# Planning Multimodal Exploratory Actions for Online Robot Attribute Learning

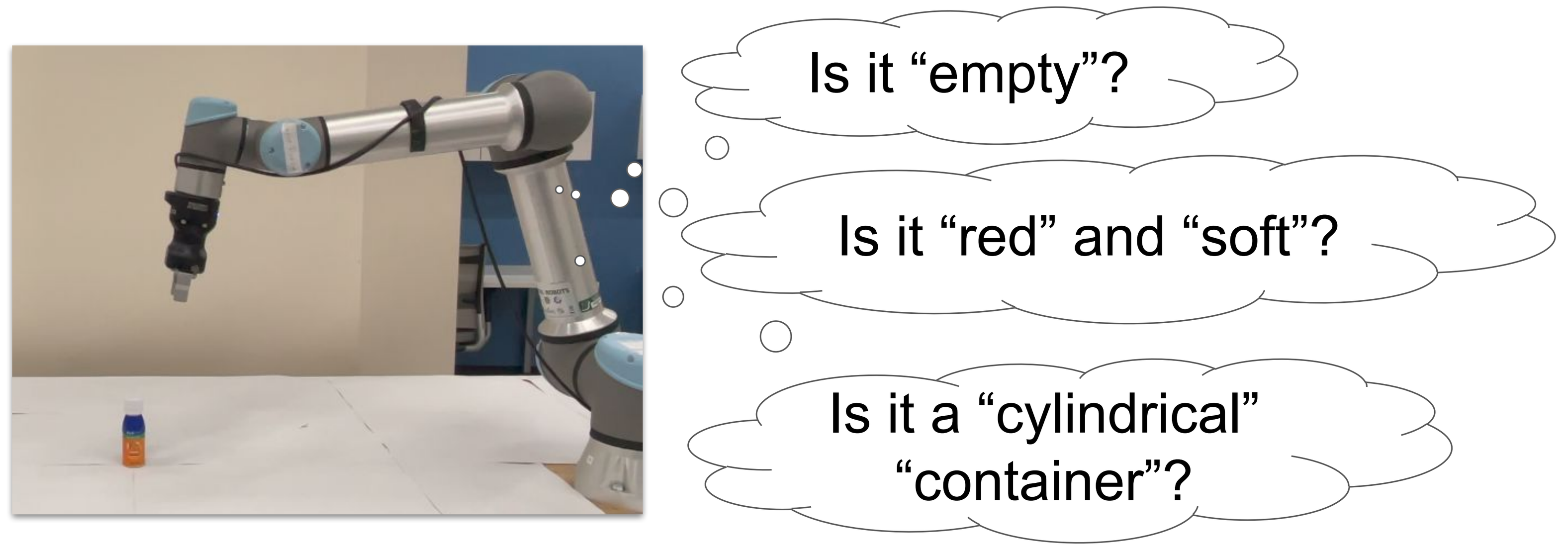

Xiaohan Zhang[1], Jivko Sinapov[2], Shiqi Zhang[1]

[1]State University of New York at Binghamton, Binghamton, NY, USA [2]Tufts University, Medford, MA, USA



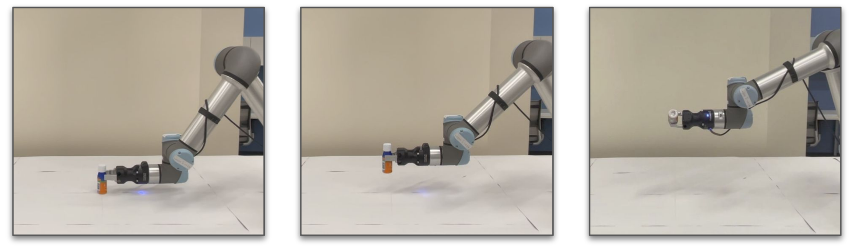
*Published at the Robotics: Science and Systems (RSS) Conference in 2021*


## Scenario




➔ Objects: 

➔ Attributes: blue, full, cylindrical, mediumweight, container…

➔ Actions:  …

*grasp* *lift* *shake*

➔ Sensory modalities: vision, audio, haptics…

## Problem Definitions

➔ Offline Robot Attribute Learning (Off-RAL)

- ◆ produces a binary classifier $\Psi$ for action-attribute pair
- ◆ estimates if an attribute applies to an object given streaming sensory data

*Off-RAL requires large amounts of exploration data at training time.*

➔ Robot Attribute Identification (RAI)

- ◆ produces an action policy $\pi$ given $\Psi$
- ◆ aims to sequentially select actions to identify the attribute(s) in **each** identification task while minimizing the action cost

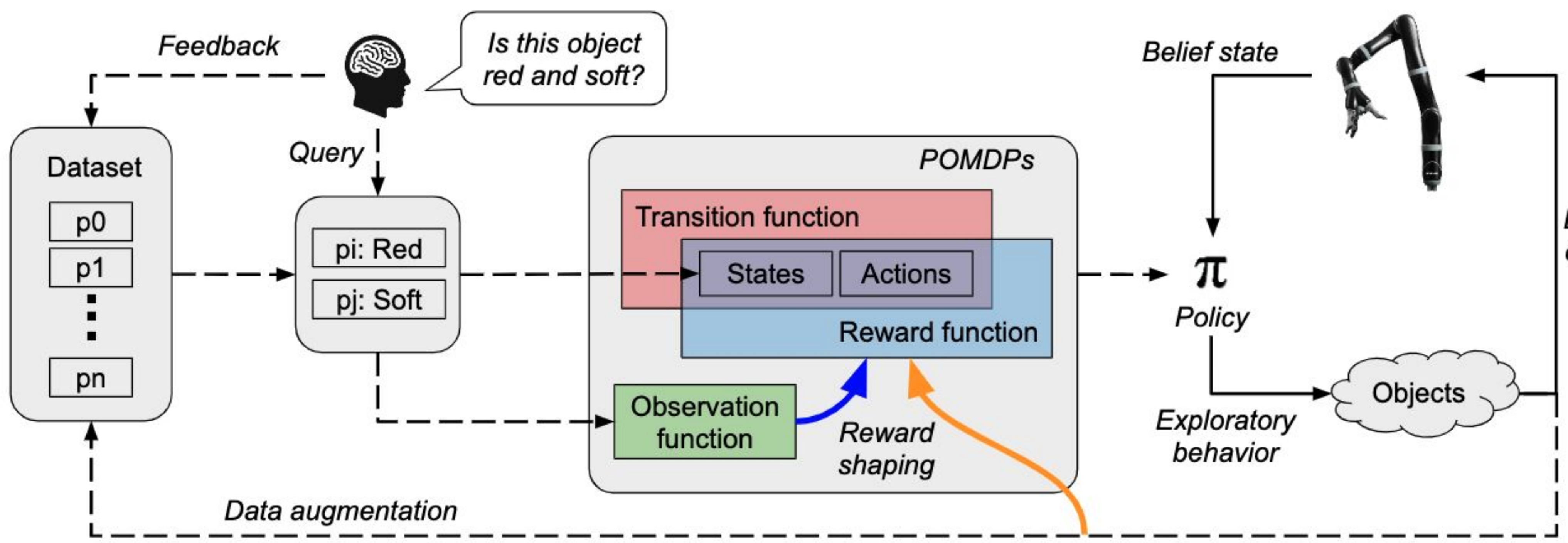
★ **Online Robot Attribute Learning (On-RAL)**

- ◆ produces an action policy $\pi$ while learning $\Psi$
- ◆ aims to minimize the discounted cumulative action cost and maximize the success rate of attribute identification

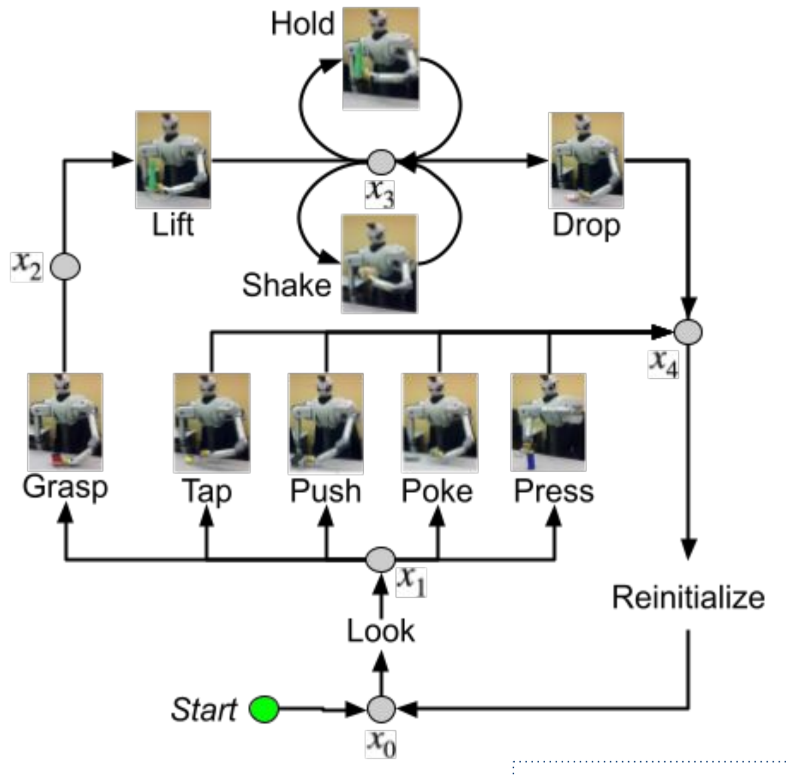
*Rational On-RAL agents learn from data collected in early tasks, trading off early-phase performance for long-term performance.*

## Methods

➔ Information-Theoretic Reward Shaping (ITRS) Overview:




➔ Transition Diagram:




➔ State Space $X \times Y$:

- ◆ $X$ is specified by fully observable domain variables (current states of the robot-object system, e.g, whether *grasp* and *drop* are successful or not)
- ◆ $Y$ is specified by partially domain variables ($N$ queried attributes)

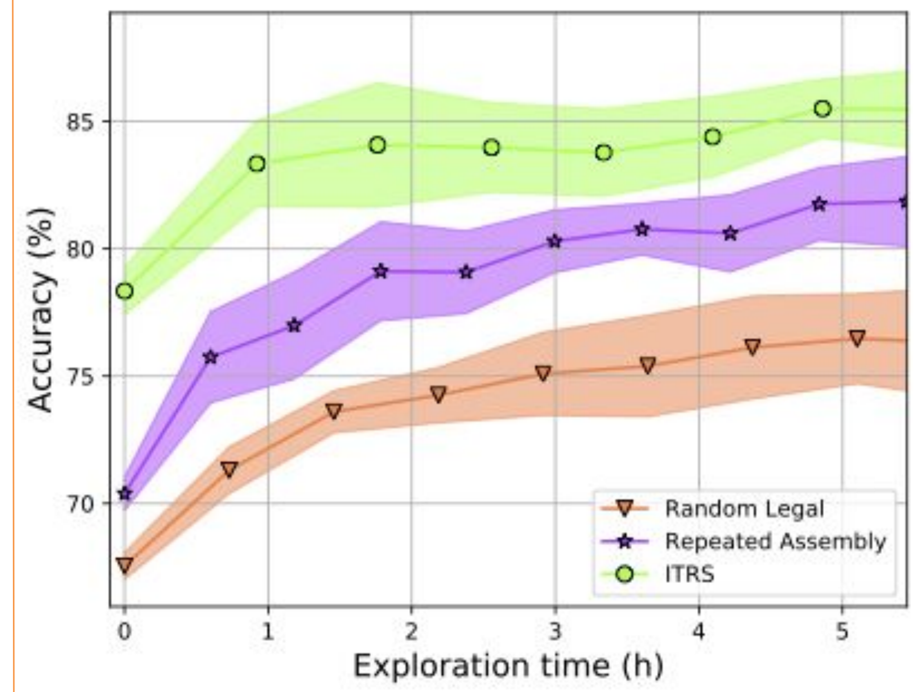
➔ Observation Functions: (Amiri et al., 2018)

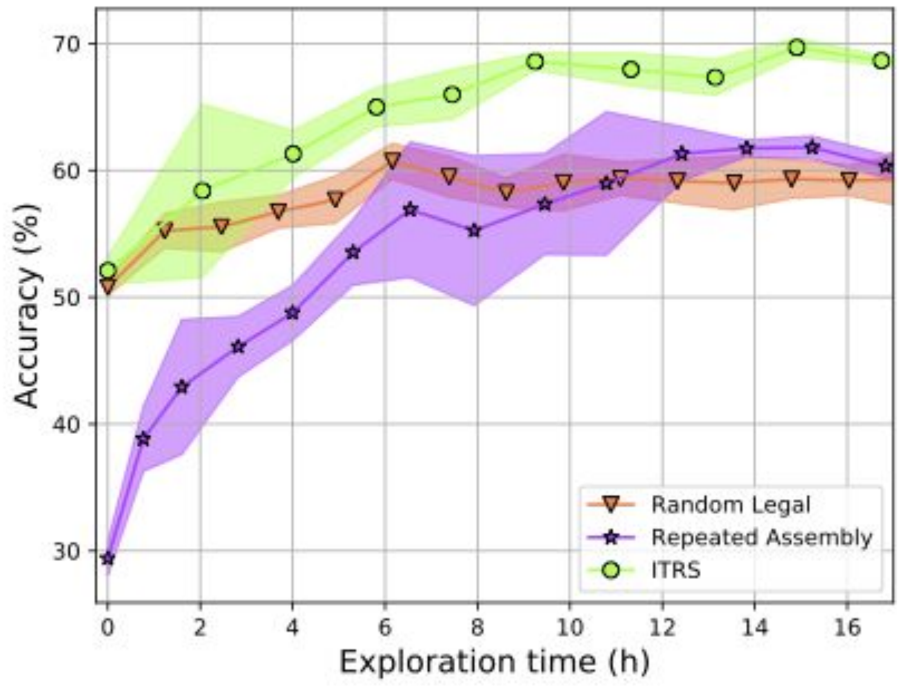
$$O(s,a,z) = \mathbf{Pr}(\mathbf{p}^z|\mathbf{p}^s,a) = \prod_{i=0}^{N-1} \Theta^a_{p_i}(p^s_i, p^z_i)$$

➔ Reward Shaping: $R(s,a,s') = R^{real}(s,a,s') + \alpha \cdot Ent(s,a) - \beta \cdot IE(p,a)$

- ◆ Real-world reward function $R^{real}$
  - rewards (or penalizes) successful (or unsuccessful) identifications
  - reflects exploratory action cost (e.g., 0.5 for *look* and 22.0 for *press*)
- ◆ Perception quality: $Ent(s,a) = -\sum_{z_i \in Z} O(z_i|s,a)\log_2 O(z_i|s,a)$
- ◆ Interaction experience: $IE(p,a) = \frac{1}{\delta}\cdot|\mathcal{F}^c, \text{ where } c \in \mathcal{C}_a, p \in \Lambda(\mathcal{F}^c)|$

## Results




➔ Datasets:

- ◆ **CY101** (Tatiya and Sinapov, 2019)
- ◆ **ISPY32** (Thomason et at., 2016)

➔ Baselines:

- ◆ Random Legal (Thomason et al., 2018)
- ◆ Repeated Assembly (Amiri et al., 2018)

➔ Experimental results show that ITRS enables the robot to complete attribute identification tasks at a higher accuracy using the same amount of training time compared to baselines.

## Conclusion

➔ We focus on a novel On-RAL problem where the robot is required to complete attribute identification tasks and learn its observation model for each attribute at the same time

➔ We propose an algorithm called ITRS that achieves the trade-off between exploration (actively collecting data for attribute learning) and exploitation (using the learned attributes for identification tasks)

## Acknowledgements


A portion of this research has taken place at the Autonomous Intelligent Robotics (AIR) Group, SUNY Binghamton. AIR research is supported in part by grants from the National Science Foundation (NRI-1925044), Ford Motor Company (URP Awards 2019-2021), OPPO (Faculty Research Award 2020), and SUNY Research Foundation.