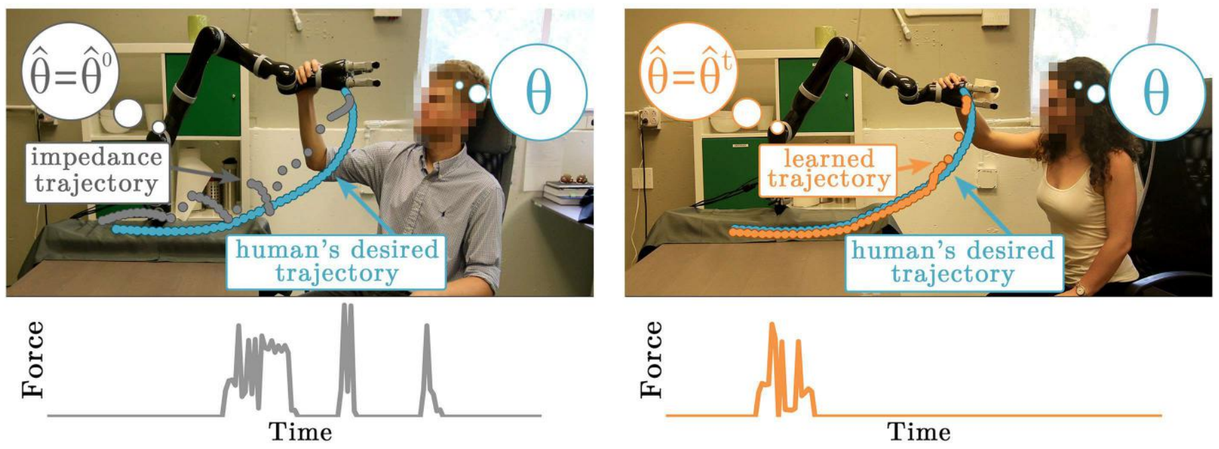
Learning Robot Objectives from Physical Human Interaction

Andrea Bajcsy*, Dylan P. Losey*, Marcia K. O'Malley, Anca D. Dragan



Traditionally, robots treat physical interaction as a **disturbance**, and resume their original behavior after the interaction ends.

We argue physical human interaction is **intentional** and thus **informative**: it is useful info about how the robot should be doing its task.

Use **pHRI to correct the robot's objective function** *while the robot is performing its current task*

Robot trying to optimize:

$$r(x, u_R, u_H; \theta) = \theta^T \phi(x, u_R, u_H) - \lambda ||u_H||^2$$

hidden variable — task reward — human effort

Observation model of human interaction forces:

$$P(u_H | x, u_R; \theta) \propto e^{Q(x, u_R + u_H; \theta)}$$

Assume noisy-rational human selecting actions so robot behaves optimally after current step

QMDP Approximation

$$Q(x, u_H + u_R, b) = \int b(\theta) Q(x, u_H + u_R, \theta) d\theta$$

Assume robot gets full observability at next time step.

From Policies to Trajectories

I. Action. **Plan** trajectory with MAP of θ (not b)

$$\xi_R^t = \arg\max_{\xi} \hat{\theta}^t \cdot \Phi(\xi)$$

II. Estimation. Relate human force to θ via intended trajectory by **deforming** original trajectory

$$\xi_H = \xi_R + \mu A^{-1} U_H$$

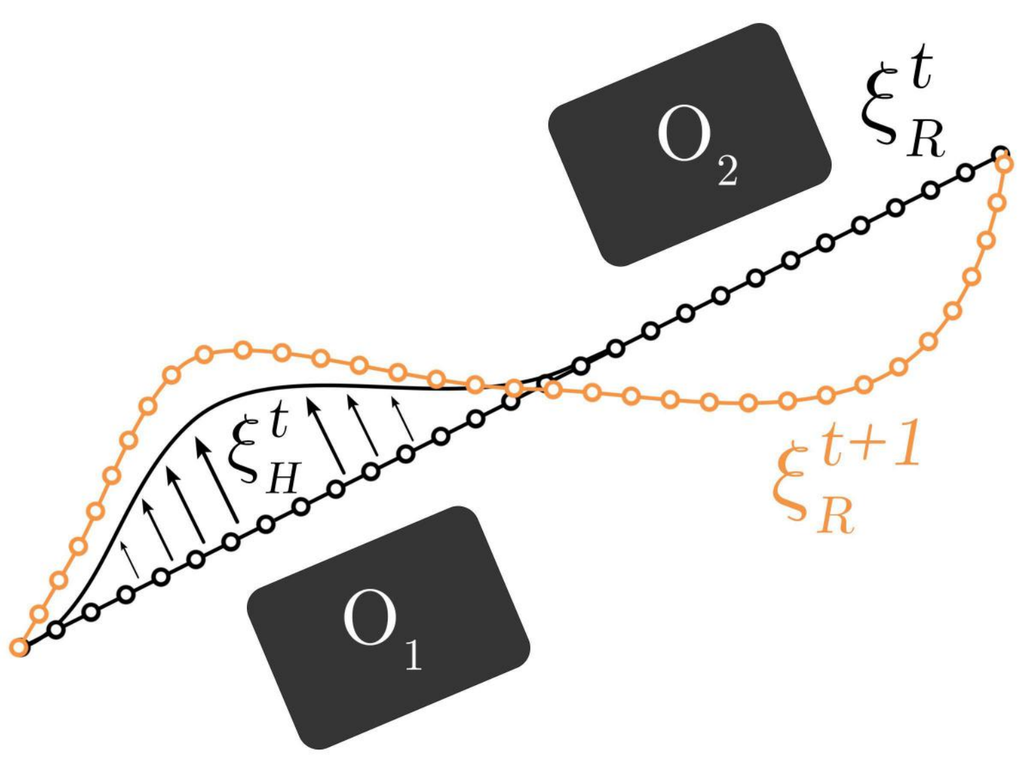
Observation model of human's intended trajectory

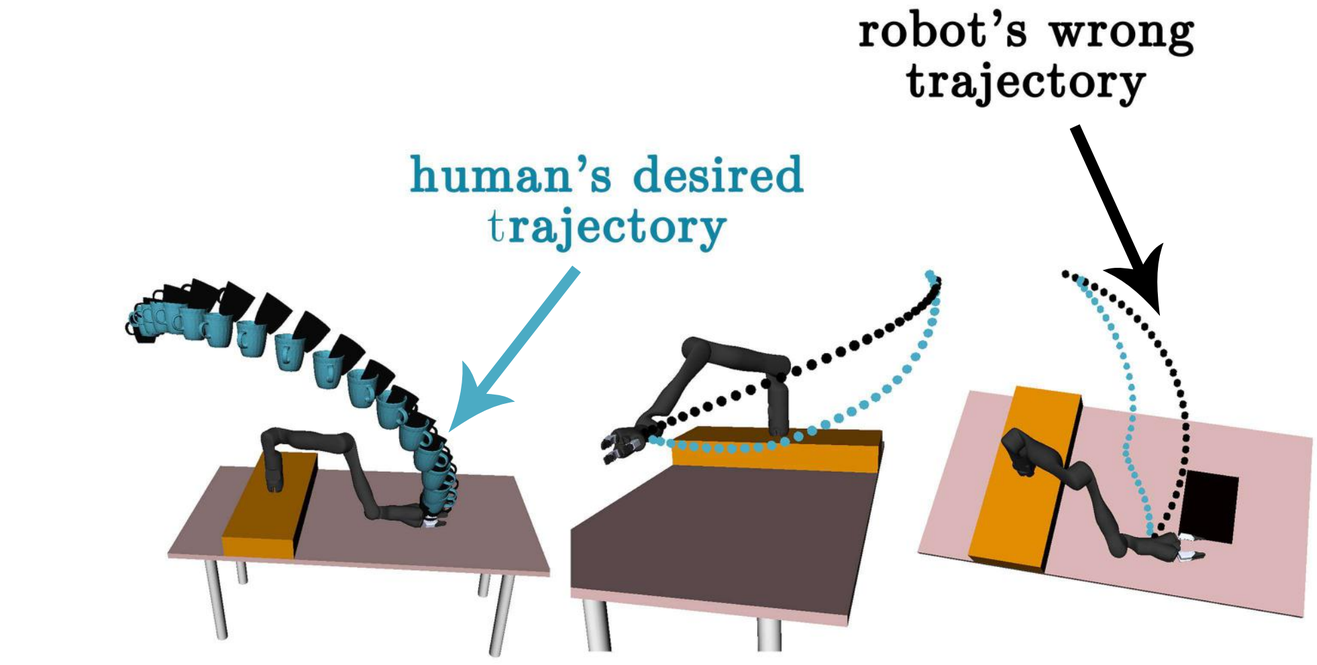
$$P(\xi_H | \xi_R, \theta) \approx e^{\theta^T \Phi(\xi_H) - \lambda ||\xi_H - \xi_R||^2}$$

Maximum a posteriori **estimate of θ**

$$\hat{\theta}^{t+1} = \hat{\theta}^t + \alpha\big(\Phi(\xi_H^t) - \Phi(\xi_R^t)\big)$$

Algorithm for Online Learning from pHRI





User Study

- Investigated the benefits of *in-task* learning
- Within-subjects experiment with 10 participants
- Manipulated pHRI strategy with 2 levels: *learning* & *impedance*