Systematic Failures: Long horizon contexts

Environment

Room Connectivity

Room 0 is connected to room 3 and 5. Room 3 is connected to room 0. Room 5 is connected to room 0.

Observation

You are now in Room 3 with Bomb 5. What is your next action?

Alpha

Action selection: Move to Room 5.

Message to Team: "I am moving to Room 5 to help Charlie with

Bomb 3."

Environment

Results: You can not directly move to Room 5 because it is not adjacent to your current location; Room 3. Consider taking a detour to another room first and then move to your destination.

ChatGPT: Negative ToM Inferences

Introspection

Yes, the current contents of Room 6 are Bomb 4 and Alpha.

1st-order ToM

No, player charlie does not know the current contents of room 6.

2rd-order ToM

No, player charlie is not aware that I know the current contents of room 6.

Emergent Collaborative Behaviors

Alpha

I am inspecting Bomb 1 in Room 0. Bravo; please move to Room 3. Charlie; please move to Room 5.

Bravo

I am moving to Room 3 as instructed. Alpha; please defuse Bomb 1 in Room 0 using the red tool.

Charlie

I am moving to Room 5 to search for more bombs.

GPT-4+Belief: Positive ToM Inferences

Introspection

Yes, the current contents of Room 5 are as follows: Player alpha (you) and Bomb 3.

1st-order ToM

No, Player Charlie does not know the current contents of Room 5 since they moved to Room 6. They are only aware of the information they had before leaving Room 5.

2rd-order ToM

Yes, player Charlie is aware that you know the current contents of Room 5 because you mentioned in your previous message that you were moving to Room 5 to assist with Bomb 3.