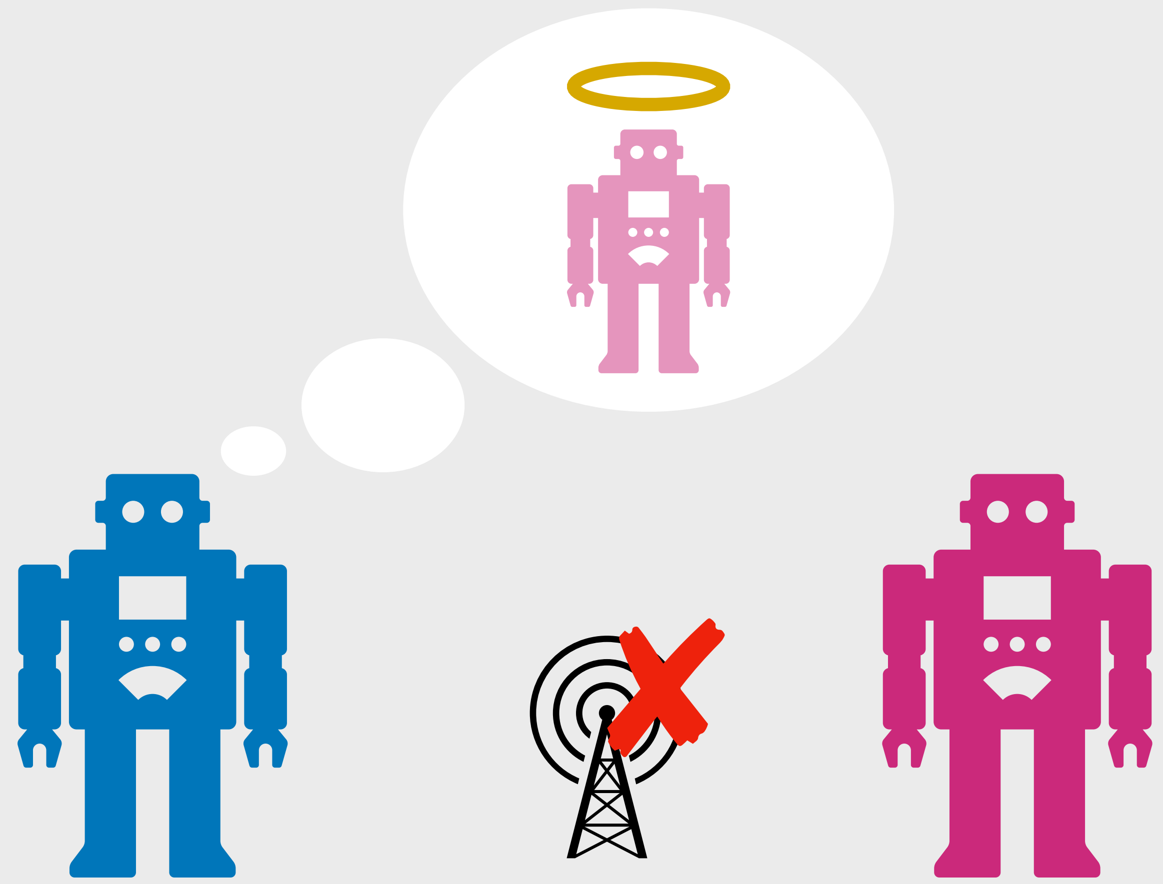
Planning Not to Talk:
Multiagent Systems that are Robust to Communication Loss

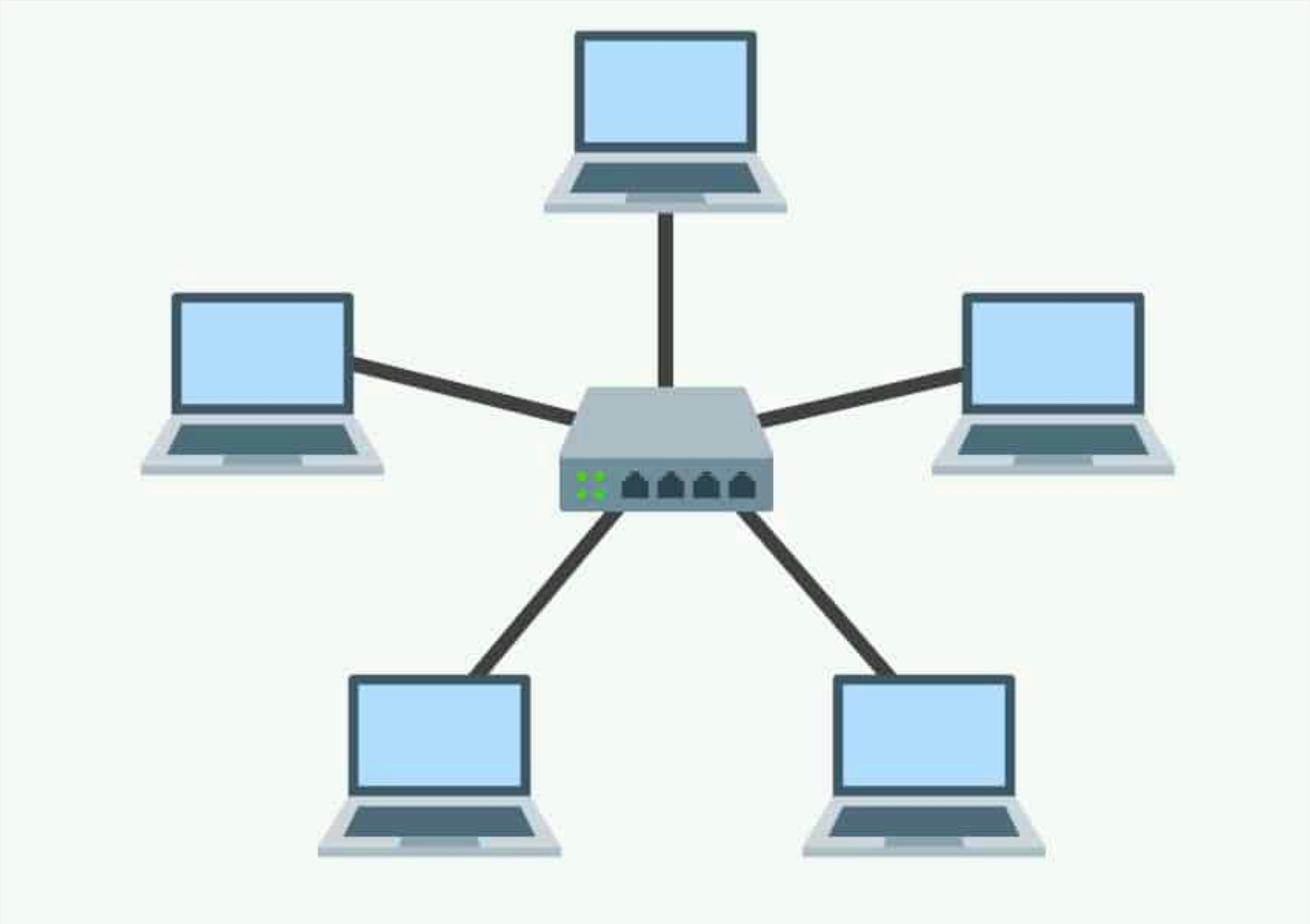


Mustafa O. Karabag & Cyrus Neary

TEXAS
The University of Texas at Austin

autonomous
SYSTEMS GROUP

The need of communication robust strategies in multi-agent systems

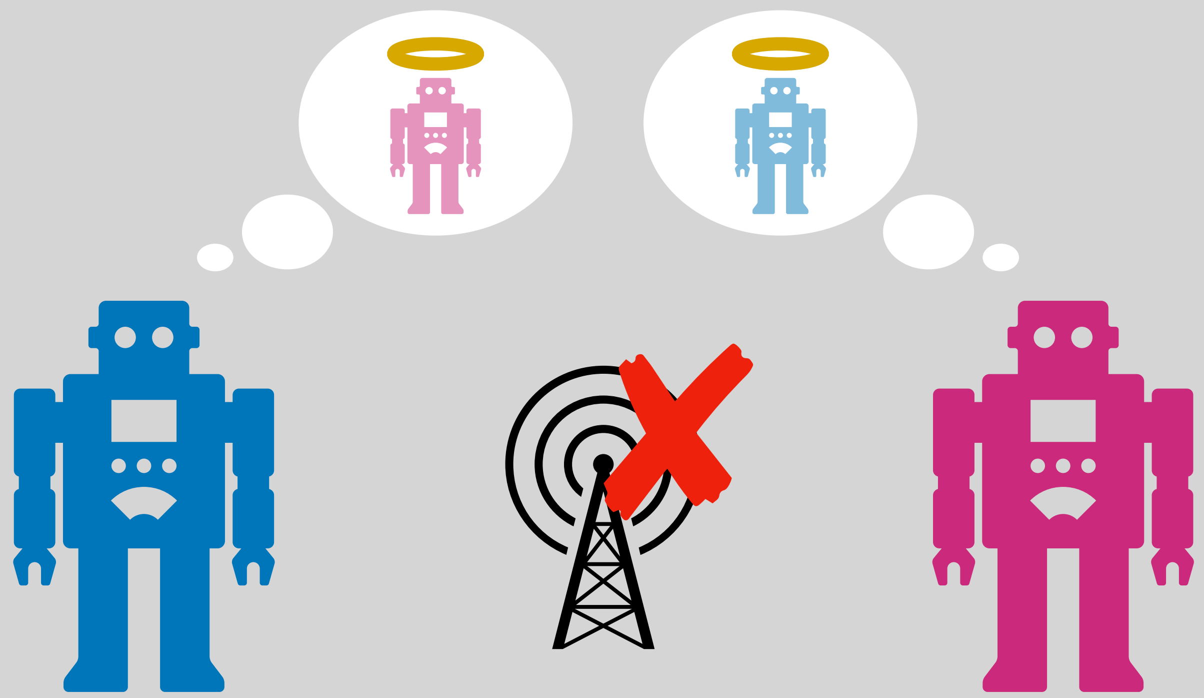


High success is achievable with high dependencies,
but communication loss leads to catastrophe.

Need high success with low dependencies!

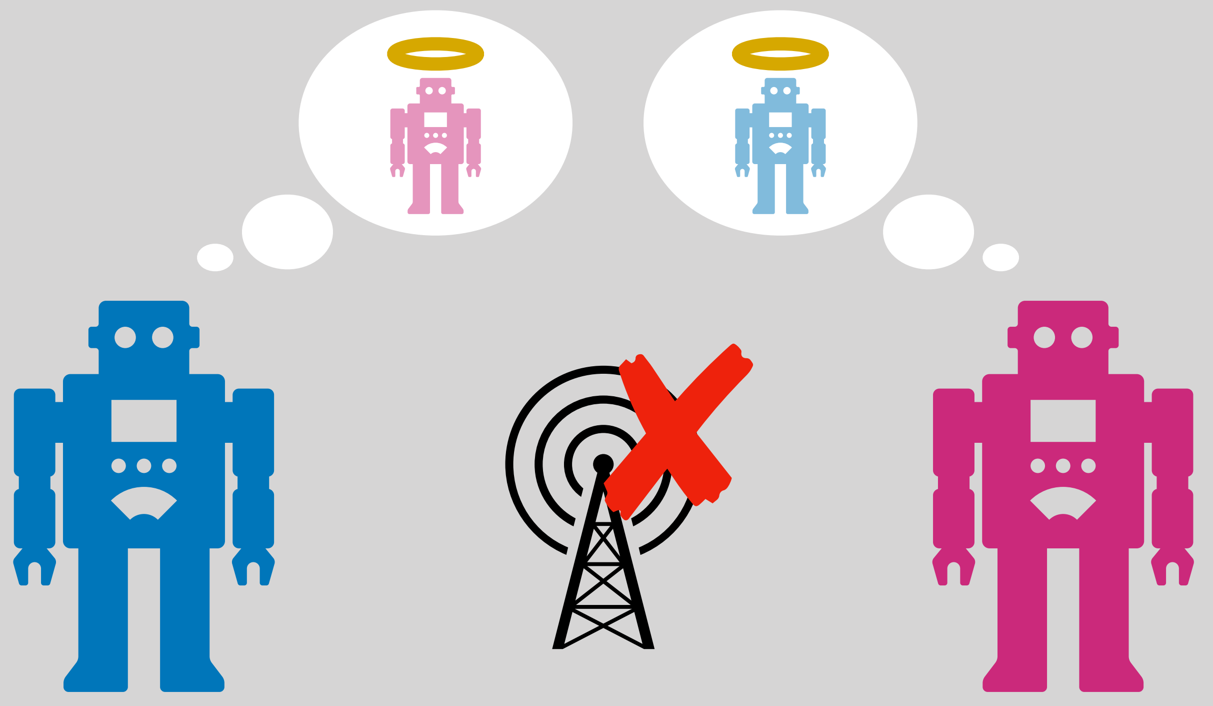
By the end of this talk

(De)centralized Policy Execution



By the end of this talk

(De)centralized Policy Execution

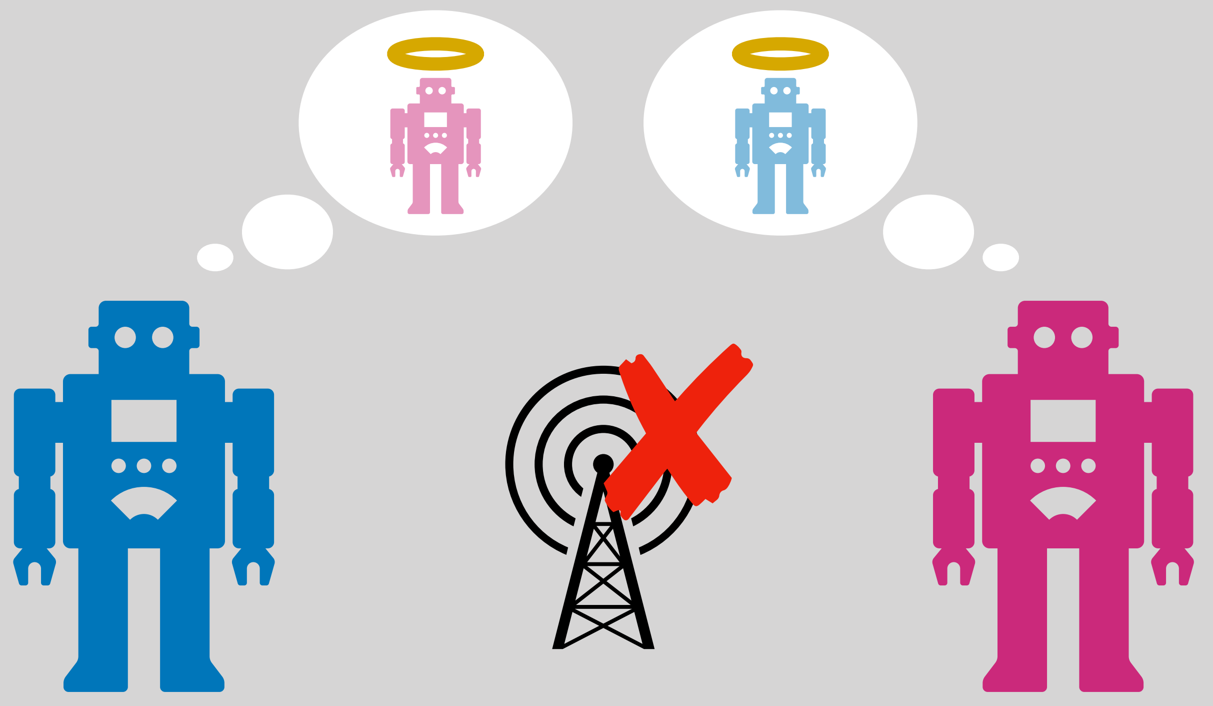


Performance Guarantees under Communication Loss

$$\text{Value}^{full} - \text{Value}^{loss} \geq g(\text{Dependencies})$$

By the end of this talk

(De)centralized Policy Execution



Performance Guarantees under Communication Loss

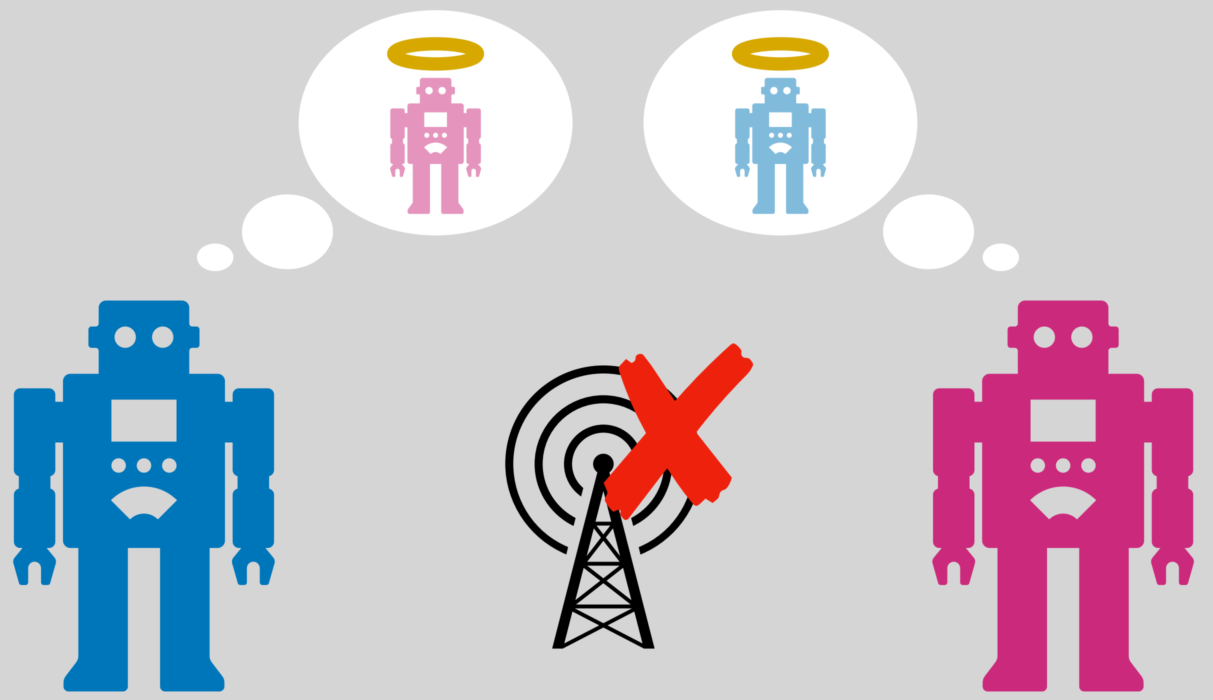
$$\text{Value}^{full} - \text{Value}^{loss} \geq g(\text{Dependencies})$$

Policy Optimization for Communication Loss

$$\max_{\pi} \text{Value}^{loss}$$

By the end of this talk

(De)centralized Policy Execution



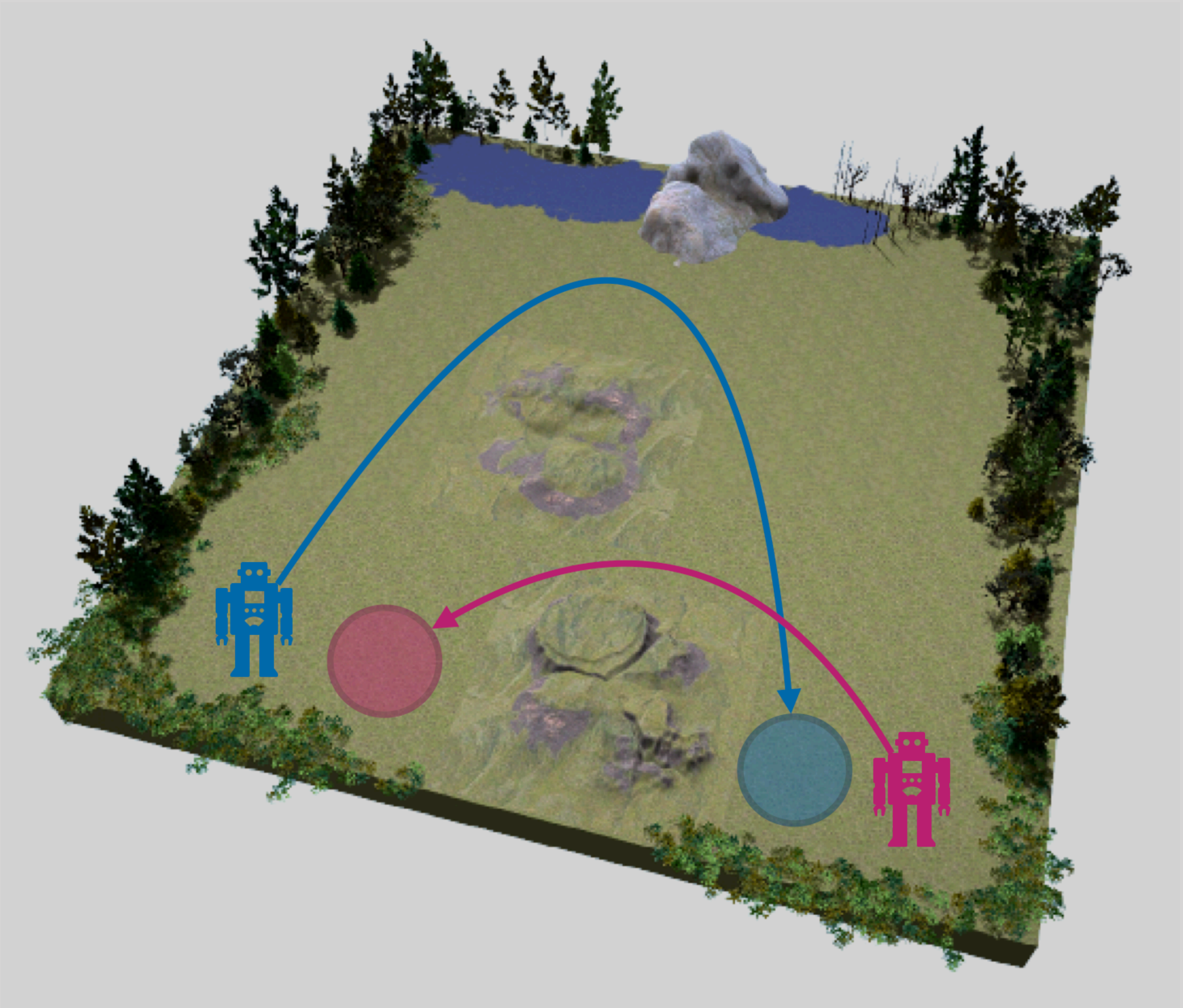
Performance Guarantees under Communication Loss

$$\text{Value}^{full} - \text{Value}^{loss} \geq g(\text{Dependencies})$$

Policy Optimization for Communication Loss

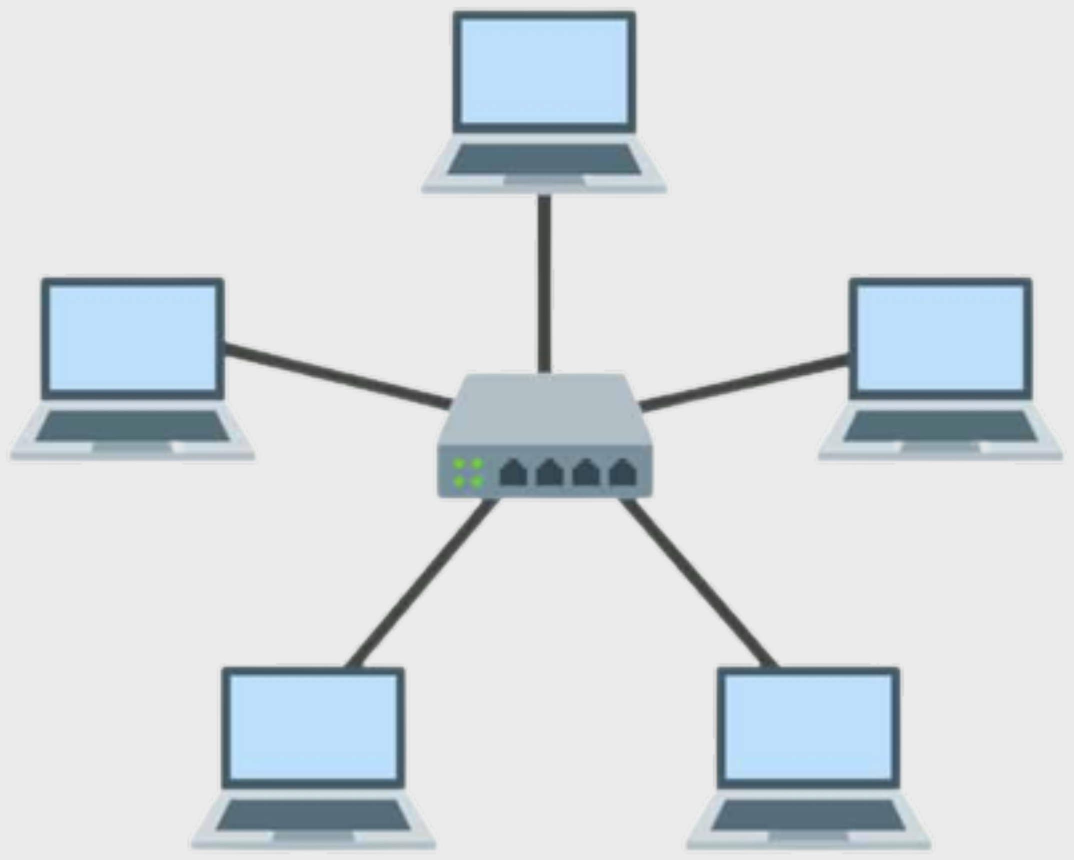
$$\max_{\pi} \text{Value}^{loss}$$

Minimally Dependent Behavior



Modeling of multi-agent systems

N agents with independent dynamics.



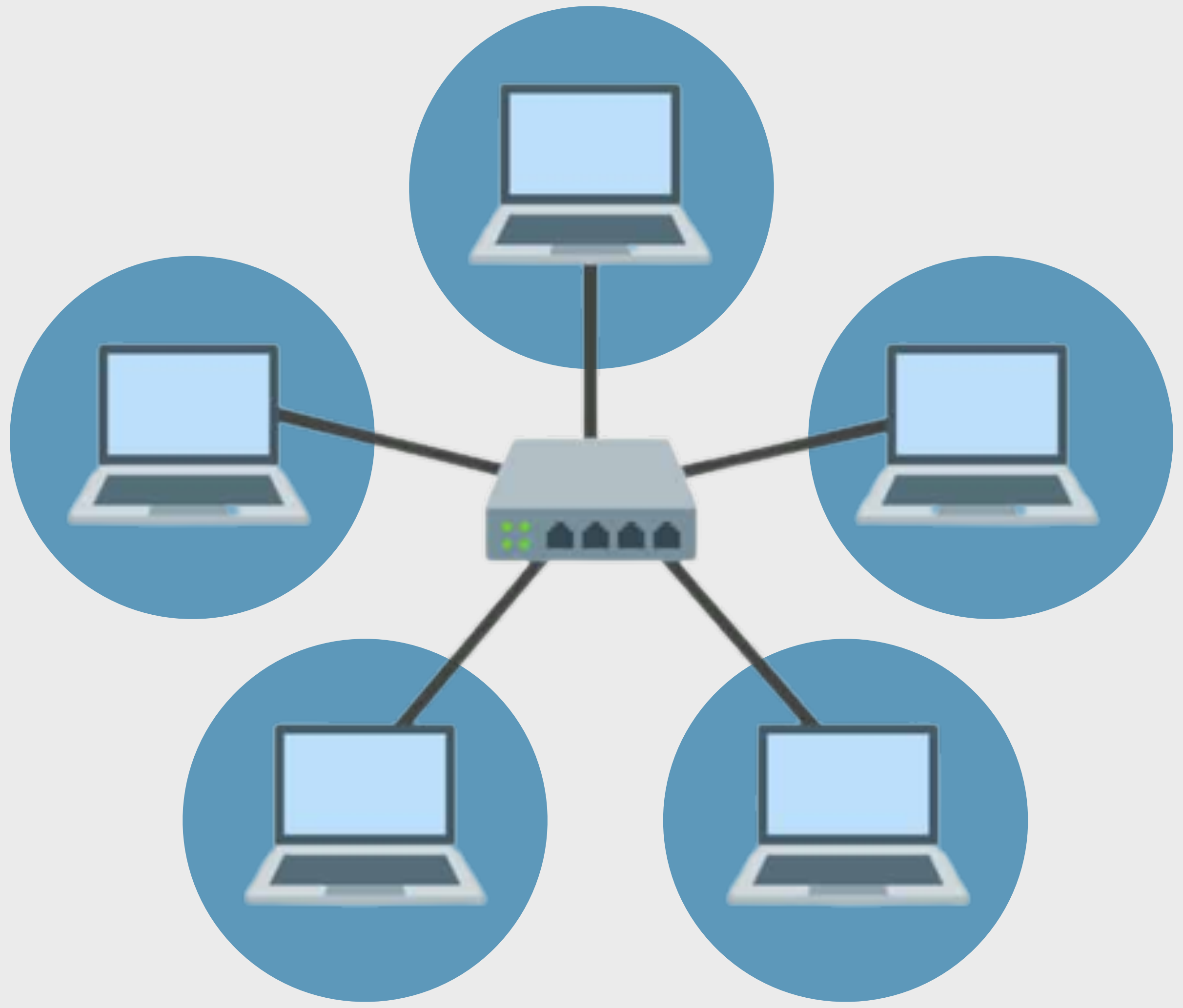
Modeling of multi-agent systems

N agents with independent dynamics.

Markov decision process (MDP) for Agent i

$$\mathscr{M}^i = (S^i, A^i, P^i, s_0^i)$$

- S^i: States
- A^i: Actions
- P^i: Transition probability function
- s_0^i: Initial state



Modeling of multi-agent systems

N agents with independent dynamics.

Markov decision process (MDP) for Agent i

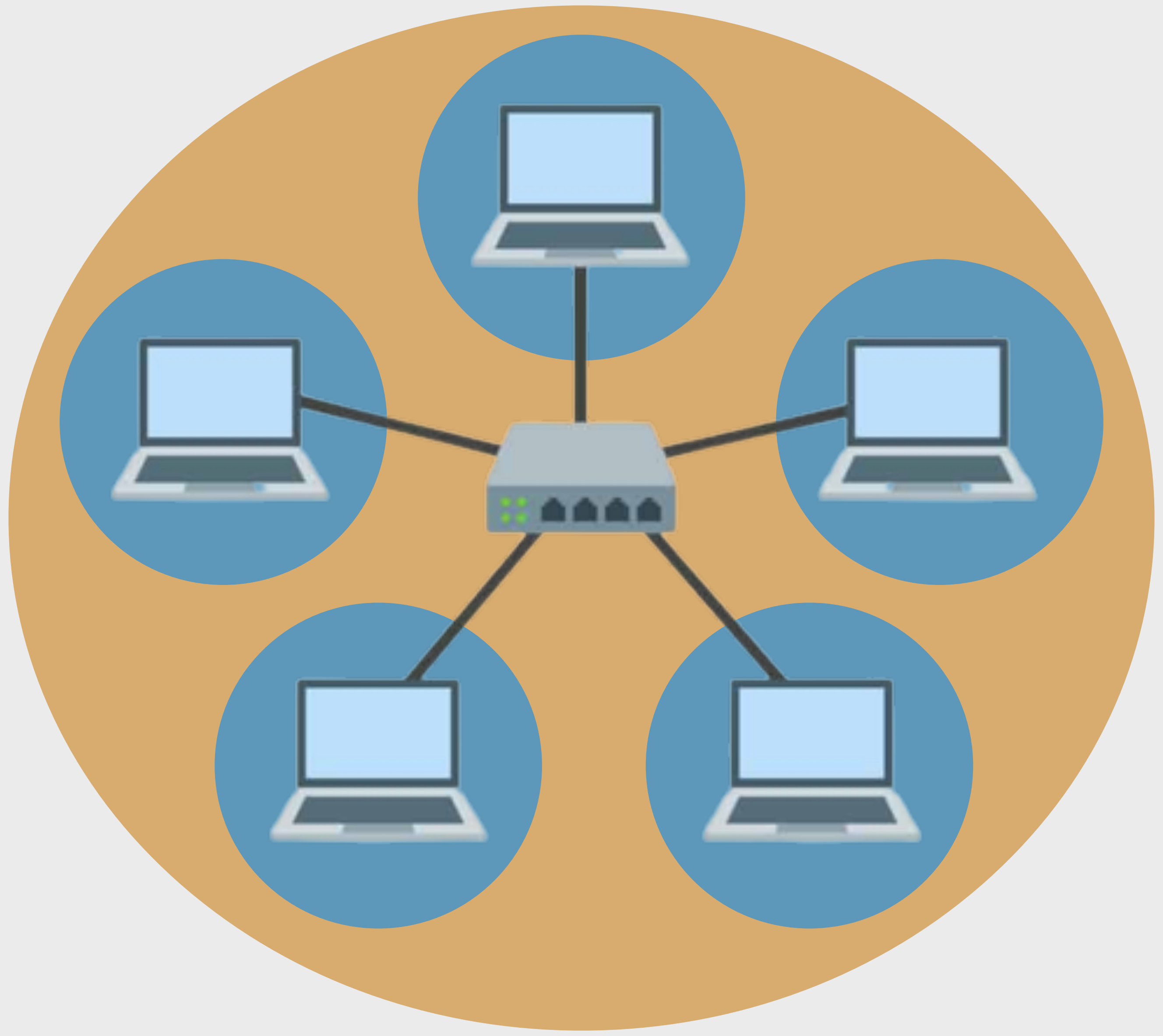
$$\mathscr{M}^i = (S^i, A^i, P^i, s_0^i)$$

States

Actions

Transition probability function

Initial state



Joint MDP

$$\mathscr{M} = (S, A, P, s_0)$$

$S = S^1 \times \ldots \times S^N$

$A = A^1 \times \ldots \times A^N$

$P(s, a, q) = \prod_{i=1}^{N} P^i(s^i, a^i, q^i)$

$s_I = (s_I^1, \ldots, s_I^N)$

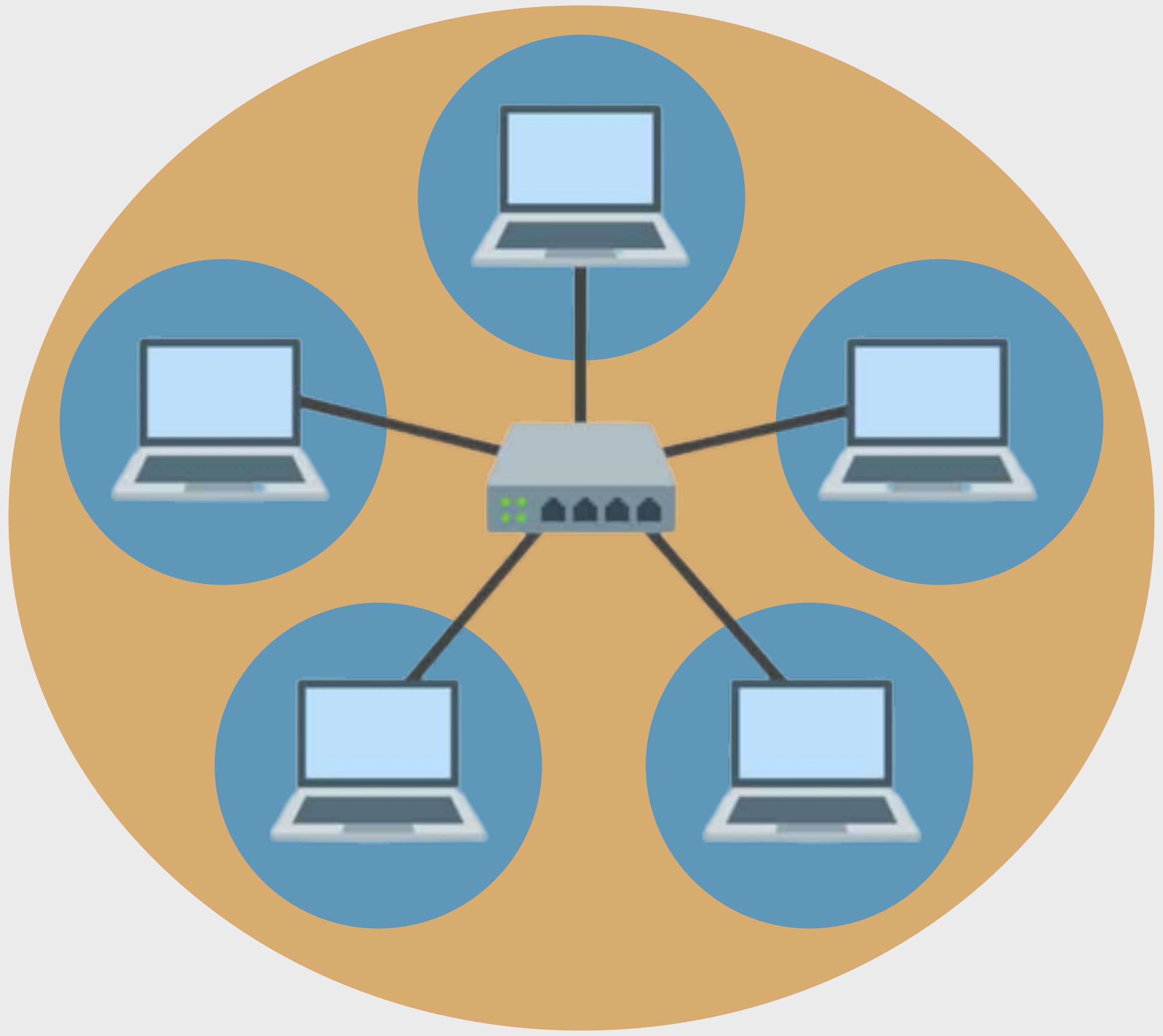
Modeling of multi-agent systems

N agents with independent dynamics.

Markov decision process (MDP) for Agent i

$\mathscr{M}^i = (S^i, A^i, P^i, s_0^i)$

States — Actions — Transition probability function — Initial state



Joint MDP $\mathscr{M} = (S, A, P, s_0)$

$S = S^1 \times \ldots \times S^N$ $A = A^1 \times \ldots \times A^N$ $P(s, a, q) = \prod_{i=1}^{N} P^i(s^i, a^i, q^i)$ $s_I = (s_I^1, \ldots, s_I^N)$

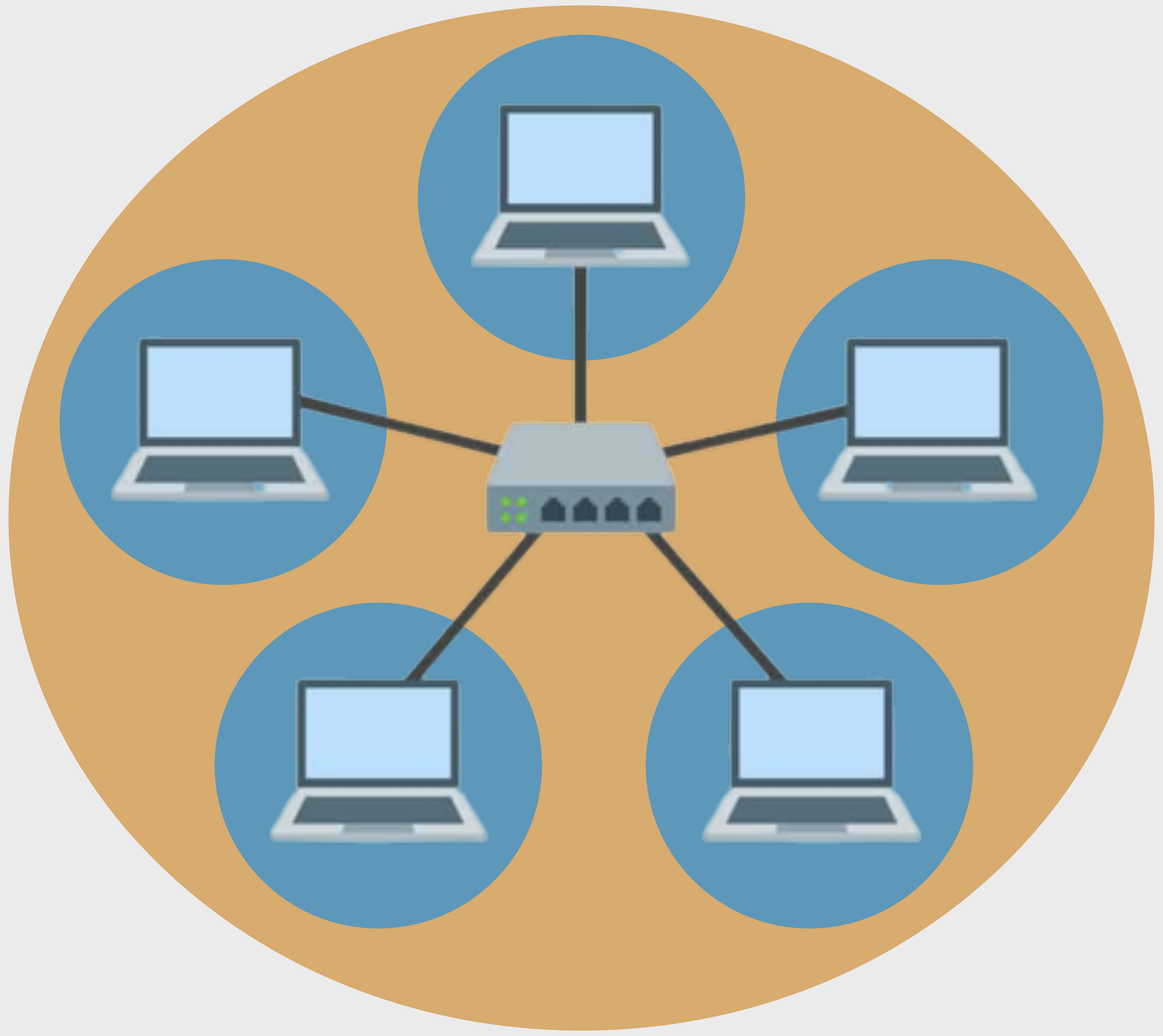
Team task: Eventually reach a target set S_T. The reachability probability is v.

Modeling of multi-agent systems

N agents with independent dynamics.

Markov decision process (MDP) for Agent i $\mathscr{M}^i = (S^i, A^i, P^i, s_0^i)$

States — Actions — Transition probability function — Initial state



Joint MDP $\mathscr{M} = (S, A, P, s_0)$

$S = S^1 \times \ldots \times S^N$ $\quad A = A^1 \times \ldots \times A^N$ $\quad P(s, a, q) = \prod_{i=1}^{N} P^i(s^i, a^i, q^i)$ $\quad s_I = (s_I^1, \ldots, s_I^N)$

Team task: Eventually reach a target set S_T. The reachability probability is v.

Joint policy $\pi_{joint}(s) \in \Delta(A)$ → Action distribution given the state

Spectrum of coordination in multi-agent systems

Spectrum of coordination in multi-agent systems

Full coordination

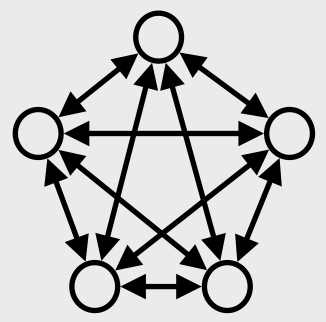
Better performance

High dependencies

No coordination

Worse performance

No dependencies



Fully centralized:

Share s
Jointly decide on a

Spectrum of coordination in multi-agent systems

Full coordination

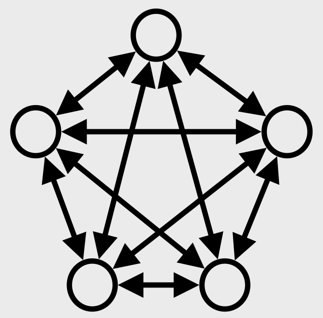
Better performance

High dependencies

No coordination

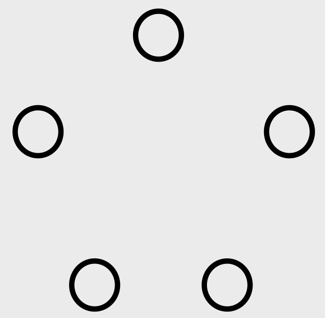
Worse performance

No dependencies



Fully centralized:

Share s
Jointly decide on a



Fully decentralized:

Share nothing

Spectrum of coordination in multi-agent systems

Full coordination

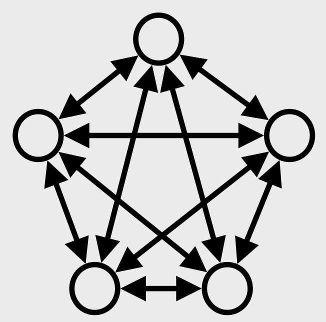
Better performance

High dependencies

No coordination

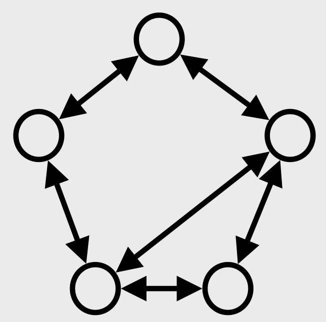
Worse performance

No dependencies



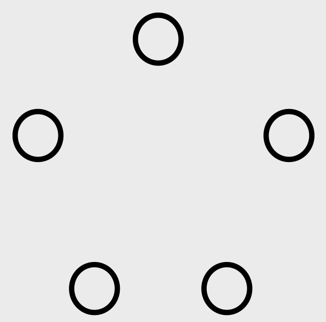
Fully centralized:

Share s
Jointly decide on a



Fixed communication graph:

Share s with few others
Jointly decide on a with few others



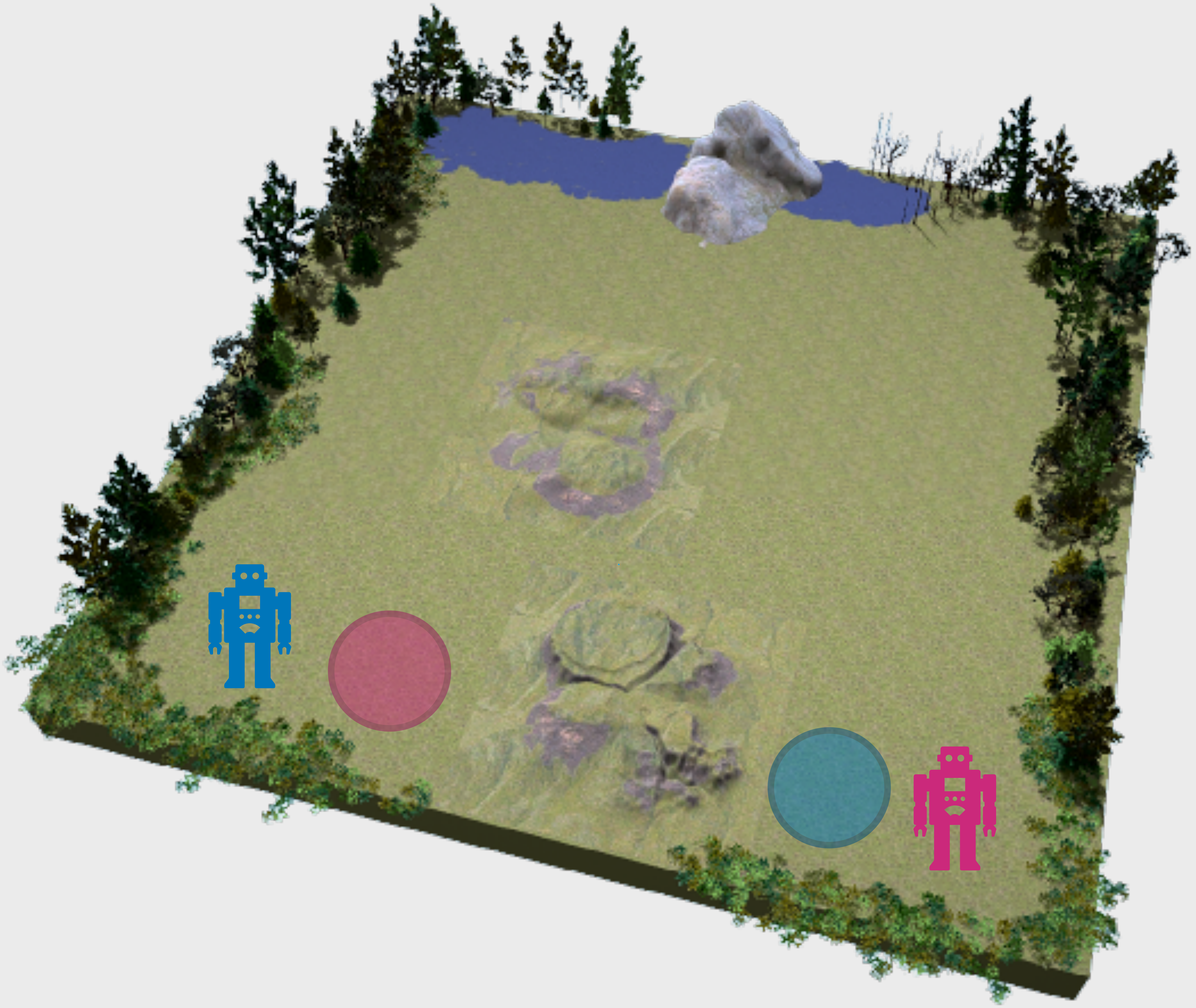
Fully decentralized:

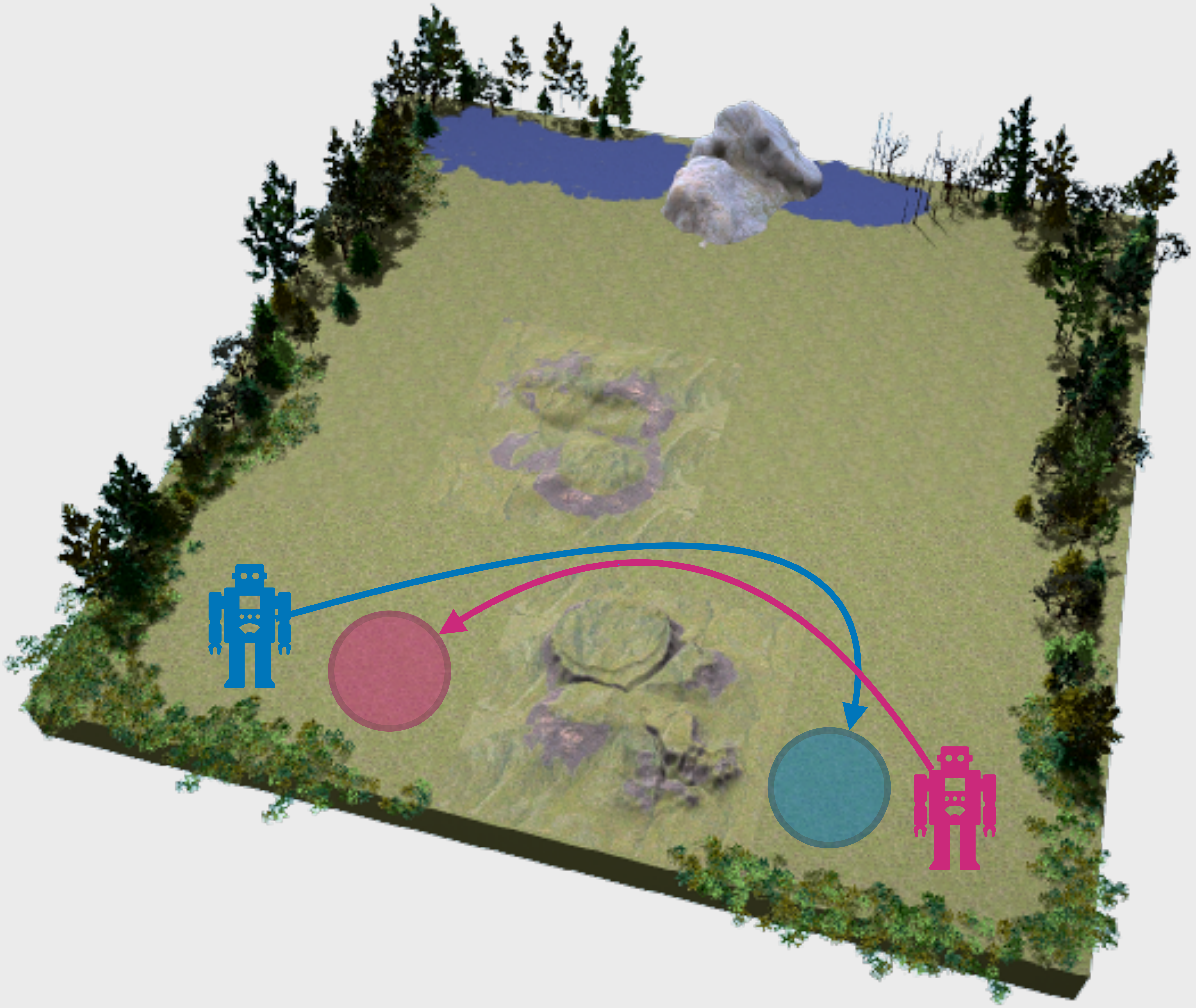
Share nothing

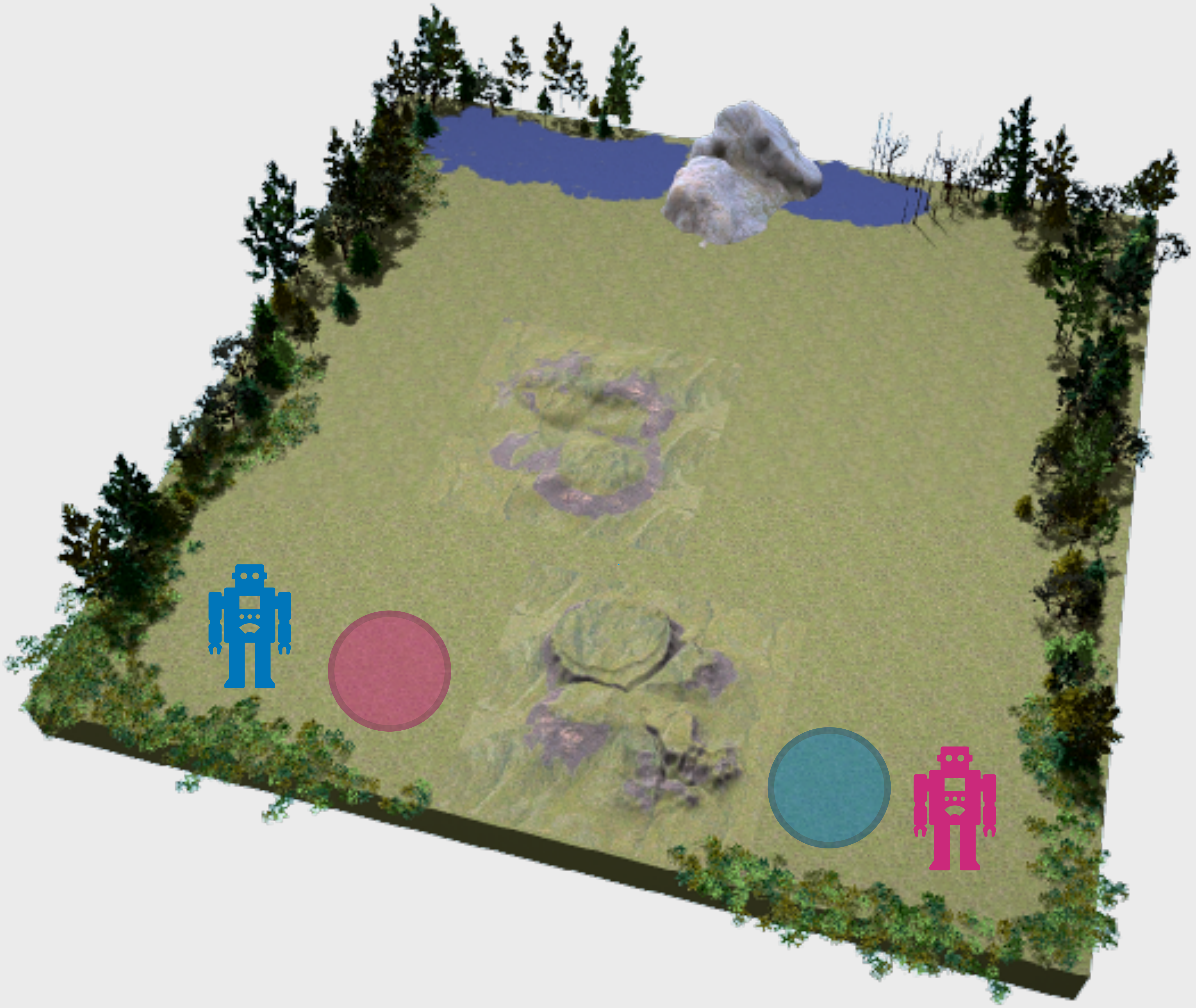
What is done? What is needed?

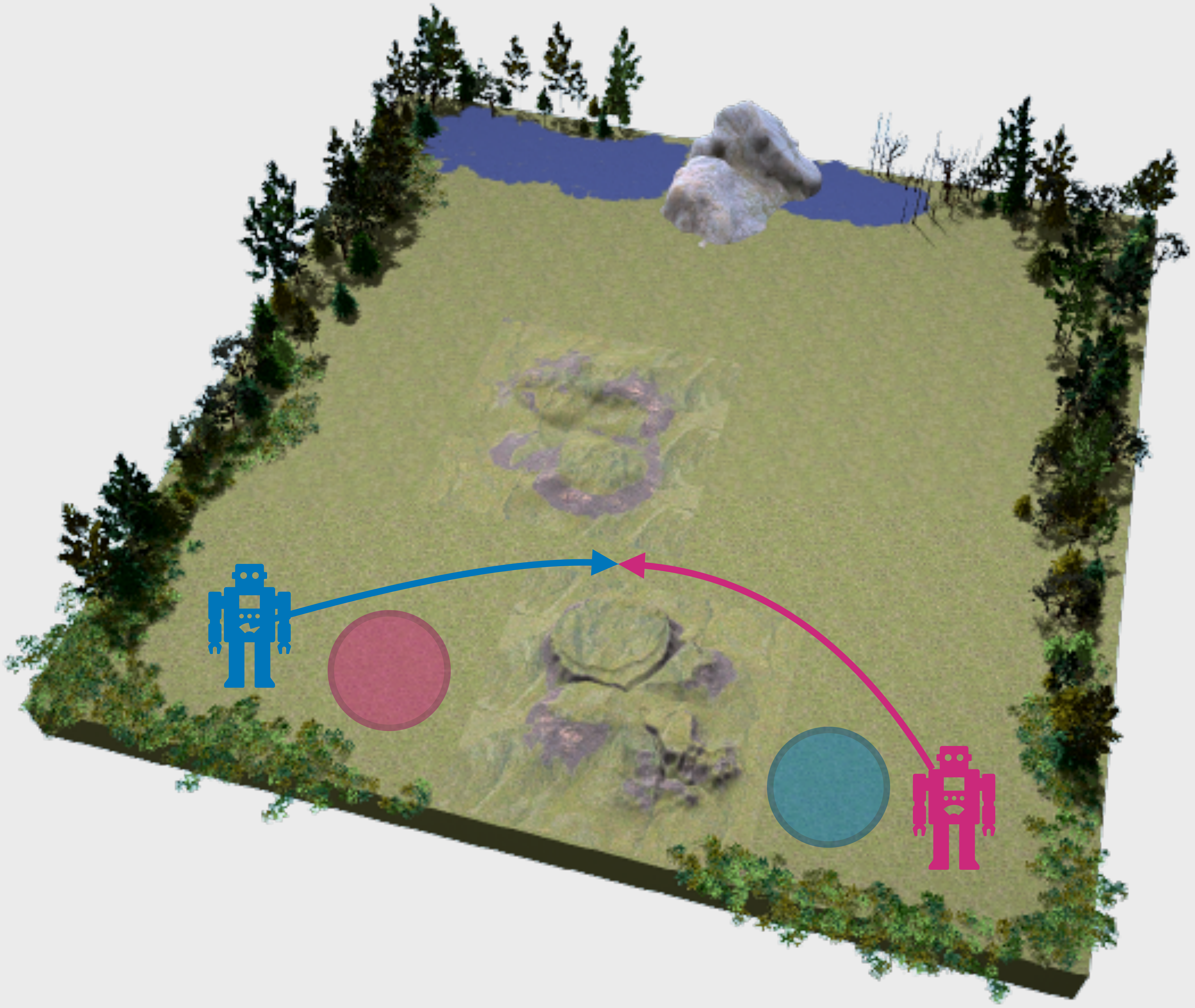
Existing methods are either **oblivious** to dependencies or **restricted** to explicit communication graphs.

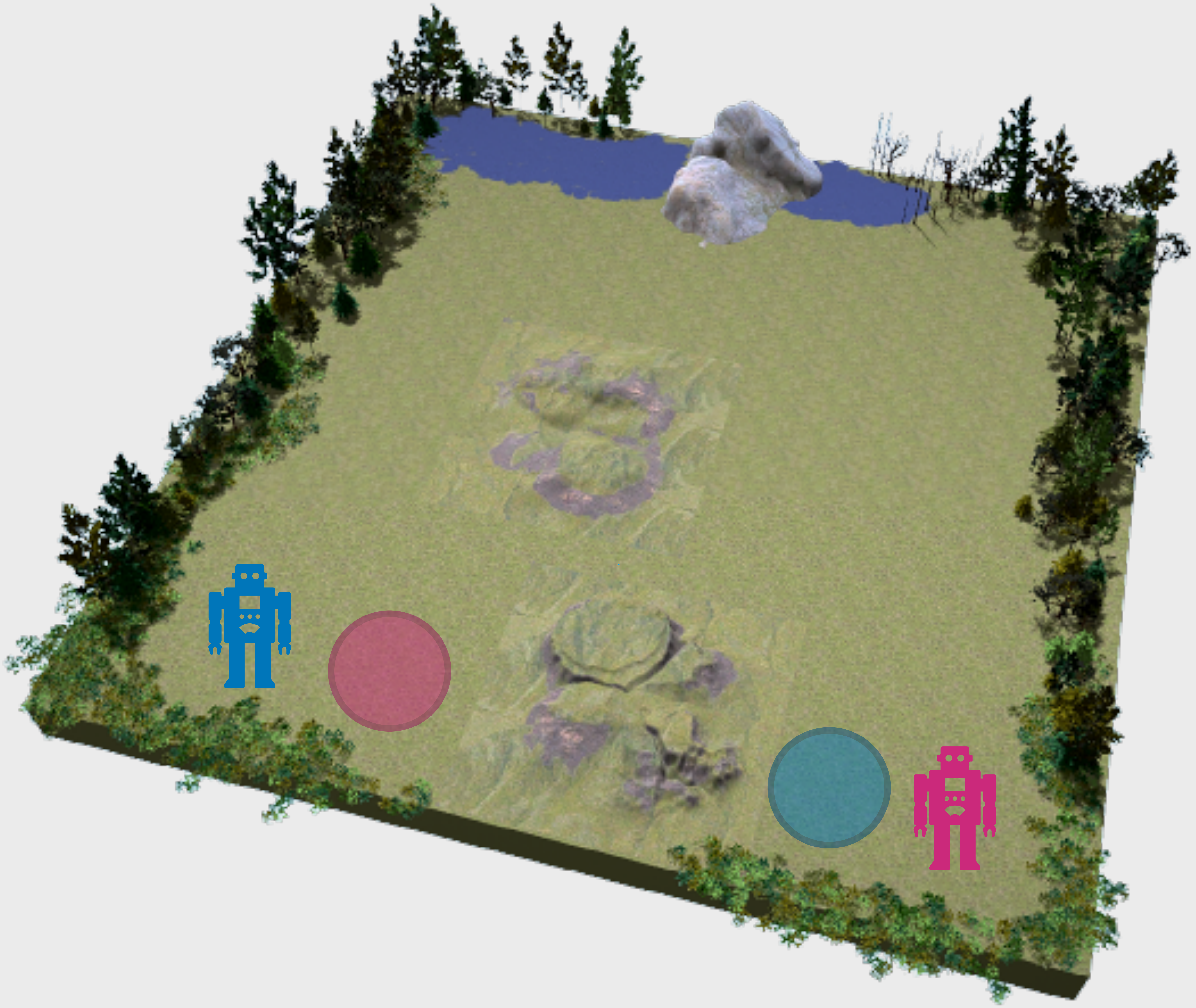
We want **performant** policies that are **robust** to communication losses.

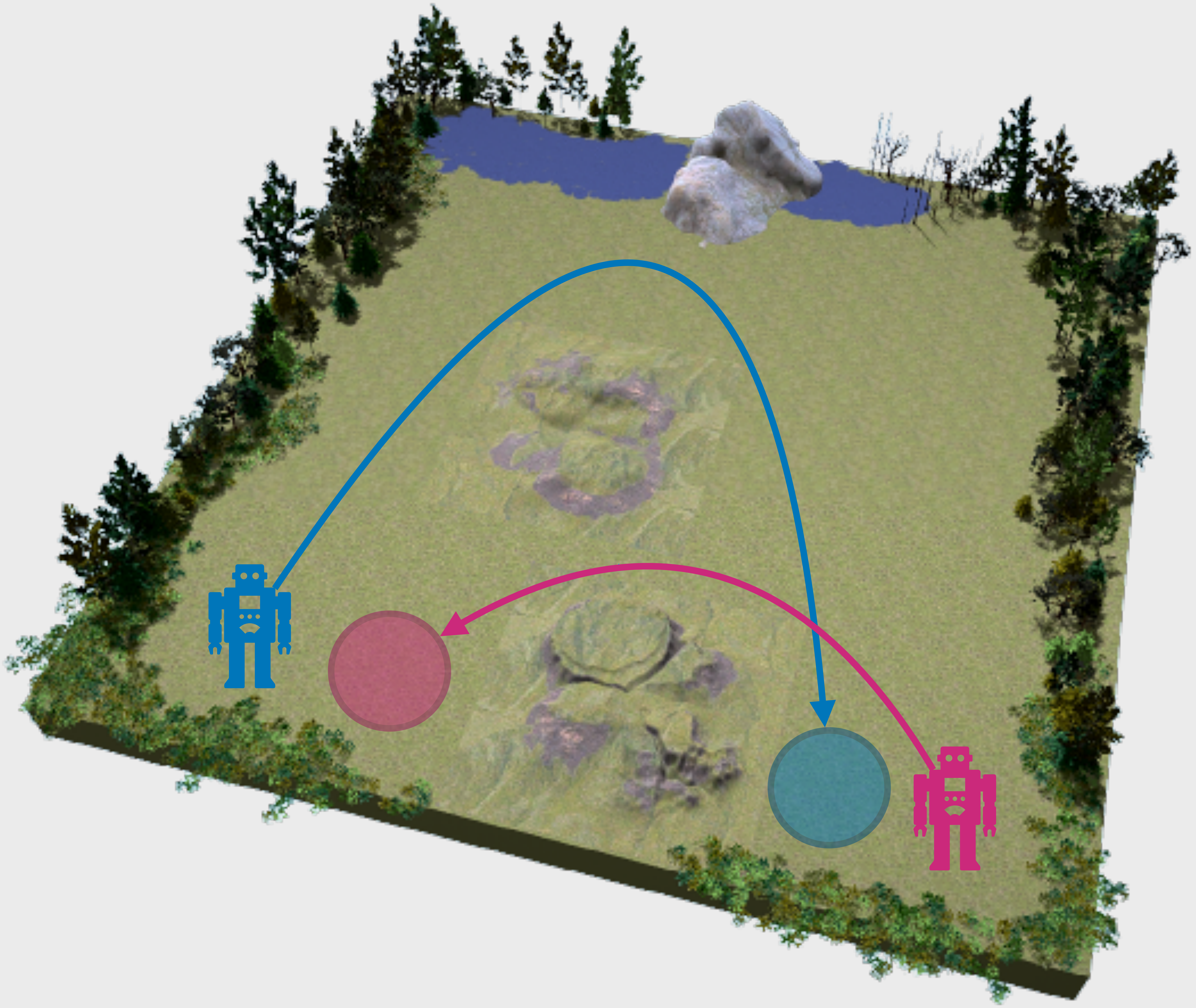






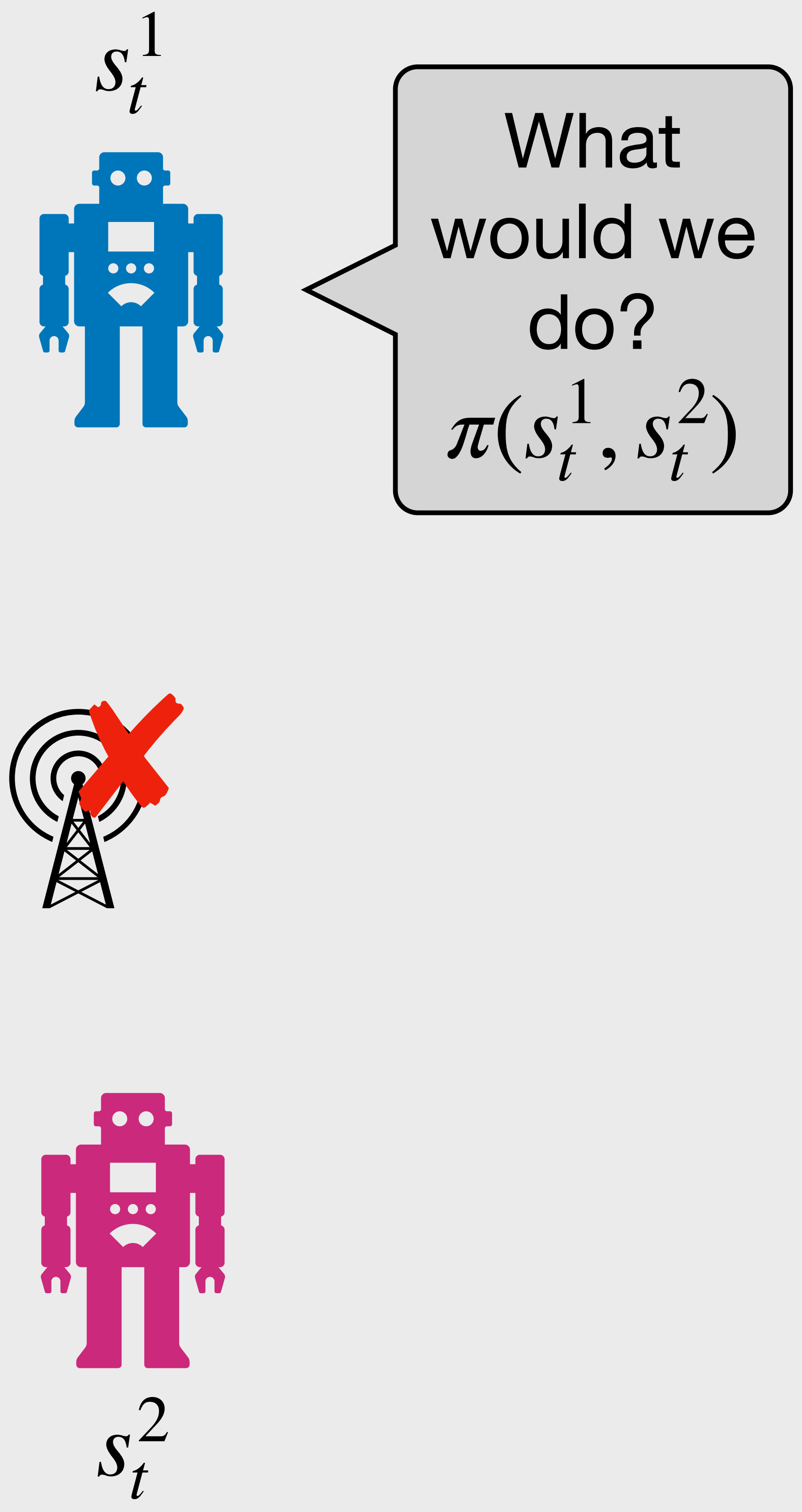




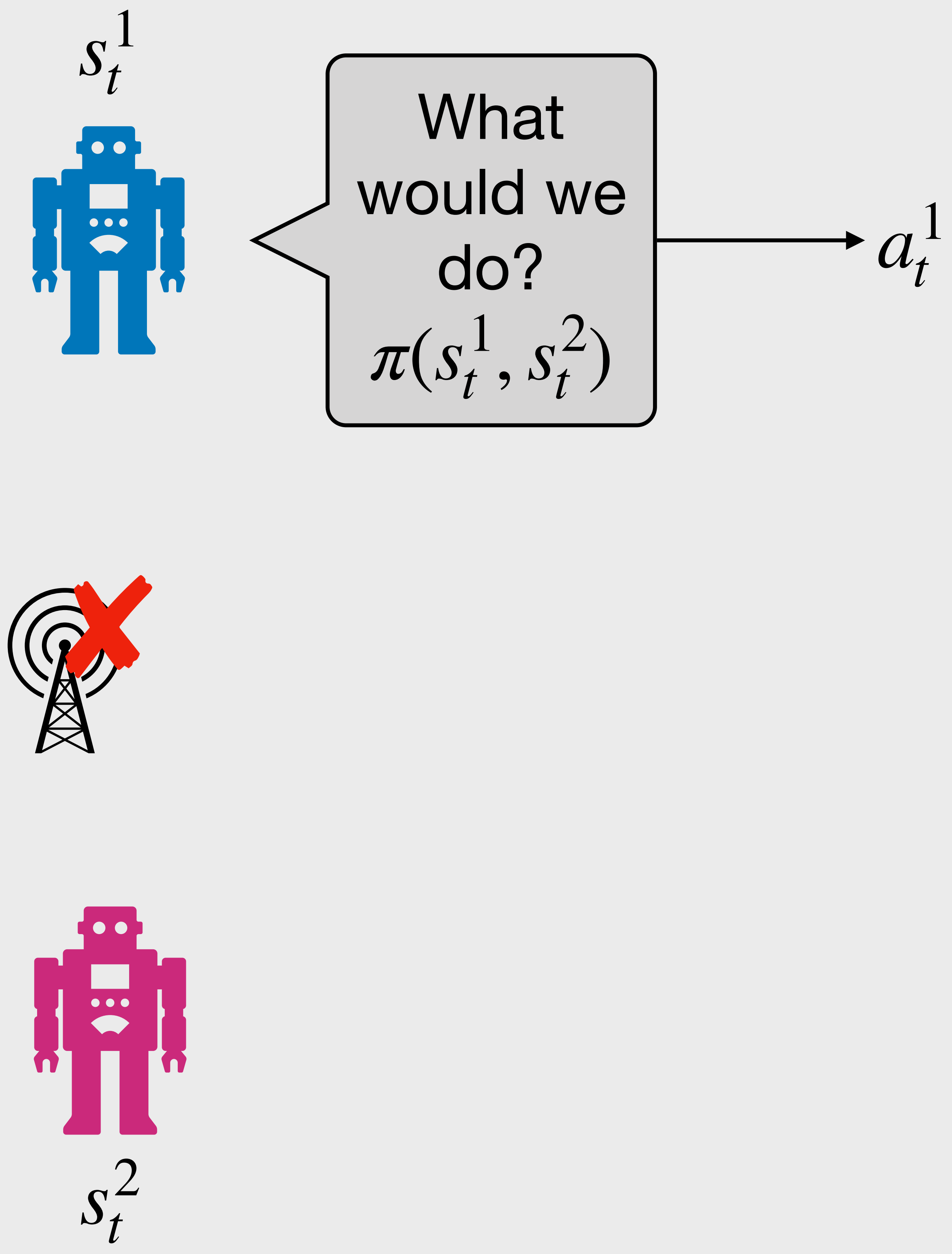


What if agents cannot communicate?

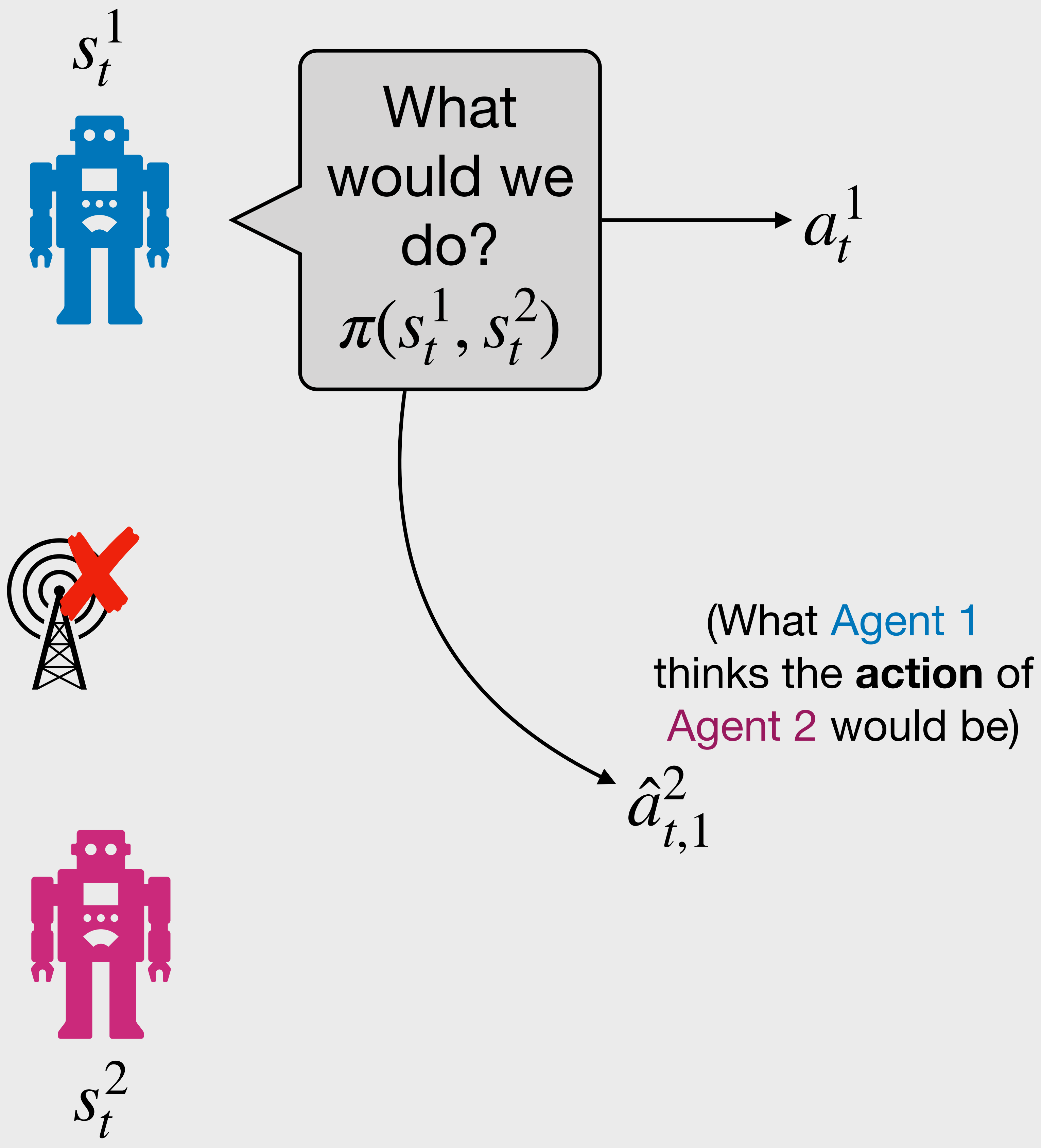
What if agents cannot communicate?



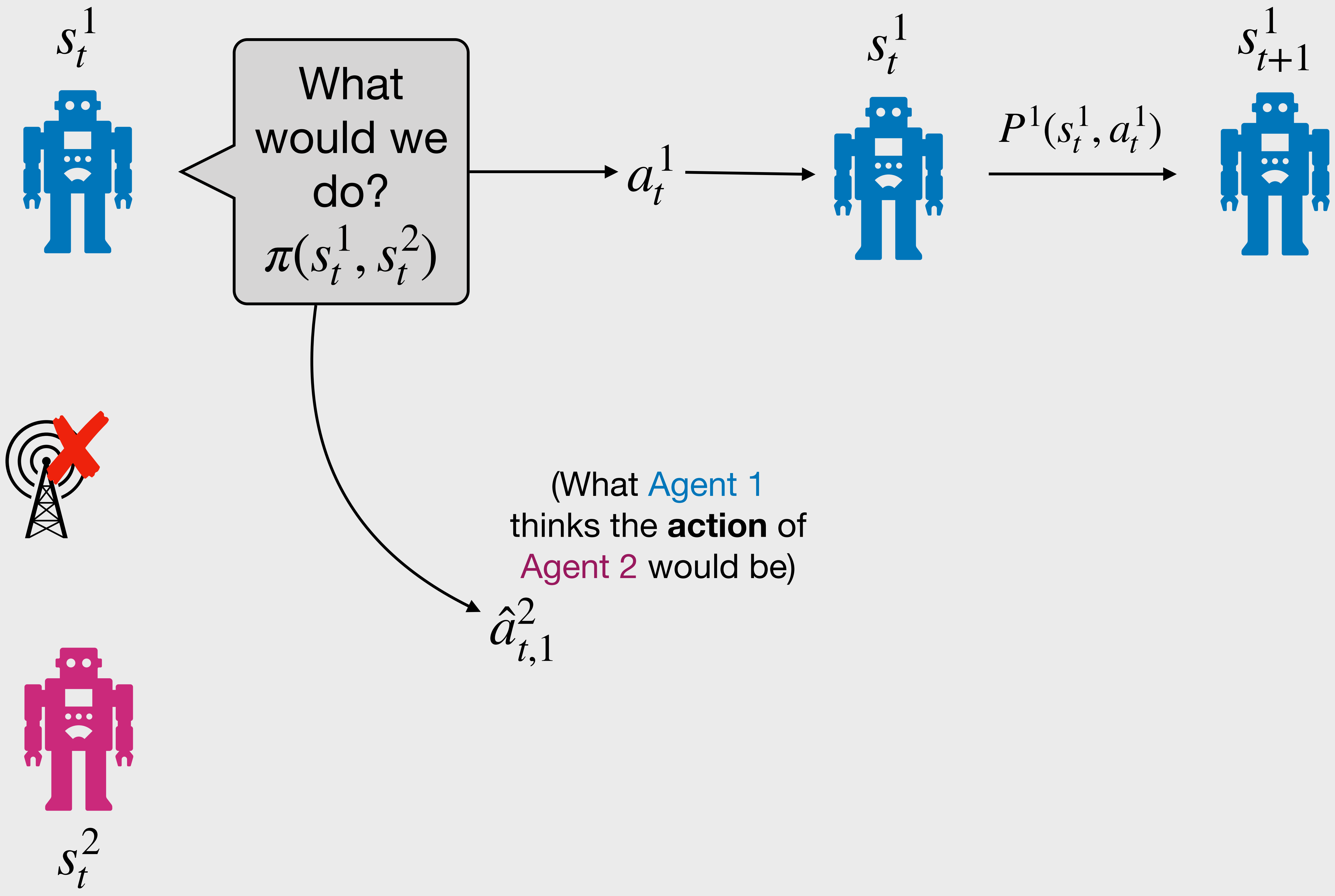
What if agents cannot communicate?



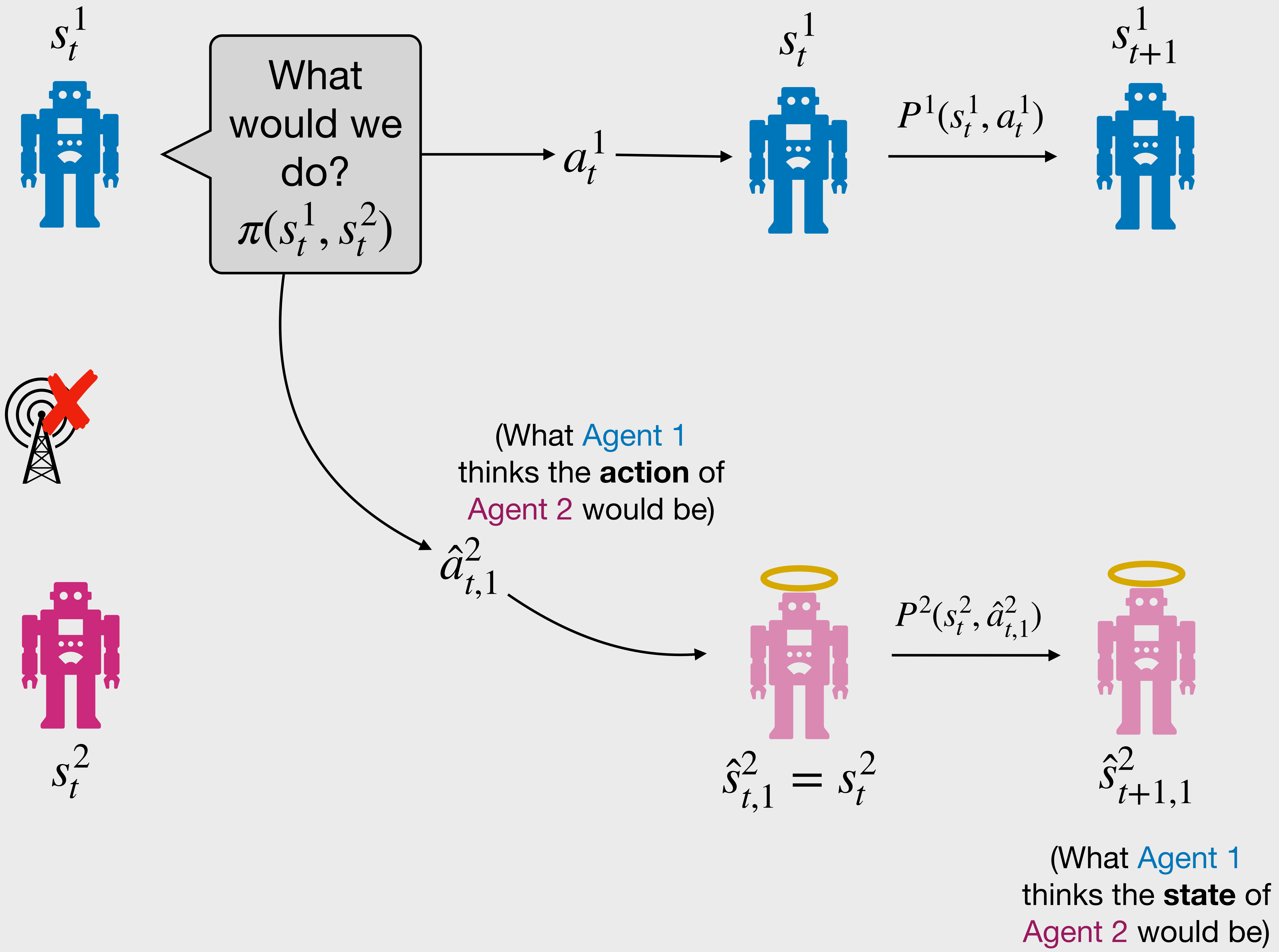
What if agents cannot communicate?



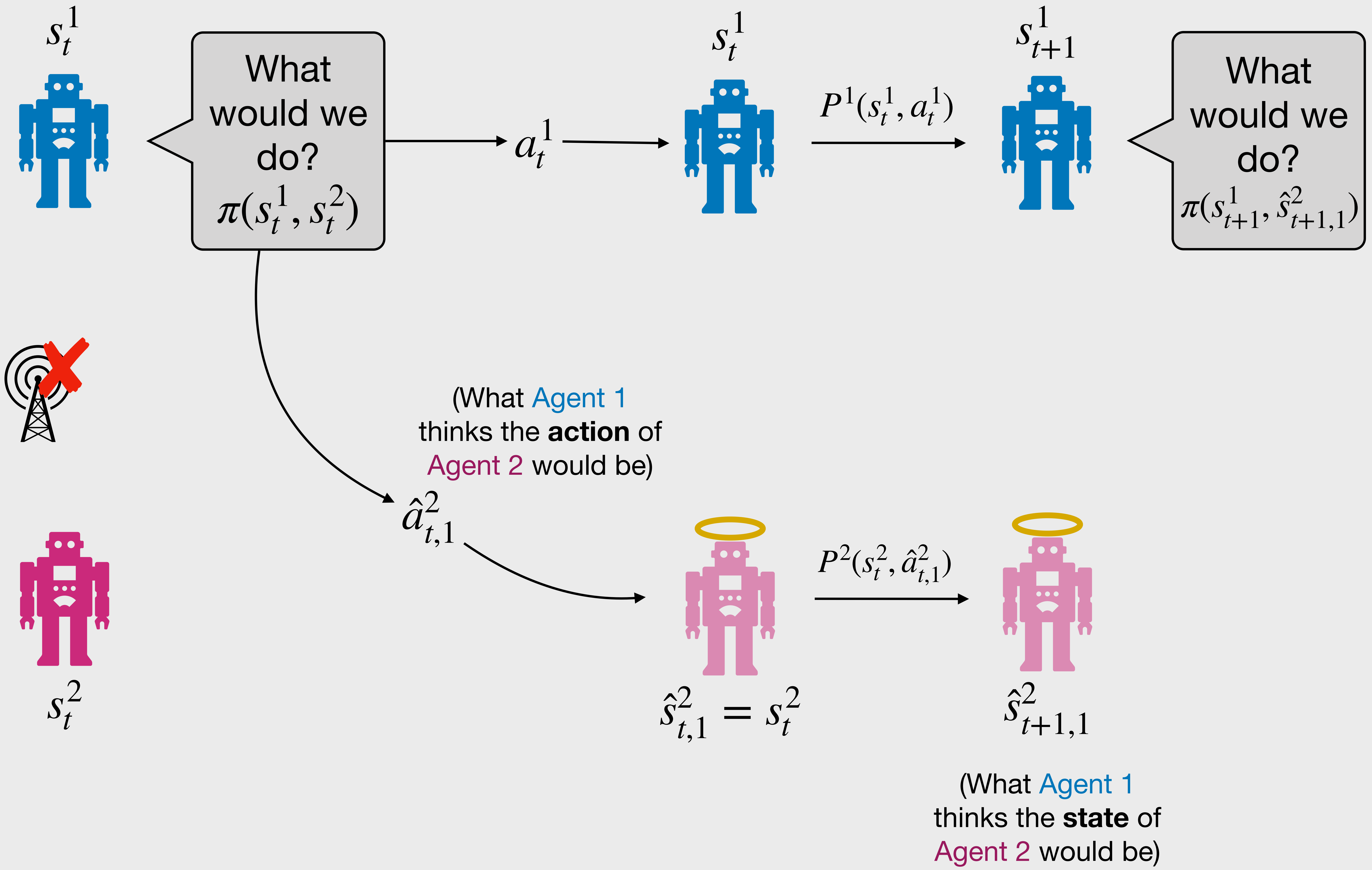
What if agents cannot communicate?



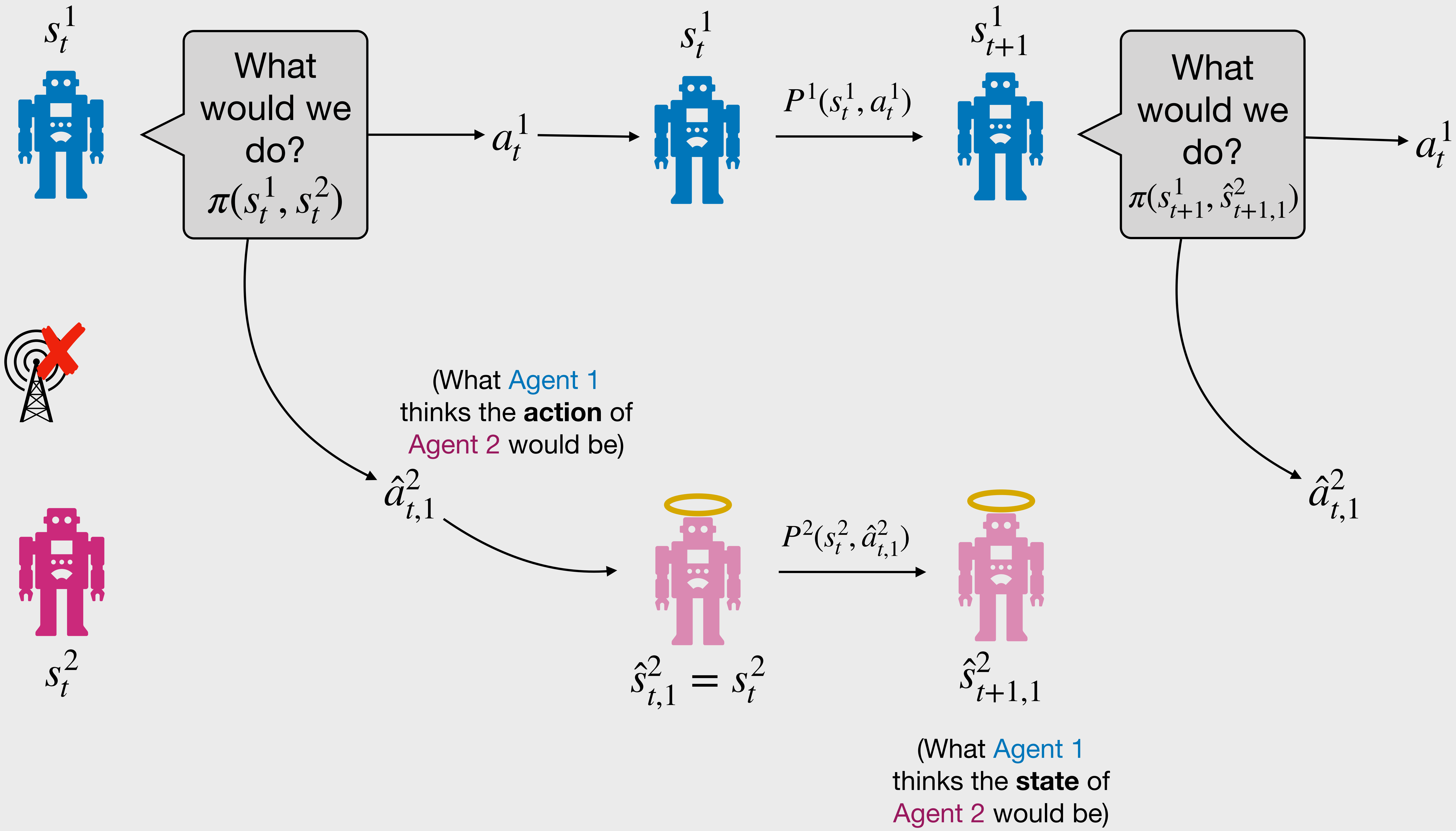
What if agents cannot communicate?



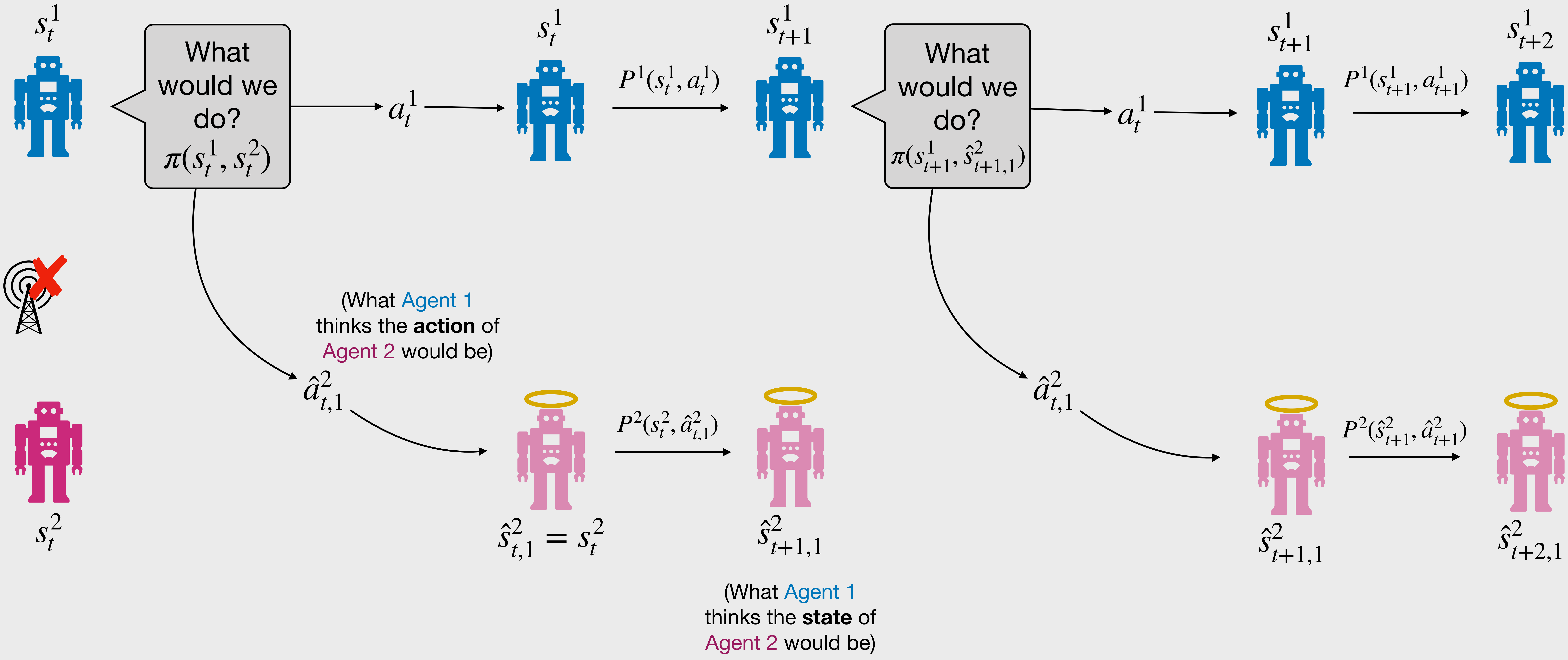
What if agents cannot communicate?



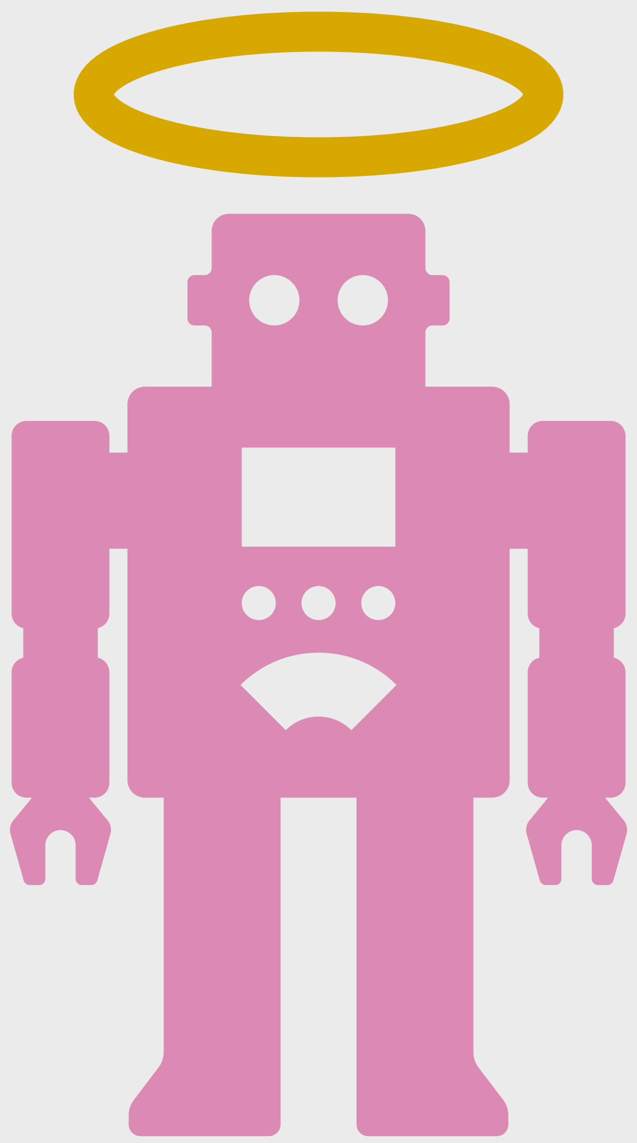
What if agents cannot communicate?



What if agents cannot communicate?



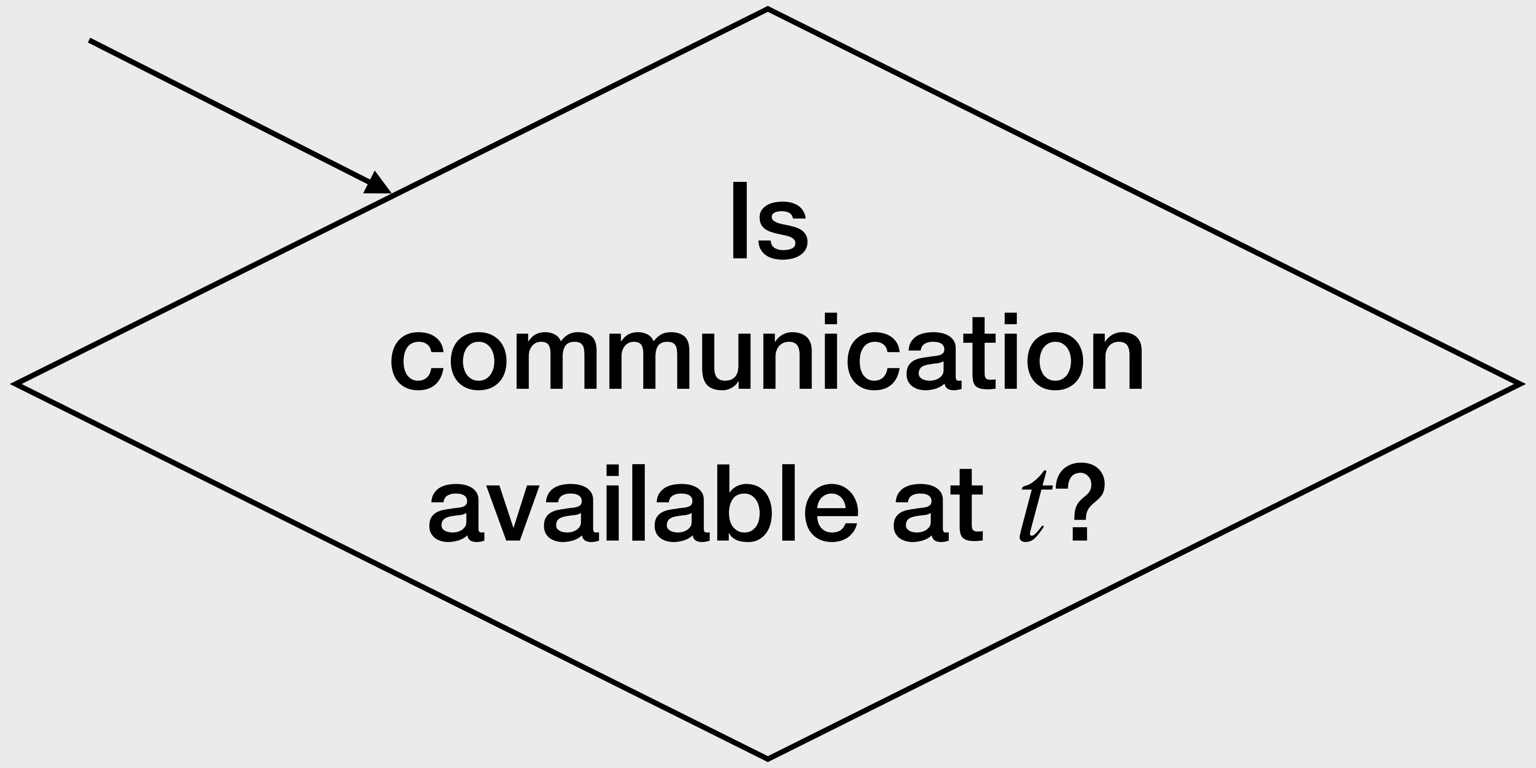
What if agents cannot communicate?



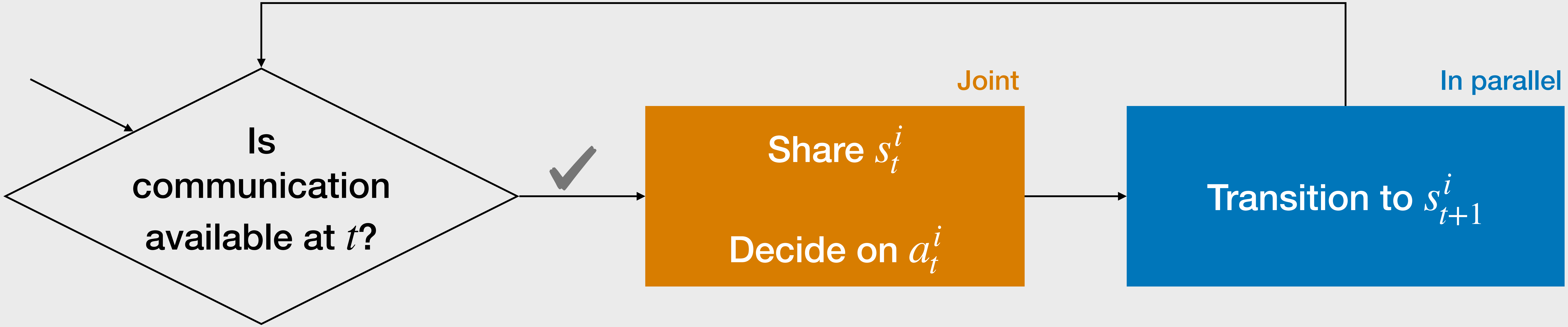
Imaginary play

Policy execution under permanent or intermittent communication loss

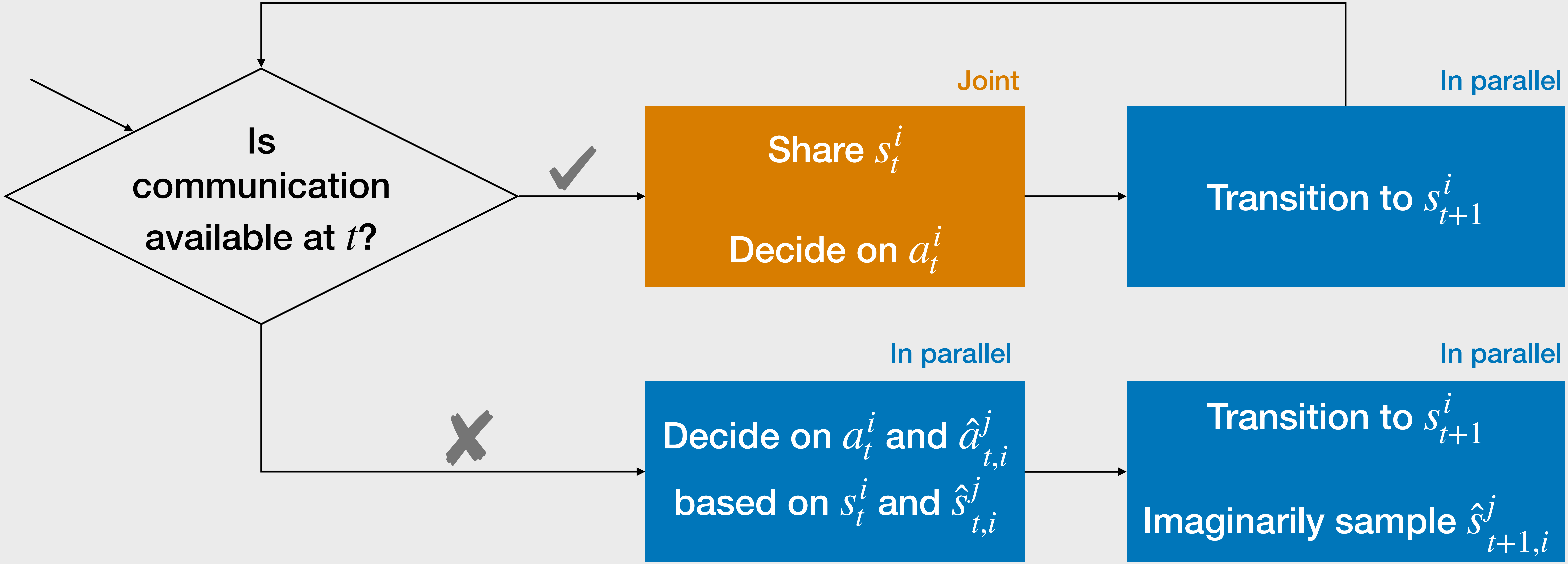
Policy execution under permanent or intermittent communication loss



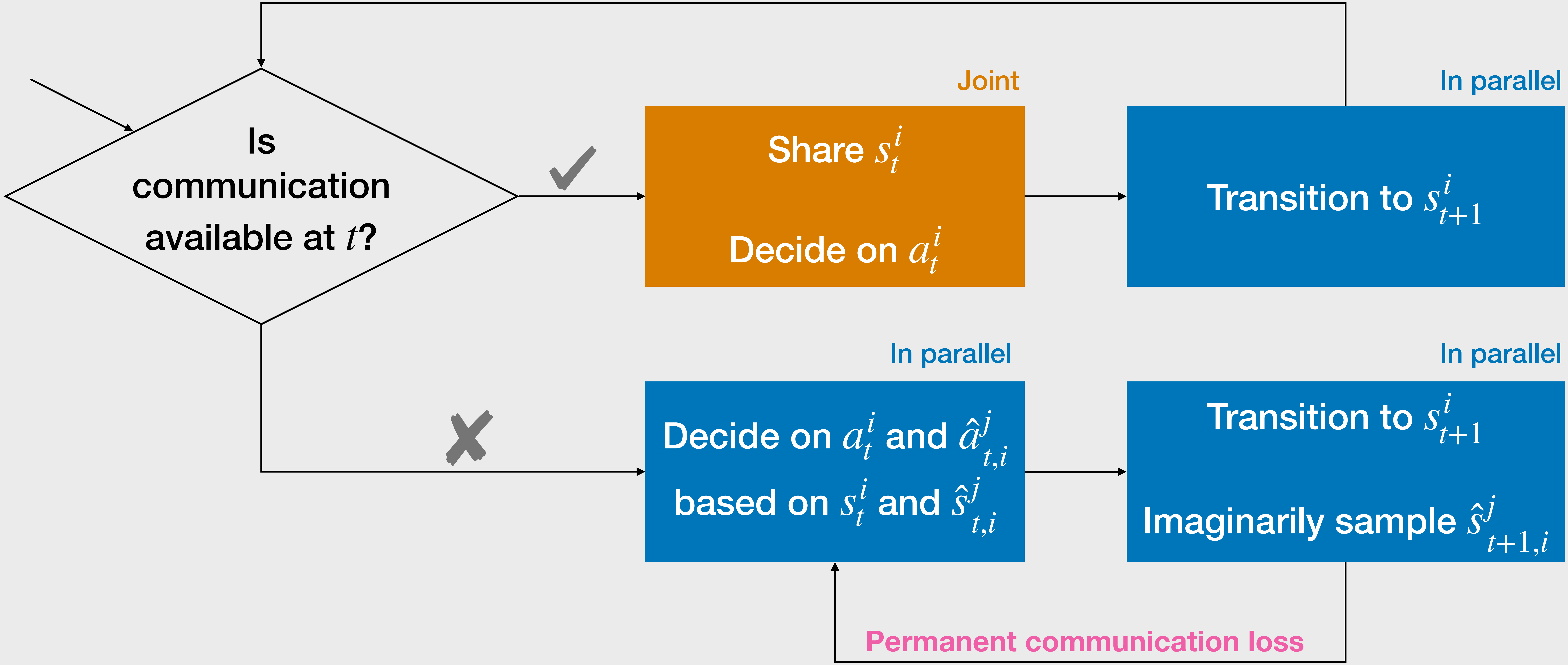
Policy execution under permanent or intermittent communication loss



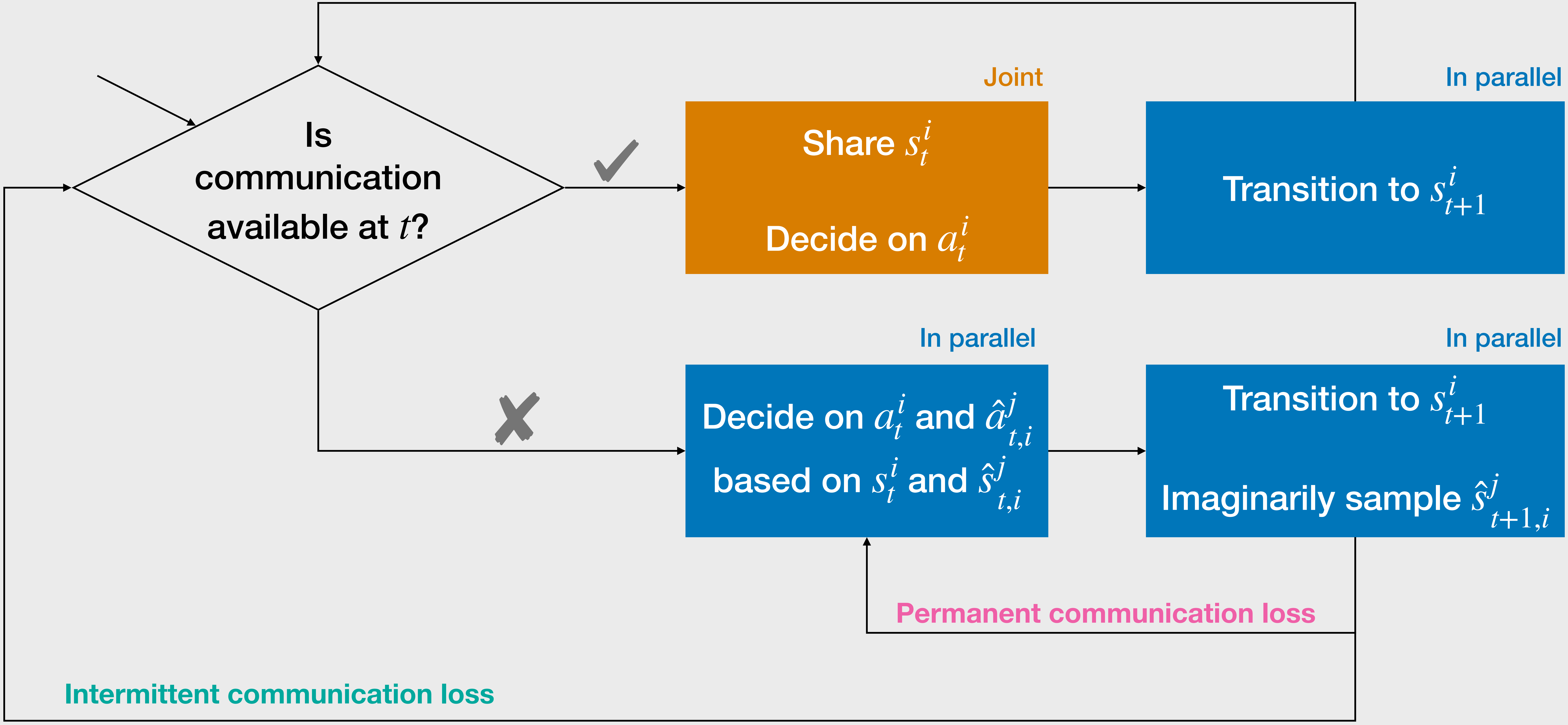
Policy execution under permanent or intermittent communication loss



Policy execution under permanent or intermittent communication loss



Policy execution under permanent or intermittent communication loss



Measuring intrinsic dependencies between agents

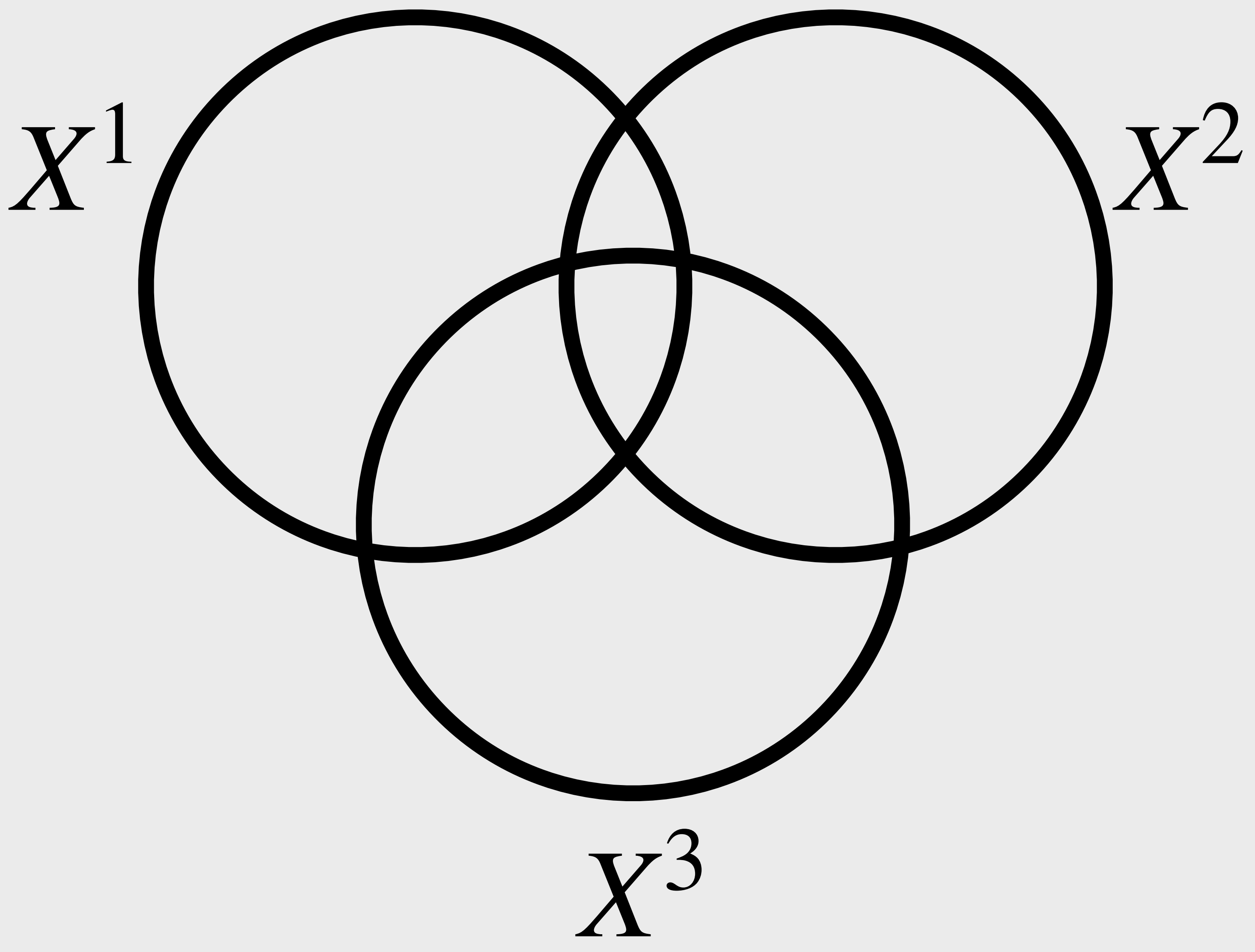
State-action
processes of agents

$$\mathbf{X} = (X^1, \ldots, X^N)$$

Joint measure: μ

Individual measures: $\mu^1, \ldots, \mu^N$

Product measure: $\mu^{prod} = \mu^1 \times \ldots \times \mu^N$



Measuring intrinsic dependencies between agents

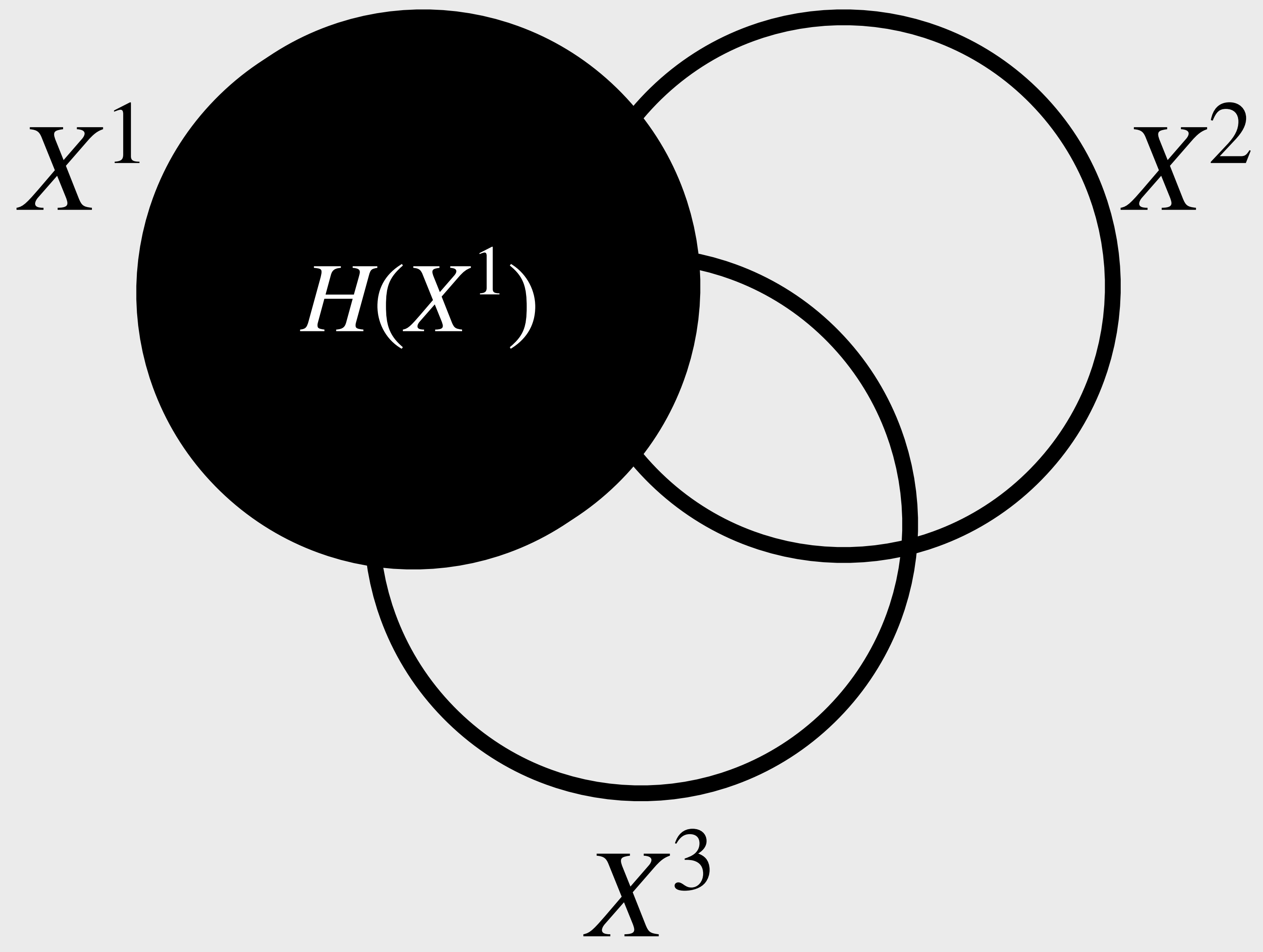
State-action processes of agents

$$\mathbf{X} = (X^1, \ldots, X^N)$$

Joint measure: μ

Individual measures: $\mu^1, \ldots, \mu^N$

Product measure: $\mu^{prod} = \mu^1 \times \ldots \times \mu^N$



Entropy
=
Information

$$\textbf{Entropy} = H(X^i) = \sum_{x \in Support(X^i)} \mu^i(x) \log\left(\frac{1}{\mu^i(x)}\right)$$

Measuring intrinsic dependencies between agents

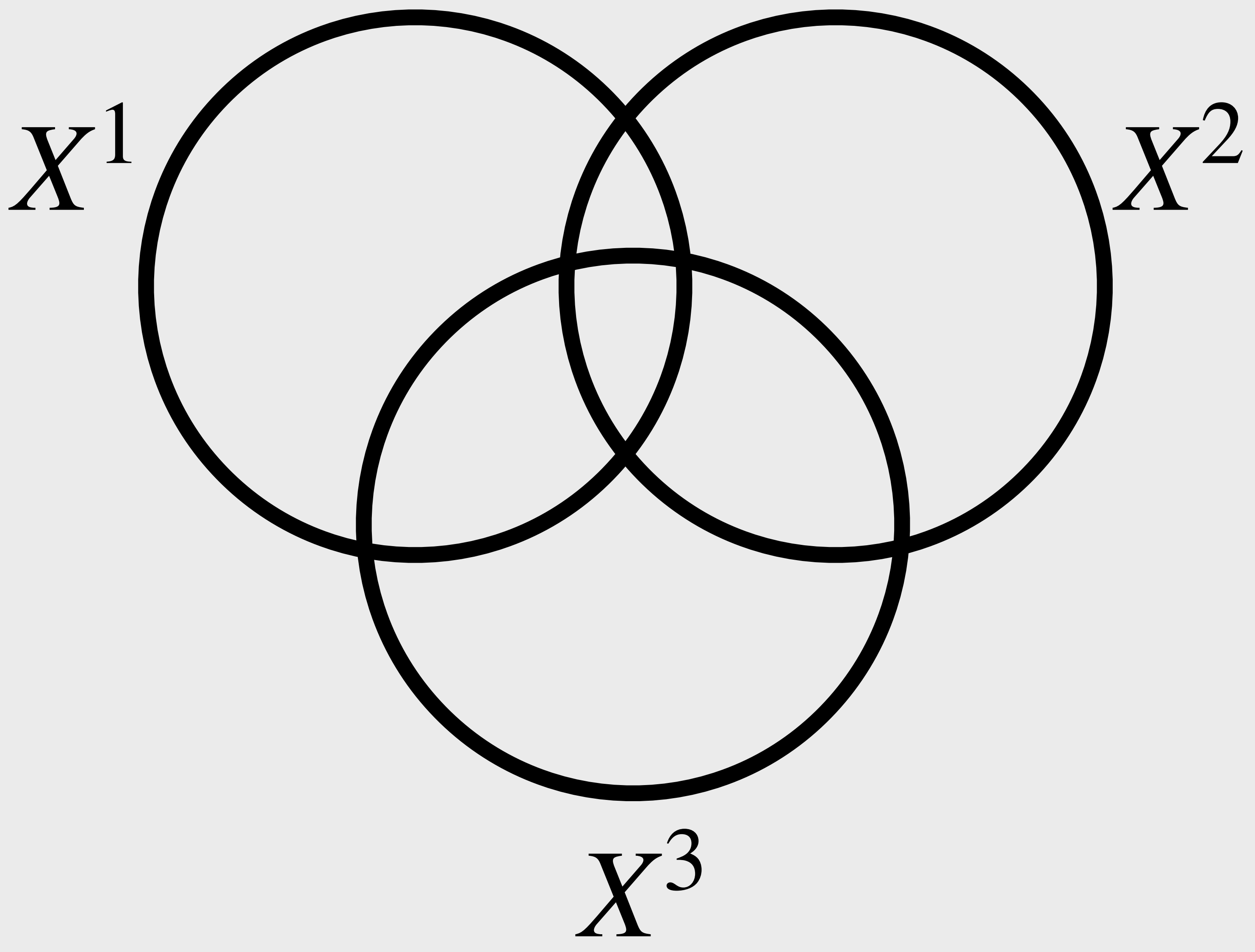
State-action
processes of agents

$$\mathbf{X} = (X^1, \ldots, X^N)$$

Joint measure: μ

Individual measures: $\mu^1, \ldots, \mu^N$

Product measure: $\mu^{prod} = \mu^1 \times \ldots \times \mu^N$



Entropy
=
Information

Measuring intrinsic dependencies between agents

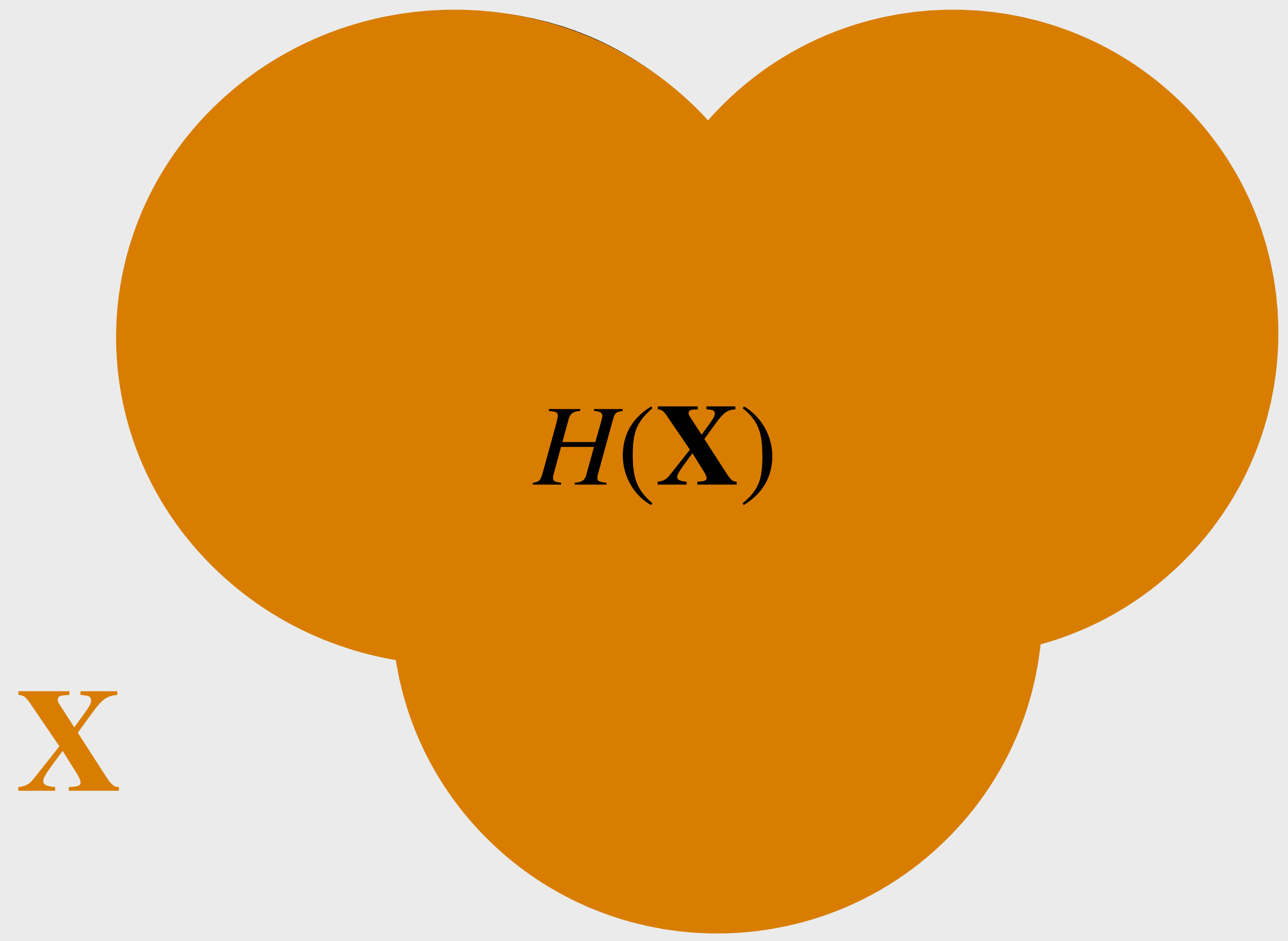
State-action
processes of agents

$$\mathbf{X} = (X^1, \ldots, X^N)$$

Joint measure: μ

Individual measures: $\mu^1, \ldots, \mu^N$

Product measure: $\mu^{prod} = \mu^1 \times \ldots \times \mu^N$



Entropy
=
Information

$$\textbf{Entropy} = H(\mathbf{X}) = \sum_{x \in Support(\mathbf{X})} \mu(x) \log\left(\frac{1}{\mu(x)}\right)$$

Measuring intrinsic dependencies between agents

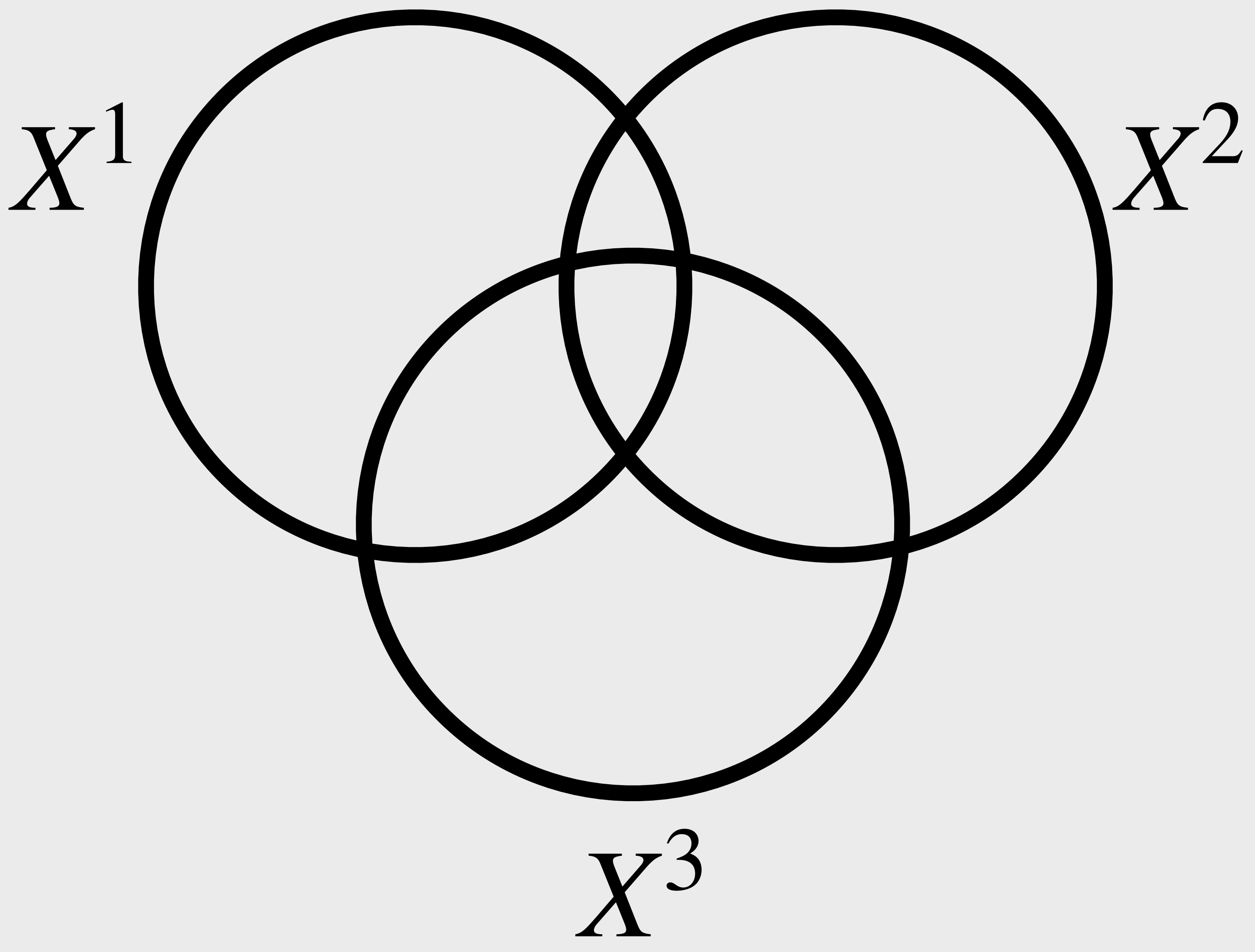
State-action
processes of agents

$$\mathbf{X} = (X^1, \ldots, X^N)$$

Joint measure: μ

Individual measures: $\mu^1, \ldots, \mu^N$

Product measure: $\mu^{prod} = \mu^1 \times \ldots \times \mu^N$



Measuring intrinsic dependencies between agents

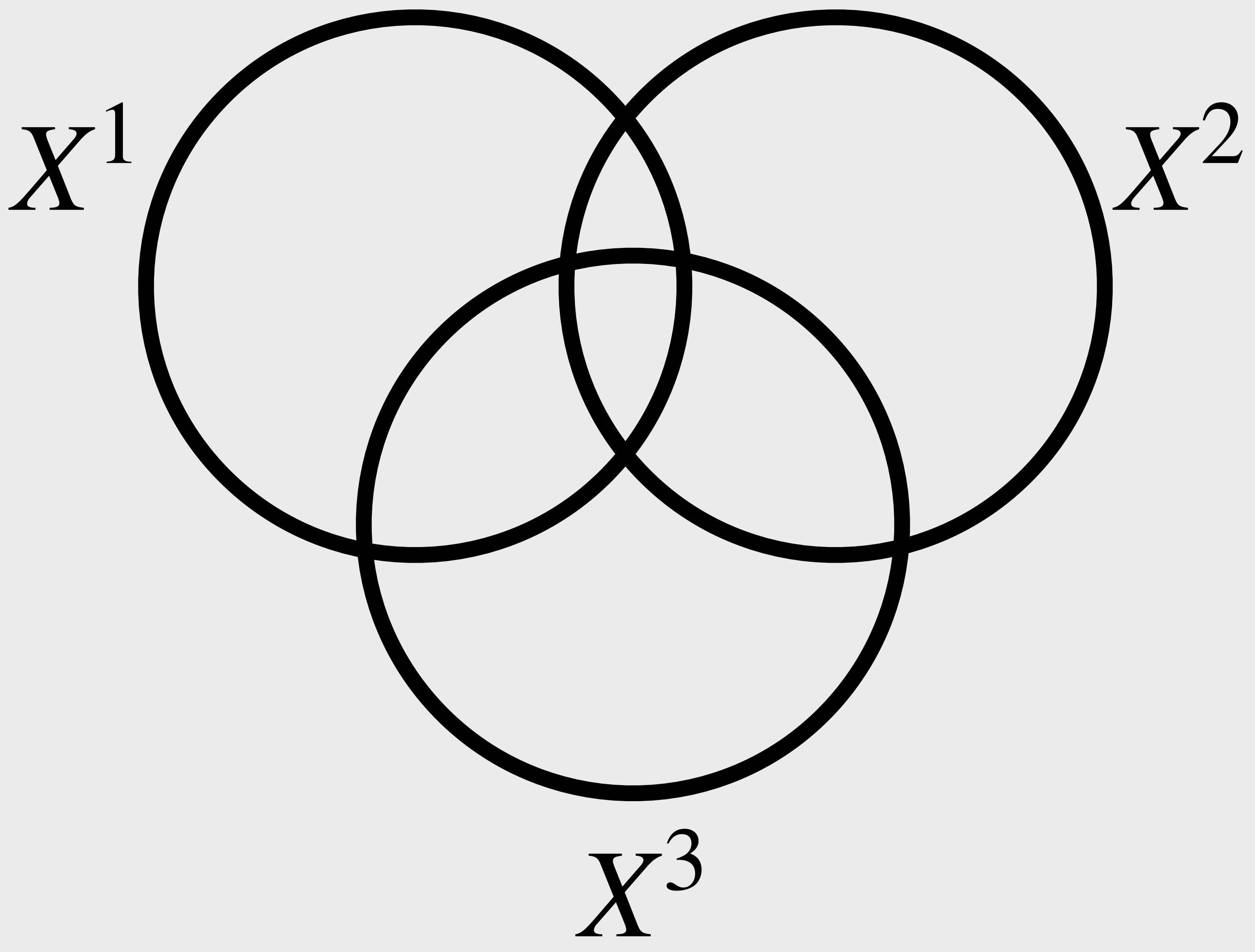
State-action processes of agents

$$\mathbf{X} = (X^1, \ldots, X^N)$$

Joint measure: μ

Individual measures: $\mu^1, \ldots, \mu^N$

Product measure: $\mu^{prod} = \mu^1 \times \ldots \times \mu^N$



Total correlation
=
Shared Information

Total correlation $= C(X^1, \ldots, X^N) = \left(\sum_{i=1}^{N} H(X^i) \right) - H(\mathbf{X})$

Measuring intrinsic dependencies between agents

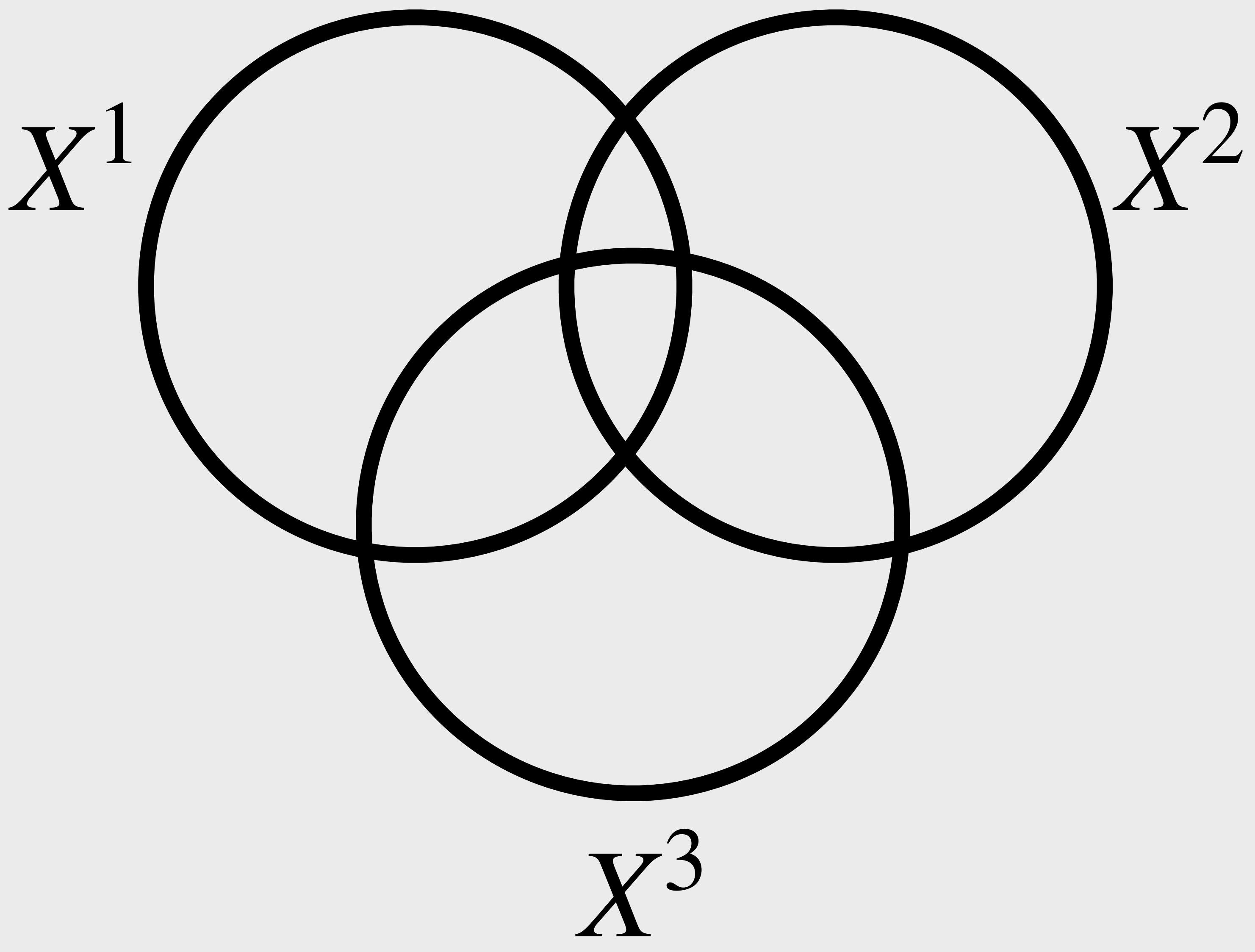
State-action processes of agents

$$\mathbf{X} = (X^1, \ldots, X^N)$$

Joint measure: μ

Individual measures: $\mu^1, \ldots, \mu^N$

Product measure: $\mu^{prod} = \mu^1 \times \ldots \times \mu^N$



Total correlation
=
Shared Information
=
Dissimilarity between the joint and product measures

Total correlation $= C(X^1, \ldots, X^N) = KL\left(\mu \,||\, \mu^{prod}\right)$

Total correlation is the difference between full communication and fully imaginary play

Total correlation $= C_{\pi^{joint}} = \left(\sum_{i=1}^{N} H(X^i) \right) - H(\mathbf{X}) = KL\left(\mu \,||\, \mu^{prod}\right)$

t_{loss} : when the communication loss starts

$\mu_{t_{loss}}^{img}$: the probability measure induced by imaginary play

Joint measure μ **= Full communication** μ^{full}

Product measure μ^{prod} **= No communication (imaginary play)** μ_0^{img}

Roadmap to theoretical guarantees

Lemma:

Behavior difference under
full communication
and
no communication

Property:

Total
Correlation

Roadmap to theoretical guarantees

Lemma:

Property:

Total
Correlation

Behavior difference under
full communication
and
no communication

Behavior difference under
full communication
and
some communication

Roadmap to theoretical guarantees

Property:

Total Correlation

Lemma:

Behavior difference under
full communication
and
no communication

Behavior difference under
full communication
and
some communication

Theorem:

Performance difference under
full communication
and
some communication

Lemma: Any extra communication at the beginning does not hurt.

$$C_{\pi^{joint}} = KL\left(\mu^{full} || \mu_0^{img}\right) \geq KL\left(\mu^{full} || \mu_{t_{loss}}^{img}\right)$$

Lemma: Any extra communication at the beginning does not hurt.

$$C_{\pi^{joint}} = KL\left(\mu^{full} || \mu_0^{img}\right) \geq KL\left(\mu^{full} || \mu_{t_{loss}}^{img}\right)$$

Stronger lemma: Any extra communication does not hurt.

Λ : a binary sequence of communication availability

μ_Λ^{int}: the probability measure induced by intermittent play

$$C_{\pi^{full}} = KL\left(\mu^{full} || \mu_0^{img}\right) \geq KL\left(\mu^{full} || \mu_\Lambda^{int}\right)$$

Lemma: Any extra communication at the beginning does not hurt.

$$C_{\pi^{joint}} = KL\left(\mu^{full} || \mu_0^{img}\right) \geq KL\left(\mu^{full} || \mu_{t_{loss}}^{img}\right)$$

Stronger lemma: Any extra communication does not hurt.

Λ : a binary sequence of communication availability

μ_Λ^{int}: the probability measure induced by intermittent play

$$C_{\pi^{full}} = KL\left(\mu^{full} || \mu_0^{img}\right) \geq KL\left(\mu^{full} || \mu_\Lambda^{int}\right)$$

Even stronger lemma: Frequent communication is better.

Λ : a Bernoulli(q) process of communication availability

$$C_{\pi^{full}} = KL\left(\mu^{full} || \mu_0^{img}\right) \geq KL\left(\mu^{full} || \mu_\Lambda^{int}\right) / q .$$

Performance guarantees: Imaginary play with adversarial communication loss

Theorem: f is an arbitrary function that determines the communication availability based on the team's joint history.

Communication loss does not affect much if total correlation is low:

$$v^{img} \geq v^{full} - \sqrt{1 - \exp\left(-C_{\pi^{joint}}\right)}.$$

Reachability probability of imaginary play under f

Reachability probability of full communication

Total correlation

Performance guarantees: Imaginary play with structured communication loss

Theorem: Consider a communication system that permanently fails with probability p at every time step.

$$v^{img} \geq \max\left(v^{full} - \sqrt{1 - \exp\left(-C_{\pi^{joint}}\right)},\; v^{full}(1-p)^{\frac{l^{full}}{v^{full}}} \right).$$

Reachability probability of imaginary play

Reachability probability of full communication

Total correlation

Function of expected path length l^{full} under full communication

Performance guarantees:
Intermittent communication with structured communication loss

Theorem: Communication system that fails
with a probability q at any communication step

$$v^{img} \geq \max\left(v^{full} - \sqrt{1 - \exp\left(-qC_{\pi^{joint}}\right)},\quad v^{full}(1-q)^{\frac{l^{full}}{v^{full}}} \right)$$

Reachability probability of imaginary play

Reachability probability of full communication

Effective total correlation

Function of expected path length l^{full} under full communication

Improving the performance:
How to synthesize minimum-dependency policies?

Until this point, π_{joint} is given.

Now, find a good π_{joint}, i.e., a minimum-dependency policy.

Improving the performance: How to synthesize minimum-dependency policies?

Until this point, π_{joint} is given.

Now, find a good π_{joint}, i.e., a minimum-dependency policy.

Ideally maximize

$$\max\left(v^{full} - \sqrt{1 - \exp\left(-C_{\pi^{joint}}\right)}, v^{full}(1-p)^{\frac{l^{full}}{v^{full}}}\right)$$

Improving the performance:
How to synthesize minimum-dependency policies?

Until this point, π_{joint} is given.

Now, find a good π_{joint}, i.e., a minimum-dependency policy.

Ideally maximize

$$\max\left(v^{full} - \sqrt{1 - \exp\left(-C_{\pi^{joint}}\right)}, v^{full}(1-p)^{\frac{l^{full}}{v^{full}}}\right)$$

Too ugly to optimize!

Monotone in all variables.

Improving the performance:
How to synthesize minimum-dependency policies?

Until this point, π_{joint} is given.

Now, find a good π_{joint}, i.e., a minimum-dependency policy.

Ideally maximize

$$\max\left(v^{full} - \sqrt{1 - \exp\left(-C_{\pi^{joint}}\right)}, v^{full}(1-p)^{\frac{l^{full}}{v^{full}}}\right)$$

Too ugly to optimize!

Monotone in all variables.

Instead maximize $v^{full} - \delta\ C_{\pi^{joint}} - \beta\ l^{full}$ where $\delta > 0$ and $\beta > 0$ are constants.

Improving the performance: How to synthesize minimum-dependency policies?

Occupation measure = The expected number of times that a state-action pair is used

$$v^{full} - \delta \; C_{\pi^{joint}} - \beta \; l^{full}$$

Improving the performance:
How to synthesize minimum-dependency policies?

Occupation measure = The expected number of times that a state-action pair is used

$$v^{full} - \delta \; C_{\pi^{joint}} - \beta \; l^{full}$$

can be represented with occupancy measures (under v^{full})

can be represented with occupancy measures (under l^{full})

Improving the performance: How to synthesize minimum-dependency policies?

Occupation measure = The expected number of times that a state-action pair is used

$$v^{full} - \delta \; C_{\pi^{joint}} - \beta \; l^{full}$$

can be represented with occupancy measures (v^{full})

can be represented with occupancy measures (l^{full})

$$\left(\sum_{i=1}^{N} H(X^i) \right) - H(\mathbf{X})$$

entropy of a hidden Markov model has no analytical form due to non-stationarity

can be represented with occupancy measures if π_{joint} is stationary

Improving the performance:
How to synthesize minimum-dependency policies?

$\bar{X}^i$ = the stationary process that shares the same occupancy measures with X^i

Fact: $\bar{C}_{\pi^{joint}} := \left(\sum_{i=1}^{N} H(\bar{X}^i) \right) - H(\mathbf{X}) \geq C_{\pi^{joint}} = \left(\sum_{i=1}^{N} H(X^i) \right) - H(\mathbf{X})$

Improving the performance:
How to synthesize minimum-dependency policies?

$\bar{X}^i$ = the stationary process that shares the same occupancy measures with X^i

Fact: $\bar{C}_{\pi^{joint}} := \left(\sum_{i=1}^{N} H(\bar{X}^i) \right) - H(\mathbf{X}) \geq C_{\pi^{joint}} = \left(\sum_{i=1}^{N} H(X^i) \right) - H(\mathbf{X})$

$$v^{full} - \delta\ \bar{C}_{\pi^{joint}} - \beta\ l^{full}$$

can be represented with occupancy measures

can be represented with occupancy measures

$$\left(\sum_{i=1}^{N} H(\bar{X}^i) \right) - H(\mathbf{X})$$

can be represented with occupancy measures

can be represented with occupancy measures if π_{joint} is stationary

Improving the performance:
Synthesize via non-convex optimization

Linear | Concave | Convex

$$\max \; v^{full} - \delta \left(\sum_{i=1}^{N} H(\bar{X}^i) \right) + \delta \, H(\mathbf{X}) - \beta \; l^{full}$$

subject to dynamics

Use convex-concave procedure for synthesis.

Back to the valley example:
Optimal centralized policy (baseline) with full communication

Back to the valley example:
Optimal centralized policy (baseline) with full communication

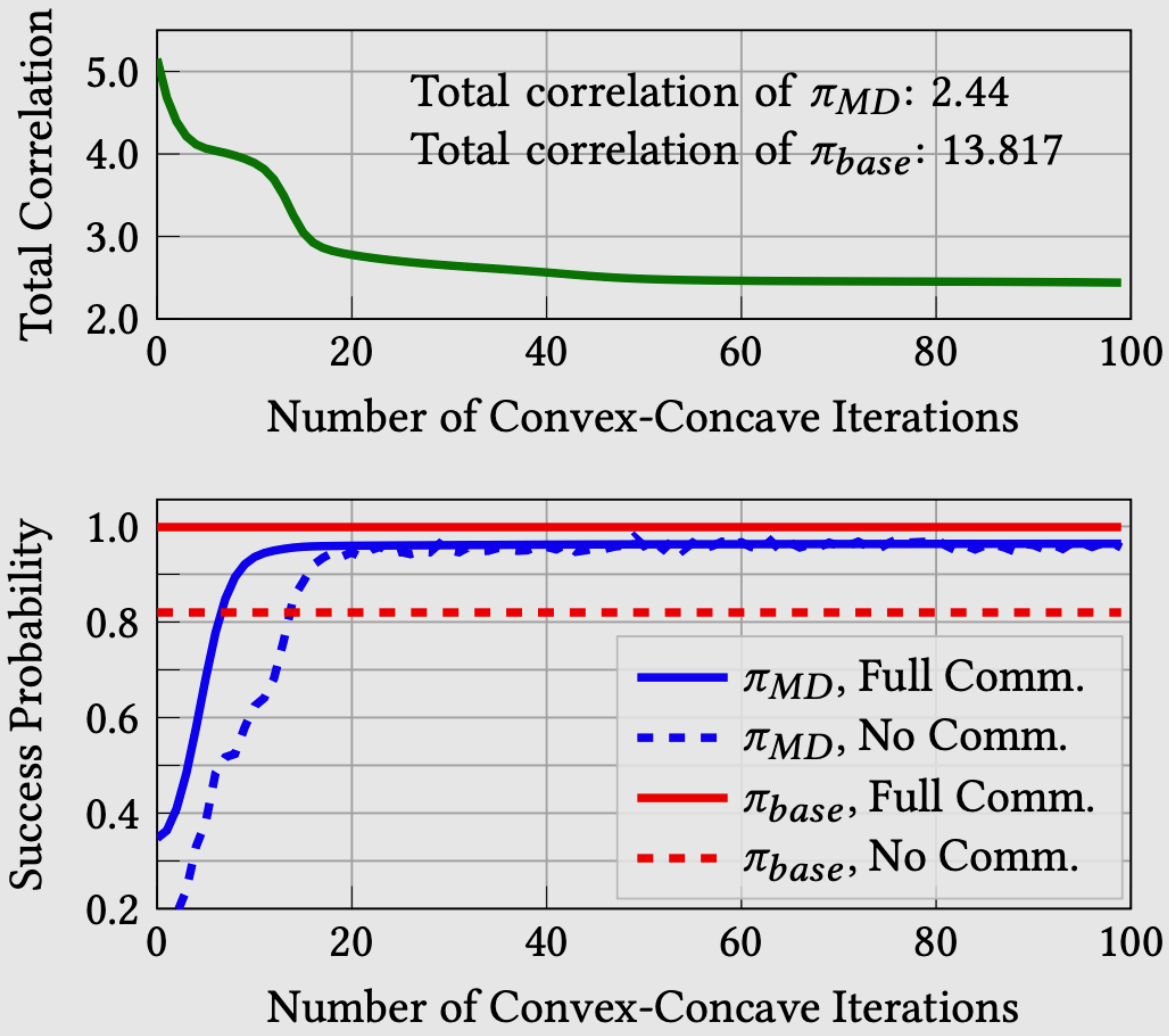
Back to the valley example:
Optimal centralized policy (baseline) with no communication

Back to the valley example:
Optimal centralized policy (baseline) with no communication

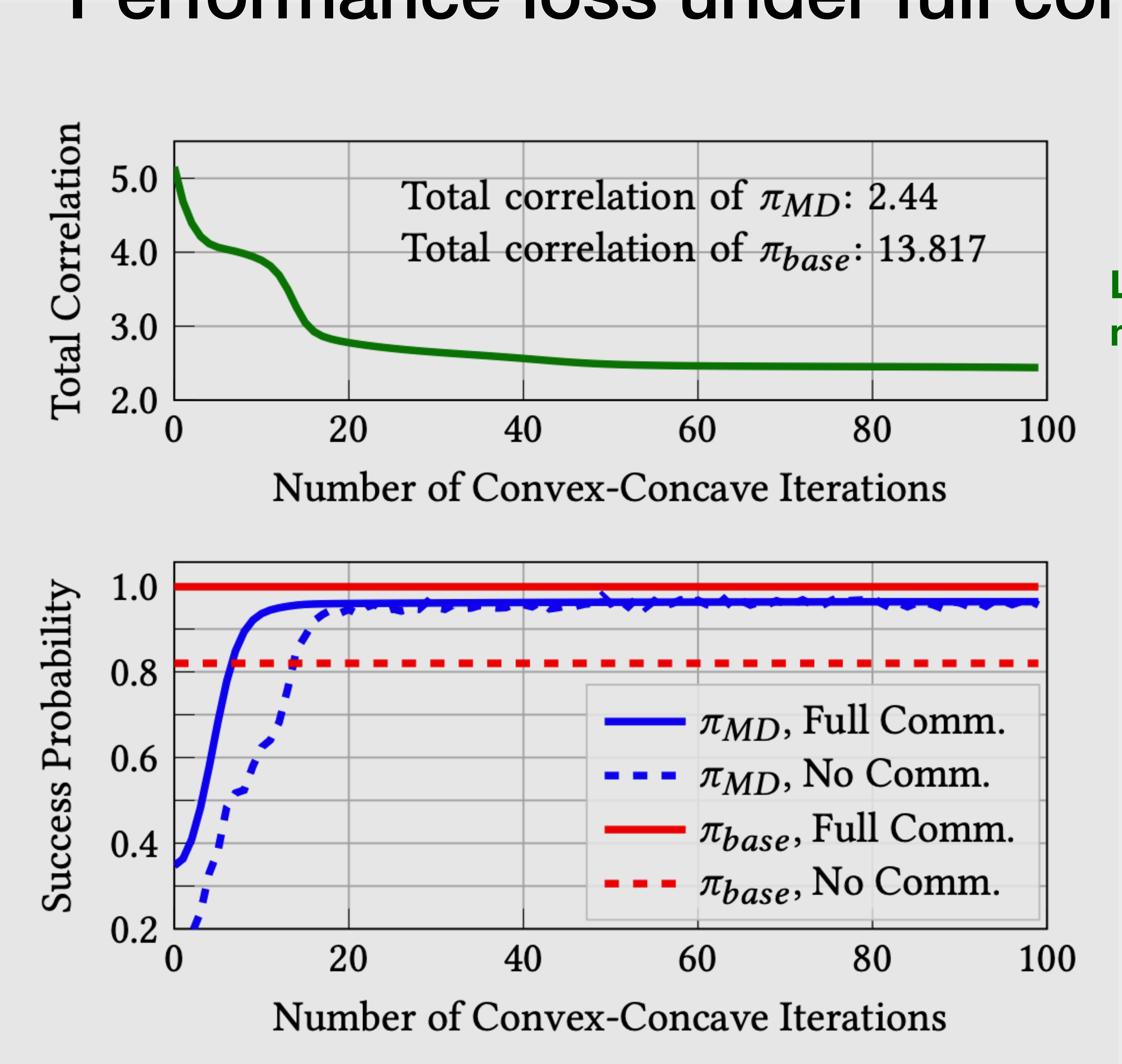
Back to the valley example: Minimum-dependency policy (ours) with no communication

Back to the valley example:
Minimum-dependency policy (ours) with no communication

Performance loss under full communication loss

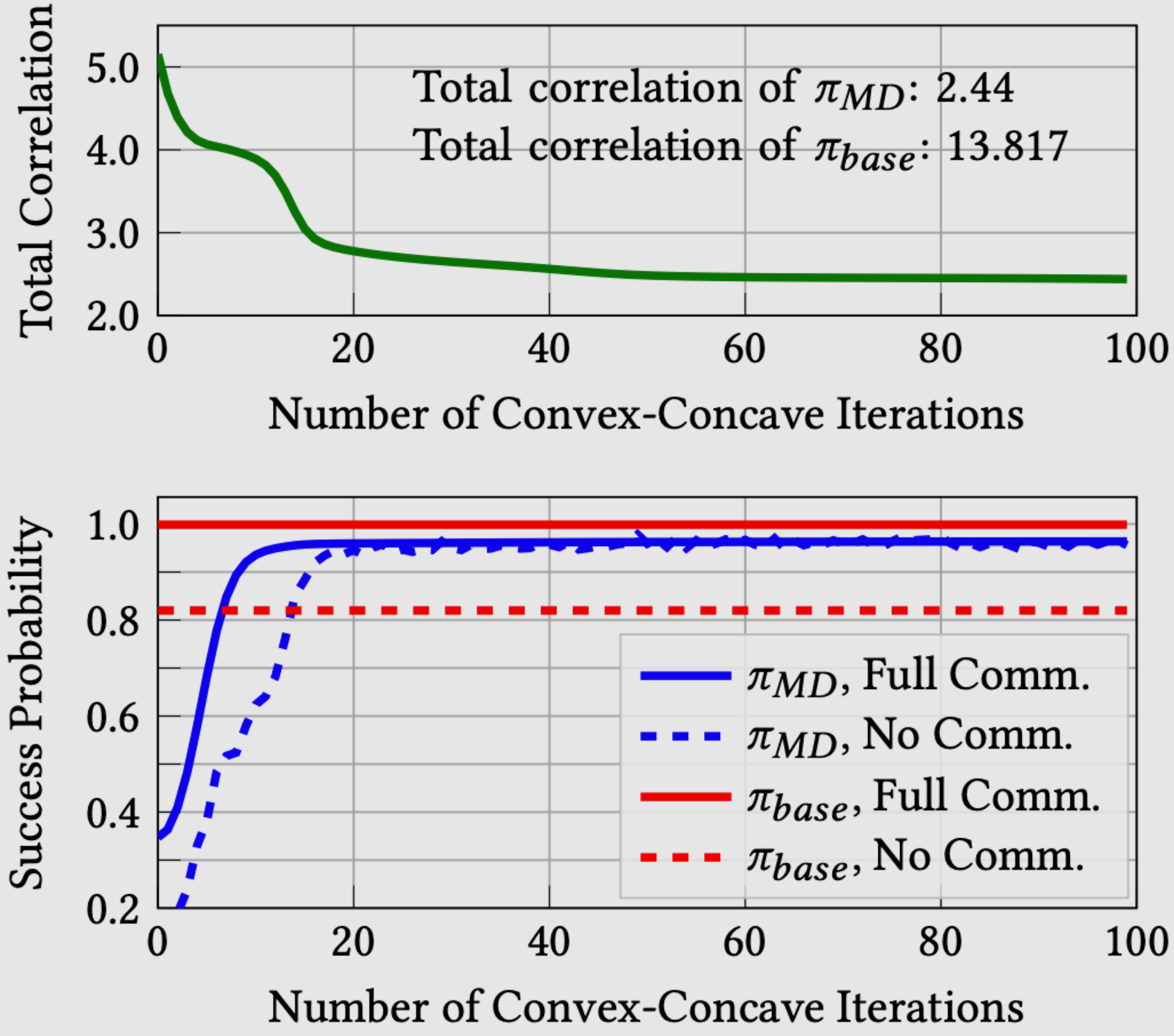


Performance loss under full communication loss



Low total correlation for minimum-dependency policy

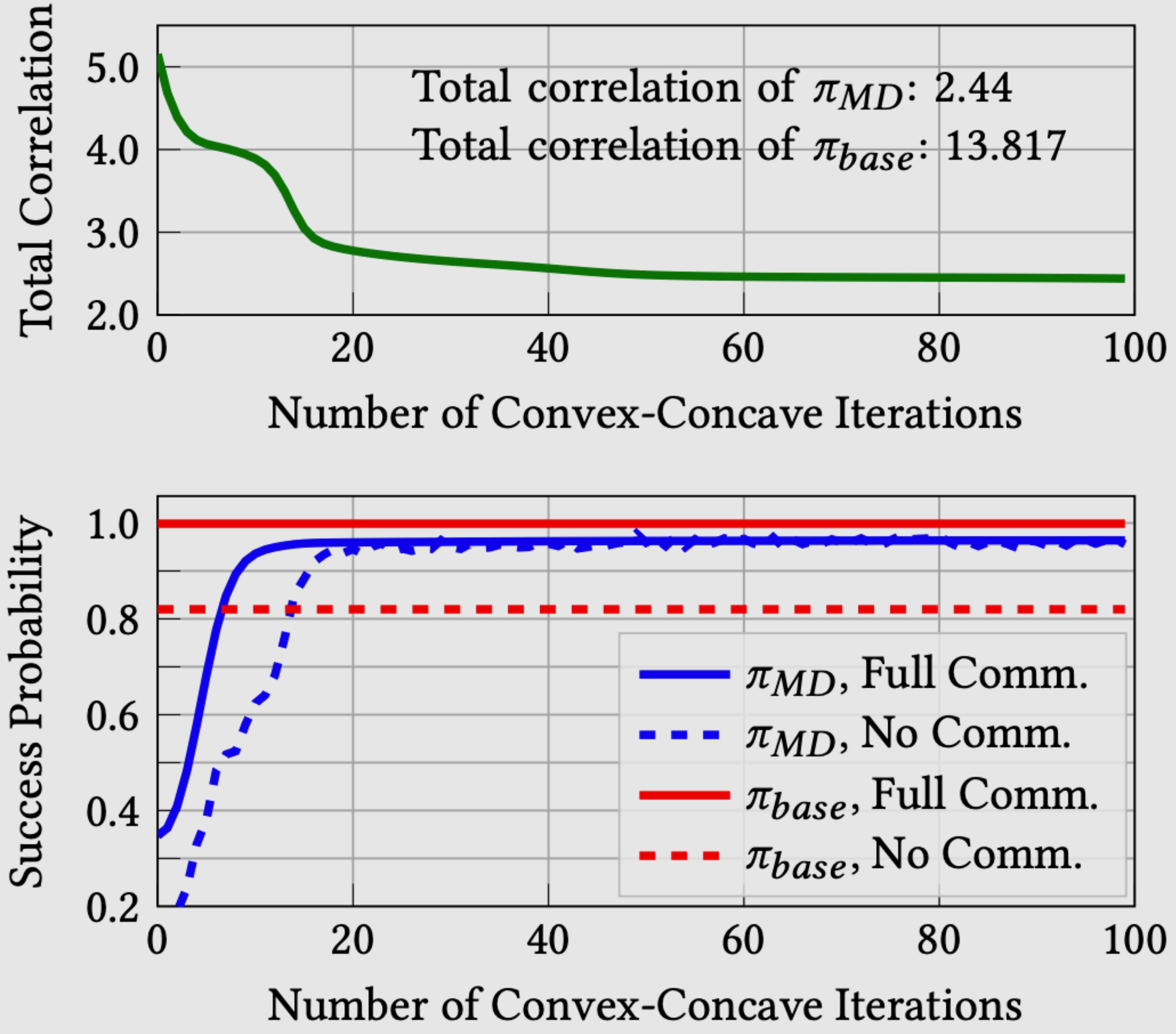
Performance loss under full communication loss



Low total correlation for minimum-dependency policy

Consistent performance for minimum-dependency policy

Performance loss under full communication loss

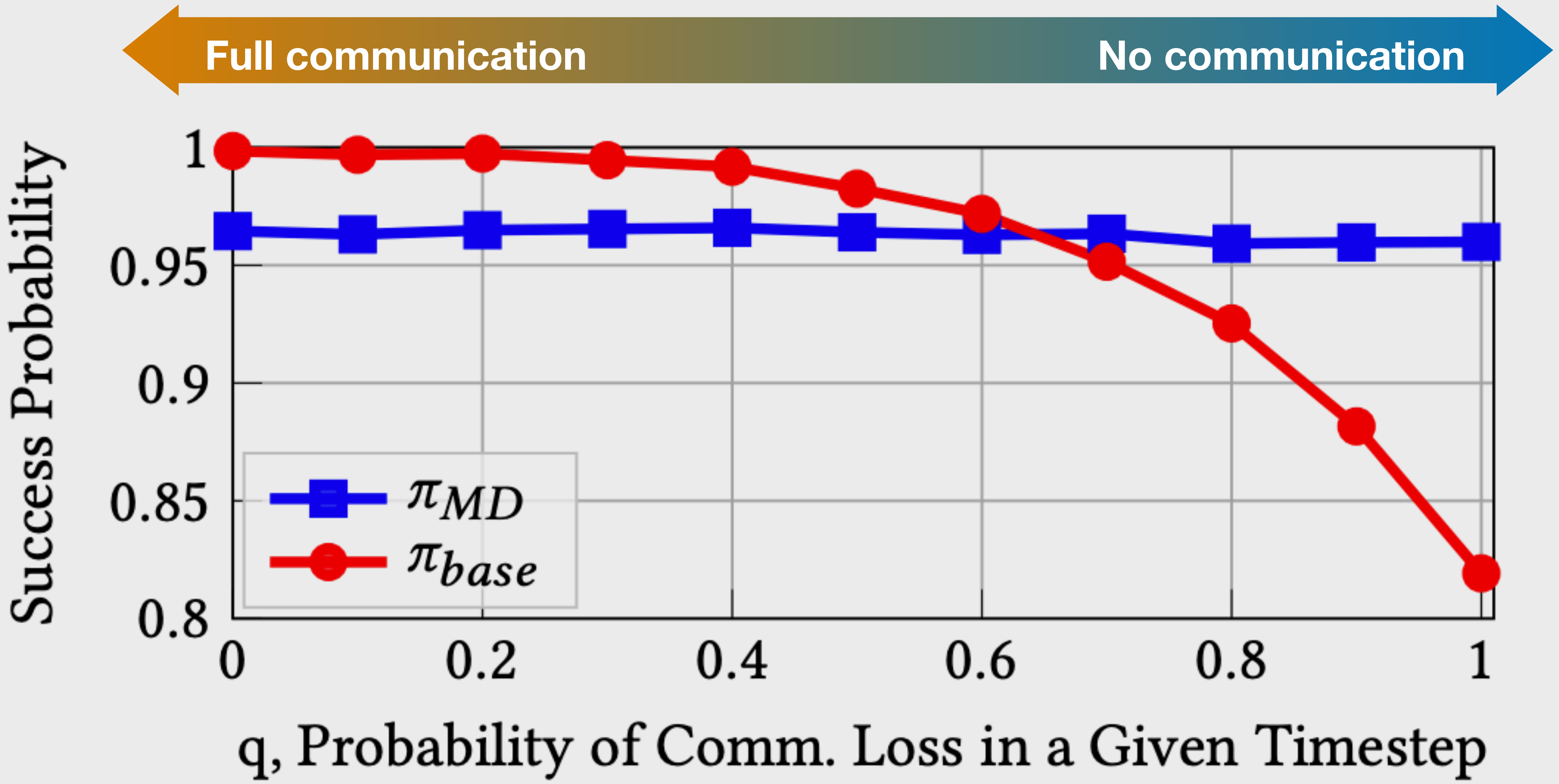


Low total correlation for minimum-dependency policy

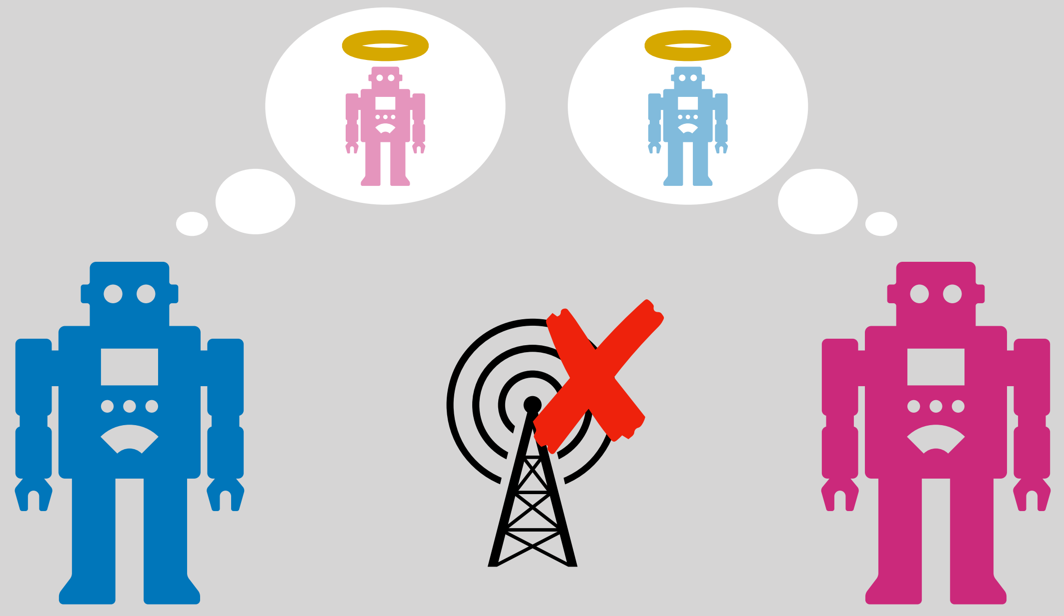
Consistent performance for minimum-dependency policy

20% performance drop for baseline policy

Performance loss under intermittent communication loss



Policy Execution Algorithm



Performance Guarantees

$$v^{img} \geq v^{full} - \sqrt{1 - \exp\left(-C_{\pi^{joint}}\right)}$$

Resulting Behavior

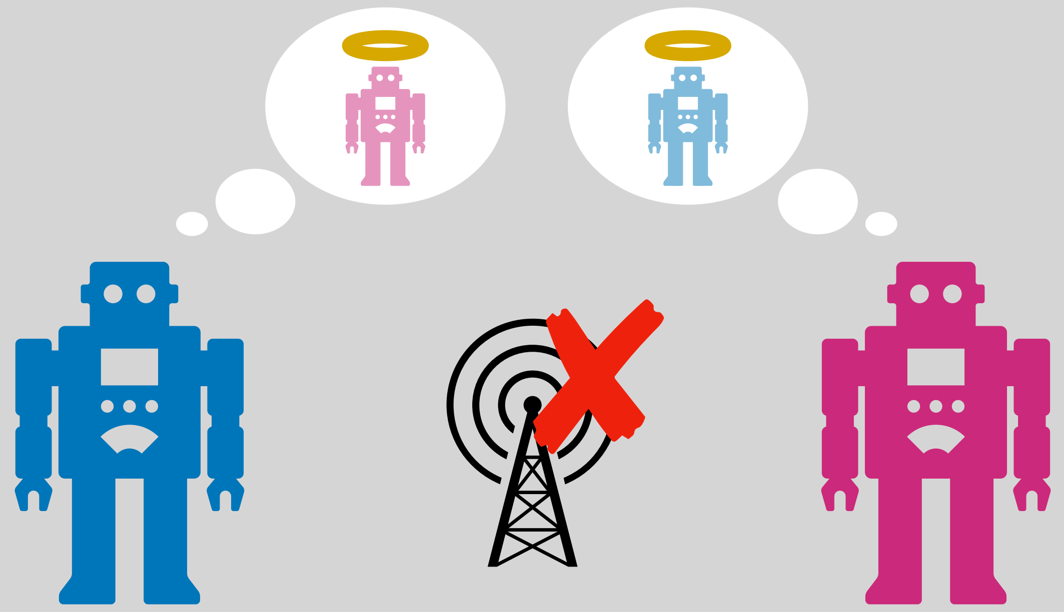
Policy Optimization

$$\max \; v^{full} - \delta \left(\sum_{i=1}^{N} H(\bar{X}^i) \right) + \delta \, H(\mathbf{X}) - \beta \; l^{full}$$

subject to dynamics

Mustafa O. Karabag (karabag@utexas.edu) & Cyrus Neary (cneary@utexas.edu)

Policy Execution Algorithm



Performance Guarantees

$$v^{img} \geq v^{full} - \sqrt{1 - \exp\left(-C_{\pi^{joint}}\right)}$$

Resulting Behavior

Policy Optimization

$$\max \; v^{full} - \delta \left(\sum_{i=1}^{N} H(\bar{X}^{i}) \right) + \delta \; H(\mathbf{X}) - \beta \; l^{full}$$

subject to dynamics

Mustafa O. Karabag (karabag@utexas.edu) & Cyrus Neary (cneary@utexas.edu)