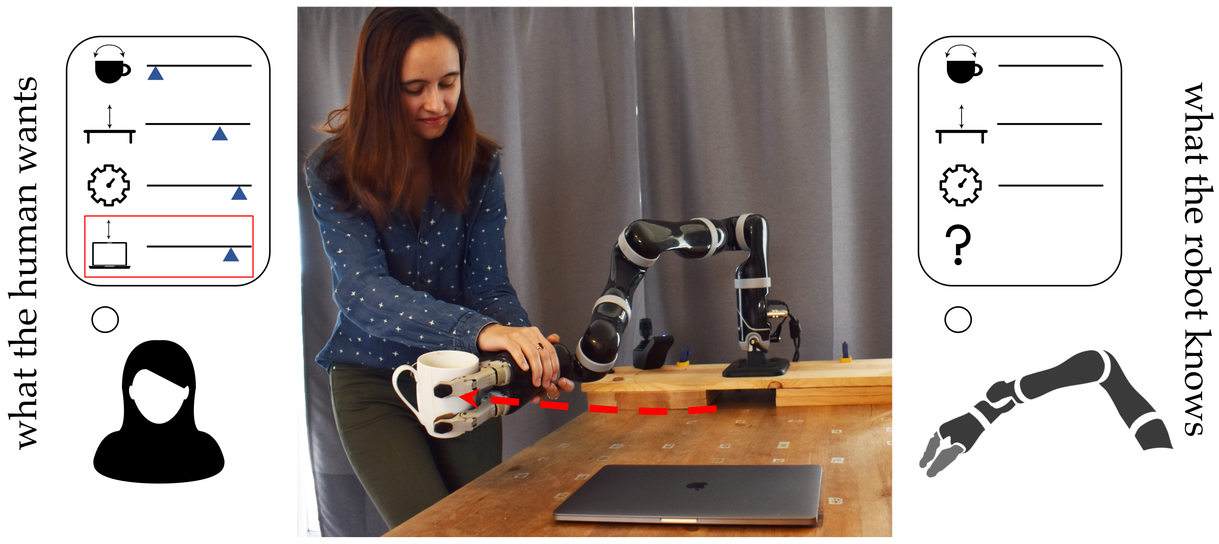
Berkeley UNIVERSITY OF CALIFORNIA Feature Expansive Reward Learning: Rethinking Human Input

BAIR BERKELEY ARTIFICIAL INTELLIGENCE RESEARCH

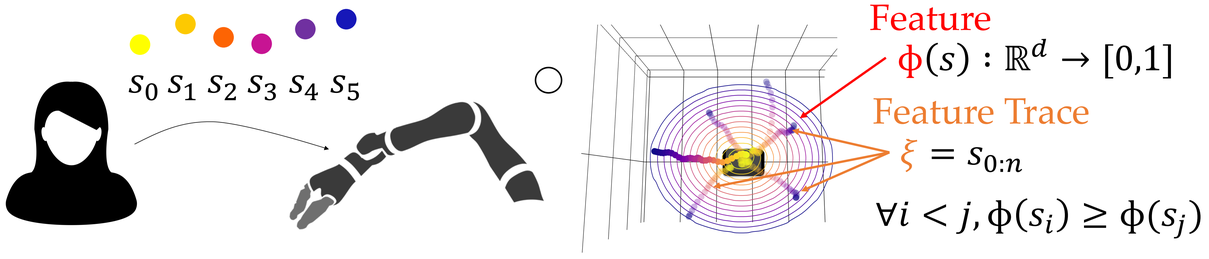
Andreea Bobu*, Marius Wiggert*, Claire Tomlin, Anca D. Dragan

Problem Statement: How can the robot update its reward R_θ from human input even when it doesn't understand what the human input refers to?



Key Insight: Instead of learning about the missing feature(s) *implicitly*, the robot should ask for data that *explicitly* teaches it what is missing.

Feature Traces: A New Type of Human Input

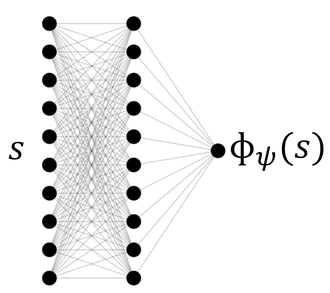


Learning a Feature Function

1. Monotonicity
$\phi_\psi(s_i) \geq \phi_\psi(s_j)$
$(s_i^k, s_j^k, 1), \forall\, i < j, \forall k$

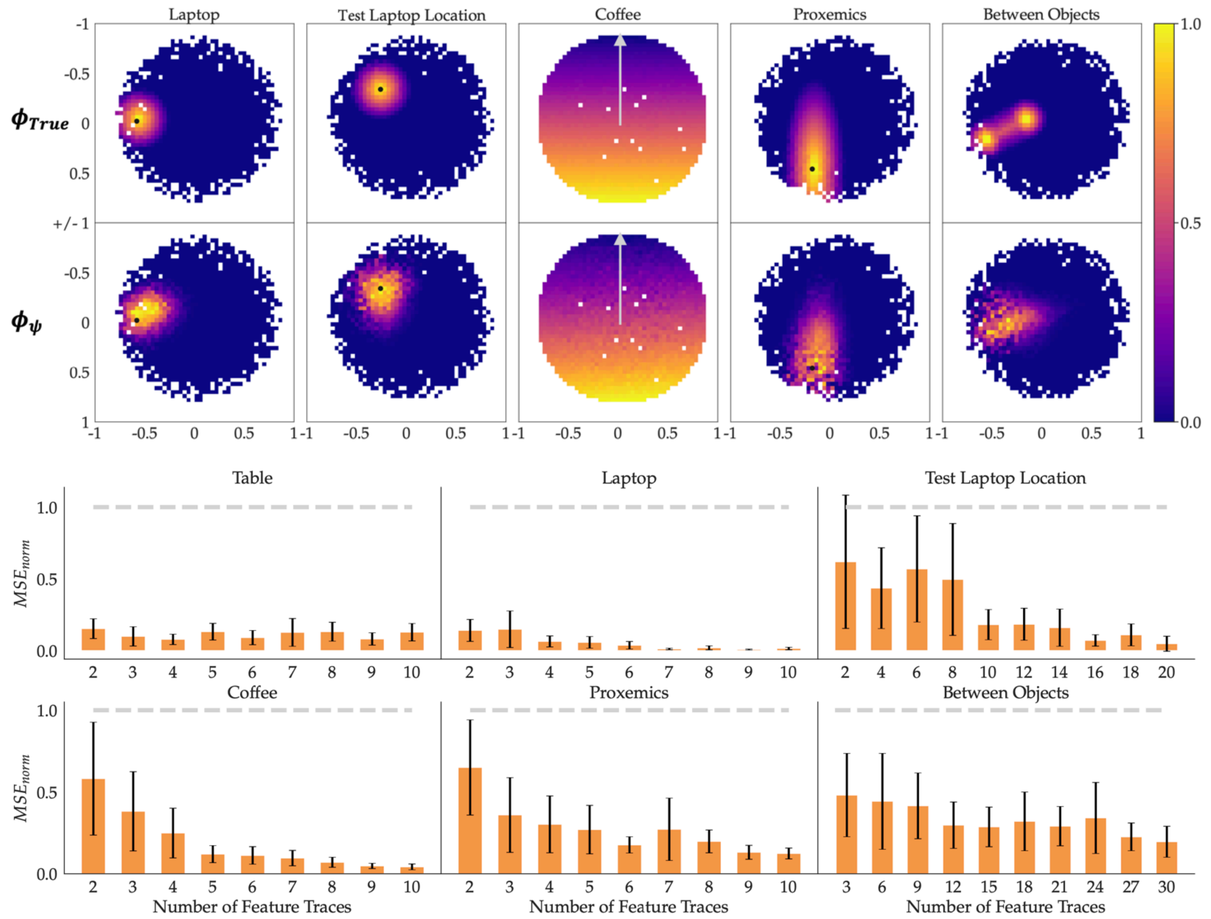
2. Start/End Labels
$\phi_\psi(s_0^k) \approx 1, \phi_\psi(s_{n^k}^k) \approx 0$
$(s_0^i, s_0^j, 0.5), (s_{n^i}^i, s_{n^j}^j, 0.5), \forall\, i, j$

$(s, s', y) \in \mathcal{D}, y \in \{0, 0.5, 1\}$



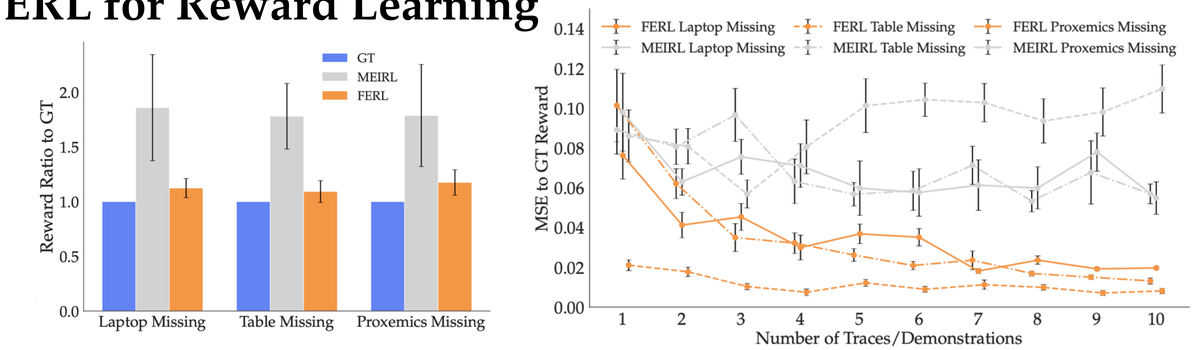
$$L(\psi) = -\sum_{(s,s',y)\in\mathcal{D}} y \log\big(P(s \succ s')\big) + (1-y)\log(1 - P(s \succ s'))$$

FERL for Feature Learning



With enough data, FERL learns good features, and, with more data, it both learns increasingly better features, and becomes less input-sensitive.

FERL for Reward Learning



FERL learns rewards that better generalize to the state space, are less input-sensitive, and produce trajectories that are preferred over deep IRL rewards.