

Inferring Belief States in Partially-Observable Human-Robot Teams

Jack Kolb, Karen M. Feigh



People have an internal representation of the world, called a **mental model**.



When in teams, people **predict** the mental models of their teammates, forming a **team mental model**.



We use team mental models for a variety of downstream tasks:

Telegraphing.

Activity recognition.

Planning and navigation.

Communication.

... and much more.

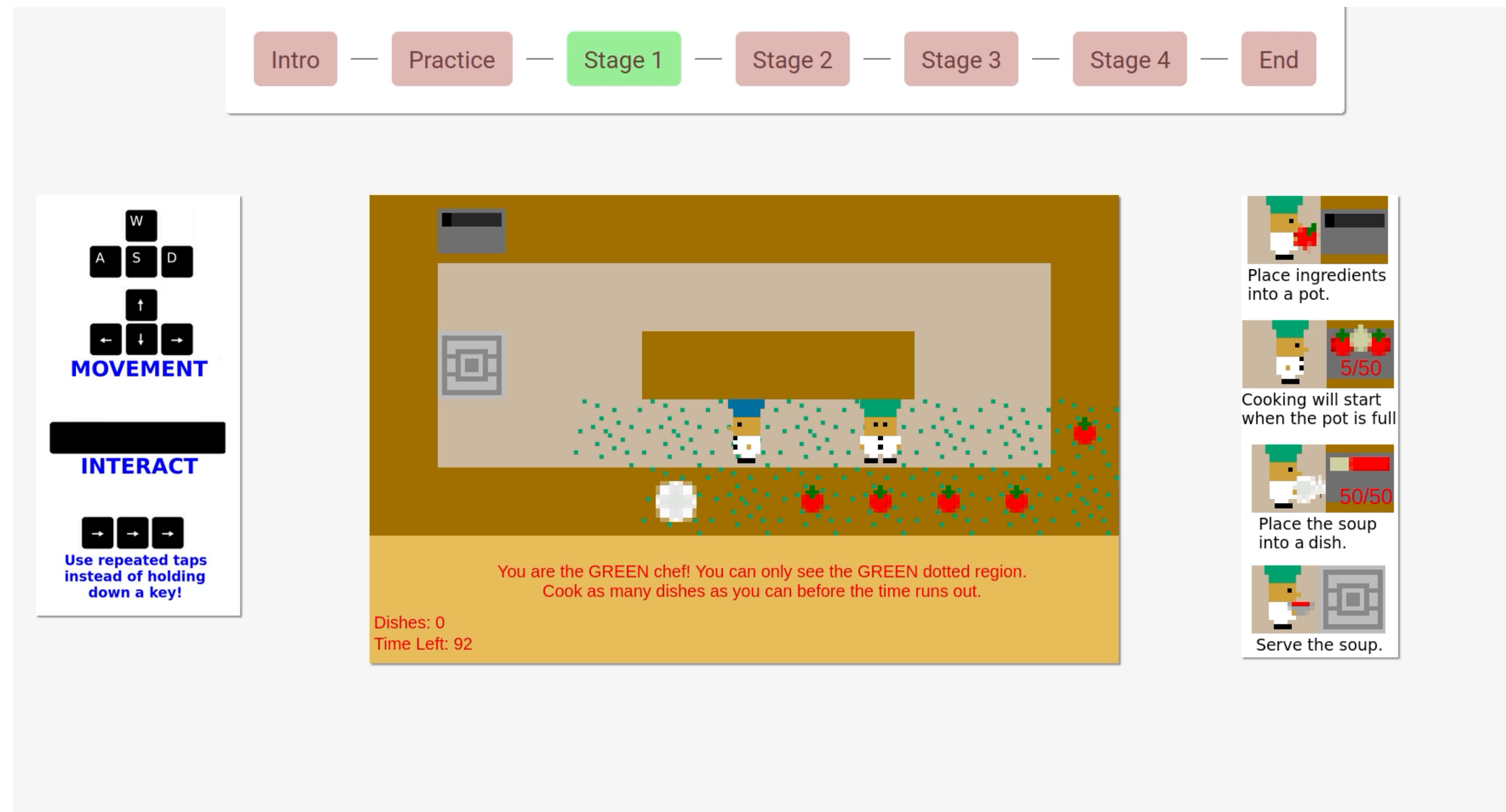


How can we give robots this capability?

How well can current methods predict user responses to situation awareness questions?

We ran a user study where participants played a 2D, collaborative, partially-observable cooking game with an autonomous agent.

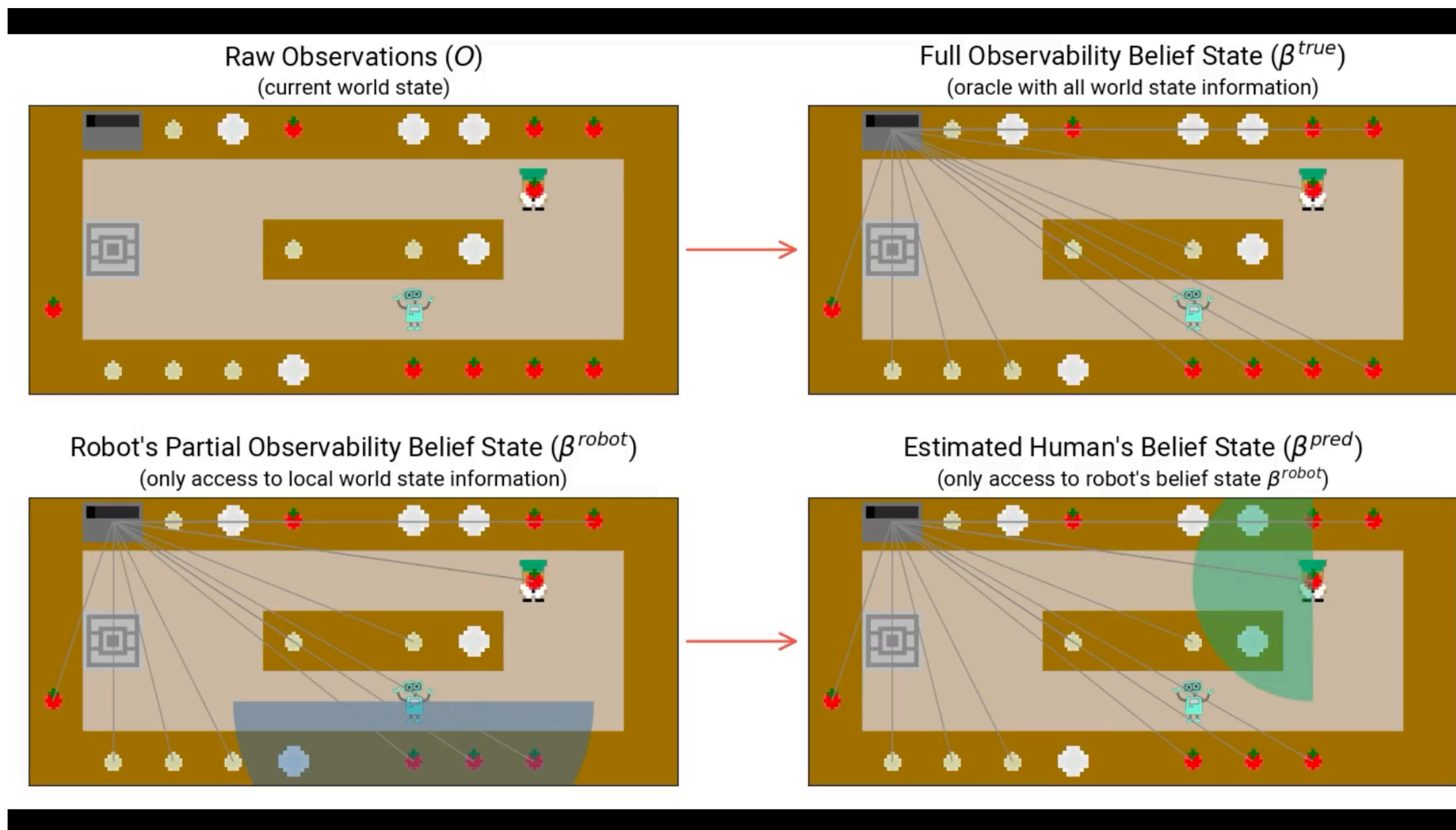
Participants were regularly asked situation awareness questions.



How well can current methods predict user responses to situation awareness questions?

In post-hoc playthroughs, we constructed a scene graph for the robot agent and predicted the scene graph for the human.

We then reasoned about what the user knew and predicted their responses.

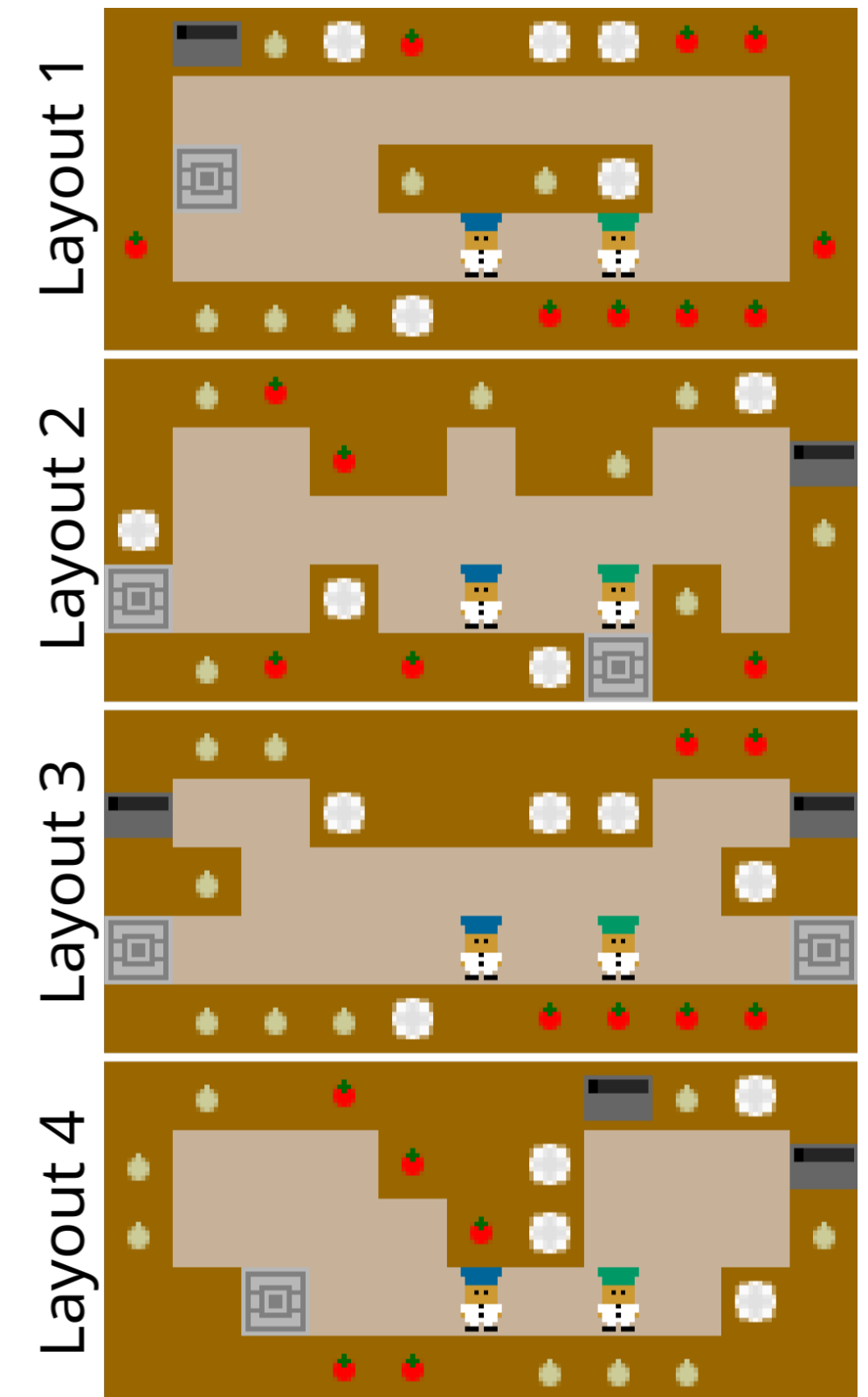
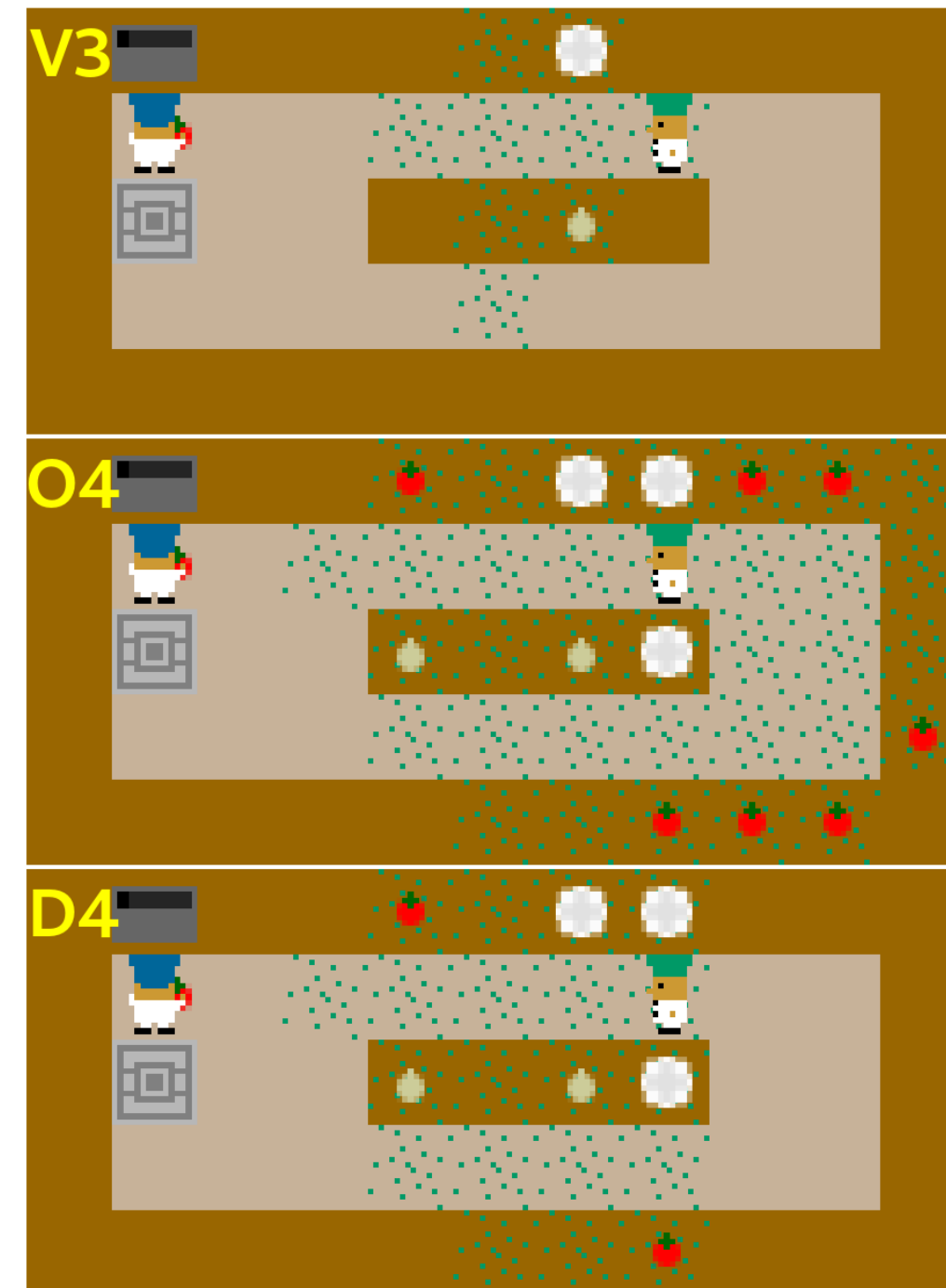


Results

We evaluated two methods for predicting user responses:

1. **Logical Predicates:** Hand-crafted rules used the user's scene graph to produce the best response to the situation awareness question.
2. **LLM:** Fed the scene graph and game description into GPT4 and asked it to choose the user's most likely response.

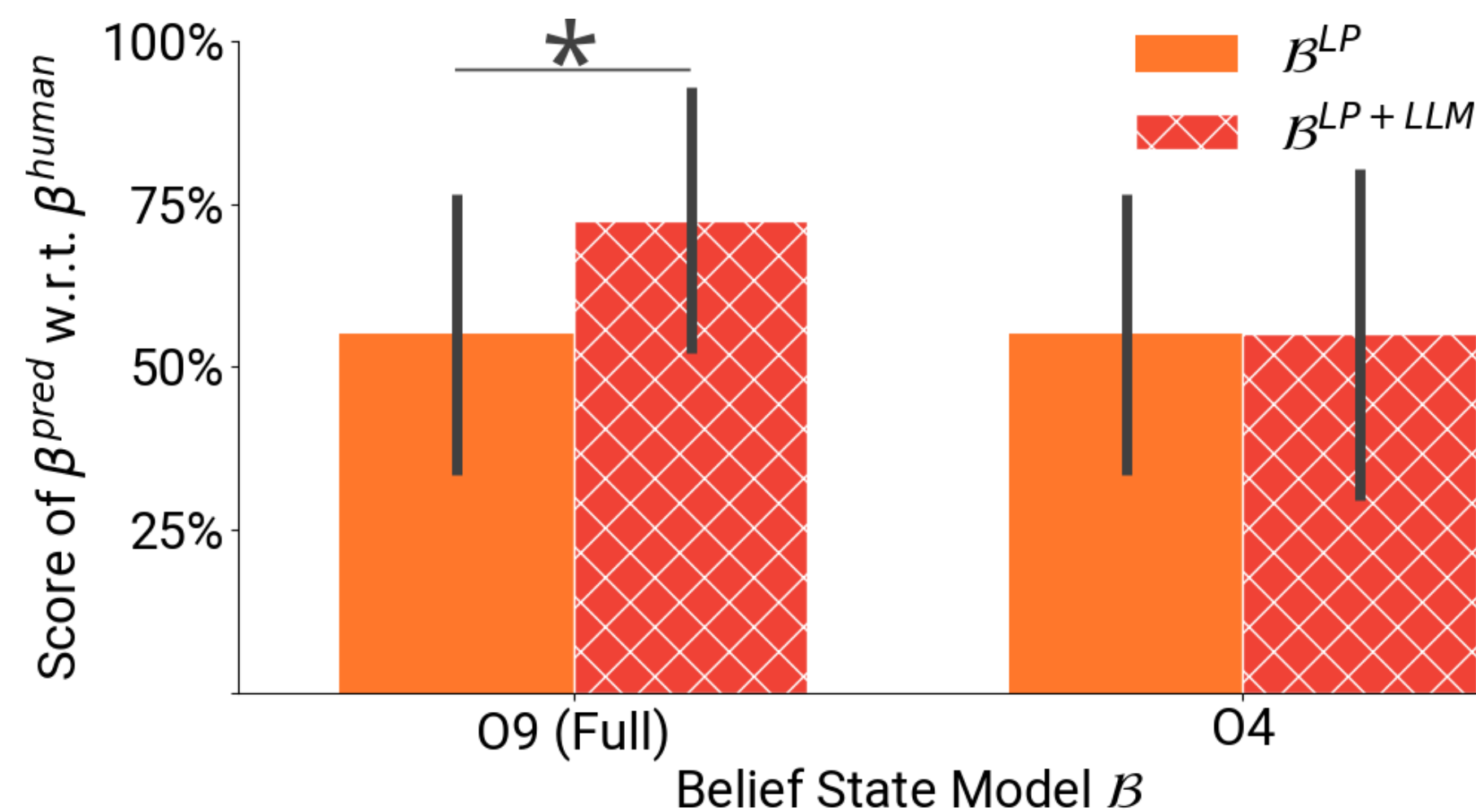
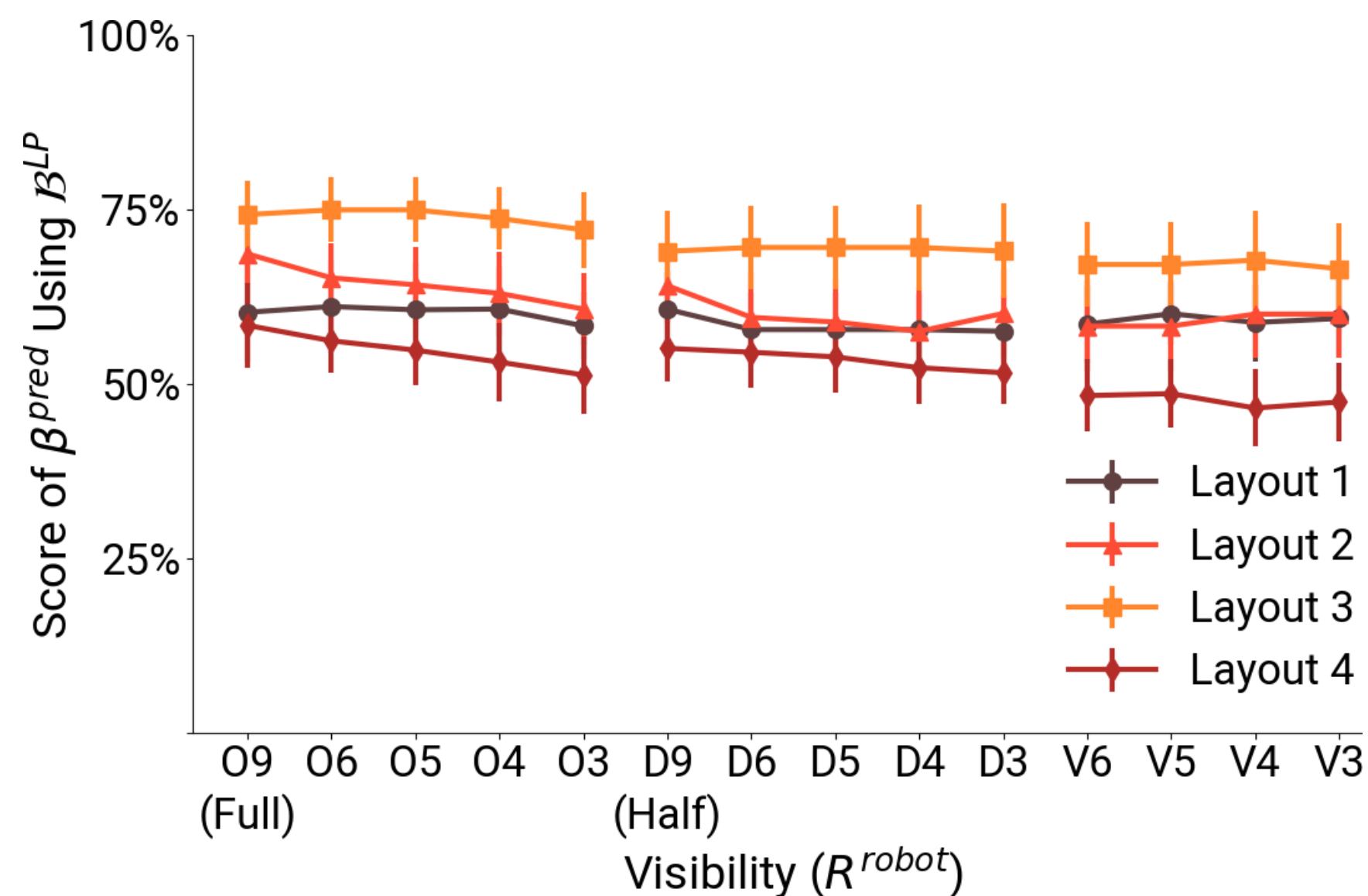
Across several types of observability and game layouts:



Results

Key takeaways include:

1. The models were only ~50-75% accurate at predicting user responses, **motivating future work** and opening ample **room for improvement**.
2. The two models were **resilient to low observability**, however the LLM model outperformed in high observability.



Our environment is easily accessible and modifiable!

All source code and user data is available at:

`https://github.com/gt-cec/tmm-hai`

