ICLR

Multi-Agent Debate with Memory Masking

Hongduan Tian[1],
Xiao Feng[1], Ziyuan Zhao[2], Xiangyu Zhu[2], Rolan Yan[2], Bo Han[1]
[1]TMLR Group, Hong Kong Baptist University
[2]Search Algorithm Group, WeChat, Tencent

TMLR
TRUSTWORTHY MACHINE LEARNING AND REASONING

香港浸會大學
HONG KONG BAPTIST UNIVERSITY

WeChat

ICLR 2026 Paper: https://arxiv.org/pdf/2603.20215

GitHub Repository: https://github.com/tmlr-group/MAD-MM

Outline

- Introduction of Multi-Agent Debate
- Motivation: Erroneous Memories Deteriorate MAD
- Multi-Agent Debate with Memory Masking (MAD-M^2)
- Summary

Published as a conference paper at ICLR 2026

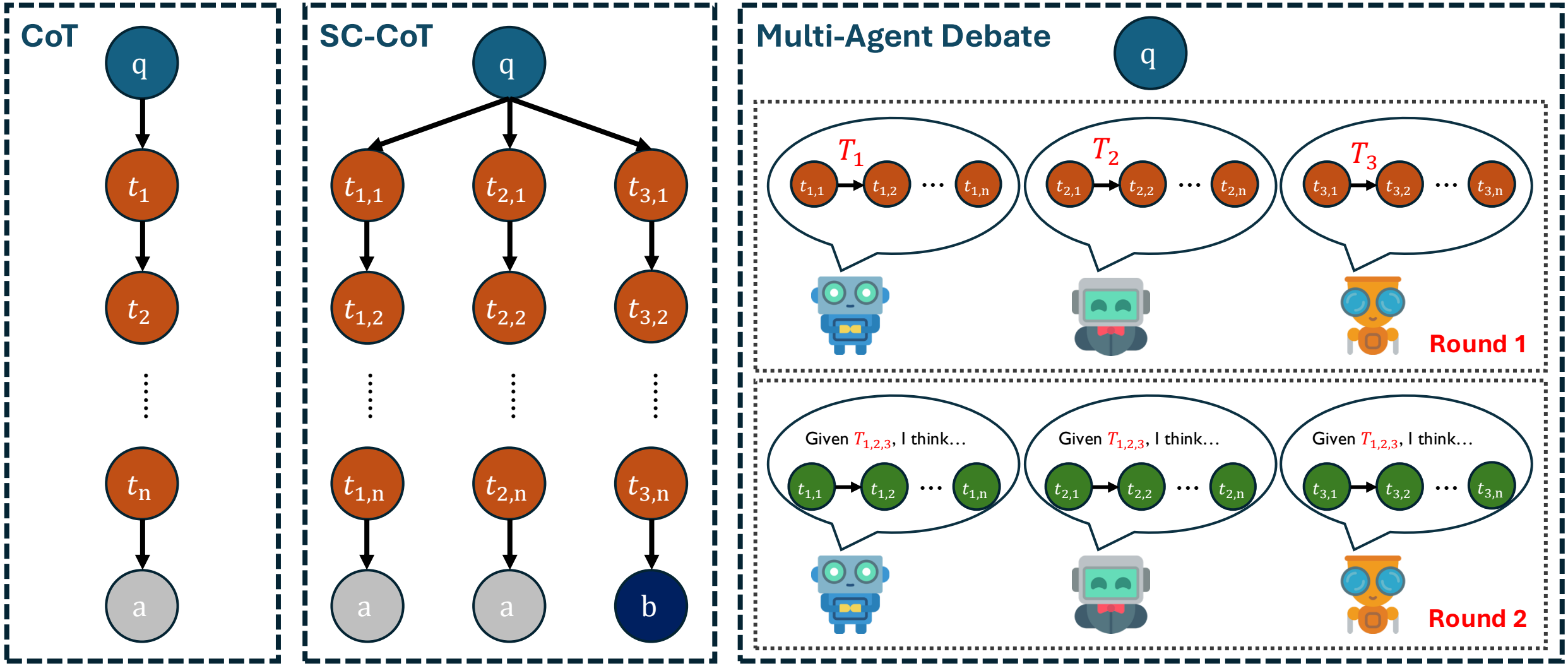
MULTI-AGENT DEBATE WITH MEMORY MASKING

Hongduan Tian[1] **Xiao Feng**[1] **Ziyuan Zhao**[2] **Xiangyu Zhu**[2] **Rolan Yan**[2] **Bo Han**[1] †
[1]TMLR Group, Hong Kong Baptist University [2]Search Algorithm Group, WeChat, Tencent
{cshdtian, xiaofeng, bhanml}@comp.hkbu.edu.hk
{joshuazhao, xiangyuzhu, rolanyan}@tencent.com

Outline

- Introduction of Multi-Agent Debate
- Motivation: Erroneous Memories Deteriorate MAD
- Multi-Agent Debate with Memory Masking (MAD-M^2)
- Summary

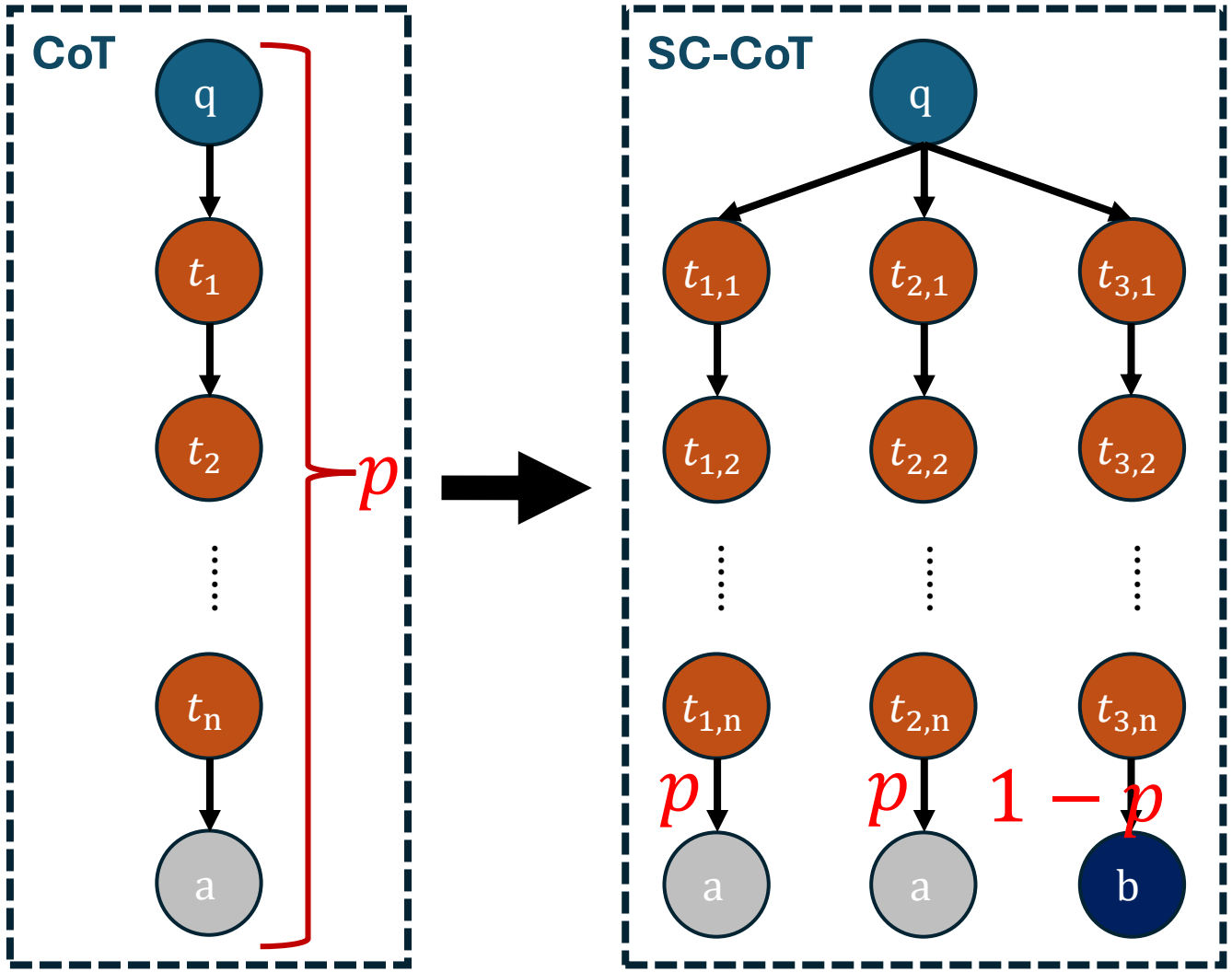
Multi-Agent Debate: A Reasoning Paradigm



Outline

- Introduction of Multi-Agent Debate
- Motivation: Erroneous Memories Deteriorate MAD
- Multi-Agent Debate with Memory Masking (MAD-M^2)
- Summary

Analysis on SC-CoT



Assume the probability that an agent can generate the correct answer of the given query q is p and N_{sc} answers are sampled, then the probability that the final answer is correct is bounded as

hard reasoning

$$P(N_{\text{cor}} > \frac{N_{\text{sc}}}{2}) \leq \exp\left(-2N_{\text{sc}}(\frac{1}{2}-p)^2\right), \qquad p < \frac{1}{2},$$

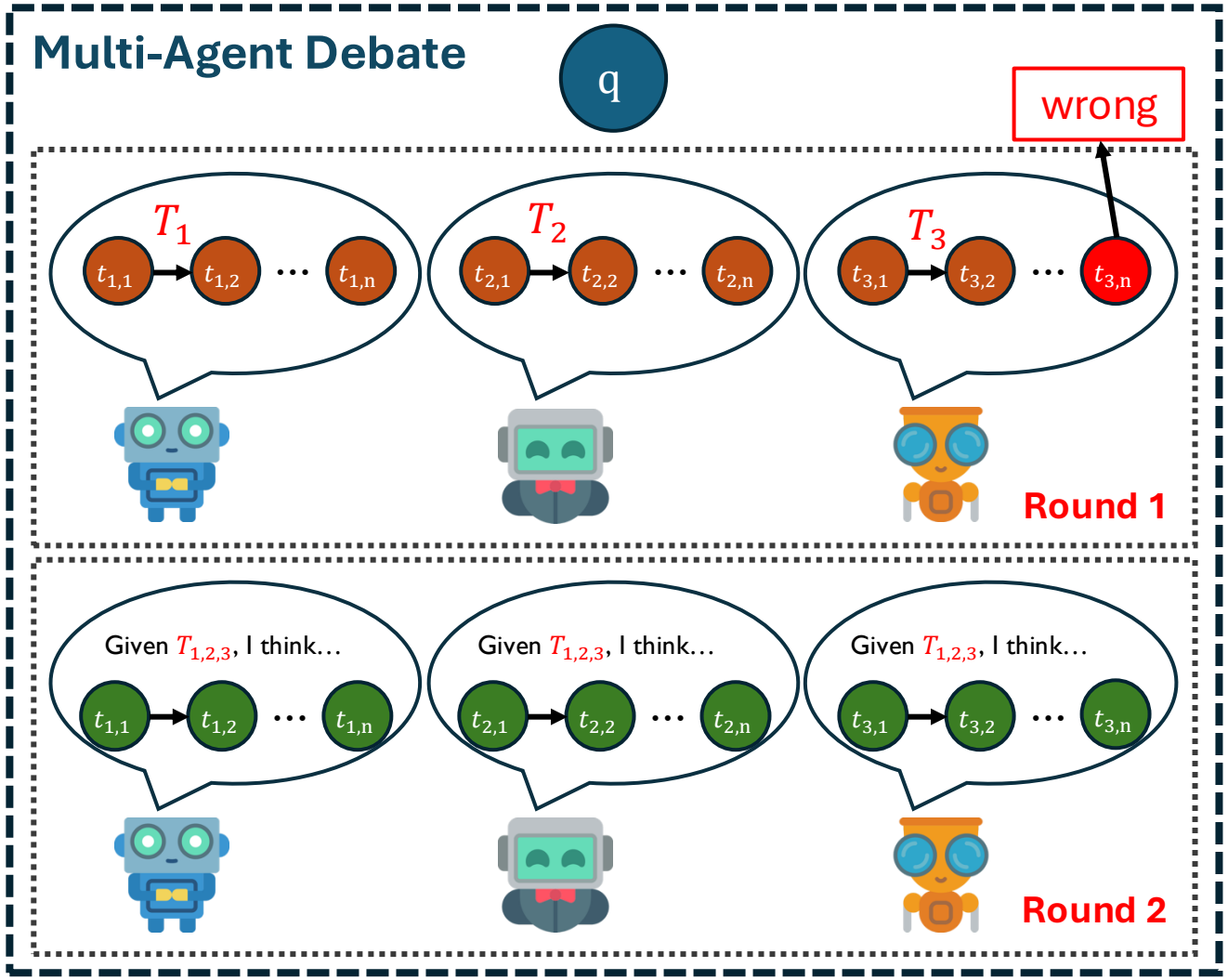
$$P(N_{\text{cor}} > \frac{N_{\text{sc}}}{2}) \geq 1-\exp\left(-2N_{\text{sc}}(\frac{1}{2}-p)^2\right), \quad p \geq \frac{1}{2}.$$

easy reasoning

Observation 1: SC-CoT fails to achieve good performance in all cases.

Observation 2: Increasing the sampling size ***deteriorates*** the performance of SC-CoT, when the reasoning is hard.

Multi-Agent Debate



- Consider an MAD framework with N_a agents and 2 debate rounds.
- **Assume** the probability that an agent can generate the correct answer in the zero-shot way of the given query q is p.
- **Assume** the probability that an agent can infer the correct answer based on a set of memories with j correct memories is $e^{-\alpha(N_a-j)}$.

Proposition 2.3 (MAD). *Consider a 2-round MAD reasoning, where N_a agents are involved in each debate round. With Assumption 2.1, the probability that the final answer is correct is bounded by:*

$$P(N_{cor}^{(2)} > \frac{N_a}{2}) \leq \sum_{j=0}^{j^*} \omega_j \exp\left(-2N_a\left(\frac{1}{2}-e^{\alpha(j-N_a)}\right)^2\right) + \sum_{j=j^*+1}^{N_a} \omega_j, \quad e^{\alpha(j-N_a)} < \frac{1}{2},$$

$$P(N_{cor}^{(2)} > \frac{N_a}{2}) \geq \sum_{j=j^*+1}^{N_a} \omega_j \left(1-\exp\left(-2N_a\left(\frac{1}{2}-e^{\alpha(j-N_a)}\right)^2\right)\right), \quad e^{\alpha(j-N_a)} \geq \frac{1}{2},$$

where $\omega_j = \binom{N_a}{j} p^j (1-p)^{N_a-j}$, $j^* = \lfloor N_a - \frac{\ln 2}{\alpha} \rfloor$. For simplicity, the corresponding lower and upper bounds of situations $e^{\alpha(j-N_a)} < \frac{1}{2}$ and $e^{\alpha(j-N_a)} \geq \frac{1}{2}$ are trivial bounds 0 and 1, respectively.

Observation 3: The reasoning capability of MAD is "covered" by SC-CoT.

Observation 4: Increasing the number of agents does **NOT** help improve the performance of MAD, when the reasoning is hard.

Unreliable Memories in MAD

Gerry goes to sleep at 11 pm each night. But he doesn't sleep very well, so he awakens at a random time between 1:00 am and 3:59 am, with each minute equally likely (including 1:00 and 3:59). He then reads the hour and minute shown on his clock as a three digit number, so 2:56 am would be 256. What is the probability that that number will be divisible by 7?"

Debate Round 1

Agent1

- Gerry can wake up anytime between 1:00 and 3:59 → 180 possible minutes.
- Each minute is written as a three-digit number (e.g., 1:00 → 100).
- Count how many of these are divisible by 7:
  - 1:00–1:59 → 8 numbers,
  - 2:00–2:59 → 9 numbers,
  - 3:00–3:59 → 9 numbers.
- Total favorable = 26.
- Probability = 26/180 = 13/90.

Agent 2

- Gerry can wake up anytime from 1:00 to 3:59 → 180 possible minutes.
- Each time is written as a three-digit number (HMM).
- Find numbers divisible by 7 between 100 and 359.
- Smallest is 105, largest is 357 → total 37 valid times.
- Probability = 37/180.

Take both responses as references

Debate Round 2

Agent1

- Gerry can wake between 1:00 and 3:59 → 180 possible minutes.
- Each time is written as a three-digit number (100–359).
- Numbers divisible by 7 in this range go from 105 to 357 → 37 valid cases.
- Probability = 37/180.

Agent 2

- Gerry can wake between 1:00–3:59 → 180 total minutes (180 possible numbers from 100–359).
- Numbers divisible by 7 in this range are from 105 to 357.
- There are 37 such valid times.
- Probability = 37/180.

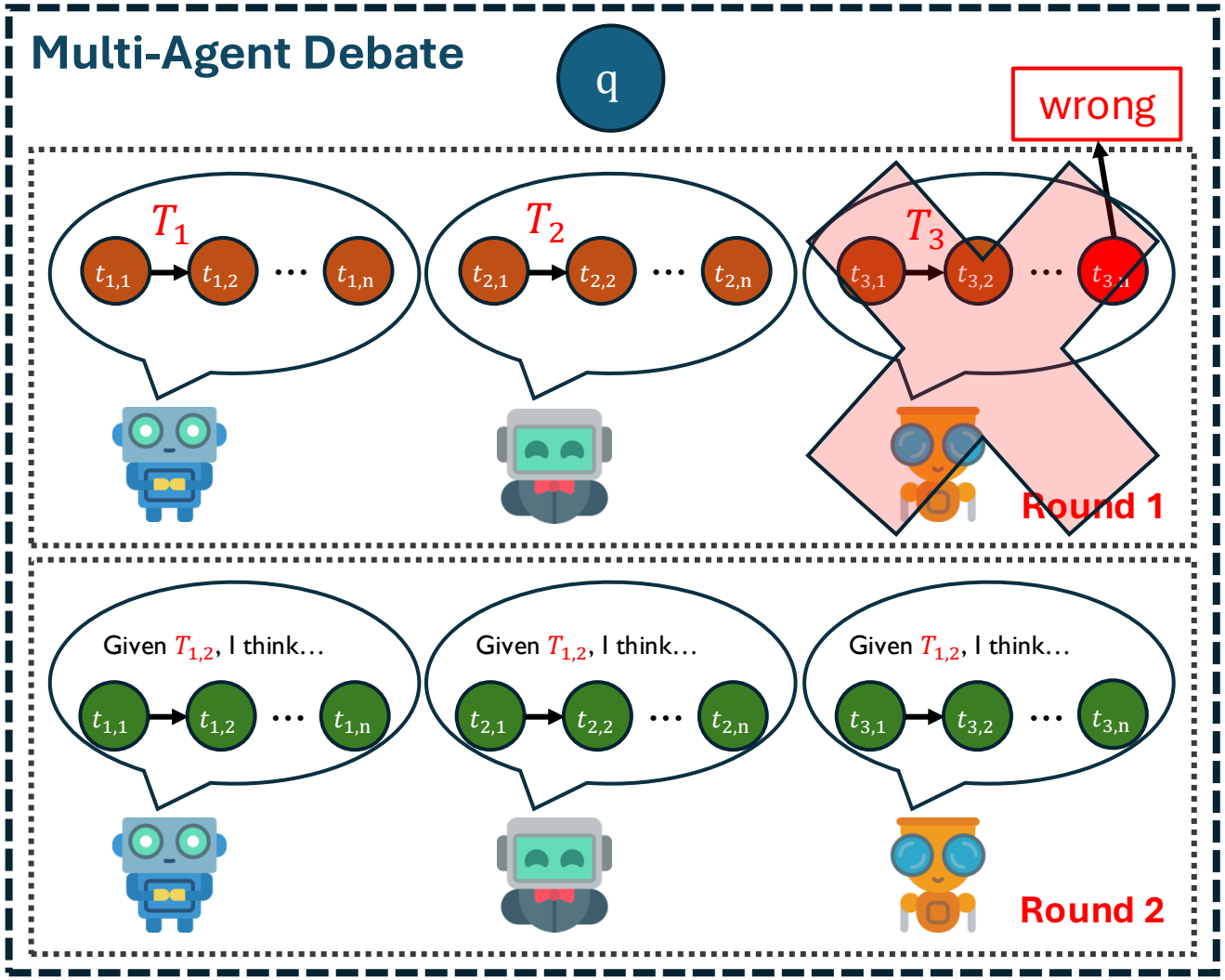
Agent 1 infers the answer in a more careful way, while agent 2 makes a mistake due to the rough reasoning

Empirical Observation: Unreliable memories among agents may ***misguide*** agents towards the wrong directions.

Outline

- Introduction of Multi-Agent Debate
- Motivation: Erroneous Memories Deteriorate MAD
- Multi-Agent Debate with Memory Masking (MAD-M^2)
- Summary

What if erroneous memories are removed...?



Proposition 2.3 (MAD). *Consider a 2-round MAD reasoning, where N_a agents are involved in each debate round. With Assumption 2.1, the probability that the final answer is correct is bounded by:*

$$P(N_{cor}^{(2)} > \frac{N_a}{2}) \leq \sum_{j=0}^{j^*} \omega_j \exp\left(-2N_a\left(\frac{1}{2} - e^{\alpha(j-N_a)}\right)^2\right) + \sum_{j=j^*+1}^{N_a} \omega_j, \quad e^{\alpha(j-N_a)} < \frac{1}{2},$$

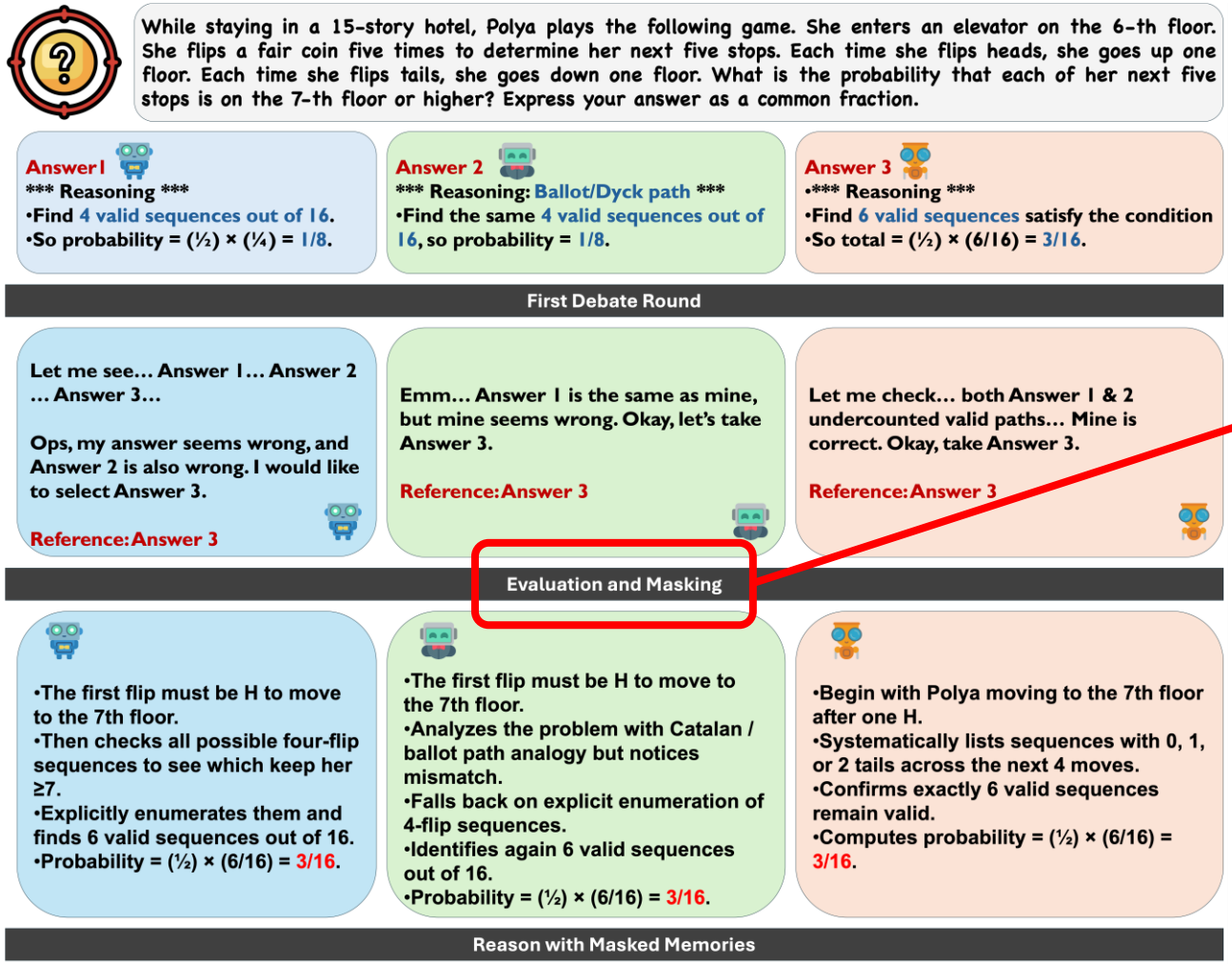
$$P(N_{cor}^{(2)} > \frac{N_a}{2}) \geq \sum_{j=j^*+1}^{N_a} \omega_j \left(1 - \exp\left(-2N_a\left(\frac{1}{2} - e^{\alpha(j-N_a)}\right)^2\right)\right), \quad e^{\alpha(j-N_a)} \geq \frac{1}{2},$$

where $\omega_j = \binom{N_a}{j} p^j (1-p)^{N_a - j}$, $j^* = \lfloor N_a - \frac{\ln 2}{\alpha} \rfloor$. For simplicity, the corresponding lower and upper bounds of situations $e^{\alpha(j-N_a)} < \frac{1}{2}$ and $e^{\alpha(j-N_a)} \geq \frac{1}{2}$ are trivial bounds 0 and 1, respectively.

1. The hard reasoning is turned into the easy reasoning, with $j \to N_a^{'}$.
2. In this case, the probability will approach to 1.

Better In-context Information

Multi-Agent Debate with Memory Masking



Step 1: At the first debate round, given a query, each agent generate its own answers;

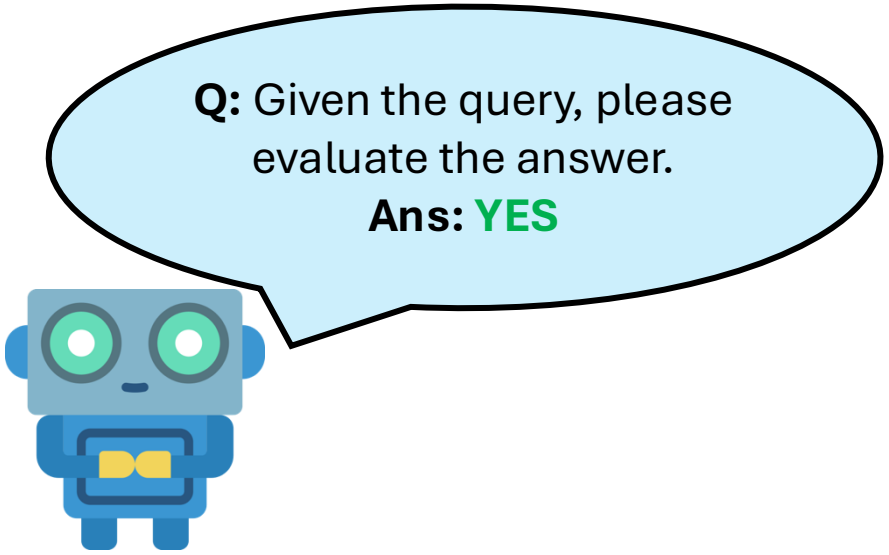
Step 2: *From the **second** debate round, each agent **evaluate**s the memories derived from the last debate round, and **mask**s those potential erroneous memories;*

Step 3: Based on the retained memories, each agent generate new answers towards the query.

Two Masking Strategies

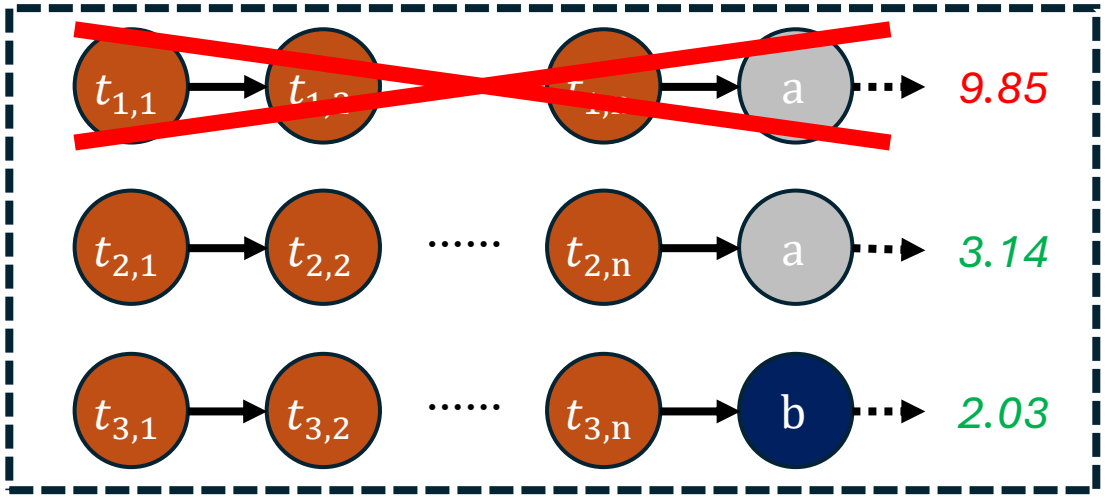
Subjective Masking

- Allow LLM Agents to evaluate the memories with the flags of "YES", "NO", and "NOT SURE".



Objective Masking

- Preserve the memories with low perplexity.



Empirical Results

Experimental Settings

- **Models:**
  - **General Model:** Qwen2.5-7B
  - **Math Model:** Qwen2.5-Math-7B, DeepSeek-Math-7B, QwQ-32B
- **Benchmarks:**
  - **Math:** GSM8K, MATH, AIME 2024, AIME 2025
  - **Language Understanding:** MMLU_Pro,
- **Baselines:** CoT, SC-CoT, (Naive) MAD

Empirical Results

Main Results

Table 1: Empirical results of accuracy (with standard deviation) and token consumption ($T.$). We evaluate four mainstream open-source LLMs on both mathematical reasoning and language understanding benchmarks. We highlight the best performance in **bold**, and the second-best performance in underline. For fairness, all results are the average of five trials on different seeds (i.e., 41-45).

Methods	AIME24		AIME25		MMLU_Pro		MATH		GSM8K	
	Acc. (%) ↑	$T.$ ↓	Acc. (%) ↑	$T.$ ↓	Acc. (%) ↑	$T.$ ↓	Acc. (%) ↑	$T.$ ↓	Acc. (%) ↑	$T.$ ↓
Qwen2.5-7B-Instruct										
CoT	3.3	×0.07	3.3	×0.08	36.0±6.3	×0.08	49.2±3.8	×0.07	64.8±5.4	×0.09
CoT-SC	10.0	×0.43	**10.0**	×0.45	39.2±4.0	×0.51	**58.0±1.0**	×0.45	83.6±3.9	×0.51
MAD	**13.3**	×1.00	6.7	×1.00	43.0±2.2	×1.00	55.6±3.7	×1.00	**91.8±2.4**	×1.00
MAD-M^2(S)	**13.3**	×1.13	3.3	×1.21	**43.6±2.3**	×1.17	56.8±2.1	×1.20	89.0±4.0	×1.25
MAD-M^2(O)	6.7	×0.68	6.7	×0.65	42.4±5.3	×0.69	54.2±2.6	×0.67	89.0±2.0	×0.72
Qwen2.5-Math-7B-Instruct										
CoT	13.3	×0.08	10.0	×0.08	39.6±0.9	×0.08	77.8±5.2	×0.08	95.2±1.6	×0.08
CoT-SC	**23.3**	×0.53	10.0	×0.48	**41.4±5.1**	×0.47	**82.0±4.7**	×0.44	**96.4±1.7**	×0.45
MAD	6.7	×1.00	6.7	×1.00	34.2±2.9	×1.00	71.2±3.3	×1.00	95.2±1.8	×1.00
MAD-M^2(S)	6.7	×1.37	6.7	×1.37	35.0±2.2	×1.36	71.2±3.3	×1.41	95.2±1.8	×1.44
MAD-M^2(O)	13.3	×0.67	**13.3**	×0.62	37.0±2.9	×0.62	80.2±3.8	×0.62	95.4±1.7	×0.60
DeepSeek-Math-7B-Instruct										
CoT	0.0	×0.07	0.0	×0.09	27.8±7.9	×0.17	34.2±4.5	×0.08	79.0±3.8	×0.09
CoT-SC	**3.3**	×0.44	0.0	×0.46	**32.2±4.3**	×0.99	**44.4±3.9**	×0.47	**88.8±2.6**	×0.52
MAD	0.0	×1.00	0.0	×1.00	31.2±5.4	×1.00	38.6±2.6	×1.00	81.2±2.7	×1.00
MAD-M^2(S)	0.0	×1.32	0.0	×1.30	30.8±5.2	×1.66	37.0±5.1	×1.31	80.8±3.5	×1.33
MAD-M^2(O)	0.0	×0.67	0.0	×0.67	30.8±6.4	×0.75	39.8±3.6	×0.68	82.2±4.4	×0.71
QwQ-32B										
CoT	**80.0**	×0.13	56.7	×0.14	75.2±4.9	×0.11	80.8±1.6	×0.10	97.4±2.3	×0.08
CoT-SC	**80.0**	×0.85	**80.0**	×0.85	**76.4±6.8**	×0.63	**81.6±0.9**	×0.61	97.4±2.3	×0.44
MAD	76.7	×1.00	73.3	×1.00	75.4±4.2	×1.00	79.2±2.8	×1.00	97.2±2.3	×1.00
MAD-M^2(S)	76.7	×1.28	73.3	×1.25	75.8±6.3	×1.22	79.6±2.3	×1.27	**97.8±1.9**	×1.32
MAD-M^2(O)	**80.0**	×0.67	76.7	×0.90	75.2±5.9	×0.69	75.0±3.9	×0.67	96.6±1.8	×0.56

Observation 1: MAD-M^2 can outperform naive MAD in almost all cases of models and datasets.

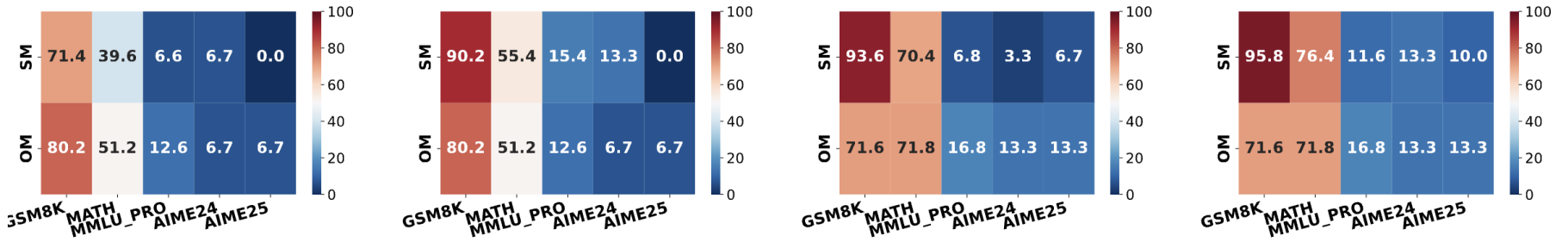
Observation 2: MAD with powerful LLM agents tends to achieve better performance in reasoning tasks.

Observation 3: Generally, powerful LLM agents prefer the objective masking strategy while relatively weak LLM agents prefer the subjective masking strategy.

Empirical Results

Erroneous Memory Detection

- ***Strict Rule:*** All erroneous are identified.
- ***Loose Rule:*** The correct memory takes a dominant place.

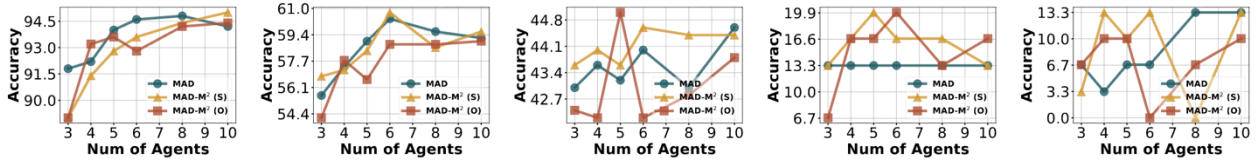


(a) Qwen2.5-7B (S) (b) Qwen2.5-7B (L) (c) Qwen2.5-Math-7B (S) (d) Qwen2.5-Math-7B (L)

- **Observation 1:** MAD-M^2 achieves better performance in erroneous memory identification on simple reasoning tasks (e.g., GSM8K & MATH) than on hard reasoning tasks (AIME).
- **Observation 2:** The subjective masking strategy helps ensure that correct memories take a dominant place (i.e., loose rule) in the memory set.

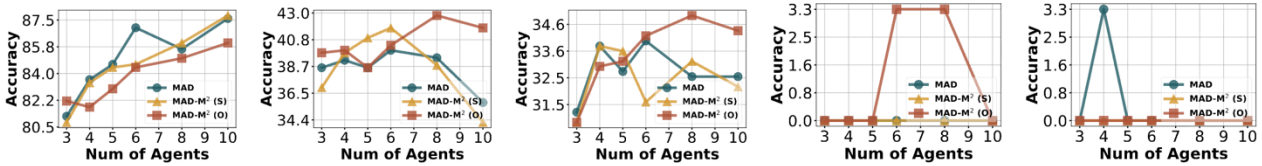
Empirical Results

Scaling Analysis of Agent Size



(a) GSM8K (b) MATH (c) MMLU_Pro (d) AIME24 (e) AIME25

Figure 4: **Effect of scaling the number of agents in the case of Qwen2.5-7B-Instruct.** The number of agents is increased from 3 to 10. According to the figures, both frameworks benefit from the increase of the number of agents and MAD-M^2(S) tends to achieve better performance in most cases.



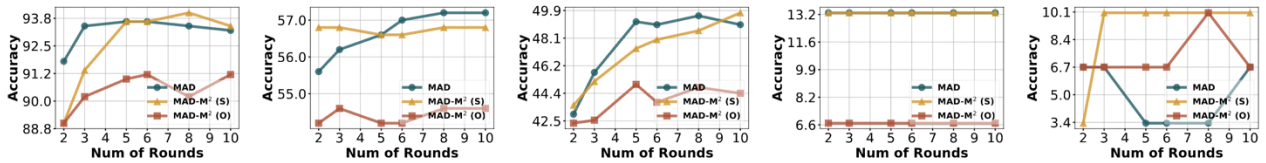
(a) GSM8K (b) MATH (c) MMLU_Pro (d) AIME24 (e) AIME25

Figure 5: **Effect of scaling the number of agents in the case of DeepSeek-Math-7B-Instruct.** The number of agents is increased from 3 to 10. According to the figures, both frameworks act differently when the number of agents increases and MAD-M^2(O) achieves better performance in most cases.

- **Observation 1:** All MAD frameworks generally benefit from the increasing of the number of agents.
- **Observation 2:** MAD with weak LLM agents benefits more from the subjective masking strategies, while MAD with powerful LLM agents benefits more from the objective masking strategy.

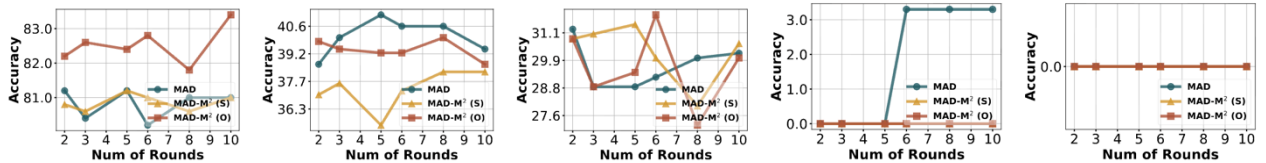
Empirical Results

Scaling Analysis of Debate Round



(a) GSM8K (b) MATH (c) MMLU_Pro (d) AIME24 (e) AIME25

Figure 6: **Effect of scaling the number of debate rounds in the case of Qwen2.5-7B-Instruct.** The number of debate rounds increases from 2 to 10. According to the figures, both frameworks basically benefit from the increase in debate rounds. In most cases, MAD-M^2(S) outperforms MAD-M^2(O).



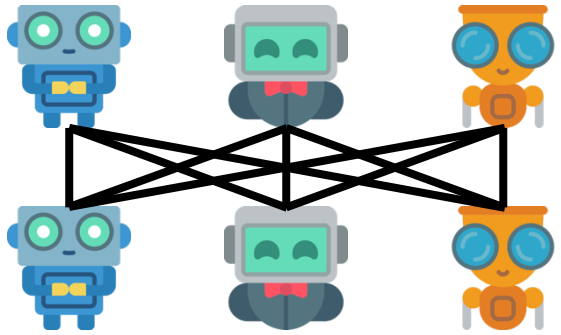
(a) GSM8K (b) MATH (c) MMLU_Pro (d) AIME24 (e) AIME25

Figure 7: **Effect of scaling the number of debate rounds w.r.t. DeepSeek-Math-7B-Instruct.** The number of debate rounds increases from 2 to 10. According to the figures, the performance of different cases varies. In most cases, MAD-M^2(O) achieves better performance than MAD-M^2(S).

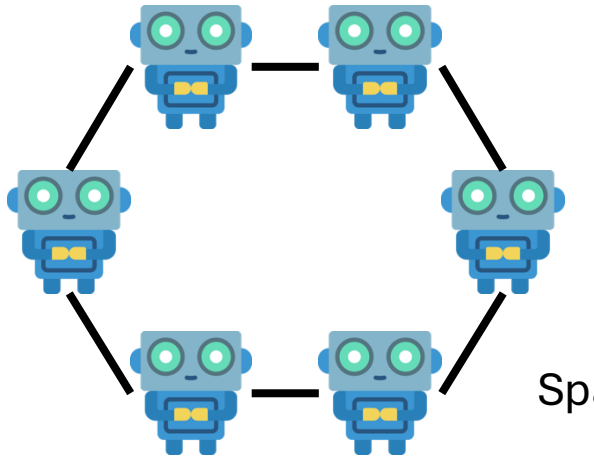
- **Observation 1:** Increasing the number of debate rounds does not consistently improve the performance MAD frameworks.
- **Observation 2:** MAD with weak LLM agents benefits more from the subjective masking strategies, while MAD with powerful LLM agents benefits more from the objective masking strategy.

Empirical Results

Comparison between MAD-M^2 and Sparse MAD



MAD-M^2



Sparse MAD

Table 3: **Comparison between MAD-M^{2}and Sparse MAD.** Average accuracy with standard deviation is reported. We compare MAD-M^2(S), MAD-M^2(O), and Sparse MAD on Qwen2.5-7B-Instruct and Qwen2.5-Math-7B-Instruct with both mathematical reasoning and language understanding benchmarks. For fairness, all results are the average of five trials on different seeds (i.e., 41-45).

Method	AIME24	AIME25	MMLU_Pro	MATH	GSM8K
Qwen2.5-7B-Instruct					
MAD-M^2(S)	16.7	**13.3**	**44.6±5.7**	**60.8±1.9**	93.6±1.1
MAD-M^2(O)	**20.0**	0.0	42.2±4.2	58.8±5.6	92.8±1.3
Sparse MAD	13.3	6.7	44.2±3.4	59.2±2.7	**94.2±1.9**
Qwen2.5-Math-7B-Instruct					
MAD-M^2(S)	**16.7**	13.3	66.2±4.6	62.8±4.3	**96.4±1.8**
MAD-M^2(O)	13.3	**16.7**	**67.4±7.0**	63.8±3.0	95.8±2.7
Sparse MAD	10.0	10.0	38.6±2.3	**76.4±3.8**	96.2±1.8

Observation: The ***dynamic*** MAD-M2 achieves better reasoning performance than the ***static*** Sparse MAD

Outline

- Introduction of Multi-Agent Debate
- Motivation: Erroneous Memories Deteriorate MAD
- Multi-Agent Debate with Memory Masking (MAD-M^2)
- Summary

Summary

- **Empirically**, we find that there exist erroneous memories in the typical multi-agent debate, and the erroneous memories will misguide other agents towards the wrong directions.
- **Theoretically**, we demonstrate that MAD is vulnerable to erroneous memories and removing the erroneous memories facilitates better reasoning performance.
- **Technically**, we propose a simple yet effective framework, MAD-M^2 to mask the potential erroneous memories in MAD.
- **Empirically**, extensive experiments and analyses indicate the efficacy of our proposed MAD-M^2 framework.

Thank You!

Paper

Codes