# Learning to cooperate with Multi-Agent Reinforcement Learning

Maciej Wiatrak

ML in PL, November 23rd, 2019

University of Edinburgh

# Introduction

**Why is this research important?**

Building and observing the interactions between artificial agents could help us to gain a better understanding about human behaviour.

# Introduction

**Why is this research important?**

Building and observing the interactions between artificial agents could help us to gain a better understanding about human behaviour.

**What we know what we don’t know**

Human Intelligence encapsulates social aspects, but how do we incorporate them into machines?

# Introduction

**Why is this research important?**

Building and observing the interactions between artificial agents could help us to gain a better understanding about human behaviour.

**What we know what we don't know**

Human Intelligence encapsulates social aspects, but how do we incorporate them into machines?

**Experiment**

The study of the emergence of collective behaviour among artificial intelligence agents.

# Introduction

**Why is this research important?**

Building and observing the interactions between artificial agents could help us to gain a better understanding about human behaviour.

**What we know what we don't know**

Human Intelligence encapsulates social aspects, but how do we incorporate them into machines?

**Experiment**

The study of the emergence of collective behaviour among artificial intelligence agents.

**Hypothesis**

Observing how agents learn to cooperate could have promising applications in both social sciences and artificial intelligence.

# What is intelligence?

***"Intelligence measures an agent's ability to achieve goals in a wide range of environments"***

(Legg and Hutter, 2007)

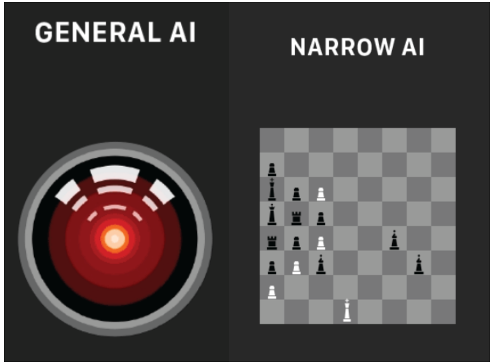
Inspired by: Slides from Thore Graepel

# What is intelligence?

***"Intelligence measures an agent's ability to achieve goals in a wide range of environments"***

(Legg and Hutter, 2007)

**Generality $\succ$ Complexity**



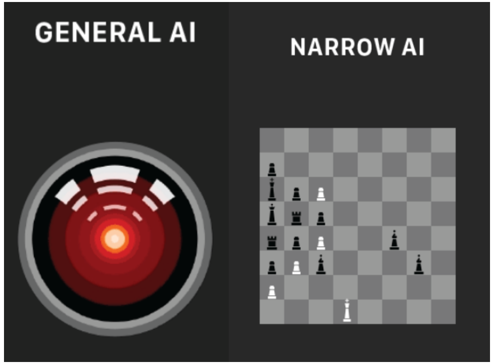

Inspired by: Slides from Thore Graepel

# What is intelligence?

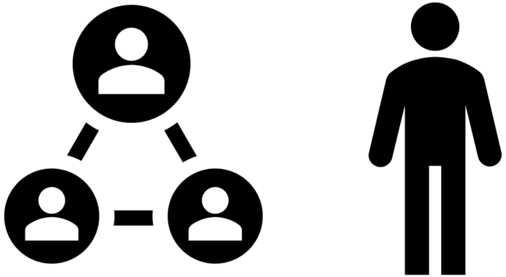
***"Intelligence measures an agent's ability to achieve goals in a wide range of environments"***

(Legg and Hutter, 2007)

**Generality ≻ Complexity**



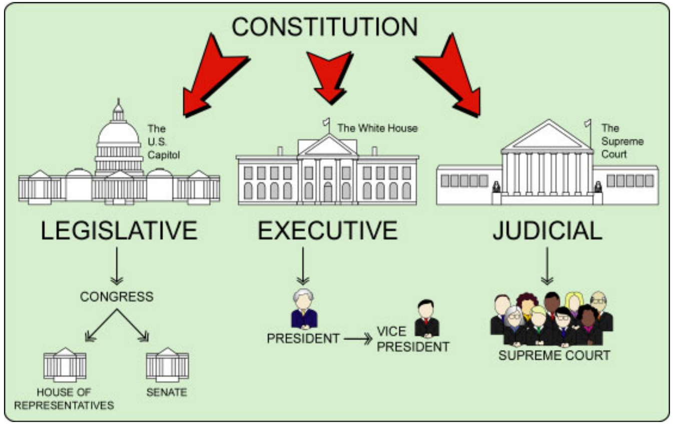

**Multi-agent ≻ Single-agent**



Inspired by: Slides from Thore Graepel

# Why should we care about multi-agent design?

1. **We live in a multi-agent world…**
   - Examples: government, market, traffic, family
   - …in order to succeed, an agent needs to consider the actions of other agents.

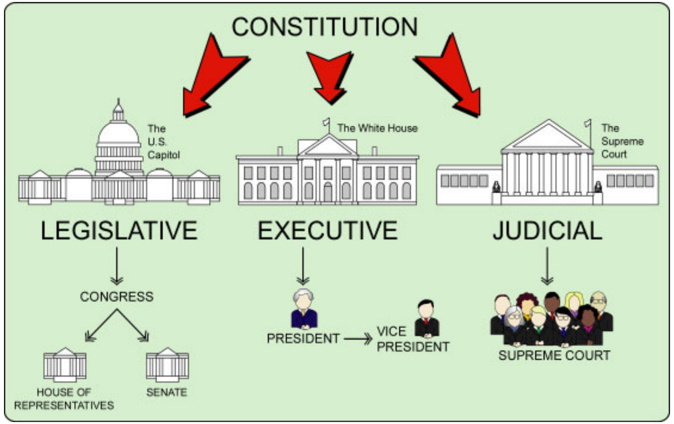



Inspired by: Slides from Thore Graepel

# Why should we care about multi-agent design?

1. **We live in a multi-agent world…**
   - Examples: government, market, traffic, family
   - …in order to succeed, an agent needs to consider the actions of other agents.

2. **Multi-agent design provides robustness scalability and flexibility.**



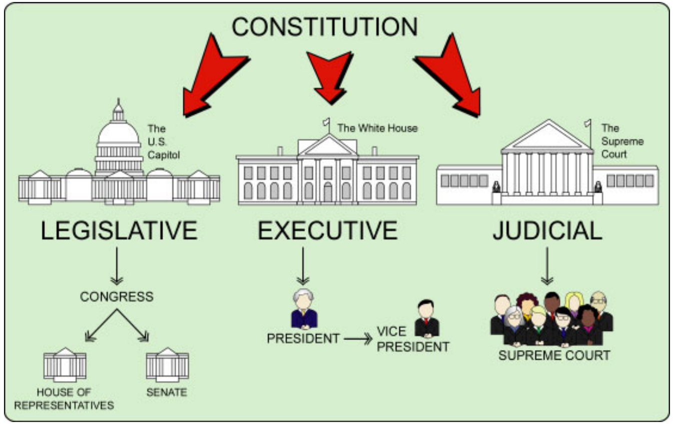

Inspired by: Slides from Thore Graepel

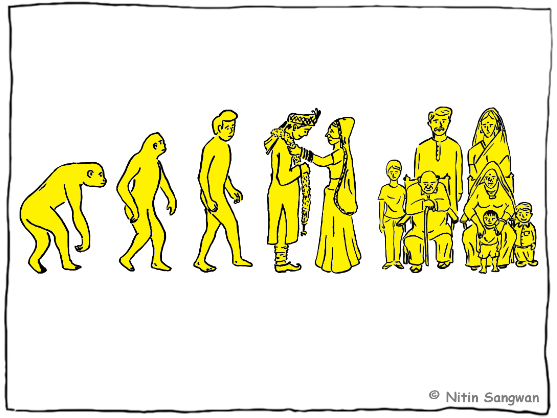
# Why should we care about multi-agent design?

1. **We live in a multi-agent world…**
   - Examples: government, market, traffic, family
   - …in order to succeed, an agent needs to consider the actions of other agents.

2. **Multi-agent design provides robustness scalability and flexibility.**

3. **Human Intelligence did not evolve in isolation…**
   - …it's a result of cumulative cultural evolution.
   - Why should it be possible to create AI in a single-agent framework?
   - "It takes a village to raise a child" (African proverb)




Inspired by: Slides from Thore Graepel

# Social dilemmas

***"Social dilemmas expose tensions between collective and individual rationality"***

Situations where an individual may profit from selfishness, unless too many individuals choose the selfish option, in which case the whole group loses.

(Rapoport, 1974)

Inspired by: Slides from Thore Graepel

# Social dilemmas

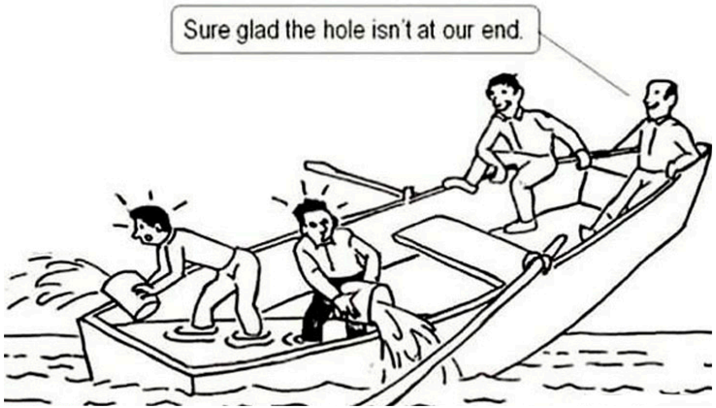
***"Social dilemmas expose tensions between collective and individual rationality"***

Situations where an individual may profit from selfishness, unless too many individuals choose the selfish option, in which case the whole group loses.

(Rapoport, 1974)

Examples:

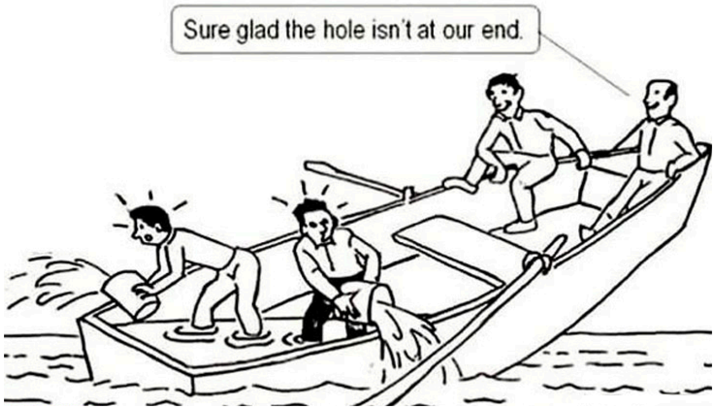
1. Free-riding
2. Voter turnout
3. Public goods



# Social dilemmas

***"Social dilemmas expose tensions between collective and individual rationality"***
Situations where an individual may profit from selfishness, unless too many individuals choose the selfish option, in which case the whole group loses.

(Rapoport, 1974)




Examples:
1. Free-riding
2. Voter turnout
3. Public goods
4. **The tragedy of the commons**

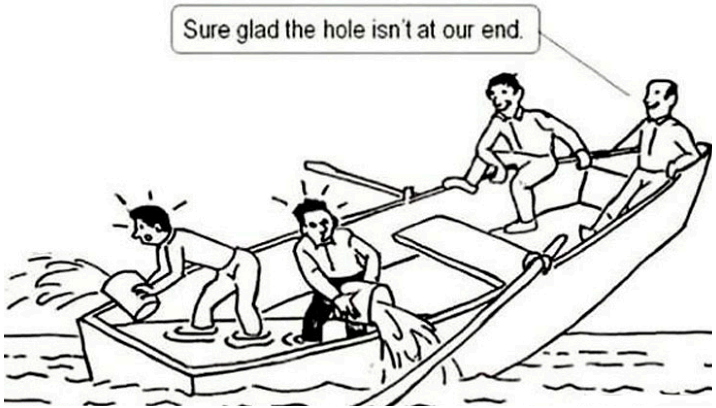
Inspired by: Slides from Thore Graepel

# Social dilemmas

***"Social dilemmas expose tensions between collective and individual rationality"***
Situations where an individual may profit from selfishness, unless too many individuals choose the selfish option, in which case the whole group loses.

(Rapoport, 1974)

Examples:
1. Free-riding
2. Voter turnout
3. Public goods
4. **The tragedy of the commons**



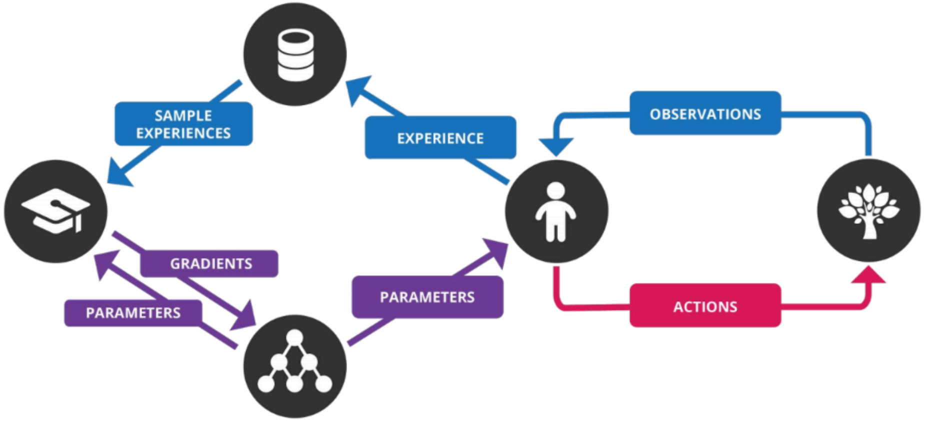

***Despite all these obstacles, how can cooperation emerge and be stable?***

Inspired by: Slides from Thore Graepel

# Deep Reinforcement Learning - DQN

**Deep Q-network:**

- Q-learning

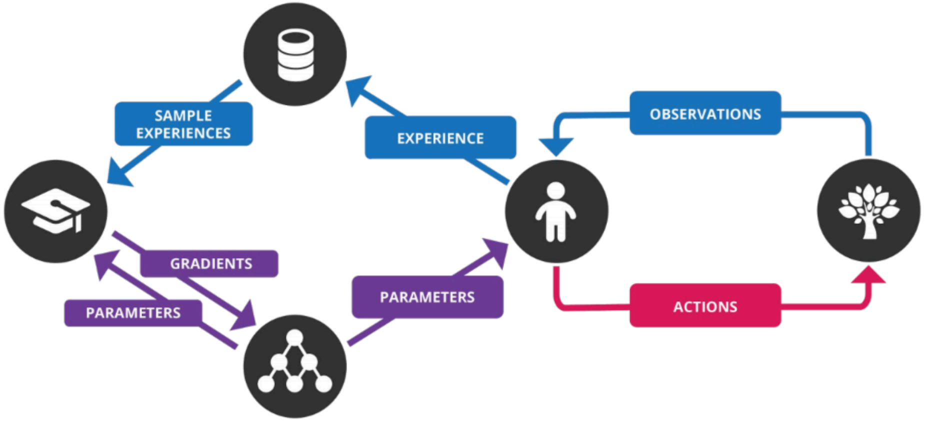



(Mnih et al., 2015)

# Deep Reinforcement Learning - DQN

**Deep Q-network:**

- Q-learning
- Off-policy

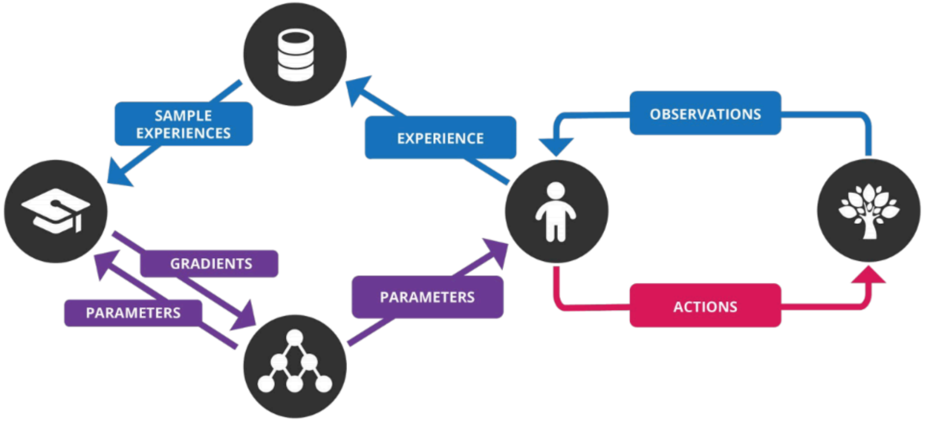



(Mnih et al., 2015)

# Deep Reinforcement Learning - DQN

**Deep Q-network:**

- Q-learning
- Off-policy
- Experience replay

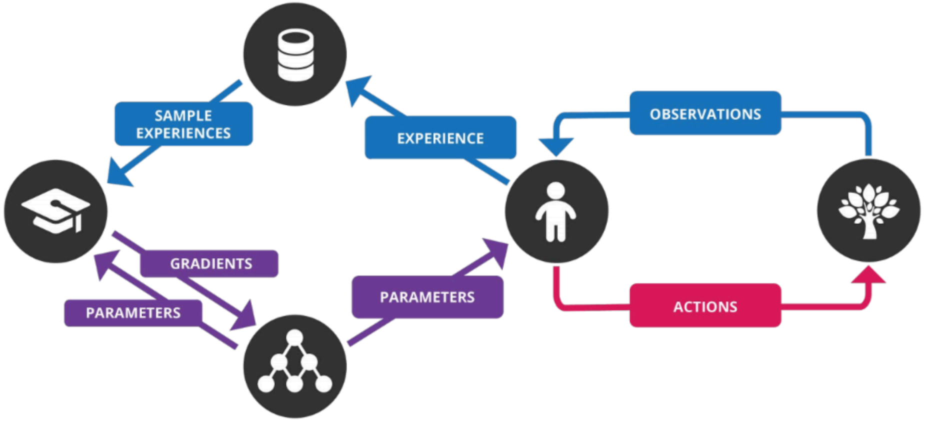



(Mnih et al., 2015)

# Deep Reinforcement Learning - DQN

**Deep Q-network:**

- Q-learning
- Off-policy
- Experience replay
- Target network

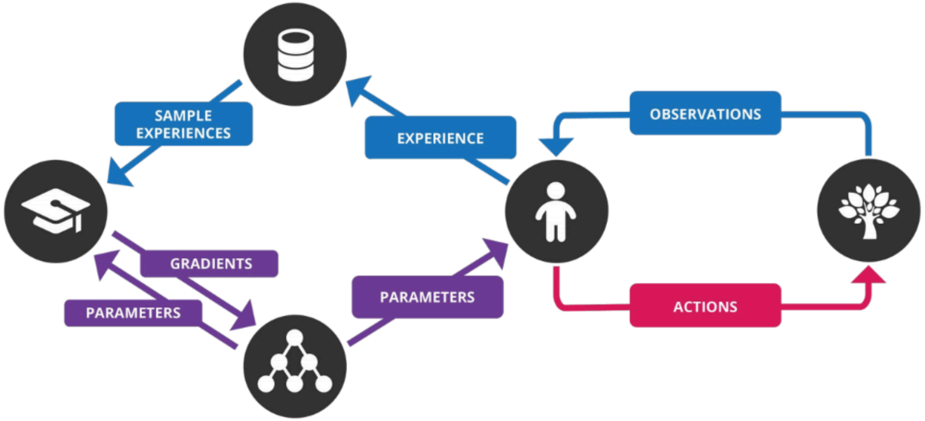



(Mnih et al., 2015)

# Deep Reinforcement Learning - DQN

**Deep Q-network:**

- Q-learning
- Off-policy
- Experience replay
- Target network
- …multiple improvements



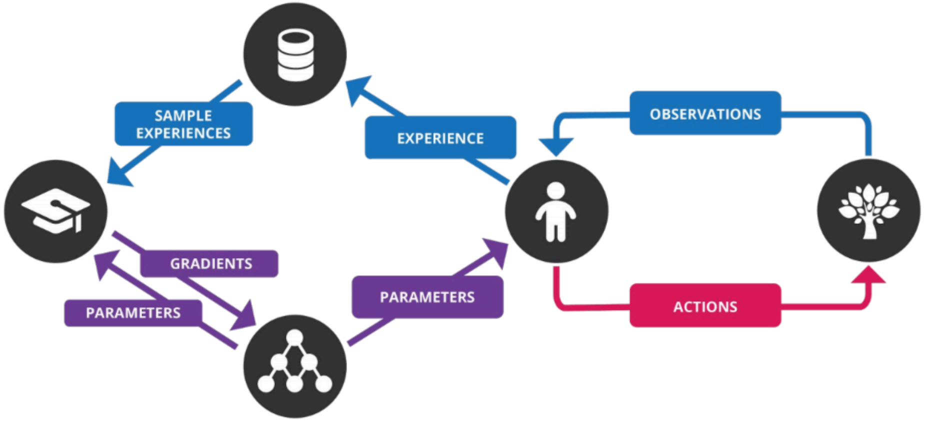

(Mnih et al., 2015)

# Deep Reinforcement Learning - DQN

**Deep Q-network:**

- Q-learning
- Off-policy
- Experience replay
- Target network
- …multiple improvements

**Decentralized training decentralized execution:**




(Mnih et al., 2015)

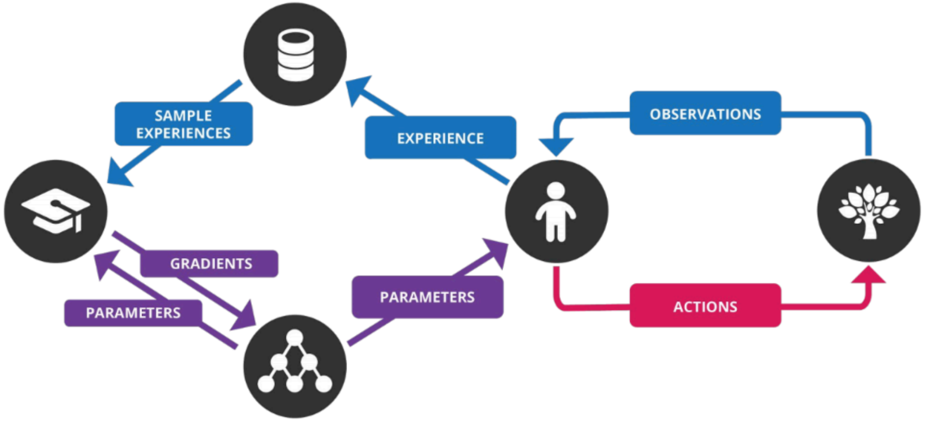
# Deep Reinforcement Learning - DQN

**Deep Q-network:**

- Q-learning
- Off-policy
- Experience replay
- Target network
- …multiple improvements

**Decentralized training decentralized execution:**

- All training is individual




(Mnih et al., 2015)

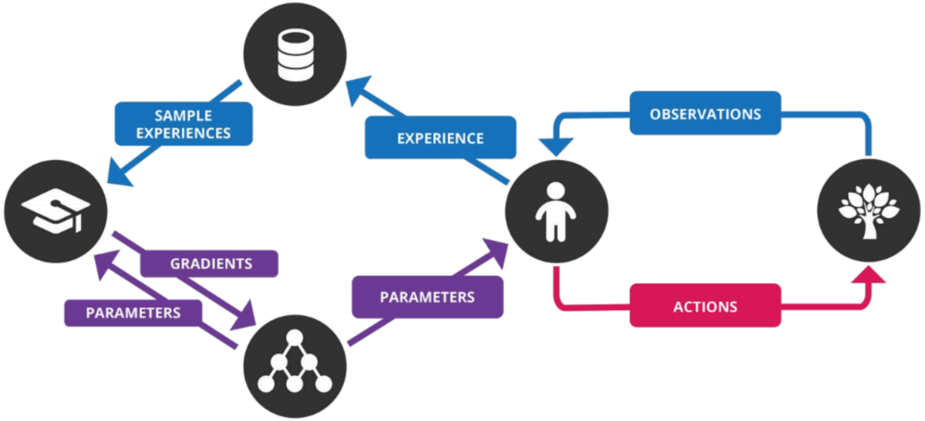
# Deep Reinforcement Learning - DQN

**Deep Q-network:**

- Q-learning
- Off-policy
- Experience replay
- Target network
- …multiple improvements

**Decentralized training decentralized execution:**

- All training is individual
- The agents regard other agents as part of the environment

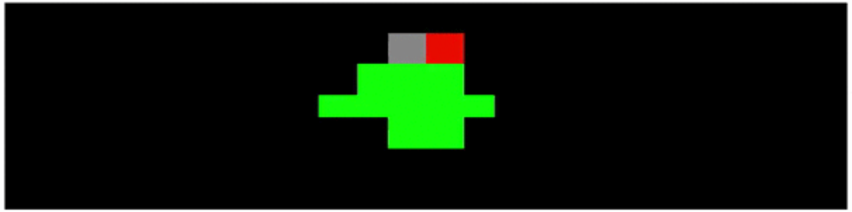



(Mnih et al., 2015)

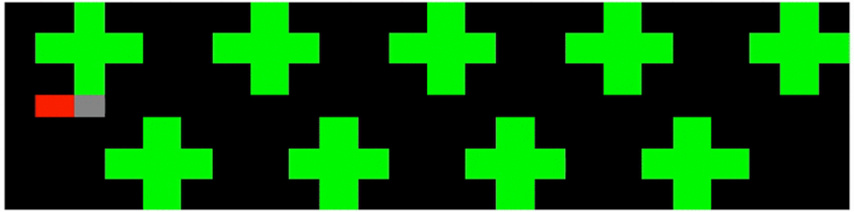
# Achieving sustainability

## Single-agent case/s

**Map 1:**



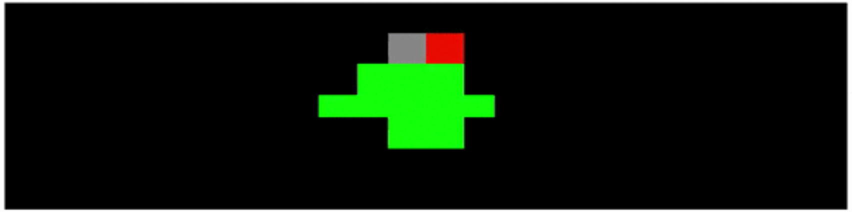
**Map 2:**



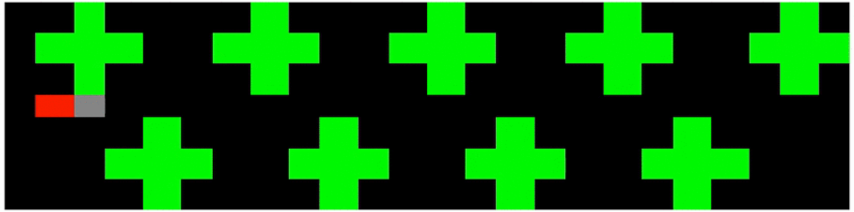
# Achieving sustainability

## Single-agent case/s

**Map 1:**



**Map 2:**

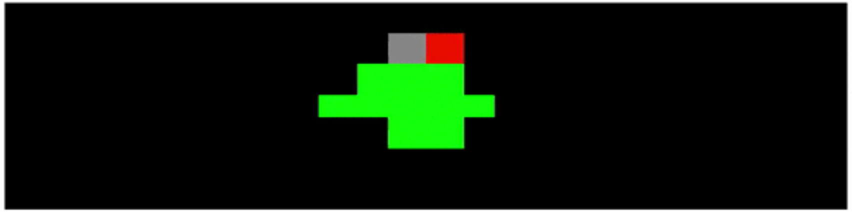


- Agent needs to learn a sustainable strategy to maximize its reward.

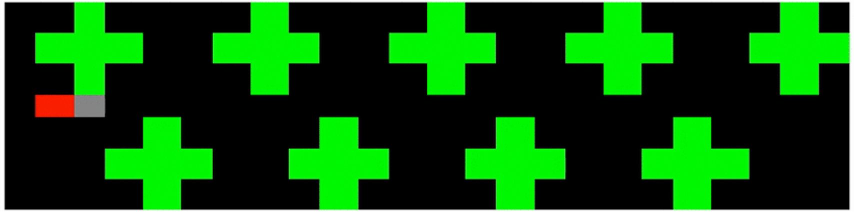
# Achieving sustainability

## Single-agent case/s

**Map 1:**

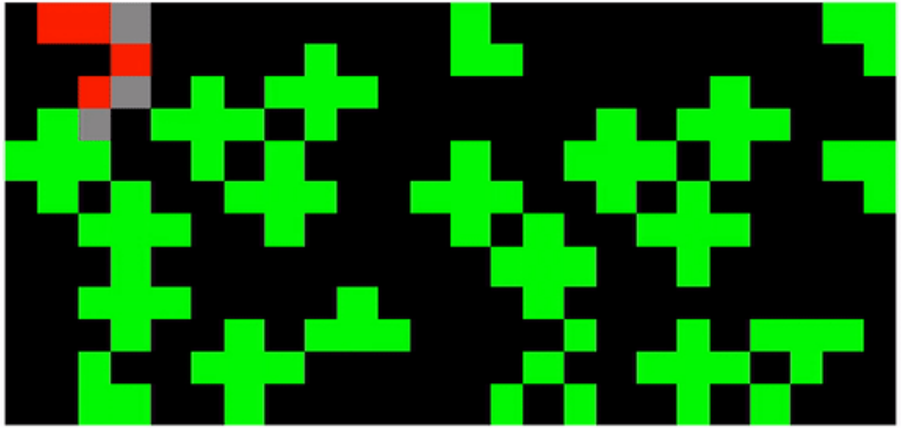


**Map 2:**



- Agent needs to learn a sustainable strategy to maximize its reward.

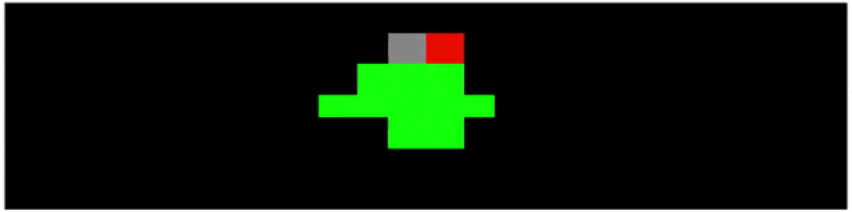
## Multi-agent case (the tragedy of the commons model)



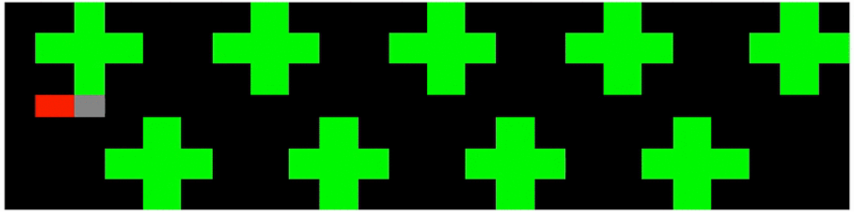
# Achieving sustainability

## Single-agent case/s

**Map 1:**

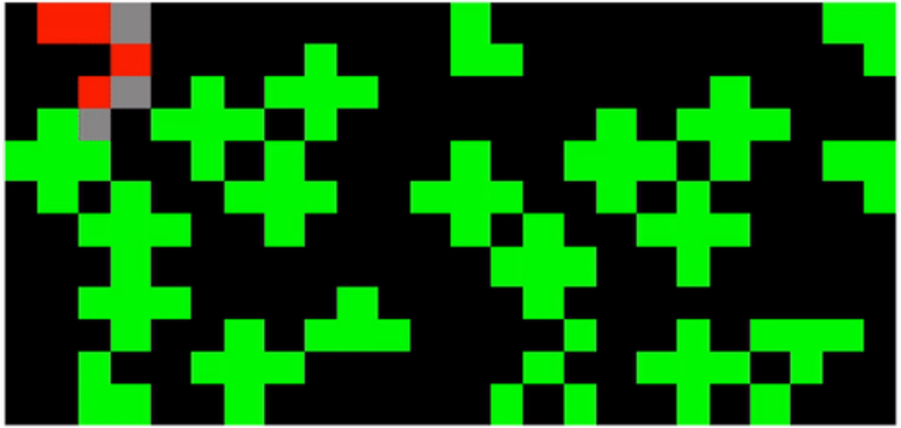


**Map 2:**



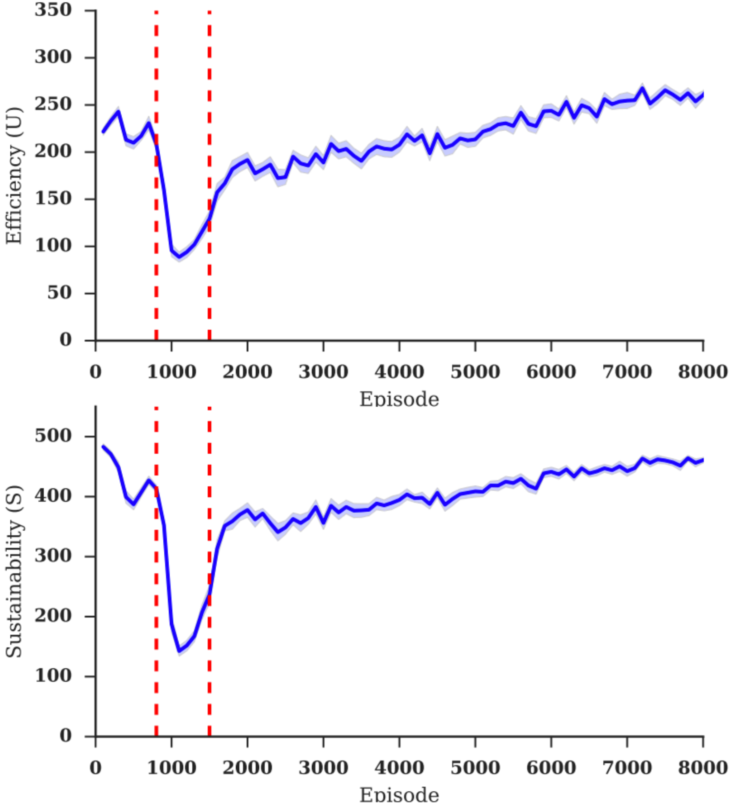
- Agent needs to learn a sustainable strategy to maximize its reward.

## Multi-agent case (the tragedy of the commons model)



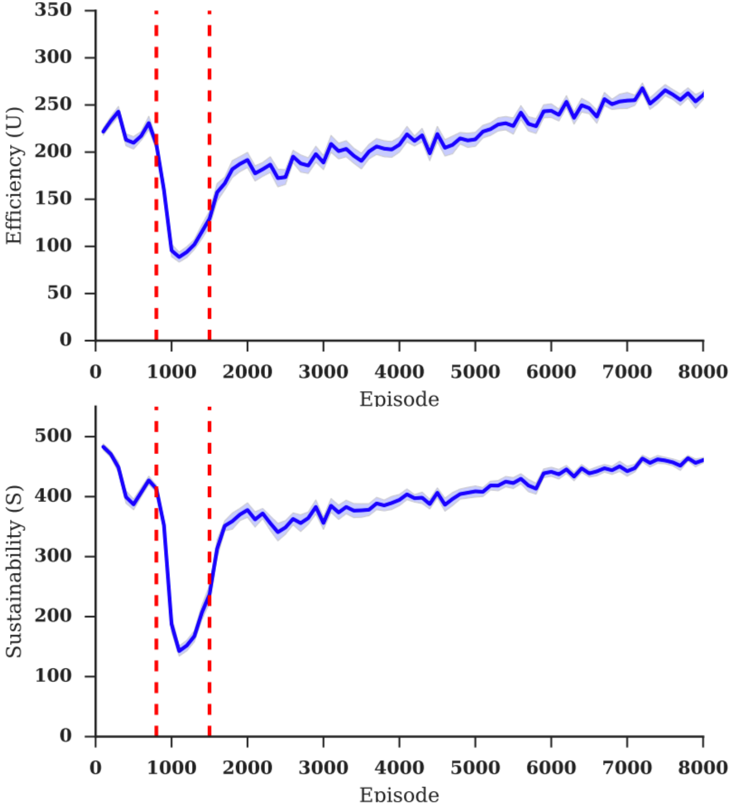
- Agents need to learn to cooperate with each other to prevent resource depletion and maximize their rewards.
- Agents can attack each other by *freezing* other agents with a laser beam.

# Results




(Perolat et al., 2017)

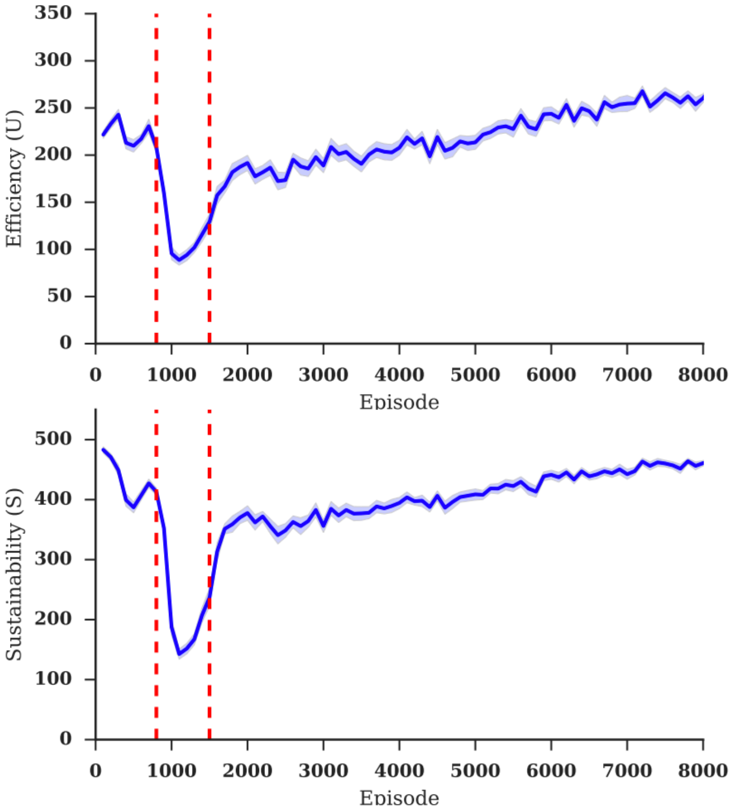
# Results




(Perolat et al., 2017)

1. **Agents with limited cognitive capabilities are capable of cooperation in resource management problem.**

# Results

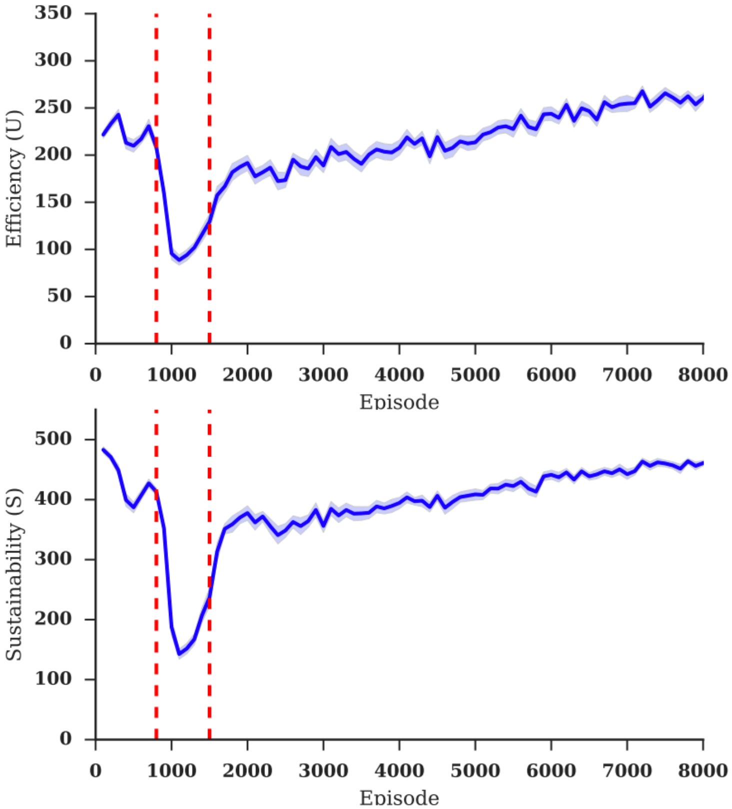



(Perolat et al., 2017)

1. **Agents with limited cognitive capabilities are capable of cooperation in resource management problem.**

2. **It opens avenues for further research in:**

# Results

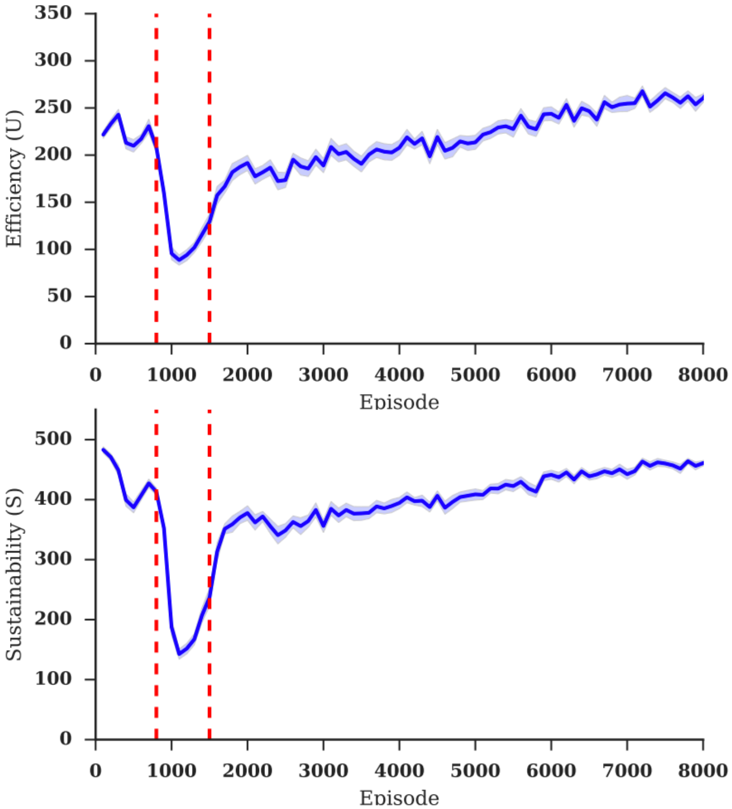



(Perolat et al., 2017)

1. **Agents with limited cognitive capabilities are capable of cooperation in resource management problem.**

2. **It opens avenues for further research in:**
   - **Social Sciences**
     - Allows for monitoring how different game parameters influence the outcome.
     - Could be potentially applied to aiding cooperative behaviour among humans.

# Results



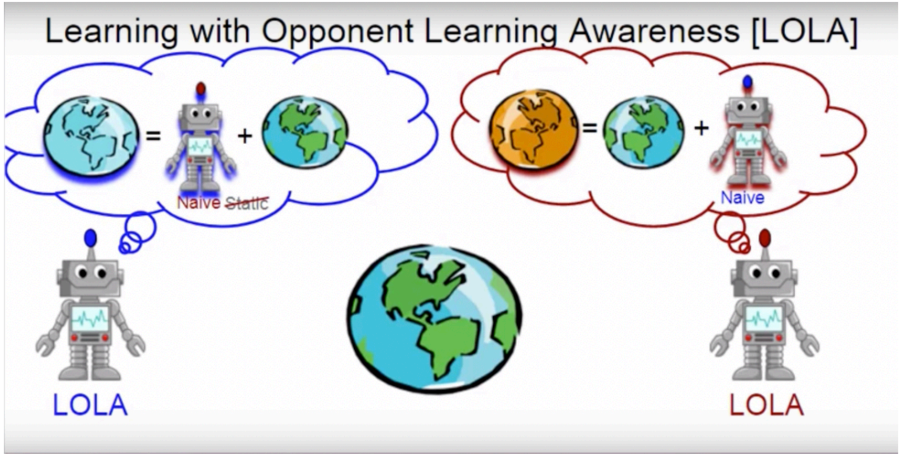

(Perolat et al., 2017)

1. **Agents with limited cognitive capabilities are capable of cooperation in resource management problem.**

2. **It opens avenues for further research in:**
   - **Social Sciences**
     - Allows for monitoring how different game parameters influence the outcome.
     - Could be potentially applied to aiding cooperative behaviour among humans.
   - **Artificial Intelligence**
     - Captures more information such as inequality and peacefulness.

# Developing better algorithms - LOLA

**Learning with Opponent-Learning Awareness (LOLA)**



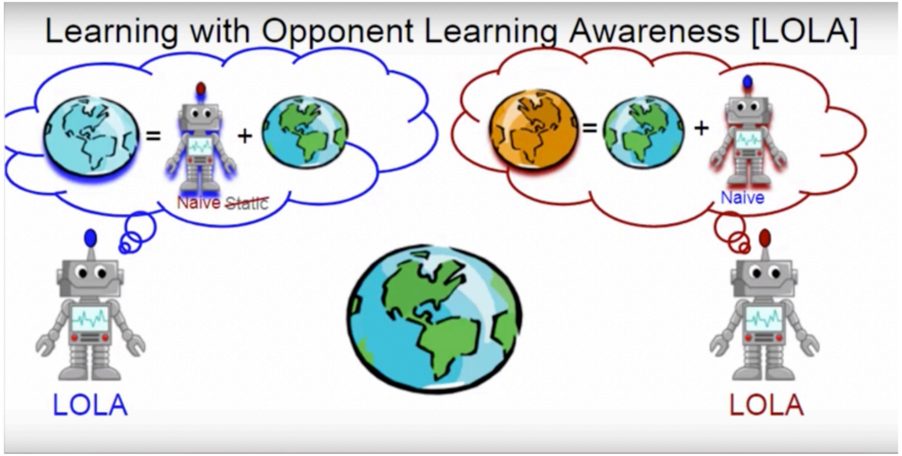

(Foerster et al., 2016)

$$\left(\frac{\partial V^1(\theta_i^1,\theta_i^2)}{\partial \theta_i^2}\right)^T \frac{\partial^2 V^2(\theta_i^1,\theta_i^2)}{\partial \theta_i^1 \partial \theta_i^2} \cdot \delta\eta,$$

# Developing better algorithms - LOLA

**Learning with Opponent-Learning Awareness (LOLA)**

- Opponent Modelling method



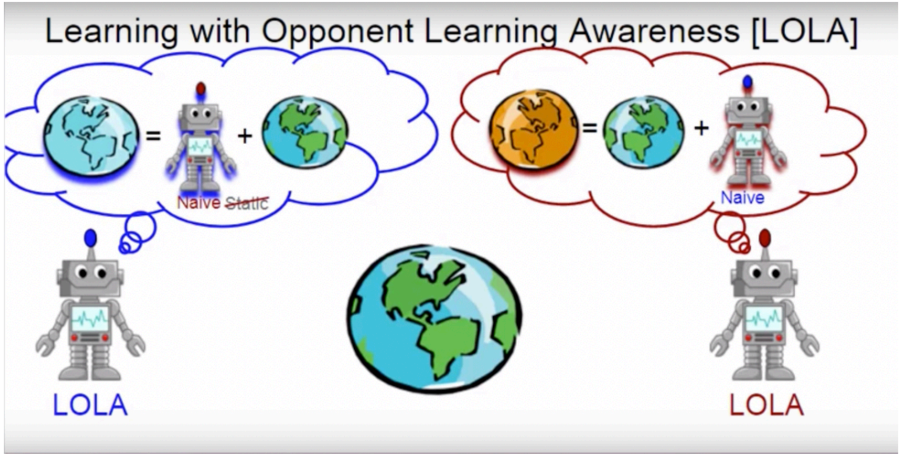

(Foerster et al., 2016)

$$\left(\frac{\partial V^1(\theta_i^1, \theta_i^2)}{\partial \theta_i^2}\right)^T \frac{\partial^2 V^2(\theta_i^1, \theta_i^2)}{\partial \theta_i^1 \partial \theta_i^2} \cdot \delta\eta,$$

# Developing better algorithms - LOLA

**Learning with Opponent-Learning Awareness (LOLA)**

- Opponent Modelling method
- Allows to account for the learning of other agents



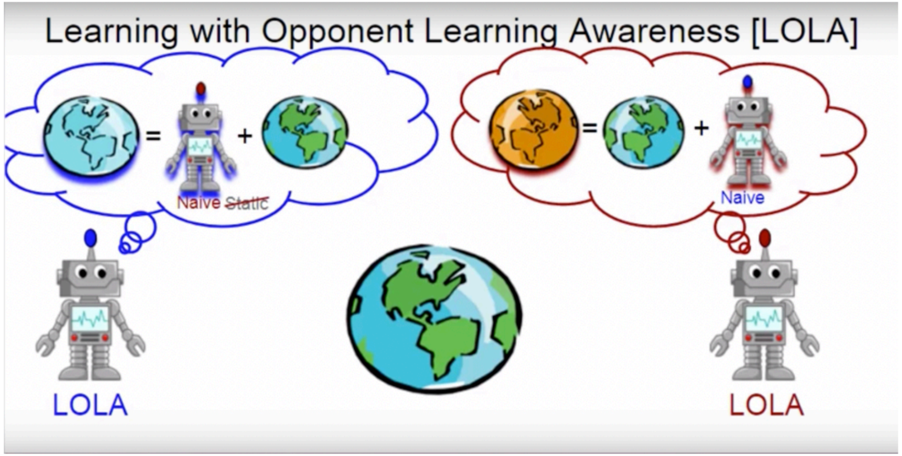

(Foerster et al., 2016)

$$\left(\frac{\partial V^1(\theta_i^1, \theta_i^2)}{\partial \theta_i^2}\right)^T \frac{\partial^2 V^2(\theta_i^1, \theta_i^2)}{\partial \theta_i^1 \partial \theta_i^2} \cdot \delta\eta,$$

# Developing better algorithms - LOLA

**Learning with Opponent-Learning Awareness (LOLA)**

- Opponent Modelling method
- Allows to account for the learning of other agents
- Adjusts its policy in order to shape the learning of other agents



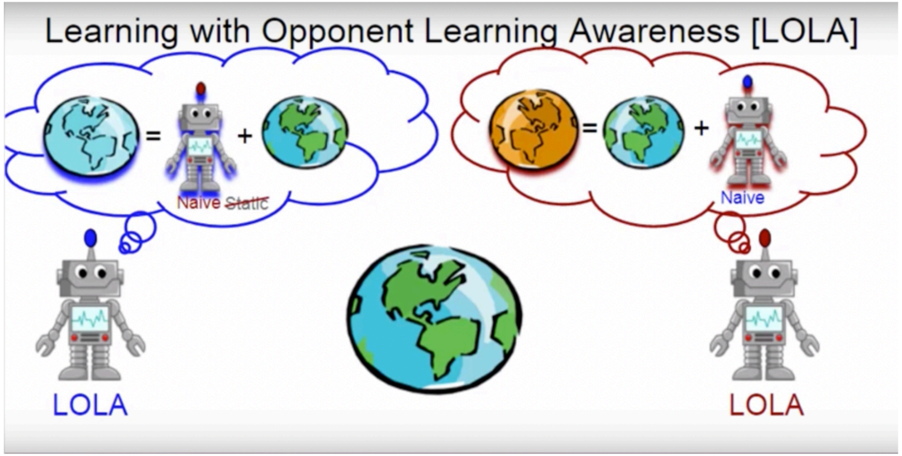

(Foerster et al., 2016)

$$\left(\frac{\partial V^1(\theta_i^1, \theta_i^2)}{\partial \theta_i^2}\right)^T \frac{\partial^2 V^2(\theta_i^1, \theta_i^2)}{\partial \theta_i^1 \partial \theta_i^2} \cdot \delta\eta,$$

# Developing better algorithms - LOLA

**Learning with Opponent-Learning Awareness (LOLA)**

- Opponent Modelling method
- Allows to account for the learning of other agents
- Adjusts its policy in order to shape the learning of other agents
- SOTA in cooperative game theory games



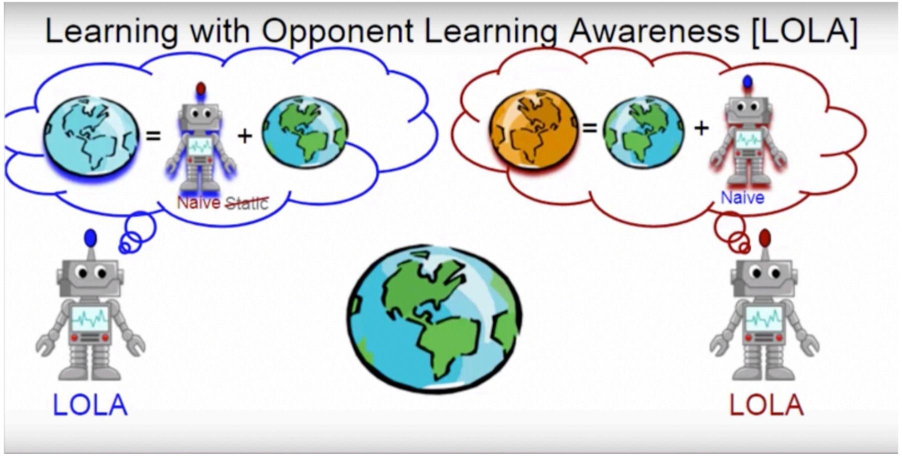

(Foerster et al., 2016)

$$\left(\frac{\partial V^1(\theta_i^1,\theta_i^2)}{\partial\theta_i^2}\right)^T \frac{\partial^2 V^2(\theta_i^1,\theta_i^2)}{\partial\theta_i^1\partial\theta_i^2}\cdot\delta\eta,$$

# Developing better algorithms - LOLA

**Learning with Opponent-Learning Awareness (LOLA)**

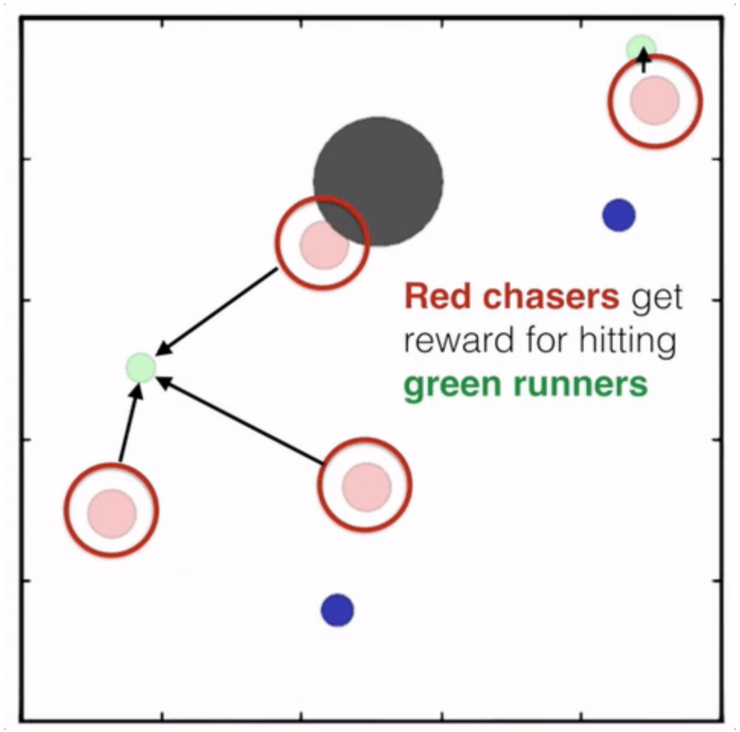
- Opponent Modelling method
- Allows to account for the learning of other agents
- Adjusts its policy in order to shape the learning of other agents
- SOTA in 5cooperative game theory games
- …but is memory and compute intensive




(Foerster et al., 2016)

$$\left(\frac{\partial V^1(\theta_i^1,\theta_i^2)}{\partial \theta_i^2}\right)^T \frac{\partial^2 V^2(\theta_i^1,\theta_i^2)}{\partial \theta_i^1 \partial \theta_i^2} \cdot \delta\eta,$$

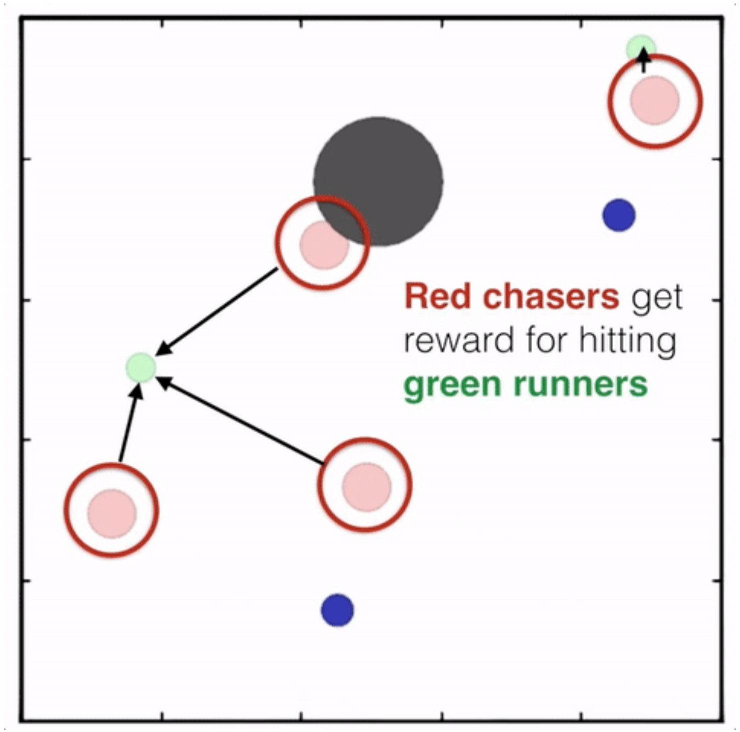
# Developing better algorithms - MADDPG




(Lowe et al., 2017)

**Multi-Agent Deep Deterministic Policy Gradient:**

# Developing better algorithms - MADDPG

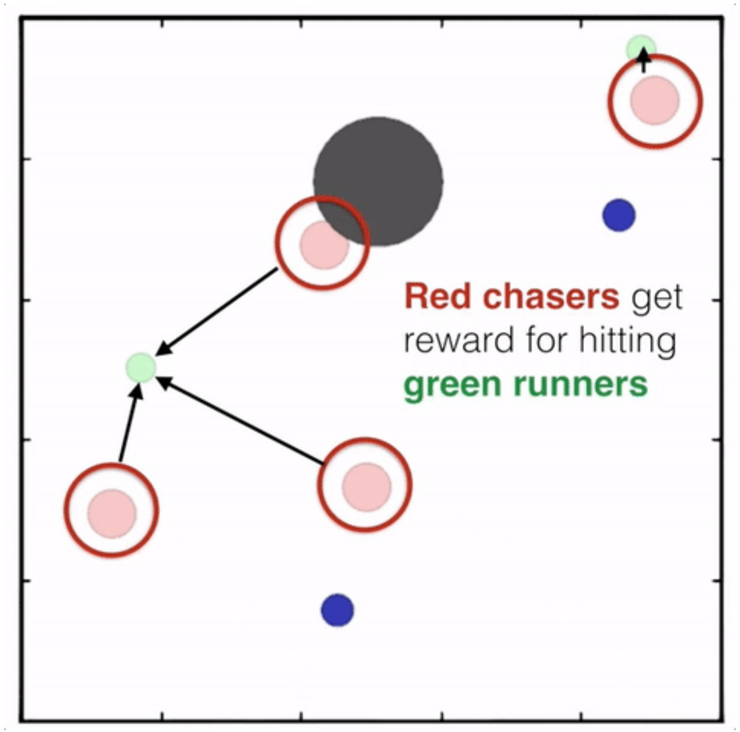



(Lowe et al., 2017)

**Multi-Agent Deep Deterministic Policy Gradient:**

- Centralized training decentralized execution

# Developing better algorithms - MADDPG

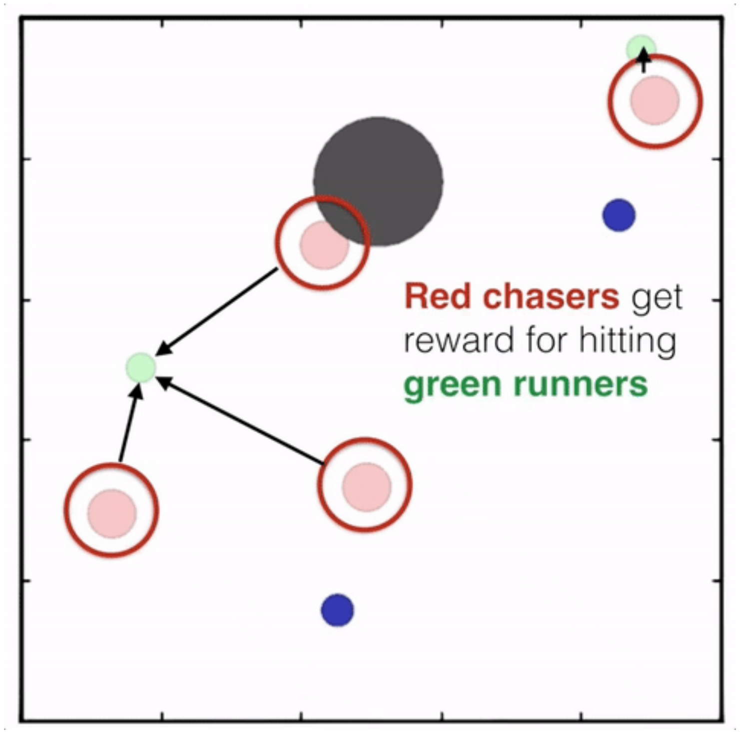



(Lowe et al., 2017)

**Multi-Agent Deep Deterministic Policy Gradient:**

- Centralized training decentralized execution
- Actor-critic architecture
  - Critics have the access to observations of all agents

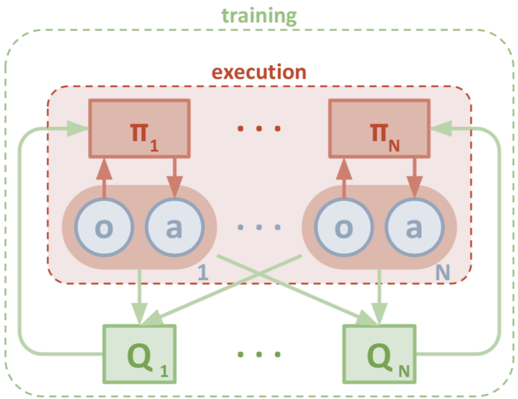
# Developing better algorithms - MADDPG




(Lowe et al., 2017)

**Multi-Agent Deep Deterministic Policy Gradient:**

- Centralized training decentralized execution
- Actor-critic architecture
  - Critics have the access to observations of all agents



# Challenges

- **Non-stationarity**
- **Open Multi-Agent Systems**
- **Multi-Agent Credit Assignment**
- **Transfer learning**
- **Limited Access to Open information**

# Thank you!

## References


Foerster, J., Chen, R., Al-Shedivat, M., Whiteson, S., Abbeel, P., Mordatch, I. (2016). ‘Learning with Opponent-Learning Awareness’. AAMAS.

Graepel, T. (2017), ‘The role of Multi-Agent Learning in Artificial Intelligence Research’. (*The Alan Turing Institute)*

Lowe, R., Wu, Y., Tamar, A., Harb, J., Abbeel, P., Mordatch, I. (2017). ‘Multi-Agent Actor-Critic for Mixed Cooperative-Competitive Environments. NIPS.

Legg, S. and Hutter, M. (2007). ‘Universal Intelligence: A Definition of Machine Intelligence’. *Minds and Machines.*

Leibo, J. Z., Zambaldi, V., Lanctot, M., Marecki, J. and Graepel, T. (2017), ‘Multi-agent Reinforcement Learning in Sequential Social Dilemmas’.

Mnih, V., Kavukcuoglu, K., Silver, D., Rusu, A. A., Veness, J., Bellemare, M. G., Graves, A., Riedmiller, M., Fidjeland, A. K., Ostrovski, G., Petersen, S., Beat- tie, C., Sadik, A., Antonoglou, I., King, H., Kumaran, D., Wierstra, D., Legg, S. and Hassabis, D. (2015), ‘Human-level control through deep reinforce- ment learning.’, *Nature* .

Niv, Y. (2009), ‘Reinforcement learning in the brain’, *Journal of Mathematical Psychology*

Ostrom, E., Gardner, R. and Walker, J. (1994), Rules, Games, and Common Pool Source problems, *in* ‘Rules, Games, and Common Pool Resources’.

Perolat, J., Leibo, J. Z., Zambaldi, V., Beattie, C., Tuyls, K. and Graepel, T. (2017), ‘A multi-agent reinforcement learning model of common-pool re- source appropriation’, *NIPS*.