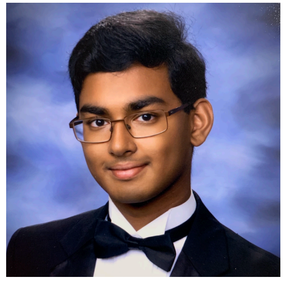
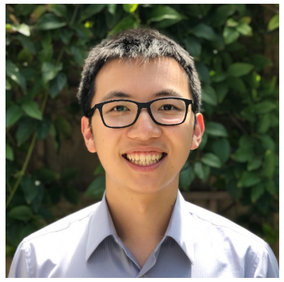
Diverse Conventions for Human-AI Collaboration



Bidipta Sarkar



Andy Shih

Dorsa Sadigh

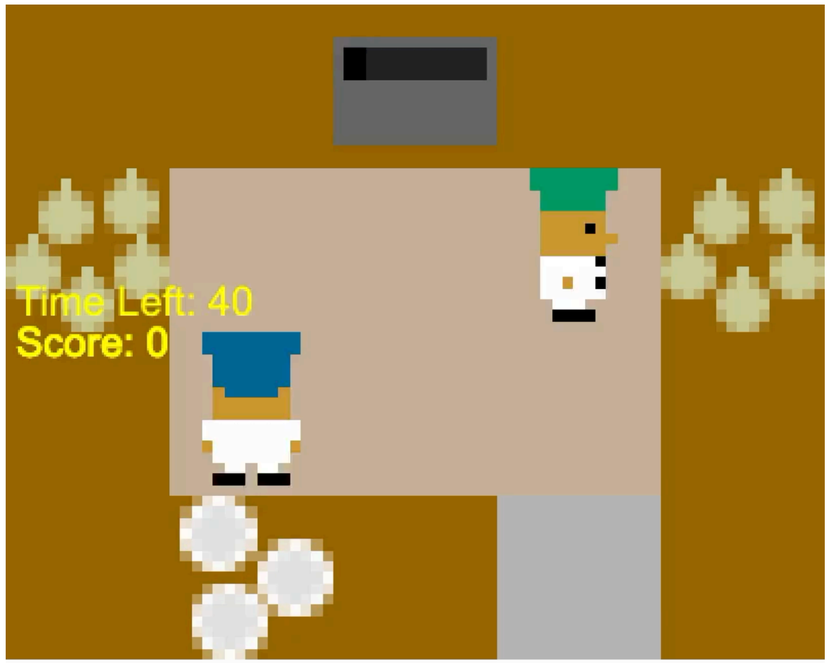
Conventions

Arbitrary solutions to recurring coordination problems

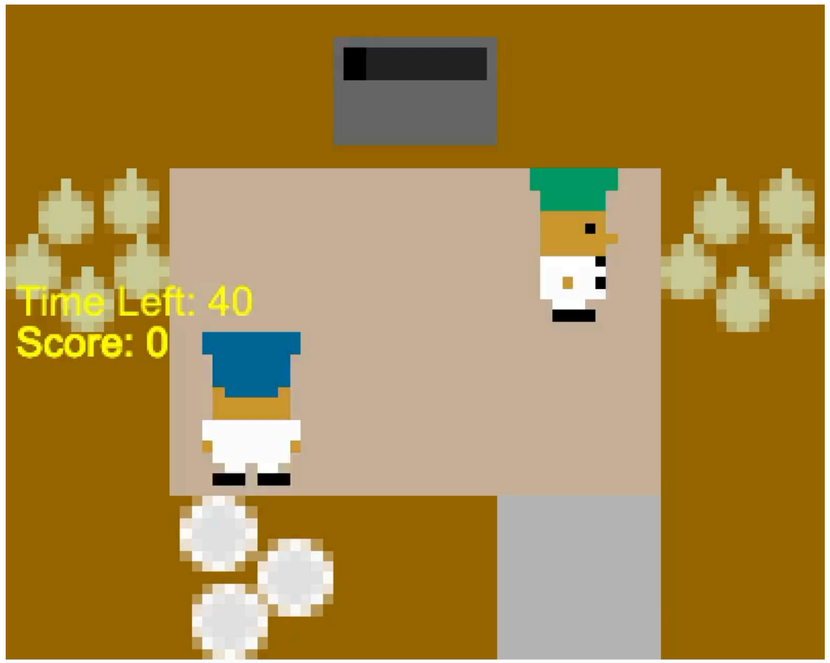
Conventions

Arbitrary solutions to recurring coordination problems

AI Conventions

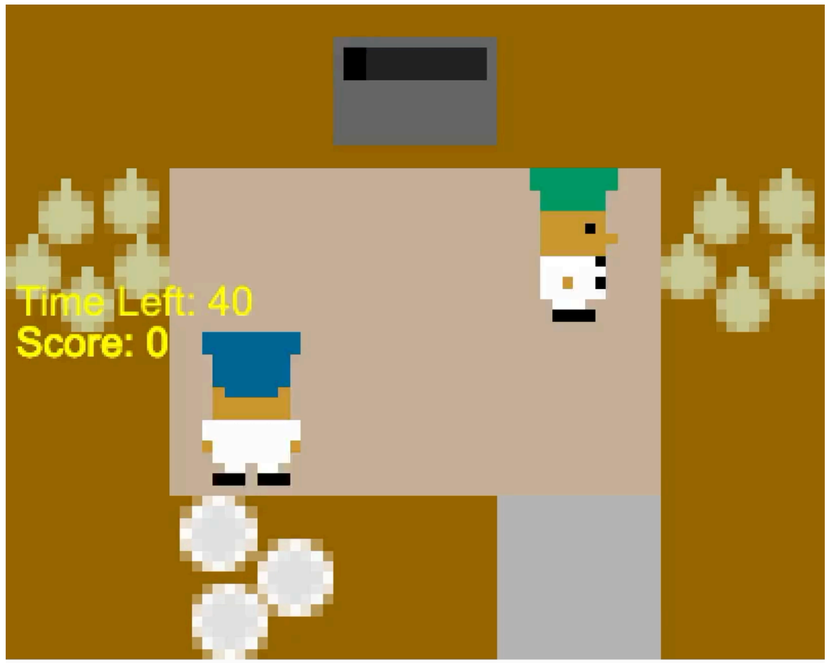


AI Conventions

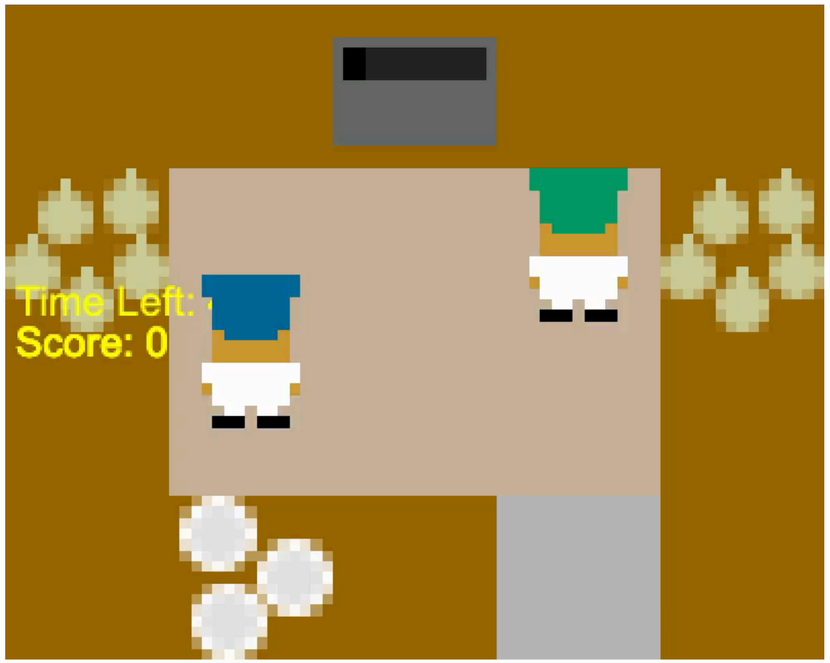


Two AI Agents achieve expert human performance

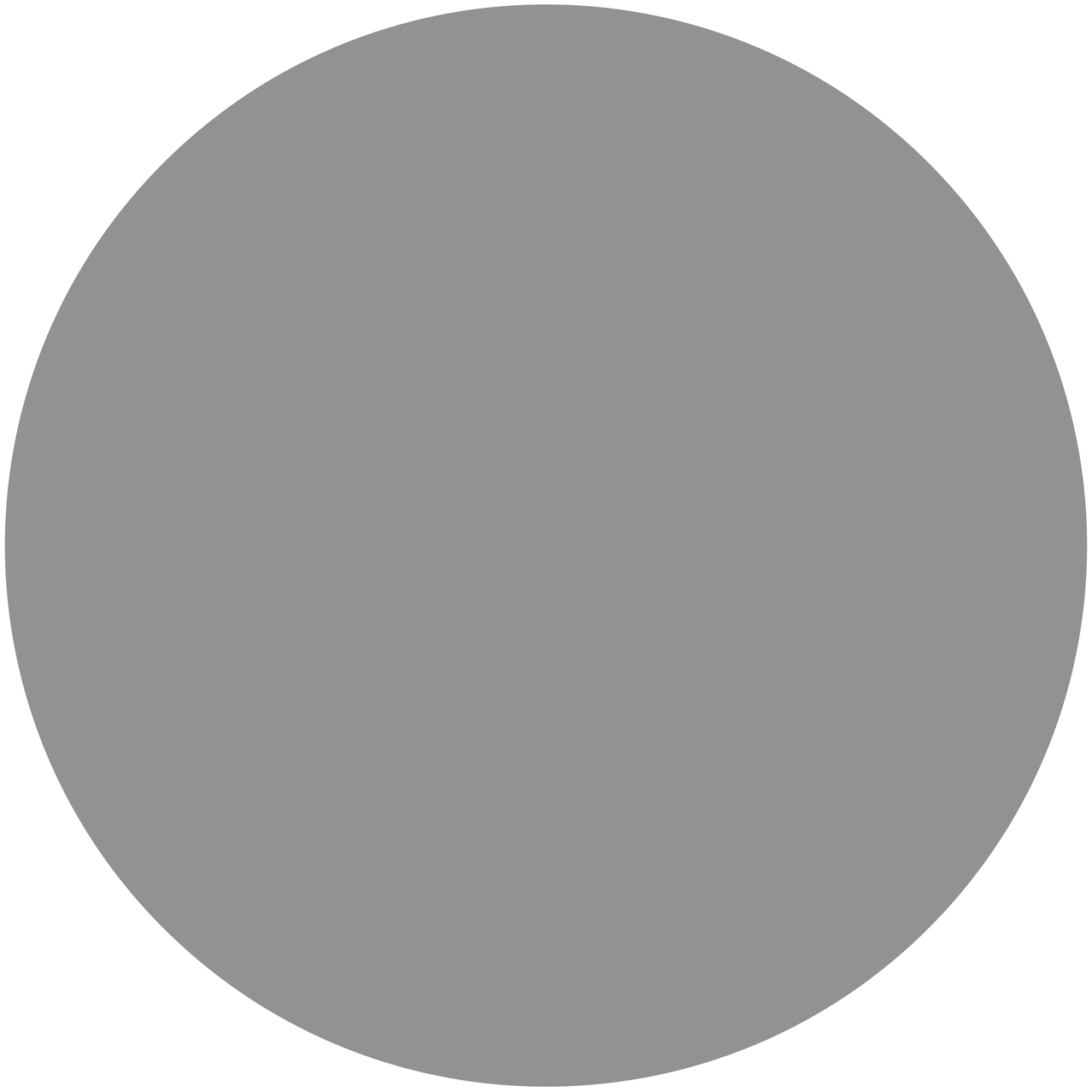
AI Conventions

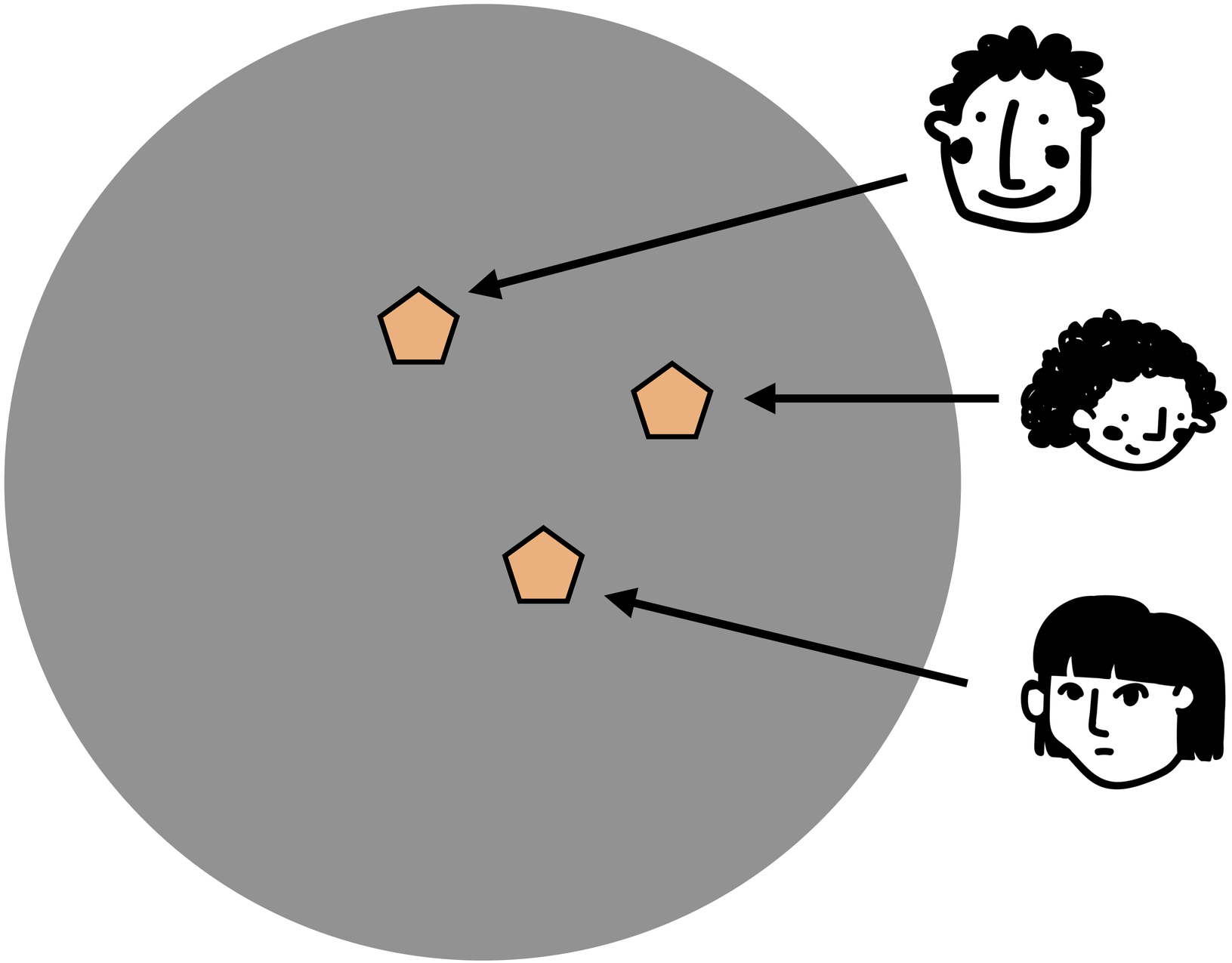


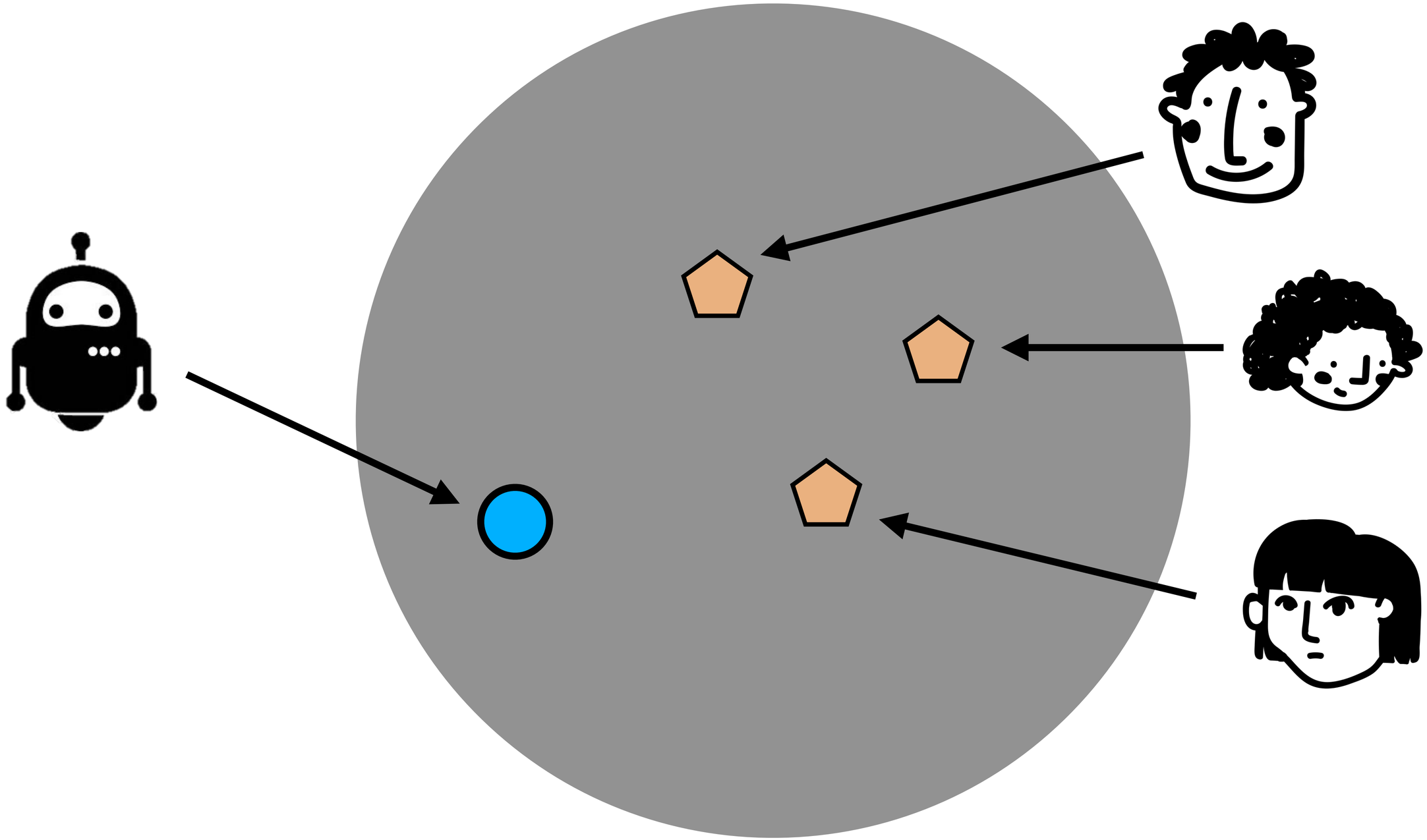
Two AI Agents achieve expert human performance

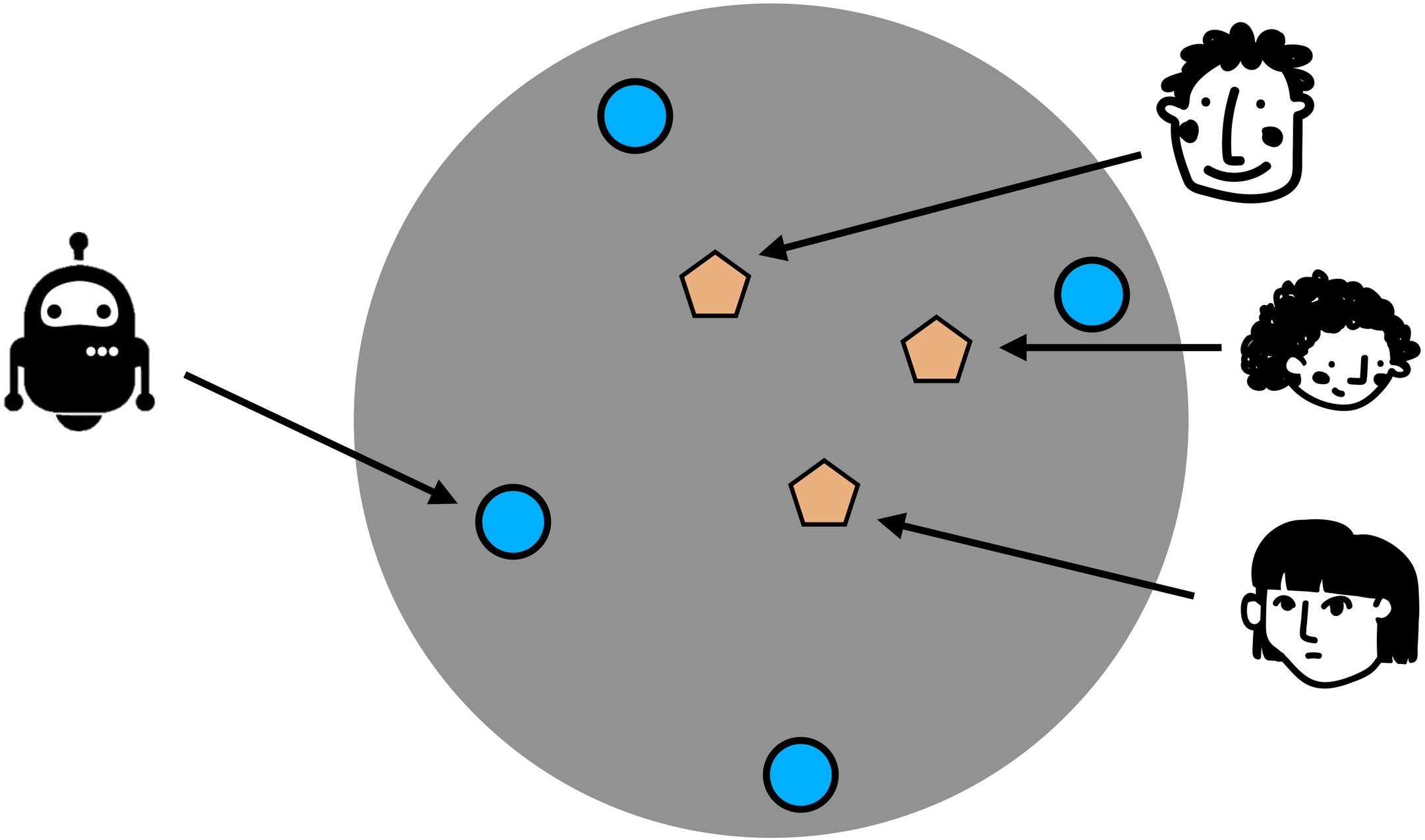


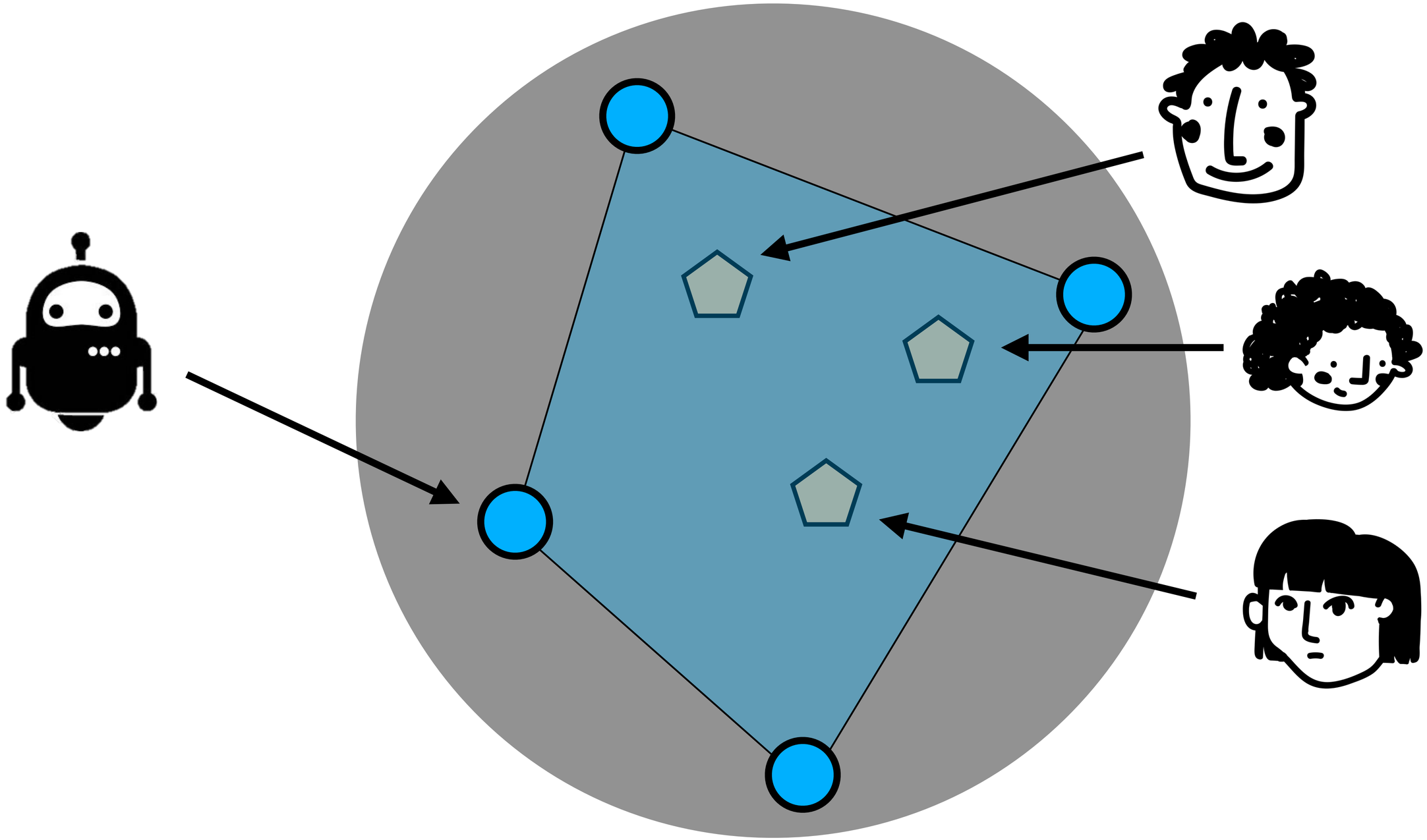
Green is a human player. Blue is confused!

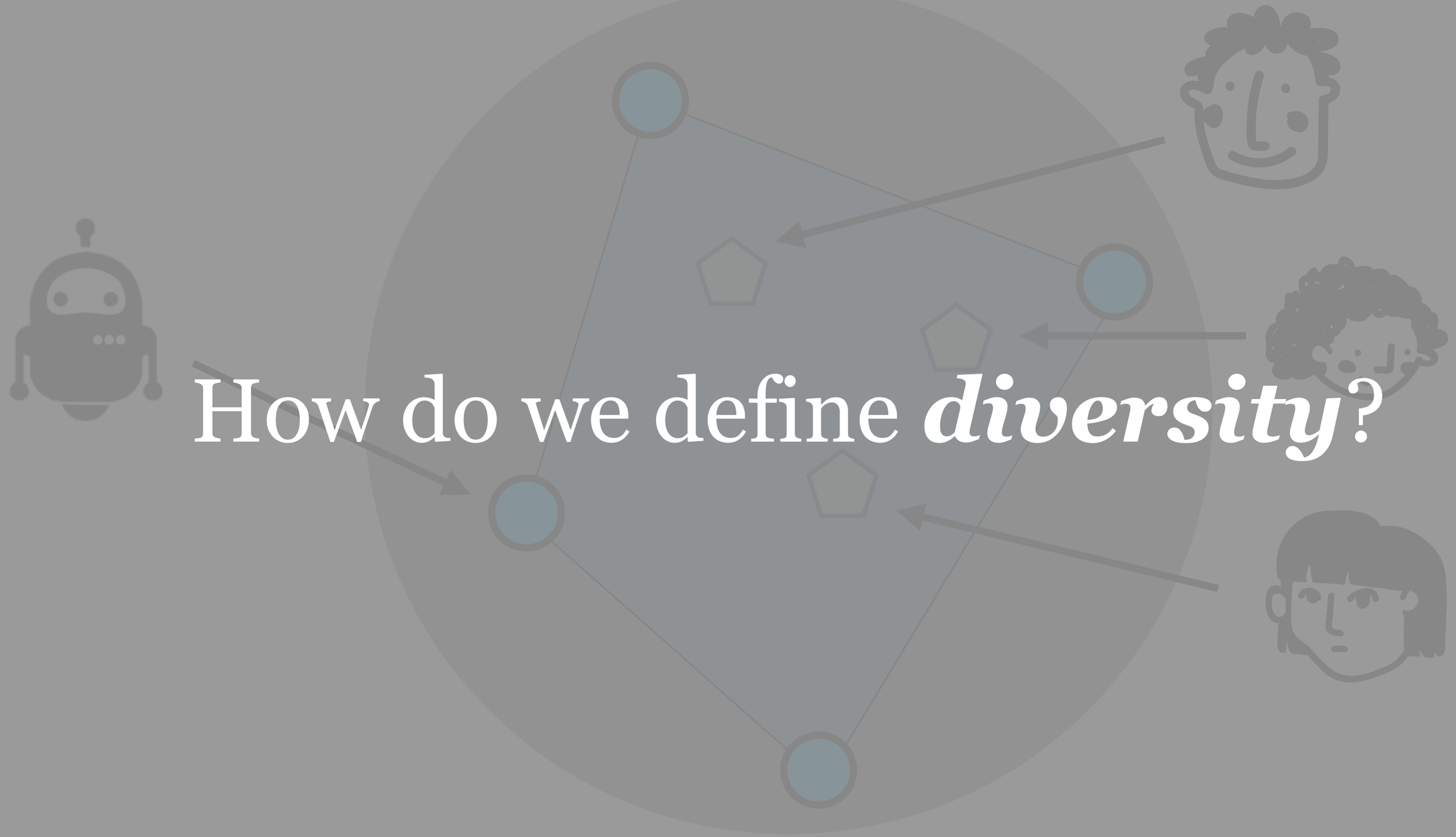












How do we define ***diversity***?

Statistical Diversity

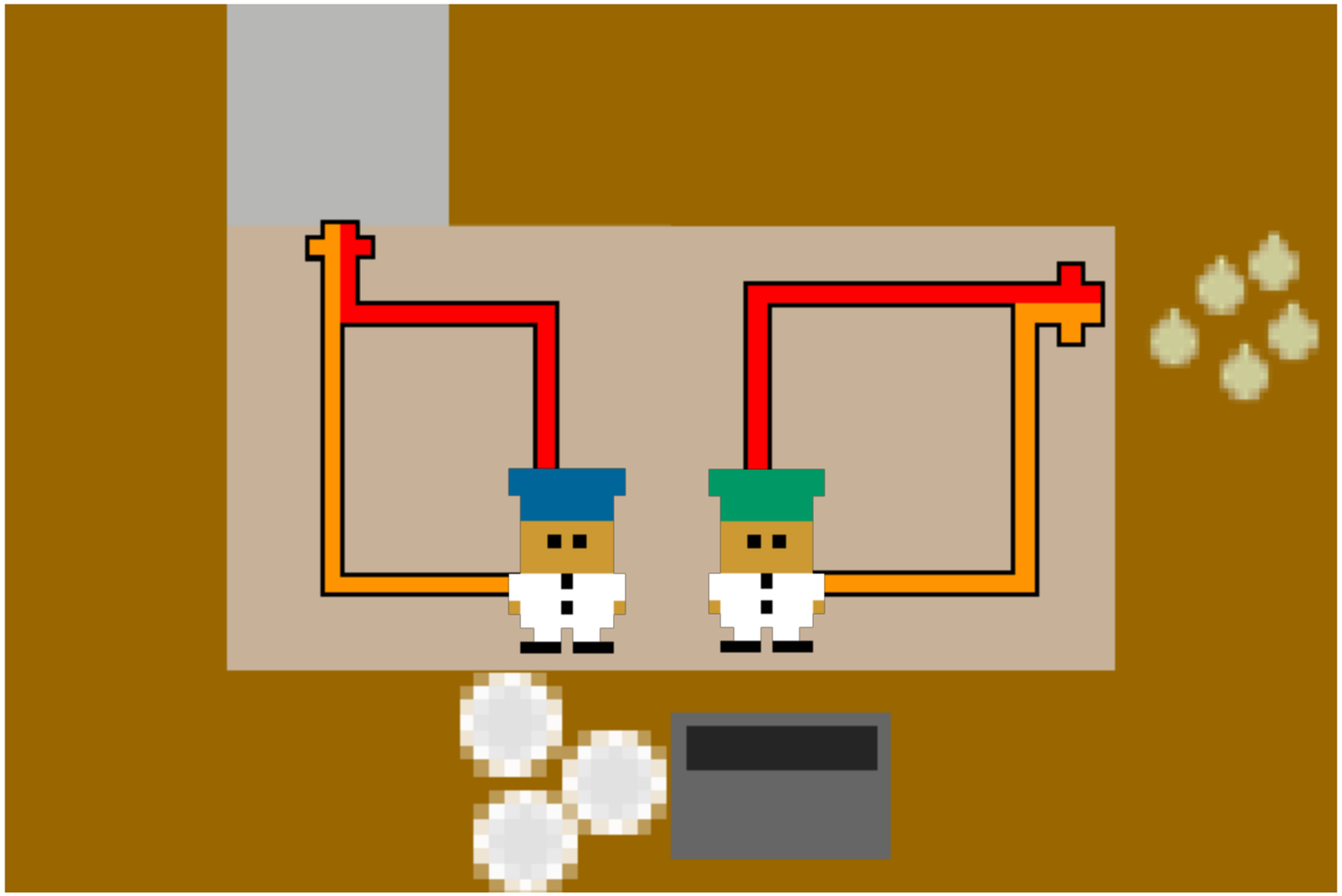
Maximize the difference in trajectories

Statistical Diversity

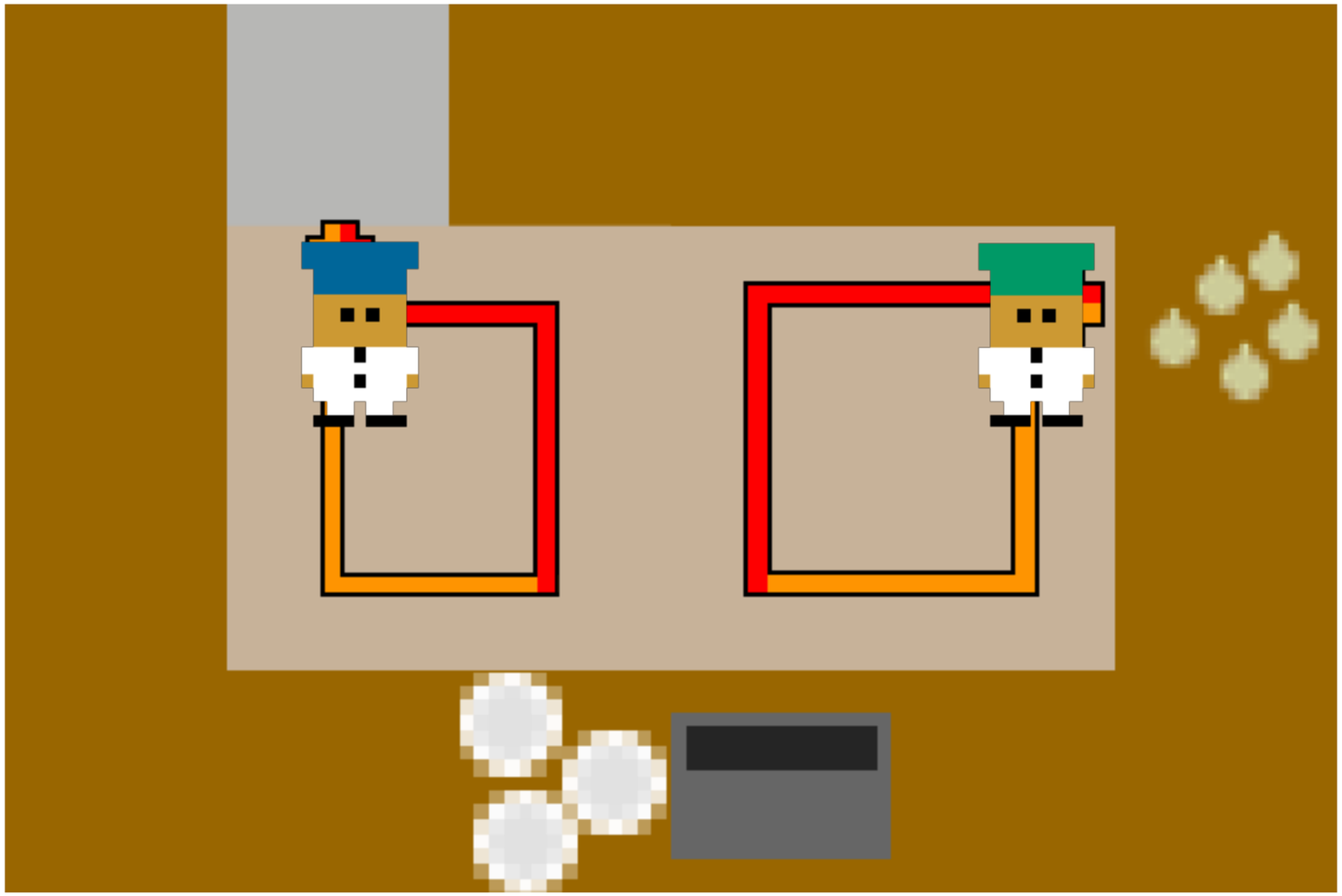
Maximize the difference in trajectories

$$\text{Maximize} \quad C_{ADAP}(D) = \mathop{\mathbb{E}}_{s \in S} \left[\mathop{\mathbb{E}}_{\substack{\pi_1, \pi_2 \in D \\ \pi_1 \neq \pi_2}} \exp(-D_{\mathsf{KL}} \left(\pi_1(s) \quad \pi_2(s) \right) \right]$$

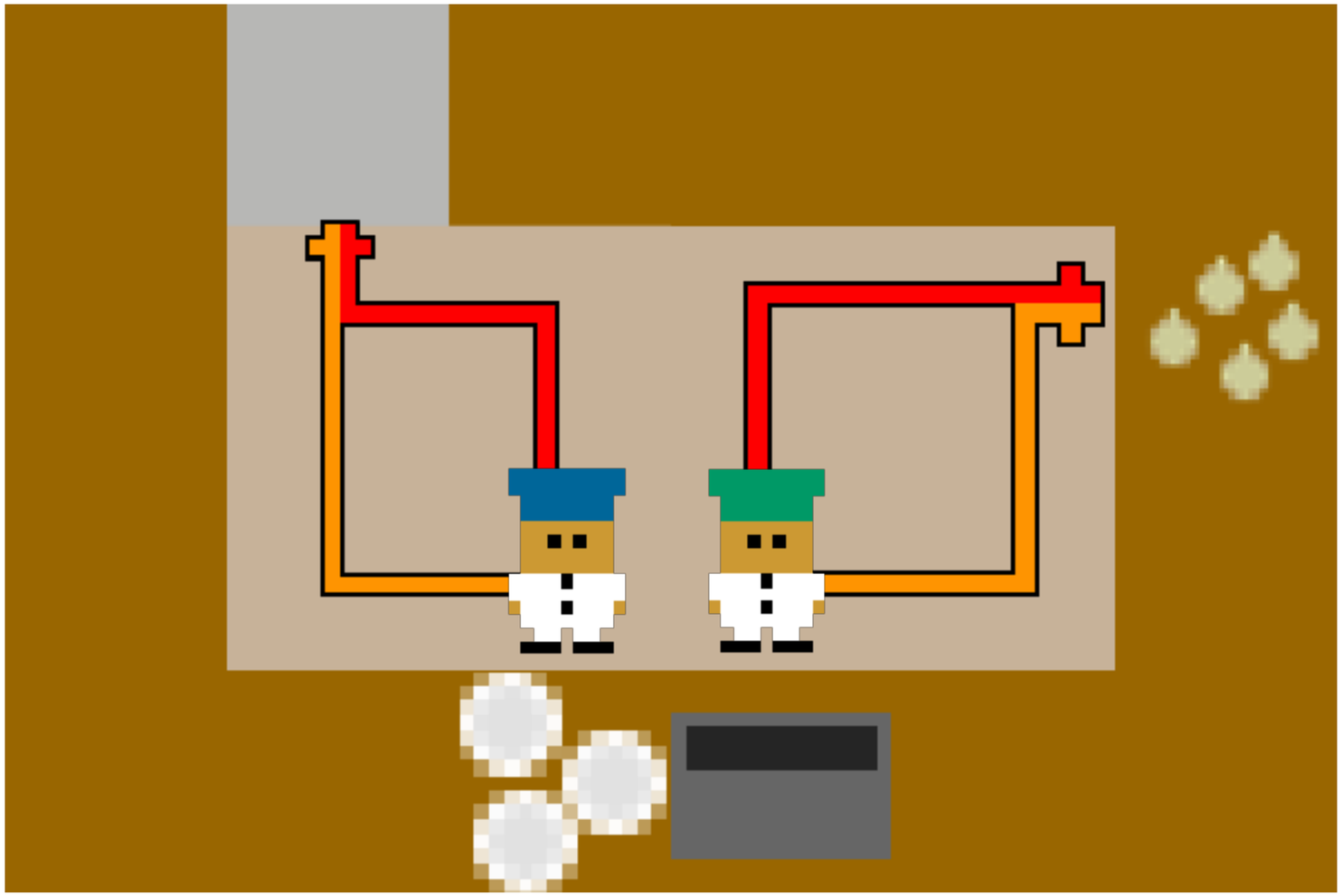
Statistical Diversity



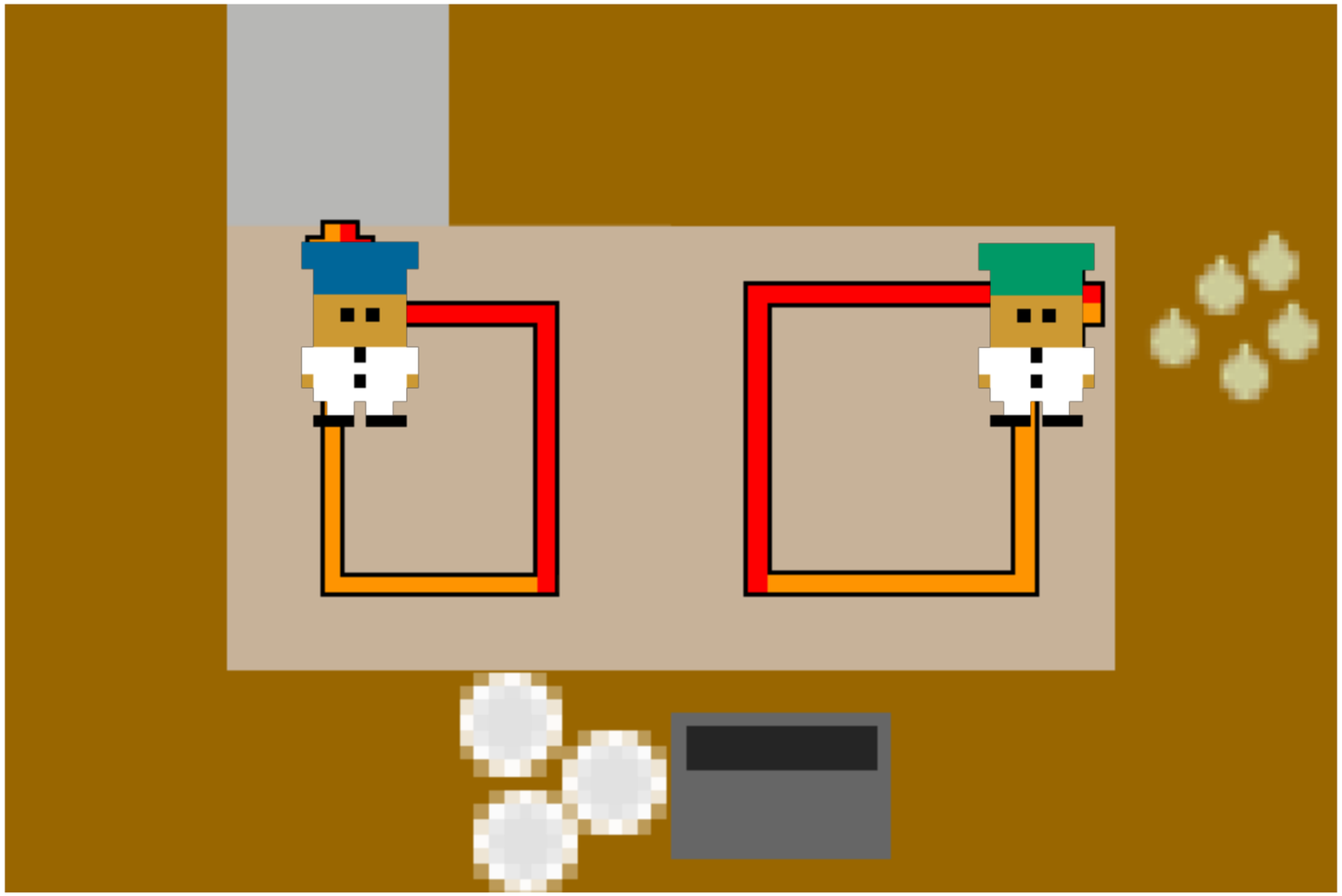
Statistical Diversity



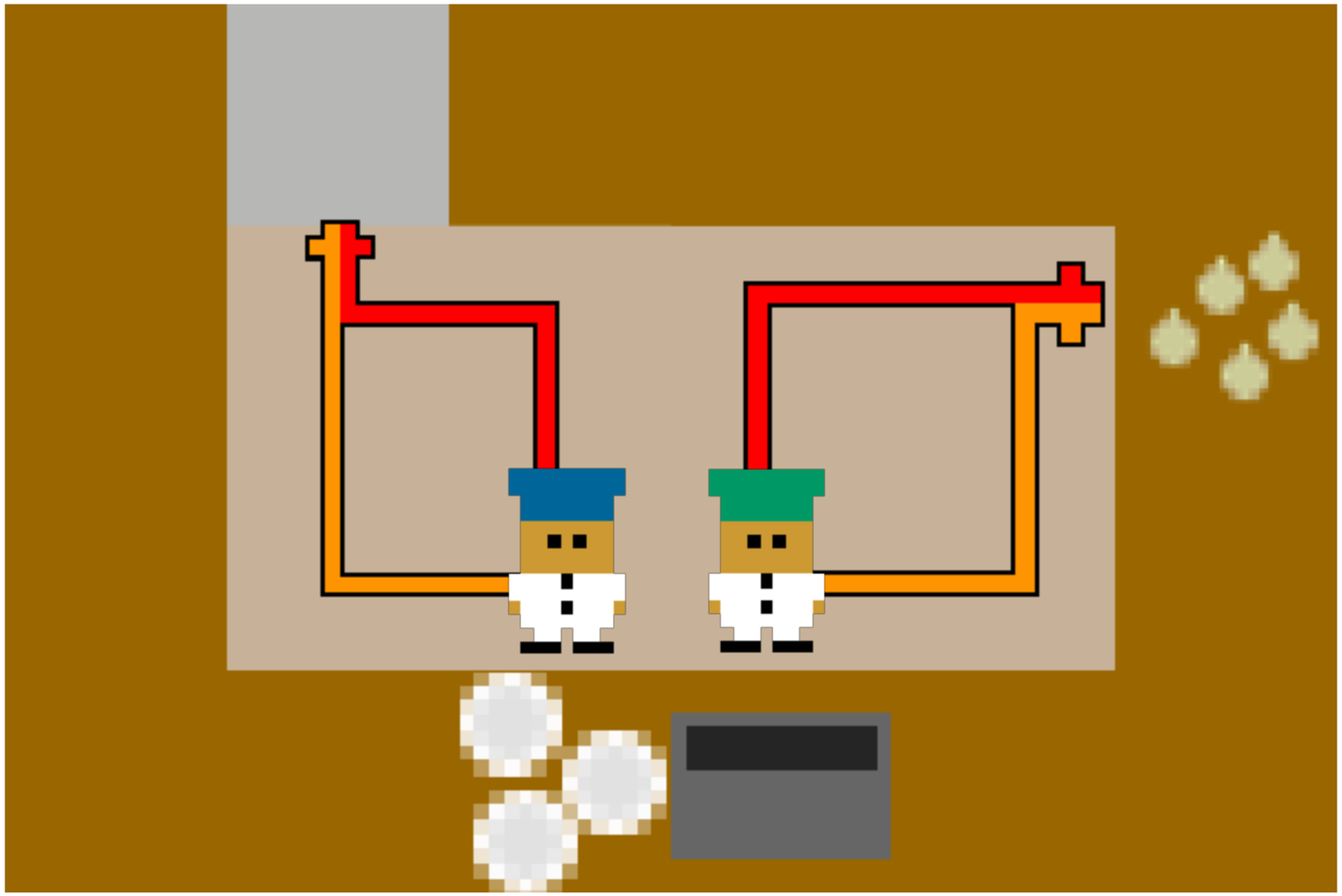
Statistical Diversity



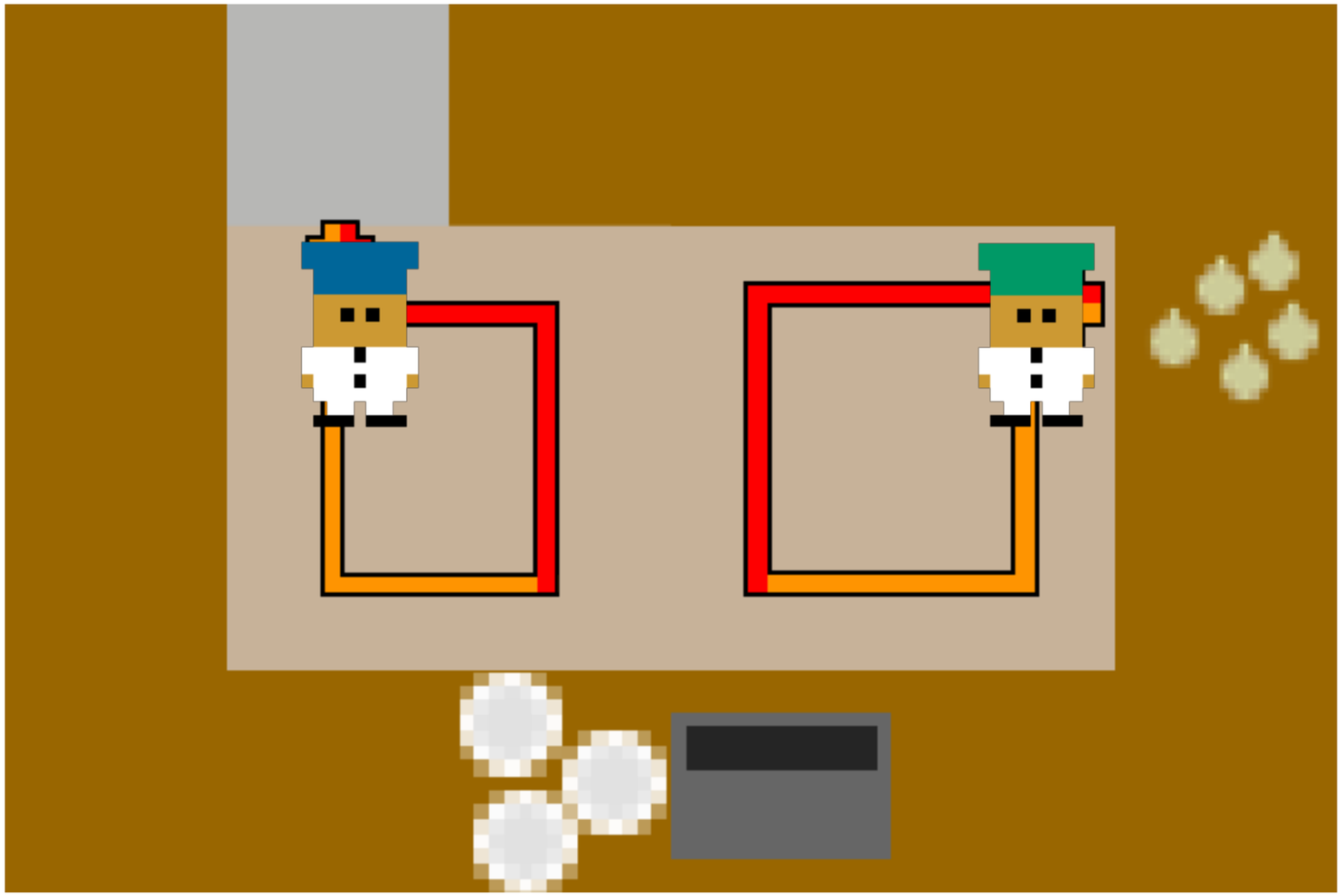
Statistical Diversity



Cross-Play

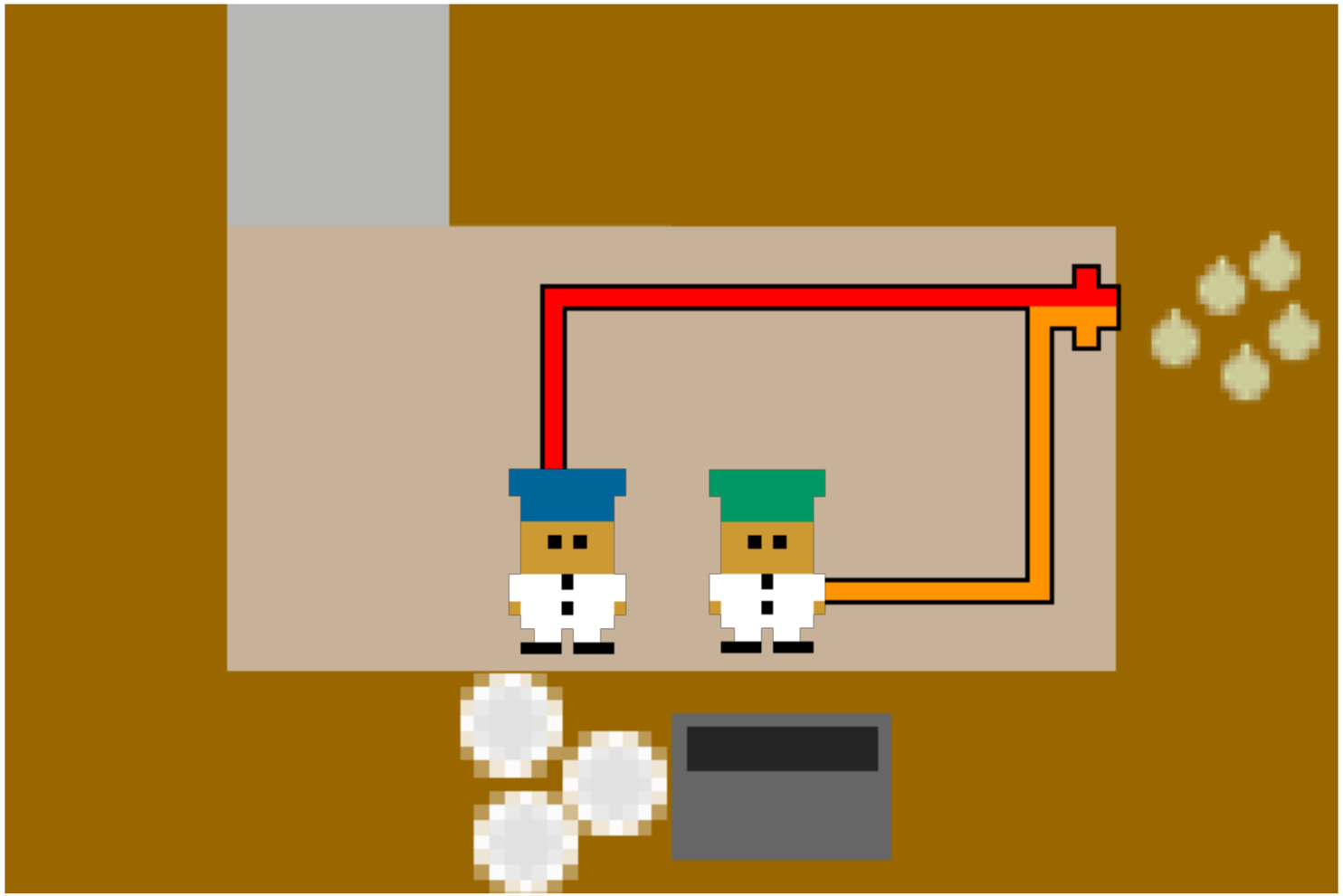


Cross-Play

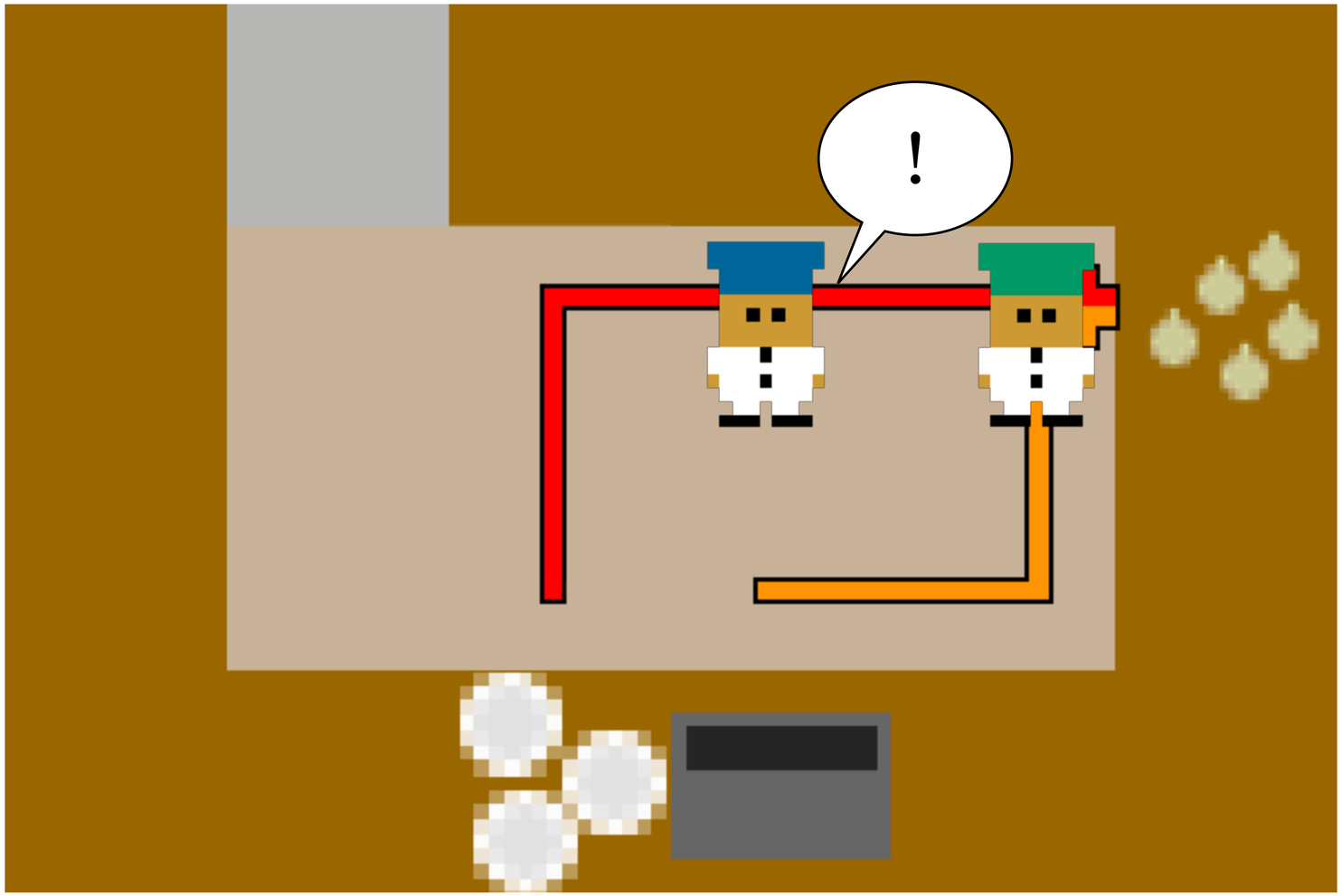


These policies are *compatible*, and are therefore *similar*

Cross-Play



Cross-Play



These policies are *incompatible*, and are therefore *different*

Cross-Play

Minimize the cross-play score between different conventions

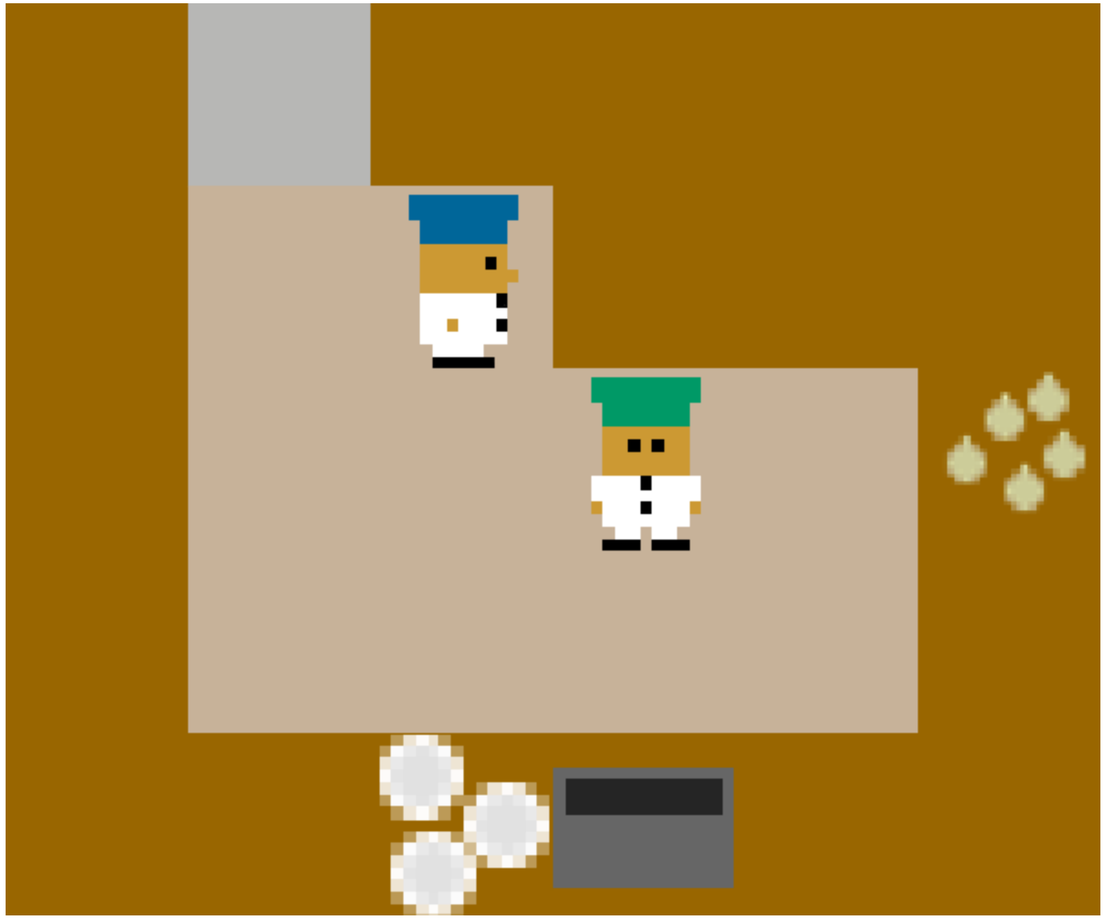
Cross-Play

Minimize the cross-play score between different conventions

$$\textit{Minimize} \quad L(\pi_2) = -J(\pi_2, \pi_2) + J(\pi_2, \pi_1)$$

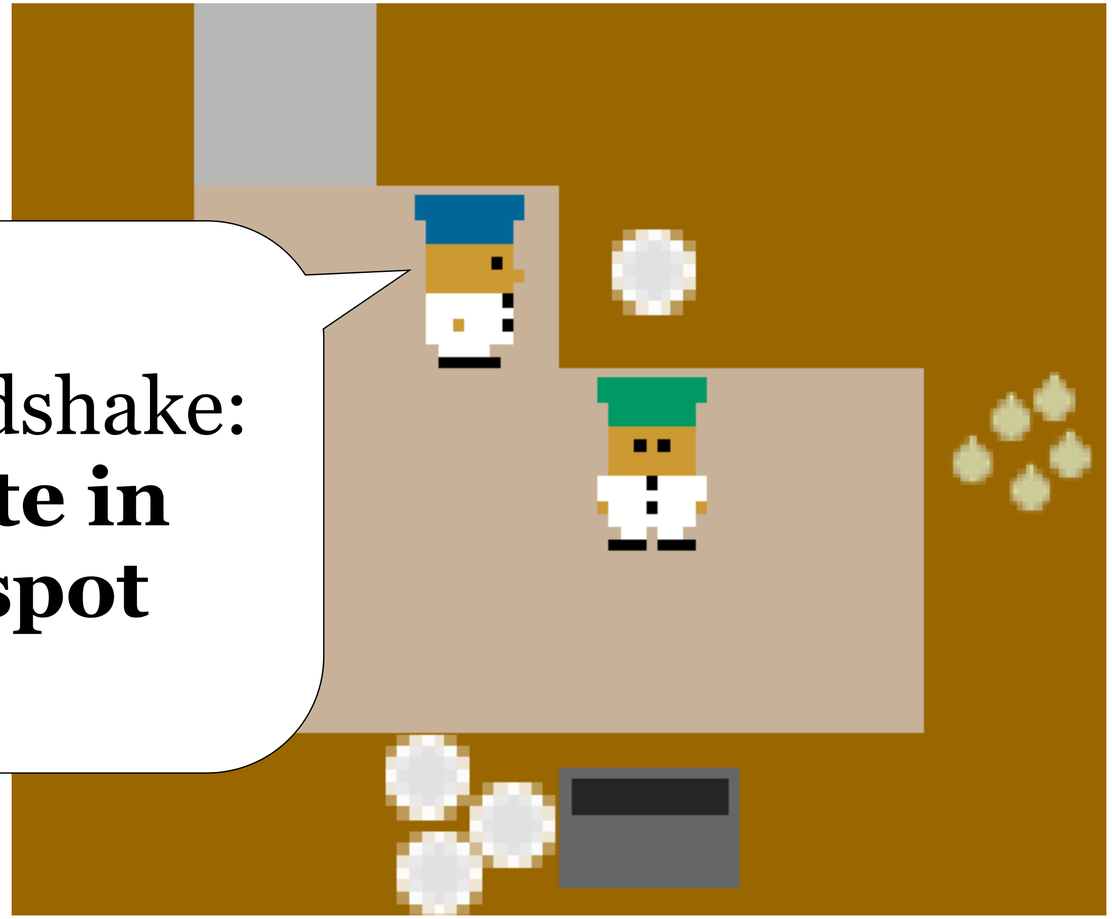
Where J is the expected reward when two policies play together.

Handshakes

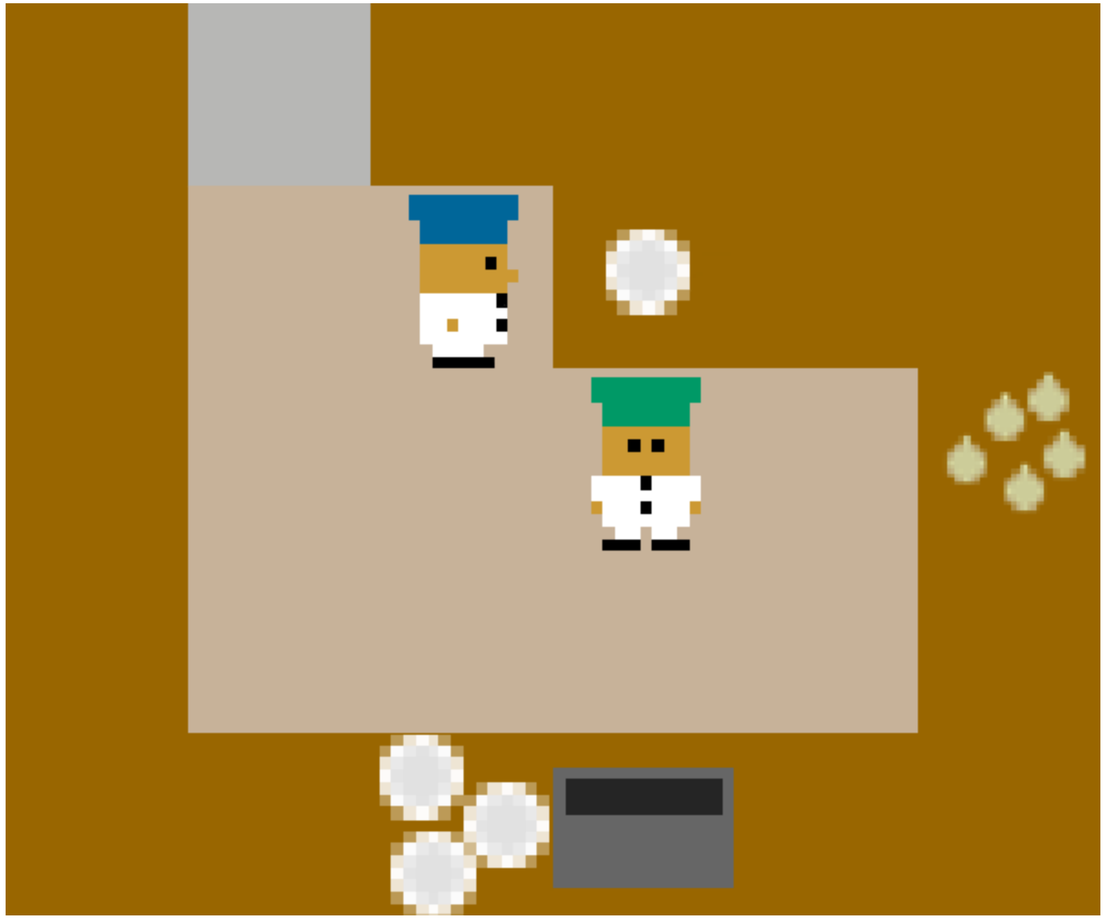


Handshakes

Initiate Handshake:
Place plate in strange spot

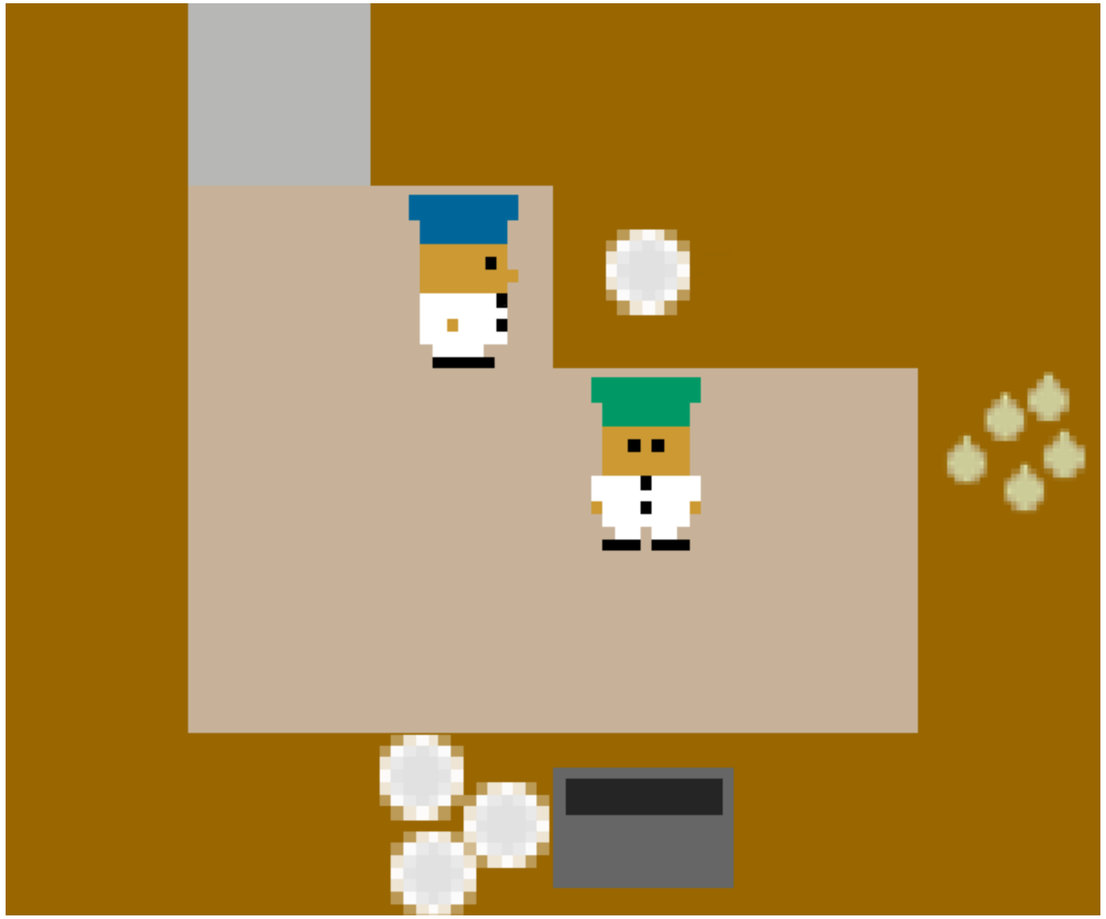


Handshakes



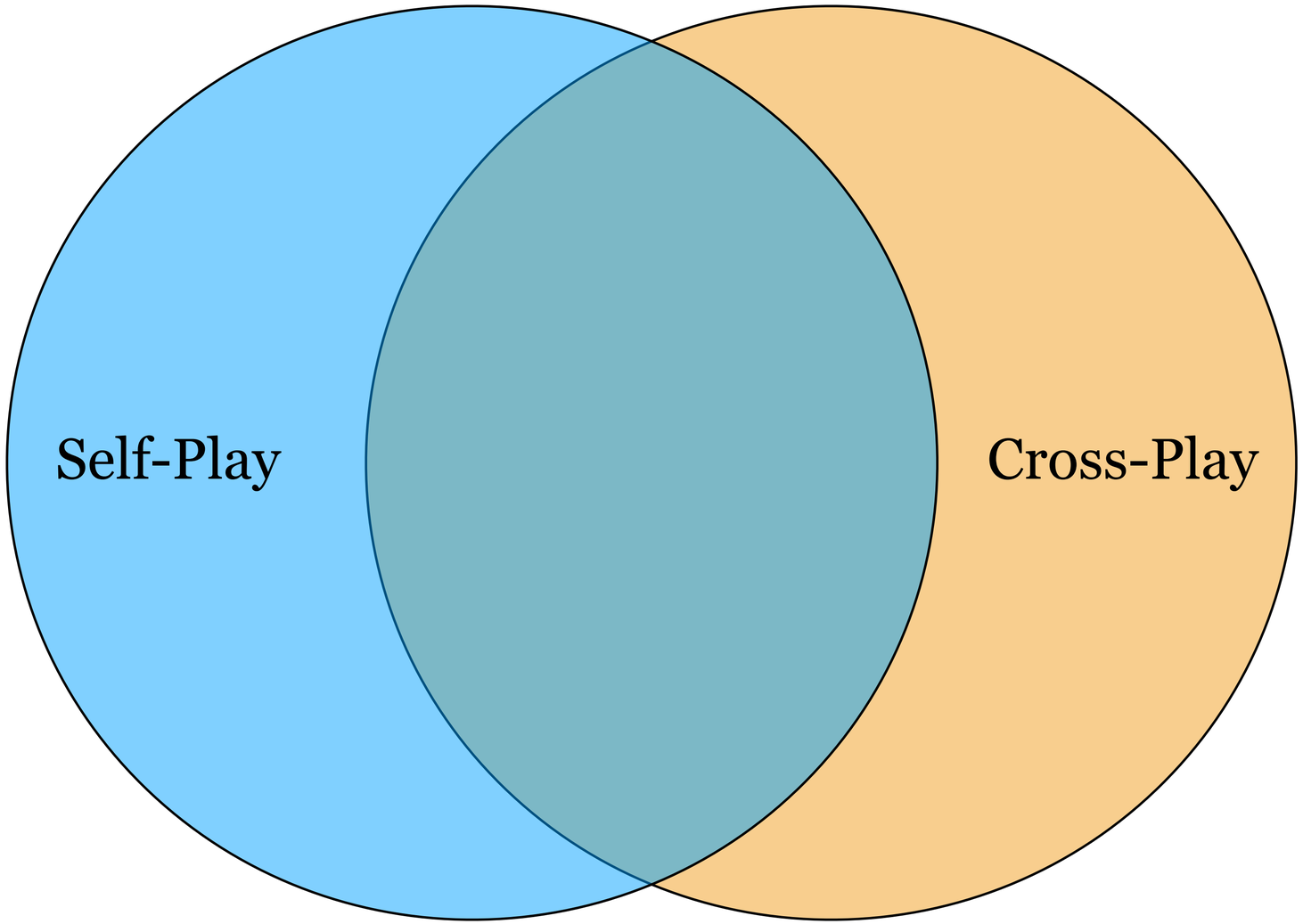
Handshakes

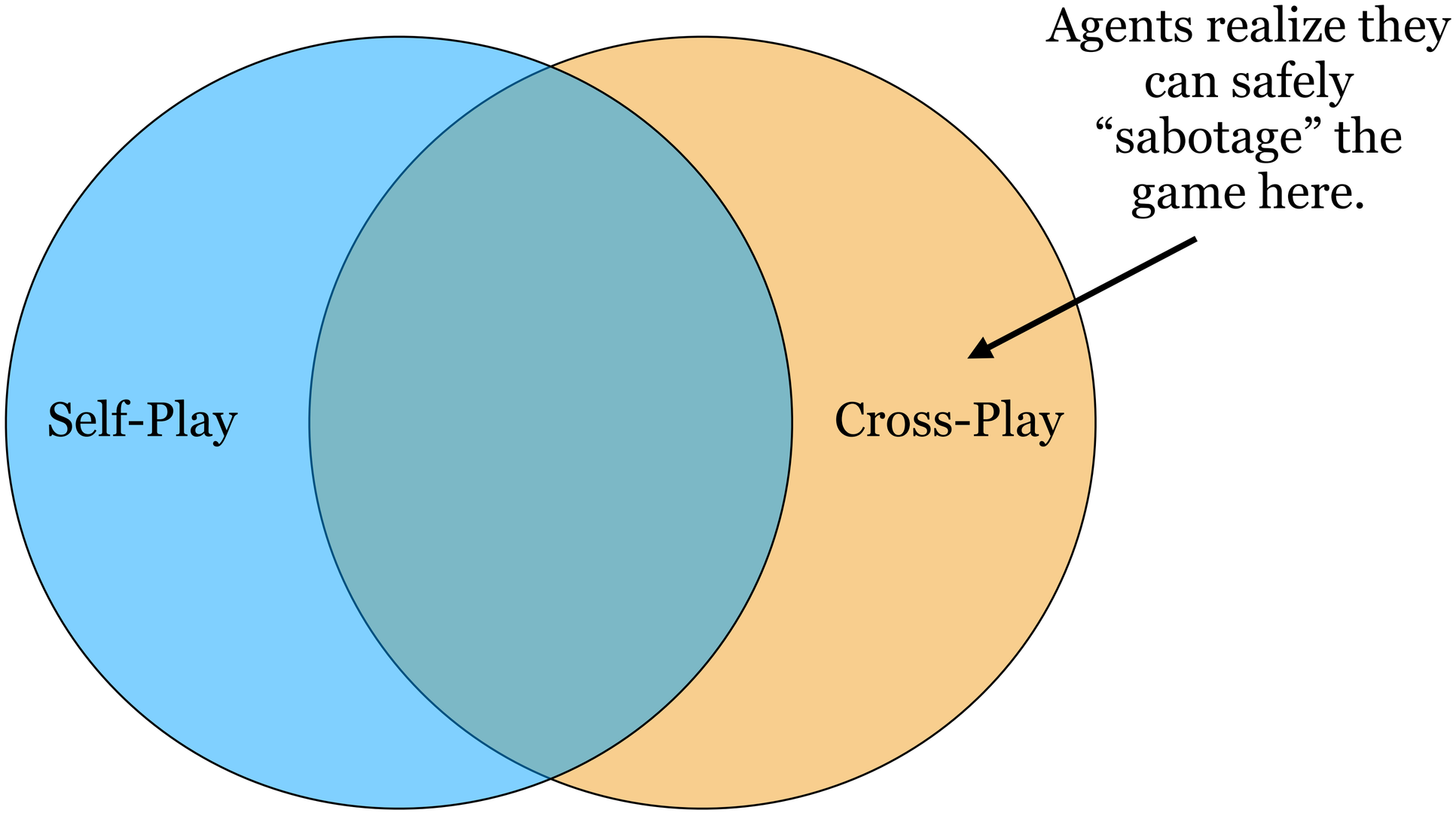
Handshakes

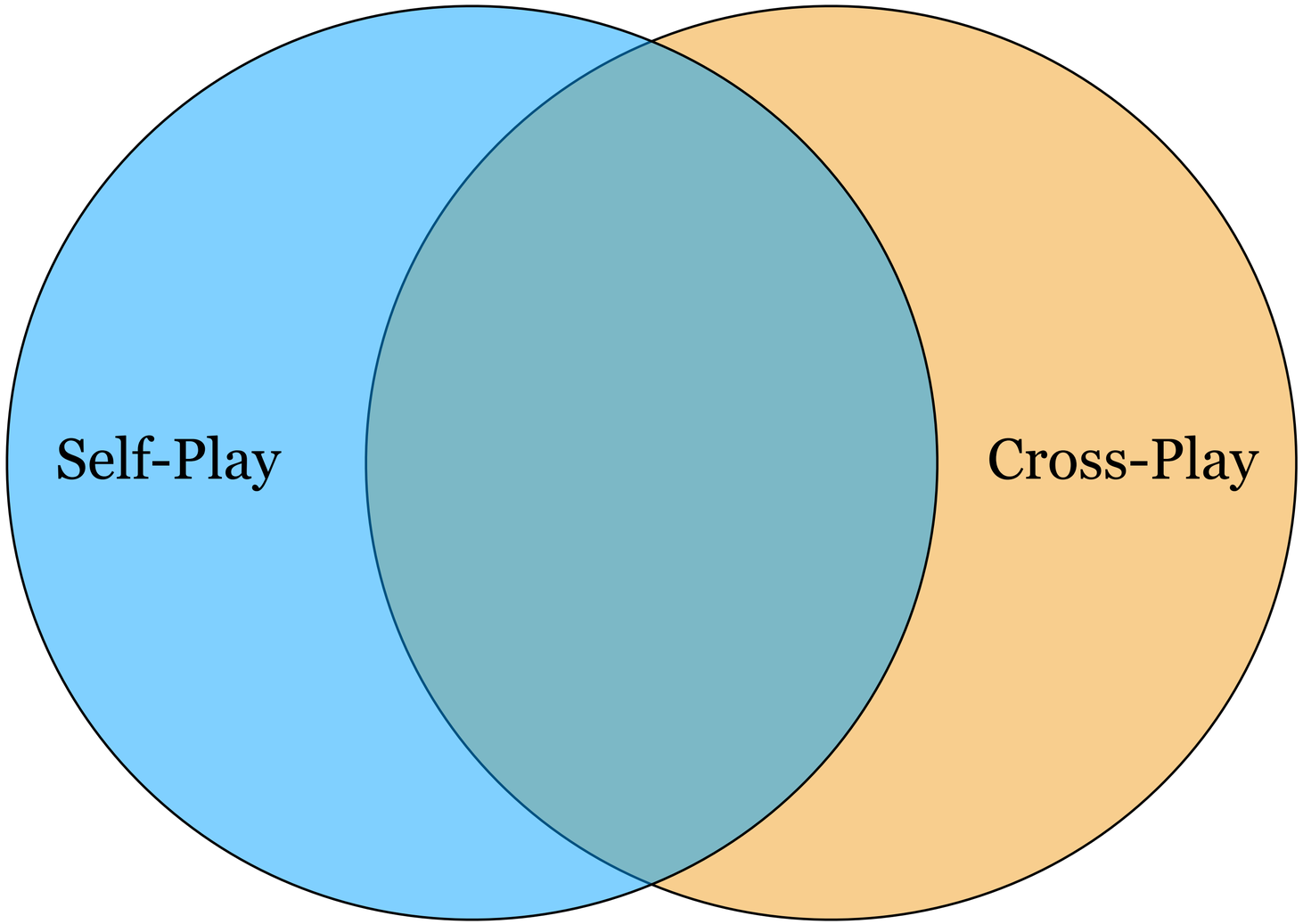


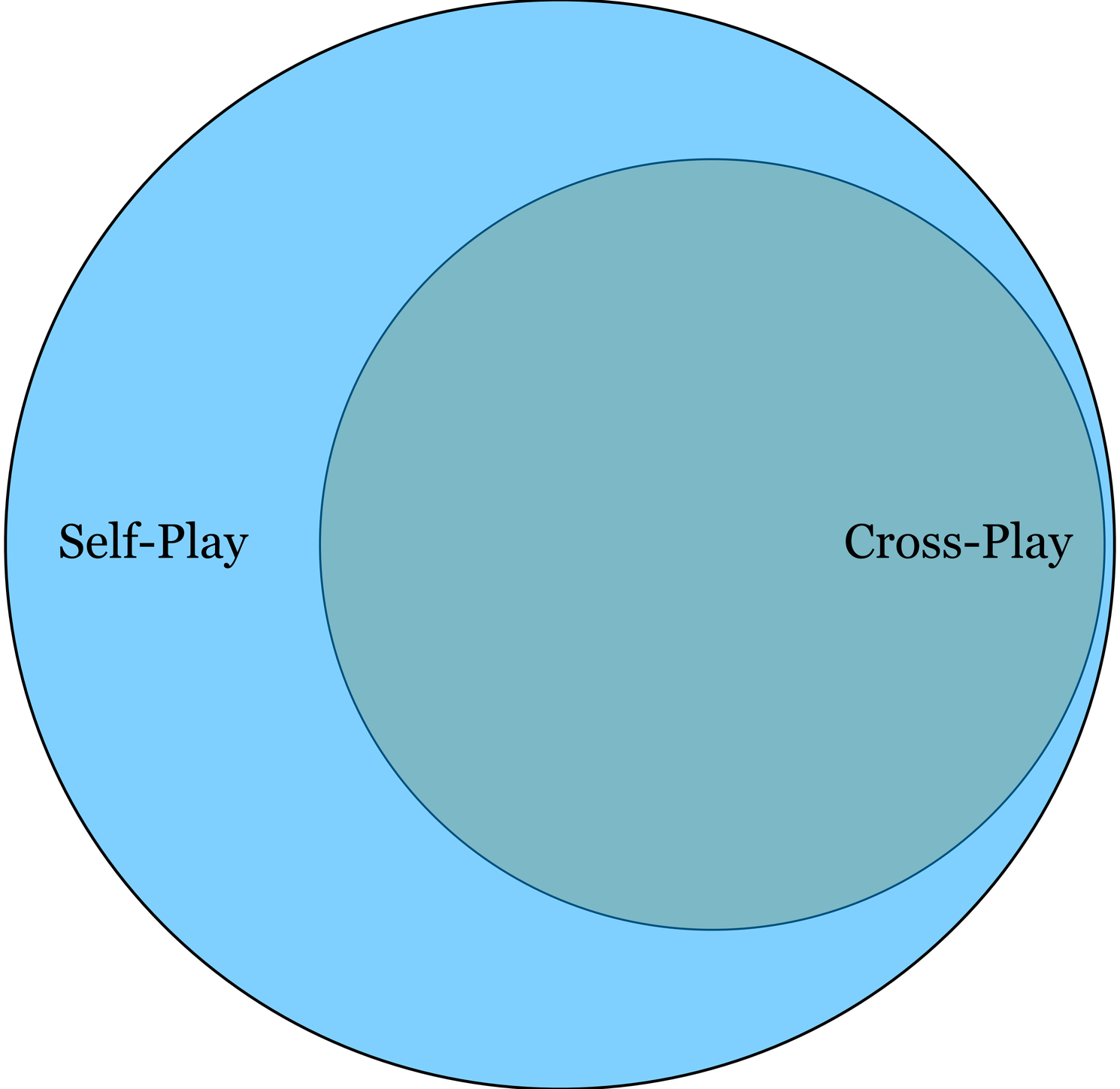
Handshakes

Cross-play incorrectly values this convention!


Self-Play
Cross-Play

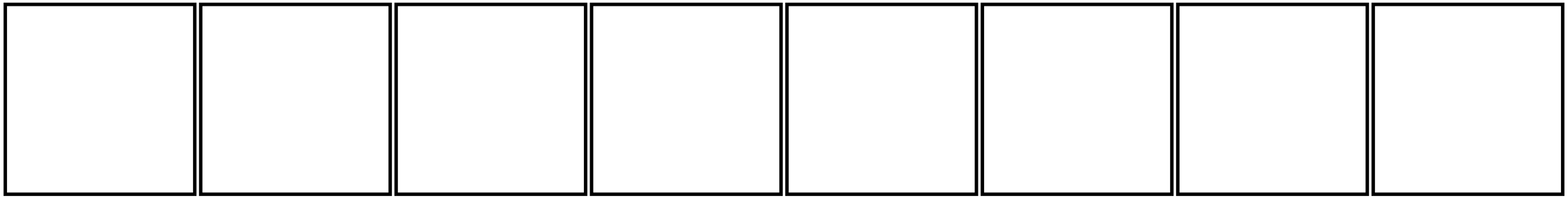

Agents realize they can safely “sabotage” the game here.
Self-Play
Cross-Play


Self-Play
Cross-Play


Self-Play
Cross-Play

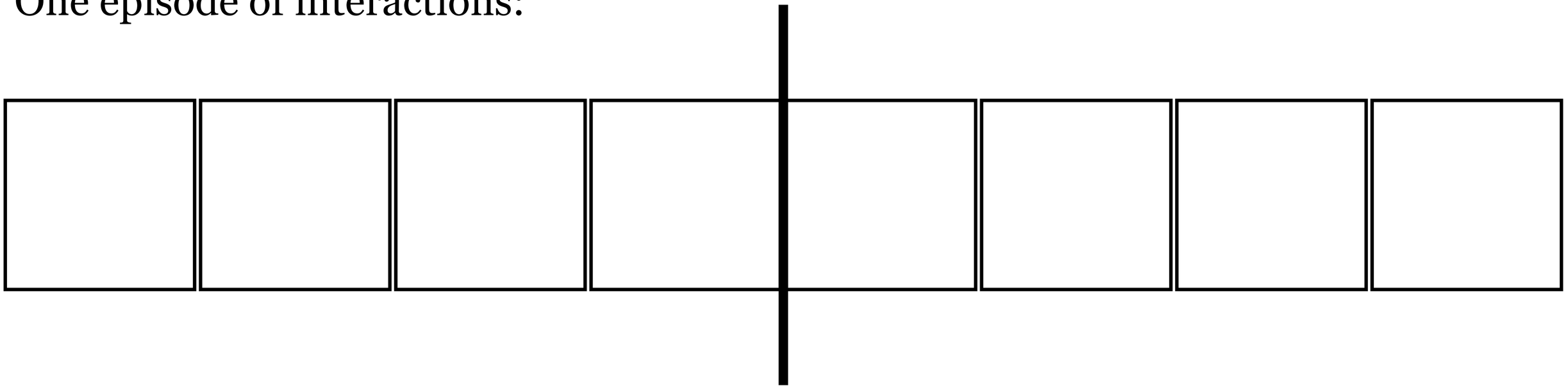
Mixed-Play

One episode of interactions:



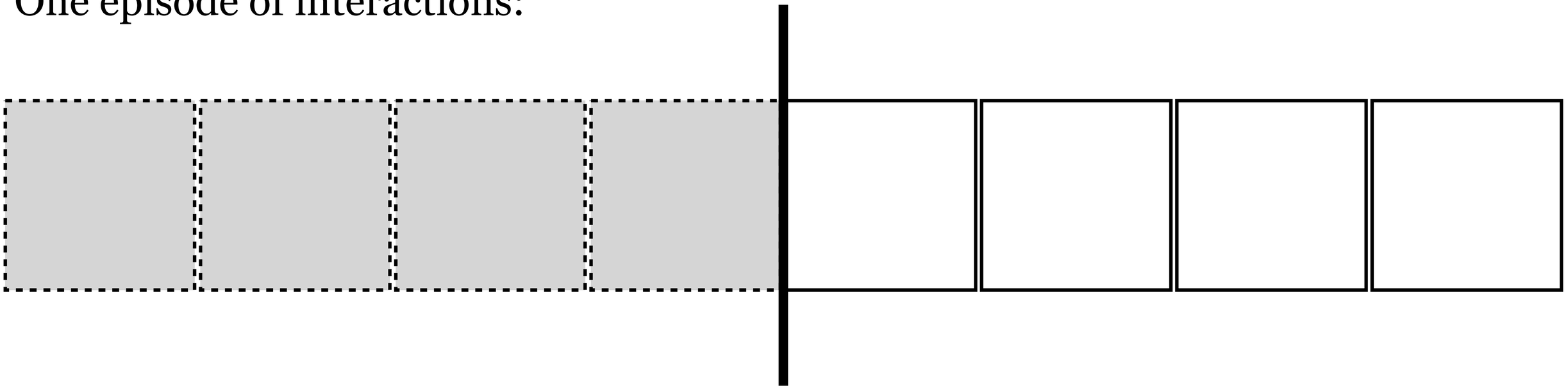
Mixed-Play

One episode of interactions:



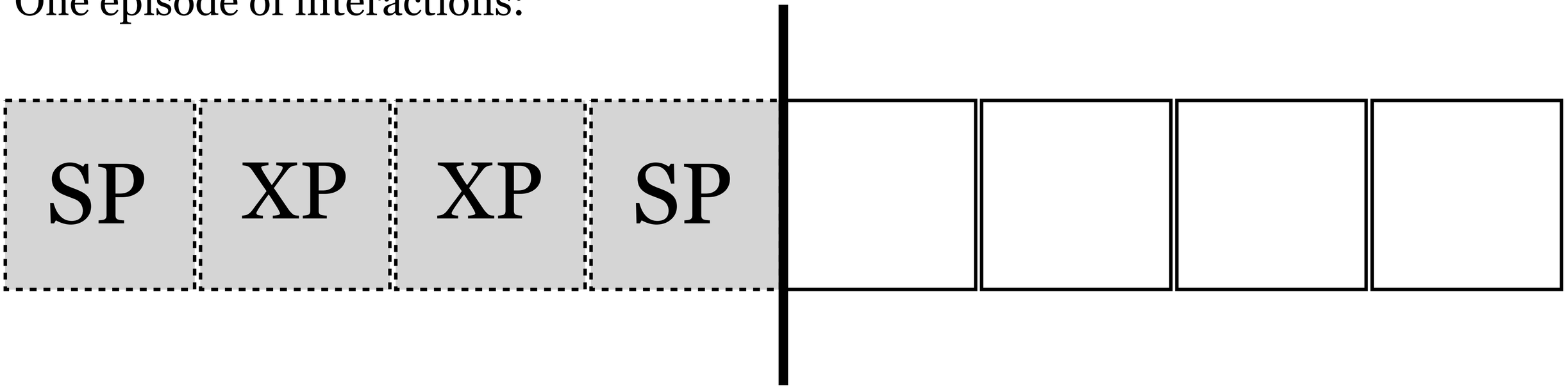
Mixed-Play

One episode of interactions:



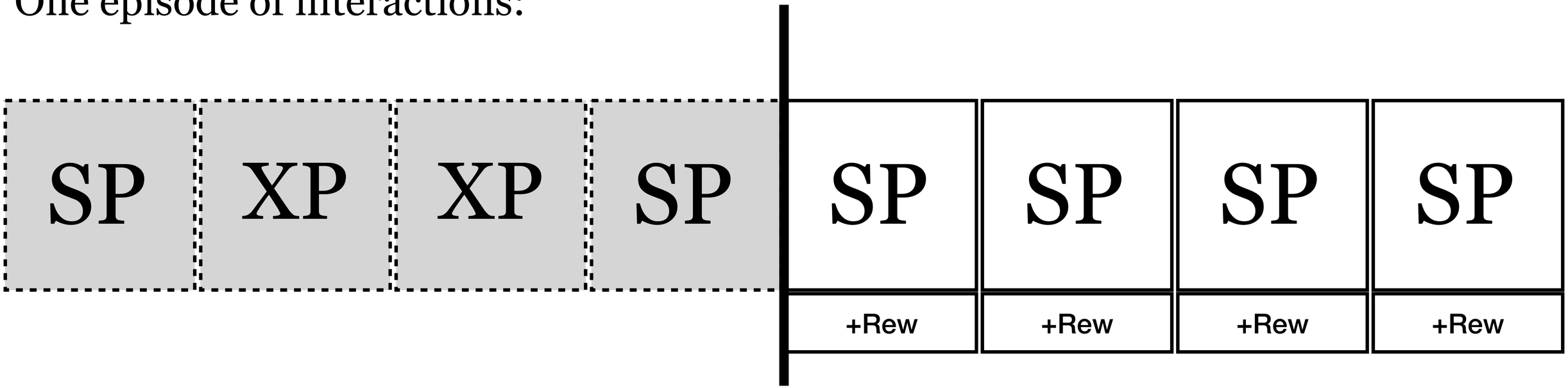
Mixed-Play

One episode of interactions:

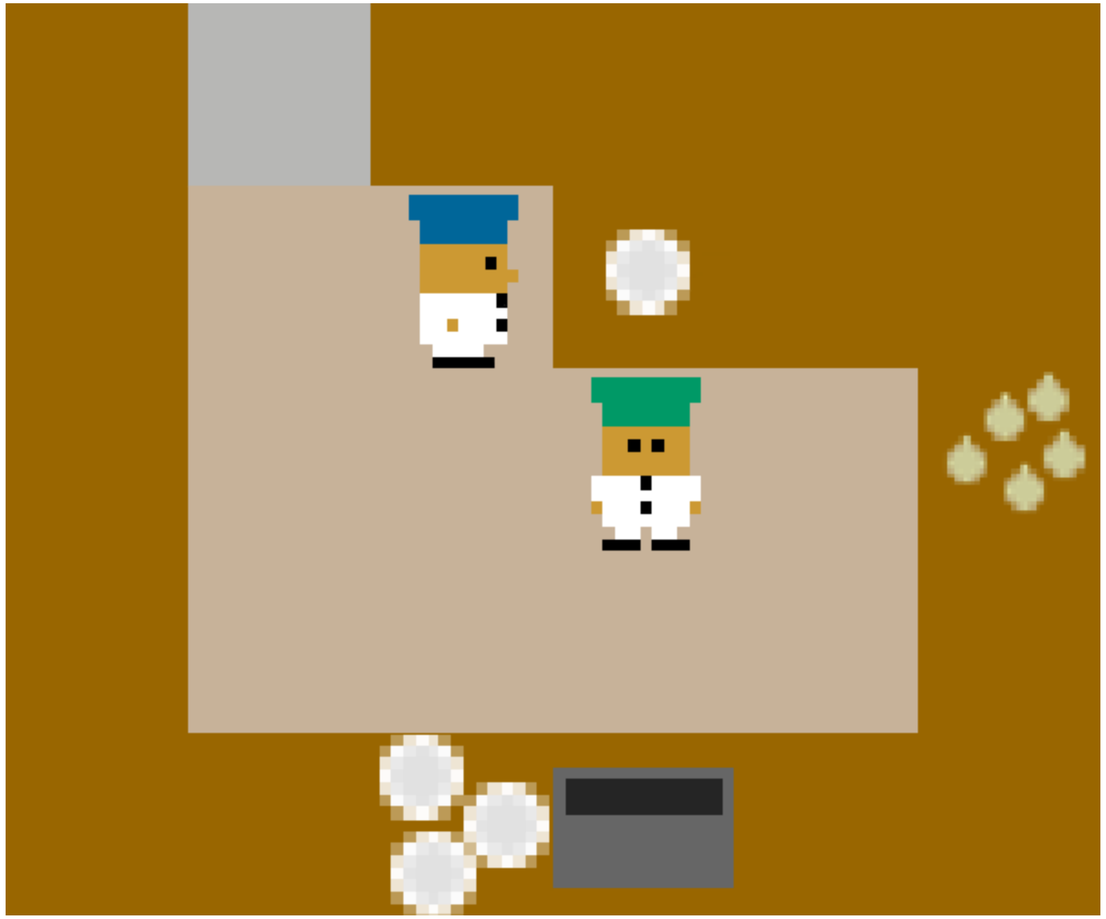


Mixed-Play

One episode of interactions:



Mixed-Play



Mixed-Play

Mixed-Play fixes handshakes!

CoMeDi

Minimize loss function to generate the n^{th} convention:

$$L(\pi_n) = -J(\pi_n, \pi_n) + \alpha J(\pi_n, \pi^*) - \beta J_{\mathsf{M}}(\pi_n, \pi^*)$$

π^* is the most-compatible previously discovered convention

J_M is the expected return from the self-play phase of mixed-play

CoMeDi

Minimize loss function to generate the n^{th} convention:

$$L(\pi_n) = -J(\pi_n, \pi_n) + \alpha J(\pi_n, \pi^*) - \beta J_{\mathsf{M}}(\pi_n, \pi^*)$$

Maximize Self-Play

π^* is the most-compatible previously discovered convention

J_M is the expected return from the self-play phase of mixed-play

CoMeDi

Minimize loss function to generate the n^{th} convention:

$$L(\pi_n) = -J(\pi_n, \pi_n) + \alpha J(\pi_n, \pi^*) - \beta J_{\mathsf{M}}(\pi_n, \pi^*)$$

Maximize Self-Play

Minimize Cross-Play

π^* is the most-compatible previously discovered convention

J_M is the expected return from the self-play phase of mixed-play

CoMeDi

Minimize loss function to generate the n^{th} convention:

$$L(\pi_n) = -J(\pi_n, \pi_n) + \alpha J(\pi_n, \pi^*) - \beta J_{\mathsf{M}}(\pi_n, \pi^*)$$

Maximize Self-Play

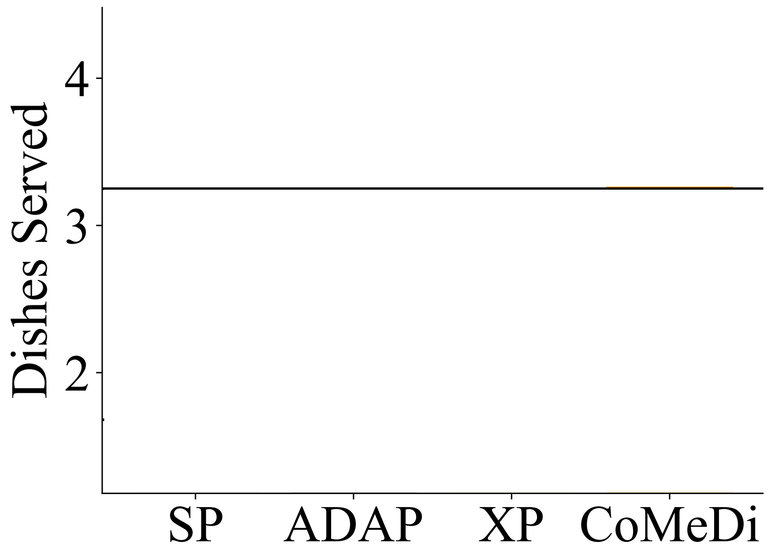
Minimize Cross-Play

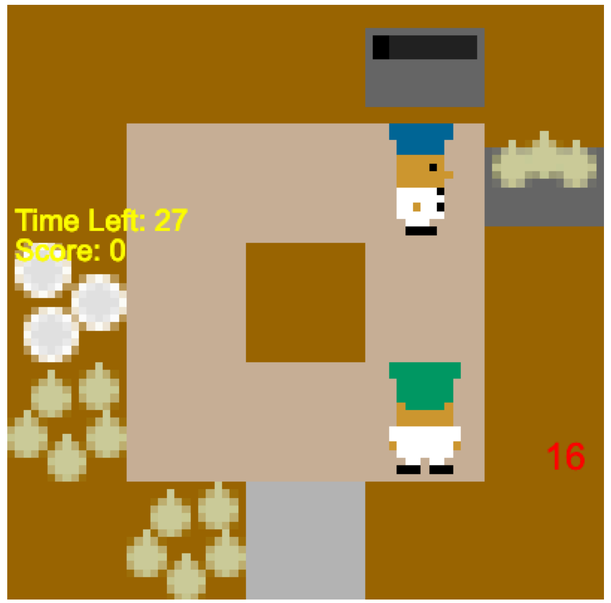
Maximize Mixed-Play

π^* is the most-compatible previously discovered convention

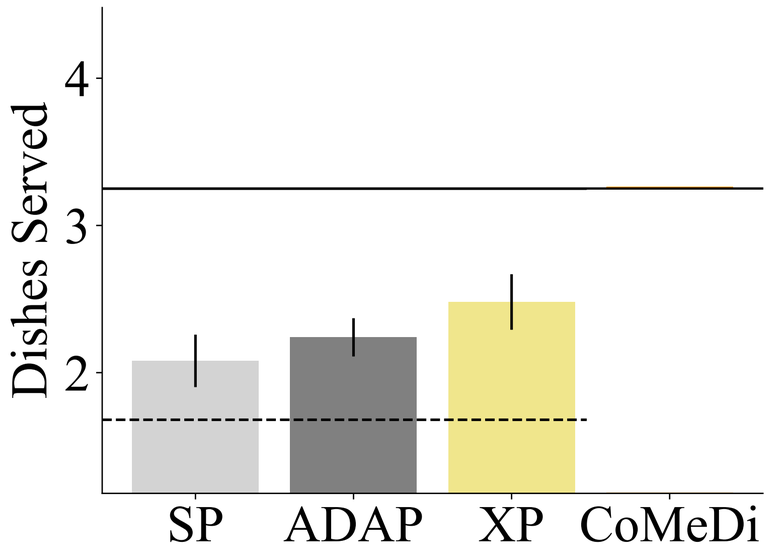
J_M is the expected return from the self-play phase of mixed-play

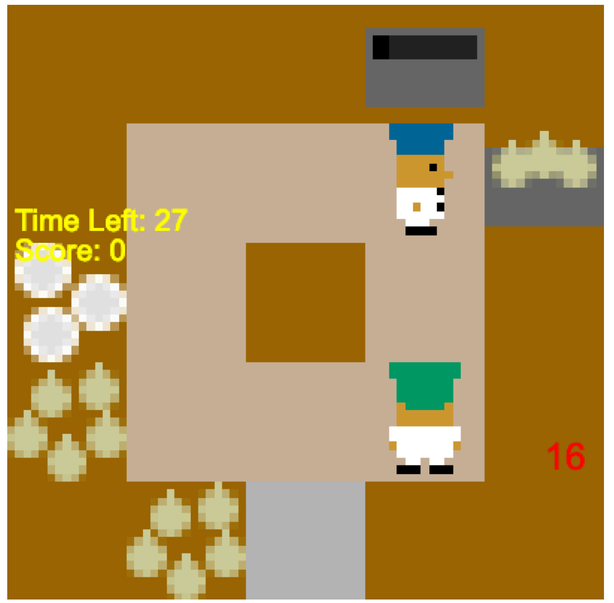
User Study



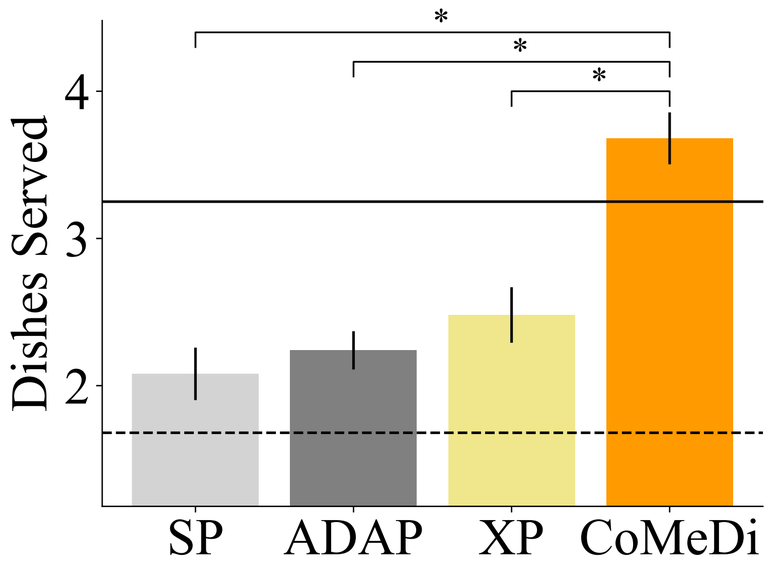


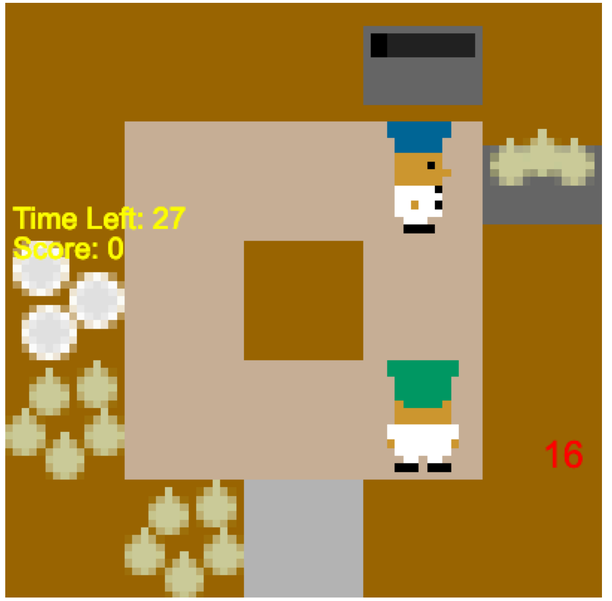
User Study



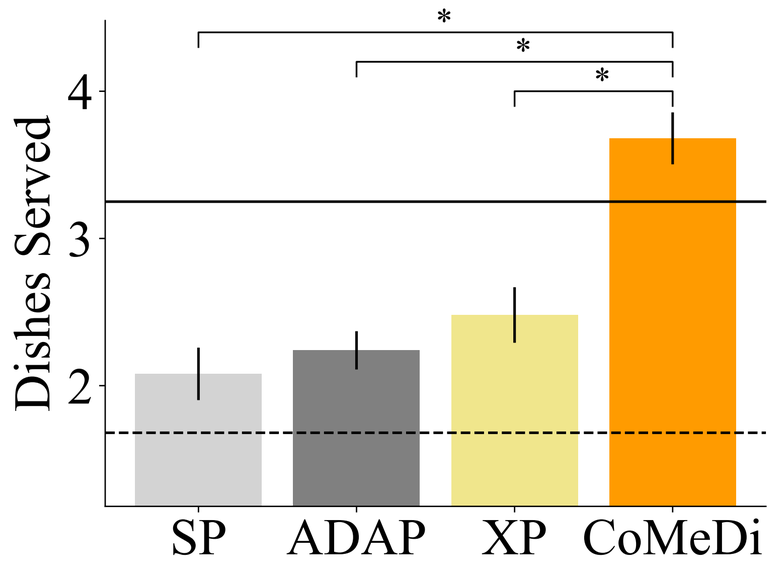


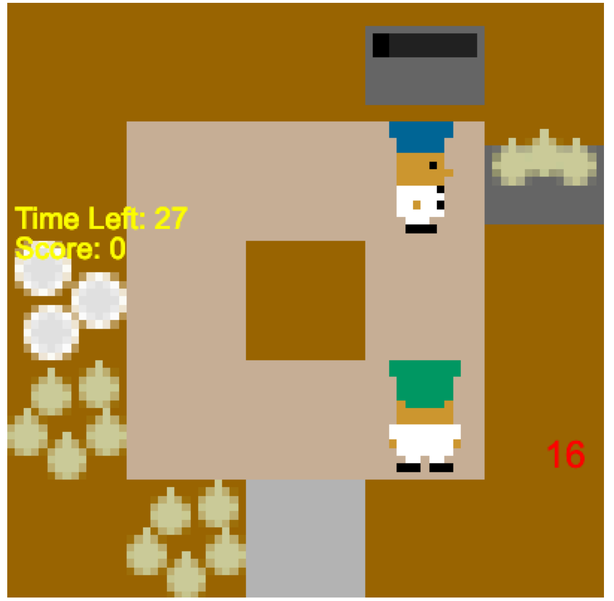
User Study

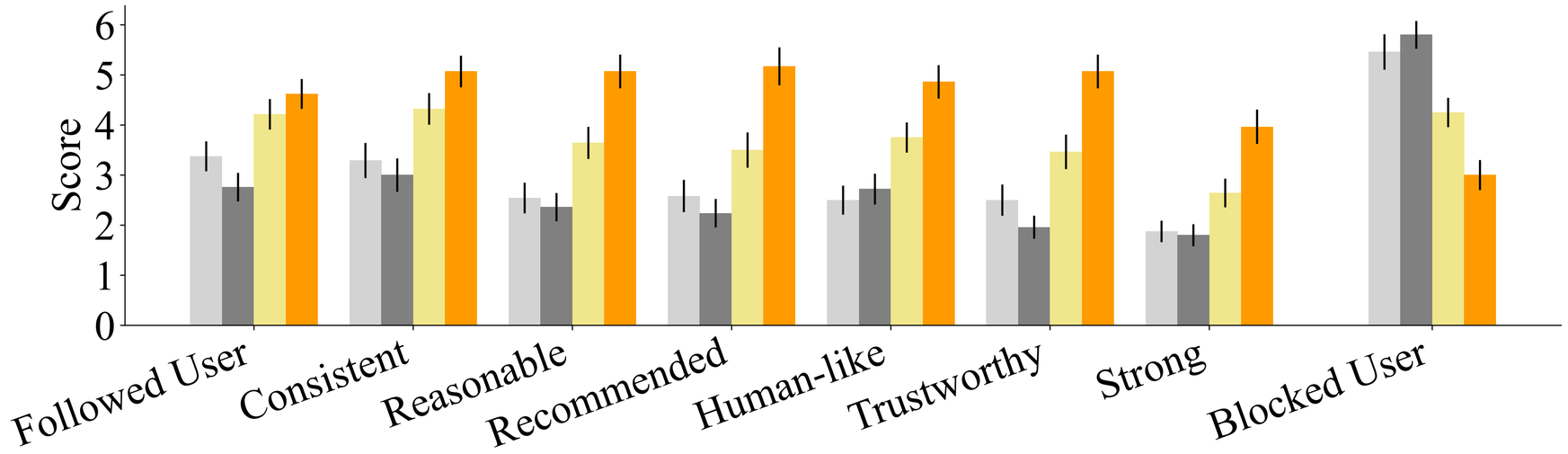




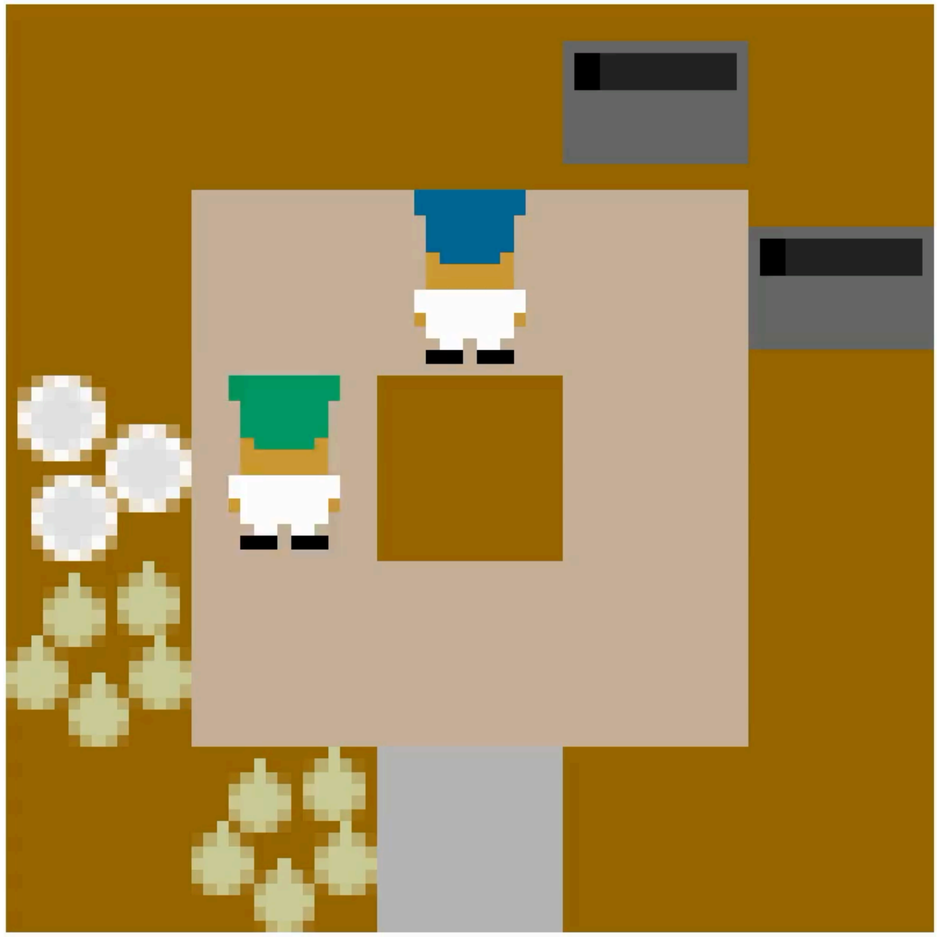
User Study

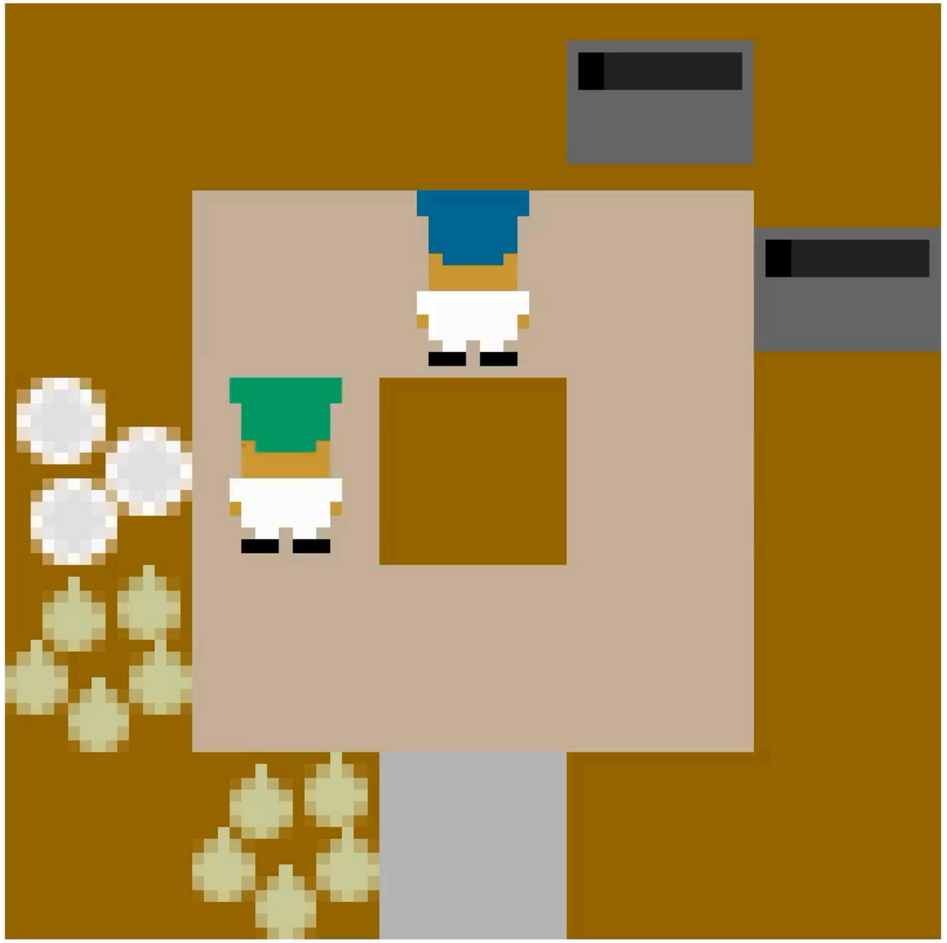

*
*
*
Dishes Served
4
3
2
SP
ADAP
XP
CoMeDi


Time Left: 27
Score: 0
16

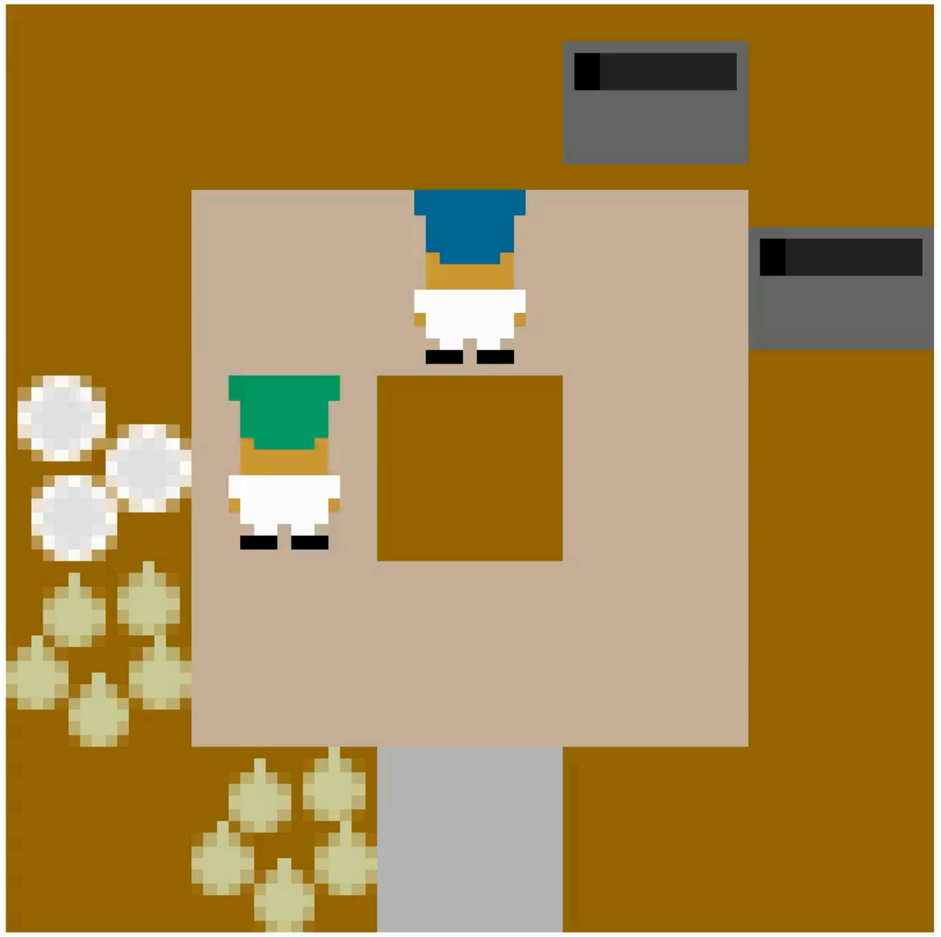

Score
6
5
4
3
2
1
0
Followed User
Consistent
Reasonable
Recommended
Human-like
Trustworthy
Strong
Blocked User

User Study

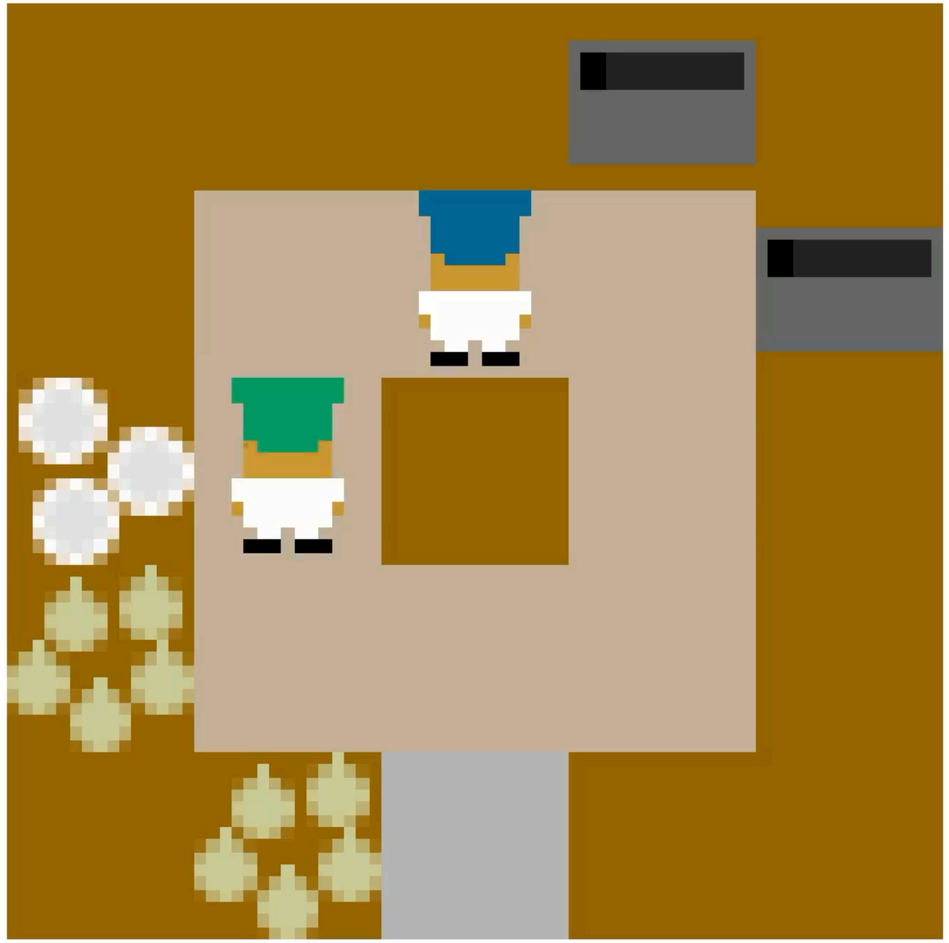




User Study



CoMeDi



Self-Play

CoMeDi enables us to train a diverse set of conventions, which can be used to create strong convention-aware agents for human-AI collaboration.