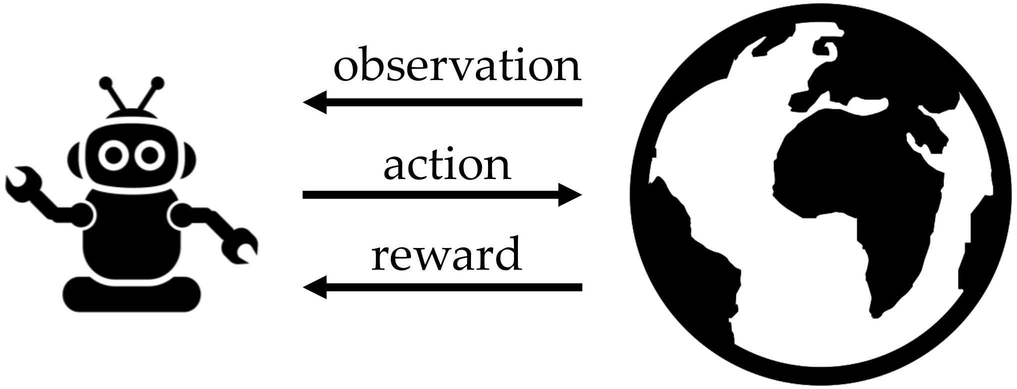
Robot Learning

Multi-agent Learning

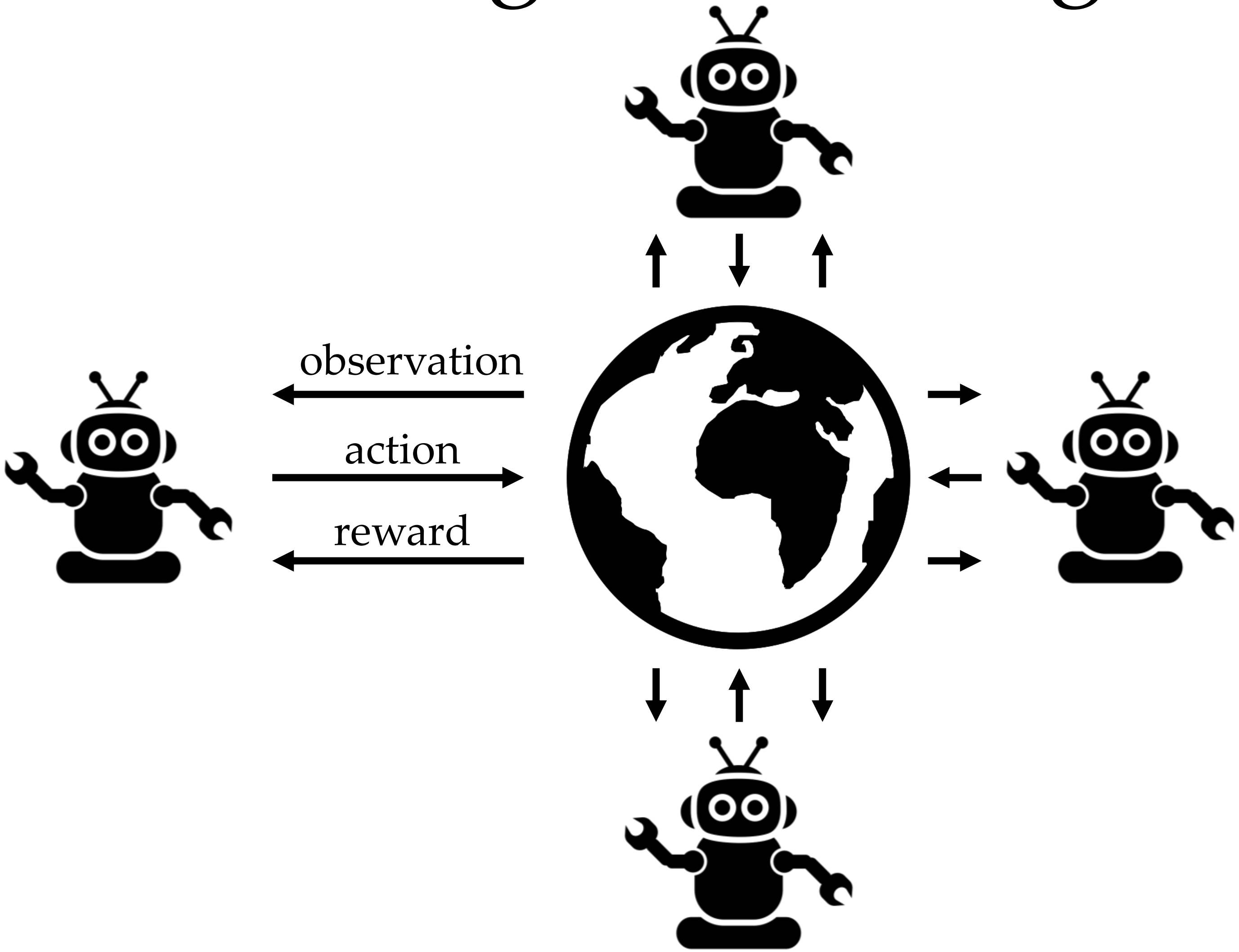
USC University of Southern California

LiraLab

What is multi-agent learning?



What is multi-agent learning?



Today…

- Taxonomy

- Fundamental problems (and some solutions)

Why multi-agent learning?

Real world applications often have multiple agents.

Can each agent learn independently?

Why multi-agent learning?

Real world applications often have multiple agents.

Can each agent learn independently? Sometimes.

Taxonomy: centralized vs. decentralized

- CTCE: This is single-agent learning.
- CTDE: Interesting!
- DTCE: Not realistic.
- DTDE: Independent agents (almost), but still interesting!

Taxonomy: centralized vs. decentralized

Why does centralized vs. decentralized training matter?

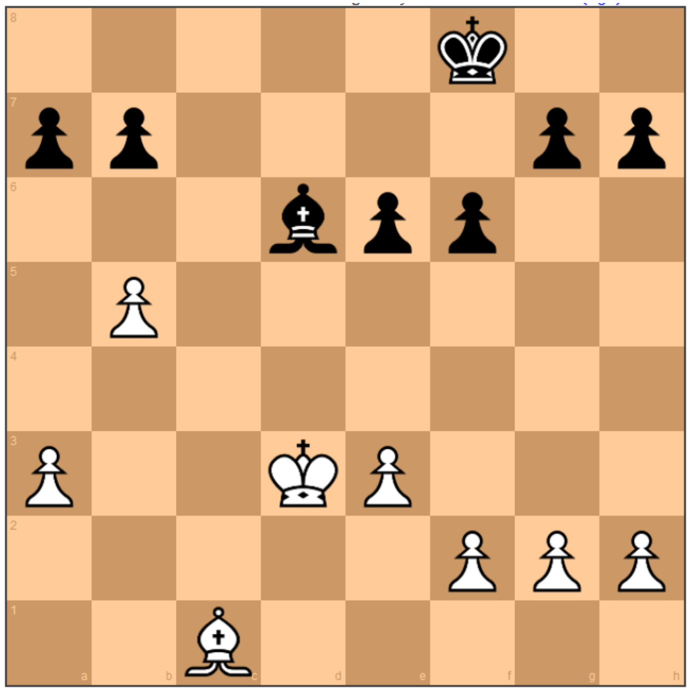
- Agents can share information if the training is centralized.
  - They can share their observations to predict the state more accurately.
  - They can share reward information.
  - We can *sometimes* have a shared critic to train individual actors better.

Taxonomy: cooperative vs. competitive

Fully competitive (zero-sum) systems

Matching pennies game

	Heads	Tails
Heads	2,0	0,2
Tails	0,2	2,0



From: ChessBase

Taxonomy: cooperative vs. competitive

Fully cooperative systems

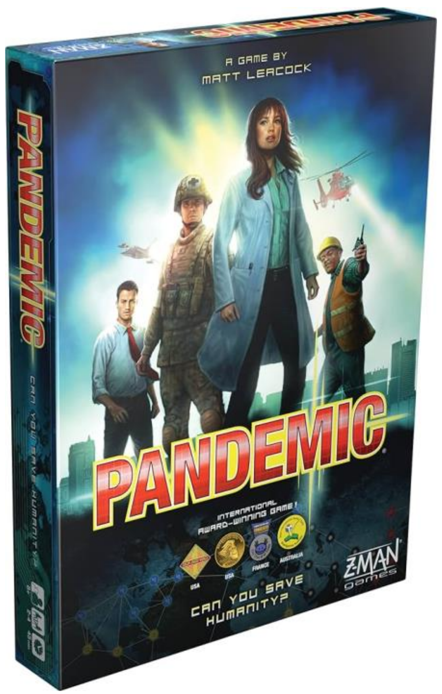
Pure coordination game

	Left	Right
Left	8,8	0,0
Right	0,0	8,8

Assurance game

	Contribute	Defect
Contribute	8,8	0,0
Defect	0,0	5,5

Is this really fully cooperative?



Estimating Human Intent for Physical Human-Robot Co-Manipulation
Townsend et al., 2017

Taxonomy: cooperative vs. competitive

General-sum games (a superset)

Prisoner's dilemma

	Cooperate	**Defect**
Cooperate	5,5	-2,8
Defect	8,-2	0,0

Stag hunt

	Hunt	**Forage**
Hunt	2,2	-2,1
Forage	1,-2	1,1

Bach or Stravinsky

	Bach	**Stravinsky**
Bach	3,2	1,1
Stravinsky	1,1	2,3

Chicken game

	Dare	**Chicken**
Dare	-3,-3	4,-1
Chicken	-1,4	3,3

Taxonomy: cooperative vs. competitive

General-sum games (a superset)

From: Wikipedia

Taxonomy: cooperative vs. competitive

Why does this matter?

- In a two-player zero-sum game, maximizing the rewards is equivalent to minimizing the rewards of the opponent.
  - In a zero-sum game with >2 players, there will be situations where some agents should cooperate.
- In a fully cooperative game, all rewards are shared: if training is centralized, we can have a common critic to train all agents.

Taxonomy: sequence of actions

- Simultaneous: agents decide their actions at the same time.
  - This is usually the case in robotics.
- Hierarchical: some agents have the privilege of seeing others' actions before deciding their actions.
  - This is often used in human-robot collaboration settings, e.g., when one of them takes corrective actions over the other.
- Turn-taking: agents take actions one by one and the state of the system changes after each and every action.
  - Many board games fall under this category.

Taxonomy: symmetric vs asymmetric

Pure coordination game

	Left	Right
Left	8,8	0,0
Right	0,0	8,8

Prisoner's dilemma

	Cooperate	Defect
Cooperate	5,5	-2,8
Defect	8,-2	0,0

Stag hunt

	Hunt	Forage
Hunt	2,2	-2,1
Forage	1,-2	1,1

Assurance game

	Contribute	Defect
Contribute	8,8	0,0
Defect	0,0	5,5

Chicken game

	Dare	Chicken
Dare	-3,-3	4,-1
Chicken	-1,4	3,3

These are all symmetric between the agents.

Taxonomy: symmetric vs asymmetric

In a symmetric multi-agent system,

$$r_1(s, a_1, a_2) = r_2(s, a_2, a_1)$$

which can be extended to any number of agents.

This does not hold in asymmetric systems.

You could also create symmetric vs. asymmetric distinction over the dynamics of the system.

Taxonomy: symmetric vs asymmetric

Why does this matter?

- Symmetry can be exploited for data augmentation.
- It also enables training a single critic in zero-sum systems if training is centralized.

Taxonomy: communication

How do we formulate communication?

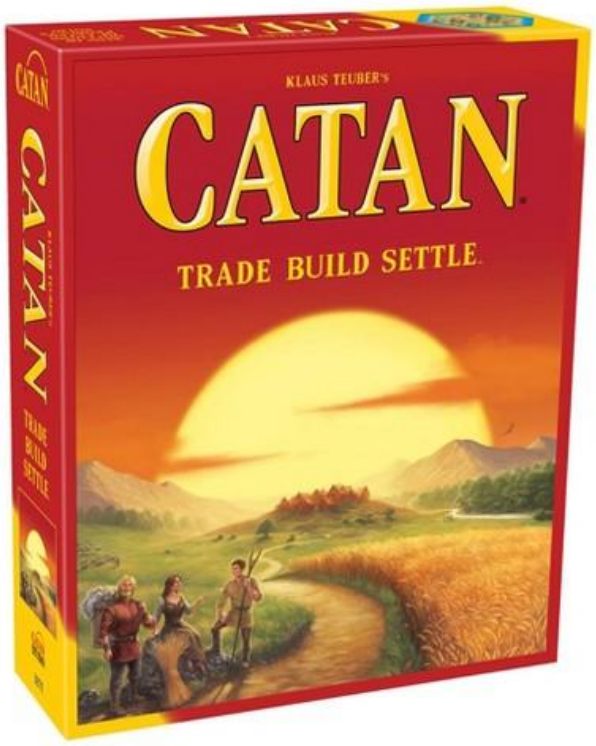
- Communication can be action that affects the observation of (some) other agents.
  - No need to change the formulation.

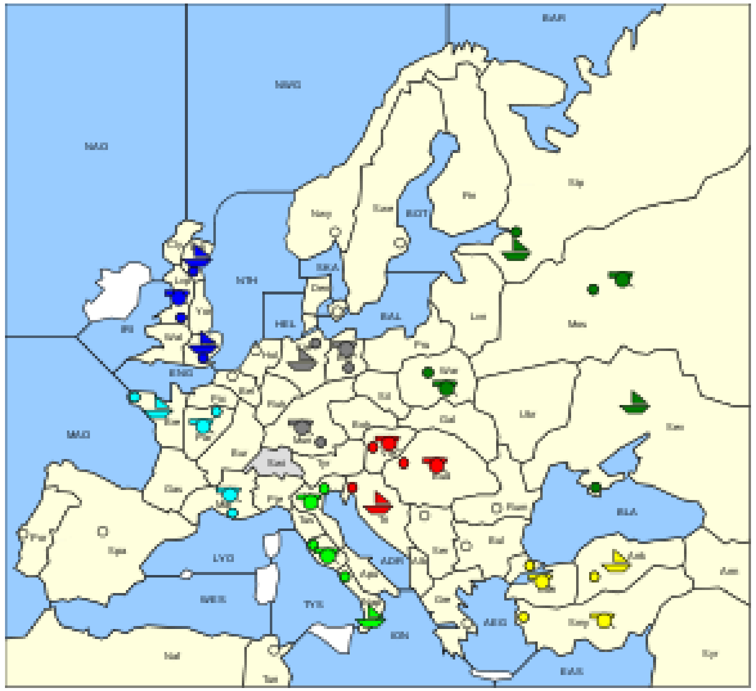
- Communication can be an action that is taken and received after getting the observation but before taking the actual action.
  - Divide each timestep into two. The first is for communication, and the second is for the actual action.

Taxonomy: communication

Is communication relevant in a zero-sum (competitive) game?

Yes!





From: Wikipedia

Taxonomy: communication

Is communication relevant in a setting where all agents get the same observation?

Depends on how the agents were trained.

	Left	**Right**
Left	8,8	0,0
Right	0,0	8,8

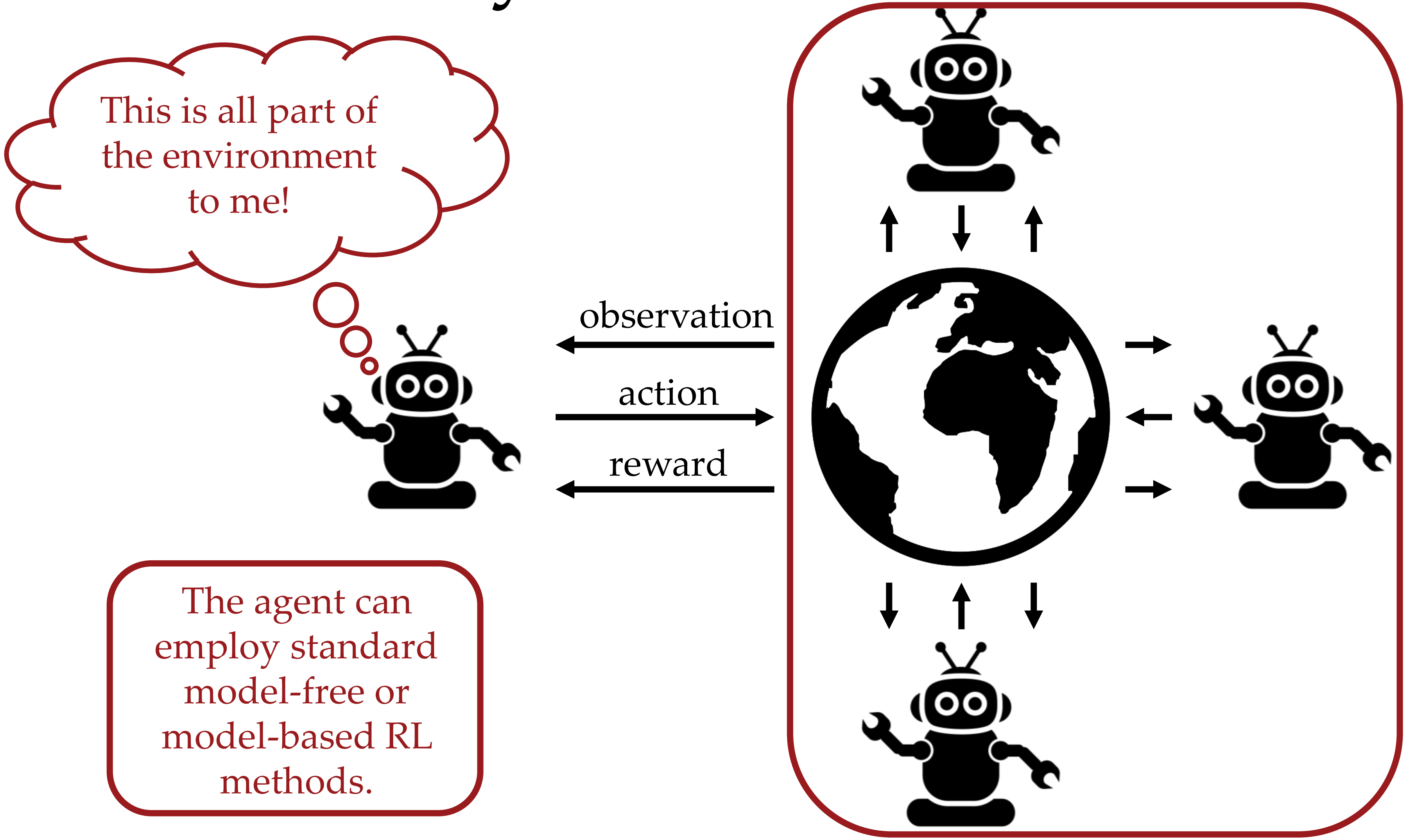
Taxonomy: communication

Unlimited communication in cooperative systems significantly eases the problem: agents can share all information they have. The problem can be reduced to a single-agent problem.

Communication is limited in more interesting research problems:

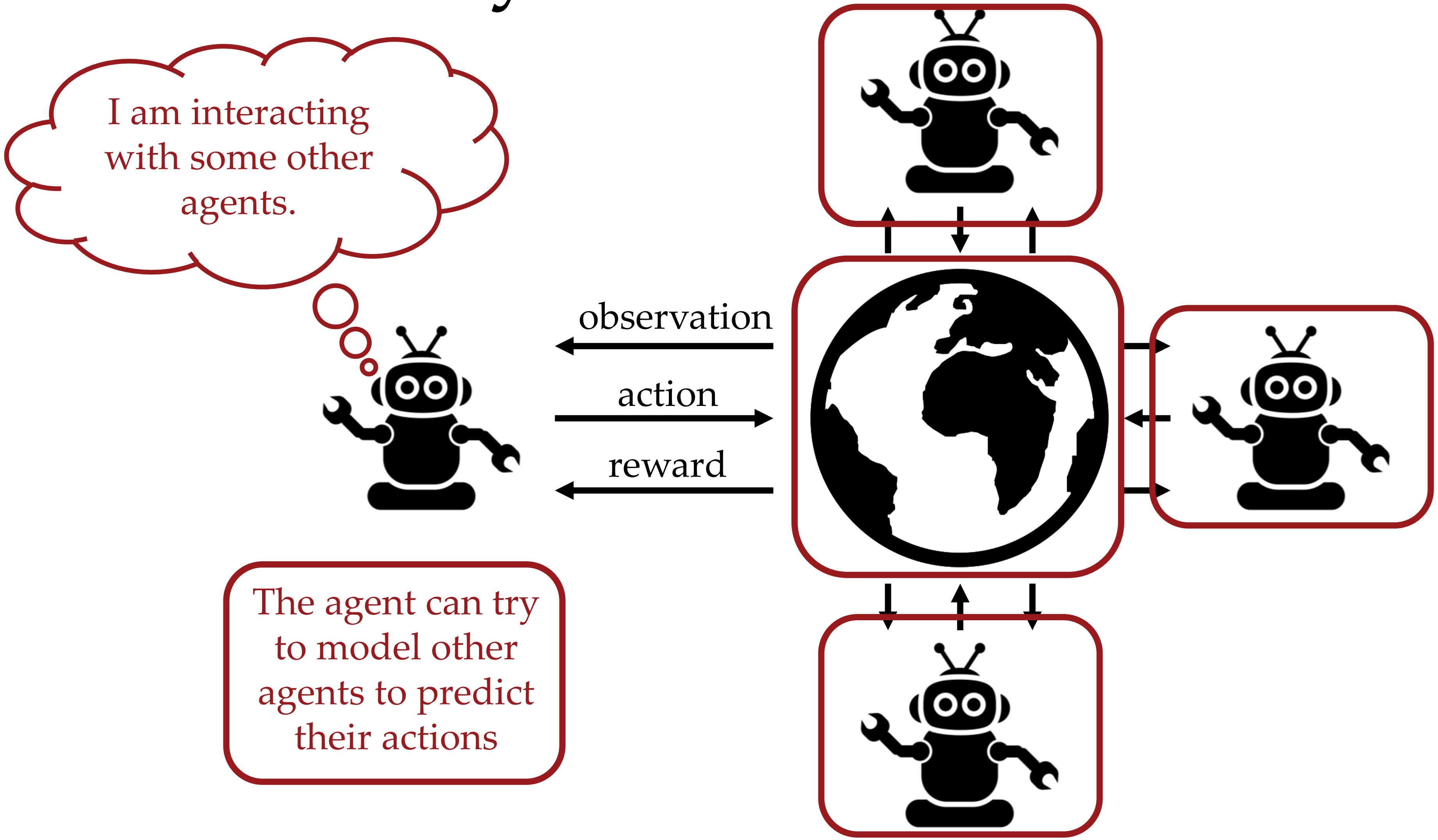
- Information bandwidth (size, noise)
- Receivers

Taxonomy: model-free vs. model-based



The agent can employ standard model-free or model-based RL methods.

Taxonomy: model-free vs. model-based



Taxonomy: model-free vs. model-based

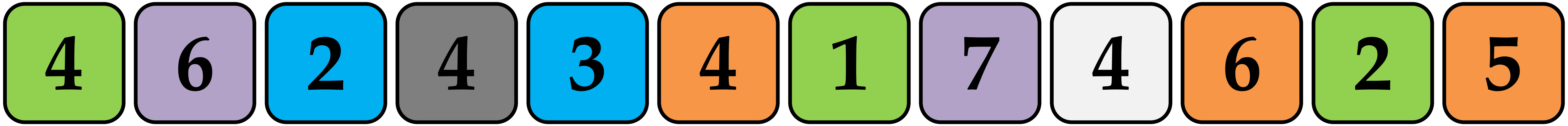
Theory of Mind (ToM)

Theory of mind refers to the capacity to understand other people by ascribing mental states to them.

Humans aren't the only great apes that can 'read minds'
Morell, Science Magazine 2016

Taxonomy: model-free vs. model-based

Theory of Mind (ToM)



All cards are visible to both Alice and Bob. One card is chosen and revealed to Alice. They win the game if Bob can predict the chosen card after Alice gives a single piece of clue (either color or number).

Taxonomy: model-free vs. model-based

Theory of Mind (ToM)

4 6 2 4 3 4 1 7 4 6 2 5

The chosen card is green!

Alice said green. It must be either 4, 1 or 2.

Level 0 ToM

Taxonomy: model-free vs. model-based

Theory of Mind (ToM)

4	6	2	4	3	4	1	7	4	6	2	5

The chosen card is green!

Let's say Alice models me as Level 0 ToM.
Alice would say one to let Level 0 ToM know it is **1**.
So it must be **4** or **2**.

Level 1 ToM

Taxonomy: model-free vs. model-based

Theory of Mind (ToM)

4	6	2	4	3	4	1	7	4	6	2	5

The chosen card is green!

Let's say Alice models me as Level 1 ToM.
- Alice would say one to let Level 0 ToM know it is 1.
- Alice would say three to let Level 0 ToM know it is 3.
- Alice would say blue to let Level 1 ToM know it is 2.
- This means Alice would say two if it was 2.

So it must be 4.

Level 2 ToM

Taxonomy: model-free vs. model-based

Theory of Mind (ToM)

4	6	2	4	3	4	1	7	4	6	2	5

We can keep going to higher levels but prior work has found that humans often do not go beyond level 2.

It is also rare that it will matter for decision making after level 2.

Level 3 ToM

Today…

- Taxonomy

- Fundamental problems (and some solutions)

Credit assignment problem

In a fully cooperative multi-agent system, the agents do not know whose actions cause the high (or low) rewards.

One solution is counterfactual rewards: what would happen if an agent took some baseline action? This is the COMA algorithm.

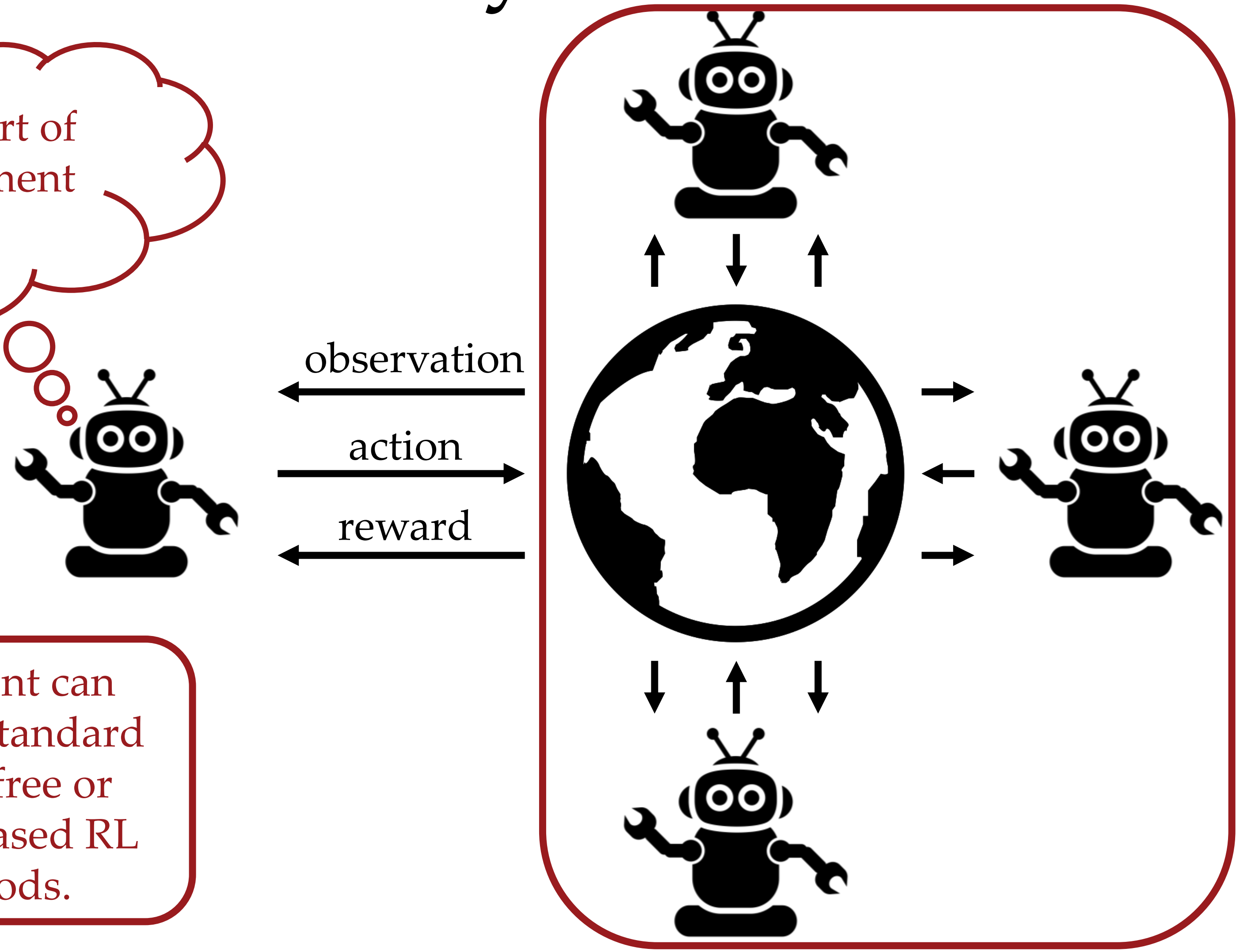
Advantage function in:

- Single-agent RL: $A(s, a) = Q(s, a) - V(s)$
- Multi-agent RL: $A_i(s, \boldsymbol{a}) = Q(s, \boldsymbol{a}) - \mathbb{E}_{a' \sim \pi_i}\left[Q\left(s, (\boldsymbol{a}_{-\boldsymbol{i}}, a')\right)\right]$

Counterfactual multi-agent policy gradients
Foerster et al., AAAI 2018

Non-stationarity

This is all part of the environment to me!



The agent can employ standard model-free or model-based RL methods.

But it is not really an MDP, because other agents are also learning!

Role/convention selection

	Left	Right
Left	**8,8**	0,0
Right	0,0	8,8

	Left	Right
Left	8,8	0,0
Right	0,0	**8,8**

	Left	Right
Left	8,8	0,0
Right	**0,0**	8,8

Role/convention selection

Role/convention selection

This problem does not exist in two-player zero-sum games.

AlphaZero beats all humans in chess, not just the agents it played against during training.

This is because best actions in a two-player zero-sum game do not depend on any conventions.

Today…

- Taxonomy

- Fundamental problems (and some solutions)

Next time…

We have only 2 lectures left!

November 10^{th}	November 17^{th}	November 24^{th}	December 1^{st}
Veterans' day (no lecture)	Class presentations on multi-agent learning and RoboNLP	Thanksgiving (no lecture)	Project presentations

Project presentations

- For each project: 7 minutes of presentation + 2 minutes of Q&A
- There is no requirement that all members must present.
  - But at least one member must present – there is no online option!

- No presentation template – be creative! :)
- Show the audience that you worked on something interesting.
  - Share your great results.
  - …or your insights about why you got negative results.

Peer review

- Each project will get 2 reviews. We will announce the matchings via email.
- Each review should be no longer than 1 page.
- Do NOT write your name in the review – we will share your review with the project members.

- The review you write will only affect your grade and not the members of the project. So you should be as critical as possible.