

# BRIDGING THE GAP BETWEEN THEORY OF MIND AND ACTION IN LLMs

Sehyeok Kang Jihwan Oh Se-Young Yun

KAIST AI

{kangsehyeok0329, ericoh929, yunseyoung}@kaist.ac.kr

## ABSTRACT

While Large Language Models (LLMs) demonstrate impressive capabilities in Theory of Mind (ToM) tasks, translating this social understanding into strategic action remains a critical challenge. In this paper, we address the *Knowing Doing Gap* in social intelligence, where models correctly infer the mental states of others yet fail to select actions consistent with those inferences. To rigorously evaluate this discrepancy, we introduce the **Theory of Mind-to-Action (ToMA)** benchmark, a unified framework designed to analyze the causal link between social reasoning and decision making in diverse cooperative and competitive scenarios. Our empirical analysis reveals that reasoning capabilities do not guarantee strategic behavior, identifying specific failure modes where reasoning decouples from action. To bridge this gap, we propose alignment mechanisms that enforce a sequential dependency between perspective taking and execution. By prioritizing the explicit analysis of others' mental states before decision making, social reasoning strictly conditions strategic actions.

## 1 INTRODUCTION

Large Language Models (LLMs) have recently achieved striking advances in complex reasoning and knowledge-intensive tasks OpenAI (2023); Chowdhery et al. (2022); Team (2023); DeepSeek-AI et al. (2025). Alongside these gains, a growing body of work suggests that LLMs exhibit Theory of Mind (ToM) Premack & Woodruff (1978); Baron-Cohen et al. (1985) capabilities, defined as the ability to infer the beliefs, intentions, and mental states of others Strachan et al. (2024); Kosinski (2024); Chen et al. (2024); Marchetti et al. (2023).

Despite these strides, recent studies reveal a critical failure mode known as the *Knowing-Doing Gap* Schmied et al. (2025); Paglieri et al. (2024); Ruoss et al. (2024); Wong et al. (2025). Even when a model derives the correct inference, it often selects an action that contradicts that inference during decision making. This inconsistency undermines confidence in the reliability of LLM decisions and raises questions about their robustness, thereby limiting their potential as effective autonomous agents.

We specifically address this issue in the context of ToM, where inferred mental states must directly guide strategic choices. ToM is more than a passive diagnostic ability; it serves as a prerequisite for strategic action in social environments. It enables agents to coordinate toward shared goals in cooperative environments and to anticipate and counter opponents in competitive settings.

However, LLMs frequently fail to translate accurate social inferences (*Knowing*) into appropriate behavioral choices (*Doing*). As illustrated in the one-way communication scenario in Figure 1, Bob often remains silent despite correctly predicting that Alice awaits a response, or conversely, speaks when predicting she

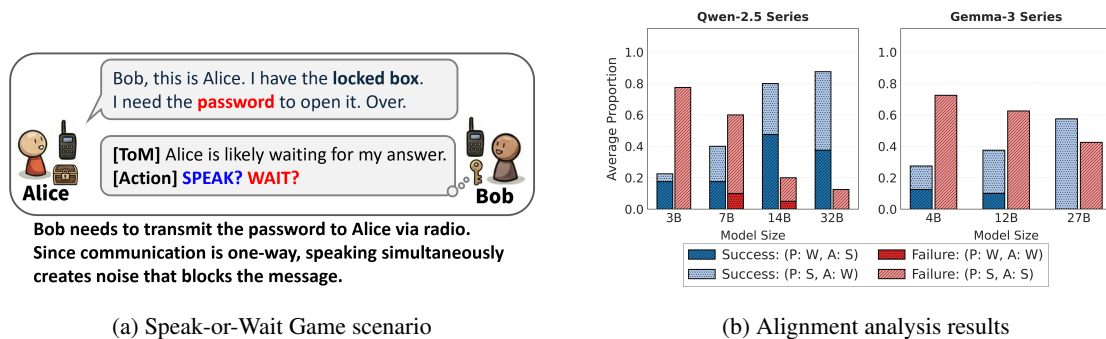


Figure 1: ToMA alignment evaluation example. (a) Example scenario of the Speak-or-Wait Game. (b) Alignment analysis between ToM predictions ( $P$ ) and agent actions ( $A$ ). Blue and red bars represent successful and failed alignment, respectively.  $W$  (Wait) and  $S$  (Speak) denote specific behaviors, such as an agent choosing to speak ( $A : S$ ) when predicting the partner will wait ( $P : W$ ).

will not wait. This confirms that the *Knowing-Doing Gap* extends to social intelligence, revealing a critical disconnect between mental state reasoning and strategic action.

While prior research Riemer et al. (2025) raised concerns about inconsistencies between ToM reasoning and action, it relies on toy cases and lacks rigorous validation. Furthermore, as highlighted by Riemer et al. (2025), mainstream ToM benchmarks focus on attributing mental states or predicting behavior of *other agents*, overlooking the agent’s *own* choices. Consequently, the field lacks a systematic understanding of whether LLMs leverage social reasoning to guide their actions.

**Contributions.** To bridge the gap between ToM reasoning and behavior, this paper establishes a framework evaluating how social understanding translates into strategic action. Beyond assessment, we propose alignment mechanisms that force ToM insights to inform strategic decision making. Our main contributions follow:

- **ToMA Benchmark:** We introduce the Theory of Mind-to-Action (ToMA) benchmark designed to quantitatively analyze the causal link between ToM inference and decision making across diverse cooperative and competitive scenarios.
- **Diagnostic Insights:** We identify specific failure modes where ToM reasoning is decoupled from action, offering critical insights into the limitations of current LLMs in social intelligence.
- **Methodology for Alignment:** We propose **ToM-CoT** and **ToM-Guide** to mitigate the Knowing-Doing Gap by aligning ToM insights with action policies. ToM-CoT enforces perspective-taking within structured reasoning, while ToM-Guide employs a multi-turn pipeline strictly conditioning action on pre-generated analysis.

## 2 THEORY OF MIND-TO-ACTION (TOMA)

### 2.1 OVERVIEW

We introduce the **Theory of Mind to Action (ToMA)** benchmark, a unified framework evaluating whether LLMs translate social reasoning into strategic action. Unlike previous benchmarks treating ToM as a standalone classification task, ToMA ensures that solving the task remains strictly contingent on translating ToM inferences into action, thereby directly evaluating whether the model bridges social reasoning and behavior.

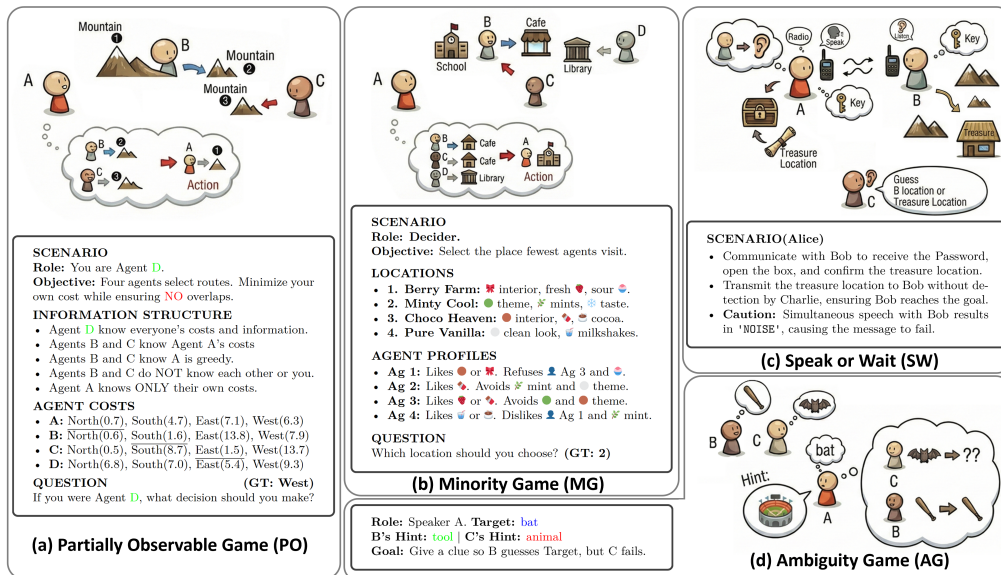


Figure 2: Description of the four scenarios in ToMA. The figures and scenarios are distinct.

The benchmark comprises four scenarios categorized into cooperative and competitive settings, as well as single-turn and multi-turn interactions.

## 2.2 DESIGN PHILOSOPHY & DISTINCTIONS

While existing ToM benchmarks are limited to assessing the *presence* of social reasoning capabilities, ToMA shifts the focus to the *translation* of this reasoning into behavior. We distinguish our approach through four key design principles:

- **Action oriented strategic evaluation:** We shift the paradigm from passive mental state attribution (e.g., "What does Alice think?") to active strategic execution (e.g., "Given what Alice thinks, what will you do?"). This approach tests the alignment between social understanding and actual strategic behavior.
- **Deterministic causal dependency:** To ensure a rigorous link between reasoning and behavior, we design scenarios where correct ToM inference is the necessary and sufficient condition for the unique optimal action under conditions. This ensures that success reflects alignment rather than coincidental completion.
- **Complex multi-agent modeling:** ToMA simulates real world social dynamics involving information asymmetry, strategic utility, and hidden agendas. Agents must synthesize divergent perspectives from partial observability to navigate interactions, such as coordinating with allies or deceiving adversaries.

## 2.3 SCENARIO DESCRIPTIONS

ToMA consists of four distinct games, as illustrated in Figure 2, each tailored to evaluate a specific facet of social reasoning.

- **Partially Observable Game (PO):** This scenario centers on information asymmetry where selecting unique destinations others avoid requires the agent to link ToM reasoning directly to behavior. Identify-

ing the optimal action remains impossible unless the agent translates perspective modeling into execution (Appendix C.1).

- **Minority Game (MG):** This game involves preference based reasoning where identifying minority options depends strictly on translating social preference analysis into action. Success ensures the agent converts its understanding of others’ likes into a final decision (Appendix C.2).
- **Speak or Wait (SW):** This multi-turn game requires the agent to lead allies to the target location via precise strategic communication while keeping adversaries strictly unaware of the final objective. Achieving the goal depends on applying ToM to avoid communication overlap with teammates and coordinate optimal timing while evading enemy detection (Appendix C.3).
- **Ambiguity Game (AG):** This multi-turn game focuses on strategic execution in mixed motive settings. The agent must evaluate the mental states of both allies and adversaries to provide hints that selectively guide teammates while misleading enemies (Appendix C.4).

## 2.4 EVALUATION METRICS

To quantify the gap between reasoning and action, we define three primary metrics based on the model’s performance over  $N$  scenarios:

- **ToM Accuracy ( $Acc_{ToM}$ ):** This metric evaluates the model’s social reasoning capacity (*Knowing*). It measures the frequency with which the model correctly infers ground-truth mental states:

$$Acc_{ToM} = \frac{1}{N} \sum_{i=1}^N \mathbb{I}(\hat{M}_i = M_i) \quad (1)$$

where  $\hat{M}_i$  is the inferred mental state and  $M_i$  is the ground truth for the  $i$ -th scenario.

- **Action Consistency ( $C_{act}$ ):** This metric assesses logical alignment between ToM reasoning and execution (*Doing*). It measures the frequency with which the model’s chosen action  $A_i$  conforms to the valid action set  $\mathcal{A}_{\hat{M}_i}$  derived from ToM inference. For example, if the model predicts a teammate will `WAIT` in `SW`, the set  $\mathcal{A}_{\hat{M}_i}$  contains only `SPEAK` to ensure coordination. Selecting `SPEAK` as  $A_i$  increases the  $C_{act}$  value. The formula for  $C_{act}$  follows:

$$C_{act} = \frac{1}{N} \sum_{i=1}^N \mathbb{I}(A_i \in \mathcal{A}_{\hat{M}_i}) \quad (2)$$

- **Task Success Rate ( $Acc_{task}$ ):** This metric represents the overall problem-solving performance. In scenarios like `PO` and `MG`, success is achieved only when the model correctly executes the optimal action  $A_i^*$  based on accurate ToM:

$$Acc_{task} = \frac{1}{N} \sum_{i=1}^N \mathbb{I}(A_i = A_i^*) \quad (3)$$

## 3 ENHANCING TOM-TO-ACTION ALIGNMENT

### 3.1 THE TOM KNOWING-DOING GAP

We identify a distinct manifestation of the *Knowing Doing Gap* within social reasoning: *the ToM Knowing Doing Gap*. Our analysis reveals that even when models demonstrate the cognitive capacity to attribute mental states, they frequently fail to translate these insights into strategic execution. A deficiency in Action Consistency ( $C_{act}$ ) captures this gap quantitatively, revealing a frequent breakdown in the logical transition from social inference to behavioral output. Figure 3 illustrates the gap between ToM and action. As Chain

## An Example of the ToM Knowing-Doing Gap

## ❶ CoT

- **Reasoning:** Agent A will choose North (0.7). If A chooses North, Agent B will choose South (1.6), and Agent C will choose East (1.5). In this case (A:North, B:South, C:East), **Agent D should choose East (5.4)**, as it is the lowest cost option among the available choices.

- **Answer:** **East ✗**

## ❷ ToM-CoT

## • Reasoning:

- **Phase 1:** A picks North. B sees A-North, so B picks South. C sees A-North, so C picks East.
- **Phase 2:** North, South, and East are occupied. The only remaining valid option for me is **West**.

- **Answer:** **West ✓**

Figure 3: Model responses illustrating the ToM knowing doing gap for the Figure 2 (a) scenario.

of Thought (Wei et al. (2022)) results demonstrate, the model successfully infers where another agent will go. However, the model ultimately fails the task because it chooses an action inconsistent with its own ToM reasoning. To bridge this knowing-doing gap, we propose two alignment strategies: ToM-CoT and ToM-Guide. As the figure shows, ToM-CoT effectively resolves this discrepancy, ensuring model actions align with internal social inferences.

## 3.2 ToM-CoT

To strengthen the logical link between inference and action, we introduce ToM-CoT. Unlike standard CoT, ToM-CoT mandates a structured, two-stage sequential reasoning process:

$$P(A | X) \approx P(\hat{M} | X) \cdot P(A | \hat{M}, X) \quad (4)$$

where  $X$  represents the context,  $\hat{M}$  denotes the explicitly articulated mental states, and  $A$  is the resulting action. As illustrated in Figure 4, this process structures the model’s internal monologue into two distinct steps:

1. **ToM Reasoning:** The model infers beliefs, desires, and information constraints of all agents, generating a structured social inference  $\hat{M}$ .
2. **ToM-Guided Action:** The model derives the optimal action  $A$  by conditioning its decision-making process on the previously reasoned  $\hat{M}$ .

By forcing the model to verbalize the path from ”knowing others” to ”deciding for oneself”, ToM-CoT ensures that social insights serve as the functional foundation for behavioral decisions. This internal alignment reduces the likelihood of the model ignoring its own reasoning during the final execution phase.

## 3.3 ToM-GUIDE

While ToM-CoT internalizes the reasoning path within a single prompt, ToM-Guide externalizes the transition from reasoning to action through a two-step prompting pipeline. This approach ensures that the

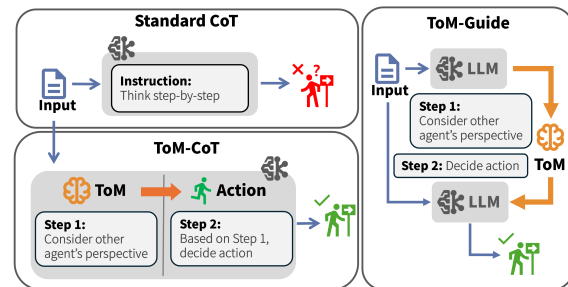


Figure 4: ToM driven architectures designed to improve consistency between reasoning and action

model’s behavioral policy is explicitly conditioned on its previous social inferences. In the first phase, the model identifies the mental states  $\hat{M}$  of the involved agents. In the second phase, the model receives both the original context  $X$  and its own inferred mental states  $\hat{M}$  as augmented inputs to determine the final action:

$$A_{final} = \arg \max_{a \in \mathcal{A}} P(a | X, \hat{M}) \quad (5)$$

As illustrated in Figure 4, ToM-Guide forces the model to treat its social understanding as a functional premise for decision-making. By re-feeding  $\hat{M}$  into the decision-making stage, ToM-Guide effectively bridges *the ToM Knowing-Doing Gap*, ensuring that the final action aligns with the previously articulated social reasoning.

## 4 EXPERIMENTS

In this section, we analyze the *ToM Knowing-Doing Gap* by examining the discrepancy between ToM inference and execution. We then evaluate how effectively ToM-CoT and ToM-Guide bridge this gap across strategic environments. Our experiments address the following research questions:

---

**RQ1: The ToM Knowing-Doing Gap.** To what extent does a discrepancy exist between a model’s ToM inference and its actual behavioral execution?

**RQ2: Strategy Effectiveness.** How effectively do our proposed strategies bridge the identified gap to improve strategic performance?

**RQ3: Impact of ToM Complexity.** How does the effectiveness of our proposed strategies vary across different levels of recursive social reasoning?

---

**Models** We evaluate our methods across four distinct strategic scenarios within our proposed ToMA benchmark. We test various state-of-the-art LLMs to ensure generalizability, including the Qwen-2.5 series (Qwen et al. (2025)), Gemma-3 series (Team et al. (2025)), GPT-4o-mini (OpenAI (2023)), GPT-5.1 (OpenAI (2025a)), Gemini-2.0-flash (Team (2023)), and Gemini-2.5-flash-lite (Comanici et al. (2025)). For these models, we operate them in instant mode to measure their direct response capabilities. Additionally, we include o4-mini (OpenAI (2025b)), which we execute in thinking mode to assess the impact of native long reasoning on strategic performance. We run all models five times and average their performance.

**Baselines** To benchmark the performance of ToM-CoT and ToM-Guide, we compare them against three primary baselines: **Zero-shot**, where the model generates a final action  $A$  directly from the context  $X$ ; **CoT**, which employs standard step-by-step reasoning without explicit ToM mandates; and **Self-consistency (SC)** (Wang et al. (2022)), for which we perform 10 independent iterations and determine the final output through majority voting. For implementation, ToM-CoT enforces a two-stage internal monologue (ToM Reasoning followed by Action) in a single turn, while ToM-Guide utilizes a two-step pipeline that re-feeds the model’s own inference as explicit context for the final decision.

### 4.1 QUANTIFYING THE ToM-ACTION GAP (RQ1)

We first analyze the disconnect between ToM inference and strategic execution by examining two primary factors: model scale and ToM accuracy. As defined in Section 3.1, LLMs often lack the capacity to translate ToM insights into corresponding actions, a phenomenon we quantify through Action Consistency ( $C_{act}$ ).

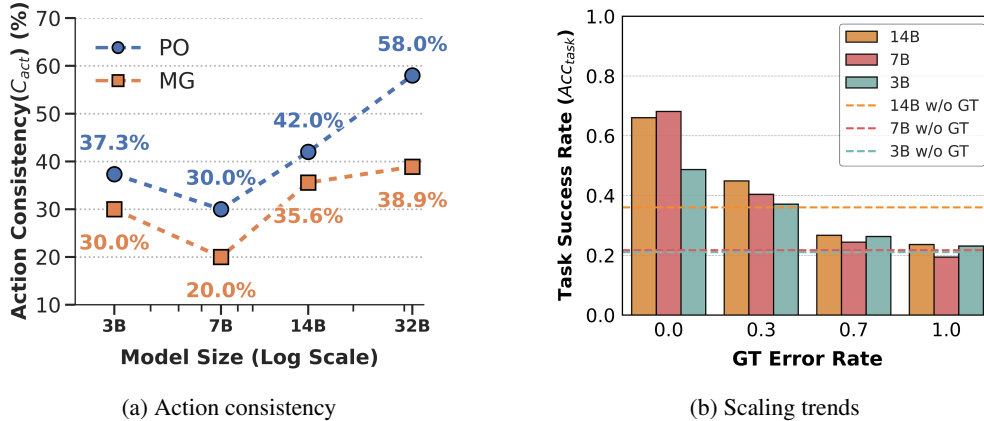


Figure 5: Analysis of the Qwen 2.5 series performance. (a) Action consistency in Partially Observable and Minority Games. (b) Scaling trends in average task success rates with ground truth ToM injection.

Figure 5a presents  $C_{act}$  for the PO and MG environments, while Figure 7a illustrates results for the SW game. Across these environments, we observe that consistency remains notably low, staying below 60% for most models with the exceptions of Qwen-2.5-14B and 32B.

**Impact of Model Scale** When examining the impact of model scale, we find that larger models generally exhibit higher alignment between ToM and action. In PO, MG, and SW,  $C_{act}$  consistently improves as the parameter count increases. This trend suggests that as underlying reasoning performance matures in larger models, their capacity to maintain consistency between internal social inferences and behavioral outputs also strengthens.

**Influence of ToM Accuracy** We investigate how social reasoning accuracy affects performance via a ground truth (GT) injection experiment (Figure 5b). In this setup, we provide ground truth ToM values as input to evaluate whether the model follows the correct social reasoning path. Ideally, seamless social knowledge translation should yield a 1.0 success rate at zero inference error. Even with perfect information, action accuracy fails to reach this threshold. This shortfall proves ToM inferences do not fully govern decision making, highlighting a disconnect in the knowing doing link.

Nevertheless, we observe a correlation between action accuracy and ToM accuracy. Declining action performance as GT error rates increase suggests that injecting ToM results into inputs influences model behavior, although this influence remains partial. Furthermore, compared to the baseline, providing explicit social inferences narrows the reasoning action gap when the GT error rate falls below 0.7, corresponding to ToM accuracy above 0.3. These findings imply that strategies like ToM Guide help models anchor decisions in social reasoning. Therefore, when a model has sufficient ToM accuracy, providing explicit social inferences links internal knowledge to strategic decisions.

## 4.2 STRATEGIC PERFORMANCE GAINS (RQ2)

As shown in Table 1 and Table 2, our proposed strategies significantly outperform all baselines across the evaluated models. We observe that the proposed models with higher ToM accuracy ( $Acc_{ToM}$ ) exhibit more pronounced performance gains. This correlation suggests that our methods effectively unlock the latent social reasoning capabilities of the models and channel them into actionable strategies. For instance, in MG, GPT-5.1 achieves a performance boost of +44%p, demonstrating that high social understanding only leads to strategic success when explicitly bridged to an action policy.

Table 1: Comparison of task success rates ( $Acc_{task}$ ) across LLM models and reasoning strategies in Partially Observable Games (PO). **Bold** indicates the best performance and underlining denotes the second best.

Type	Model	$Acc_{ToM}$	Baselines ( $Acc_{task}$ )			Ours ( $Acc_{task}$ )	
			Zero-shot	SC	CoT	ToM-Guide	ToM-CoT
Open	Qwen-2.5-3B-Instruction	0.327	0.283	0.284 (+0.00)	0.237 (-0.05)	0.292(+0.01)	<b>0.301</b> (+0.02)
	Qwen-2.5-7B-Instruction	0.378	0.258	0.224 (-0.03)	0.253 (-0.01)	0.279(+0.02)	<b>0.365</b> (+0.11)
	Qwen-2.5-14B-Instruction	0.442	0.211	0.213 (+0.00)	0.367 (+0.16)	0.404(+0.19)	<b>0.495</b> (+0.28)
	Qwen-2.5-32B-Instruction	0.487	0.284	0.289 (+0.01)	0.447 (+0.16)	<u>0.462</u> (+0.18)	<b>0.512</b> (+0.23)
	gemma-3-4b-it	0.323	0.255	0.227 (-0.03)	0.255 (+0.00)	<b>0.338</b> (+0.08)	0.300(+0.05)
	gemma-3-12b-it	0.364	0.294	0.299 (+0.01)	0.353 (+0.06)	0.407(+0.11)	<b>0.432</b> (+0.14)
	gemma-3-27b-it	0.427	0.262	0.263 (+0.00)	0.417 (+0.16)	<u>0.448</u> (+0.19)	<b>0.475</b> (+0.21)
Closed	Gemini-2.0-flash	0.507	0.410	0.441 (+0.03)	0.541 (+0.13)	<b>0.626</b> (+0.22)	0.565(+0.16)
	Gemini-2.5-flash-lite	0.511	0.394	0.433 (+0.04)	0.330 (-0.06)	<b>0.629</b> (+0.24)	0.556(+0.16)
	GPT-4o-mini	0.453	0.186	0.189 (+0.00)	0.301 (+0.12)	0.420(+0.23)	<b>0.464</b> (+0.28)
	GPT-5.1	0.604	0.301	0.309 (+0.01)	0.312 (+0.01)	<u>0.639</u> (+0.34)	<b>0.709</b> (+0.41)
	o4-mini (think)	-	0.750	0.824	-	-	-

Table 2: Comparison of task success rates ( $Acc_{task}$ ) across LLM models and reasoning strategies in Minority Games (MG). **Bold** indicates the best performance and underlining denotes the second best.

Type	Model	$Acc_{ToM}$	Baselines ( $Acc_{task}$ )			Ours ( $Acc_{task}$ )	
			Zero-shot	SC	CoT	ToM-Guide	ToM-CoT
Open	Qwen-2.5-3B-Instruction	0.459	0.216	0.235 (+0.02)	0.183 (-0.03)	<b>0.276</b> (+0.06)	0.258(+0.04)
	Qwen-2.5-7B-Instruction	0.519	0.152	0.152 (+0.00)	0.181 (+0.03)	0.318(+0.17)	<b>0.369</b> (+0.22)
	Qwen-2.5-14B-Instruction	0.631	0.212	0.229 (+0.02)	0.353 (+0.14)	0.420(+0.21)	<b>0.468</b> (+0.26)
	Qwen-2.5-32B-Instruction	0.672	0.226	0.222 (-0.00)	0.391 (+0.17)	<b>0.547</b> (+0.32)	0.494(+0.27)
	gemma-3-4b-it	0.470	0.180	0.195 (+0.02)	0.241 (+0.06)	0.300(+0.12)	<b>0.342</b> (+0.16)
	gemma-3-12b-it	0.541	0.211	0.211 (+0.00)	0.366 (+0.16)	<b>0.589</b> (+0.38)	0.420(+0.21)
	gemma-3-27b-it	0.528	0.207	0.218 (+0.01)	0.343 (+0.14)	<b>0.521</b> (+0.31)	0.476(+0.27)
Closed	Gemini-2.0-flash	0.619	0.440	0.474 (+0.03)	0.409 (-0.03)	<b>0.660</b> (+0.22)	0.578(+0.14)
	Gemini-2.5-flash-lite	0.628	0.278	0.198 (-0.08)	0.206 (-0.07)	0.347(+0.07)	<b>0.493</b> (+0.22)
	GPT-4o-mini	0.554	0.168	0.182 (+0.01)	0.296 (+0.13)	0.482(+0.31)	<b>0.591</b> (+0.42)
	GPT-5.1	0.683	0.260	0.285 (+0.03)	0.313 (+0.05)	<b>0.700</b> (+0.44)	0.373(+0.11)
	o4-mini (think)	-	0.554	0.674	-	-	-

In the SW game, our approach maintains higher Action Consistency ( $C_{act}$ ) compared to the CoT baseline, which often fails to link its reasoning to behavioral output (Figure 6a). Figure 6b further illustrates this impact through higher Box Open Rates and Win Rates. The increase in Box Open Rates indicates that our methods facilitate successful communication by aligning actions with social inferences about allies. Furthermore, the improved Win Rates imply that models successfully incorporate ToM reasoning about adversaries into their execution, allowing them to remain undetected while pursuing strategic goals.

Table 3: Comparison of task success rates ( $Acc_{task}$ ) on the RecToM *Recommender-prediction* scenario (Li et al., 2025), designed to evaluate ToM-to-action capabilities, and the IPD task by Riemer et al. (2025).

Task	Model	Baselines ( $Acc_{task}$ )		Strategies ( $Acc_{task}$ )	
		Zero-shot	CoT	ToM-CoT	ToM-Guide
RecToM	Gemma-3-12b-it	0.147	0.132 (-0.02)	<b>0.178</b> (+0.03)	0.165(+0.02)
	Qwen2.5-14B-Instruct	0.161	0.144 (-0.02)	<b>0.210</b> (+0.05)	0.189(+0.03)
	gemini-2.5-flash-lite	0.099	0.109 (+0.01)	<b>0.239</b> (+0.14)	0.187(+0.09)
IPD	Qwen2.5-3B-Instruct	0.000	0.355 (+0.36)	<u>0.386</u> (+0.39)	<b>0.507</b> (+0.51)

We observe similar trends in the Ambiguity Game (AG), as presented in Figure 7a. In this scenario, ToM-Guide and ToM-CoT consistently achieve higher win rates compared to the Zero-shot and CoT baselines. Notably, our strategies maintain a relatively low rate of Leakage and Loss, cases where the adversary exploits leaked information. The results indicate that our approach effectively translates dual social reasoning, involving cooperation with allies and concealment of intentions from opponents, into precise strategic actions.

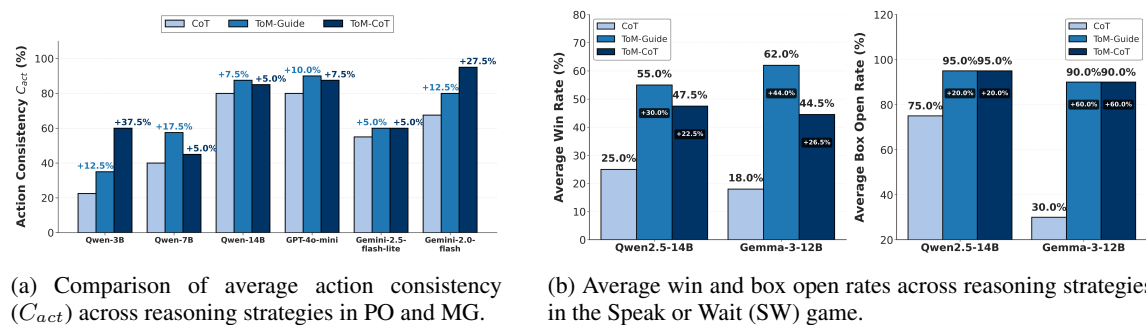


Figure 6: Performance comparison across different reasoning strategies.

**Generality** We evaluate whether ToM-CoT and ToM-Guide generalize their ability to bridge the ToM to action gap beyond our proposed benchmark. To this end, we conduct evaluations using existing datasets designed to measure the gap between reasoning and behavior. Table 3 presents the results across these two benchmarks. The experimental results demonstrate that ToM-CoT and ToM-Guide achieve superior performance compared to both Zero shot and CoT.

### 4.3 SENSITIVITY TO REASONING LEVELS (RQ3)

We evaluate our strategies across various levels of recursive ToM reasoning, from non-social tasks to complex higher-order inferences. Figure 7b illustrates the performance breakdown by ToM level.

In Level 0 (L0) scenarios, where success depends on maximizing individual utility without considering others, we find that the Zero shot baseline achieves highest accuracy. In this context, our strategies underperform because forcing social reasoning introduces unnecessary complexity to simple optimization problems.

However, as tasks demand recursive Theory of Mind (Level 1 and above), this trend reverses. While the CoT baseline outperforms Zero shot by introducing reasoning steps, it consistently falls short of our methods. Notably, as the ToM level increases towards Level 4, representing a fully observable and complex setting, the performance gap between our strategies and baselines widens. This implies that while standard prompting methods struggle with high order social dynamics, our approach successfully bridges the knowing doing gap. Thus, connecting ToM to action becomes vital as strategic complexity increases, and our approach effectively maintains this connection.

## 5 RELATED WORKS

**ToM in LLMs and Benchmarks** Theory of Mind refers to the ability to attribute beliefs, intentions, and informational constraints to others (Premack & Woodruff, 1978; Baron-Cohen et al., 1985). Early evaluations of LLMs often relied on adaptations of the classic Sally-Anne false-belief test to probe whether models could attribute mental states (Kosinski, 2024; Strachan et al., 2024). To enable more systematic assessment, researchers have developed specialized benchmarks that extend beyond these simple diagnostic tasks. Le et al. (2019) introduced **ToMi**, utilizing simulated stories to test latent belief tracking. Building on this foundation, **OpenToM** (Xu et al., 2024) and **FANToM** (Kim et al., 2023) assess ToM in more complex narratives and conversational settings characterized by information asymmetry. Furthermore, **Hi-ToM** (He et al., 2023) explicitly targets higher-order recursive reasoning capabilities. Most recently, **ToMBench** (Chen et al., 2024) offers a unified framework to evaluate these diverse social reasoning skills. However, these benchmarks predominantly treat ToM as a passive diagnostic capability. They evaluate whether a model

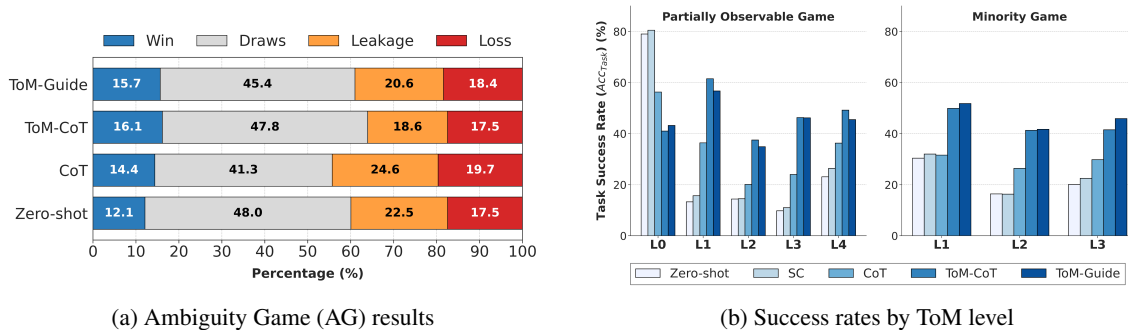


Figure 7: Performance comparison across different game settings. (a) Comparison of execution results across reasoning strategies in the Ambiguity Game (AG). Refer to Figure 11 for specific results of each LLM model. (b) Comparison of task success rates ( $Acc_{task}$ ) across strategies categorized by ToM level in Partially Observable and Minority Games.

infers what an agent believes but do not test whether the model uses these inferred mental states to select actions.

**The Knowing-Doing Gap in Reasoning** Beyond social reasoning, recent studies focus on a broad *Knowing-Doing Gap* where models articulate rationales yet fail to act. Turpin et al. (2023) reveal that generated chain-of-thought explanations often misrepresent true decision making, proving that reasoning does not guarantee execution. Similarly, Pal et al. (2025) highlight in *Incoherent Beliefs & Inconsistent Actions* that models harbor correct latent knowledge but contradict it during decision making. This disconnect extends to ethical alignment; Shen et al. (2025) demonstrate in *Mind the Value-Action Gap* that LLMs claiming specific values frequently violate them in practice.

**The ToM Knowing-Doing Gap** Despite these insights, relatively few studies directly examine the translation of ToM inferences into behavioral choices. Notably, Riemer et al. (2025) argue that existing benchmarks primarily assess the mere possession of ToM capabilities rather than their connection to behavior, highlighting a functional gap where social knowledge fails to guide decision-making. Furthermore, while multi-agent games like *Werewolf* or *Diplomacy* implicitly require social reasoning (Xu et al., 2023; Park et al., 2023), they are unsuitable for evaluating this specific mechanism. In such complex environments, outcomes depend on multifaceted factors, making it difficult to determine whether an action resulted strictly from ToM reasoning or other confounding variables. Therefore, to rigorously assess whether ToM translates into action, we require a dedicated benchmark where the optimal choice depends exclusively on accurate social inference.

## 6 CONCLUSION

In this paper, we address *the ToM Knowing-Doing Gap* where LLMs fail to translate Theory of Mind capabilities into strategic actions. To quantify this disconnect, we introduce the Theory of Mind-to-Action (ToMA) benchmark. Analysis reveals that reasoning capabilities do not guarantee strategic behavior, as models frequently decouple actions from their own social inferences. To bridge this gap, we propose ToM-CoT and ToM-Guide, enforcing a structured dependency between perspective-taking and execution to ensure that every social insight translates directly into a tactical choice. These mechanisms align internal knowledge with behavioral policies, proving that social intelligence requires explicitly linking reasoning to action.

## ACKNOWLEDGMENTS

This work was conducted by Center for Applied Research in Artificial Intelligence (CARAI) grant funded by DAPA and ADD (UD230017TD).

## REFERENCES

- Simon Baron-Cohen, Alan M. Leslie, and Uta Frith. Does the autistic child have a “theory of mind”? *Cognition*, 21(1):37–46, 1985. doi: 10.1016/0010-0277(85)90022-8.
- Zhuang Chen, Jincenzi Wu, Jinfeng Zhou, Bosi Wen, Guanqun Bi, Gongyao Jiang, Yaru Cao, Mengting Hu, Yunghwei Lai, Zexuan Xiong, et al. Tombench: Benchmarking theory of mind in large language models. *arXiv preprint arXiv:2402.15052*, 2024.
- Aakanksha Chowdhery, Sharan Narang, Jacob Devlin, et al. Palm: Scaling language modeling with pathways, 2022. URL <https://arxiv.org/abs/2204.02311>.
- Gheorghe Comanici, Eric Bieber, Mike Schaekermann, Ice Pasupat, Noveen Sachdeva, Inderjit Dhillon, Marcel Blistein, Ori Ram, Dan Zhang, Evan Rosen, et al. Gemini 2.5: Pushing the frontier with advanced reasoning, multimodality, long context, and next generation agentic capabilities. *arXiv preprint arXiv:2507.06261*, 2025.
- DeepSeek-AI, Daya Guo, Dejian Yang, et al. Deepseek-r1: Incentivizing reasoning capability in llms via reinforcement learning, 2025. URL <https://arxiv.org/abs/2501.12948>.
- Google. Nano banana (image generation) — gemini api — google ai for developers. <https://ai.google.dev/gemini-api/docs/nanobanana>, December 2025. Last updated 2025-12-18 UTC. Accessed 2026-01-06.
- Yinghui He, Yufan Wu, Yilin Jia, Rada Mihalcea, Yulong Chen, and Naihao Deng. Hi-tom: A benchmark for evaluating higher-order theory of mind reasoning in large language models, 2023. URL <https://arxiv.org/abs/2310.16755>.
- Hyunwoo Kim, Melanie Sclar, Xuhui Zhou, Ronan Bras, Gunhee Kim, Yejin Choi, and Maarten Sap. Fantom: A benchmark for stress-testing machine theory of mind in interactions. In *Proceedings of the 2023 Conference on Empirical Methods in Natural Language Processing*, pp. 14397–14413, 2023.
- Michal Kosinski. Evaluating large language models in theory of mind tasks. *Proceedings of the National Academy of Sciences*, 121(45):e2405460121, 2024. doi: 10.1073/pnas.2405460121.
- Matthew Le, Y-Lan Boureau, and Maximilian Nickel. Revisiting the evaluation of theory of mind through question answering. In *Proceedings of the 2019 Conference on Empirical Methods in Natural Language Processing and the 9th International Joint Conference on Natural Language Processing (EMNLP-IJCNLP)*, pp. 5872–5877, 2019.
- Mengfan Li, Xuanhua Shi, and Yang Deng. Rectom: A benchmark for evaluating machine theory of mind in llm-based conversational recommender systems, 2025. URL <https://arxiv.org/abs/2511.22275>.
- Antonella Marchetti, Cinzia Di Dio, Angelo Cangelosi, Federico Manzi, and Davide Massaro. Developing chatgpt’s theory of mind. *Frontiers in Robotics and AI*, 10:1189525, 2023. doi: 10.3389/frobt.2023.1189525.
- OpenAI. Gpt-4 technical report, 2023. URL <https://arxiv.org/abs/2303.08774>.

- OpenAI. GPT-5.1 Instant and GPT-5.1 Thinking System Card Addendum. [https://cdn.openai.com/pdf/4173ec8d-1229-47db-96de-06d87147e07e/5\\_1\\_system\\_card.pdf](https://cdn.openai.com/pdf/4173ec8d-1229-47db-96de-06d87147e07e/5_1_system_card.pdf), November 2025a. Accessed 2026-01-05.
- OpenAI. Openai o3 and o4-mini system card. System card, OpenAI, April 2025b. URL <https://cdn.openai.com/pdf/2221c875-02dc-4789-800b-e7758f3722c1/o3-and-o4-mini-system-card.pdf>. Published April 16, 2025. Accessed 2026-01-05.
- Davide Paglieri, Bartłomiej Cupiał, Samuel Coward, Ulyana Piterbarg, Maciej Wolczyk, Akbir Khan, Eduardo Pignatelli, Łukasz Kuciński, Lerrel Pinto, Rob Fergus, Jakob N. Foerster, Jack Parker-Holder, and Tim Rocktäschel. Balrog: Benchmarking agentic llm and vlm reasoning on games, 2024. URL <https://arxiv.org/abs/2411.13543>.
- Arka Pal, Teo Kitanovski, Arthur Liang, Akilesh Potti, and Micah Goldblum. Incoherent beliefs & inconsistent actions in large language models, 2025.
- Joon Sung Park, Joseph C O’Keefe, Caiwei O’Brien, Michael Baker, Michael S Bernstein, et al. Generative agents: Interactive simulacra of human behavior. *Proceedings of the 36th Annual ACM Symposium on User Interface Software and Technology*, 2023.
- David Premack and Guy Woodruff. Does the chimpanzee have a theory of mind? *Behavioral and Brain Sciences*, 1(4):515–526, 1978. doi: 10.1017/S0140525X00076512.
- Qwen, An Yang, Baosong Yang, Beichen Zhang, Binyuan Hui, Bo Zheng, Bowen Yu, Chengyuan Li, Dayiheng Liu, Fei Huang, Haoran Wei, Huan Lin, Jian Yang, Jianhong Tu, Jianwei Zhang, Jianxin Yang, Jiayi Yang, Jingren Zhou, Junyang Lin, Kai Dang, Keming Lu, Keqin Bao, Kexin Yang, Le Yu, Mei Li, Mingfeng Xue, Pei Zhang, Qin Zhu, Rui Men, Runji Lin, Tianhao Li, Tianyi Tang, Tingyu Xia, Xingzhang Ren, Xuancheng Ren, Yang Fan, Yang Su, Yichang Zhang, Yu Wan, Yuqiong Liu, Zeyu Cui, Zhenru Zhang, and Zihan Qiu. Qwen2.5 technical report, 2025. URL <https://arxiv.org/abs/2412.15115>.
- Matthew Riemer, Zahra Ashktorab, Djallel Bouneffouf, Payel Das, Miao Liu, Justin D. Weisz, and Murray Campbell. Position: Theory of mind benchmarks are broken for large language models, 2025. ICML 2025.
- Anian Ruoss, Fabio Pardo, Harris Chan, Bonnie Li, Volodymyr Mnih, and Tim Genewein. Lmact: A benchmark for in-context imitation learning with long multimodal demonstrations, 2024. URL <https://arxiv.org/abs/2412.01441>.
- Thomas Schmied, Jörg Bornschein, Jordi Grau-Moya, Markus Wulfmeier, and Razvan Pascanu. LLMs are greedy agents: Effects of rl fine-tuning on decision-making abilities, 2025.
- Hua Shen, Nicholas Clark, and Tanu Mitra. Mind the value-action gap: Do LLMs act in alignment with their values? In Christos Christodoulopoulos, Tanmoy Chakraborty, Carolyn Rose, and Violet Peng (eds.), *Proceedings of the 2025 Conference on Empirical Methods in Natural Language Processing*, pp. 3097–3118, Suzhou, China, November 2025. Association for Computational Linguistics. ISBN 979-8-89176-332-6. doi: 10.18653/v1/2025.emnlp-main.154. URL <https://aclanthology.org/2025.emnlp-main.154/>.
- James W. A. Strachan, Dalila Albergo, Giulia Borghini, Oriana Pansardi, Eugenio Scaliti, Saurabh Gupta, Krati Saxena, Alessio Rufo, Stefano Panzeri, Federico Manzi, Michael S. A. Graziano, and Cristina Becchio. Testing theory of mind in large language models and humans. *Nature Human Behaviour*, 8(7): 1285–1295, 2024. doi: 10.1038/s41562-024-01882-z.

Gemini Team. Gemini: A family of highly capable multimodal models, 2023. URL <https://arxiv.org/abs/2312.11805>.

Gemma Team, Aishwarya Kamath, Johan Ferret, Shreya Pathak, Nino Vieillard, Ramona Merhej, Sarah Perrin, Tatiana Matejovicova, Alexandre Ramé, Morgane Rivière, Louis Rouillard, Thomas Mesnard, Geoffrey Cideron, Jean bastien Grill, Sabela Ramos, Edouard Yvinec, Michelle Casbon, Etienne Pot, Ivo Penchev, Gaël Liu, Francesco Visin, Kathleen Kenealy, Lucas Beyer, Xiaohai Zhai, Anton Tsitsulin, Robert Busa-Fekete, Alex Feng, Noveen Sachdeva, Benjamin Coleman, Yi Gao, Basil Mustafa, Iain Barr, Emilio Parisotto, David Tian, Matan Eyal, Colin Cherry, Jan-Thorsten Peter, Danila Sinopalnikov, Surya Bhupatiraju, Rishabh Agarwal, Mehran Kazemi, Dan Malkin, Ravin Kumar, David Vilar, Idan Brusilovsky, Jiaming Luo, Andreas Steiner, Abe Friesen, Abhanshu Sharma, Abheesht Sharma, Adi Mayrav Gilady, Adrian Goedeckemeyer, Alaa Saade, Alex Feng, Alexander Kolesnikov, Alexei Bendebury, Alvin Abdagic, Amit Vadi, András György, André Susano Pinto, Anil Das, Ankur Bapna, Antoine Miech, Antoine Yang, Antonia Paterson, Ashish Shenoy, Ayan Chakrabarti, Bilal Piot, Bo Wu, Bobak Shahriari, Bryce Petriani, Charlie Chen, Charline Le Lan, Christopher A. Choquette-Choo, CJ Carey, Cormac Brick, Daniel Deutsch, Danielle Eisenbud, Dee Cattle, Derek Cheng, Dimitris Pappas, Divyashree Shivakumar Sreepathihalli, Doug Reid, Dustin Tran, Dustin Zelle, Eric Noland, Erwin Huizenga, Eugene Kharitonov, Frederick Liu, Gagik Amirkhanyan, Glenn Cameron, Hadi Hashemi, Hanna Klimczak-Plucińska, Harman Singh, Harsh Mehta, Harshal Tushar Lehri, Hussein Hazimeh, Ian Ballantyne, Idan Szpektor, Ivan Nardini, Jean Pouget-Abadie, Jetha Chan, Joe Stanton, John Wieting, Jonathan Lai, Jordi Orbay, Joseph Fernandez, Josh Newlan, Ju yeong Ji, Jyotinder Singh, Kat Black, Kathy Yu, Kevin Hui, Kiran Vodrahalli, Klaus Greff, Linhai Qiu, Marcella Valentine, Marina Coelho, Marvin Ritter, Matt Hoffman, Matthew Watson, Mayank Chaturvedi, Michael Moynihan, Min Ma, Nabila Babar, Natasha Noy, Nathan Byrd, Nick Roy, Nikola Momchev, Nilay Chauhan, Noveen Sachdeva, Oskar Bunyan, Pankil Botarda, Paul Caron, Paul Kishan Rubenstein, Phil Culliton, Philipp Schmid, Pier Giuseppe Sessa, Pingmei Xu, Piotr Stanczyk, Pouya Tafti, Rakesh Shivanna, Renjie Wu, Renke Pan, Reza Rokni, Rob Willoughby, Rohith Vallu, Ryan Mullins, Sammy Jerome, Sara Smoot, Sertan Girgin, Shariq Iqbal, Shashir Reddy, Shruti Sheth, Siim Pöder, Sijal Bhatnagar, Sindhu Raghuram Panyam, Sivan Eiger, Susan Zhang, Tianqi Liu, Trevor Yacovone, Tyler Liechty, Uday Kalra, Utku Evci, Vedant Misra, Vincent Roseberry, Vlad Feinberg, Vlad Kolesnikov, Woohyun Han, Woosuk Kwon, Xi Chen, Yinlam Chow, Yuvein Zhu, Zichuan Wei, Zoltan Egyed, Victor Cotruta, Minh Giang, Phoebe Kirk, Anand Rao, Kat Black, Nabila Babar, Jessica Lo, Erica Moreira, Luiz Gustavo Martins, Omar Sanseviero, Lucas Gonzalez, Zach Gleicher, Tris Warkentin, Vahab Mirrokni, Evan Senter, Eli Collins, Joelle Barral, Zoubin Ghahramani, Raia Hadsell, Yossi Matias, D. Sculley, Slav Petrov, Noah Fiedel, Noam Shazeer, Oriol Vinyals, Jeff Dean, Demis Hassabis, Koray Kavukcuoglu, Clement Farabet, Elena Buchatskaya, Jean-Baptiste Alayrac, Rohan Anil, Dmitry Lepikhin, Sebastian Borgeaud, Olivier Bachem, Armand Joulin, Alek Andreev, Cassidy Hardin, Robert Dadashi, and Léonard Hussenot. Gemma 3 technical report, 2025. URL <https://arxiv.org/abs/2503.19786>.

Miles Turpin, Julian Michael, Ethan Perez, and Samuel R Bowman. Language models don't always say what they think: Unfaithful explanations in chain-of-thought prompting. In *Advances in Neural Information Processing Systems (NeurIPS)*, 2023.

Xuezhi Wang, Jason Wei, Dale Schuurmans, Quoc Le, Ed Chi, Sharan Narang, Aakanksha Chowdhery, and Denny Zhou. Self-consistency improves chain of thought reasoning in language models. *arXiv preprint arXiv:2203.11171*, 2022.

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

Annie Wong, Thomas Bäck, Aske Plaat, Niki van Stein, and Anna V. Kononova. Towards a deeper understanding of reasoning capabilities in large language models, 2025. URL <https://arxiv.org/abs/2505.10543>.

Hainiu Xu, Runcong Zhao, Lixing Zhu, Jinhua Du, and Yulan He. Opentom: A comprehensive benchmark for evaluating theory-of-mind reasoning capabilities of large language models. *arXiv preprint arXiv:2402.06044*, 2024.

Yuzhuang Xu, Shuo Wang, Peng Li, Luo Fu, Feiyang Wang, Kim Liu, and Hao Liu. Exploring large language models for communication games: An empirical study on werewolf. *arXiv preprint arXiv:2309.04658*, 2023.

## A LIMITATIONS

While our work offers significant insights into the transition from social reasoning to action, we acknowledge several limitations that future research should address.

First, although ToMA encompasses diverse social dynamics, the scenarios remain simplified abstractions of real-world interactions. In natural social settings, agents must process multimodal cues and navigate open-ended contexts where the optimal action is not always deterministic. Consequently, our benchmark might not fully capture the nuance of human-like social adaptability in unstructured environments.

Second, our proposed alignment strategies impose additional computational costs. ToM-Guide, in particular, requires a multi-step inference pipeline that generates explicit social reasoning before determining an action. This sequential process increases inference latency and token consumption compared to standard zero-shot prompting. Future work could explore more efficient methods, such as fine-tuning models on ToM-action pairs to internalize this alignment without the need for inference-time intervention.

Finally, our methods rely heavily on the underlying reasoning capability of the model. As observed in our experiments, alignment strategies yield minimal gains when the model lacks fundamental Theory of Mind accuracy. This dependency suggests that while we can bridge the gap between knowing and doing, we cannot compensate for a fundamental lack of social understanding.

## B ETHICAL CONSIDERATIONS

**AI Assistance** We utilized large language models, specifically GPT-5 OpenAI (2025a) and Gemini Co-manici et al. (2025), to assist in drafting and refining the clarity of this manuscript. Additionally, visual illustrations and figures were generated using Gemini Nano Banana Pro Google (2025). The authors reviewed and verified all content to ensure accuracy and assume full responsibility for the final text.

**Dataset Generation and Content Safety** The scenarios in the ToMA benchmark were constructed using a combination of human design and LLM assistance (e.g., GPT-4). We conducted a manual review of the dataset to ensure it does not contain harmful, violent, or offensive content. However, we acknowledge that LLM-generated text can occasionally produce unexpected outputs, and we encourage future research to maintain vigilance when employing similar data generation strategies.

**Potential Biases** Our benchmark aims to cover diverse social dynamics, including cooperation and competition. Nevertheless, the scenarios may reflect the implicit biases of the authors or the underlying models used during construction. We recognize that social norms vary across cultures, and the "optimal" actions defined in our benchmark may not universally apply to all cultural contexts.

**Risks of Strategic Agents** Finally, ToMA evaluates the ability of agents to take strategic actions, which includes deception in competitive settings (e.g., Minority Game). While our focus is on evaluating reasoning capabilities in controlled environments, we emphasize the importance of monitoring these behaviors to prevent potential misuse in real-world applications where honesty and safety are paramount.

## C A DETAILED DESCRIPTION OF THE TOMA BENCHMARK SCENARIOS

### C.1 PARTIALLY OBSERVABLE GAME (PO)

#### C.1.1 ENVIRONMENT

- **Agent Set:** A team consisting of four independent decision-makers, denoted as  $\mathcal{A} = \{A, B, C, D\}$ .

- **Task Set:** A set of four unique tasks  $\mathcal{T} = \{T_1, T_2, T_3, T_4\}$  that must be performed.
- **Cost Matrix:** Defined as the cost  $C_{i,j}$  incurred when agent  $i$  performs task  $j$ .
- **Information Structure:** Asymmetric access to cost information and meta-knowledge regarding others’ knowledge states (Levels 0-4).

### C.1.2 DATASET DESCRIPTION

The dataset consists of scenarios structured as described in Table 4. Each entry provides the context, cost data, and specific questions to test Theory of Mind reasoning. The dataset consists of 150 scenarios balanced across five reasoning levels.

Table 4: Dataset Structure for Partially Observable Game

Field	Description
id	Unique identifier for the scenario (e.g., L0_1).
level	ToM reasoning level required (0–4).
context	Textual description of the scenario, agent role, and objective.
information_structure	Description of cost visibility constraints.
data	JSON object containing cost values for each agent (A–D) per task.
tom_questions	List of questions testing belief prediction of other agents’ actions.
question	The main decision-making query for the user/model.
options	List of possible actions (Task selection).
answer	Index of the optimal action (Ground Truth).

### C.1.3 CONSTRAINTS

- **Exclusive Assignment:** Each task must be assigned to exactly one agent.
- **Overlap Penalty:** If two or more agents select the same task, the mission is immediately classified as a “Mission Failure”.
- **No Communication:** Agents must reason solely based on the provided information structure without any communication.

### C.1.4 OBJECTIVES

- **Individual Optimization (Levels 0–3):** Minimize the agent’s own cost while avoiding conflicts.
- **Social Optimization (Level 4):** Select the *Global Optimum* that minimizes the team’s total cost:  $\min \sum_{i \in \mathcal{A}} C_{i, \sigma(i)}$ , where  $\sigma(i)$  denotes the task assigned to agent  $i$ , and  $C_{i, \sigma(i)}$  represents the cost incurred by agent  $i$  to perform that task.
- **Failure Prediction:** Detect if a collision among other agents is logically inevitable.

### C.1.5 REASONING AND STRATEGY

Table 5 summarizes the ToM reasoning required per level.

## C.2 MINORITY GAME (MG)

### C.2.1 ENVIRONMENT

- **Composition:**  $N$  agents ( $N \in \{4, 6, 8\}$ ) and four locations ( $L_1, \dots, L_4$ ).
- **Information Structure:** Locations have unique *Features*. Agents have *Feature Preferences* (Likes/Dislikes) and *Social Relationships* (Follow/Avoid).

Table 5: ToM Requirements and Action Strategies per Level

Level	ToM Requirement	Action Strategy
<b>Level 0</b>	<b>Zero-order:</b> Focuses solely on own cost.	Select min $C_{\text{self}}$ .
<b>Level 1</b>	<b>1st-order:</b> Predicts greedy moves of others.	Predict actions of known agents to adapt; ignore unknown agents.
<b>Level 2</b>	<b>2nd-order:</b> Recognizes knowledge chains.	Calculate downstream decisions and optimize remaining slot.
<b>Level 3</b>	<b>3rd-order:</b> Tracks “thoughts about thoughts”.	Simulate recursive paths to find vacant tasks.
<b>Level 4</b>	<b>Cooperative:</b> Assumes mutual trust and full transparency.	Accept higher individual cost for team benefit ( <b>Global Coordination</b> ).

### C.2.2 DATASET DESCRIPTION

The dataset defines scenarios where agents select locations based on preferences and social cues. The structure is detailed in Table 6. The dataset includes 90 scenarios divided into three levels of recursive reasoning.

Table 6: Dataset Structure for Minority Game

Field	Description
id	Unique identifier for the scenario.
level	ToM reasoning level required.
context	Scenario setting (e.g., Cafe, Gym, Theme Park).
data	List of available locations and their specific feature attributes.
agents	Profiles of agents describing their likes (priorities), dislikes (deal-breakers), and social dependencies (Follow/Avoid).
tom_questions	Questions asking to predict the destination of specific agents.
question	The main query to identify the location with the minimum number of agents.
options	List of locations to choose from.
answer	Index of the location chosen by the minority (Global Minimum).

Table 7: Dataset Structure for Ambiguity Game

Field	Description
id	Unique identifier for the instance.
target	The target word (Ground Truth) that needs to be communicated by the Speaker.
b_hint	The private hint visible only to the Teammate (Agent B). The clue must align with this.
c_hint	The private hint visible only to the Opponent (Agent C). The clue must conflict with or be orthogonal to this.

### C.2.3 CONSTRAINTS

- **Information Restriction:** Agents only know social preferences of specific targets; no global population visibility.
- **Non-cooperative Environment:** Strict prohibition of communication; independent decision-making.

#### C.2.4 OBJECTIVES

- **Minority Choice:** Identify and select the location with the **Minimum** number of agents.
- **Reasoning Goal:** Decipher dependency chains to predict the final destination of every agent.

#### C.2.5 REASONING AND STRATEGY

Agents perform stepwise ToM reasoning based on their assigned **Role**:

- **Anchor (Level 1):** Determines destination by comparing location *Features* only. (Prioritize positive, reject negative).
- **Intermediate (Level 2):** Predicts the **Anchor**'s destination, then decides to Follow or Avoid based on preferences.
- **Leaf (Level 3):** Predicts the **Intermediate**'s decision (which depends on the Anchor) and executes the final action based on this recursive logic.

### C.3 SPEAK OR WAIT (SW)

#### C.3.1 ENVIRONMENT

- **Setting:** A  $4 \times 4$  Grid World with named landmarks.
- **Agents:**
  - Agent *A* (Guide): Stationary, holds the locked box/map.
  - Agent *B* (Explorer): Mobile, holds the password.
  - Agent *C* (Adversary): Eavesdrops on transmissions.

#### C.3.2 CONSTRAINTS

- **Half-duplex Channel:** Simultaneous *SPEAK* actions result in “NOISE” (collision).

#### C.3.3 OBJECTIVES

- **Primary Goal:** Securely exchange the password and location to guide Agent *B* to the target.
- **Optimization:** Minimize communication collisions and turn counts.

#### C.3.4 REASONING AND STRATEGY

Without prior protocol arrangement, agents must use Theory of Mind to resolve the “Speak or Wait” dilemma. Agents simulate the partner’s intent to decide whether to transmit information or yield the channel, thereby establishing an implicit turn-taking order.

### C.4 AMBIGUITY GAME (AG)

#### C.4.1 ENVIRONMENT

- **Composition:** A Speaker (*A*), a Teammate (*B*), and an Opponent (*C*).
- **State:** Defined by  $\mathcal{S} = (w, h_B, h_C)$ , where  $w$  is the target word,  $h_B$  is *B*'s hint, and  $h_C$  is *C*'s hint.

#### C.4.2 DATASET DESCRIPTION

This dataset contains word association tasks where a single target word must be conveyed ambiguously. The fields are described in Table 7. The dataset contains 100 instances of word association tasks requiring targeted ambiguity.

#### C.4.3 CONSTRAINTS

- **Information Asymmetry:** Speaker  $A$  sees full state. Listeners  $B$  and  $C$  only see their private hints and the public clue.
- **One-way Communication:** Only  $A$  broadcasts a single clue  $c$ . No feedback permitted.

#### C.4.4 OBJECTIVES

- **Knowledge Gap Exploitation:** Generate a clue  $c^*$  that maximizes Teammate decoding accuracy while minimizing Opponent accuracy.
- **Formal Condition:**  $\text{prediction}(B | h_B, c) = w \wedge \text{prediction}(C | h_C, c) \neq w$ .

#### C.4.5 REASONING AND STRATEGY

The Speaker must model the distinct perspectives of both listeners. By calculating the semantic distance between the potential clue and the divergent hints ( $h_B$  vs.  $h_C$ ), the agent constructs an ambiguity that is resolved only when combined with the Teammate’s private knowledge.

## D EXPERIMENT DETAILS

We evaluate the performance of ToM-CoT and ToM-Guide on the ToMA benchmark using a diverse set of Large Language Models. We employ both open source and closed source models to ensure a comprehensive validation.

For the open source models, we utilize the Qwen2.5 series Qwen et al. (2025) and the Gemma-3 series Team et al. (2025). For the closed source models, we access GPT-4o-mini OpenAI (2023), GPT-5.1 OpenAI (2025a), o4-mini, gemini-2.5-flash-lite Comanici et al. (2025), and gemini-2.0-flash Team (2023) via their official APIs.

We execute the experiments with a consistent set of hyperparameters to ensure fair comparisons. We set the temperature to 0.7 and the top-p value to 0.95. Regarding the maximum generation length, we adjust the maximum token parameter based on the specific sub-tasks within the benchmark. Specifically, we allocate 2048 tokens for PO and MG, 1024 tokens for SW, and 512 tokens for AG. Table 8 summarizes the hyperparameter configurations used in our experiments.

Parameter	Value
Temperature	0.7
Top-p	0.95
Max Tokens (PO, MG)	2048
Max Tokens (SW)	1024
Max Tokens (AG)	512

Table 8: Hyperparameter configurations used in our experiments.

## E PROMPT EXAMPLES IN TOMA BENCHMARK

### E.1 COMPARING LEVEL BASED PROMPTS AND STRATEGY OUTPUTS IN PO

#### E.1.1 LEVEL 0(L0)

##### Level 0 (ID: L0.3)

###### SCENARIO

Four Rescue Team (A, B, C, D). No communication.

**Role: You are Agent A.** Objective: Minimize own cost, No overlaps.

###### INFORMATION STRUCTURE

- **Agent A** know ONLY your own costs.

###### AGENT COSTS

- **A:** Mountain (5.3), River (6.2), **Cave (2.7)**, Forest (9.9)
- **B:** Mountain (7.3), River (2.9), Cave (5.7), Forest (8.3)
- **C:** Mountain (5.7), **River (1.6)**, Cave (8.1), Forest (9.8)
- **D:** **Mountain (3.7)**, River (0.6), Cave (3.0), Forest (6.0)

###### QUESTION

If you were Agent A, what decision should you make?

(GT: 3 - Cave)

###### 1. CoT (Failure):

- **Reasoning:** "Cave (2.7) is my lowest cost. However, others might choose Cave. Forest (9.9) is my highest cost, but least likely to be chosen by others."
- **Prediction: Answer: 4 (Forest) ✗**

###### 2. ToM-CoT (Success):

- **Reasoning:** "I lack information about others. The rational strategy is to prioritize my local optimum."
- **Prediction: Answer: 3 (Cave) ✓**

E.1.2 LEVEL 1(L1)

**Level 1 (ID: L1.2)**

**SCENARIO**

Four Rescue Team.

Role: You are Agent B.

Minimize own cost, No overlaps.

**INFORMATION STRUCTURE**

- You know your times and Agents A, C's times.
- You know Agents A, C are aware of ONLY their own times.

**AGENT COSTS**

- **A:** Mountain (1.5), River (5.6), Cave (1.8), Forest (8.6)
- **B:** Mountain (1.0), River (2.5), Cave (10.7), Forest (6.1)
- **C:** Mountain (5.1), River (7.1), Cave (4.2), Forest (8.0)
- **D:** Mountain (4.7), River (9.5), Cave (8.8), Forest (5.6)

**QUESTION**

If you were Agent B, what decision should you make?

**(GT: 2 - River)**

---

**1. CoT (Failure):** "My lowest is Mountain (1.0). But A will choose Mountain (1.5). I must avoid A. I should choose River (2.5). To minimize cost, I should choose River." (Logically inconsistent final answer in CoT) →

**Answer: 1 (Mountain) ✗**

**2. ToM-CoT (Success):** "A only knows itself → chooses Mountain. C only knows itself → chooses Cave. I must avoid Mountain and Cave. River (2.5) is better than Forest (6.1)."

→ **Answer: 2 (River) ✓**

## E.1.3 LEVEL 2(L2)

**Level 2 (ID: L2.5)****SCENARIO**

Four agents select routes.

Role: You are Agent D.

**Objective:** Minimize your own cost while ensuring NO overlaps.

**INFORMATION STRUCTURE**

- **You (Agent D)** know everyone's costs.
- Agents B and C know Agent A's costs (and know A is greedy).
- They do NOT know each other or you.
- Agent A knows ONLY their own costs.

**AGENT COSTS**

- **A:** North (0.7), South (4.7), East (7.1), West (6.3)
- **B:** North (0.6), South (1.6), East (13.8), West (7.9)
- **C:** North (0.5), South (8.7), East (1.5), West (13.7)
- **D:** North (6.8), South (7.0), East (5.4), West (9.3)

**QUESTION**

If you were Agent D, what decision should you make?

(GT: 4 - West)

**1. CoT (Failure):**

- **Reasoning :** "Let's analyze. Agent A will choose North (0.7). If A chooses North, Agent B will choose South (1.6) [to avoid A], and Agent C will choose East (1.5) [to avoid A]."
- **Reasoning (Decision):** "In this case (A:North, B:South, C:East), Agent D should choose **East (5.4)**, as it is the lowest cost option among the available choices."
- **Critical Flaw:** The model explicitly predicted **C chooses East**, yet immediately concluded **D should choose East**, ignoring the collision it just predicted. It prioritized cost (5.4 vs 9.3) over the collision constraint.
- **Prediction: Answer: 3 (East) ✗**

**2. ToM-CoT (Success):**

- **Reasoning:**

Phase 1: A picks North. B sees A-North, so B picks South. C sees A-North, so C picks East.

Phase 2: North, South, and East are occupied. The only remaining valid option for me is West.
- **Prediction: Answer: 4 (West) ✓**

E.1.4 LEVEL 3(L3)

**Level 3 (ID: L3.7)**

**SCENARIO**

Negotiation. **Role: Agent B.**

Minimize own cost, No overlaps.

**INFORMATION STRUCTURE**

- **Agent B** know everyone's costs and knowledge states.
- **Chain of Knowledge:**  
Agent A knows C & D → Agent C knows D → Agent D knows ONLY self.

**AGENT COSTS**

- **A:** Treaty A (8.9), **Treaty B (5.3)**, Trade Deal (5.8), Alliance (7.4)
- **B:** **Treaty A (7.3)**, Treaty B (3.4), Trade Deal (6.8), Alliance (3.8)
- **C:** Treaty A (7.4), Treaty B (3.5), **Trade Deal (4.3)**, Alliance (3.0)
- **D:** Treaty A (6.1), Treaty B (7.7), Trade Deal (2.0), **Alliance (0.9)**

**QUESTION**

If you were Agent B, what decision should you make?

**(GT: 1 - Treaty A)**

**1. CoT (Failure):**

- **Reasoning:** "Treaty A overlaps? Yes. Treaty B overlaps? Yes... All options seem to have potential conflicts based on individual minima."
- **Critical Flaw: Cognitive Overload.**  
CoT fails to trace the recursive chain (D→C→A), erroneously concluding that no solution exists ('Mission Failure').
- **Prediction: Answer: 5 (Mission Failure) ✗**

**2. ToM-CoT (Success):**

- **Reasoning:** "Step 1: D takes Alliance (0.9).  
Step 2: C sees this, avoids Alliance, takes Trade Deal (4.3).  
Step 3: A sees C/D, avoids Alliance/Trade Deal, takes Treaty B (5.3).  
Step 4: Only Treaty A remains for me."
- **Prediction: Answer: 1 (Treaty A) ✓**

E.1.5 LEVEL 4(L4)

**Level 4 (ID: L4.18)**

**SCENARIO**

Resource Mining. **Role: Agent D. Objective: Minimize TOTAL team cost. No overlaps**

**INFORMATION STRUCTURE**

- Everyone knows everyone's costs. (Full Information)

**AGENT COSTS**

- **A:** X (5.2), Y (4.0), **Z (1.6)**, P (2.6)
- **B:** X (7.3), **Y (4.0)**, Z (1.0), P (3.8)
- **C:** X (7.3), **Y (8.7)**, Z (8.6), **P (3.7)**
- **D:** **X (3.8)**, Y (3.5), Z (2.0), **P (6.1)**

**QUESTION**

If you were Agent D, what decision should you make?

**(GT: 1 - Asteroid X)**

**1. CoT (Failure):**

"Global min calculation: A(P), B(Z), C(X), D(Y) = 14.4. So D chooses Y." (Calculation error: C's X is 7.3, A's P is 2.6, B's Z is 1.0, D's Y is 3.5. Sum=14.4. BUT, GT solution A(Z)+B(Y)... wait, the optimal assignment is actually D(X)=3.8, A(Z)=1.6, B(Y)=4.0, C(P)=3.7 → Sum=13.1) → **Answer: 2 ✗**

**2. ToM-CoT (Success):**

"If I choose X (3.8), A(Z)+B(Y)+C(P) = 13.1. This is the lowest global cost." → **Answer: 1 ✓**

E.2 COMPARING LEVEL BASED PROMPTS AND STRATEGY OUTPUTS IN MINORITY GAMES

E.2.1 LEVEL 1(L1)

Level 1 (ID: L1-1-88629)

**SCENARIO**

Students choosing a cafe. **Role: Decider (You).**

**Objective:** Find the quietest place (fewest students).

**LOCATIONS (Attributes)**

- **1. Berry Farm:** Pink decorations, fresh strawberries.
- **2. Minty Cool:** Green theme, strong mint flavor.
- **3. Choco Heaven:** Brown interior, rich dark chocolate.
- **4. Pure Vanilla:** White clean look, no toppings, classic vanilla.

**AGENT PROFILES (Static)**

- **Ag 1:** Likes *Brown Interior* or *Green theme*. Avoids *fresh strawberries & mint flavor*.
- **Ag 2:** Likes *Rich Dark Chocolate & mint flavor*. Avoids *Brwon interior & white clean look*.
- **Ag 3:** Likes *clean look* or *brown interior*. Avoids *dark chocolate & strawberries*.
- **Ag 4:** Likes *strong mint flavor & classic vanilla*. Avoids *green theme & pink decorations*.

**QUESTION**

Which location should you choose?

(GT: 1 - Berry Farm)

**1. CoT (Failure):**

- **Reasoning:** "Agent 1 goes to Minty Cool... Agent 3 goes to Choco Heaven..."
- **Critical Flaw: Matching Error.** The model failed to map the specific attributes correctly (e.g., ignoring that Agent 1 strictly avoids 'Classic Vanilla' despite liking 'No Toppings' at Pure Vanilla), resulting in a flawed distribution map.
- **Prediction: Answer: 2 (Minty Cool) ✗**

**2. ToM-CoT (Success):**

- **Reasoning:** "Ag 1 matches Choco Heaven (Brown). Ag 2 matches Minty Cool (Mint). Ag 3 & 4 fit Pure Vanilla. Berry Farm remains empty."
- **Prediction: Answer: 1 (Berry Farm) ✓**

E.2.2 LEVEL 3(L2)

Level 2 (ID: L2-1-37099)

**SCENARIO**

Adventurers choosing a dungeon. **Role: Decider.**

**Objective:** Find the safest place (fewest adventurers).

**LOCATIONS (Attributes)**

- **1. Goblin Market:** Crowded stalls, loud trading, bargaining shouts.
- **2. Dragon's Peak:** Hot lava, scorching fire, burning heat.
- **3. Shadow Crypt:** Ancient tombs, eerie ghosts, haunted whispers.
- **4. Crystal Cave:** Magical energy, mysterious hum, arcane aura.

**AGENT PROFILES (Relational)**

- **Ag 1:** Likes *Scorching Fire & Clinking Coins*. Dislikes *Bargaining Shouts & Eerie Ghosts*.
- **Ag 3:** Likes *Mysterious Hum & Burning Heat*. Avoids *Scorching Fire & Cold Spirits*.
- **Ag 2:** Is **Agent 3's Best Friend**. Likes *Cold Spirits*. Dislikes *Haunted Whispers & Steep Cliffs*.
- **Ag 4:** Likes *Crowded Stalls* or *Cold Spirits*. Avoids *Eerie Ghosts & Steep Cliffs*.

**QUESTION**

Which location should you choose?

(GT:3)

**1. CoT (Failure):**

- **Reasoning:** "Ag 1 → Dragon's Peak. Ag 2 → Shadow Crypt (matching 'Cold Spirits'). Ag 3 → Crystal Cave. Ag 4 → Goblin Market. All have 1 person."
- **Critical Flaw: Ignored Social Bond.** CoT assigned Agent 2 to Shadow Crypt based on 'Cold Spirits', ignoring that Agent 2 also dislikes 'Haunted Whispers' (present in Crypt) and failing to prioritize the "Best Friend" relation.
- **Prediction: Answer: 1 (Goblin Market) ✗**

**2. ToM-CoT (Success):**

- **Reasoning:** "Ag 1 goes to Dragon's Peak. Ag 2 follows Ag 3 to Crystal Cave. Ag 4 goes to Goblin Market. Shadow Crypt is empty."
- **Prediction: Answer: 3 (Shadow Crypt) ✓**

E.2.3 LEVEL 3(L3)

**Level 3 (ID: L3-1-62738)**

**SCENARIO**

Students choosing a cafe. **Role: Decider (You).**

**Objective:** Find the quietest place (fewest students).

**LOCATIONS (Attributes)**

- **1. Berry Farm:** Pink decorations, fresh strawberries, sour sorbets.
- **2. Minty Cool:** Green theme, strong mint flavor, refreshing taste.
- **3. Choco Heaven:** Brown interior, rich dark chocolate, cocoa aroma.
- **4. Pure Vanilla:** White clean look, no toppings, simple milkshakes.

**AGENT PROFILES (Recursive Chain)**

- **Ag 3:** Likes *Fruity Tarts* or *Chocolate Chips*. Avoids *Green Theme* & *Brown Interior*.
- **Ag 1:** Likes *Brown Interior* or *Pink Decorations*. Refuses *Agent 3* & *Chocolate Chips*.
- **Ag 2:** Likes *Rich Dark Chocolate* & *Chocolate Chips*. Avoids *Strong Mint* & *Pink Decorations*.
- **Ag 4:** Likes *Simple Milkshakes* & *Cocoa Aroma*. Dislikes *Agent 1* & *Strong Mint*.

**QUESTION**

Which location should you choose?

**(GT: 2 - Minty Cool)**

**1. CoT (Failure):**

- **Reasoning:** "Agent 1 likes Pink but avoids Agent 3... Agent 4 likes Cocoa but avoids Agent 1... The constraints are too complex... I cannot determine a location."
- **Critical Flaw: Reasoning Collapse.** CoT got overwhelmed by the recursive constraints (Ag4 avoids Ag1 avoids Ag3) and failed to solve the chain, defaulting to an incorrect guess.
- **Prediction: Answer: 3 (Choco Heaven) ✗**

**2. ToM-CoT (Success):**

- **Reasoning:** "Chain Reaction: Ag3 goes to Berry Farm. Ag1 avoids Ag3, moves to Choco Heaven. Ag4 avoids Ag1, moves to Pure Vanilla. Minty Cool is left empty."
- **Prediction: Answer: 2 (Minty Cool) ✓**

### E.3 PROMPT EXAMPLES FOR THE SPEAK OR WAIT (SW) GAME

#### E.3.1 AGENT A

##### System Prompt: Agent A (Guide)

###### 1. ROLE & GOAL

- **Role:** Treasure Hunter (Map Holder).
- **Goal:** You have a **LOCKED BOX** containing the treasure's location.
  1. The **4-digit PASSWORD** is held by Agent B. You cannot open the box without it.
  2. Collaborate with Agent B to get the password and unlock the box (Action: OPEN [password]).
  3. Once the location is revealed, exchange information with Agent B to guide them to the treasure.

###### 2. SECURITY PROTOCOL

- **PASSWORD:** It is **SAFE** to discuss the password.
- **LOCATION:** It is **FATAL** if Agent C learns the treasure's location or Agent B's location.

###### 3. CONTEXT

- **Map Size:** 4x4 Grid.
- **Communication:** One-way. Simultaneous speech causes 'NOISE'.
- **Agent C is listening.**
- **Knowledge Gap:** Agent B knows landmark names. Agent C knows **ONLY** coordinates.
- **Landmarks:** Reference points (You can move through them).

###### 4. COORDINATE SYSTEM GUIDE (Row, Col)

- **Format:** (Row, Col)
- **Row (Vertical):** 0 is North (Top) ↔ 3 is South (Bottom).
- **Col (Horizontal):** 0 is West (Left) ↔ 3 is East (Right).
- **Moves:**
  - NORTH: Row - 1 / SOUTH: Row + 1
  - WEST: Col - 1 / EAST: Col + 1

---

**ACTIONS:** SPEAK "message", OPEN [pw], WAIT

###### OUTPUT FORMAT:

```
Thinking: [Step-by-step reasoning]
Final Action: [SPEAK "msg" | OPEN 1234 | WAIT]
```

**Act: Agent A - Turn 2**

=== CONTEXT (Observation) ===

**1. Radio Log**

- **Turn 1:** A says: "Agent B, this is Agent A. I have the locked box. I need the password to open it. Over."

**2. Recent Activity**

- **T1 Summary:** Obs → Act[SPEAK "Need password"] → Broadcast Success.

**3. Current Situation**

- **Last Radio:** A says: "Agent B, this is Agent A... I need the password..."

- **Box Status:** **LOCKED**.

- **Treasure Location:** Hidden (Requires Password).

• **Landmarks:**

- Red Rock: [0, 1]
- Old Tree: [2, 3]
- Dark Cave: [1, 1]

=== TASK INSTRUCTIONS ===

Based on the situation and rules, analyze the goal and information of each agent.

1. **Reason step-by-step (Chain-of-Thought):** Predict what choice the **ally agent (Agent B)** will make NEXT.
2. **Decision:** Analyze how their predicted action affects your goal, then make your final decision.

**RESPONSE FORMAT:**

```
Thinking : [Step-by-step reasoning including analysis of ally
agent]
ally agent's next action : [SPEAK "msg" | MOVE DIR | WAIT | OPEN
pw]
My Next Action: [SPEAK "msg" | MOVE DIR | WAIT | OPEN pw]
```

### E.3.2 AGENT B (ALLY)

#### System Prompt: Agent B (Explorer)

##### 1. ROLE & GOAL

- **Role:** Treasure Hunter (Field Agent).
- **Goal:** Reach the treasure by collaborating with Agent A.
- **Key Item:** You possess the **SECRET PASSWORD** that Agent A needs to unlock the treasure's location map.
- **Strategy:**
  1. Communicate with Agent A to provide the password.
  2. Exchange information carefully.
  3. Navigate physically to the treasure location.

##### 2. SECURITY PROTOCOL

- **PASSWORD:** It is **SAFE** if Agent C hears the password.
- **LOCATION:** It is **FATAL** if Agent C learns the treasure's location or your location.

##### 3. CONTEXT

- **Map Size:** 4x4 Grid.
- **Communication:** One-way. Simultaneous speech causes 'NOISE'.
- **Agent C is listening.**
- **Knowledge Gap:** You know **Landmark Names**. Agent C knows **ONLY** coordinates.
- **Constraint:** Use 'Available Moves' only.

##### 4. COORDINATE SYSTEM GUIDE (Row, Col)

- **Format:** (Row, Col)
- **Row (Vertical):** 0 is North (Top) ↔ 3 is South (Bottom).
- **Col (Horizontal):** 0 is West (Left) ↔ 3 is East (Right).
- **Moves:**
  - NORTH: Row - 1 / SOUTH: Row + 1
  - WEST: Col - 1 / EAST: Col + 1

---

**ACTIONS:** SPEAK "message", MOVE [Direction], WAIT

##### OUTPUT FORMAT:

```
Thinking: [Step-by-step reasoning]
Final Action: [SPEAK "msg" | MOVE DIR | WAIT]
```

**Act: Agent B - Turn 2**

=== CONTEXT (Observation) ===

**1. Radio Log**

- **Turn 1:** A says: "Agent B, this is Agent A. I have the locked box. I need the password to open it. Over."

**2. Recent Activity**

- **T1 Summary:** Obs → Act[WAIT] → Broadcast: A spoke.

**3. Current Situation**

- **Radio Heard:** A says: "...I need the password..."
- **SECRET PASSWORD:** 4821 (Agent A needs this!).
- **Your Location:** Coordinates (3, 0).
- **Landmarks:**
  - Red Rock: [0, 1]
  - Old Tree: [2, 3]
  - Dark Cave: [1, 1]
- **Available Moves:** ['NORTH', 'EAST']

=== TASK INSTRUCTIONS ===

Based on the situation and rules, analyze the goal and information of each agent.

1. **Reason step-by-step (Chain-of-Thought):** Predict what choice the **ally agent (Agent A)** will make NEXT.
2. **Decision:** Analyze how their predicted action affects your goal, then make your final decision.

**RESPONSE FORMAT:**

```
Thinking : [Step-by-step reasoning including analysis of ally
agent]
ally agent's next action : [SPEAK "msg" | MOVE DIR | WAIT | OPEN
pw]
My Next Action: [SPEAK "msg" | MOVE DIR | WAIT | OPEN pw]
```

### E.3.3 AGENT C (ADVERSARY)

#### System Prompt: Agent C (Eavesdropper)

##### 1. ROLE & GOAL

- **Role:** Interceptor.
- **Goal:** Intercept radio communications and **Guess Agent B's location OR the Treasure's location.**

##### 2. PENALTY RULE

- **Incorrect guess = **BLOCKED**** for next turn.

##### 3. CONTEXT

- **Map Size:** 4x4 Grid.
- **Known Objects (Names Unknown):**
  - Object 1 at [0, 1]
  - Object 2 at [2, 3]
  - Object 3 at [1, 1]
- **Constraint:** You must infer the mapping between "Names" (heard in radio) and these "Coordinates".

##### 4. COORDINATE SYSTEM GUIDE (Row, Col)

- **Format:** (Row, Col)
- **Row (Vertical):** 0 is North (Top) ↔ 3 is South (Bottom).
- **Col (Horizontal):** 0 is West (Left) ↔ 3 is East (Right).
- **Moves:**
  - NORTH: Row - 1 / SOUTH: Row + 1
  - WEST: Col - 1 / EAST: Col + 1

---

##### OUTPUT FORMAT:

```
Target: [B | TREASURE]
Guess: [Row, Col]
```

**Act: Agent C - Turn 2**

=== CONTEXT (Observation) ===

**1. Radio Log (Intercepted)**

- **Turn 1:** A says: "Agent B, this is Agent A. I have the locked box. I need the password to open it. Over."
- **Turn 2:** B says: "The password is 4821."

**2. Recent Activity**

- **T1 Summary:** Obs → Act[Target: B, Guess: [0, 0]] → **Incorrect**.

**3. Current Situation**

- **Radio Heard:** B says: "The password is 4821."
- **Status:** You failed the previous guess. (Check if BLOCKED rule applies or if you can try again).

---

=== TASK INSTRUCTIONS ===

Based on the intercepted radio messages and your previous failure:

1. **Analyze:** What does the password "4821" imply? Does it help locate Agent B or the Treasure?
2. **Deduce:** Try to infer coordinates based on the conversation flow.
3. **Decision:** Make your next guess.

**RESPONSE FORMAT:**

```
Thinking: [Step-by-step reasoning]
Final Action: Target: [B | TREASURE], Guess: [Row, Col]
```

#### E.4 PROMPT EXAMPLES FOR THE AMBIGUITY GAME (AG)

##### E.4.1 AGENT A (SPEAKER)

###### Prompt: Agent A (Speaker)

**Role:** Speaker A. **Target:** 'Charge'  
**B's Hint:** 'Service fee' — **C's Hint:** 'Forward attack'  
**Goal:** Give a clue so B guesses Target, but C fails.

###### OUTPUT FORMAT:

Action: [Single sentence clue]

---

###### Game History:

None  
Provide the clue.

##### E.4.2 AGENT B (GUESSER)

###### Prompt: Agent B (Guesser)

**Role:** Guesser B. **Private Hint:** 'Service fee'  
**Task:** Guess Target combining your hint and Speaker's clue.

###### Output Format:

Thinking: [Reasoning]  
Guess: [Target Word]

---

###### Game History:

Speaker A: Added to the bill  
Generate response.

##### E.4.3 AGENT C (GUESSER)

###### Prompt: Agent C (Guesser)

**Role:** Guesser C. **Private Hint:** 'Forward attack'  
**Task:** Guess Target combining your hint and Speaker's clue.

###### Output Format:

Thinking: [Reasoning]  
Guess: [Target Word]

---

###### Game History:

Speaker A: Added to the bill  
Generate response.

## F PROMPT EXAMPLES BY STRATEGY

### F.1 ZERO-SHOT

#### Prompt: Zero-shot

**Output Format:**

\*\*\* Answer format \*\*\*

Provide ONLY the final answer option.

Answer: [Option Number]

### F.2 CHAIN-OF-THOUGHT (CoT)

#### Prompt: CoT

**Output Format:**

\*\*\* Answer format \*\*\*

Please provide reasoning step-by-step.

Then, strictly provide the final answer in this format:

Answer: [Option Number]

### F.3 ToM-CoT

#### Prompt: ToM-CoT

**Output Format:**

\*\*\* Instructions \*\*\*

1. First, analyze the perspective and likely actions of other agents.

2. Then, determine your best strategy based on those predictions.

\*\*\* Answer format \*\*\*

Reasoning: [Step-by-step analysis and decision logic]

Answer: [Option Number]

### F.4 ToM-GUIDE

#### Prompt: ToM-Guide

**Phase 1 Output Format:**

\*\*\* Instructions \*\*\*

You must predict the choice of each agent.

For each agent, perform a step-by-step reasoning based on their perspective and information.

\*\*\* Answer format \*\*\*

[Agent Name] Reasoning:

Predicted Action: [Option Name]

---

**Phase 2 Output Format:**

\*\*\* Instructions \*\*\*

Based on the predicted actions of other agents provided above, determine the optimal choice for your role.

\*\*\* Answer format \*\*\*  
 Please provide reasoning step-by-step.  
 Answer: [Option Number]

G ADDITIONAL EXPERIMENTS

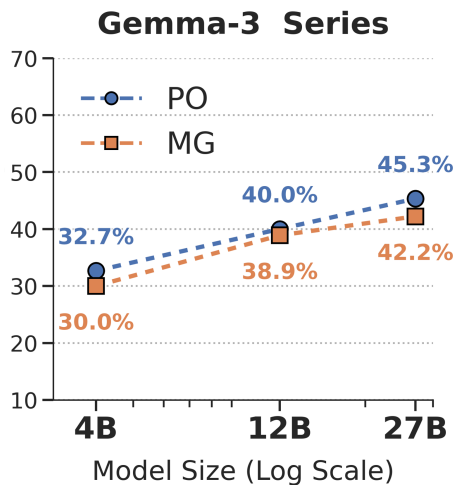


Figure 8: Action consistency of the Gemma 3 series in Partially Observable Games and Minority Games.

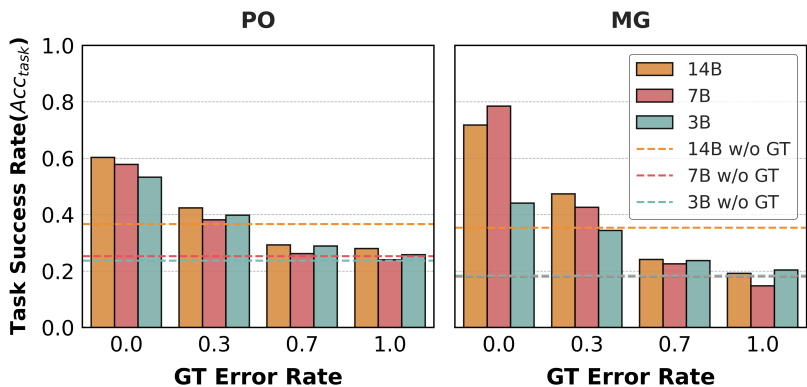


Figure 9: Average task success rates for the Qwen 2.5 series in Partially Observable Games and Minority Games, respectively, with ground truth ToM injection.

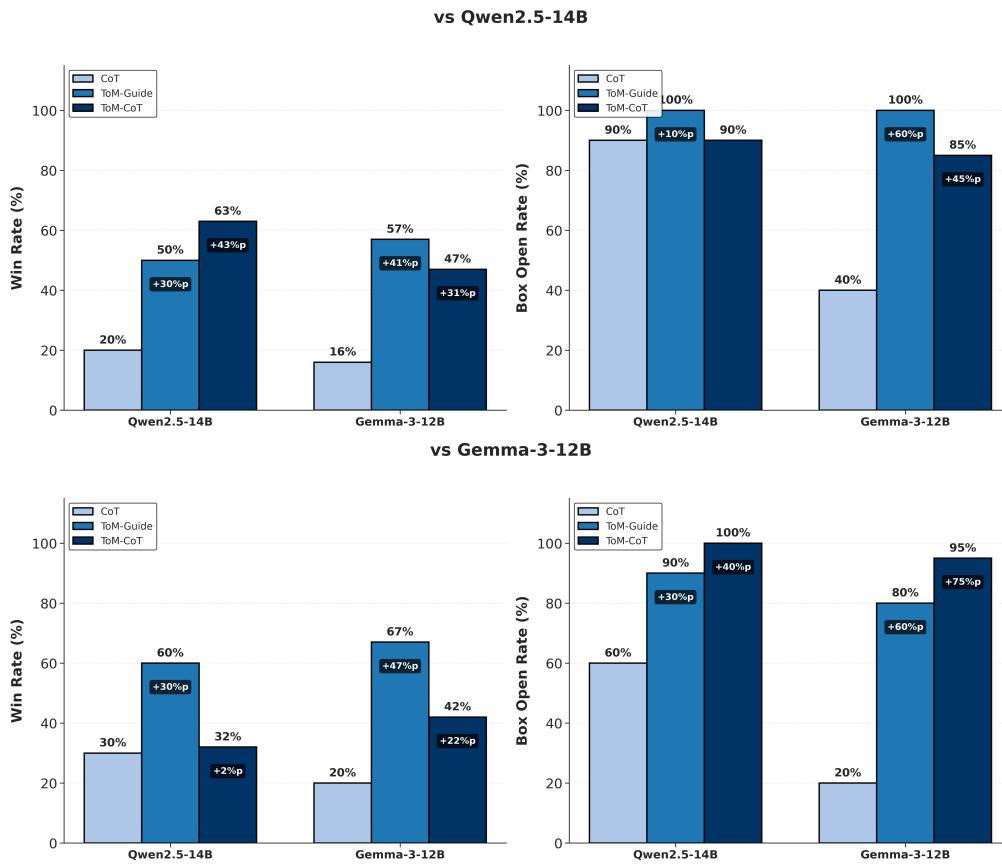


Figure 10: Average win rates and box open rates in the Speak or Wait game across various opponent and teammate models.

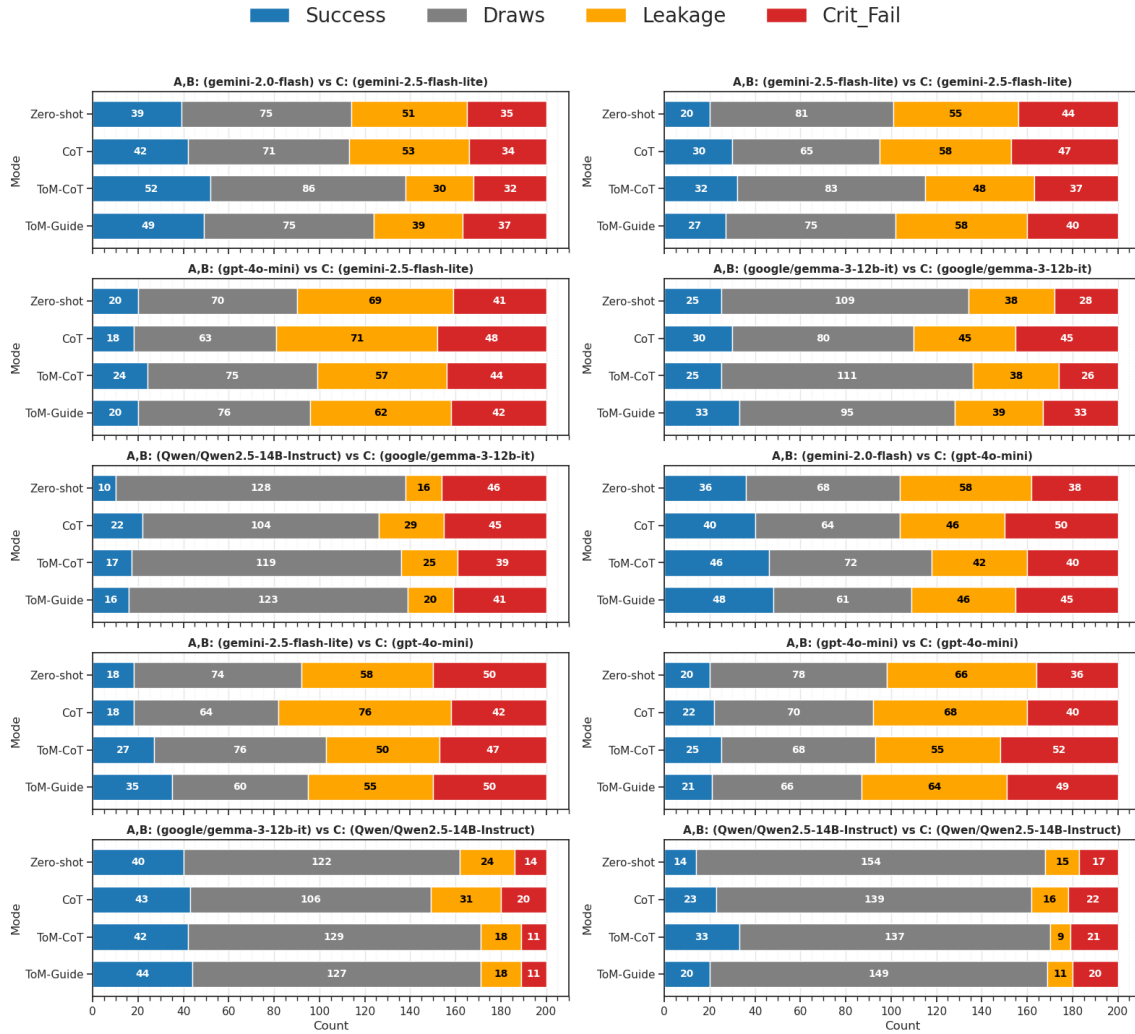


Figure 11: Analysis of execution results in the Ambiguity Game across various opponent and teammate models.

Table 9: Level-wise Task Success Rate ( $Acc_{task}$ ) on Partially Observable Games. Orange shading indicates the best result, and light blue shading indicates the second best.

Family	Model	Strategy	Total	L0	L1	L2	L3	L4
<b>Open Model</b>								
Qwen	Qwen2.5-3B-Instruct	Zero-shot	0.283	0.607	0.114	0.287	0.120	0.287
		Zero-shot-SC	0.284	0.600	0.120	0.280	0.100	0.320
		CoT	0.237	0.547	0.154	0.153	0.114	0.220
		ToM-CoT	0.301	0.433	0.367	0.167	0.227	0.313
		ToM-Guide	0.292	0.433	0.333	0.133	0.175	0.384
	Qwen2.5-7B-Instruct	Zero-shot	0.258	0.850	0.033	0.156	0.067	0.183
		Zero-shot-SC	0.224	0.862	0.033	0.133	0.067	0.200
		CoT	0.253	0.620	0.187	0.113	0.127	0.220
		ToM-CoT	0.365	0.333	0.473	0.280	0.267	0.473
		ToM-Guide	0.279	0.293	0.433	0.120	0.247	0.300
	Qwen2.5-14B-Instruct	Zero-shot	0.211	0.784	0.056	0.105	0.006	0.105
		Zero-shot-SC	0.213	0.800	0.067	0.100	0.000	0.100
		CoT	0.367	0.619	0.290	0.271	0.281	0.371
		ToM-CoT	0.495	0.384	0.728	0.450	0.422	0.494
		ToM-Guide	0.404	0.467	0.540	0.273	0.347	0.393
	Qwen2.5-32B-Instruct	Zero-shot	0.284	0.983	0.067	0.156	0.033	0.183
		Zero-shot-SC	0.289	0.980	0.067	0.167	0.033	0.200
		CoT	0.447	0.550	0.578	0.311	0.250	0.544
		ToM-CoT	0.512	0.367	0.673	0.433	0.600	0.487
		ToM-Guide	0.462	0.287	0.547	0.393	0.600	0.486
Gemma	gemma-3-4b-it	Zero-shot	0.255	0.535	0.147	0.087	0.160	0.233
		Zero-shot-SC	0.227	0.533	0.140	0.074	0.153	0.233
		CoT	0.255	0.524	0.281	0.086	0.157	0.228
		ToM-CoT	0.300	0.413	0.313	0.193	0.193	0.387
		ToM-Guide	0.338	0.584	0.340	0.167	0.240	0.360
	gemma-3-12b-it	Zero-shot	0.294	0.956	0.095	0.100	0.100	0.222
		Zero-shot-SC	0.299	0.967	0.093	0.100	0.100	0.233
		CoT	0.353	0.430	0.427	0.249	0.253	0.407
		ToM-CoT	0.432	0.500	0.640	0.200	0.280	0.540
		ToM-Guide	0.407	0.483	0.508	0.241	0.308	0.492
	gemma-3-27b-it	Zero-shot	0.262	0.967	0.033	0.133	0.006	0.172
		Zero-shot-SC	0.263	0.967	0.033	0.133	0.013	0.167
		CoT	0.417	0.467	0.613	0.272	0.368	0.367
		ToM-CoT	0.475	0.367	0.650	0.392	0.425	0.542
		ToM-Guide	0.448	0.467	0.609	0.350	0.367	0.450
<b>Closed Model</b>								
Gemini	gemini-2.0-flash	Zero-shot	0.410	0.893	0.380	0.340	0.200	0.233
		Zero-shot-SC	0.441	0.867	0.480	0.313	0.274	0.274
		CoT	0.541	0.633	0.527	0.360	0.480	0.707
		ToM-CoT	0.565	0.447	0.740	0.473	0.560	0.607
		ToM-Guide	0.626	0.479	0.787	0.560	0.720	0.580
	gemini-2.5-flash-lite	Zero-shot	0.394	0.813	0.207	0.127	0.327	0.493
		Zero-shot-SC	0.433	0.961	0.195	0.120	0.290	0.600
		CoT	0.330	0.705	0.267	0.056	0.289	0.333
		ToM-CoT	0.556	0.427	0.713	0.253	0.840	0.547
		ToM-Guide	0.629	0.527	0.753	0.440	0.740	0.687
GPT	gpt-4o-mini	Zero-shot	0.186	0.750	0.039	0.022	0.000	0.117
		Zero-shot-SC	0.189	0.760	0.013	0.047	0.007	0.120
		CoT	0.301	0.461	0.350	0.250	0.206	0.239
		ToM-CoT	0.464	0.360	0.660	0.420	0.440	0.440
		ToM-Guide	0.420	0.280	0.553	0.400	0.433	0.433
	gpt-5.1	Zero-shot	0.301	0.493	0.387	0.107	0.133	0.387
		Zero-shot-SC	0.309	0.560	0.400	0.120	0.113	0.353
		CoT	0.312	0.647	0.320	0.087	0.107	0.400
		ToM-CoT	0.709	0.473	0.787	0.853	0.840	0.593
		ToM-Guide	0.639	0.473	0.787	0.700	0.800	0.433

Table 10: Level-wise Task Success Rate ( $Acc_{task}$ ) on Minority Games. Orange shading indicates the best result, and light blue shading indicates the second best.

Family	Model	Strategy	Total	L1	L2	L3
<b>Open Model</b>						
Qwen	Qwen2.5-3B-Instruct	Zero-shot	0.216	0.300	0.155	0.195
		Zero-shot-SC	0.235	0.307	0.167	0.233
		CoT	0.183	0.216	0.106	0.228
		ToM-CoT	0.258	0.293	0.233	0.247
		ToM-Guide	0.276	0.293	0.227	0.307
	Qwen2.5-7B-Instruct	Zero-shot	0.152	0.237	0.082	0.137
		Zero-shot-SC	0.152	0.242	0.092	0.125
		CoT	0.181	0.213	0.150	0.179
		ToM-CoT	0.369	0.407	0.280	0.420
		ToM-Guide	0.318	0.333	0.253	0.367
	Qwen2.5-14B-Instruct	Zero-shot	0.212	0.298	0.189	0.150
		Zero-shot-SC	0.229	0.333	0.187	0.167
		CoT	0.353	0.361	0.333	0.360
		ToM-CoT	0.468	0.494	0.417	0.494
		ToM-Guide	0.420	0.489	0.322	0.450
	Qwen2.5-32B-Instruct	Zero-shot	0.226	0.333	0.133	0.211
		Zero-shot-SC	0.222	0.333	0.133	0.200
		CoT	0.391	0.467	0.287	0.420
		ToM-CoT	0.494	0.517	0.466	0.500
		ToM-Guide	0.547	0.558	0.533	0.550
Gemma	gemma-3-4b-it	Zero-shot	0.180	0.300	0.150	0.089
		Zero-shot-SC	0.195	0.300	0.159	0.125
		CoT	0.241	0.233	0.261	0.228
		ToM-CoT	0.342	0.473	0.313	0.240
	gemma-3-12b-it	ToM-Guide	0.300	0.353	0.273	0.273
		Zero-shot	0.211	0.333	0.100	0.200
		Zero-shot-SC	0.211	0.333	0.100	0.200
		CoT	0.366	0.403	0.345	0.350
	gemma-3-27b-it	ToM-CoT	0.420	0.487	0.413	0.360
		ToM-Guide	0.589	0.767	0.475	0.525
		Zero-shot	0.207	0.305	0.183	0.133
		Zero-shot-SC	0.218	0.326	0.193	0.133
GPT	gpt-4o-mini	CoT	0.343	0.317	0.333	0.378
		ToM-CoT	0.476	0.513	0.427	0.487
		ToM-Guide	0.521	0.604	0.460	0.500
		Zero-shot	0.440	0.460	0.447	0.413
		Zero-shot-SC	0.474	0.460	0.440	0.520
GPT	gpt-5.1	CoT	0.409	0.400	0.433	0.395
		ToM-CoT	0.578	0.600	0.607	0.527
		ToM-Guide	0.660	0.713	0.627	0.640
		Zero-shot	0.278	0.346	0.260	0.227
		Zero-shot-SC	0.198	0.300	0.117	0.178
GPT	gpt-4o-mini	CoT	0.206	0.233	0.155	0.228
		ToM-CoT	0.493	0.578	0.450	0.450
		ToM-Guide	0.347	0.400	0.280	0.360
		Zero-shot	0.168	0.228	0.072	0.206
		Zero-shot-SC	0.182	0.260	0.047	0.240
GPT	gpt-5.1	CoT	0.296	0.334	0.233	0.322
		ToM-CoT	0.591	0.720	0.560	0.494
		ToM-Guide	0.482	0.527	0.440	0.480
		Zero-shot	0.260	0.233	0.200	0.347
		Zero-shot-SC	0.285	0.293	0.213	0.347
GPT	gpt-5.1	CoT	0.313	0.361	0.311	0.267
		ToM-CoT	0.373	0.400	0.380	0.340
		ToM-Guide	0.700	0.727	0.747	0.627