Carnegie Mellon University

Clarifying Feature Overspecification in Reward Learning from State Corrections via Follow Up Questions

Ethan Villalovoz[1], Michelle Zhao[2], Henny Admoni[2], Reid Simmons[2]
[1]Washington State University, [2]Carnegie Mellon University

THE ROBOTICS INSTITUTE

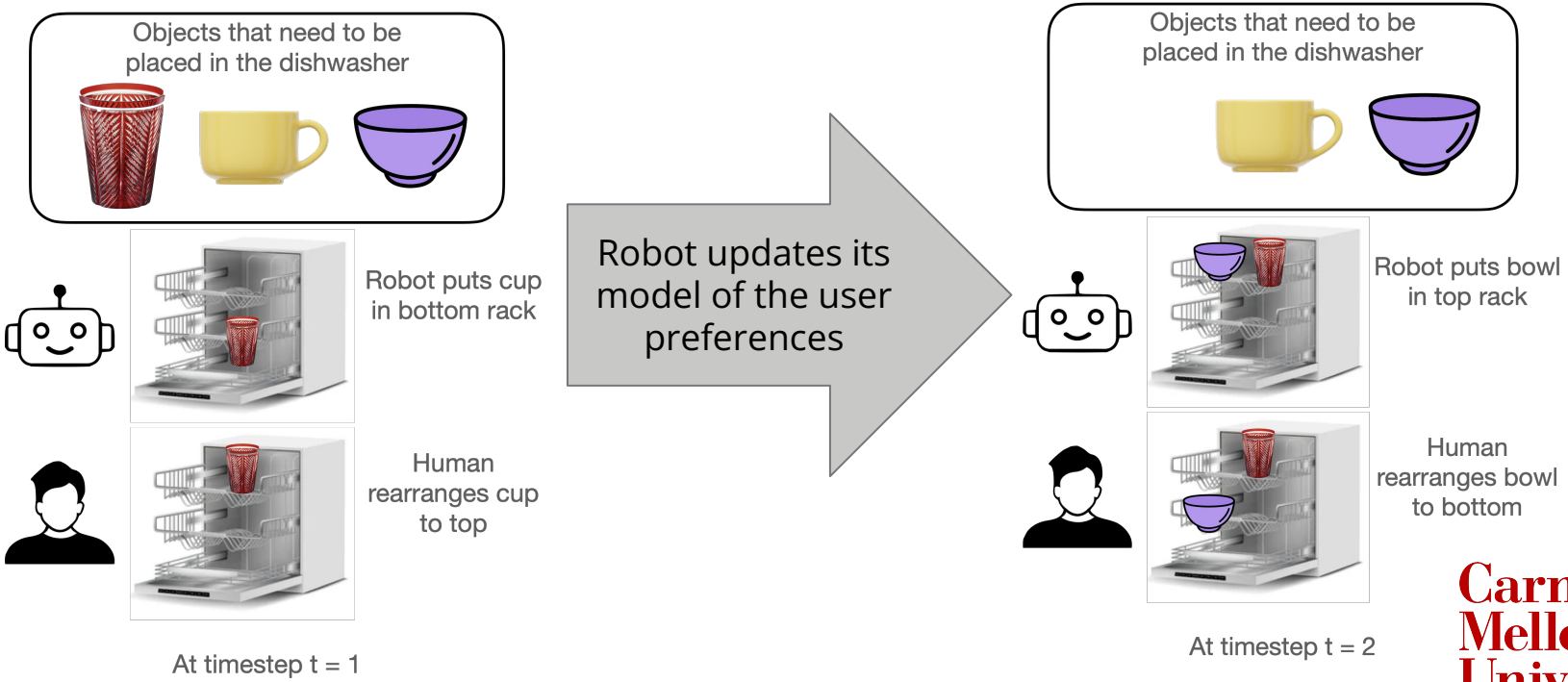
RISS | ROBOTICS INSTITUTE SUMMER SCHOLARS

Motivation

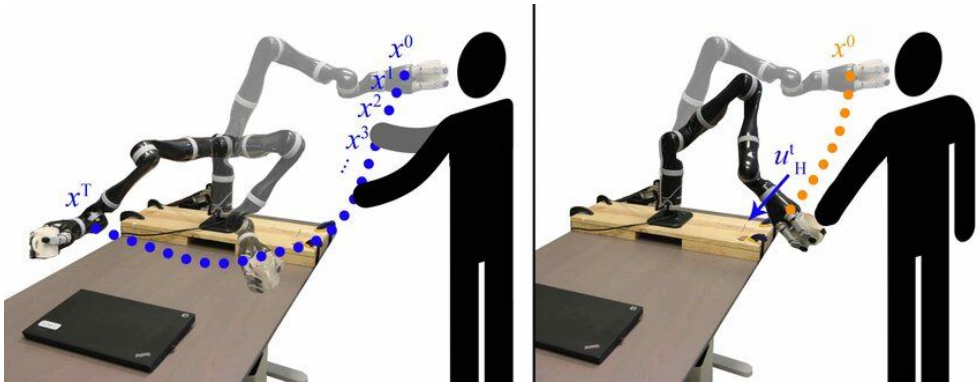
Robots designed for assisting people in household tasks need to know what people want!

Carnegie Mellon University

We envision an interaction in which the user can correct the state intermittently during the robot's task execution.



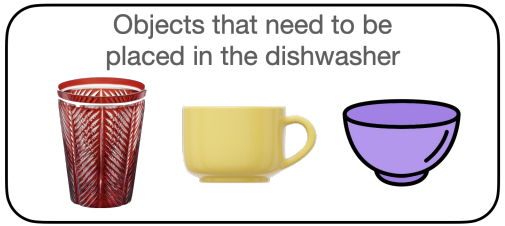
Conceptually, this differs from corrections to the robot's trajectory

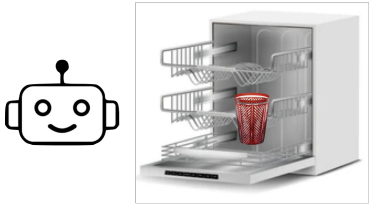


Bobu et al. Quantifying Hypothesis Space Misspecification in Learning from Human-Robot Demonstrations and Physical Corrections

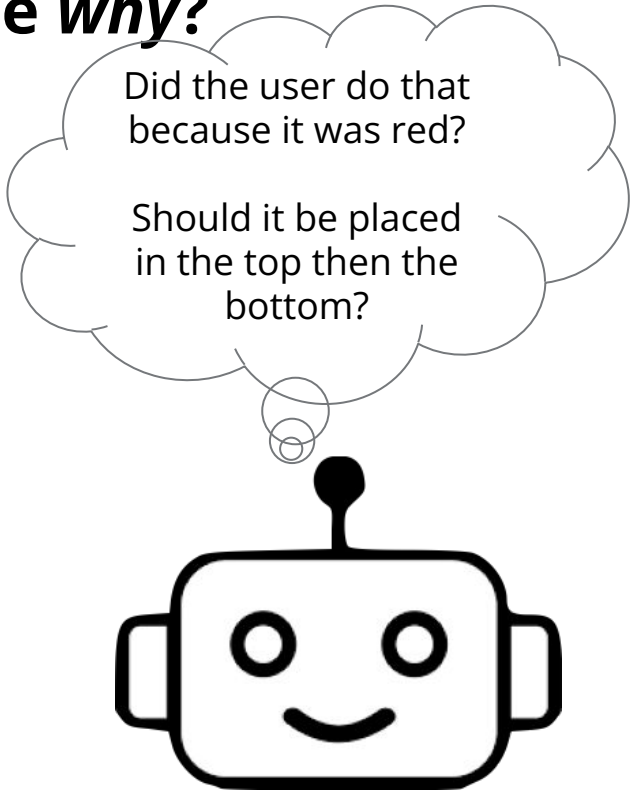
Carnegie Mellon University

Upon observing a state correction, the robot may not understand the *why*?





At timestep t = 1



Robot updates its model of the user preferences

Carnegie Mellon University

Research Question

Can we enable robot learning from iterative state corrections and bootstrap learning via clarification questions?

Contributions:

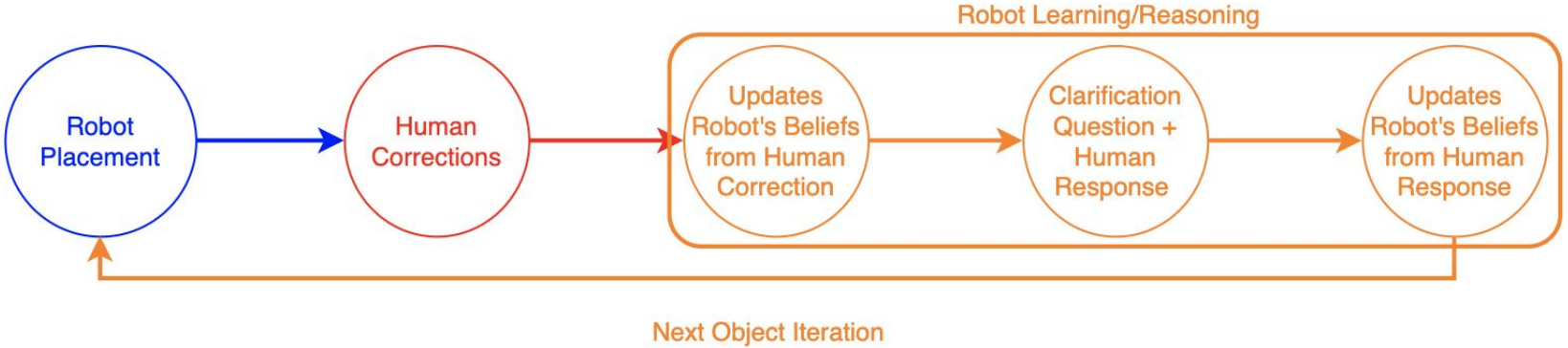
- Learning from Corrections:

  Robots learn from user state corrections to align with human preferences and enhance performance.

- Proactive Dialogue:

  Robots prompt for user guidance during uncertainty to improve efficiency and reduce errors.

Interactive Workflow

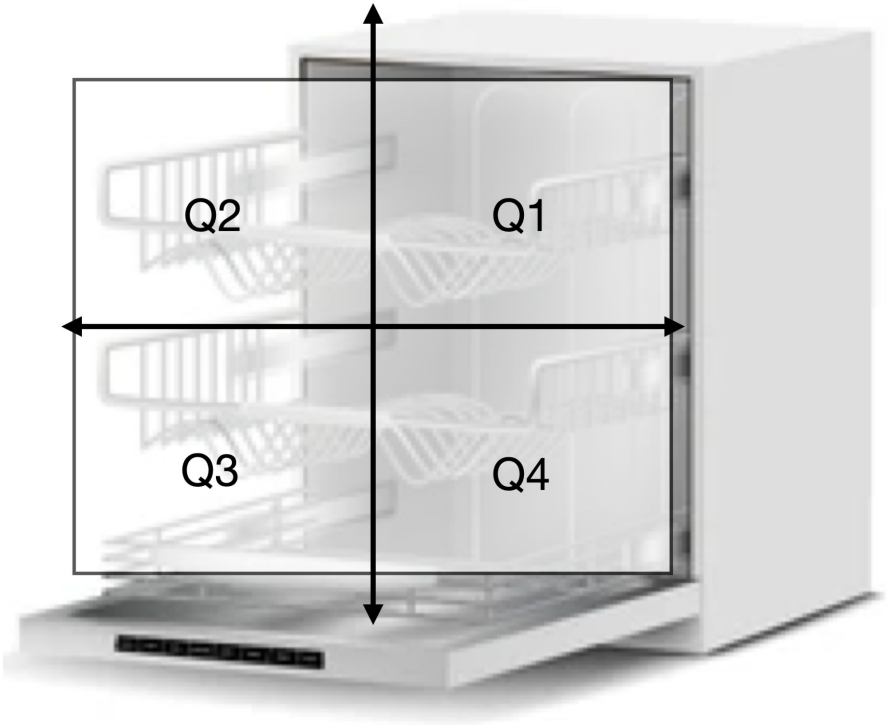


Carnegie Mellon University

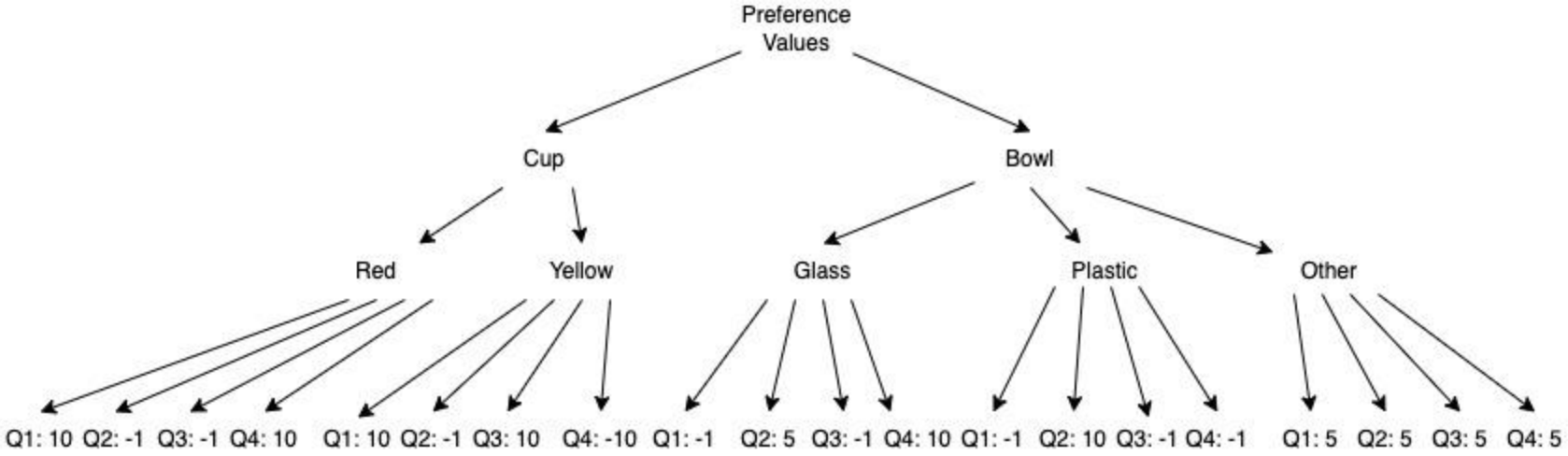
Interaction MDP

We represent the dishwasher as a grid.

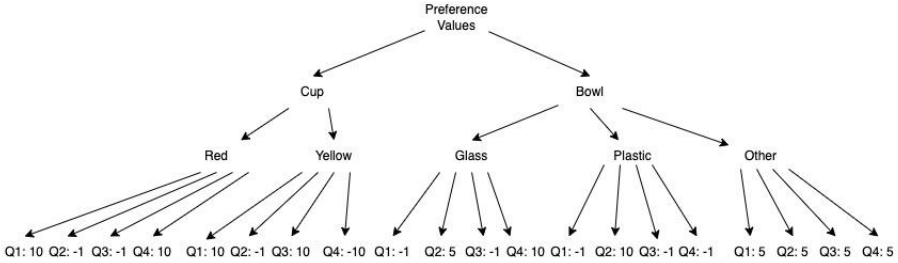
- State space S
- Robot action space A - movement of objects into quadrants
- Human action space A^h - for our context, this is identical to A.
- Transition function $T : S \times A \rightarrow S$
- Reward hypothesis space $\Theta = \{\theta_0, \ldots, \theta_H\}$
- Robot beliefs b

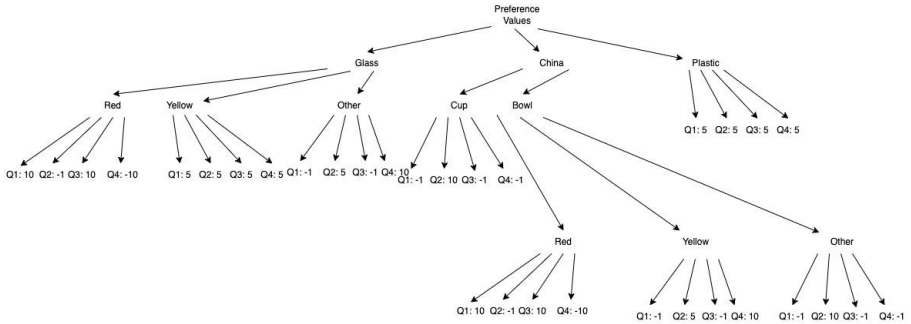


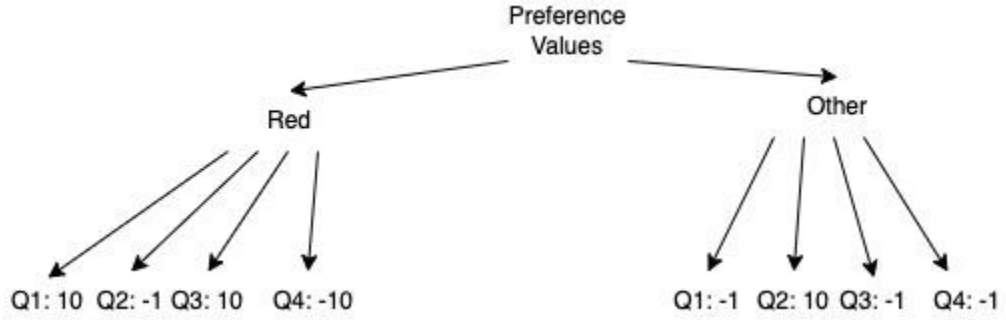
Each reward function in the Reward Hypothesis space is a tree



Reward Hypothesis space is a discrete set of trees



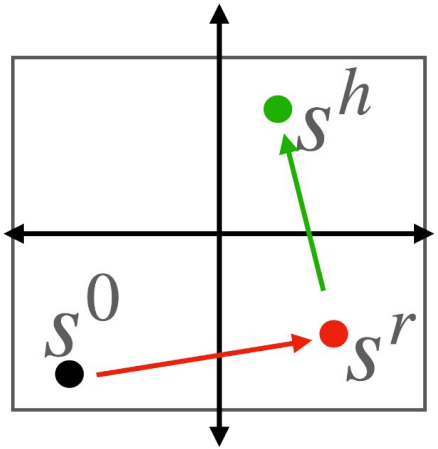




Carnegie Mellon University

Iterative Interaction

1. Initial state s_0
2. Robot takes $\pi_b^r(s_0) \rightarrow a_0^r$. State transitions to s_0^r.
3. Human corrects $\pi^h(s_0^r) \rightarrow a_0^h$. State transitions to s_0^h.
4. Robot updates beliefs $b' \leftarrow CorrectionUpdate(b, s_0, s_0^r, s_0^h)$



Bayesian Update Given Corrected State

1. Check Possible Cases:

- S_2 != S_1 # Robot Placement is not the same as Human Correction
- S_2 == S_1 and S_0 != S_1 # Robot Placement is the same as Human Correction, Robot Placement is not the same as initial placement
- S_2 != S_0 # Human Correction is not the same as initial placement

2. Update:

$P(\theta_i \mid S^h > S^r) = P(\theta_i) * P(S^h > S^r \mid \theta_i)$

$$P(s^h > s^r \mid \theta_i) = \frac{e^{\beta R_{\theta_i}(s^h)}}{e^{\beta R_{\theta_i}(s^h)} + e^{\beta R_{\theta_i}(s^r)}}$$

Carnegie Mellon University

Feature Uncertainty Question

Environment States

- Clarification
  a. Real Question
  b. Hallucinated Question

Features

- Preference
  a. “Is the reason that you performed a^h in s because of {color, type, or material}?”

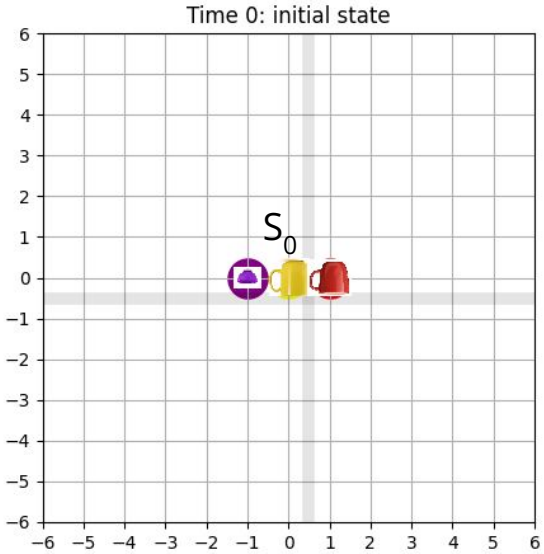
Hypothesis

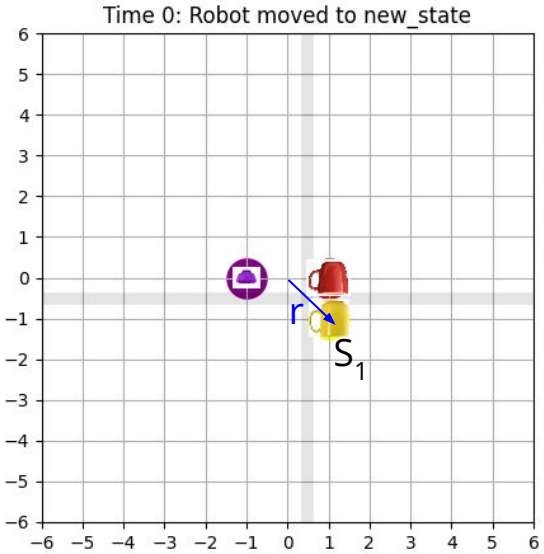
- "I think you prefer objects to be placed closer to the center. Is that correct?"

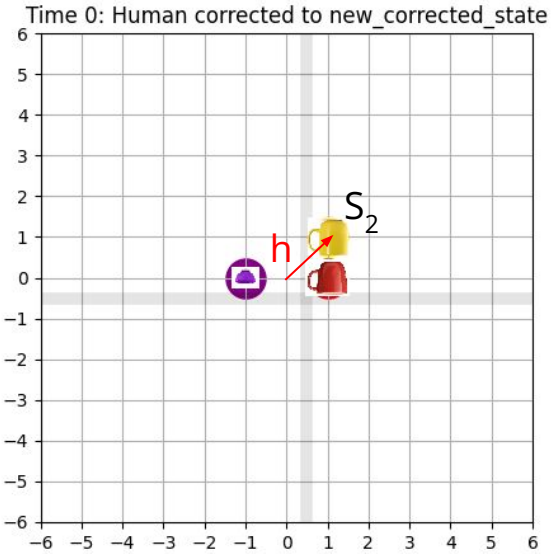
Carnegie Mellon University

Resultant Interaction

Timestep t = 0

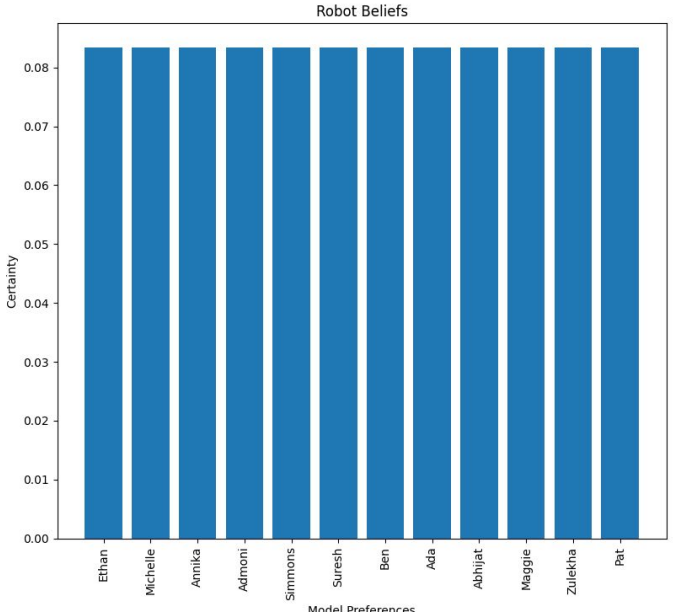


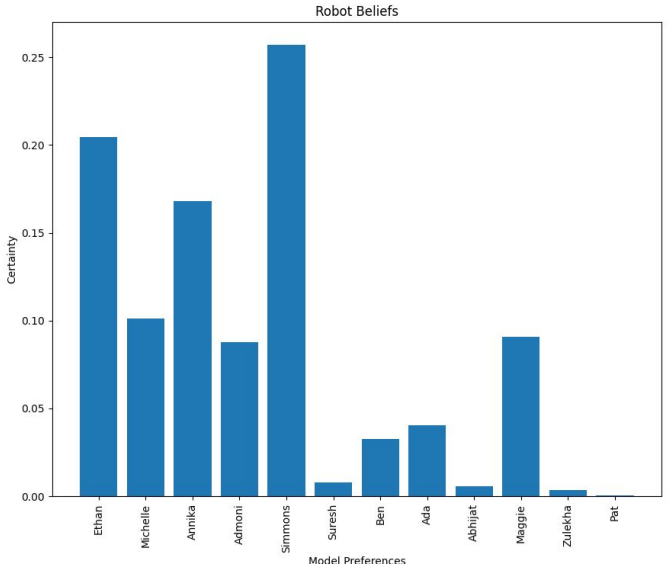




Resultant Interaction - Bayesian Update

Timestep t = 0 (inference)



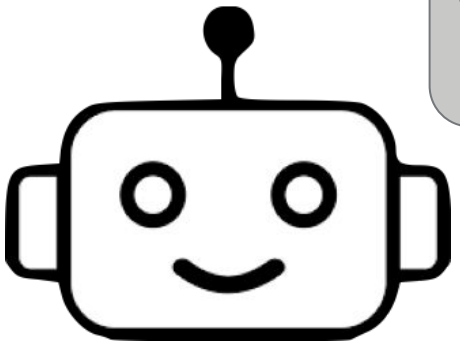


Carnegie Mellon University

Resultant Interaction

Timestep t = 0 (inference)

Clarification Question Asked:

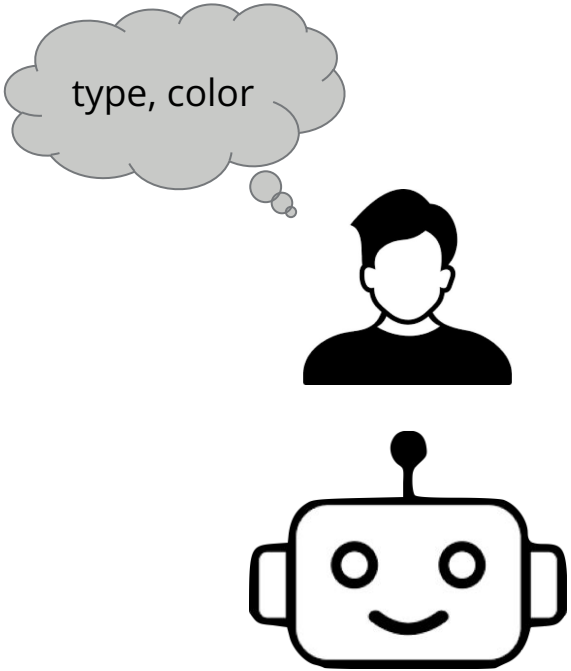


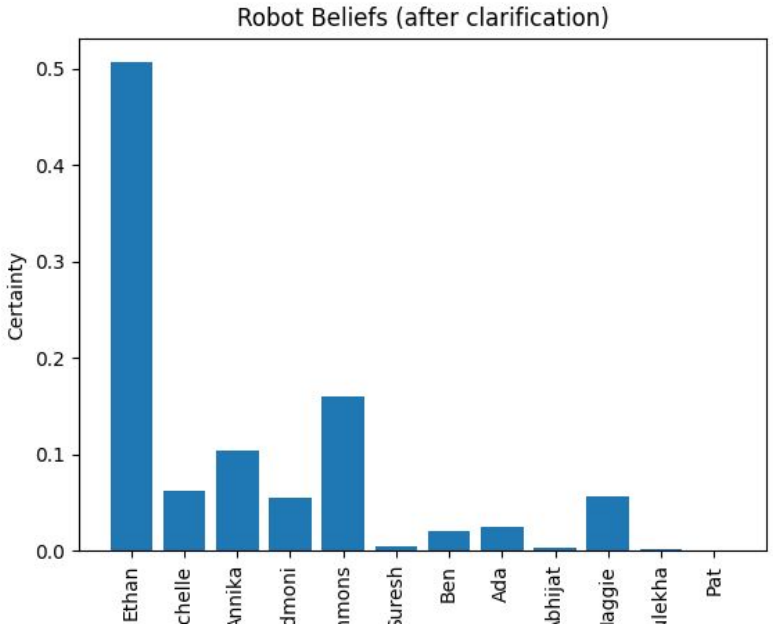
For the recent ['cup', 'yellow', 'glass'] object, which features of [color, type, and material] were relevant to the location it should be in?

Carnegie Mellon University

Resultant Interaction

Timestep t = 0 (update beliefs again)

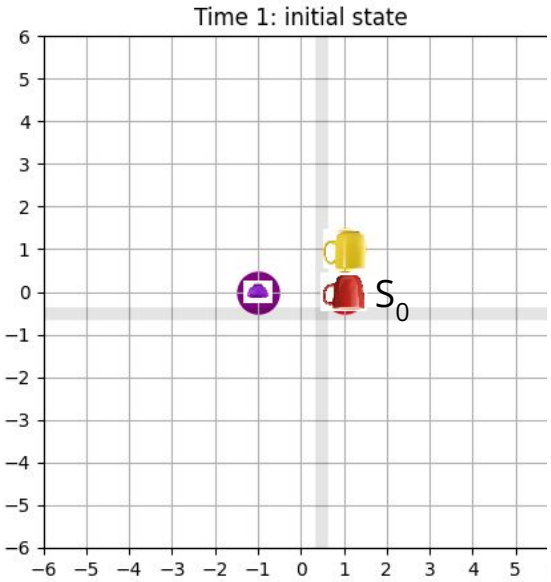


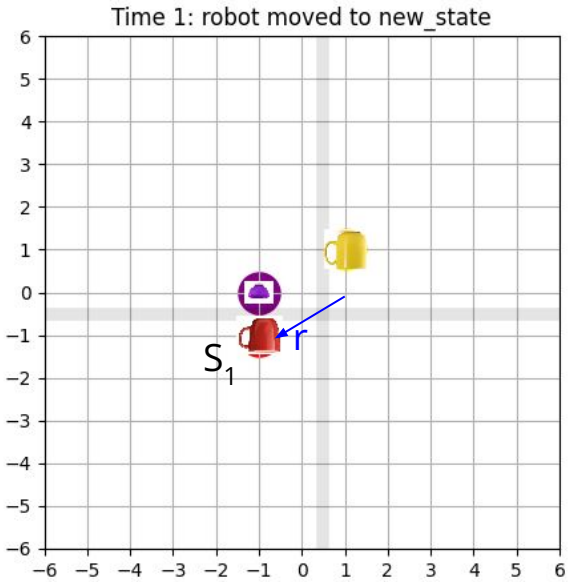


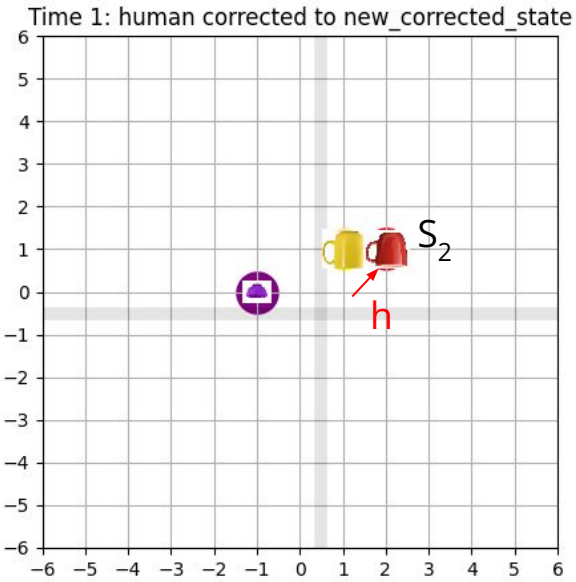
Carnegie Mellon University

Resultant Interaction

Timestep t = 1

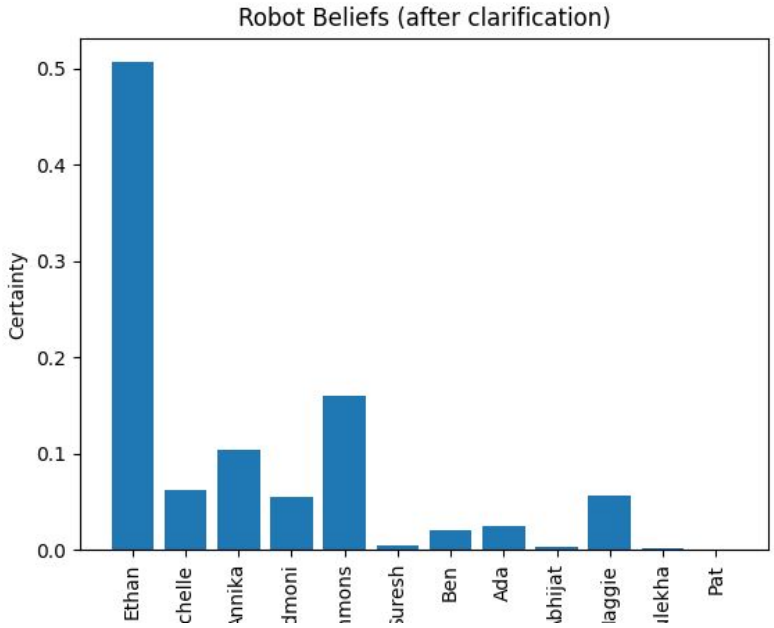




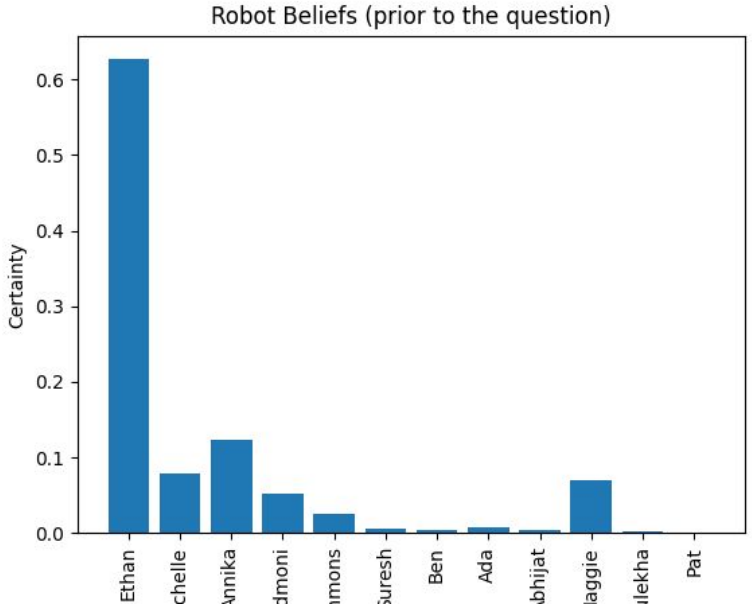


Resultant Interaction - Bayesian Update

Timestep t = 0 (inference)

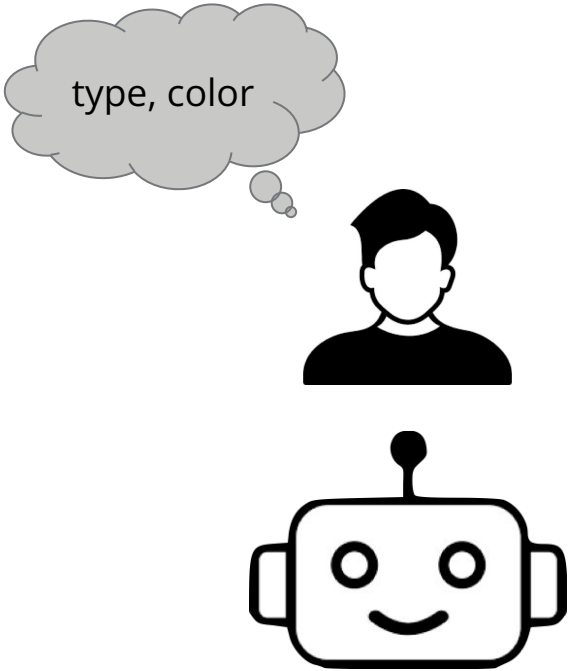


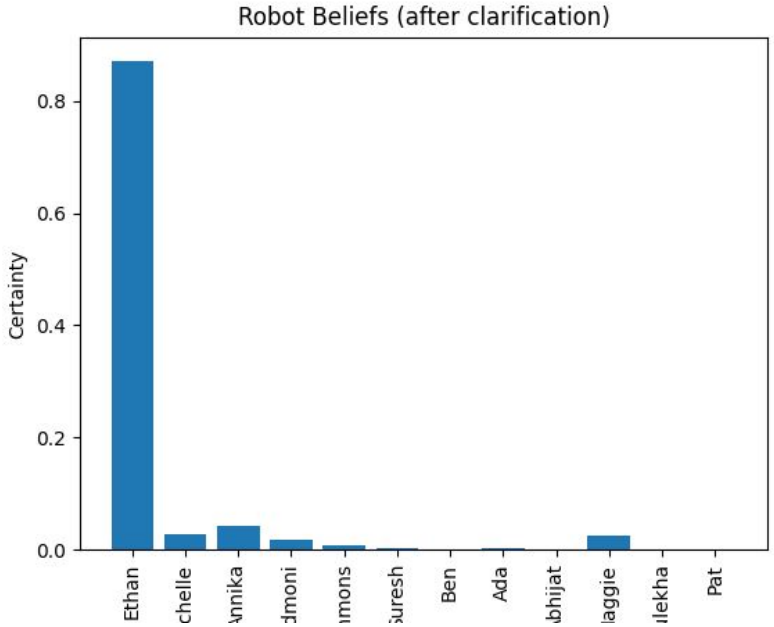
Timestep t = 1 (inference)



Resultant Interaction

Timestep t = 1 (update beliefs again)

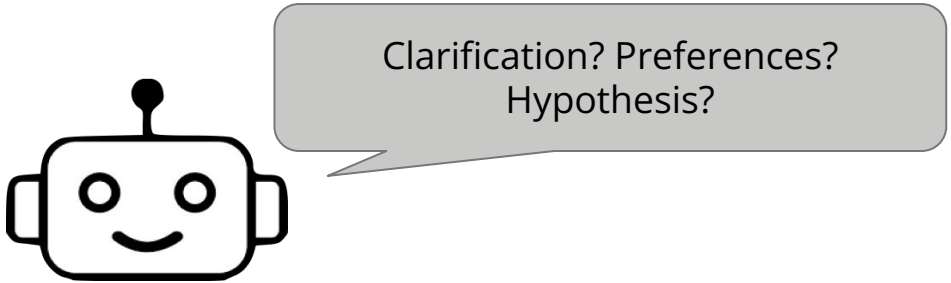




Carnegie Mellon University

Future Work

- Continue Clarification Question Implementation
- Pilot User Study



Carnegie Mellon University

Carnegie Mellon University

Thank you

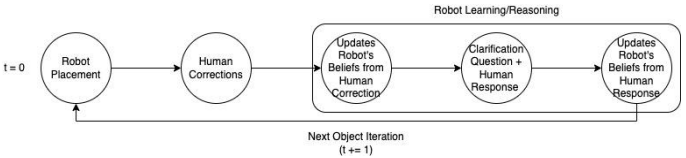
RISS: Dr. John M. Dolan, Rachel Burcin

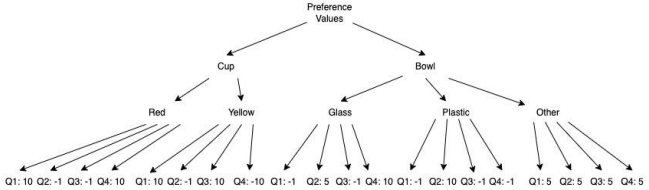
HARP & RASL Lab:
Michelle Zhao, Dr. Henny Admoni, Dr. Reid Simmons

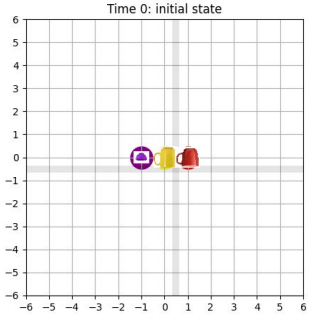
THE ROBOTICS INSTITUTE

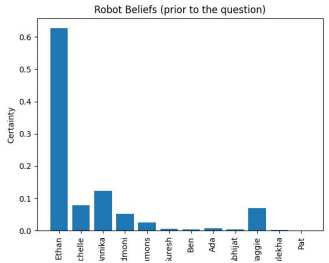
RISS | ROBOTICS INSTITUTE SUMMER SCHOLARS

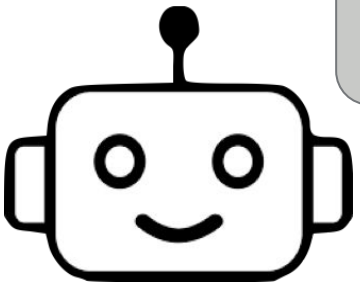
Questions?











For the recent ['cup', 'yellow', 'glass'] object, which features of [color, type, and material] were relevant to the location it should be in? (give comma separated responses):

Carnegie Mellon University