



Unreliable Memories among Agents



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

Agent 1

- 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

Agent 1

- 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 = $37/180$. ✗

The **erroneous memories** may **misguide** the agents towards the wrong directions.

Mathematical Justification

Assumption 2.1. Assume that the probability that an agent independently generates the correct answer only with the given query is $p \in [0, 1]$, while the probability that the agent generates correct answers to the query based on the previous memories is assumed to be $e^{-\alpha N_e}$, where N_e denotes the number of erroneous memories and $\alpha \in \mathbb{R}^+$ is a constant coefficient that indicates the robustness of agents to erroneous memories. The reasoning is deemed as successful when the number of correct answers N_{cor} in the final round satisfies $N_{cor} > \frac{N_{res}}{2}$, where N_{res} denotes the number of responses.

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.

The **erroneous memories** will significantly **deteriorate** the reasoning capability of MAD.

Multi-Agent Debate with Memory Masking



While staying in a 15-story hotel, Poly plays the following game. She enters an elevator on the 6-th floor. She flips a fair coin five times to determine her next five stops. Each time she flips heads, she goes up one floor. Each time she flips tails, she goes down one floor. What is the probability that each of her next five stops is on the 7-th floor or higher? Express your answer as a common fraction.

Answer 1

*** Reasoning ***

- Find 4 valid sequences out of 16.
- So probability = $(\frac{1}{2}) \times (\frac{1}{4}) = 1/8$.

Answer 2

*** Reasoning: Ballot/Dyck path ***

- Find the same 4 valid sequences out of 16, so probability = $1/8$.

Answer 3

*** Reasoning ***

- Find 6 valid sequences satisfy the condition
- So total = $(\frac{1}{2}) \times (6/16) = 3/16$.

First Debate Round

Let me see... Answer 1... Answer 2... Answer 3...

Ops, my answer seems wrong, and Answer 2 is also wrong. I would like to select Answer 3.

Reference: Answer 3

Emm... Answer 1 is the same as mine, but mine seems wrong. Okay, let's take Answer 3.

Reference: Answer 3

Let me check... both Answer 1 & 2 undercounted valid paths... Mine is correct. Okay, take Answer 3.

Reference: Answer 3

Evaluation and Masking

The first flip must be H to move to the 7th floor. Then checks all possible four-flip sequences to see which keep her ≥ 7 . Explicitly enumerates them and finds 6 valid sequences out of 16. Probability = $(\frac{1}{2}) \times (6/16) = 3/16$.

The first flip must be H to move to the 7th floor. Analyzes the problem with Catalan / ballot path analogy but notices mismatch. Falls back on explicit enumeration of 4-flip sequences. Identifies again 6 valid sequences out of 16. Probability = $(\frac{1}{2}) \times (6/16) = 3/16$.

Begin with Poly moving to the 7th floor after one H. Systematically lists sequences with 0, 1, or 2 tails across the next 4 moves. Confirms exactly 6 valid sequences remain valid. Computes probability = $(\frac{1}{2}) \times (6/16) = 3/16$.

Reason with Masked Memories

Details of Multi-Agent Debate with Memory Masking:

- Given a query X , at the first debate round, each agent A_{θ_i} generate its response $A_{\theta_i}(X)$. Then, the corresponding memory can be obtained $\mathcal{M}_1 = [A_{\theta_1}(X), \dots, A_{\theta_{N_a}}(X)]$
- From the **second** debate round, the memories obtained from the last debate will be evaluated by each agent, and a mask will be generated via a map $g_{A_{\theta_i}}^{\text{map}}: \mathcal{M}_r \rightarrow \{0, 1\}^{N_a}$ to mask the potential wrong memories: $\widehat{\mathcal{M}}_r^{(i)} = M^{(i)} \odot \mathcal{M}_r$, where $M^{(i)} = g_{A_{\theta_i}}^{\text{map}}(\mathcal{M}_r)$
- Based on the retained memories, a set of new responses are generated as the memories for the next round of debate or the final answers: $\mathcal{M}_{r+1} = [A_{\theta_1}(X, \widehat{\mathcal{M}}_r^{(1)}), \dots, A_{\theta_{N_a}}(X, \widehat{\mathcal{M}}_r^{(N_a)})]$

Two Masking Strategies

- Subjective Masking:** Agents evaluate each memory with three flags: "YES", "NO", and "NOT SURE". Then, memories with "YES" will be retained. Memories with "NOT SURE" will be retained in some cases.
- Objective Masking:** Memories with lower perplexity will be retained.

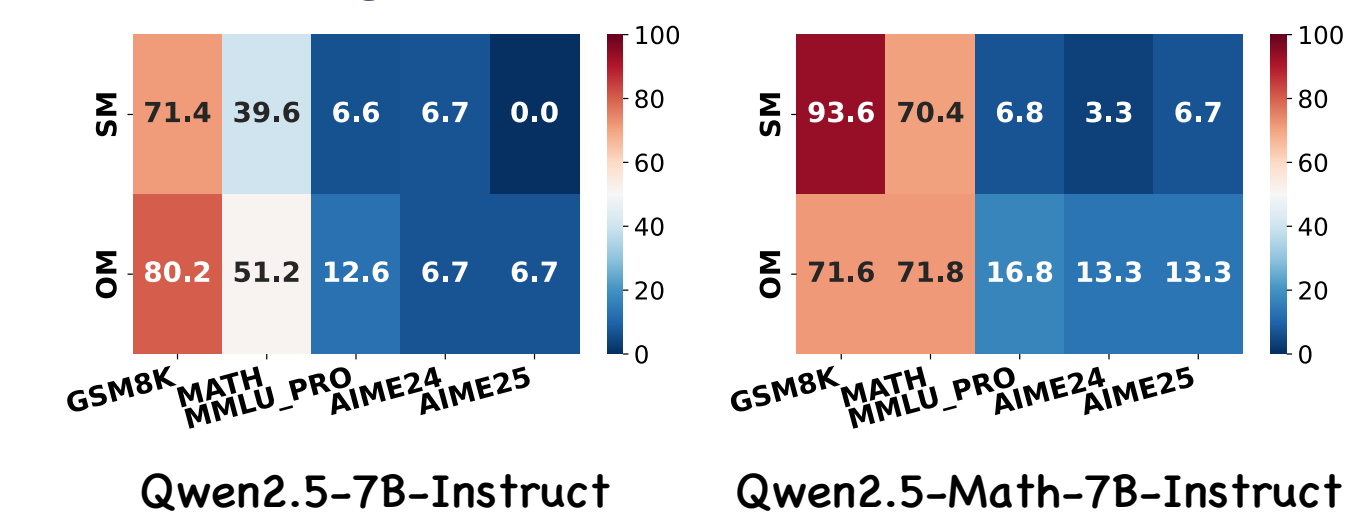
Experiments

Quantitative Results on Reasoning Benchmarks

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	<u>56.8±2.1</u>	×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	<u>37.0±2.9</u>	×0.62	80.2±3.8	×0.62	<u>95.4±1.7</u>	×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	<u>39.8±3.6</u>	×0.68	<u>82.2±4.4</u>	×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	<u>97.4±2.3</u>	×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	<u>75.8±6.3</u>	×1.22	<u>79.6±2.3</u>	×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

Investigation on Erroneous Memory Detection



- MAD-M² achieves better performance in identifying the erroneous memories on easy reasoning tasks;
- Powerful LLMs help identify erroneous memories more accurately.

Scaling Analyses

