Emergent Communication - First steps

Maxime Toquebiau

September 30th, 2021

Since last time

Emergent Communication

- Extended review of literature
- Development of the Push Environment
- Experimenting with models
- First results

Models:

- CommNet (Sukhbaatar et al., 2016)
- DIAL, RIAL (Foerster et al., 2016)
- BiCNet (Peng et al., 2017)
- Mordatch and Abbeel, 2018
- ATOC (Jiang and Lu, 2018)
- TarMAC (Das et al., 2019)
- IC3Net (Singh et al., 2019)
- VBC (Zhang et al., 2019)
- NDQ (Wang et al., 2020b)
- IMAC (Wang et al., 2020a)
- TMC (Zhang et al., 2020)

What tasks ?

What language ?

What RL algorithm ?

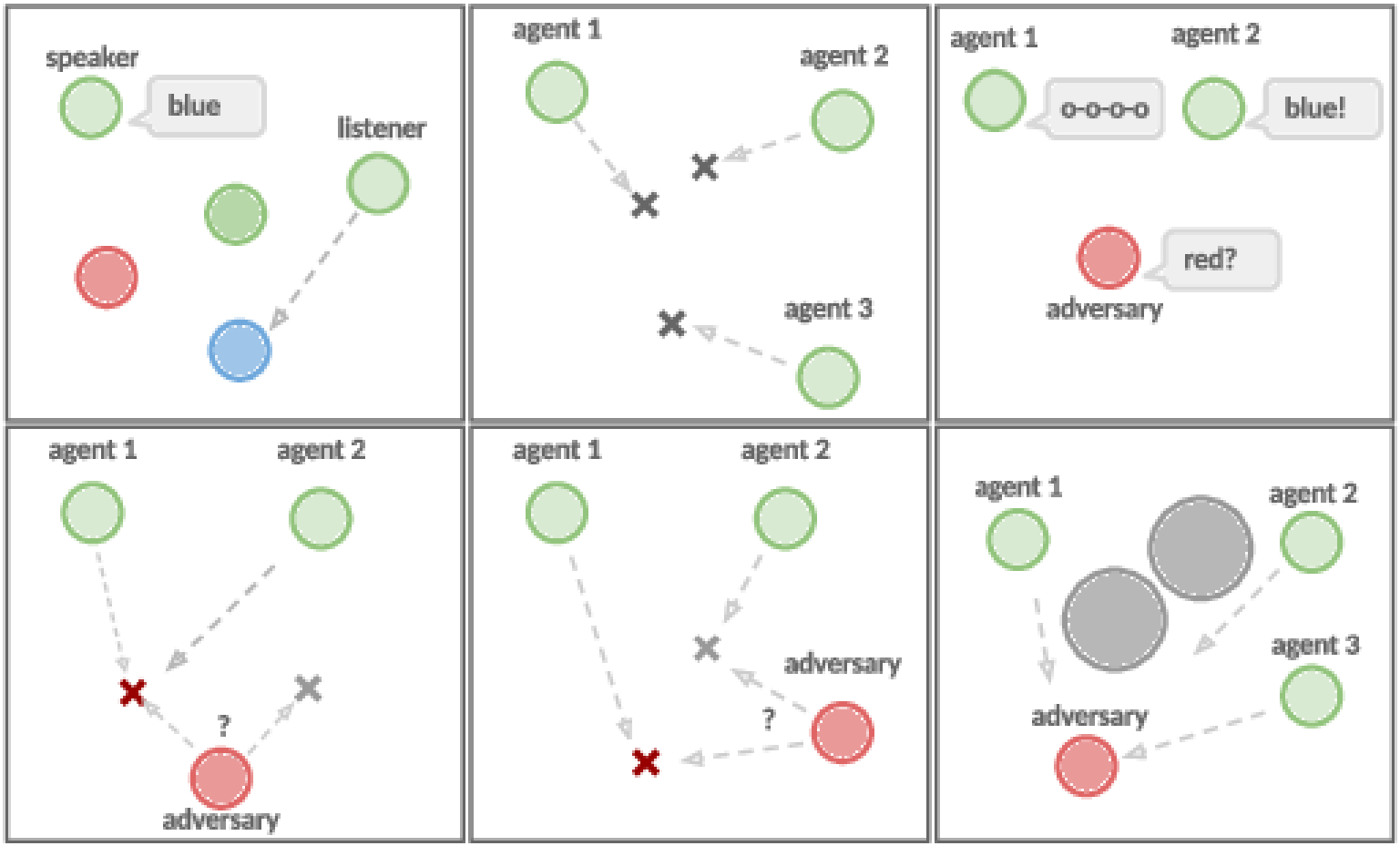
How to learn communication ?

How does communication work ?

How does it affect the agents' policy ?

How to design communication well ?

Base: Multiagent Particle Environment (MPE)



Emergence of Grounded Compositional Language in Multi-Agent Populations, Mordatch and Abeel, 2018

MPE Characteristics:

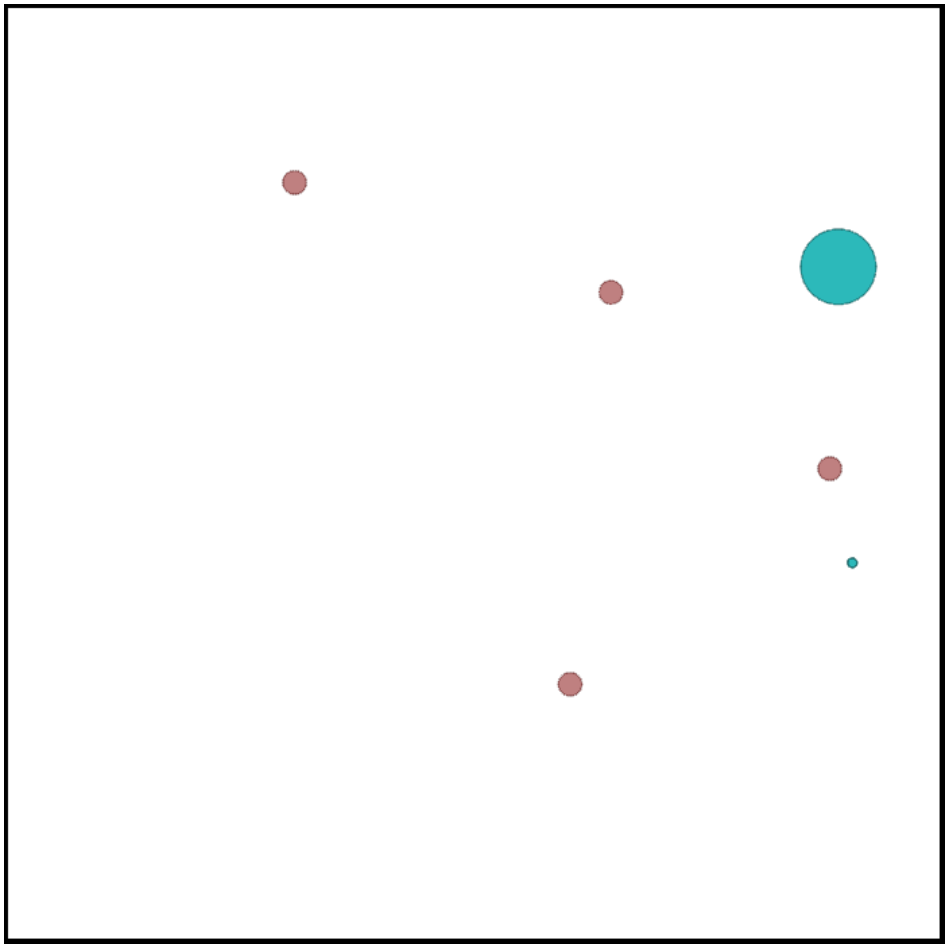
- Physics based
- Easy to use
- Easy to create custom scenario
- Several existing scenarios used in literature

Existing scenarios:

- Very simple and straightforward
- Very short
- Fully observable

Coop Push Scenario:

- 4 agents
- 1 movable object bigger and heavier
- 1 unmovable landmark
- **Goal:** Move the object on the landmark
- Partially observable



Observations:

- Own position and velocity
- Position and velocity of agents and objects in observation range
- Position of landmarks in observation range

Actions:

- Discrete: Up/Down/Left/Right
- or Continuous: $[dx, dy], (dx, dy) \in [-1, 1]^2$

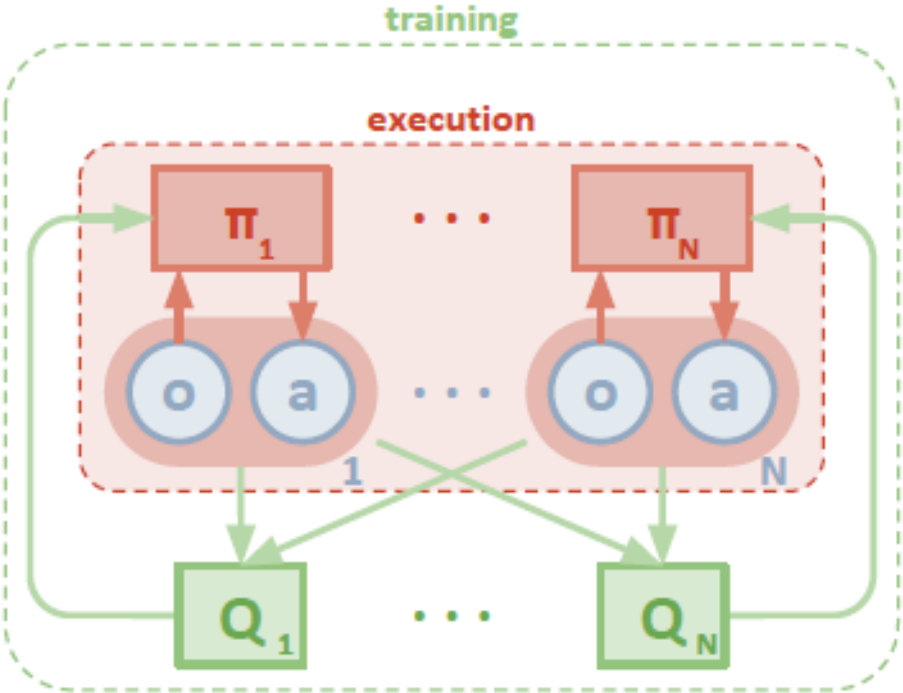
Reward:

- -1 x squared distance between object and landmark
- $+$ penalty for collision between agents

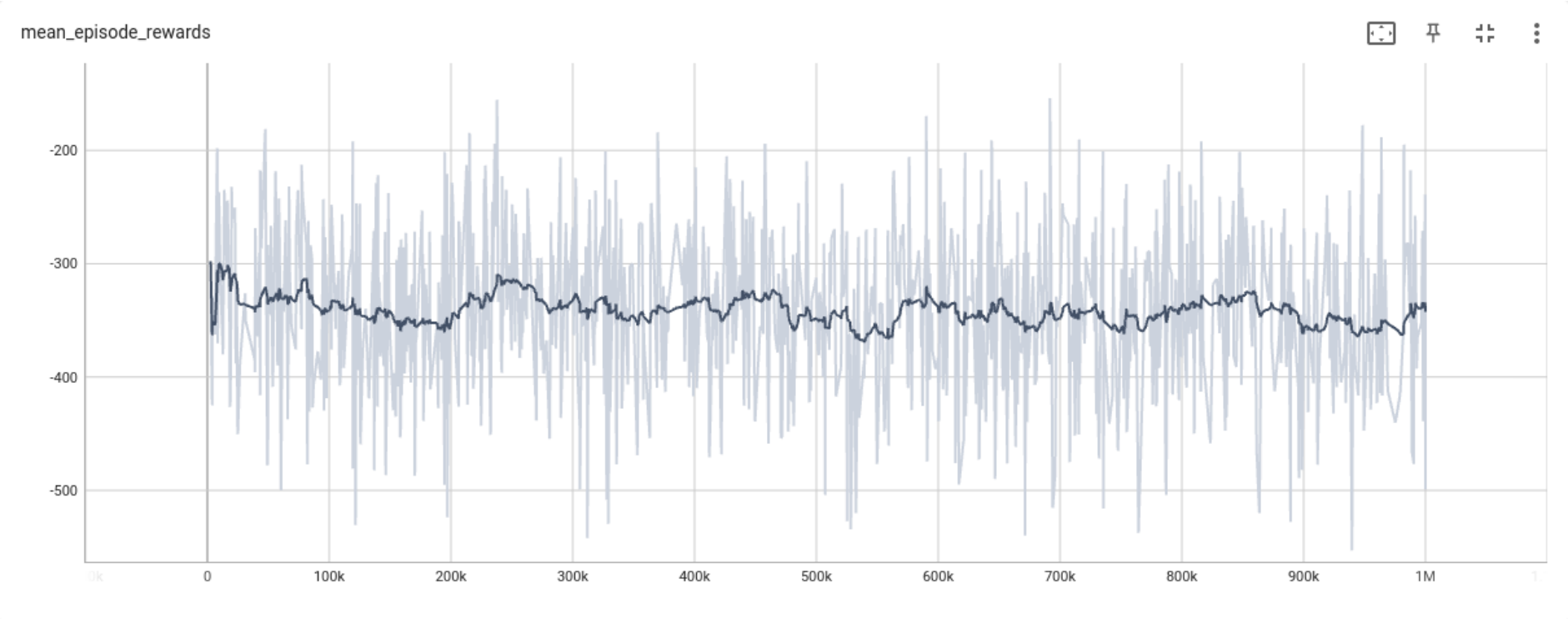
MADDPG (Lowe et al., 2017)

- Relatively simple implementation
- Originally implemented for MPE
- Quite light model
- Used to be state-of-the-art
- No communication

⇒ Good for experimenting in the environment



No convergence!



Simpler settings:

- Less agents
- Single agent
- Fully observable
- Reward for distance to the object
- No max speed

$\Rightarrow$ Still no good results

Hypotheses:

- Scenario is broken
- Scenario is way harder than previous ones
- Reward is poorly defined
- MADDPG is bad:
  - **Partially observable:** No memory, no communication
  - **Fully observable:** Exploration issue ?

Future tasks

- Keep experimenting to fix scenario
- Try training other MADRL methods (QMIX, MAVEN, COMA)
- Add memory and communication

Thank you!