m \otimes \nabla_{u_m} \widehat{\mathrm{NMD}}}{\widehat{\mathrm{NMD}}}$$

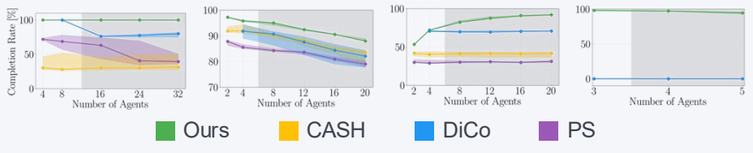
5 Results

First on all four benchmarks, ≥95% completion.



HM: HyperMARL · CH: CASH · DC: DiCo · PS: parameter sharing

Zero-shot to unseen team sizes (gray range).



This work is supported by the European Research Council (ERC) Project 949940 (gAIa), by a Leverhulme Trust Research Project Grant, and by the EPSRC funded INFORMED-AI project EP/Y028732/1. We gratefully acknowledge their support.