# HRI in Shared Autonomy: Information Theoretic Perspective

# Motivation

- The core motivation for this problem comes from a few different observations we have had from our past studies
    - Lack of understanding of what/how the robot is trying to help the human diminishes collaboration, increases user frustration and affects task performance detrimentally.
    - How should one design robot autonomy and calibrate parameters that affect robot actions? In our first paper, we assumed an optimal control problem setting in which the cost function is unknown. Typical approaches in literature, seeks to learn cost functions using IRL methods, or hand engineer cost functions suitable for a task. Once cost functions are designed then an MDP type framework is used to solve for optimal control policies. However, the question remains, how general is this approach and is there be a more fundamental aspect or metric of HRI that we should be trying to to optimize for rather than task based metrics?

# Transparency in HRI?

- What is transparency in HRI and why is it important?
  - In literature, transparency in a human-robot context can be defined as a method to establish shared intent and shared awareness. However, quantifying transparency in HRI in concrete mathematical terms (possibly using information theoretic concepts) is not to be seen in literature.
  - Transparency is important because it makes each agent’s actions and intent unambiguous to the other, can help in the emergence of cooperative and coordinated behavior and as a result can possibly lead to higher user satisfaction and task performance.
  - Transparency in the human-robot interaction can possibly be used as a **task-agnostic, platform-agnostic metric that can the joint system should try to optimize**. The consequence of which will likely be better task performance.

**Can proper quantification of transparency in HRI and maximization of transparency in HRI lead to better performance in shared autonomy systems?**

# Perception-Action Loops

- Can we possibly treat the human-robot system as a coupled perception action loop?
  - For the human the robot can be part of the human's environment and from the robot's perspective the human will be part of the robot's environment.
- If this is possible, then an (PO)MDP type framework can possibly be used to describe how the joint system evolves.
  - An (PO)MDP that evolves in time is an unrolled **Dynamic Bayesian Network (DBN).**
  - There is a good body of work which studies **'Information Flow'** in Dynamic Bayesian Nets (Ay and Polani). Also related is the notion of **'Empowerment'** (Salge and Polani) which amounts to the influence an agent's actions has on its on future sensor measurements.
  - By relying on metrics from this body of work, we can likely quantify information flow between the different components in a coupled human-robot perception action loop (treated as a DBN).

**Can information flow in the perception-action loop be considered as a concrete mathematical quantification of the notion of transparency?**

# Shaping Information Flow

- Once we quantify information flow, we can possibly shape it.
  - We can possibly treat the design of robot autonomy as a control problem in which the objective is to **maximize information flow (and thereby improve transparency)** between the agents in a shared autonomy system.
  - Maximizing information flow may correspond to and result in the emergence of better communication and expression of bidirectional intent, task capabilities and internal state.
  - How does the choice of state variables affect the computation of information flow or equivalently is there a information flow metric that is invariant to the choice of state representation?

# Possible concrete steps

- Step 1
  - In a simple shared autonomy context, quantify information flow between human and the agent. 2D world with a point robot that is being teleoperated to reach a goal. Autonomy also helps out in task accomplishment. Different levels of autonomy could be tested and the information flow could be computed. One could establish a correlation between the computed metrics and user's perceived transparency of the interaction.
- Step 2
  - With the above notion of transparency and self-knowledge of the robot's actions, active inference based approaches could possibly be used to understand and characterize humans (whether they are cooperative, neutral or adversarial and how they choose to interact with the robot.
- Step 3
  - Once human characteristics are determined, shape robot autonomy in such a way the joint system moves towards states of high information flow.