Ignorance is bliss: the role of noise and heterogeneity in training and deployment of:

Single Agent Policies for the Multi-Agent Persistent Surveillance Problem

Tom Kent

Collective Dynamics Seminar

30-10-19

Bio

Undergraduate
University of Edinburgh (2007-2011)
Mathematics Msc

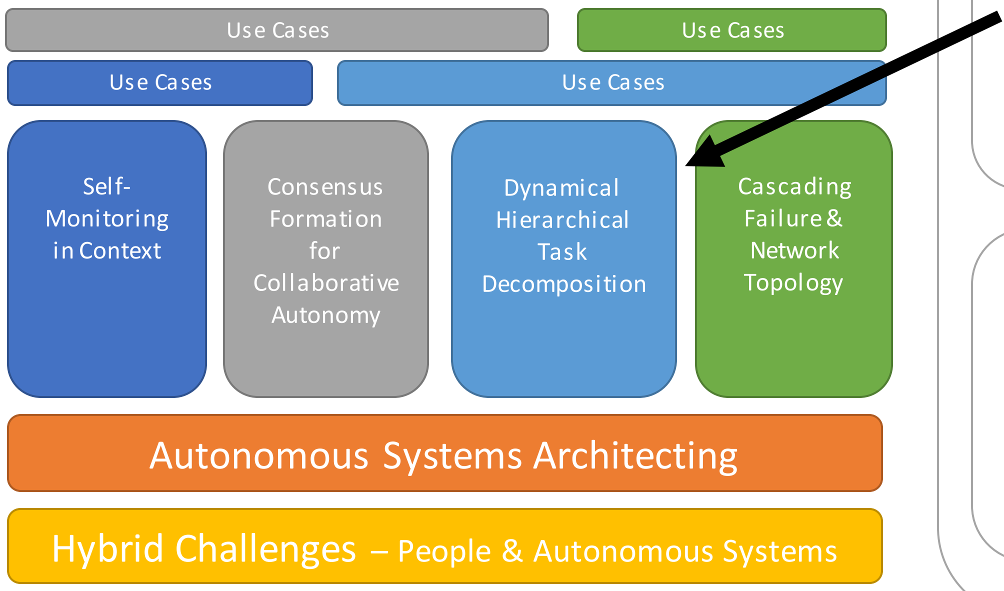
PhD
University of Bristol (2011-2015)
Aerospace Engineering
Optimal Routing and Assignment for Commercial Formation Flight

Post Doc
University of Bristol (2015-Present)
Venturer Project
Path Planning & Decision Making for Driverless Cars

T-B PHASE

T-B Partnership in Hybrid Autonomous Systems Engineering

- **Five-year project** (2017-22) fundamental autonomous system design problems
- **Hybrid Autonomous Systems Engineering** ‘R3 Challenge’:
  - **Robustness, Resilience, and Regulation**.
- Innovate **new design principles and processes**
- Build **new tools** for analysis and design
- Engaging with **real Thales use cases**:
  - Hybrid Low-Level Flight
  - Hybrid Rail Systems
  - Hybrid Search & Rescue.
- **Engaging stakeholders** within Thales
- Finding a balance between academic and industrial outputs



University of BRISTOL

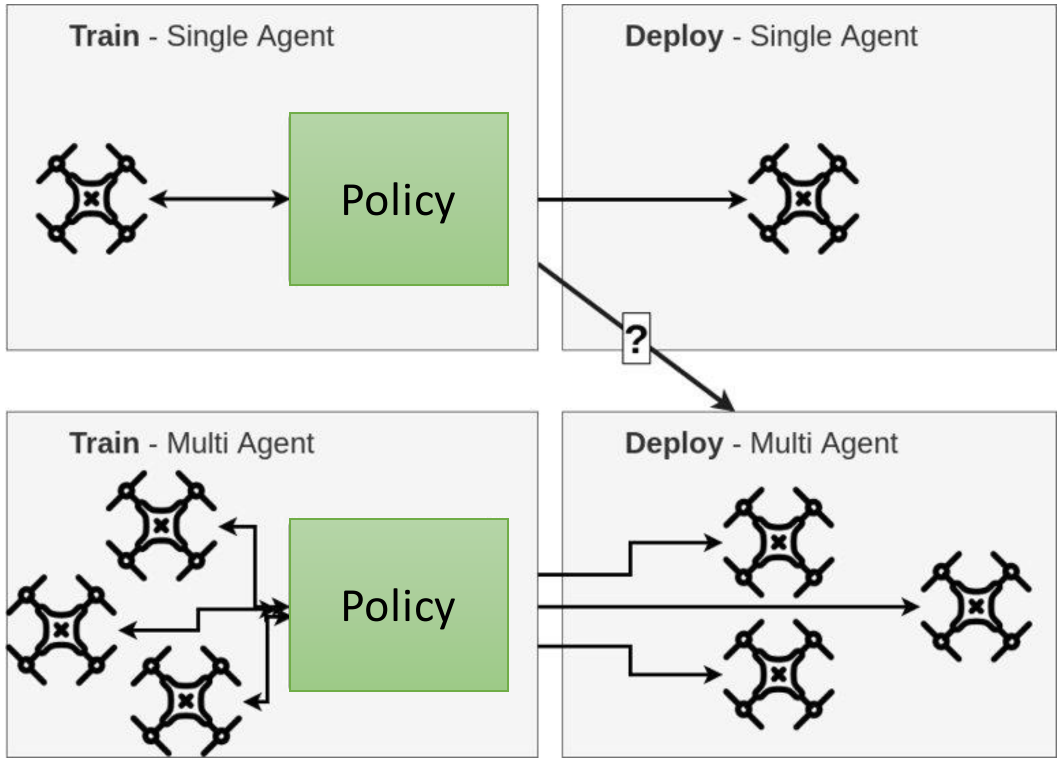
Academic PIs
Seth Bullock
Eddie Wilson
Jonathan Lawry
Arthur Richards

Post-Docs
Tom Kent
Michael Crosscombe
Debora Zanatto

PhDs
Elliot Hogg
Will Bonnell
Chris Bennett
Charles Clarke

Motivating Question

Can we train single-agent policies in isolation that can be successfully deployed in multi-agent scenarios?



- **Tricky to train/model end-to-end** for large multi-agent problems – lots of samples required

- Evaluation Loss:
  Single-Agent Environment =
  ~ (Noise, under-modelling, uncertainty)
  Multi-Agent Environment =
  ~ (Noise, under-modelling, uncertainty)^(No. Agents)
  + interactions

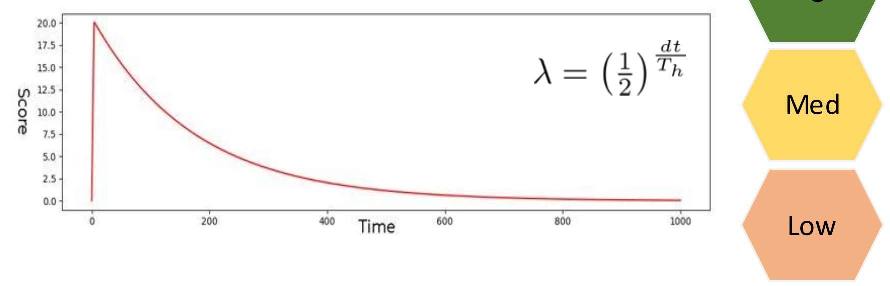
- **Enormous design-space and parameter-space**

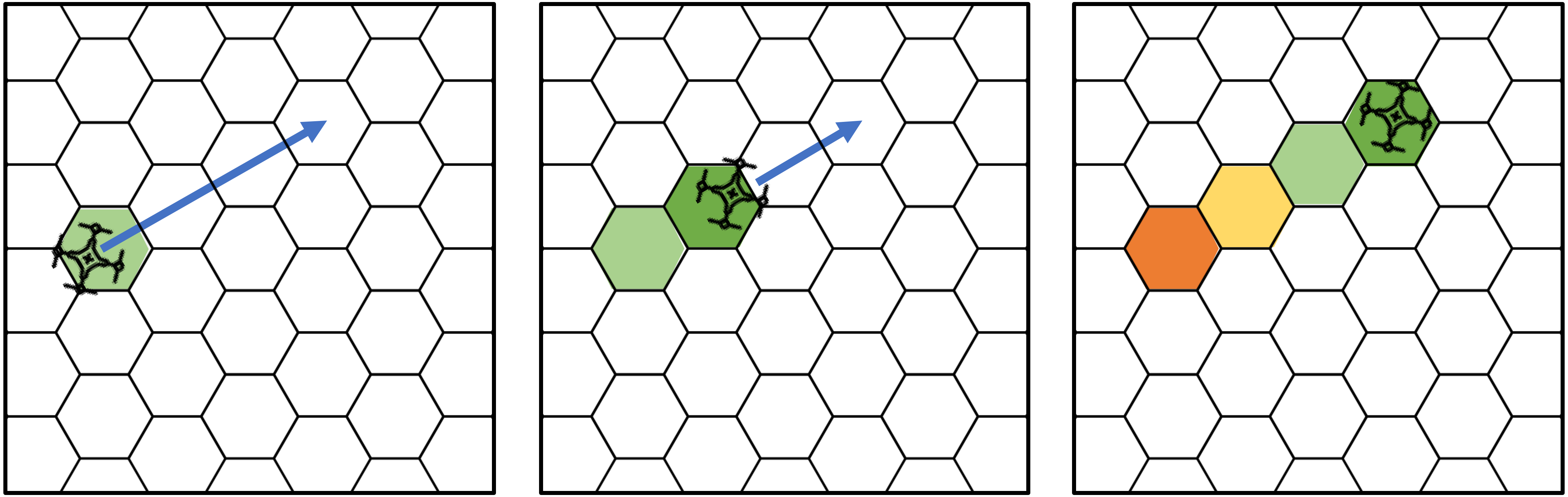
- Do we **need** to solve the entire problem at once?

Persistent Surveillance

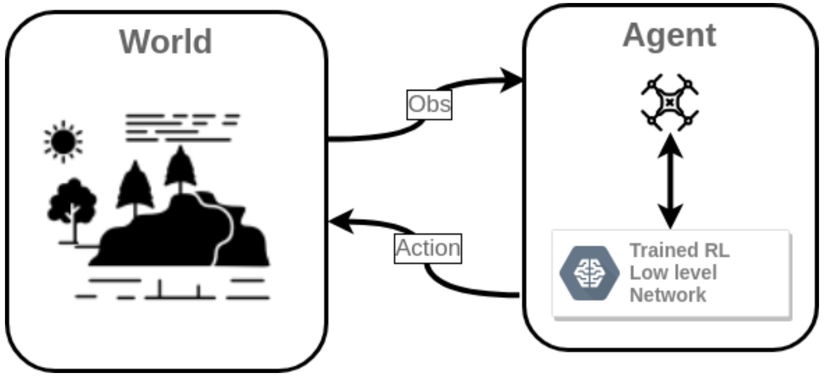
Objective: Maximise Surveillance Score (Sum of all hexes)
Method: Continuously visit hexes to increase score
Hex score: Increases quickly then decays

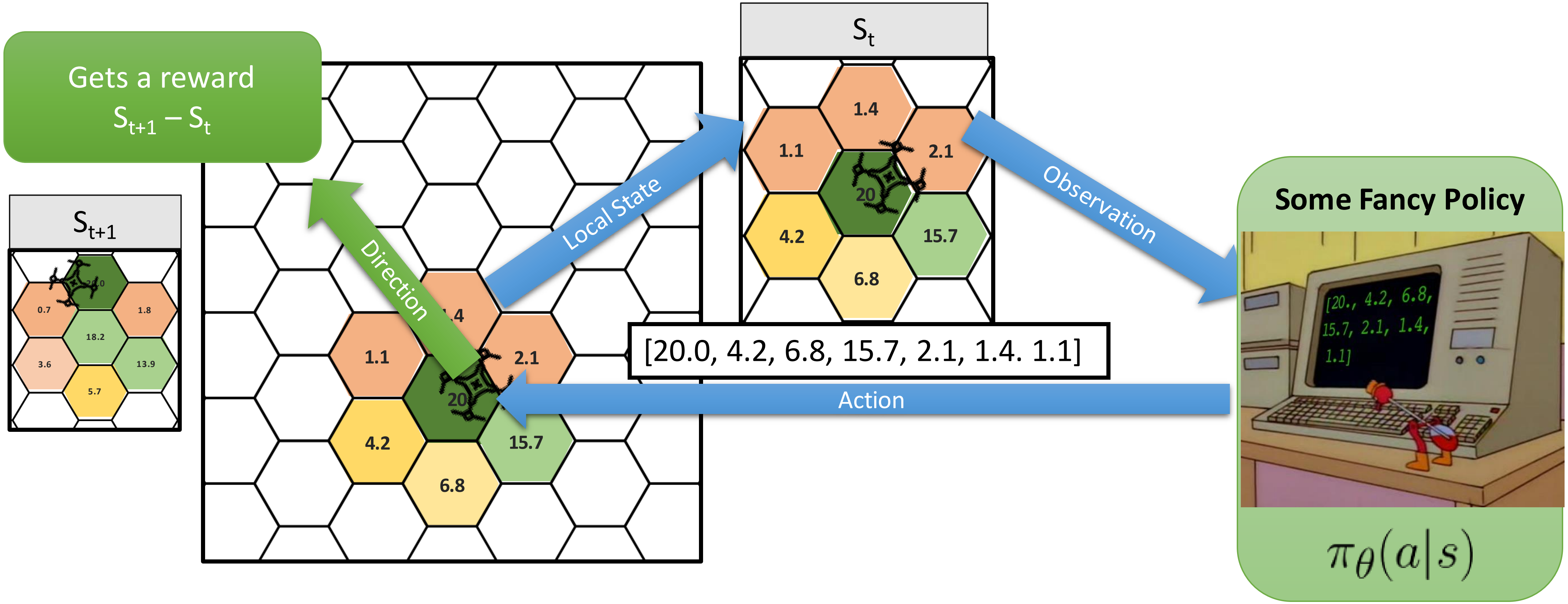
Hex Score





Local Policies



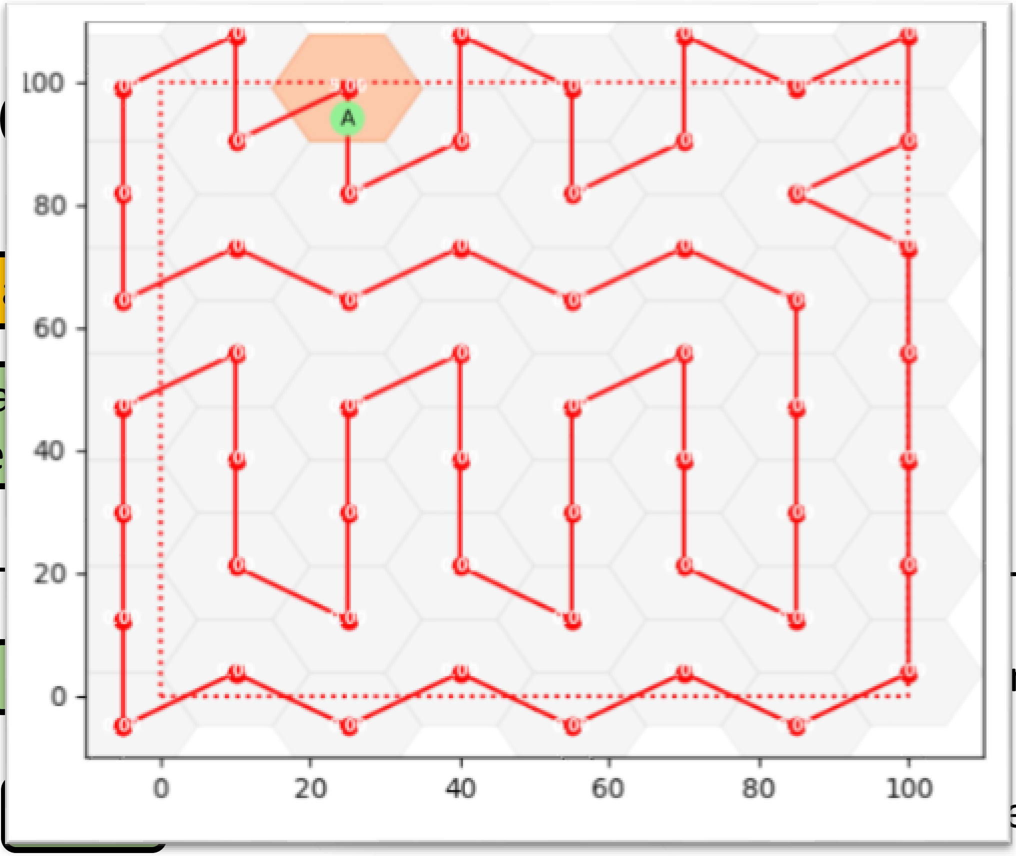


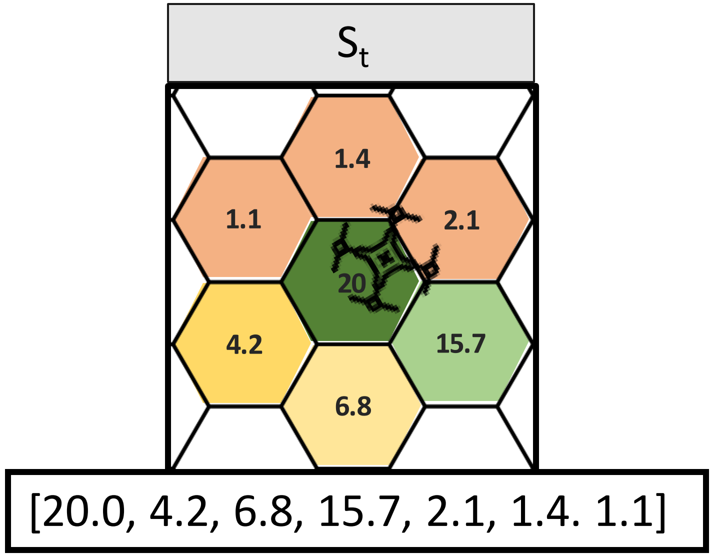
L

Heuristics

Ra

Gra
De



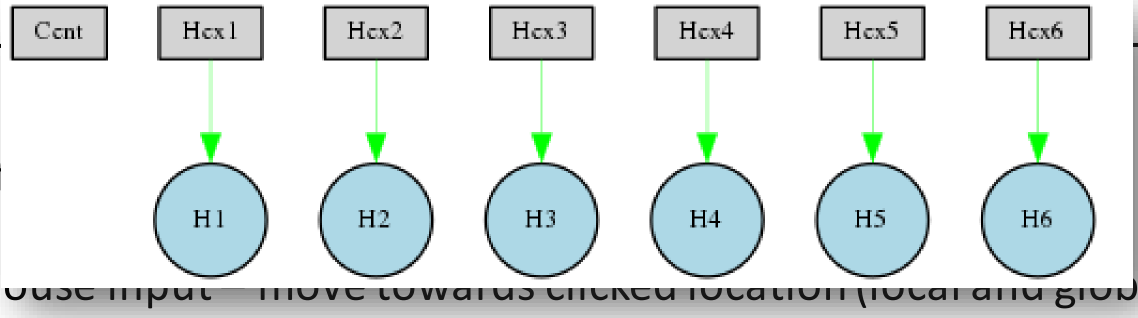


'AI'

ned neural net – Deterministic policy

es – hand crafted approximates gradient descent

Benchmarks



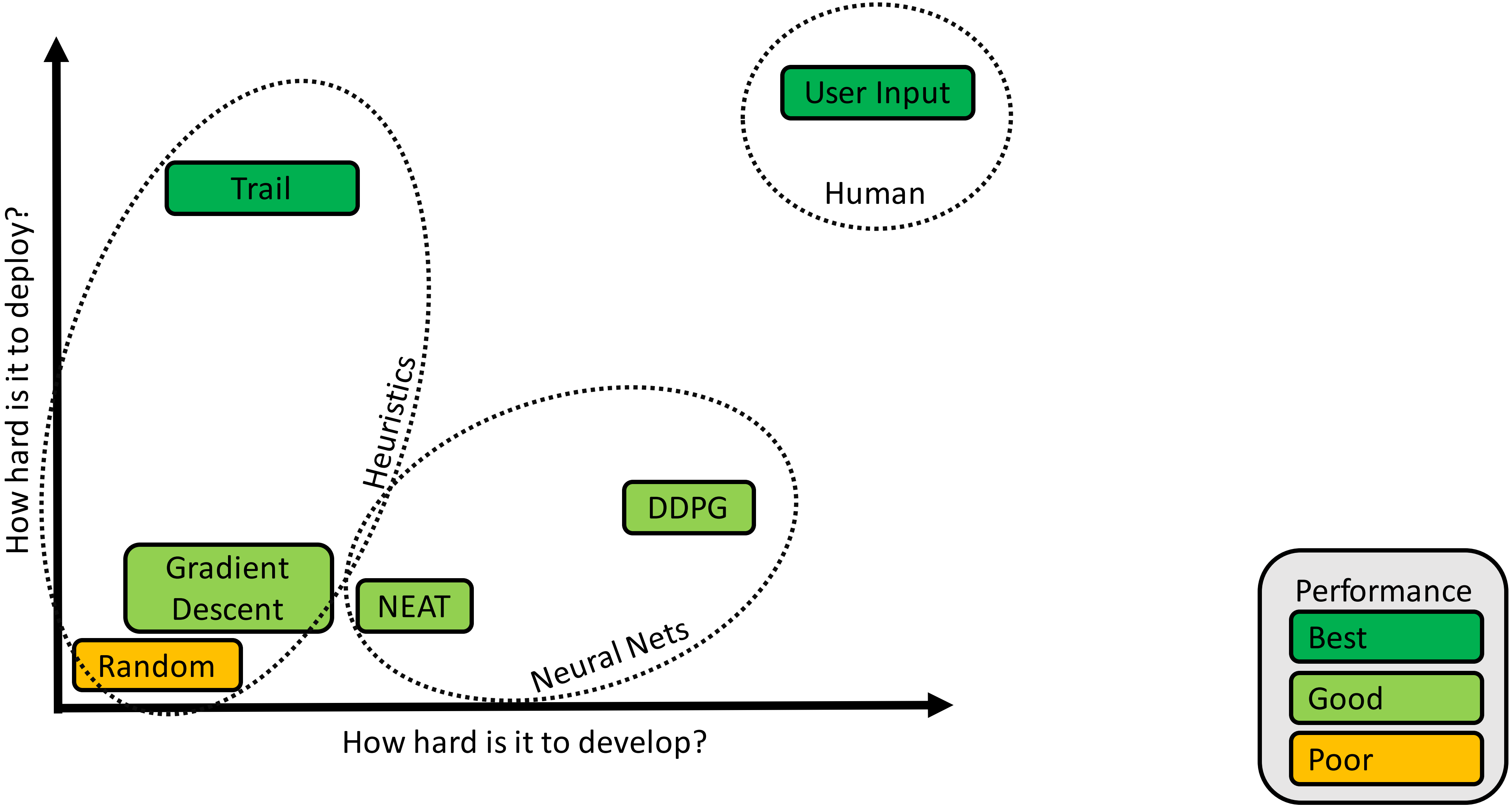
Trail — Pre-def ng in a loop

User Input — User Mouse input – move towards clicked location (local and global version)

Performance

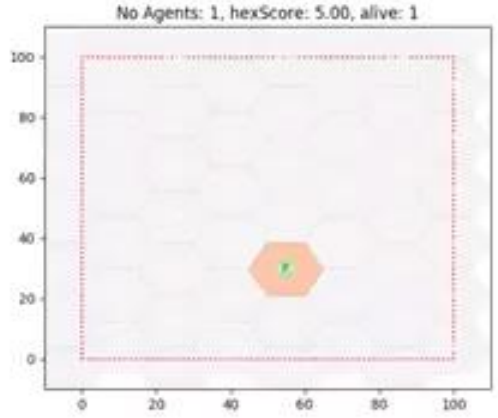
- Best
- Good
- Poor

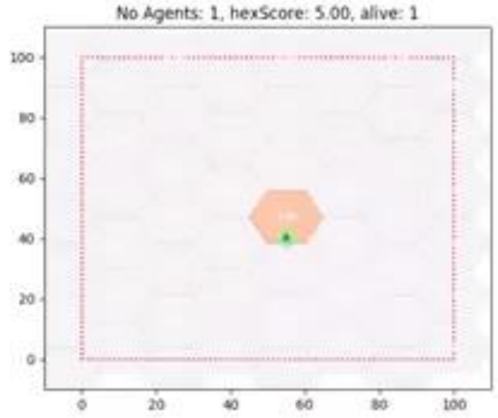
Comparison of Local Policies



Comparison of Local Policies

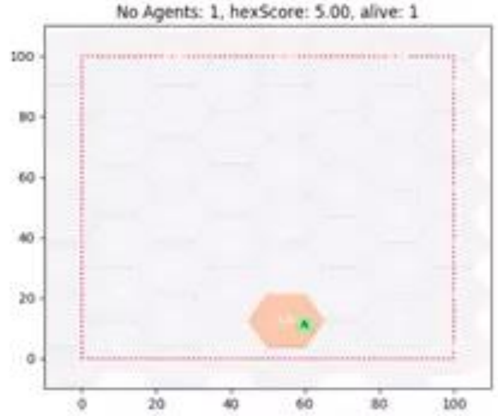
Random

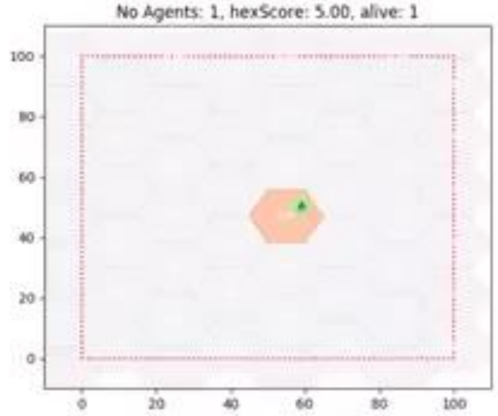




Gradient Descent

DDPG



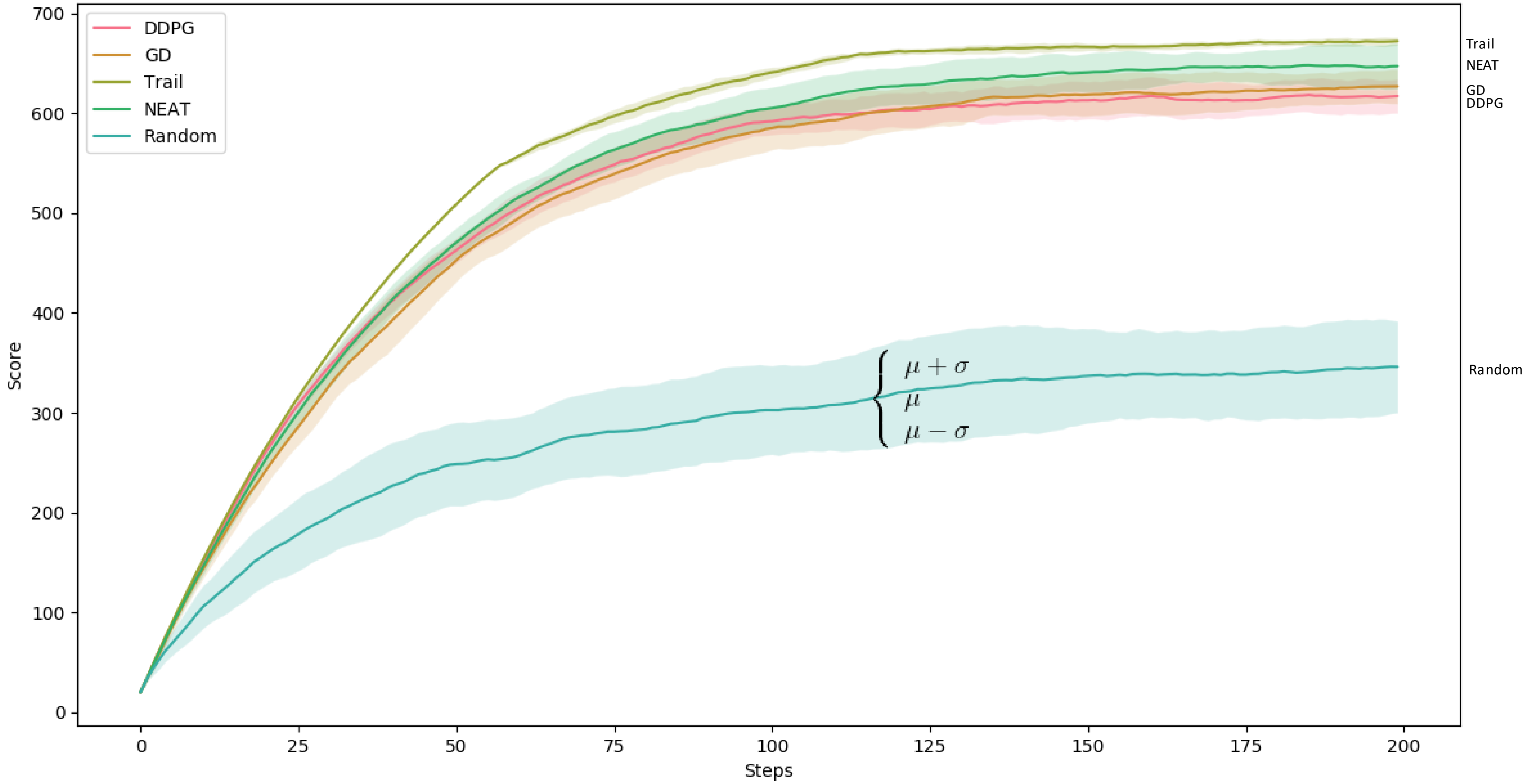


Trail

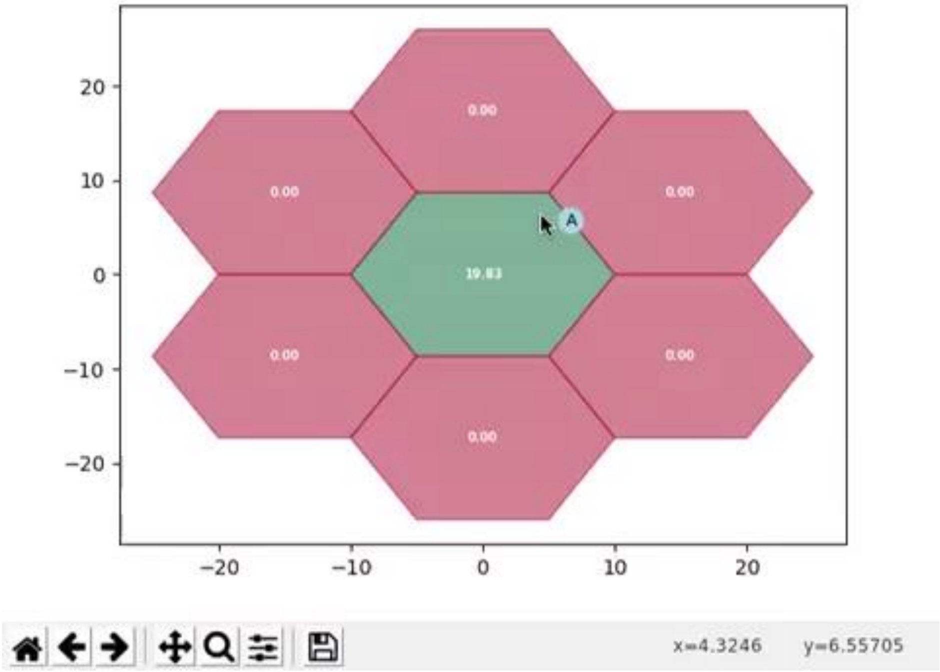
Performance

- Best
- Good
- Poor

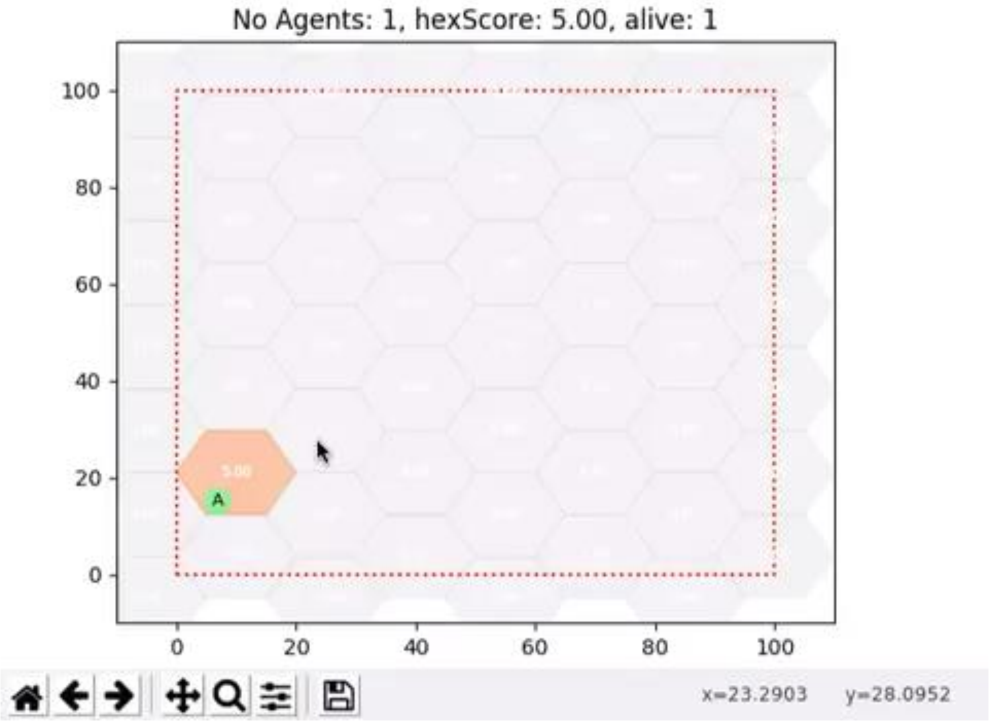
Policy Performance – 1 Agent



Human input (aka graduate descent)

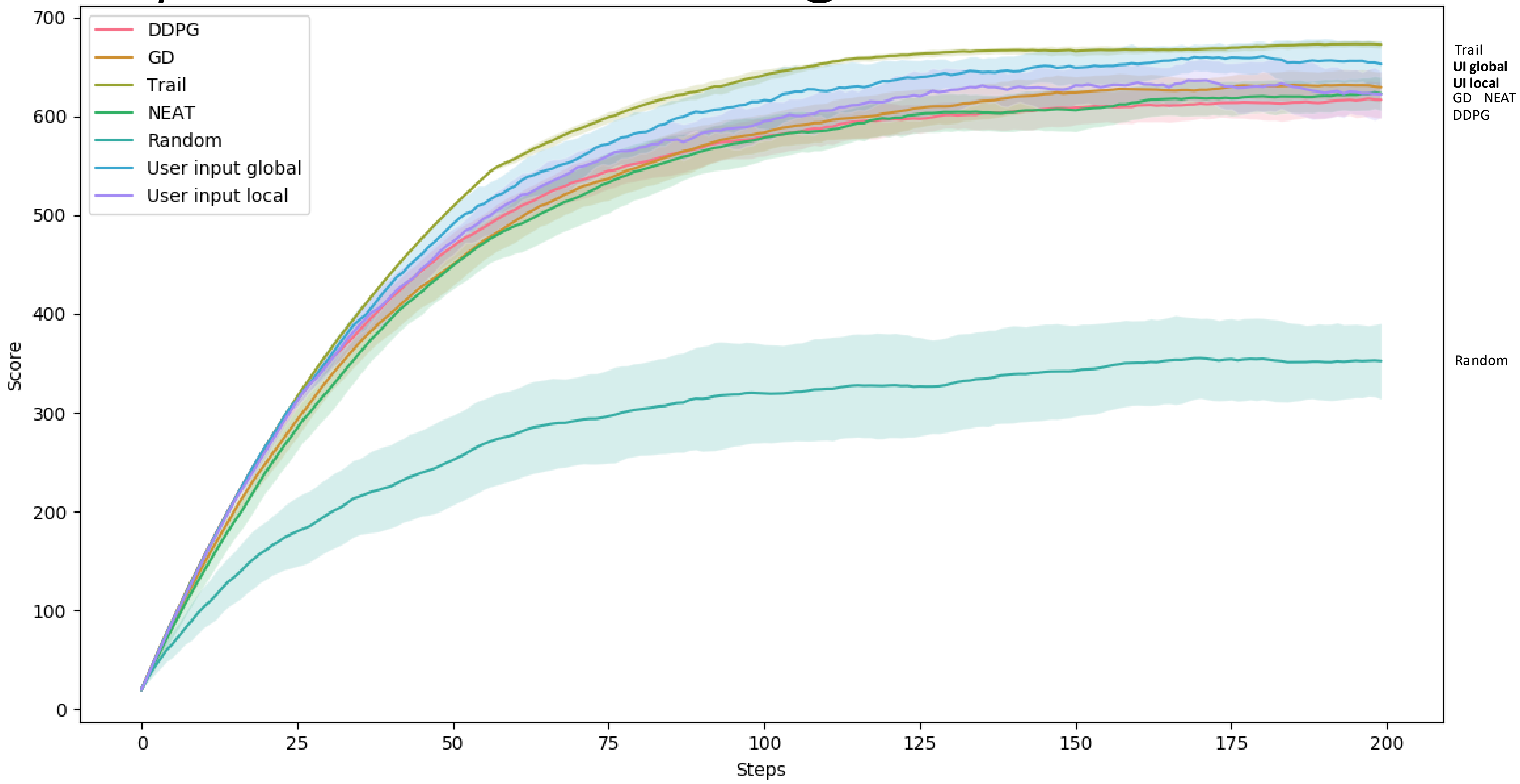


Local view
- User clicks hex
- Agent moves in direction of cursor
- Attempt to build global picture & localise
- Users tend to do gradient descent



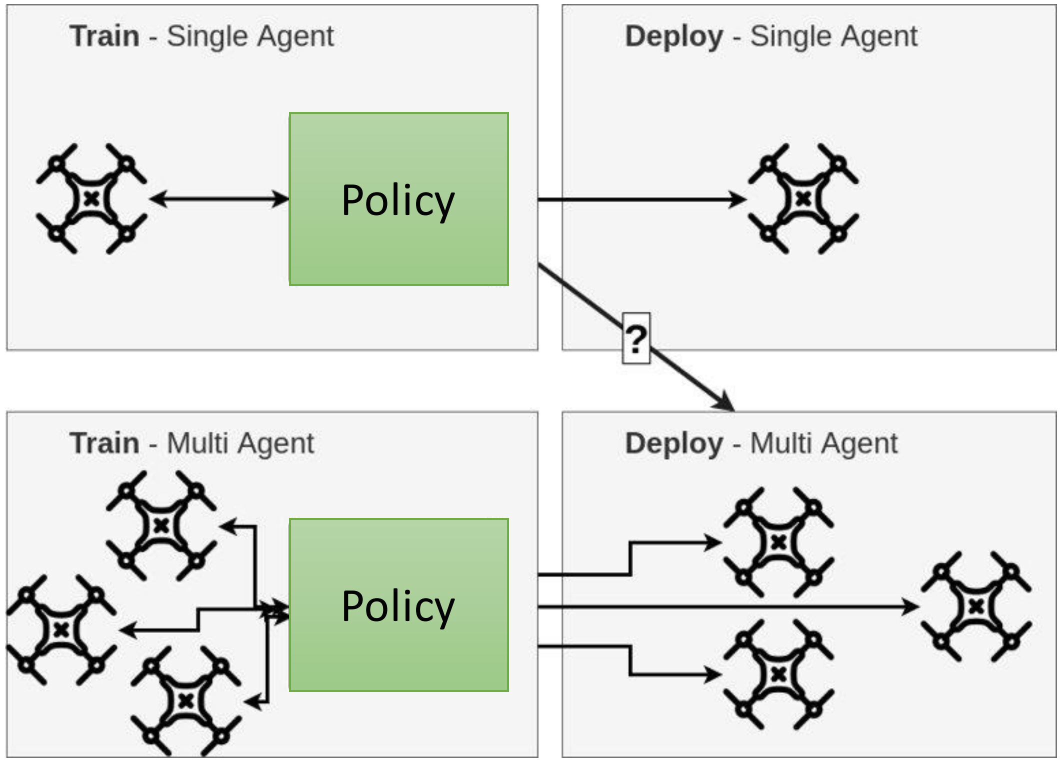
Global view
- User clicks hex
- Agent moves in direction of cursor
- Can more easily plan ahead
- Users tend to attempt a trail

Policy Performance – 1 Agent



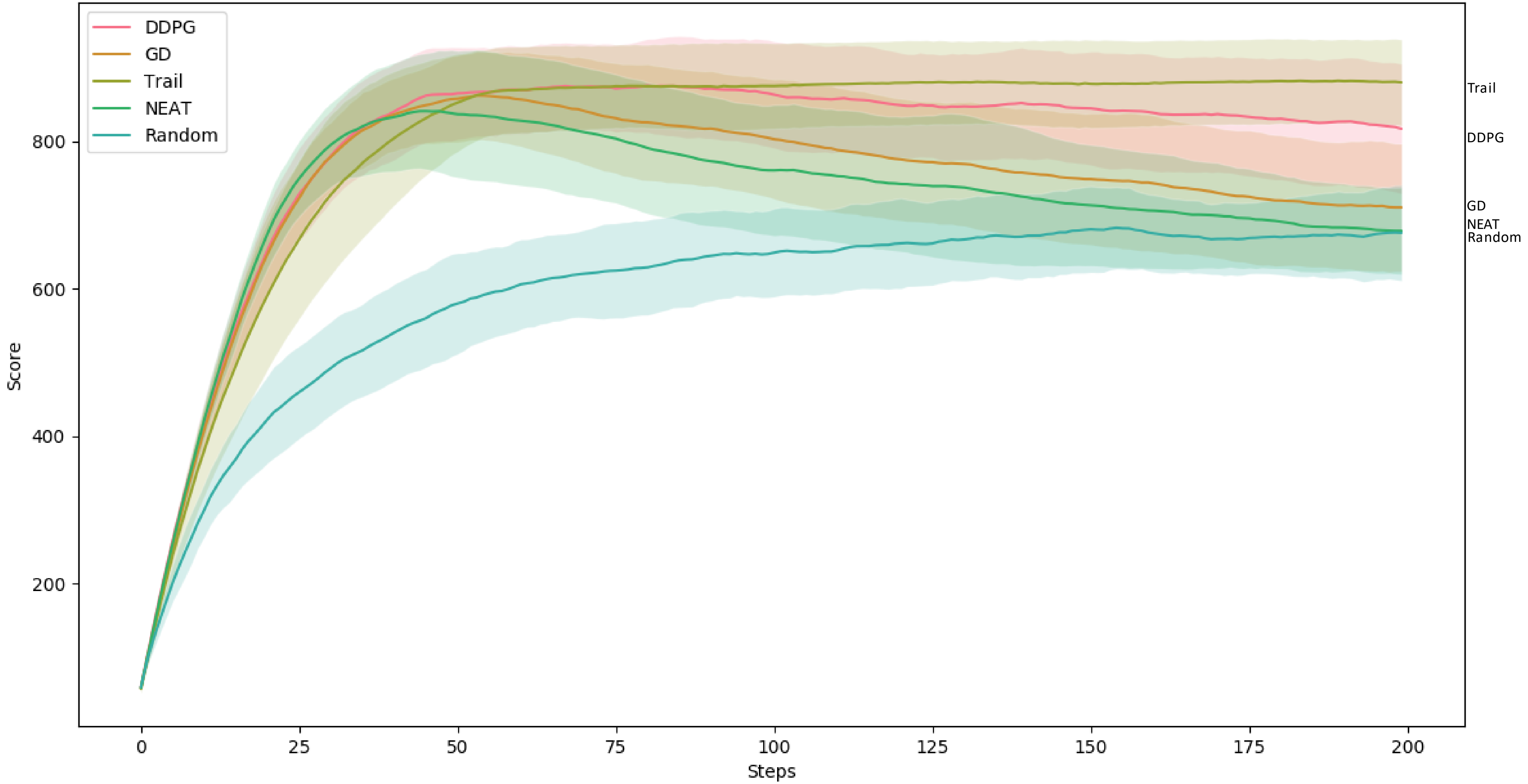
Multiple agents

Can we train single-agent policies in isolation that can be successfully deployed in multi-agent scenarios?

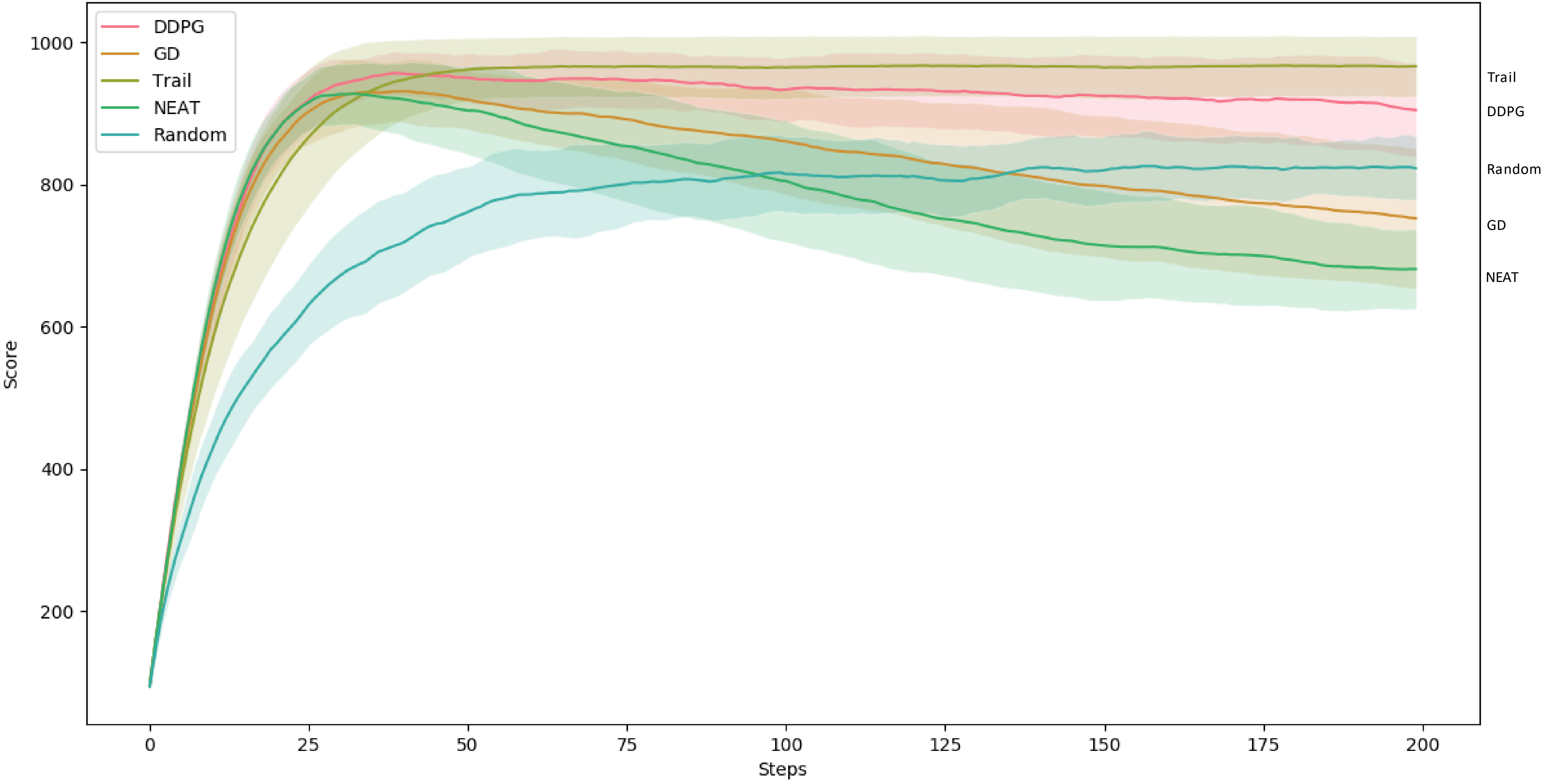


- All Agents have identical policies
- Agents all have perfect global state knowledge
- Agents observe their local state and decide action
- Agents then all move simultaneously
- No communications
- No cooperation or planning for other agents
- Other agents appear as 'obstacles'

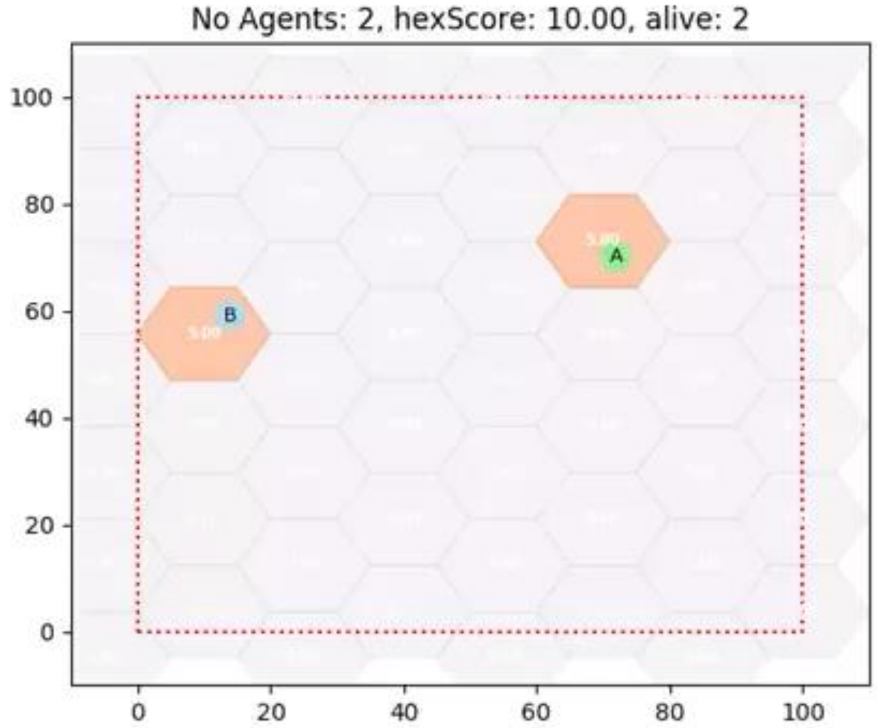
Policy Performance – 3 Agents



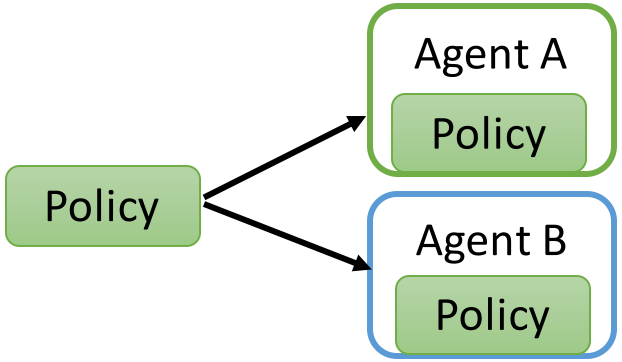
Policy Performance – 5 Agents

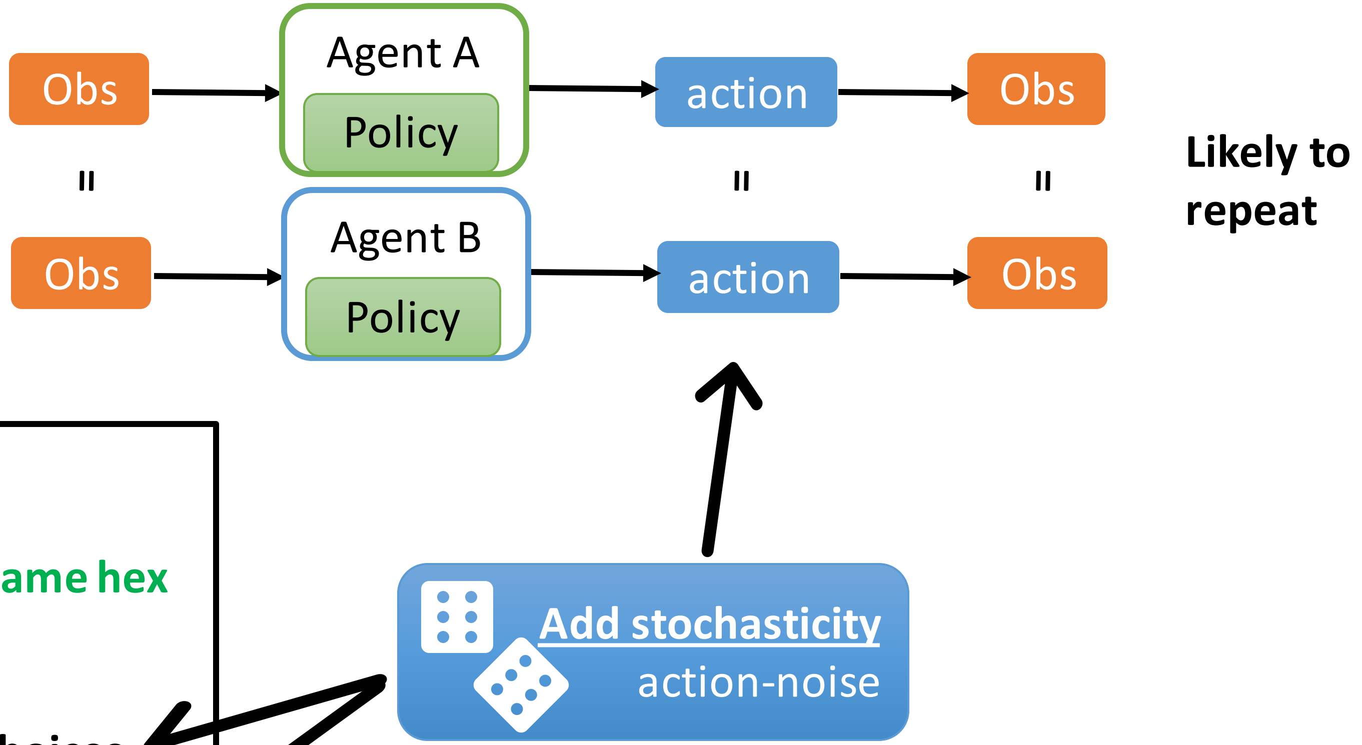


Homogeneous-policy convergence problem



Homogeneous-policy convergence problem



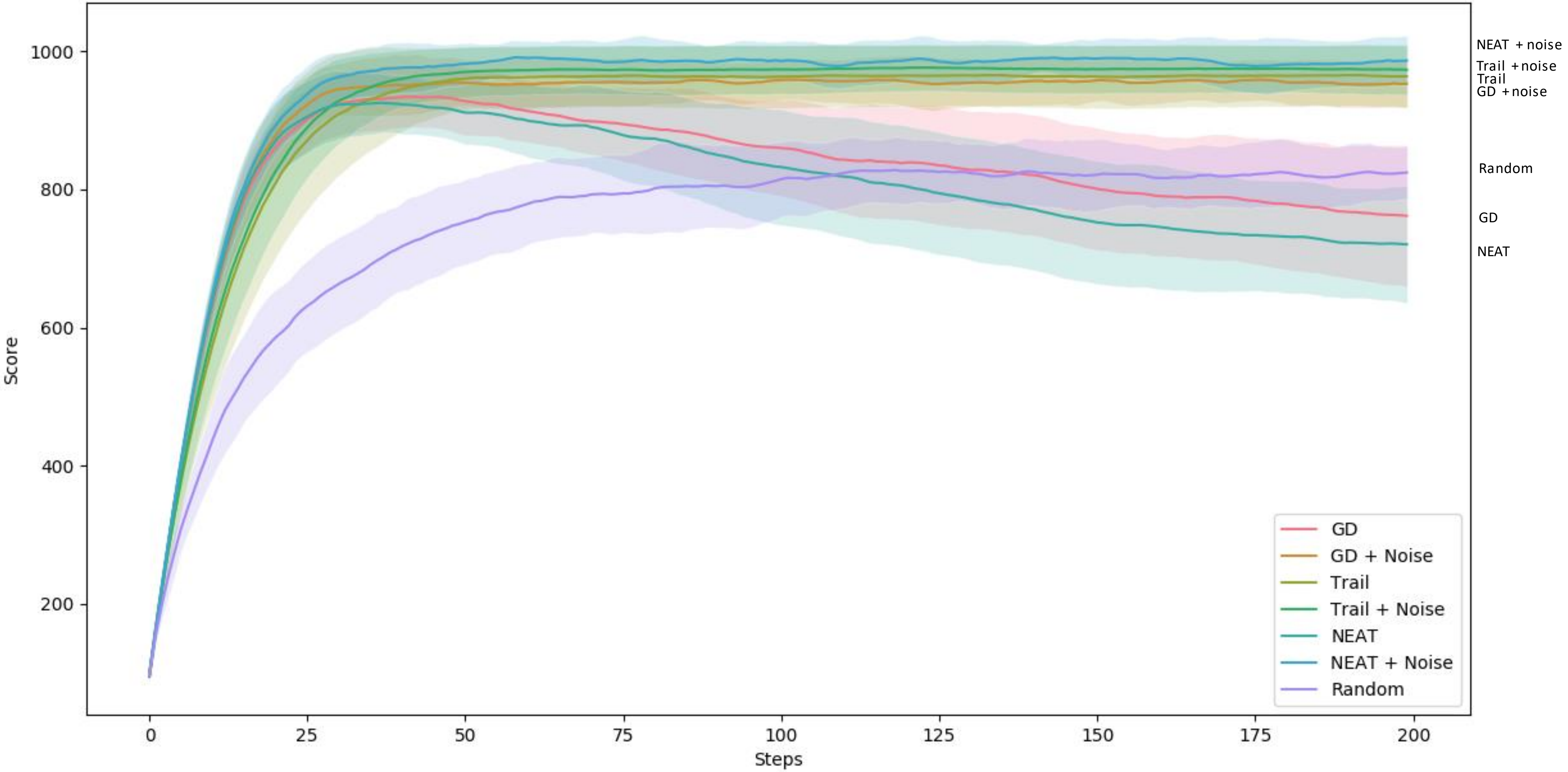


<u>The convergence cycle</u>

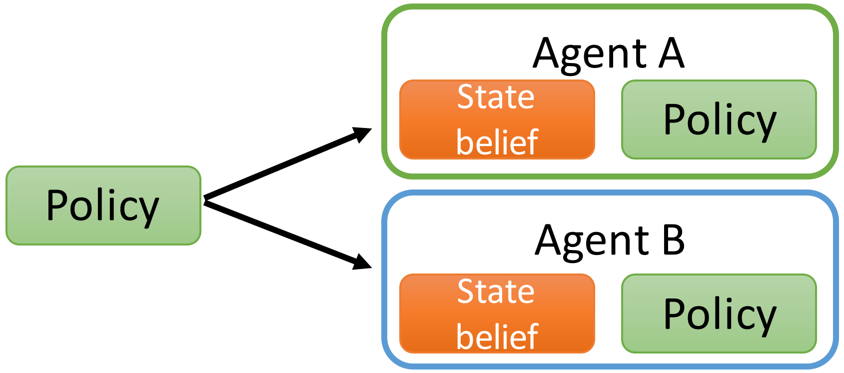
1) Agents move into the **same hex**
   - **Cooperate to stop agents occupying the same hex**
2) Get an **identical state observation**
   - **Have differing state beliefs**
3) Identical policies returns **identical action choices**
   - **Make policies non-deterministic**
4) Identical actions lead to **high chance of repeating 1)**
   - **Have agents take turns**

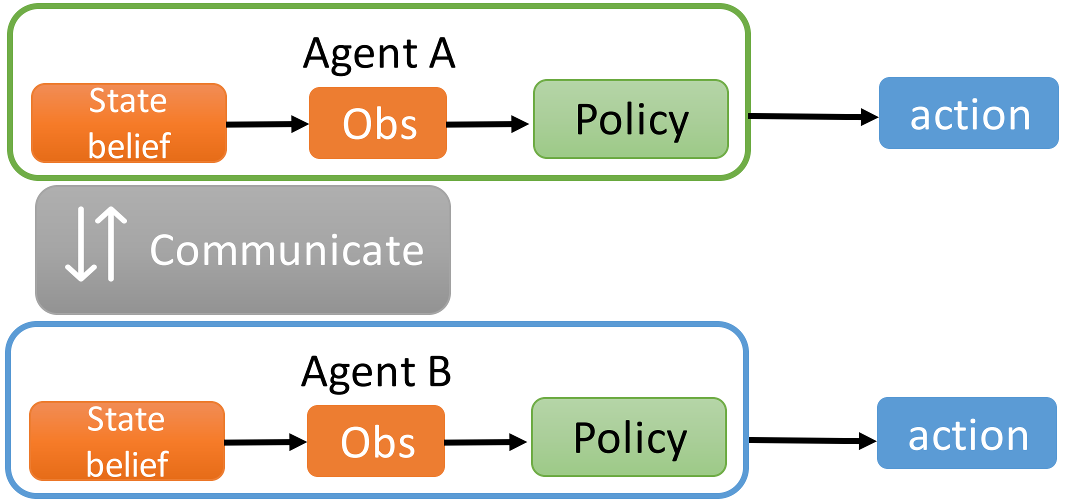
We can break this cycle at any of these points!

Policy Performance & action noise - 5 agents



Decentralised State

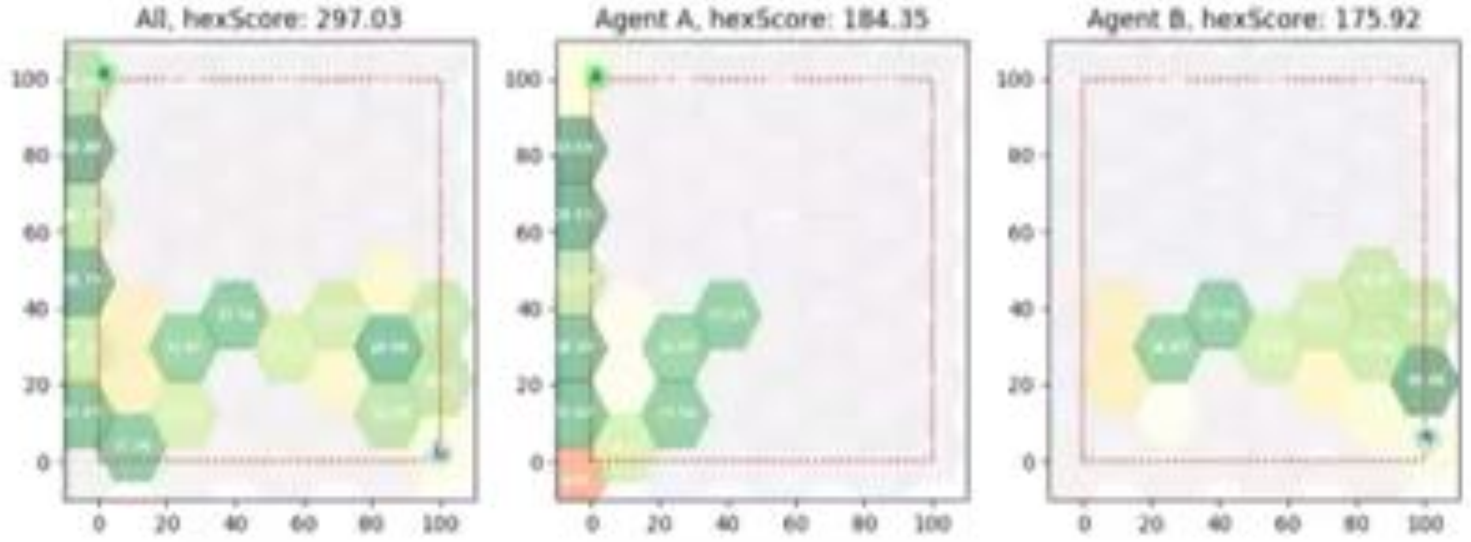




<u>The convergence cycle</u>

1) Agents move into the **same hex**
2) Get an **identical state observation**
3) Identical policies returns **identical action choices**
4) Identical actions lead to **high chance of repeating 1)**

<u>Add stochasticity</u>
individual state beliefs
Comms for state consensus



Centralised State

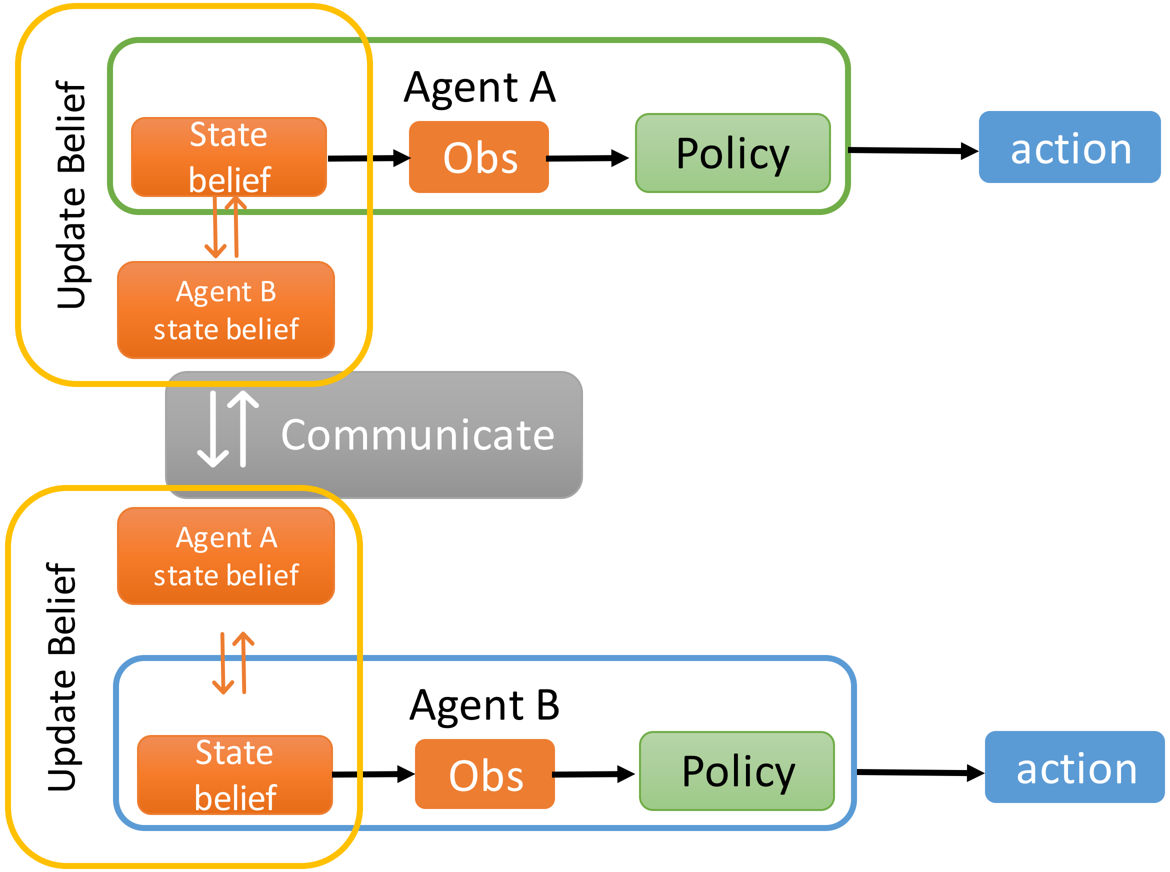
Local States

Belief Updating

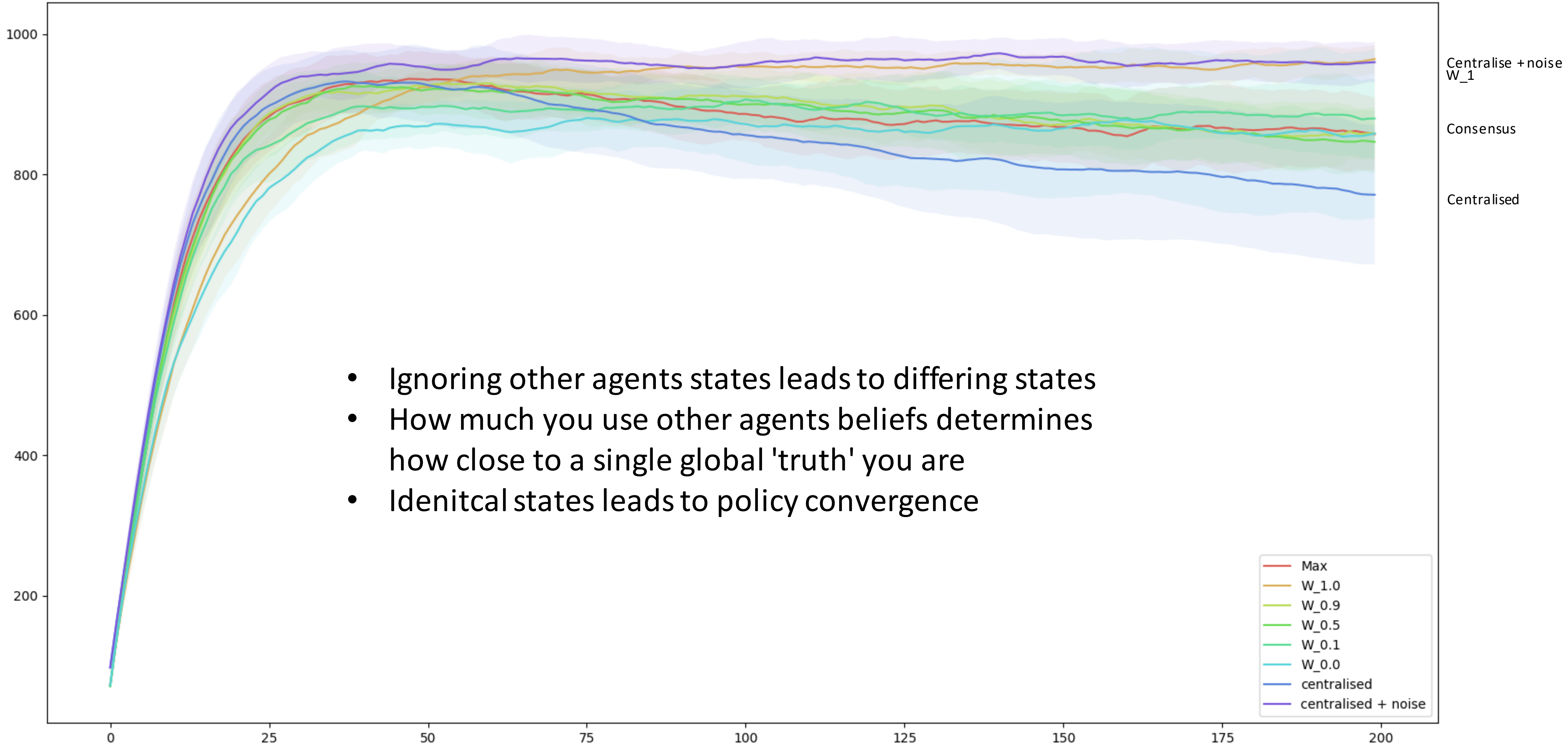
- Agents communicate ***their*** state-belief
- Agents update their belief to form global '*true*' state
- How should agents incorporate these other agents' beliefs?

<u>Update functions</u>

1) **Max:**
   The max value of own and other's beliefs
2) **Average:**
   Average of own belief and other agents' beliefs
3) **Weighted Average:**
   Proportionally weight own belief and others
   1) **W_0.9 -> 0.9*(own belief) + 0.1*(others)**
   2) **W_1.0 -> 1.0*(own belief)**
   3) **W_0.0 -> 1.0*(others belief)**

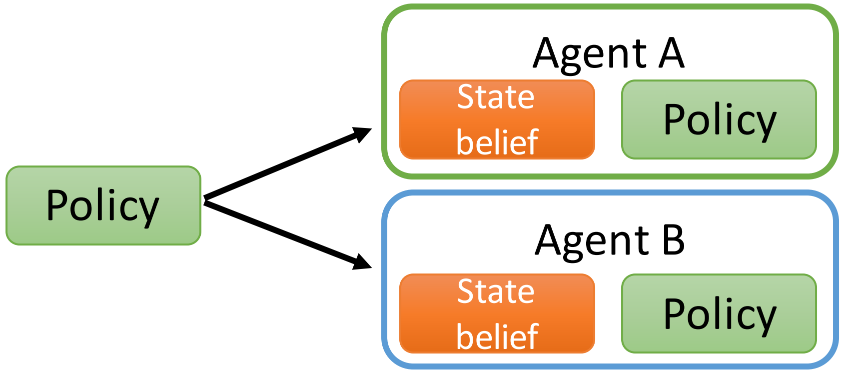


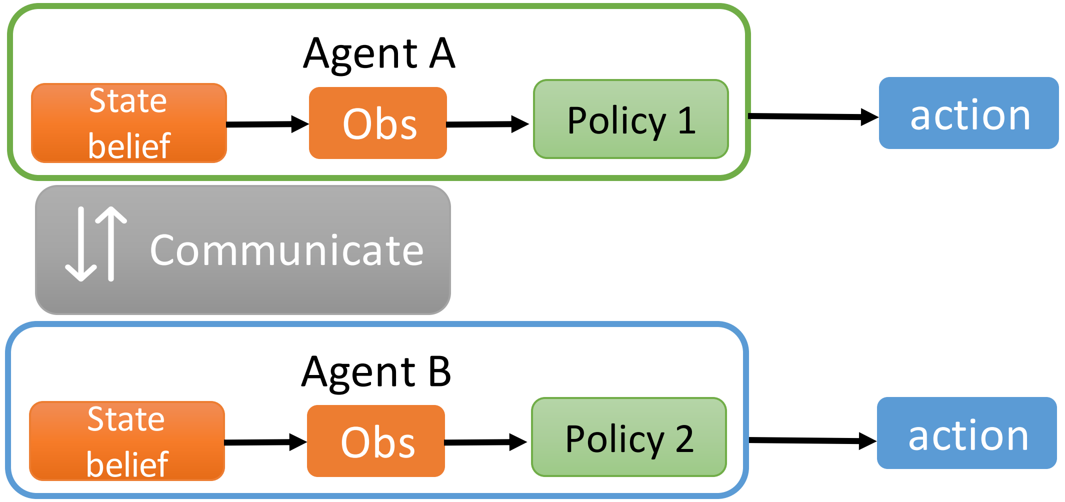
State belief Consensus results



- Ignoring other agents states leads to differing states
- How much you use other agents beliefs determines how close to a single global 'truth' you are
- Idenitcal states leads to policy convergence

Decentralised State Heterogeneous Policies





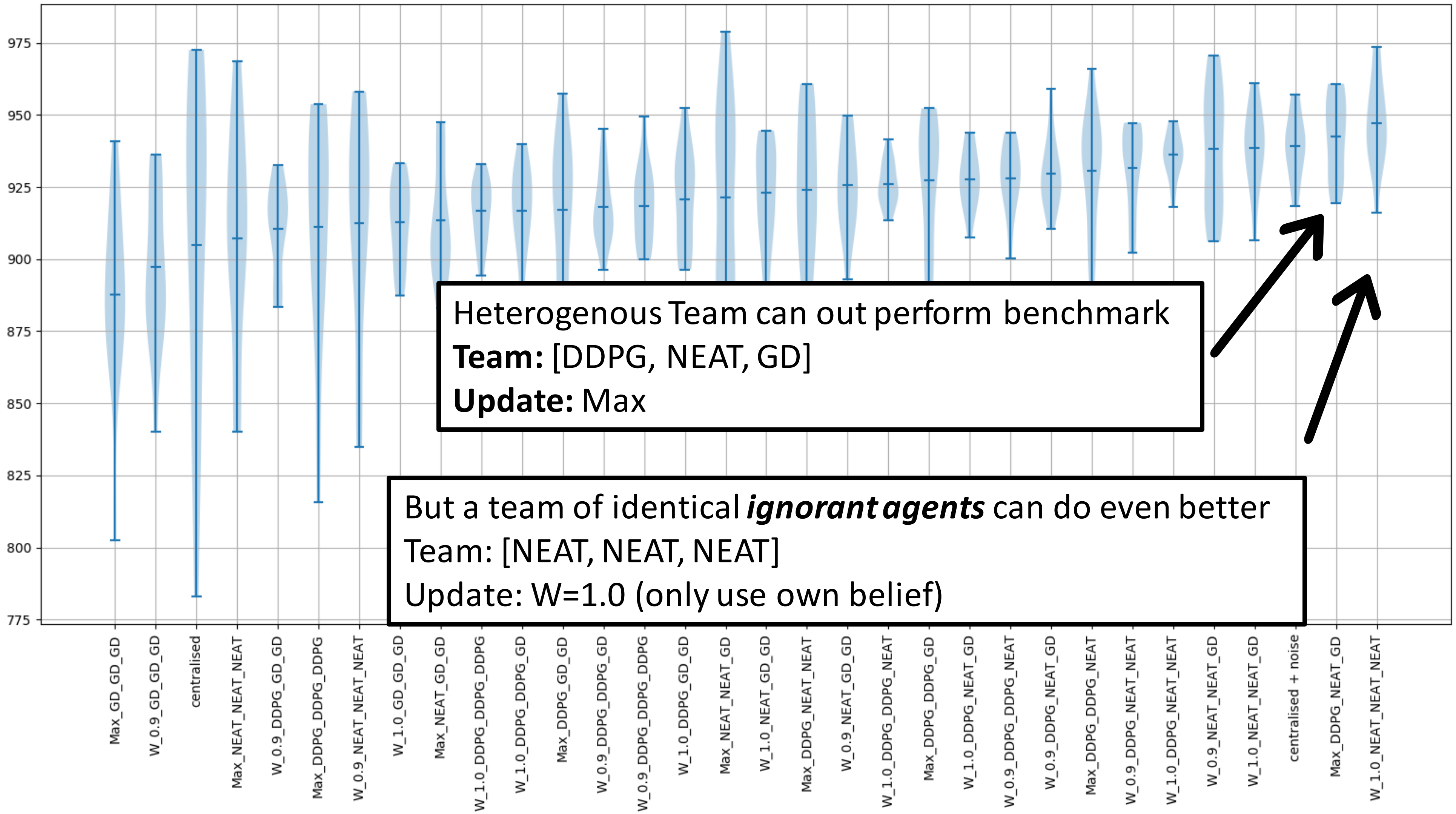
The convergence cycle

1) Agents move into the **same hex**
2) Get an **identical state observation**
3) Identical policies returns **identical action choices**
4) Identical actions lead to **high chance of repeating 1)**

Add stochasticity
individual state beliefs
Comms for state consensus

Heterogeneous Teams
Different agent policies

Decentralised State Heterogeneous Policies



Team Size
3

Policies
Gradient Descent
DDPG
NEAT

Belief Update
Max
W = 1.0
W = 0.9

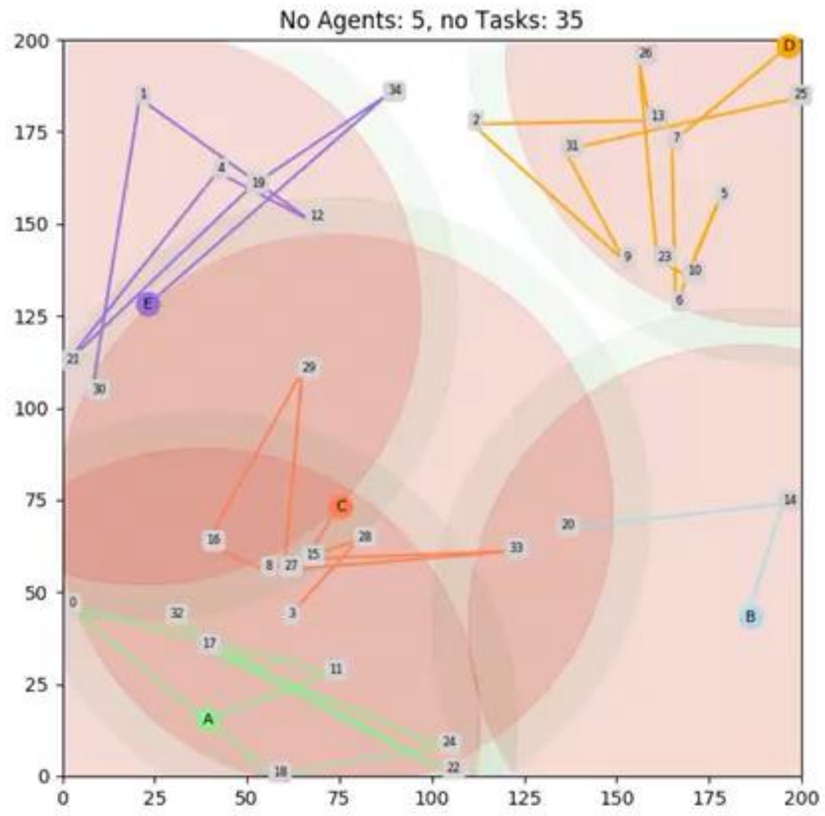
Benchmark
Centralised + action noise
Centralised

Local Policies: Take away

- The multi-agent persistent surveillance problem is somewhat simplistic
  - Short-term planning is often sufficient
- **Agents trained in isolation can still perform in a multi-agent scenario**
  - Global 'trail' policies perfom better
  - Simplistic gradient descent approaches perform pretty well
- **Homogeneous-policy convergence cycle is a problem** and can be avoided by essentially becoming more heterogeneous
  - **Action stochasticity** – adding noise
  - **State/observation stochasticity** – agent specific state beliefs
  - **Heterogenous policies** – teams of different agents
- Decentralised case with agents having partial knowledge can be benificial
- Different methods of state consensus indicate that communication, that is being closer to the *global truth,* can be detrimental to performance

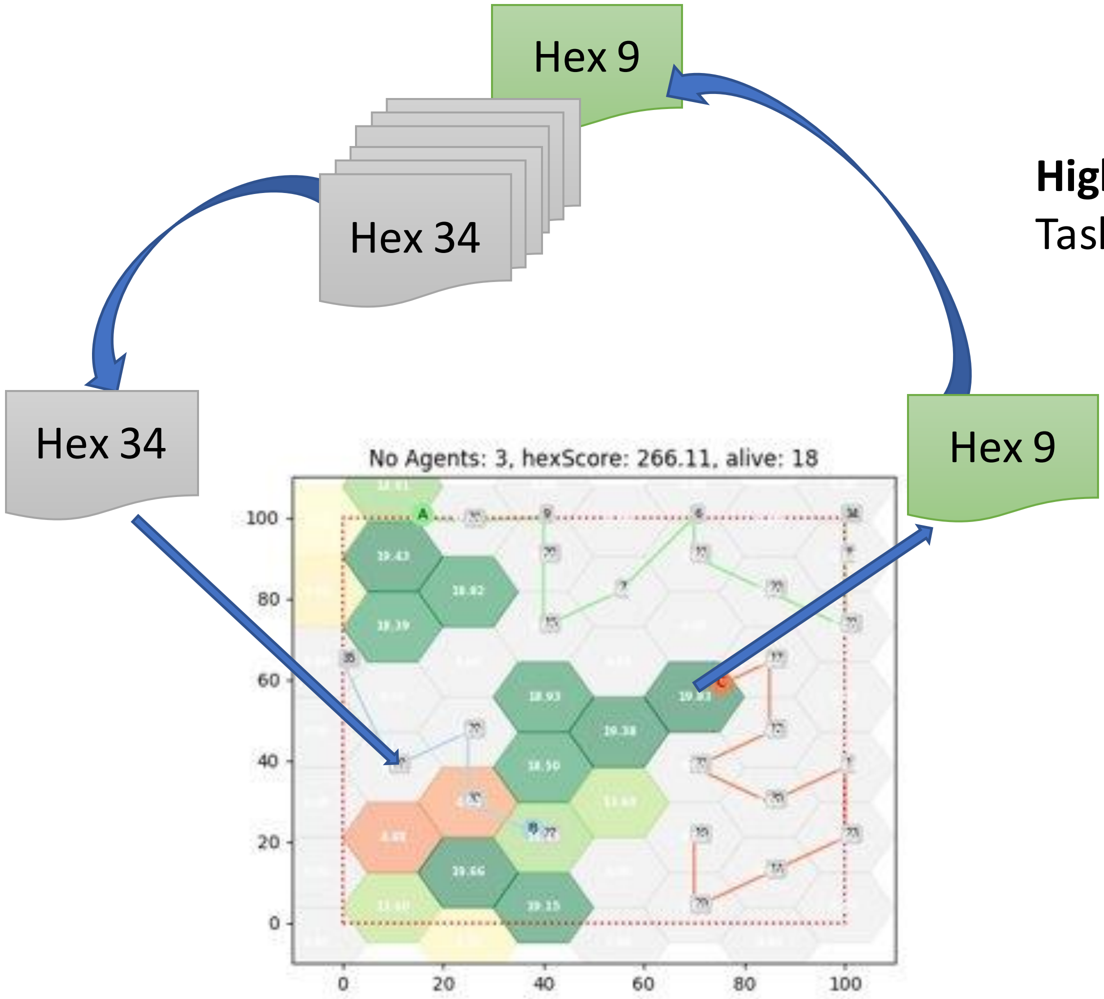
Higher Level Decisions

- What if we moved up the decision making hierarchy?
- Previous work [1]:
  Decentralised Co-Evoultionary Algorithm to solve decentralised **Multi-Agent Travelling Salesman (DEA)**
- **Make Persistent surveillance a higher-level goal**
  - the agents do not consider it
- What if we instead **place tasks in order to maximise the surveillance score**?
- MATSP and shortest path problems lead to essentially **decentralised trails**

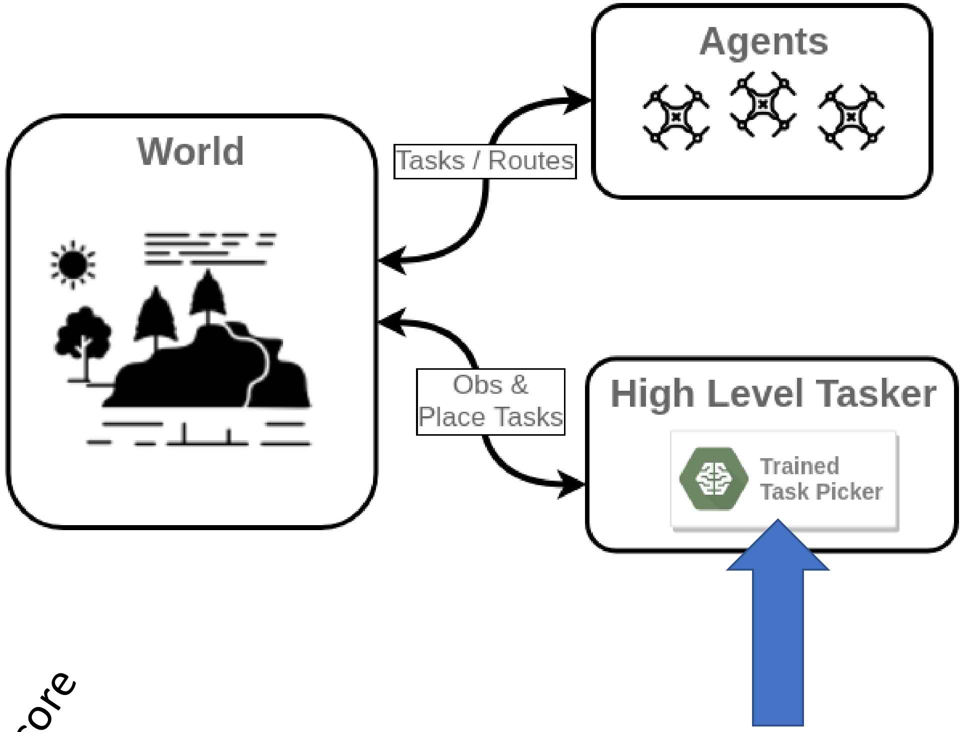


[1] Thomas E. Kent and Arthur G. Richards. "**Decentralised multi-demic evolutionary approach to the dynamic multi-agent travelling salesman problem**". In: Proceedings of the Genetic and Evolutionary Computation Conference Companion on - GECCO '19. doi: 10.1145/3319619.3321993

Combining Persistent Surveillance and MATSP

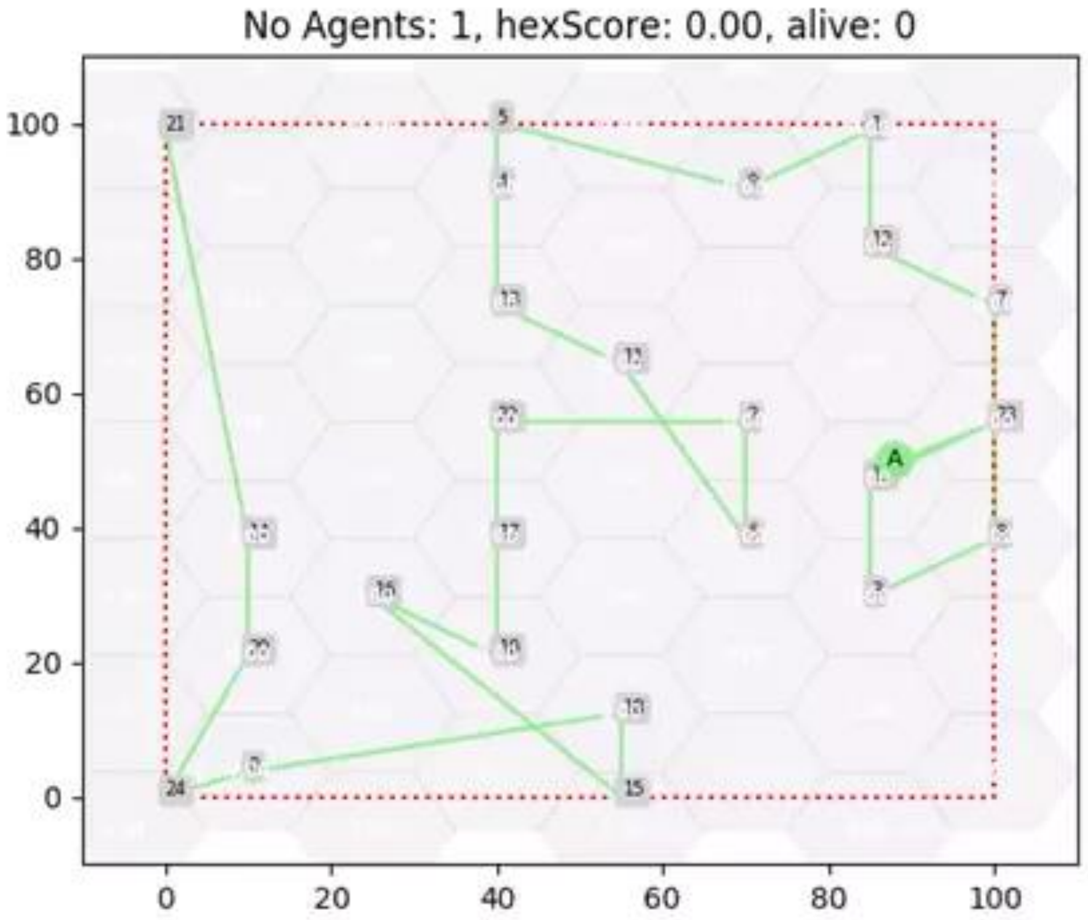


High Level
Tasking

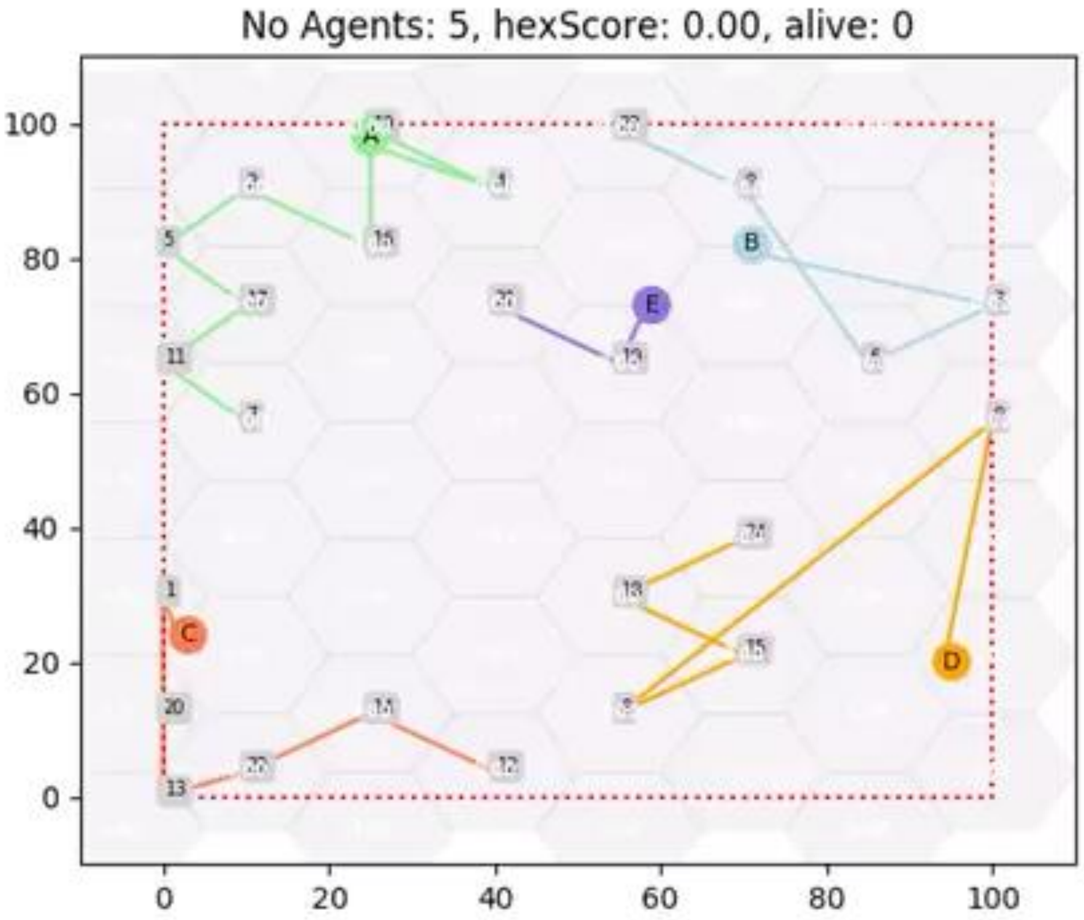


Persistent
Surveillance Tasker

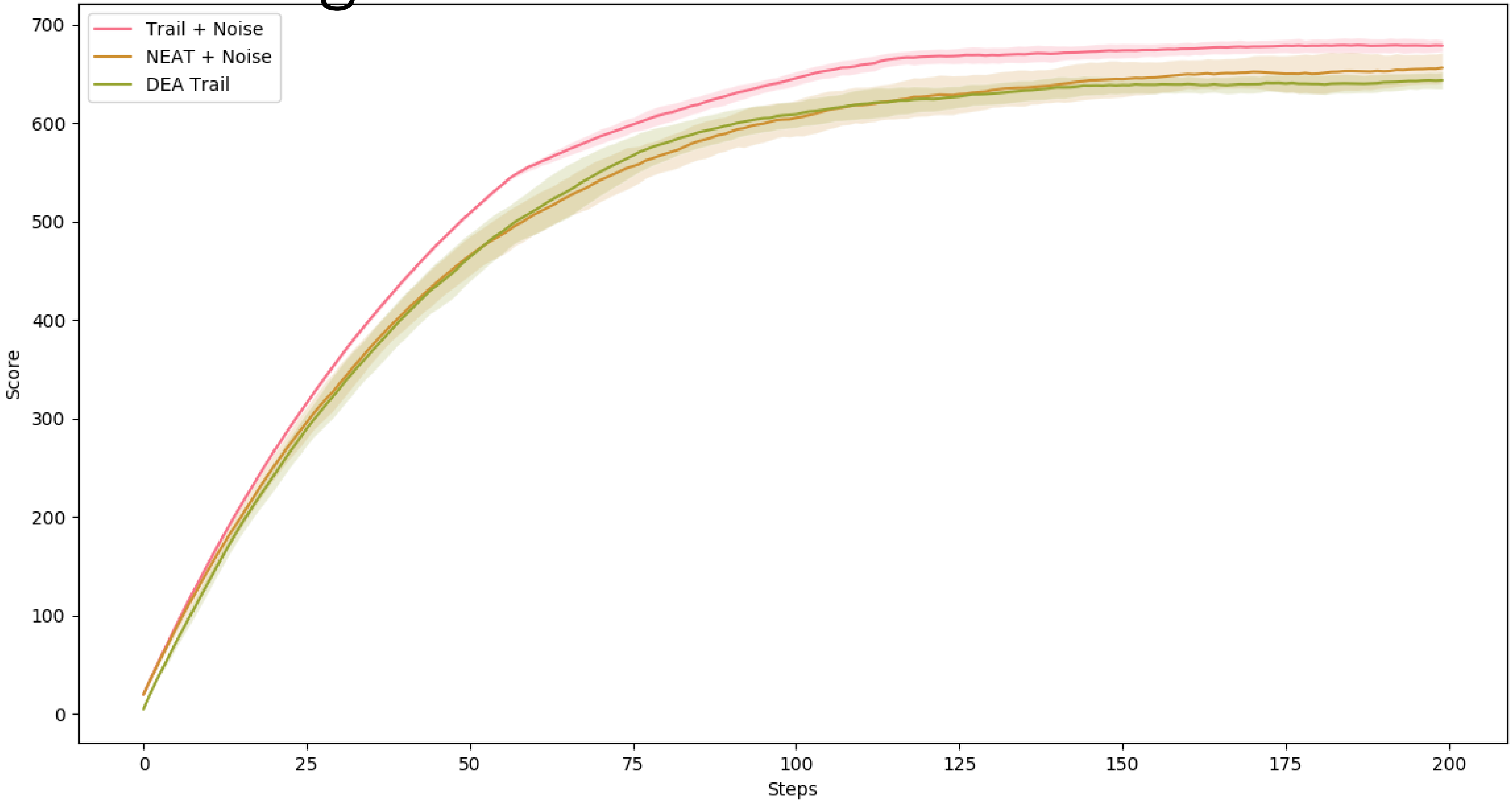
1 Agent



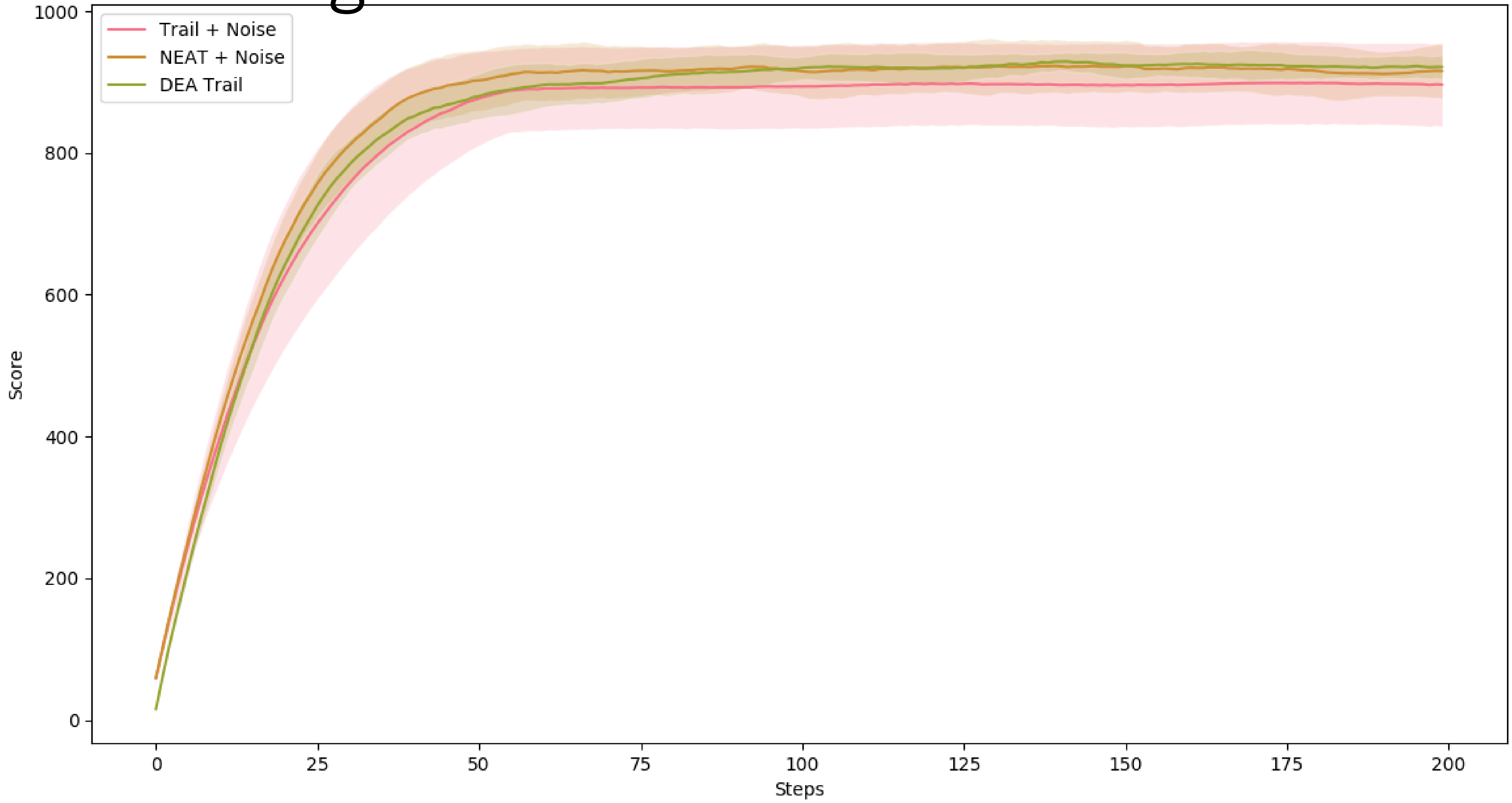
5 Agents



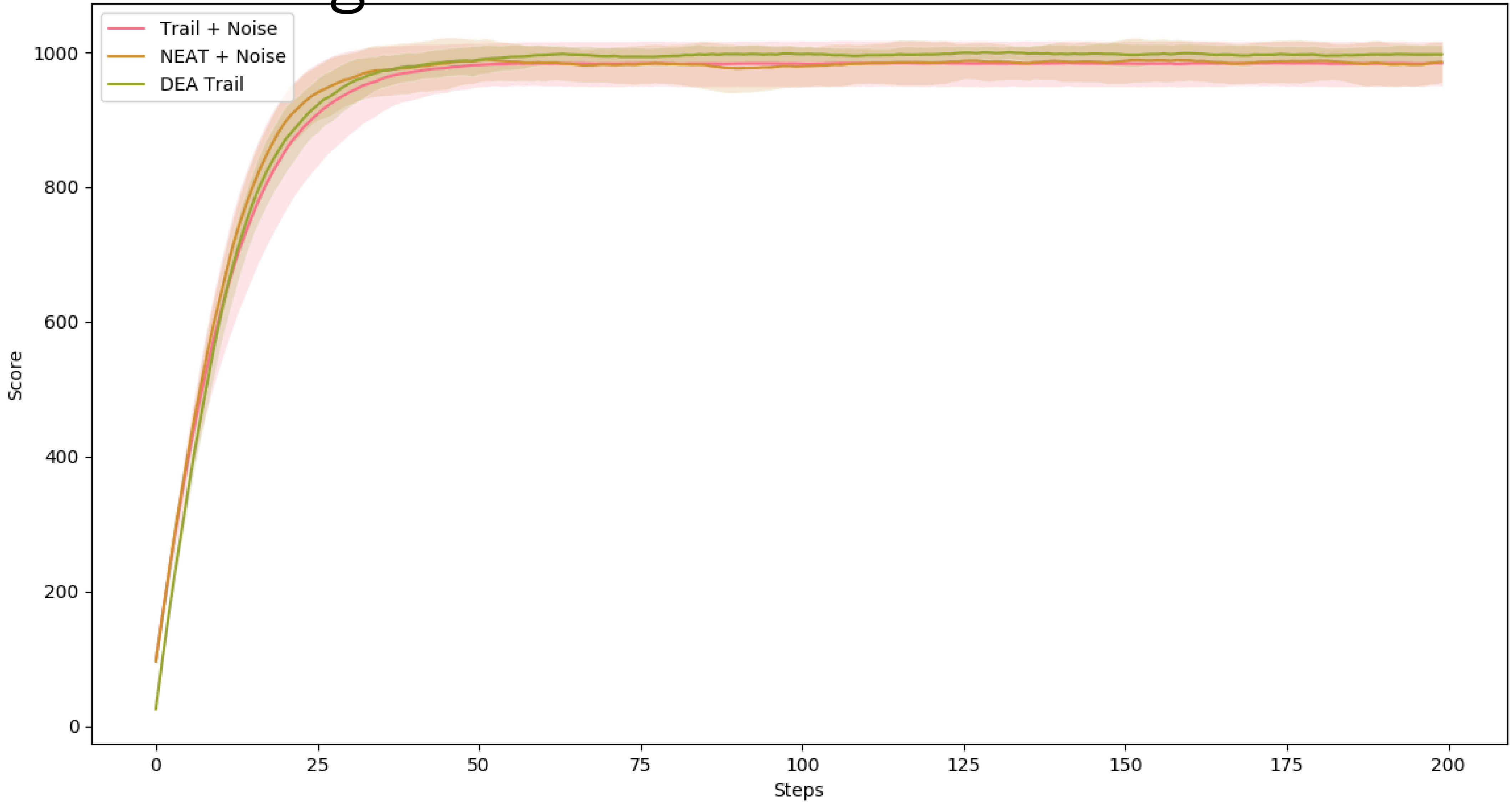
Combining Persistent Surveillance and MATSP



Combining Persistent Surveillance and MATSP



Combining Persistent Surveillance and MATSP



Task Assignment for MATSP: Take away

- *Hopefully without making the entire presentation irrelevant*
- Higher level tasking can be more effective than local policies
  - Requires communication and coordination
  - Implicit coordination from the MATSP problem definition

- There can often be complementary higher level objectives:
  - MATSP + Persistant surveillance

uestions

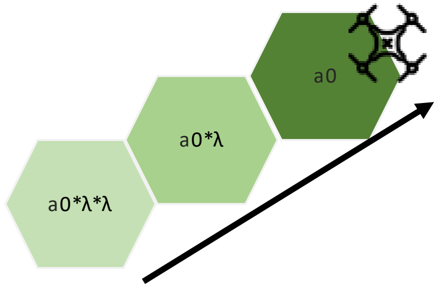
Thomas.kent@bristol.ac.uk

Appendix

Theoretical Max

- Number of hexes n = 56
- Hex height (width) = 15m
- Agent speed 5m/s => **3dt to cross**
- Linear Increase per timestep:
  ld = 5 -> adds 15 to the hex so **a0 = 15**
- Th = 120, dt = 3
- If we make a trail around all n=56 hexes we can hit **542**.
- If we continue and re-join 'tail' we can max out each hex so a0 = 20 and we can then hit **723**

$$\lambda = \left(\frac{1}{2}\right)^{\frac{dt}{T_h}}$$



Geometric Series

$$a_0^0 + a_0\lambda^1 a_0\lambda^2 + ...a_0\lambda^n = \sum_{k=0}^{n-1} a_0\lambda^k = a_0\left(\frac{1-\lambda^n}{1-\lambda}\right)$$

Multi-Agent: Geometric Series

$$a_0\left(\frac{1-\lambda^{n_1}}{1-\lambda}\right) + a_0\left(\frac{1-\lambda^{n_2}}{1-\lambda}\right) + ... + a_0\left(\frac{1-\lambda^{n_{N_a}}}{1-\lambda}\right)$$