Implicitly Aligning Humans and Autonomous Agents through Shared Task Abstractions

IJCAI 2025

Stéphane Aroca-Ouellette, Miguel Aroca-Ouellette, Katharina von der Wense, Alessandro Roncone

Motivation

Human-agent collaboration

- Autonomous agents are becoming increasingly prevalent
- We are interested in how to create autonomous agents that can better collaborate with humans

Motivation

Zero-shot coordination

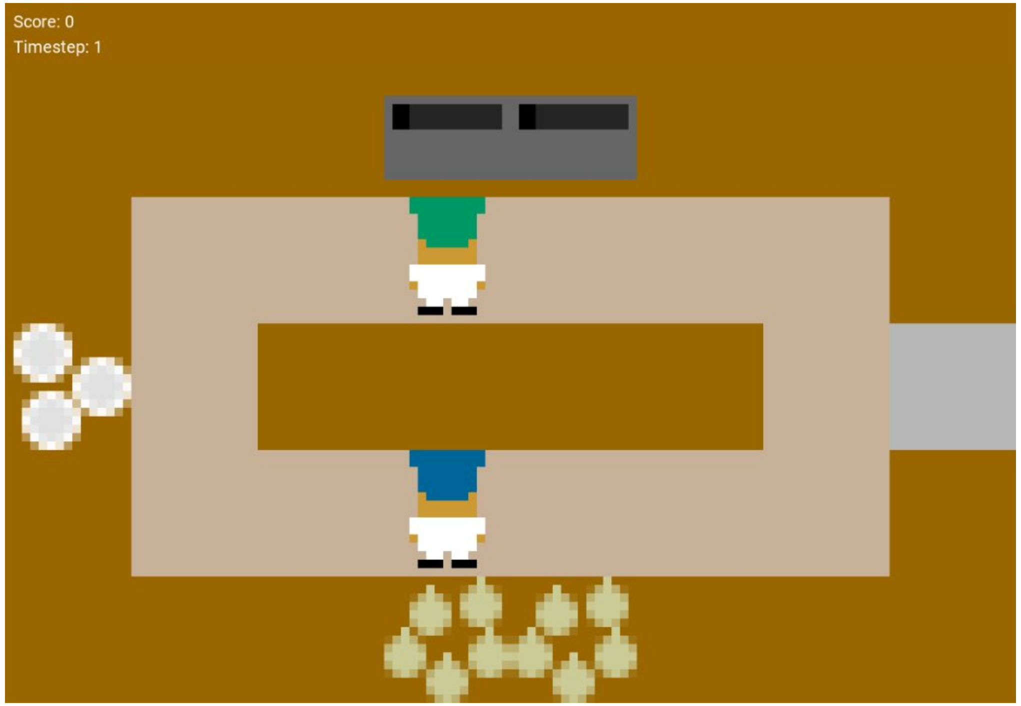
- Zero-shot coordination, or ad hoc teaming, is a domain where agents are teamed up with a new unknown teammate
- This teammate has unknown preferences, strategies, and proficiencies
- Success depends on quickly aligning toward a unified strategy

Overcooked

A collaborative cooking game

- We explore our approach in a simplified overcooked environment
- It involves two chefs cooperatively cooking onion soups
- Low level actions are:

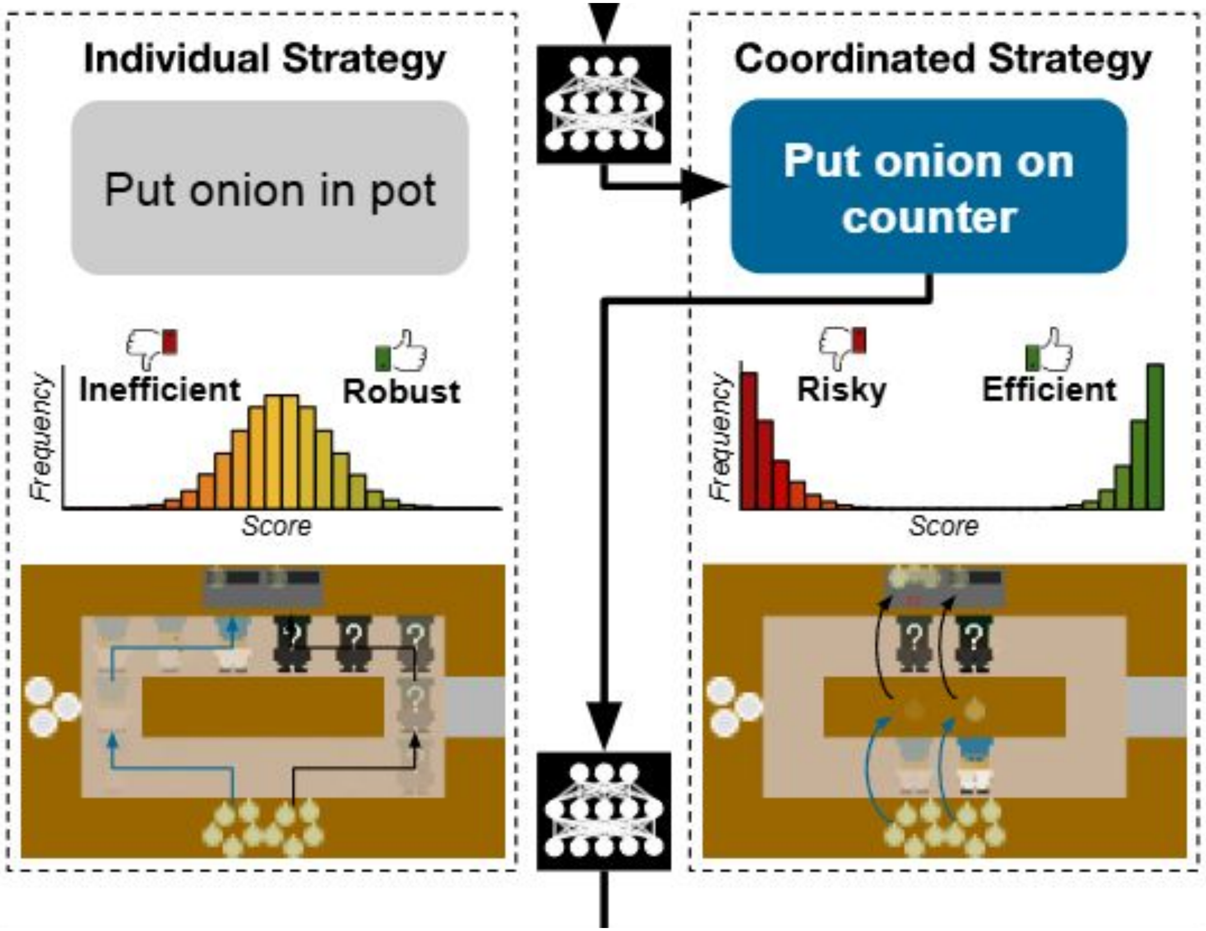
  UP, DOWN, LEFT, RIGHT, INTERACT, STAY
- Each soup rewards the whole team



Motivation

Illustrative Example - Why this problem is hard

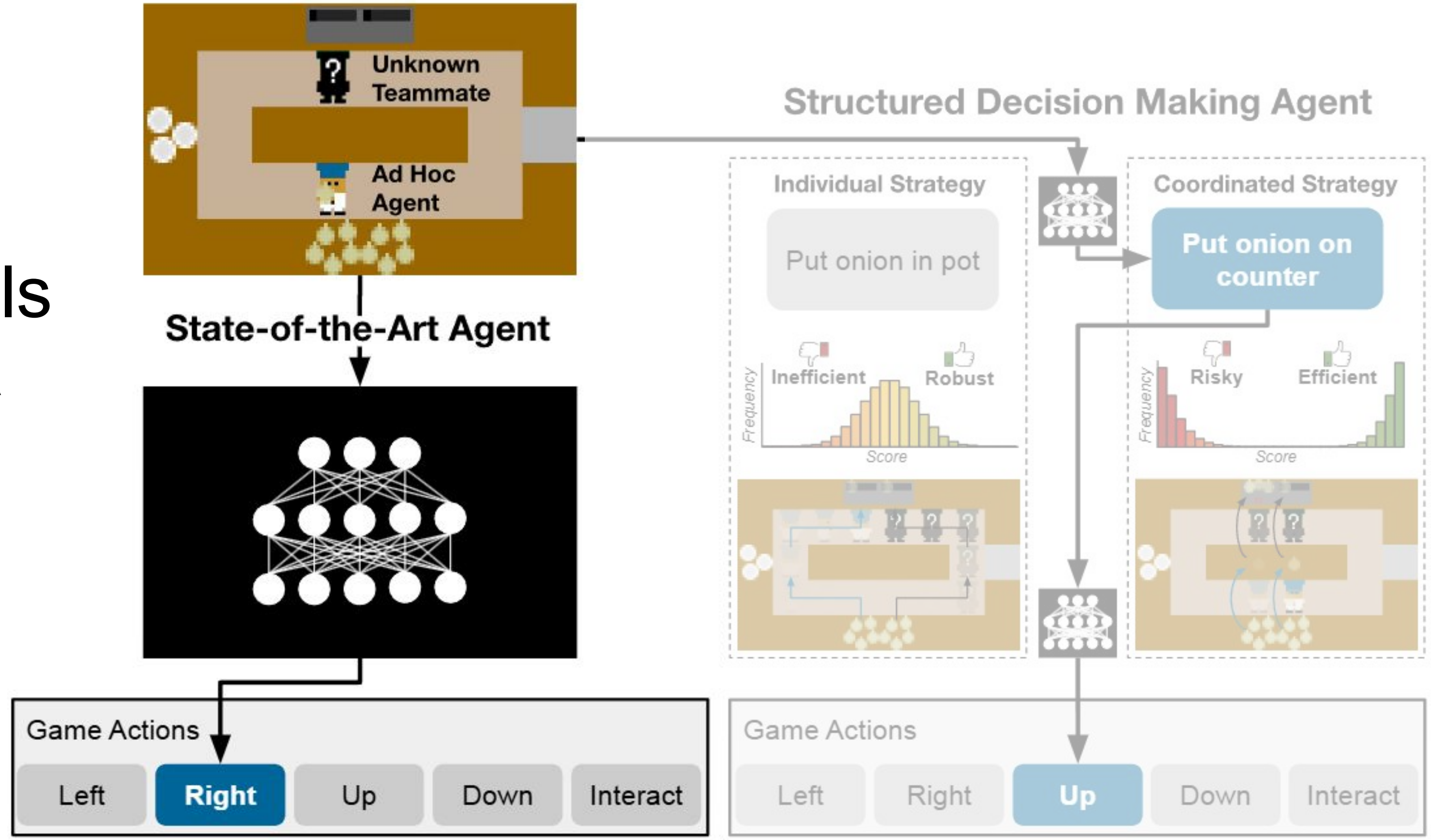
- Individual strategy is robust, but inefficient
- Coordinated strategy is efficient, but prone to failure
- Choosing the right strategy quickly is critical to success



Motivation

Current SotA

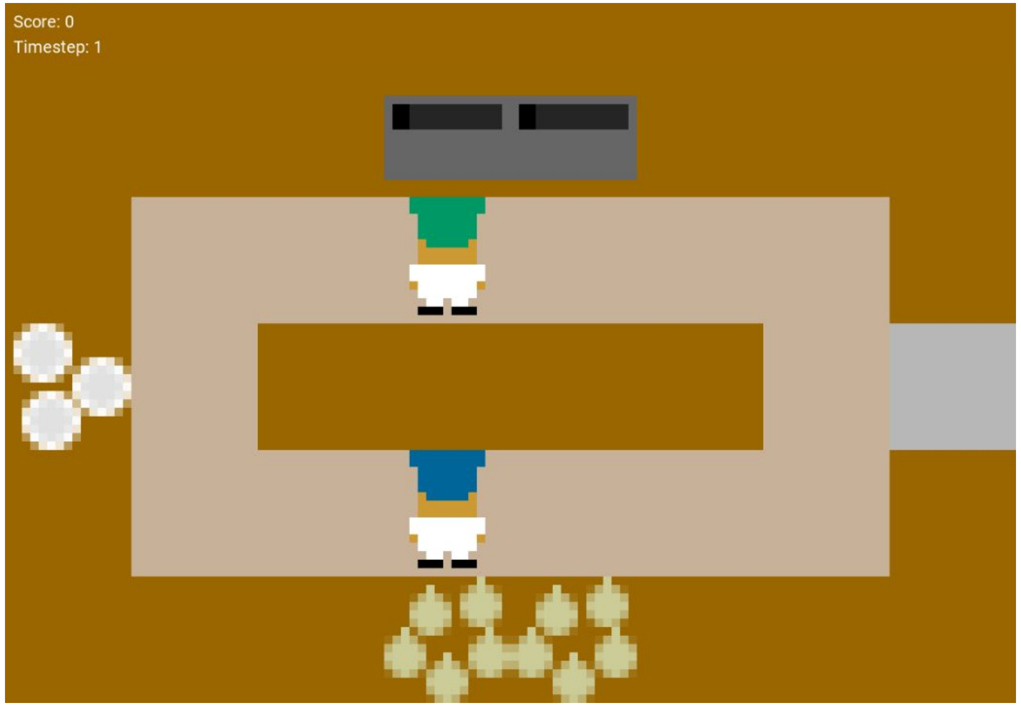
- Current state of the art models subsume all levels of decision making into a single black box



Overcooked

A collaborative cooking game

- We explore our approach in a simplified overcooked environment
- It involves two chefs cooperatively cooking onion soups
- Low level actions are:
  `UP, DOWN, LEFT, RIGHT, INTERACT, STAY`
- Each soup rewards the whole team
- Humans leverage subtasks such as:
  - *Placing an onion into a pot*
  - *Grabbing a dish from the dispenser*
  - *Serving a soup*



Motivation

Intuition for our Approach

- In collaborative tasks, humans frequently leverage task structures
- Can we develop agents that mirror this approach?
- These structure enable rapid generalization by:
  - Focusing agents on the relevant information at each level of abstraction
  - Preventing overfitting to specific patterns found in training
  - Creating more task-oriented agents
- Provides a shared foundation to anchor implicit alignment

Motivation

Intuition for our Approach

- In collaborative tasks, humans frequently leverage task structures
- Can we develop agents that mirror this human approach?
- These structure enable rapid generalization by
  - Focusing agents on the relevant information at each level of abstraction
  - Preventing overfitting to specific patterns found in training
  - Creating more task-oriented agents
- Provides a shared foundation to anchor implicit alignment

Motivation

Intuition for our Approach

- In collaborative tasks, humans frequently leverage task structures
- Can we develop agents that mirror this human approach?
- These structure enable rapid generalization by
  - Focusing agents on the relevant information at each level of abstraction
  - Preventing overfitting to specific patterns found in training
  - Creating more task-oriented agents
- Provides a shared foundation to anchor implicit alignment

Motivation

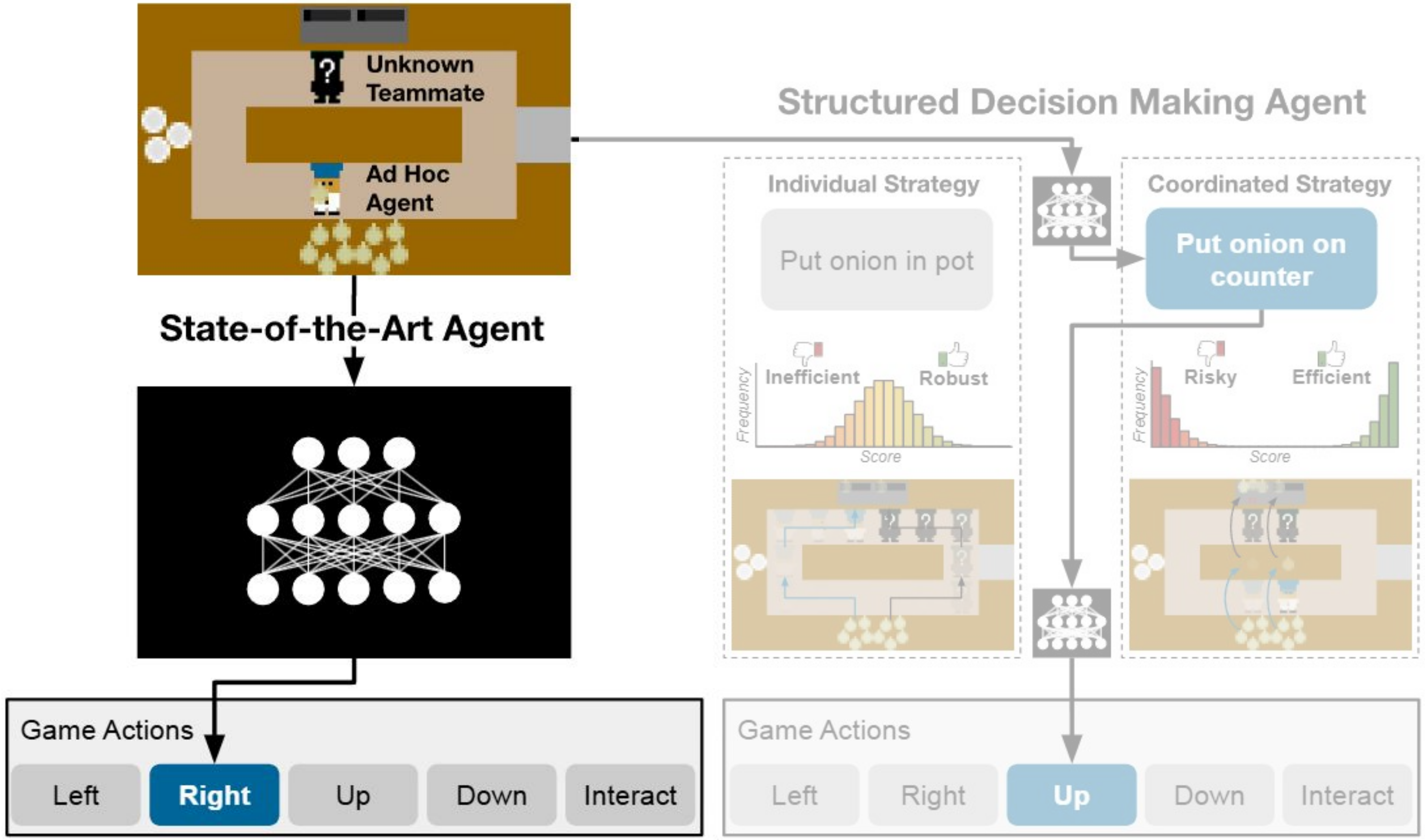
Intuition for our Approach

- In collaborative tasks, humans frequently leverage task structures
- Can we develop agents that mirror this human approach?
- These structure enable rapid generalization by:
  - Focusing agents on the relevant information at each level of abstraction
  - Preventing overfitting to specific patterns found in training
  - Creating more task-oriented agents
- Provides a shared foundation to anchor implicit alignment

Motivation

Leveraging shared task abstractions

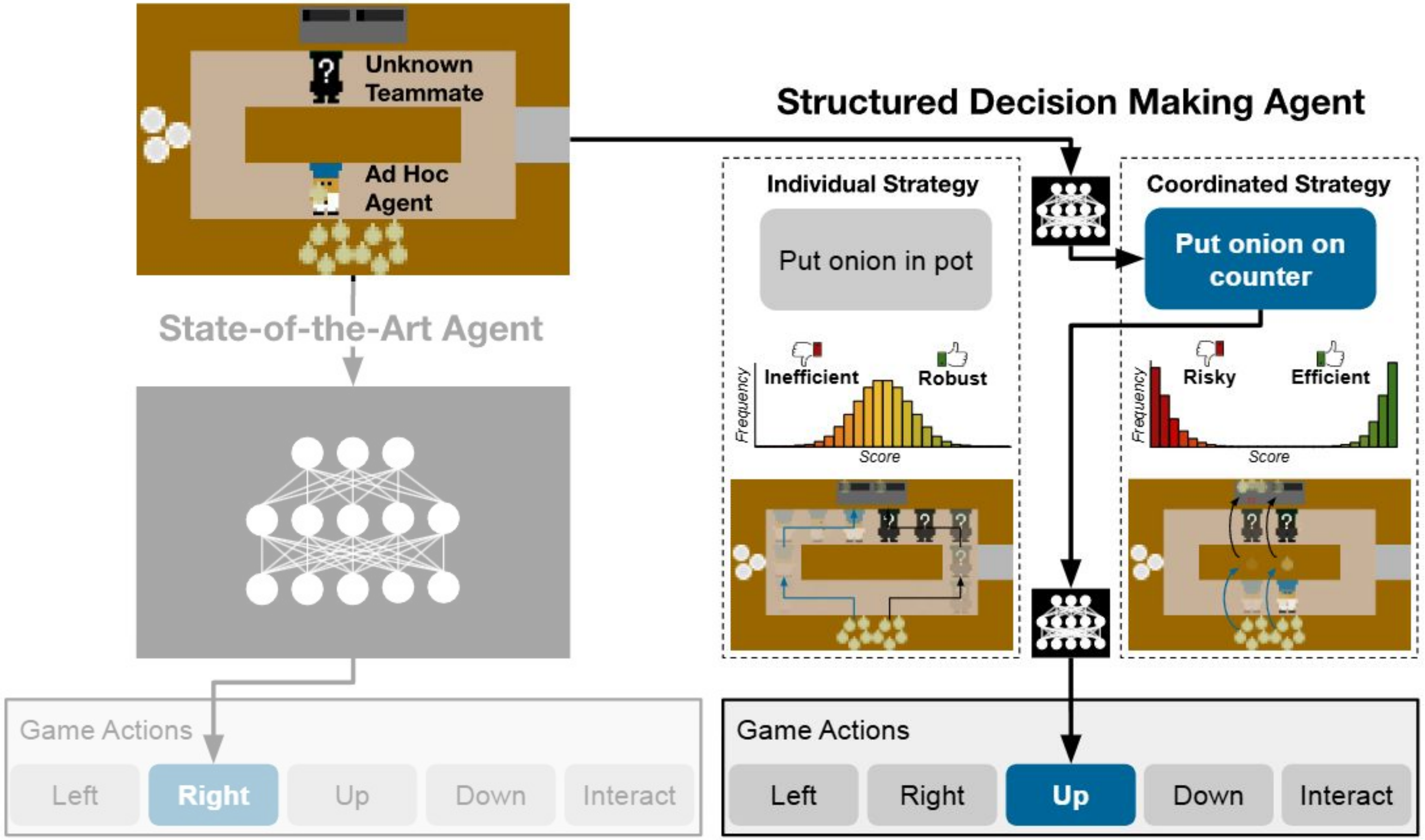
- Motivated by how humans rely on shared task abstractions for collaboration



Motivation

Leveraging shared task abstractions

- Motivated by how humans rely on shared task abstractions for collaboration
- We develop agents with human-interpretable task structures to improve implicit alignment of teammates

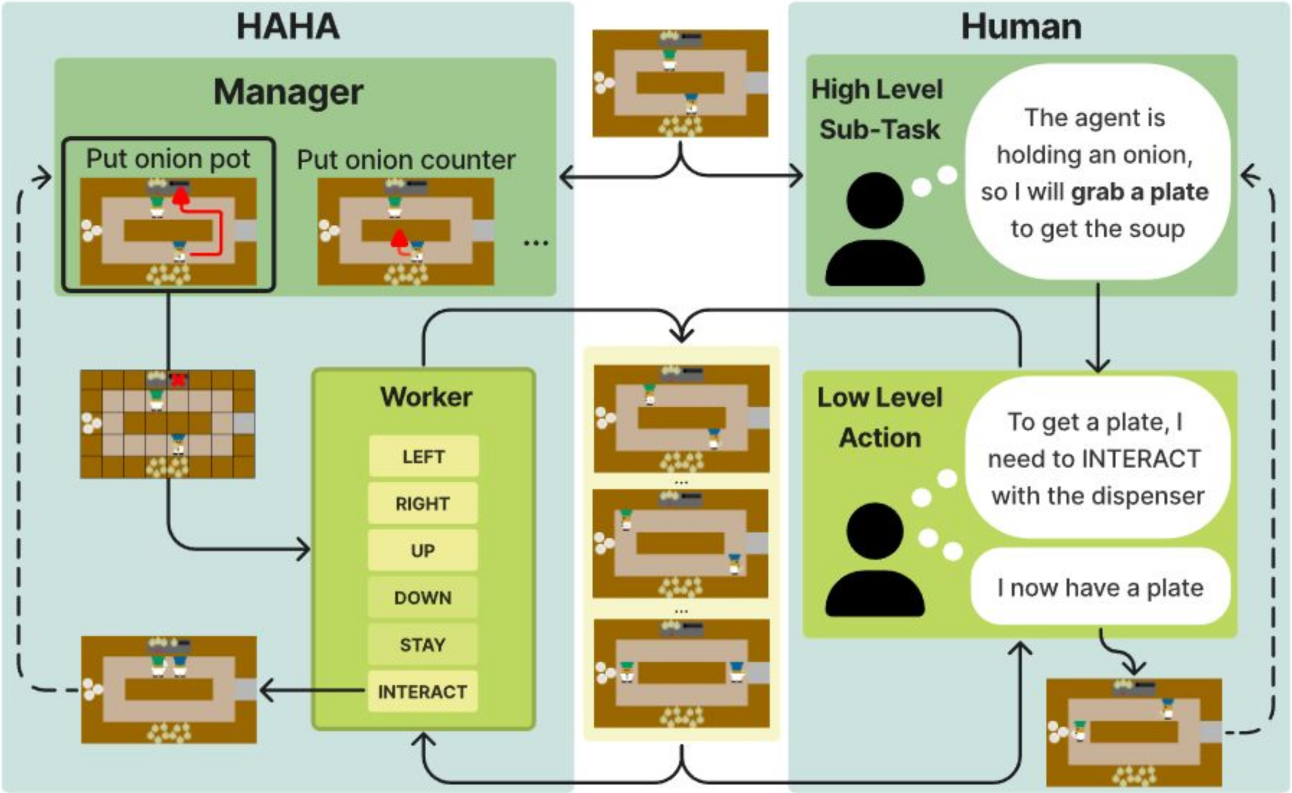


Method

Hierarchical task structure

We leverage FuN and develop HAHA (HA^2):
Hierarchical **A**d **H**oc **A**gents consisting of

1. A worker, which learns movement patterns to complete subtasks
2. A manager that focuses on which subtask to complete

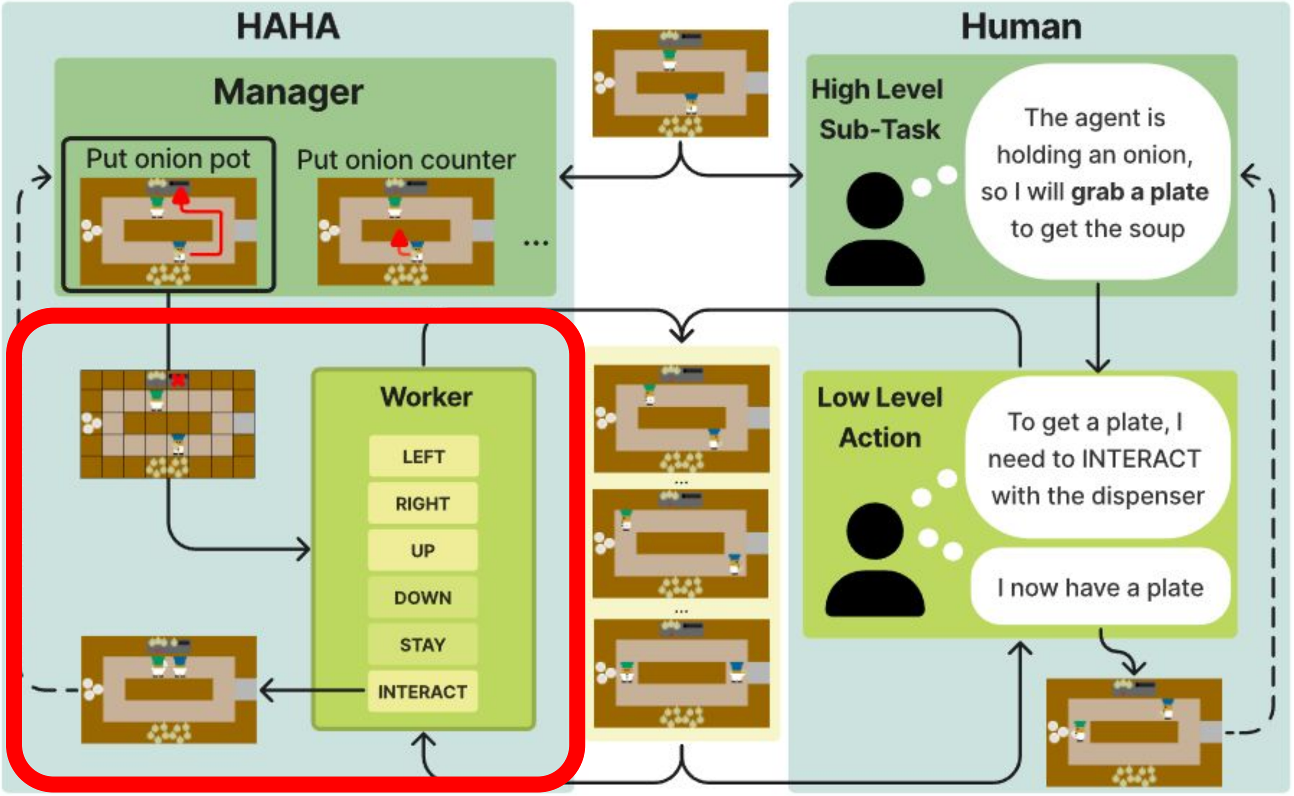


Method

Hierarchical task structure

We leverage FuN and develop HAHA (HA^2): **H**ierarchical **A**d **H**oc **A**gents consisting of

1. A worker, which learns movement patterns to complete subtasks
2. A manager that focuses on which subtask to complete

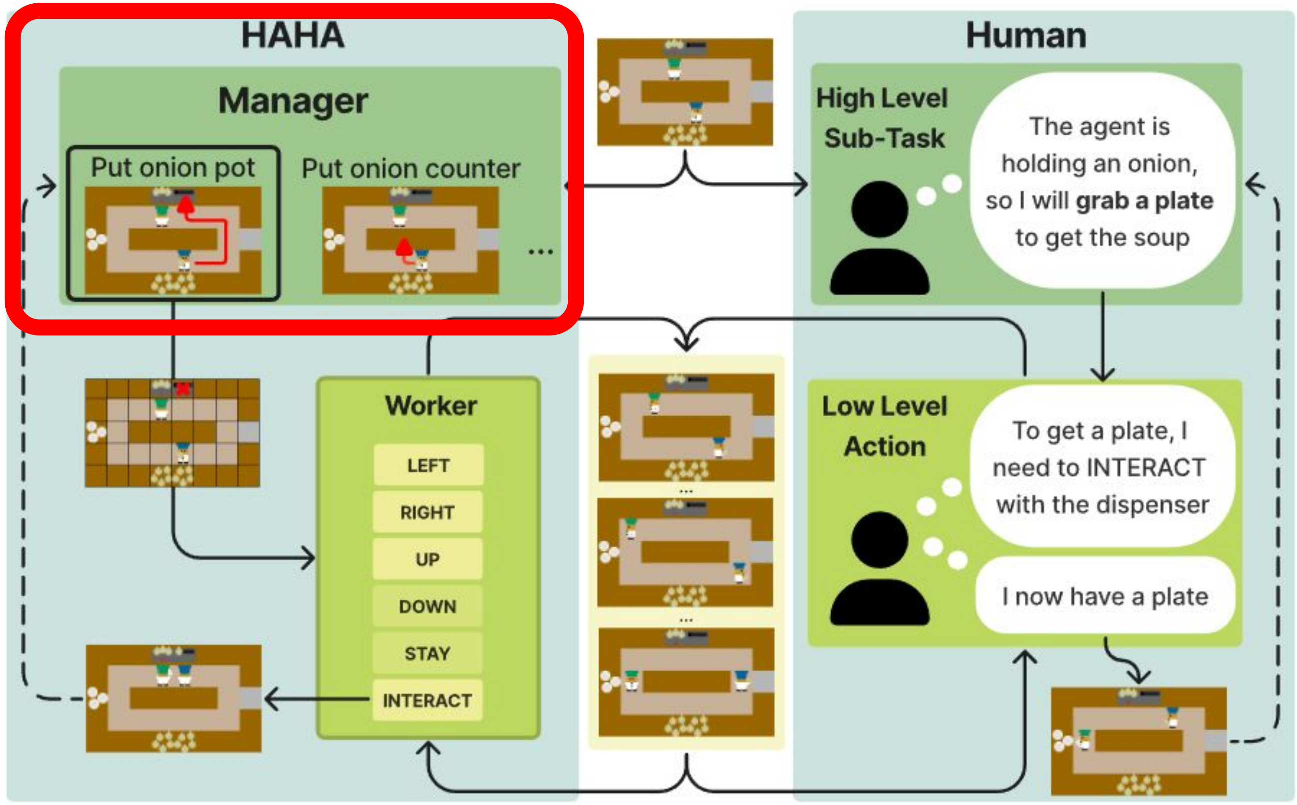


Method

Hierarchical task structure

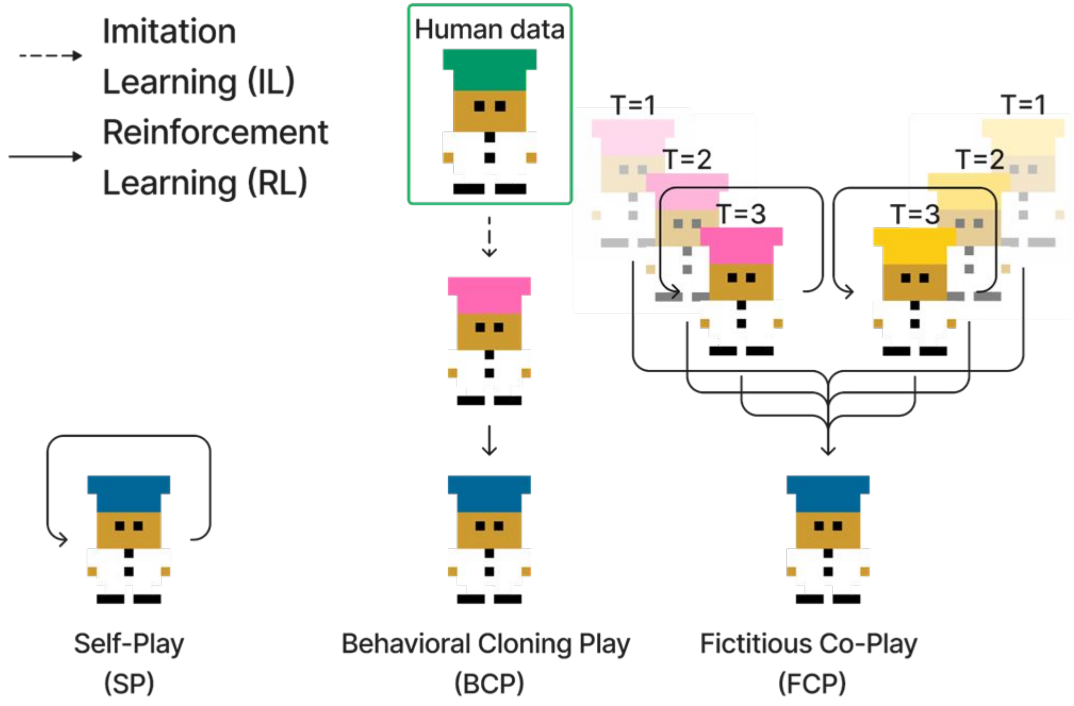
We leverage FuN and develop HAHA (HA^2):
Hierarchical **A**d **H**oc **A**gents consisting of

1. A worker, which learns movement patterns to complete subtasks
2. A manager that focuses on which subtask to complete



Method

Baselines



We train two versions of HAHA, one using the same teammates as BCP, and on using the same teammates as FCP.

Research Questions

1. Does HAHA improve performance with unseen agents?

2. Can HAHA agents generalize better to changes in the layouts?

3. Does HAHA create higher performing and more fluent human-agent teams?

Results

RQ1. Generalization to unseen agents

- Improved performance with unseen agents

	BCP	$HAHA_{BCP}$	FCP	$HAHA_{FCP}$
AA	199.9±8.0	278.3±6.3	210.8±40.0	**293.5±7.2**
CoR	79.2±4.2	133.3±3.2	138.6±2.5	**147.6±0.8**
CC	17.1±11.4	91.2±5.0	74.3±19.3	**99.9±2.8**
CrR	143.1±13.8	177.7±4.1	183.9 ±4.7	**185.5±2.3**
FC	73.1±5.6	**77.6±3.5**	56.7 ± 4.1	58.4±4.8
Avg.	102.5±4.5	151.6±2.4	133.0±8.8	**157.0±1.3**

Results

RQ1. Generalization to unseen agents

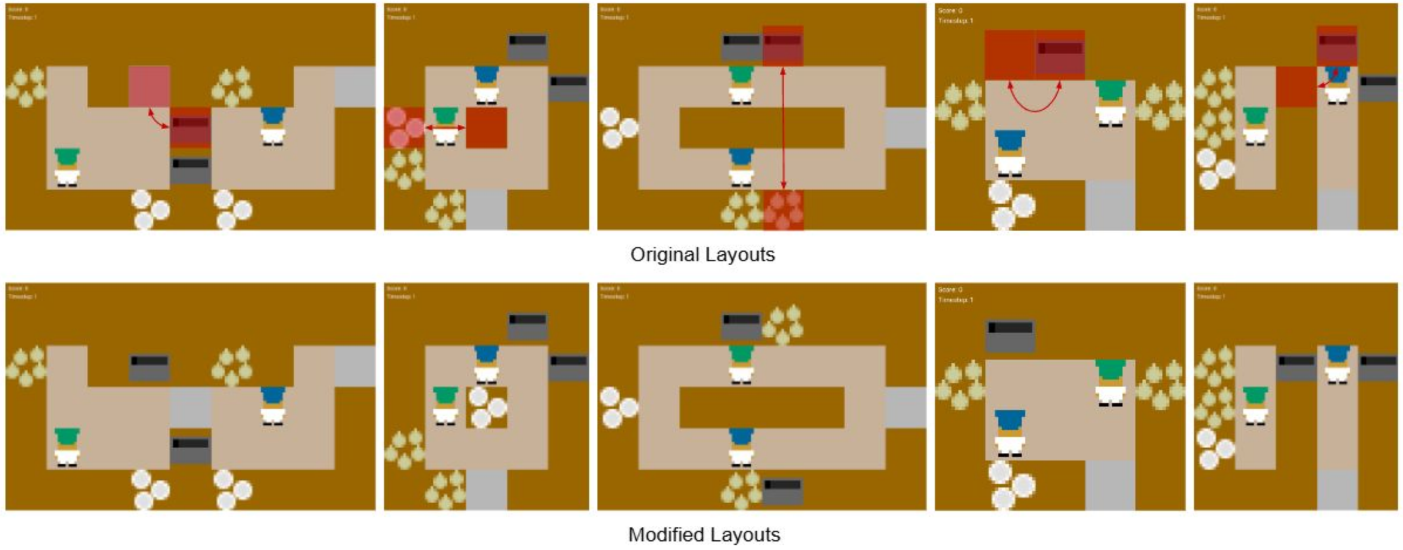
- Improved performance with unseen agents

	BCP	$HAHA_{BCP}$	FCP	$HAHA_{FCP}$
AA	199.9±8.0	278.3±6.3	210.8±40.0	**293.5±7.2**
CoR	79.2±4.2	133.3±3.2	138.6±2.5	**147.6±0.8**
CC	17.1±11.4	91.2±5.0	74.3±19.3	**99.9±2.8**
CrR	143.1±13.8	177.7±4.1	183.9 ±4.7	**185.5±2.3**
FC	73.1±5.6	**77.6±3.5**	56.7 ± 4.1	58.4±4.8
Avg.	102.5±4.5	151.6±2.4	133.0±8.8	**157.0±1.3**

Results

RQ2. Generalization to environment shifts

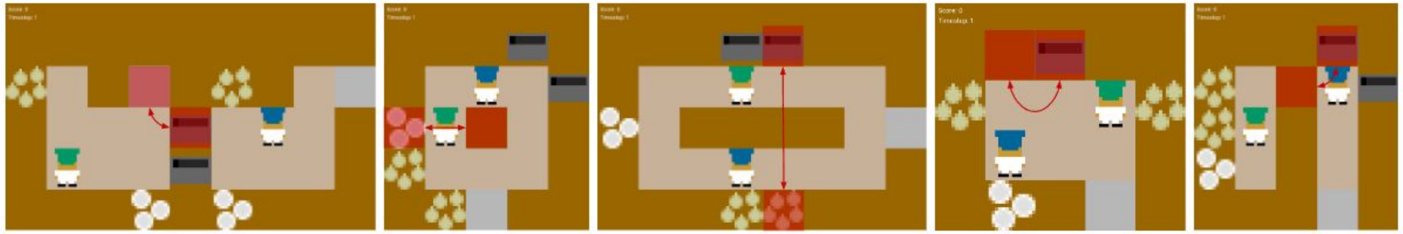
- More robust to environmental shifts



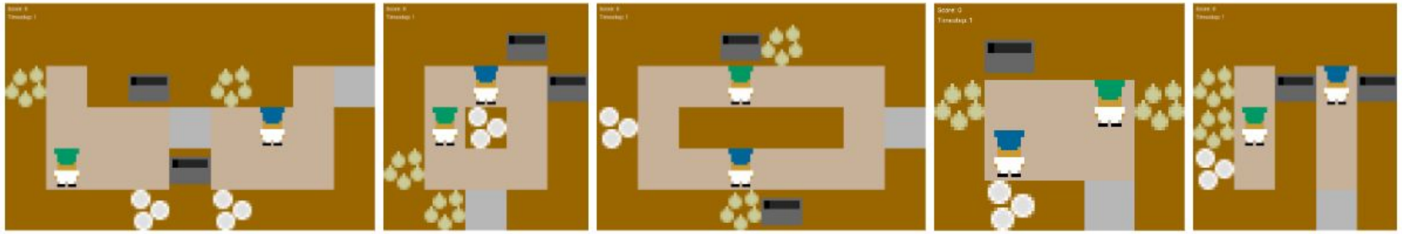
Results

RQ2. Generalization to environment shifts

- More robust to environmental shifts



Original Layouts



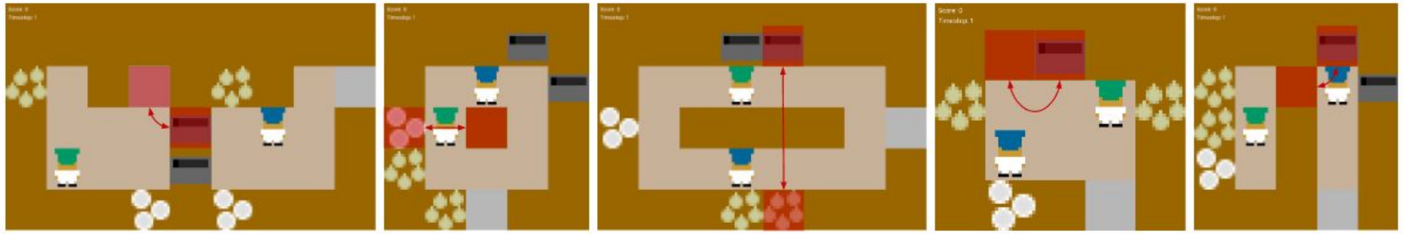
Modified Layouts

	BCP	HAHA_{BCP}	FCP	HAHA_{FCP}
AA	199.9±8.0	278.3±6.3	210.8±40.0	**293.5±7.2**
CoR	79.2±4.2	133.3±3.2	138.6±2.5	**147.6±0.8**
CC	17.1±11.4	91.2±5.0	74.3±19.3	**99.9±2.8**
CrR	143.1±13.8	177.7±4.1	183.9 ±4.7	**185.5±2.3**
FC	73.1±5.6	**77.6±3.5**	56.7 ± 4.1	58.4±4.8
Avg.	102.5±4.5	151.6±2.4	133.0±8.8	**157.0±1.3**
~ AA	23.6±41.5	157.2±40.4	7.6±14.2	**208.0±28.1**
~ CoR	11.6±11.4	**152.8±7.0**	22.8±6.4	143.2±12.6
~ CC	2.0±2.5	70.0±15.8	9.2±14.5	**110.0±35.5**
~ CrR	5.6±2.9	**162.4±15.2**	0.8±1.6	154.8±36.8
~ FC	10.4±8.9	17.2±31.5	3.2±3.0	**20.8±31.7**
~ Avg.	10.6±9.5	111.9±13.4	8.7±2.4	**127.3±7.1**

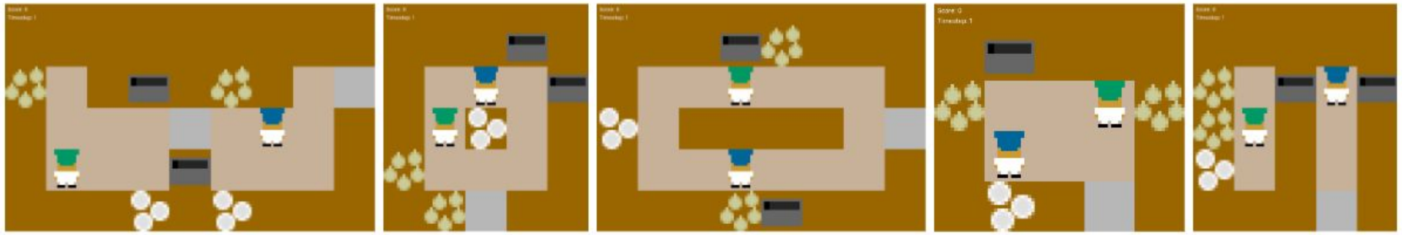
Results

RQ2. Generalization to environment shifts

- More robust to environmental shifts



Original Layouts



Modified Layouts

	BCP	HAHA$_{BCP}$	FCP	HAHA$_{FCP}$
AA	199.9±8.0	278.3±6.3	210.8±40.0	**293.5±7.2**
CoR	79.2±4.2	133.3±3.2	138.6±2.5	**147.6±0.8**
CC	17.1±11.4	91.2±5.0	74.3±19.3	**99.9±2.8**
CrR	143.1±13.8	177.7±4.1	183.9 ±4.7	**185.5±2.3**
FC	73.1±5.6	**77.6±3.5**	56.7 ± 4.1	58.4±4.8
Avg.	102.5±4.5	151.6±2.4	133.0±8.8	157.0±1.3
~ AA	23.6±41.5	157.2±40.4	7.6±14.2	**208.0±28.1**
~ CoR	11.6±11.4	**152.8±7.0**	22.8±6.4	143.2±12.6
~ CC	2.0±2.5	70.0±15.8	9.2±14.5	**110.0±35.5**
~ CrR	5.6±2.9	**162.4±15.2**	0.8±1.6	154.8±36.8
~ FC	10.4±8.9	17.2±31.5	3.2±3.0	**20.8±31.7**
~ Avg.	10.6±9.5	111.9±13.4	8.7±2.4	127.3±7.1

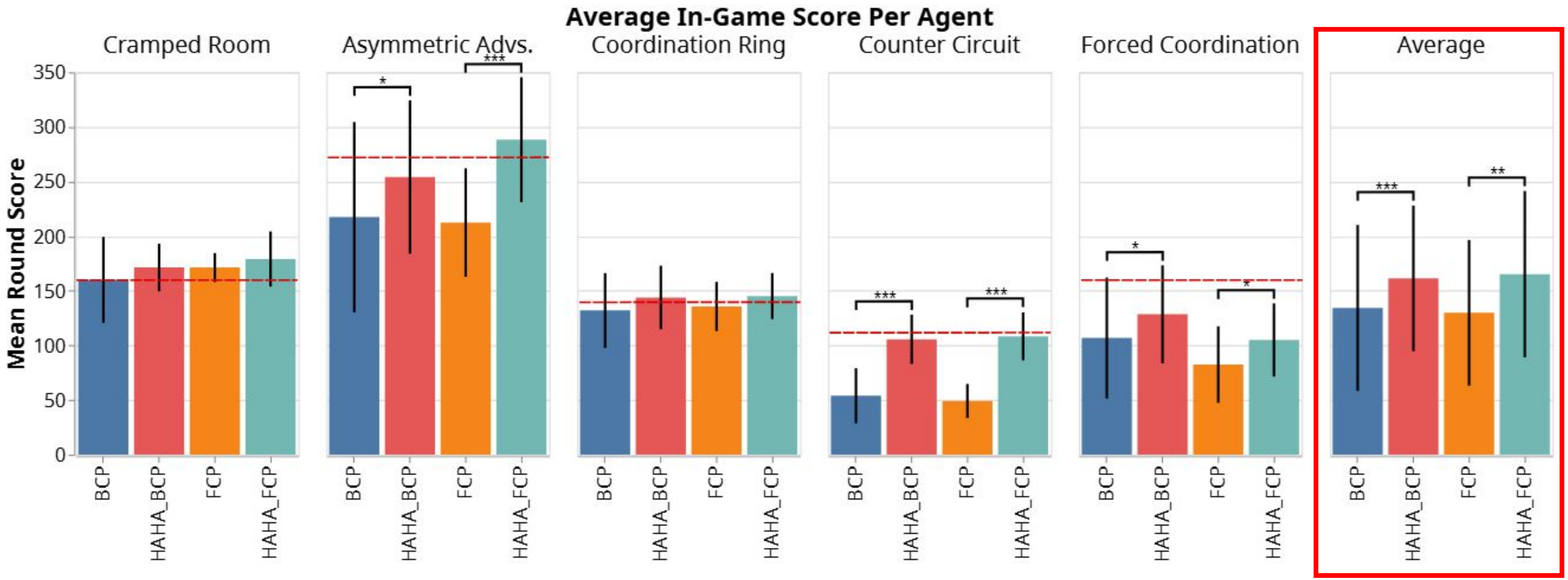
User Study

This is about humans after all

- Recruited 75 participants to run an online user study
  - Participants were recruited on Prolific.co
- Within-subject design
- Each participant played 10 rounds, playing with a HAHA agent and their respective baseline on each of the five layouts
  - Order of agents was randomized
- Rounds lasted 80s (400 timesteps @5FPS)
- Answered a survey of likert scale questions between rounds
- Selected which agent they preferred playing with

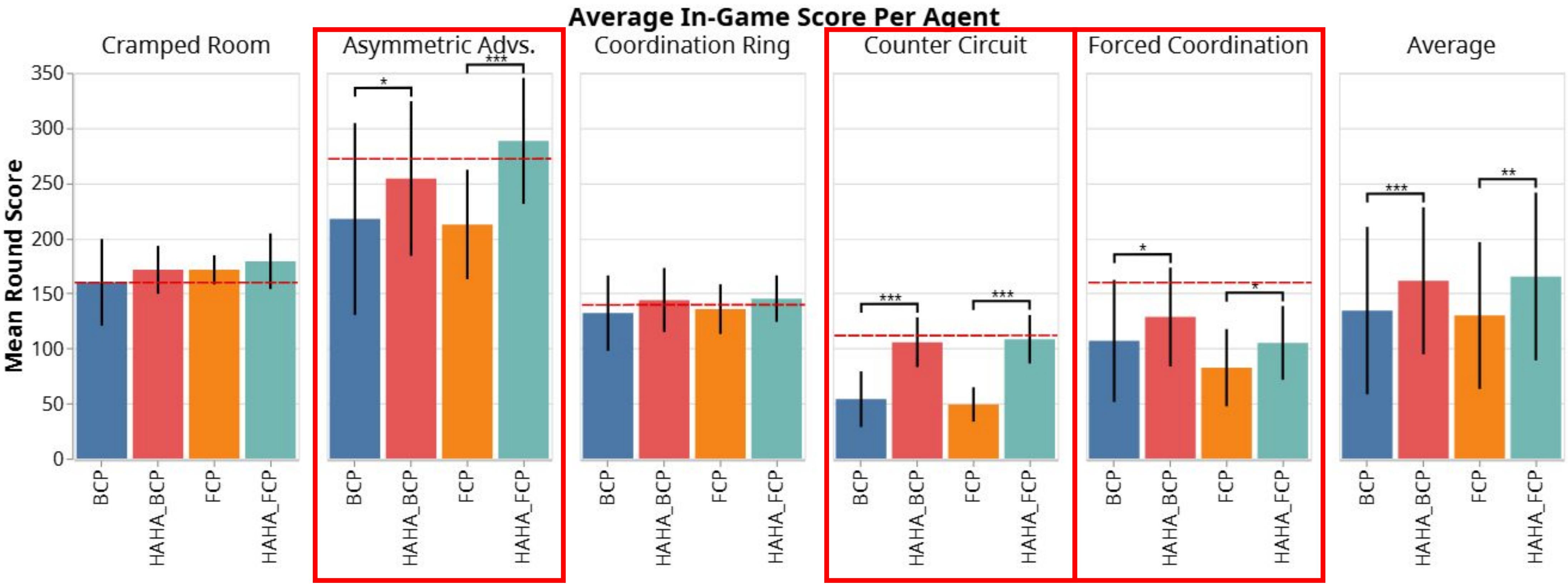
Results

RQ3. Generalization to human teammates



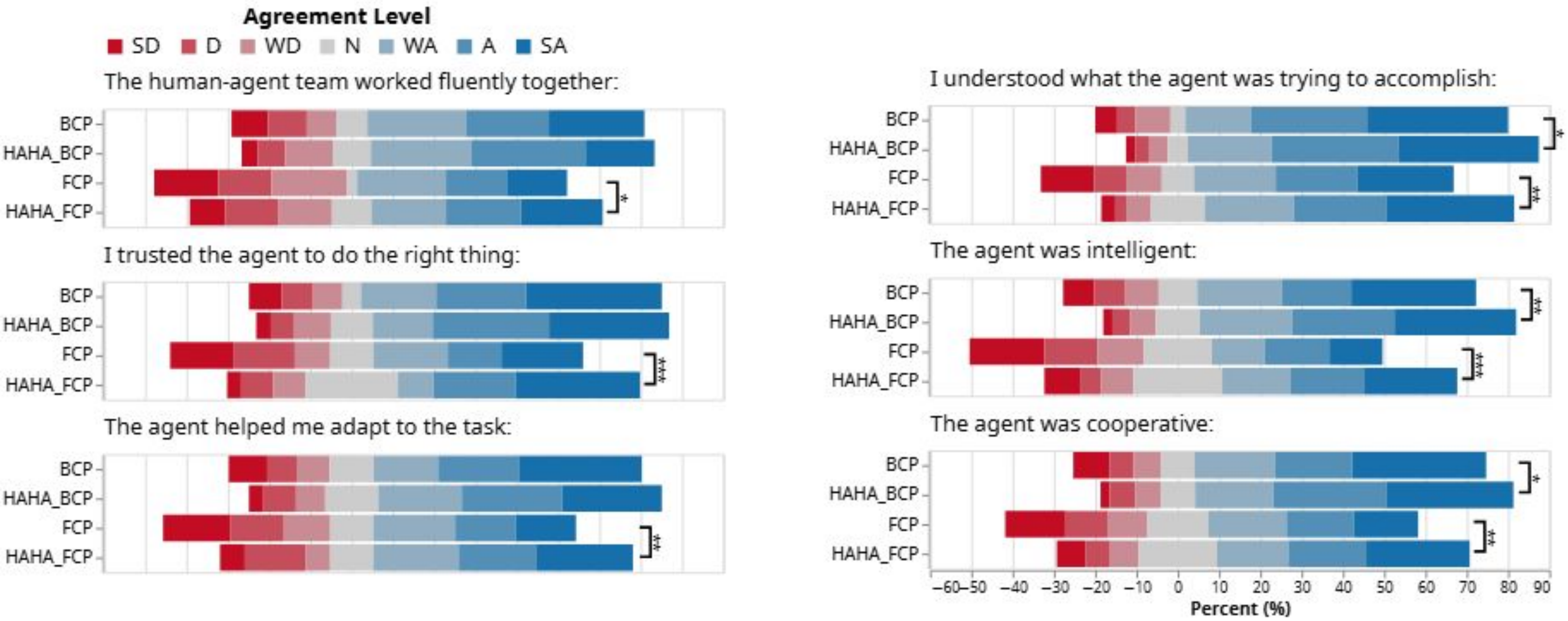
Results

RQ3. Generalization to human teammates



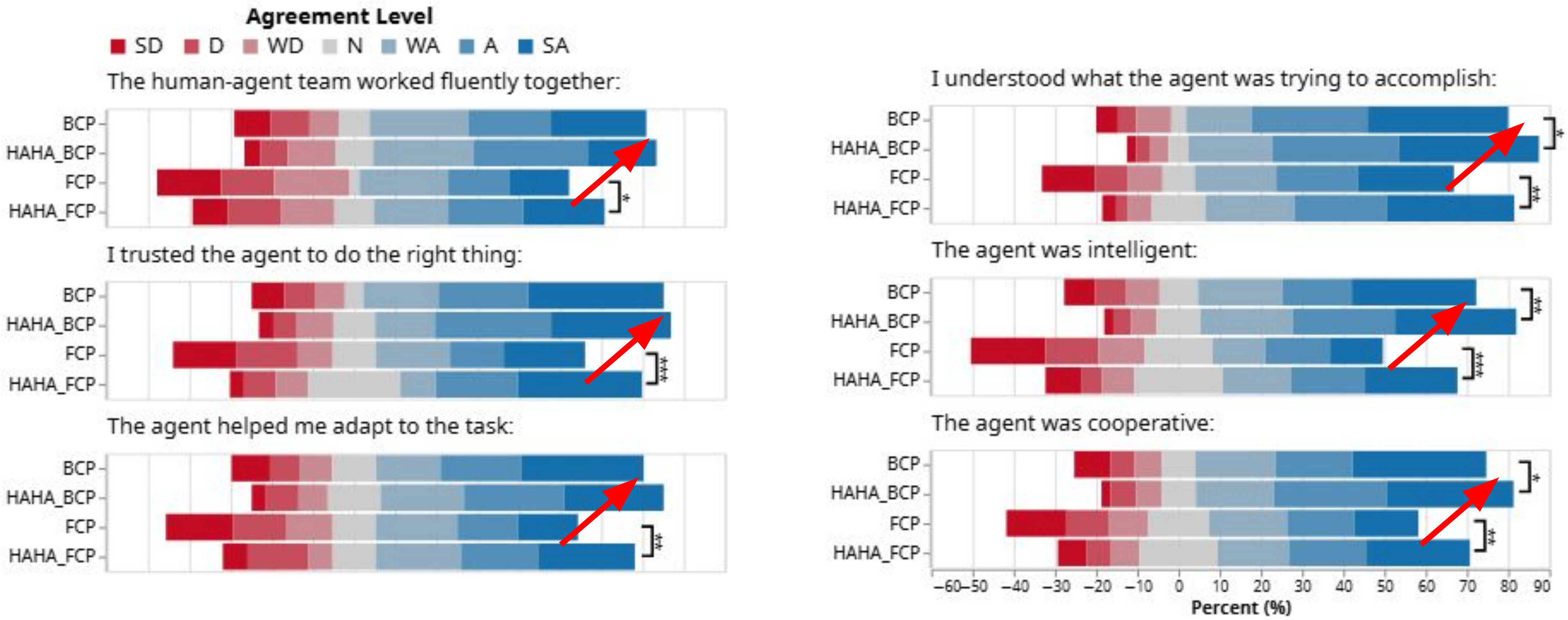
Results

RQ3. Generalization to human teammates



Results

Insight: is human gameplay still valuable?



Humans preferred HAHA

Compared to baselines

	% Preferred	p-value
HA^2_{BCP} over BCP	57.68	0.0070
HA^2_{FCP} over FCP	65.25	0.0000018

Table 2: Human preference between pairs of agents and their respective significance.

Comparison to SotA

System-level comparison

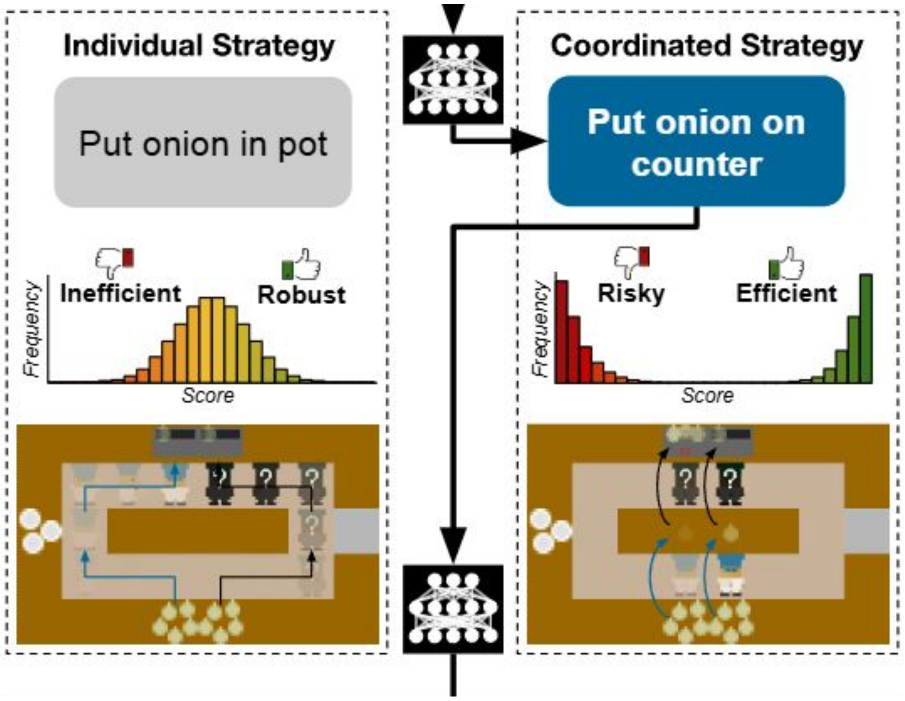
	Training Steps	W. Proxy	W. Humans
FCP	1.0e9	**157**	119
MEP	5.5e7*	98	98
TrajeDi	5.5e7*	76	87
PECAN	NR	105	134
HiPT	1.0e9	134	131
GAMMA	1.5e8	132	NR
HA^2_{FCP}	6.6e7	**157**	**165**

Table 3: Results comparing HA^2 to other published results. All results are taken from the respective works and adjusted to 400 timesteps, except for TrajeDi's results which are taken from [Zhao *et al.*, 2023]. NR=not reported. * indicates that separate agents are trained for each layout and that the cumulative step count across layouts is presented. FCP [Strouse *et al.*, 2021], MEP [Zhao *et al.*, 2023], TrajeDi [Lupu *et al.*, 2021], PECAN [Lou *et al.*, 2023], HiPT [Loo *et al.*, 2023], GAMMA [Liang *et al.*, 2024],

Tuning HA^2

Post training adjustments

- The human interpretable layer allows us to tune HA^2 post-training
- We can manually weigh certain subtasks
- Increases score with humans (104 →108)
- Increase human perception of the agent
- Even greater benefits when paired with itself, (137 → 154)



	% Preferred	p-value
HAHA_{BCP} over BCP	57.68	0.0070
HAHA_{FCP} over FCP	65.25	0.0000018
HAHA_{tuned} over HAHA	66.67	0.0093

HA^2

Takeaways

- Humans rely on shared task abstractions to establish common ground in collaboration.
- We can extend this approach to autonomous agents to improve generalization to new and unseen teammates.

Thank you for listening!

Please come see me at the poster session if you have any questions