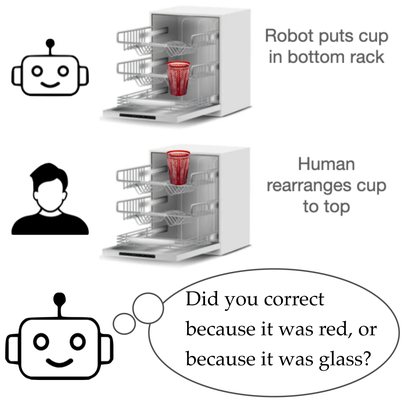
Clarifying Feature Overspecification in Reward Learning from State Corrections via Follow Up Questions

Ethan Villalovoz[1] Michelle Zhao[2] Henny Admoni[2] Reid Simmons[2]

[1]Washington State University [2]Carnegie Mellon University

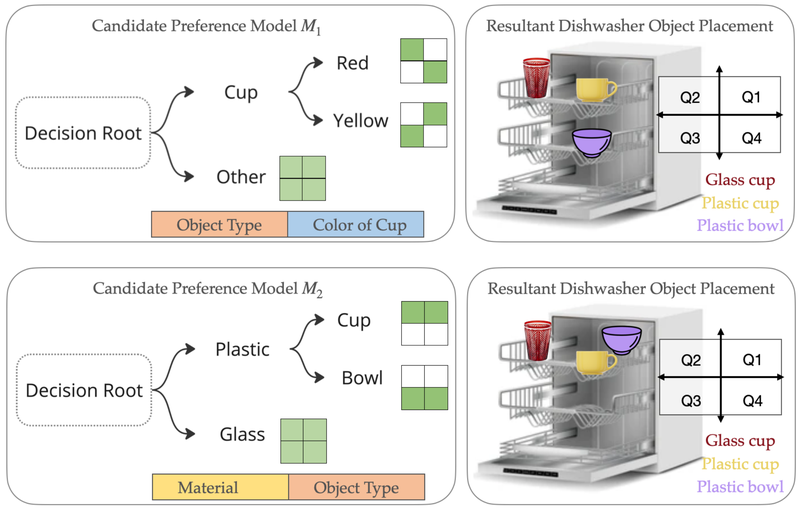
Motivation



Household robots need to know what people want. In high-level semantic tasks, the user may want to correct the goal state of the robot, rather than teleoperation [1]. The *reason why* users correct differs.

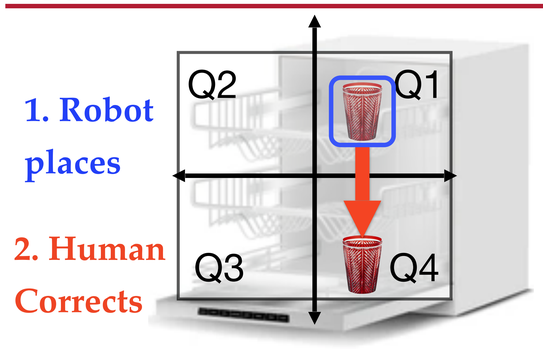
Key Challenge

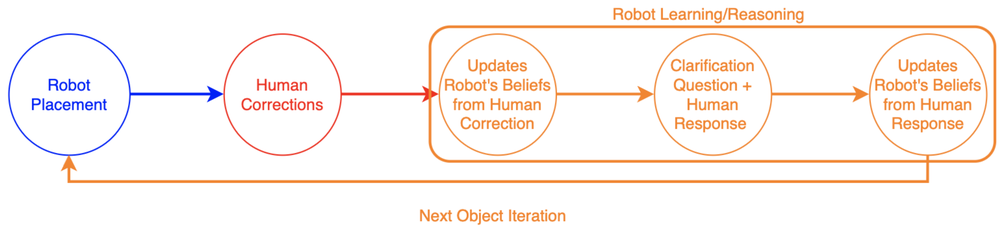
For different users, the task features relevant to their preference-guided decisions may vary.



Key Insight: Leverage follow-up questions to clarify features relevant to a human correction to speed-up reward learning.

Approach

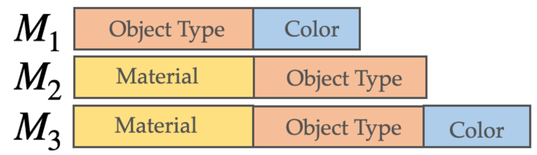




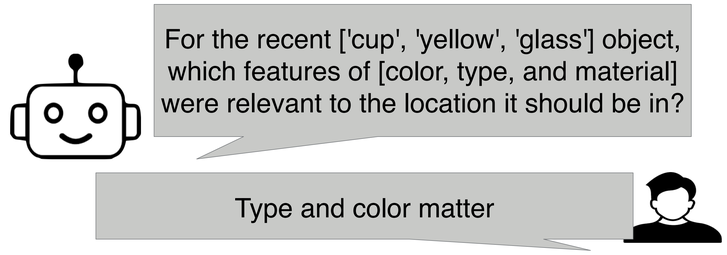
3. Update Robot Beliefs from Human Correction

$$P(M_i \mid \underset{\text{preferred over initial placement}}{\text{Corrected state}}) = P(\underset{\text{preferred over initial placement}}{\text{Corrected state}} \mid M_i)P(M_i)$$

4. Identify Each Reward Model's Relevant Features



5. Pose Followup Question to Human



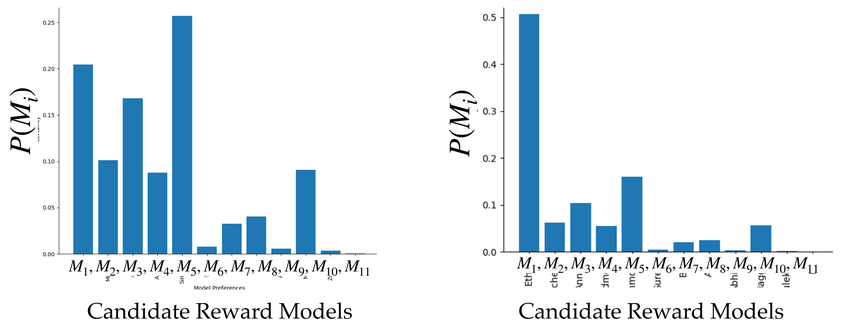
6. Update Robot Beliefs Again from Human Dialogue

$$P(M_i \mid \underset{\text{matter}}{\text{Type and color}}) = P(\underset{\text{matter}}{\text{Type and color}} \mid M_i)P(M_i)$$

Results

Clarification questions enabled the robot to more quickly find the true reward model of the human preferences.

Beliefs After Updating From Human Correction → Beliefs After Updating From Followup Dialogue



Future Work

1. **Clarification Question Implementation:**
   - Enhance the naive approach for feature uncertainty questions.
   - Develop LLM-generated questions addressing environment states, features, and hypotheses.
2. **Pilot User Study**

Acknowledgments

This work is supported by the Robotics Institute, Carnegie Mellon University. I would like to thank Dr. Henny Admoni, Dr. Reid Simmons, and especially my mentor Michelle Zhao for the transformative experience and opportunity to conduct research throughout this summer. I would like to thank the Robotics Institute Summer Scholars Program, specifically Rachel Burcin and Dr. John Dolan, for the opportunity.

References

[1] Bobu, A., Bajcsy, A., Fisac, J. F., Deglurkar, S., & Dragan, A. D. (2020). Quantifying hypothesis space misspecification in learning from human–robot demonstrations and physical corrections. IEEE Transactions on Robotics, 36(3), 835-854.