SOM: Structured Opponent Modeling for LLM-based Agents via Structural Causal Model

Shiyue Cao
School of Artificial Intelligence, University of Chinese Academy of Sciences & Institute of Automation, Chinese Academy of Sciences
Beijing, China
caoshiyue2021@ia.ac.cn

Pei Xu*
National Key Laboratory of Cognition and Decision Intelligence for Complex Systems, Institute of Automation, Chinese Academy of Sciences
Beijing, China
pei.xu@ia.ac.cn

Likun Yang
School of Artificial Intelligence, University of Chinese Academy of Sciences
Beijing, China
yanglikun2021@ia.ac.cn

Lei Cui
School of Artificial Intelligence, University of Chinese Academy of Sciences
Beijing, China
cuilei2024@ia.ac.cn

Xiaotang Chen
Institute of Automation, Chinese Academy of Sciences
Beijing, China
xtchen@nlpr.ia.ac.cn

Kaiqi Huang*
School of Artificial Intelligence, University of Chinese Academy of Sciences & Institute of Automation, Chinese Academy of Sciences
Beijing, China
kaiqi.huang@nlpr.ia.ac.cn

ABSTRACT

Accurately predicting opponents' behavior from interactions is a fundamental capability for large language model (LLM)-based agents in multi-agent and game-theoretic environments. Existing approaches often entangle opponent modeling with prediction, relying on implicit contextual reasoning and limiting adaptability in dynamic interactions. To this end, we propose **S**tructured **O**pponent **M**odeling (**SOM**), a two-stage opponent modeling framework that distinctly decouples opponent model construction and opponent prediction. At the construction stage, SOM employs a Structural Causal Model (SCM), a graph-based formalism for representing dependencies among variables, to capture directed links between opponents' observations and actions, yielding an explicit and structured opponent representation. At the prediction stage, the LLM performs structured reasoning along clear pathways derived from the SCM, improving both prediction accuracy and stability. Extensive experiments on diverse multi-agent benchmarks demonstrate that SOM consistently outperforms state-of-the-art LLM-based reasoning baselines, enabling more accurate and adaptable strategic decision-making in complex and dynamic multi-agent interactions.

KEYWORDS

Opponent Modeling; Large Language Models; Multi-agent Games

ACM Reference Format:
Shiyue Cao, Pei Xu, Likun Yang, Lei Cui, Xiaotang Chen, and Kaiqi Huang. 2026. SOM: Structured Opponent Modeling for LLM-based Agents via Structural Causal Model. In *Proc. of the 25th International Conference on Autonomous Agents and Multiagent Systems (AAMAS 2026), Paphos, Cyprus, May 25 – 29, 2026*, IFAAMAS, 9 pages. https://doi.org/10.65109/EXQH7884

*Pei Xu and Kaiqi Huang are corresponding authors.

Proc. of the 25th International Conference on Autonomous Agents and Multiagent Systems (AAMAS 2026), C. Amato, L. Dennis, V. Mascardi, J. Thangarajah (eds.), May 25 – 29, 2026, Paphos, Cyprus. © 2026 International Foundation for Autonomous Agents and Multiagent Systems (www.ifaamas.org). https://doi.org/10.65109/EXQH7884

1 INTRODUCTION

Large Language Models (LLMs) have emerged as a transformative development in artificial intelligence. By training on vast amounts of text data, they acquire extensive world knowledge [26] and exhibit strong reasoning [10] and problem-solving [24] abilities. These powerful capabilities have positioned LLMs as promising candidates for autonomous agents in complex, interactive environments such as economic simulations [8, 15], collaborative tasks [4], and strategic negotiations [2]. In these multi-agent settings, an agent's success critically hinges on its ability to model opponent behavior and adapt its own strategy accordingly [21], and a lack of deep awareness of the opponent's behavior can lead to strategies that are easily exploited or misaligned, resulting in suboptimal outcomes [3]. This is particularly crucial in strategic reasoning scenarios characterized by complex strategic interactions and continuously evolving behaviors.

However, current approaches tend to implicitly entangle the modeling—the process of identifying how opponents make decisions—with opponent prediction through LLM-based contextual reasoning [6, 7, 33, 40]. This approach lacks a clear, controllable reasoning path—it neither specifies how to systematically establish the link between raw observations and an opponent's final action, nor does it guide the language model on what key intermediate reasoning processes to include, such as inferring the opponent's beliefs or their hidden information. Without this structural guidance, the language model's inference process becomes difficult to control, often missing key information [17] or producing hallucinations [11]. While existing structured reasoning methods such as Tree-of-Thought [37] and Graph-of-Thought [1] enhance LLM reasoning in many tasks, they are primarily designed for static problem settings and lack mechanisms to incorporate external feedback, making them difficult to adapt to the non-stationary nature of strategic interactions [42]. These limitations highlight the need for new approaches that enable explicit and adaptable opponent modeling in dynamic multi-agent settings.

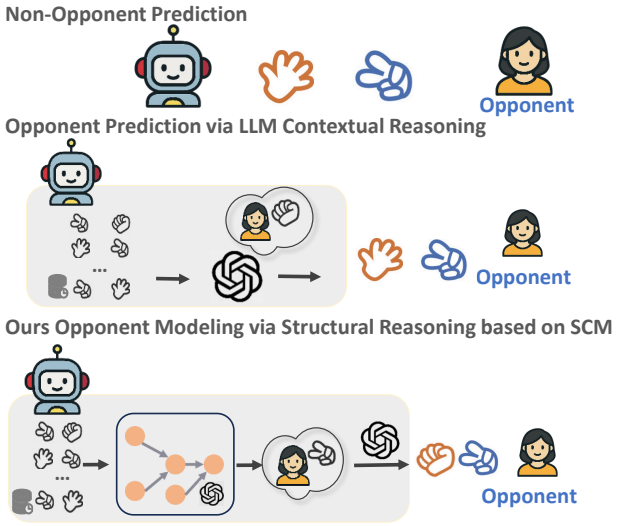


Figure 1: Illustrating different opponent modeling paradigms. Unlike baselines that ignore opponent behavior or entangle modeling within implicit reasoning, SOM explicitly constructs a structured model to guide opponent prediction.

To address these challenges, we propose **S**tructured **O**pponent **M**odeling (**SOM**), a two-stage framework that explicitly separates opponent model construction and opponent prediction. This design enables LLM-based agents to reason about opponents through a structured and controllable process rather than relying solely on implicit contextual inference. As illustrated in Figure 1, this two-stage design offers explicit and controllable reasoning pathways, in contrast to existing LLM-based approaches that entangle opponent modeling within contextual reasoning.

In the **opponent model construction** stage, SOM builds an explicit opponent model grounded in Structural Causal Models (SCMs), which provide a structural framework to organize reasoning dependencies among observable factors and opponent's decisions. After each opponent action, the LLM performs reflection to infer how the observed outcome may have arisen—linking the opponent's decisions to contextual cues and hypothesizing intermediate reasoning variables that could explain this connection. These insights are then used to progressively build and refine the SCMs, forming the explicit reasoning backbone.

In the **opponent prediction** stage, the LLM performs reasoning guided by the structured dependencies captured during the construction stage to anticipate the opponent's next action. At each step, the model draws on reasoning examples associated with the relevant dependency in the structure, which record prior successful inferences linking observed factors to opponent behavior. This allows the agent to continuously refine its reasoning with new observations, improving both the accuracy and adaptability of predictions in dynamic multi-agent interactions.

Finally, we validate the effectiveness of our approach across multiple multi-agent game environments. Extensive experiments demonstrate that our framework significantly outperforms existing baseline methods when facing different opponents. Analysis of the training process further confirms that our method accurately learns opponent strategies during interactions.

Overall, our contributions to strategic reasoning can be summarized as follows:

- We propose **SOM**, a novel opponent modeling framework that leverages Structural Causal Models (SCMs) to transform opponent prediction into a structured and controllable reasoning process.
- Within SOM, we implement two key mechanisms: a dynamic construction of the reasoning structure during interactions, and the integration of opponent-specific reasoning knowledge into the structured dependencies.
- We empirically validate SOM across diverse multi-agent environments, showing that it outperforms strong baselines and adaptively captures the behavior of different opponents over time.

2 RELATED WORK

2.1 Strategic Reasoning with LLMs

Strategic reasoning [42] refers to the capability of an agent to analyze the opponent's history and the game state, infer the opponent's strategy and actions, and adjust its own strategy to select the best course of action based on these predictions. Early work like Cicero [20] combined language models with strategic reasoning, creating a conversational agent capable of playing Diplomacy. Cicero utilized an LLM to model other players' beliefs and intentions to predict their actions, enabling human-level play. Subsequent research has applied LLMs to various multi-player games. In social deduction games like Werewolf [30, 34], studies aim to enhance agents' strategic abilities by enabling them to understand game mechanics and adapt to opponents' tactics, often involving implicit opponent prediction through dialogue analysis. Theory of Mind (ToM) [7] and k-level thinking models [43] have also been adapted to recursively infer opponents' hidden beliefs and predict their behavior in strategic reasoning. The EMO [39] method simulates opponent modeling by constructing multiple agent-specific models, but it still lacks an explicit representation of the opponent's decision-making process.

While these methods leverage the powerful reasoning capabilities of LLMs and often incorporate some form of opponent action prediction, they typically treat opponent modeling as a general reasoning task. Although some approaches may use perspective-taking to simulate inferential processes, these often lack clear and controllable reasoning pathways.

2.2 Structured Prompting for Reasoning

Structured prompting, a technique that guides LLMs through multi-step reasoning by explicitly structuring the prompt format, has significantly enhanced their reasoning capabilities. A foundational approach is Chain-of-Thought (CoT) [29], which enables LLMs to generate a series of intermediate natural language reasoning steps. Building upon this, Self-Consistency (SC) [28] improves CoT's robustness by sampling diverse reasoning paths and aggregating results via majority voting. To overcome the inherent linearity of CoT, Tree-of-Thought (ToT) [37] models reasoning as a tree-like exploration, allowing for branching and backtracking. Further

generalizing this concept, Graph-of-Thought (GoT) [1] employs arbitrary graph structures to represent complex dependencies between thoughts. Building on this, Diagram-of-Thought (DoT) [41] allows a single LLM to internally construct and reason over DAGs using role-specific tokens, streamlining multi-step reasoning without external control. Logic-of-Thought (LoT) [16] further integrates formal logic into prompts to improve consistency and deductive precision.

While structured prompting has significantly enhanced the reasoning capabilities of LLMs, existing methods are predominantly designed for static problem settings and lack mechanisms to incorporate feedback or adapt their reasoning structures over time. As a result, they struggle to effectively capture opponent behavior in dynamic multi-agent environments characterized by strategic interactions and evolving behaviors. This limitation highlights the urgent need for approaches that enable more adaptive and opponent-aware reasoning in such settings.

2.3 Opponent Modeling

Opponent modeling (OM), which analyzes and predicts other agents' behaviors in multi-agent systems, is a fundamental technique. To tackle unknown and non-stationary opponents: Encoder-decoder architectures [22] identify opponent models using only the controlled agent's local information. UAOM [35] captures aleatoric and epistemic uncertainties in stochastic opponent behaviors.Meta-learned Bayesian belief inference [44] combines variational autoencoders to model opponent beliefs; the meta-multiagent policy gradient theorem [13] adapts to new agents by accounting for mutual non-stationary dynamics. GSCU [5] learns offline opponent policy embeddings and trains a universal best-response model. For diverse opponents, MBOM [38] simulates recursive reasoning via an environment model, adapting to various types by mixing improved policies. OEOM [12] continuously generates diverse opponents via population-based training and enhances robustness with in-context reinforcement learning. To exploit opponents: L2E [31] gains exploitation abilities through minimal interactions; M-FOS [18] achieves long-horizon shaping via model-free optimization; MOL [9] uses best response theory to approximate preferences for stable equilibrium improvements. Unlike these traditional opponent modeling approaches, our work focuses on opponent modeling in LLM driven decision-making scenarios, which has not been adequately explored in existing research.

3 PRELIMINARIES

3.1 Partially Observable Stochastic Game

We model the multi-agent interactions as a **Partially Observable Stochastic Game (POSG)**, a standard framework for sequential decision-making with multiple agents. A POSG is formally defined by the tuple [36]:

$$\langle N, S, \{A^i\}_{i=1}^N, P, \{R^i\}_{i=1}^N, \gamma, \{O^i\}_{i=1}^N, Q\rangle,$$

where, N is the set of agents, and S denotes the state space. Each agent i has an individual action space A^i, and the joint action space is defined as $A = \times_{i=1}^N A^i$. The state transition function is given by $P : S \times A \to \Delta(S)$, where $P(s' \mid s, a)$ denotes the probability of transitioning from state s to state s' after taking joint action a. Each agent i receives a scalar reward determined by its reward function $R^i : S \times A \times S \to \mathbb{R}$, which gives a scalar reward for the transition $(s, a) \to s'$. $\gamma \in [0, 1]$ is the discount factor.

Each agent i receives observations $o^i \in O^i$ from the environment, and the joint observation space is defined as $O = \times_{i=1}^N O^i$. The observation function $Q : S \times A \times S \to \Delta(O)$ specifies the probability of receiving a joint observation o given joint action a and next state s', i.e., $Q(o \mid a, s')$.

The agent's local history at time t is the sequence of its past observations, actions, and rewards: $h_t^i = (o_0^i, a_0^i, r_0^i, \ldots, a_{t-1}^i, r_{t-1}^i, o_t^i)$. The agent's policy maps this history to a distribution over actions: $\pi^i(a_t^i \mid h_t^i)$.

In this work, we focus on the perspective of the self-agent (the agent under our control), denoted by superscript i. All other agents, collectively denoted by $-i$, are treated as opponents. Each opponent's policy π^{-i} is sampled from a predefined and diverse policy set Π^{opp}, which includes fixed, rule-based, and adaptive strategies.

During adaptation, the self-agent i interacts repeatedly with opponents over M episodes of POSG. The objective is to derive a policy π^i that maximizes the expected cumulative reward over the time horizon T and across all M episodes:

$$\max_{\pi^i} \mathbb{E}_{\pi^{-i} \sim \Pi^{\text{opp}}} \left[\sum_{m=1}^{M} \sum_{t=0}^{T} R_t^i \right].$$

3.2 Structural Causal Models

A **Structural Causal Model (SCM)** [23] provides a formal framework for representing causal relationships, comprising a set of variables, a causal graph, and structural equations.

Causal Graph. The causal relationships among variables are represented by a Causal Graph, denoted as $\mathcal{G}(\mathcal{V}, \mathcal{E})$.

- $\mathcal{V}$ is a set of variables (nodes) in the model.
- $\mathcal{E}$ is a set of directed edges, where an edge $V_i \to V_j$ signifies that V_i is a direct causal factor for V_j.

The graph $\mathcal{G}$ is a Directed Acyclic Graph (DAG), ensuring no causal cycles.

Structural Equations. For each variable $V_j \in \mathcal{V}$, its value is determined by its direct causal parents—denoted as $Pa(V_j)$—which are the set of variables in $\mathcal{V}$ with directed edges pointing to V_j, along with an exogenous disturbance variable U_j that accounts for external influences not explained by the model. Each such relationship is captured by a structural function f_j:

$$V_j = f_j(Pa(V_j), \ldots, U_j).$$

These structural functions f_j define the mechanism by which the value of each variable is determined by its direct causes.

In our work, we adopt SCM to formalize the opponent's decision-making process. The variables V encompass not only observable states, but also crucial latent variables representing the opponent's internal state (e.g., beliefs). The causal graph $\mathcal{G}$ structures the reasoning flow from observations to beliefs and then to actions, with each step governed by a structural function f_j that represents a decision process. While the graph and functions are unknown, the core premise of our work is that they can be dynamically inferred and approximated by LLMs. Our framework, SOM, is designed to

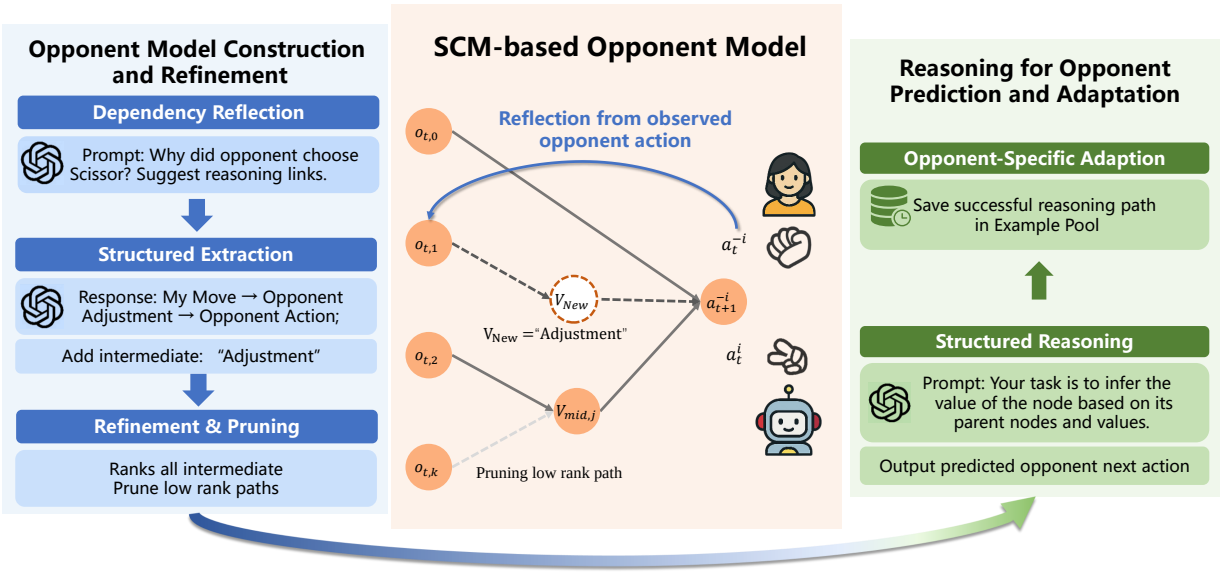


Figure 2: Illustration of the opponent modeling pipeline of SOM. SOM operates in two explicit stages. First, it constructs the SCM representation of the opponent by building a structured causal graph that captures key decision-relevant variables and their dependencies. Second, it populates the structural relationships of this SCM using personalized reasoning examples derived from past interactions. During inference, SOM traverses the graph to simulate the opponent's reasoning process step by step, enabling explicit and adaptive opponent modeling.

instantiate this SCM, using an LLM to both discover the causal structure and execute the reasoning within it.

4 METHOD

To overcome the limitations of unstructured opponent modeling, we propose SOM, a framework that grounds the modeling process in Structural Causal Models and enables structured reasoning from observations to opponent actions.

4.1 Overview of SOM Framework

In multi-agent environments, a self-agent's success critically hinges on accurately predicting an opponent's next action based on its own observations and interaction history. However, achieving precision and adaptively accurate opponent modeling in dynamic settings remains a significant challenge.

To address this, we propose SOM, a novel framework that leverages the principles of SCMs for precision and adaptive opponent behavior prediction. SOM's overall architecture is designed to enhance modeling adaptability by uncovering the underlying logic of an opponent's decisions.

As shown in Figure 2, the framework consists of two interconnected mechanisms: Dynamic SCM Construction and Refinement, which builds and updates the structured representation of the opponent's decision process through a causal graph; and Reasoning for Opponent Prediction and Adaptation, which performs structured inference within this SCM using personalized reasoning knowledge to predict opponent actions. The following sections detail these two components and their interactions.

4.2 Dynamic SCM Construction and Refinement

To systematically establish a structured link between observations and opponent behavior, SOM grounds opponent model construction in the framework of Structural Causal Models (SCMs), where this structured dependency is explicitly represented through a causal graph $\mathcal{G}(\mathcal{V}, \mathcal{E})$. Accordingly, SOM dynamically constructs and continuously refines this graph to capture how observable factors and latent reasoning variables jointly shape the opponent's decisions over time. The process follows an "observation—reflection—extraction—consolidation—pruning" cycle, enabling the model to adaptively update its representation of the opponent's decision logic as interactions unfold.

Graph Initialization. SOM begins by constructing a minimal directed acyclic graph $\mathcal{G}_0 = (\mathcal{V}_0, \mathcal{E}_0)$ before interaction. The initial node set $\mathcal{V}_0$ includes all observable variables $\{o^i_{t,1}, \ldots, o^i_{t,k}\}$ and the opponent's action a_t^{-i}. Edges are added from each observation variable to the opponent action $\mathcal{E}_0 = \{(o^i_{t,k}, a_t^{-i}) \mid \forall k\}$, representing an initial hypothesis that all observations may directly influence the opponent's action.

Reflection Phase. As the interaction proceeds, after observing the opponent's actual action a_t^{-i} in each round, SOM prompts the LLM to generate a natural language reflection. This reflection, based on the interaction history, the agent's current observation o^i_t, and the opponent's action a_t^{-i}, hypothesizes potential intermediate reasoning steps or latent beliefs that led the opponent from observations to the final action.

Structured Extraction. Another LLM module parses this reflective text, transforming unstructured natural language into structured causal chains. This process extracts intermediate nodes V_{mid}

that lie between observations and actions, along with the specific causal pathways they form (e.g., $o^i_{t,k} \rightarrow V_{\text{mid}} \rightarrow a_t^{-i}$).

Graph Update and Consolidation. After extraction, the system executes Graph Update and Consolidation. For each newly extracted intermediate node V_{new}, SOM queries an LLM to determine if it is semantically equivalent to any existing node in the graph's node set $\mathcal{V}$. To do this, the LLM receives the description of V_{new} and a list of all existing nodes in $\mathcal{V}$, and then makes a matching decision. Concurrently, the system maintains a reinforcement count $c(V)$ for each intermediate node $V \in \mathcal{V}$: if V_{new} matches an existing node V_{exist}, its count is incremented ($c(V_{\text{exist}}) \leftarrow c(V_{\text{exist}}) + 1$). If no match is found, V_{new} is added to $\mathcal{V}$ as a new node with $c(V_{\text{new}}) = 1$, and the edge set $\mathcal{E}$ is updated according to the extracted causal chain.

Graph Refinement and Pruning. To control complexity and retain critical causal hypotheses, the framework performs Graph Expansion and Refinement. After each update, SOM ranks all intermediate nodes based on their reinforcement counts $c(V)$ and retains only the top-K nodes. Nodes below this rank are pruned from the graph, ensuring the graph remains concise and effective by preserving repeatedly validated decision logic and discarding transient or refuted hypotheses.

4.3 Reasoning for Opponent Prediction and Adaptation

Given the constructed SCM that represents the opponent's decision process, SOM performs opponent prediction through structured reasoning along the dependencies encoded in the model. Specifically, reasoning proceeds in three stages—topological inference, example-guided reasoning, and personalized adaptation. By simulating the functional relationships among variables defined in the SCM, SOM predicts the opponent's next action a^{-i}_{t+1} and continually updates its understanding of the opponent as interactions unfold.

SOM traverses the causal graph $\mathcal{G}$ in topological order, ensuring that each node V_j is inferred only after all of its parent nodes $Pa(V_j)$ have been determined. The root nodes, typically the agent's observations o^i_{t+1}, obtain their values directly from the environment, whereas intermediate and action nodes are computed through a step-by-step inference process that depends on their parent nodes. For each node V_j, its value is determined by a structural equation $V_j = f_j(Pa(V_j))$, which is implemented by an LLM equipped with dynamically updated knowledge to simulate the parent-to-child causal mapping.

To enable accurate node inference, SOM constructs a tailored prompt for the LLM. The prompt consists of two components: (i) the current inferential context, namely the determined values of all parent nodes $Pa(V_j)$, and (ii) relevant reasoning examples retrieved from an opponent-specific example pool $\mathcal{P}_{\textbf{opponent}}$. The retrieval process first converts the parent nodes and their values into a textual query, and then performs semantic-similarity search in $\mathcal{P}$ to identify the top-M most similar examples. Combining the context with these retrieved examples, the LLM performs example-guided reasoning to infer the most likely value of V_j and generate the corresponding reasoning text. This personalized process enhances prediction stability.

SOM maintains a shared causal graph $\mathcal{G}$ while achieving personalized adaptation to different opponents through dynamically maintained, opponent-specific example pools. For each opponent, a distinct example pool $\mathcal{P}_{\text{opponent}}$ is incrementally populated with parent-to-child reasoning steps generated by the LLM (only when the predictions are correct). Each example e is formally represented as a four-tuple, $e = \langle$parent values, child value, reasoning text, target link$\rangle$, capturing one validated reasoning event. A strict credit-assignment policy ensures the quality of the stored knowledge: only when the predicted action $\hat{a}^{-i}_{t+1}$ matches the observed action a^{-i}_{t+1} are all intermediate reasoning steps accepted, formatted as examples, and stored in the corresponding pool. This mechanism accumulates high-quality, opponent-specific reasoning knowledge, enabling SOM to perform highly personalized and adaptive opponent modeling. The complete process of SOM is detailed in Algorithm 1.

Algorithm 1 SOM Opponent Modeling Loop

Initialize: Minimal Causal Graph $\mathcal{G}$; Example Pool $\mathcal{P}_{\text{opponent}} \leftarrow \emptyset$
for each interaction round $t = 1$ to T **do**
  Observe current observation o^i_t and opponent's actual action a_t^{-i}
  Construct/Update Graph: Prompt LLM to hypothesize causal links from o^i_t to a_t^{-i}
  Extract nodes and edges to update the graph $\mathcal{G}$
  Update Example Pool: If prediction for round $t-1$ ($\hat{a}^{-i}_{t-1}$) was correct, add its successful parent-to-child reasoning steps to $\mathcal{P}_{\text{opponent}}$
  Predict Next Action: Traverse $\mathcal{G}$ in topological order
  for each node V_j in topological order **do**
    Retrieve examples based on parent nodes $Pa(V_j)$
    Infer node value $V_j = f_j(Pa(V_j), \text{examples})$
  end for
  Output predicted action $\hat{a}^{-i}_{t+1}$ = value of $V_{a^{-i}}$
  Agent selects own action based on $\hat{a}^{-i}_{t+1}$
end for

5 EXPERIMENT

5.1 Experiment Setup

Environments. We evaluate our approach in three distinct multi-agent game environments:

- **G08A** [43]: A multi-round number-guessing game in which players choose a number between 1 and 100 in each round, aiming to be closest to 80% of the group average. This is a variant of the classic "Guess 2/3 of the Average" game proposed by Ledoux [14], where success hinges on accurately anticipating others' choices.
- **Survival Auction Game (SAG)** [19]: A multi-round sealed-bid auction game, adapted from the classic sealed-bid auction game [27], where players bid for water to restore health points. In each round, players submit bids privately, and the highest bidder wins the water. Success hinges on accurately anticipating opponents' bids to acquire water at the lowest possible cost.
- **Undercover Game** [32]: A social deduction game where players are Civilians or Undercovers with different words. Players infer their own roles from clues. Civilians aim to

Table 1: Win rates of different reasoning methods against various opponents in the G0.8A game. Rows represent the evaluated agent, and columns represent the opponent type. SOM achieves the highest overall average win rate, particularly excelling against the Mixed opponent group that aggregates diverse reasoning strategies.

Evaluated Method	Opponent Method							Avg.
	LLM only	CoT	ToT	K-R	Reflexion	Ours	Mixed	
LLM only	0.19	0.04	0.12	0.02	0.10	0.03	0.07	0.08
CoT [29]	0.68	0.16	0.54	0.28	0.32	0.09	0.36	0.35
ToT [37]	0.46	0.18	0.22	0.12	0.22	0.11	0.26	0.22
K-R [43]	**0.84**	0.54	0.48	0.24	0.45	0.17	0.42	0.45
Reflexion [25]	0.64	0.10	0.40	0.20	0.26	**0.23**	0.54	0.34
Ours	0.80±0.14	**0.61**±0.11	**0.59**±0.09	**0.39**±0.12	**0.47**±0.08	0.19±0.10	**0.64**±0.13	**0.53**

identify Undercovers, who try to conceal their roles. The core is reasoning about others' roles based on their behaviors.

Enhanced Reasoning Baselines. Recent advances in prompting techniques have significantly improved the reasoning capabilities of large language models. We focus on four representative baseline methods:

- **Chain of Thought (CoT)** [29] is a prompting method that guides LLMs to generate explicit intermediate reasoning steps, enabling them to decompose complex problems into simpler parts.
- **Tree of Thoughts (ToT)** [37] generalizes CoT by allowing LLMs to explore multiple reasoning paths. It facilitates deliberate decision-making through evaluating multiple reasoning paths, self-evaluating progress, and applying lookahead and backtracking strategies.
- **K-Level Reasoning (K-R)** [43] equips LLMs with recursive strategic reasoning, enabling agents to form higher-order beliefs about others' beliefs and adapt dynamically in multi-agent environments.
- **Reflexion** [25] enables LLM agents to improve through linguistic feedback instead of weight updates, by verbally reflecting on task feedback and storing reflections in episodic memory for better future decisions.

Meanwhile, we introduce an additional baseline named Mixed Opponent (**Mixed**), which is composed by randomly sampling opponent behaviors from CoT, ToT, K-R, and Reflexion agents. This baseline is designed to simulate a more diverse and uncertain opponent environment.

For a fair comparison, all methods are provided with a warm-up phase of 5 episodes prior to evaluation, during which interaction histories are collected. These histories are supplied as contextual input to the baseline methods during evaluation. During the evaluation phase, no additional cross-episode history is provided. Similarly, our method also fixes the SCM structure during evaluation and does not perform any cross-episode updates or adaptation, in order to ensure consistency and reproducibility across multi-runs. Unless otherwise specified, all methods and results are evaluated using GPT-4o as the base model.

More detailed experimental settings can be found in the supplementary materials.

5.2 Results

G0.8A Game and Survival Auction Game. Tables 1 and 2 summarize performance across G0.8A and SAG environments. In G0.8A, SOM achieves the highest average win rate (0.53). Against Mixed opponents, SOM substantially outperforms ToT and CoT, demonstrating its adaptability to heterogeneous strategies. While K-R excels against single LLM-only opponents, its performance declines against diverse groups, showing the limitations of fixed k-level assumptions under non-stationary behaviors. Reflexion shows moderate gains but slightly higher win rates when SOM is the opponent; this asymmetry reflects that SOM's stable, equilibrium-oriented reasoning may increase tie frequency in a game where win rates are theoretically low (0.2–0.3). In contrast, ToT and CoT struggle with dynamic mixed strategies, confirming that explicit two-stage modeling provides a tangible advantage.

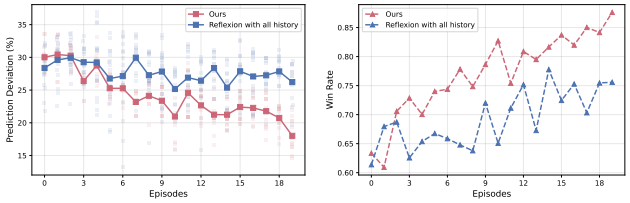
In SAG, SOM consistently leads all baselines with an average survival of 7.8 rounds. Notably, SOM surpasses K-R in nearly all matchups, highlighting its advantage when optimal bidding requires continuous adjustment. CoT and ToT exhibit variable performance, with ToT struggling against heterogeneous strategies, underscoring the limitations of static tree-based reasoning. While Reflexion improves via episodic feedback, it lacks SOM's stability due to the absence of structured model construction. Across both environments, SOM's two-stage approach—separating model construction from prediction and integrating opponent-specific knowledge—demonstrates robust adaptability and superior performance against diverse or dynamically changing opponents.

Undercover Game. In the Undercover Game (Figure 4), which requires linguistic reasoning and implicit role inference, SOM again demonstrates consistent superiority over all baselines. It achieves higher win rates both as a Civilian (Figure 4a), when identifying deceptive language patterns, and as an Undercover (Figure 4b), when strategically concealing its role. This performance improvement highlights SOM's ability to integrate structural knowledge about discourse patterns—such as topic shifts and semantic divergence—into its reasoning process. While CoT and ToT often overfit to surface-level linguistic cues, SOM's structured model allows it to capture how utterances functionally depend on hidden role intent.

Multi-Round Interaction Analysis. To investigate SOM's long-term performance, we analyze its prediction deviation and win rate in G0.8A over continuous episodes. Both SOM and Reflexion are initialized with full historical context—SOM through state refinement

Table 2: Average survival rounds of different reasoning methods against various opponents in the Survival Auction Game (SAG). Rows denote the evaluated reasoning method, while columns denote the opponent type. SOM achieves the longest survival across most opponent types, especially under the Mixed setting, demonstrating its robustness and adaptability in dynamic auction interactions.

Evaluated Method	Opponent Method							Avg.
	LLM only	CoT	ToT	K-R	Reflexion	Ours	Mixed	
LLM only	5.7	4.0	5.4	4.7	6.8	4.2	4.9	5.1
CoT [29]	6.5	5.0	7.8	5.6	5.6	4.9	5.5	5.8
ToT [37]	6.0	5.8	3.7	5.1	6.4	5.5	4.6	5.3
K-R [43]	8.1	8.4	7.8	5.6	7.4	**6.1**	6.2	7.1
Reflexion [25]	3.7	3.7	7.2	6.6	4.4	5.8	5.2	5.2
Ours	**9.1**±0.73	**8.8**±0.80	**8.3**±0.69	**7.9**±0.81	**8.1**±0.80	4.9±0.72	**7.4**±0.83	**7.8**



(a) Prediction deviation. **(b) Win rate over episodes.**

Figure 3: Action prediction deviation and win rate over episodes in G0.8A. (a) Prediction Deviation: SOM maintains higher accuracy and stability than Reflexion. (b) Win Rate: SOM exhibits superior learning progress and a higher final win rate over extended episodes.

Table 3: Ablation study of SOM components. We incrementally add SOM's core modules to evaluate their impact on Prediction Deviation and Win Rate in the G0.8A game.

Model Variant	Prediction Deviation ↓ (%)	Win Rate ↑
LLM-only	43.0	0.04
+ Static Graph	30.4	0.19
+ Intermediate Nodes	27.1	0.51
+ Graph Refine	26.9	0.54
+ Reasoning Examples (SOM)	**25.3**	**0.61**

and Reflexion via retrieved historical reflections—against LLM-only opponents.

As shown in Figure 3a, SOM's prediction deviation steadily decreases and stabilizes below Reflexion, indicating that the dynamic SCM refinement effectively captures opponent patterns. This superior accuracy translates into a strategic advantage: Figure 3b shows SOM's win rate consistently increases in tandem with error reduction, eventually significantly outperforming all baselines. Conversely, Reflexion plateaus due to the absence of a structured, continuous modeling process. These results validate SOM's core design—improving decision quality through adaptive opponent modeling—and demonstrate its robust capacity for progressive reasoning and adaptation.

5.3 Analysis of SOM's Components

5.3.1 ***Ablation Study.*** To validate the effectiveness of each core component of SOM, we conduct a series of ablation experiments by incrementally adding key modules and evaluating their impact on model performance. The experiments are carried out in the G0.8A game environment, and the results are summarized in Table 3.

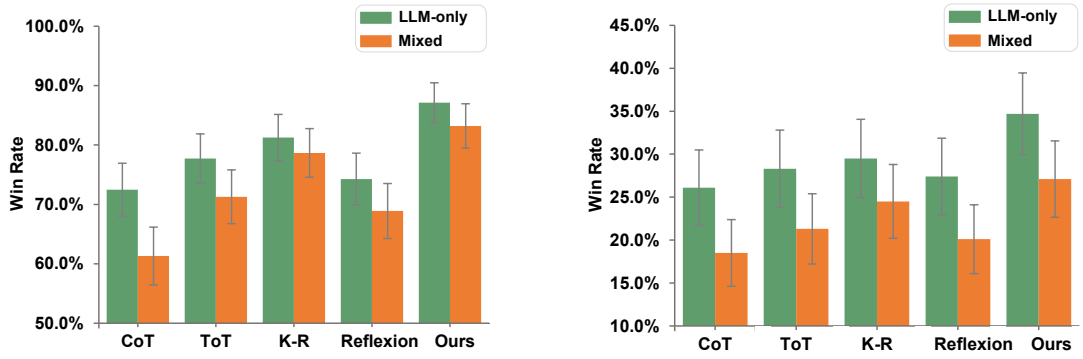
LLM-only: As the baseline setting, this variant involves no structured modeling. It yields the highest prediction deviation (43.0%) and the lowest win rate, indicating the limitations of relying solely on end-to-end language model reasoning without structural guidance.

+ Static Graph: When a static causal graph is introduced, consisting only of direct edges from observation nodes to the action node—the prediction deviation drops significantly, and win rate improves. This demonstrates that even a basic hypothesized reasoning structure can provide meaningful guidance for the LLM's inference process, improving both stability and directionality.

+ Intermediate Nodes: We then incorporate the mechanism for dynamically extracting intermediate variables from LLM-generated reflections. This leads to a substantial boost in both prediction accuracy and win rate, highlighting that the key to effective modeling lies not merely in surface-level observation-action mappings, but in uncovering intermediate reasoning steps that reflect the opponent's underlying decision process. These steps guide the LLM to reason in a structured, step-by-step manner.

+ Graph Refine: The addition of the graph refinement and pruning mechanism—based on reinforcement counts—helps retain reliable causal paths and eliminate spurious or hallucinated connections. This further stabilizes performance by reducing redundancy and preserving the most consistent decision logic.

+ Reasoning Examples (SOM): Finally, we enable the full SOM framework by adding the personalized example-pool mechanism. This module retrieves and leverages previously successful reasoning trajectories that are semantically similar to the current context, effectively simulating the structural equations $V_j = f_j(Pa(V_j))$ defined in the SCM framework. This step validates the central advantage of our two-stage design: while the causal graph captures the general structure of an opponent's decision logic, the reasoning examples instantiate the functional mappings within that structure, enabling highly personalized and adaptive opponent modeling. This results in the lowest prediction deviation (25.3%) and the highest win rate (0.61) among all variants.



(a) As Civilian in the game

(b) As Undercover in the game

Figure 4: Win Rate against different opponents in Undercover game. The performance of SOM and baseline methods is evaluated against LLM-only opponents and Mixed opponents. SOM consistently outperforms all baselines in both scenarios.

Table 4: Knowledge transfer test. Each agent plays against the same strong opponent: CoT (GPT-4o). SOM-T denotes a transferred SOM model originally constructed by a GPT-4o agent during its interaction with the CoT (GPT-4o) opponent.

Agent Variant	Prediction Deviation ↓ (%)	Win Rate ↑
GPT-4o + SOM	25.3	0.61
LLaMA-3-8B + CoT	76.1	0.07
LLaMA-3-8B + SOM-T	**45.8**	**0.31**
Mixtral-8B + CoT	95.6	0.02
Mixtral-8B + SOM-T	**65.2**	**0.27**

Overall, the ablation results clearly demonstrate that each component of SOM contributes significantly to its final performance. In particular, the introduction of intermediate variables and the use of personalized reasoning examples are critical to improving both predictive accuracy and strategic decision-making.

5.3.2 ***Structured Knowledge Transfer Analysis.*** One core advantage of SOM is its ability to construct structured opponent models that generalize across different LLMs. To validate this, we conduct a knowledge transfer experiment (Table 4), testing whether the SOM-generated model—comprising a causal graph $\mathcal{G}$ and an opponent-specific example pool $\mathcal{P}_{\text{opponent}}$—can be effectively reused by other agents.

We begin by allowing a strong agent (GPT-4o + SOM) to interact with a strong opponent (CoT-driven GPT-4o) and store the final constructed opponent model, including the causal graph and the reasoning examples. Then, we test two weaker open-source models (LLaMA-3-8B and Mixtral-8B) by directly loading this constructed model (denoted as **SOM-T**) and using it to play against the same CoT (GPT-4o) opponent.

As shown in Table 4, SOM-T substantially improves the performance of weaker models without any additional training. When LLaMA-3-8B uses the transferred SOM model, its prediction deviation falls from 76.1% to 45.8% and its win rate rises from 0.07 to 0.31. Mixtral-8B shows the same pattern. Importantly, the improvement is achieved by directly loading the SOM-based opponent model into the target agent, without fine-tuning the target LLM. This demonstrates that the structured representation learned by SOM encodes reusable behavioral regularities that benefit different model architectures. At the same time, transferred models do not fully close the gap to the high-capacity SOM instantiation, indicating that recipient model capacity and inference ability still constrain final performance. In short, SOM produces structured opponent knowledge that is transferable and practically useful across LLMs, while remaining complementary to improvements in base model capacity.

6 CONCLUSION

We introduce **SOM**, a novel two-stage opponent modeling framework inspired by structural causal modeling principles. By dynamically constructing and refining a structured reasoning graph, SOM explicitly decouples the process of opponent model construction from that of behavior prediction. Comprehensive experiments demonstrate that this structured reasoning approach substantially improves prediction accuracy and decision-making performance across diverse game-theoretic environments. Furthermore, the modular opponent models produced by SOM can be seamlessly transferred to empower other agents, highlighting its generality and practical utility.

Nevertheless, we acknowledge a limitation: the "causal" structures discovered by SOM represent functional dependencies inferred from observational data, rather than verified causal mechanisms of the opponent's cognition. Bridging this gap requires integrating more principled causal discovery techniques or controlled interaction settings in future work.

Overall, SOM offers a promising step toward building LLM-based agents that are more adaptive, interpretable, and robust in complex multi-agent environments.

ACKNOWLEDGMENTS

This work was supported by the National Science and Technology Major Project under Grant No. 2022ZD0116403, and in part by the Beijing Natural Science Foundation under Grant No. 4264131.

REFERENCES

[1] Maciej Besta, Nils Blach, Ales Kubicek, Robert Gerstenberger, Michal Podstawski, Lukas Gianinazzi, Joanna Gajda, Tomasz Lehmann, Hubert Niewiadomski, Piotr Nyczyk, et al. 2024. Graph of thoughts: Solving elaborate problems with large language models. In *Proceedings of the AAAI conference on artificial intelligence*, Vol. 38. 17682–17690.

[2] Federico Bianchi, Patrick John Chia, Mert Yuksekgonul, Jacopo Tagliabue, Dan Jurafsky, and James Zou. 2024. How well can llms negotiate? negotiationarena platform and analysis. *arXiv preprint arXiv:2402.05863* (2024).

[3] Micah Carroll, Rohin Shah, Mark K Ho, Tom Griffiths, Sanjit Seshia, Pieter Abbeel, and Anca Dragan. 2019. On the utility of learning about humans for human-ai coordination. *Advances in neural information processing systems* 32 (2019).

[4] Pei Chen, Boran Han, and Shuai Zhang. 2024. CoMM: Collaborative multi-agent, multi-reasoning-path prompting for complex problem solving. *arXiv preprint arXiv:2404.17729* (2024).

[5] Haobo Fu, Ye Tian, Hongxiang Yu, Weiming Liu, Shuang Wu, Jiechao Xiong, Ying Wen, Kai Li, Junliang Xing, Qiang Fu, et al. 2022. Greedy when sure and conservative when uncertain about the opponents. In *International Conference on Machine Learning*. PMLR, 6829–6848.

[6] Zhenyu Guan, Xiangyu Kong, Fangwei Zhong, and Yizhou Wang. 2024. Richelieu: Self-evolving llm-based agents for ai diplomacy. *Advances in Neural Information Processing Systems* 37 (2024), 123471–123497.

[7] Jiaxian Guo, Bo Yang, Paul Yoo, Bill Yuchen Lin, Yusuke Iwasawa, and Yutaka Matsuo. 2023. Suspicion-agent: Playing imperfect information games with theory of mind aware gpt-4. *arXiv preprint arXiv:2309.17277* (2023).

[8] John J Horton. 2023. *Large language models as simulated economic agents: What can we learn from homo silicus?* Technical Report. National Bureau of Economic Research.

[9] Yudong Hu, Congying Han, Haoran Li, and Tiande Guo. 2023. Modeling opponent learning in multiagent repeated games. *Applied Intelligence* 53, 13 (2023), 17194–17210.

[10] Shima Imani, Liang Du, and Harsh Shrivastava. 2023. Mathprompter: Mathematical reasoning using large language models. *arXiv preprint arXiv:2303.05398* (2023).

[11] Ziwei Ji, Nayeon Lee, Rita Frieske, Tiezheng Yu, Dan Su, Yan Xu, Etsuko Ishii, Ye Jin Bang, Andrea Madotto, and Pascale Fung. 2023. Survey of hallucination in natural language generation. *ACM computing surveys* 55, 12 (2023), 1–38.

[12] Yuheng Jing, Kai Li, Bingyun Liu, Haobo Fu, Qiang Fu, Junliang Xing, and Jian Cheng. 2026. An Open-Ended Learning Framework for Opponent Modeling. In *Proceedings of the AAAI Conference on Artificial Intelligence*, Vol. 39. 23222–23230.

[13] Dong Ki Kim, Miao Liu, Matthew D Riemer, Chuangchuang Sun, Marwa Abdulhai, Golnaz Habibi, Sebastian Lopez-Cot, Gerald Tesauro, and Jonathan How. 2021. A policy gradient algorithm for learning to learn in multiagent reinforcement learning. In *International Conference on Machine Learning*. PMLR, 5541–5550.

[14] Alain Ledoux. 1981. Concours résultats complets. *Les victimes se sont plu à jouer le* 14 (1981), 10–11.

[15] Nian Li, Chen Gao, Mingyu Li, Yong Li, and Qingmin Liao. 2024. Econagent: large language model-empowered agents for simulating macroeconomic activities. In *Proceedings of the 62nd Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*. 15523–15536.

[16] Naiqi Li, Peiyuan Liu, Zheng Liu, Tao Dai, Yong Jiang, and Shu-Tao Xia. 2025. Logic-of-Thought: Empowering Large Language Models with Logic Programs for Solving Puzzles in Natural Language. *arXiv preprint arXiv:2505.16114* (2025). https://arxiv.org/abs/2505.16114

[17] Nelson F Liu, Kevin Lin, John Hewitt, Ashwin Paranjape, Michele Bevilacqua, Fabio Petroni, and Percy Liang. 2023. Lost in the middle: How language models use long contexts. *arXiv preprint arXiv:2307.03172* (2023).

[18] Christopher Lu, Timon Willi, Christian A Schroeder De Witt, and Jakob Foerster. 2022. Model-free opponent shaping. In *International Conference on Machine Learning*. PMLR, 14398–14411.

[19] Shaoguang Mao, Yuzhe Cai, Yan Xia, Wenshan Wu, Xun Wang, Fengyi Wang, Tao Ge, and Furu Wei. 2024. ALYMPICS: LLM Agents Meet Game Theory–Exploring Strategic Decision-Making with AI Agents. *arXiv preprint arXiv:2311.03220* (2024).

[20] Meta Fundamental AI Research Diplomacy Team (FAIR), Anton Bakhtin, Noam Brown, Emily Dinan, Gabriele Farina, Colin Flaherty, Daniel Fried, Andrew Goff, Jonathan Gray, Hengyuan Hu, et al. 2022. Human-level play in the game of Diplomacy by combining language models with strategic reasoning. *Science* 378, 6624 (2022), 1067–1074.

[21] Samer Nashed and Shlomo Zilberstein. 2022. A survey of opponent modeling in adversarial domains. *Journal of Artificial Intelligence Research* 73 (2022), 277–327.

[22] Georgios Papoudakis, Filippos Christianos, and Stefano Albrecht. 2021. Agent modelling under partial observability for deep reinforcement learning. *Advances in Neural Information Processing Systems* 34 (2021), 19210–19222.

[23] J. Pearl. 2000. *Causality: Models, Reasoning, and Inference*. Cambridge University Press, New York, NY, USA.

[24] Sumedh Rasal. 2024. Llm harmony: Multi-agent communication for problem solving. *arXiv preprint arXiv:2401.01312* (2024).

[25] Noah Shinn, Federico Cassano, Ashwin Gopinath, Karthik Narasimhan, and Shunyu Yao. 2023. Reflexion: Language agents with verbal reinforcement learning. *Advances in Neural Information Processing Systems* 36 (2023), 8634–8652.

[26] Kai Sun, Yifan Ethan Xu, Hanwen Zha, Yue Liu, and Xin Luna Dong. 2023. Head-to-tail: How knowledgeable are large language models (llm)? AKA will llms replace knowledge graphs? *arXiv preprint arXiv:2308.10168* (2023).

[27] William Vickrey. 1961. Counterspeculation, auctions, and competitive sealed tenders. *The Journal of finance* 16, 1 (1961), 8–37.

[28] Xuezhi Wang, Jason Wei, Dale Schuurmans, Quoc Le, Ed Chi, Sharan Narang, Aakanksha Chowdhery, and Denny Zhou. 2022. Self-Consistency Improves Chain of Thought Reasoning in Language Models. *arXiv preprint arXiv:2203.11171* (2022). https://arxiv.org/abs/2203.11171

[29] Jason Wei, Xuezhi Wang, Dale Schuurmans, Maarten Bosma, Fei Xia, Ed Chi, Quoc V Le, Denny Zhou, et al. 2022. Chain-of-thought prompting elicits reasoning in large language models. *Advances in neural information processing systems* 35 (2022), 24824–24837.

[30] Shuang Wu, Liwen Zhu, Tao Yang, Shiwei Xu, Qiang Fu, Yang Wei, and Haobo Fu. 2024. Enhance reasoning for large language models in the game werewolf. *arXiv preprint arXiv:2402.02330* (2024).

[31] Zhe Wu, Kai Li, Hang Xu, Yifan Zang, Bo An, and Junliang Xing. 2022. L2E: Learning to exploit your opponent. In *2022 International Joint Conference on Neural Networks (IJCNN)*. IEEE, 1–8.

[32] Lin Xu, Zhiyuan Hu, Daquan Zhou, Hongyu Ren, Zhen Dong, Kurt Keutzer, See Kiong Ng, and Jiashi Feng. 2023. Magic: Investigation of large language model powered multi-agent in cognition, adaptability, rationality and collaboration. *arXiv preprint arXiv:2311.08562* (2023).

[33] Yuzhuang Xu, Shuo Wang, Peng Li, Fuwen Luo, Xiaolong Wang, Weidong Liu, and Yang Liu. 2023. Exploring large language models for communication games: An empirical study on werewolf. *arXiv preprint arXiv:2309.04658* (2023).

[34] Zelai Xu, Chao Yu, Fei Fang, Yu Wang, and Yi Wu. 2023. Language agents with reinforcement learning for strategic play in the werewolf game. *arXiv preprint arXiv:2310.18940* (2023).

[35] Likun Yang, Pei Xu, Shiyue Cao, Yongjian Ren, Xiaotang Chen, and Kaiqi Huang. 2025. Uncertainty-Aware Opponent Modeling for Deep Reinforcement Learning. In *Proceedings of the 24th International Conference on Autonomous Agents and Multiagent Systems*. 2217–2225.

[36] Yaodong Yang and Jun Wang. 2020. An overview of multi-agent reinforcement learning from game theoretical perspective. *arXiv preprint arXiv:2011.00583* (2020).

[37] Shunyu Yao, Dian Yu, Jeffrey Zhao, Izhak Shafran, Tom Griffiths, Yuan Cao, and Karthik Narasimhan. 2023. Tree of thoughts: Deliberate problem solving with large language models. *Advances in neural information processing systems* 36 (2023), 11809–11822.

[38] Xiaopeng Yu, Jiechuan Jiang, Wanpeng Zhang, Haobin Jiang, and Zongqing Lu. 2022. Model-based opponent modeling. *Advances in Neural Information Processing Systems* 35 (2022), 28208–28221.

[39] XiaoPeng Yu, Wanpeng Zhang, and Zongqing Lu. 2025. LLM-Based Explicit Models of Opponents for Multi-Agent Games. In *Proceedings of the 2025 Conference of the Nations of the Americas Chapter of the Association for Computational Linguistics: Human Language Technologies (Volume 1: Long Papers)*. 892–911.

[40] Ceyao Zhang, Kaijie Yang, Siyi Hu, Zihao Wang, Guanghe Li, Yihang Sun, Cheng Zhang, Zhaowei Zhang, Anji Liu, Song-Chun Zhu, et al. 2024. Proagent: building proactive cooperative agents with large language models. In *Proceedings of the AAAI Conference on Artificial Intelligence*, Vol. 38. 17591–17599.

[41] Yilun Zhang, Yujun Cai, Yifei Li, and Yaodong Yang. 2024. On the Diagram of Thought. *arXiv preprint arXiv:2409.10038* (2024). https://arxiv.org/abs/2409.10038

[42] Yadong Zhang, Shaoguang Mao, Tao Ge, Xun Wang, Adrian de Wynter, Yan Xia, Wenshan Wu, Ting Song, Man Lan, and Furu Wei. 2024. LLM as a Mastermind: A Survey of Strategic Reasoning with Large Language Models. *arXiv preprint arXiv:2404.01230* (2024).

[43] Yadong Zhang, Shaoguang Mao, Tao Ge, Xun Wang, Yan Xia, Man Lan, and Furu Wei. 2024. K-Level Reasoning: Establishing Higher Order Beliefs in Large Language Models for Strategic Reasoning. *arXiv preprint arXiv:2402.01521* (2024).

[44] Luisa Zintgraf, Sam Devlin, Kamil Ciosek, Shimon Whiteson, and Katja Hofmann. 2021. Deep interactive bayesian reinforcement learning via meta-learning. *arXiv preprint arXiv:2101.03864* (2021).