Sequential Decision Making for Cooperative Agents

Part II: Multiagent Decision-Making under Uncertainty

Stefan J. Witwicki
Robotic Systems Laboratory (LSRO)
École Polytechnique Fédérale de Lausanne

João V. Messias
Institute for Systems and Robotics (ISR)
IST, Universidade de Lisboa

EASSS-2014

July 15, 2014

Multiagent Decision-Making

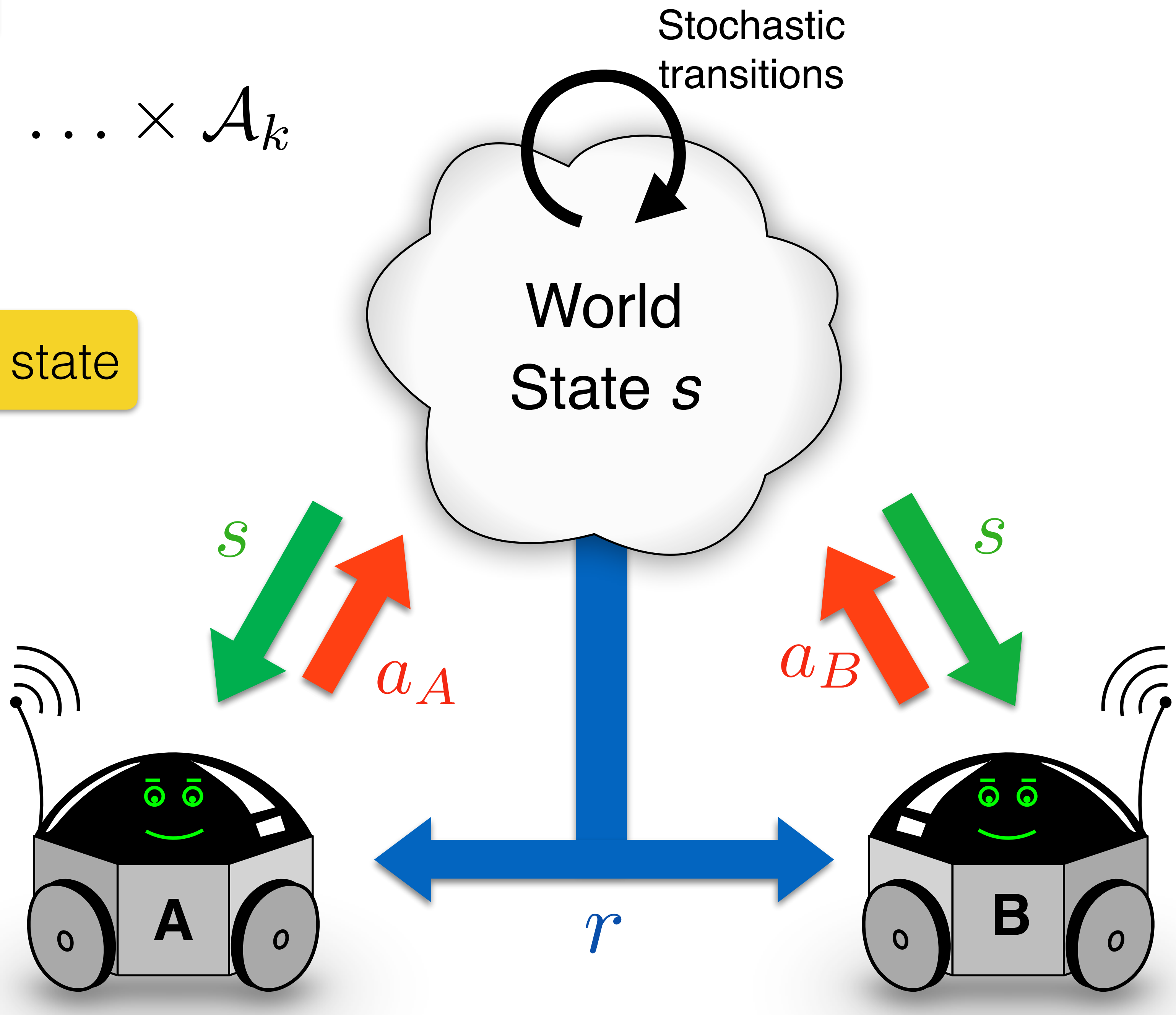
Multiagent MDPs

Distributed Actions

$$\mathcal{A} = \mathcal{A}_1 \times \mathcal{A}_2 \times \ldots \times \mathcal{A}_k$$

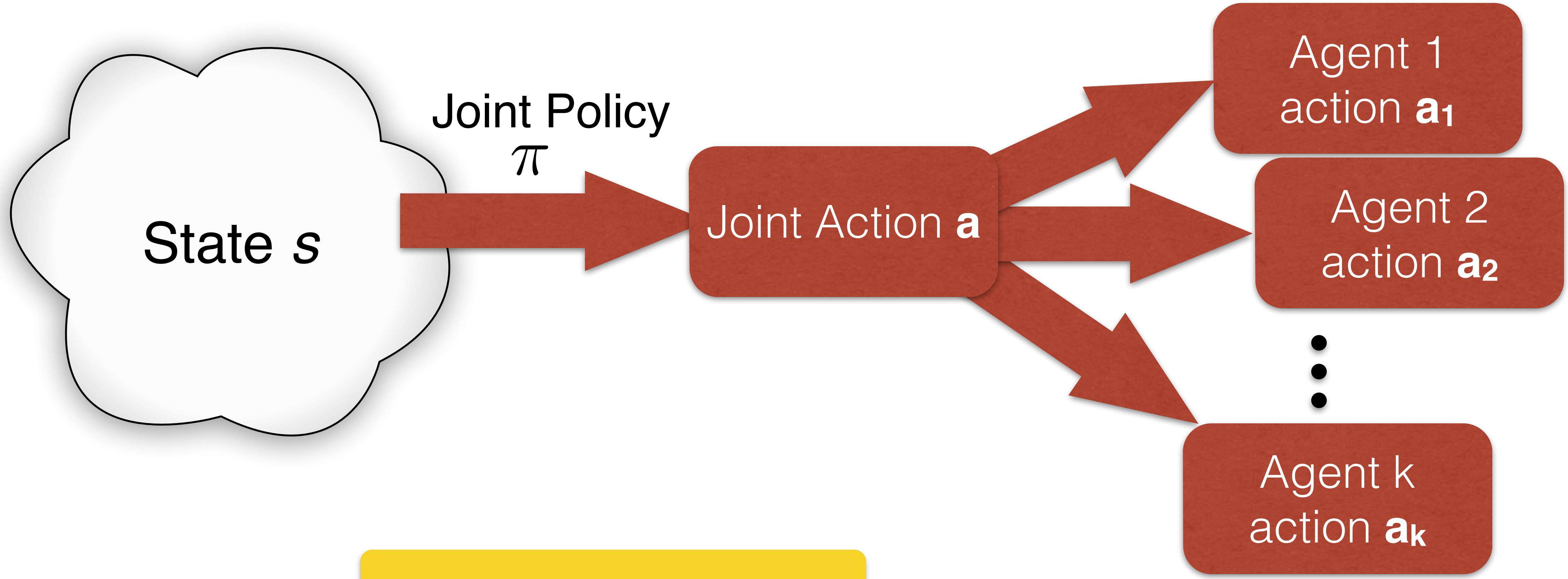
All agents observe the state

Shared reward



Joint Policies

We now need to find a **joint** policy
that specifies actions for all agents:



Centralized Planning

MMDP Solutions

Single-agent MDP solution methods are applicable
(Value Iteration, Policy Iteration, RL, etc.)

But is this scalable?

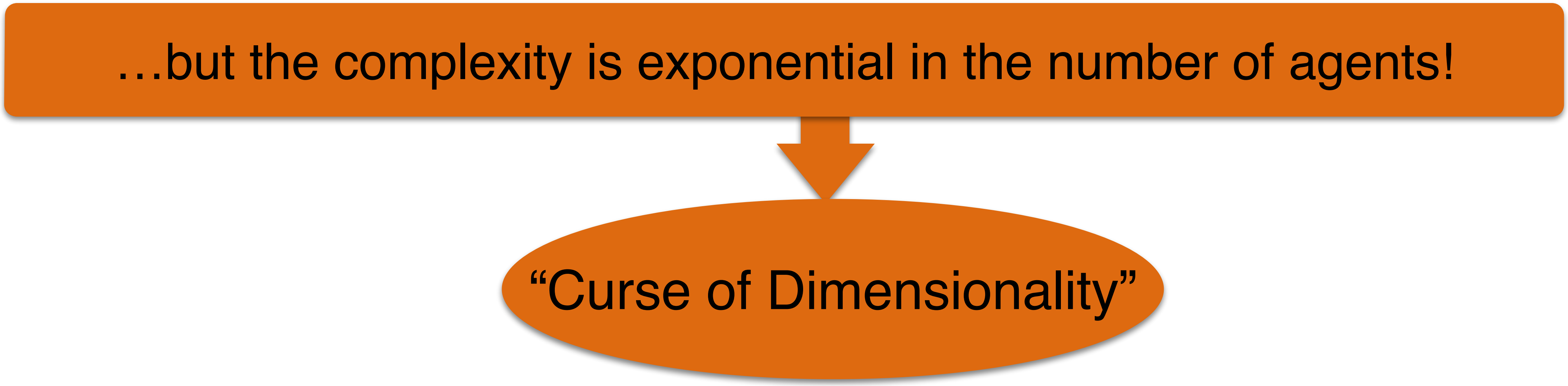
Consider an example Multiagent MDP with:

- 4 actions per agent (e.g. N/S/E/W) $|\mathcal{A}_i| = 4$
- 1 state factor with 8 states per agent $|\mathcal{X}_i| = 8$

#Agents	#Actions	#States
2	16	64
4	256	4096
8	65536	~16.7 million
16	~4.3 billion	~280 trillion

MMDP Solutions

MDPs are P-complete…



…but the complexity is exponential in the number of agents!

“Curse of Dimensionality”

More advanced solution methods:

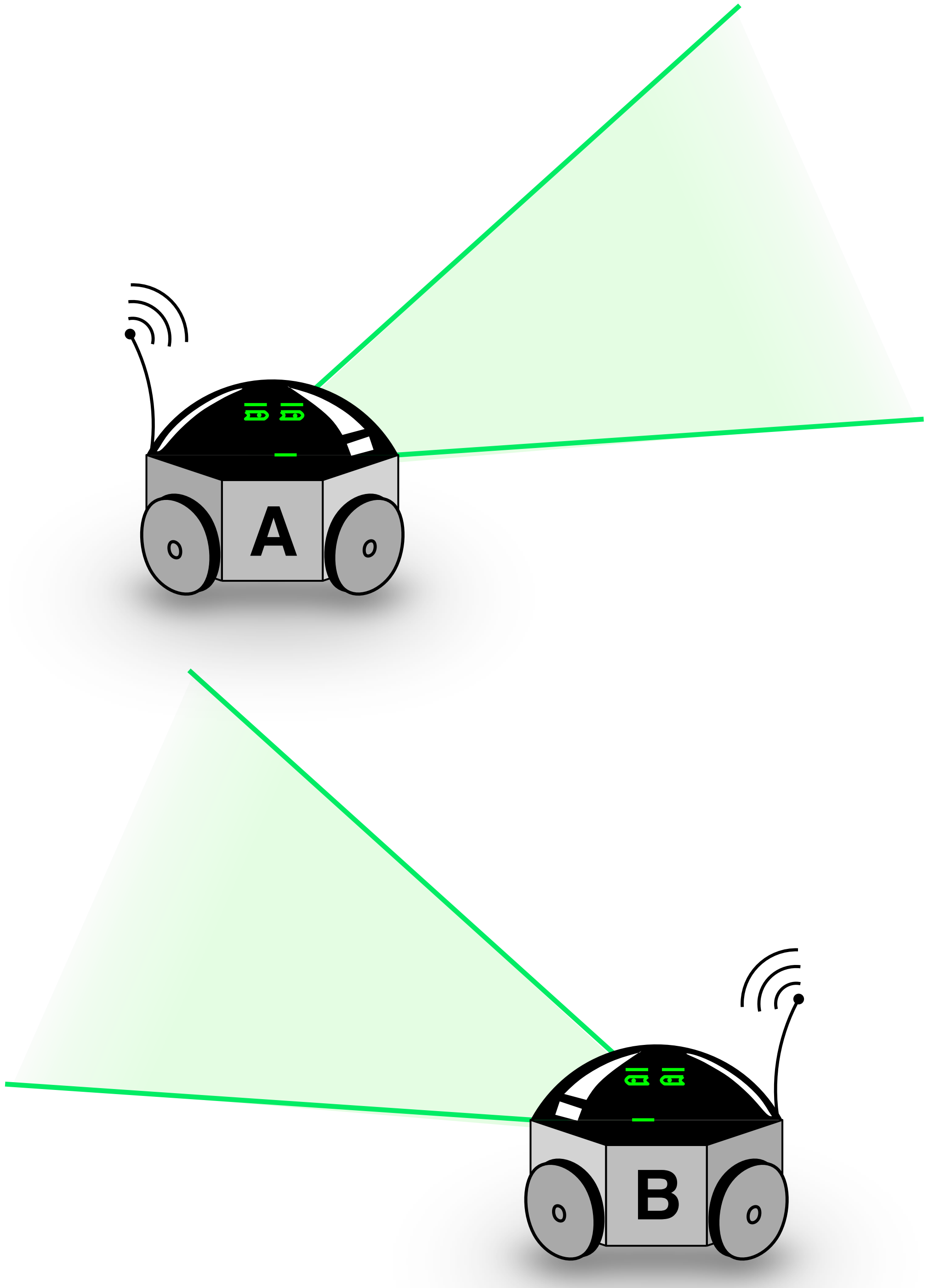
- By exploiting the problem structure (SPUDD)
- Monte-Carlo planning methods (UCT)

Multiagent POMDPs

It is rarely the case that a single agent can observe the whole system.

The partial information of each agent needs to be taken into account

Communication makes a huge difference!



Multiagent POMDPs (MPOMDPs)

Distribution Actions

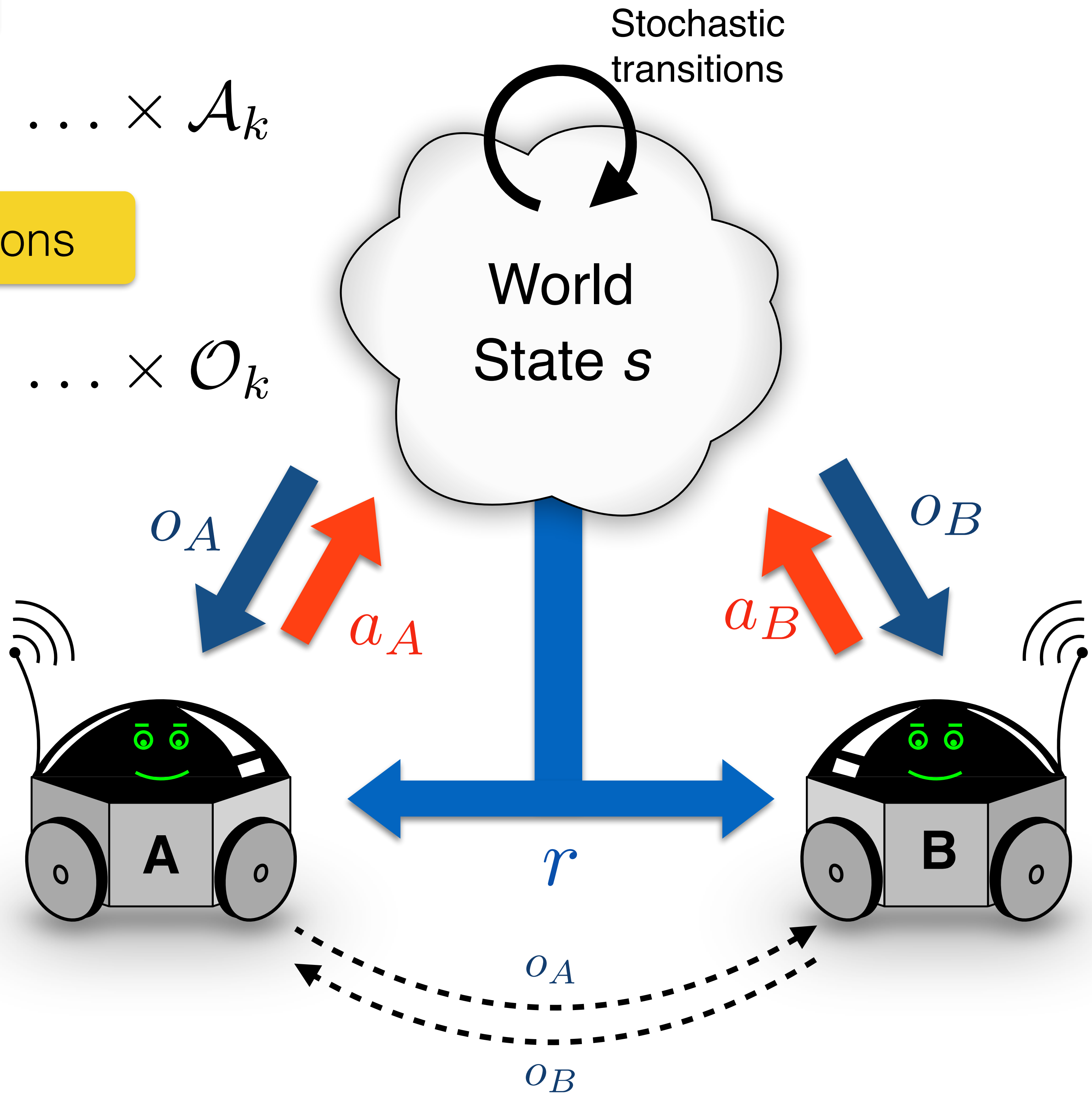
$$\mathcal{A} = \mathcal{A}_1 \times \mathcal{A}_2 \times \ldots \times \mathcal{A}_k$$

Distributed Observations

$$\mathcal{O} = \mathcal{O}_1 \times \mathcal{O}_2 \times \ldots \times \mathcal{O}_k$$

Shared reward

Free communication

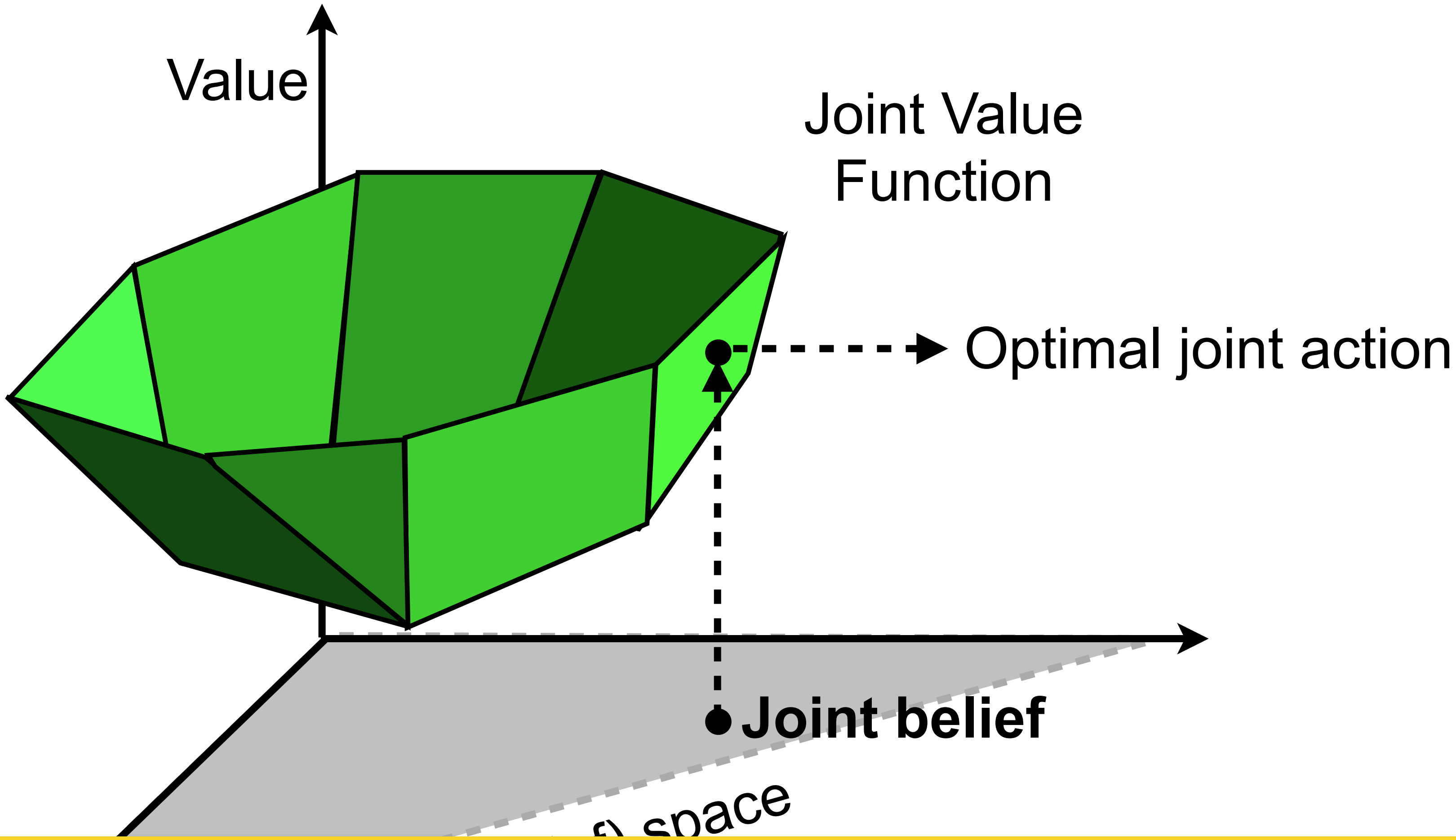


Joint Belief States

Each agent knows the *joint* observation

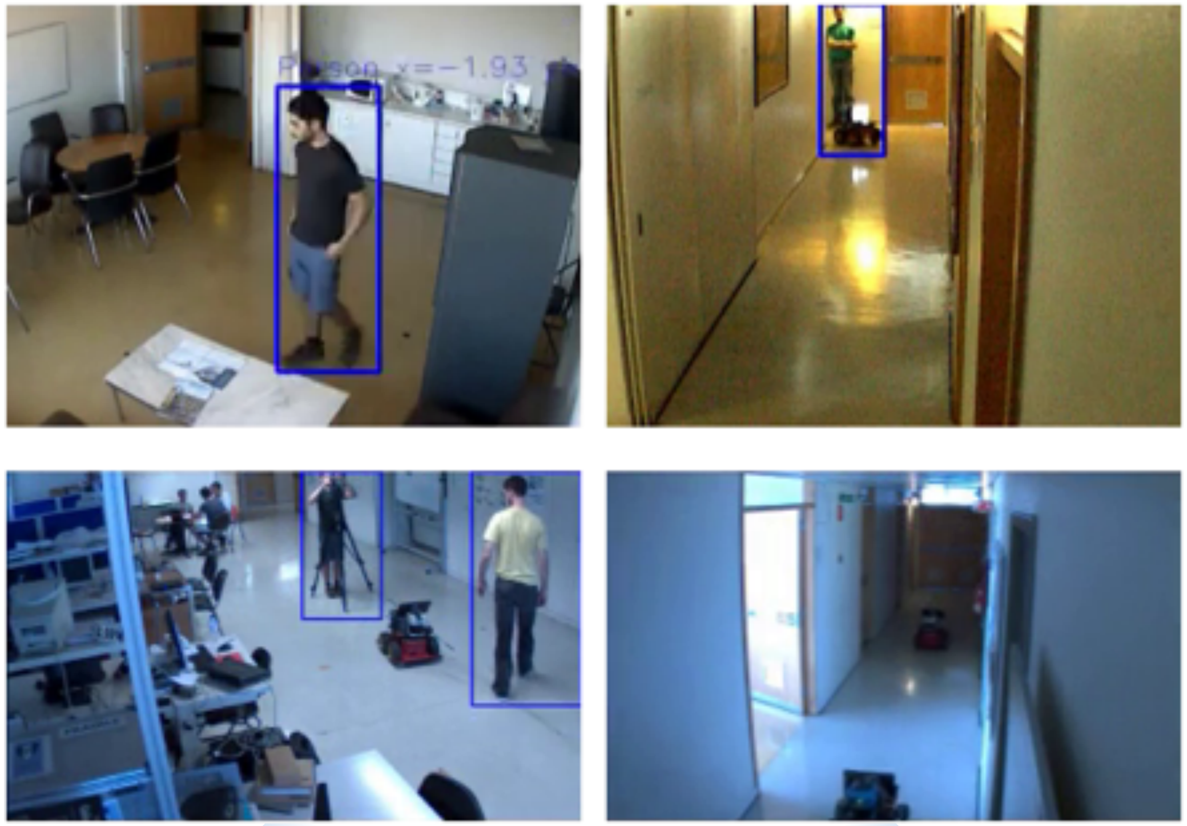
$$\mathbf{o} = \langle o_1, o_2, \ldots, o_k \rangle$$

It is possible to maintain (and update) a *joint* belief state

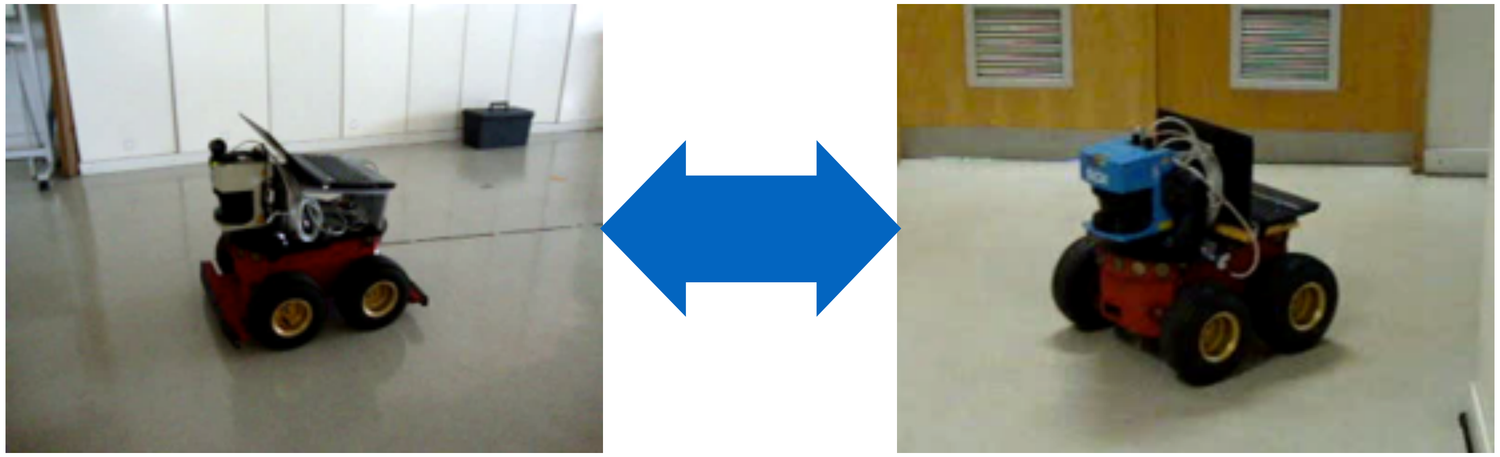


Essentially one big POMDP with many actions and observations!

MPOMDP Example



Automatic Event Detection
(visitors, intruders, emergencies)



Cooperative
Event Response

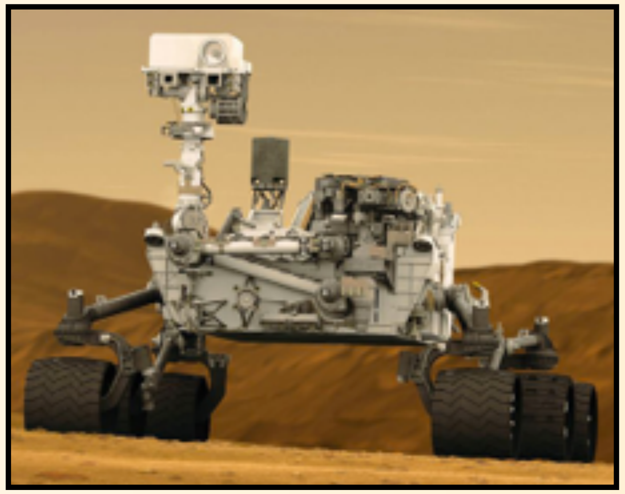
MPOMDP Example

Communication Reduction

MPOMDPs assume **perfect** communication!

Bandwidth can be limited;

Communication can expend energy;



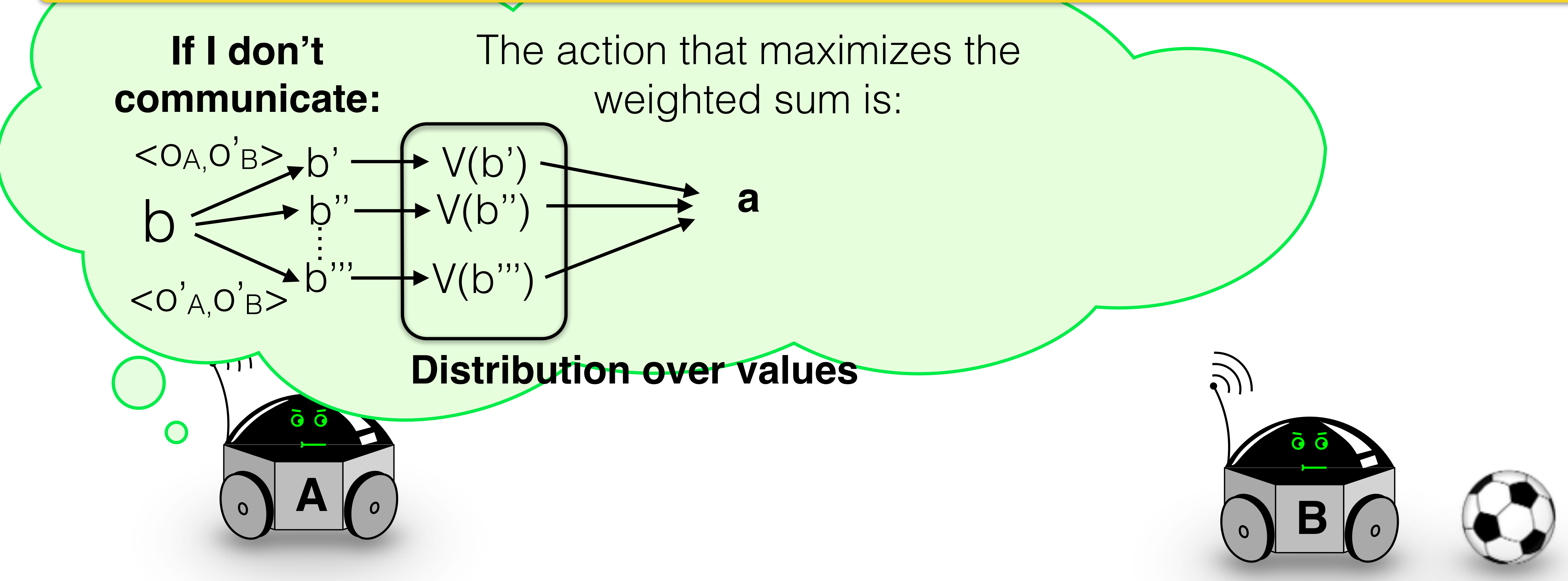
Communication can carry sensitive information.

OK, so what can we do about it?

Common trick:

1. Plan with free communication
2. Try to minimize comms. during execution

Approach 1: Reason about possible joint beliefs during execution (Roth et. al, ’05)



If I don’t communicate:

The action that maximizes the weighted sum is:

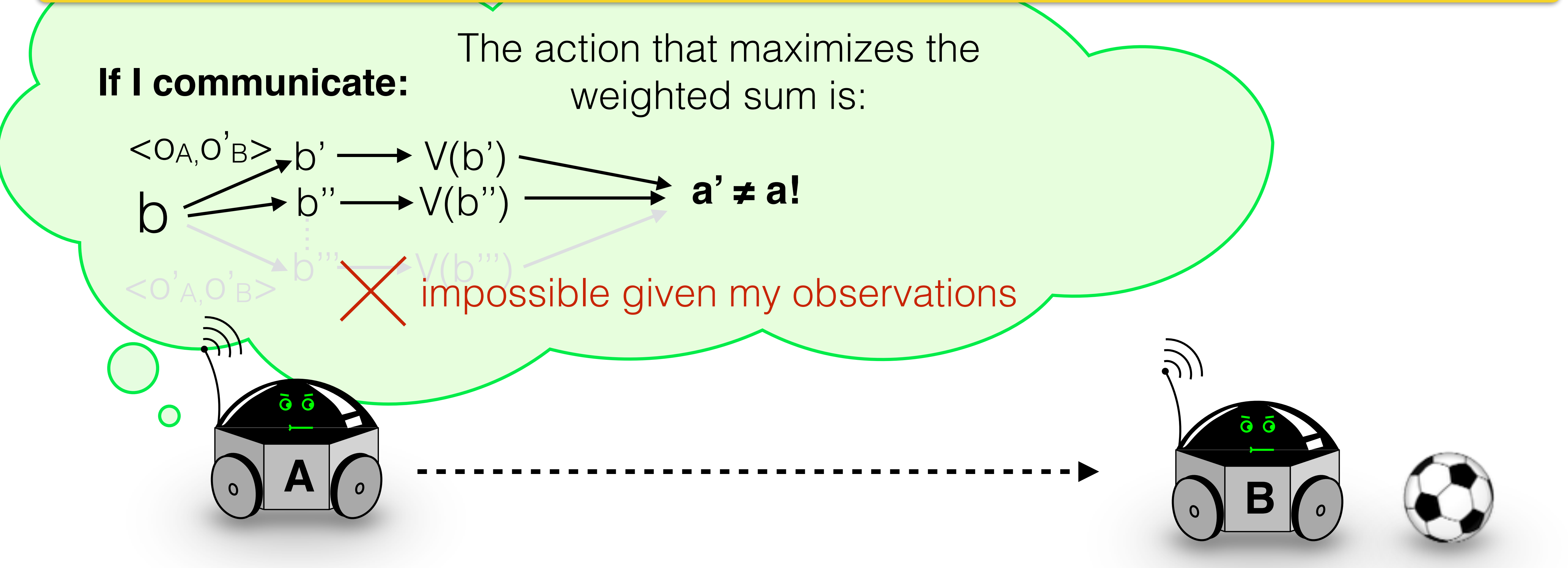
Distribution over values

OK, so what can we do about it?

Common trick:

1. Plan with free communication
2. Try to minimize comms. during execution

Approach 1: Reason about possible joint beliefs during execution (Roth et. al, '05)



If I communicate:

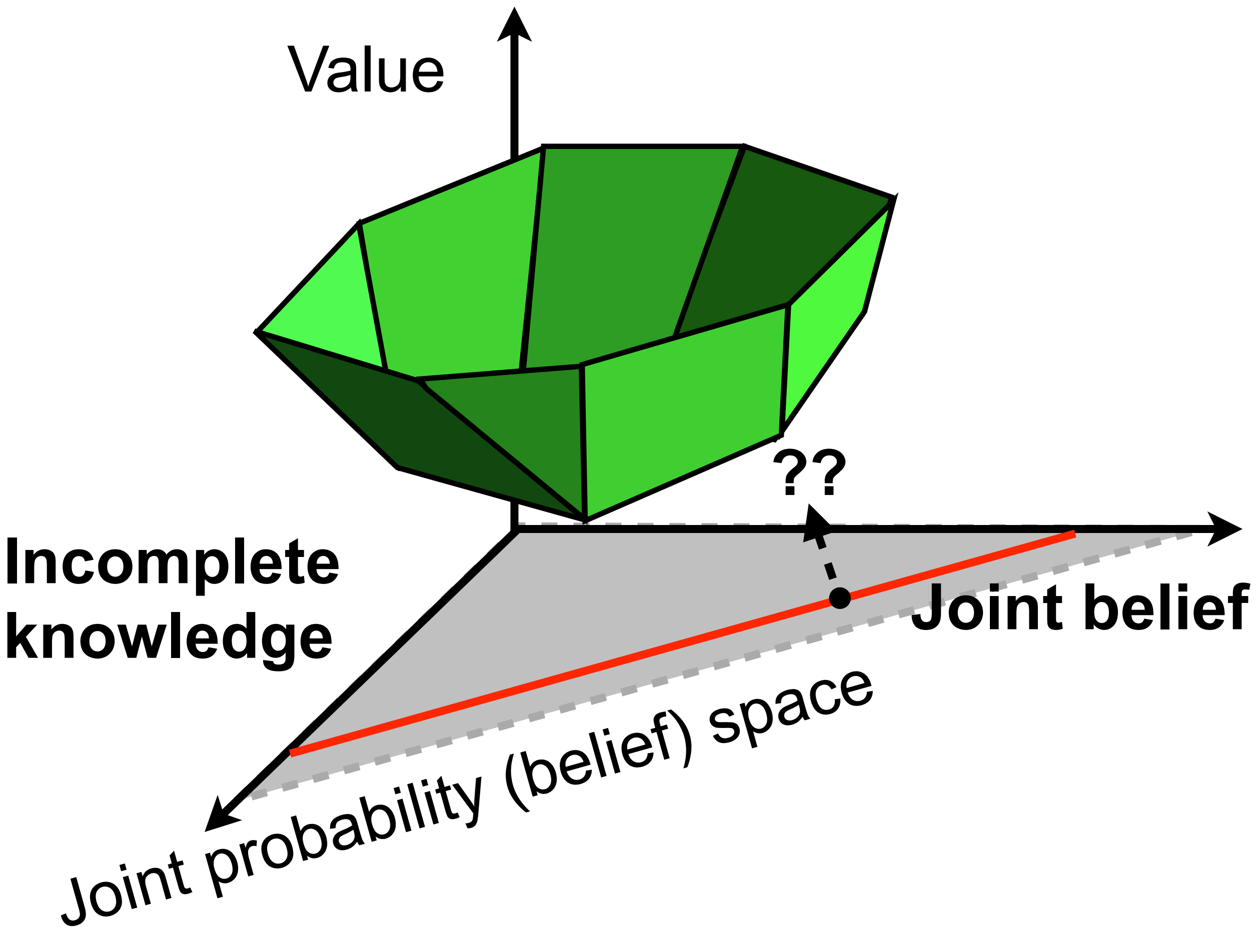
The action that maximizes the weighted sum is:

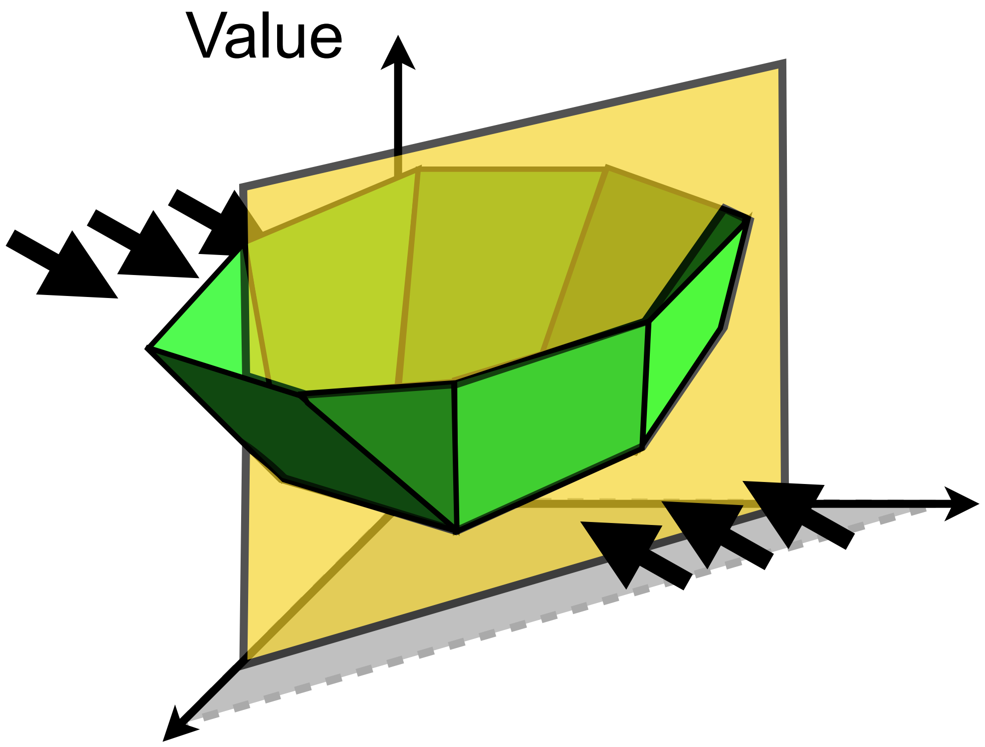
a' ≠ a!

impossible given my observations

Communication Reduction

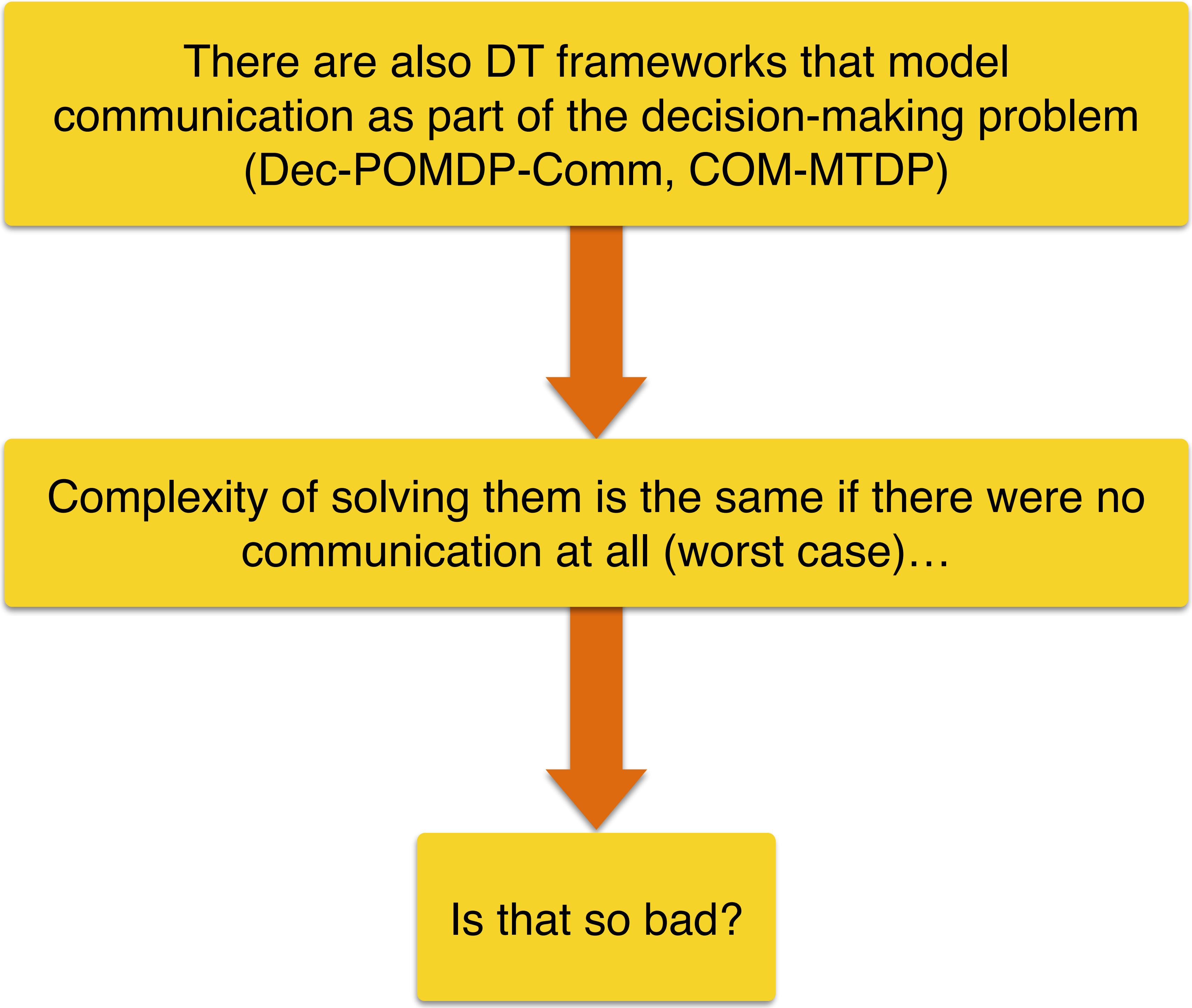
Approach 2: Reason about local knowledge before execution (Messias et. al, ’11)





try to get a local policy, by analyzing the joint value function from the perspective of each agent.

Communication Reduction



There are also DT frameworks that model communication as part of the decision-making problem (Dec-POMDP-Comm, COM-MTDP)

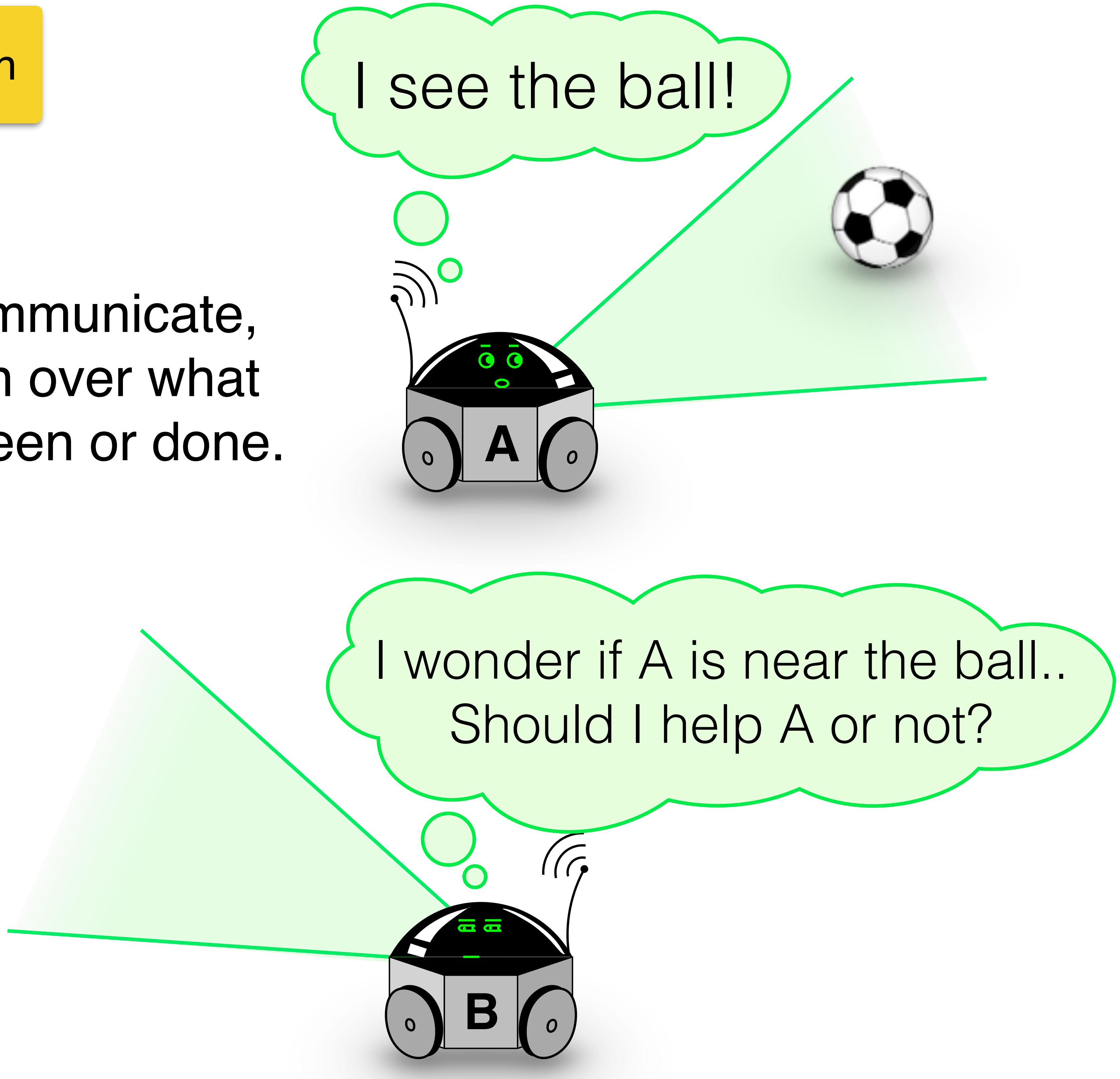
Complexity of solving them is the same if there were no communication at all (worst case)…

Is that so bad?

Non-communicative Decision-Making

Without **explicit** communication

When agents cannot communicate, they are forced to reason over what other agents **may** have seen or done.



Decentralized POMDPs (Dec-POMDPs)

Distributed Actions

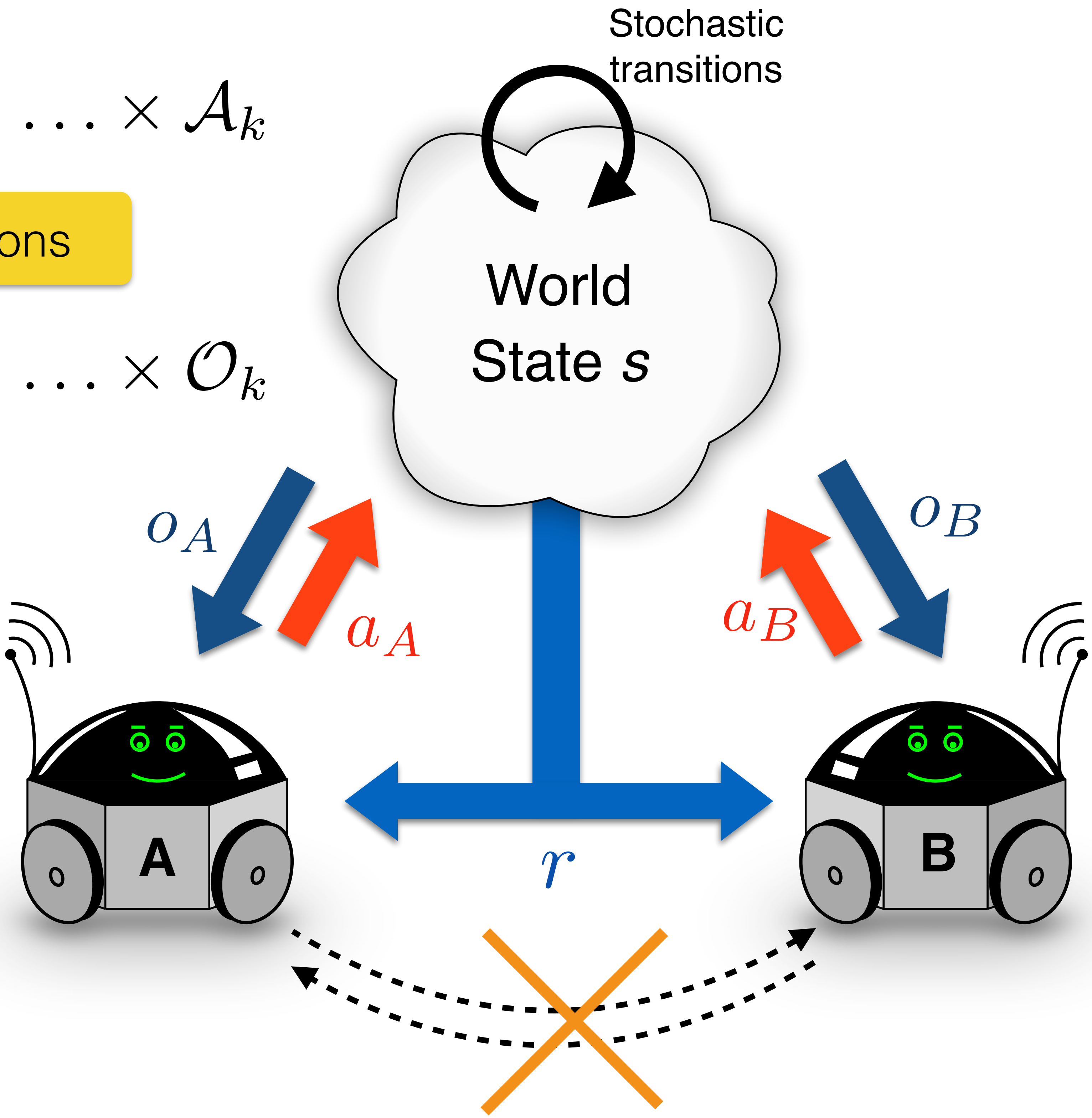
$$\mathcal{A} = \mathcal{A}_1 \times \mathcal{A}_2 \times \ldots \times \mathcal{A}_k$$

Distributed Observations

$$\mathcal{O} = \mathcal{O}_1 \times \mathcal{O}_2 \times \ldots \times \mathcal{O}_k$$

Shared reward

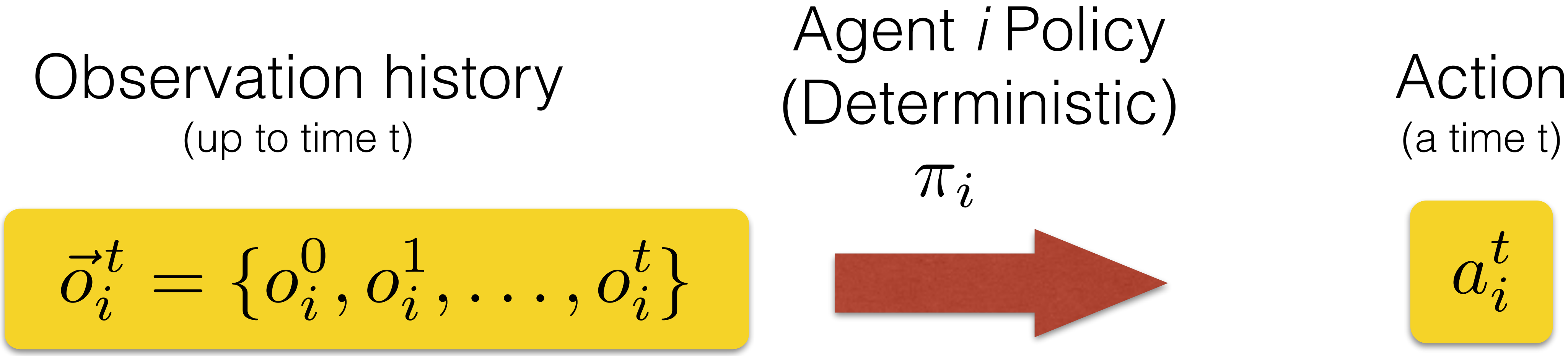
No communication



Dec-POMDP Policies

In a Dec-POMDP, it is no longer possible to maintain a joint belief!

Each agent must consider all of its previous observations…



Observation history

(up to time t)

$\vec{o}_i^{\,t} = \{o_i^0, o_i^1, \ldots, o_i^t\}$

Agent *i* Policy

(Deterministic)

π_i

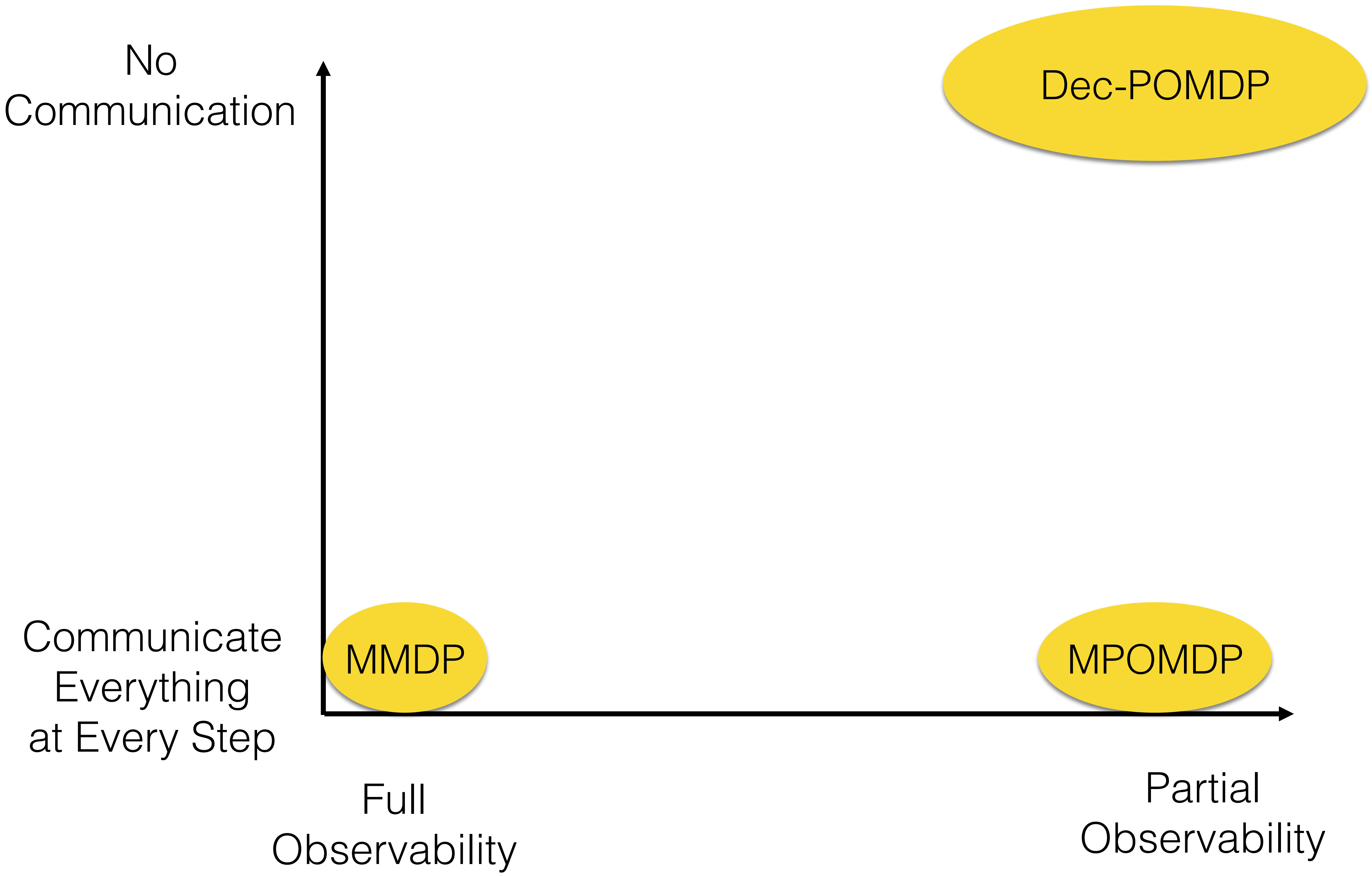
Action

(a time t)

a_i^t

$|\mathcal{A}_i|^{\frac{|\mathcal{O}_i|^h - 1}{|\mathcal{O}_i| - 1}}$ possible policies…

Properties of the discussed multiagent frameworks


No Communication
Dec-POMDP
Communicate Everything at Every Step
MMDP
MPOMDP
Full Observability
Partial Observability