Attention manipulation in reinforcement learning agents

Oriol Corcoll, Abdullah Makkeh, Jaan Aru, Dirk Oliver Theis, Raul Vicente Zafra
Tartu University

Introduction

Joint attention is the ability of individuals to focus on a **common goal** and is believed to be foundational to many of our social competencies like theory of mind.

Can deep reinforcement learning (RL) agents also benefit from joint attention?
As in Kaplan & Hafner (2004)*, we use their four prerequisites for joint attention:

- **Attention tracking:** the ability to recognize what goal others are attending to.
- **Attention manipulation:** use of communication to direct the attention of others to a common goal.
- **Social coordination:** engagement in coordinated interaction with others agents to accomplish a goal.
- **Intentional stance:** acknowledge that agents act towards a set goal.

We evaluate how to enable **better social coordination through attention manipulation** and provide a **measure of attention manipulation** that captures the degree to which an agent is manipulating others to achieve its goal.

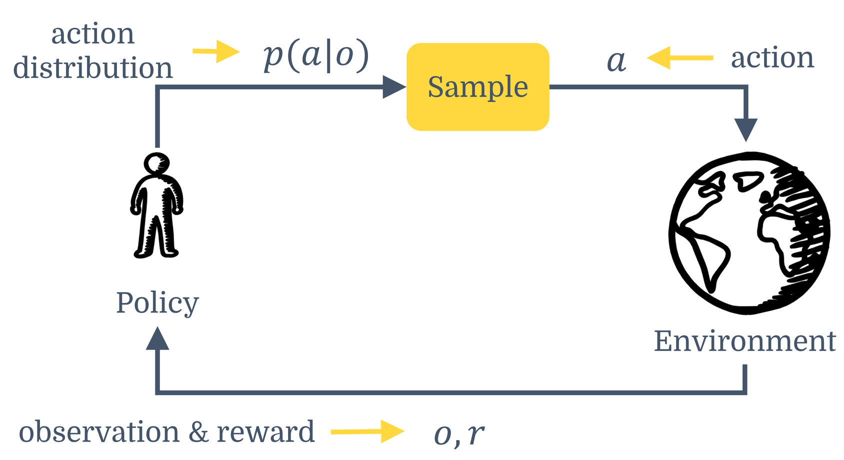
*Kaplan, F., & Hafner, V. (2004). The Challenges of Joint Attention (Tech. Rep.).

Attention Manipulation

Many Multi-agent RL research focus on developing **agents that solve tasks individually**. In To a large extent, this is due to the lack of environments providing the necessary pressures for these agent to need to collaborate with others. We found that when the **environment has the following characteristics**, agents learn to manipulate each other to cooperate:

- **Signalling:** Agents should have a way to communicate i.e. signals or complex messages.
  outcome: allows coordination and better exploration.
- **Specialization:** Agents need to have particular information or skill not available to others.
  outcome: division of labour and collaboration
- **Limited field of vision:** Cooperating becomes useful when there is the need for shared knowledge. If agents have complete vision of the environment, this need for sharing knowledge diminishes.
  outcome: promote cooperation by signals instead of vision.
- **Time pressure:** Attention manipulation can be used to complete tasks faster, e.g. agent would be able to pick two boxes on its own but it would take less time if it can use another agent to pick one of them.
  outcome: movement becomes expensive compared to signalling.
- **Collective goals:** Agents can have their own goals and rewards but, in order to cooperate, some of them need to be shared with other agents.
  outcome: for two agents to cooperate both should benefit from it.

Section you can skip if you know about RL



- In deep reinforcement learning, the policy is implemented with an **artificial neural network**.
- In our case the environment produces also a **signal "s"** and the policy takes this signal as input.

Measuring Attention Manipulation

Attention manipulation can be understood as influence over others to set a common goal. As shown by Jaques et al. (2018)*, influence can be computed in terms of counterfactuals i.e. answering questions like "would an agent do the same action if the other had not sent a signal?"

What would the action be if the other agent had not sent the signal?

$$p_{\sim s_i^s}(a^r|o^r) \equiv \sum_{\forall j \neq i} p(a^r|s_j^s, o^r)p(s_j^s|o^s)$$

How much did the signal influence my next action?

$$AM(r, s) \equiv D_{\mathrm{KL}}\left(p(a^r|s_i^s, o^r)||p_{\sim s_i^s}(a^r|o^r)\right)$$

signal receiver — signal sender

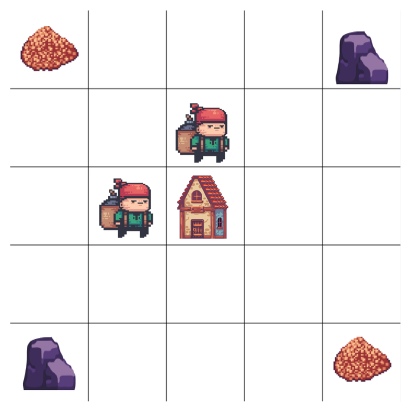
Instead of relying on a **collective reward that can be confounded** on other factors of the environment, this formula allows the measurement of an important aspect of the cooperation between agents, **attention manipulation.**

*Jaques, N., Lazaridou, A., Hughes, ... de Freitas, N. (2018). Social Influence as Intrinsic Motivation for Multi-Agent Deep Reinforcement Learning.

Environments and Model

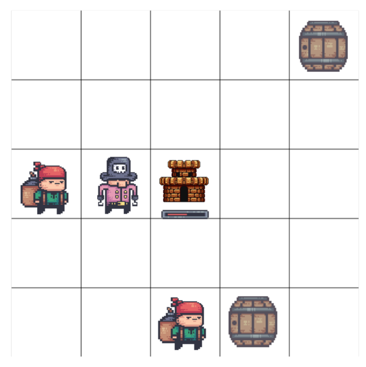
Resource Mining

Two agents need to mine coal and gold and take them to a nearby storage. The reward is bigger if they unload resources together.



- **Signaling:** agents receive a signal when other's backpack is full.
- **Specialization:** mine coal or gold.
- **Limited field of vision:** 3x3 of the total 6x6.
- **Time pressure:** the game ends if they run out of resource.
- **Collective goals:** agents need to unload resources together.

Defend the Fort

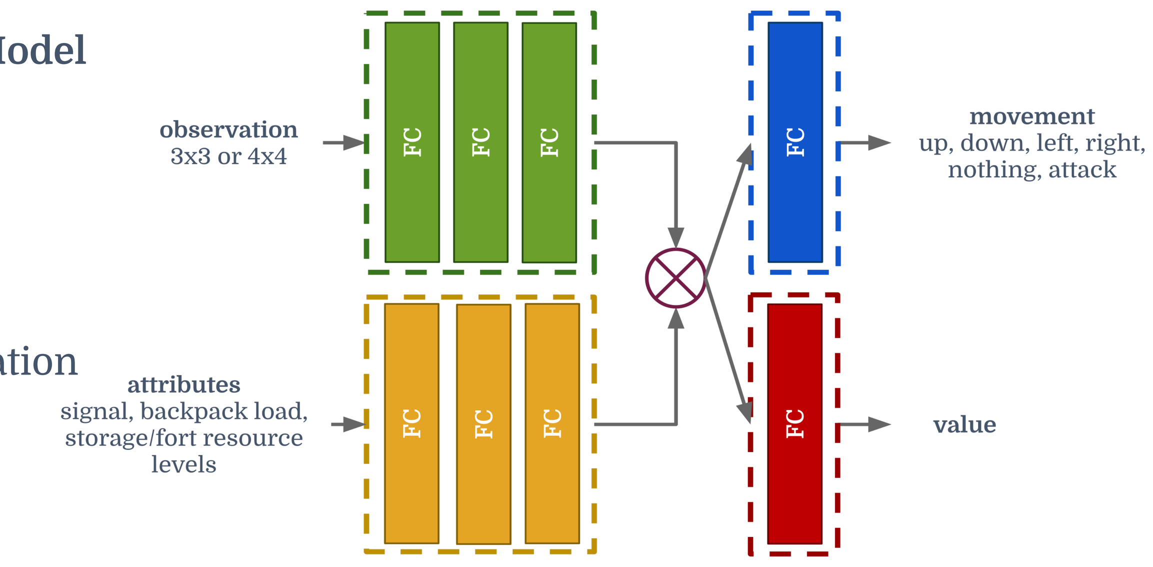


Two agents share the duty of protecting a fort and to gather food to be alive.

- **Signaling:** they receive a signal when one of them see an enemy.
- **Specialization:** defend or collect food.
- **Limited field of vision:** 4x4 of the total 10x10.
- **Time pressure:** food in the fort's storage decreases per time step.
- **Collective goals:** defending increases the chances of survival.

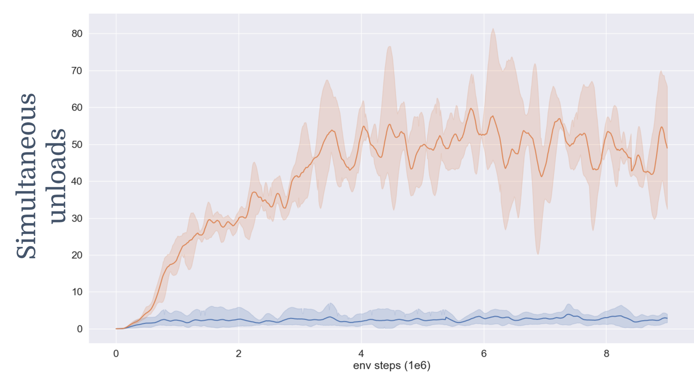
Reinforcement Learning Model

- Fully connected layers
- 32 units each
- ReLU
- Proximal Policy Optimization
- Adam optimizer



Results

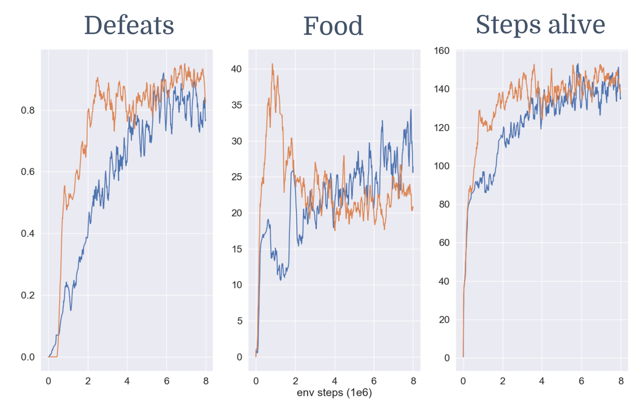
Resource Mining



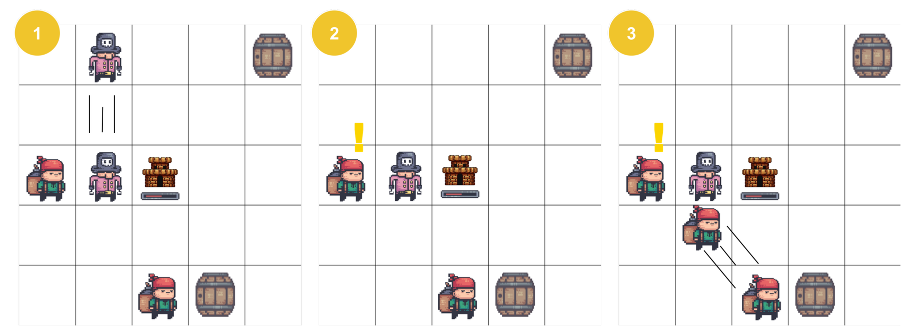
Simultaneous unload of resources by agents using signals (orange) and without using them (blue).

Evolution of the attention manipulation measurement during training, with signaling (orange) and without (blue).

Defend the Fort



Agents learn a strategy to defend the fort quicker with attention manipulation. Defeated enemies (left), collected food (middle) and steps alive (right).



Example of agents' reaction during an attack after training. An agent is guarding the fort and the other agent is collecting food. Once an attack starts (1) the guard agent sends a signal (2) to change the goal of the other agent from "collect" to "defend" (3).

Future Work

- In order to have a controllable setting, signals are produced by the environment. We will use our measurement of attention manipulation as **intrinsic reward** to learn to produce these signals.
- Original AM measure works w.r.t. actions but this measure could be more accurate if agents would **model goals**.

$$AM(r, s) \equiv \underbrace{D_{\mathrm{KL}}\left(p(g^r|s_i^s, o^r)\,[g^s]\,||p_{\sim s_i^s}(g^r|o^r)\,[g^s]\right)}$$

Did the agent change its goal due to When there is a signal, are the goals of both agents aligned?

- Does classical environments like **stag-hunt** require attention manipulation? We will analyze if this environment provides the necessary pressures for agents to use this skill.
- Both AM measures can be extended to delayed manipulation i.e. the swap between actions/goals due to a signal is not immediate in time.
- Currently, attention is positional but ideally we would use **gaze** (as in humans) to define what an agent is attending to.