Collaborative Belief Reasoning with LLMs for Efficient Multi-Agent Collaboration

Anonymous ACL submission

Abstract

Effective real-world multi-agent collaboration requires not only accurate planning but also the ability to reason about collaborators' intents–a crucial capability for avoiding miscoordination and redundant communication under partial observable environments. Due to their strong planning and reasoning capabilities, large language models (LLMs) have emerged as promising autonomous agents for collaborative task solving. However, existing collaboration frameworks for LLMs overlook their reasoning potential for *dynamic intent inference*, and thus produce inconsistent plans and redundant communication, reducing collaboration efficiency. To bridge this gap, we propose ***CoBel-World***, a novel framework that equips LLM agents with a ***Co**llaborative **Bel**ief **World***–an internal representation jointly modeling the physical environment and collaborators' mental states. CoBel-World enables agents to parse external open-world knowledge into structured beliefs via a symbolic belief representation module, and perform zero-shot Bayesian-style belief updates through LLM reasoning. This allows agents to proactively detect potential miscoordination (e.g., conflicting plans) and communicate adaptively. Evaluated on challenging embodied benchmarks (i.e., TDW-MAT and C-WAH), CoBel-World significantly reduces communication cost by **64-79%** and improves task completion efficiency by **4-28%** compared to the strongest baseline. Our results show that explicit, intent-aware belief modeling is essential for efficient and human-like collaboration in LLM-based multi-agent systems.

1 Introduction

In recent years, large language models (LLMs) have demonstrated remarkable capabilities in reasoning, planning, and decision-making (Liu et al., 2024a; OpenAI, 2023; Comanici et al., 2025; Wu et al., 2025a,b), highlighting their growing potential to act as autonomous agents in collaborative problem-solving. While these advances are promising, the effectiveness of existing LLM-based collaboration frameworks has been largely confined to simple text-based domains with high environmental certainty (Hong et al., 2023; Qian et al., 2024; Li et al., 2023a). Real-world collaboration, by contrast, requires agents to coordinate actions under uncertainty and adapt to dynamic, partially observable environments characterized by incomplete and misaligned information (Bernstein et al., 2000; Foerster et al., 2019). In such scenarios, communication becomes essential for synchronizing internal states, sharing observations, and aligning intents across agents (Pan et al., 2025; Chan et al., 2023; Han et al., 2024; Chen et al., 2025).

As shown in Figure 1, recent approaches have explored various communication protocols to enable information sharing and consensus in multi-agent systems. However, these methods typically rely on predefined collaboration schemes and fixed communication protocols-such as step-by-step message generation (Zhang et al., 2023), dense discussion (Mandi et al., 2024), or event-triggered multi-round discussion (Liu et al., 2024b). Crucially, they lack the ability to dynamically identify potential miscoordination and communicate adatively. As a result, repetitive communication and inconsistent planning frequently occur, leading to high communication costs and redundant physical actions.

We argue that this limitation stems from the lack of explicit belief modeling. In multi-agent systems, beliefs refer to the agent's internal representation of the world, including the external environment and mental states (e.g., intents, knowledge) of collaborators (Kominis and Geffner, 2015; Geffner and Bonet, 2013). In decentralized multi-agent reinforcement learning (DEC-MARL), belief modeling has proven critical for collaboration under partial observation, enabling agents to infer and align with others' internal states (Pritz and Leung, 2025; Wen et al., 2019; Zhai et al., 2023). With accurate be-

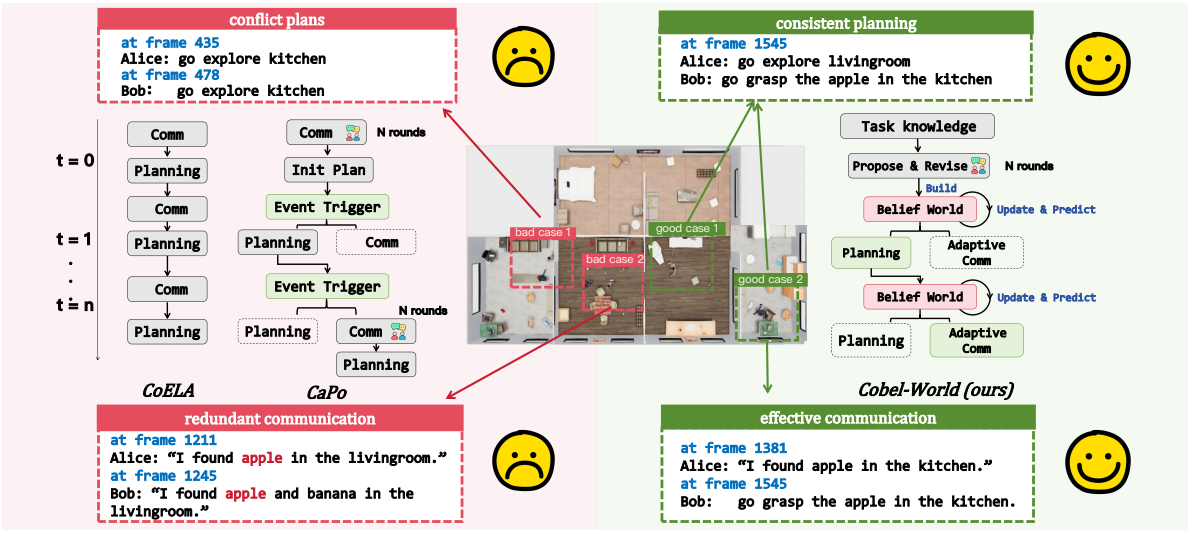

Figure 1: **Comparison of existing communication protocols for multi-agent collaboration with our work.** From left to right: (a) CoELA (Zhang et al., 2023): Fixed templates for step-by-step message generation and planning. (b) CaPo (Liu et al., 2024b): Event-driven multi-round discussion. (c) CoBel-World (ours): Belief modeling and adaptive collaboration. Our method enables consistent planning and effective communication.

lief estimation, agents can selectively communicate only the valuable information to achieve efficient communication and reach consensus, thus promoting consistent collaboration.

Despite its advantages, belief modeling for LLM agents faces two fundamental challenges. First, LLM agents operate in open-ended physical environments featured with high-dimensional, compositional action space and free-form natural language communication. These characteristics make it difficult to ground linguistic instructions into structured, explicit belief representations. Second, collecting real-world interaction trajectories to fine-tune LLMs for inferring others' intents is prohibitively expensive and often infeasible. Consequently, LLM agents must construct and update beliefs in a zero-shot manner, without access to annotated interaction data during pretraining or downstream adaptation.

To address these challenges, we propose ***CoBel-World***, a novel framework that equips LLM agents with a ***Co**llaborative **Bel**ief **World***–an internal representation of the external environment and mental states of collaborators. By leveraging the advanced reasoning capabilities of LLMs, CoBel-World enables agents to reason about the internal states of collaborators and predict the future states of the environment, thus facilitating more efficient and human-like collaboration. Specifically, CoBel-World incorporates two core components. First, inspired by symbolic planning languages such as PDDL (Fox and Long, 2003; Fabiano et al., 2021), we introduce a symbolic belief representation module to translate natural language descriptions of the open-ended world into symbolic, structured representations of beliefs. Agents then can use these symbolic beliefs to derive belief rules that guide task execution through a collaborative propose-and-revise process. Second, we design a Bayesian belief collaboration protocol that operates in the spirit of Bayesian filter. This protocol harnesses LLM reasoning to predict possible beliefs and detect potential miscoordination in a zero-shot manner, without requiring additional data for LLM fine-tuning. CoBel-World uses this protocol to dynamically update each agent's belief world model, ensuring consistent multi-agent collaboration even in partially observable environments.

To summarize, this work makes the following contributions:

- We propose CoBel-World, a novel framework that integrates a collaborative belief world into LLM agents, enabling efficient communication and consistent planning.
- We design a symbolic belief representation module to represent the world knowledge in a structured and explicit form to guide collaboration. We further design a Bayesian belief collaboration protocol in a Bayesian filter manner, demonstrating how to leverage LLM reasoning capabilities to predict beliefs and detect miscoordination in a zero-shot manner.

- We evaluate CoBel-World on challenging embodied collaboration benchmarks TDW-MAT and C-WAH with open-ended environments (Zhang et al., 2023). Results show that CoBel-World reduces the average communication cost by ***64-79%*** while improving the average task completion efficiency by ***4-28%*** compared to state-of-the-art methods, demonstrating the efficacy of belief-driven collaboration.

2 Related Work

LLM-based multi-agent collaboration. Recent efforts such as MetaGPT (Hong et al., 2023) and ChatDev (Qian et al., 2024) have explored the use of LLMs for collaborative task solving. In particular, CoELA (Zhang et al., 2023), CaPo (Liu et al., 2024b), and RoCo (Mandi et al., 2024) integrate LLMs with perception and action modules to support collaborative embodied tasks in open-ended environments. However, these approaches typically rely on fixed communication protocols, such as tepby-step message generation (Zhang et al., 2023), event-driven multi-round discussion (Liu et al., 2024b), or dense discussion (Guo et al., 2024), leading to excessive communication overhead and poor scalability under partial observability. In contrast, our work introduces a belief-driven communication mechanism that enables LLM agents to dynamically identify and exchange only the most valuable information, significantly reducing communication redundancy while improving collaboration efficiency.

Belief modeling in multi-agent systems. In decentralized partially observable Markov decision process (DEC-POMDP), belief modeling is central to enabling agents to maintain and update probabilistic estimates over hidden states and other agents' intents (Kominis and Geffner, 2015; Moreno et al., 2021). Techniques such as Bayesian reasoning (Foerster et al., 2019) and probabilistic recursive reasoning (Wen et al., 2019) allow agents to infer unobserved variables and align internal states through belief estimation. More recent approaches leverage pretrained belief models (Zhai et al., 2023; Pritz and Leung, 2025), achieving improved collaboration in cooperative games such as Hanabi and Overcooked. Wu et al. (2020) leverages inverse planning to infer collaborators' beliefs, allowing agents to dynamically switch between task division and joint collaboration. Jha et al. (2024) enables agents to perform higher-order belief modeling with significantly reduced computational cost. Cao et al. (2024) incorporates logical rules to infer human goals and beliefs from demonstrations, thereby guiding hierarchical human–AI collaboration. While promising, these methods are largely limited to low-dimensional, discrete-state environments with handcrafted features or require extensive training data. Our work bridges this gap by leveraging the zero-shot reasoning capabilities of LLMs to construct and update structured belief representations in high-dimensional, open-ended physical environments without environment-specific training.

Recent works (Yi et al., 2025; Zhang et al., 2024) attempt to incorporate belief modeling into LLM-based multi-agent systems to guide decision and strategy selection. However, these works primarily operate under communication-free settings, which limits their scalability in real-world partially observable environments. In contrast, CoBel-World leverages structured belief modeling to guide communication behaviors. Agents with such collaborative belief world can proactively determine when, who and how to communicate.

3 Method

In this section, we present ***CoBel-World***, a principled framework leveraging belief modeling to mitigate communication redundancy and collaborative misalignment in multi-agent systems. Following the paradigm of belief modeling in traditional MARL, we decompose CoBel-World into two seamlessly integrated components: **symbolic belief representation** (detailed in §3.1) for belief construction and **Bayesian belief collaboration** (detailed in §3.2) for belief update. The overall framework of CoBel-World is depicted in Figure 2.

3.1 Symbolic Belief Representation

In this section, we delve into the detailed design of symbolic belief representation, which consists of a symbolic belief language for structured belief representations and a collaborative belief initialization scheme for belief construction.

Symbolic belief language. Inspired by classical planning languages (Fox and Long, 2003; Fabiano et al., 2021), we formalize beliefs as tuples consisting of entities, attributes, and predicates. In particular, since beliefs are inherently higher-order (e.g., "Bob believes that Alice believes the apple is in the living room"), we explicitly introduce a

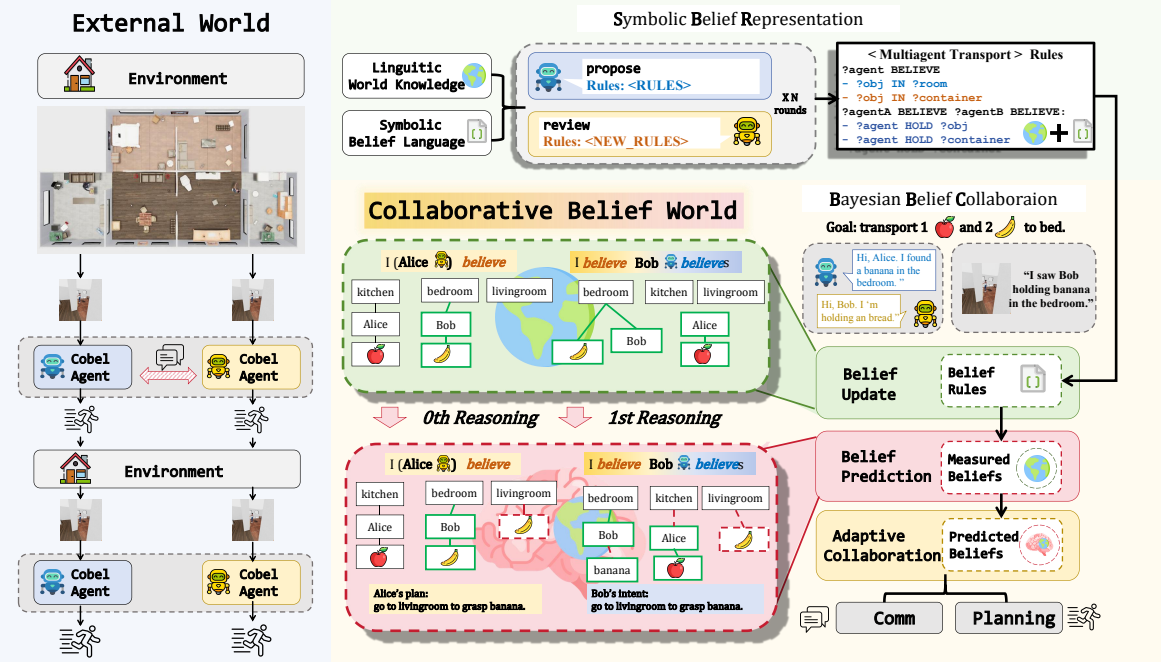

Figure 2: **Overview of CoBel-World.** Cobel-World comprises two key components: (1) **Symbolic belief representation**: All agents are organized in a collaborative reasoning process to analyze the requirements of the task and summarize the rules in a structured format. The resulting consensus set of belief rules forms the collaborative belief world. (2) **Bayesian belief collaboration**: After the belief world is constructed, each agent updates it through **belief update** and **belief prediction**, both of which are facilitated by LLM reasoning. Adaptive collaborative decisions will be made based on the beliefs.

recursive belief predicate to capture the collaborators' mental states. Specifically, we design the symbolic belief language as follows:

Atomic state:

$$s ::= \langle e_i, \textsc{Pred}, e_j \rangle \mid \langle e_i, \textsc{Prop}, v \rangle, \quad (1)$$

Zero-order belief:

$$b^0 ::= n_i \ \textsc{Bel} \ s, \quad (2)$$

First-order belief:

$$b^1 ::= n_i \ \textsc{Bel} \ (n_j \ \textsc{Bel} \ s), \quad (3)$$

where entity $e \in \mathcal{E}$ and $\mathcal{E}$ denotes the set of entities including all agents (e.g., `Alice`) and objects (e.g., `apple`); PRED represents relational descriptors (e.g. `In`, `Hold`); PROP denotes entity properties (e.g., exploration status); and $v \in \mathcal{V}$ defines the discrete values (e.g., `part`, `all`). The operator BEL serves as a connective representing an agent's mental state, mapping an agent $n_i \in \mathcal{N}$ to a state s or another belief b. Here, $\mathcal{N}$ denotes the set of all agents, with n_i, n_j representing specific instances.

Figure 3 provides a concrete example for our belief representations. When agent Alice observes the other agent Bob holding a banana, Alice constructs a zero-order belief as: `Alice BELIEVE Bob HOLD banana`. When Alice receives the message "I found an apple in the bedroom" from Bob, this information is interpreted as a first-order belief: `Alice BELIEVE Bob BELIEVE apple IN bedroom`.

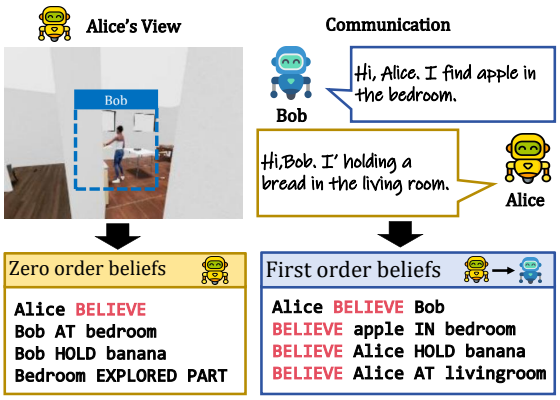

Figure 3: Examples of the transformation from unstructured observations to structured beliefs.

Collaborative belief initialization. Solely relying on the agent itself to construct beliefs is prone to hallucination issues. Hence, we introduce a collaborative belief initialization mechanism where all agents collaborate on belief construction. Specifically, rather than constructing beliefs independently, agents jointly refine a shared set of symbolic

belief rules through iterative proposal and cross-agent review. These rules encode task constraints, agent capabilities, and logical dependencies. The resulting consensus set of belief rules forms the collaborative belief world that lays the foundation for the subsequent Bayesian update process.

3.2 Bayesian Belief Collaboration

In DEC-POMDP, belief modeling follows a Bayesian filter (Chen et al., 2003) paradigm (detailed in Appendix B): a **update** step that incorporates posterior observation, followed by a **prediction** step based on prior beliefs. We adopt this well-grounded mathematical structure for CoBel-World. In the update phase, the agent updates its beliefs in response to partial observations. In the prediction phase, we leverage the reasoning capabilities of LLMs to predict the potential states of the external environment and infer collaborator's intents. The design details are elaborated as follows.

Belief update. Given the belief world which consists of a set of belief rules $\mathcal{R}$ constructed in the first phase, the agent updates two levels of beliefs via LLM reasoning.

$$b_t^0 = LLM_{\text{update\_zero}}(\mathcal{R}, o_t), \tag{4}$$

$$b_t^1 = LLM_{\text{update\_first}}(\mathcal{R}, o_c), \tag{5}$$

where $o_t = (o_v, o_c)$ is the partial observation acquired from the environment and collaborators, o_v represents the ego-centric visual perception (e.g., object positions), and o_c denotes the communication message explicitly transmitted by other agents. Notably, during the update of first-order beliefs, we employ theory-of-mind (ToM) reasoning (Li et al., 2023b; Ma et al., 2023; Strachan et al., 2024) to prompt the agent to interpret messages from the collaborator's perspective. This prevents the agent from confusing personal information with public information, ensuring a more accurate belief estimation (Shi et al., 2024; Zhang et al., 2025).

Belief prediction. Based on the updated beliefs, the agent proactively predicts the possible beliefs in the future to maintain the knowledge about the external environment and the intents of collaborators. Specifically, the agent predicts the future zero-order belief to reason the possible states of environment (Wei et al., 2022). Based on these predicted beliefs, agents then generate plans that maximize task efficiency by prioritizing high-utility, low-uncertainty exploration or manipulation steps. This process is formally defined as:

$$b_{t+1}^0 = LLM_{\text{reason}}(b_t^0, G, P), \tag{6}$$

$$\pi_{t+1} = LLM_{\text{plan}}(\bar{b}_{t+1}^0, G, P), \tag{7}$$

where b_{t+1}^0 represents the predicted zero-order beliefs, π_{t+1} represents the plan the agent will take, G is the task goal and P is the description of task progress which are both provided by the environment.

Next, the agent predicts the first-order beliefs to infer the intents of collaborators and then reasons about their possible plans to avoid conflicts and miscoordination.

$$b_{t+1}^1 = LLM_{\text{reason}}(b_t^1, G, P), \tag{8}$$

$$\bar{\pi}_{t+1} = LLM_{\text{plan}}(b_{t+1}^1, G, P), \tag{9}$$

where b_{t+1}^1 represents the predicted first-order beliefs, $\bar{\pi}_{t+1}$ represents the predicted possible plans of collaborators.

Adaptive collaboration. After updating and predicting the collaborative belief world, each agent obtains an estimation about the external world, and thus can adaptively decide the collaboration strategies. For example, when potential miscoordination (e.g. conflicting plans) is detected, they send context-aware messages to promote consensus and consistent planning among collaborators. In contrast, when serious conflicts are unlikely to occur, agents can prefer directly executing actions to improve overall efficiency.

To be specific, we implement the adaptive collaboration scheme with two steps. First, we prompt the agents to explicitly reason over two key aspects: i) belief misalignment (e.g., only Bob knows the apple's location), and ii) potentially conflicting actions (e.g., Alice and Bob plan to explore the same room). Second, if agents detect the potential miscoordination, they construct a message with the misaligned beliefs and share their intents. Based on this reasoning analysis, agents autonomously adjust their collaboration strategies, thus achieving efficient, adaptive, and intent-aware collaboration. Details are illustrated in Figure 2.

4 Experienment

4.1 Setup

Benchmarks. Recent efforts have established several benchmarks to evaluate LLM-based multi-agent systems in open-ended environments (Chang et al., 2024; Zhang et al., 2023). To demonstrate

Table 1: Performance comparison on TDW_MAT benchmark. "↑/↓" means higher/lower is better. Results in pink and underlined denote the best and second-best performance in each LLM group, respectively, Subscripts ± indicate standard deviation across three independent trials.

Task Category	Classic Agents		Qwen3-32B Agents			DeepseekV3.2 Agents			GPT-4o Agents		
	RHP	RHP+RHP	CoELA	CaPo	**CoBel-World**	CoELA	CaPo	**CoBel-World**	CoELA	CaPo	**CoBel-World**
Transport Rate (↑)											
Food-low-cap	$50.00_{\pm 0.00}$	$75.00_{\pm 0.00}$	$73.89_{\pm 3.47}$	$62.22_{\pm 1.92}$	**84.44**$_{\pm 4.19}$	$61.67_{\pm 2.89}$	$58.89_{\pm 2.51}$	**86.11**$_{\pm 0.96}$	$79.44_{\pm 3.47}$	$79.44_{\pm 4.20}$	**87.22**$_{\pm 4.20}$
Stuff-low-cap	$50.00_{\pm 0.00}$	$75.00_{\pm 0.00}$	$73.89_{\pm 4.19}$	$58.33_{\pm 4.41}$	**85.00**$_{\pm 0.00}$	$59.44_{\pm 5.85}$	$50.56_{\pm 5.09}$	**76.11**$_{\pm 1.92}$	$77.22_{\pm 2.55}$	$75.56_{\pm 7.88}$	**80.00**$_{\pm 0.00}$
Low-cap Avg	$50.00_{\pm 0.00}$	$75.00_{\pm 0.00}$	$73.89_{\pm 3.47}$	$60.28_{\pm 3.15}$	**84.72**$_{\pm 2.10}$	$60.56_{\pm 1.92}$	$54.72_{\pm 1.27}$	**81.11**$_{\pm 0.48}$	$78.33_{\pm 1.44}$	$77.50_{\pm 4.64}$	**83.61**$_{\pm 2.10}$
Food-high-cap	$40.00_{\pm 0.00}$	$73.33_{\pm 0.00}$	$75.00_{\pm 8.66}$	$67.22_{\pm 9.62}$	**86.11**$_{\pm 8.22}$	$63.89_{\pm 5.85}$	$59.44_{\pm 0.96}$	**75.56**$_{\pm 0.96}$	$86.67_{\pm 1.67}$	$82.22_{\pm 5.09}$	**87.78**$_{\pm 1.92}$
Stuff-high-cap	$46.67_{\pm 0.00}$	$80.00_{\pm 0.00}$	**86.11**$_{\pm 3.85}$	$70.56_{\pm 15.12}$	$82.78_{\pm 5.36}$	$66.67_{\pm 2.89}$	$59.44_{\pm 6.74}$	**78.33**$_{\pm 0.00}$	$79.44_{\pm 5.36}$	$80.56_{\pm 1.92}$	**85.56**$_{\pm 1.92}$
High-cap Avg	$43.33_{\pm 0.00}$	$76.67_{\pm 0.00}$	$80.56_{\pm 5.55}$	$68.89_{\pm 2.93}$	**84.44**$_{\pm 4.28}$	$65.28_{\pm 1.73}$	$59.44_{\pm 2.93}$	**76.94**$_{\pm 0.48}$	$83.06_{\pm 2.68}$	$81.39_{\pm 2.55}$	**86.67**$_{\pm 1.67}$
Total Average	$46.67_{\pm 0.00}$	$75.83_{\pm 0.00}$	$77.22_{\pm 1.05}$	$64.58_{\pm 3.00}$	**84.58**$_{\pm 1.10}$	$62.92_{\pm 1.10}$	$57.08_{\pm 1.44}$	**79.72**$_{\pm 1.05}$	$80.69_{\pm 1.05}$	$79.44_{\pm 2.06}$	**85.14**$_{\pm 1.05}$
Communication Cost (↓)											
Food-low-cap	—	—	$3921.33_{\pm 384.22}$	$7936.28_{\pm 924.44}$	**1537.00**$_{\pm 124.25}$	$3174.67_{\pm 595.31}$	$6902.06_{\pm 779.59}$	**616.22**$_{\pm 75.44}$	$2569.83_{\pm 196.92}$	$4572.03_{\pm 932.33}$	**937.44**$_{\pm 75.13}$
Stuff-low-cap	—	—	$3304.44_{\pm 147.34}$	$7424.33_{\pm 703.84}$	**1354.06**$_{\pm 198.50}$	$2901.06_{\pm 138.85}$	$6832.39_{\pm 1297.08}$	**809.67**$_{\pm 232.38}$	$2586.44_{\pm 149.51}$	$3950.36_{\pm 844.17}$	**833.11**$_{\pm 15.42}$
Low-cap Avg	—	—	$3612.89_{\pm 238.95}$	$7680.31_{\pm 466.14}$	**1445.53**$_{\pm 88.19}$	$3037.86_{\pm 332.14}$	$6867.22_{\pm 466.80}$	**712.94**$_{\pm 153.91}$	$2578.14_{\pm 173.15}$	$4261.69_{\pm 888.25}$	**885.28**$_{\pm 43.62}$
Food-high-cap	—	—	$3895.78_{\pm 189.59}$	$7786.89_{\pm 800.99}$	**1579.39**$_{\pm 330.99}$	$3331.56_{\pm 884.38}$	$7142.50_{\pm 323.73}$	**510.67**$_{\pm 16.17}$	$2523.83_{\pm 118.60}$	$4673.69_{\pm 1140.50}$	**937.67**$_{\pm 178.62}$
Stuff-high-cap	—	—	$3817.33_{\pm 86.31}$	$7697.72_{\pm 875.93}$	**1430.44**$_{\pm 181.53}$	$2593.67_{\pm 20.89}$	$7330.17_{\pm 649.90}$	**627.17**$_{\pm 24.83}$	$2366.89_{\pm 249.92}$	$4529.42_{\pm 1065.67}$	**894.83**$_{\pm 148.18}$
High-cap Avg	—	—	$3856.56_{\pm 52.39}$	$7742.31_{\pm 807.86}$	**1504.92**$_{\pm 94.08}$	$2962.61_{\pm 446.63}$	$7236.33_{\pm 451.23}$	**568.92**$_{\pm 4.33}$	$2445.36_{\pm 175.10}$	$4602.06_{\pm 1103.08}$	**916.25**$_{\pm 162.25}$
Total Average	—	—	$3734.72_{\pm 112.63}$	$7711.31_{\pm 182.13}$	**1475.22**$_{\pm 35.22}$	$3000.24_{\pm 196.27}$	$7051.78_{\pm 158.52}$	**640.93**$_{\pm 74.79}$	$2511.75_{\pm 151.99}$	$4432.71_{\pm 995.67}$	**900.76**$_{\pm 86.77}$

CoBel-World's efficiency in communication, we follow CoELA (Zhang et al., 2023) and adopt the two challenging embodied multi-agent benchmarks for our experiments: TDW-MAT (Zhang et al., 2023), and the C-WAH (Zhang et al., 2023). TDW-MAT is built on the general purpose virtual world simulation platform TDW (Gan et al., 2020), and requires agents to move objects to the destination by their hands or containers. In C-WAH, agents are requested to complete five types of household tasks, represented as various predicates with specific counts that must be satisfied. By default, we use two agents for collaboration. More details about TDW-MAT and C-WAH environments are provided in Appendix C.1 and C.2, respectively.

Metrics. Our evaluation metrics span two dimensions: task completion efficiency and communication cost. For task completion efficiency, we use different metrics for the two benchmarks. On TDW-MAT, we adopt *transport rates* as the primary performance metric, which refers to the fraction of subtasks successfully completed within 3,000 time steps (frames). Note that a single action step may span multiple time steps (e.g., arm resetting). On C-WAH, we report the *average steps* required to complete all tasks, which reflects the efficiency of collaborative coordination. For communication cost, we compute *the average number of tokens* generated by all agents per episode for communication. Higher transport rates, fewer average steps, and fewer tokens indicate better performance.

Baselines. We select two types of baselines for performance comparison: traditional LLM-free agents and LLM-based agents. The traditional agents include: (i) MCTS-based Hierarchical Planner (MHP) (Zhang et al., 2023): A hierarchical planning approach that features a Monte Carlo Tree Search (MCTS)-based high-level planner and a regression-based low-level planner. (ii) Rule-based Hierarchical Planner (RHP) (Zhang et al., 2023): A heuristic approach that uses a rule-based high-level planner combined with an A-start-based low-level planner for navigation. The LLM-based baselines include: (iii) CoELA (Zhang et al., 2023): A collaboration framework based on step-by-step templated message generation and planning. (iv) CaPo (Liu et al., 2024b): A collaboration framework based on event-driven multi-round discussions.

LLM selection. To comprehensively evaluate the effectiveness of CoBel-World across different LLMs, we adopt three state-of-the-art LLMs for CoBel-World and other LLM-based baselines: Qwen3-32B (Yang et al., 2025), DeepseekV3.2 (Liu et al., 2025) and GPT-4o (Hurst et al., 2024). We set the temperature as 0.7, top-p as 1, and maximum token limit as 512 for all LLMs.

4.2 Main Results

Performance. Table 1 and Table 2 compare the performance of different methods on the TDW-MAT and C-WAH benchmarks, respectively. Our CoBel-World framework achieves superior task efficiency over all baseline methods while significantly reducing communication cost. On TDW-MAT, CoBel-World improves average transport rate by ***4%*** over the best baseline results; on C-WAH, it reduces

Table 2: Performance comparison on C-WAH benchmark. "↑/↓" means higher/lower is better. Results in pink and underlined denote the best and second-best performance in each LLM group, respectively, Subscripts ± indicate standard deviation across three independent trials.

Task / Obs.		Classic Agents		Qwen3-32B Agents			DeepseekV3.2 Agents			GPT-4o Agents		
		MHP	MHP+MHP	CoELA	CaPo	**CoBel-World**	CoELA	CaPo	**CoBel-World**	CoELA	CaPo	**CoBel-World**
Average Step (↓)												
Prepare tea	Sym.	155.67$_{\pm 7.37}$	94.00$_{\pm 2.50}$	**87.17**$_{\pm 13.66}$	98.50$_{\pm 2.18}$	94.00$_{\pm 8.67}$	84.83$_{\pm 8.39}$	72.67$_{\pm 1.04}$	**62.33**$_{\pm 7.09}$	79.50$_{\pm 6.08}$	85.00$_{\pm 3.50}$	**64.50**$_{\pm 10.40}$
	Vis.	211.50$_{\pm 11.17}$	121.50$_{\pm 8.41}$	169.67$_{\pm 15.12}$	181.17$_{\pm 18.66}$	**114.67**$_{\pm 21.83}$	167.17$_{\pm 27.97}$	114.17$_{\pm 27.81}$	**76.83**$_{\pm 4.95}$	135.00$_{\pm 10.44}$	119.00$_{\pm 0.87}$	**54.00**$_{\pm 4.09}$
Wash dishes	Sym.	94.50$_{\pm 7.09}$	56.33$_{\pm 4.31}$	55.67$_{\pm 7.11}$	57.17$_{\pm 14.78}$	**50.33**$_{\pm 15.00}$	65.67$_{\pm 8.37}$	64.33$_{\pm 3.75}$	**62.83**$_{\pm 4.04}$	**43.00**$_{\pm 2.78}$	55.17$_{\pm 4.19}$	44.83$_{\pm 3.01}$
	Vis.	118.50$_{\pm 14.08}$	118.17$_{\pm 48.79}$	**105.67**$_{\pm 17.19}$	133.83$_{\pm 19.30}$	112.50$_{\pm 12.12}$	108.00$_{\pm 8.00}$	**96.17**$_{\pm 14.87}$	118.83$_{\pm 16.97}$	**71.83**$_{\pm 9.46}$	105.67$_{\pm 5.84}$	77.50$_{\pm 17.76}$
Prepare meal	Sym.	105.83$_{\pm 1.26}$	67.33$_{\pm 5.58}$	63.33$_{\pm 10.26}$	71.67$_{\pm 1.53}$	**59.67**$_{\pm 0.29}$	69.50$_{\pm 5.89}$	63.83$_{\pm 16.43}$	**44.00**$_{\pm 4.00}$	57.33$_{\pm 9.57}$	50.33$_{\pm 5.75}$	**50.00**$_{\pm 6.50}$
	Vis.	147.83$_{\pm 16.78}$	99.17$_{\pm 7.52}$	108.83$_{\pm 10.79}$	158.67$_{\pm 5.01}$	**100.00**$_{\pm 18.03}$	144.83$_{\pm 26.42}$	92.83$_{\pm 17.24}$	**85.00**$_{\pm 3.54}$	91.67$_{\pm 8.31}$	107.67$_{\pm 14.15}$	**68.83**$_{\pm 5.77}$
Put groceries	Sym.	112.50$_{\pm 0.50}$	72.33$_{\pm 4.86}$	81.83$_{\pm 5.03}$	72.33$_{\pm 20.31}$	**52.17**$_{\pm 1.53}$	71.33$_{\pm 7.91}$	57.50$_{\pm 6.26}$	**56.00**$_{\pm 4.44}$	63.17$_{\pm 7.01}$	**53.50**$_{\pm 2.18}$	54.67$_{\pm 1.89}$
	Vis.	158.17$_{\pm 9.00}$	89.17$_{\pm 10.20}$	118.83$_{\pm 11.00}$	146.00$_{\pm 25.24}$	**76.17**$_{\pm 11.27}$	144.50$_{\pm 17.41}$	112.67$_{\pm 6.53}$	**67.33**$_{\pm 7.07}$	105.67$_{\pm 30.71}$	89.50$_{\pm 12.26}$	**83.83**$_{\pm 3.88}$
Set up table	Sym.	85.67$_{\pm 6.93}$	54.50$_{\pm 8.53}$	68.50$_{\pm 4.09}$	64.33$_{\pm 3.55}$	**55.17**$_{\pm 5.75}$	59.67$_{\pm 4.04}$	**51.00**$_{\pm 1.00}$	53.33$_{\pm 0.58}$	59.67$_{\pm 3.62}$	55.50$_{\pm 5.00}$	**47.67**$_{\pm 1.04}$
	Vis.	109.17$_{\pm 9.70}$	80.17$_{\pm 9.46}$	98.50$_{\pm 30.74}$	130.17$_{\pm 18.11}$	**93.00**$_{\pm 17.26}$	110.83$_{\pm 6.66}$	119.83$_{\pm 15.33}$	**84.00**$_{\pm 4.24}$	90.67$_{\pm 14.74}$	**71.50**$_{\pm 12.13}$	72.17$_{\pm 5.25}$
Sym. Avg		110.83$_{\pm 1.97}$	68.90$_{\pm 4.06}$	71.30$_{\pm 2.70}$	72.80$_{\pm 0.61}$	**65.07**$_{\pm 1.88}$	70.20$_{\pm 1.39}$	59.07$_{\pm 1.12}$	**55.70**$_{\pm 0.78}$	60.53$_{\pm 2.25}$	59.90$_{\pm 2.55}$	**52.33**$_{\pm 1.56}$
Vis. Avg		149.03$_{\pm 5.88}$	101.63$_{\pm 1.39}$	120.30$_{\pm 3.34}$	149.97$_{\pm 7.11}$	**99.27**$_{\pm 5.23}$	135.07$_{\pm 7.35}$	107.13$_{\pm 9.13}$	**86.40**$_{\pm 1.41}$	98.97$_{\pm 3.91}$	98.67$_{\pm 2.10}$	**71.27**$_{\pm 1.26}$
Comm. Cost (↓)												
Prepare tea	Sym.	—	—	1137.83$_{\pm 140.59}$	5759.17$_{\pm 839.20}$	**299.33**$_{\pm 20.53}$	1248.67$_{\pm 493.40}$	6437.67$_{\pm 847.38}$	**413.33**$_{\pm 249.15}$	1089.67$_{\pm 55.25}$	13337.50$_{\pm 2432.89}$	**303.50**$_{\pm 44.27}$
	Vis.	—	—	1657.50$_{\pm 371.05}$	5293.67$_{\pm 854.03}$	**306.83**$_{\pm 26.00}$	1639.33$_{\pm 638.09}$	4842.00$_{\pm 595.82}$	**301.67**$_{\pm 59.38}$	972.83$_{\pm 106.81}$	8291.83$_{\pm 1206.14}$	**308.00**$_{\pm 14.26}$
Wash dishes	Sym.	—	—	988.83$_{\pm 103.83}$	5649.83$_{\pm 993.33}$	**277.50**$_{\pm 13.76}$	1028.67$_{\pm 354.98}$	4304.83$_{\pm 742.29}$	**241.67**$_{\pm 4.75}$	683.00$_{\pm 113.51}$	11443.83$_{\pm 2119.44}$	**336.50**$_{\pm 43.08}$
	Vis.	—	—	1096.67$_{\pm 182.01}$	3298.83$_{\pm 1404.99}$	**304.67**$_{\pm 13.53}$	1629.50$_{\pm 439.98}$	2954.00$_{\pm 304.11}$	**224.00**$_{\pm 43.01}$	600.67$_{\pm 100.35}$	6814.83$_{\pm 552.12}$	**281.33**$_{\pm 6.79}$
Prepare meal	Sym.	—	—	1667.33$_{\pm 356.22}$	9594.00$_{\pm 1986.36}$	**292.00**$_{\pm 14.00}$	1227.67$_{\pm 246.36}$	7686.50$_{\pm 1268.34}$	**245.00**$_{\pm 8.19}$	1116.67$_{\pm 62.50}$	15426.33$_{\pm 1687.59}$	**305.83**$_{\pm 43.54}$
	Vis.	—	—	1651.83$_{\pm 114.86}$	7015.00$_{\pm 801.68}$	**262.00**$_{\pm 21.15}$	1988.33$_{\pm 692.49}$	5807.33$_{\pm 1464.37}$	**255.00**$_{\pm 11.72}$	1222.17$_{\pm 199.35}$	10131.33$_{\pm 4368.30}$	**269.67**$_{\pm 2.02}$
Put groceries	Sym.	—	—	1384.33$_{\pm 216.43}$	4791.00$_{\pm 850.51}$	**322.83**$_{\pm 34.16}$	922.00$_{\pm 86.64}$	4389.17$_{\pm 456.48}$	**293.50**$_{\pm 31.19}$	918.00$_{\pm 172.44}$	11486.67$_{\pm 1507.05}$	**314.33**$_{\pm 52.78}$
	Vis.	—	—	1415.83$_{\pm 196.08}$	5574.33$_{\pm 1574.59}$	**316.00**$_{\pm 50.47}$	1671.33$_{\pm 428.42}$	4898.83$_{\pm 356.46}$	**253.17**$_{\pm 21.01}$	904.17$_{\pm 115.67}$	5815.17$_{\pm 2296.64}$	**314.00**$_{\pm 25.10}$
Set up table	Sym.	—	—	1500.83$_{\pm 61.31}$	4931.50$_{\pm 1097.43}$	**330.33**$_{\pm 70.67}$	1126.50$_{\pm 420.35}$	3568.17$_{\pm 467.03}$	**284.50**$_{\pm 44.72}$	1157.50$_{\pm 148.05}$	4927.67$_{\pm 1020.22}$	**307.83**$_{\pm 47.81}$
	Vis.	—	—	1396.33$_{\pm 139.80}$	2163.83$_{\pm 768.13}$	**302.83**$_{\pm 32.20}$	1086.67$_{\pm 470.86}$	2944.17$_{\pm 607.16}$	**271.67**$_{\pm 27.19}$	956.00$_{\pm 258.02}$	4440.67$_{\pm 652.17}$	**277.83**$_{\pm 29.20}$
Sym. Avg		—	—	1335.83$_{\pm 96.90}$	6145.10$_{\pm 330.51}$	**304.40**$_{\pm 18.74}$	1110.70$_{\pm 153.74}$	5277.27$_{\pm 323.72}$	**295.60**$_{\pm 58.58}$	992.97$_{\pm 11.82}$	11324.40$_{\pm 699.15}$	**313.60**$_{\pm 16.48}$
Vis. Avg		—	—	1443.63$_{\pm 84.60}$	4669.13$_{\pm 72.39}$	**298.47**$_{\pm 17.85}$	1603.03$_{\pm 248.26}$	4289.27$_{\pm 343.18}$	**261.10**$_{\pm 14.46}$	931.17$_{\pm 65.02}$	7098.77$_{\pm 1241.07}$	**290.17**$_{\pm 10.42}$

average steps by ***6-28%*** compared to the strongest baseline. In terms of communication cost, CoBel-World reduces token usage by ***64-79%*** across all settings. These results indicate that belief-driven collaboration not only reduces redundant communication but also enhances collaboration consistency. We also notice that LLM-based agents do not always outperform traditional agents. When driven by small LLMs like Qwen3-32B, CaPo is surpassed by RHP on TDW-MAT and both CaPo and CoELA are surpassed by MHP on C-WAH. In constrast, unlike these methods that highly rely on LLMs capabilities, CoBel-World consistently achieves better performance, demonstrating its robustness.

Qualitative analysis. Figure 4 illustrates the advantages of CoBel-World over baselines in terms of collaboration consistency and communication efficiency. As shown in Figure 4 (left), at the initial stage of the task, agents will first make the plan. CoELA follows a fixed pipeline of communication-then-planning, which often fails to reach consensus with collaborators and leads to conflicting plans. In contrast, CoBel-World performs belief prediction to infer the collaborators' intents, detect potential miscoordination, and proactively initiate communication to reach consensus. For instance, Bob infers that Alice might explore his current room and thus proactively shares his intent and beliefs with her, enabling more consistent planning. CaPo relies on event-triggered multi-round discussions to reach consensus with collaborators. However, when the triggering event provides little or no benefit to collaboration, this mechanism incurs unnecessary communication cost. As illustrated in Figure 4 (right), CaPo's discussions fail to yield better plans, resulting in redundant communication. In contrast, CoBel-World leverages belief modeling to autonomously assess the necessity of communication and dynamically decides whether to communicate to inform intents or directly execute a plan to maximize task efficiency.

4.3 Ablation Study

Effects of each component. We analyze the contributions of two key components in Cobel-World to collaboration: symbolic belief representation (SBR) and Bayesian belief collaboration (BBC). As shown in Table 3, after removing the SBR module, Cobel-World exhibits a slight performance drop. This indicates that representing beliefs using unstructured natural language introduces redundant information, impairing LLMs' planning capabilities. In contrast, emoving the BBC module leads to a severe performance drop. This phenomenon demonstrates that inferring collaborators' intents fosters more proactive collaboration.

Scaling to more agents. To validate the scal-

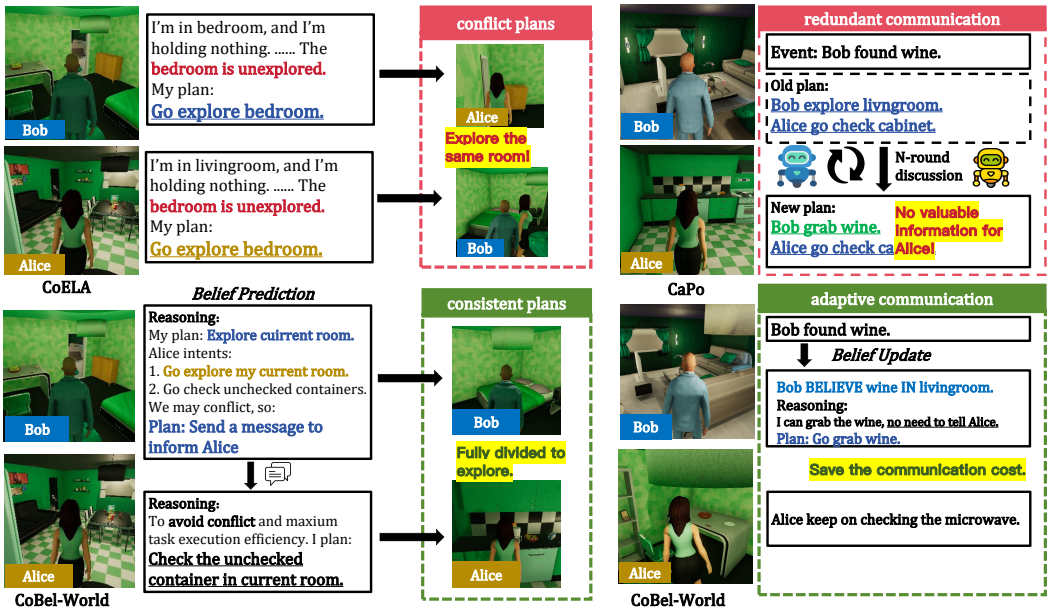

Figure 4: **Illustration of the advantages of CoBel-World in terms of planning consistency and communication efficiency on C-WAH benchmark.** All methods are powered by GPT-4o. The left part illustrates CoBel-World's superior planning consistency over CoELA, while the right panel highlights its reduced communication cost compared to CaPo.

Table 3: Effects of the components in CoBel-World using GPT-4o on C-WAH benchmark. Average steps required to complete task are reported. "SBR" denotes "symbolic belief representation" and "BBC" denotes "Bayesian belief collaboration".

Method	Symbolic Obs (↓)
CoBel-World	**52**
CoBel-World (No SBR)	55
CoBel-World (No BBC)	68

Table 4: Benefits of increasing agent number in our CoBel-World using GPT-4o on C-WAH benchmark. Average steps required for task completion are reported.

Method	Symbolic Obs (↓)
CoBel-World×2	52
CoBel-World×3	47
CoBel-World×4	**43**

Table 5: Performance comparison of heterogeneous agent collaboration on the TDW-MAT benchmark.

Method	Food	Stuff	Total
CoELA + CoELA	63	63	63
CoELA + CoBel-World	75	71	73
CoBel-World + CoBel-World	81	79	**80**

ability of CoBel-World in scenarios with larger teams, we report its performance as the number of agents increases. As shown in Table 4, CoBel-World achieves significant performance gains when adding more agents, showing strong scalability.

Cobel-World with heterogeneous agents. As shown in Table 5, a team comprising one CoBel-World agent and one CoELA agent substantially outperforms the homogeneous CoELA-CoELA setting, though it does not fully match the performance of the pure CoBel-World team. This demonstrates that CoBel-World's structured belief modeling enables effective coordination with heterogeneous partners, thereby supporting flexible collaborative scenarios.

5 Concluding Remarks

In this work, we introduce CoBel-World, a framework that equips LLM-based agents with a collaborative belief world to enable efficient and consistent multi-agent collaboration under partial observability. CoBel-World first uses a symbolic belief representation module to translate linguistic descriptions of open-ended world into structured beliefs, then harnesses LLM reasoning to perform Bayesian-style belief updates in a zero-shot manner. With CoBel-World, LLM agents can proactively infer teammates' intentions, adaptively communicate with others and detect potential miscoordination, thereby reducing redundant dialogue and physical actions. Extensive experiments show that CoBel-World reduces communication cost by 64-79% while consistently improving task completion efficiency over state-of-the-art baselines.

6 Limitations

Sensitivity to hallucinations. Since our framework uses LLM reasoning to perform zero-shot belief updates, hallucinations during the reasoning process may degrade performance. As shown in our failure analysis (Appendix D.3), hallucinations can lead to incorrect symbolic beliefs, which then influence the subsequent task execution. Future research may incorporate self-reflection or automated verification mechanisms to enhance the robustness of belief generation.

Lack of multimodal reasoning. Although the evaluation benchmarks provide high-dimensional visual observations, CoBel-World adopts the common practice in prior work–converting raw visual inputs into textual scene descriptions to serve as agent observations. While this abstraction simplifies reasoning, it fails to directly exploit the fine-grained visual cues. Future work could leverage multimodal LLMs to enable agents to perform belief updates directly from visual inputs to facilitate more effective collaborative behaviors in complex, open-world environments.

7 Ethical Considerations

In developing CoBel-World, we have considered several ethical points regarding the deployment of LLM-based agents:

Potential for model bias: Since our framework relies on LLMs for intent inference, the agents may inherit social or behavioral biases from the models' pre-training data. We encourage developers to monitor these behaviors in human-robot interaction scenarios.

Environmental impact: The use of large-scale LLMs for continuous reasoning and planning requires significant computational power. We suggest that future research explore the use of smaller, task-specific models to reduce the energy consumption and carbon footprint of these systems.

References

Daniel S. Bernstein, Shlomo Zilberstein, and Neil Immerman. 2000. The complexity of decentralized control of markov decision processes. In *Conference on Uncertainty in Artificial Intelligence*.

Chengzhi Cao, Yinghao Fu, Sheng Xu, Ruimao Zhang, and Shuang Li. 2024. Enhancing human-ai collaboration through logic-guided reasoning. In *The Twelfth International Conference on Learning Representations*.

Chi-Min Chan, Weize Chen, Yusheng Su, Jianxuan Yu, Wei Xue, Shanghang Zhang, Jie Fu, and Zhiyuan Liu. 2023. Chateval: Towards better llm-based evaluators through multi-agent debate. *arXiv preprint arXiv:2308.07201*.

Matthew Chang, Gunjan Chhablani, Alexander Clegg, Mikael Dallaire Cote, Ruta Desai, Michal Hlavac, Vladimir Karashchuk, Jacob Krantz, Roozbeh Mottaghi, Priyam Parashar, Siddharth Patki, Ishita Prasad, Xavi Puig, Akshara Rai, Ram Ramrakhya, Daniel Tran, Joanne Truong, John M. Turner, Eric Undersander, and Tsung-Yen Yang. 2024. Partnr: A benchmark for planning and reasoning in embodied multi-agent tasks. *ArXiv*, abs/2411.00081.

Weize Chen, Jiarui Yuan, Chen Qian, Cheng Yang, Zhiyuan Liu, and Maosong Sun. 2025. Optima: Optimizing effectiveness and efficiency for llm-based multi-agent system. In *Findings of the Association for Computational Linguistics: ACL 2025*, pages 11534–11557.

Zhe Chen and 1 others. 2003. Bayesian filtering: From kalman filters to particle filters, and beyond. *Statistics*, 182(1):1–69.

Gheorghe Comanici, Eric Bieber, Mike Schaekermann, Ice Pasupat, Noveen Sachdeva, Inderjit Dhillon, Marcel Blistein, Ori Ram, Dan Zhang, Evan Rosen, and 1 others. 2025. Gemini 2.5: Pushing the frontier with advanced reasoning, multimodality, long context, and next generation agentic capabilities. *arXiv preprint arXiv:2507.06261*.

Francesco Fabiano, Biplav Srivastava, Jonathan Lenchner, Lior Horesh, Francesca Rossi, and Marianna Bergamaschi Ganapini. 2021. E-pddl: A standardized way of defining epistemic planning problems. *arXiv preprint arXiv:2107.08739*.

Jakob Foerster, Francis Song, Edward Hughes, Neil Burch, Iain Dunning, Shimon Whiteson, Matthew Botvinick, and Michael Bowling. 2019. Bayesian action decoder for deep multi-agent reinforcement learning. In *International Conference on Machine Learning*, pages 1942–1951. PMLR.

Maria Fox and Derek Long. 2003. Pddl2. 1: An extension to pddl for expressing temporal planning domains. *Journal of artificial intelligence research*, 20:61–124.

Chuang Gan, Jeremy Schwartz, Seth Alter, Damian Mrowca, Martin Schrimpf, James Traer, Julian De Freitas, Jonas Kubilius, Abhishek Bhandwaldar, Nick Haber, and 1 others. 2020. Threedworld: A platform for interactive multi-modal physical simulation. *arXiv preprint arXiv:2007.04954*.

Hector Geffner and Blai Bonet. 2013. *A concise introduction to models and methods for automated planning*. Morgan & Claypool Publishers.

Xudong Guo, Kaixuan Huang, Jiale Liu, Wenhui Fan, Natalia Vélez, Qingyun Wu, Huazheng Wang,

Thomas L Griffiths, and Mengdi Wang. 2024. Embodied llm agents learn to cooperate in organized teams. *arXiv preprint arXiv:2403.12482*.

Shanshan Han, Qifan Zhang, Yuhang Yao, Weizhao Jin, and Zhaozhuo Xu. 2024. Llm multi-agent systems: Challenges and open problems. *arXiv preprint arXiv:2402.03578*.

Sirui Hong, Xiawu Zheng, Jonathan Chen, Yuheng Cheng, Jinlin Wang, Ceyao Zhang, Zili Wang, Steven Ka Shing Yau, Zijuan Lin, Liyang Zhou, and 1 others. 2023. Metagpt: Meta programming for multi-agent collaborative framework. *arXiv preprint arXiv:2308.00352*, 3(4):6.

Aaron Hurst, Adam Lerer, Adam P Goucher, Adam Perelman, Aditya Ramesh, Aidan Clark, AJ Ostrow, Akila Welihinda, Alan Hayes, Alec Radford, and 1 others. 2024. Gpt-4o system card. *arXiv preprint arXiv:2410.21276*.

Kunal Jha, Tuan Anh Le, Chuanyang Jin, Yen-Ling Kuo, Joshua B Tenenbaum, and Tianmin Shu. 2024. Neural amortized inference for nested multi-agent reasoning. In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 38, pages 530–537.

Filippos Kominis and Hector Geffner. 2015. Beliefs in multiagent planning: From one agent to many. In *Proceedings of the International Conference on Automated Planning and Scheduling*, volume 25, pages 147–155.

Guohao Li, Hasan Hammoud, Hani Itani, Dmitrii Khizbullin, and Bernard Ghanem. 2023a. Camel: Communicative agents for" mind" exploration of large language model society. *Advances in Neural Information Processing Systems*, 36:51991–52008.

Huao Li, Yu Quan Chong, Simon Stepputtis, Joseph Campbell, Dana Hughes, Michael Lewis, and Katia P. Sycara. 2023b. Theory of mind for multi-agent collaboration via large language models. In *Conference on Empirical Methods in Natural Language Processing*.

Aixin Liu, Bei Feng, Bing Xue, Bingxuan Wang, Bochao Wu, Chengda Lu, Chenggang Zhao, Chengqi Deng, Chenyu Zhang, Chong Ruan, and 1 others. 2024a. Deepseek-v3 technical report. *arXiv preprint arXiv:2412.19437*.

Aixin Liu, Aoxue Mei, Bangcai Lin, Bing Xue, Bingxuan Wang, Bingzheng Xu, Bochao Wu, Bowei Zhang, Chaofan Lin, Chen Dong, and 1 others. 2025. Deepseek-v3. 2: Pushing the frontier of open large language models. *arXiv preprint arXiv:2512.02556*.

Jie Liu, Pan Zhou, Yingjun Du, Ah-Hwee Tan, Cees G. M. Snoek, Jan Jakob Sonke, and Efstratios Gavves. 2024b. Capo: Cooperative plan optimization for efficient embodied multi-agent cooperation. *ArXiv*, abs/2411.04679.

Ziqiao Ma, Jacob Sansom, Run Peng, and Joyce Chai. 2023. Towards a holistic landscape of situated theory of mind in large language models. *arXiv preprint arXiv:2310.19619*.

Zhao Mandi, Shreeya Jain, and Shuran Song. 2024. Roco: Dialectic multi-robot collaboration with large language models. In *2024 IEEE International Conference on Robotics and Automation (ICRA)*, pages 286–299. IEEE.

Pol Moreno, Edward Hughes, Kevin R McKee, Bernardo Avila Pires, and Théophane Weber. 2021. Neural recursive belief states in multi-agent reinforcement learning. *arXiv preprint arXiv:2102.02274*.

Frans A. Oliehoek and Chris Amato. 2016. A concise introduction to decentralized pomdps. In *Springer-Briefs in Intelligent Systems*.

R OpenAI. 2023. Gpt-4 technical report. arxiv 2303.08774. *View in Article*, 2(5):1.

Melissa Z Pan, Mert Cemri, Lakshya A Agrawal, Shuyi Yang, Bhavya Chopra, Rishabh Tiwari, Kurt Keutzer, Aditya Parameswaran, Kannan Ramchandran, Dan Klein, and 1 others. 2025. Why do multiagent systems fail? In *ICLR 2025 Workshop on Building Trust in Language Models and Applications*.

Paul J Pritz and Kin K Leung. 2025. Belief states for cooperative multi-agent reinforcement learning under partial observability. *arXiv preprint arXiv:2504.08417*.

Xavier Puig, Tianmin Shu, Shuang Li, Zilin Wang, Yuan-Hong Liao, Joshua B Tenenbaum, Sanja Fidler, and Antonio Torralba. 2020. Watch-and-help: A challenge for social perception and human-ai collaboration. *arXiv preprint arXiv:2010.09890*.

Chen Qian, Wei Liu, Hongzhang Liu, Nuo Chen, Yufan Dang, Jiahao Li, Cheng Yang, Weize Chen, Yusheng Su, Xin Cong, and 1 others. 2024. Chatdev: Communicative agents for software development. In *Proceedings of the 62nd Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*, pages 15174–15186.

Haojun Shi, Suyu Ye, Xinyu Fang, Chuanyang Jin, Layla Isik, Yen-Ling Kuo, and Tianmin Shu. 2024. Muma-tom: Multi-modal multi-agent theory of mind. In *AAAI Conference on Artificial Intelligence*.

Matthijs T. J. Spaan, Geoffrey J. Gordon, and Nikos A. Vlassis. 2006. Decentralized planning under uncertainty for teams of communicating agents. In *Adaptive Agents and Multi-Agent Systems*.

James W. A. Strachan, Dalila Albergo, Giulia Borghini, Oriana Pansardi, Eugenio Scaliti, Saurabh Gupta, Krati Saxena, Alessandro Rufo, Stefano Panzeri, Guido Manzi, Michael S. A. Graziano, and Cristina Becchio. 2024. Testing theory of mind in large language models and humans. *Nature Human Behaviour*, 8:1285 – 1295.

Jason Wei, Xuezhi Wang, Dale Schuurmans, Maarten Bosma, Fei Xia, Ed Chi, Quoc V Le, Denny Zhou, and 1 others. 2022. Chain-of-thought prompting elicits reasoning in large language models. *Advances in neural information processing systems*, 35:24824–24837.

Ying Wen, Yaodong Yang, Rui Luo, Jun Wang, and Wei Pan. 2019. Probabilistic recursive reasoning for multi-agent reinforcement learning. *arXiv preprint arXiv:1901.09207*.

Duo Wu, Linjia Kang, Zhimin Wang, Fangxin Wang, Wei Zhang, Xuefeng Tao, Wei Yang, Le Zhang, Peng Cui, and Zhi Wang. 2025a. Large language models as generalist policies for network optimization. *arXiv preprint arXiv:2512.11839*.

Duo Wu, Jinghe Wang, Yuan Meng, Yanning Zhang, Le Sun, and Zhi Wang. 2025b. Catp-llm: Empowering large language models for cost-aware tool planning. In *IEEE/CVF International Conference on Computer Vision (ICCV)*. IEEE.

Sarah A Wu, Rose E Wang, James A Evans, Joshua B Tenenbaum, David C Parkes, and Max Kleiman-Weiner. 2020. Too many cooks: Coordinating multi-agent collaboration through inverse planning. In *Proceedings of the annual meeting of the cognitive science society*, volume 42.

An Yang, Anfeng Li, Baosong Yang, Beichen Zhang, Binyuan Hui, Bo Zheng, Bowen Yu, Chang Gao, Chengen Huang, Chenxu Lv, and 1 others. 2025. Qwen3 technical report. *arXiv preprint arXiv:2505.09388*.

Xie Yi, Zhanke Zhou, Chentao Cao, Qiyu Niu, Tongliang Liu, and Bo Han. 2025. From debate to equilibrium: Belief-driven multi-agent llm reasoning via bayesian nash equilibrium. *arXiv preprint arXiv:2506.08292*.

Yunpeng Zhai, Peixi Peng, Chen Su, and Yonghong Tian. 2023. Dynamic belief for decentralized multi-agent cooperative learning. In *IJCAI*, pages 344–352.

Hongxin Zhang, Weihua Du, Jiaming Shan, Qinhong Zhou, Yilun Du, Joshua B. Tenenbaum, Tianmin Shu, and Chuang Gan. 2023. Building cooperative embodied agents modularly with large language models. *ArXiv*, abs/2307.02485.

Hongxin Zhang, Zeyuan Wang, Qiushi Lyu, Zheyuan Zhang, Sunli Chen, Tianmin Shu, Yilun Du, and Chuang Gan. 2024. Combo: Compositional world models for embodied multi-agent cooperation. *ArXiv*, abs/2404.10775.

Zhining Zhang, Chuanyang Jin, Mung Yao Jia, Shunchi Zhang, and Tianmin Shu. 2025. Autotom: Scaling model-based mental inference via automated agent modeling. In *The Thirty-ninth Annual Conference on Neural Information Processing Systems*.

A Code Availability and Reproducibility

To facilitate the reproducibility of our results, we provide the full source code and implementation details of **CoBel-World** in the following anonymous repository:
https://anonymous.4open.science/r/CoBel_World

B Theoretical Analysis of CoBel-World

B.1 Multi-Agent Collaboration Formulation

We model the multi-agent collaboration task as a *decentralized partially observable Markov decision process (DEC-POMDP)* (Oliehoek and Amato, 2016; Bernstein et al., 2000; Spaan et al., 2006), defined by the tuple:

$$\mathcal{M} = \langle I, \mathcal{S}, \{\mathcal{A}_i\}, \{\mathcal{O}_i\}, T, \{O_i\}, R, h \rangle,$$

where:

- $\mathcal{N} = \{n_0, n_1, ...n_m\}$ is a finite set of m agents;
- $\mathcal{S}$ is a finite state space representing the environment;
- $\mathcal{A}_i$ is the action set of agent n_i, with $\mathcal{A} = \times_{i \in I} \mathcal{A}_i$ the joint action space;
- $\mathcal{O}_i$ is the observation set of agent n_i, encompassing partial ego-centric visual inputs and received messages;
- $T(s' \mid s, \mathbf{a}) = p(s' \mid s, \mathbf{a})$ is the transition function, denoting the probability of transitioning to state $s' \in \mathcal{S}$ from $s \in \mathcal{S}$ under joint action $\mathbf{a} \in \mathcal{A}$;
- $O_i(o_i \mid s', \mathbf{a}) = p(o_i \mid s', \mathbf{a})$ is the observation model for agent i, giving the probability of observing $o_i \in \mathcal{O}_i$ upon reaching s' after executing $\mathbf{a}$;
- $R(s, \mathbf{a})$ is the global reward function shared by all agents;
- h is the finite planning horizon.

The objective is for the team to maximize the expected cumulative reward $\mathbb{E}\left[\sum_{t=0}^{h-1} R(s^t, \mathbf{a}^t)\right]$ through decentralized execution of a joint policy $\pi = \{\pi_i\}_{i \in I}$, where each agent i selects actions $a_i^t \sim \pi_i(\cdot \mid \tau_i^t)$ based only on its local observation-action history $\tau_i^t = (o_i^0, a_i^0, \ldots, o_i^t)$.

B.2 Belief Update with Bayesian filter.

Due to partial observability, each agent n_i maintains a *belief state* $b_i : \mathcal{S} \to [0, 1]$, which represents its subjective probability distribution over the true state $s \in \mathcal{S}$. The belief b_i^t at time t is conditioned on the agent's local history $\tau_i^t = (o_i^0, a_i^0, \ldots, o_i^t)$. Upon executing action $a^t \in \mathcal{A}$ and receiving observation $o_i^{t+1} \in \mathcal{O}_i$, agent i updates its belief using a Bayesian filter:

$$b_i'(s') \propto O_i(o_i' \mid s', \mathbf{a}) \sum_{s \in \mathcal{S}} T(s' \mid s, \mathbf{a})\, b_i(s), \tag{10}$$

where b_i is the current belief, b_i' is the updated belief, $\mathbf{a}$ is the joint action, o_i' is the new observation, and T and O_i are the transition and observation models, respectively. This update decomposes into two conceptually distinct stages:

Prediction. The agent predict possible beliefs based on its current belief:

$$\bar{b}_i(s') = \sum_{s \in \mathcal{S}} T(s' \mid s, \mathbf{a})\, b_i(s), \tag{11}$$

resulting in a prior belief $\bar{b}_i$ that captures the expected state distribution after the action. In our framework, this step is enhanced by *theory of mind reasoning* (Li et al., 2023b; Ma et al., 2023), enabling agents to anticipate teammates' intentions.

Measurement update. The agent conditions the prior on the new observation o_i' (including visual input and messages):

$$b_i'(s') \propto O_i(o_i' \mid s', \mathbf{a}) \cdot \bar{b}_i(s'), \tag{12}$$

yielding a posterior belief b_i' that incorporates direct evidence. This step enables rapid belief alignment through perception and communication.

This Bayesian-style process–predicting future states and update based on observations–forms the theoretical foundation of CoBel-World.

C Additional Environment Details

Following CoELA (Zhang et al., 2023), we evaluate our methods and baselines on two challenging embodied multi-agent benchmarks with open-ended environments: ThreeDWorld Multi-Agent Transport (TDW-MAT) and Communicative Watch-And-Help (C-WAH). Detailed descriptions of these benchmarks are provided below.

C.1 TDW-MAT

Tasks. The test set of TDW-MAT consists of 24 episodes, evenly divided into two task categories: food transportation and object transportation. The food transportation tasks involves:

- 6 types of target objects: apple, banana, orange, bread, loaf, and burger.
- 3 types of containers: bowl, plate, and tea tray.

The object transportation tasks includes:

- 6 types of target objects: calculator, mouse, pen, lighter, purse and iPhone.
- 3 types of containers: plastic, wooden, and wicker baskets.

In each task instance, the environment contains 10 target objects and 2-5 containers. Scenes are instantiated across four semantically coherent room types: living room, office, kitchen, and bedroom, with object placements adhering to real-world contextual plausibility. Agents are required to maximize the number of target objects delivered to a designated goal location within a time budget of 3,000 simulation frames. Containers serve as transport tools, each capable of carrying up to three objects; Without a container, each agent can carry at most two objects simultaneously.

Observation space. The embodied agent receives a variety of observations, with the primary ones being an egocentric RGB image and a depth image. Additionally, there are several auxiliary observations. The observation space includes:

- **RGB image.** An ego-centric image captured by a forward-facing camera, with a resolution of 512 × 512 and a field of view of 90 degrees.
- **Depth image.** This image shares the same camera intrinsic parameters as RGB image.
- **Oracle perception (optional).** An image where each object ID is represented by a distinct color, using the same camera intrinsic parameters as the RGB image.
- **Agent position and rotation.** The position and rotation of the agent within the simulation environment.
- **Messages.** Information sent by all agents.

- **Held objects.** Information about the objects currently held by the agent.
- **Opponent held objects.** Information about objects held by another agent, if the agent is within view.

Action space. In TDW-MAT, agents can perform 7 distinct types of actions to interact with the environment or communicate with each other. Each action spans multiple frames. The detailed action space is outlined below:

- **Move forward.** The agent advances by 0.5m.
- **Turn left.** The agent rotates left by 15 degrees.
- **Turn right.** The agent rotates right by 15 degrees.
- **Grasp.** The agent grasps an object, successfully performing this action only when in close proximity to the object. The object can be either a target or a container.
- **Put in.** The agent places a target into a container, an action that is possible only when the agent is holding a target in one hand and a container in the other.
- **Drop.** The agent releases the objects held in hand.
- **Send message.** The agent sends a message to others, with a limit of 500 characters per frame.

Extended TDW-MAT tasks. Building upon the classic TDW-MAT benchmark introduced by CoELA (Zhang et al., 2023), we extend the evaluation along task difficulty dimension to enable a more comprehensive comparison between CoBel-World and various baselines. Specifically, tasks are categorized into low-capacity and high-capacity settings based on the number of containers available to the agent in the environment. Each difficulty level comprises half of both the food-transportation and stuff-transportation tasks. Task details are provided in Table 6.

C.2 C-WAH

C-WAH builds upon the Watch-And-Help challenge (Puig et al., 2020) by incorporating the ability for agents to send messages to one another. Sending messages, like other actions, consumes one timestep and is subject to a maximum length constraint.

Tasks. C-WAH comprises five household-oriented tasks: preparing afternoon tea, Washing dishes, preparing a meal, putting groceries, and setting up a dinner table. The test set contains 10 episodes, including both symbolic and visual observation settings. Each task involves multiple subtasks, expressed through predicates in the form "ON/IN(x, y)", which correspond to actions like "Place x ON/IN y". Detailed information is provided in Table 7. The primary objective is to complete all given subtasks within 250 timesteps, with each task containing 3-5 subtasks.

Observation space. C-WAH provides two observation modalities: symbolic observation and visual observation. In symbolic observation, agents have full access to all object-related information in the same room, including object names, locations, states, and relational attributes. In visual observation, agents receive ego-centric RGB and depth images along with auxiliary observations. Detailed observations include:

- **RGB image.** An ego-centric image from a forward-facing camera, with a resolution of 256×512 and a field of view of 60 degrees.
- **Depth image.** An image with the same camera intrinsic parameters as the RGB image.
- **Oracle Perception.** An image where each object ID is mapped to a color, sharing the same camera intrinsic parameters as the RGB image.
- **Agent position.** The agent's position within the simulation world.
- **Messages.** Information sent by all agents.
- **Held objects.** Information about the objects currently held by the agent.
- **Opponent held objects.** Information about objects held by another agent, if visible.

Action space. The action space in C-WAH includes:

- **Walk towards.** Move towards an object in the same room or towards a specific room.
- **Turn left.** Rotate left by 30 degrees.
- **Turn right.** Rotate right by 30 degrees.

Table 6: TDW_MAT tasks extended with capacity dimension

Task Type	Container Num	Container Name
Food-low-capacity	2	tea tray, bowl, plate
Food-high-capacity	5	tea tray, bowl, plate
Stuff-low-capacity	2	wood basket, wicker basket, plastic basket
Stuff-high-capacity	5	wood basket, wicker basket, plastic basket

Table 7: Detailed description of C-WAH tasks

Task Name	Oject Set
Prepare afternoon tea	ON(cupcake,coffeetable), ON(pudding,coffeetable), ON(apple,coffeetable), ON(juice,coffeetable), ON(wine,coffeetable)
Wash dishes	IN(plate,dishwasher), IN(fork,dishwasher)
Prepare a meal	ON(coffeepot,dinnertable),ON(cupcake,dinnertable), ON(pancake,dinnertable), ON(poundcake,dinnertable), ON(pudding,dinnertable), ON(apple,dinnertable), ON(juice,dinnertable), ON(wine,dinnertable)
Put groceries	IN(cupcake,fridge), IN(pancake,fridge), IN(poundcake,fridge), IN(pudding,fridge), IN(apple,fridge), IN(juice,fridge), IN(wine,fridge)
Set up a dinner table	ON(plate,dinnertable), ON(fork,dinnertable)

- **Grasp.** Grasp an object, which can be successfully performed only when the agent is close to the object.
- **Open.** Open a closed container, performable only when the agent is near the container.
- **Close.** Close an open container, performable only when the agent is near the container.
- **Put.** Place held objects into an open container or onto a surface, performable only when the agent is near the target position.
- **Send message.** Communicate with others, with a limit of 500 characters per message.

D Cobel-World Details

D.1 Belief Symbolic Representation

Prompt templates. We list the belief rules construction prompts for the two agents Alice and Bob in the benchmarks, as shown in Figure 5 and Figure 7, respectively.

Belief rules. Figure 8 illustrates the belief rules of CoBel-World.

D.2 Bayesian Belief Collaboration

In this part, we list the prompts used in the Bayesian belief collaboration module on TDW-MAT benchmark. Figure 10 and Figure 12 illustrate the prompts for zero-order belief update and prediction, respectively. Figure 9 and Figure 11 illustrate the prompts for first-order belief update and prediction, respectively. Figure 13, Figure 14, Figure 15 and Figure 16 depict the prompts for adaptive collaboration, communication, planning and replanning, respectively.

D.3 Failure cases

The failures from CoBel-World primarily stem from inherent hallucinations in LLMs. Although our symbolic belief language substantially reduces such hallucinations, it cannot fully eliminate them. We give detailed failure cases in Figure 6.

Belief Rules Construction Prompt of Alice

Init Prompt: You are Alice, you and Bob are constructing beliefs rules to denote the zero and first order belief of the world. You should first extract entity types and predicates in a specific domain given a task description and the belief symbolic language below. After that you should use the belief symbolic language to describe the possible belief types in this task domain and send to bob for discussion.

Belief symbolic language: $BELIEF_LANGUAGE $

Task description: $TASK_DESCRIPTION$
Note that the zeroth-order belief denote my knowledge of the world, first-order belief denote my knowledge of others belief.
DO NOT generate beliefs that go beyond the information specified in the task description. Consider ONLY zero-order and first-order beliefs.
The belief rules should be in syntax format with entity represented with a "?" prefix, and without any additional comment and analysis and explanation: You should output strictly in the format of the following structure:

Entity and predicate reasoning:
Zero order belief rules:
First order belief rules:

Refine Prompt: You are Alice, you and Bob are constructing beliefs rules to denote the zero and first order belief of the world. Given a task description and the belief symbolic language below, you should refine the belief rules according to Bob's suggestions.

Belief symbolic language: $BELIEF_LANGUAGE$
Task description: $TASK_DESCRIPTION$
previous content: $PREVIOUS_CONTENT$
Bob's suggestions: $SUGGESTIONS$

DO NOT generate beliefs that go beyond the information specified in the task description.
Consider ONLY zeroth-order and first-order beliefs.
Note that the zeroth-order belief denote my knowledge of the world, first-order belief denote my knowledge of others belief.
Now try to refine your previous output according to Bob's suggestions. The belief rules should be in syntax formatwith entity represented with a ? prefix, and without any additional comment and analysis and explanation: You should output strictly in the format of the following structure:

Reasoning:
Zero order belief rules:
First order belief rules:

Figure 5: Alice's belief rules construction prompt

Failure Cases

Failure type: belief update errors
Failure log:

```
Alice: low-level-plan: send a message: Hi Bob, I've explored
part of the <Office>(3000). My subplan is to continue exploring
the <Office>(3000) and transport objects to the bed.
belief_episode_0 - INFO - first belief update:
first-order beliefs:
Bob BELIEVE Alice BELIEVE <bed> IN <Office> (3000)
```

Analysis: In this case, LLM mistakenly associated `bed` and `office (3000)` in a sentence as belief, and finally got the wrong belief.

Failure type: belief prediction errors
Failure log:

```
Bob predict_first: The <Kitchen> (5000) is commonly associated
with food items like oranges, breads, and bananas. Exploring
this room might yield a higher chance of finding the goal
objects.
Bob oppo_subplans:
- go to <Kitchen>(5000) - explore current room <Kitchen>(5000)
```

Analysis: LLM mistakenly used common sense as a priori knowledge to aid belief prediction, but there is no setting in the experimental setting that "food is more likely to be in the kitchen", as a result, LLM suggest go to kitchen first.

Failure type: action-selection errors
Failure log:

```
Alice intuitive_planning: The last time you saw **Bob was in
the <Office> (1000)**, which hasn't been explored yet. It may
contain target objects or provide useful information for the
goal.
Answer: A. go to <Office> (1000)
```

Analysis: Although LLM believed that Bob is in the `<office>(1000)` , LLM still chose to go to the `<office>(1000)` which is a redundant action.

Failure type: communication content generation error
Failure log:

```
Bob: low-level-plan:send a message: Hi Alice, I've explored
all of the <Kitchen> (5000) and the <Livingroom> (8000), finding
<orange> (556740) there. I'm holding <loaf_bread> (12849971).
My subplan is to put <loaf_bread> (12849971) into <orange>
(556740), grasp <orange> (556740), and transport to the bed.
```

Analysis: `put <loaf_bread> (12849971) into <orange> (556740)` is inconsistent with common sense and the simulation environment (Items are randomly initialized in each room) , which is the hallucination of LLM.

Figure 6: Failure cases of CoBel-World.

Belief Rules Construction Prompt of Bob

Discuss Prompt: You are bob, you and Alice are constructing belief rules to denote the zero and first order belief of the world. You are required to check the belief rules made by Alice given the challenge description below. Give your reasoning progress in the reasoning:. And then give your comments: Satisfied or Unsatisfied. If Unsatisfied, you should give your suggestions to Alice on how to refine the construction.

These suggestions may include:
Missing logical relationships among key beliefs, such as omitting the agent's belief about its position.
Formatting errors, failing to comply with the prescribed format of the belief language.

Belief symbolic language: $BELIEF_LANGUAGE$ Task description: $TASK_DESCRIPTION$ Alice content: $ALICE_CONTENT$ Check if Alice's construction satisfy the need. Make deletion advice when occurring repeat syntagma. DO NOT provide suggestions that go beyond the information specified in the task description.
Consider ONLY zeroth-order and first-order beliefs.
Note that the zeroth-order belief denote my knowledge of the world, first-order belief denote my knowledge of others belief.

You should output strictly in the format of the following structure:

Reasoning:
Suggestions:
Satisfied:(yes or no)

Figure 7: Bob's belief rules construction prompt

Belief Rules

zero-order belief rules:

```
?agent BELIEVE ?object IN ?room
?agent BELIEVE ?bed IN ?room
?agent BELIEVE ?container IN ?room
?agent BELIEVE ?agent HOLD ?object
?agent BELIEVE ?agent HOLD ?container
?agent BELIEVE ?container CONTAIN ?object
?agent BELIEVE ?room EXPLORED ?exploration_state
?agent BELIEVE ?agent AT ?room
```

first-order belief rules:

```
?agentA BELIEVE ?agentB BELIEVE ?object IN ?room
?agentA BELIEVE ?agentB BELIEVE ?bed IN ?room
?agentA BELIEVE ?agentB BELIEVE ?container IN ?room
?agentA BELIEVE ?agentB BELIEVE ?agent HOLD ?object
?agentA BELIEVE ?agentB BELIEVE ?agent HOLD ?container
?agentA BELIEVE ?agentB BELIEVE ?container CONTAIN ?object
?agentA BELIEVE ?agentB BELIEVE ?room EXPLORED ?exploration_-
state
?agentA BELIEVE ?agentB BELIEVE ?agent AT ?room
```

Figure 8: Illustration of belief rules.

Prompt for First-order Beliefs Update

Assume you are an expert in multi-agent theory-of-mind reasoning. You task is to analyze and extract the information from a multi-agent dialogue. You should take the perspective of $AGENT_NAME$ and reason that what information $OPPO_NAME$ can get from the dialogue, which is defined as the first-order beliefs and translate these information into structured form.

You should follow the steps: Firstly, extract the information from the dialogue:
- What information can $OPPO_NAME$ get from the dialogue history(which is defined as the first-order beliefs)?
- What $OPPO_NAME$ plans to do?

Secondly, translate the extracted information (excluding $OPPO_NAME$'s plan) into structured first-order beliefs in the form of belief rules without any additional explanation.

Notice:
1.Maintain the structured beliefs in the format of Belief Rules.
2.DO NOT generate information not be mentioned both in Dialogue.
3.All entitiess are denoted as <name> (id), such as <table> (712) except the agents' names(e.g. Alice, Bob).
4.The exploration state of rooms MUST be part/all/none.

Following are provided information for you:
Dialogue History: $MESSAGE$
Belief Rules: $RULE$

Answer strictly in this format:
$OPPO_NAME$ knows:
$OPPO_NAME$'s plan:
structured first order beliefs:

Figure 9: Prompt for the update of first-order beliefs.

Prompt for Zero-Order Beliefs Update

Assume you are a expert good at analyze and extract information from dialogue history. You task is to analyze and extract the information from a multi-agent dialogue and translate these information into structured form.

You should follow the steps:
Firstly, extract the information from the dialogue:
- What information can $AGENT_NAME$ get from the dialogue history(which is defined as the zero-order beliefs)?
- What $OPPO_NAME$ plans to do?

Secondly, translate the extracted information (excluding $OPPO_NAME$'s plan) into structured zero-order beliefs in the form of belief rules without any additional explanation.

Notice:
1.Maintain the structured beliefs in the format of Belief Rules.
2.DO NOT generate information not be mentioned both in Dialogue.
3.All entitiess are denoted as <name> (id), such as <table> (712) except the agents' names(e.g. Alice, Bob).
4.The exploration state of rooms MUST be part/all/none.

Following are provided information for you:
Dialogue History: $MESSAGE$
Belief Rules: $RULE$

Answer strictly in this format:
$AGENT_NAME$ knows:
$OPPO_NAME$'s plan:
structured zero order beliefs:

Figure 10: Prompt for the zero-order belief update.

Prompt for First-Order Beliefs Prediction

I am $OPPO_NAME$. I want to transport as many target objects as possible to the bed with the help of containers.

First, please reason over $OPPO_NAME$'s state to answer the following question:
What the possible locations of goal objects which haven't been transported based on the room exploration state?
Goal objects is more likely to be in the rooms which are not fully explored. Put your reasoning behind the 'reasoning:'. Give your analysis in at most two reasons.
Second, based on your reasoning, please generate the best three plans $OPPO_NAME$ will take to transport goal objects as soon as possible.

The generated plans must meet following requirements:
- One single plan can be broken down into 1 to 3 actions.
- There are 5 allowed actions you can use to construct the plan.
1) 'go to': move to a specified room. 2) 'explore current room <room>(id)': explore current room(is not fully explored) for underlying target objects. 3) 'go grasp': go to grasp a specified target object. 4) 'put': Place an object into a specified container. 5) 'transport': Transport holding objects or containers to the bed and drop them on the bed.

Here is an example of a single plan:'go to <Livingroom>(4000), go grasp <apple>(5548447), and transport holding things to the bed', it can be broken down to 3 actions- 'goto <Livingroom>(4000)', 'go grasp <apple>(5548447)' and 'transport holding things to the bed'.

Actions take several steps to finish. It may be costly to go to another room or transport to the bed, use these actions sparingly. It will be more efficient to use a container to hold more objects objects and transport to bed at a time.
Notice: Represent objects, container and room strictly in the format <name>(id) like <livingroom>(1000) <wicker_basket>(5388017).

Following are provided information for you:
Goal: $GOAL$
State: $OPPO_PROGRESS$

What I can do:
I can hold two things at a time, and they can be objects or containers. I can grasp ONLY one container at a time and put objects into the holding container to hold more objects at a time. With a container, I can hold at most four objects (three in the container hold by one hand and one object on the other hand). Note that a container can contain three objects, and will be lost once transported to the bed. The room can be explored none/part/all.

Answer strictly in this format:
reasoning:
plans:
plan1:
plan2:
plan3:

Figure 11: Prompt for first-order beliefs prediction.

Prompt for Zero-Order Beliefs Prediction

I am $AGENT_NAME$. I want to transport as many target objects as possible to the bed with the help of containers.

First, please reason over $AGENT_NAME$'s state to answer the following question:
Goal objects is more likely to be in the rooms which are not fully explored. Put your reasoning behind the 'reasoning:'. Give your analysis in at most two reasons.
Second, based on your reasoning, please generate one best plan $AGENT_NAME$ will take to transport goal objects as soon as possible.

The generated plan must meet following requirements:
- This plan can be broken down into 1 to 3 actions.
- There are 5 allowed actions you can use to construct the plan.
1) 'go to': move to a specified room. 2) 'explore current room <room>(id)': explore current room(is not fully explored) for underlying target objects. 3) 'go grasp': go to grasp a specified target object. 4) 'put': Place an object into a specified container. 5) 'transport': Transport holding objects or containers to the bed and drop them on the bed.

Here is an example of a single plan:'go to <Livingroom>(4000), go grasp <apple>(5548447), and transport holding things to the bed', it can be broken down to 3 actions- 'goto <Livingroom>(4000)', 'go grasp <apple>(5548447)' and 'transport holding things to the bed'.

Actions take several steps to finish. It may be costly to go to another room or transport to the bed, use these actions sparingly. It will be more efficient to use a container to hold more objects objects and transport to bed at a time.

Notice: Represent objects, container and room strictly in the format <name>(id) like <livingroom>(1000) <wicker_basket>(5388017).

What I can do:
I can hold two things at a time, and they can be objects or containers. I can grasp ONLY one container at a time and put objects into the holding container to hold more objects at a time. With a container, I can hold at most four objects (three in the container hold by one hand and one object on the other hand). Note that a container can contain three objects, and will be lost once transported to the bed. The room can be explored none/part/all.

Following are provided information for you:
Goal: $GOAL$
State: $MY_PROGRESS$
Answer strictly in this format:
reasoning:
plan:

Figure 12: Prompt for zero-order beliefs prediction.

Prompt for Adaptive Collaboration

I am $AGENT_NAME$. My teammate $OPPO_NAME$ and I want to transport as many target objects as possible to the bed with the help of containers.

Please answer the following questions:
1. Is there any potential miscoordination between my plan and $OPPO_NAME$'s plans or between $AGENT_NAME$'s state and $OPPO_NAME$'s state? Please analyze the miscoordination in two aspects:
(1) conflicting plans: where my plan and $OPPO_NAME$'s plans may conflict in actions or locations. Such as $OPPO_NAME$ and $AGENT_NAME$ both plan to explore the same livingroom.
(2) important misaligned information: where some information my state and $OPPO_NAME$'s state may misaligned which may lead to miscoordination.
Give your analysis in at most two reasons.

2. If there exists heavy miscoordination, please answer Yes; Otherwise, answer No. Allow for a certain degree of information misalignment which can not leads to heavy miscoordination.

3. If yes, then please find the misaligned information between my state and $OPPO_NAME$'s state. Please list these misaligned pieces of information item by item. Such as I know <apple>(12123). Just list what I know, don't need to list what $OPPO_NAME$ knows.

4. If no miscoordination, just answer NO.
Following are provided information for you:
$AGENT_NAME$'s state: $MY_PROPGRESS$
$OPPO_NAME$'s state: $OPPO_PROGRESS$
$AGENT_NAME$'s plan: $MY_SUBPLAN$
$OPPO_NAME$'s plan: $OPPO_SUBPLAN$
Answer in this format:
reasons:
answer:
misaligned information:

Figure 13: Prompt for adaptive collaboration.

Prompt for Communication Module

I am $AGENT_NAME$. My teammate $OPPO_NAME$ and I want to transport as many target objects as possible to the bed with the help of containers.

Please help me generate a message to inform $OPPO_NAME$ of the misaligned information i know but he don't know and inform $OPPO_NAME$ of my plan to achieve our shared goal collaboratively. The message should meet following requirements:
1.The message has to be concise, reliable, and helpful for assisting $OPPO_NAME$ and me to collaborate efficiently, and transport as many objects to the bed as possible.
2.The message must strictly contain two parts of contents : 1. information only $AGENT_NAME$ know and 2. my plan

Here is an example of generated massage for you:
Example:
Message:Hi $OPPO_NAME$, I' ve explored all of the <kitchen>(2000) and found <apple>(12123) there. I'm holding <banana>(12234). My plan is to grasp <apple>(12123) and transport holding things to the bed.
Just send what $AGENT_NAME$ know, don't need to send what $OPPO_NAME$ knows.

Following are provided information for you:
Misaligned information: $MISALIGNED INFORMATION$
My plan: $MY_SUBPLAN$

Figure 14: Prompt for communication module.

Prompt for Planning Module

I am $AGENT_NAME$. My teammate $OPPO_NAME$ and I want to transport as many target objects as possible to the bed with the help of containers. I can hold two things at a time, and they can be objects or containers. I can grasp containers and put objects into them to hold more objects at a time. Actions take several steps to finish.

Assume that you are an expert decision maker. Given our shared goal, my plan, my state and previous actions, please analyze the previous action and plan, judge whether the plan has been completed, and if so, respond with 'PLAN DONE'. If plan not be completed, please help me choose the best available action to achieve the goal as soon as possible. Note that a container can contain three objects, and will be lost once transported to the bed. If i'm holding nothing, just grasp a object i found and then keep on exploring or grasping another object.

You MUST select a action from the action list.

Following are provided information for you:
Goal: $GOAL$
My plan: $MY_SUBPLAN$
Previous action: $PREVIOUS_ACTIONS$
My state: $PROGRESS$
Action list: $ACTION_LIST$

Answer strictly in this format:
'answer: your choice'

Figure 15: Prompt for planning module.

Prompt for Replanning Module

I am $AGENT_NAME$. My teammate $OPPO_NAME$ and I want to transport as many target objects as possible to the bed with the help of containers.

First, please reason over $AGENT_NAME$'s state to answer the following question:
What the possible locations of goal objects which haven't been transported based on the room exploration state?
Goal objects is more likely to be in the rooms which are not fully explored. Put your reasoning behind the 'reasoning:'. Give your analysis in at most two reasons.

Second, based on your reasoning and the $OPPO_NAME$'s plan, please generate one best plan $AGENT_NAME$ will take to transport goal objects as soon as possible while avoiding conflicts with $OPPO_NAME$'s plan. The plan should collaborate with $OPPO_NAME$ to maximize execution efficiency.

The generated plans must meet following requirements:
- This plans can be broken down into 1 to 3 actions.
- There are 5 allowed actions you can use to construct the plan.
1) 'go to': move to a specified room. 2) 'explore current room <room>(id)': explore current room(is not fully explored) for underlying target objects. 3) 'go grasp': go to grasp a specified target object. 4) 'put': Place an object into a specified container. 5) 'transport': Transport holding objects or containers to the bed and drop them on the bed.

Actions take several steps to finish. It may be costly to go to another room or transport to the bed, use these actions sparingly.

It will be more efficient to use a container to hold more objects objects and transport to bed at a time. If i'm holding nothing, just grasp an object i found and then keep on exploring or grasping another object. Avoid transport only one object to bed which cost more time to transport all objects except no more goal objects need to transport.

Notice: Represent objects, container and room strictly in the format <name>(id) like <livingroom>(1000) <wicker_basket>(5388017).

What I can do: I can hold two things at a time, and they can be objects or containers. I can grasp ONLY one container at a time and put objects into the holding container to hold more objects at a time. With a container, I can hold at most four objects (three in the container hold by one hand and one object on the other hand). Note that a container can contain three objects, and will be lost once transported to the bed. The room can be explored none/part/all.
Following are provided information for you:

Goal: $GOAL$
$OPPO_NAME$'s plan: $OPPO_SUBPLAN$
State: $MY_PROGRESS$

Answer strictly in this format:
reasoning:
plan:

Figure 16: Prompt for replanning module.