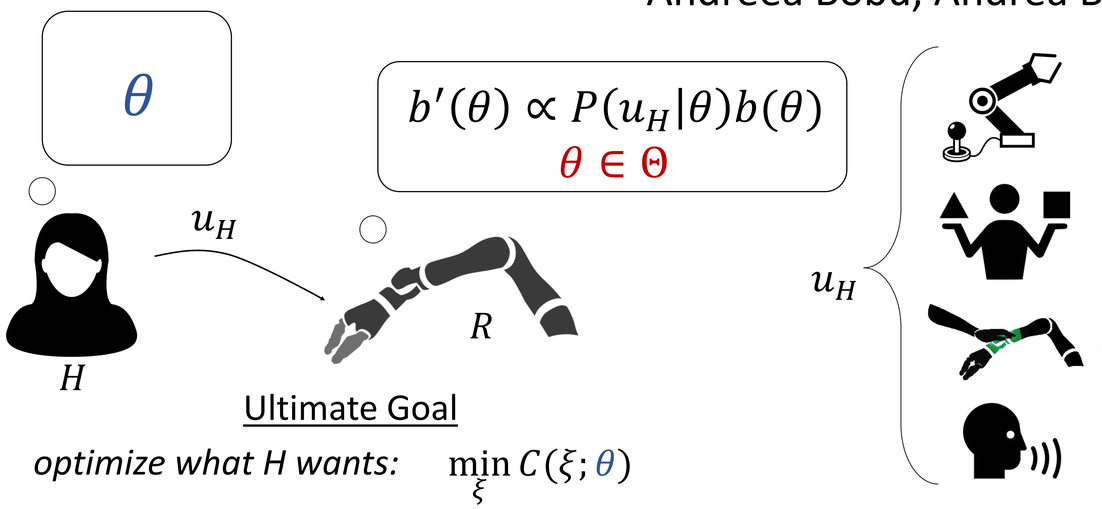
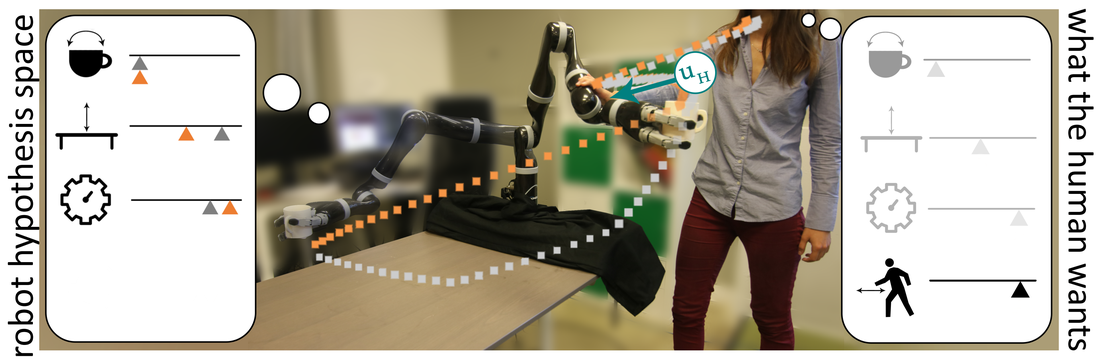
Learning Under Misspecified Objective Spaces

Andreea Bobu, Andrea Bajcsy, Jaime F. Fisac, Anca D. Dragan



Ultimate Goal

optimize what H wants: $\min_{\xi} C(\xi;\theta)$

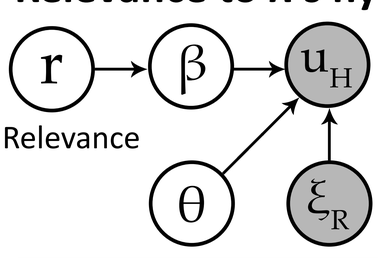


Challenge

What if what H wants is outside the robot's hypothesis space $\boldsymbol{\Theta}$*?*

Key Insight: If the human seems to be suboptimal for any hypothesis, chances are we don't have the right hypothesis space.

Relevance to *R*'s hypothesis space dictates apparent rationality



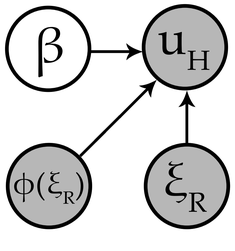
Action Weight

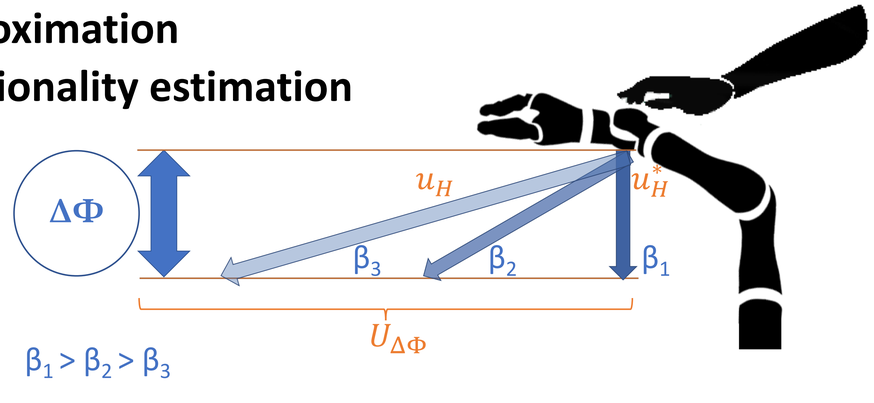
$$P(u_H|\xi_R;\beta,\theta) = \frac{e^{-\beta(\theta^T\Phi(\xi_H)+\lambda||u_H||^2)}}{\int e^{-\beta(\theta^T\Phi(\bar{\xi}_H)+\lambda||\bar{u}_H||^2)}d\bar{u}_H}$$

Rationality

Real-time approximation

a) Apparent rationality estimation





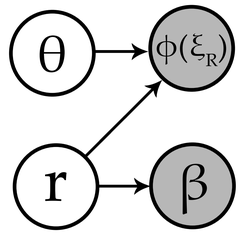
$\beta_1 > \beta_2 > \beta_3$

$$||u_H^*||^2 = ||u_H^1||^2 < ||u_H^2||^2 < ||u_H^3||^2 = ||u_H||^2$$

$$P(u_H|\beta,\Phi(\xi_H),\xi_R) = \frac{e^{-\beta||u_H||^2}}{\int_{u\in U_{\Delta\Phi}} e^{-\beta||u||^2}du} \approx e^{-\beta(||u_H||^2-||u_H^*||^2)}\sqrt{\frac{\beta^k|H_{u_H^*}|}{2\pi^k}}$$

$$\hat{\beta} = \text{argmax}P(u_H|\beta,\Phi(\xi_H),\xi_R) \approx \frac{k}{2(||u_H||^2-||u_H^*||^2)}$$

b) Relevance-aware approximate MAP estimate:

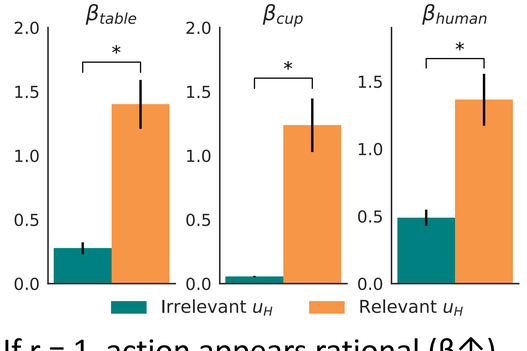


$$P(\Phi(\xi_H)|\theta,r) = \begin{cases} \left(\frac{\lambda}{\pi}\right)^{k/2} e^{-\lambda||\Phi(\xi_H)-\Phi(\xi_R)||^2}, & r=0 \\ \frac{e^{-\theta^T\Phi(\xi_H)}}{\int e^{-\theta^T\Phi(\bar{\xi}_H)}d\bar{\xi}_H}, & r=1 \end{cases}$$

Fit on offline data

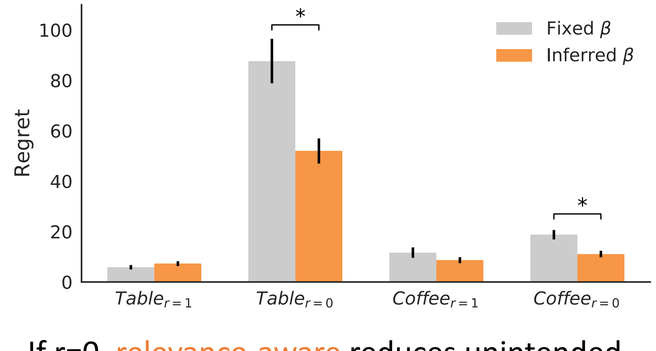
$$\hat{\theta}' = \hat{\theta} - \alpha\frac{P(r=1|\hat{\beta})P(\Phi(\xi_H)|\hat{\theta}',r=1)}{\sum_r P(r|\hat{\beta})P(\Phi(\xi_H)|\hat{\theta}',r)}(\Phi(\xi_H)-\Phi(\xi_R))$$

Offline β-estimation



If r = 1, action appears rational (β↑).
If r = 0, action appears irrational (β↓).

User study



If r=0, relevance-aware reduces unintended learning, while keeping good accuracy if r=1.