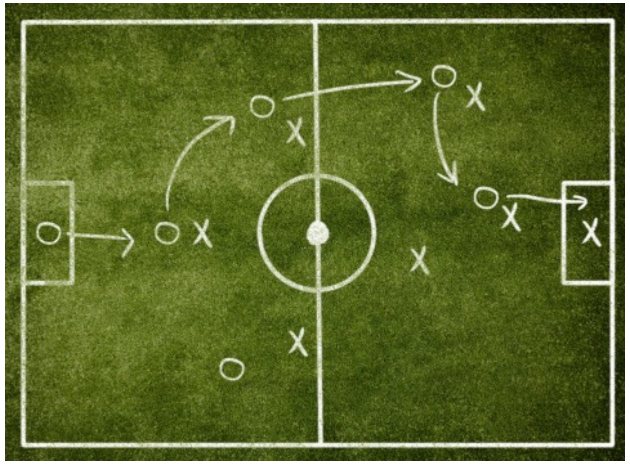
Collaborative Embodied Agents

AI Agents that can collaborate in (virtual) visual worlds

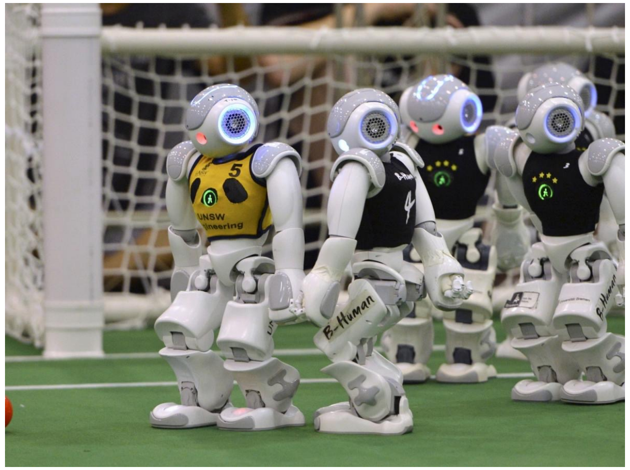
Unnat Jain

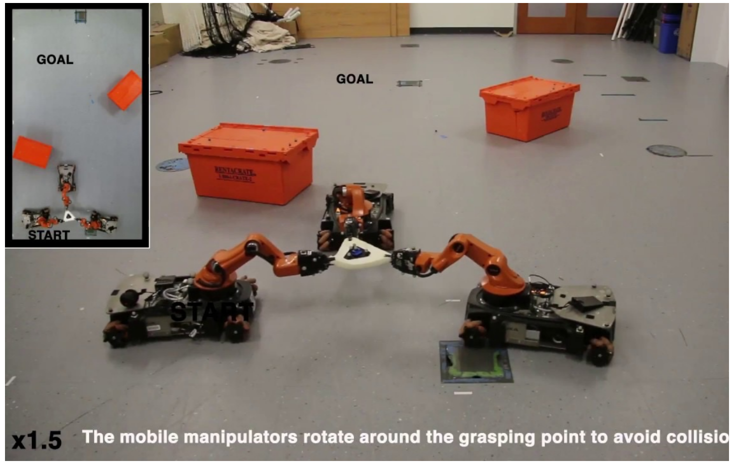
ILLINOIS

We live in a collaborative world



Collaboration in Robotics



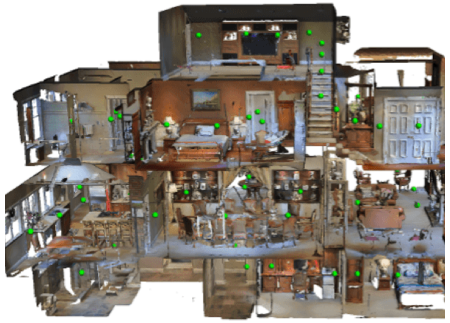


✅ **Low-level control**

⛔ **High-level behaviors**:
Collaboration and communication
Long-horizon planning
Complex interactive tasks

(Virtual) Embodied Environments

AI2THOR (Kolve et al. 2017)



Matterport3D (Chang et al. 2017)

Gibson (Xia et al. 2018)

Low-level control

[Abstracted away]

High-level behaviors:
Collaboration and communication
Long-horizon planning
Complex interactive tasks

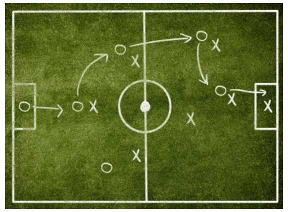
Embodied Agents

Image credits: AIHabitat

Visual Navigation	Zhu et al. 2016, Gupta et al. 2017
Instruction Following	Anderson et al. 2018
Question Answering	Das et al. 2018, Gordon et al. 2018

Low-level control

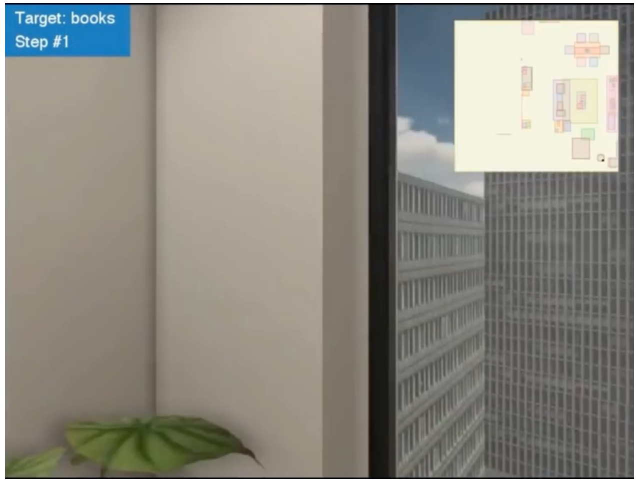
[Abstracted away]



High-level behaviors:
Collaboration and communication
Long-horizon planning
Complex interactive tasks

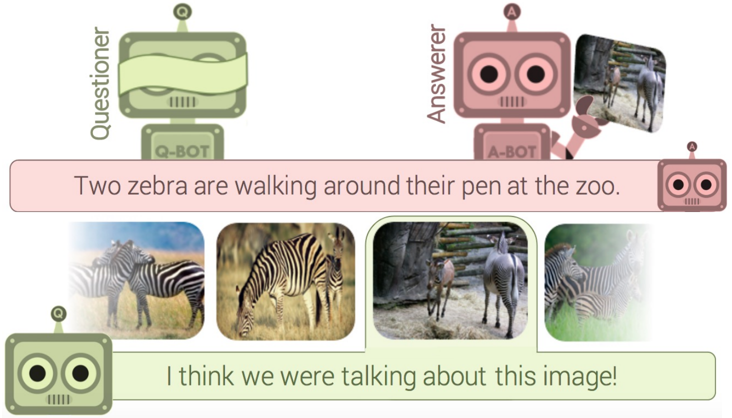
Prior work

Visually navigate



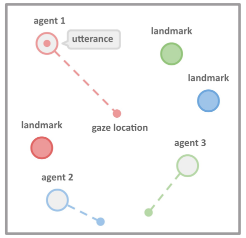
Single agent

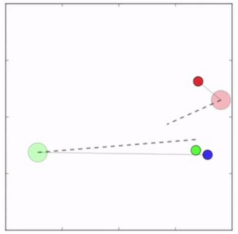
Communicate



Static

Coordinate





Simple/abstract

Zhu et al. ICRA 2017
Gupta et al. CVPR 2017
Das et al. ICCV 2017
Jain et al. CVPR 2018
Lowe et al. NeurIPS 2017
Mordatch and Abbeel AAAI 2018

Collaborative Embodied Agents

Two Body Problem
CVPR 2019

SYNC Policies
ECCV 2020

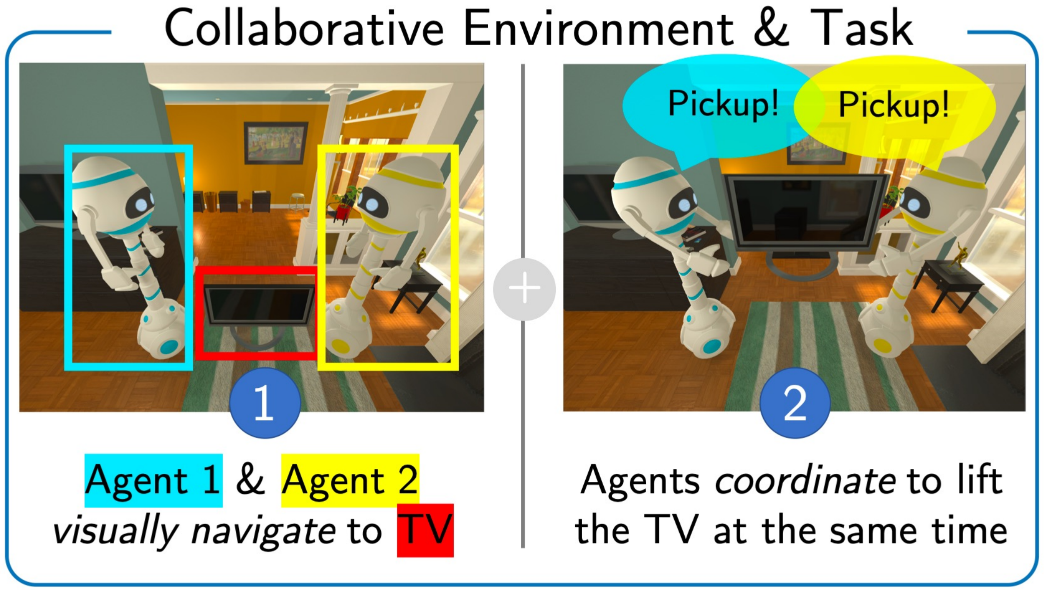
GridToPix
arXiv

Collaborative Embodied Agents

Two Body Problem
CVPR 2019

SYNC Policies
ECCV 2020

GRIDTOPIX
arXiv



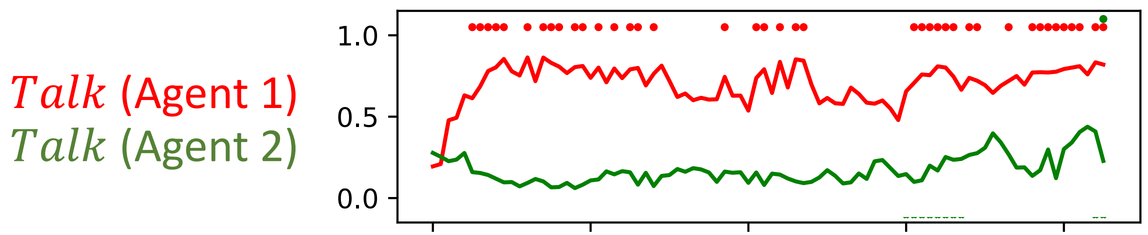
1. First collaborative embodied task - FurnLift

Collaborative Embodied Agents

Two Body Problem
CVPR 2019

SYNC Policies
ECCV 2020

GRIDTOPIX
arXiv



"I am near TV!"

2. Interpretation of emergent communication

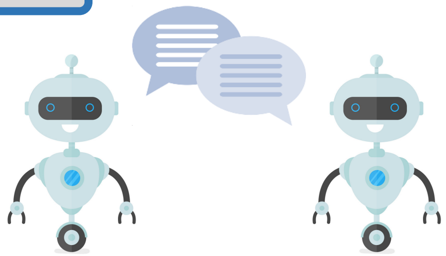
Collaborative Embodied Agents

Two Body Problem
CVPR 2019

SYNC Policies
ECCV 2020

GRIDTOPIX
arXiv

Explicit



Explicitly sending messages to communicate

Implicit

Visibility of other agent communicates information

3. Effect of communication

Collaborative Embodied Agents

Two Body Problem
CVPR 2019

SYNC Policies
ECCV 2020

GRIDTOPIX
arXiv

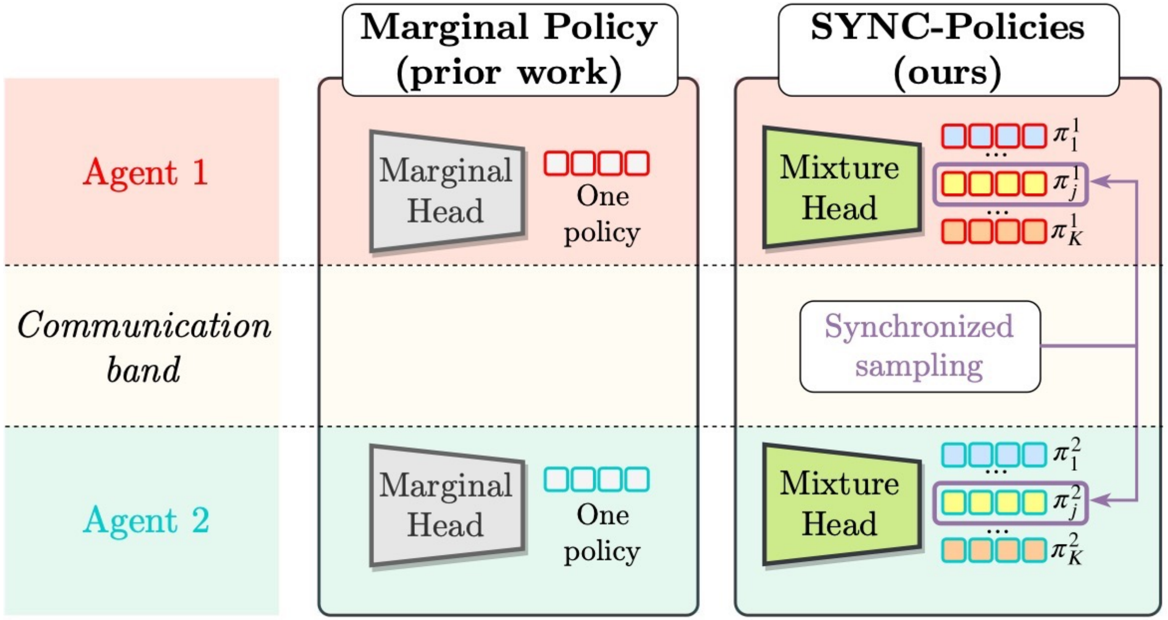
4. Intricately coordinated embodied task - FurnMove

Collaborative Embodied Agents

Two Body Problem
CVPR 2019

SYNC Policies
ECCV 2020

GRIDTOPIX
arXiv



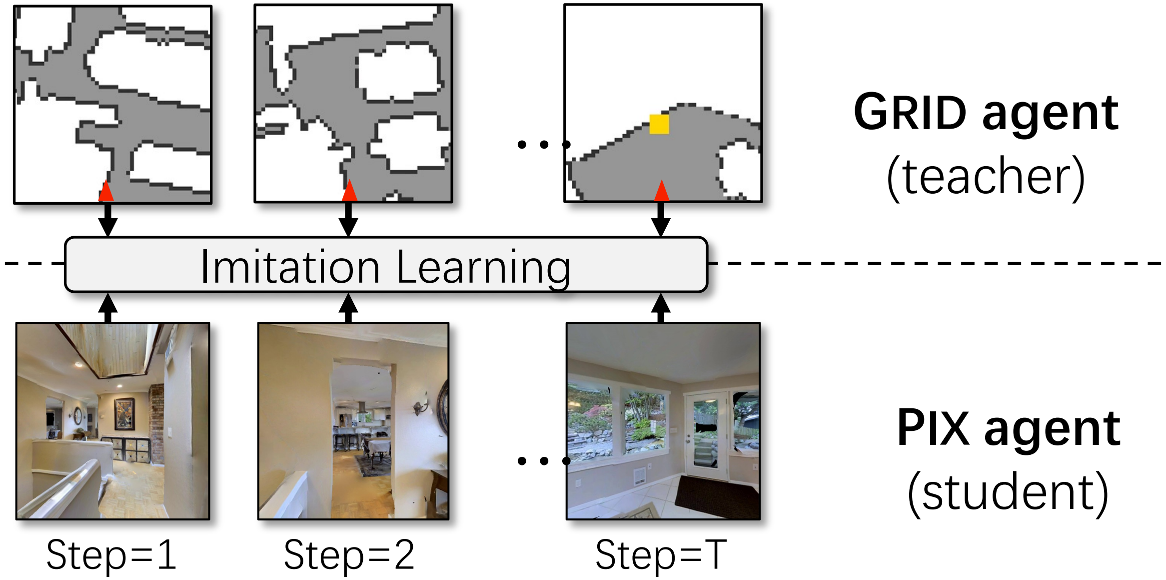
5. Richer representation of multi-agent policy

Collaborative Embodied Agents

Two Body Problem
CVPR 2019

SYNC Policies
ECCV 2020

GRIDTOPIX
arXiv



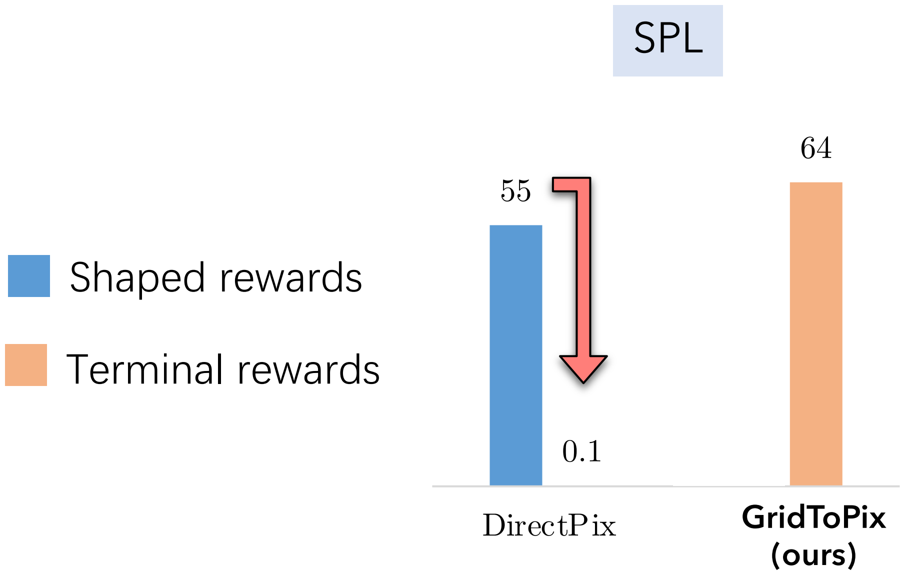
6. Teacher-Student learning

Collaborative Embodied Agents

Two Body Problem
CVPR 2019

SYNC Policies
ECCV 2020

GRIDTOPIX
arXiv



6. Learning policies from minimal supervision

Collaborative Embodied Agents

Two Body Problem
CVPR 2019

1. First collaborative embodied task – FurnLift
2. Interpretation of emergent communication
3. Effect of communication

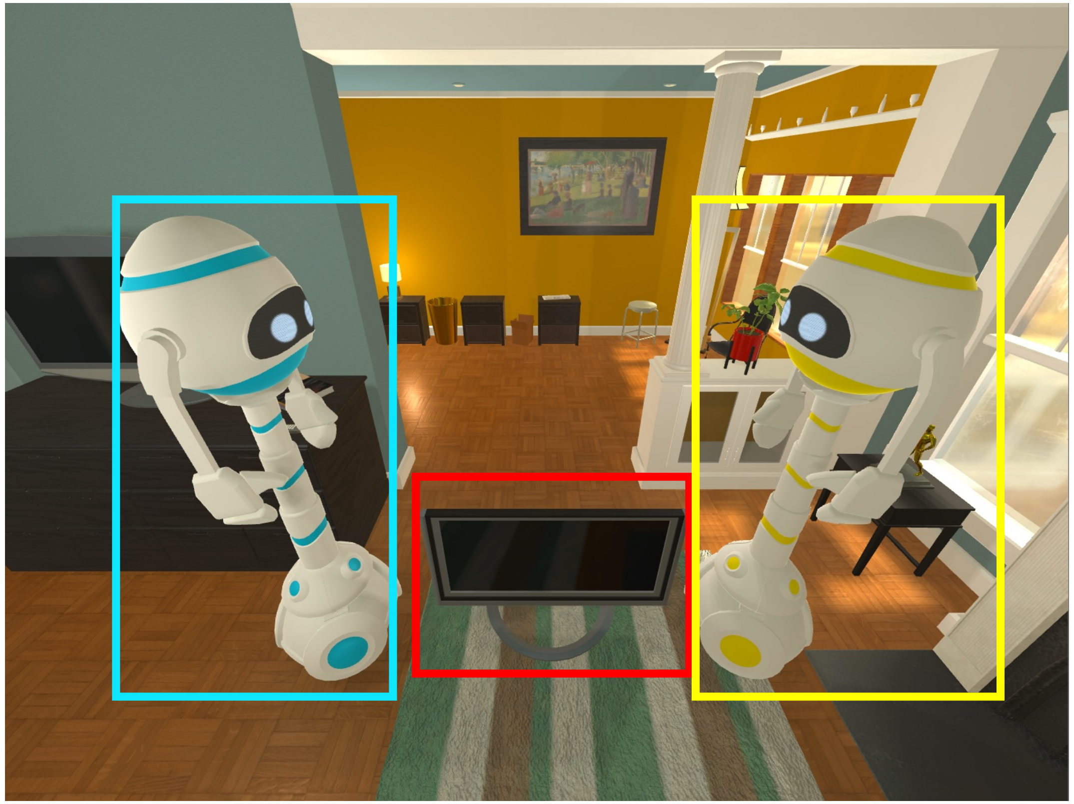
SYNC Policies
ECCV 2020

4. Intricately coordinated embodied task – FurnMove
5. Richer representation of multi-agent policy

GRIDTOPIX
arXiv

6. Learning policies from minimal supervision
7. Leveraging perfect-perception gridworlds for training

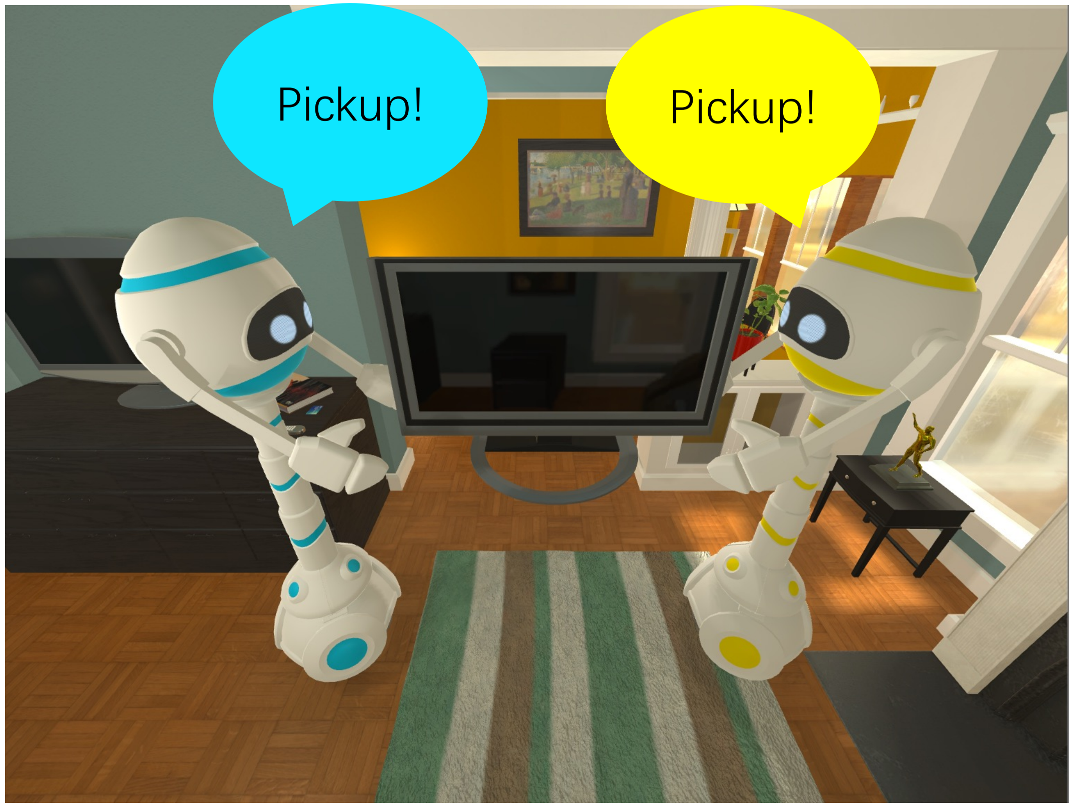
FurnLift Task



* Agents have only egocentric visual inputs

Agent 1 and Agent 2 *visually navigate* to the TV

FurnLift Task



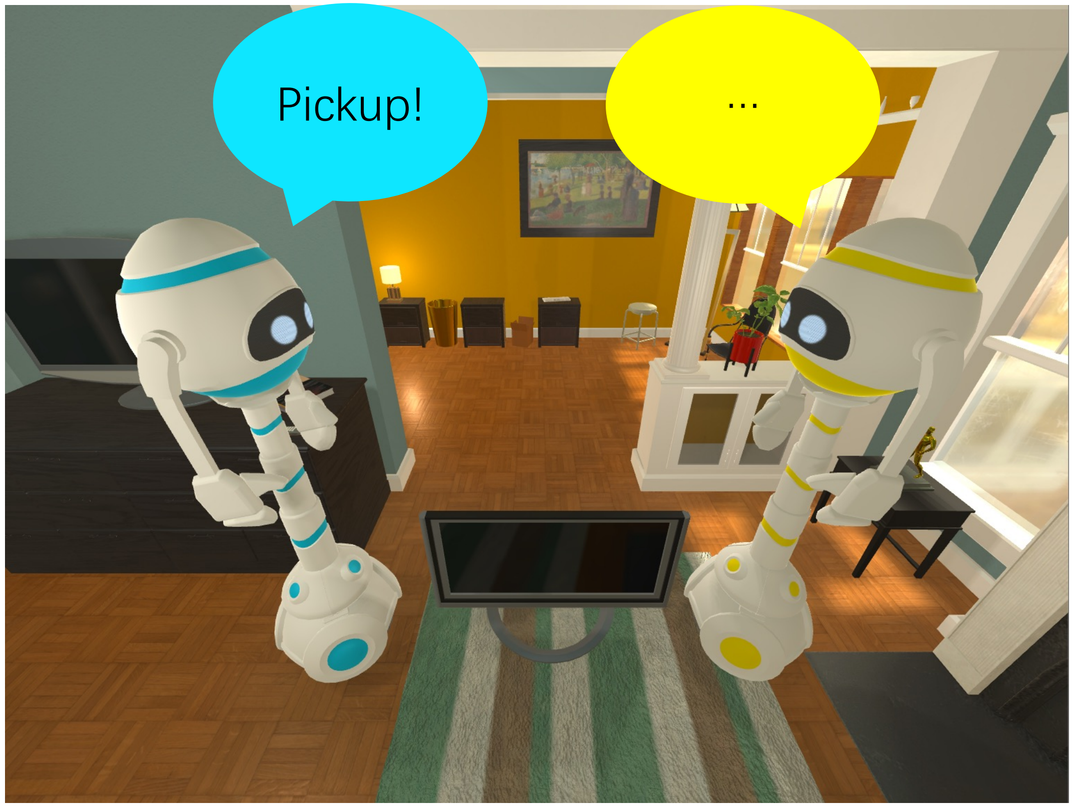
* Agents have only egocentric visual inputs

Agent 1 and Agent 2 *visually navigate* to the TV

+

Agents *coordinate* to lift/pickup the TV at the same time

FurnLift Task



* Agents have only egocentric visual inputs

Previous tasks:

- Navigation in static mesh
- Stationary environments

Furniture Lifting

✓ Interaction & coordination
✓ Non-stationary environment & credit assignment

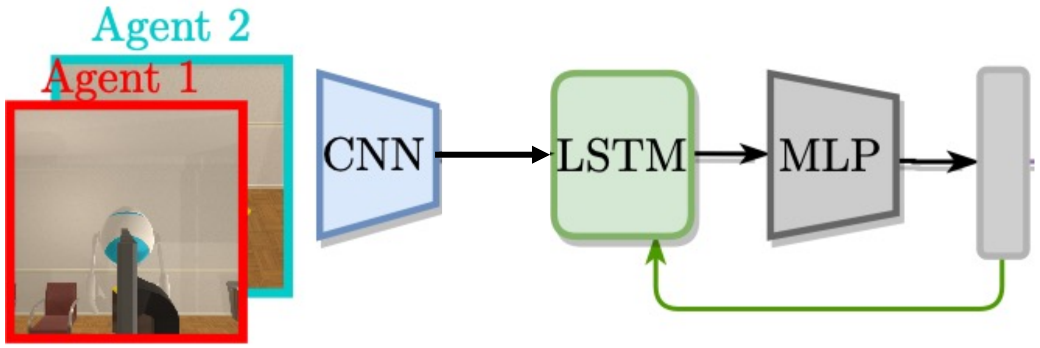
Agent observations

Top-down view

Not available to the agents

(for illustration only)

Agent Policy for FurnLift



(Pooled observations)

Central agent

Decentralized agent

Does not scale

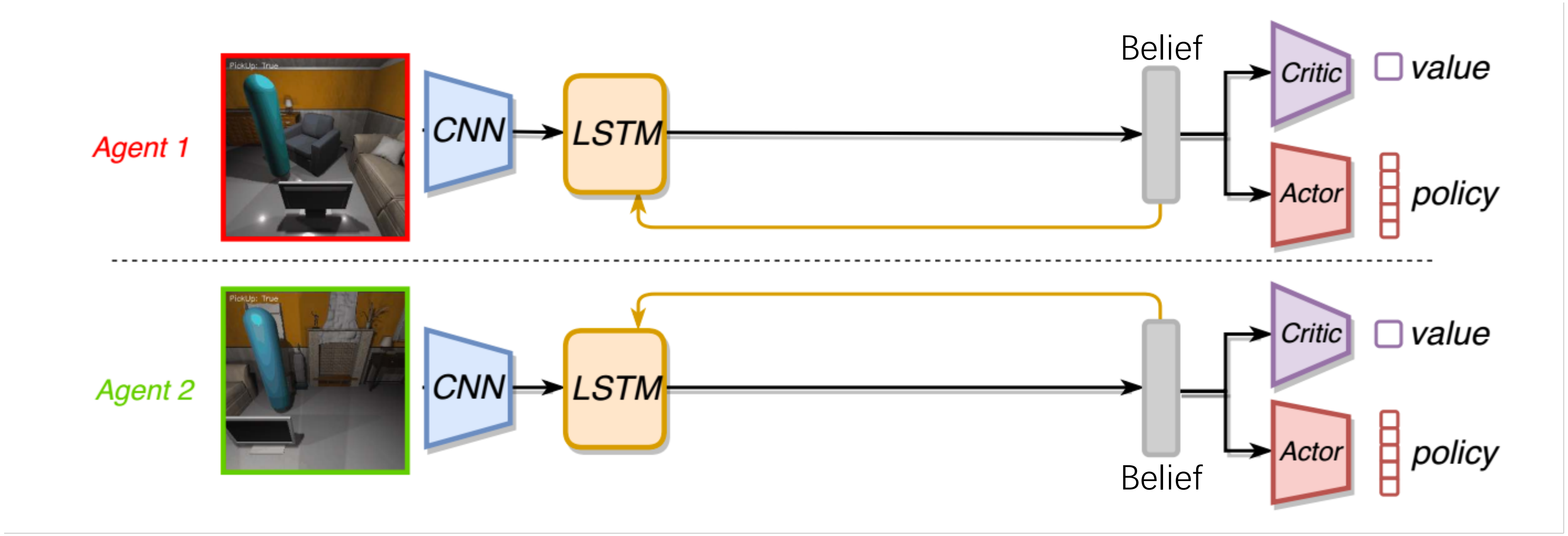
- ✗ Model complexity
- ✗ Policy parameters

(Outer product space)

Communication issues

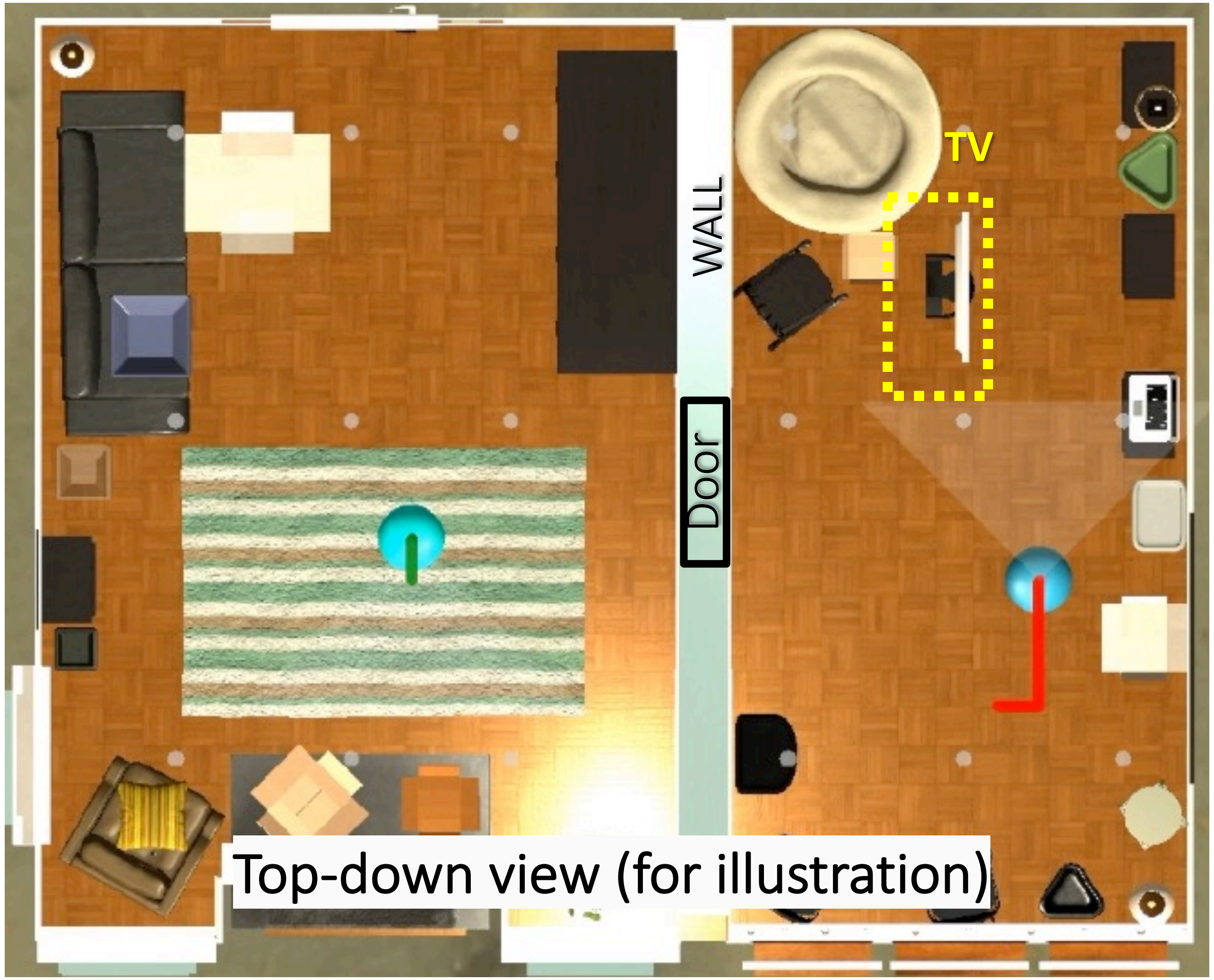
- ✗ High bandwidth
- ✗ Lost packets

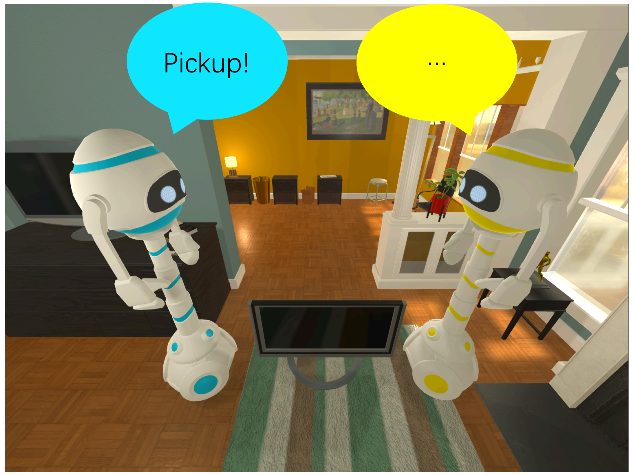
Agent Policy for FurnLift



Decentralized agent

FurnLift Task





Agent 1 and Agent 2

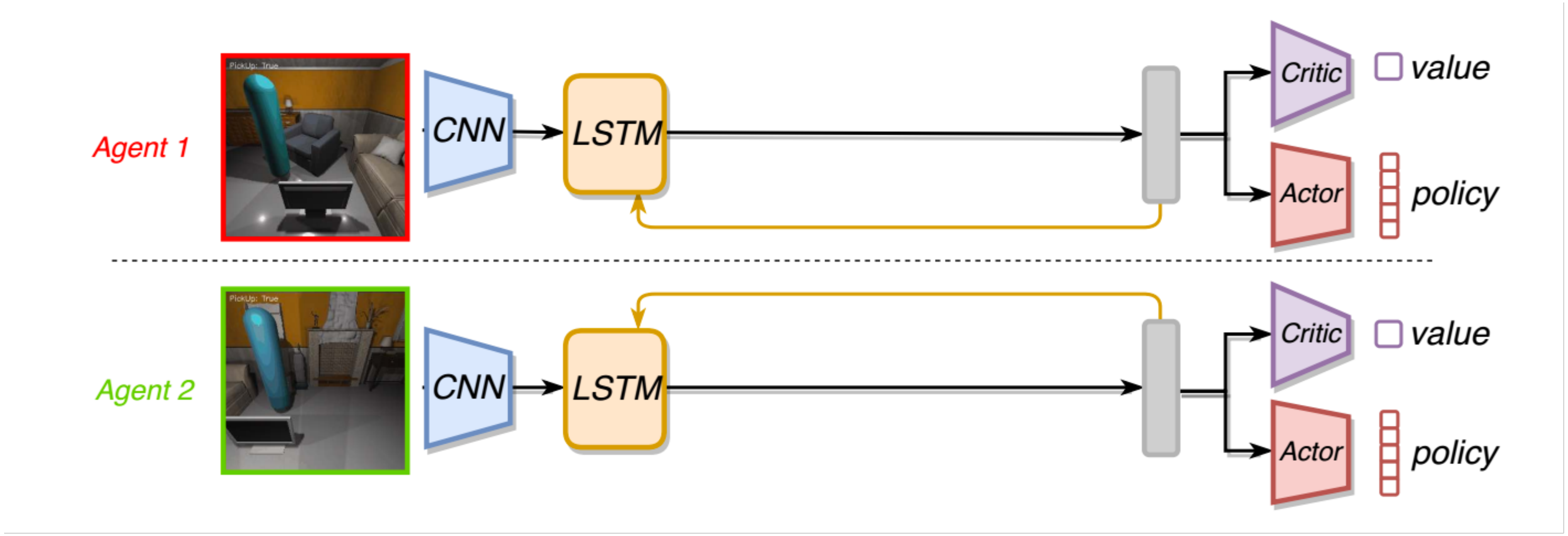
1. Navigate to TV
2. Team Lift

Agent 1 quickly finds it

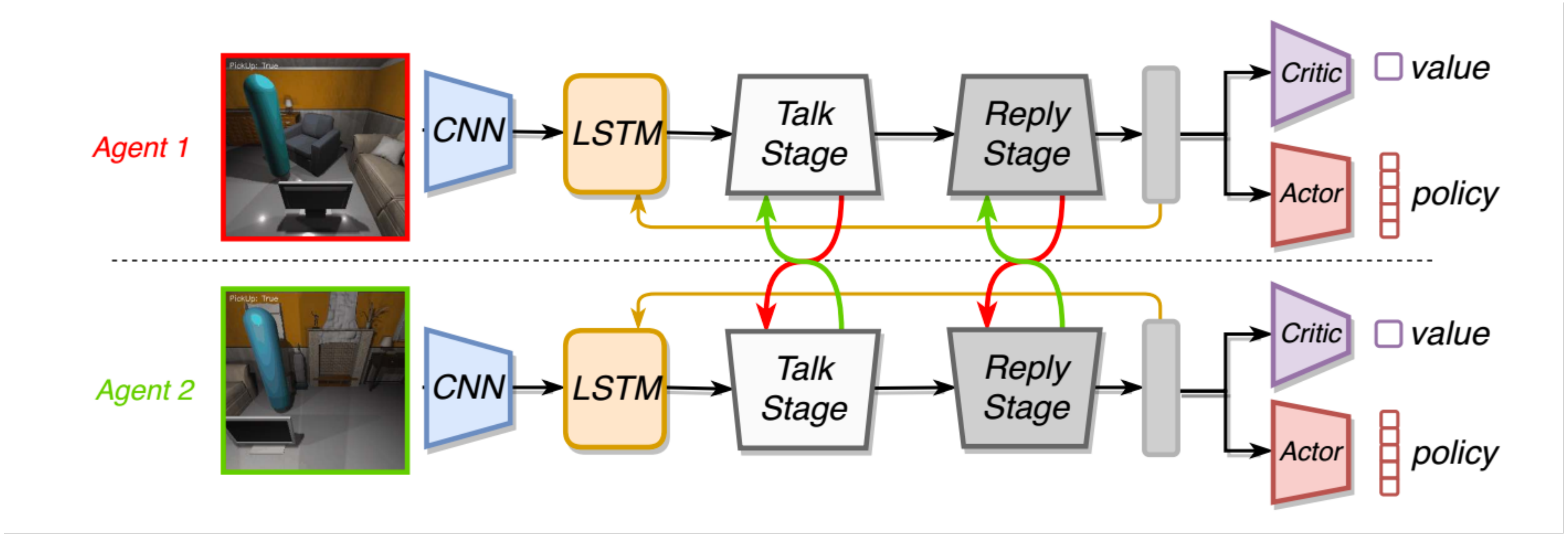
Agent 2 is on the wrong side

Need for communication

Two Body Network

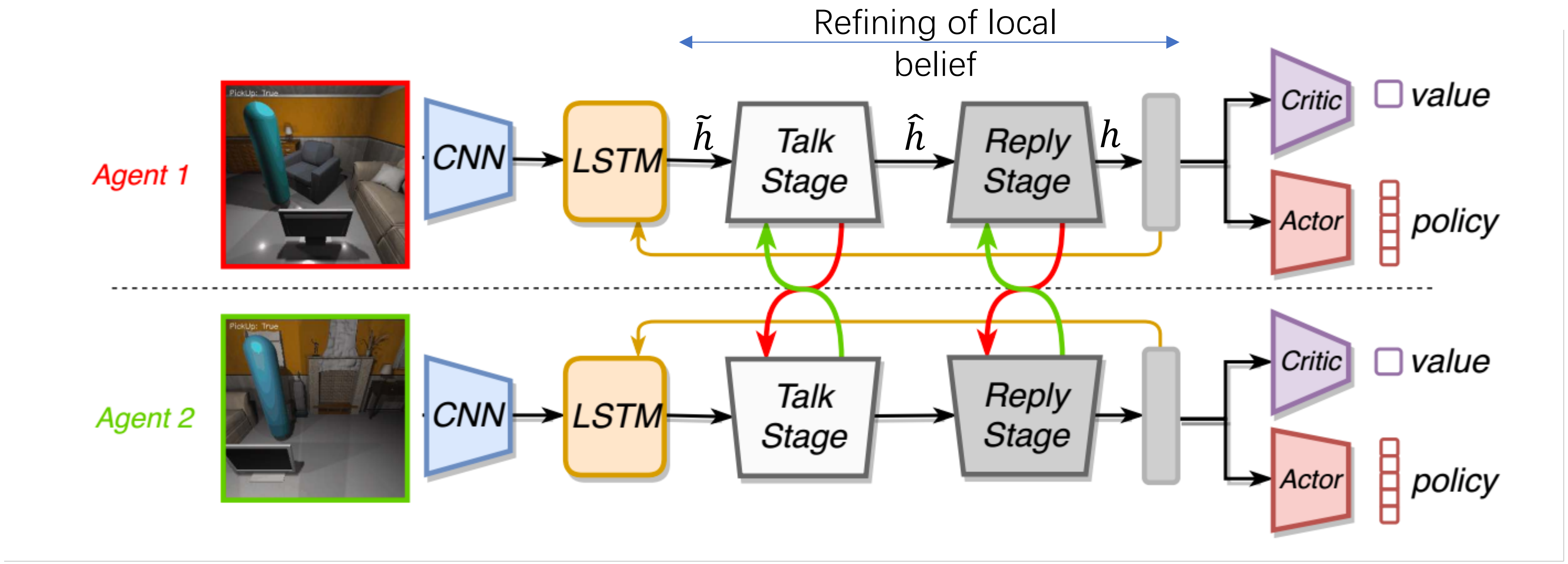


Two Body Network



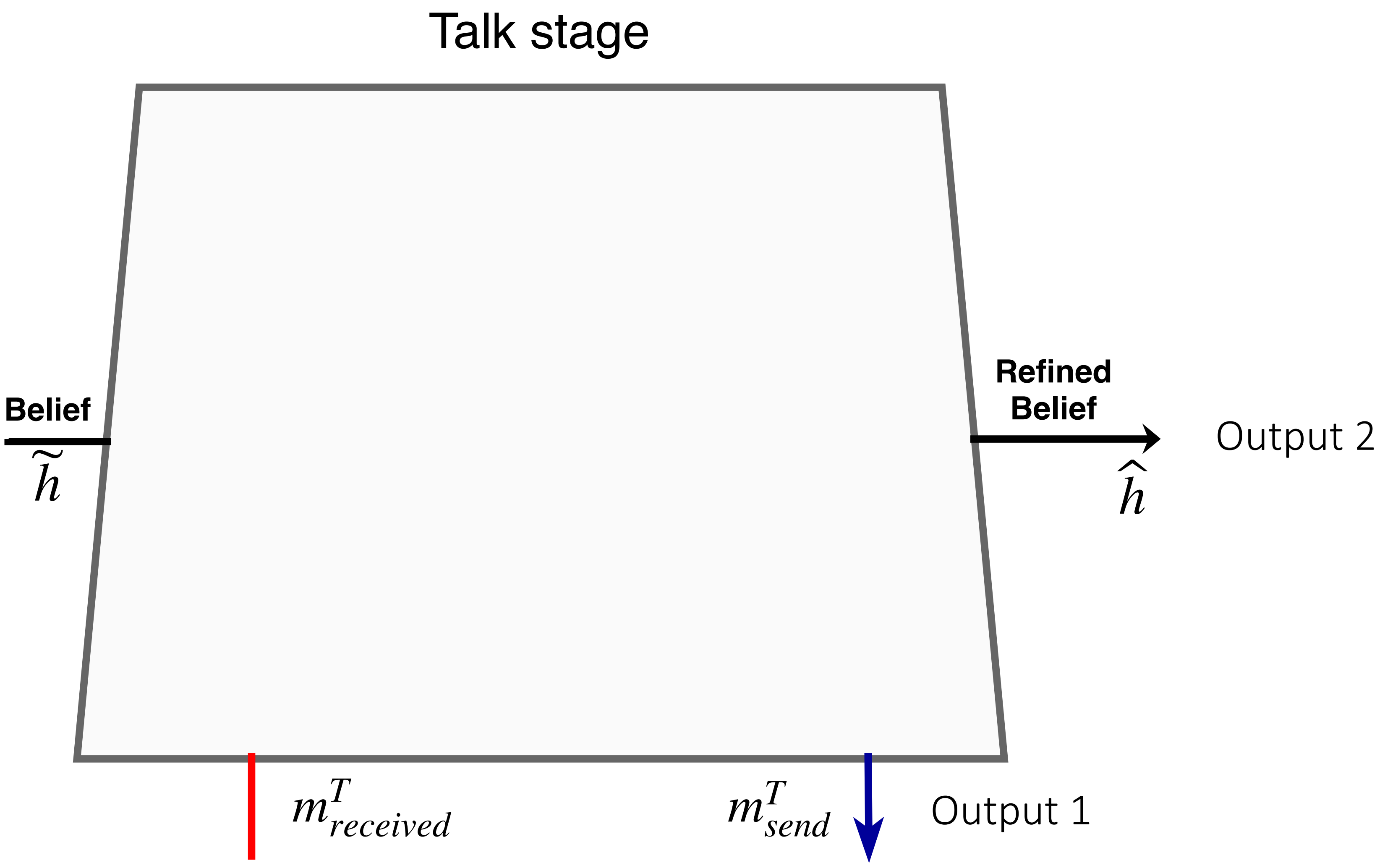
Message-based or 'explicit communication'

Two Body Network

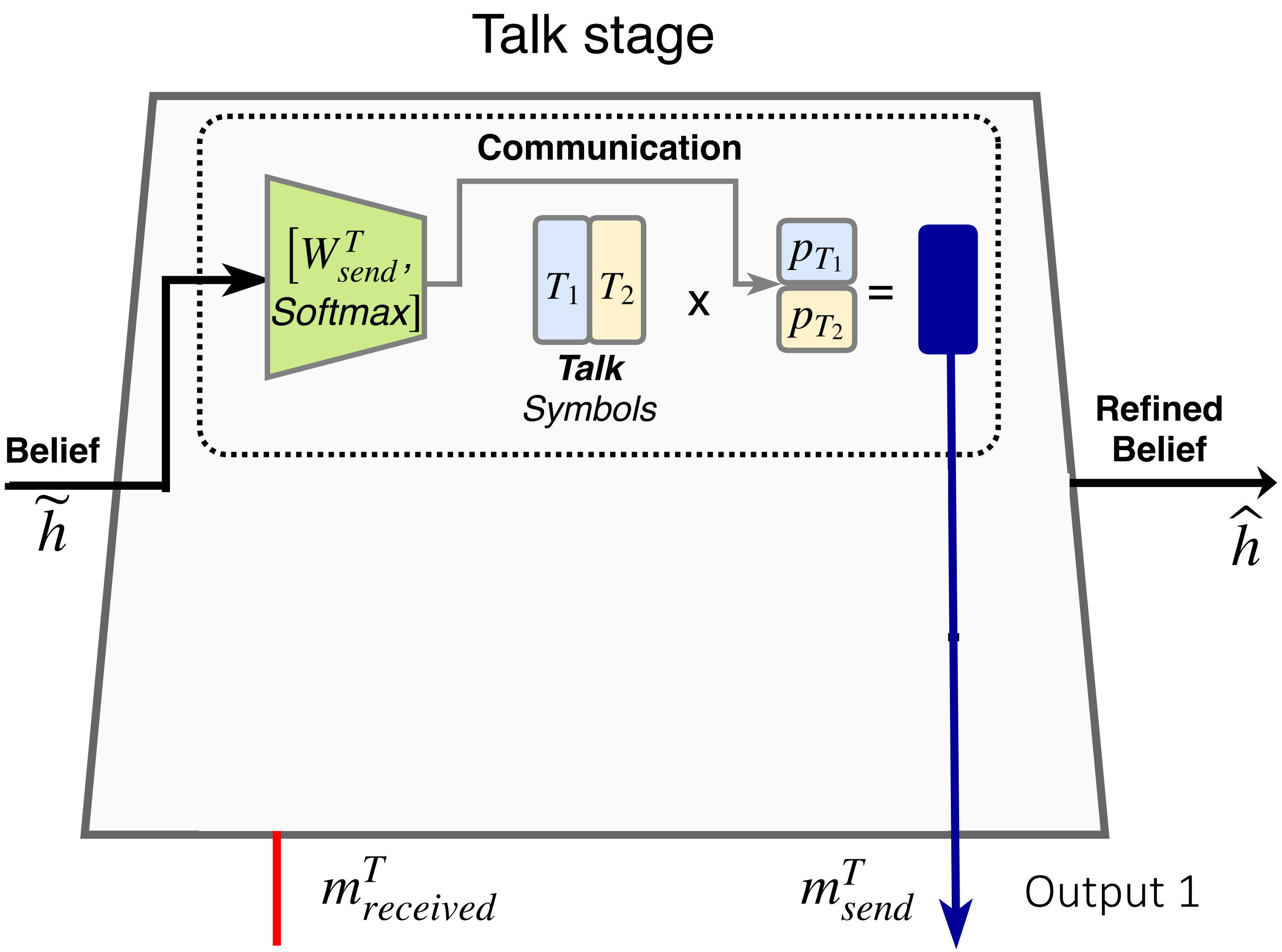


Message-based or 'explicit communication'

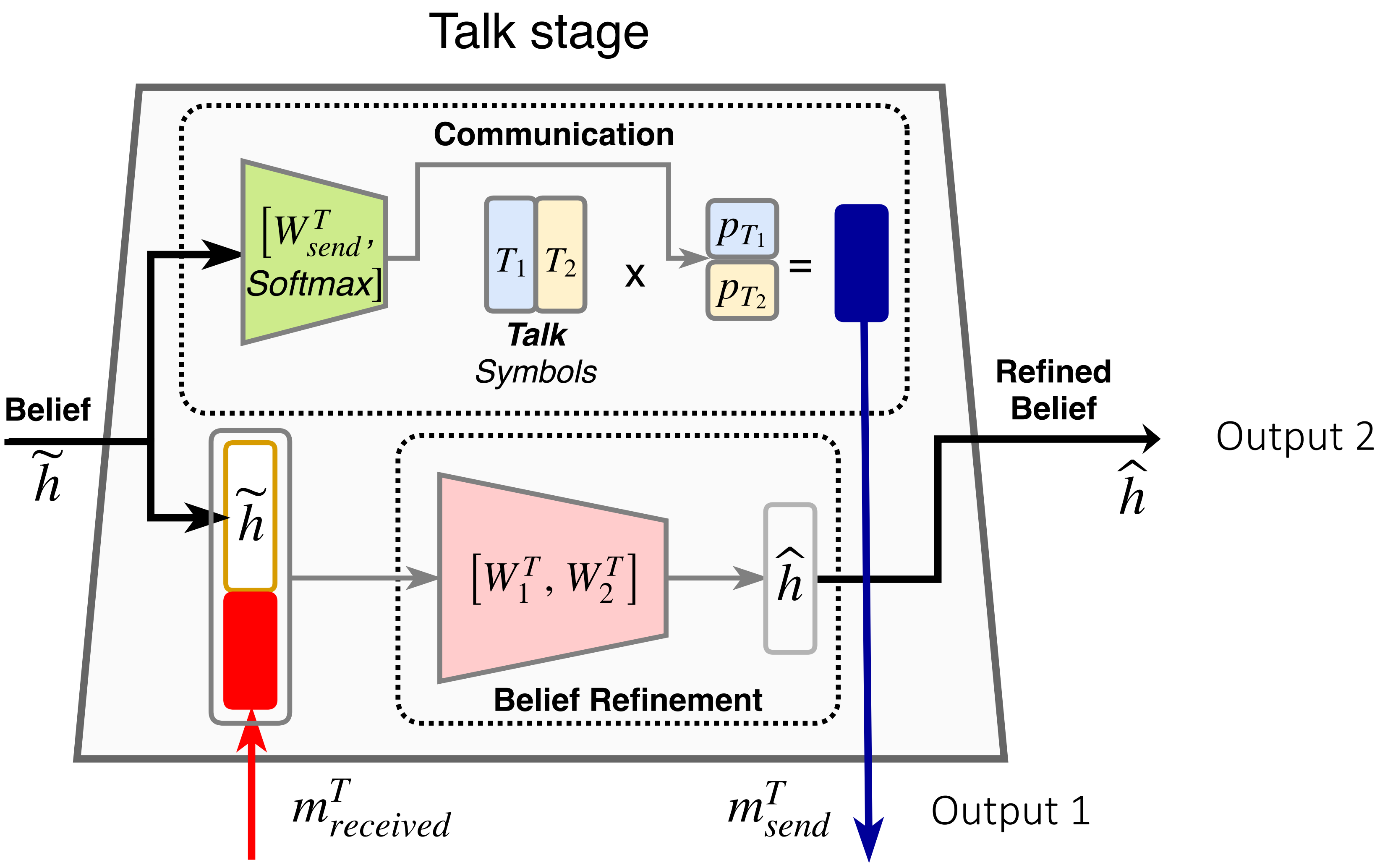
Communication and Belief Refinement



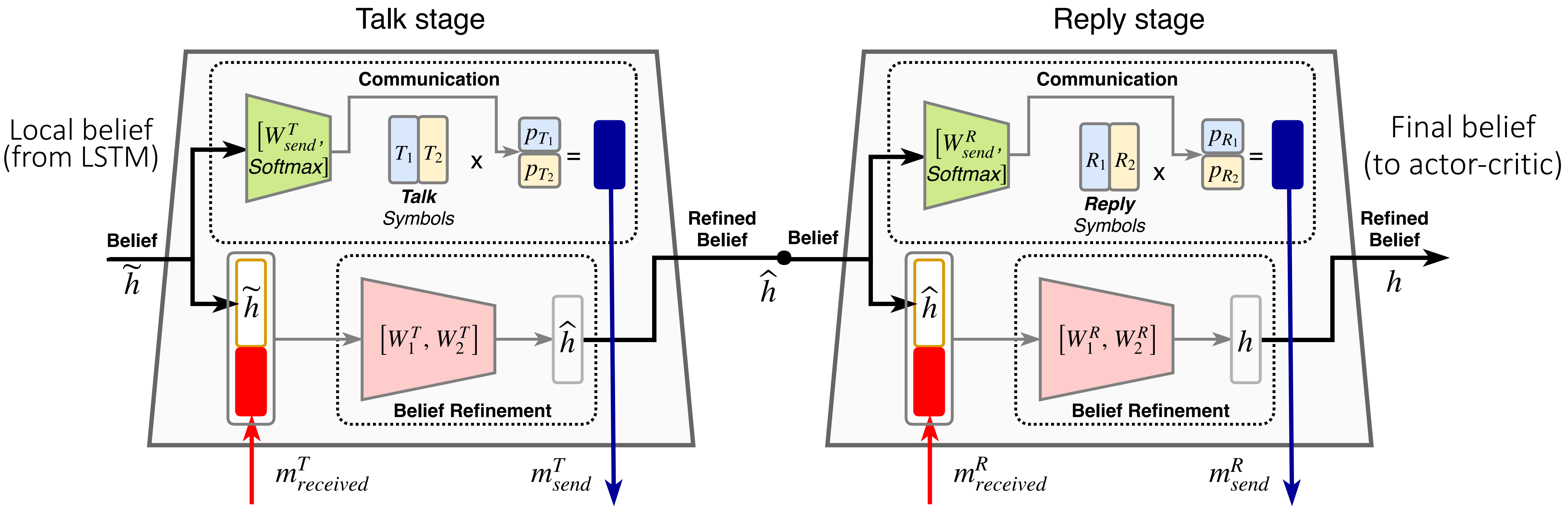
Communication and Belief Refinement



Communication and Belief Refinement

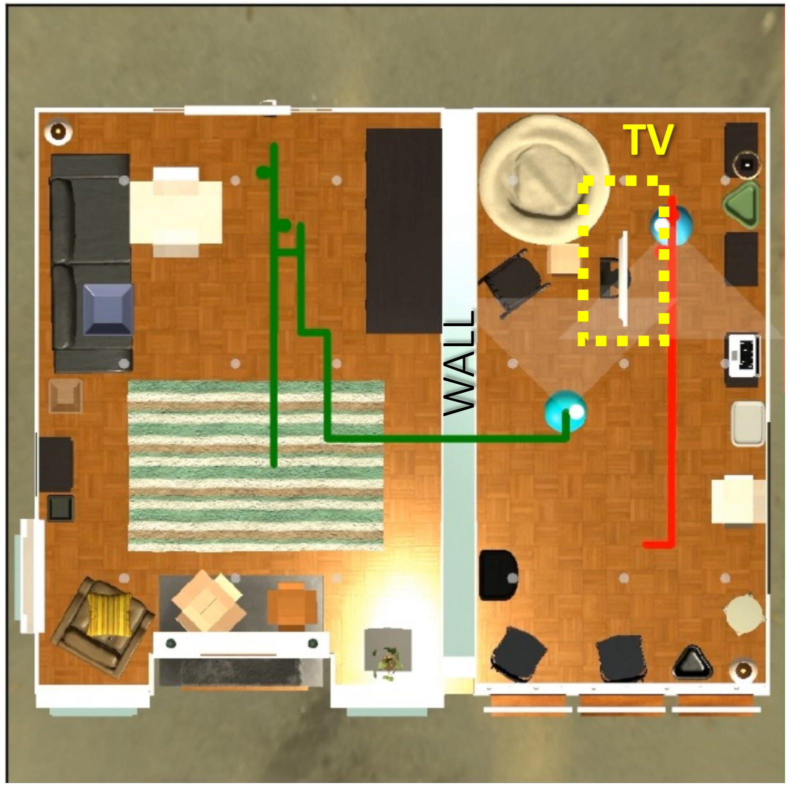


Talk and reply modules



Explicit Communication Helps

❌ *Without* explicit communication

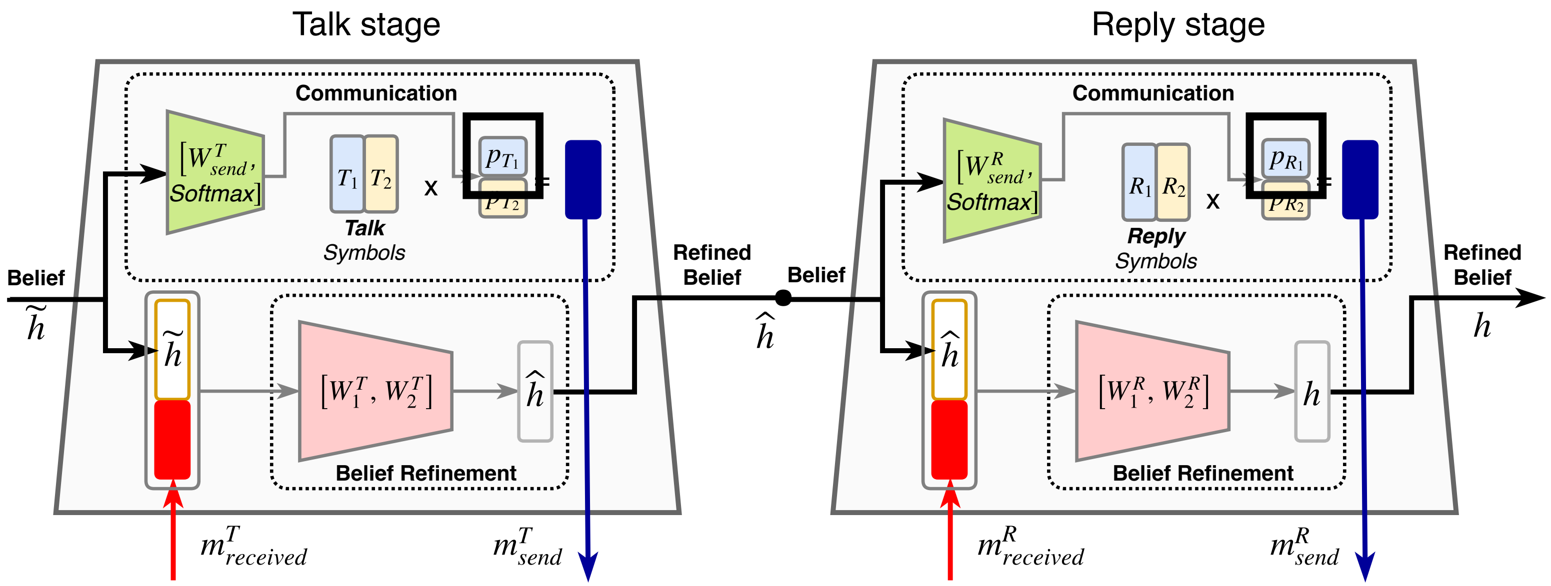


Total steps: 165
Unsuccessful pickups: 6

✔ *With* explicit communication

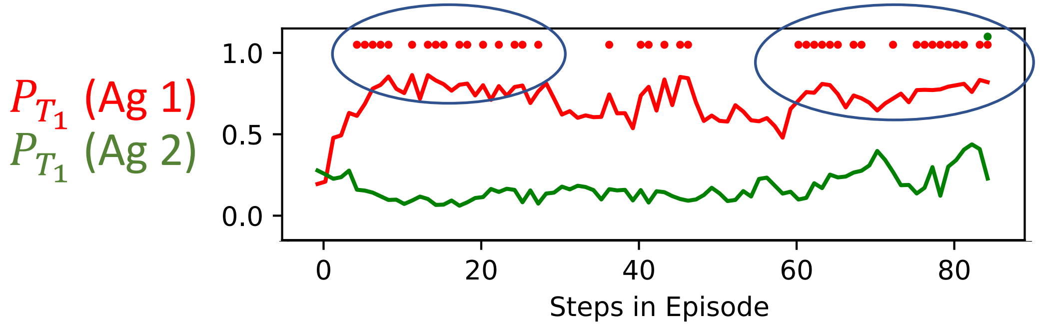
Total steps: 86
Unsuccessful pickups: 0

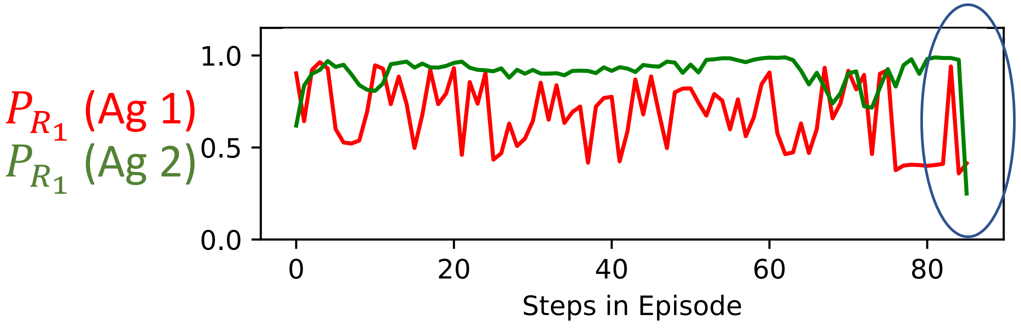
Interpretation of messages



"I am near TV!"

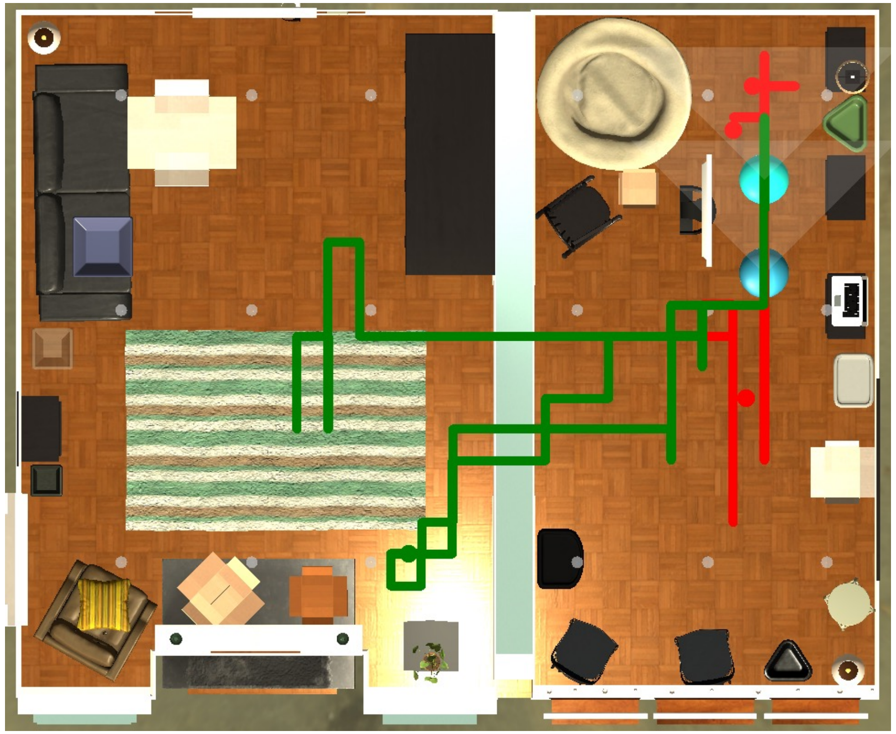
"Let us Pickup!"





Implicit Communication Helps

Visibility of other agent communicates information

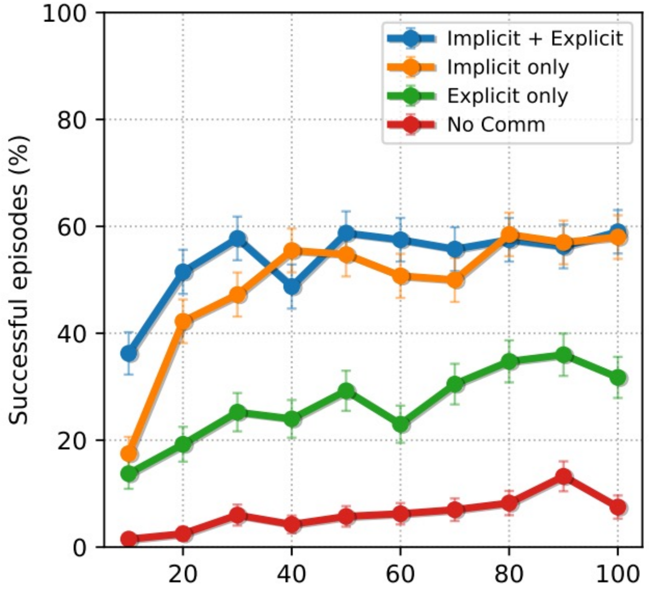


Without any communication:

Episode Unsuccessful

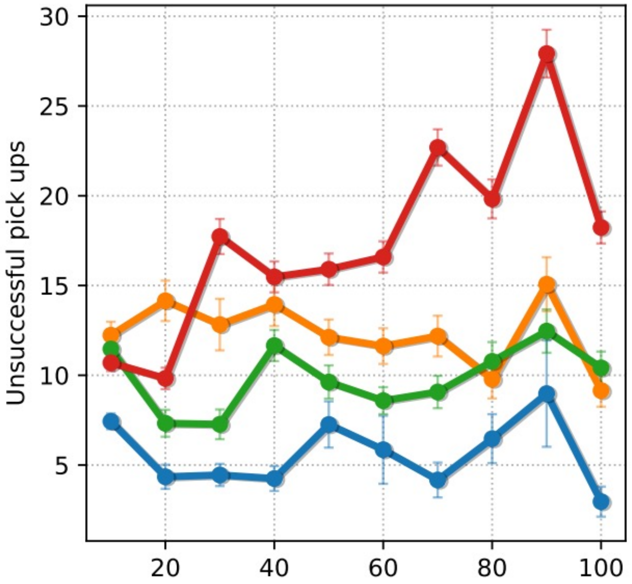
Effect of communication

Successful episodes ↑



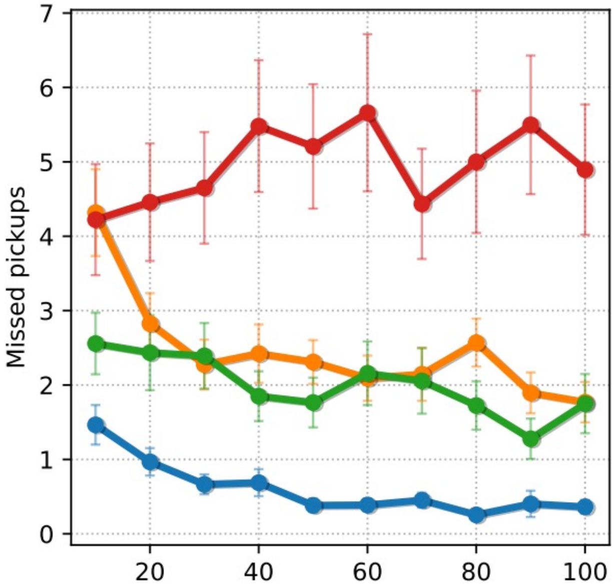
Training episodes
(x 1000)

Unsuccessful pickups ↓



Training episodes
(x 1000)

Missed pickups ↓



Training episodes
(x 1000)

Implicit + Explicit
Implicit only
Explicit only
No Comm

Collaborative Embodied Agents

Two Body Problem
CVPR 2019

SYNC Policies
ECCV 2020

GRIDTOPIX
arXiv

Takeaways

- Study collaborative behavior in visual environments
- Explicit and implicit communication are helpful
- Emergence of interpretable communication pattern

Collaborative Embodied Agents

Two Body Problem
CVPR 2019

1. First collaborative embodied task – FurnLift
2. Interpretation of emergent communication
3. Effect of communication

SYNC Policies
ECCV 2020

4. Intricately coordinated embodied task – FurnMove
5. Richer representation of multi-agent policy

GRIDTOPIX
arXiv

6. Learning policies from minimal supervision
7. Leveraging perfect-perception gridworlds for training

Intricately coordinated embodied task

Designing a harder test for collaborative agents:

- Furniture Lifting requires only one step of action coordination.
- Get agents to coordinate at **every** step.

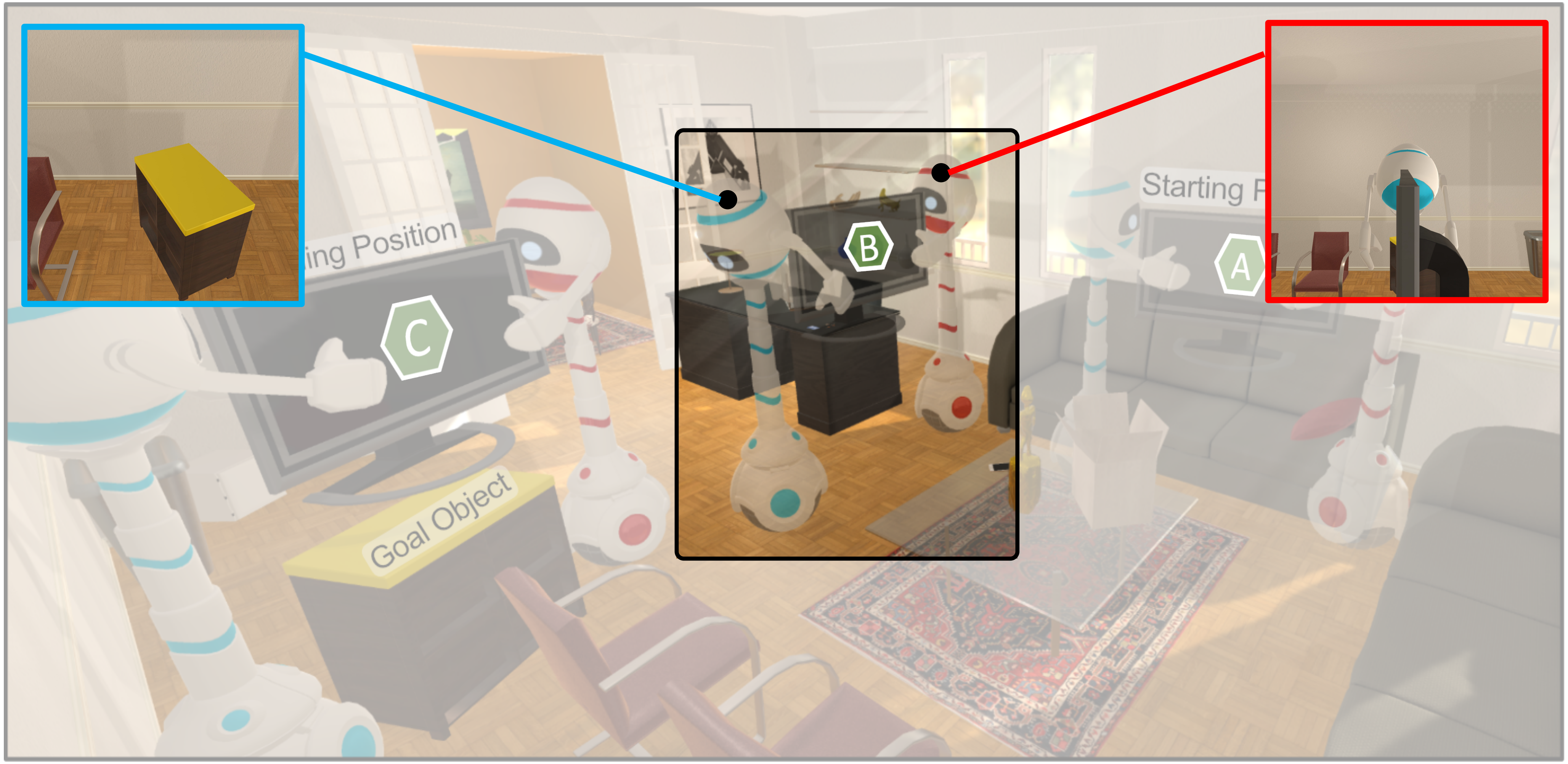
FurnMove task

FurnMove task

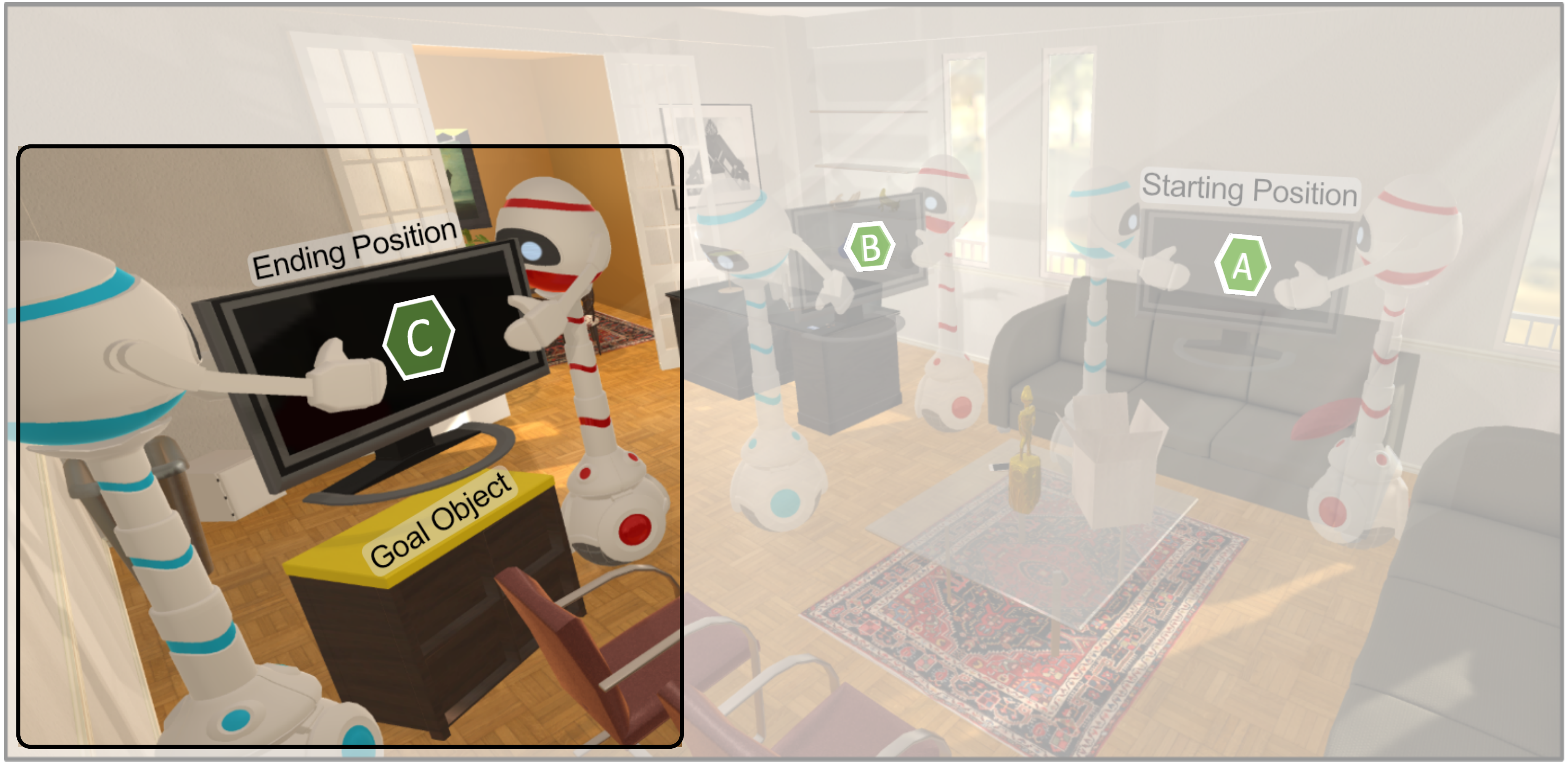
Agents must

- Remain near the TV
- Move the TV together

FurnMove task

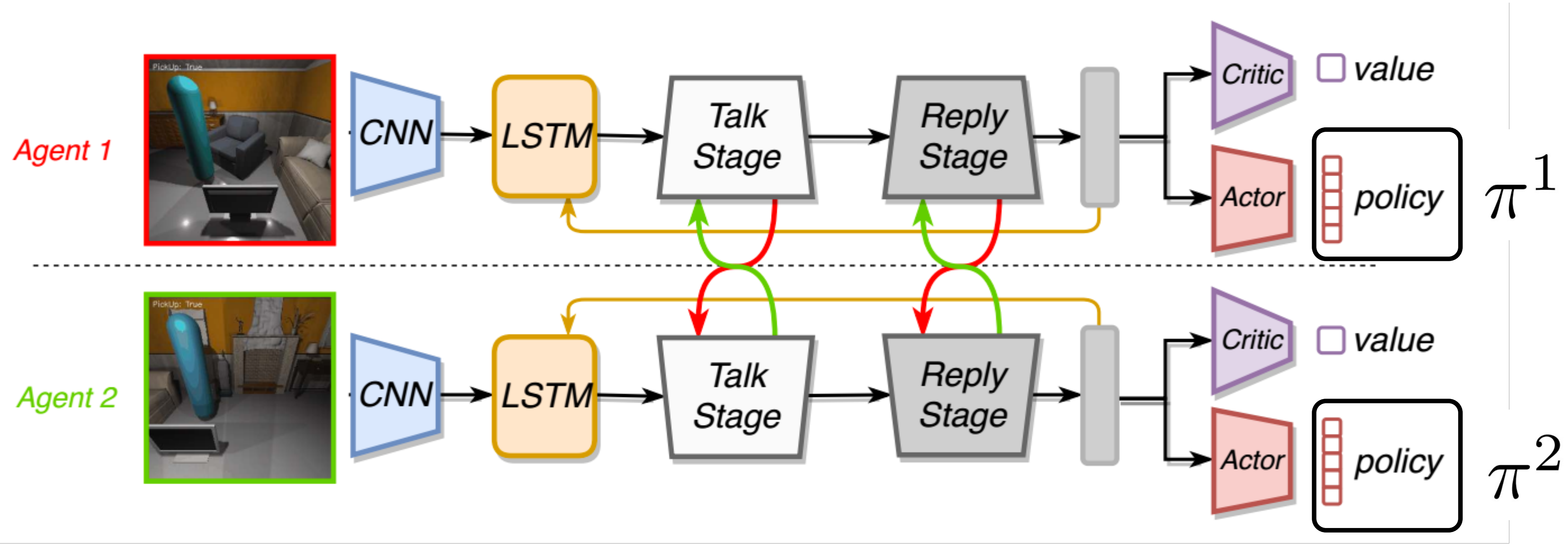


FurnMove task



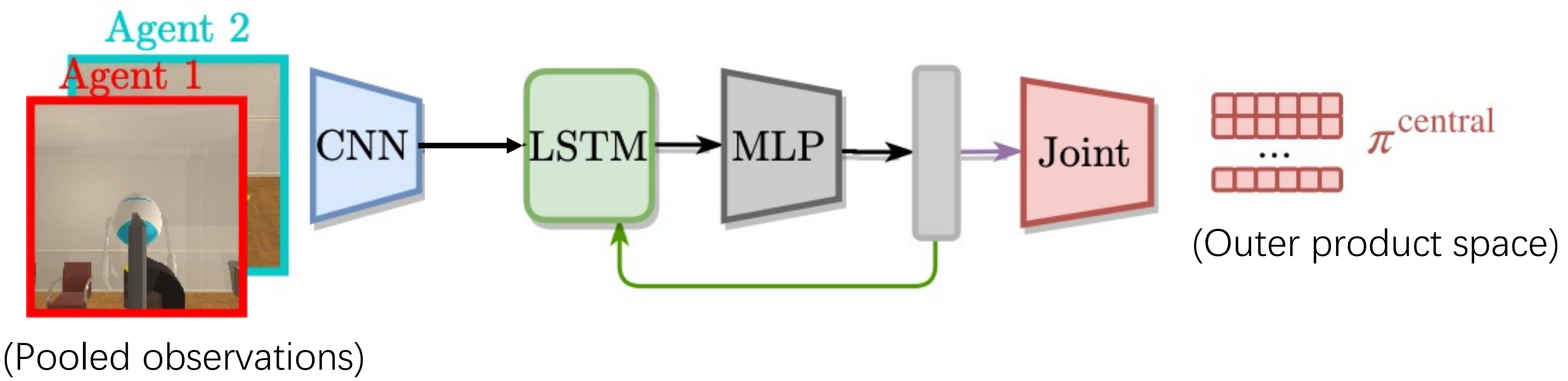
Solving FurnMove is Challenging

Single (marginal) policy per agent – 'Marginal' Agents



$$\Pi = \pi^1 \otimes \pi^2 \qquad \Pi = \pi^1 \otimes \cdots \otimes \pi^N$$

Central Model?



Does not scale

- ❌ Model complexity
- ❌ Policy parameters

Communication issues

- ❌ High bandwidth
- ❌ Lost packets

Example

Optimal Joint

$\Pi^* =$

0.2	0	0	0
0	0	0.1	0
0	0.6	0	0
0	0	0	0.1

Marginal Agents

Agent 1 Policy (π^1)

0.2	0.1	0.6	0.1

Agent 2 Policy (π^2)

0.2	0.6	0.1	0.1

Effective Joint Policy

$\pi^1 \otimes \pi^2 =$

0.04	0.12	0.02	0.02
0.02	0.06	0.01	0.01
0.12	0.36	0.06	0.06
0.02	0.06	0.01	0.01

Rank 1

Example

Optimal Joint

$\Pi^* =$

0.05	0	0	0.05
0.05	0	0.15	0
0	0	0	0.15
0	0.4	0.05	0

Mixture of Marginals

Agent 1 Policies

π_1^1	0	0.3	0	0.7
π_2^1	0.9	0	0.1	0
π_3^1	0	0	0	1
π_4^1	0.33	0.33	0.33	0

Agent 2 Policies

π_1^2	0	0	0	1
π_2^2	0.4	0	0	0.6
π_3^2	0	0.5	0.5	0
π_4^2	0.8	0.2	0	0

α
0.1
0.6
0.2
0.1

Effective Joint Policy

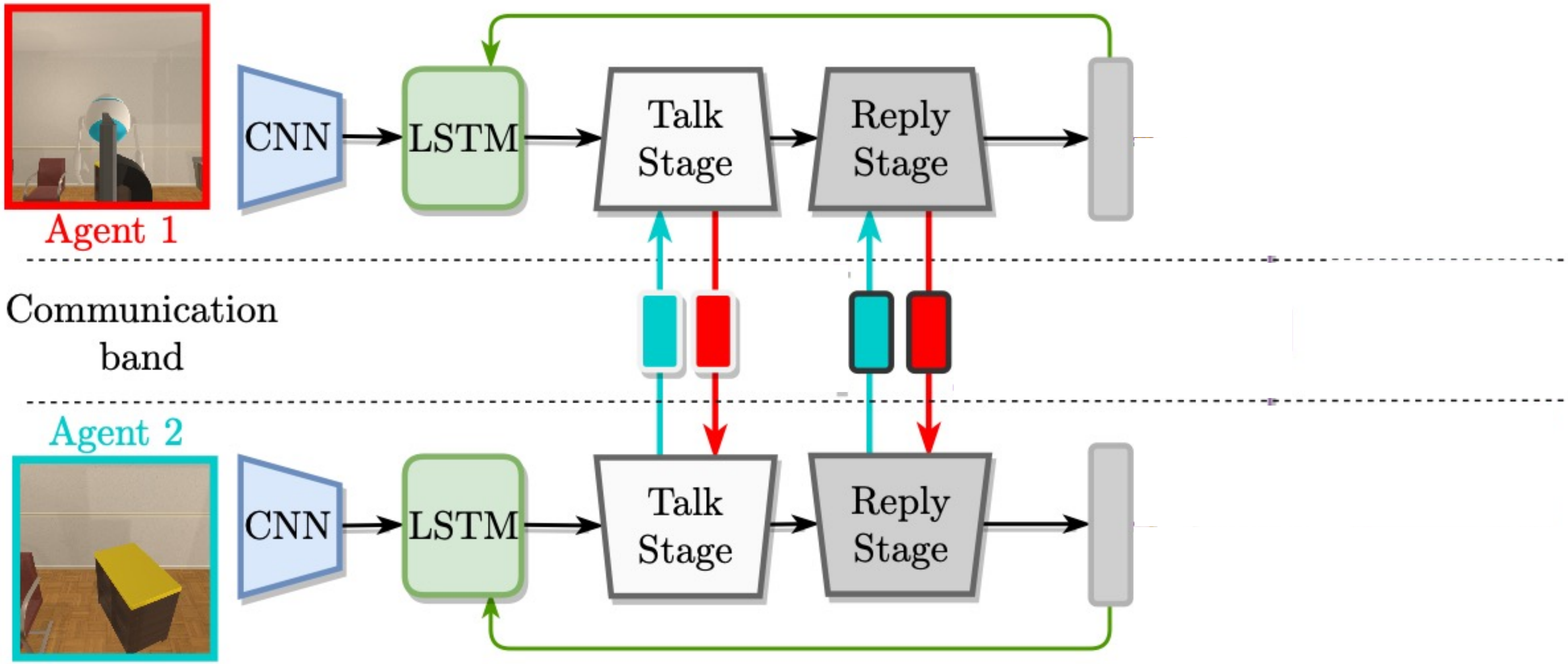
$$\sum_{i=1}^{4} \alpha_i \cdot (\pi_i^1 \otimes \pi_i^2) =$$

0.05	0	0	0.05
0.05	0	0.15	0
0	0	0	0.15
0	0.4	0.05	0

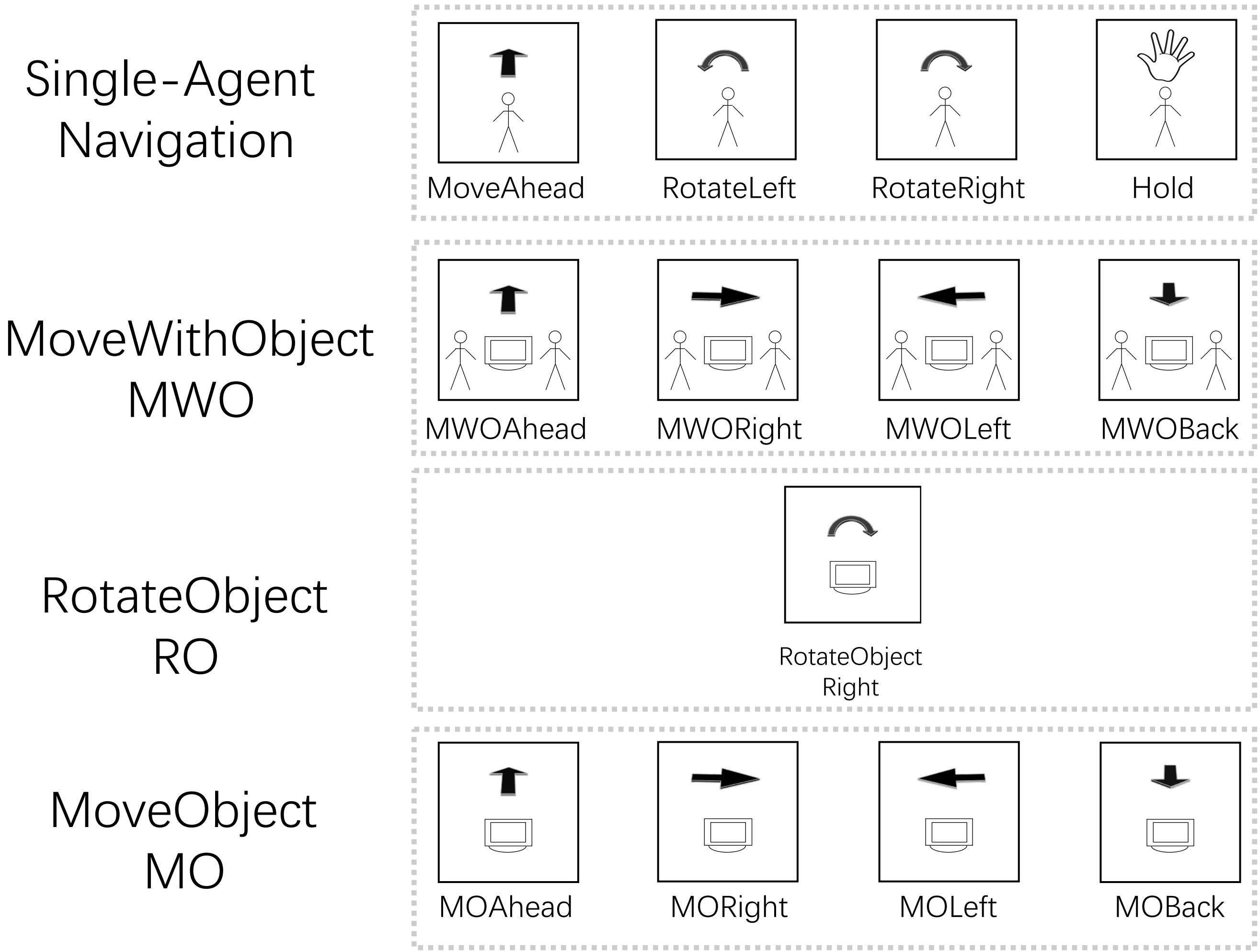
SYNC-Policies

How to sample from $\sum_{i=1}^{K} \alpha_i \cdot (\pi_i^1 \otimes \pi_i^2)$ in practice?

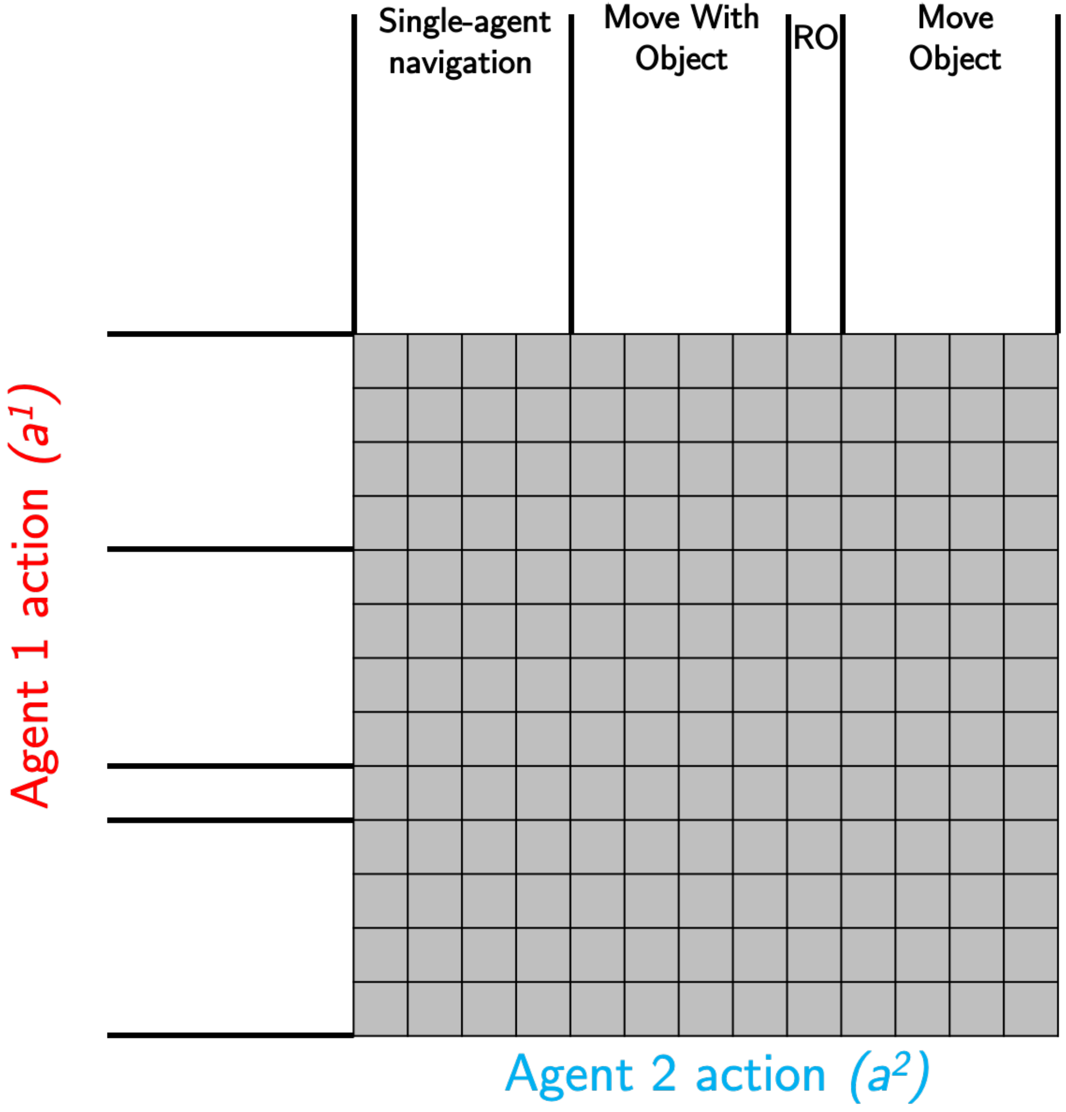
1. Compute α and K policies per agent.
2. Sample $1 \leq i \leq K$ with probability α_i. Use a shared seed so both agents sample the same i.
3. Sample actions from π_i^1 and π_i^2 independently.

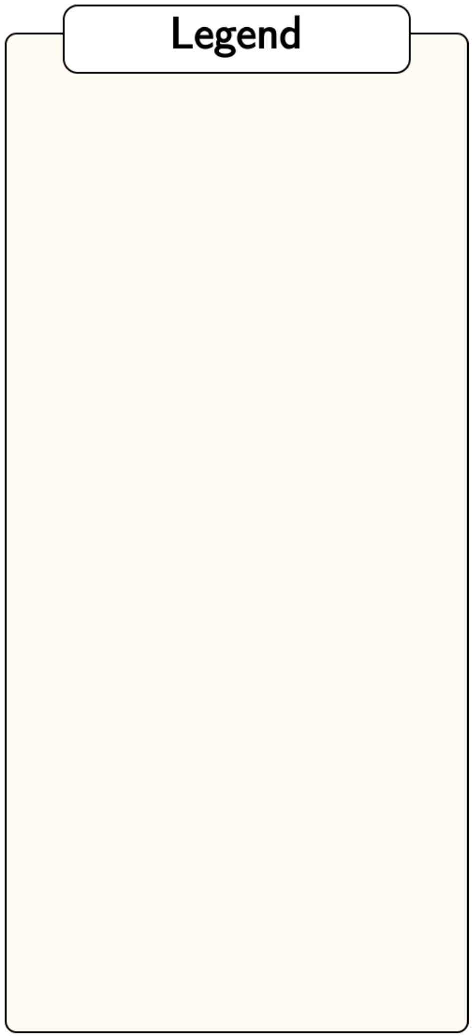


Action Space per Agent

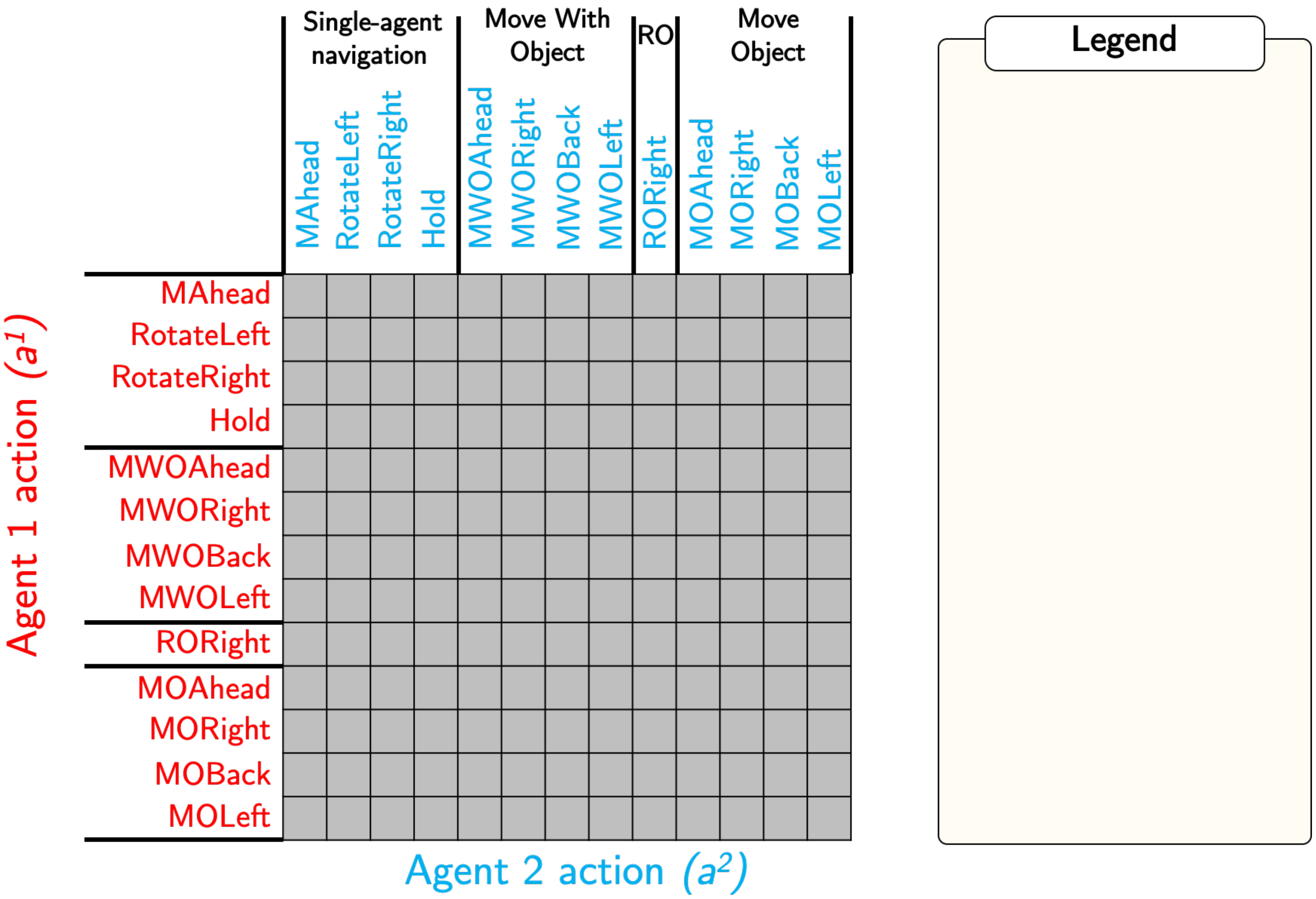


Joint action space





Joint action space



Joint action space

Agent 1 action (a^1) \ Agent 2 action (a^2)	Single-agent navigation				Move With Object				RO	Move Object			
	MAhead	RotateLeft	RotateRight	Hold	MWOAhead	MWORight	MWOBack	MWOLeft	RORight	MOAhead	MORight	MOBack	MOLeft
MAhead	X	X	X		X	X	X	X	X	X	X	X	X
RotateLeft	X	X	X		X	X	X	X	X	X	X	X	X
RotateRight	X	X	X		X	X	X	X	X	X	X	X	X
Hold					X	X	X	X	X	X	X	X	X
MWOAhead	X	X	X	X	X	X	X	X	X	X	X	X	X
MWORight	X	X	X	X	X	X	X	X	X	X	X	X	X
MWOBack	X	X	X	X	X	X	X	X	X	X	X	X	X
MWOLeft	X	X	X	X	X	X	X	X	X	X	X	X	X
RORight	X	X	X	X	X	X	X	X	X	X	X	X	X
MOAhead	X	X	X	X	X	X	X	X	X	X	X	X	X
MORight	X	X	X	X	X	X	X	X	X	X	X	X	X
MOBack	X	X	X	X	X	X	X	X	X	X	X	X	X
MOLeft	X	X	X	X	X	X	X	X	X	X	X	X	X

Agent 1 action (a^1)

Agent 2 action (a^2)

Legend

- X - Always Invalid
- Always Valid

Joint action space

Agent 1 action (a^1) \ Agent 2 action (a^2)	Single-agent navigation				Move With Object				RO	Move Object			
	MAhead	RotateLeft	RotateRight	Hold	MWOAhead	MWORight	MWOBack	MWOLeft	RORight	MOAhead	MORight	MOBack	MOLeft
MAhead	X	X	X		X	X	X	X	X	X	X	X	X
RotateLeft	X	X	X		X	X	X	X	X	X	X	X	X
RotateRight	X	X	X		X	X	X	X	X	X	X	X	X
Hold					X	X	X	X	X	X	X	X	X
MWOAhead	X	X	X	X	X	X	X	X	X	X	X	X	X
MWORight	X	X	X	X	X	X	X	X	X	X	X	X	X
MWOBack	X	X	X	X	X	X	X	X	X	X	X	X	X
MWOLeft	X	X	X	X	X	X	X	X	X	X	X	X	X
RORight	X	X	X	X	X	X	X	X		X	X	X	X
MOAhead	X	X	X	X	X	X	X	X	X	X	X	X	X
MORight	X	X	X	X	X	X	X	X	X	X	X	X	X
MOBack	X	X	X	X	X	X	X	X	X	X	X	X	X
MOLeft	X	X	X	X	X	X	X	X	X	X	X	X	X

Agent 1 action (a^1)

Agent 2 action (a^2)

Legend

X - Always Invalid

Always Valid

Joint action space

Agent 1 action (a^1) \ Agent 2 action (a^2)	Single-agent navigation				Move With Object				RO	Move Object			
	MAhead	RotateLeft	RotateRight	Hold	MWOAhead	MWORight	MWOBack	MWOLeft	RORight	MOAhead	MORight	MOBack	MOLeft
MAhead	X	X	X		X	X	X	X	X	X	X	X	X
RotateLeft	X	X	X		X	X	X	X	X	X	X	X	X
RotateRight	X	X	X		X	X	X	X	X	X	X	X	X
Hold					X	X	X	X	X	X	X	X	X
MWOAhead	X	X	X	X	X	X	X	X	X	X	X	X	X
MWORight	X	X	X	X	X	X	X	X	X	X	X	X	X
MWOBack	X	X	X	X	X	X	X	X	X	X	X	X	X
MWOLeft	X	X	X	X	X	X	X	X	X	X	X	X	X
RORight	X	X	X	X	X	X	X	X		X	X	X	X
MOAhead	X	X	X	X	X	X	X	X	X	X	X	X	X
MORight	X	X	X	X	X	X	X	X	X	X	X	X	X
MOBack	X	X	X	X	X	X	X	X	X	X	X	X	X
MOLeft	X	X	X	X	X	X	X	X	X	X	X	X	X

Legend

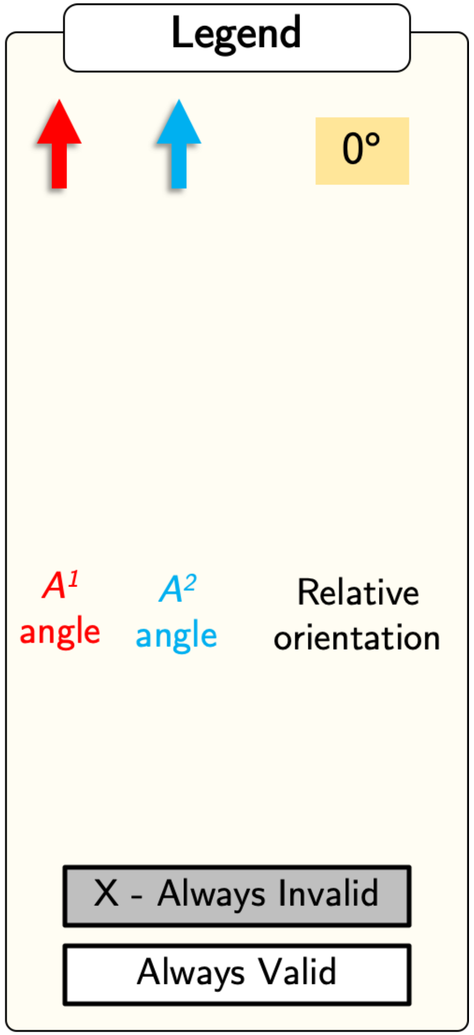
X - Always Invalid

Always Valid

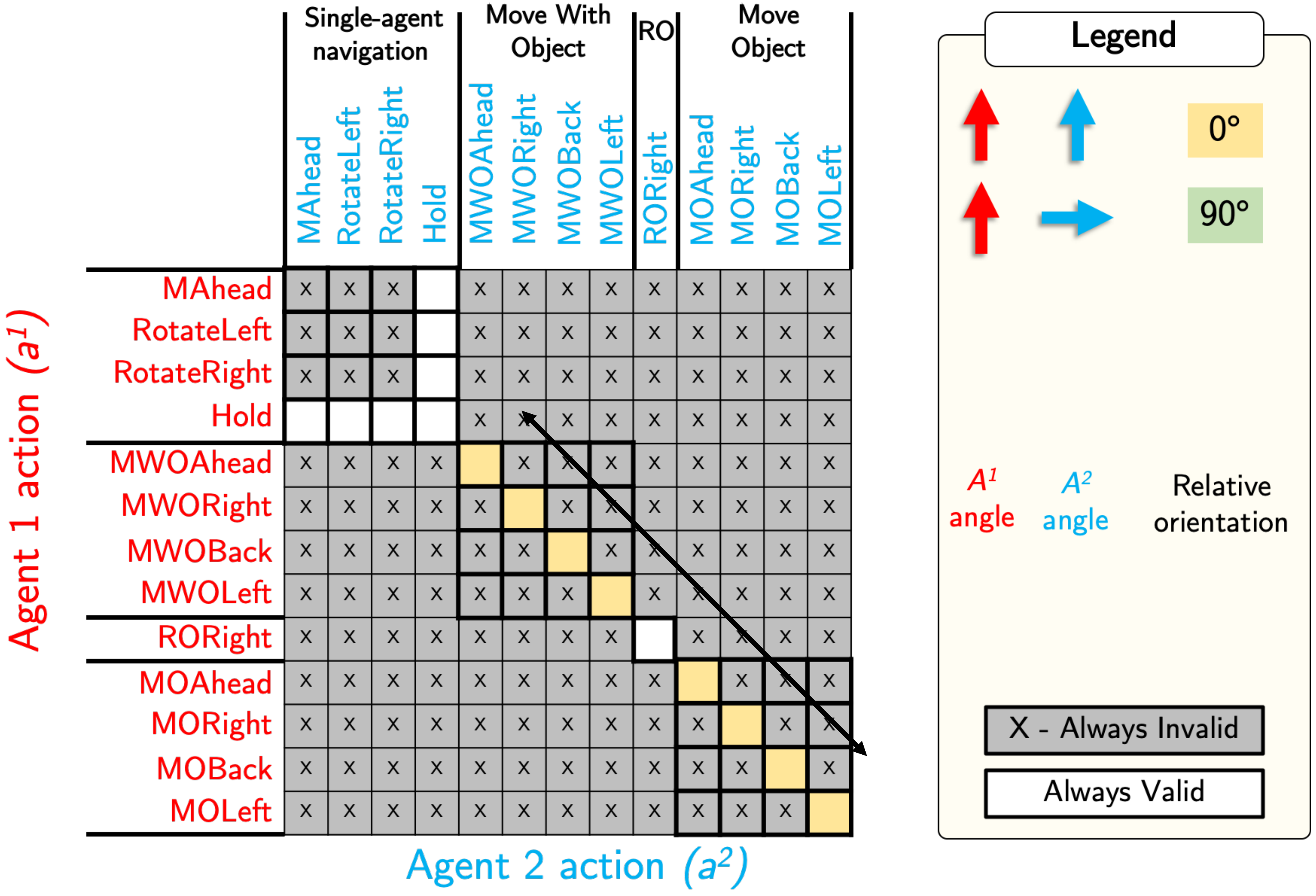
Joint action space

Agent 1 action (a^1) × Agent 2 action (a^2)

	Single-agent navigation				Move With Object				RO	Move Object			
	MAhead	RotateLeft	RotateRight	Hold	MWOAhead	MWORight	MWOBack	MWOLeft	RORight	MOAhead	MORight	MOBack	MOLeft
MAhead	X	X	X		X	X	X	X	X	X	X	X	X
RotateLeft	X	X	X		X	X	X	X	X	X	X	X	X
RotateRight	X	X	X		X	X	X	X	X	X	X	X	X
Hold					X	X	X	X	X	X	X	X	X
MWOAhead	X	X	X	X		X	X	X	X	X	X	X	X
MWORight	X	X	X	X	X		X	X	X	X	X	X	X
MWOBack	X	X	X	X	X	X		X	X	X	X	X	X
MWOLeft	X	X	X	X	X	X	X		X	X	X	X	X
RORight	X	X	X	X	X	X	X	X		X	X	X	X
MOAhead	X	X	X	X	X	X	X	X	X	X	X	X	X
MORight	X	X	X	X	X	X	X	X	X	X	X	X	X
MOBack	X	X	X	X	X	X	X	X	X	X	X	X	X
MOLeft	X	X	X	X	X	X	X	X	X	X	X	X	X



Joint action space

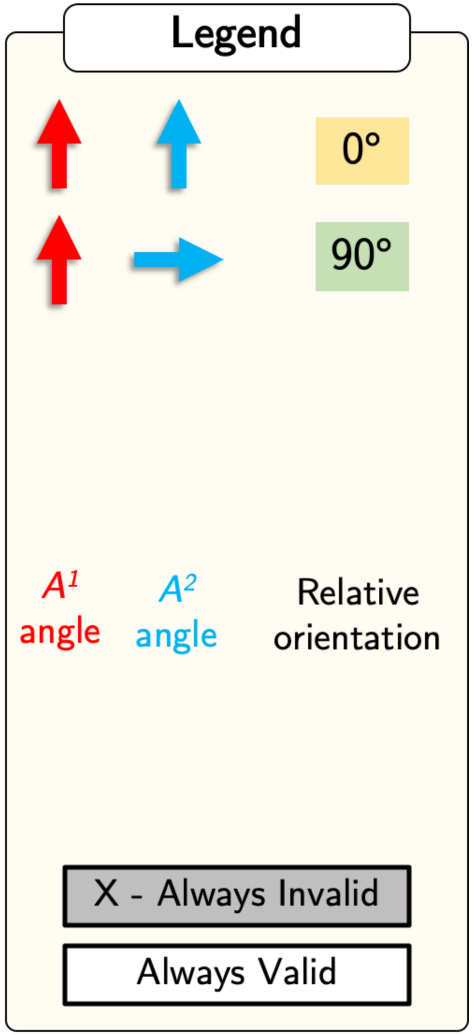


High Rank

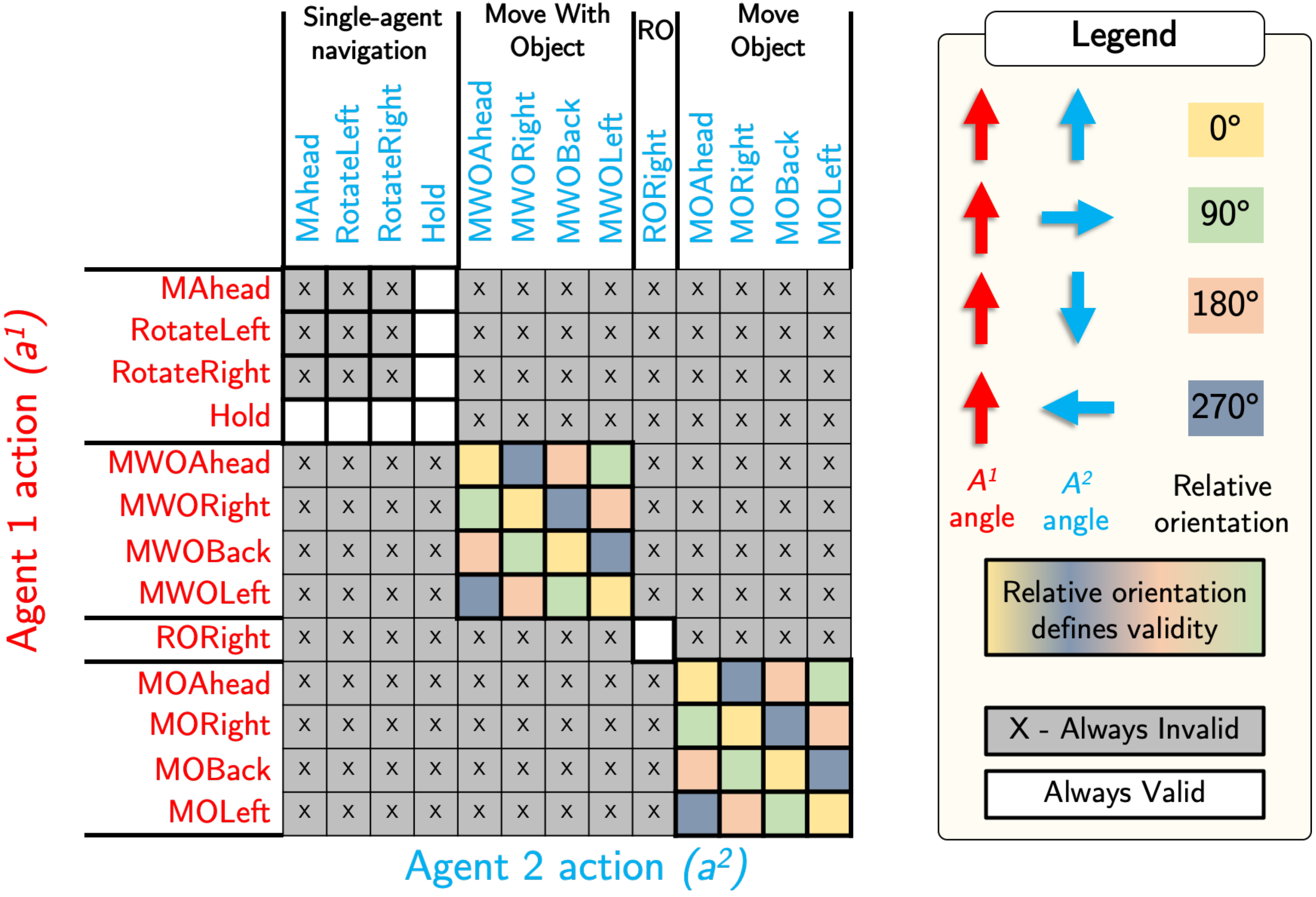
10% actions are valid

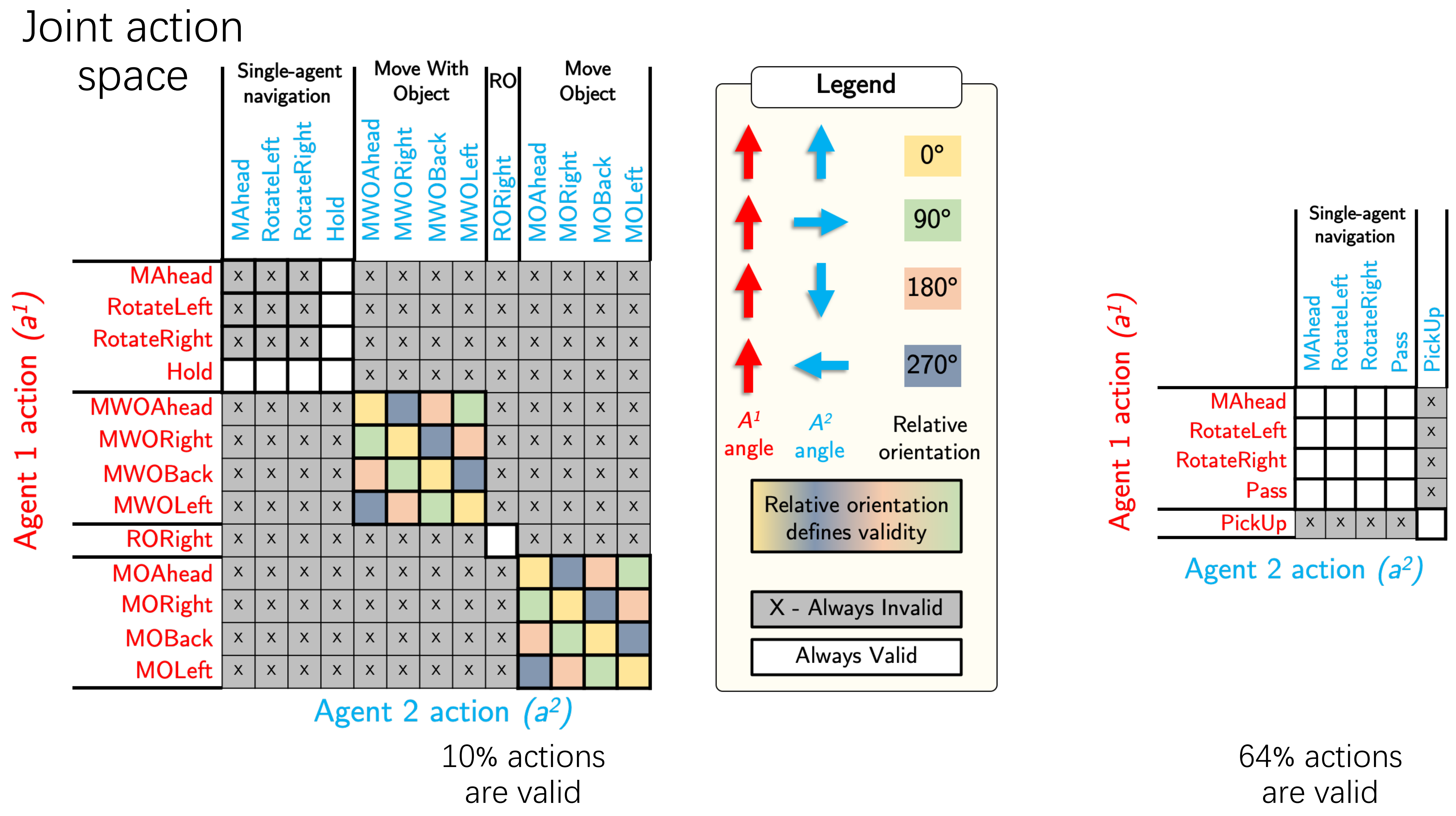
Joint action space

Agent 1 action (a^1) \ Agent 2 action (a^2)	Single-agent navigation: MAhead	RotateLeft	RotateRight	Hold	Move With Object: MWOAhead	MWORight	MWOBack	MWOLeft	RO: RORight	Move Object: MOAhead	MORight	MOBack	MOLeft
MAhead	X	X	X		X	X	X	X	X	X	X	X	X
RotateLeft	X	X	X		X	X	X	X	X	X	X	X	X
RotateRight	X	X	X		X	X	X	X	X	X	X	X	X
Hold					X	X	X	X	X	X	X	X	X
MWOAhead	X	X	X	X	X	X	X		X	X	X	X	X
MWORight	X	X	X	X		X	X	X	X	X	X	X	X
MWOBack	X	X	X	X	X		X	X	X	X	X	X	X
MWOLeft	X	X	X	X	X	X		X	X	X	X	X	X
RORight	X	X	X	X	X	X	X	X		X	X	X	X
MOAhead	X	X	X	X	X	X	X	X	X	X	X	X	
MORight	X	X	X	X	X	X	X	X	X		X	X	X
MOBack	X	X	X	X	X	X	X	X	X	X		X	X
MOLeft	X	X	X	X	X	X	X	X	X	X	X		X



Joint action space





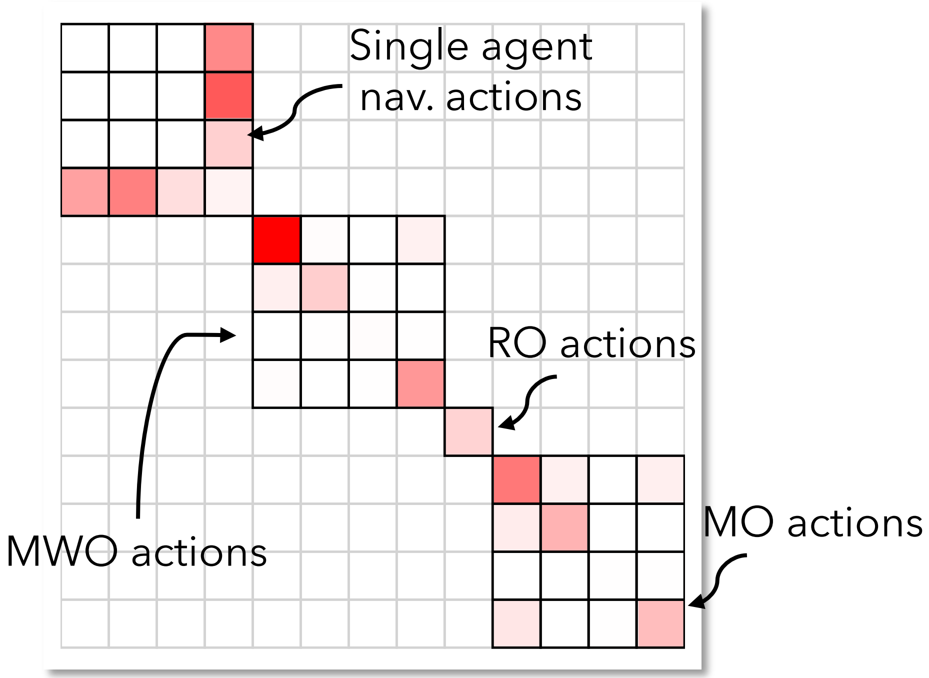
Joint action space
Single-agent navigation
Move With Object
RO
Move Object
MAhead
RotateLeft
RotateRight
Hold
MWOAhead
MWORight
MWOBack
MWOLeft
RORight
MOAhead
MORight
MOBack
MOLeft
Agent 1 action (a^1)
Agent 2 action (a^2)
10% actions are valid
Legend
0°
90°
180°
270°
A^1 angle
A^2 angle
Relative orientation
Relative orientation defines validity
X - Always Invalid
Always Valid
Single-agent navigation
MAhead
RotateLeft
RotateRight
Pass
PickUp
Agent 1 action (a^1)
Agent 2 action (a^2)
64% actions are valid

How coordinated is FurnMove?

Central Agent ⟷ Marginal Agents

	Success	Failed Pickups
FurnLift	0.6%	5.1 vs. 8.9
FurnMove	32.0%	

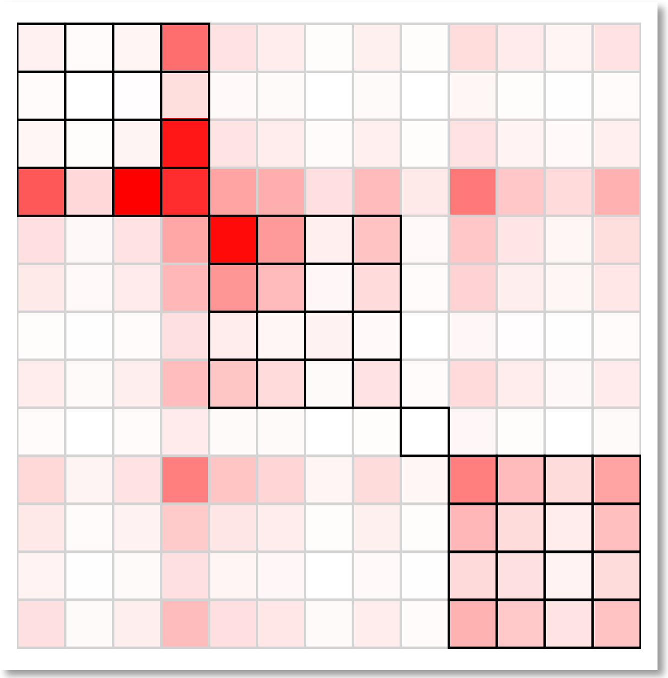
Joint Policy Summary



Central Model

65% task success
7% actions fail

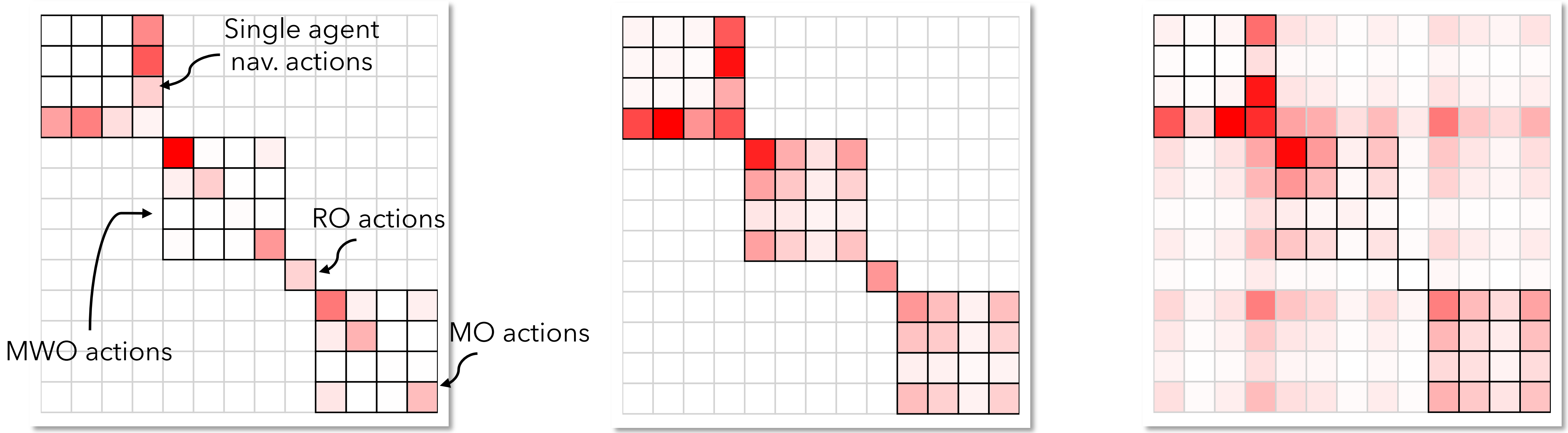
vs



Marginal Model

33% task success
65% actions fail

Joint Policy Summary



Central Model

65% task success
7% actions fail

SYNC Model

59% success rate
31% actions fail

Marginal Model

33% task success
65% actions fail

Qualitative runs

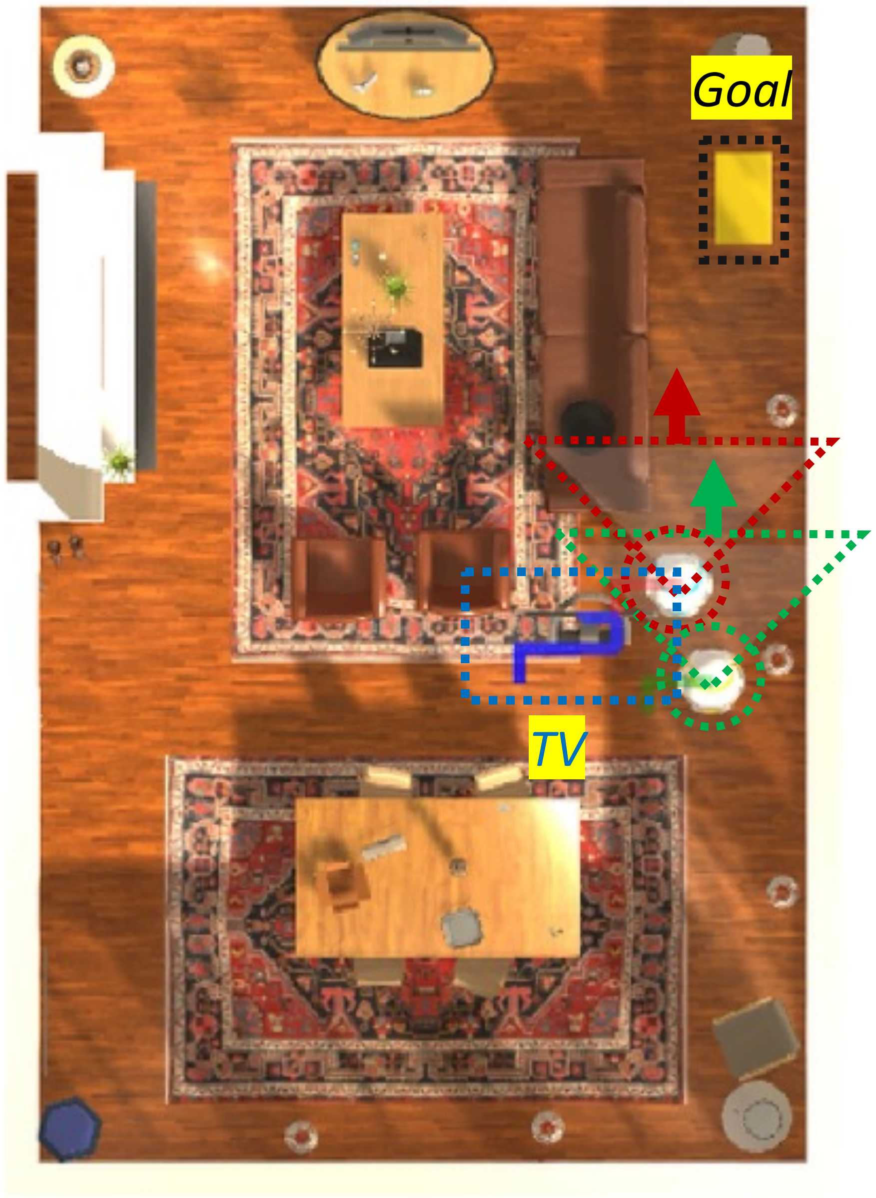
Field of view:

Triangles denote field of view & orientation of agents

Trajectories:

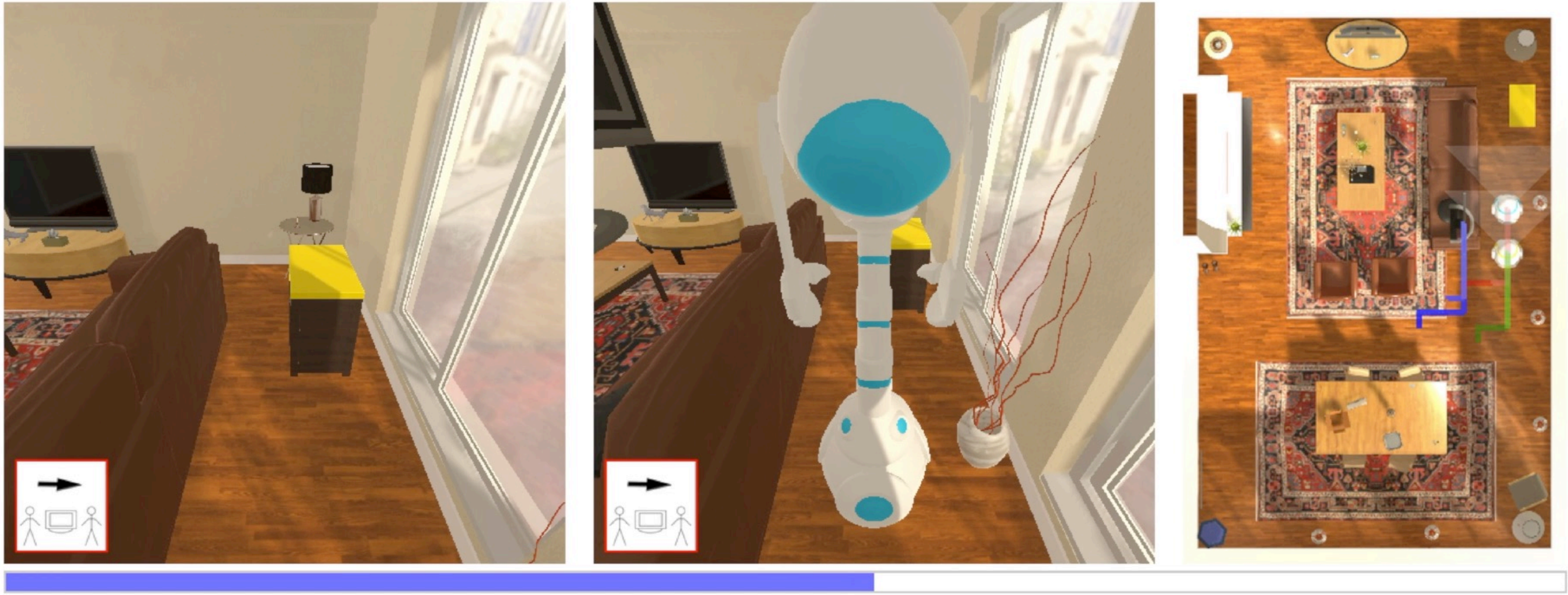
- Agent 1 trajectory in red
- Agent 2 trajectory in green
- TV trajectory in blue
- Trajectory shades become *lighter* as episode progresses

Top-down view



Marginal Agents

SYNC Agents



How many mixtures components in SYNC?

Diminishing returns from additional mixture components

# Mixture Components	Success ↑	Final Distance ↓
1 component	33	1.83
2 components	50	1.23
4 components	57	**1.08**
13 components	**59**	1.15

Collaborative Embodied Agents

Two Body Problem
CVPR 2019

SYNC Policies
ECCV 2020

GRIDTOPIX
arXiv

Takeaways

- FurnMove needs intricate coordination (high-rank joint)
- Independent and decentral execution ⇒ Rank-1
- SYNC-policies can capture a mixture-of-marginals

Collaborative Embodied Agents

Two Body Problem
CVPR 2019

1. First collaborative embodied task – FurnLift
2. Interpretation of emergent communication
3. Effect of communication

SYNC Policies
ECCV 2020

4. Intricately coordinated embodied task – FurnMove
5. Richer representation of multi-agent policy

GRIDTOPIX
arXiv

6. Learning policies from minimal supervision
7. Leveraging perfect-perception gridworlds for training

Findings about RL and Vision

(1) Visual Agents Need Shaped Rewards

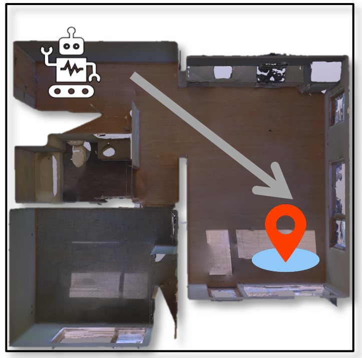
'Shaped rewards'
Dense indicators of success

- Furniture Lifting
  - Warm start with optimal actions
- Furniture Moving
  - Furn. moved closer to the goal
- PointGoal Navigation
  - Geodesic distance to goal
- Google Football
  - Checkpoint reward

Furniture Lifting
(AI2-THOR)

Furniture Moving
(AI2-THOR)

PointGoal Navigation
(Habitat+Gibson)



3 vs. 1 with Keeper
(Google Football)

Jain et al. CVPR 2019
Jain et al. ECCV 2020
Savva et al. ICCV 2019 / aihabitat.org
Kurach et al. AAAI 2020

(2) Visual Agents Fail With Terminal Rewards

'Terminal rewards'
Goal dependent or success rewards available at termination of episode

General way to supervise complex policies

Shaped rewards

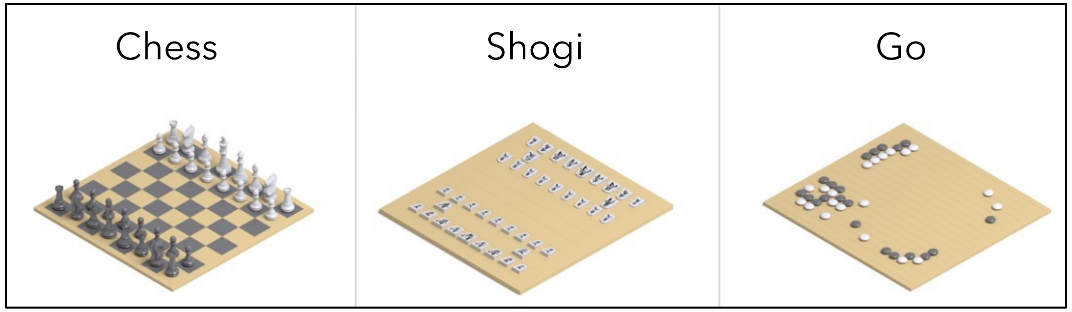
Terminal rewards

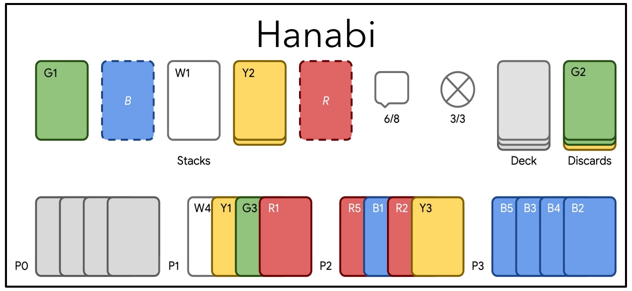
(3) RL Agents Work With Terminal Rewards

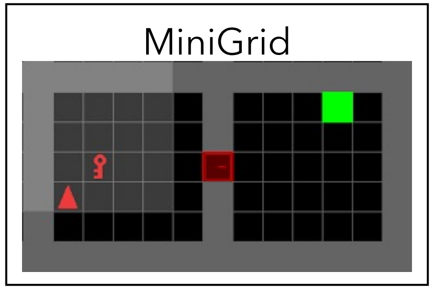
Perfect Perception!!

- Semantics are provided directly in the observations
- No perceptual reasoning is needed

- Current and previous positions of board pieces
- The cards in players' hands and previous moves
- Positions, state (open/close), and color







Silver et al. Science 2018
Bard et al. AI 2020
Lerer et al. AAAI 2020

Findings about RL and Vision

(1) Visual agents need shaped rewards

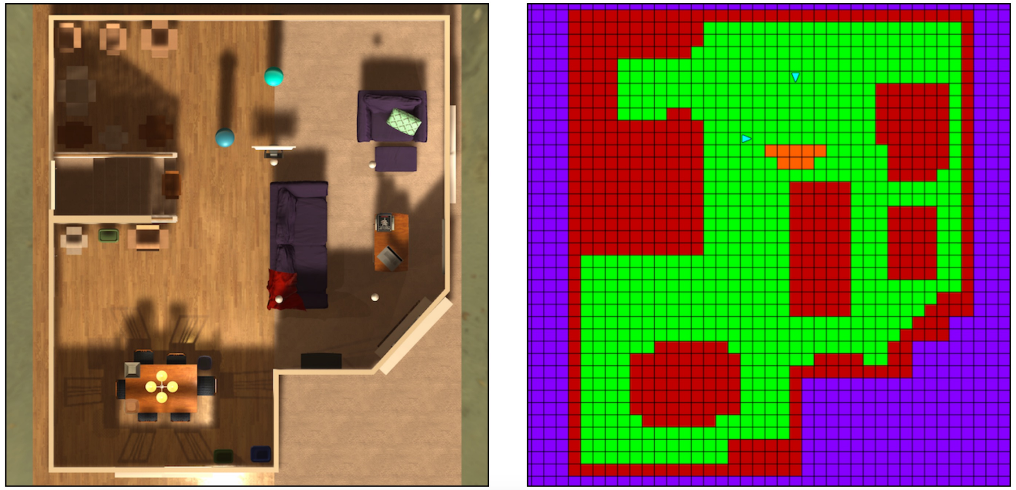
(2) Visual agents fail with terminal rewards

(3) Perfect-perception agents can learn from terminal rewards

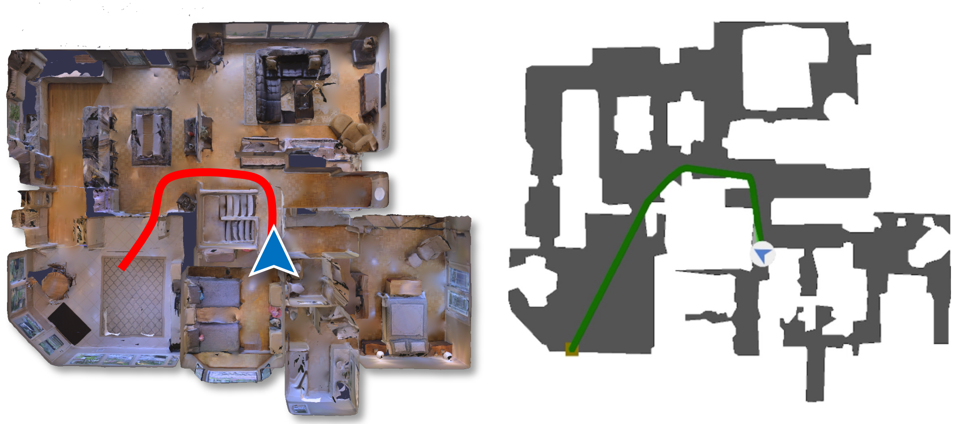
Idea: GRIDTOPIX

- Create mirroring gridworlds for embodied environments
- Decoupling of planning and perception
  - Step 1: Learn planning in gridworlds
  - Step 2: Now learn perception in visual worlds

Create 'Perfect-Perception' Gridworlds



AI2THOR-Mirroring Gridworld



AIHabitat-Mirroring Gridworld

Positions and states of objects
Can the furniture fit somewhere
Any other semantics

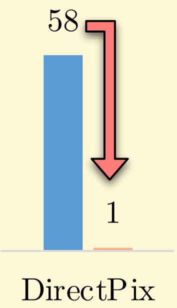
Grid Experts Can Learn from Terminal Rewards

Shaped rewards

Terminal rewards

Furniture Moving
(AI2-THOR)

Success

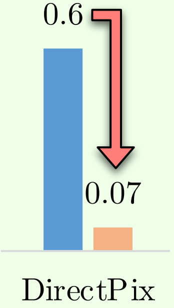


PointGoal Navigation
(Habitat+Gibson)

SPL

3 vs. 1 with Keeper
(Google Football)

Game Score



Gridworld:
1 → 56

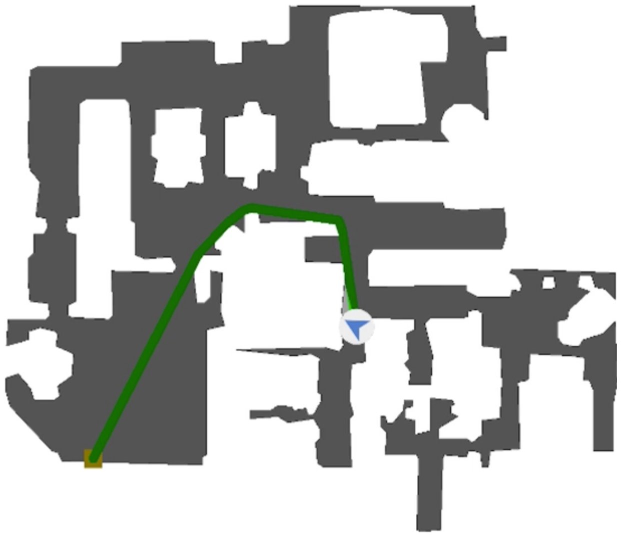
Gridworld:
0.1 → 78

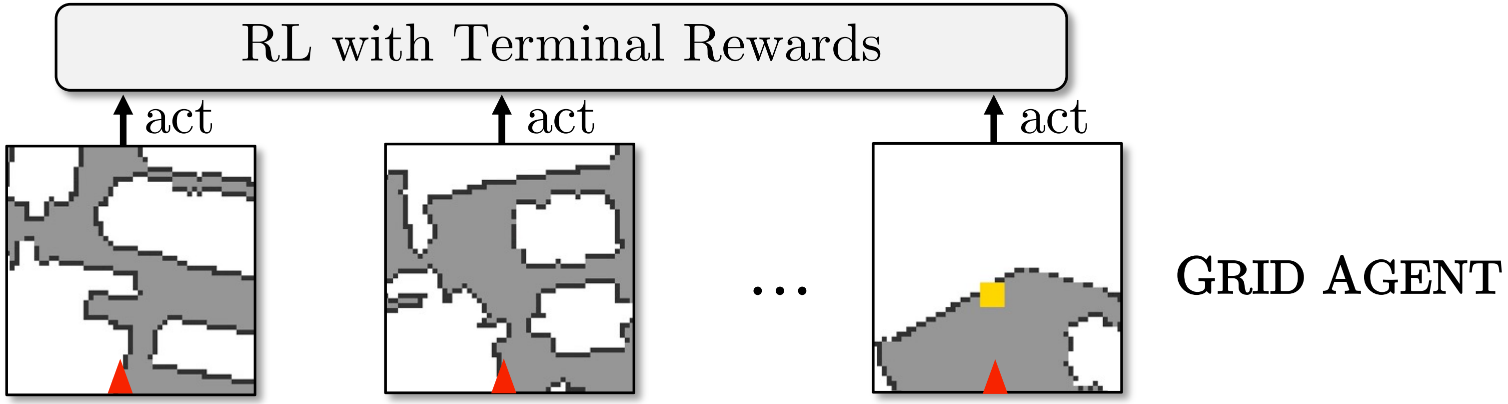
Gridworld:
0.07 → 0.9

Decoupling of Planning and Perception

Train stage 1

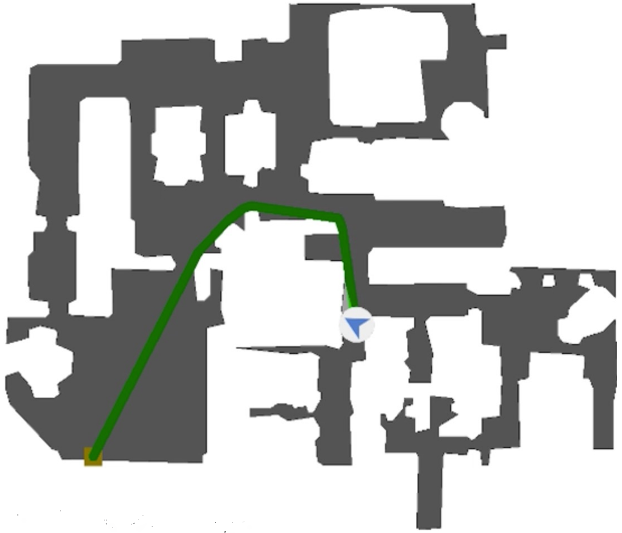
Decoupling of Planning and Perception



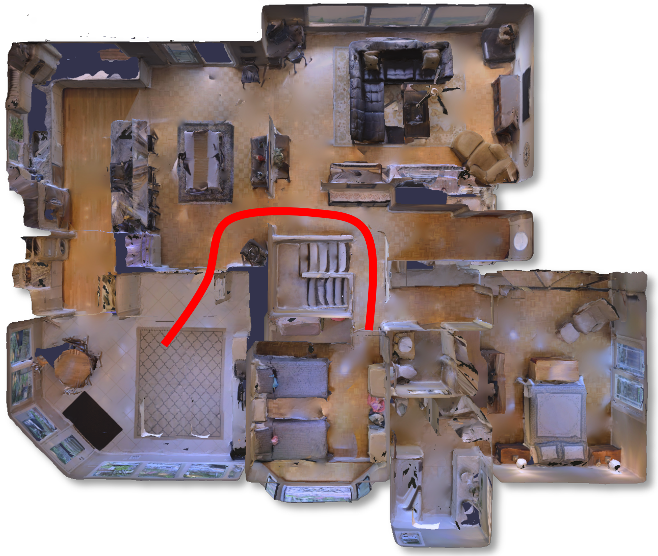


Train stage 2

Decoupling of Planning and Perception



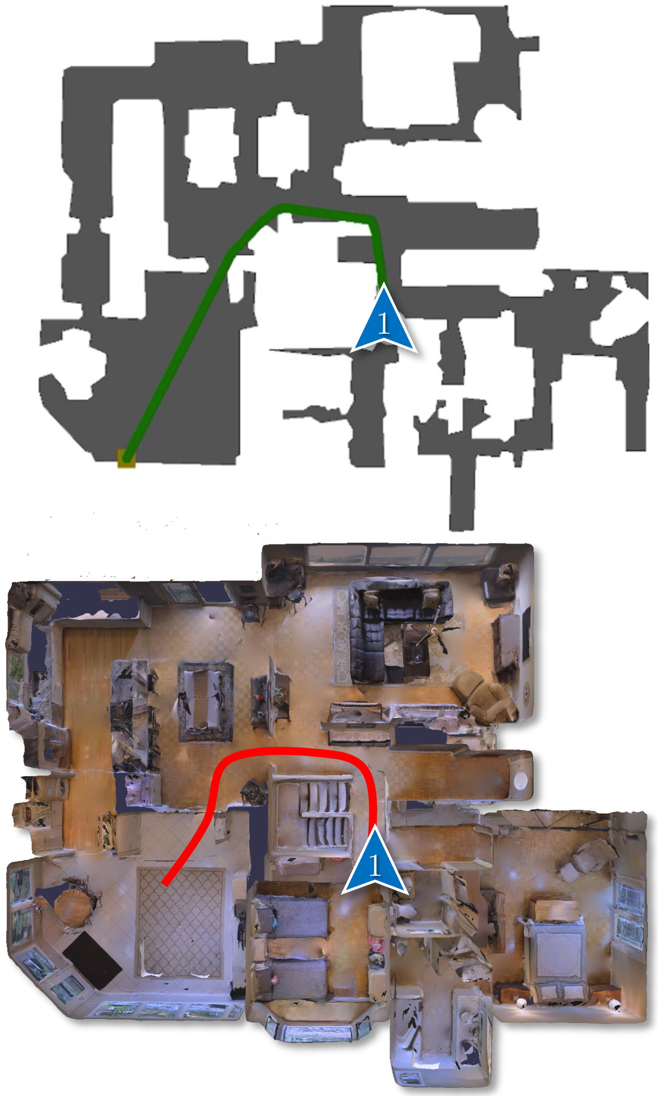
GRID AGENT
(teacher)



PIX AGENT
(student)

Train stage 2

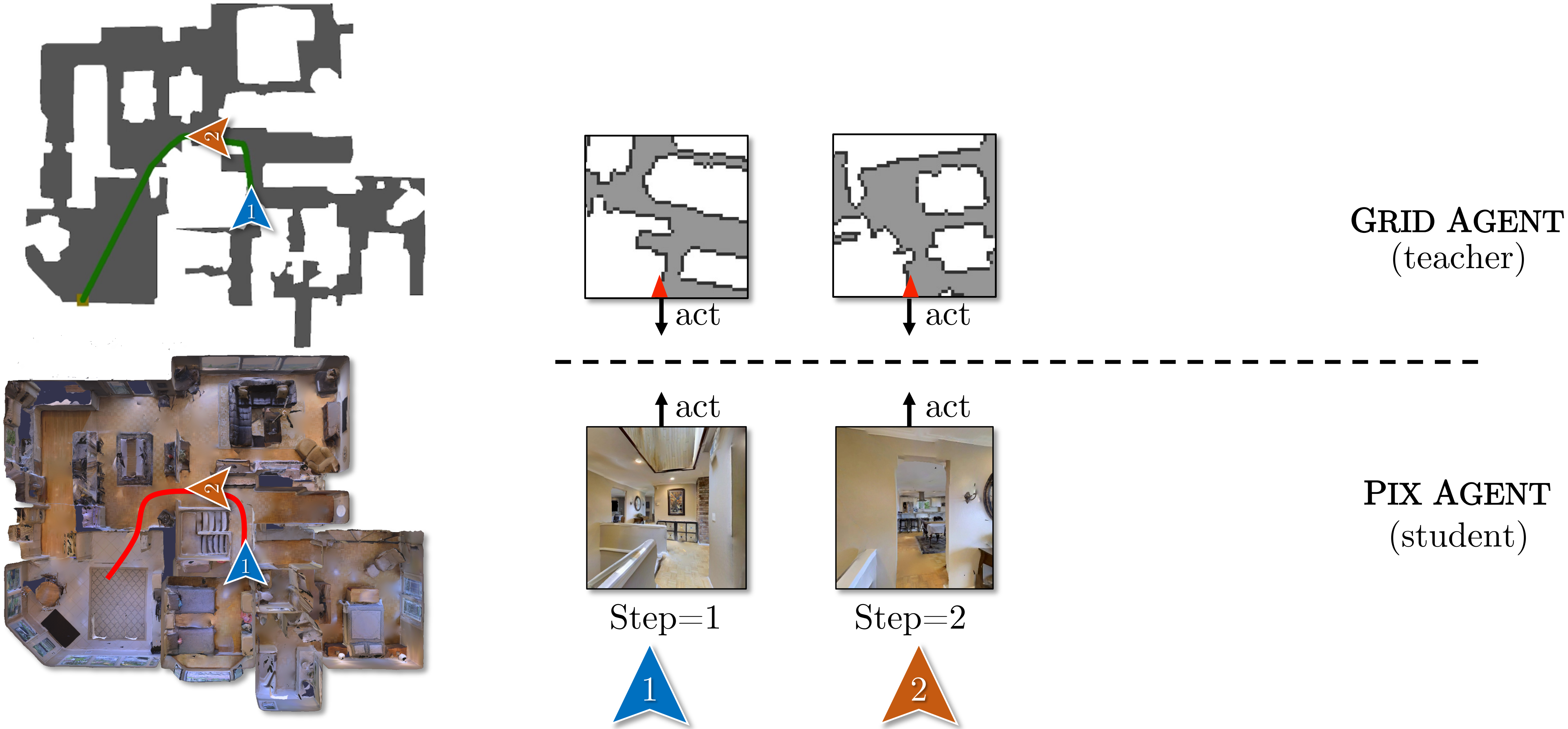
Decoupling of Planning and Perception



GRID AGENT
(teacher)

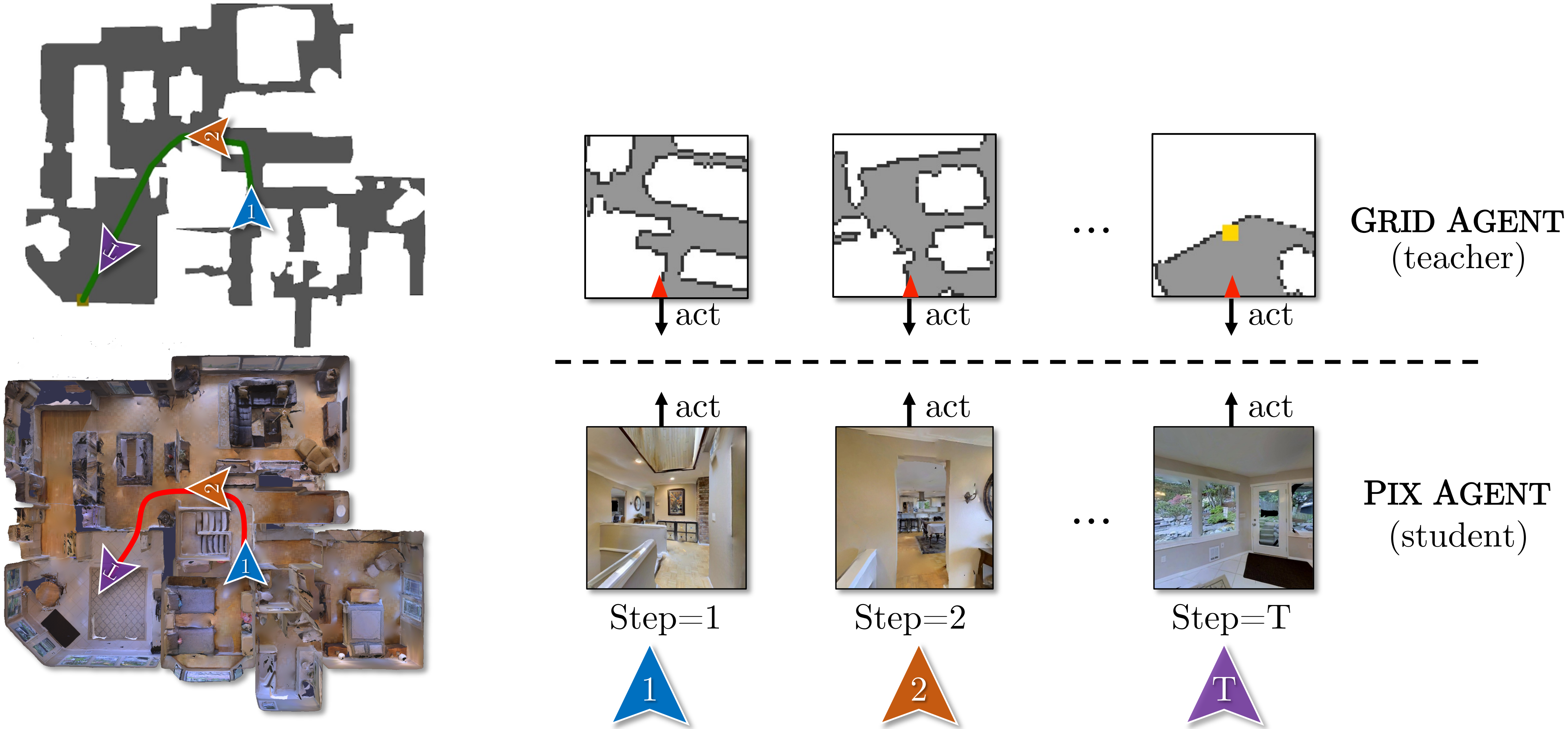
PIX AGENT
(student)

Decoupling of Planning and Perception



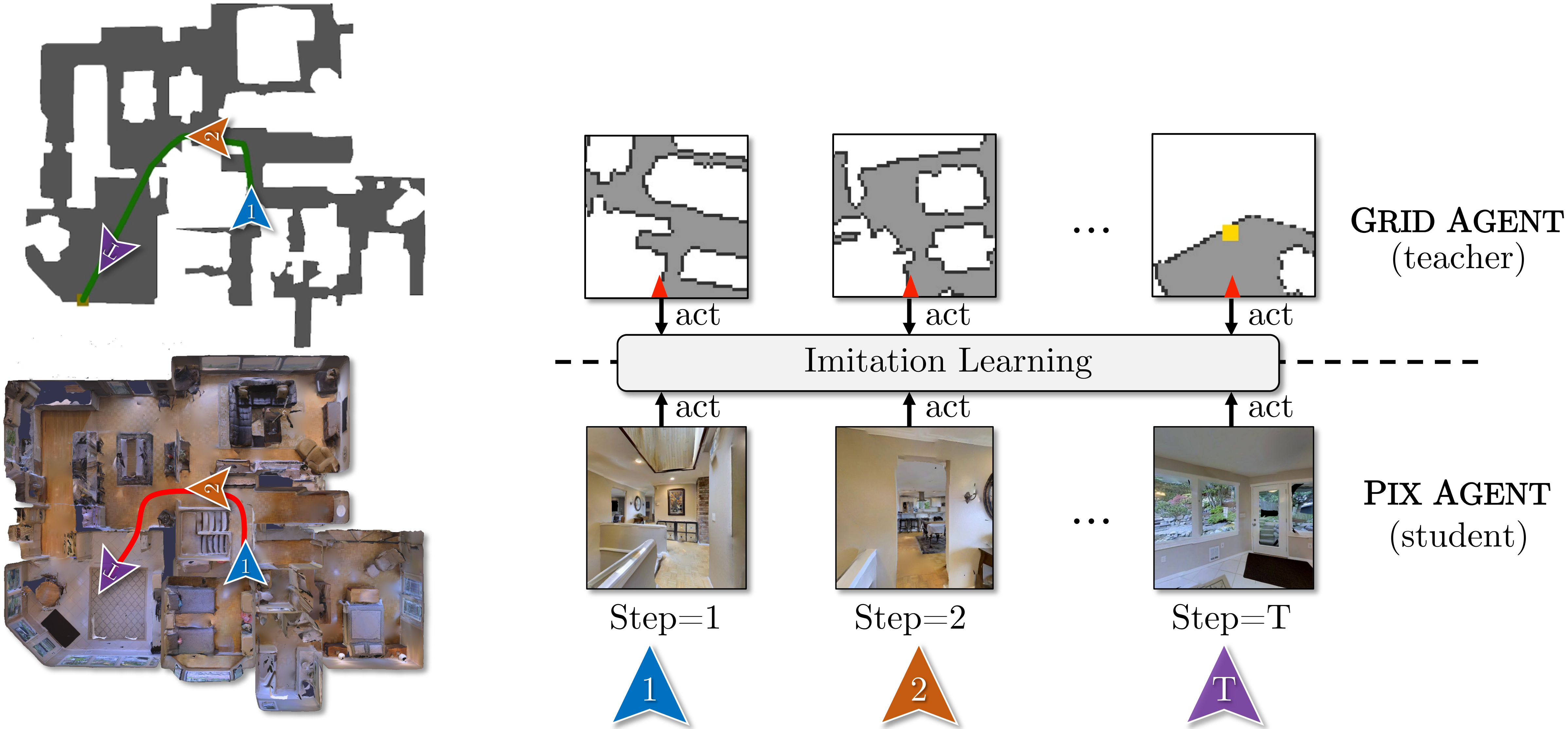
Train stage 2

Decoupling of Planning and Perception



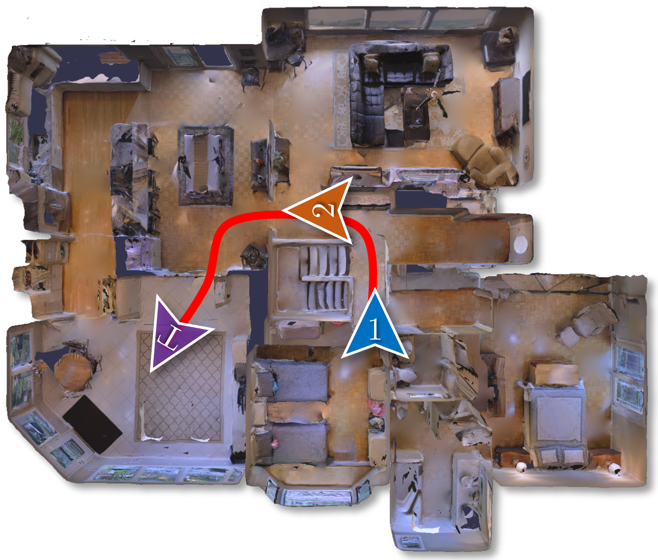
Train stage 2

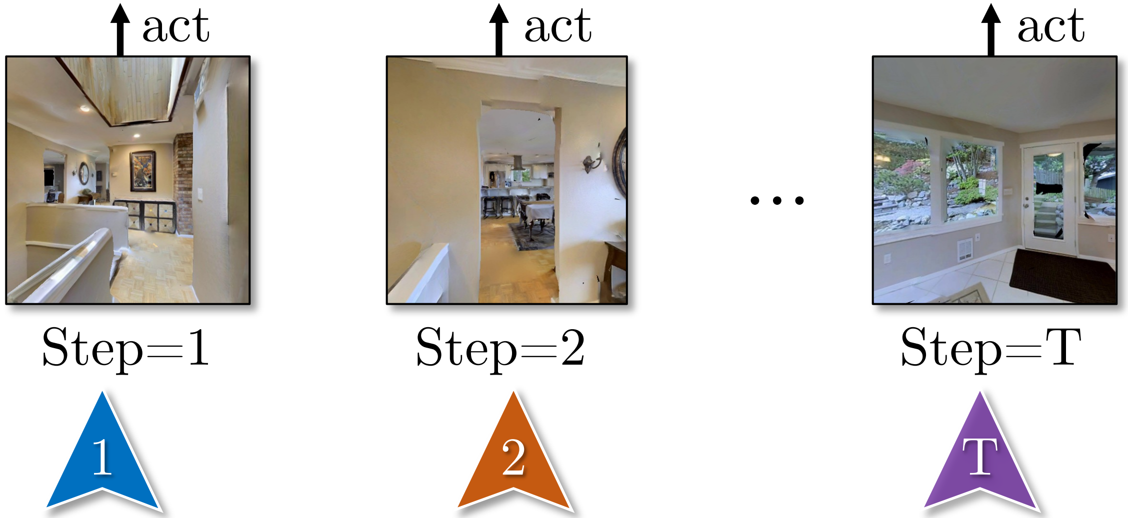
Decoupling of Planning and Perception



Test stage

Decoupling of Planning and Perception





PIX AGENT
(student)

Results

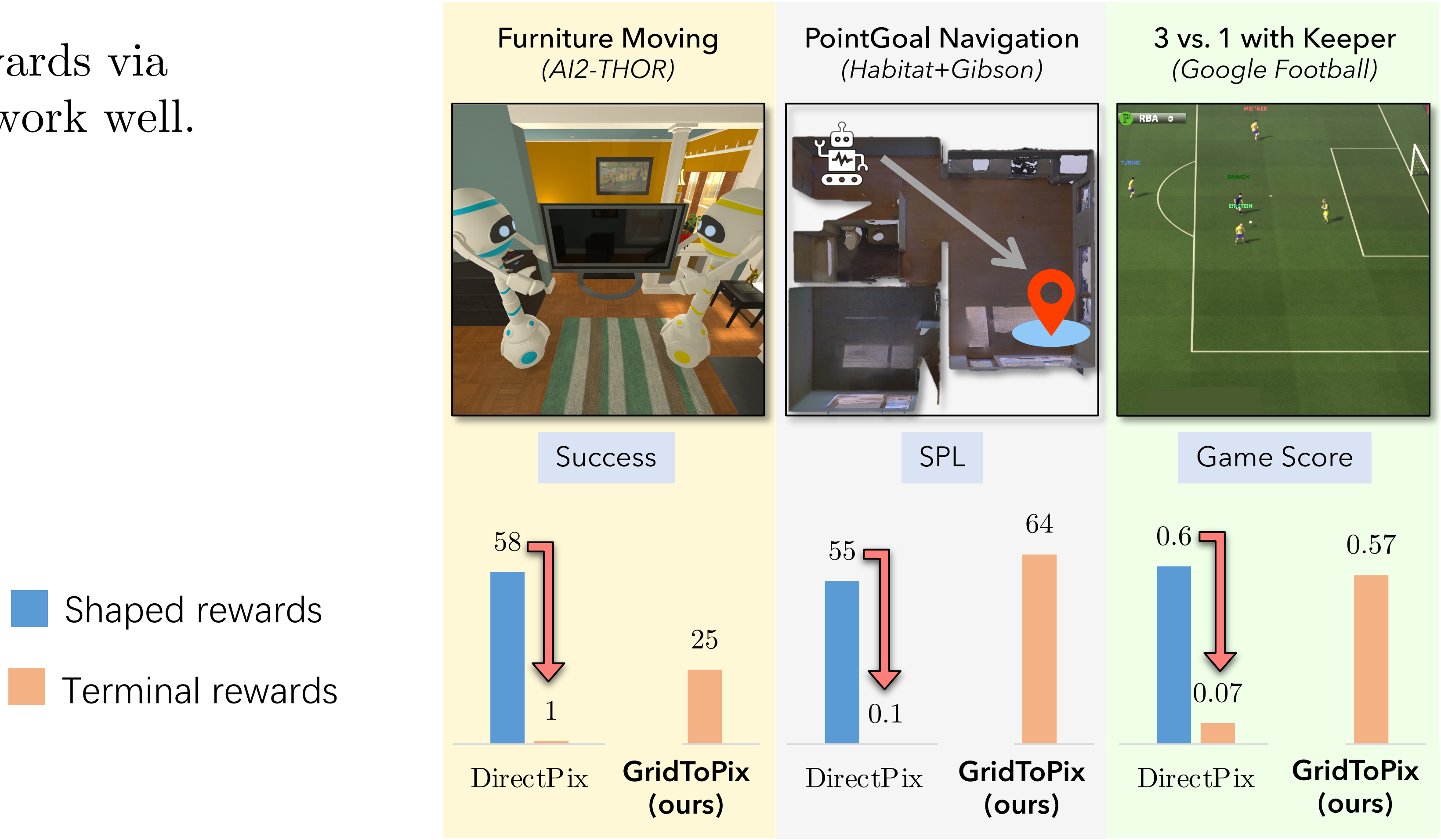
Terminal rewards do not work off-the-shelf.

Shaped rewards

Terminal rewards

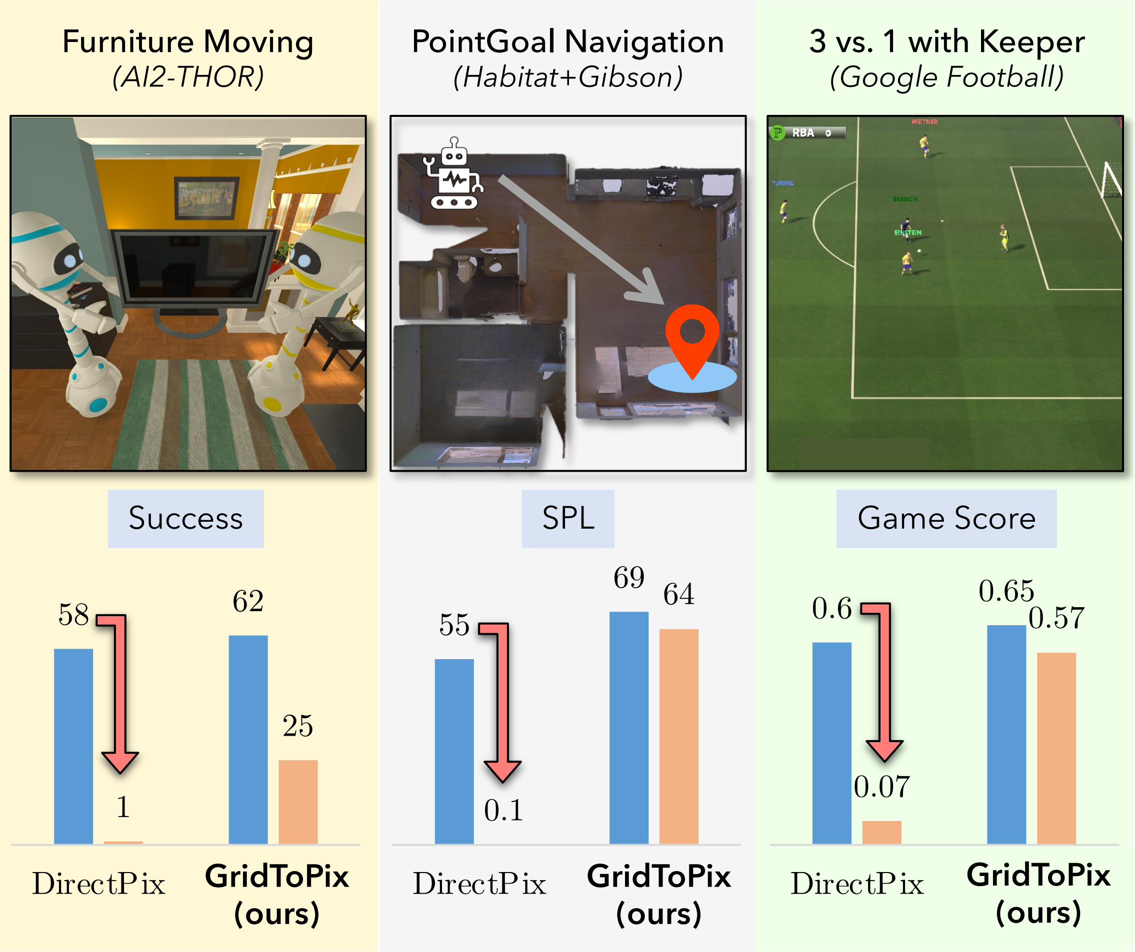
Results

Terminal rewards via GRIDTOPIX work well.



Results

Shaped rewards via GRIDTOPIX is better than a direct training.



Shaped rewards

Terminal rewards

Collaborative Embodied Agents

Takeaways

Two Body Problem
CVPR 2019

- Study collaborative behavior in visual environments
- Explicit and implicit communication are helpful
- Emergence of interpretable communication pattern

SYNC Policies
ECCV 2020

- FurnMove needs intricate coordination (high-rank joint)
- Independent and decentral execution ⇒ Rank-1
- SYNC-policies can capture a mixture-of-marginals

GRIDTOPIX
arXiv

- Visual RL agents crave dense and shaped rewards
- GRIDTOPIX leverages gridworlds for free supervision
- Improved results in terminal and shaped reward settings

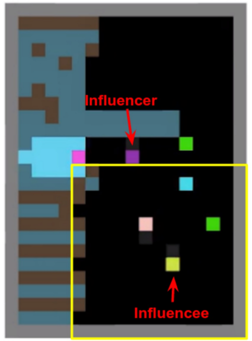
Steps Forward

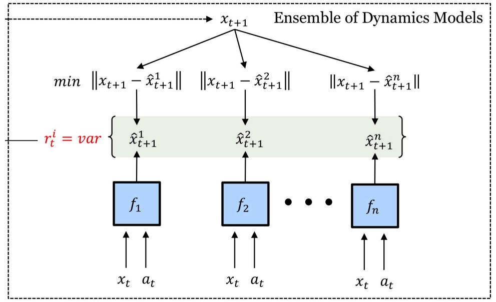
Visual Learning

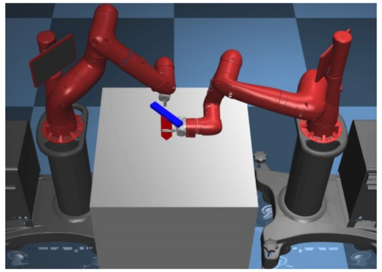
Multi-Agent RL

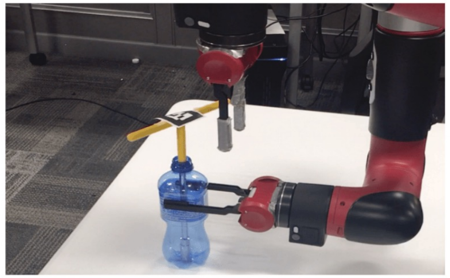
Collaborative Visual Agents

Robotic Agents









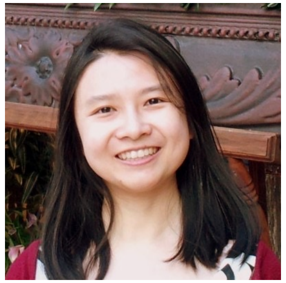
Chitnis et al. 2020
Jaques et al. 2018
Pathak et al. 2019

Thanks!

Alexander Schwing

Svetlana Lazebnik

Ali Farhadi
(UW, Apple)



Angel Chang
(SFU)

Aniruddha Kembhavi
(AI2, UW)

Kristen Grauman
(UT Austin, FAIR)

Luca Weihs
(AI2)

Manolis Savva
(SFU)