Phenomena

- **Phenomena**
  - Human-Robot Interaction in the context of shared autonomy.
  - There are two (or maybe more) ‘agents’ (human and the system) interacting and working together to control a physical device, the robot (could be an arm/wheelchair etc).
  - Goal: Mathematical characteriation softer aspects of HRI such as transparency, coordination, legibility etc?
  - Hypothesis: Optimization of transparency leads to successful human robot interaction in terms of task success, user satisfaction, acceptance etc.
- **Discussion Questions**
  - How justified is the hypothesis? Is task success an emergent consequence of transparent and legible human robot interaction? Is the causal relation justified?
  - Will optimization of transparency-type quantities can only help if a baseline objective performance has already been established?

Model

- **Model**
  - What models are commonly used?
    - MDP/POMDP framework - Javdani, Nikolaidis, Dragan, Polani et al.
    - Dynamical Systems/Optimal Control - Sadigh, Murphey et al.
    - Probabilistic Shared Control - Trautman et al.
- **Discussion questions:**
  - What modeling framework best suits the domain?
  - When are MDP/POMDP frameworks not appropriate for modeling?
  - Are there other mathematical modeling frameworks that are more appropriate for quantifying agent-agent interactions?
  - *The choice of modeling framework, I think, is closely associated with the type of questions one wants to answer using the model. Here we are interested in understanding interaction 'strength' (information flow) between the agents (human and the system) and shaping the interaction.*
  - Causal Bayesian Networks?
    - What variables need to be explicitly modeled? This is SUPER critical. (MAIN DISCUSSION)
- THE NEXT 4 SLIDES HAVE 4 DIFFERENT MODEL OPTIONS, EACH WITH DIFFERENT ASSUMPTIONS.

Causal Bayesian Networks (Coupled PA loops) - 1

U^s - System control

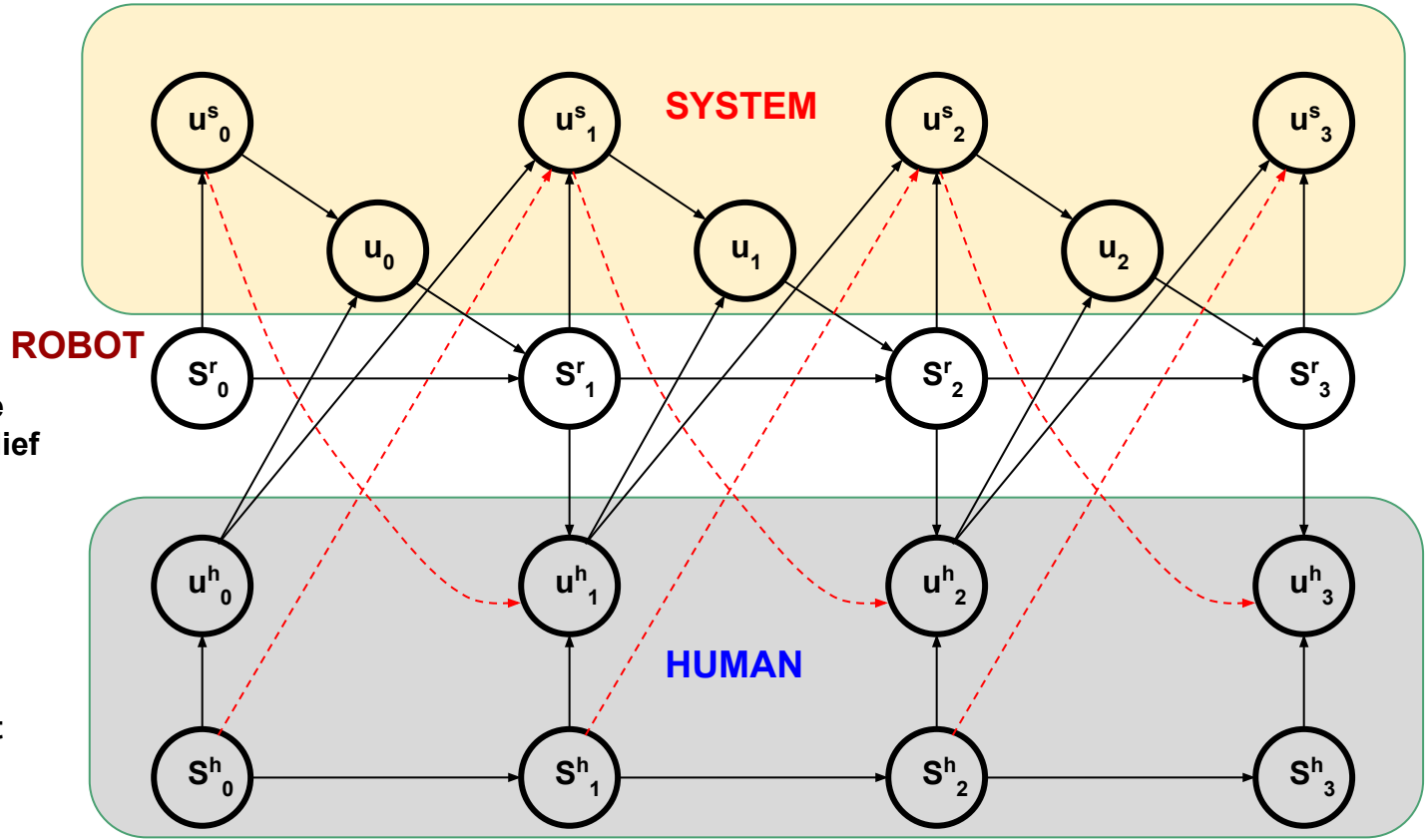
U^h - Human control

U - Shared control

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

S^r - Robot state

Red arrows indicate implicit dependence that need to perceived from available 'observations'



EXPLICIT SYSTEM STATE
- Symmetric Structure with respect to system and human

U^s - System control

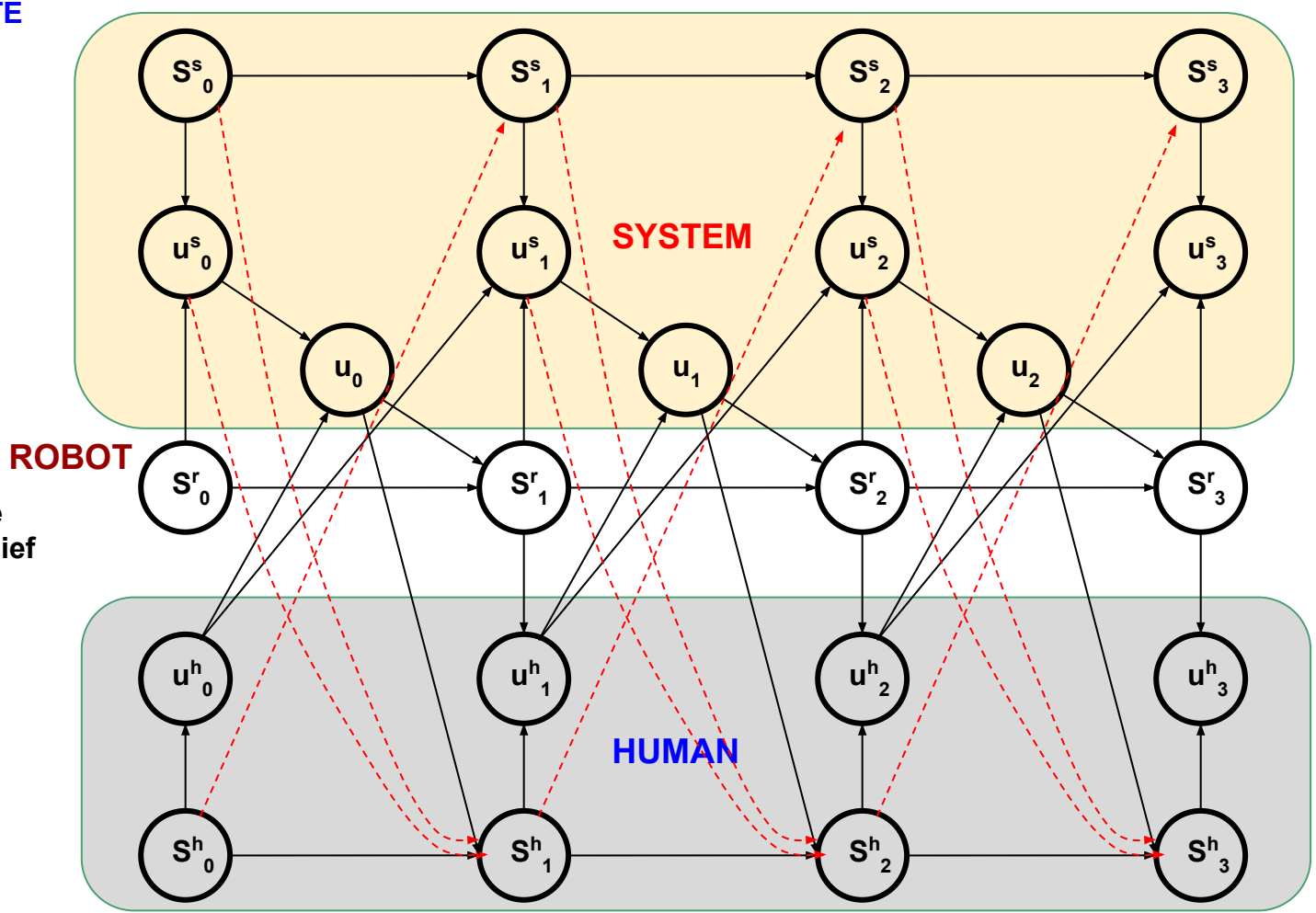
U^h - Human control

U - Shared control

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

S^r - Robot state

S^s - System state



U^h affects U^s indirectly via S^s.

U^s - System control

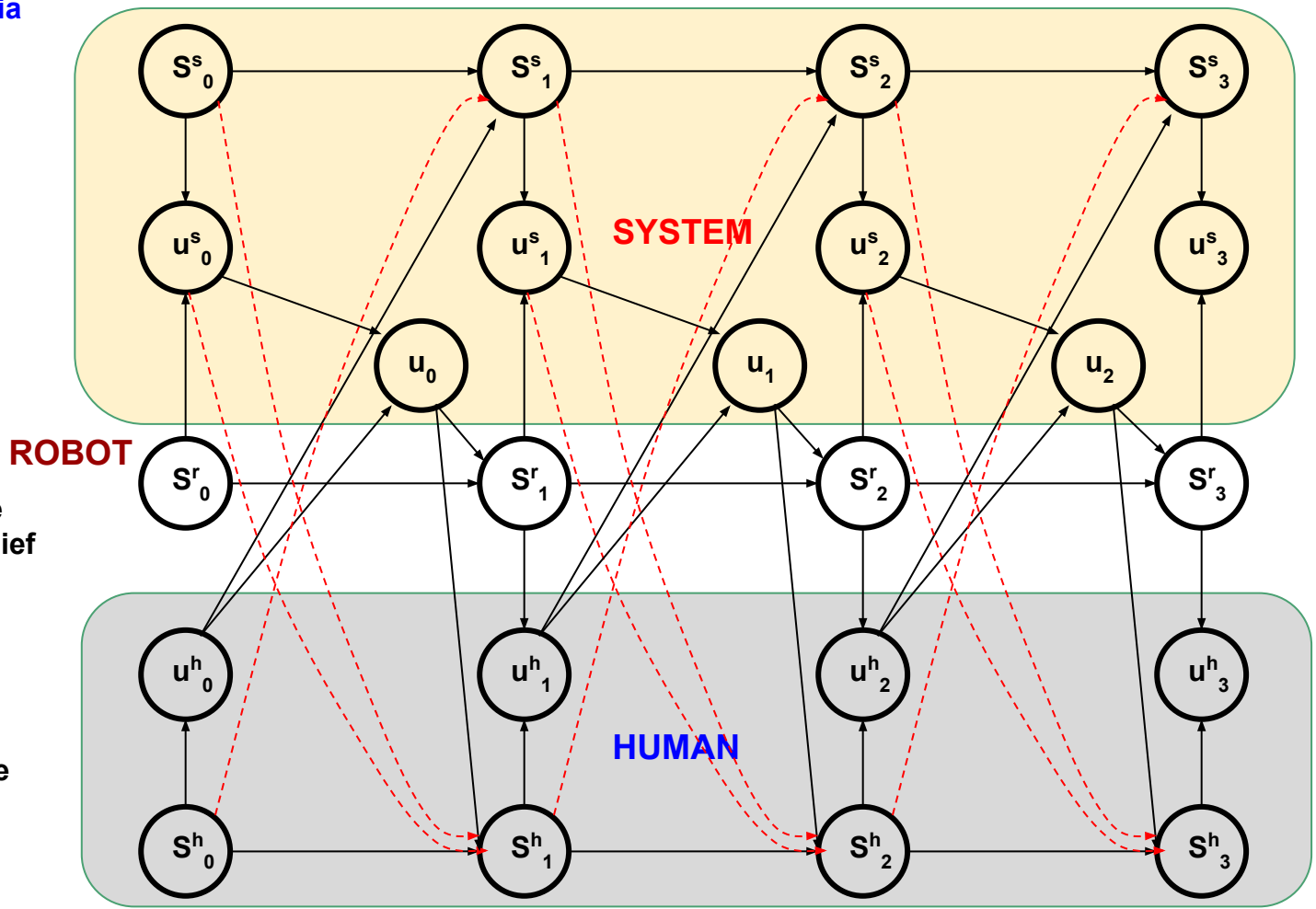
U^h - Human control

U - Shared control

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

S^r - Robot state

S^s - System internal state



U^h affects U^s indirectly via S^s.

U^s - System control

U^h - Human control

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

S^r - Robot state

S^s - System internal state