Too many cooks: Bayesian inference for coordinating multi-agent collaboration

Rose E. Wang* (MIT), Sarah A. Wu* (MIT), James A. Evans (UChicago), Joshua B. Tenenbaum (MIT), David C. Parkes (Harvard), Max Kleiman-Weiner (MIT & Harvard)

Paper: https://arxiv.org/abs/2003.11778 **Code:** https://github.com/rosewang2008/gym-cooking/

Introduction

- Many current deep learning system for multi-agent coordination are specialists (brittle with new team players) and sample-inefficient (take long to train).
- Humans, including children,[1] have theory-of-mind, i.e. a commonsense ability to coordinate on the fly, even with complete strangers.
- There are at least three coordination challenges:

1. **Divide-and-conquer**: work on separate tasks in parallel

2. **Cooperation**: work together if necessary or more efficient

3. **Spatio-temporal movement**: avoid collisions & other obstacles

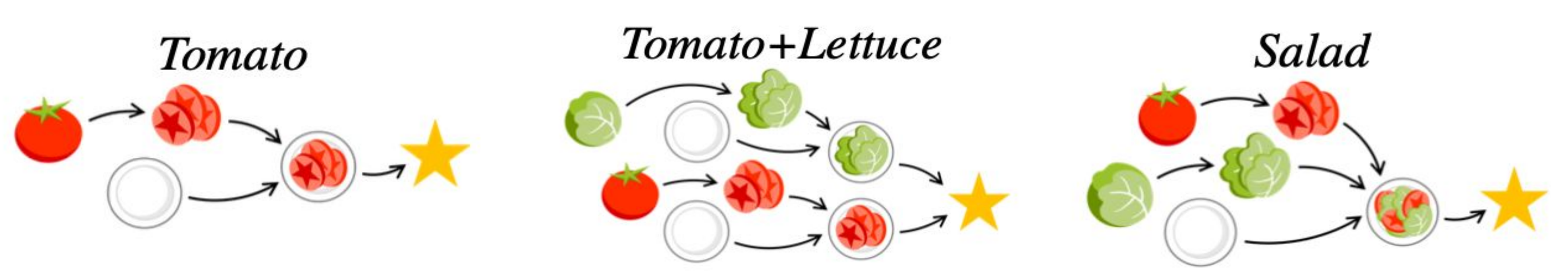
How do we build this type of ad-hoc mental state inference into machines?

Formalism

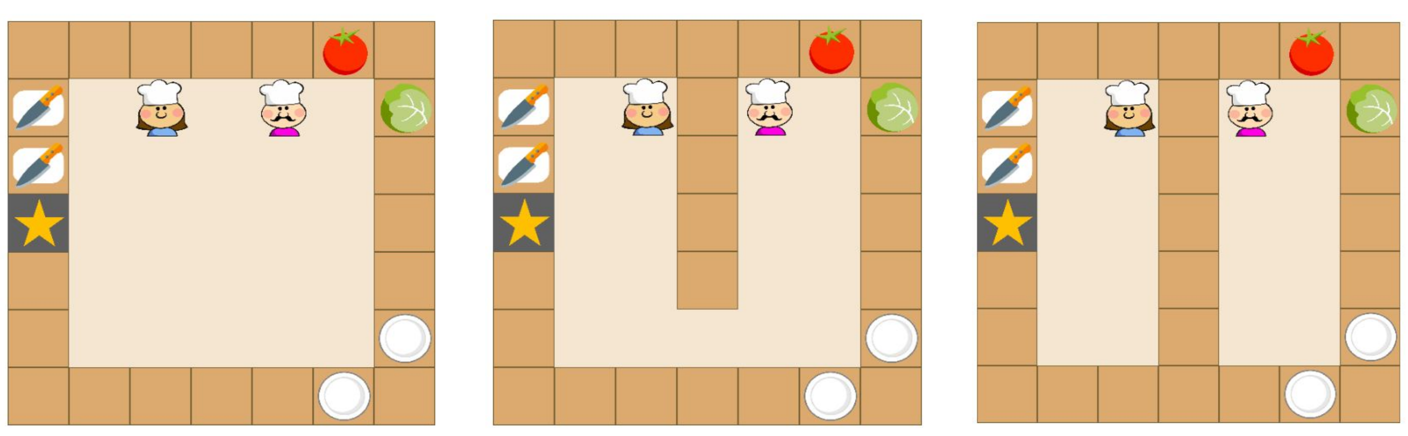
- Our work is inspired by the complex multiplayer video game *Overcooked*.
- We formalize these settings as Multi-Agent Markov Decision Processes (MMDPs)[2]
  - Add on a set of partially ordered **sub-tasks**

$$\langle n, \mathcal{S}, \mathcal{A}_{1\dots n}, T, R, \gamma, \mathcal{T} \rangle \qquad \mathcal{T} = \{\mathcal{T}_0 \dots \mathcal{T}_{|\mathcal{T}|}\}$$

- Compositional recipes (each arrow is a sub-task)



- Compositional kitchens (counters present navigation challenges & opportunities)

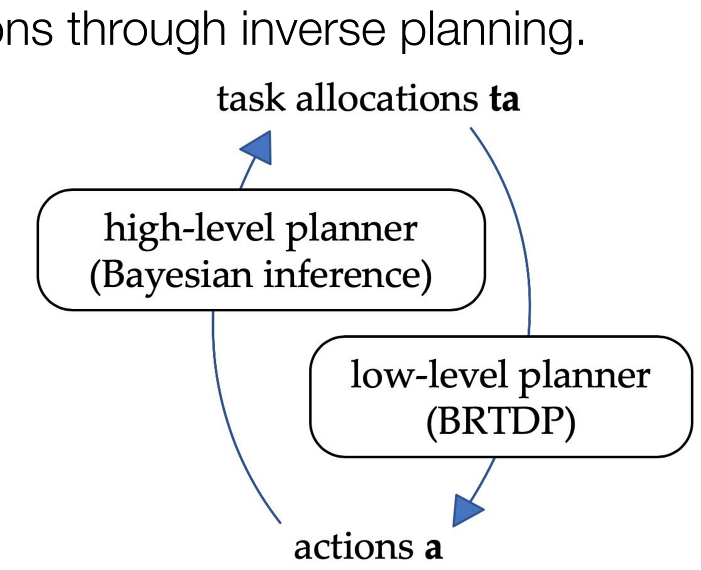


Bayesian Delegation

High-level planner (Bayesian inference)

- Uses actions to update beliefs over task allocations through inverse planning.
- Example of sub-tasks and task allocations:
  - Two sub-tasks $[\mathcal{T}_1, \mathcal{T}_2]$ and two agents $[i, j]$
  - Four possible task allocations:

$$ta = [(i:\mathcal{T}_1, j:\mathcal{T}_2),\ \ (i:\mathcal{T}_2, j:\mathcal{T}_1),\ \ (i:\mathcal{T}_1, j:\mathcal{T}_1),\ \ (i:\mathcal{T}_2, j:\mathcal{T}_2)]$$



- Each agent selects ***ta*** with maximum likelihood posterior computed via Bayes inverse planning.

$$ta^* = \arg\max_{ta} P(ta|H_{0:T}) \quad \text{where} \quad P(ta|H_{0:T}) \propto P(ta)P(H_{0:T}|ta)$$

$$P(ta) \propto \frac{1}{\text{expected time for } ta}$$

Low-level planner (BRTDP)

- Generates actions from task allocations using model-based reinforcement learning (Bounded Real-Time Dynamic Programming[3] in our model).
- Handles low-level coordination problems for each agent ***i***:
  1. Divides and conquers when $\mathcal{T}_i \neq \mathcal{T}_{-i}$ i.e. agent ***i*** has an individual task.
     - Agents best-respond to each other.
  2. Enables cooperation when $\mathcal{T}_i = \mathcal{T}_{-i}$, i.e. agent ***i*** has a joint task.
     - Agents each simulate an ideal joint planner.

Alternative model baselines

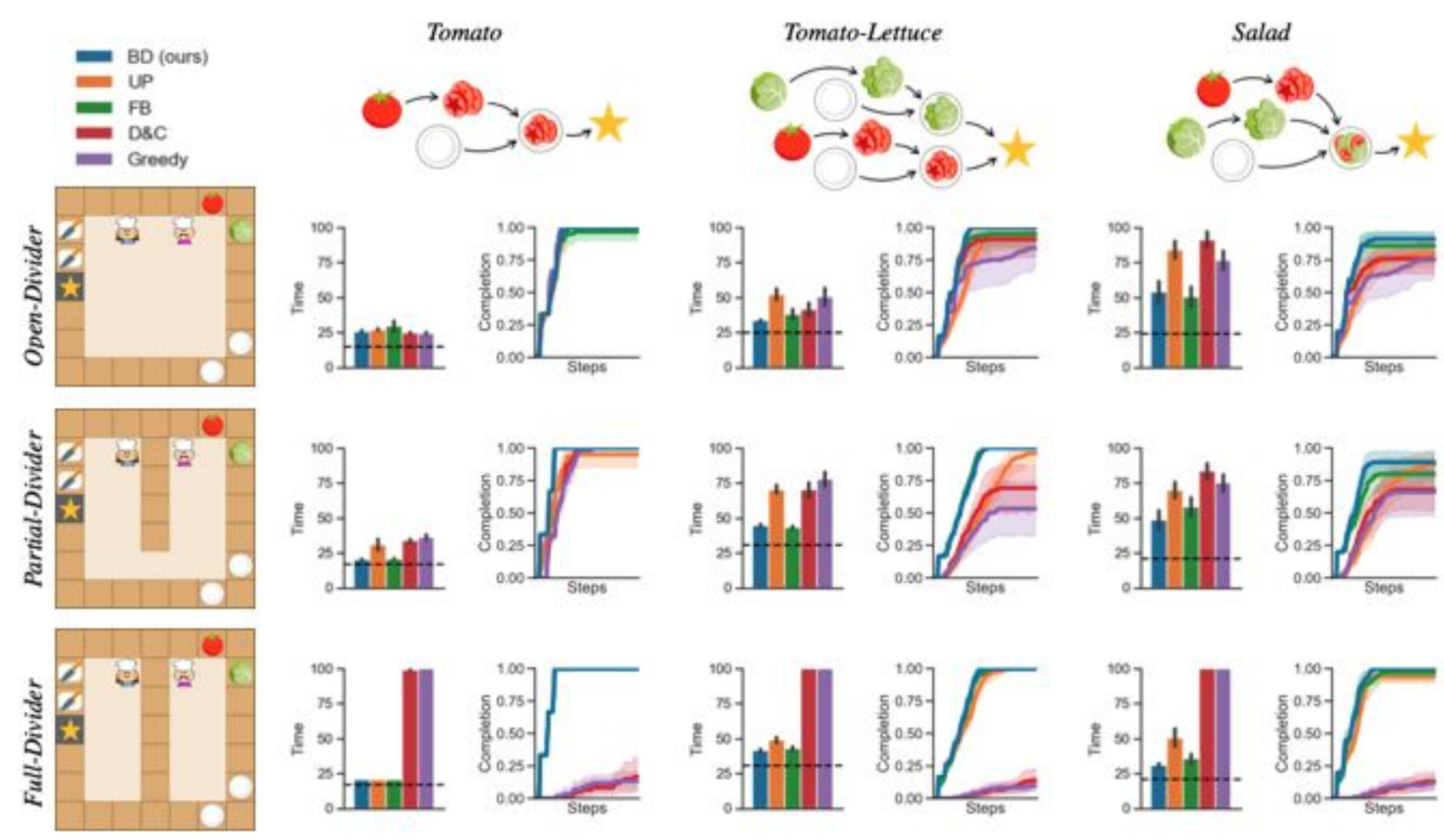
1. Uniform priors (UP): places uniform prior over all possible task allocations.
2. Fixed beliefs (FB): never updates beliefs about task allocations, i.e. keeps priors.
3. Divide & Conquer (D&C): no joint planning, i.e. only works on tasks in parallel.
4. Greedy: only considers tasks for itself, i.e. makes no inferences about others.

Experiments

1. How well does our model perform in self-play?

- Computationally simulated self-play with 2- and 3-agent teams of each model type on all 3 recipes x 3 levels.
- Found that BD agents were most successful at coordinating with each other.

		Time Steps	Completion	Shuffles
two agents	BD (ours)	**35.29 ± 1.40**	**0.98 ± 0.06**	**1.01 ± 0.05**
	UP	50.42 ± 2.04	0.94 ± 0.05	5.32 ± 0.03
	FB	37.58 ± 1.60	0.95 ± 0.04	2.64 ± 0.03
	D&C	71.57 ± 2.40	0.61 ± 0.07	13.08 ± 0.05
	Greedy	71.11 ± 2.41	0.57 ± 0.08	17.17 ± 0.06
three agents	BD (ours)	**34.52 ± 1.66**	**0.96 ± 0.08**	1.64 ± 0.05
	UP	56.84 ± 2.12	0.91 ± 0.22	5.02 ± 0.12
	FB	41.34 ± 2.27	0.92 ± 0.08	**1.55 ± 0.05**
	D&C	67.21 ± 2.31	0.67 ± 0.15	4.94 ± 0.09
	Greedy	75.87 ± 2.32	0.62 ± 0.22	12.04 ± 0.13

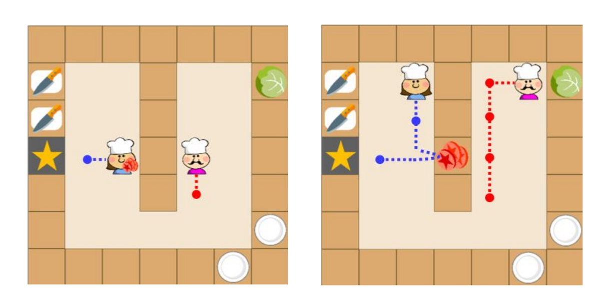


2. How well does our model perform in ad-hoc coordination?

- Computationally simulated ad-hoc play with 2-agent teams of all possible pairings among all five model types.
- Found that BD agents were most successful at coordinating ad-hoc with others.

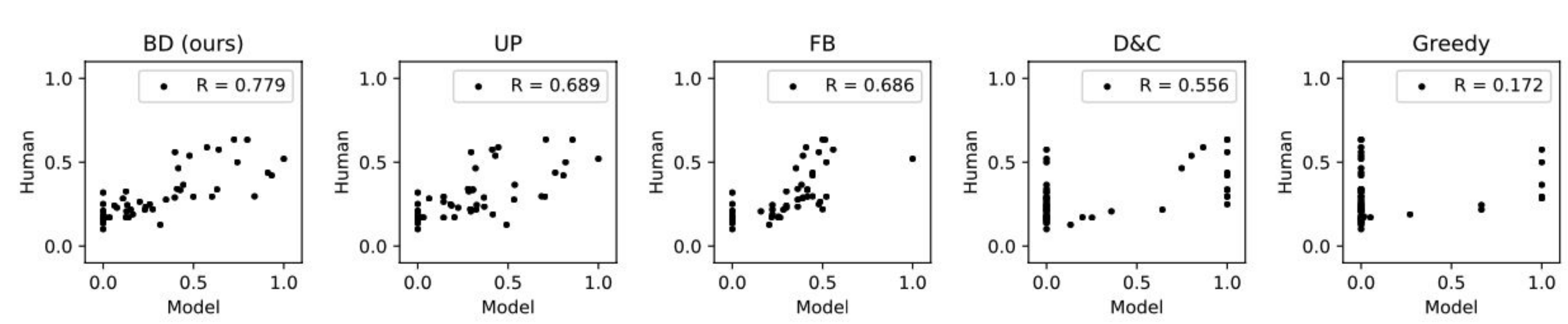
3. How do our model predictions compare with human intuitions about coordination?

- Asked participants to make inferences about 2 agents interacting over time in a behavioral task.
- Found that BD model predictions align most closely with human judgements.



Judge the likelihood that: the blue chef is plating the tomato and the red chef is chopping the lettuce.

not likely at all — certainly



Discussion & Conclusion

Using *theory-of-mind* and building on *decentralized planning*, Bayesian Delegation:

- Allows agents to rapidly infer the sub-tasks of others in group environments.
- Enables agents to decide when to cooperate and when to divide & conquer.
- Aligns with human intuitions about collaboration.

References:

1. Warneken, F., & Tomasello, M. (2006). Altruistic helping in human infants and young chimpanzees. *Science (New York, N.Y.)*, *311*(5765), 1301–1303.
2. Boutilier, C. (1996). Sequential Optimality and Coordination in Multiagent Systems. *Proceedings of TARK VI* (pp. 195–210).
3. McMahan, H. B., Likhachev, M., & Gordon, G. J. (2005). Bounded real-time dynamic programming: RTDP with monotone upper bounds and performance guarantees. In *Proceedings of ICML* (pp. 569-576).

This project was supported by:

HDSI CRCS TEMPLETON WORLD CHARITY FOUNDATION future of life INSTITUTE DARPA CBMM