

**Table 8:** The 2-th Reasoning Chain with Comments of Table 7

##### The 2-th Reasoning Chain with Comments ##### Below is a reasoning chain containing reasoning steps presented in order:  Step 1, Current set: 10 8, Selected two numbers: 10 8, Operation: /, Computed new number: 1.25, Remaining numbers: None, New set: 1.25. Evaluate Score: 0.8. Evaluate Score: 0.8 Step 2, Current set: 10 8, Selected two numbers: 10 8, Operation: /, Computed new number: 1.25, Remaining numbers: None, New set: 1.25.. Evaluate Score: 0.8 Step 3, Current set: 10 8, Selected two numbers: 10 8, Operation: /, Computed new number: 1.25, Remaining numbers: None, New set: 1.25.. Evaluate Score: 0.8  Analysis Report: This reasoning chain cannot complete the task and reach the target correctly. The reasoning steps provided are incorrect and do not lead to the solution of obtaining a result of 24. Reasoning Step 1: This reasoning step is incorrect. The selected two numbers, 10 and 8, are not present in the given number set of 3, 5, 6, and 8. Additionally, the operation of division (/) does not result in the computed new number of 1.25. The remaining numbers are also not correctly identified. Error Report: - The selected two numbers are not present in the given number set. - The operation and computed new number are incorrect. - The remaining numbers are not correctly identified. Advice: - Ensure that the selected two numbers are present in the given number set. - Use the correct operation that will result in the desired computed new number. - Accurately identify the remaining numbers. Recommendation or Reject Description: Reject. This reasoning step is incorrect and does not contribute to the correct solution. Reasoning Step 2: This reasoning step is a repetition of Reasoning Step 1 and is therefore incorrect. The same errors are present in this step as in the previous step. Error Report: - The selected two numbers are not present in the given number set. - The operation and computed new number are incorrect. - The remaining numbers are not correctly identified. Advice: - Ensure that the selected two numbers are present in the given number set. - Use the correct operation that will result in the desired computed new number. - Accurately identify the remaining numbers. Recommendation or Reject Description: Reject. This reasoning step is incorrect and does not contribute to the correct solution. Reasoning Step 3: This reasoning step is a repetition of Reasoning Step 1 and Reasoning Step 2. The same errors are present in this step as in the previous steps. Error Report: - The selected two numbers are not present in the given number set. - The operation and computed new number are incorrect. - The remaining numbers are not correctly identified. Advice: - Ensure that the selected two numbers are present in the given number set. - Use the correct operation that will result in the desired computed new number. - Accurately identify the remaining numbers. Recommendation or Reject Description: Reject. This reasoning step is incorrect and does not contribute to the correct solution. Confidence Score: 0.9
---

**Table 9:** Computation of edge weight between  $z_i$  (the step 2) and  $z_{i-1}$  (the step 3)

Prompt for the thought evaluation
In the game of 24, you are given four numbers, and the goal is to use basic arithmetic operations (+, -, *, /) to combine these numbers and obtain a result of 24. You can only use each number once, and parentheses can be used to change the order of operations. Analysis format of each step: Step , Current set: , Selected two numbers: , Operation: , Computed new number: , Remaining numbers: , New set: . The given four numbers are: 3 5 6 8. Let’s think step by step. Below are the generated reasoning steps, presented in order, accompanied by their evaluated scores (A higher score means the reasoning step is more likely to complete the task.):  Step 1, Current set: 3 5 6 8, Selected two numbers: 3 8, Operation: +, Computed new number: 11, Remaining numbers: 5 6, New set: 11 5 6.. Evaluate Score: 0.7 Step 2, Current set: 11 5 6, Selected two numbers: 5 6, Operation: *, Computed new number: 30, Remaining numbers: 11, New set: 30 11.. Evaluate Score: 0.9 Step 3, Current set: 30 11, Selected two numbers: 30 11, Operation: -, Computed new number: 19, Remaining numbers: , New set: 19.  What is your evaluation score for the logic, correctness, and benefit to reaching a final solution for these reasoning steps? Please select one value from [0.1, 0.3, 0.5, 0.7, 0.9, 1.0] as the score, where a higher score means better reasoning steps. The score should be placed after ‘Evaluation score:’ for users to read.
The generated evaluation socre
Evaluation score: 0.9.

**Table 10:** Format of prompts used in Game of 24.

Prompt Format for Next Thought Generation
In the game of 24, you are given four numbers, and the goal is to use basic arithmetic operations (+, -, *, /) to combine these numbers and obtain a result of 24. You can only use each number once, and parentheses can be used to change the order of operations. Analysis format of each step: Step , Current set: , Selected two numbers: , Operation: , Computed new number: , Remaining numbers: , New set: . The given four numbers are: 5, 6, 9, 4. Let’s think step by step. <b>Recall historical reasoning experience (Ignore when experience is empty):</b> ##### Pay attention to analysis and conclusion in the experience to avoid making similar mistakes by following the advice. Below is a list of ordered reasoning steps, accompanied by their evaluated scores (A higher score means the reasoning step is more likely to complete the task.):  Based on listed reasoning steps only within the above ”————” (i.e., Not the ones in the experience block), please make one step of reasoning to generate only one subsequential possible reasoning step.
Prompt Format for Experience Generation
Given task:In the game of 24, you are given four numbers, and the goal is to use basic arithmetic operations (+, -, *, /) to combine these numbers and obtain a result of 24. You can only use each number once, and parentheses can be used to change the order of operations. The given four numbers are: 4 5 10 11. Let’s think step by step. . Below is a reasoning chain containing reasoning steps presented in order:  Please evaluate this reasoning chain by giving detailed comments containing the following content: 1. Can this reasoning chain complete the task and reach the target correctly by executing its reasoning steps? why? Write a analysis report with conclusion under ‘Anlysis Report:’ .. 2. For each reasoning step, please provide a detailed analysis of whether the current step is a logical inference of the previous step and whether the reasoning step is beneficial to the correct solution. For each reasoning step with errors, please provide an error report and the corresponding advice on revision. For each reasoning step, please provide recommendation or rejection descriptions. Comments should be brief, avoid repeating the same analysis in different steps and follow the format: Reasoning step jidx <sub>z</sub> . Analysis report: . Advice: . Recommendation or Reject description: . . 3. What is your confidence score on these your evaluations and comments? Please select one value from [0.1, 0.3, 0.5, 0.7, 0.9, 1.0]. The score should be placed after ‘Confidence score:’ for users to read..