NEURAL INFORMATION PROCESSING SYSTEMS

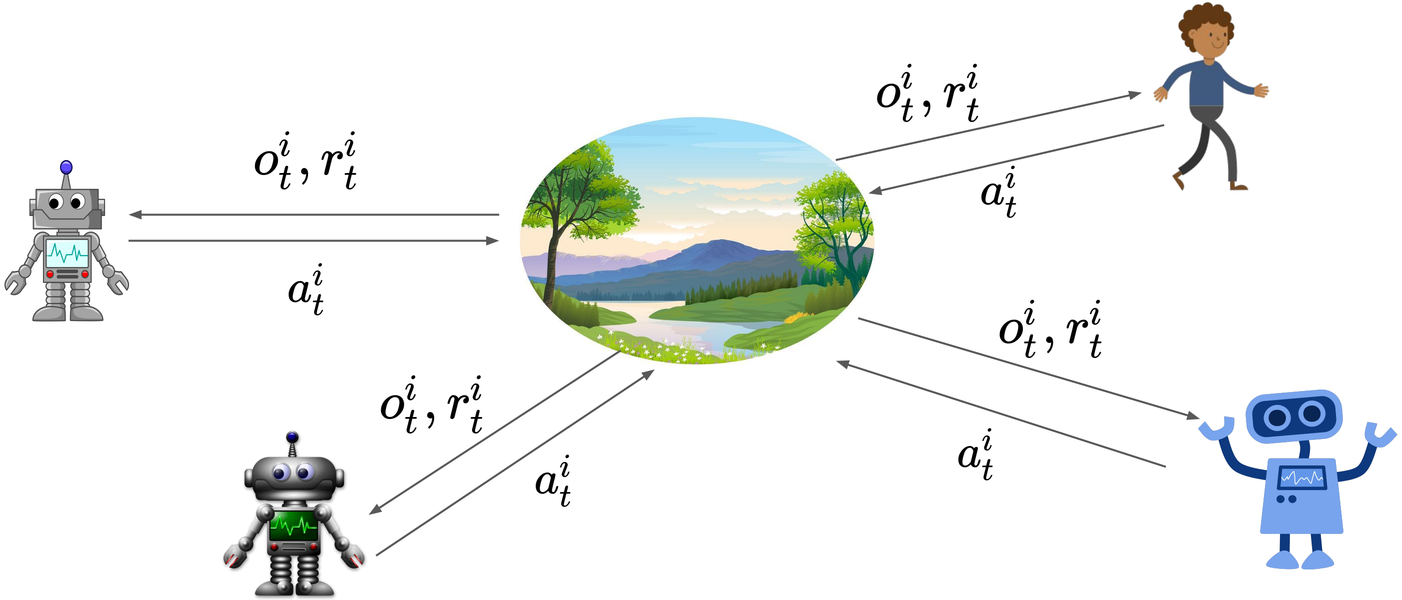
HyperMARL: Adaptive Hypernetworks for Multi-Agent RL

Kale-ab Abebe Tessera, Arrasy Rahman, Amos Storkey, Stefano V. Albrecht

THE UNIVERSITY of EDINBURGH informatics

TEXAS The University of Texas at Austin

Setting: Cooperative MARL - Dec-POMDP



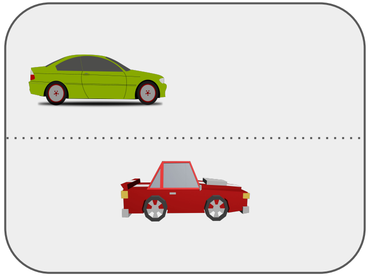
Goal:

$$\boldsymbol{\pi}^* = \arg\max_{\boldsymbol{\pi}} \mathbb{E}_{s_0 \sim \mu,\, \mathbf{a}_t \sim \boldsymbol{\pi}} \left[\sum_{t=0}^{\infty} \gamma^t R(s_t, \mathbf{a}_t) \right].$$

Adaptivity: Individual Specialisation vs Shared Behaviours

Specialisation

Shared Behaviours

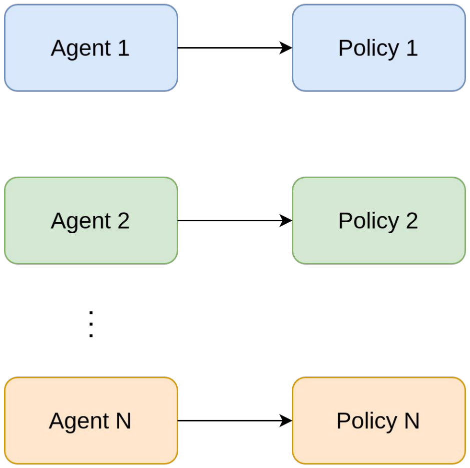


Source: Google Football, Search and Rescue, Robots, Birds, Fish.

Efficient Adaptivity in MARL

No Parameter Sharing (NoPS)

$\pi^i(a^i|o^i)$

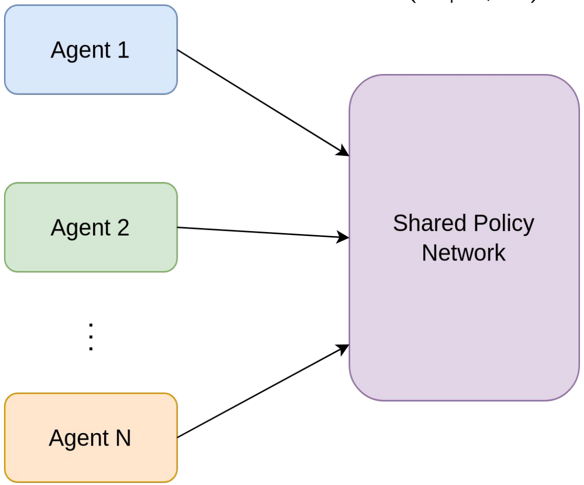


Individual Policies

✅Diverse Behaviour
❌ Sample Inefficient

Full *Parameter Sharing* (FuPS)

$\pi(a^i|o^i, id)$



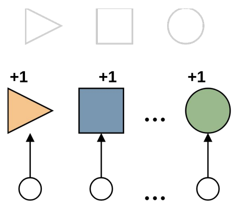
ID =[1,0,0]

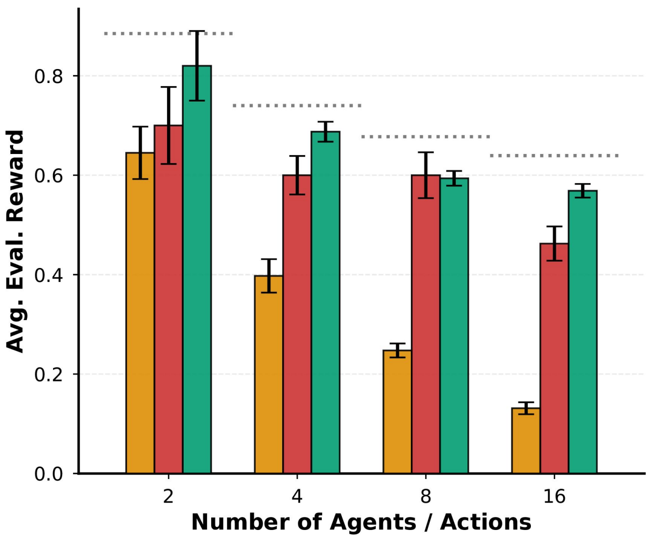
Shared Policy

✅ Scalable
✅ Sample efficient
❌ Diverse behaviours – specialisation.

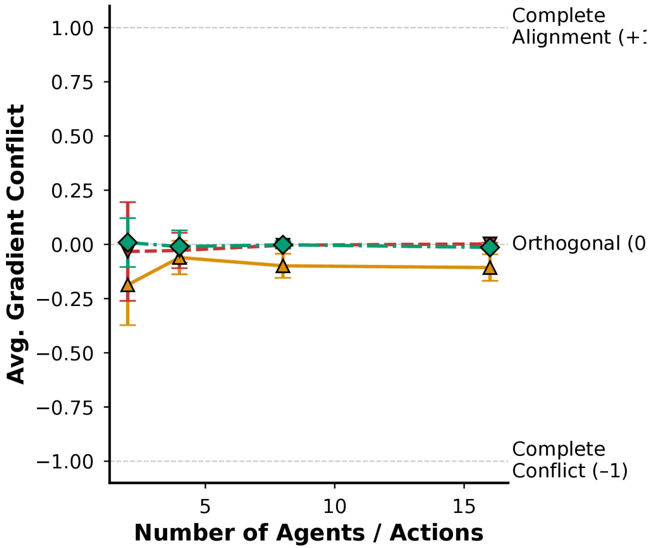
Challenges: Parameter Sharing + Specialisation

Specialisation Game





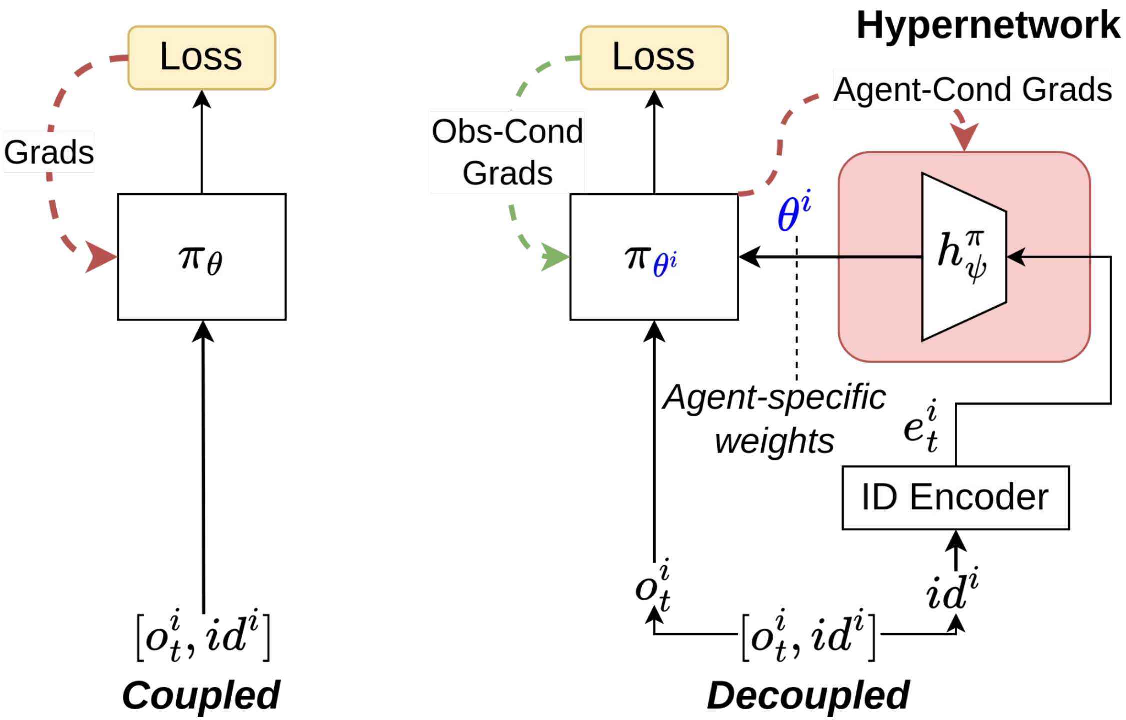
Avg. Evaluation Reward



Avg. Grad Conflict

💡**Coupling agent-IDs** and **observations** = (correlates) **higher gradient interference**

Study Adaptivity in MARL - HyperMARL



FuPS HyperMARL

Does not:
✅ change the **learning objective**.
✅ knowing the **optimal diversity level**.
✅ require **sequential updates**.

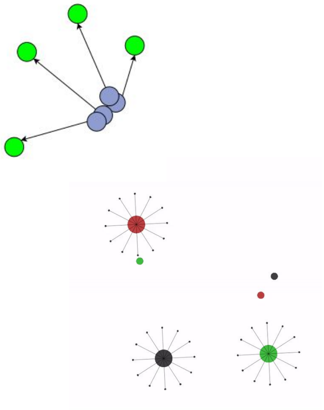
HyperMARL: Gradient Decoupling

Weights of the hypernetwork:

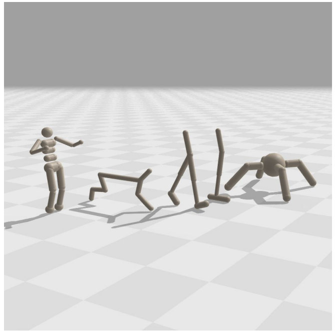
$$\nabla_\psi J(\psi) = \sum_{i=1}^{I} \underbrace{\nabla_\psi h_\psi^\pi(e^i)}_{\mathbf{J}_i \text{ (agent-conditioned)}} \underbrace{\mathbb{E}_{\mathbf{h}_t, \mathbf{a}_t \sim \boldsymbol{\pi}} \left[A(\mathbf{h}_t, \mathbf{a}_t) \nabla_{\theta^i} \log \pi_{\theta^i}(a_t^i \mid h_t^i) \right]}_{Z_i \text{ (observation-conditioned)}}.$$

- ***Agent-conditioned:*** deterministic wrt to mini-batch samples, separating agent identity from traj noise.
- ***Observation-conditioned***: averages trajectory noise **per agent**.

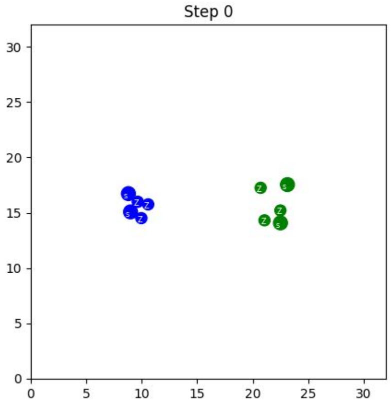
Environments - Specialisation, Homogenous Behaviours, Mixed - 22 scenarios, 6 baselines



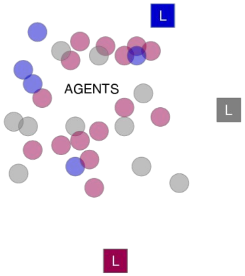
Dispersion, Navigation - Up to 8 agents



Multi-Agent Mujoco - Up to 17 agents

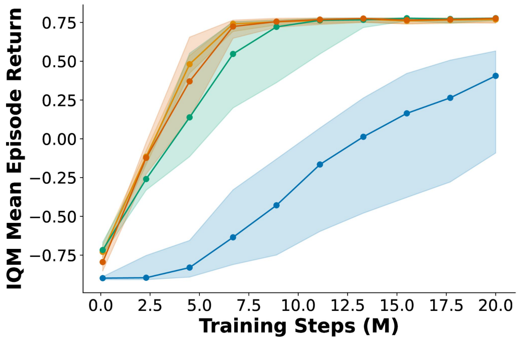


SMAX (Starcraft) - Up to 20 agents

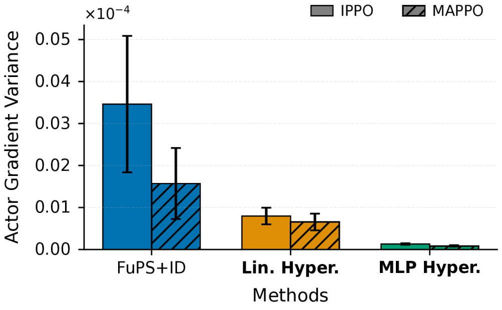


Blind-Particle Spread - Up to 30 agents

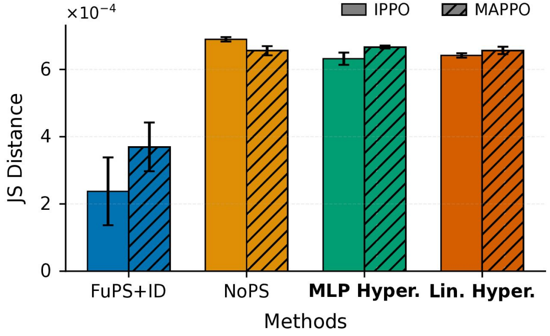
Results: 1. Competitive to NoPS and Baselines in Specialised Tasks



Mean Return



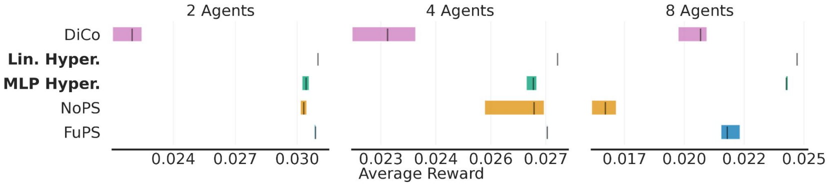
Actor Grad. Variance

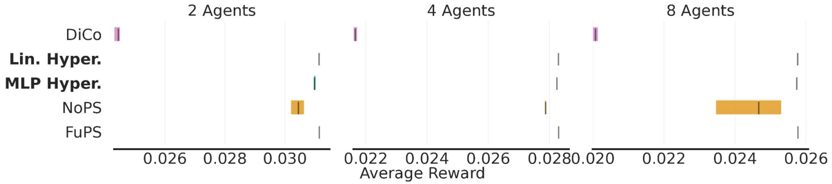


Diversity

FuPS+ID NoPS **MLP Hyper.** **Lin. Hyper.**

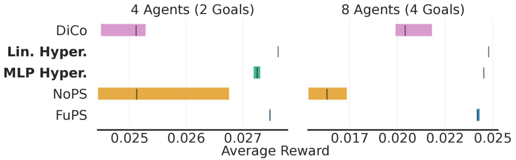
Results: 2. Adaptive Tasks





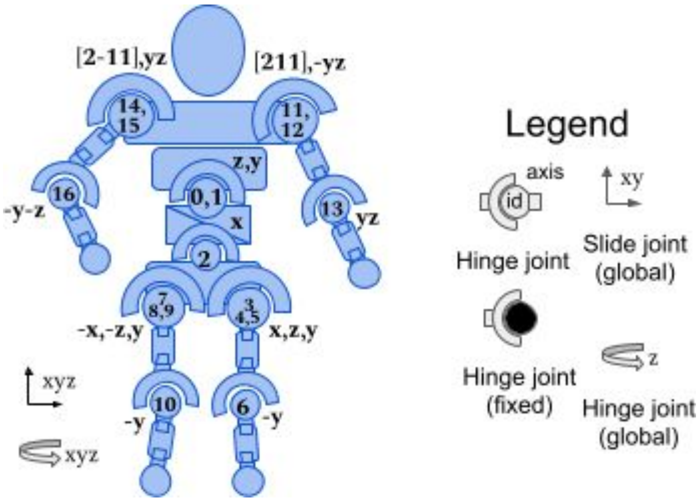
a) Different Goals

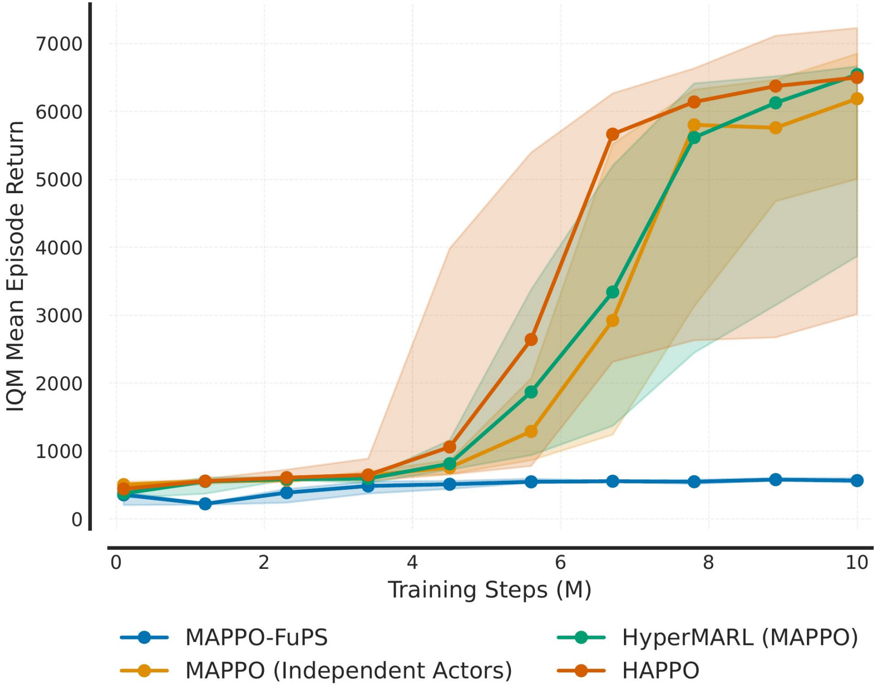
b) Same Goals



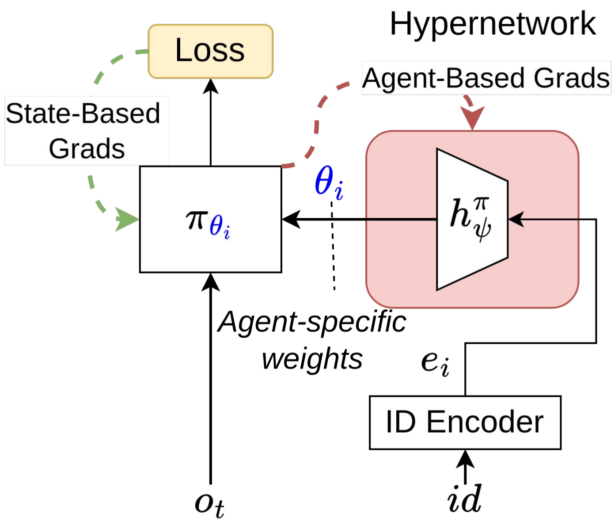
c) Mixed Goals

Results: 3. Large Scale - Humanoid-v2 17x1

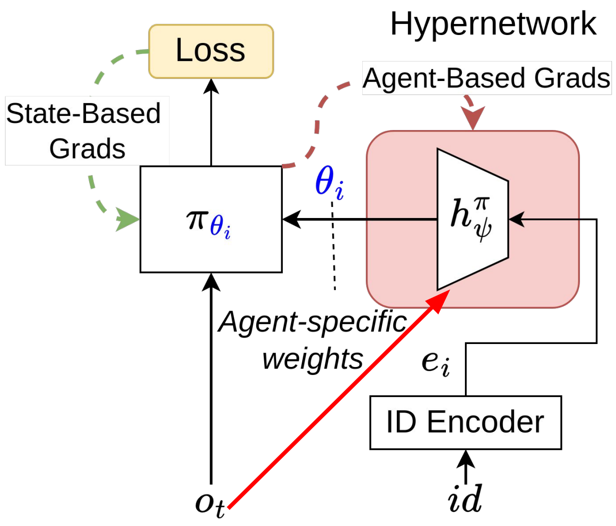




Ablations - Decoupled Grads

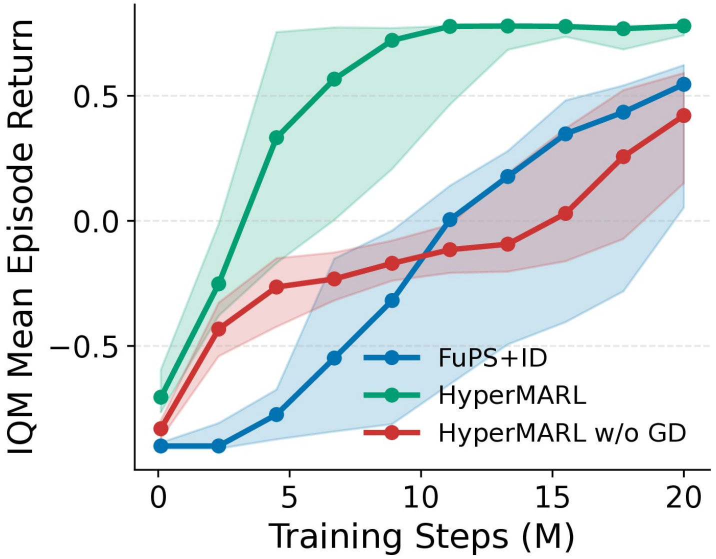


HyperMARL

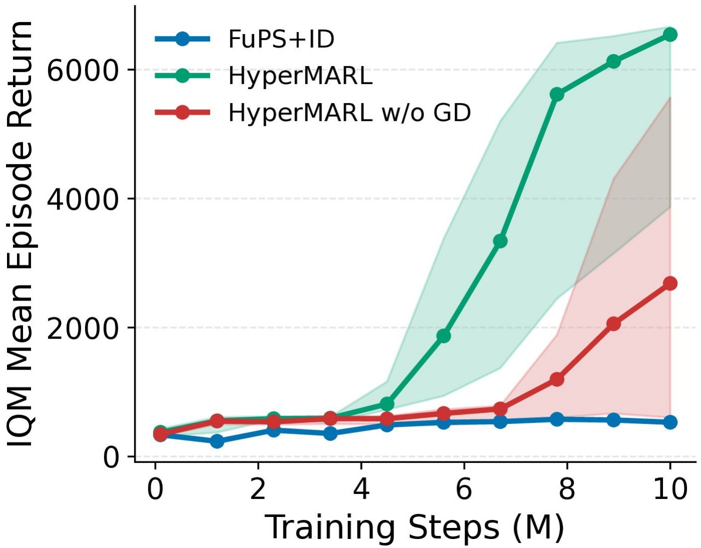


HyperMARL w/o decoupled grads

Ablations - Decoupled Grads + Initialisation



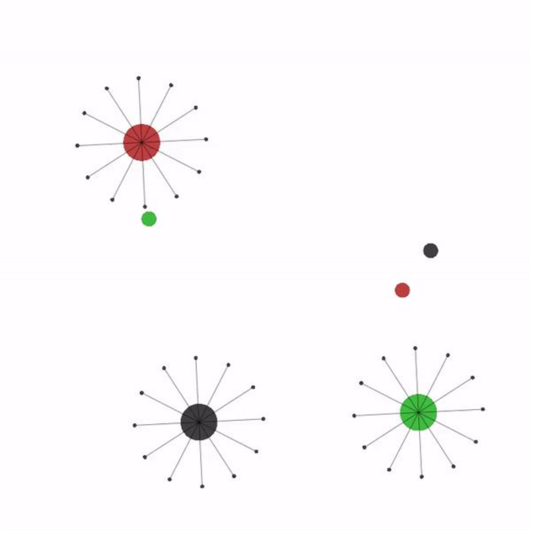
Dispersion

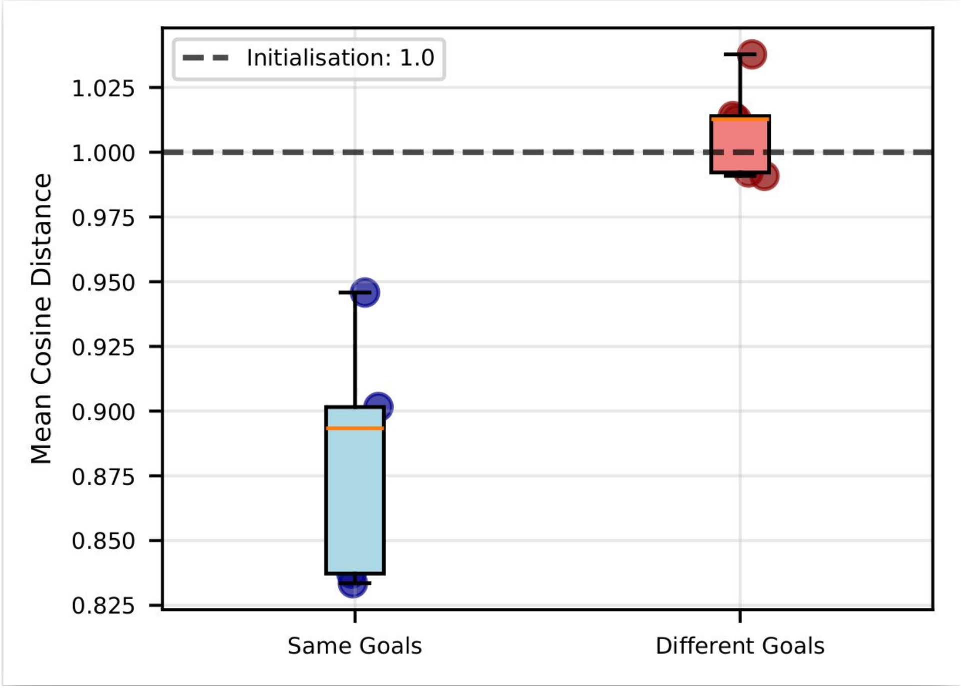


Humanoid MAMuJoCo

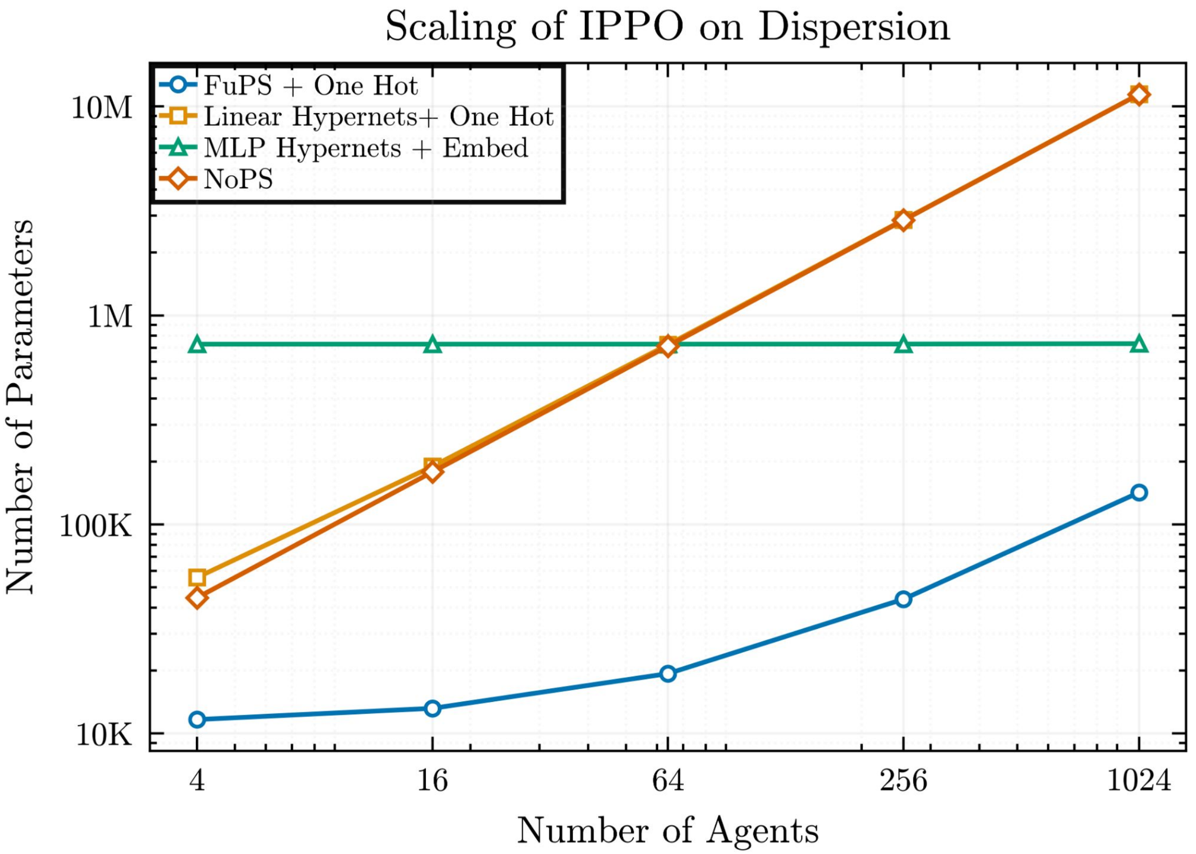
* **Gradient decoupling** and **initialisation** is important.

Ablations - Agent Embeddings - Match Goal





Challenges and Future Work - Num. of Params



*More efficient hypernet architecture like chunked hypernetworks.

* When parameter count is matched, HyperMARL > FuPS

Takeaway & Thanks

- Simple Idea:
  - 💡 ***Gradient decoupling*** is a key ingredient in mitigating cross-agent interference & *obs-ID coupling* – correlates with a lower grad variance.

- Instantiation of this gradient decoupling:
  - 💭 **HyperMARL** - Agent-conditioned Hypernetworks can achieve gradient decoupling – possible to learn adaptive behaviour, across diverse tasks, *without altering the learning objective, preset diversity levels or sequential updates.*

- Future work and challenges:
  - 🔍 Number of params -> chunked hypernets or low-rank approximations.

Thank you!

NEURAL INFORMATION PROCESSING SYSTEMS

Email: k.tessera@ed.ac.uk

Paper

Code