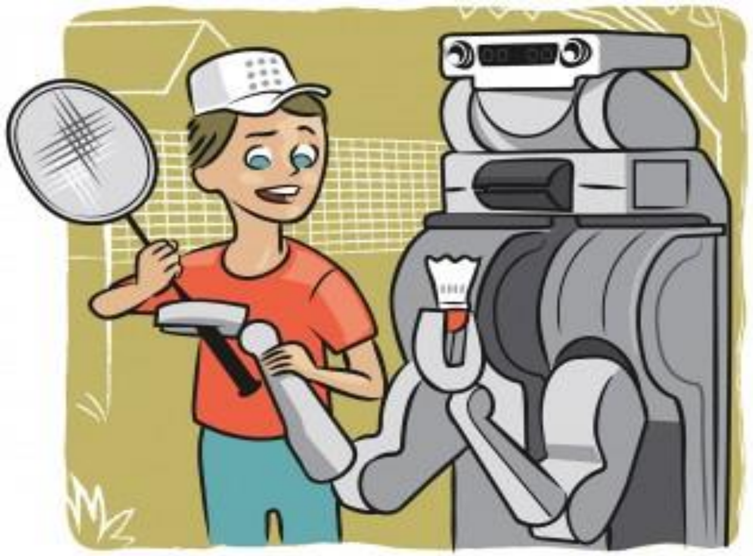
Interactive Imitation Learning

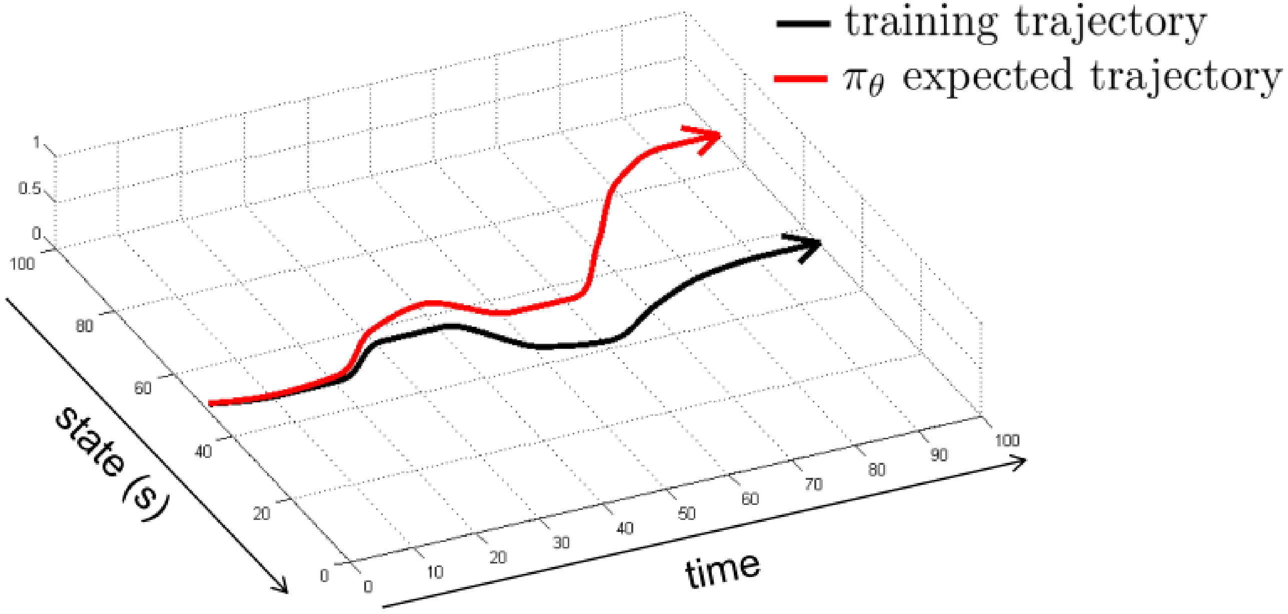


Instructor: Daniel Brown

[Some slides adapted from Sergey Levine (CS 285) and Alina Vereshchaka (CSE4/510)]

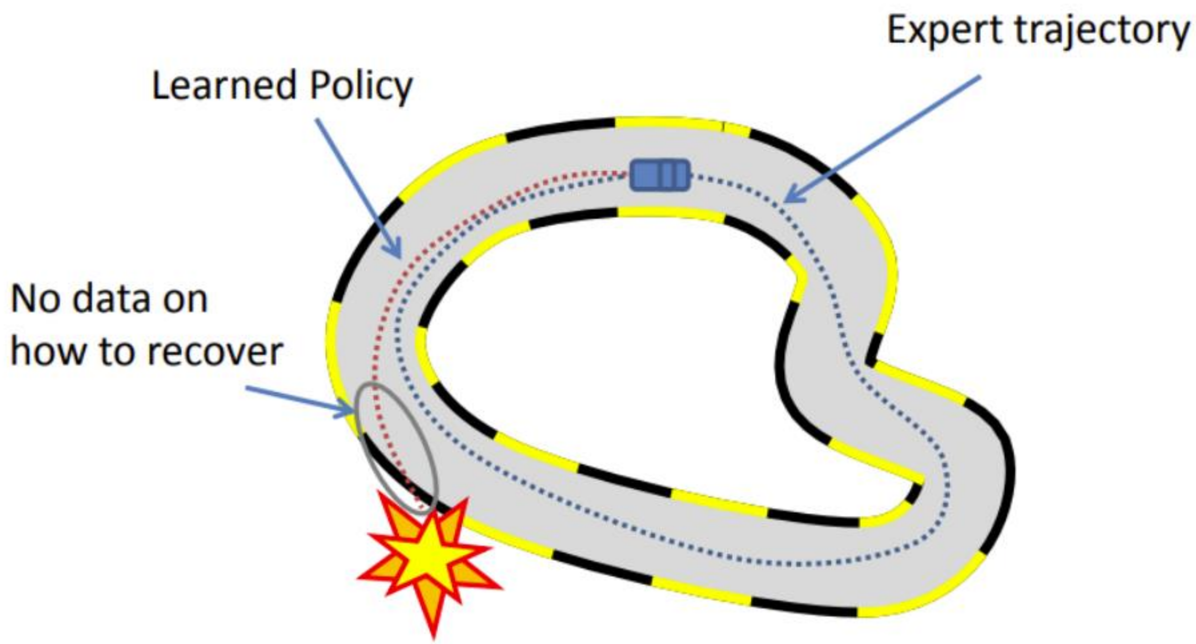
What could go wrong?

$\mathbf{o}_t$ $\mathbf{a}_t$ → training data → supervised learning → $\pi_\theta(\mathbf{a}_t|\mathbf{o}_t)$



Distribution Shift

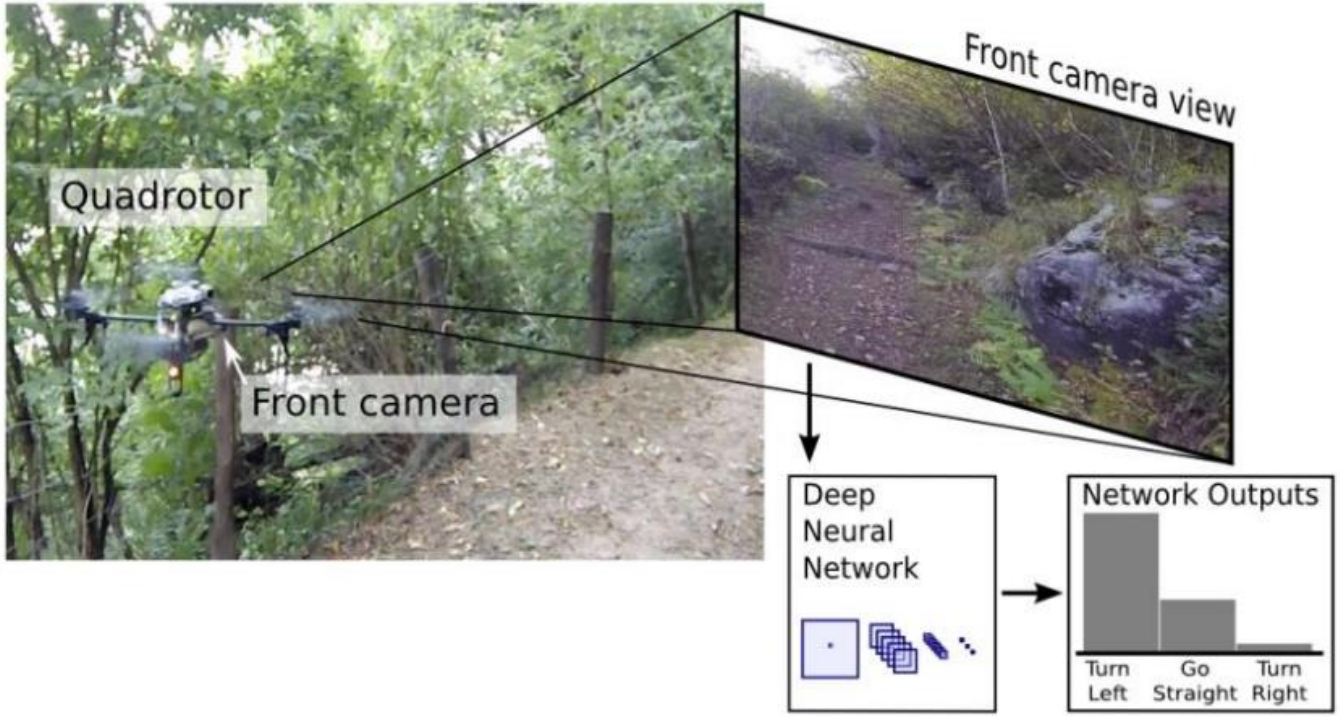
$$p_{\pi^*}(o_t) \neq p_{\pi_\theta}(o_t)$$



	Supervised Learning	Supervised Learning + Control
Train	$(x, y) \sim D$	$s \sim P(\cdot\,\|s, \pi^*(s))$
Test	$(x, y) \sim D$	$s \sim P(\cdot\,\|s, \pi(s))$

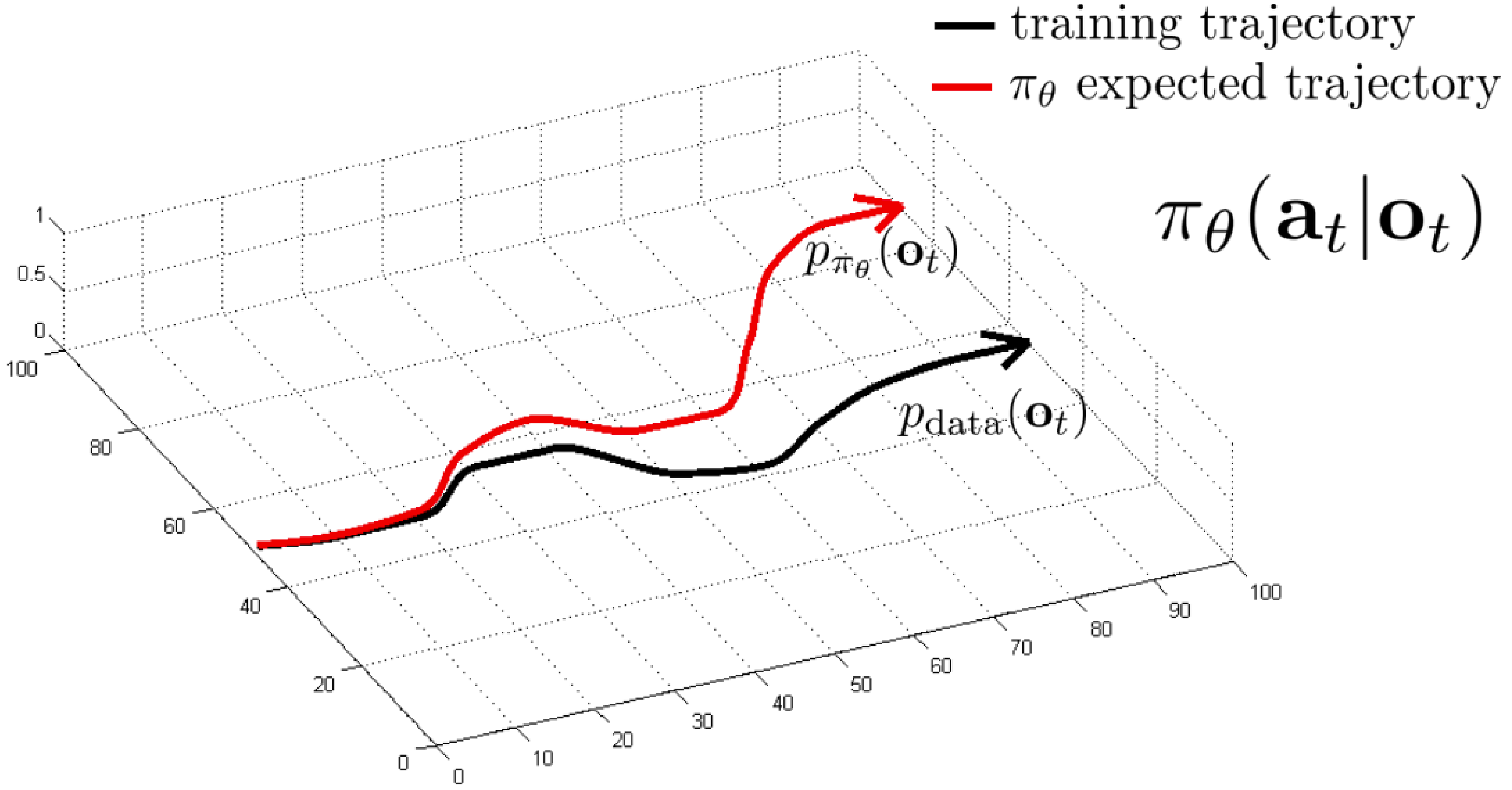
A Machine Learning Approach to Visual Perception of Forest Trails for Mobile Robots

Alessandro Giusti[1], Jérôme Guzzi[1], Dan C. Cireşan[1], Fang-Lin He[1], Juan P. Rodríguez[1]
Flavio Fontana[2], Matthias Faessler[2], Christian Forster[2]
Jürgen Schmidhuber[1], Gianni Di Caro[1], Davide Scaramuzza[2], Luca M. Gambardella[1]



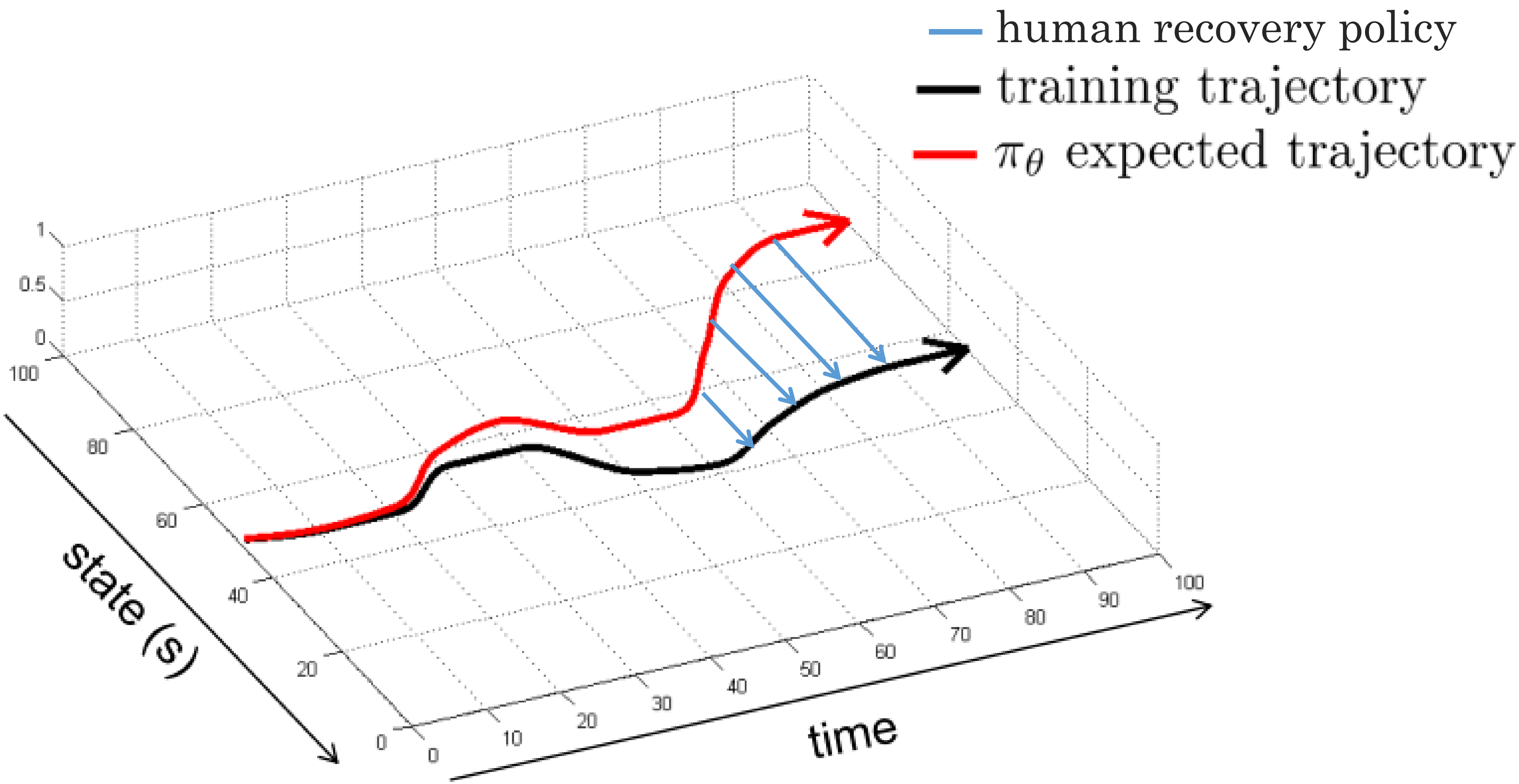

Deep Neural Network
Control Signal

Can we make it work more often?



$$\pi_\theta(\mathbf{a}_t|\mathbf{o}_t)$$

can we make $p_{\text{data}}(\mathbf{o}_t) = p_{\pi_\theta}(\mathbf{o}_t)$?


human recovery policy
training trajectory
π_θ expected trajectory
1
0.5
0
100
80
60
40
20
0
state (s)
0
10
20
30
40
50
60
70
80
90
100
time

DAgger

can we make $p_{\text{data}}(\mathbf{o}_t) = p_{\pi_\theta}(\mathbf{o}_t)$?

idea: instead of being clever about $p_{\pi_\theta}(\mathbf{o}_t)$, be clever about $p_{\text{data}}(\mathbf{o}_t)$!

DAgger: Dataset **Agg**regation

goal: collect training data from $p_{\pi_\theta}(\mathbf{o}_t)$ instead of $p_{\text{data}}(\mathbf{o}_t)$

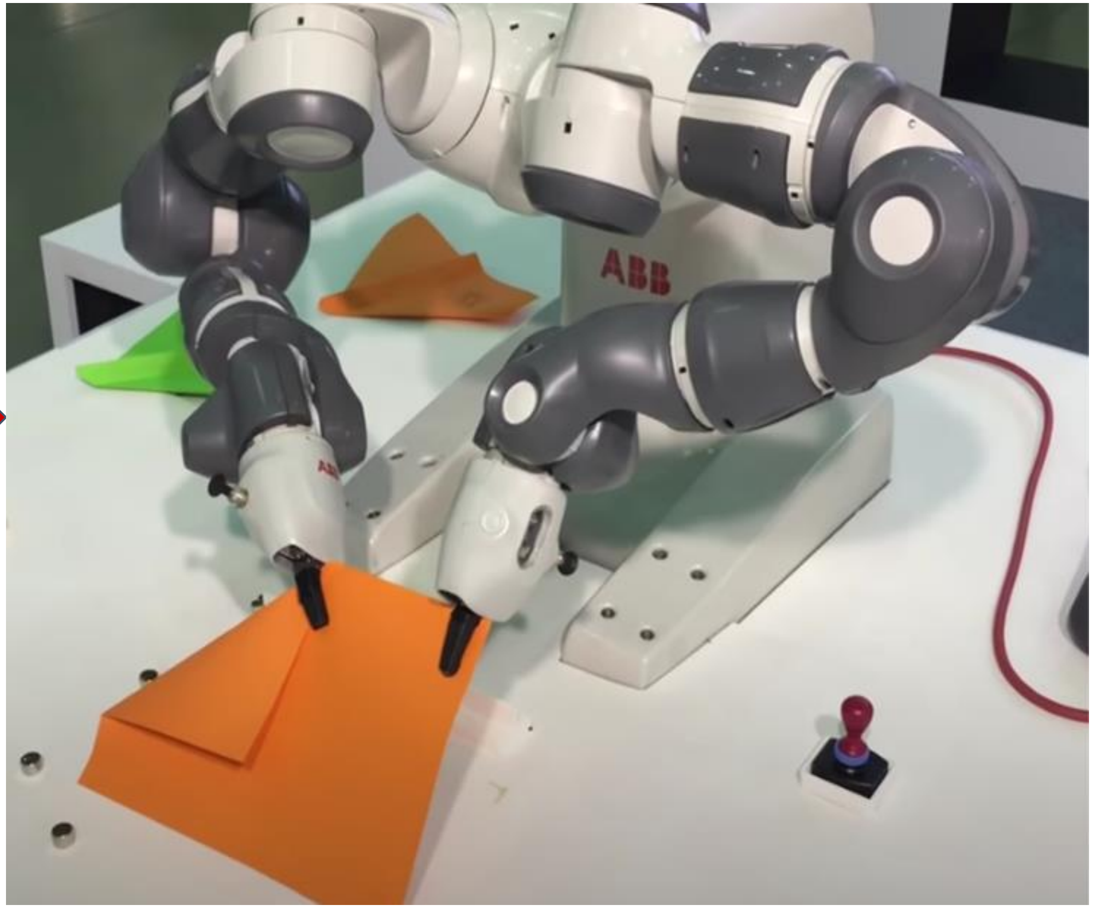
how? just run $\pi_\theta(\mathbf{a}_t|\mathbf{o}_t)$

but need labels $\mathbf{a}_t$!

1. train $\pi_\theta(\mathbf{a}_t|\mathbf{o}_t)$ from human data $\mathcal{D} = \{\mathbf{o}_1, \mathbf{a}_1, \dots, \mathbf{o}_N, \mathbf{a}_N\}$
2. run $\pi_\theta(\mathbf{a}_t|\mathbf{o}_t)$ to get dataset $\mathcal{D}_\pi = \{\mathbf{o}_1, \dots, \mathbf{o}_M\}$
3. Ask human to label $\mathcal{D}_\pi$ with actions $\mathbf{a}_t$
4. Aggregate: $\mathcal{D} \leftarrow \mathcal{D} \cup \mathcal{D}_\pi$

Ross et al. '11

Interactive IL



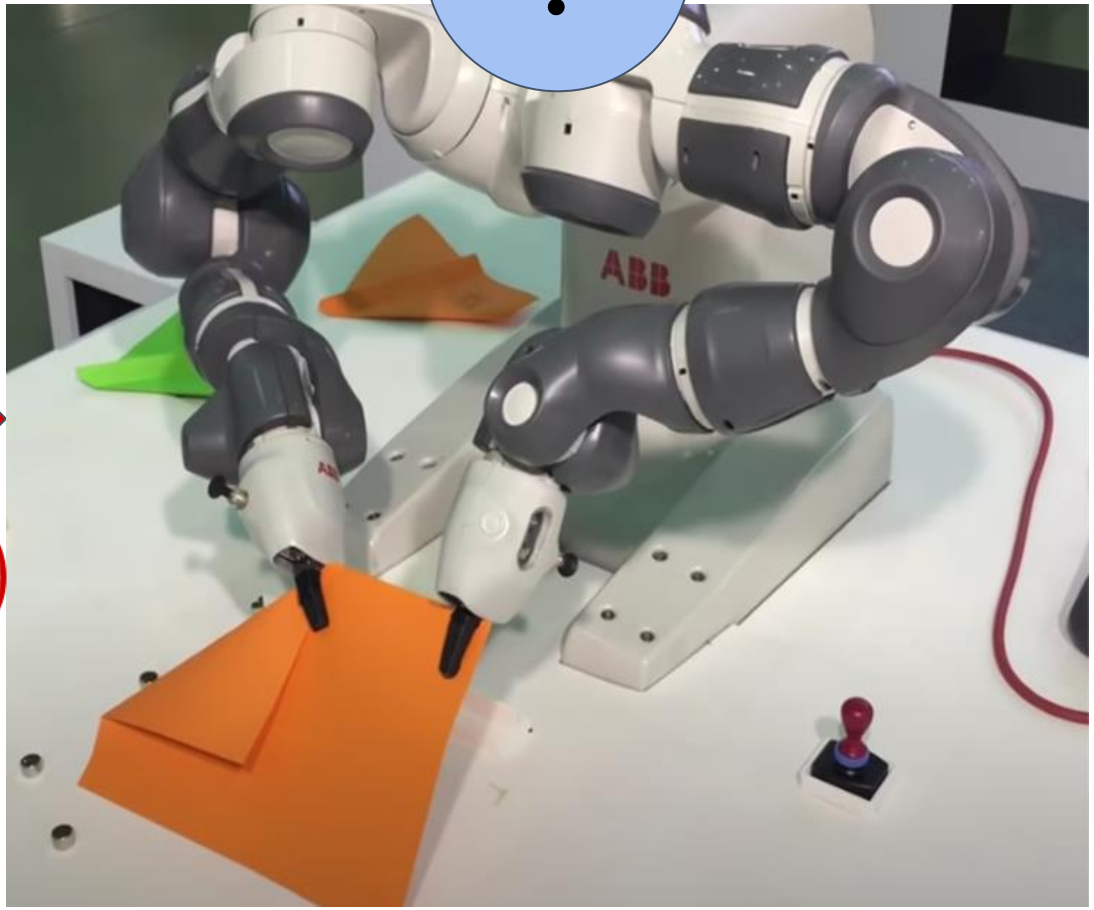
Interactive IL

$\pi_H(s)$

$\pi_{\text{meta}}(s)$

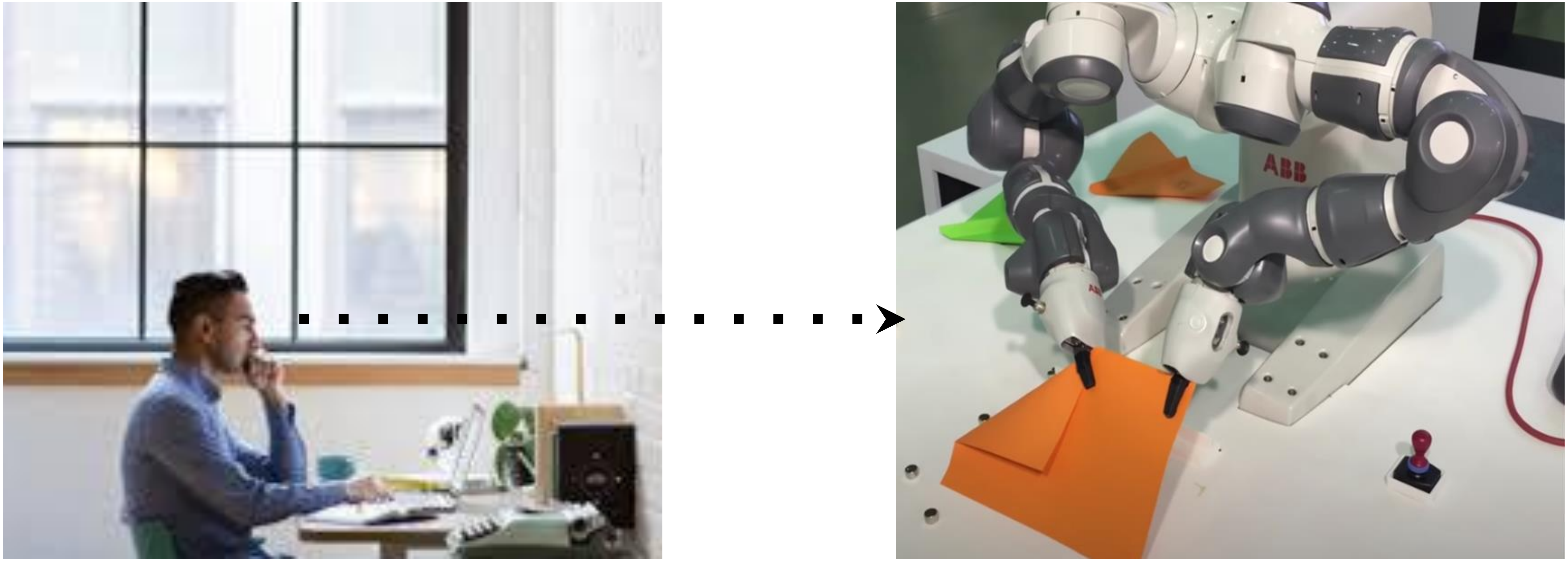
???

?



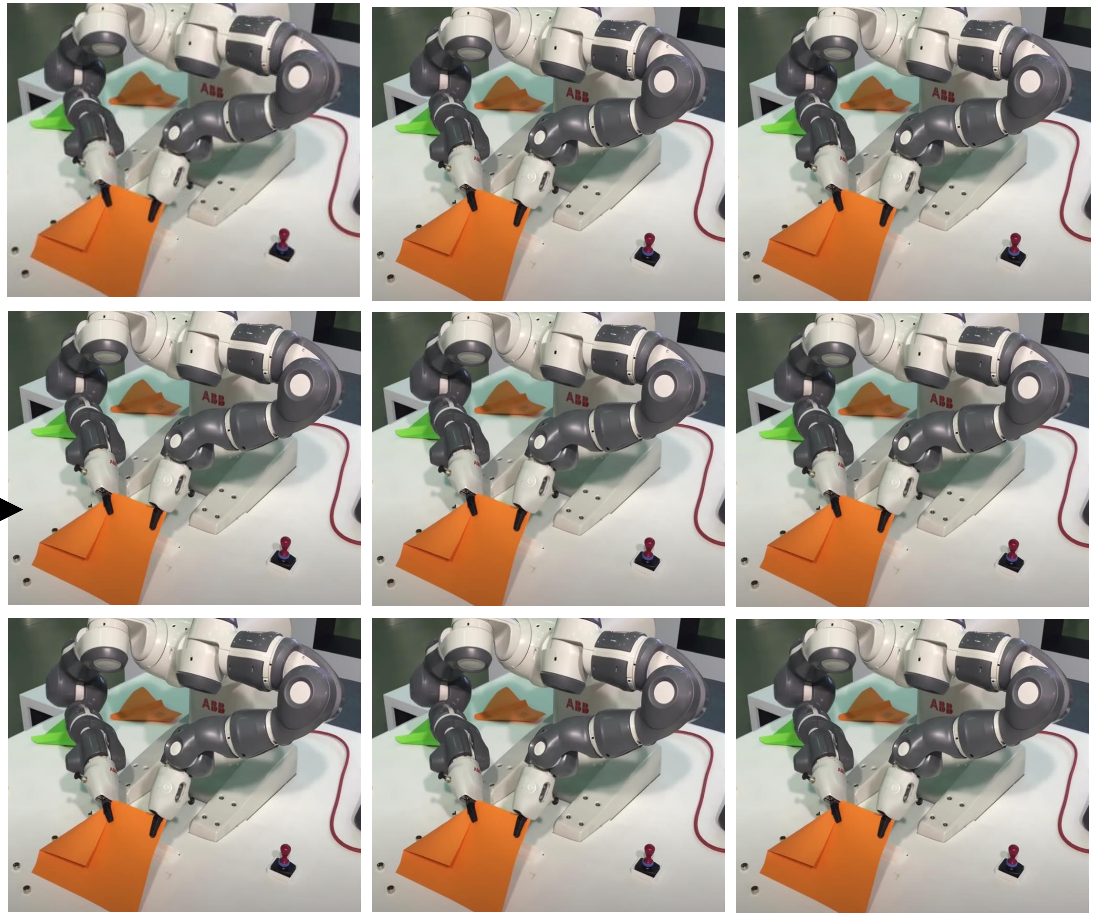
$\pi_R(s)$

Human-Gated Interactive IL



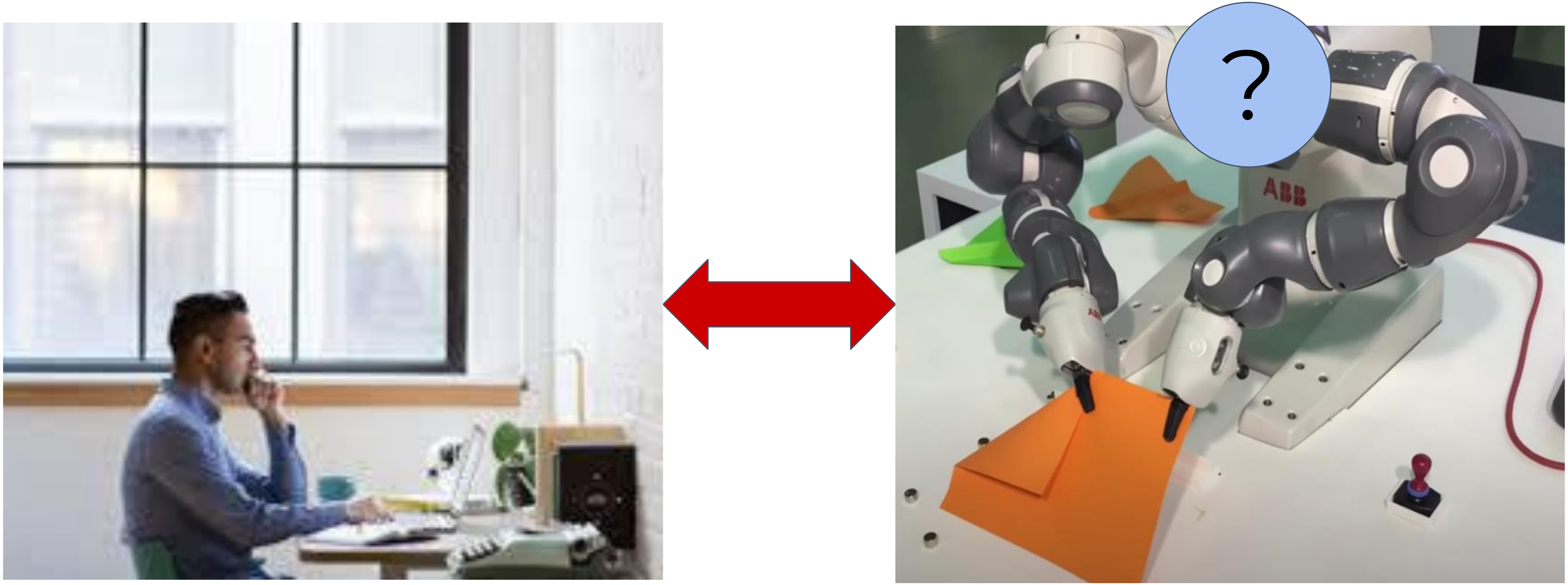
[3] M. Kelly, C. Sidrane, K. Driggs-Campbell, and M. J. Kochenderfer. HG-DAgger: Interactive Imitation Learning with Human Experts. ICRA 2019.

Human-Gated Interactive IL



[3] M. Kelly, C. Sidrane, K. Driggs-Campbell, and M. J. Kochenderfer. HG-DAgger: Interactive Imitation Learning with Human Experts. ICRA 2019.

Robot-Gated Interactive IL



[4] J. Zhang, K. Cho. Query-Efficient Imitation Learning for End-to-End Autonomous Driving. AAAI 2017.
[5] K. Menda, K. Driggs-Campbell, M. Kochenderfer. EnsembleDAgger: A Bayesian Approach to Safe Imitation Learning. IROS 2019.

When should a robot ask for help?

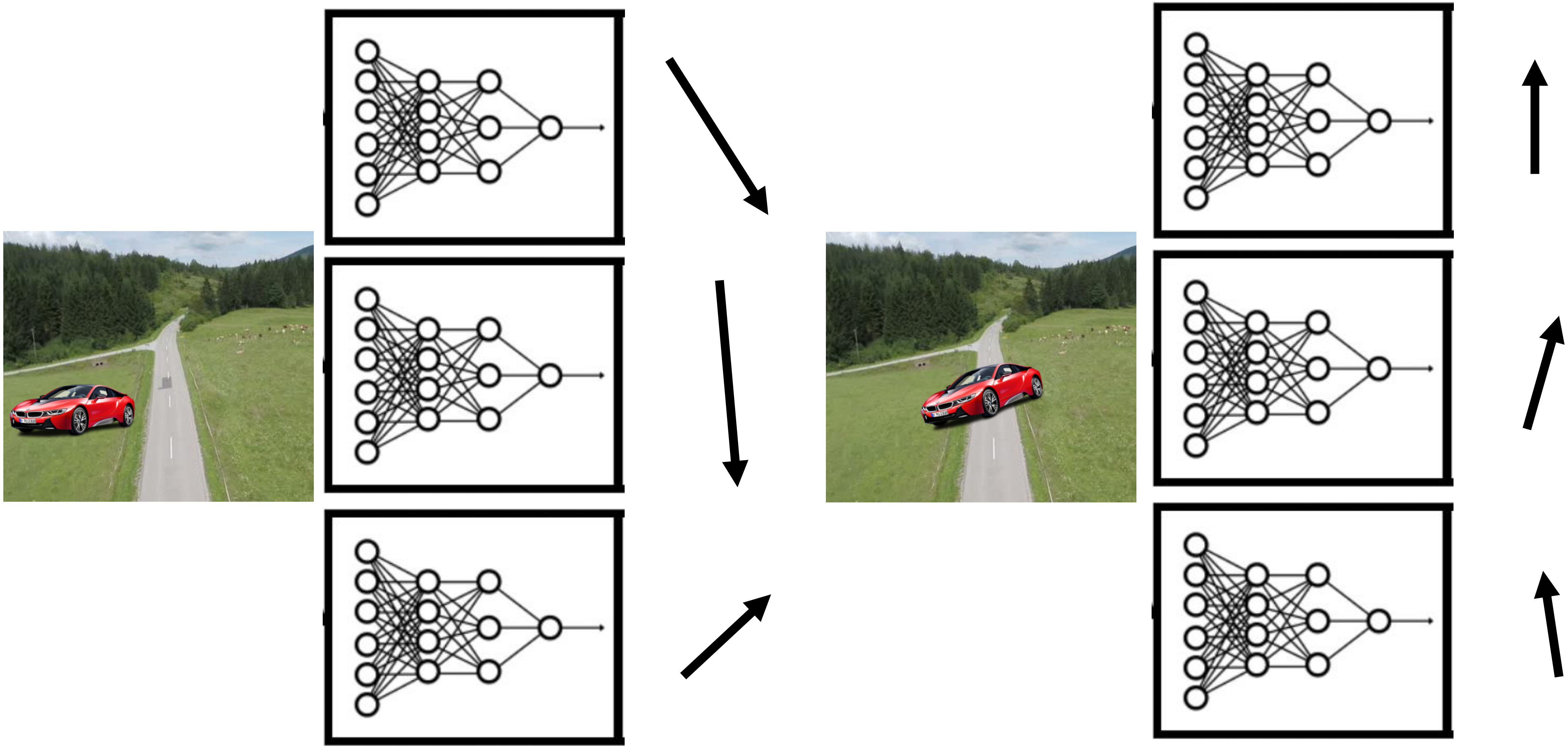
Novel (and risky)

When should a robot ask for help?

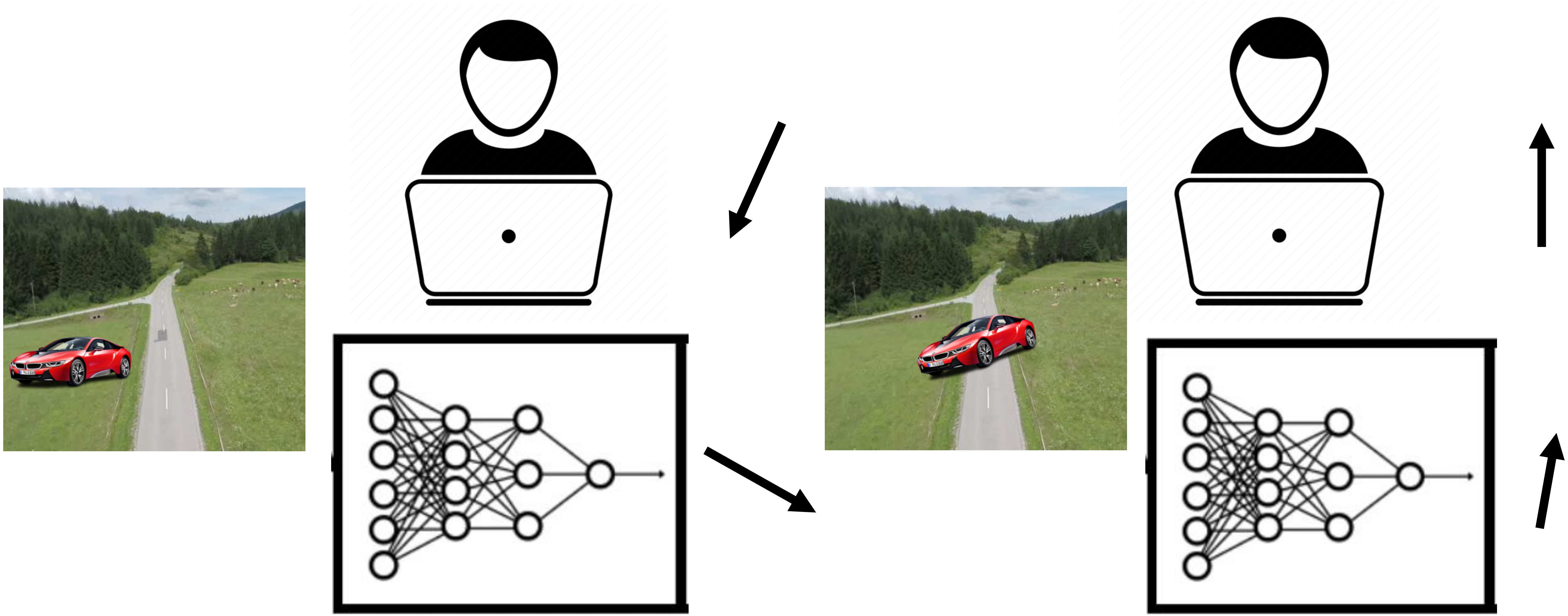
Novel (and risky)

Risky (but not novel)

Novelty Estimation

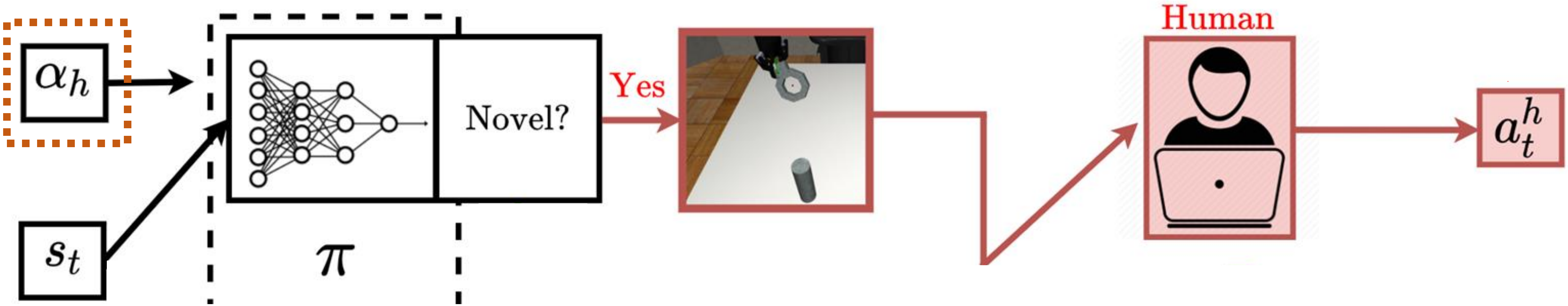


Novelty Estimation: Supervisor Mode



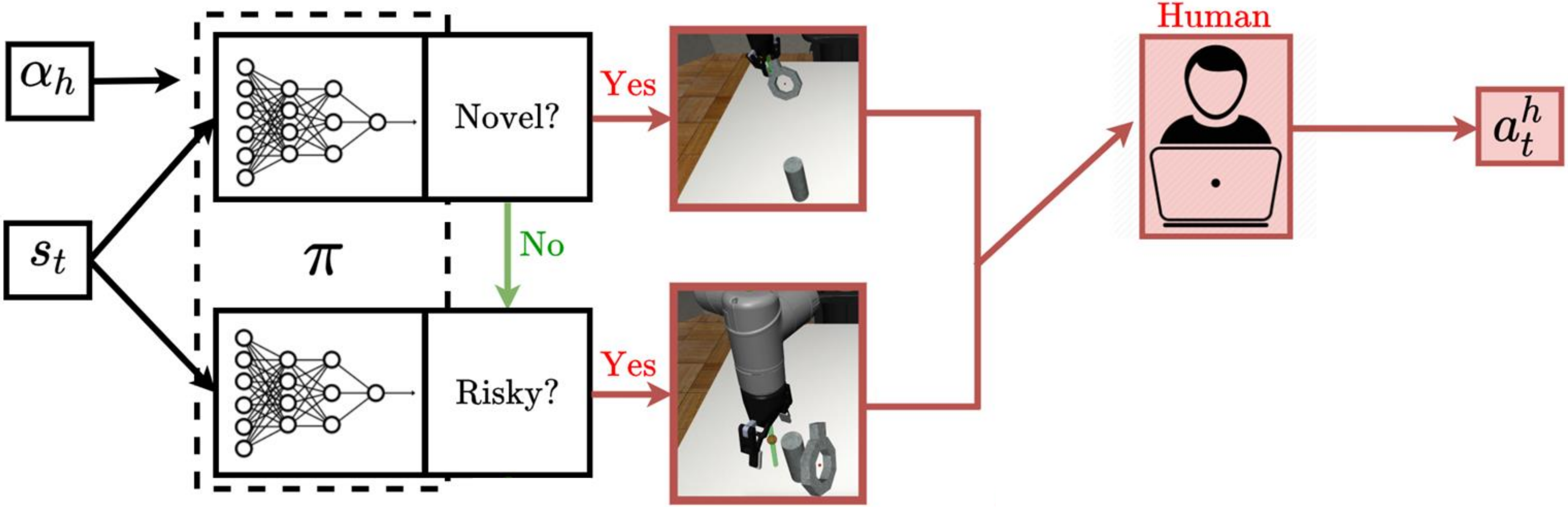
ThriftyDAgger

Target percent of time human wants to give interventions.



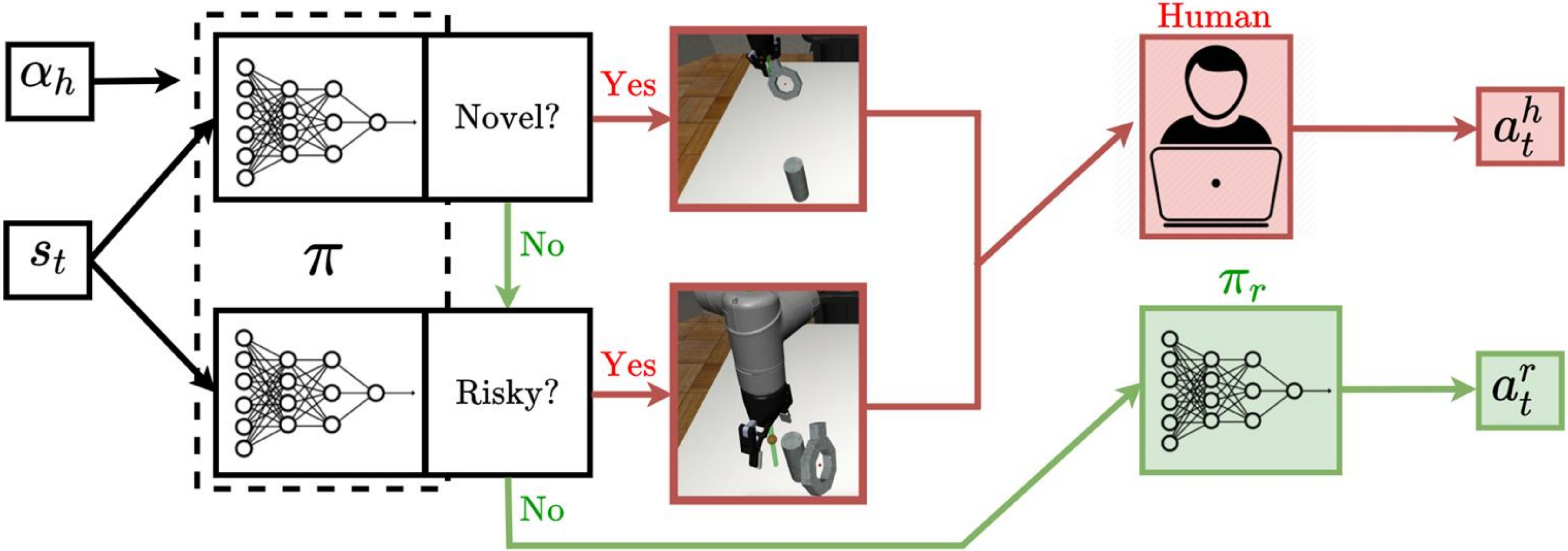
Hoque et al. "ThriftyDAgger: Budget-Aware Novelty and Risk Gating for Interactive Imitation Learning." CoRL 2021.

ThriftyDAgger

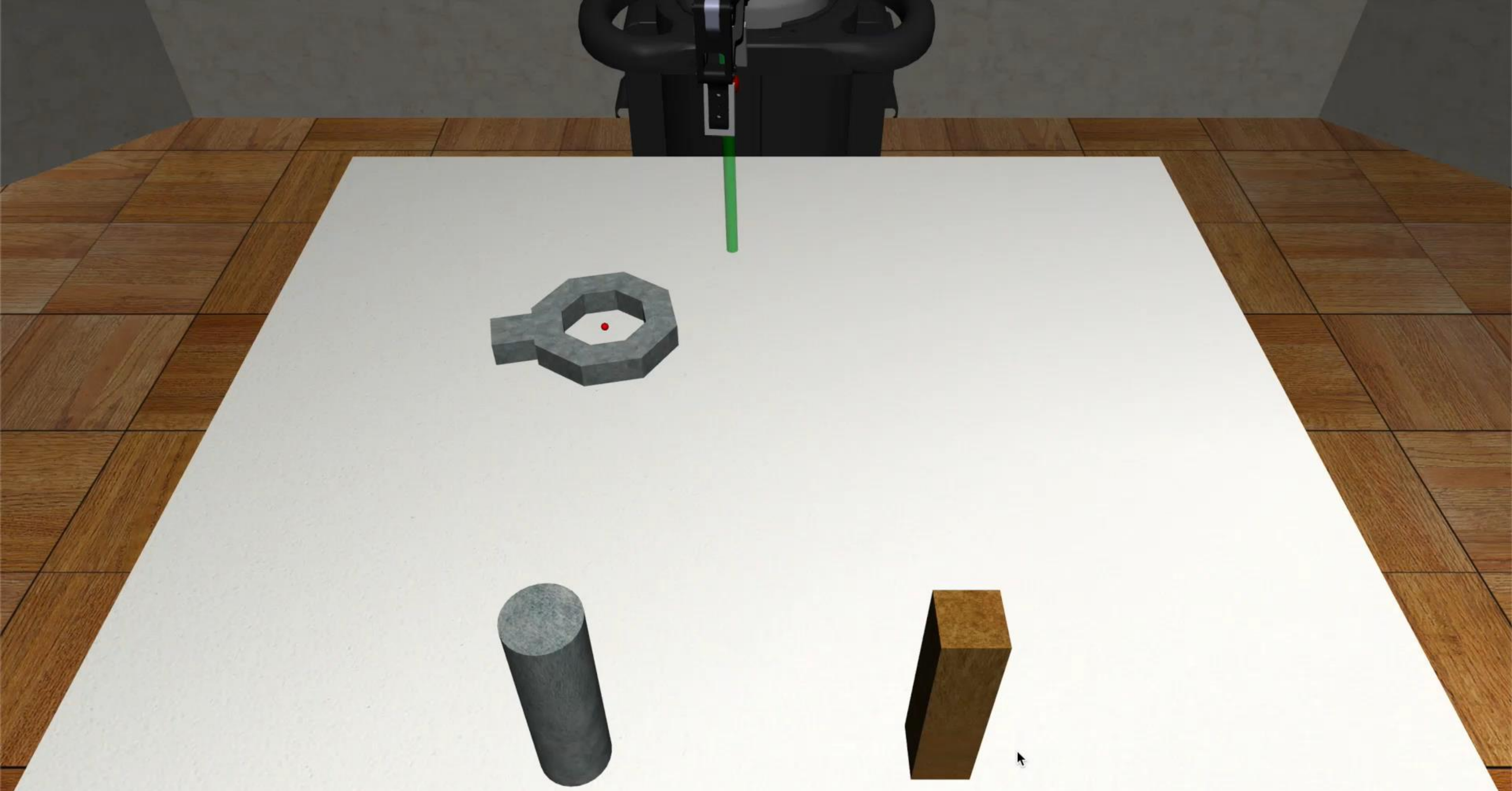


Hoque et al. "ThriftyDAgger: Budget-Aware Novelty and Risk Gating for Interactive Imitation Learning." CoRL 2021.

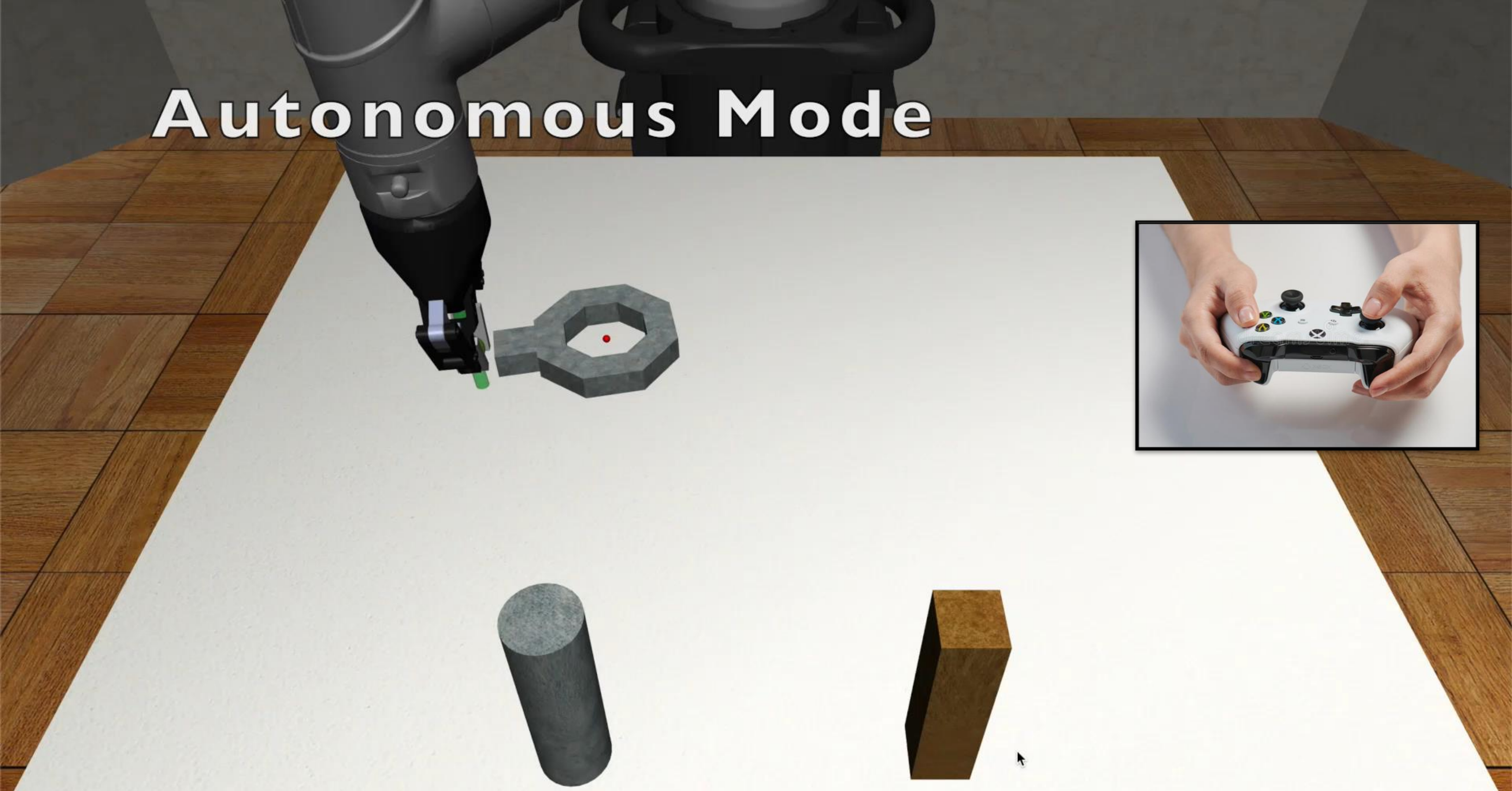
ThriftyDAgger



Hoque et al. "ThriftyDAgger: Budget-Aware Novelty and Risk Gating for Interactive Imitation Learning." CoRL 2021.

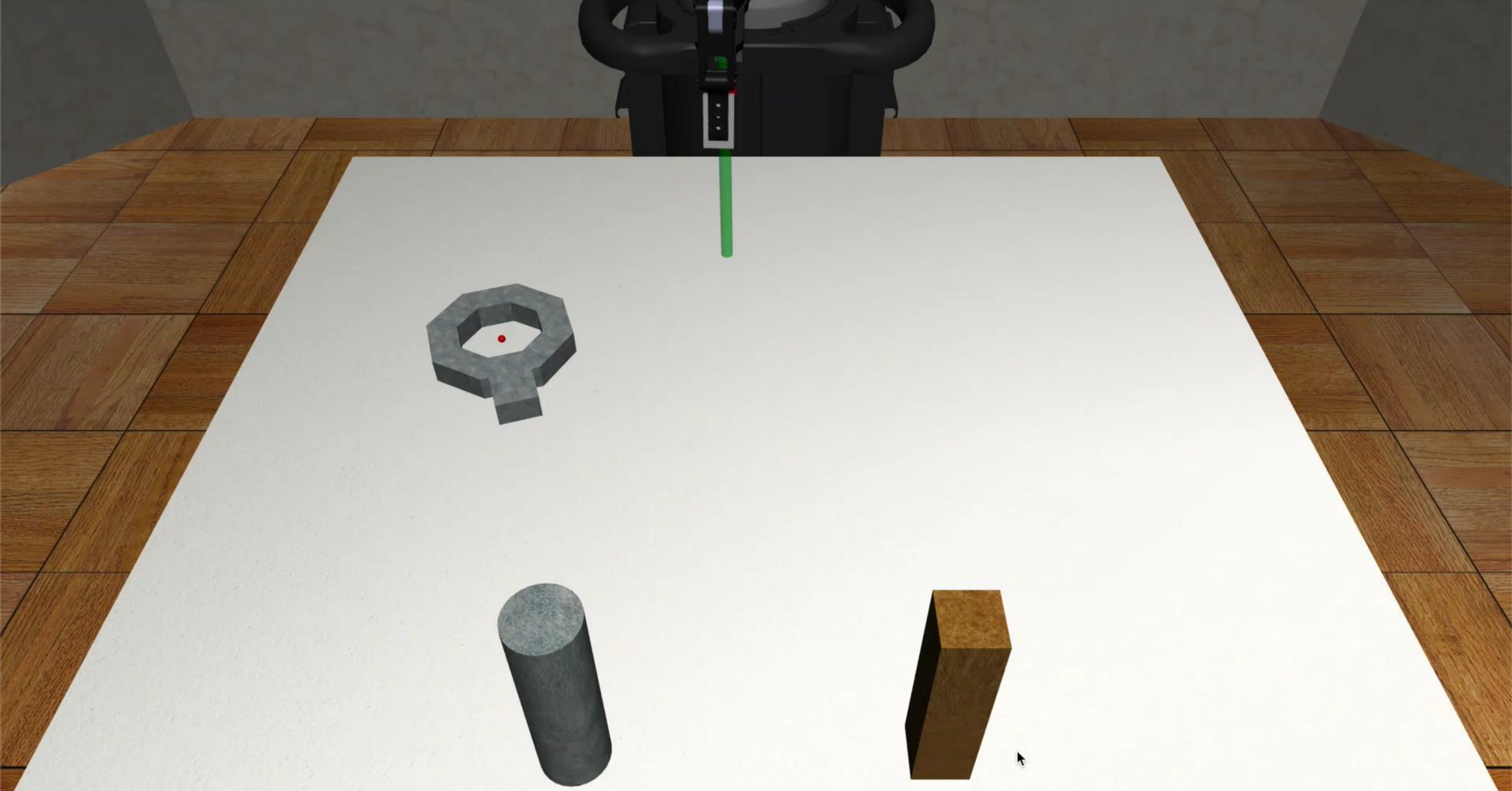


Hoque et al. "ThriftyDAgger: Budget-Aware Novelty and Risk Gating for Interactive Imitation Learning." CoRL 2021.

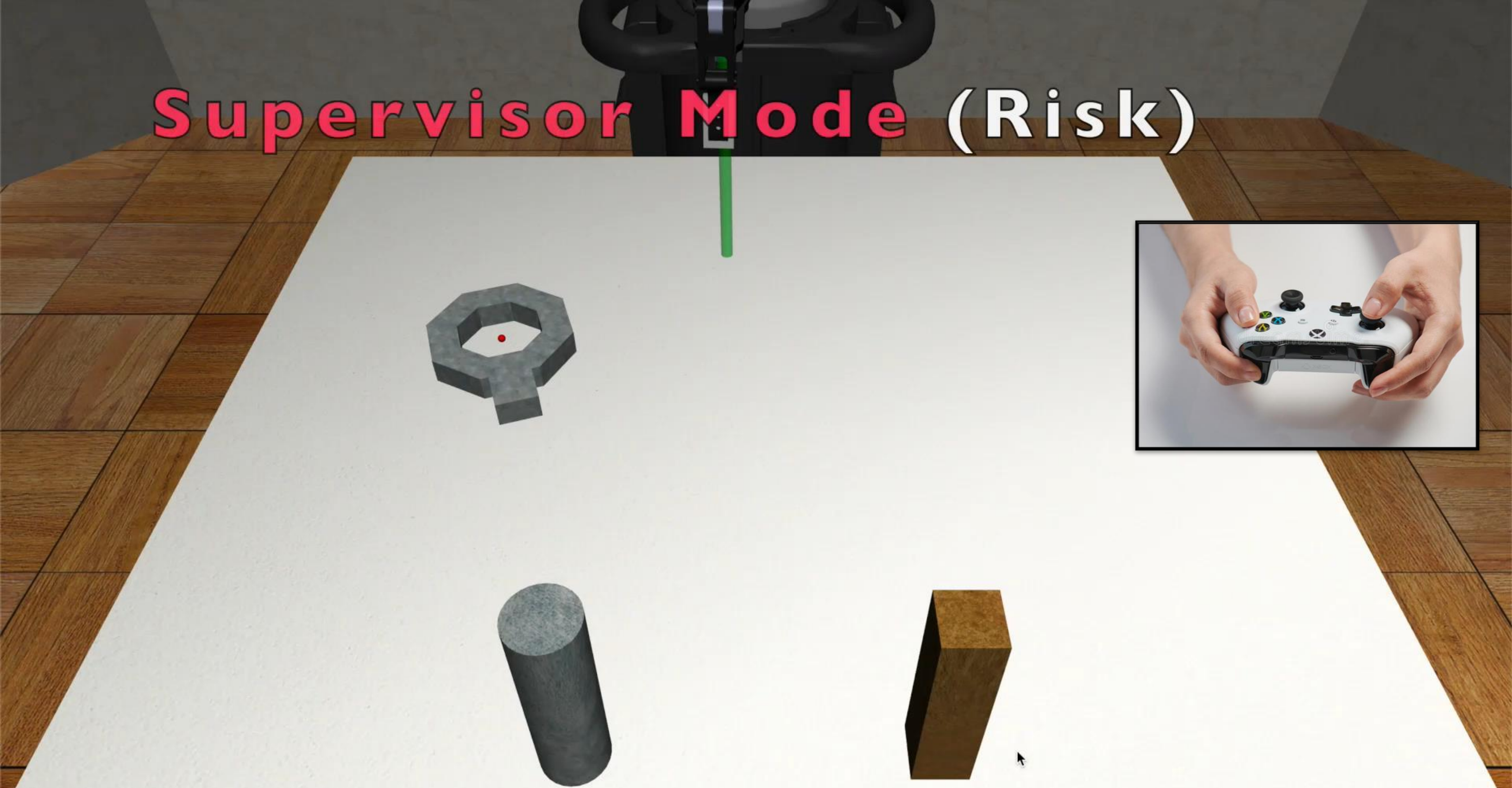


Hoque et al. "ThriftyDAgger: Budget-Aware Novelty and Risk Gating for Interactive Imitation Learning." CoRL 2021.

Hoque et al. "ThriftyDAgger: Budget-Aware Novelty and Risk Gating for Interactive Imitation Learning." CoRL 2021.

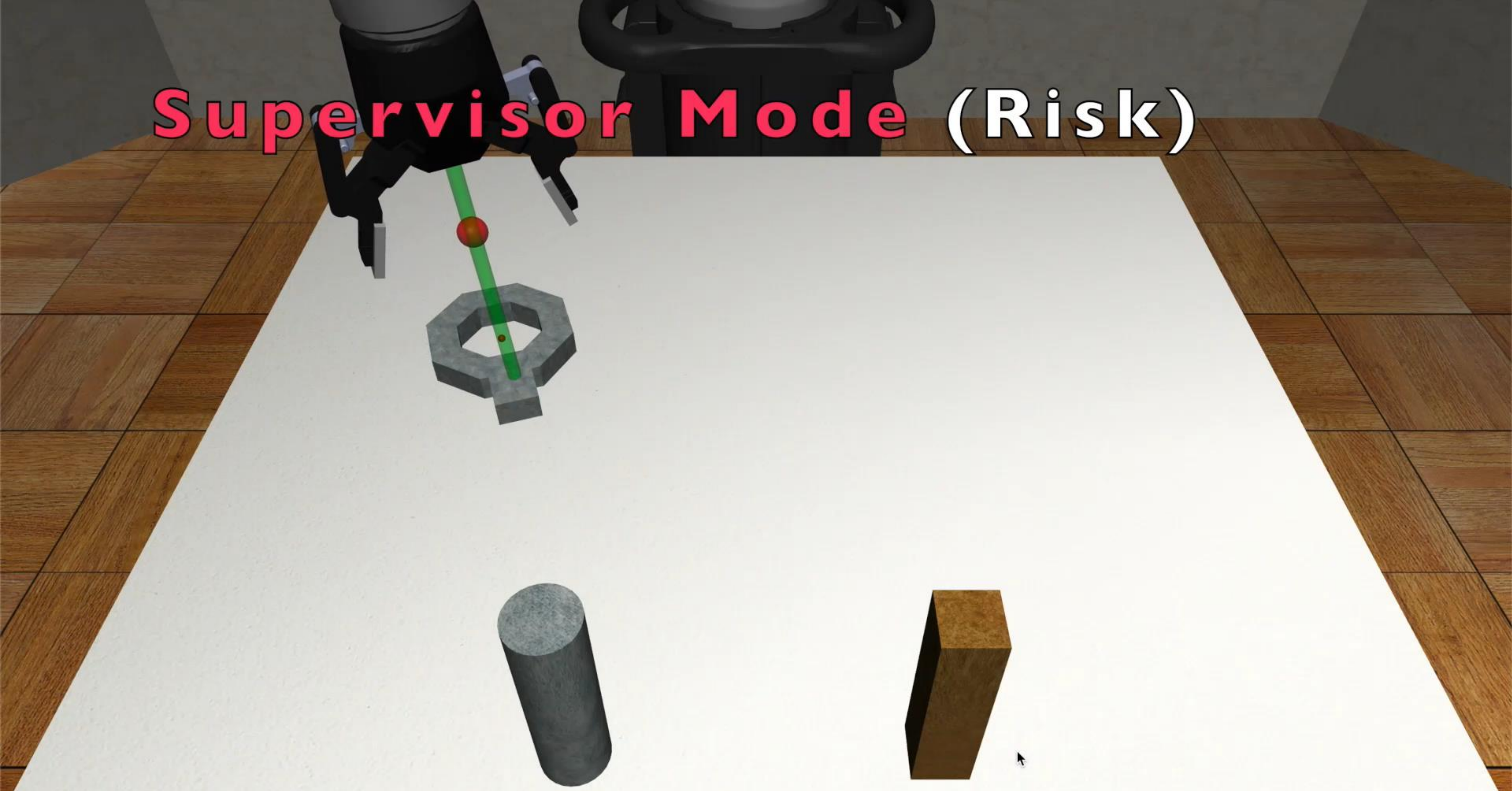


Hoque et al. "ThriftyDAgger: Budget-Aware Novelty and Risk Gating for Interactive Imitation Learning." CoRL 2021.



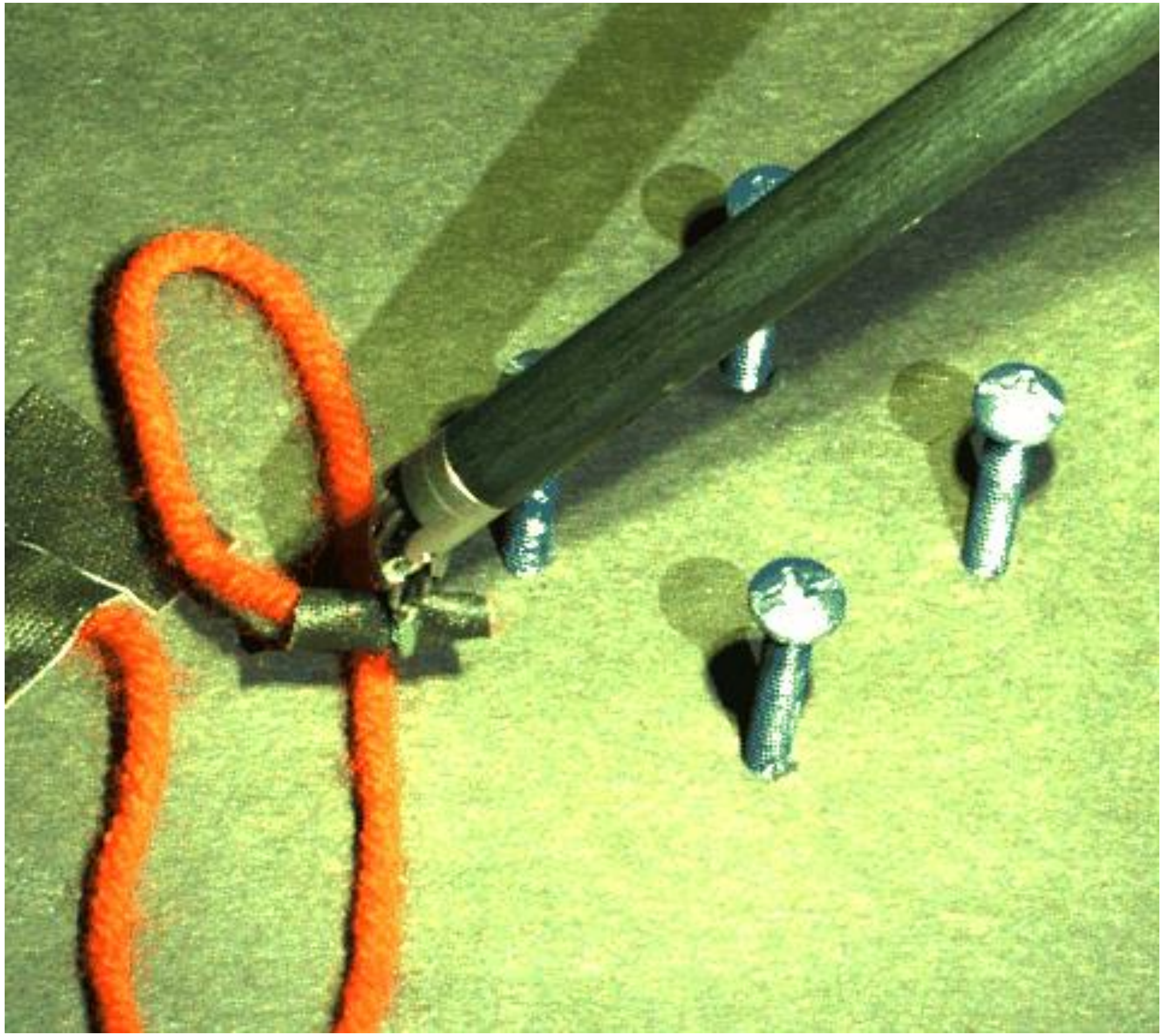
Hoque et al. "ThriftyDAgger: Budget-Aware Novelty and Risk Gating for Interactive Imitation Learning." CoRL 2021.

Supervisor Mode (Risk)

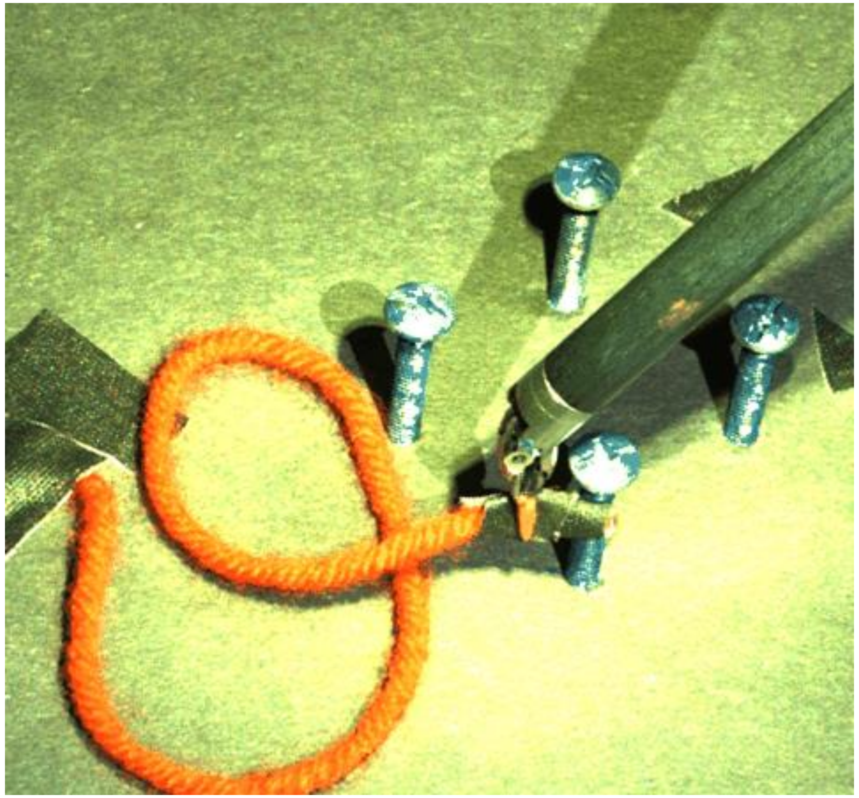


Hoque et al. "ThriftyDAgger: Budget-Aware Novelty and Risk Gating for Interactive Imitation Learning." CoRL 2021.

Human Demonstration

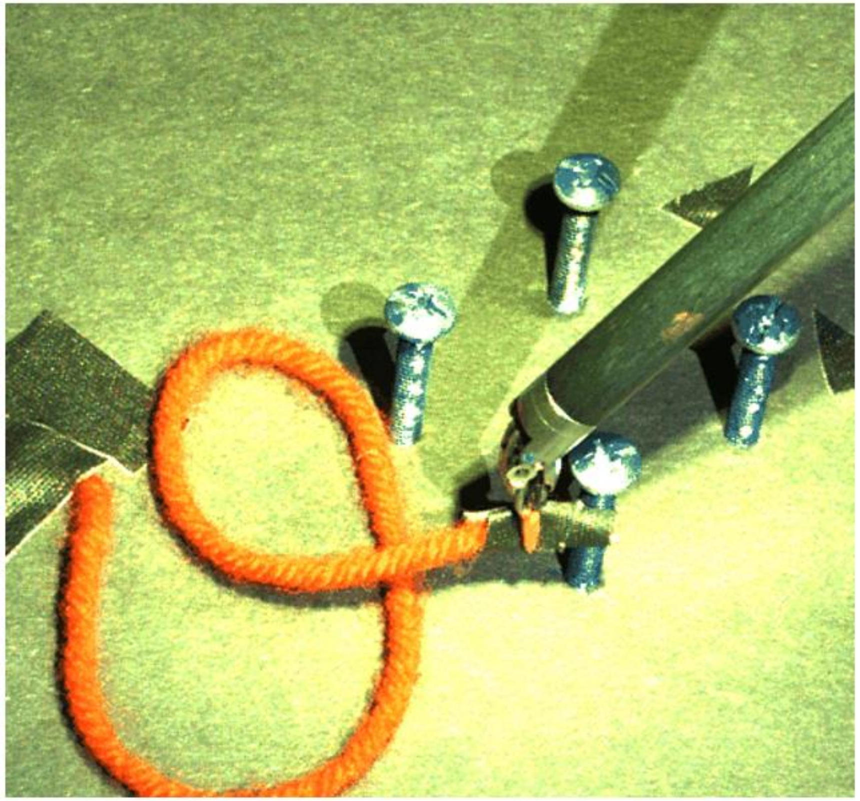


Behavior Cloning

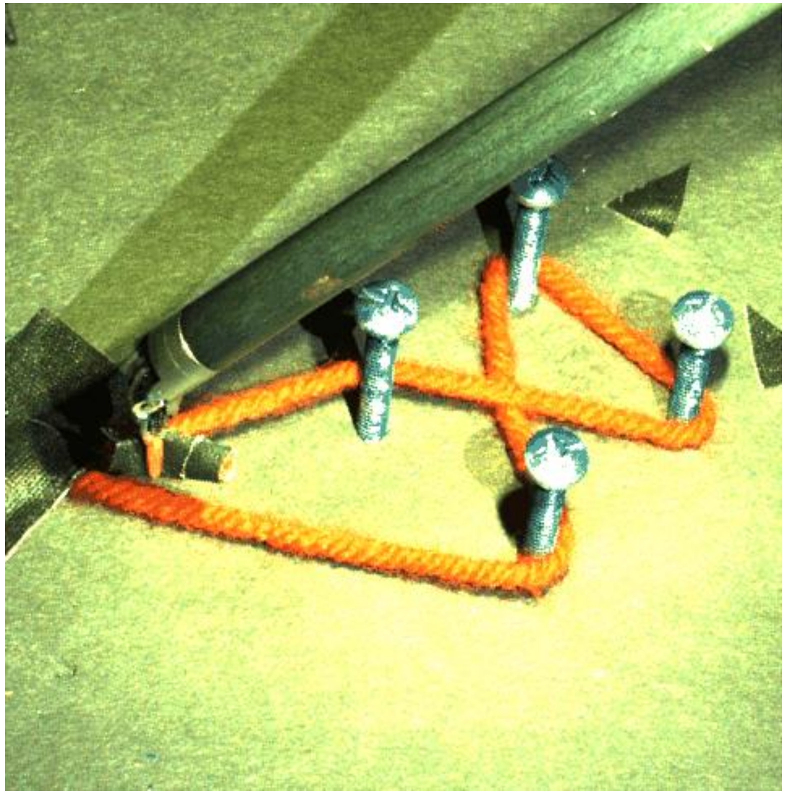


Hoque et al. "ThriftyDAgger: Budget-Aware Novelty and Risk Gating for Interactive Imitation Learning." CoRL 2021.

Behavior Cloning

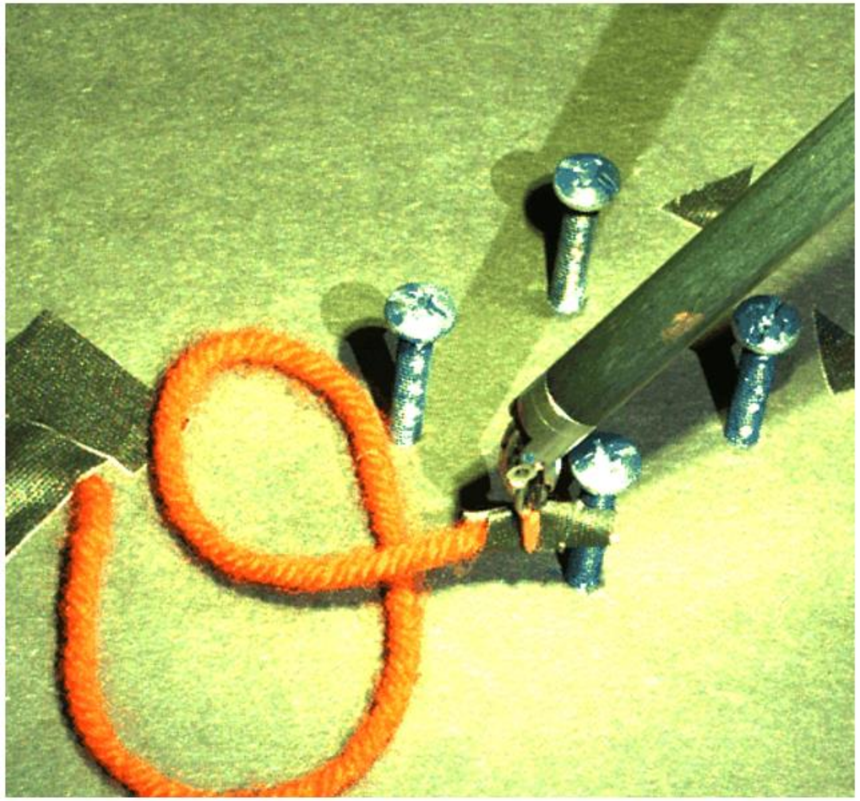


ThriftyDAgger (autonomous)

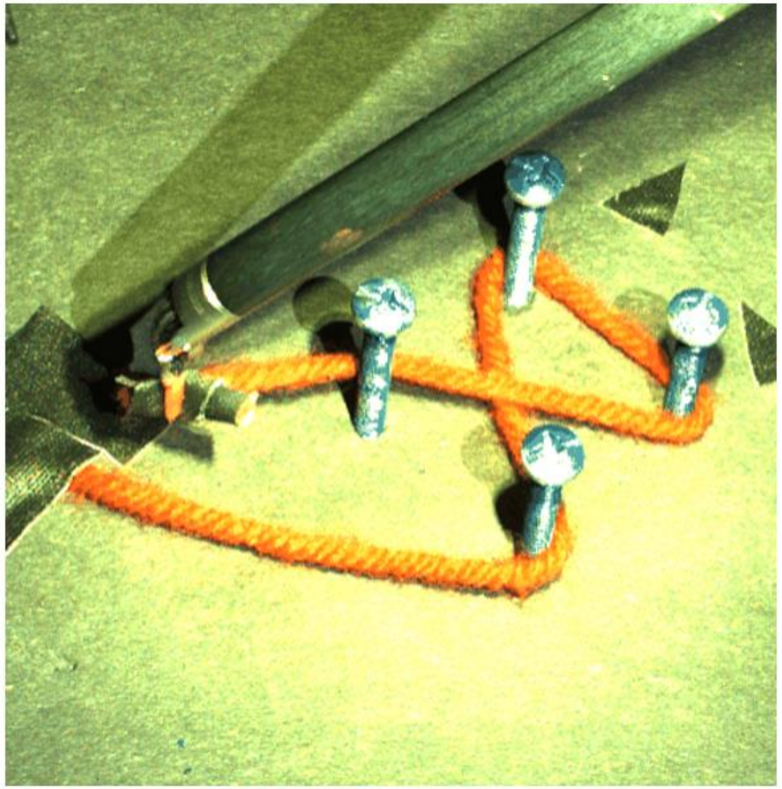


Hoque et al. "ThriftyDAgger: Budget-Aware Novelty and Risk Gating for Interactive Imitation Learning." CoRL 2021.

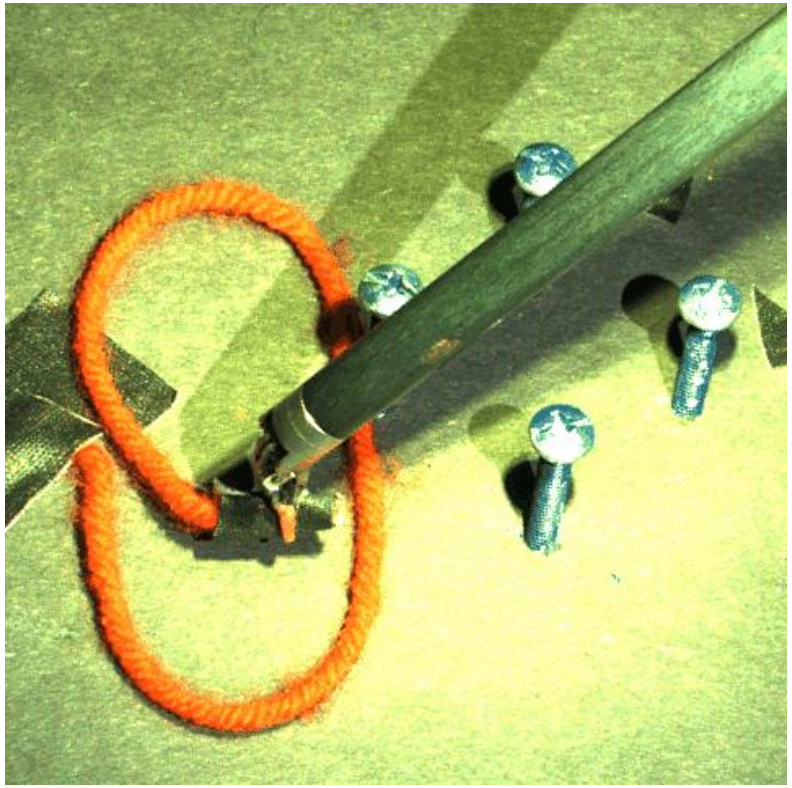
Behavior Cloning



ThriftyDAgger (autonomous)



ThriftyDAgger (+human)

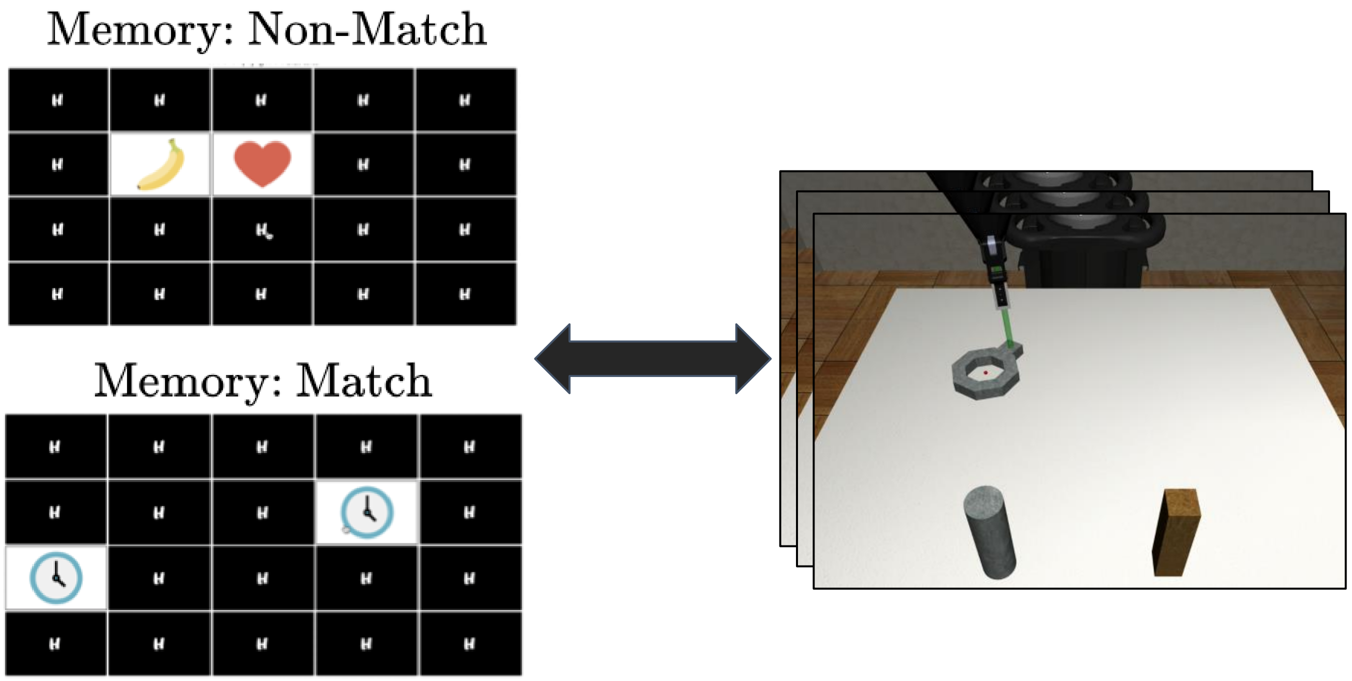


Hoque et al. "ThriftyDAgger: Budget-Aware Novelty and Risk Gating for Interactive Imitation Learning." CoRL 2021.

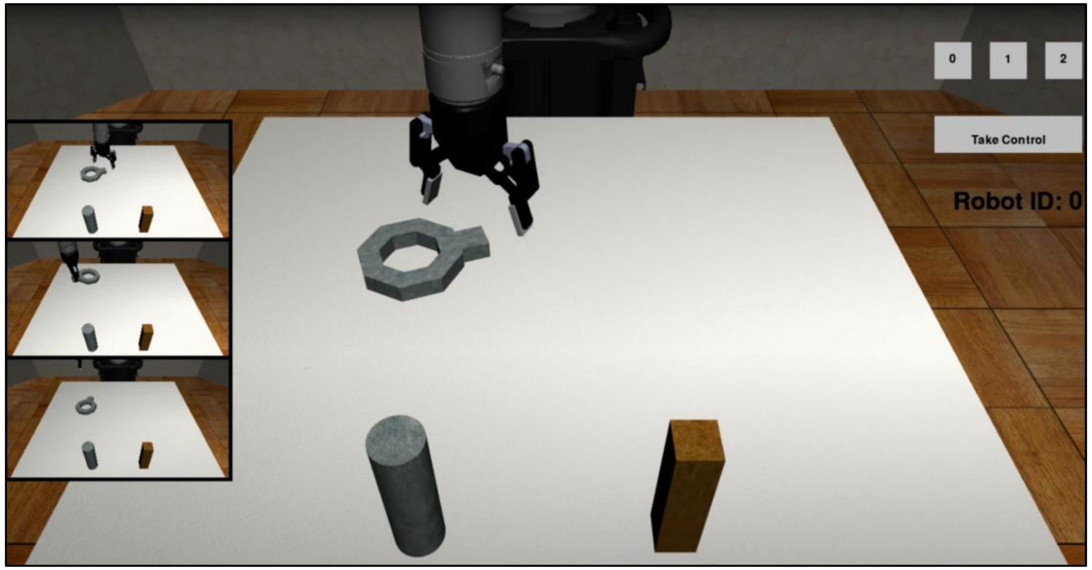
User Study

N=10 subjects each control 3 robots in simulation.

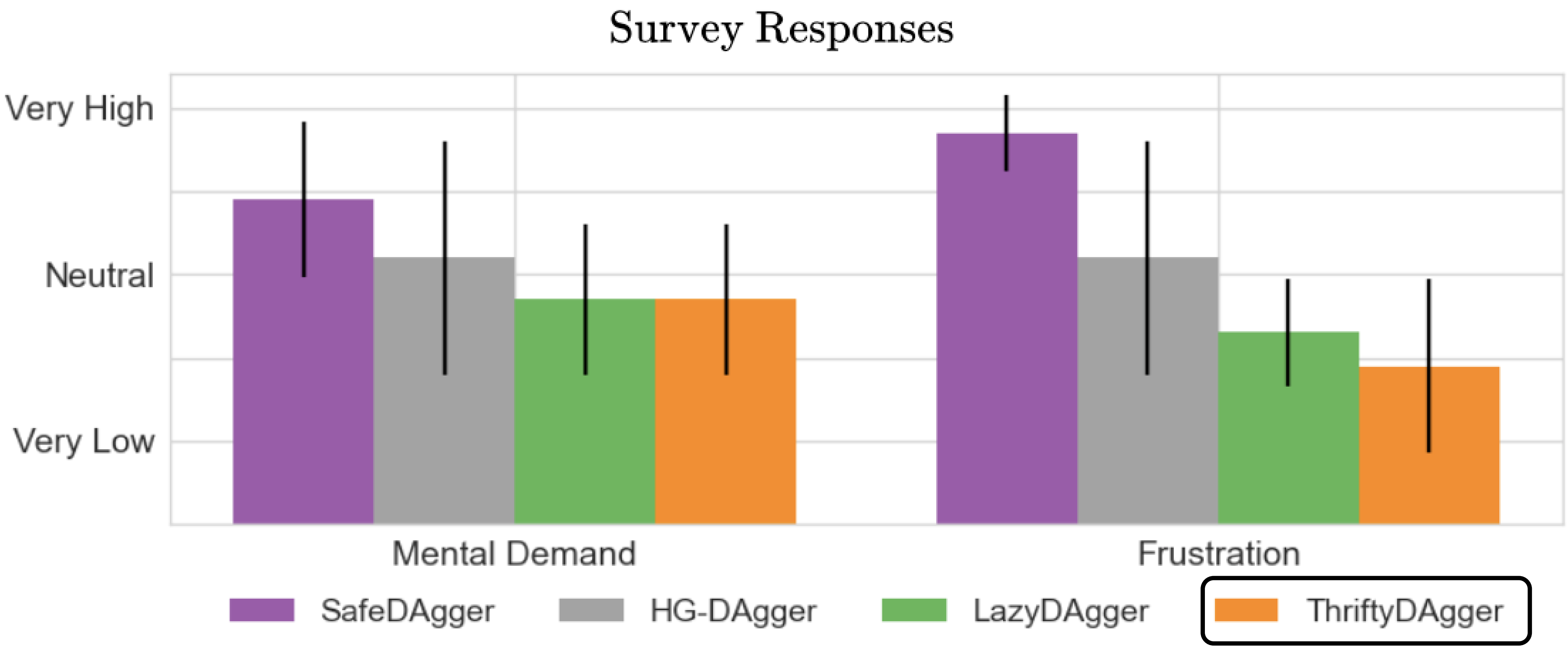
Robot-Gated



Human-Gated



ThriftyDAgger Qualitative Results



User Study Quantitative Results

ThriftyDAgger had

- 21% fewer human interventions
- 57% more concentration pairs found
- 80% more throughput