Principles of Robot Autonomy II

Human-Robot Interaction

Stanford University

iliad
intelligent and interactive autonomous systems

Announcement

Paper presentation for 4-credit students due next Wednesday.

Turn them in on Gradescope (slides for live presenters, video recordings for non-live presenters)

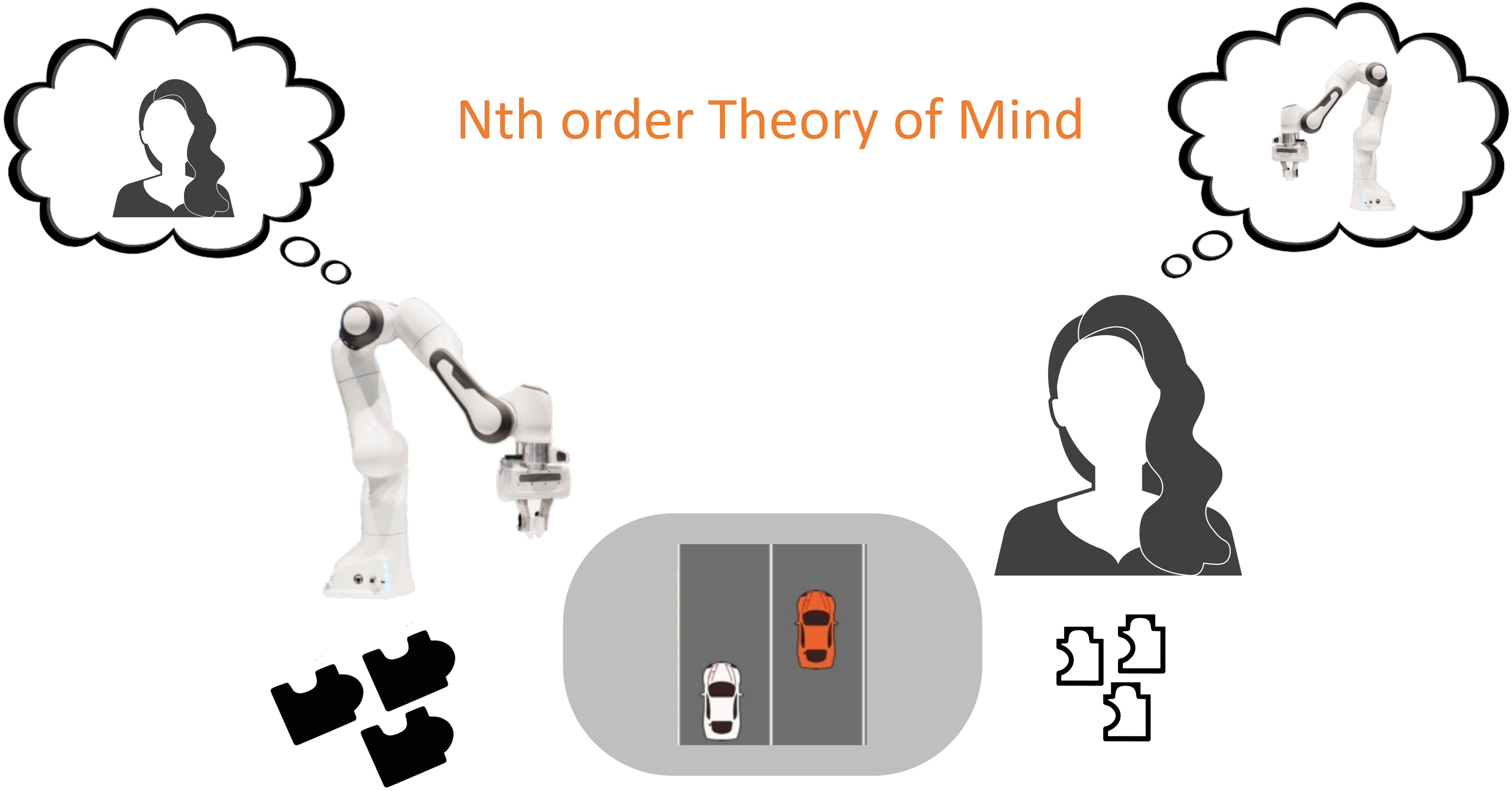
Today’s itinerary

- Game-Theoretic Views on Multi-Agent Interactions
- Partner Modeling: Active Info Gathering over Human’s Intent
- Partner Modeling: Learning and Influencing Latent Intent
- Partner Modeling: Role Assignment

Today’s itinerary

- Game-Theoretic Views on Multi-Agent Interactions
- Partner Modeling: Active Info Gathering over Human’s Intent
- Partner Modeling: Learning and Influencing Latent Intent
- Partner Modeling: Role Assignment

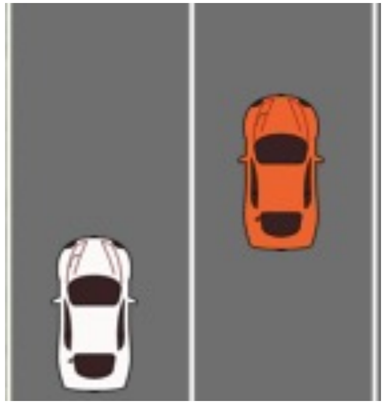
Nth order Theory of Mind



[*Sadigh, Sastry, Seshia, Dragan,* RSS 2016, IROS 2016, AURO 2018]

Interaction as a Dynamical System

$$u_{\mathcal{R}}^* = \underset{u_{\mathcal{R}}}{\operatorname{argmax}} R_{\mathcal{R}}(x, u_{\mathcal{R}}, u_{\mathcal{H}}^*(x, u_{\mathcal{R}}))$$

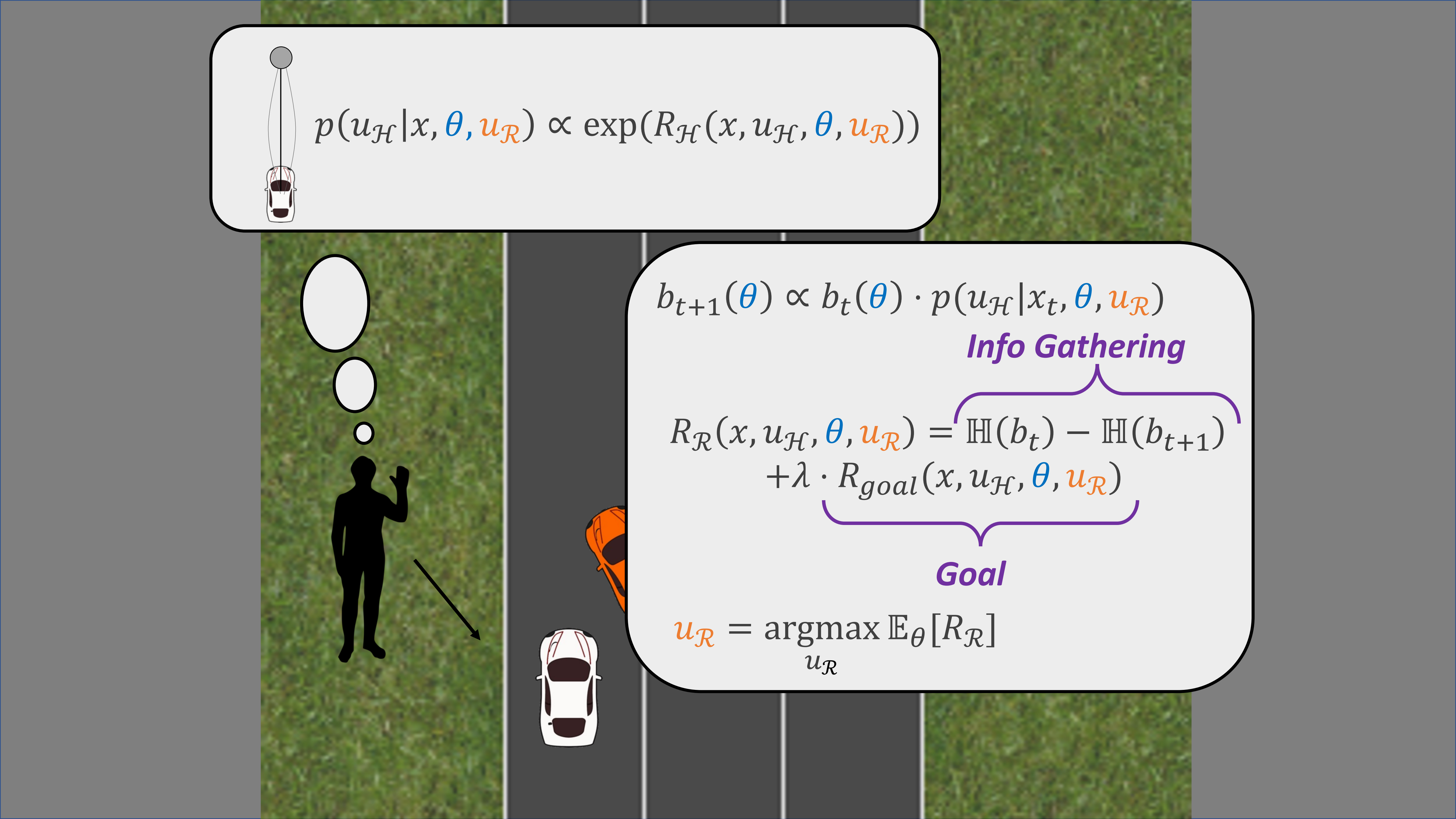


Find optimal actions for the robot while accounting for the human response $u_{\mathcal{H}}^*$.

Model $u_{\mathcal{H}}^*$ as optimizing the human reward function $R_{\mathcal{H}}$.

$$u_{\mathcal{H}}^*(x, u_{\mathcal{R}}) \approx \underset{u_{\mathcal{H}}}{\operatorname{argmax}} R_{\mathcal{H}}(x, u_{\mathcal{R}}, u_{\mathcal{H}})$$

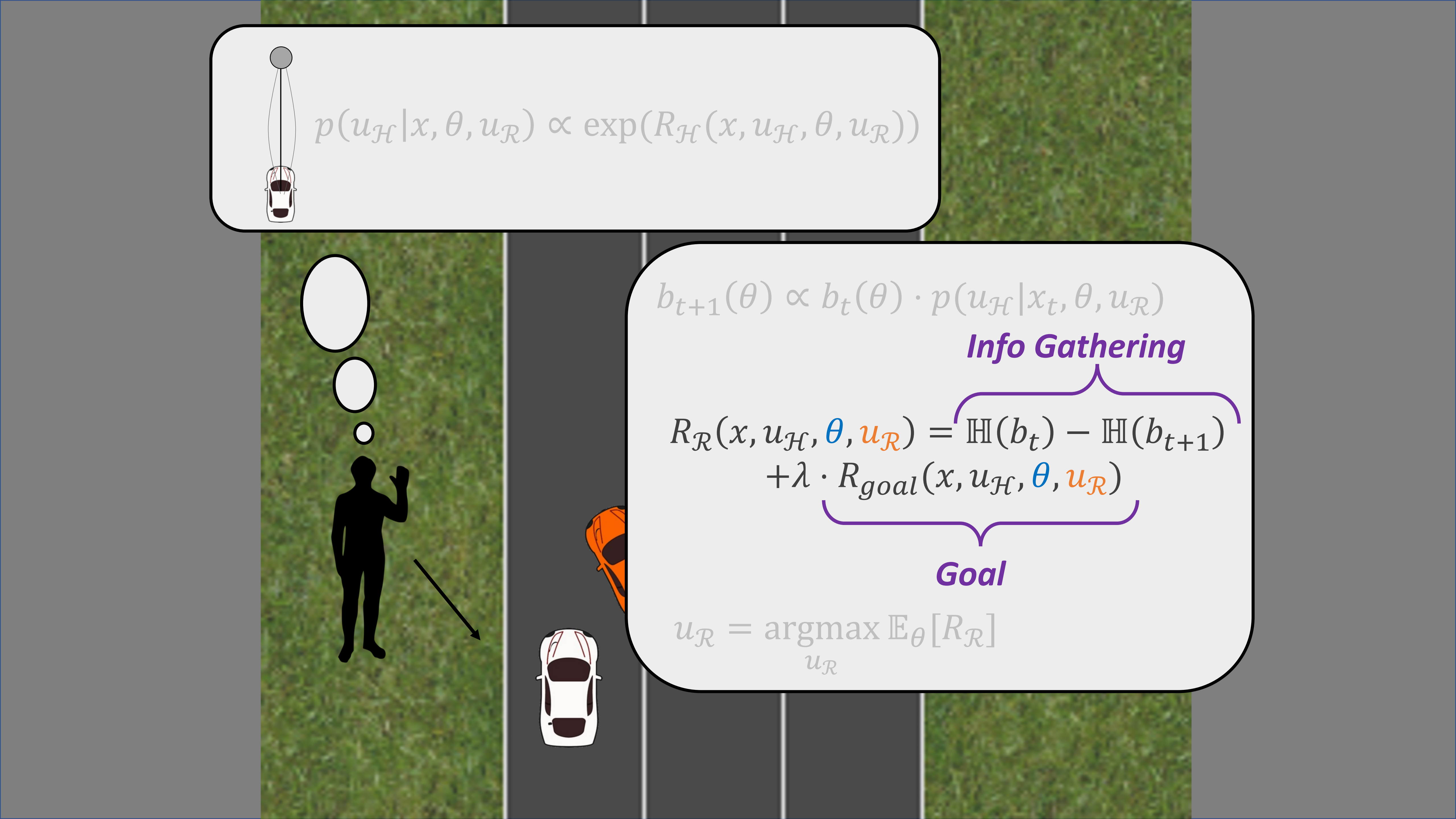
Sadigh et al. RSS 2016, AURO 2018



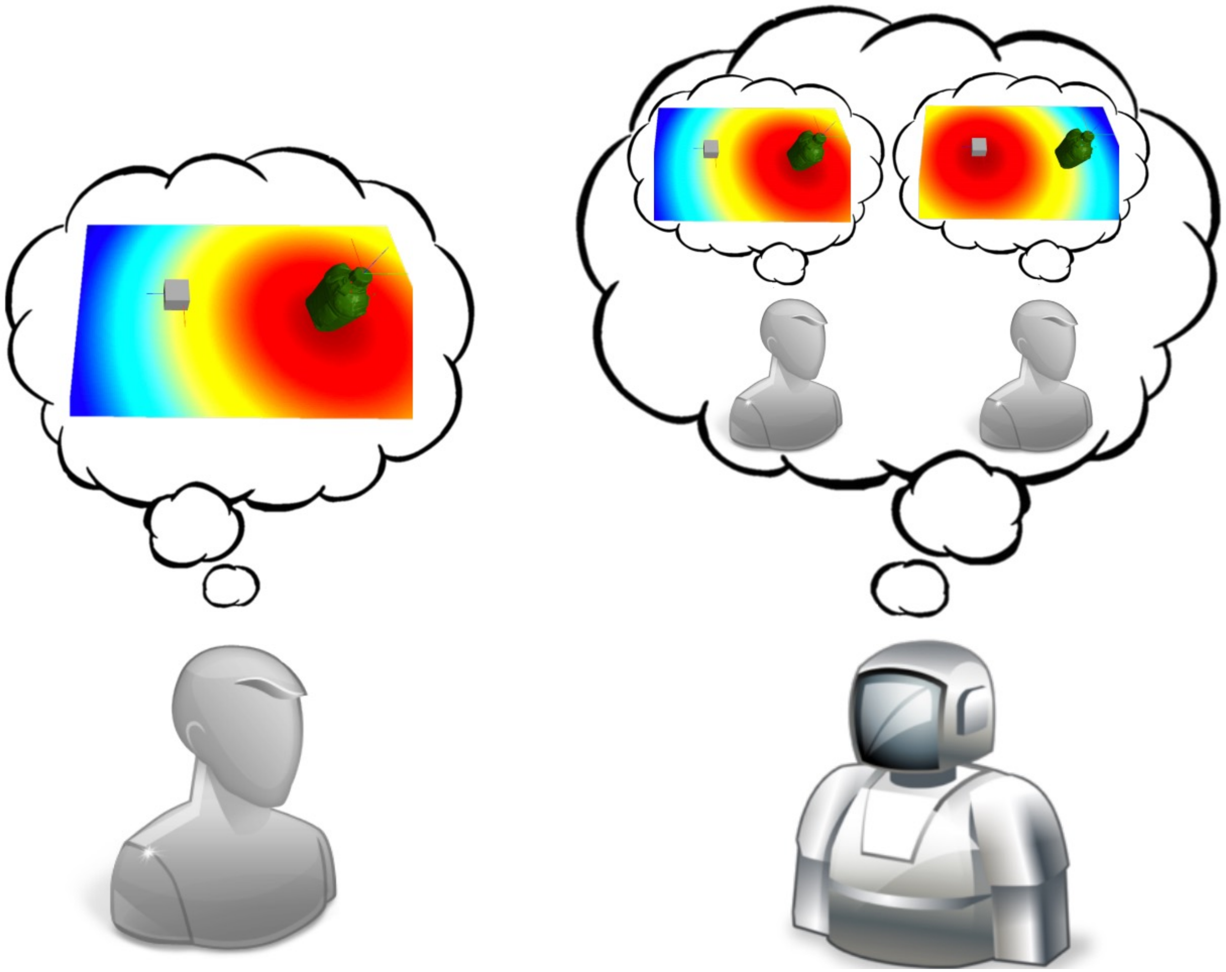
$p(u_{\mathcal{H}} | x, \theta, u_{\mathcal{R}}) \propto \exp(R_{\mathcal{H}}(x, u_{\mathcal{H}}, \theta, u_{\mathcal{R}}))$
$b_{t+1}(\theta) \propto b_t(\theta) \cdot p(u_{\mathcal{H}} | x_t, \theta, u_{\mathcal{R}})$
$R_{\mathcal{R}}(x, u_{\mathcal{H}}, \theta, u_{\mathcal{R}}) = \overbrace{\mathbb{H}(b_t) - \mathbb{H}(b_{t+1})}^{\textit{Info Gathering}} + \lambda \cdot \underbrace{R_{goal}(x, u_{\mathcal{H}}, \theta, u_{\mathcal{R}})}_{\textit{Goal}}$
$u_{\mathcal{R}} = \underset{u_{\mathcal{R}}}{\operatorname{argmax}} \, \mathbb{E}_{\theta}[R_{\mathcal{R}}]$

Today’s itinerary

- Game-Theoretic Views on Multi-Agent Interactions
- Partner Modeling: Active Info Gathering over Human’s Intent
- Partner Modeling: Learning and Influencing Latent Intent
- Partner Modeling: Role Assignment


$p(u_{\mathcal{H}}|x,\theta,u_{\mathcal{R}}) \propto \exp(R_{\mathcal{H}}(x,u_{\mathcal{H}},\theta,u_{\mathcal{R}}))$
$b_{t+1}(\theta) \propto b_t(\theta) \cdot p(u_{\mathcal{H}}|x_t,\theta,u_{\mathcal{R}})$
$R_{\mathcal{R}}(x,u_{\mathcal{H}},\theta,u_{\mathcal{R}}) = \overbrace{\mathbb{H}(b_t) - \mathbb{H}(b_{t+1})}^{\text{Info Gathering}}$
$+\lambda \cdot \underbrace{R_{goal}(x,u_{\mathcal{H}},\theta,u_{\mathcal{R}})}_{\text{Goal}}$
$u_{\mathcal{R}} = \operatorname{argmax}_{u_{\mathcal{R}}} \mathbb{E}_{\theta}[R_{\mathcal{R}}]$

Modeling Intent Inference using POMDPs



[Javdani et al.]

POMDP Formulation

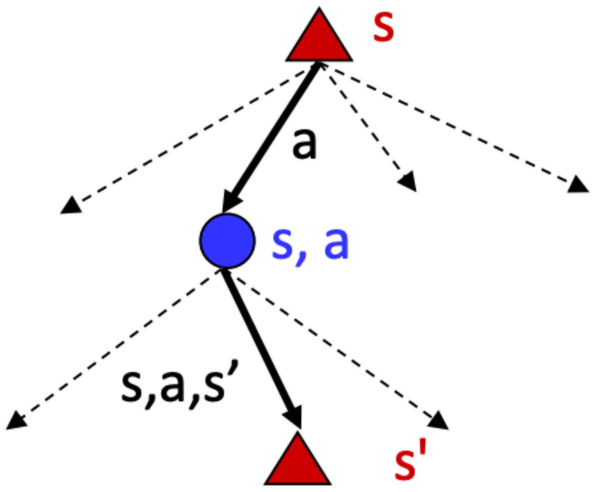
MDPs have:

States S

Actions A

Transition Function $P(s'|s,a)$

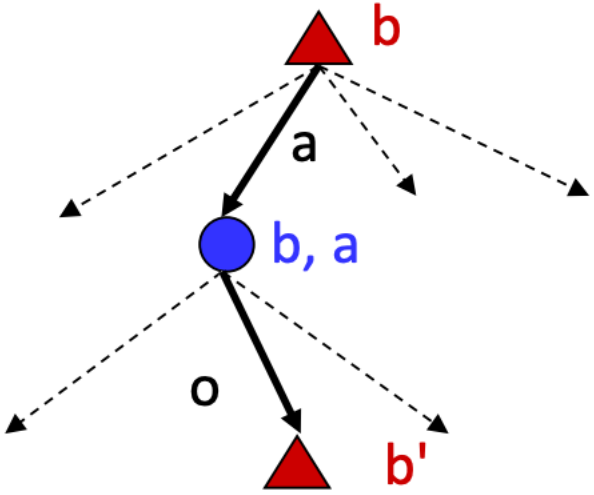
Reward $R(s,a,s')$



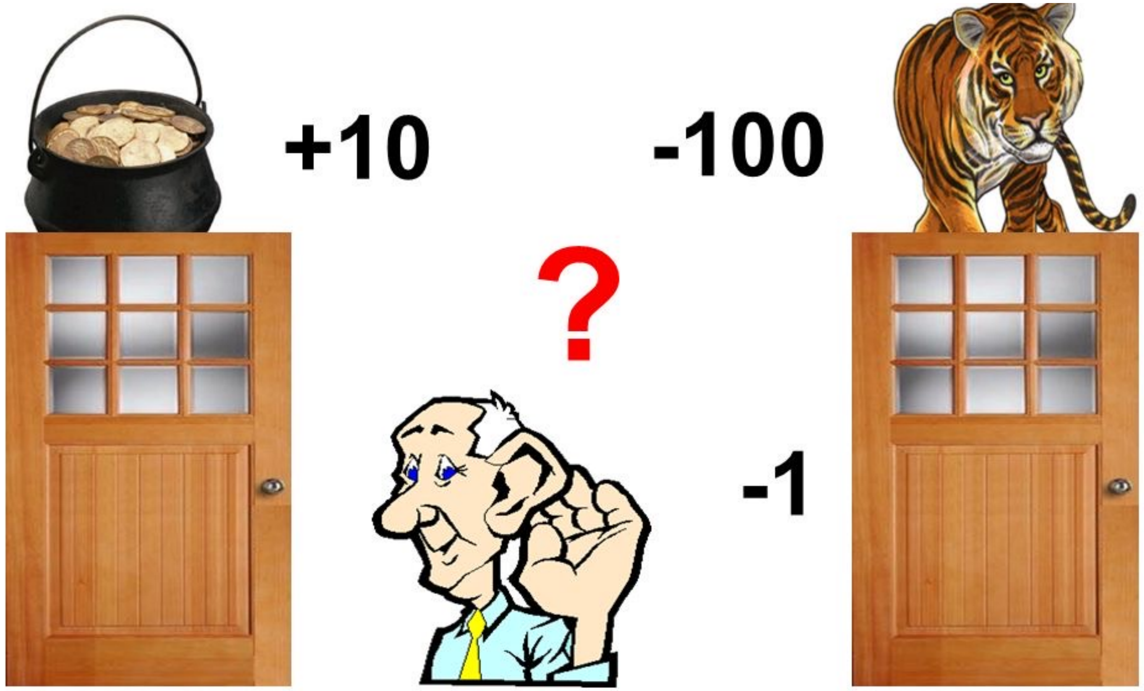
POMDPs add:

Observations O

Observation Function $P(o|s)$



Tiger Example



Actions $a = \{0, : \text{listen } 1\text{: open left}, 2\text{: open right}\}$

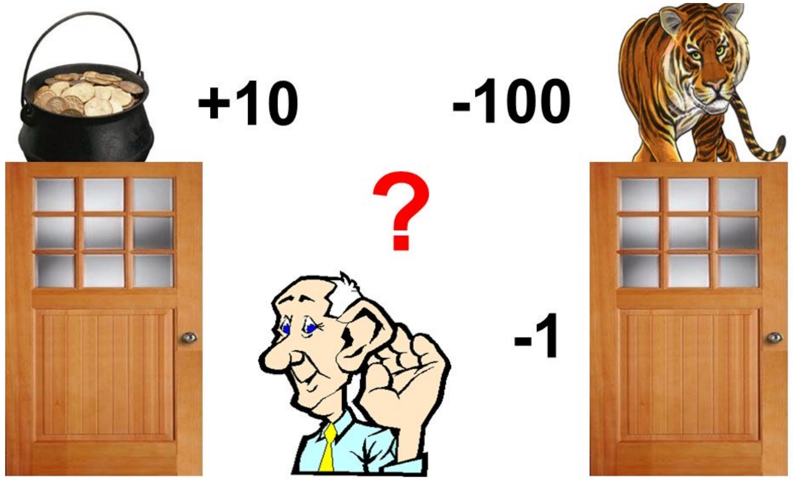
Reward Function:

- Penalty for wrong opening: -100
- Reward for correct opening: +10
- Cost of listening: -1

Observations:

- To hear the tiger on the left
- To hear the tiger on the right

Tiger Example



Belief update based on observations:

$$b_1(s_i) \propto p(o|s_i, a) \sum_{s_j \in S} p(s_i|s_j, a) \cdot b_0(s_j)$$

Immediate return *Discounted future return*

Value Iteration over Beliefs

$$V^*(b) = \max_{a \in A}[\sum_{s \in S} b(s) \cdot R(s, a) + \gamma \sum_{o \in O} P(o|b, a) \cdot V^*(b_o^a)\,]$$

Hard to compute continuous space MDPs -> Approximation

Tiger Example

Immediate return *Discounted future return*

Value Iteration over Beliefs

$$V^*(b) = \max_{a \in A}\Big[\sum_{s \in S} b(s) \cdot R(s,a) + \gamma \sum_{o \in O} P(o|b,a) \cdot V^*(b_o^a)\Big]$$

Hard to compute continuous space MDPs -> Approximation

Q-MDP Approximation

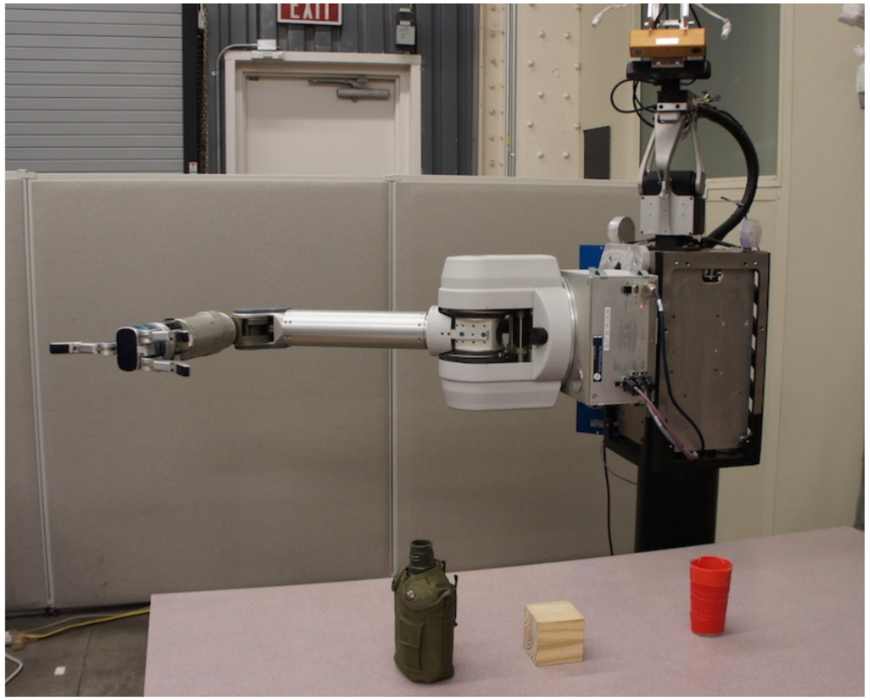
$$V^*(b) = \mathbb{E}_s[V^*(s)] = \sum_s b(s) \cdot V^*(s)$$

Intent Inference

X Robot States

A Robot Actions

$T: X \times A \rightarrow X$ Transition function



$u \in U$ Human continuous input

$D: U \rightarrow A$ Mapping between human input and robot actions

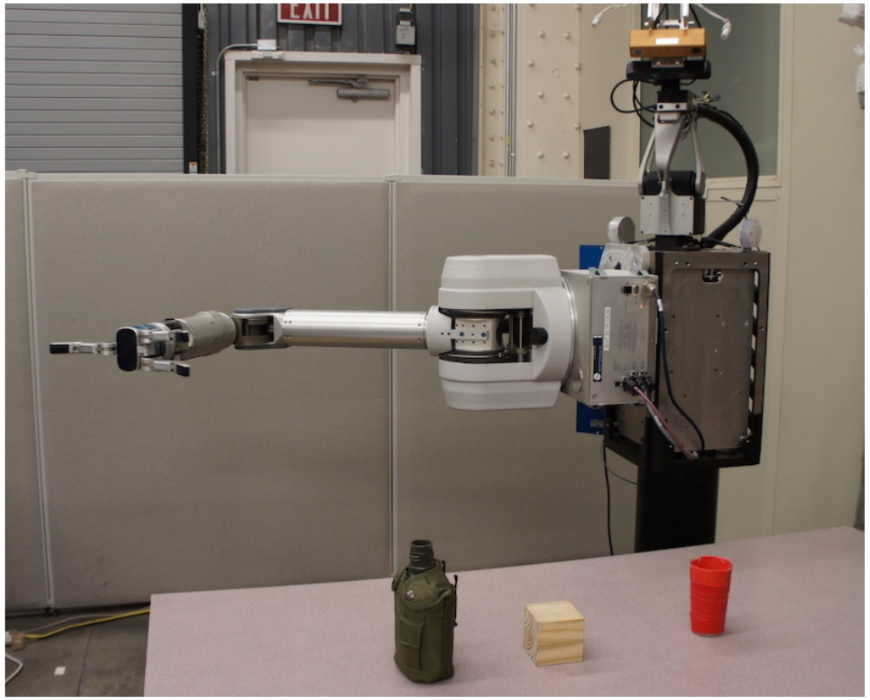
User’s Policy is Learned from IRL

$\pi_g^{usr}(x) = p(u|x, g)$ We learn a policy for each goal

$p(\xi|g) \propto \exp(-C_g^{usr}(\xi))$

$p(g|\xi) \propto p(\xi|g) \cdot p(g)$ Bayes Rule

POMDP Observation Model



Hindsight Optimization (Q-MDP)

Estimate cost-to-go of the belief by assuming full observability will be obtained at the next time step.

You never gather information, but can plan efficiently in deterministic subproblems.

$$b(s) = b(g) = p(g|\xi)$$

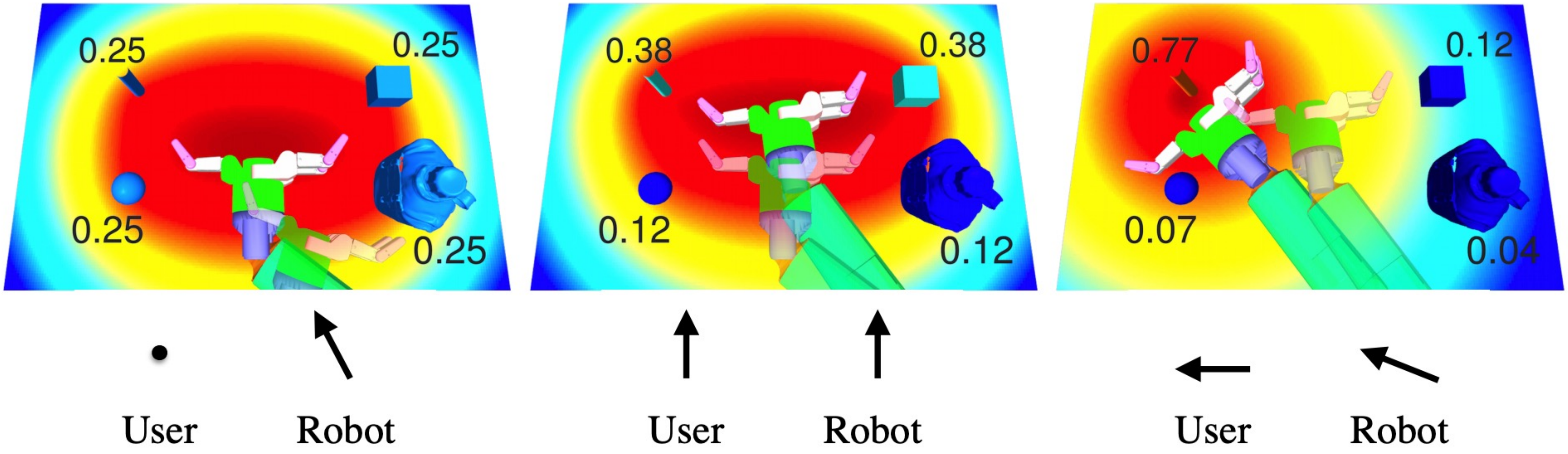
Uncertainty is only over goals

$$Q(b, a, u) = \sum_g b(g) \cdot Q_g(x, a, u)$$

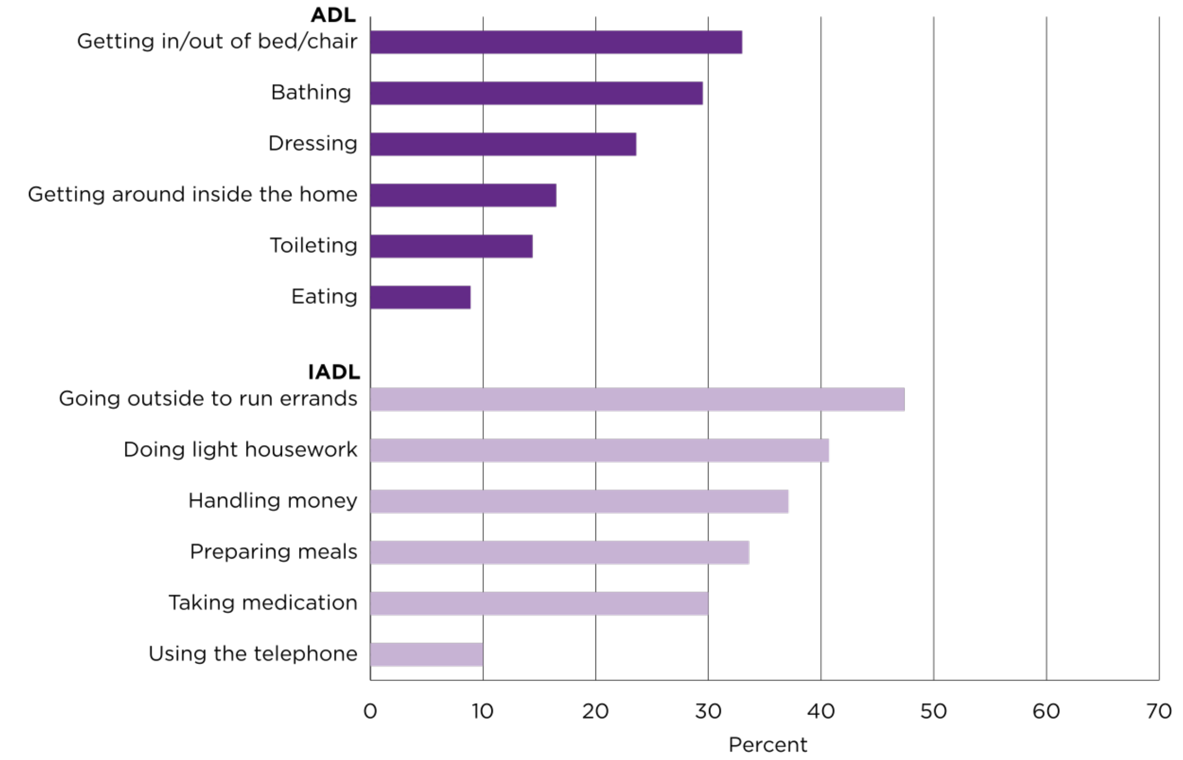
Action-Value function of the POMDP

Cost-to-Go of Acting optimally and going towards goal g

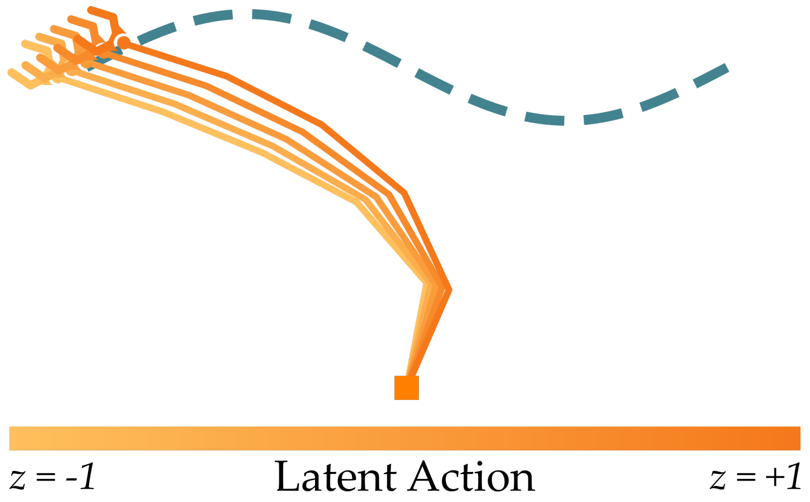
Shared Autonomy with Hindsight Optimization



Prevalence of Difficulty Performing ADLs and IADLs in Adults 18 Years and Older With One or More Selected Symptoms That Interfere With Everyday Activities: 2014

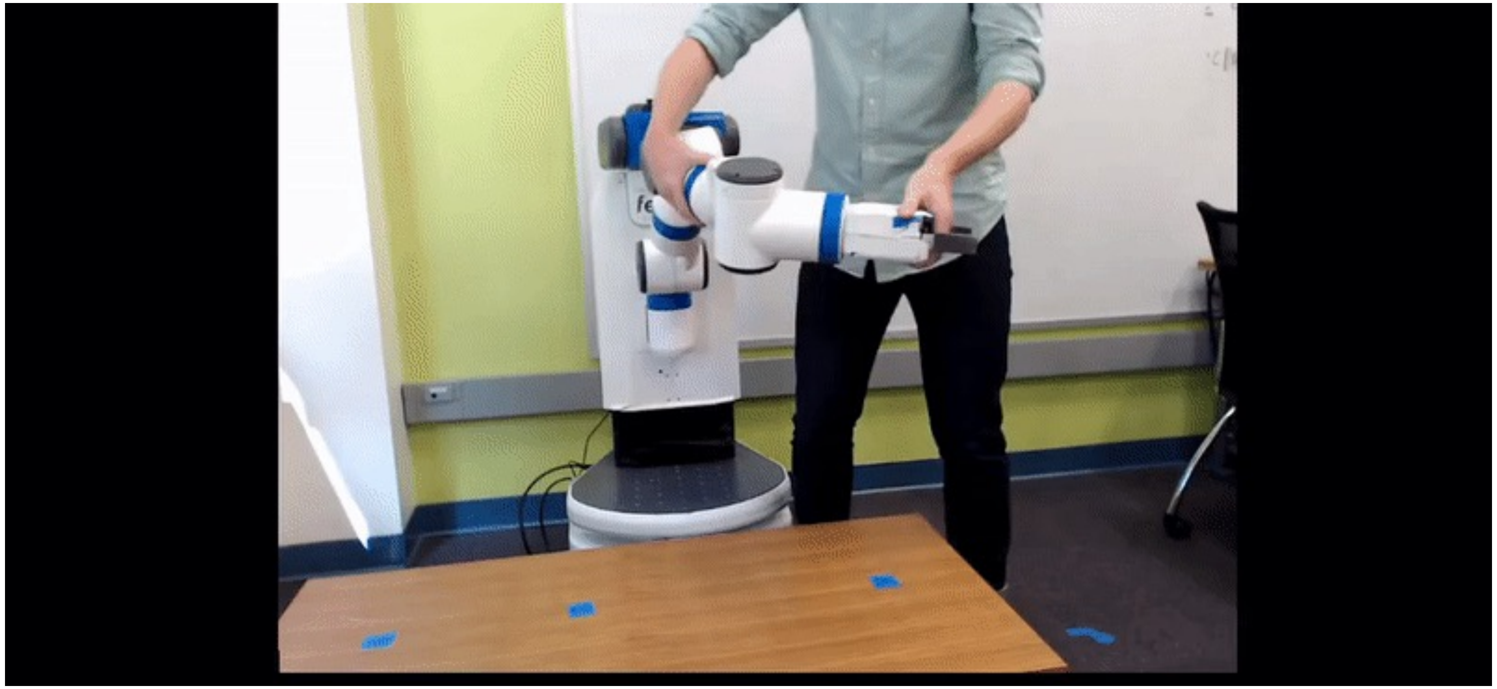


Source: U.S. Census Bureau, Social Security Administration Supplement to the 2014 Panel of the Survey of Income and Program Participation, September–November 2014.



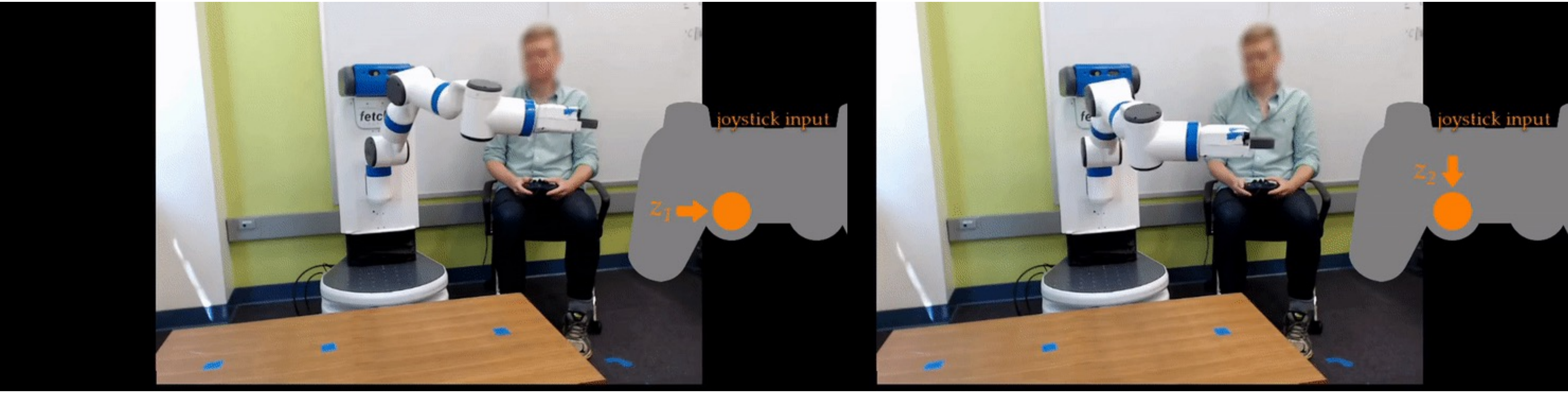
- Assistive robotic arms are ***dexterous***
- This dexterity makes it hard for users to ***control*** the robot
- How can robots ***learn*** low-dimensional representations that make controlling the robot intuitive?

Our Vision



Offline, expert demonstrations of ***high-dimensional*** motions

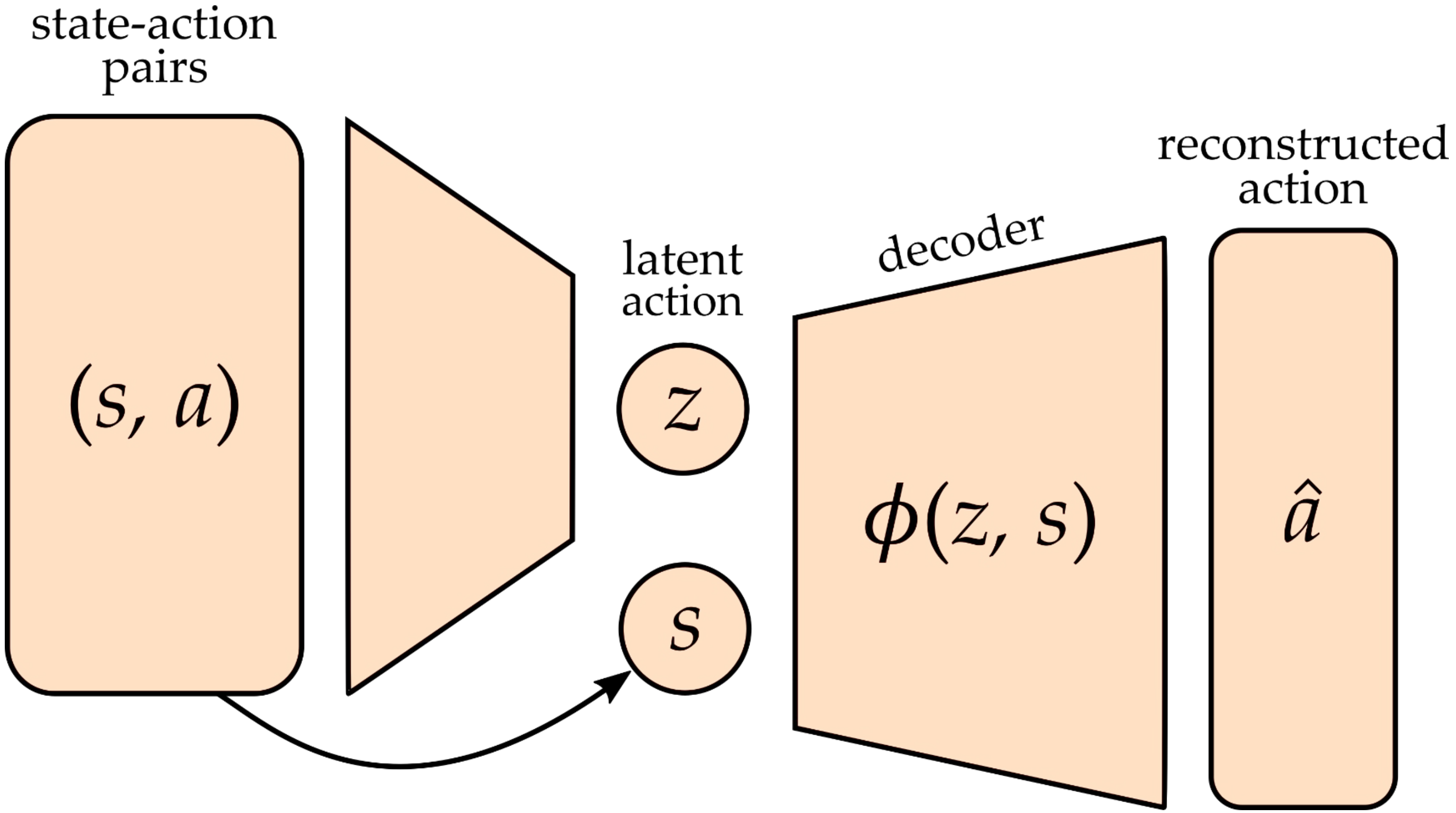
Our Vision



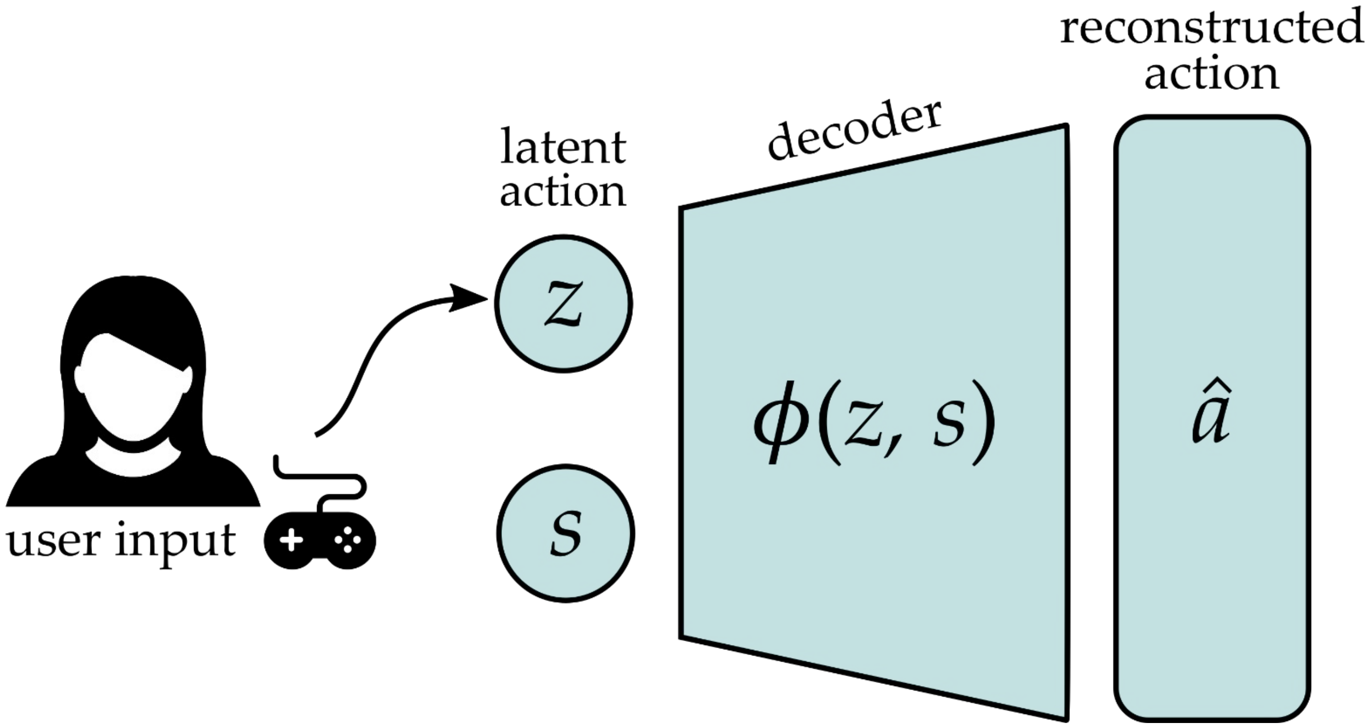
Learn *low-dimensional* latent representations for online control

We make it easier to control ***high-dimensional*** robots by ***embedding*** the robot's actions into a ***low-dimensional*** latent space.

Model Structure (cVAE)



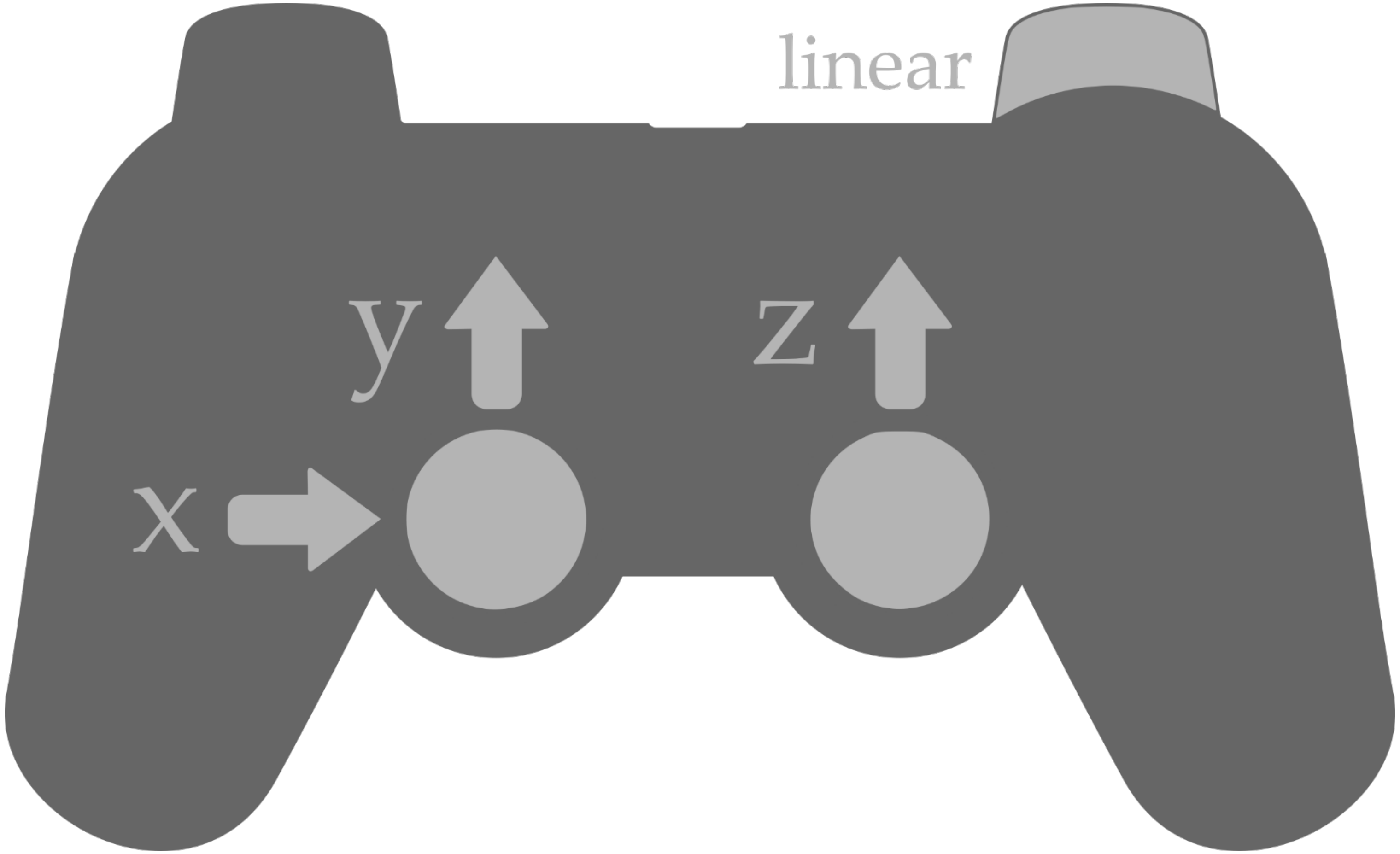
Model Structure (cVAE)

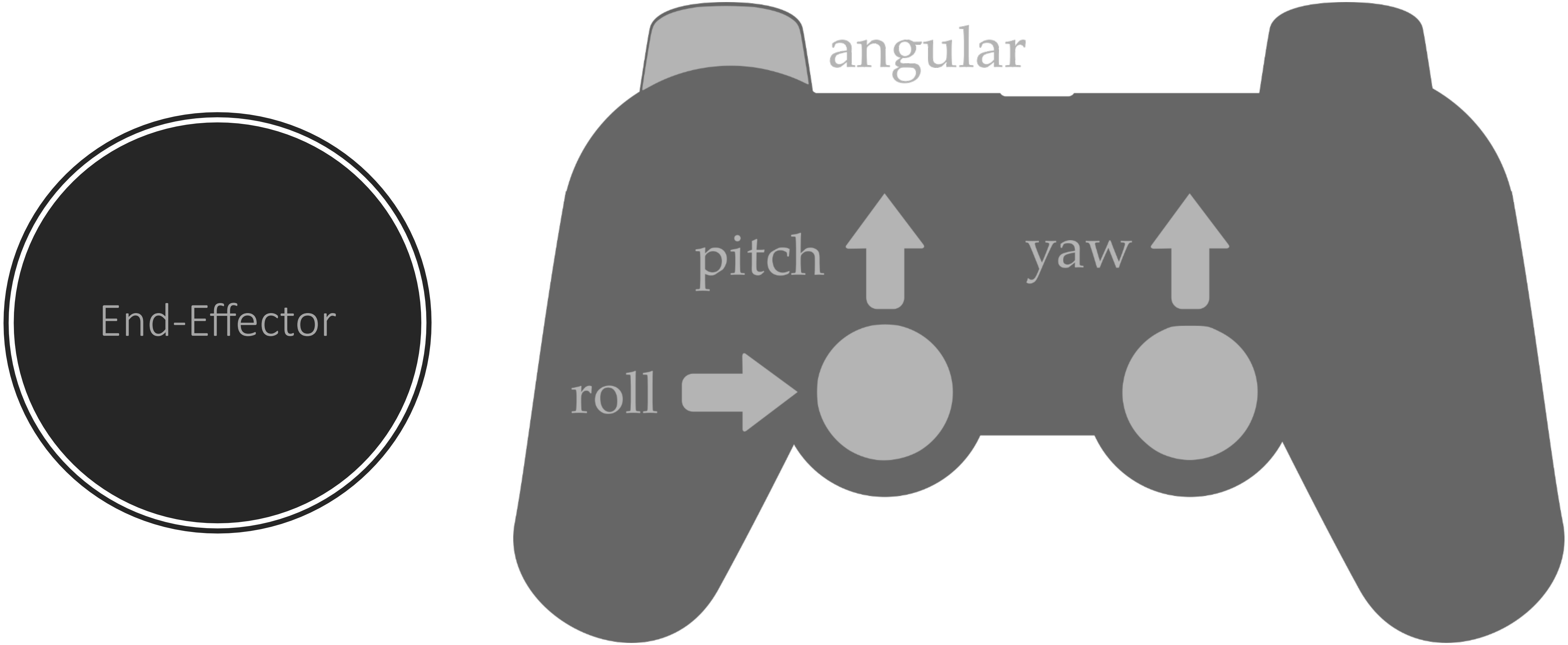


User Study

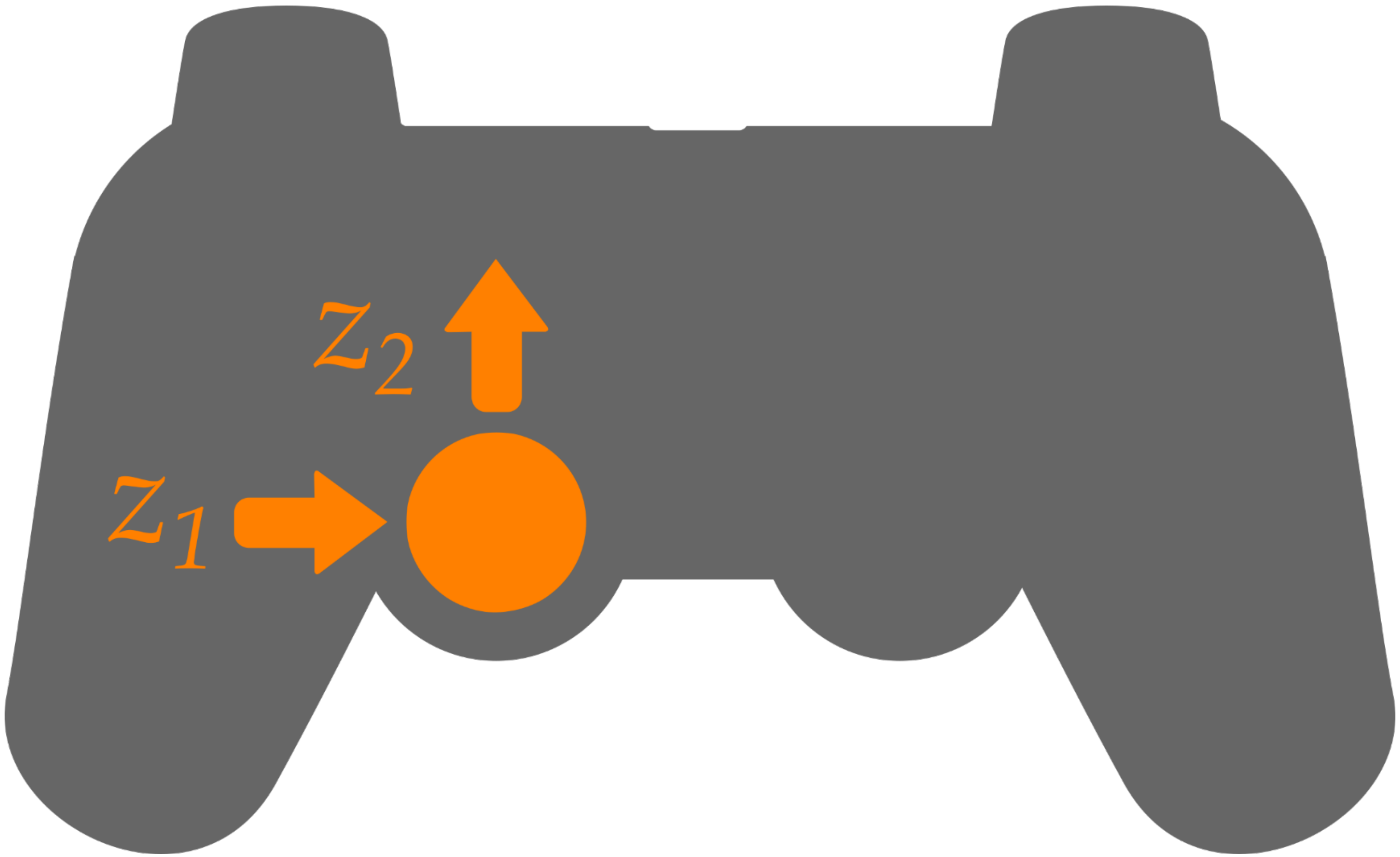
- We trained on less than ***7 minutes*** of kinesthetic demonstrations
- Demonstrations consisted of moving between shelves, pouring, stirring, and reaching motions
- We compared our ***Latent Action*** to the current method for assistive robotic arms (***End-Effector***)


End-Effector


linear
y
z
x


End-Effector
angular
pitch
yaw
roll

Latent Actions



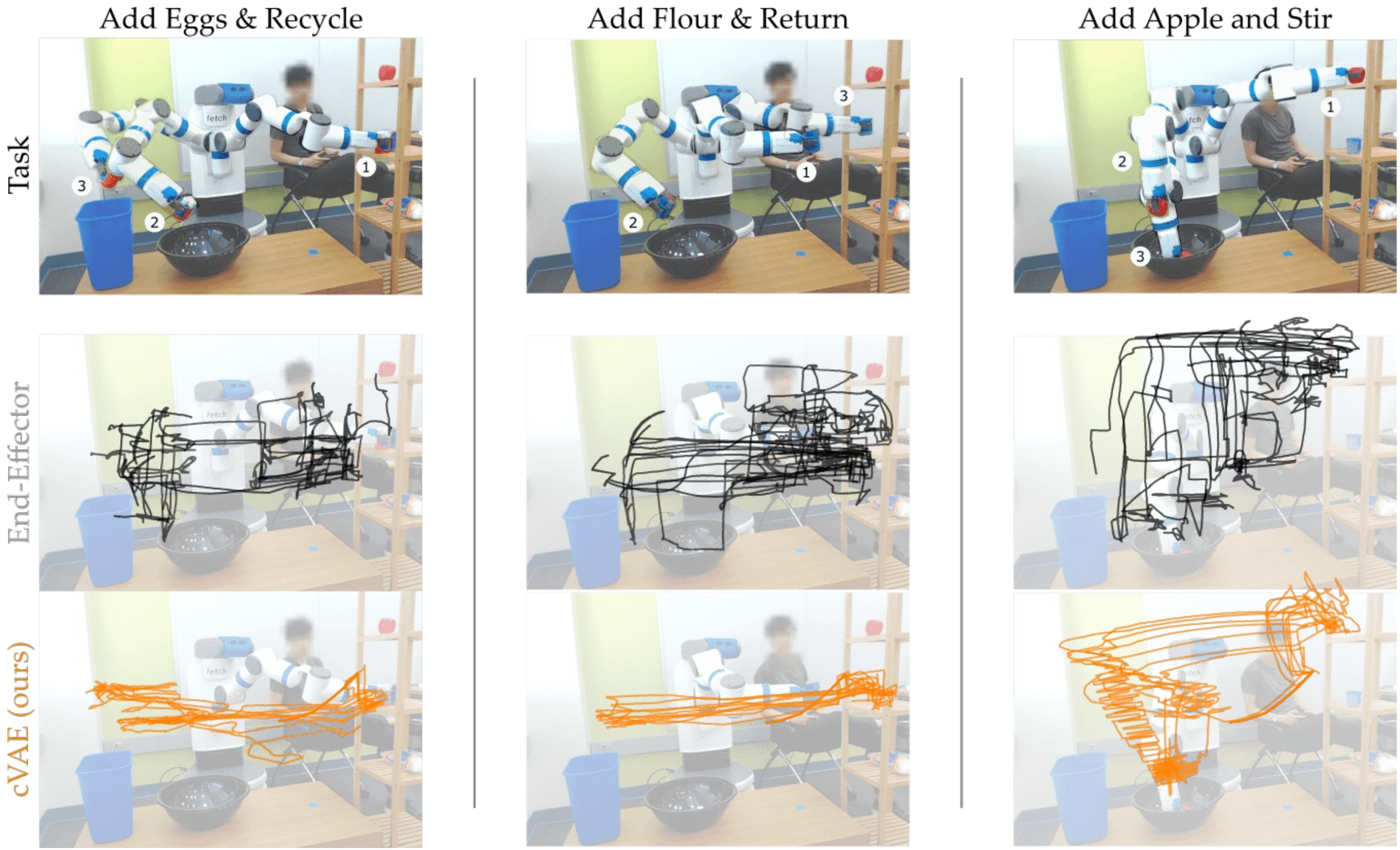
4x Speed

(1) add eggs

End-Effector

(1) add eggs

Latent Action



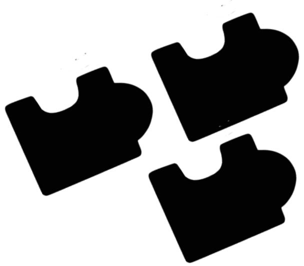
Add Eggs & Recycle
Add Flour & Return
Add Apple and Stir
Task
End-Effector
cVAE (ours)
1
2
3

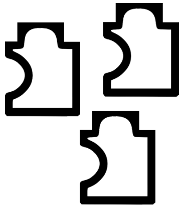
Today’s itinerary

- Game-Theoretic Views on Multi-Agent Interactions
- Partner Modeling: Active Info Gathering over Human’s Intent
- Partner Modeling: Learning and Influencing Latent Intent
- Partner Modeling: Role Assignment

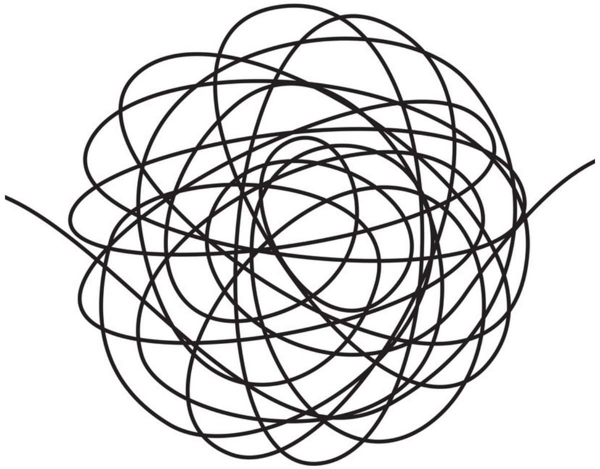
Nth order Theory of Mind

Most interactive tasks are not the same as playing chess!





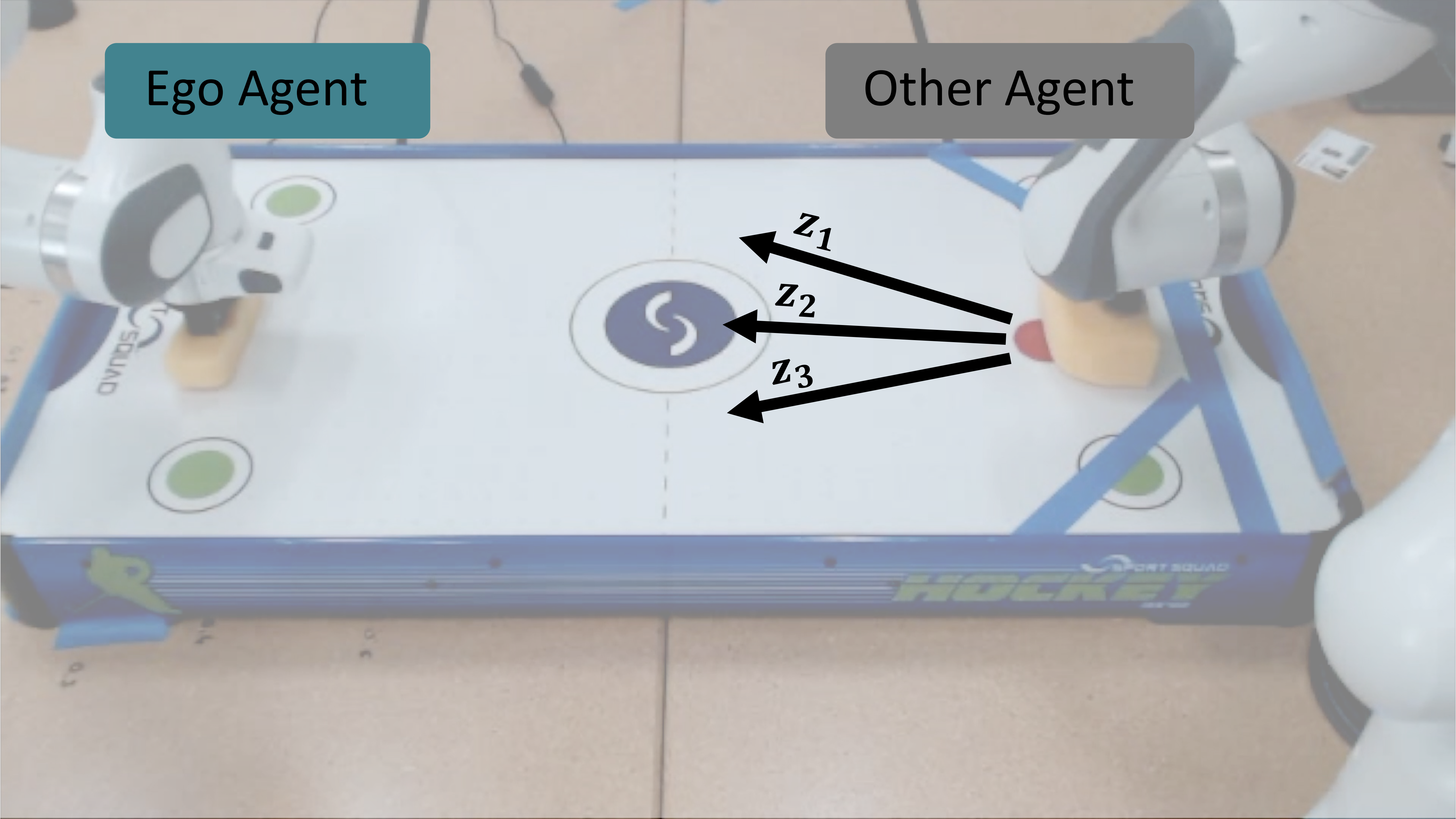
… low-dimensional shared representation
that captures the interaction and can change over time.

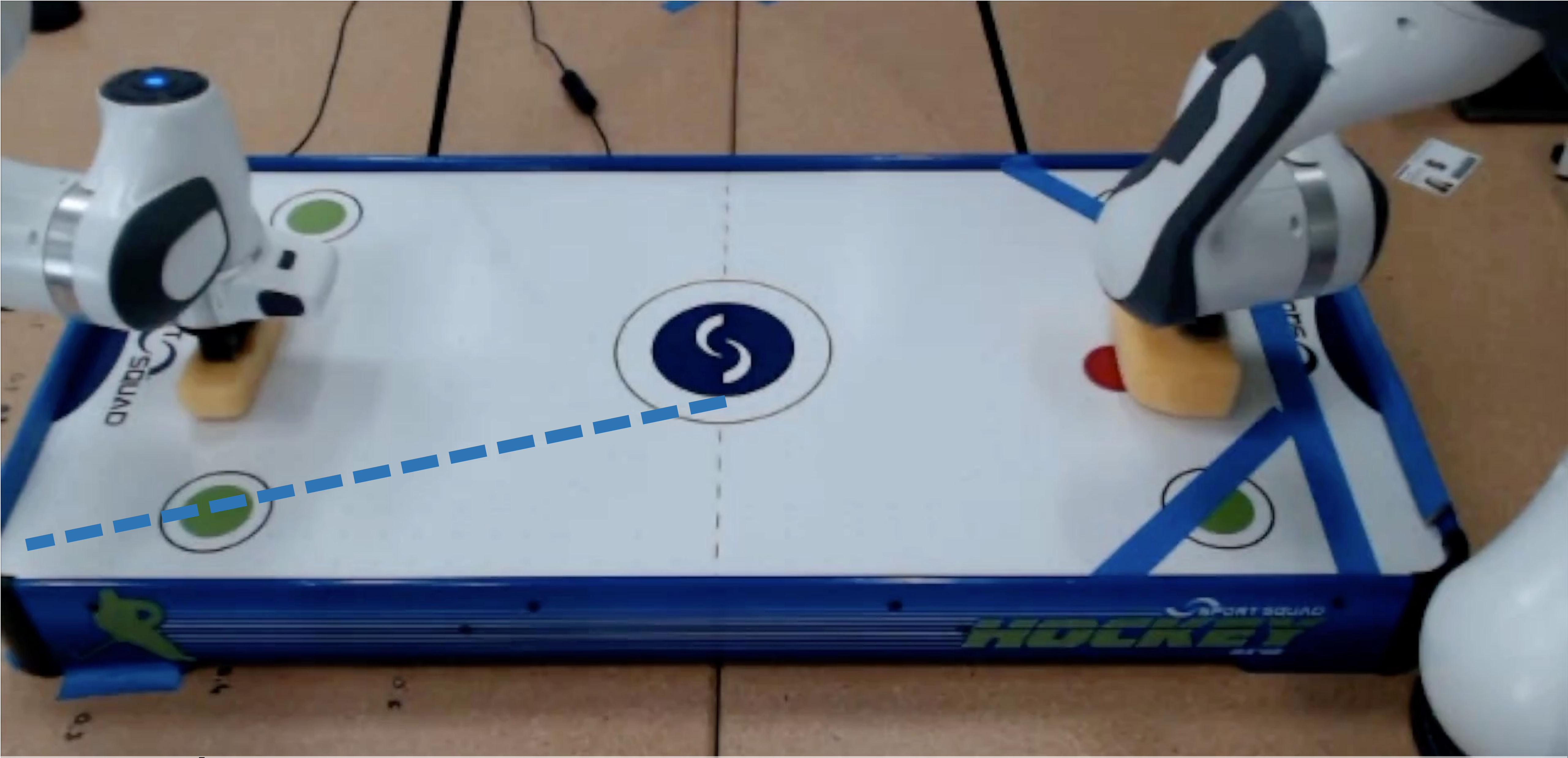


Other agents are often non-stationary:
They update their behavior in response to the robot.


Ego Agent
Other Agent
SPORT SQUAD
HOCKEY

$$a \in \mathbb{R}^7$$


Ego Agent
Other Agent
z_1
z_2
z_3

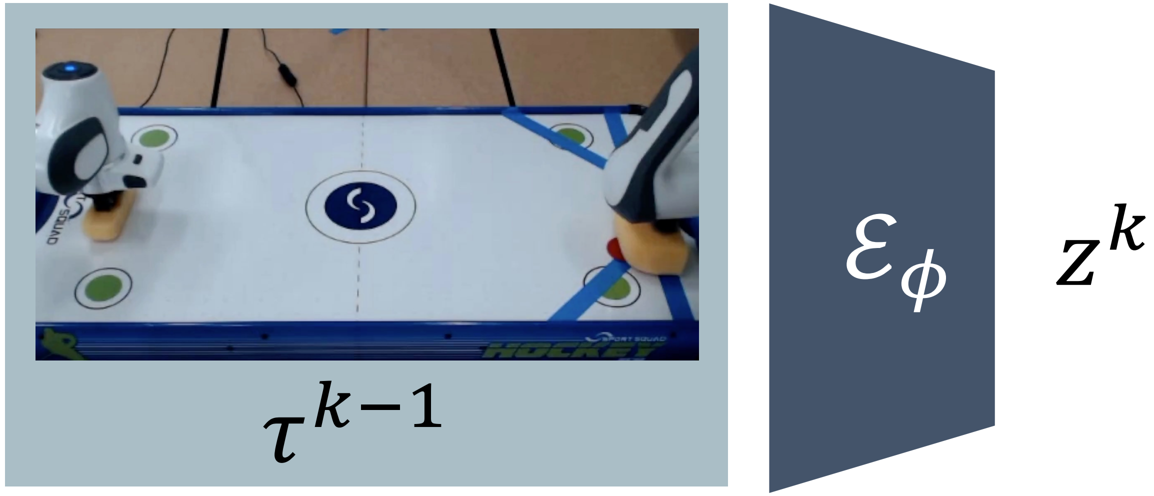


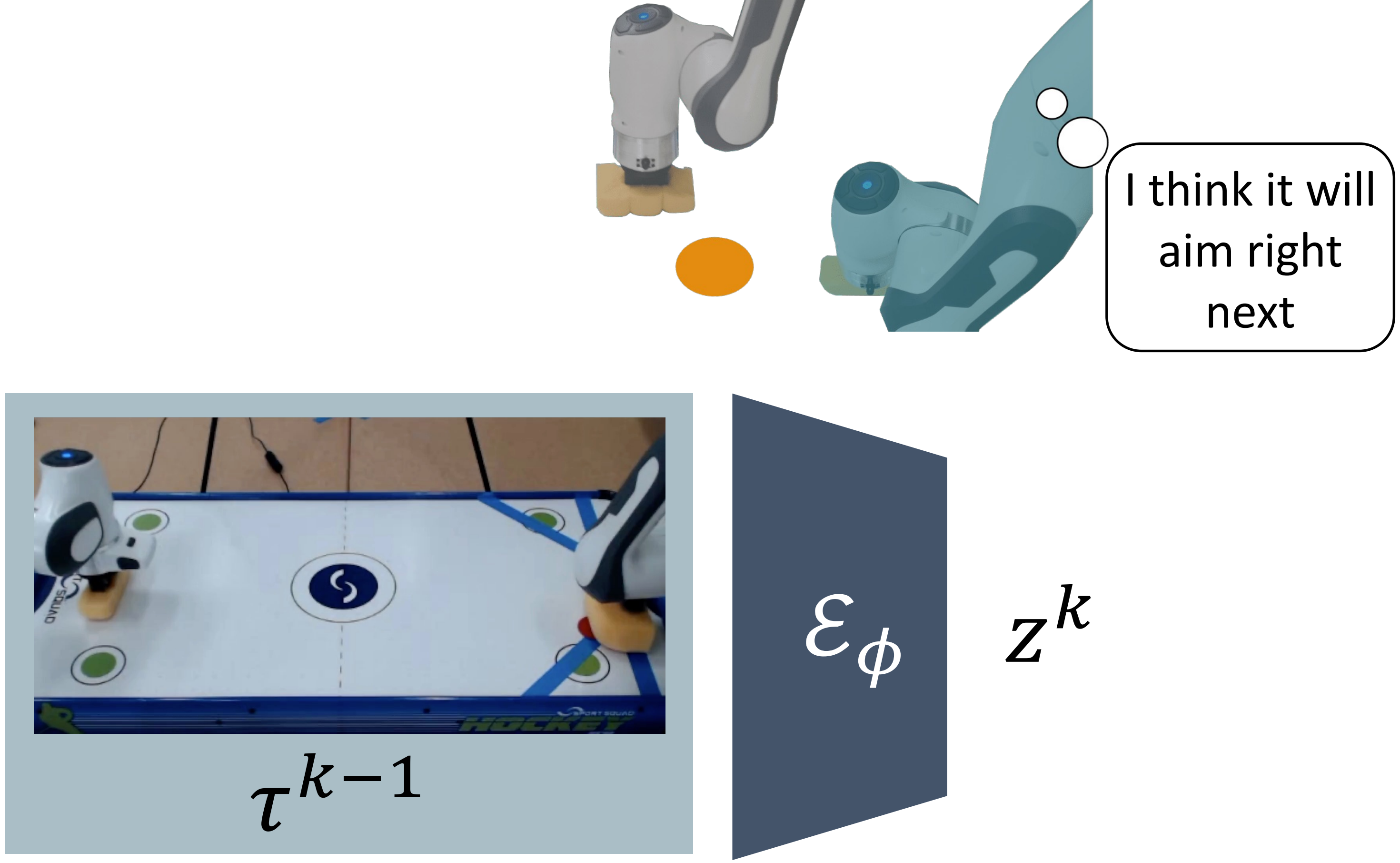
$$\tau^i = \{(s_1, a_1, r_1), \ldots, (s_H, a_H, r_H)\}$$

$$z^{i+1} \sim f(\cdot \mid z^{i}, \tau^{i})$$

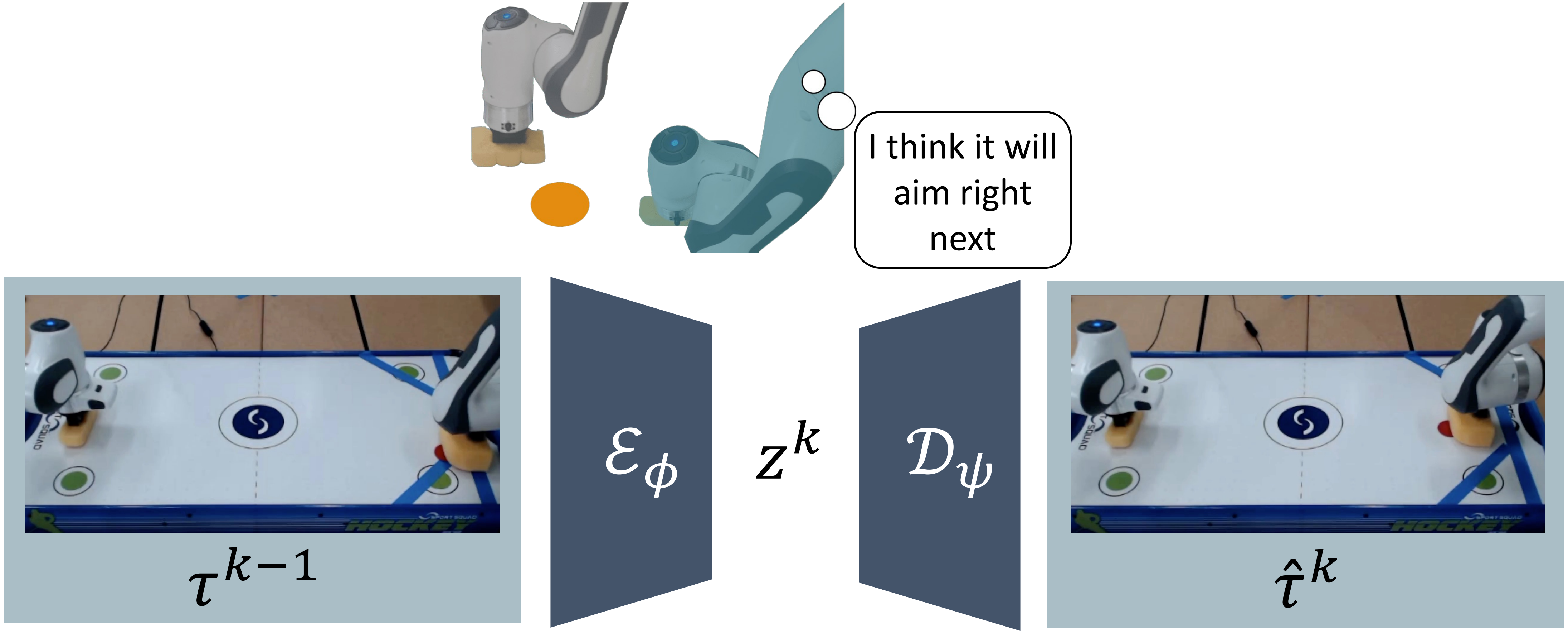
Modeling Other Agent’s Behavior

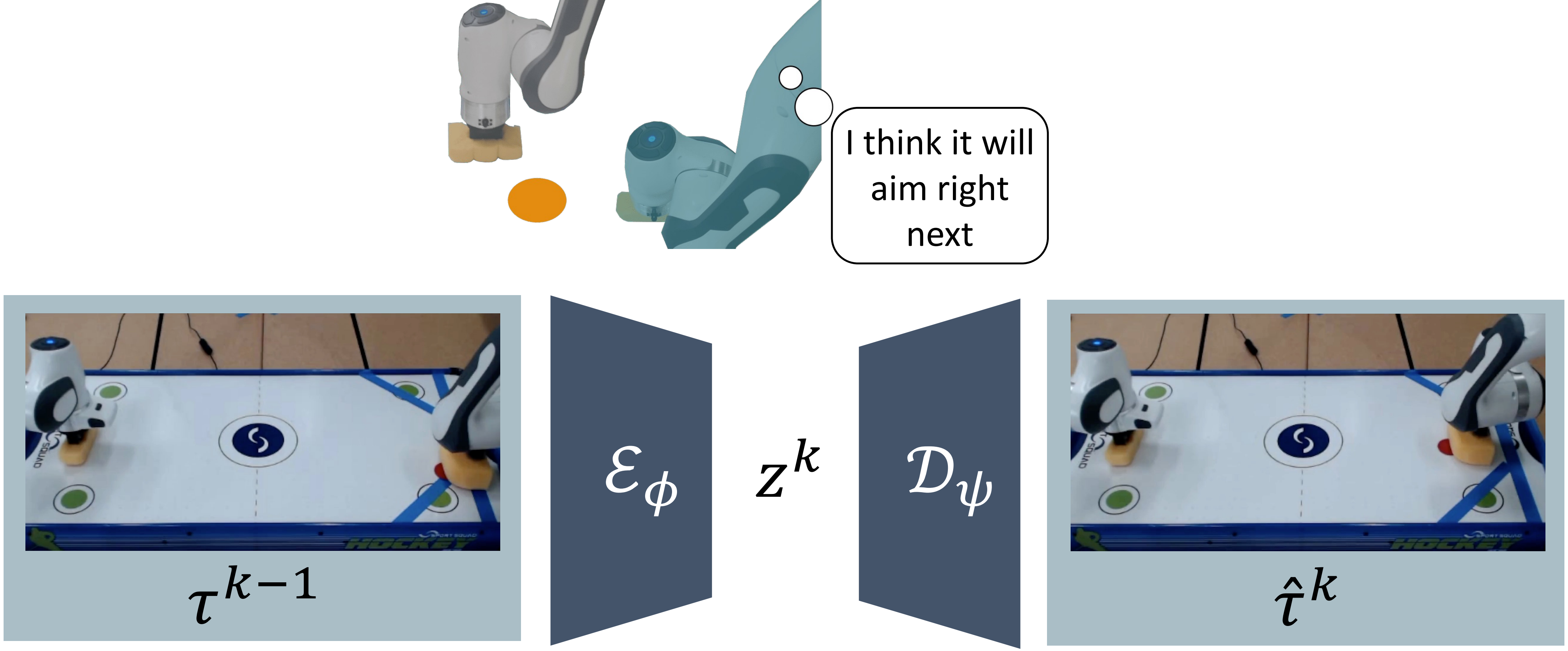
Modeling Other Agent's Behavior



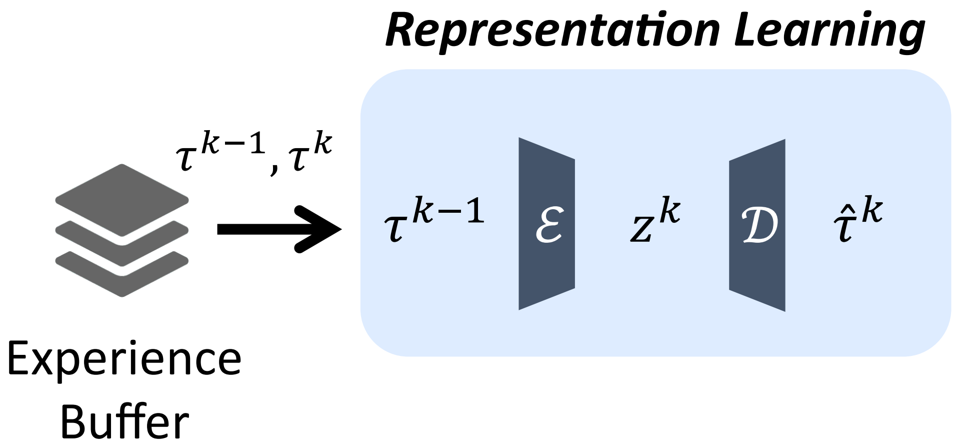


I think it will aim right next
τ^{k-1}
$\mathcal{E}_\phi$
z^k


I think it will aim right next
$\mathcal{E}_\phi$
z^k
$\mathcal{D}_\psi$
τ^{k-1}
$\hat{\tau}^k$



Learning objective: $$\max_{\phi,\psi} \sum_{i=2}^{N} \sum_{t=1}^{H} \log p_{\phi,\psi}(s_{t+1}^i, r_t^i \mid s_t^i, a_t^i, \tau^{i-1})$$



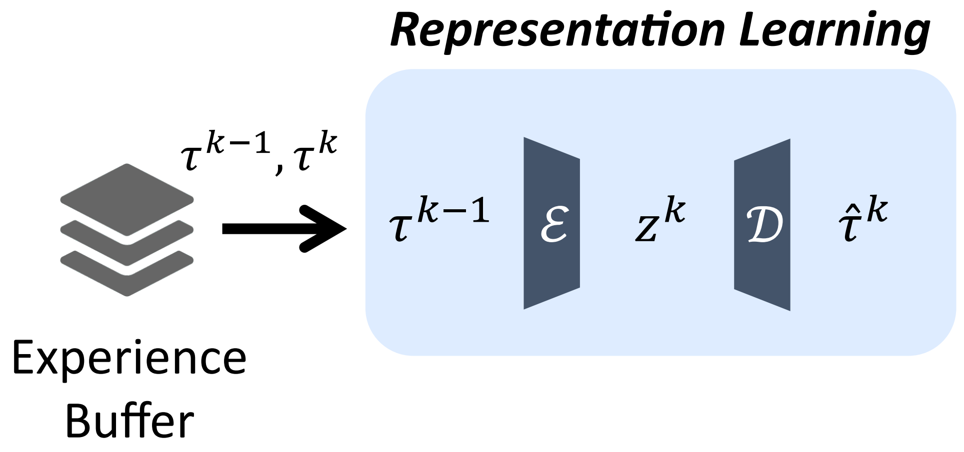
Representation Learning
τ^{k-1}, τ^k
Experience Buffer
τ^{k-1}
$\mathcal{E}$
z^k
$\mathcal{D}$
$\hat{\tau}^k$

Learning and Influencing Latent Intent

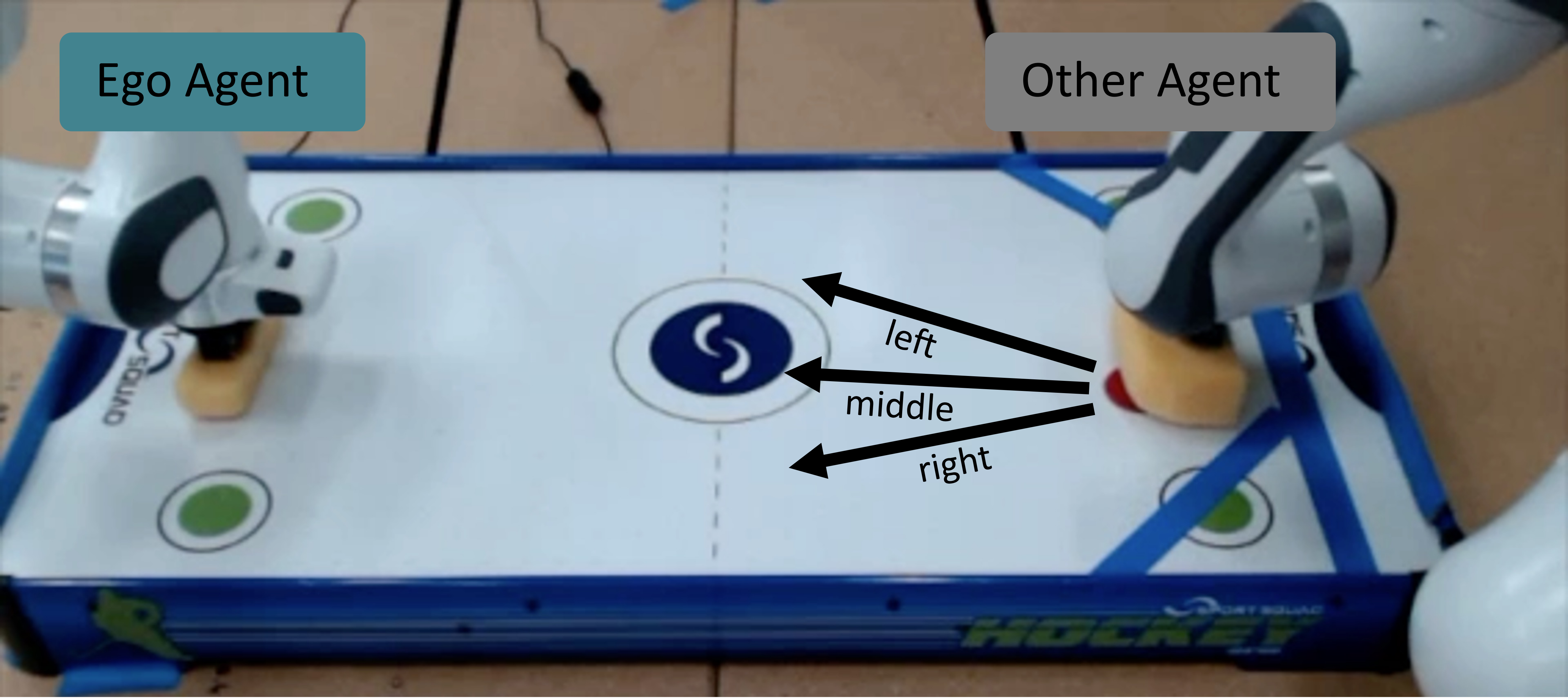
Maximize expected return
within an interaction

$$\max_{\theta} \quad \mathbb{E}_{\pi_\theta(a|s,z^i)}\left[\sum_{t=1}^{H} R(s, z^i)\right]$$

to *react* to the other agent

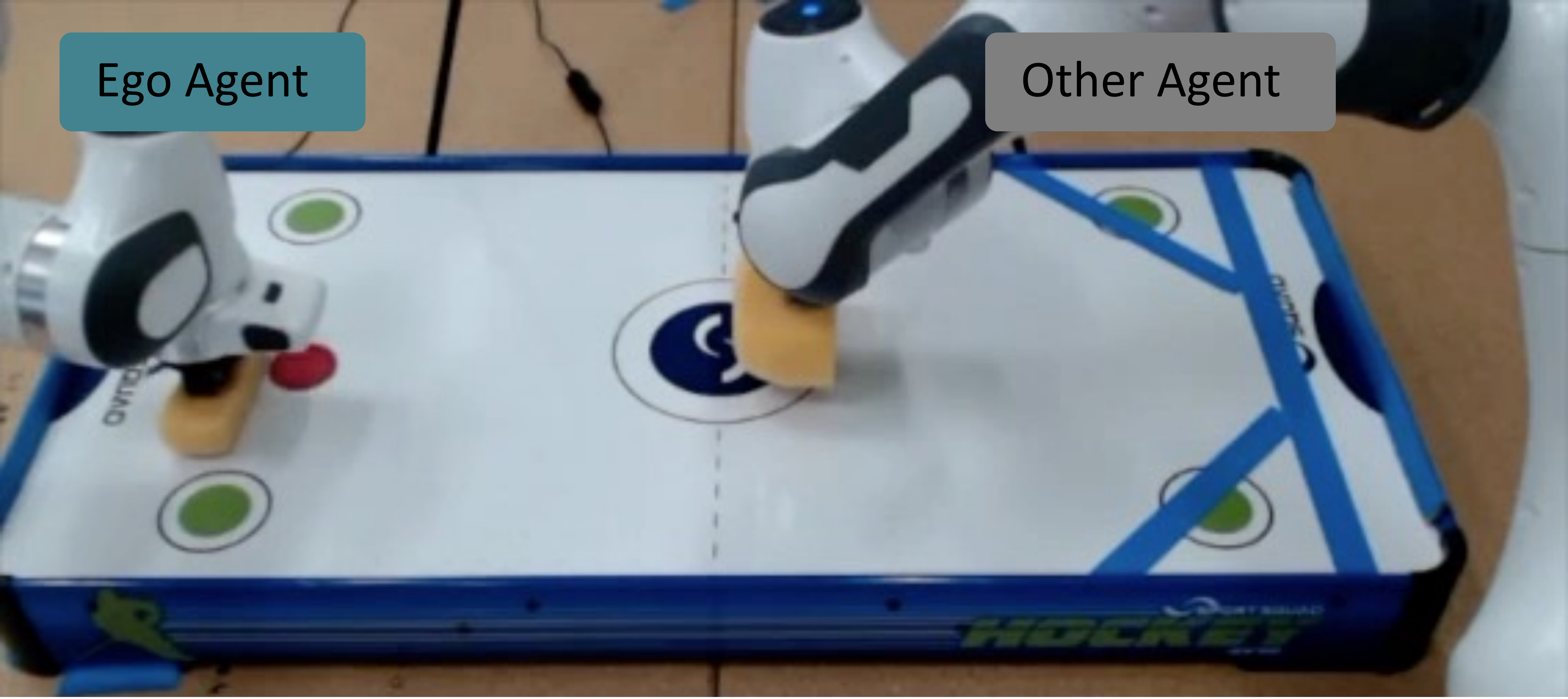


[*Xie, Losey, Tolsma, Finn, Sadigh*, CoRL 2020]



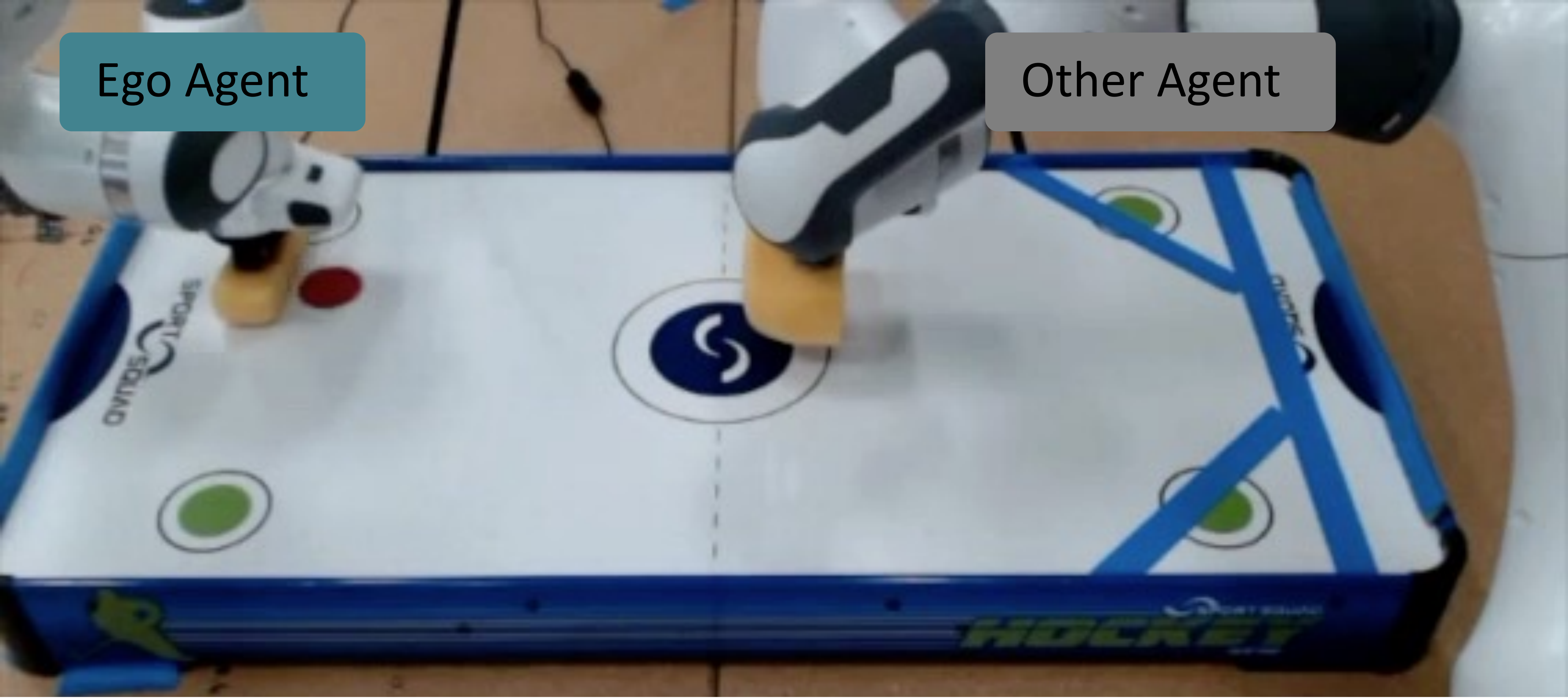
Air Hockey Results

Air Hockey Results



Air Hockey Results

Air Hockey Results



Air Hockey Results

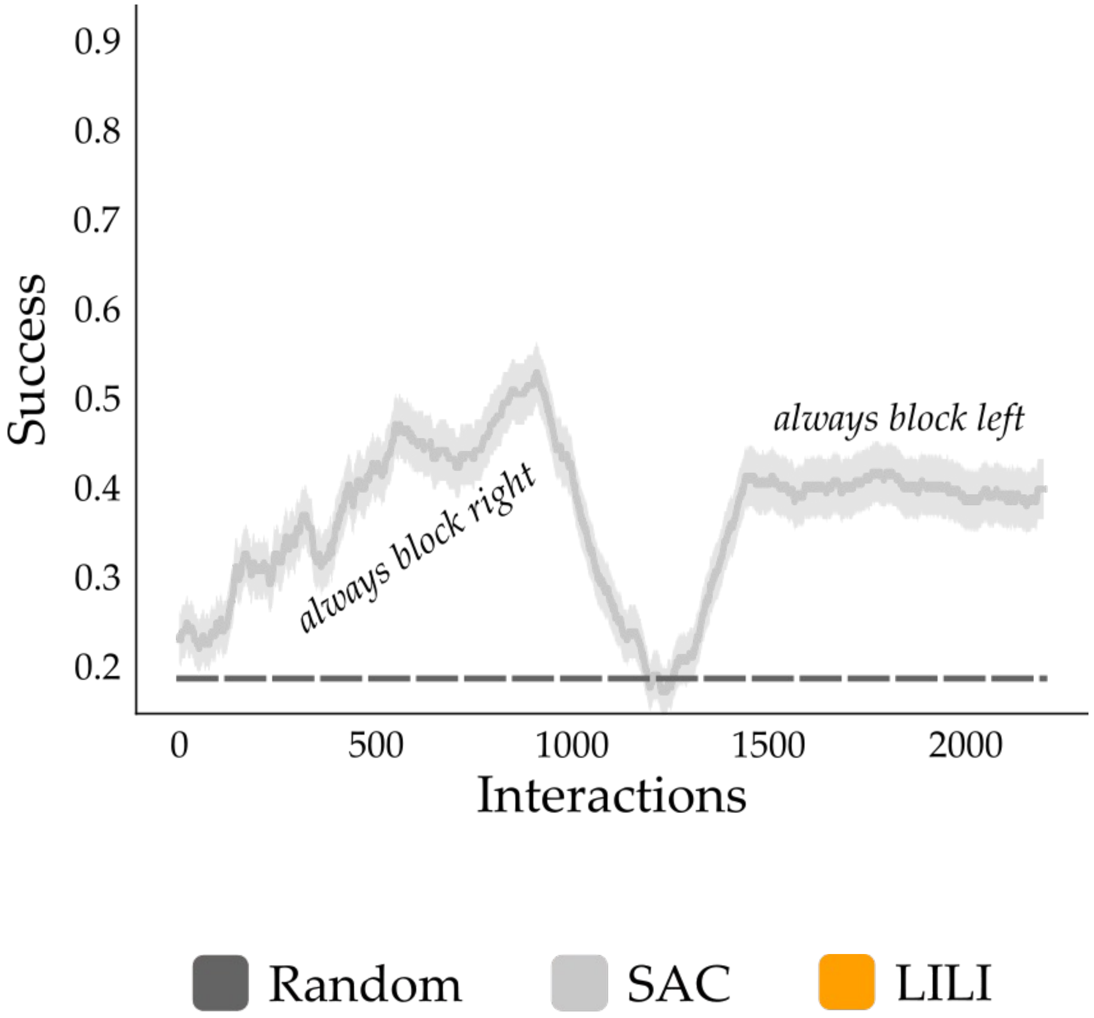

2x speed
SAC: initial policy

2x speed

SAC: 2 hours of training

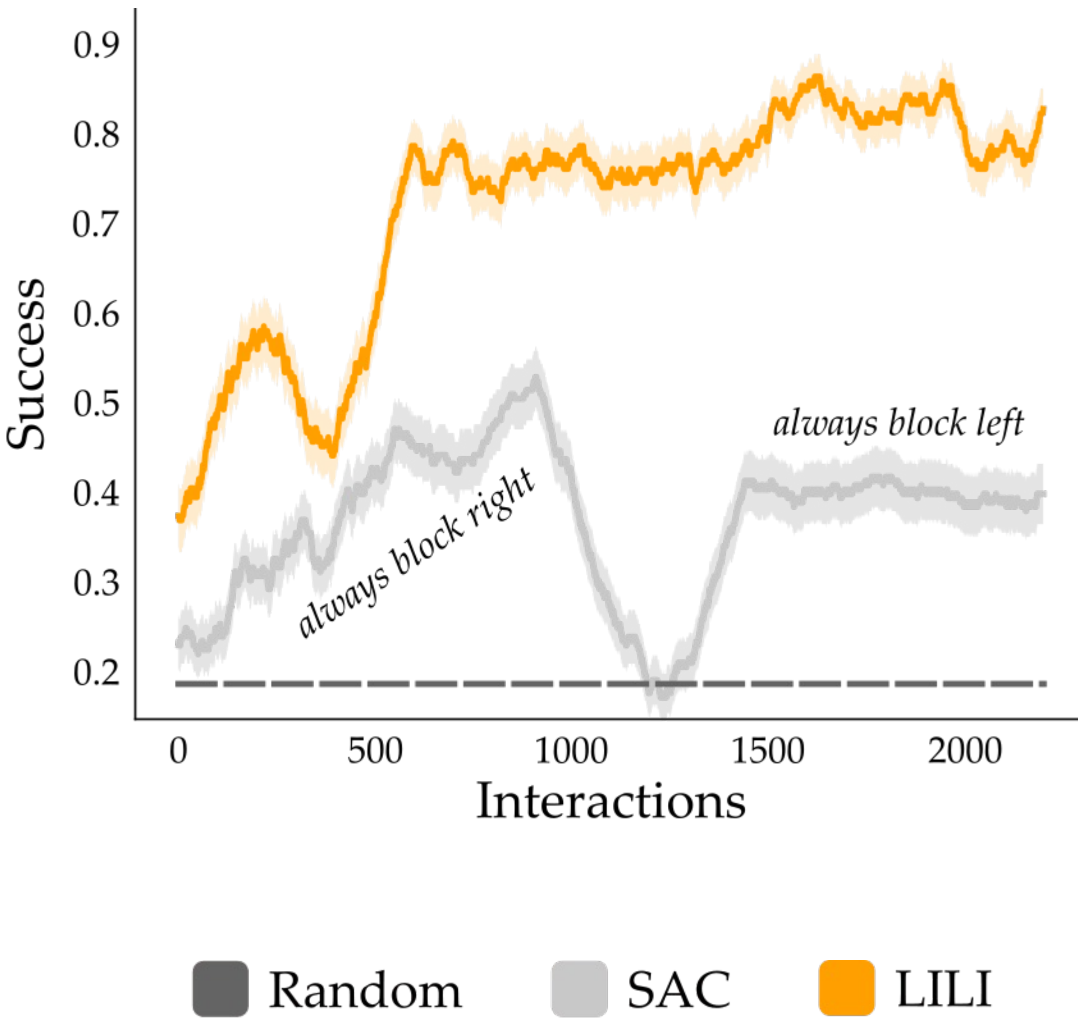

2x speed
SAC: 4 hours of training
HOCKEY

Air Hockey Results




2x speed
LILI: 4 hours of training

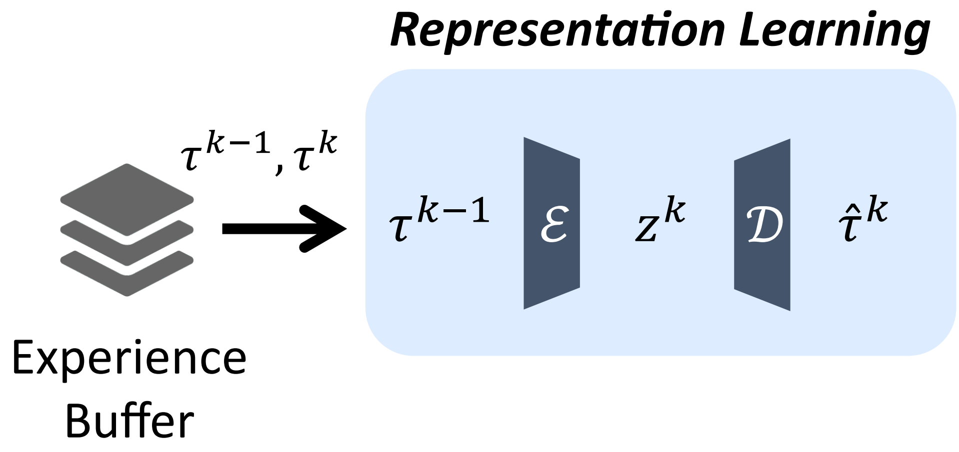
Air Hockey Results



Reacting to Other Agents

Maximize expected return
within an interaction

$$\max_{\theta} \quad \mathbb{E}_{\pi_\theta(a|s,z^i)}\left[\sum_{t=1}^{H} R(s, z^i)\right]$$

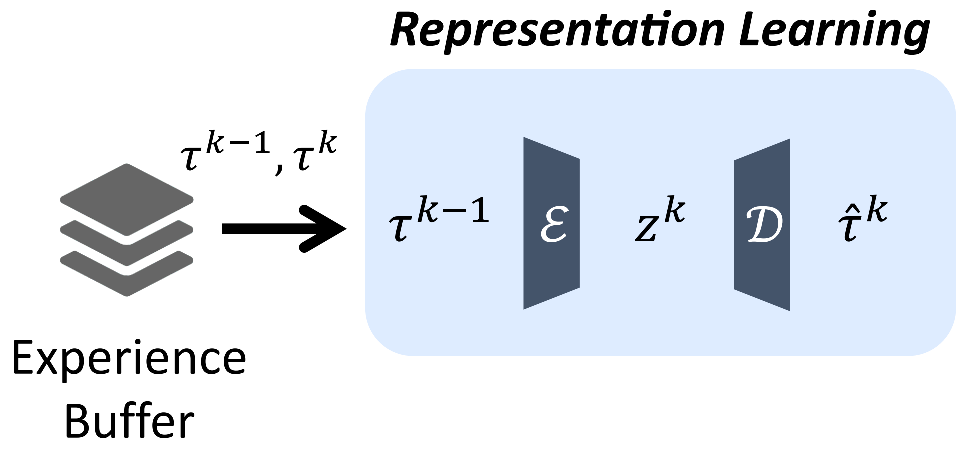


to *react to* the other agent

Influencing Other Agents

Maximize expected return *across* interactions

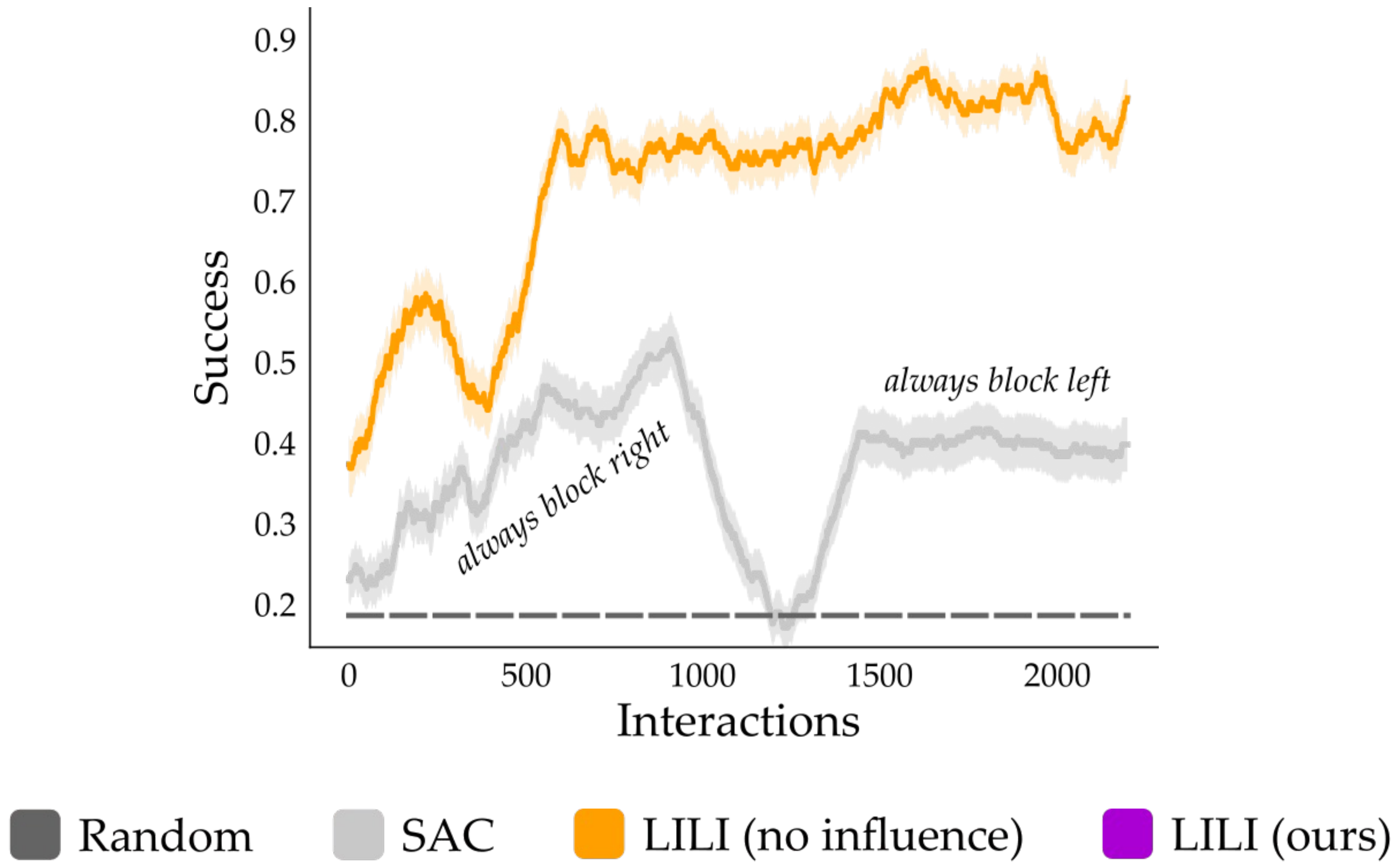
$$\max_{\theta} \sum_{i=1}^{\infty} \gamma^{i} \mathbb{E}_{\pi_\theta(a|s, z^i)} \left[\sum_{t=1}^{H} R(s, z^i) \right]$$



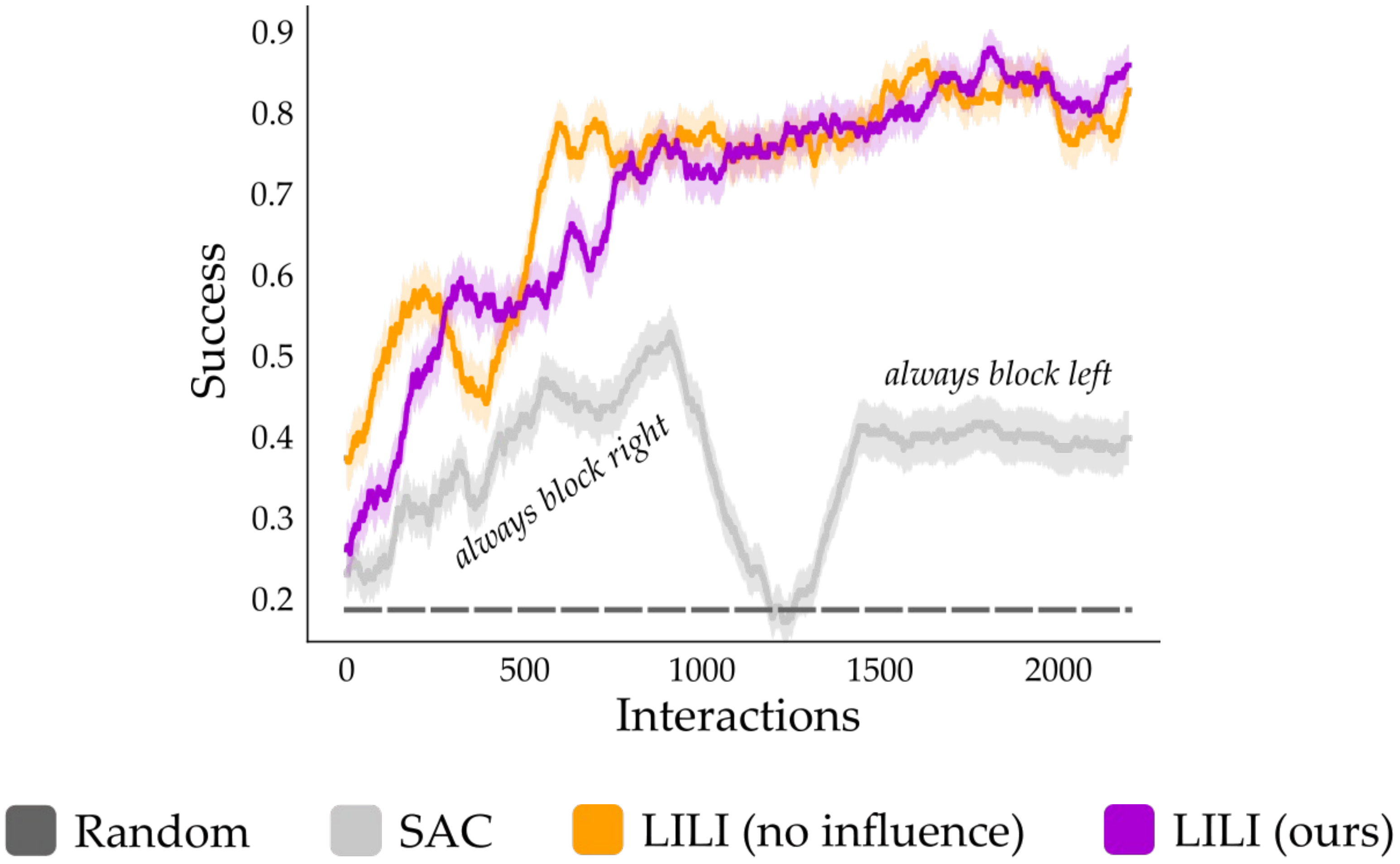
to *influence* the other agent


2x speed
LILI (with influence): 4 hours of training

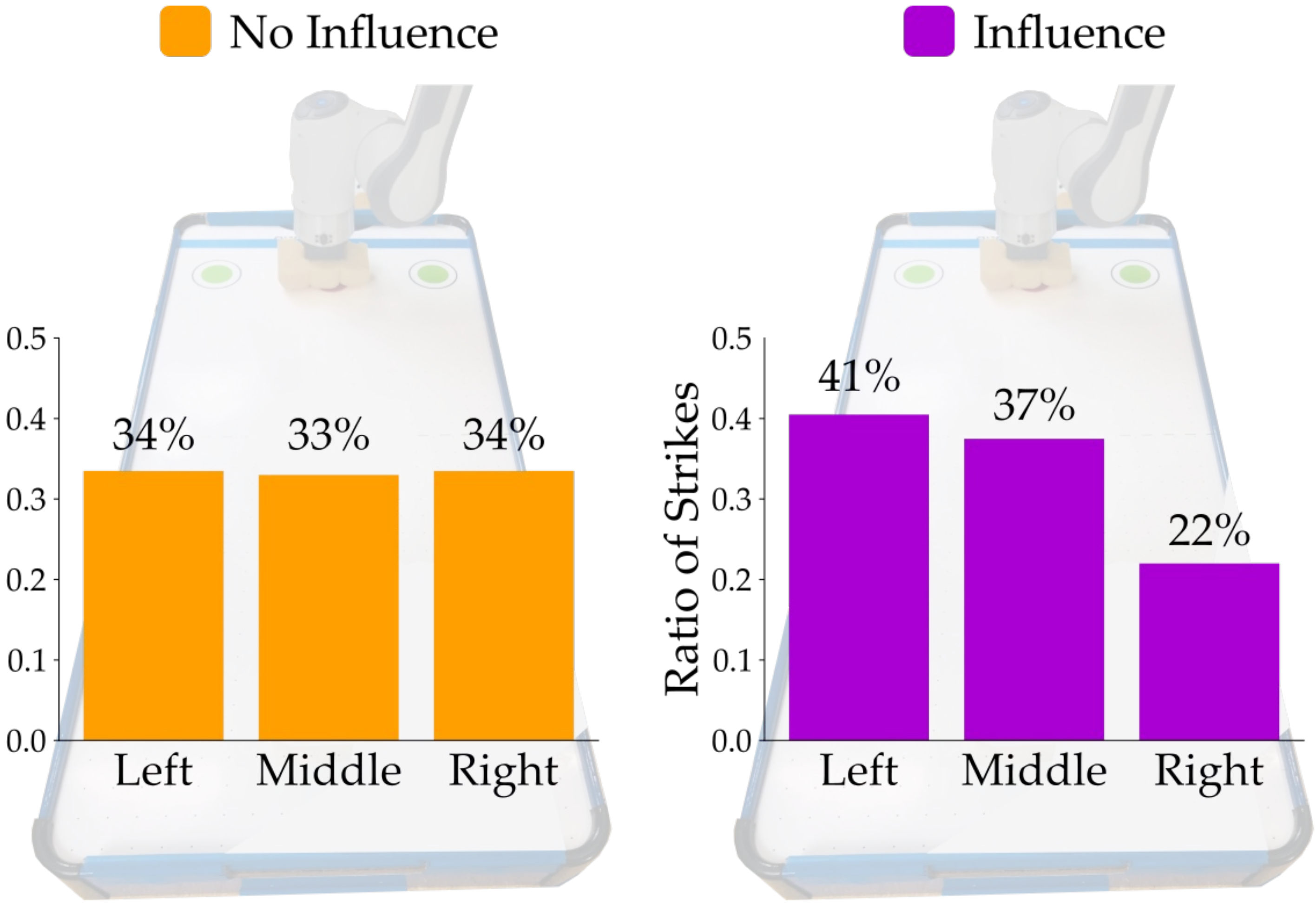
Air Hockey Results



Air Hockey Results



Air Hockey Results



[*Xie, Losey, Tolsma, Finn, Sadigh,* CoRL 2020]

Playing with a Human Expert

Playing with a Human Expert

Key Takeaways

Human partners are often non-stationary –
which can be represented by low-dimensional latent strategies.

LILI *anticipates* the partner's policies using latent strategies to *react* and *influence* the other agent.

Today’s itinerary

- Game-Theoretic Views on Multi-Agent Interactions
- Partner Modeling: Active Info Gathering over Human’s Intent
- Partner Modeling: Learning and Influencing Latent Intent
- Partner Modeling: Role Assignment