On the Critical Role of Conventions in Adaptive Human-AI Collaboration

Andy Shih, Arjun Sawhney, Jovana Kondic, Stefano Ermon, Dorsa Sadigh
Multiagent collaboration requires knowing 2 things:

1. How to be good at the task
2. How to coordinate with partners
Goal
Collaborative Block Placing
Collaborative Block Placing
Collaborative Block Placing
Collaborative Block Placing
Collaborative Block Placing
Collaborative Block Placing
For Player 1

**Rule:** Placing red block correctly

**Convention:** Signaling location of blue block
Multiagent collaboration requires knowing 2 things:

1. How to be good at the task
2. How to coordinate with partners
Conventions: Shared representation that breaks symmetry between equally optimal strategies (developed over repeated interactions)
Generalization: New Partner (Fixed Task)
Generalization: New Partner (Fixed Task)
Generalization: New Partner (Fixed Task)
Generalization: New Task (Fixed Partner)
Generalization: New Task (Fixed Partner)
Generalization: New Task (Fixed Partner)
We need a modular architecture to work with new combinations of partners and tasks.
Modular Architecture

State $s$ → task module $g^t(a|z)$
Modular Architecture

State $s$

Partner module $g^p_1$

Task module $g^t$

$g^p_1(a|z)$

$g^t(a|z)$
Multiply the action distributions

$$
\pi_1(a|s) = \frac{1}{Z_1} g^t(a|z) g^p_1(a|z)
$$
Modular Architecture

State $s$

task module

$g^t$

$z$

partner module

$g_1^p$

$g_n^p$

Multiply the action distributions

$$\pi_1(a|s) = \frac{1}{Z_1} g^t(a|z) g_1^p(a|z)$$

$$\pi_n(a|s) = \frac{1}{Z_n} g^t(a|z) g_n^p(a|z)$$
Modular Architecture

Multiply the action distributions

\[ \pi(a|s) = \frac{1}{Z_1} g^t(a|z) g^p(a|z) \]
\[ \ldots \]
\[ \pi_n(a|s) = \frac{1}{Z_n} g^t(a|z) g^p_n(a|z) \]

\[ D(s) = \sum_{a \in A} |g^t(a|s) - \frac{1}{n} \sum_{i} \pi_i(a|s)| \]
Collaborative Contextual Bandit

In each context (A and B), each partner independently picks one of the green boxes. Scores a point if the same green box is chosen.
Collaborative Contextual Bandit

In each context (A and B), each partner independently picks one of the green boxes. Scores a point if the same green box is chosen.
Collaborative Contextual Bandit

In each context (A and B), each partner independently picks one of the green boxes. Scores a point if the same green box is chosen.
In context A, there is only one choice. No conventions needed.
Human Study

In context B, it is unclear how to carry over conventions.
Human Study

In context C, conventions successfully carry over to test task.
Adapting to new partner

Ours: Colored lines
Baselines: Grey lines
Higher is better
Adapting to new partner

Ours: Colored lines
Baselines: Grey lines
Higher is better

---

Our Method  BaselineAgg  BaselineAgg EarlyStopping  BaselineModular  FOMAML

average episode reward

arms

blocks

Hanabi
Adapting to new partner

Ours: Colored lines
Baselines: Grey lines
Higher is better

Our Method, BaselineAgg, BaselineAgg EarlyStopping, BaselineModular, FOMAML

0.9
0.8
0.7
0.6
0.5
0 50 100 150 200 250 steps
average episode reward

arms
blocks
Hanabi
Adapting to new partner

Ours: Colored lines
Baselines: Grey lines
Higher is better

Our Method  BaselineAgg  BaselineAgg EarlyStopping  BaselineModular  FOMAML

arms

blocks

Hanabi
Adapting to new task

Ours: Colored lines
Baselines: Grey lines
Higher is better
Adapting to new task

Ours: Colored lines
Baselines: Grey lines
Higher is better
Takeaways

Collaborative tasks involve a combination of rules and conventions.
Takeaways

Collaborative tasks involve a combination of *rules* and *conventions*.

**Rules** carry over to different *partners*;  
*Conventions* carry over to different *tasks*;
Takeaways

Collaborative tasks involve a combination of rules and conventions.

Rules carry over to different partners;
Conventions carry over to different tasks;

Learn separate composable representations for rules and conventions to quickly adapt to new tasks and partners.