HRI in Shared Autonomy: Information Theoretic Perspective

Coupled PA loops

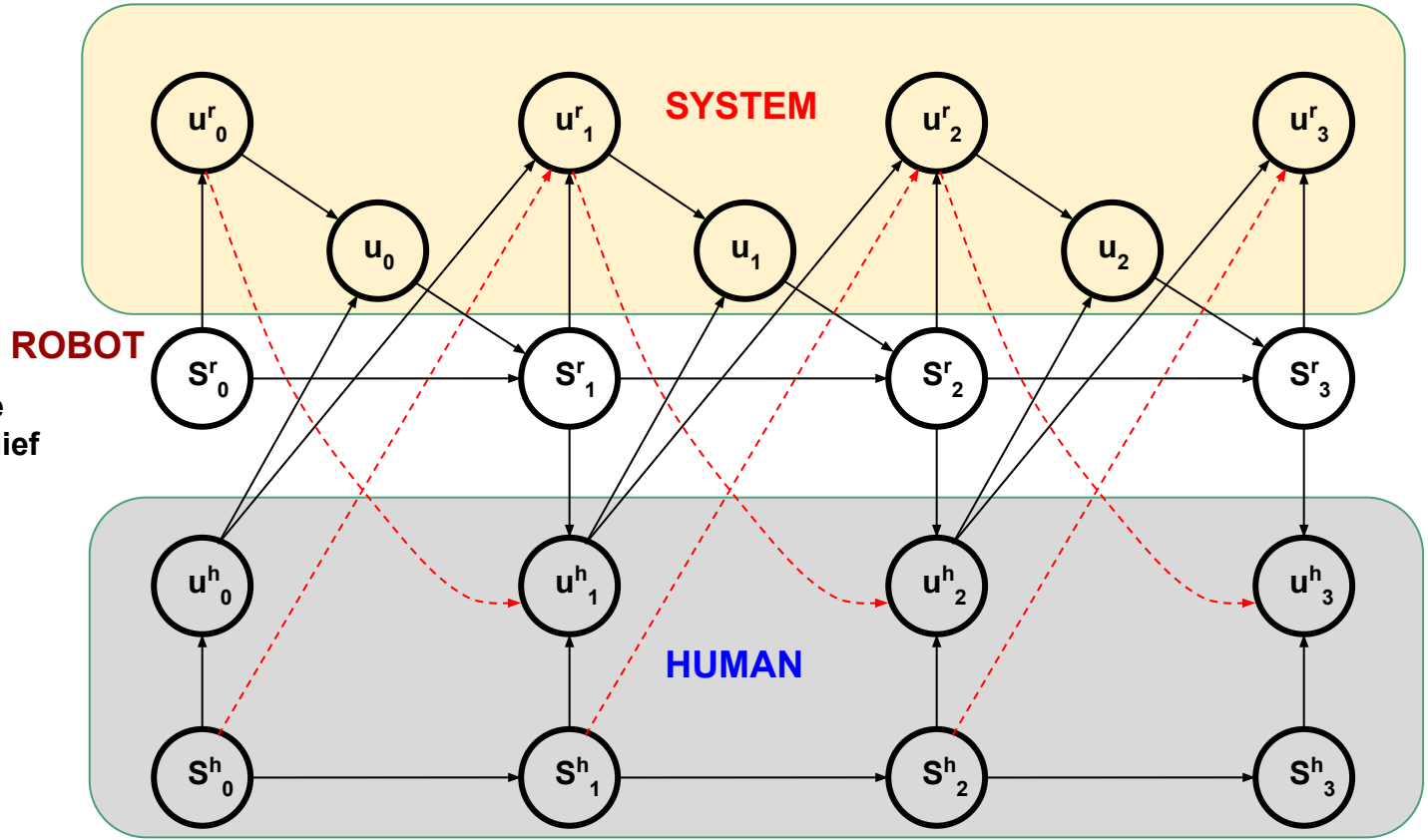
U^r - Robot control

U^h - Human control

U - Shared control

S^h - Human internal state (For example, Goal + Belief regarding robot assistance)

S^r - Robot state



Coupled PA loops

- The previous slide depicts HRI as a coupled DBN that evolves in time. The beige colored box indicates the components that belong to the system and the gray colored box indicates the human.
- The arrows that cross over from one box to the other correspond to the **perception** component. The **actions** that each agent execute are denoted as $\mathbf{u}^{r}$ and $\mathbf{u}^{h}$.
- The red lines indicate implicit dependence of control commands on variables that need to be inferred from perceived observations.
- It is in this sense that this behaves like a POMDP, because the human's internal state and robot autonomy control commands actions are not directly accessible to the system and human respectively and need to be inferred from the observations that each component receives.

POMDPs, PA Loops and Communication Channels

- Perception-Action loops can be depicted as DBNs evolving in time and depending on whether the states are directly accessible or not, can be viewed as an MDP or POMDP
- However, if we are to think of the problem in terms of POMDPs, then implicitly we are also assuming the existence of a cost function that needs to be optimized for optimal behavior. It is possible that this cost function could be cast in terms of some information theoretic quantities of interest.
- Different pathways in the coupled PA loops can be thought of a ***communication channel.*** Each communication channel also has an associated channel capacity.
- For the example, the channel between $\mathbf{s}^h$ and $\mathbf{u}^r$ can possibly capture the transparency of human intentions to the system.
- Similarly, the channel between $\mathbf{u}^r$ and $\mathbf{u}^h$ might be able to capture how transparent the robot assistance is to the human. (Note, I am a little unsure about the implicit edge robot control command to human control command. The idea is that although the human cannot directly see what the robot contribution is, s/he might infer it from the observation of robot state, self-knowledge of one's control command and dynamics of the system.