ICML 2022

Short Oral Presentation

Research conducted under:

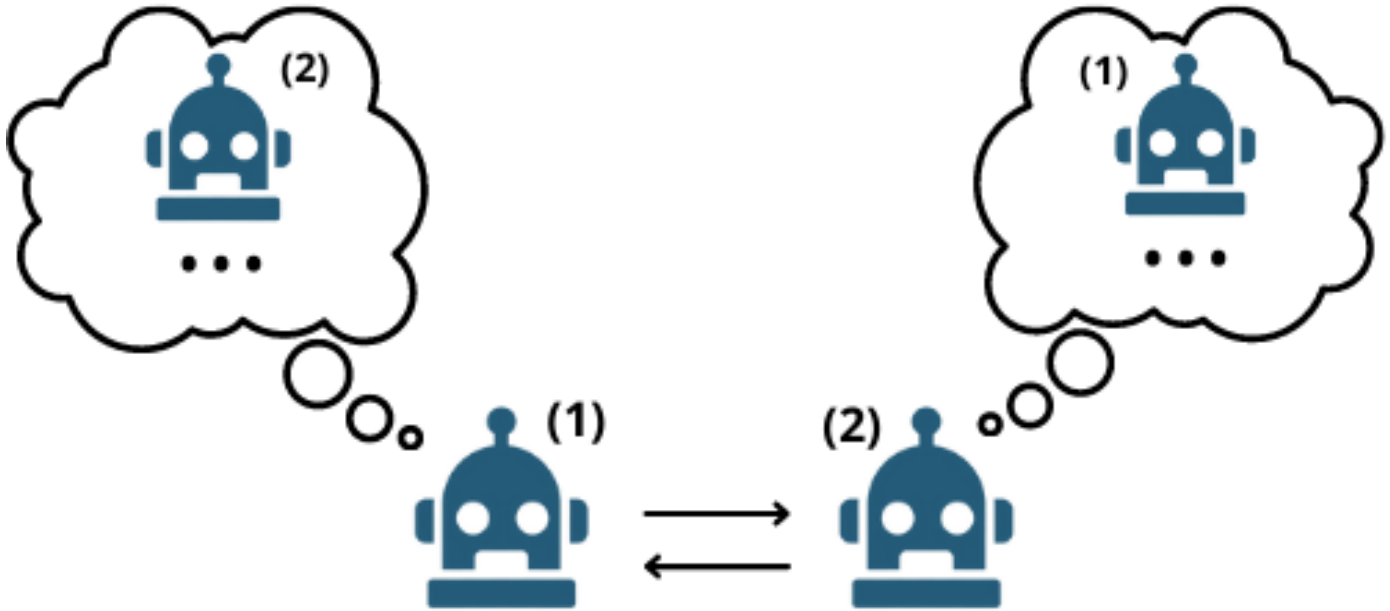
Foerster Lab for AI Research

Whiteson Research Lab

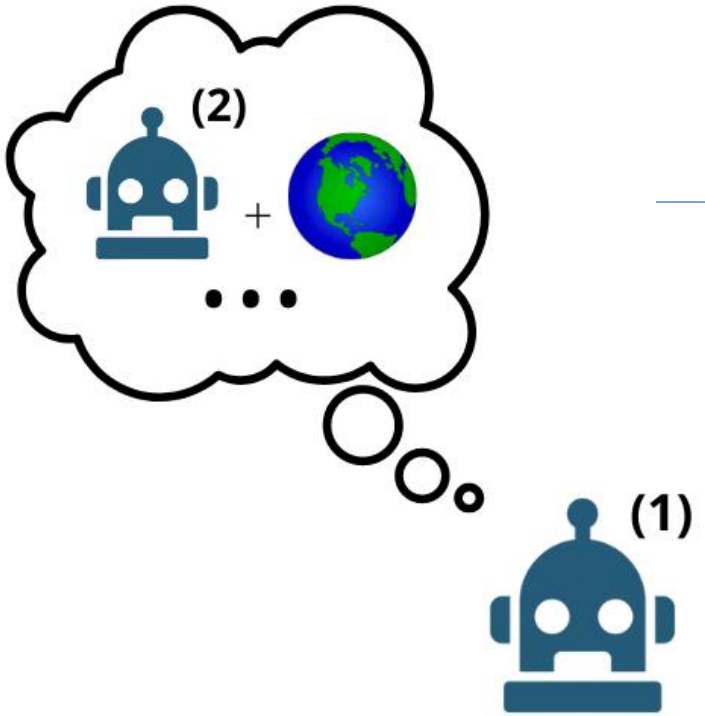
Generalized Beliefs for Cooperative AI

Darius Muglich, Luisa Zintgraf, Christian Schroeder de Witt, Shimon Whiteson, Jakob Foerster

Cooperation of Agents



Beliefs



$$
\begin{aligned}
b_t^i &= p(\tau_t \mid \tau_t^i) \\
&= p(\tau_t \mid \tau_{t-1}^i, a_t^j, o_t^i) \\
&= \frac{b_{t-1}^i \pi^j(a_t^j \mid \tau_{t-1}) p(o_t^i \mid \tau_{t-1}, a_t^j)}{\sum_{\tau'_{t-1}} b_{t-1}^i \pi^j(a_t^j \mid \tau'_{t-1}) p(o_t^i \mid \tau'_{t-1}, a_t^j)}.
\end{aligned}
$$

FLAIR

Generalized Beliefs

Assume can factor unobservable environment features from state; call these features $\{c_h\}$.

$$\psi \quad \text{s.t.} \quad c_h \sim \psi(c_h \mid \mathbf{c}_{<h}, \tau_t^i)$$

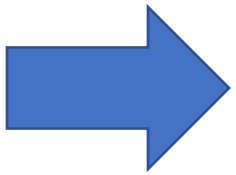
Pool of Policies

π_1 π_2

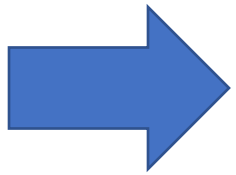
π_3 π_4

...

Generate rollouts via selfplay



Train autoregressive belief Ψ_θ with rollouts.

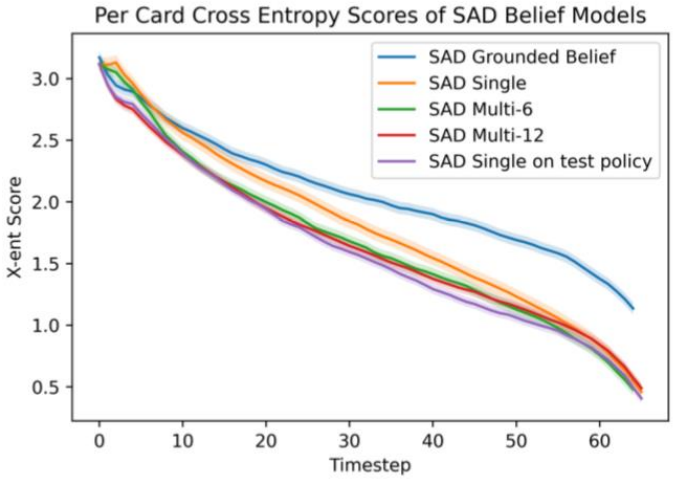


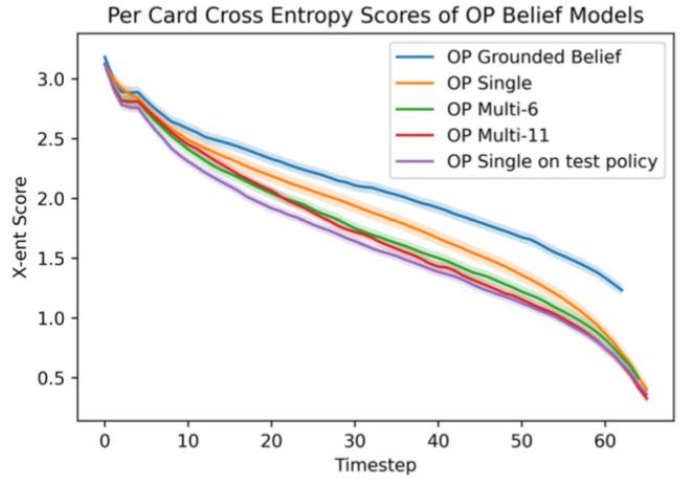
Leverage Ψ_θ for improving coordination.

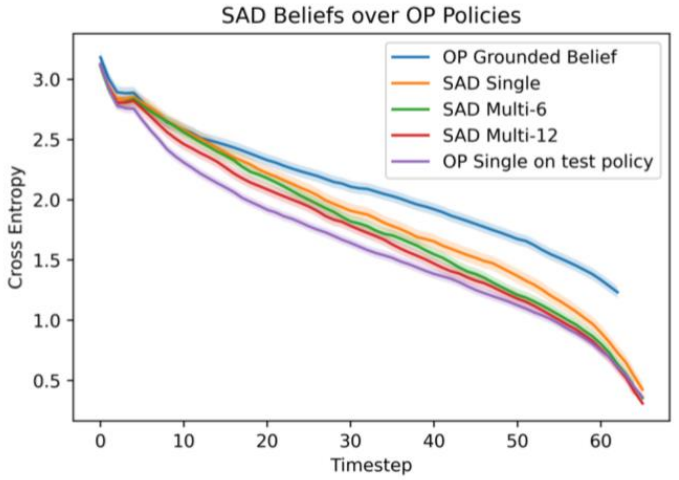
FLAIR

Generalized Beliefs Can Indeed Generalize

- Single means only one policy in training pool, Multi-x means x policies in training pool.
- Lower curve means more information deduced.
- Both intra and inter population generalization is possible.







Hu, Hengyuan, and Jakob N. Foerster. "Simplified action decoder for deep multi-agent reinforcement learning." *arXiv preprint arXiv:1912.02288* (2019).

Hu, Hengyuan, et al. "“Other-Play” for Zero-Shot Coordination." *International Conference on Machine Learning*. PMLR, 2020.

Using the Generalized Belief for Coordination

Best Response

- Train a best response over a pool of policies with the hidden state of the generalized belief as input.

Generalized Belief Search

- Use generalized belief to infer the unknown environmental features
- Run simulated rollouts on training pool policies

Results

SAD	W/O	SBS	GBS
BR W/O GEN. BELIEF	10.29± 1.05	11.32± 1.18	12.01±1.03
BR W/ GEN. BELIEF	12.36±0.96	12.03±1.11	12.47±1.02
OP	**W/O**	**SBS**	**GBS**
BR W/O GEN. BELIEF	17.49± 0.89	17.81± 0.92	18.31±0.85
BR W/ GEN. BELIEF	18.30±0.84	17.99±0.89	18.41±0.82
SAD FOR OP	**W/O**	**SBS**	**GBS**
BR W/O GEN. BELIEF	17.49± 0.89	18.47± 0.85	18.54±0.81
BR W/ GEN. BELIEF	18.11±0.82	18.68±0.87	18.99±0.84

SBS denotes "Single Belief Search", the baseline we compare Generalized Belief Search (GBS) against

Summary

- Generalized beliefs can zero-shot adapt to new policies at test time.
- There are many ways to use the generalized belief to improve cooperative performance.

Check out the paper and code on ArXiv:
https://arxiv.org/abs/2206.12765

FLAIR