Human-CPS through the Lens of Learning and Control

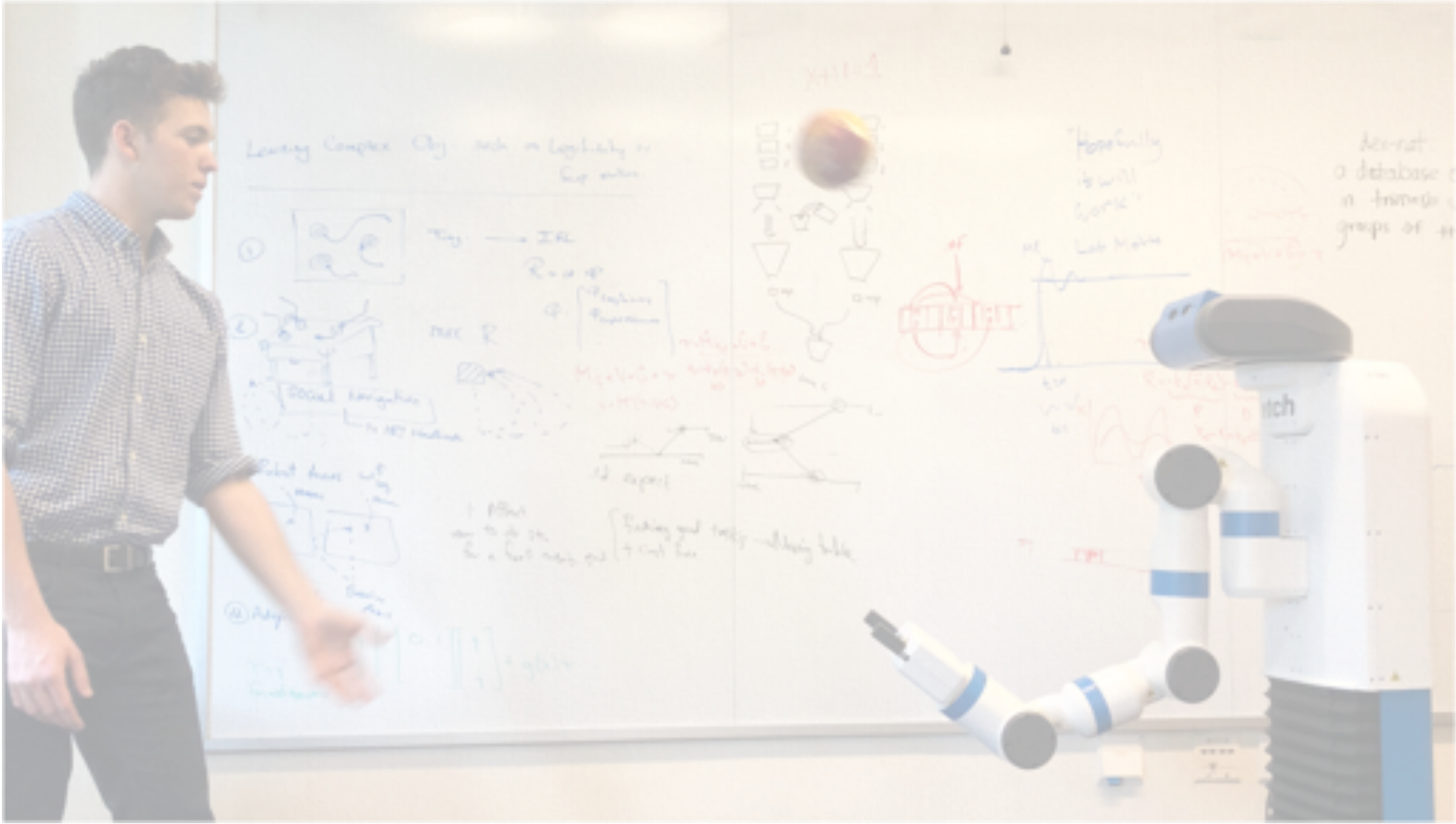
Dorsa Sadigh

iliad

intelligent and interactive autonomous systems

LELAND STANFORD JUNIOR UNIVERSITY · DIE LUFT DER FREIHEIT WEHT · 1891

Human-CPS through the Lens of Learning and Control



REAL-TIME
GOAL 1

1) There is an *opportunity* for learning and control

… to formalize and solve challenging problems of interaction with humans.

1) There is an *opportunity* for learning and control

… to formalize and solve challenging problems of interaction with humans.

2) We need to design *computational models of human* behavior

Can we rely on low-dimensional statistics that capture high-dimensional interactions?

1) There is an *opportunity* for learning and control

… to formalize and solve challenging problems of interaction with humans.

2) We need to design *computational models of human* behavior

Can we rely on low-dimensional statistics that capture high-dimensional interactions?

3) We spend a lot of effort learning what humans want or do…
… but humans constantly *change*

What can learning and control do?

1) There is an *opportunity* for learning and control

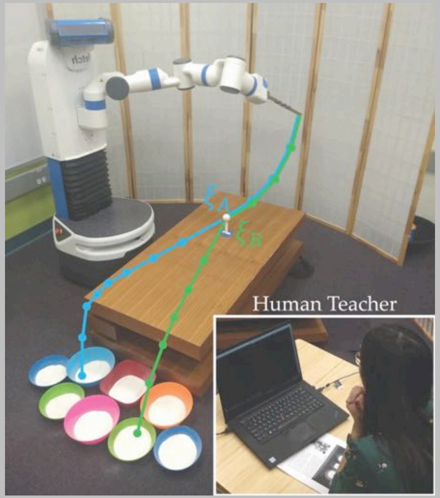
… to formalize and solve challenging problems of interaction with humans.

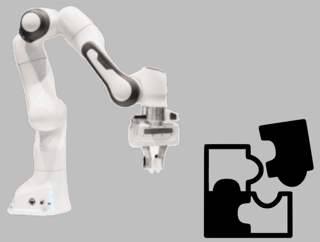
2) We need to design *computational models of human* behavior

Can we rely on low-dimensional statistics that capture high-dimensional interactions?

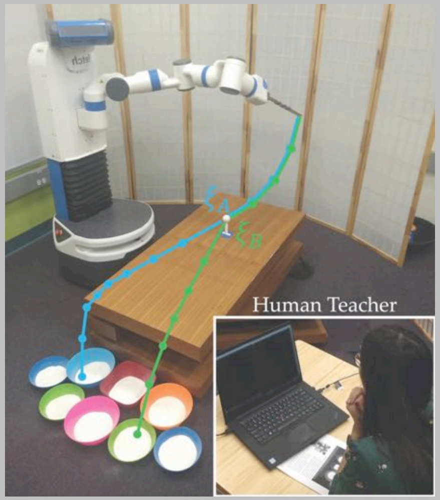
3) We spend a lot of effort learning what humans want or do…
… but humans constantly *change*

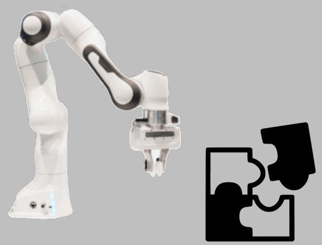
What can learning and control do?



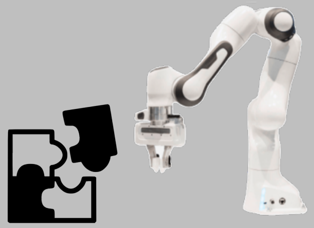


Teach *through demonstrations or comparisons*





***Teach** through demonstrations or comparisons*



***Teleoperate** the robot*

Human Teacher
Collaborative Block Stacking
Collaborative Transport

Human Models

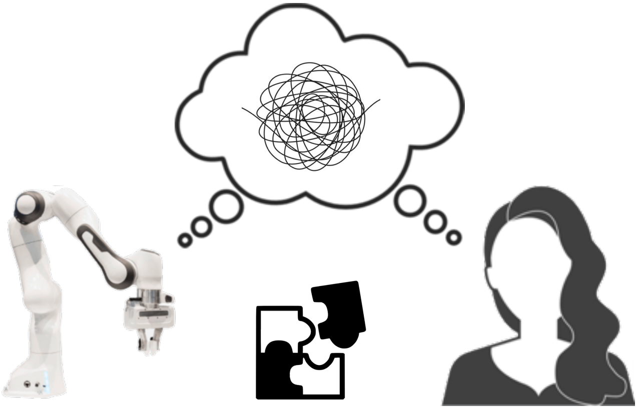
- Data-efficient learning of reward functions with different sources of data
- What happens on the ends of the risk spectrum?

Human Models

- Data-efficient learning of reward functions with different sources of data
- What happens on the ends of the risk spectrum?

Conventions

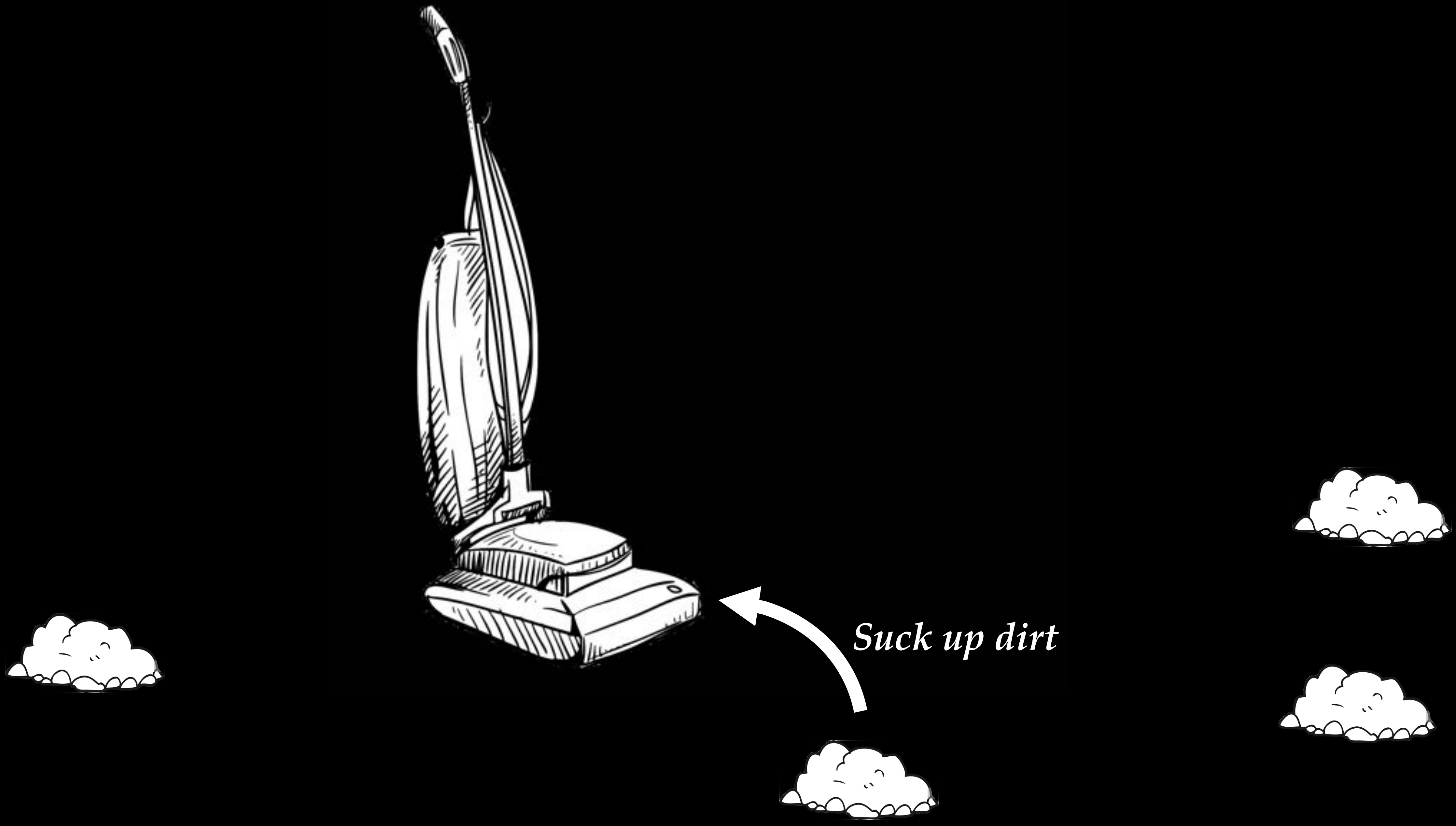
- What low dimensional representations are necessary when collaborating with humans?

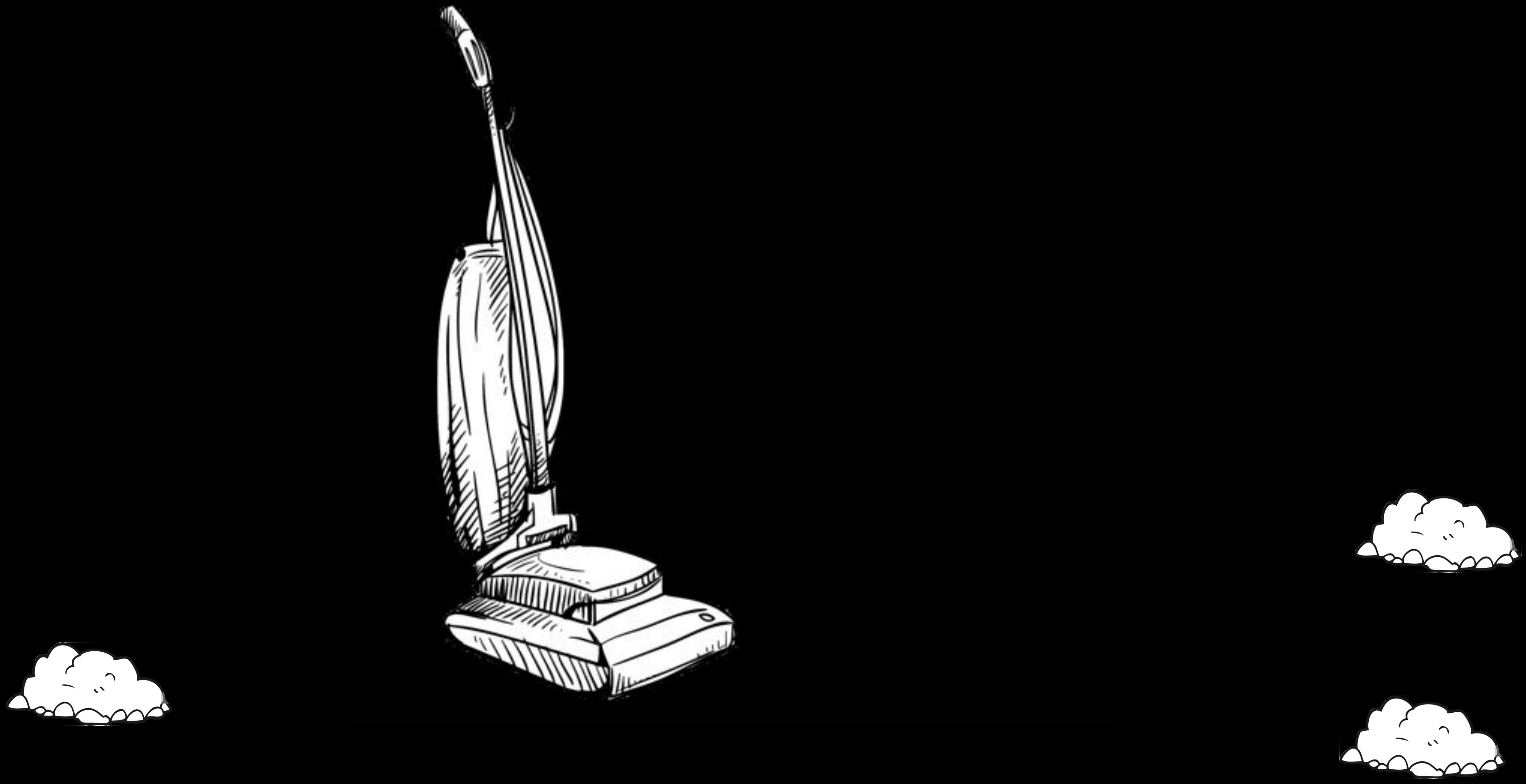


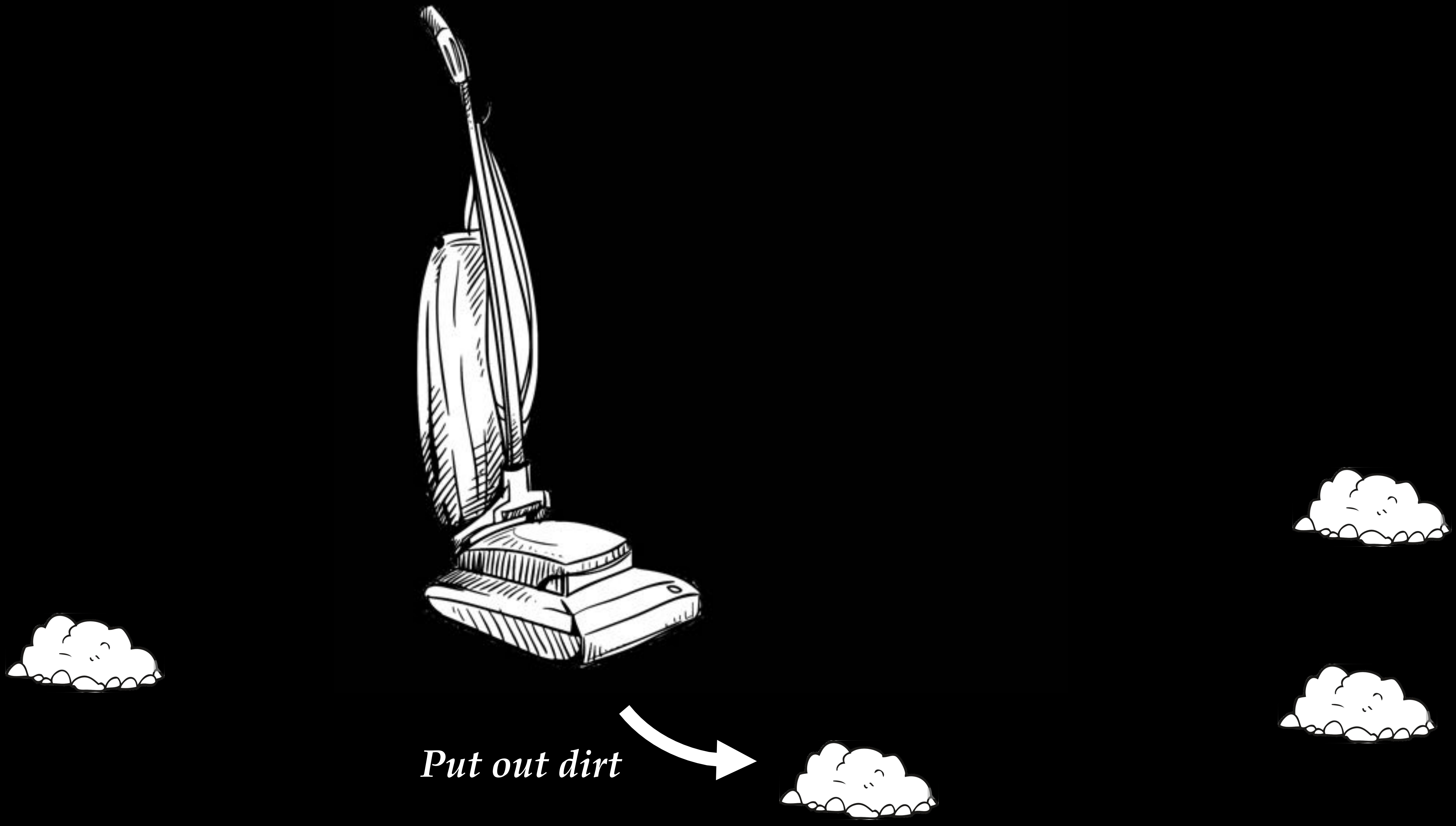

$R(\xi) =?$

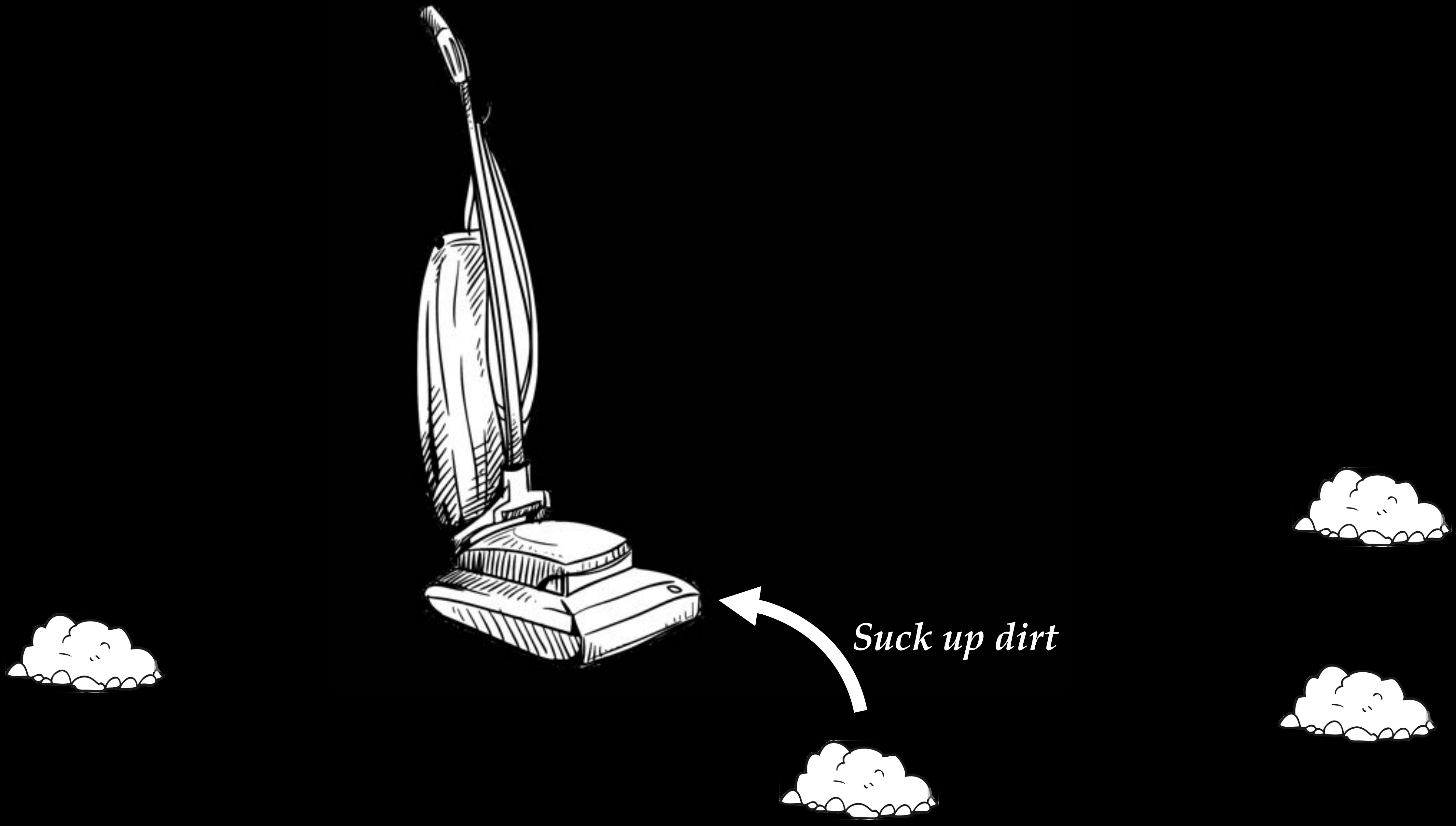
Objective:

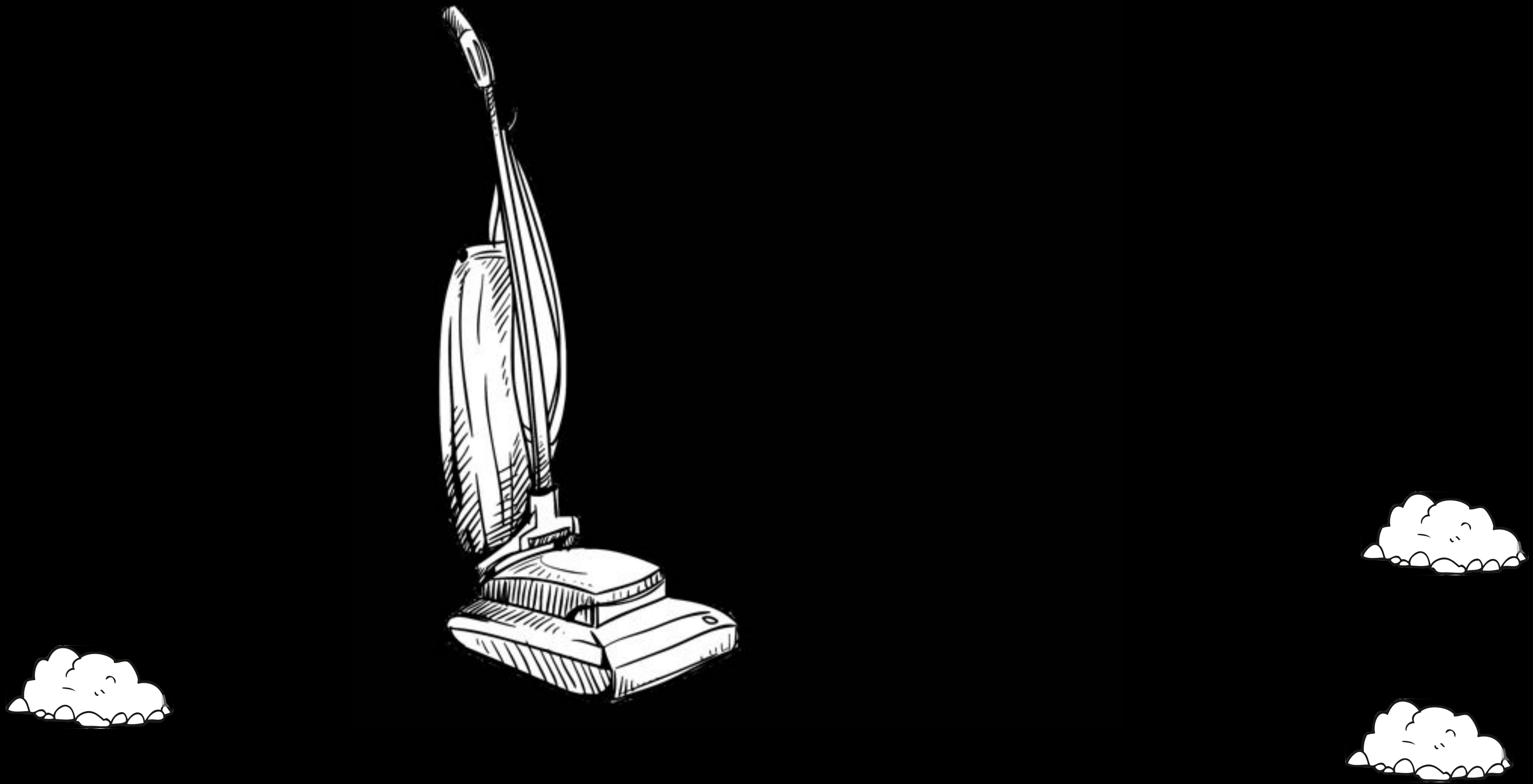
Suck up as much dirt as possible

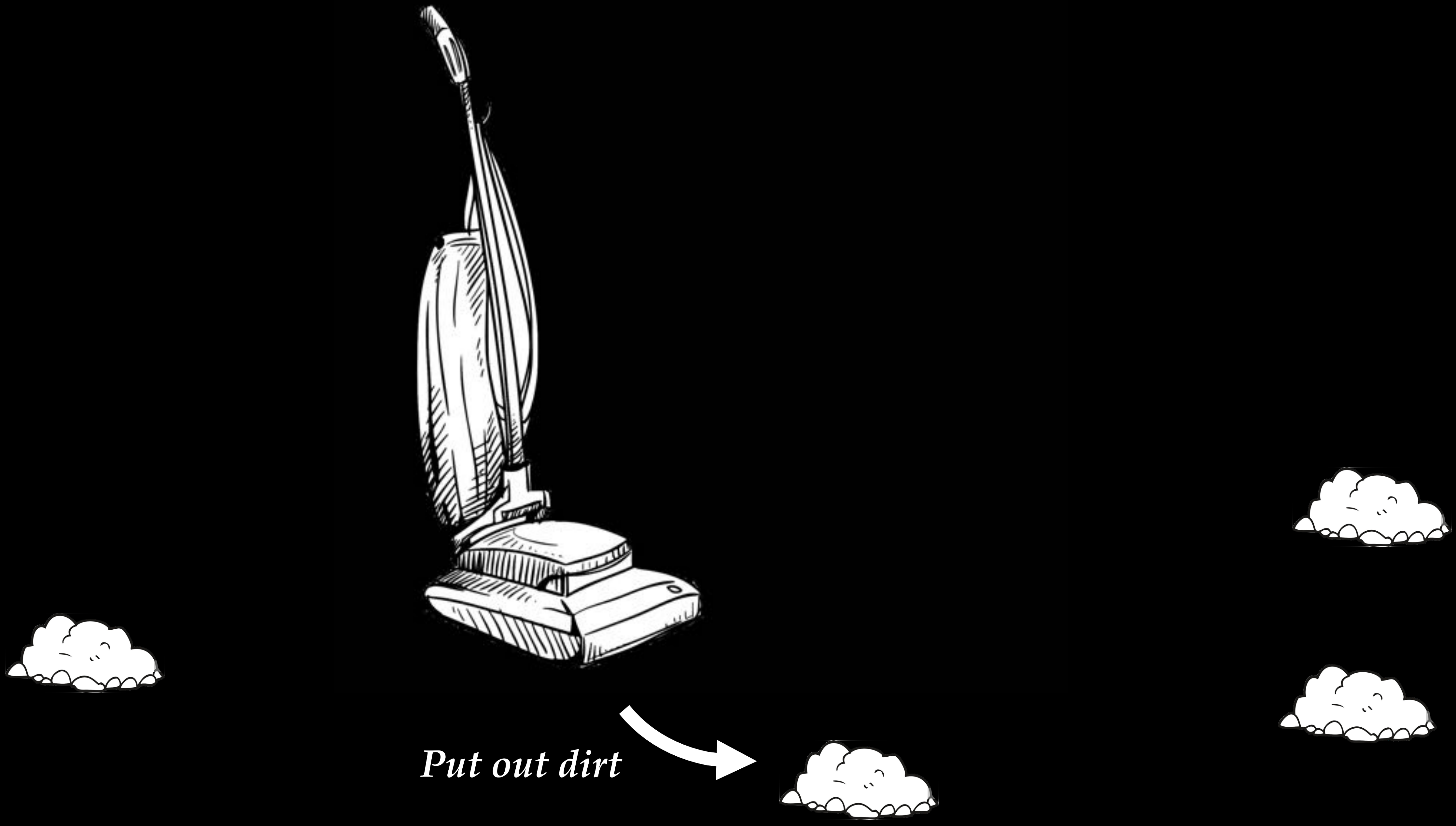

Suck up dirt




Put out dirt

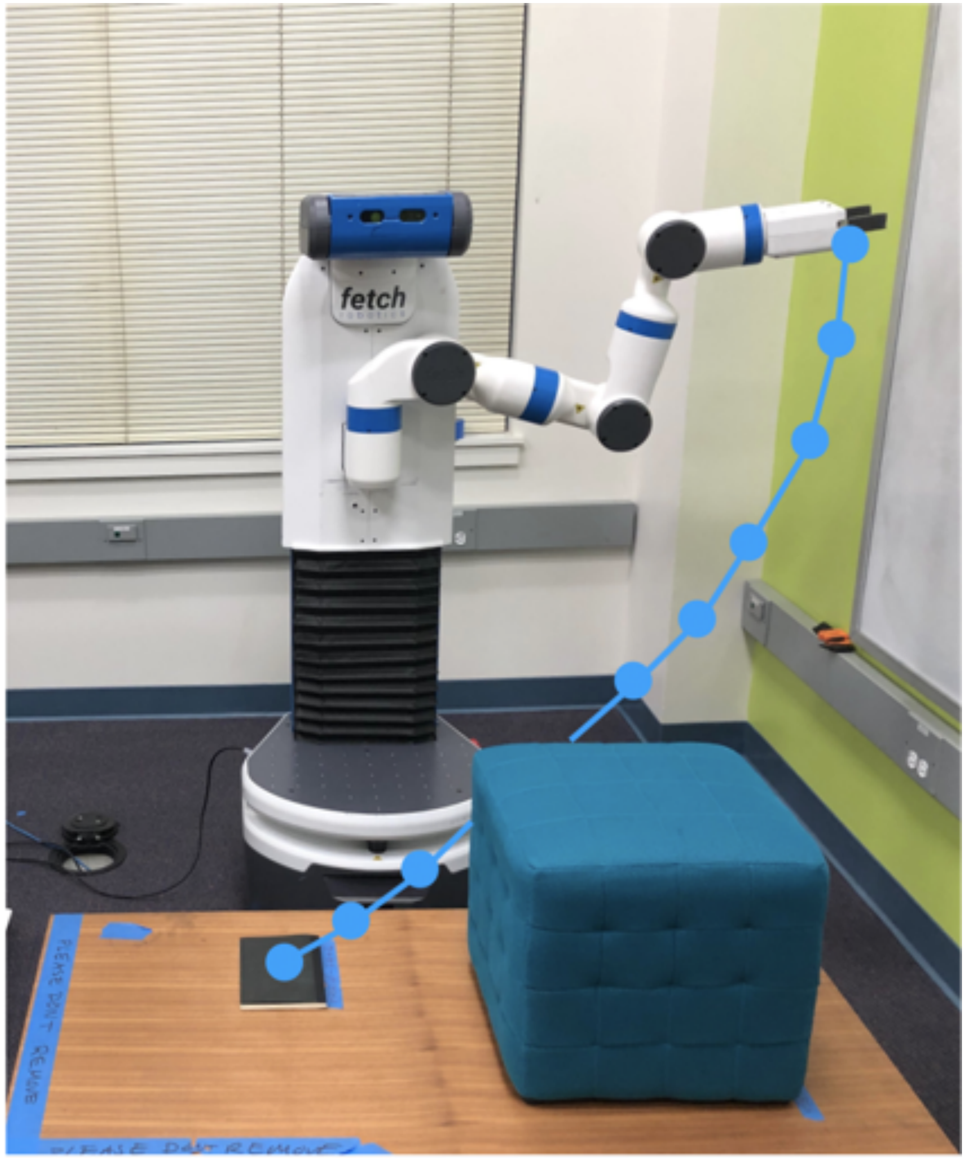

Suck up dirt



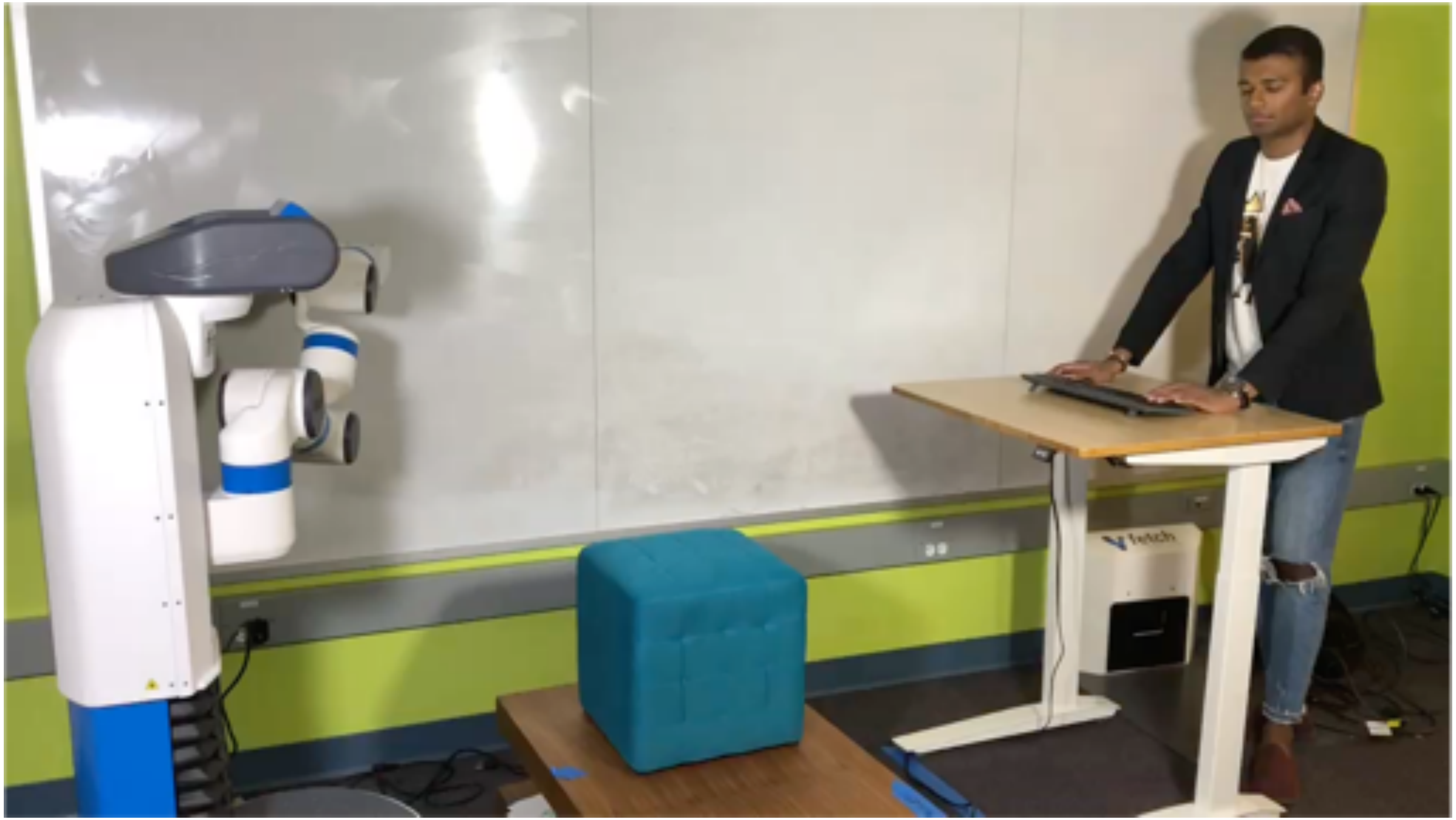

Put out dirt


$R_H(\xi)$
$\widehat{R_H(\xi)} = ?$

1. Reach the goal
2. Avoid the obstacle
3. Keep the arm low

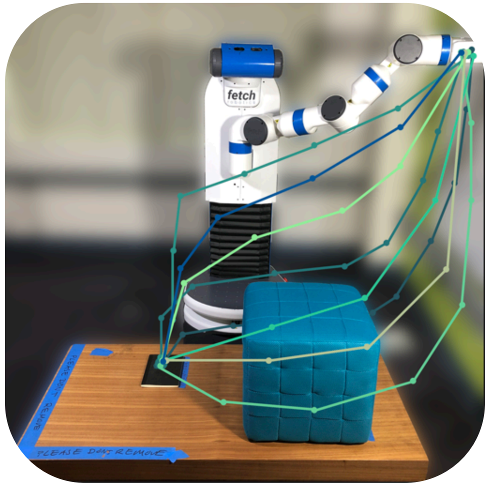


Collect Expert Demonstrations



Inverse Reinforcement Learning

Learn Human's reward function based on
Inverse Reinforcement Learning:



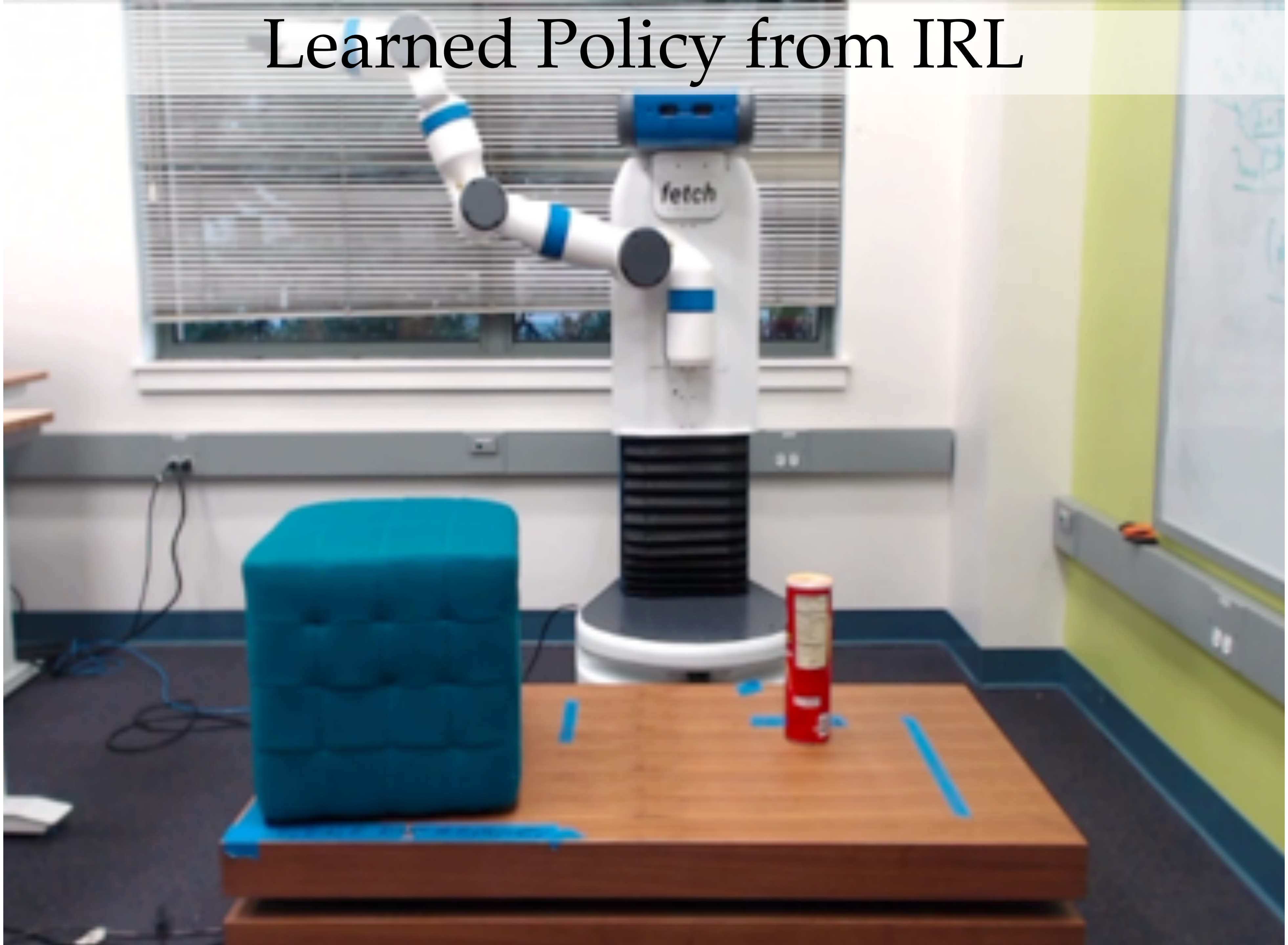
$$P(a_H|s,w) \propto \exp(R_H(s,a_H))$$

$$R_H(s,a_H) = w^\top \phi(s,a_\mathcal{H})$$

$$a_H^* = \max_{a_H} R_H(s,a_H)$$

[Ziebart' 08] [Levine'10]

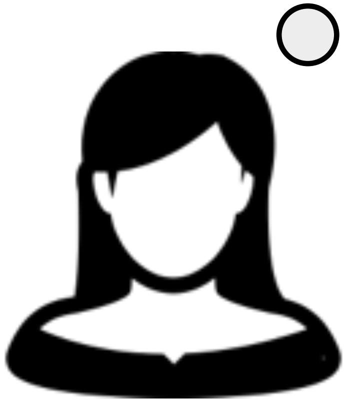
Learned Policy from IRL



Providing Demonstrations is Difficult!

“I had a hard time controlling the robot”

“I found the system difficult as someone who isn’t kinetically gifted”



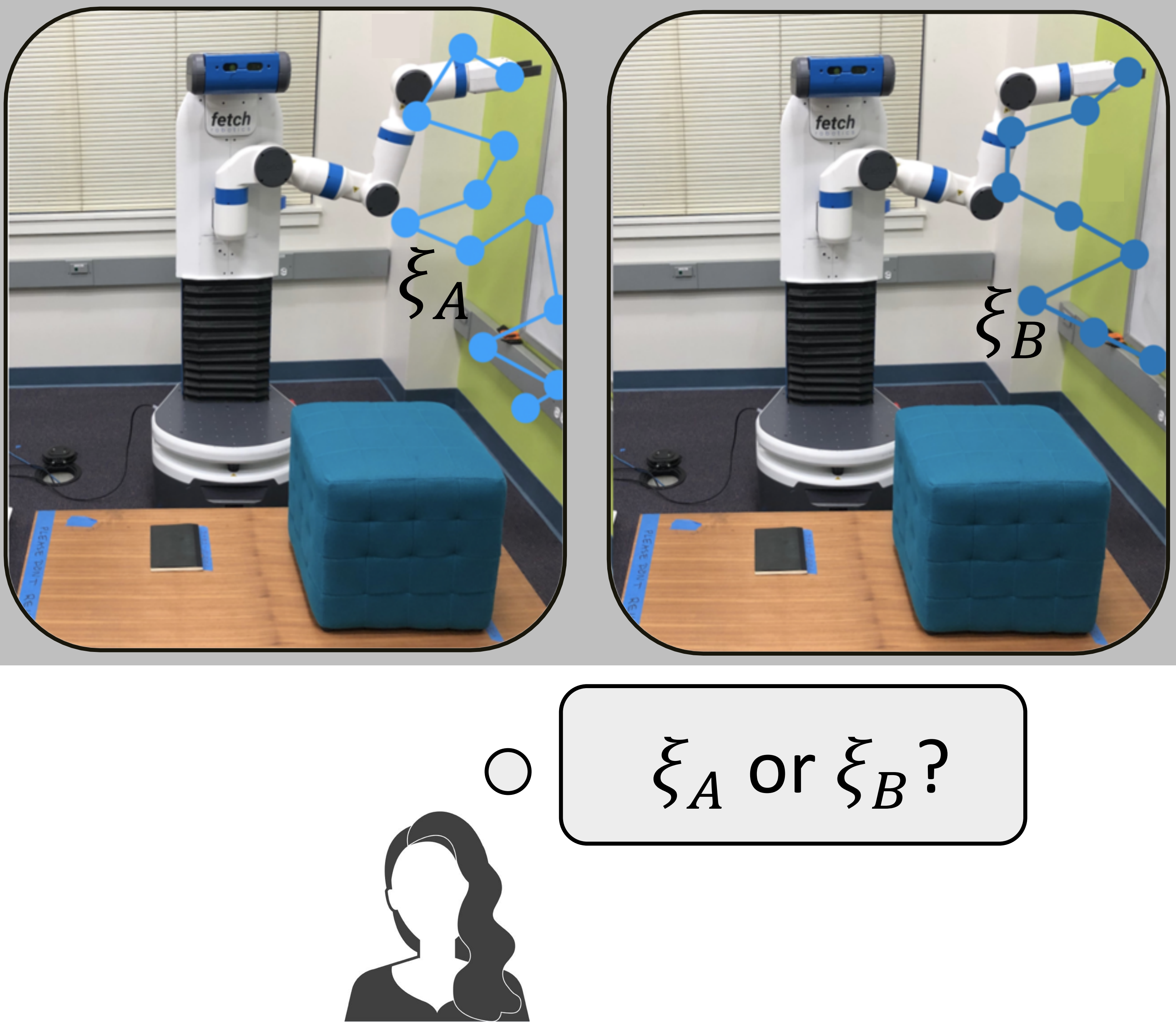
Leverage different sources of data to learn reward functions:

Demonstrations

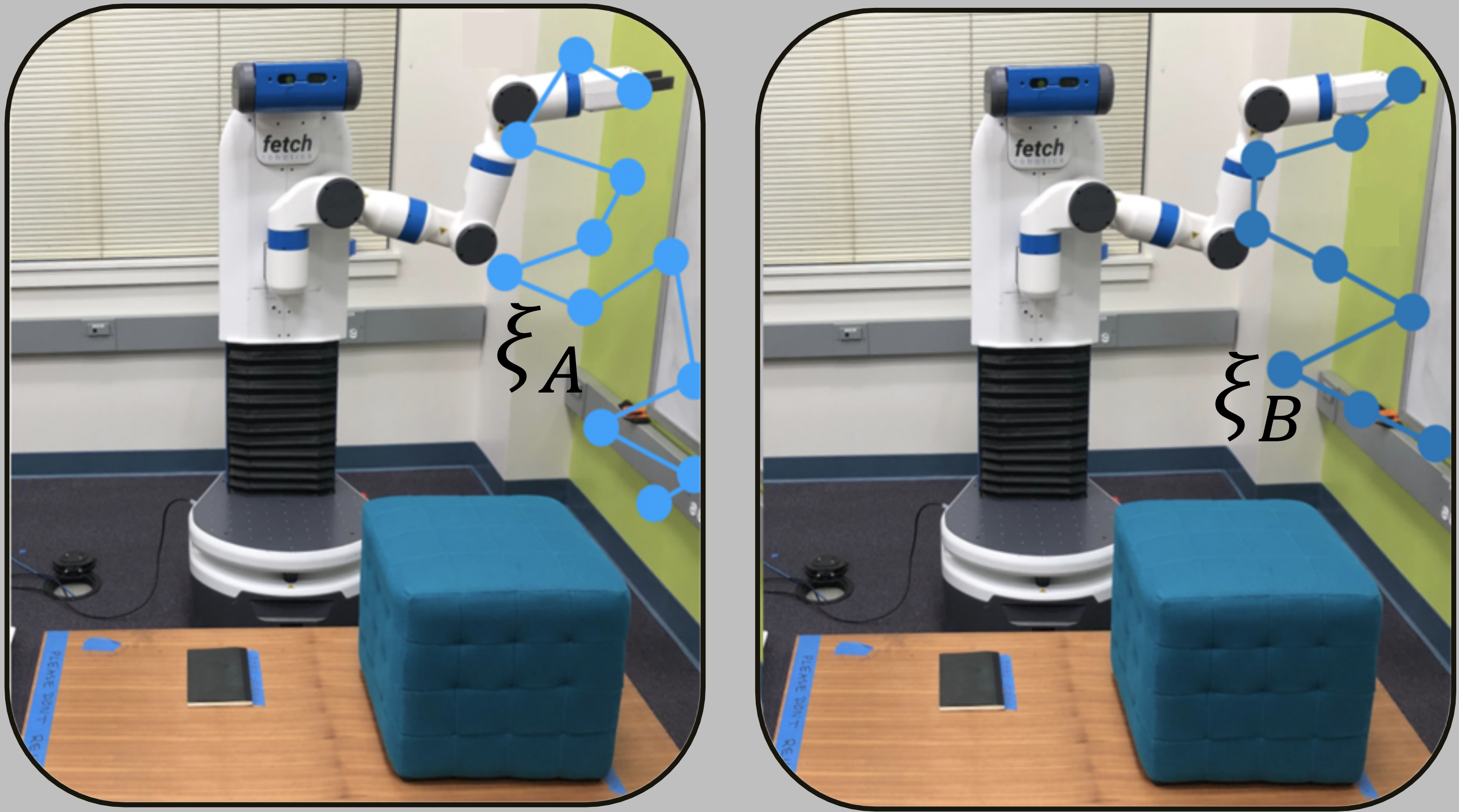
Comparisons

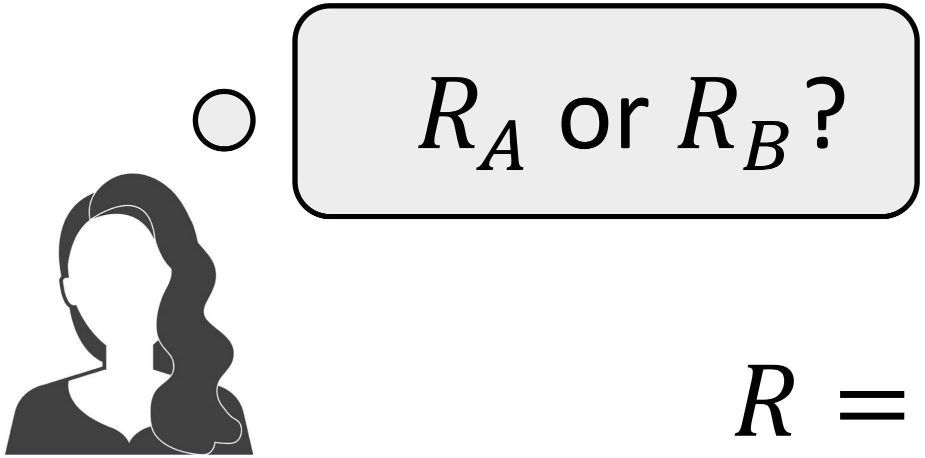
Language Instructions

Physical Feedback

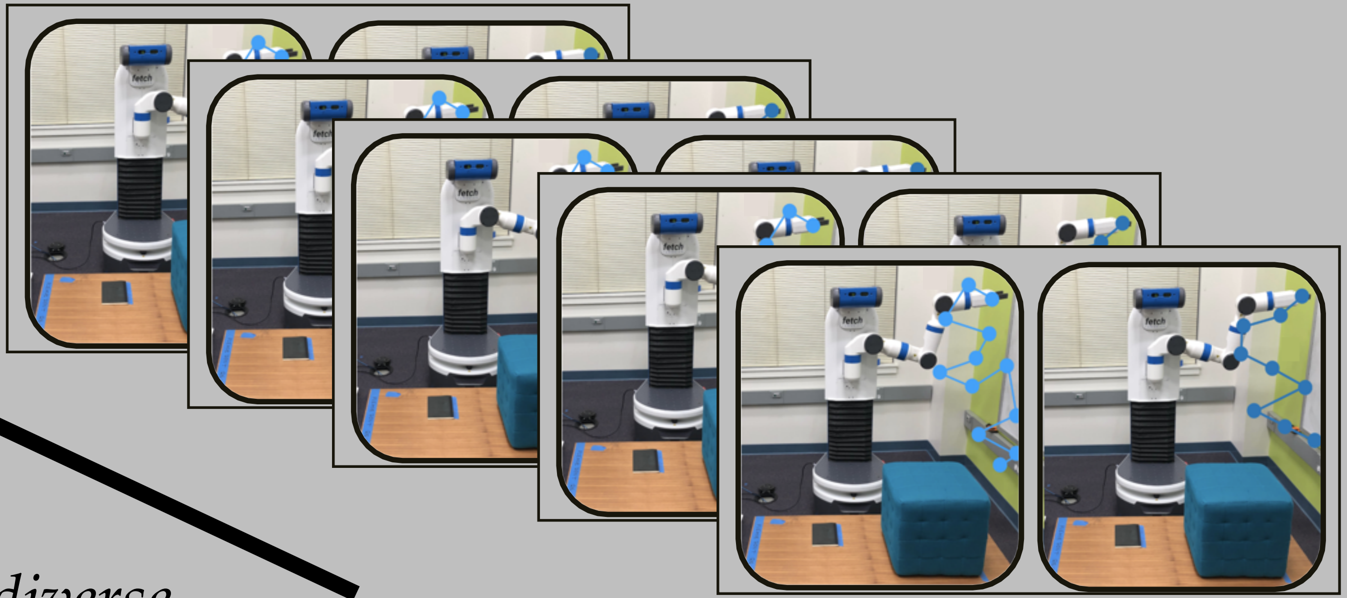


fetch
fetch
ξ_A
ξ_B
ξ_A or ξ_B?





$$R = \boldsymbol{w} \cdot \phi$$



Most informative, diverse sequence of queries

ξ_A or ξ_B ?

Actively *synthesizing* queries

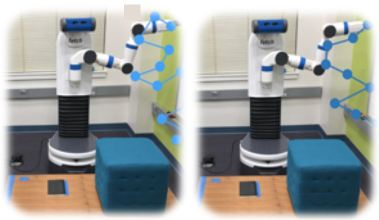
$$\max_{\varphi} \underbrace{\min\{\mathbb{E}[1 - f_{\varphi}(w)], \mathbb{E}[1 - f_{-\varphi}(w)]\}}^{\textit{minimum volume removed}}$$

Subject to $\varphi \in \mathbb{F}$

$$\mathbb{F} = \{\varphi : \varphi = \Phi(\xi_A) - \Phi(\xi_B), \xi_A, \xi_B \in \Xi\}$$

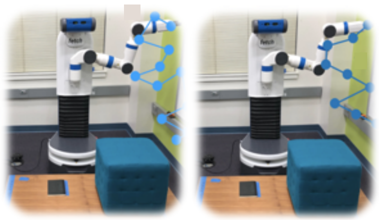
Human update function $f_{\varphi}(\boldsymbol{w}) = \min(1, \exp(I_t \boldsymbol{w}^{\top} \varphi))$

[Sadigh et al. RSS17]
[Biyik et al. CoRL18]
[Biyik et al. CDC19]



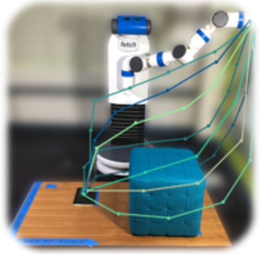
X ✓

Preferences:
Easier and more accurate to use – but *gives one bit of information*.

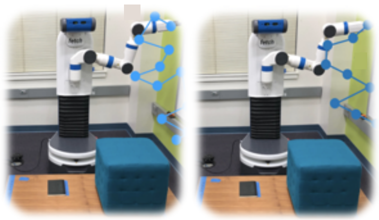


X ✓

Preferences:
Easier and more accurate to use – but *gives one bit of information*.

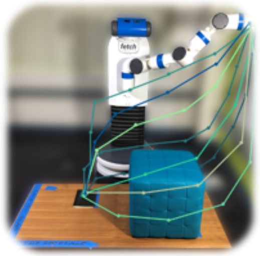


Demonstrations:
Rich and informative – but *noisy* and *inaccurate*.

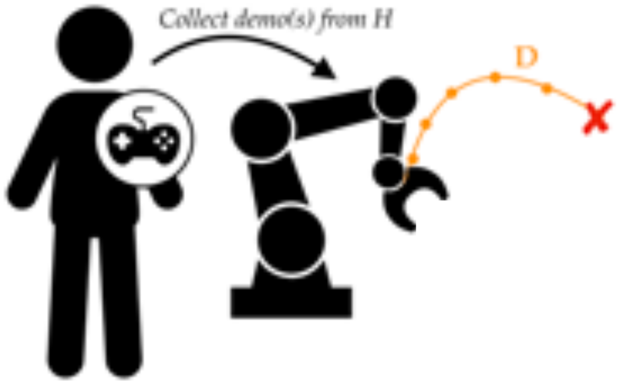


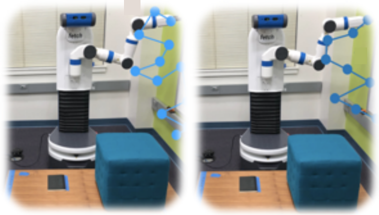
X ✓

Preferences:
Easier and more accurate to use – but *gives one bit of information*.



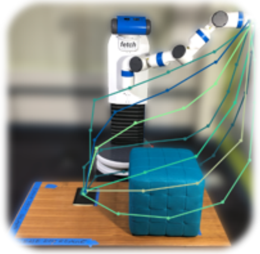
Demonstrations:
Rich and informative – but *noisy* and *inaccurate*.



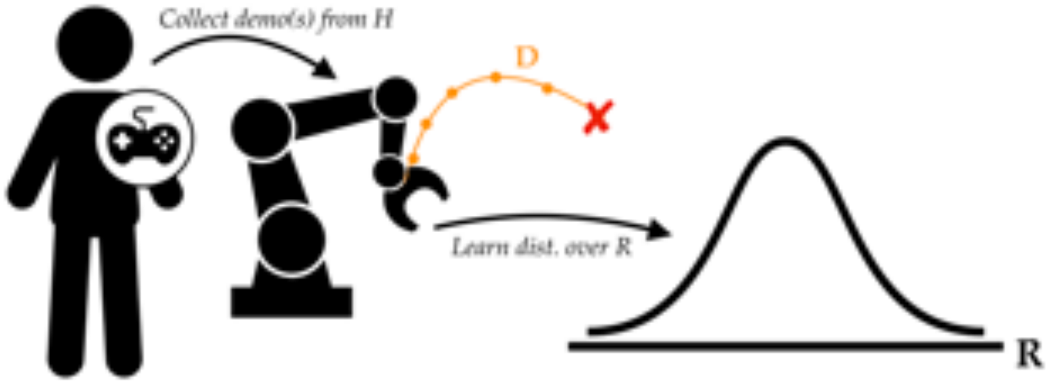


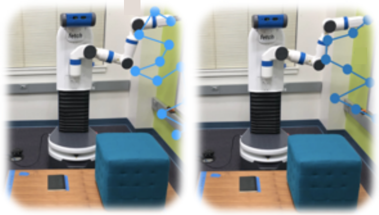
X ✓

Preferences:
Easier and more accurate to use – but *gives one bit of information*.



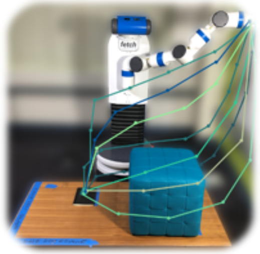
Demonstrations:
Rich and informative – but *noisy* and *inaccurate.*



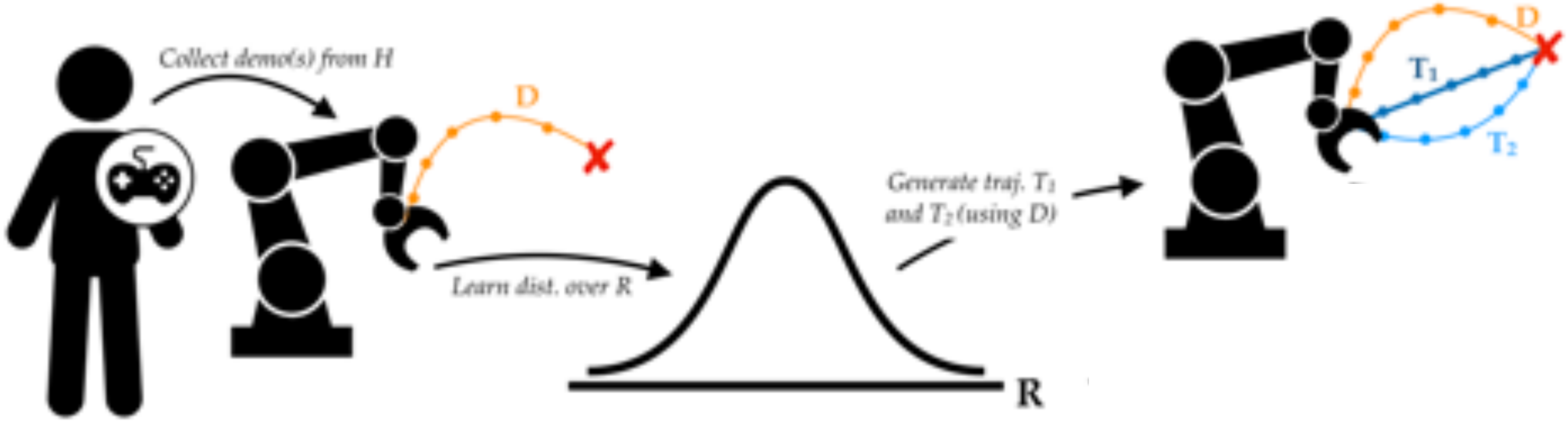


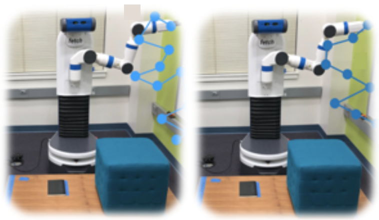
X ✓

Preferences:
Easier and more accurate to use – but *gives one bit of information*.



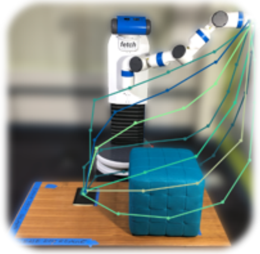
Demonstrations:
Rich and informative – but *noisy* and *inaccurate.*



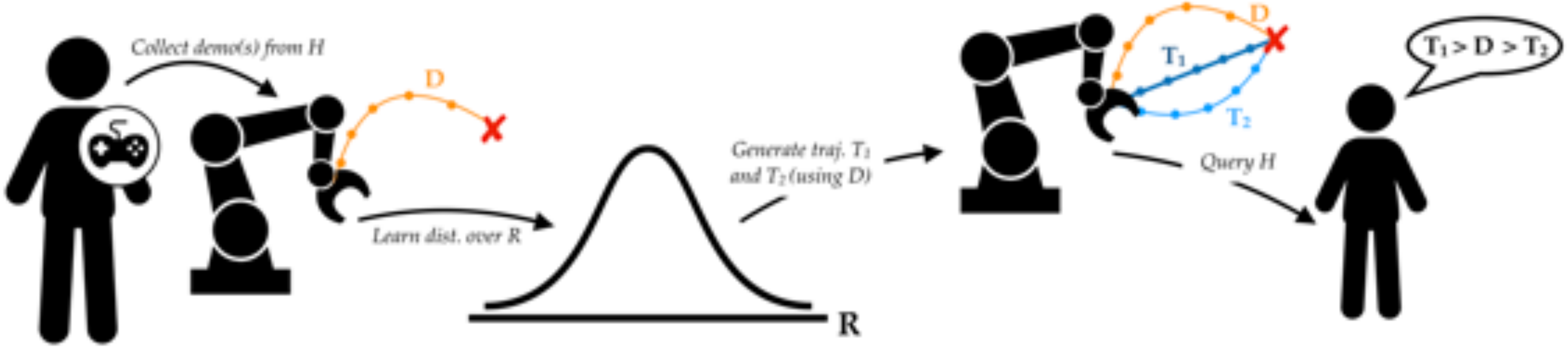


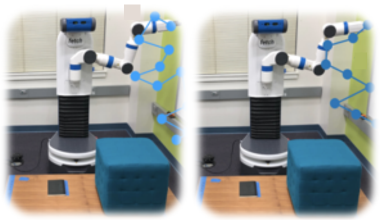
X ✓

Preferences:
Easier and more accurate to use – but *gives one bit of information*.



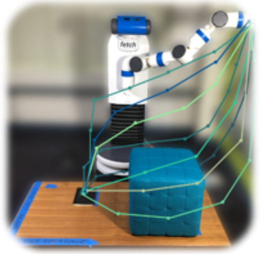
Demonstrations:
Rich and informative – but *noisy* and *inaccurate.*



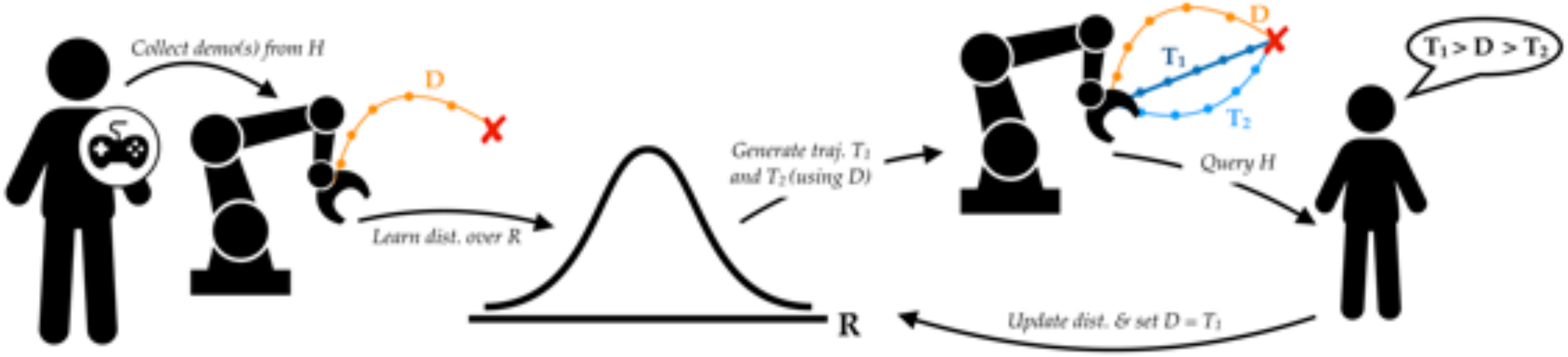


X ✓

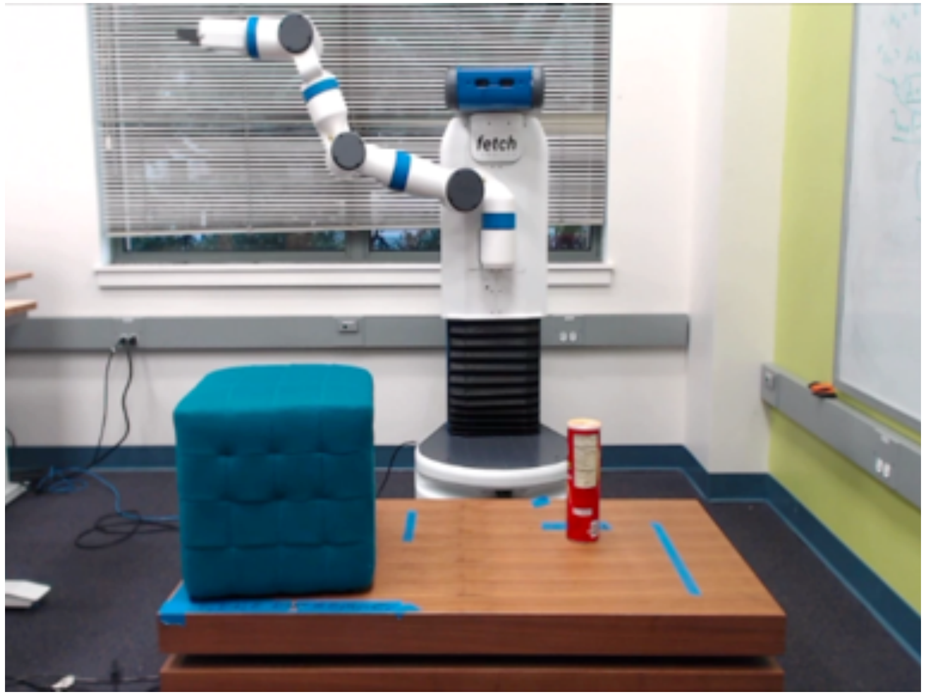
Preferences:
Easier and more accurate to use – but *gives one bit of information*.



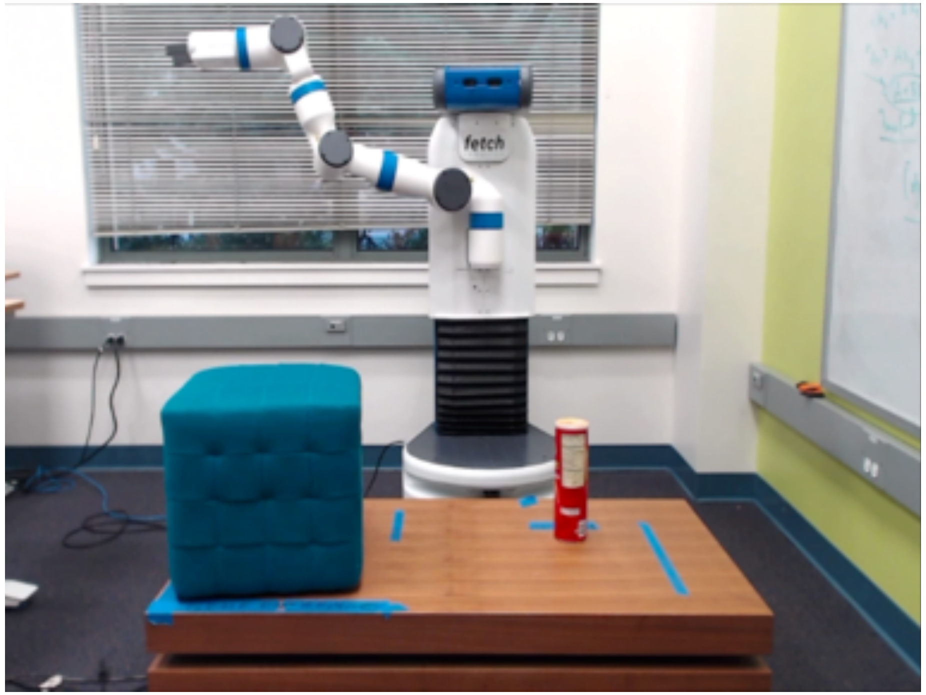
Demonstrations:
Rich and informative – but *noisy* and *inaccurate*.



Learning from Demonstration

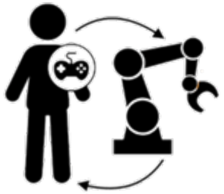


Learning from Demonstrations & Preferences



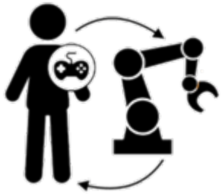
[Palan et al., RSS19]

Key Idea:



Integrating demonstrations and comparisons
to efficiently learn reward functions

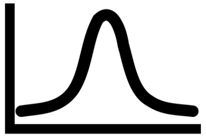
Key Idea:



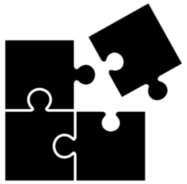
Integrating demonstrations and comparisons
to efficiently learn reward functions

Other considerations:

Dynamically changing rewards



Non-linear reward functions



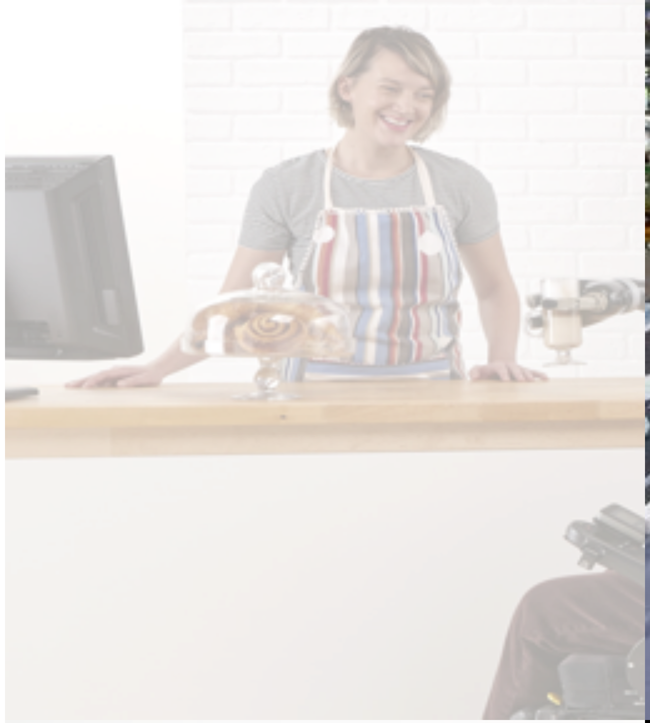
Easy active learning with info gain

[Basu et al. IROS19]
[Biyik et al., CoRL19]
[Biyik et al., submitted to RSS20]

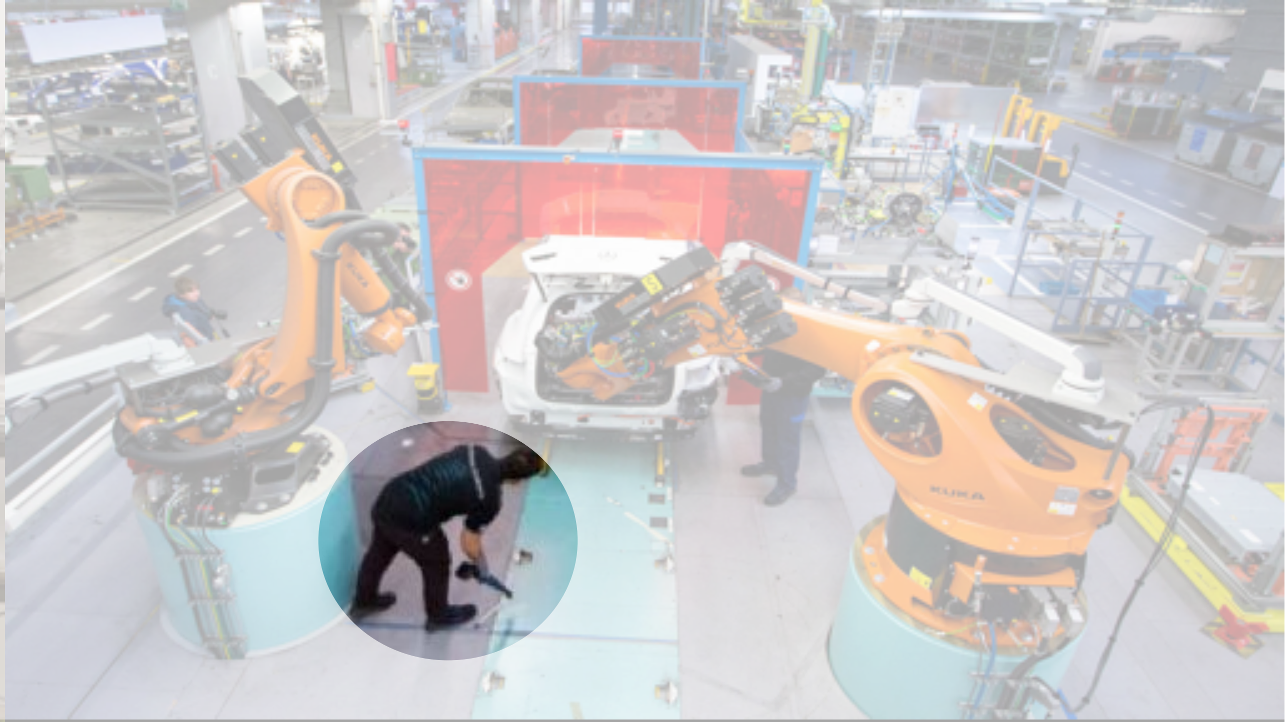
Human Models

- Data-efficient learning of reward functions with different sources of data
- What happens on the ends of the risk spectrum?

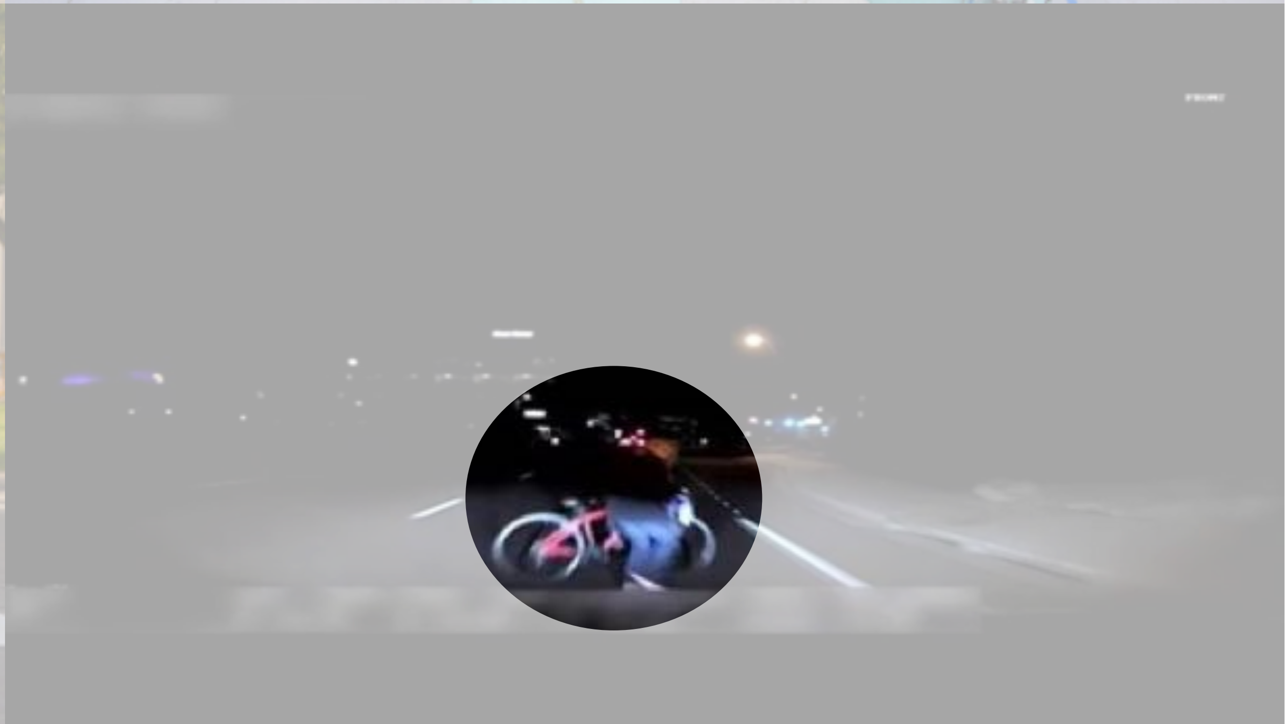

Vehiculo de investigación




KUKA


KUKA


Vehículo de Investigación



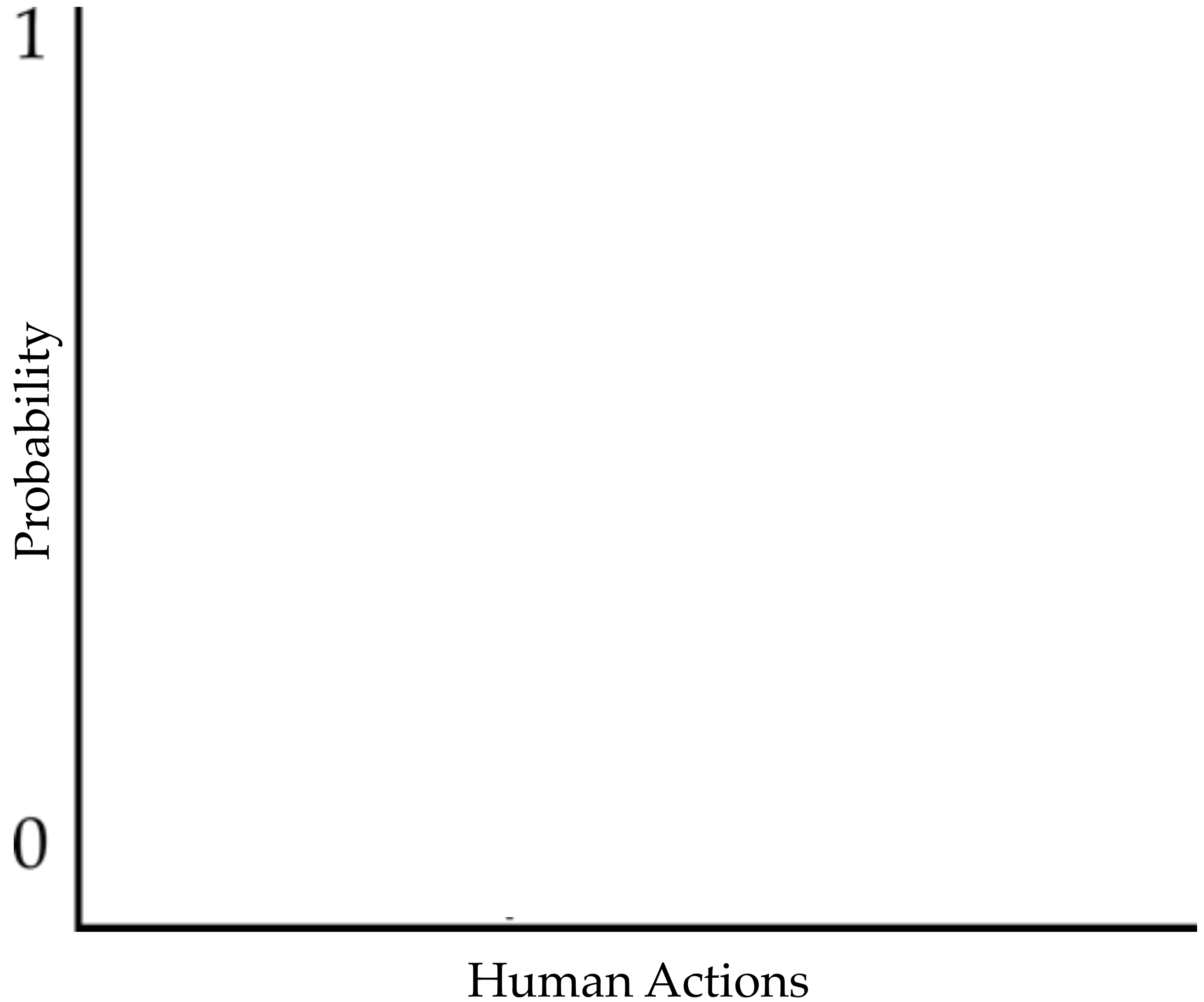

The light turns yellow for the human-driven (blue) car.
Will the blue car pass or stop?

https://www.pe.com/2013/06/09/murrieta-longer-yellow-lights-seen-as-traffic-camera-remedy/

Robots must recognize that people can behave suboptimally in risky scenarios

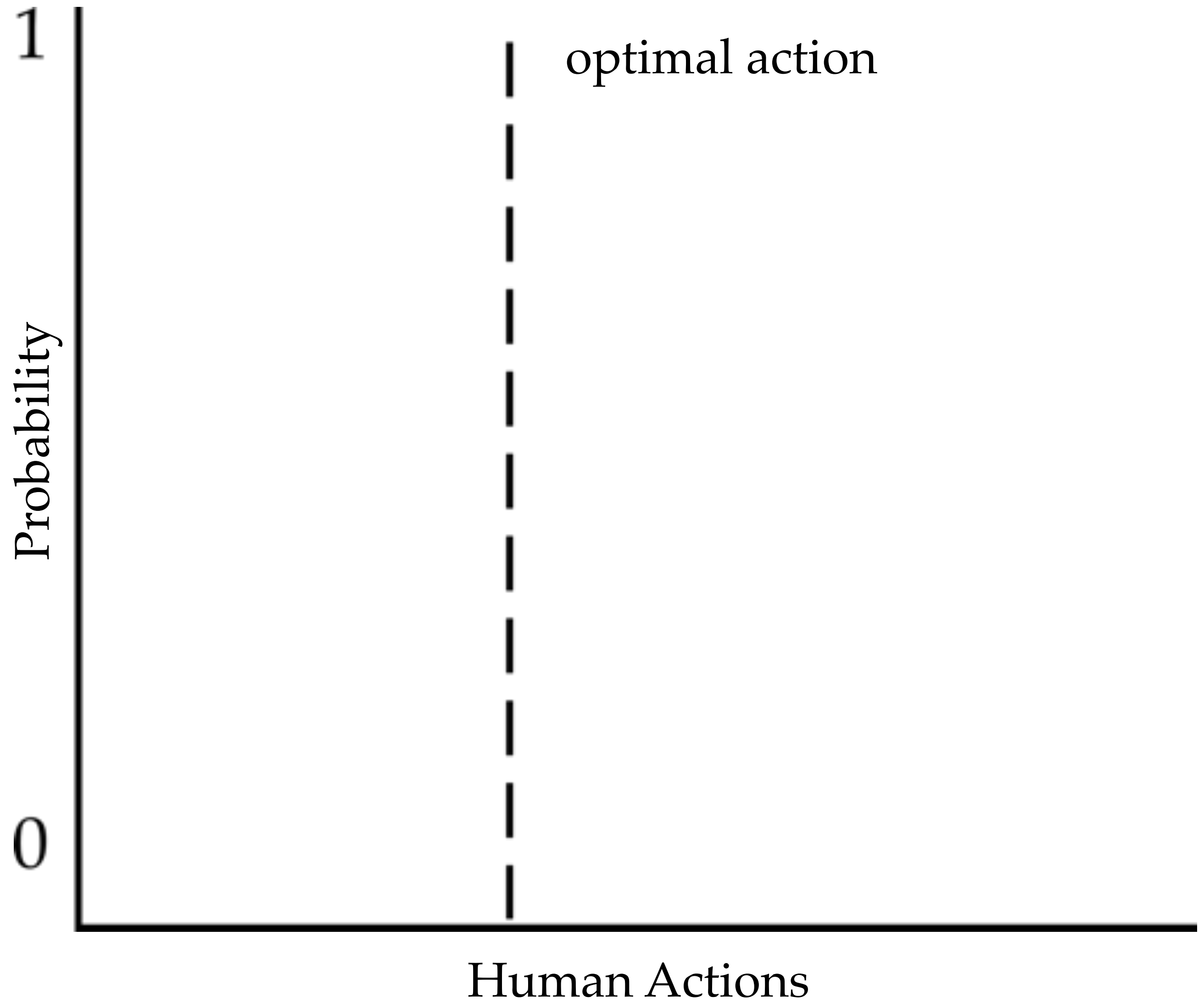
Baseline human models

$$a_H^* = \arg\max_{a_H} R_H(a_H)$$



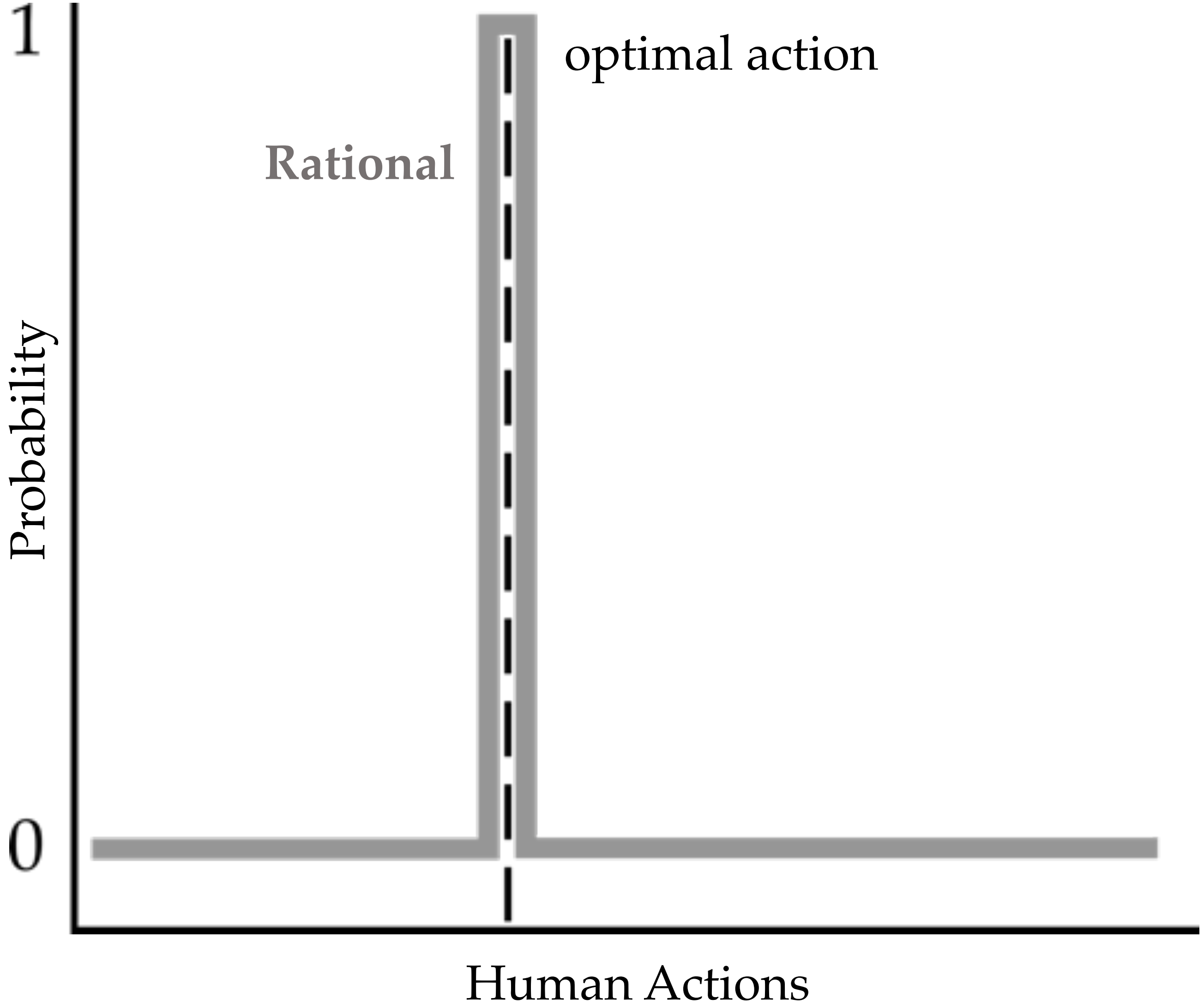
Baseline human models

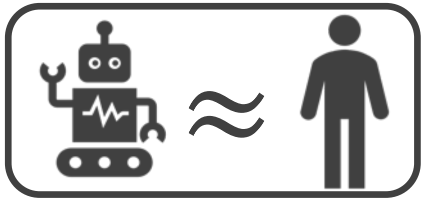
$$a_H^* = \arg\max_{a_H} R_H(a_H)$$



Baseline human models

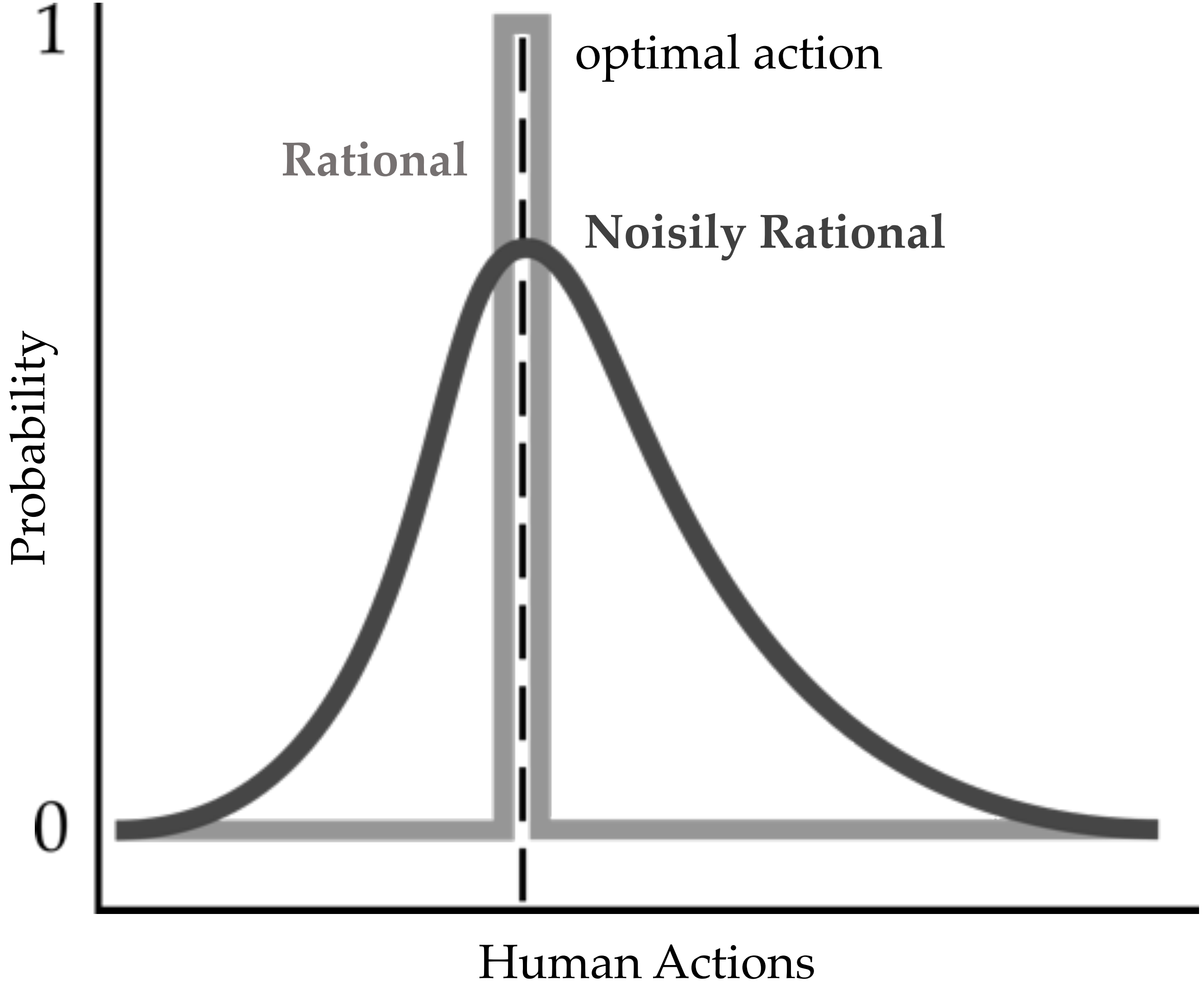
$$a_H^* = \arg\max_{a_H} R_H(a_H)$$





Baseline human models

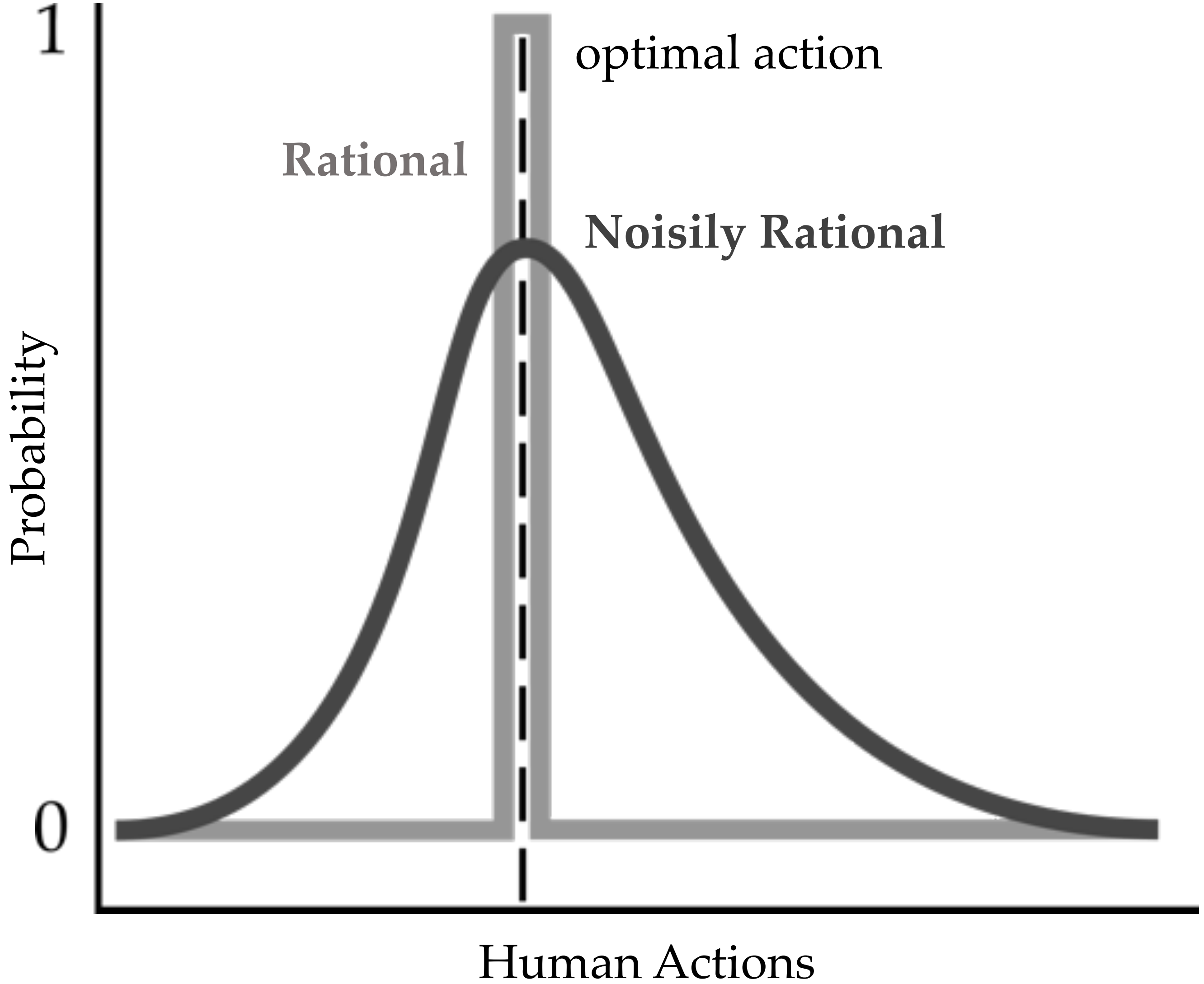
$$P(a_H) = \frac{\exp(\theta R_H(a_H))}{\sum_{a \in A_H} \exp(\theta R_H(a))}$$



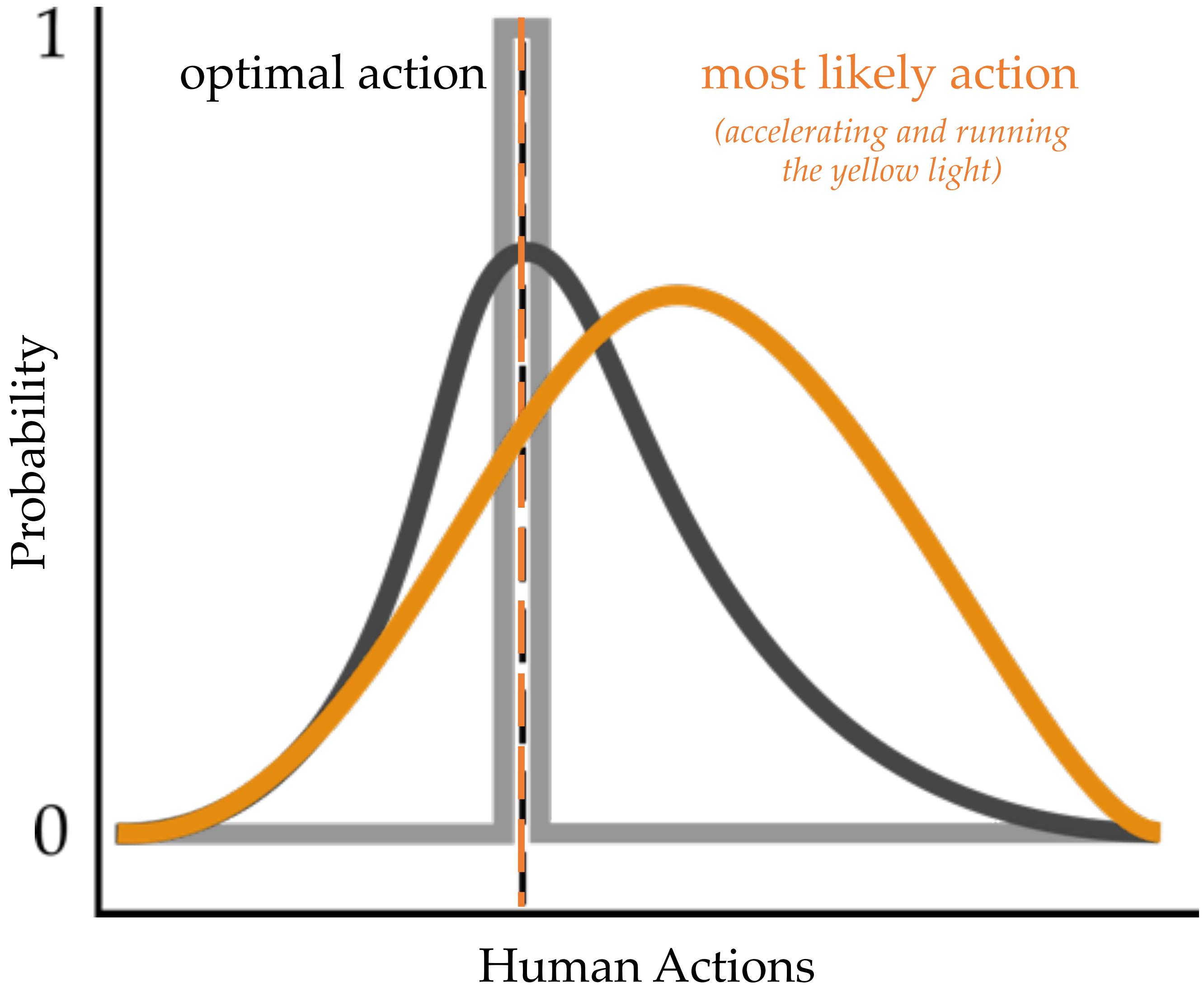
Baseline human models

rationality coefficient

$$P(a_H) = \frac{\exp(\theta R_H(a_H))}{\sum_{a \in A_H} \exp(\theta R_H(a))}$$

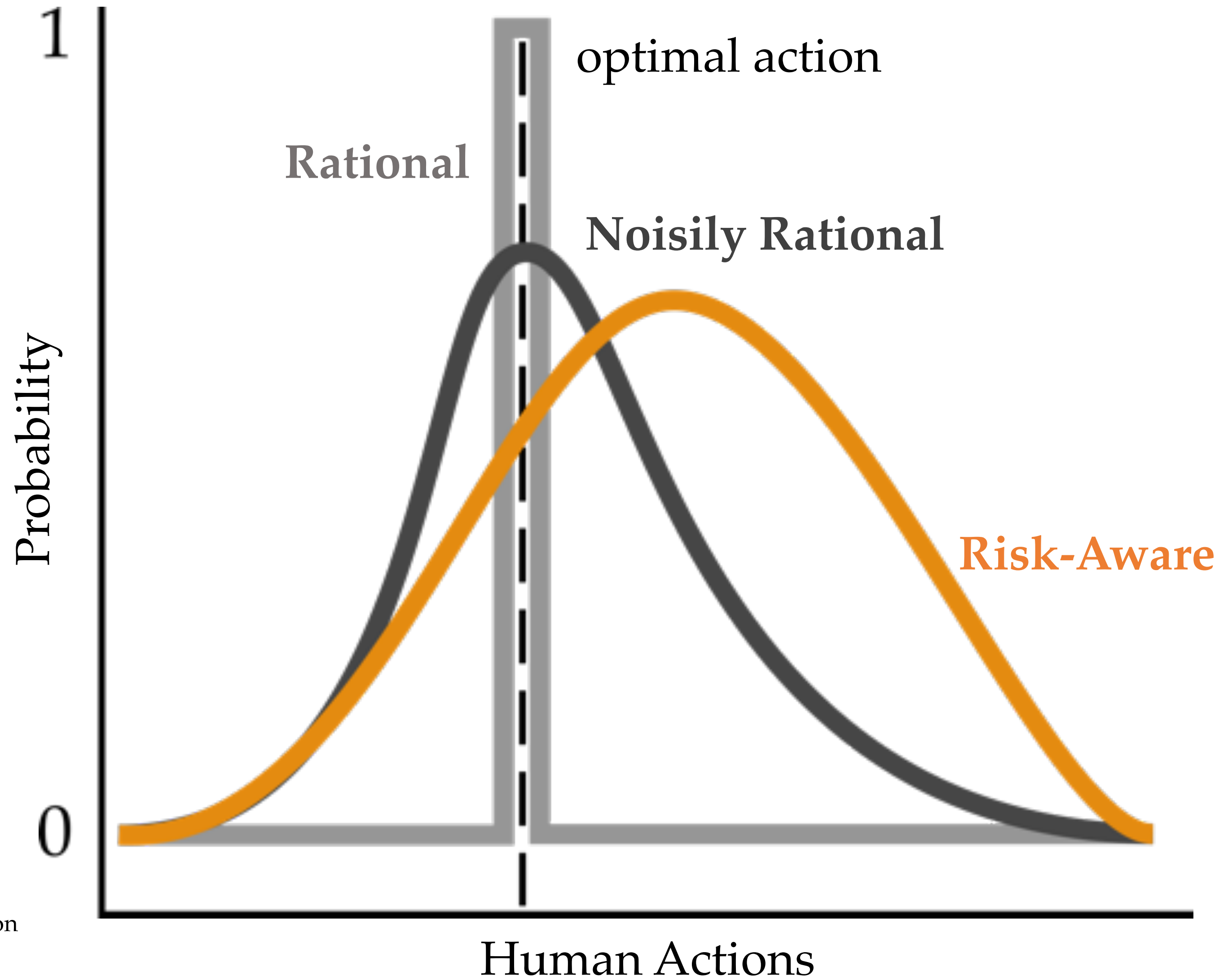


Baseline human models



Risk-aware model: Cumulative Prospect Theory

$$P(a_H) = \frac{\exp(\theta R_H^{CPT}(a_H))}{\sum_{a \in A_H} \exp(\theta R_H^{CPT}(a))}$$



Amos Tversky and Daniel Kahneman, "Advances in prospect theory: Cumulative representation of uncertainty," Journal of Risk and Uncertainty 1992.

Risk-aware model: Cumulative Prospect Theory

$$R_H(a_H) = p^{(1)} R_H^{(1)}(a_H) + \cdots + \; p^{(k)} R_H^{(k)}(a_H)$$

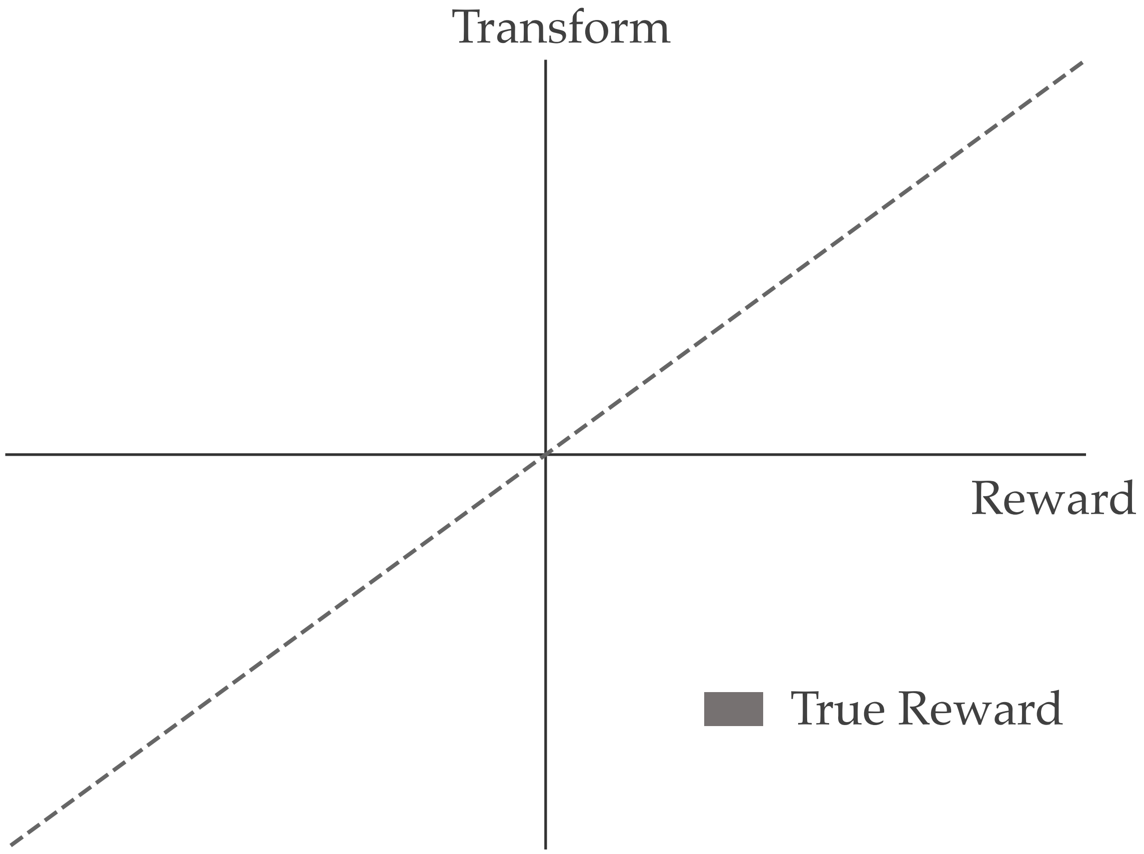
Risk-aware model: Cumulative Prospect Theory

$$R_H^{CPT}(a_H) = p^{(1)} R_H^{(1)}(a_H) + \cdots + p^{(k)} R_H^{(k)}(a_H)$$

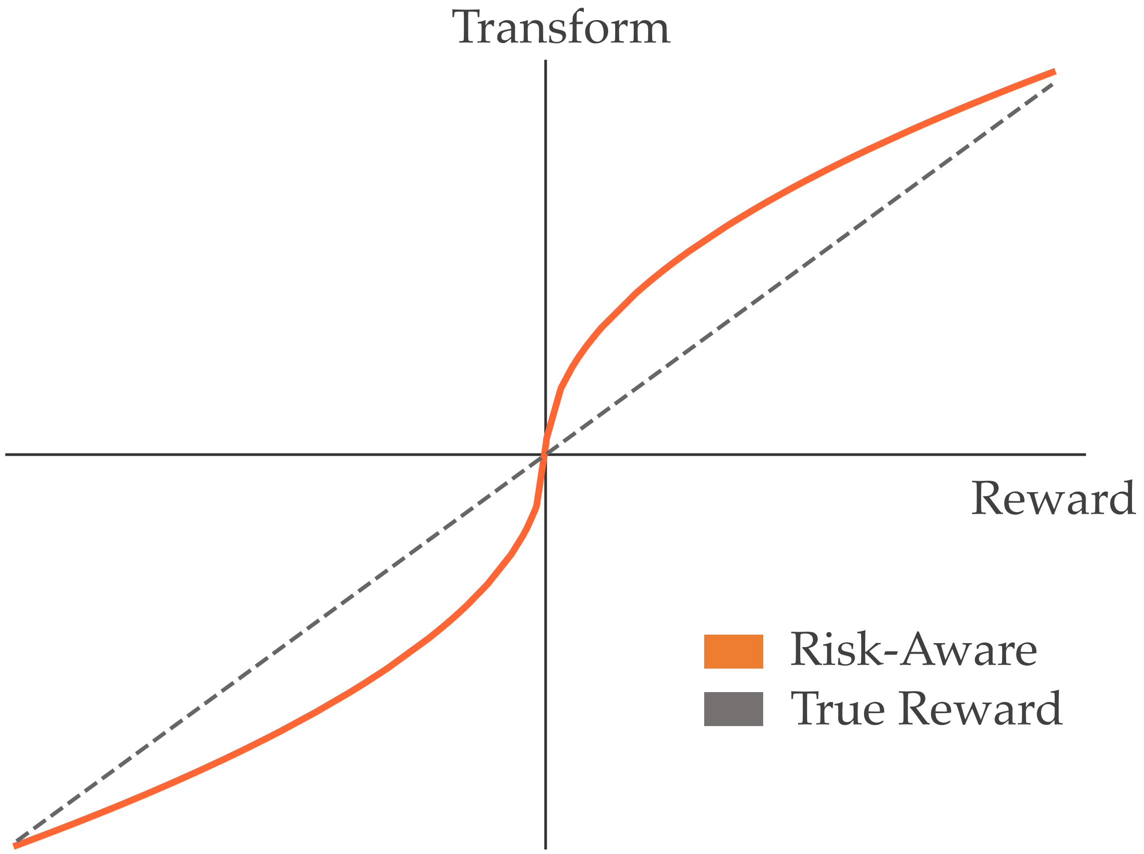
Risk-aware model: Cumulative Prospect Theory

$$R_H^{CPT}(a_H) = p^{(1)} R_H^{(1)}(a_H) + \cdots + p^{(k)} R_H^{(k)}(a_H)$$

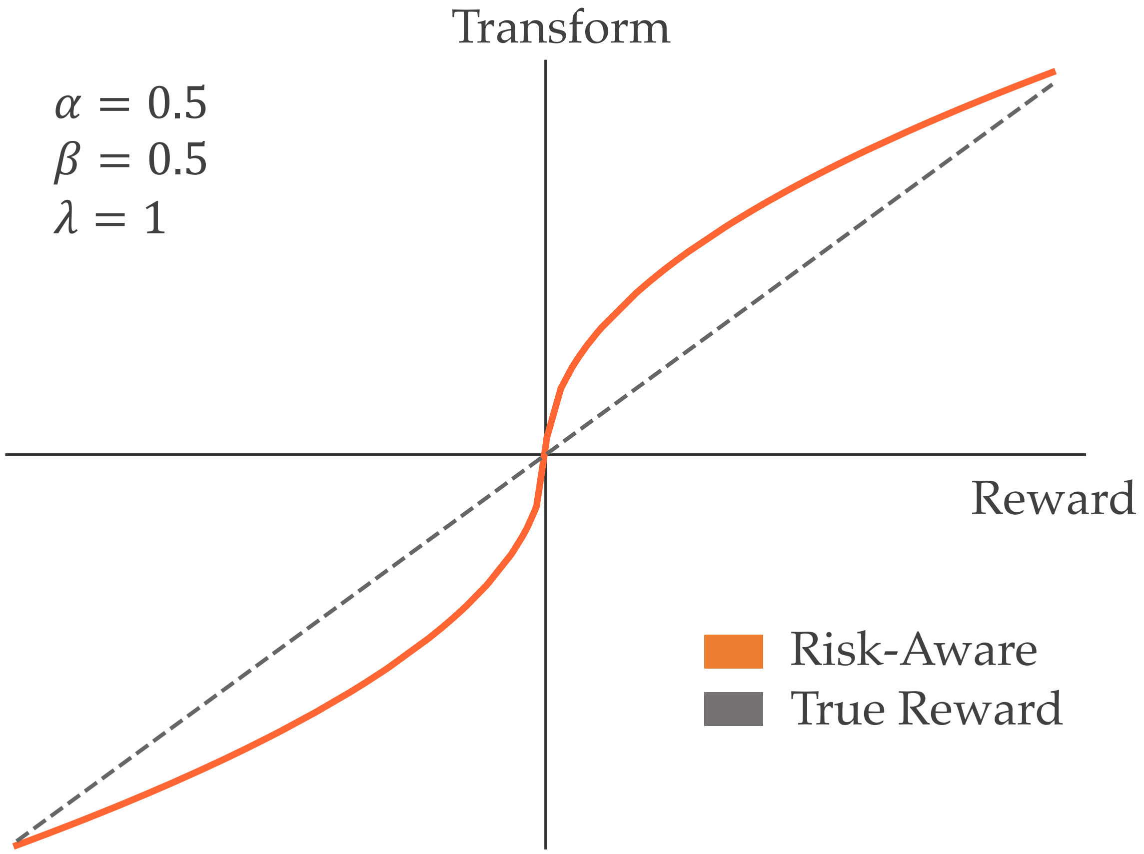
Risk-aware model: Cumulative Prospect Theory



Risk-aware model: Cumulative Prospect Theory



Risk-aware model: Cumulative Prospect Theory

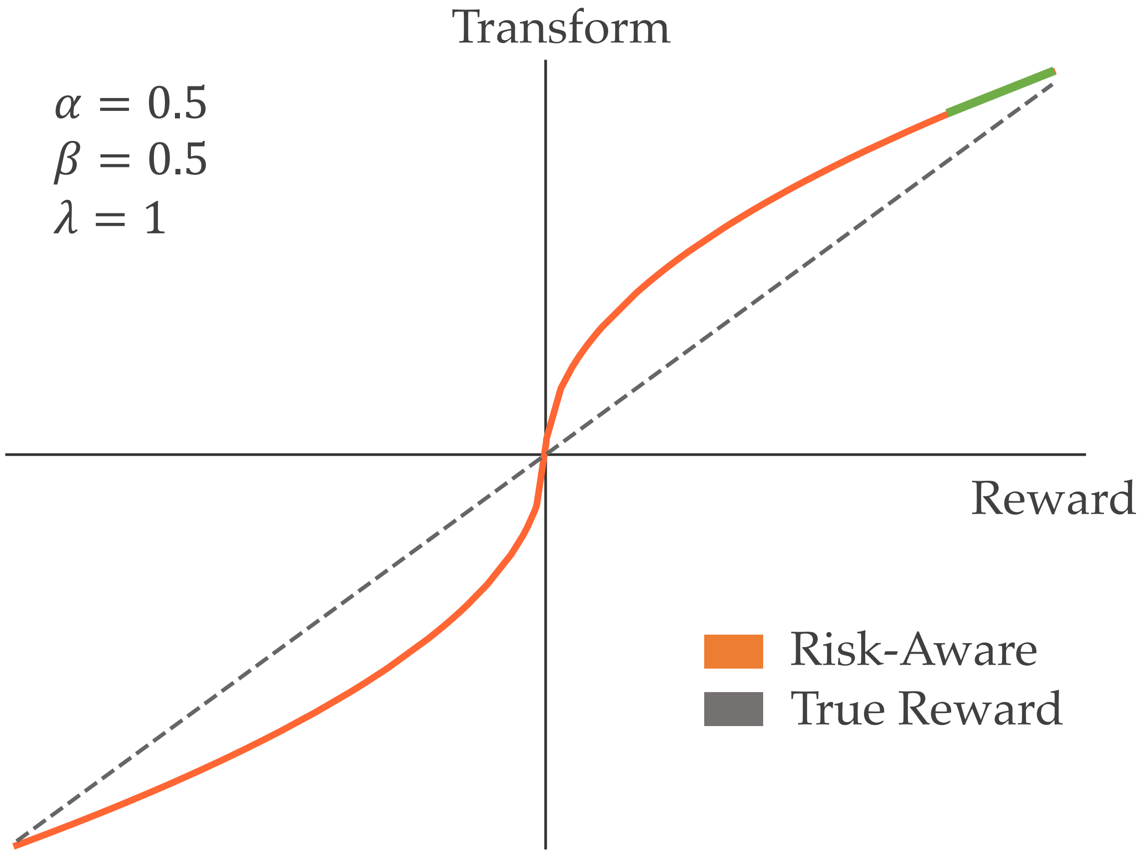


$$v(R) = \begin{cases} R^{\alpha} & , R \geq 0 \\ -\lambda(-R)^{\beta} & , R < 0 \end{cases}$$

$\alpha, \beta \in [0,1]$

$\lambda \in [0, \infty)$

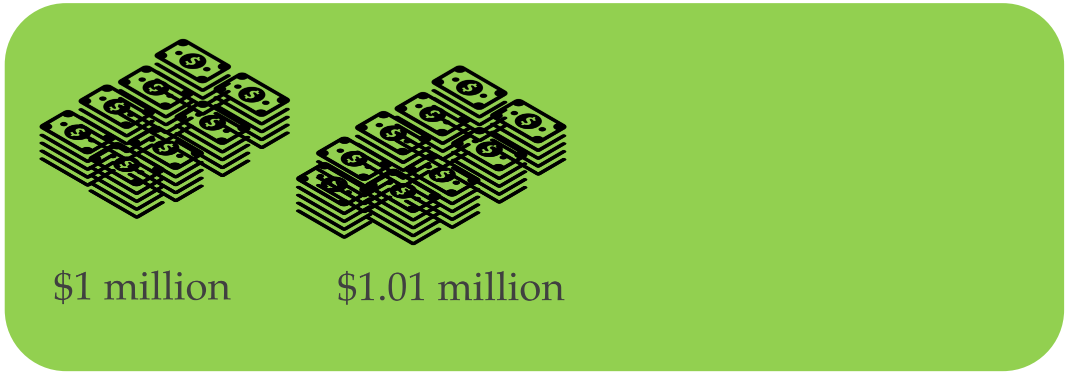
Risk-aware model: Cumulative Prospect Theory



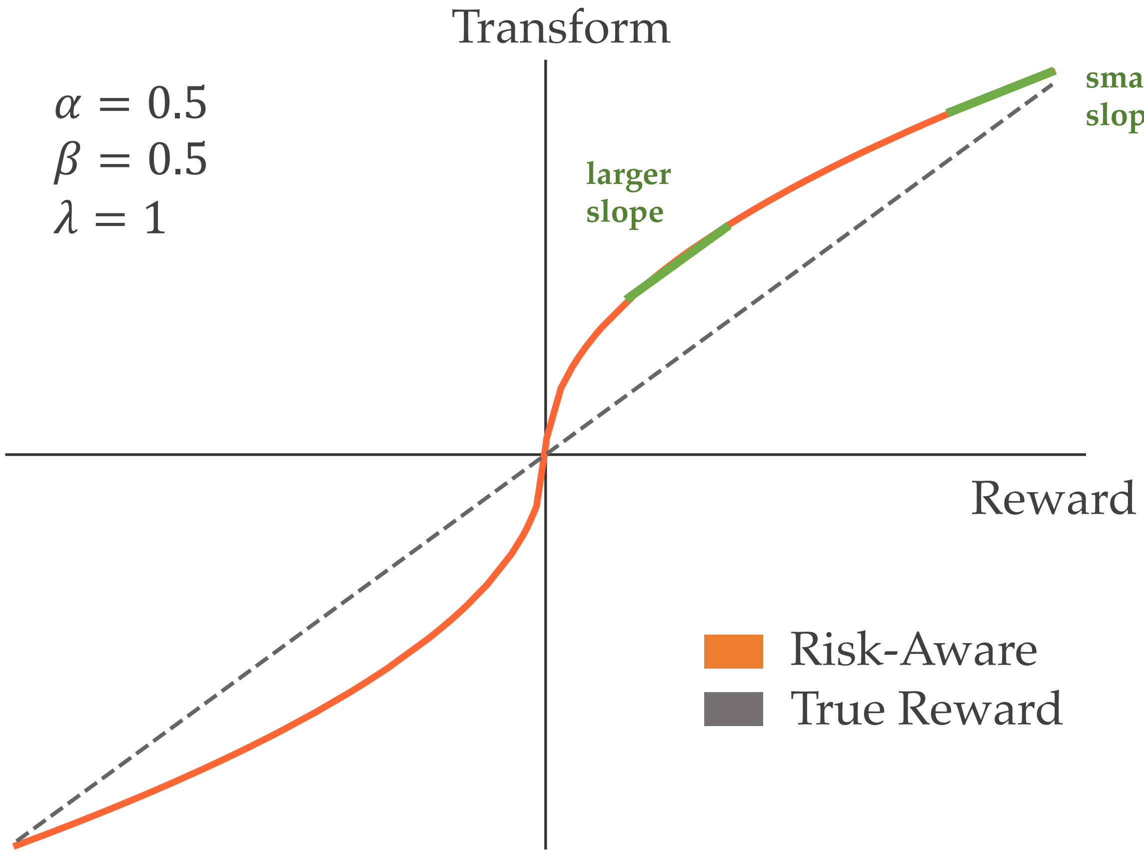
$$v(R) = \begin{cases} R^{\alpha} & , R \geq 0 \\ -\lambda(-R)^{\beta} & , R < 0 \end{cases}$$

$\alpha, \beta \in [0,1]$

$\lambda \in [0, \infty)$



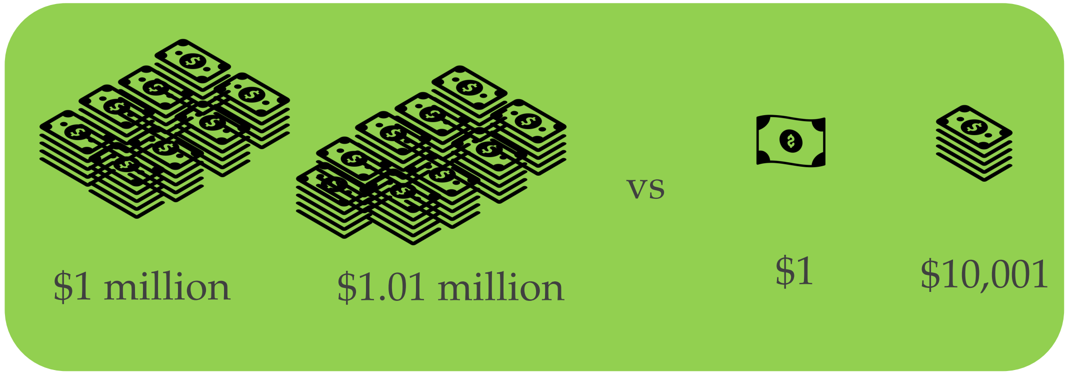
Risk-aware model: Cumulative Prospect Theory



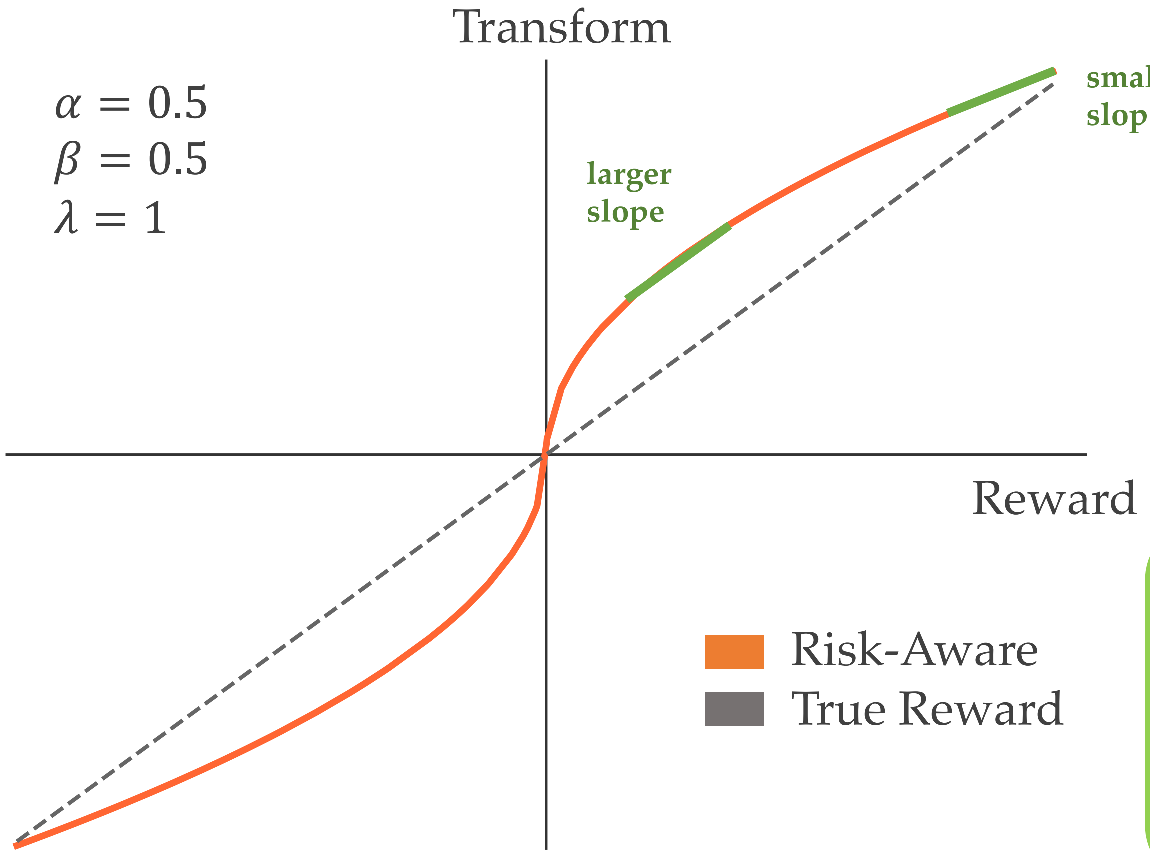
$$v(R) = \begin{cases} R^{\alpha} & , R \geq 0 \\ -\lambda(-R)^{\beta} & , R < 0 \end{cases}$$

$\alpha, \beta \in [0,1]$

$\lambda \in [0, \infty)$



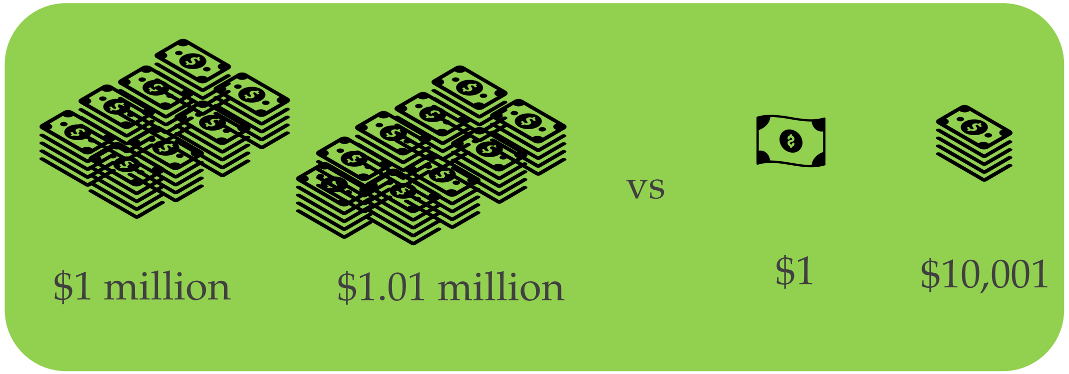
Risk-aware model: Cumulative Prospect Theory



$$v(R) = \begin{cases} R^{\alpha} & , R \geq 0 \\ -\lambda(-R)^{\beta} & , R < 0 \end{cases}$$

$\alpha, \beta \in [0,1]$

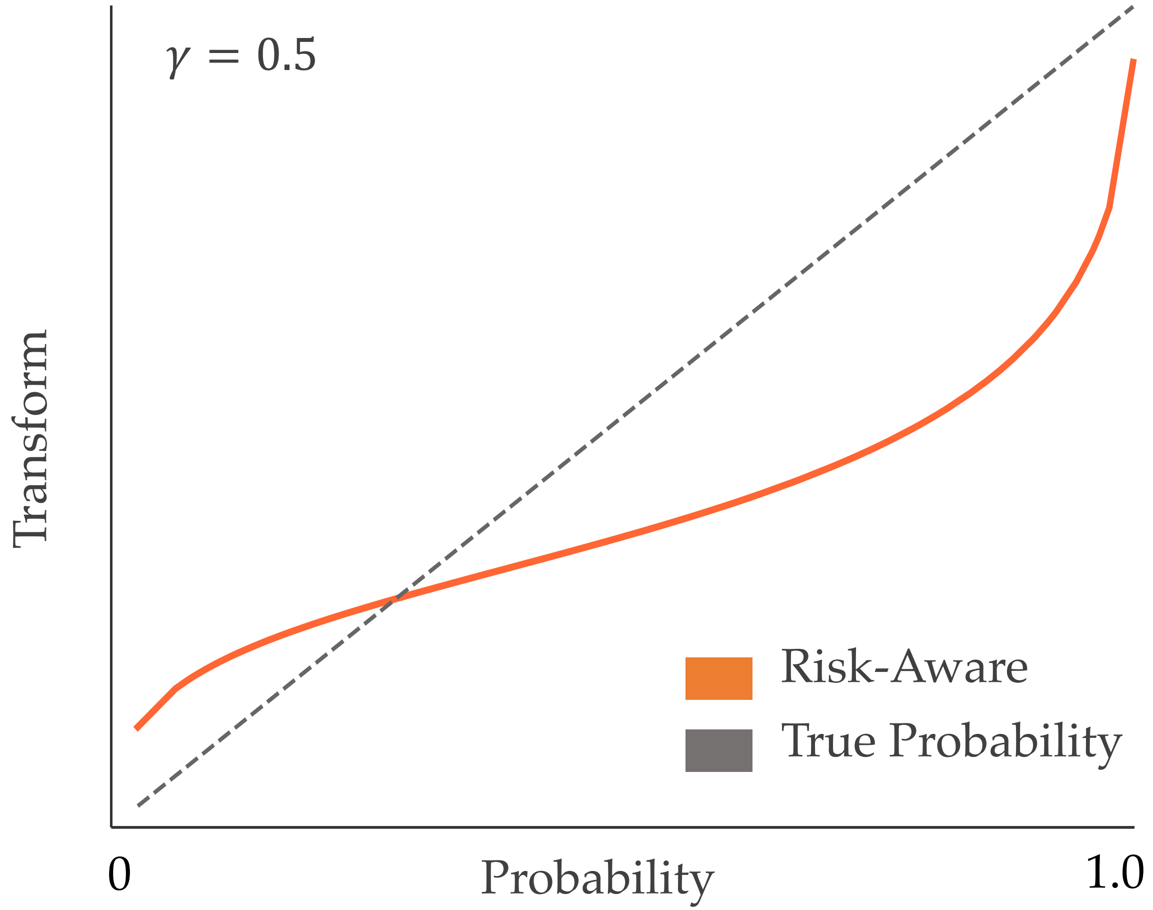
$\lambda \in [0, \infty)$



Risk-aware model: Cumulative Prospect Theory

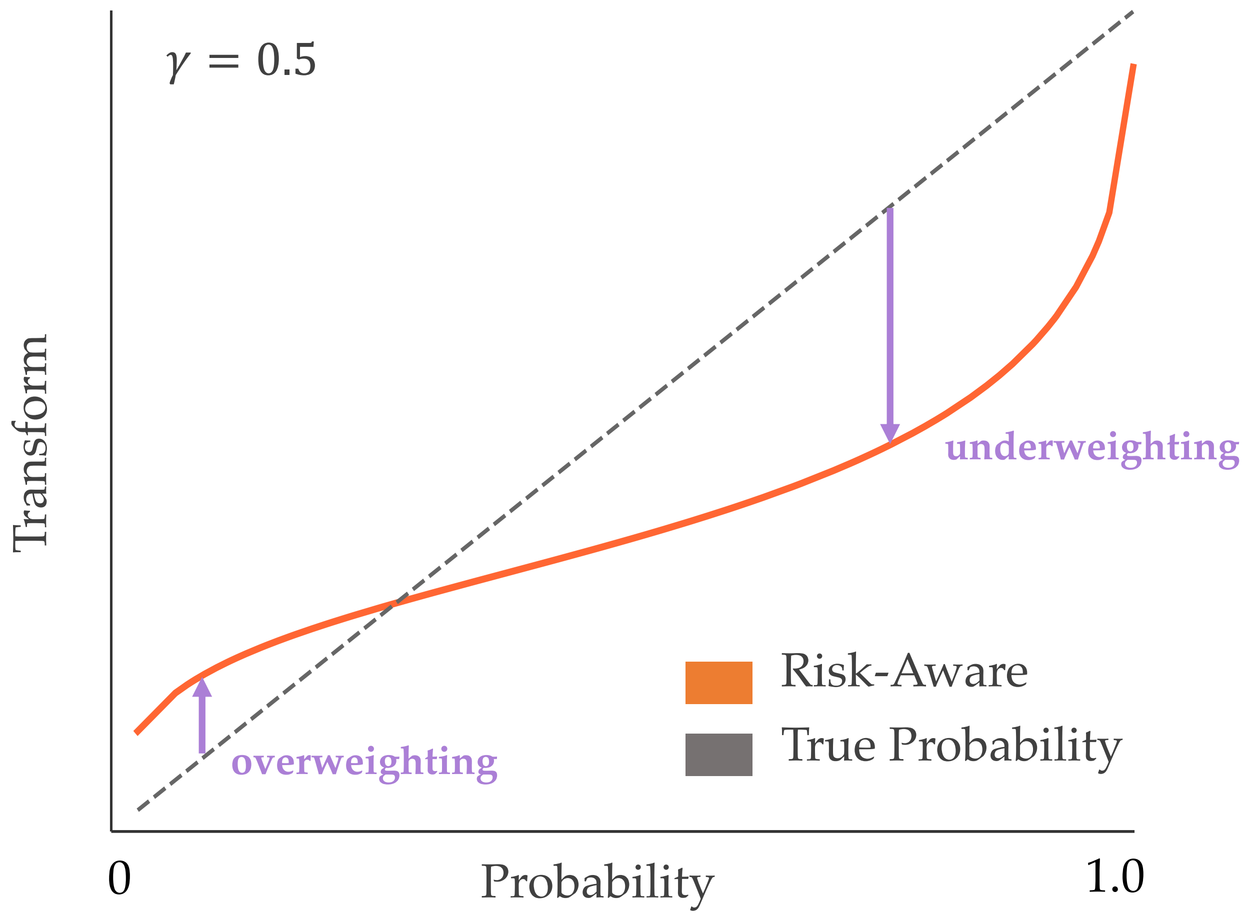
$$R_H^{CPT}(a_H) = p^{(1)} R_H^{(1)}(a_H) + \cdots + p^{(k)} R_H^{(k)}(a_H)$$

Risk-aware model: Cumulative Prospect Theory



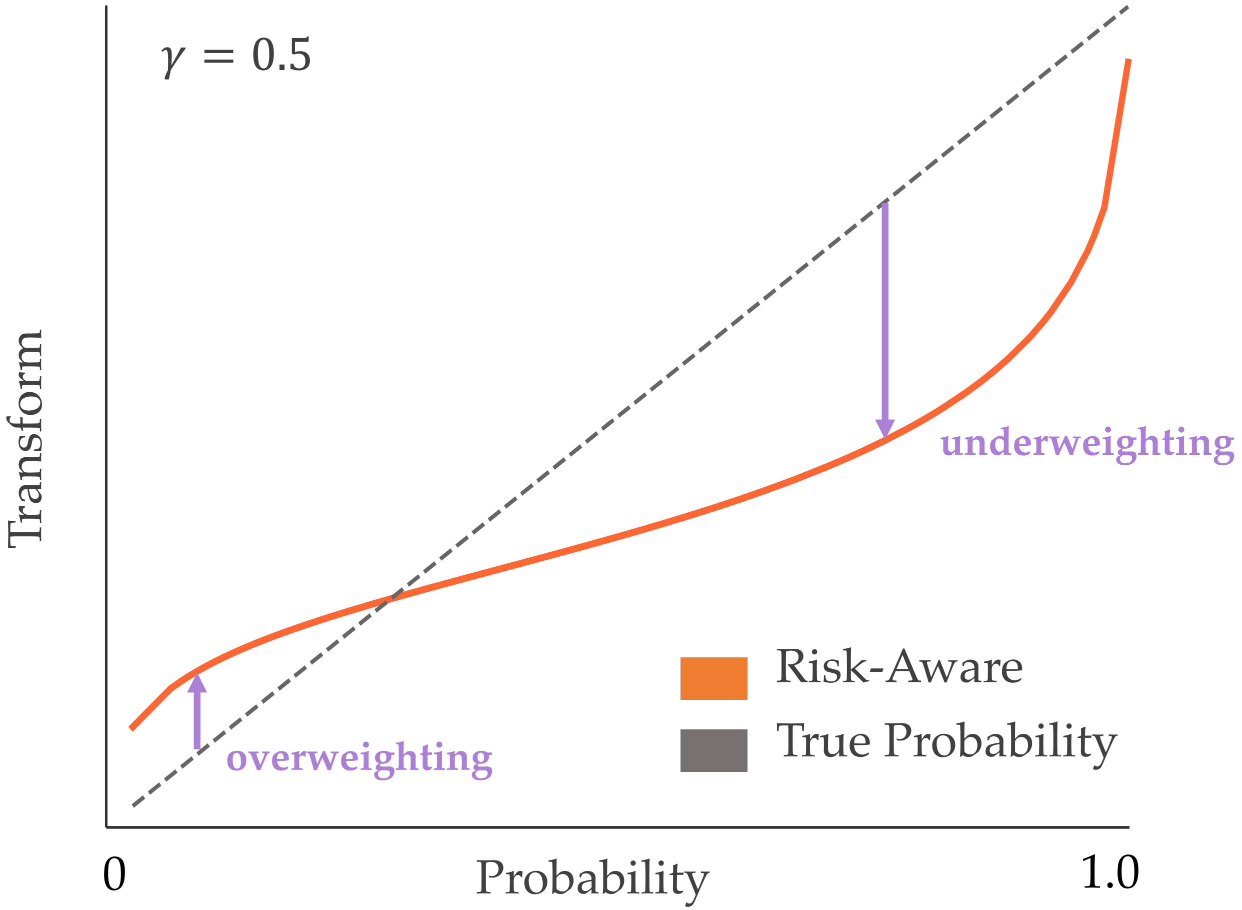
$$w(p) = \frac{p^{\gamma}}{(p^{\gamma} + (1-p)^{\gamma})^{1/\gamma}} \qquad \gamma \in [0,1]$$

Risk-aware model: Cumulative Prospect Theory

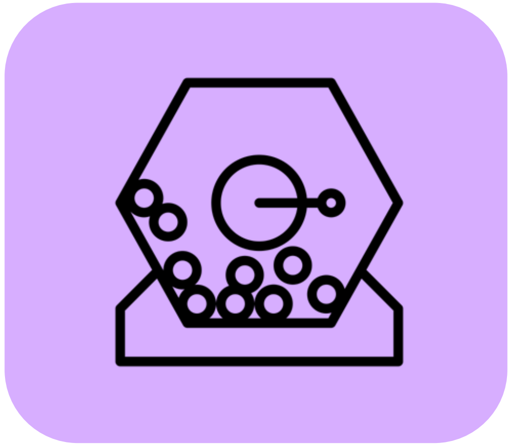


$$w(p) = \frac{p^\gamma}{(p^\gamma + (1-p)^\gamma)^{1/\gamma}} \qquad \gamma \in [0,1]$$

Risk-aware model: Cumulative Prospect Theory



$$w(p) = \frac{p^{\gamma}}{(p^{\gamma} + (1-p)^{\gamma})^{1/\gamma}} \qquad \gamma \in [0,1]$$

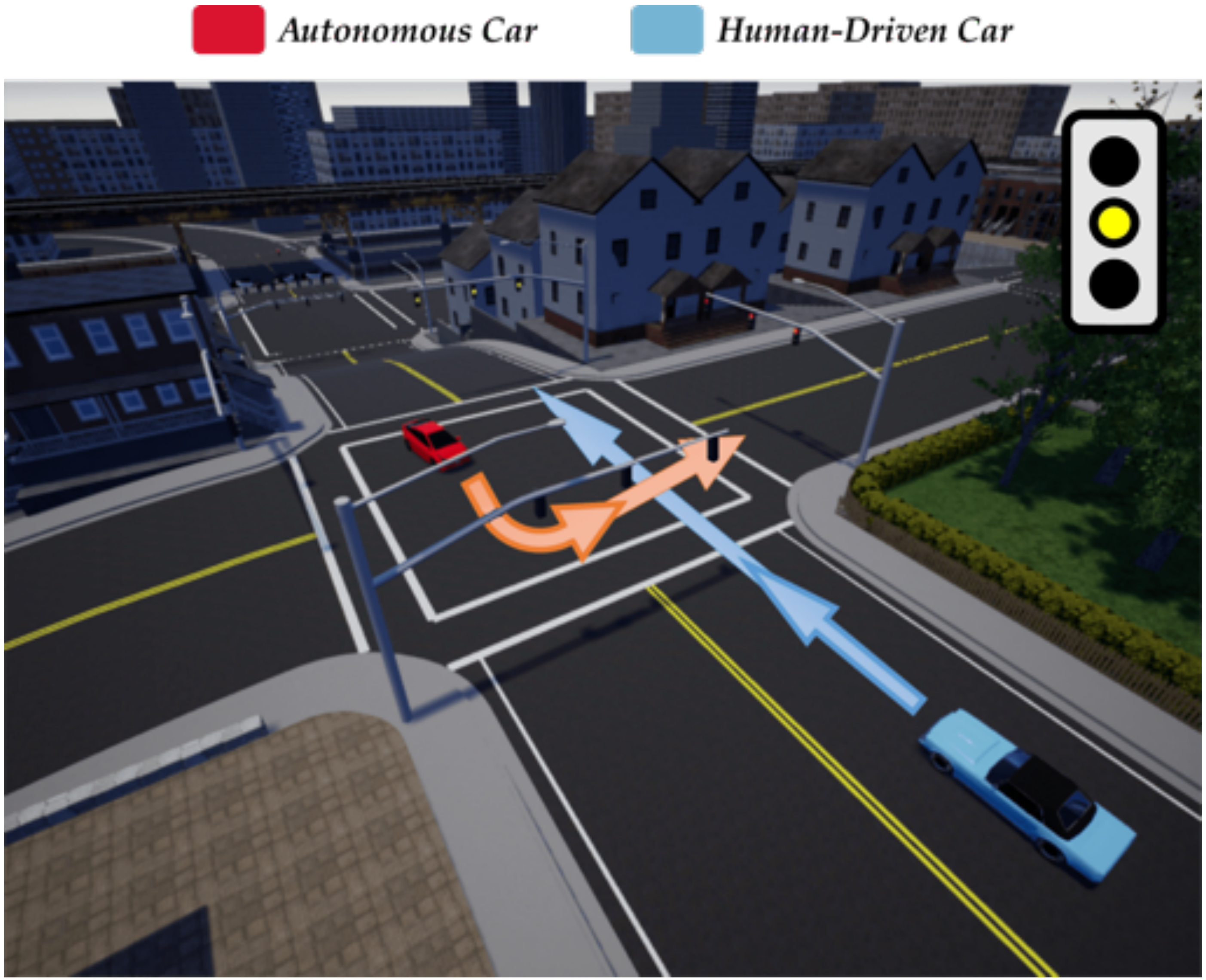


Risk-aware model: Cumulative Prospect Theory

$$R_H^{CPT}(a_H) = p^{(1)} R_H^{(1)}(a_H) + \cdots + p^{(k)} R_H^{(k)}(a_H)$$

When do we behave suboptimally?

Autonomous driving task



Study results

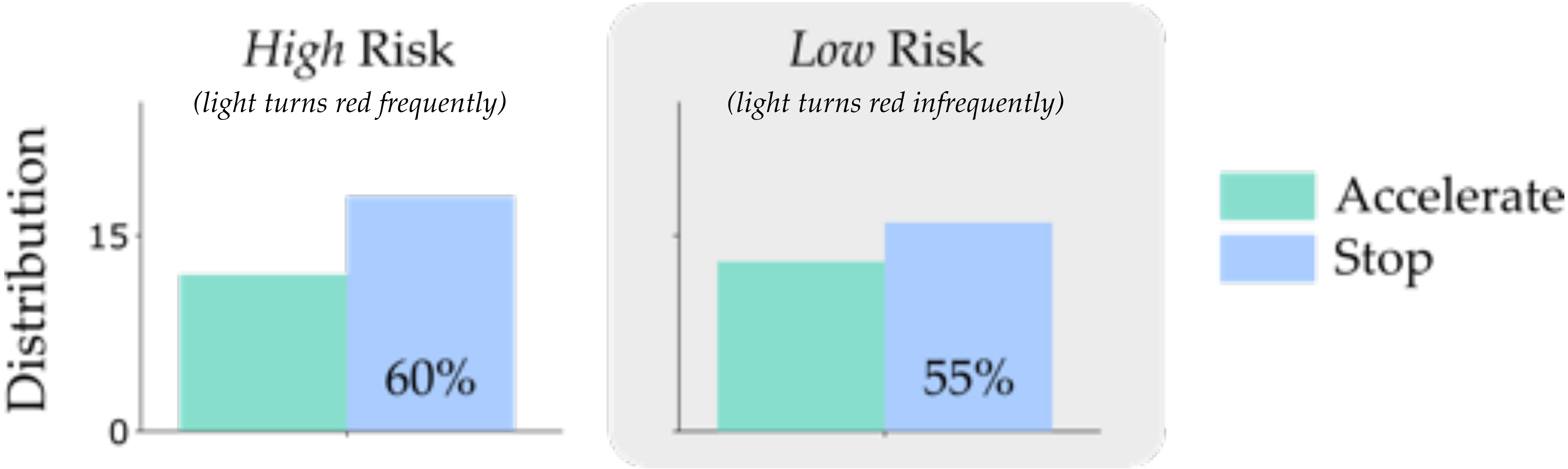
Study results

N=30

High Risk: light turns red 95% of time
Low Risk: light turns red 5% of time

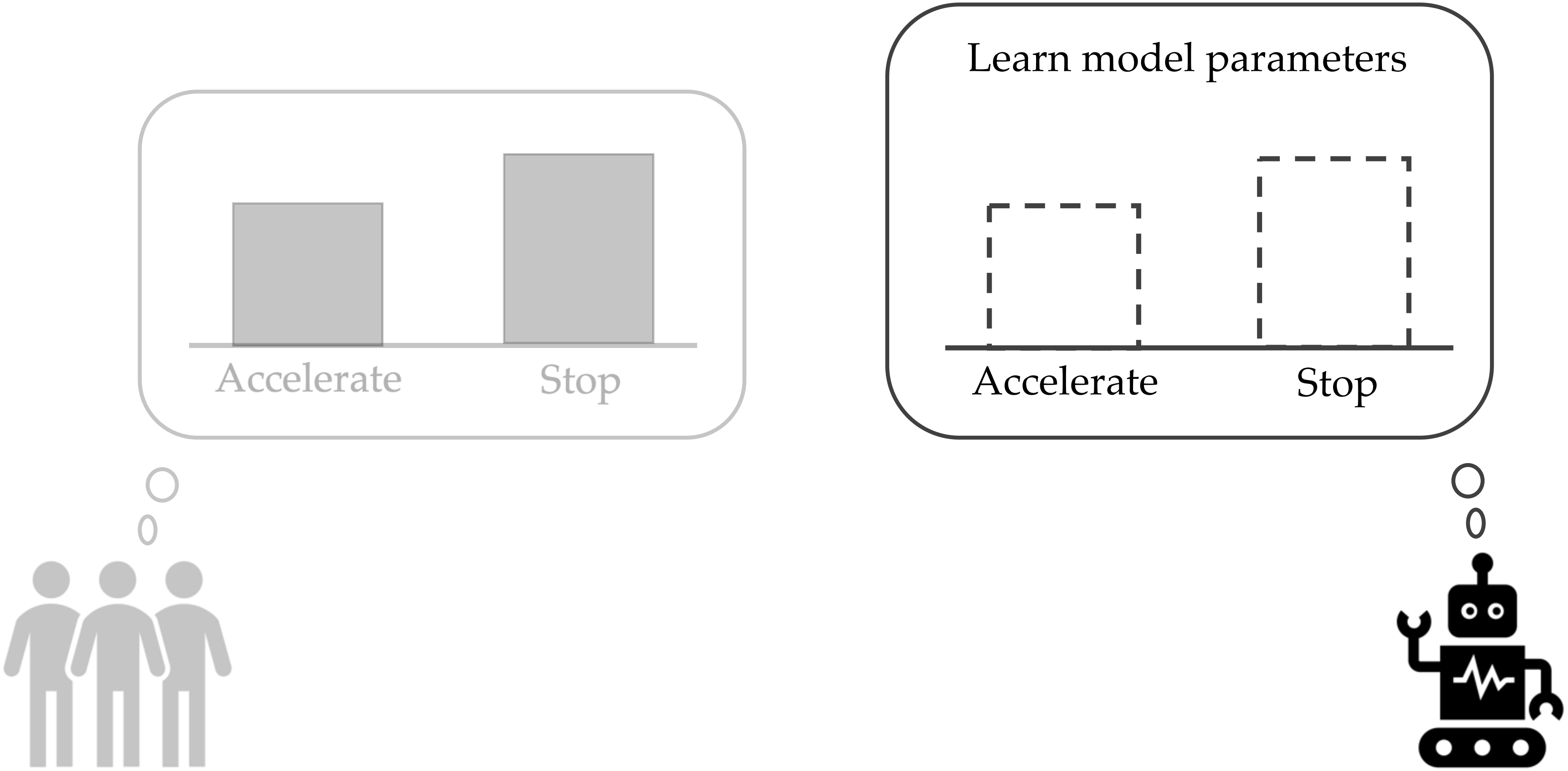
Study results

N=30



Majority of people preferred the suboptimal action!

Experiment



Modeling results

Modeling results

N=30

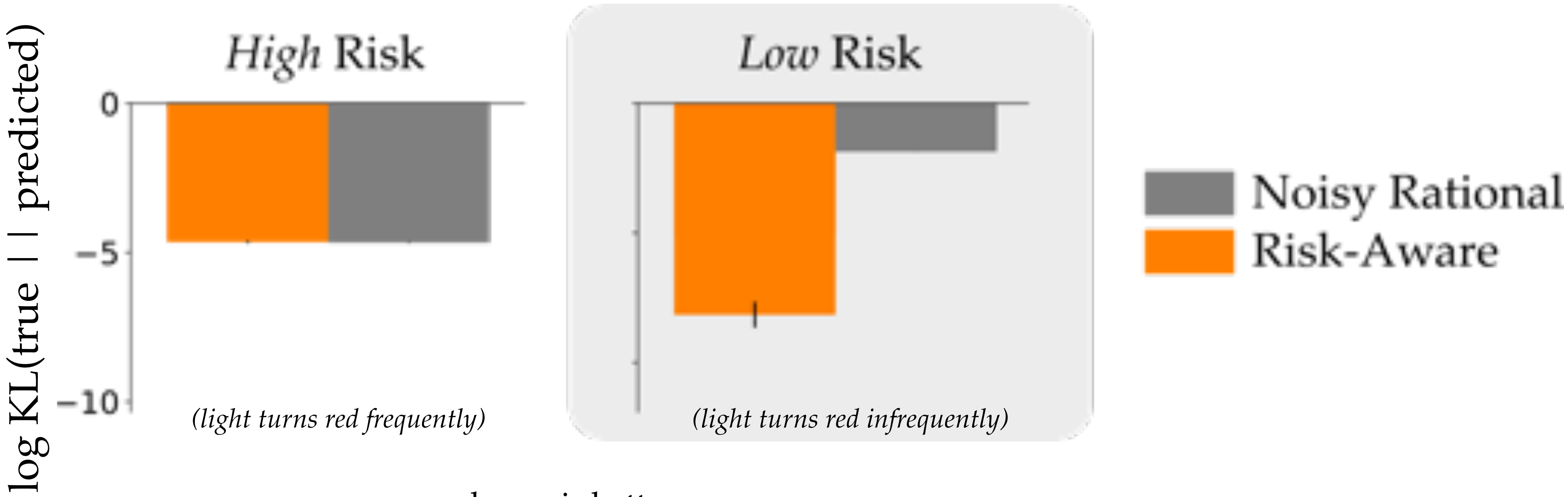
log KL(true || predicted)

(light turns red frequently)

lower is better

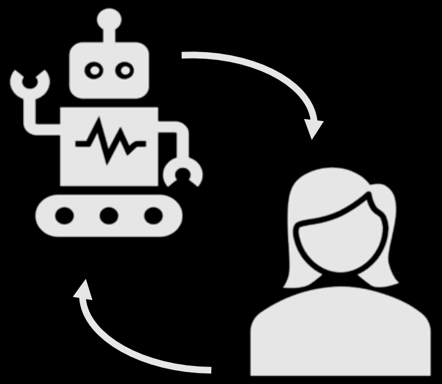
Modeling results

N=30



lower is better

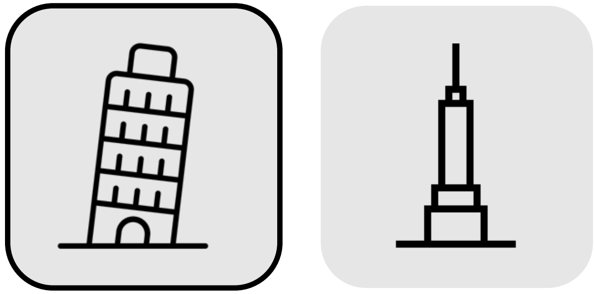
Robots that plan with risk-aware models



Collaborative cup stacking task

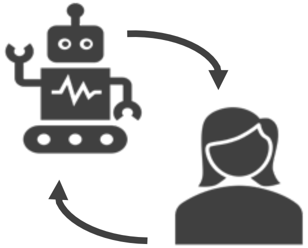
Collaborative cup stacking task

Efficient but unstable tower

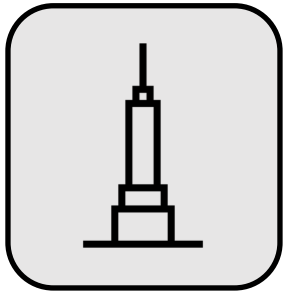


- Awarded 105 points
- Remains upright 20% of the time

Collaborative cup stacking task



Inefficient but stable tower



- Awarded 20 points
- Never falls

Collaborative cup stacking task

Efficient but unstable tower

- Awarded 105 points
- Remains upright 20% of the time

105 * 0.2 = 21

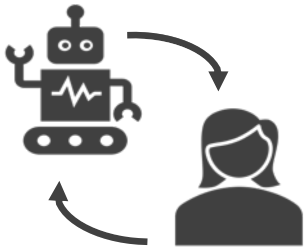
>

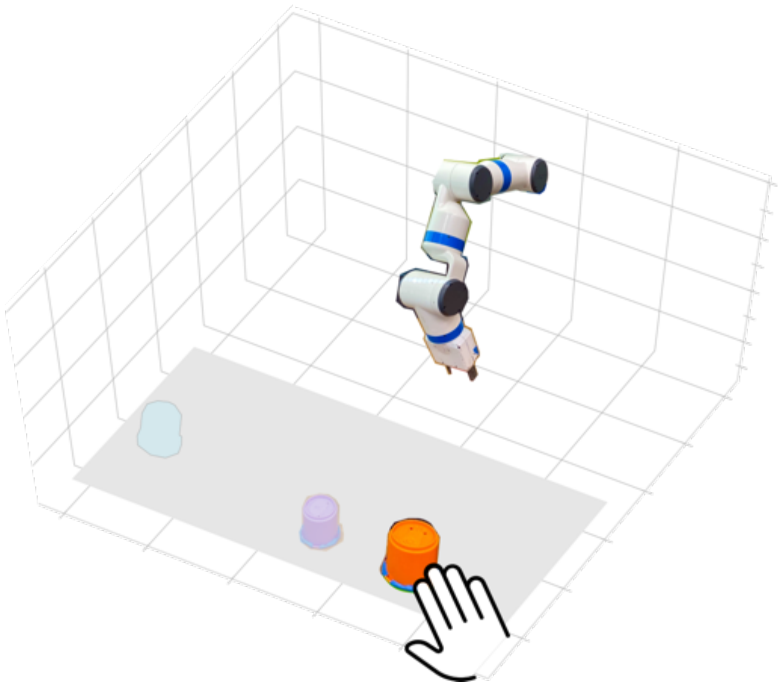
Inefficient but stable tower

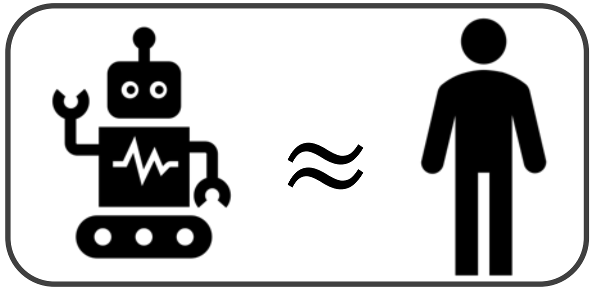
- Awarded 20 points
- Never falls

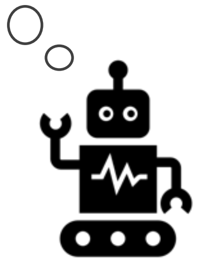
20

Noisily rational robot

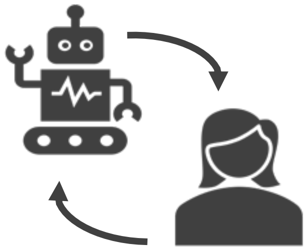


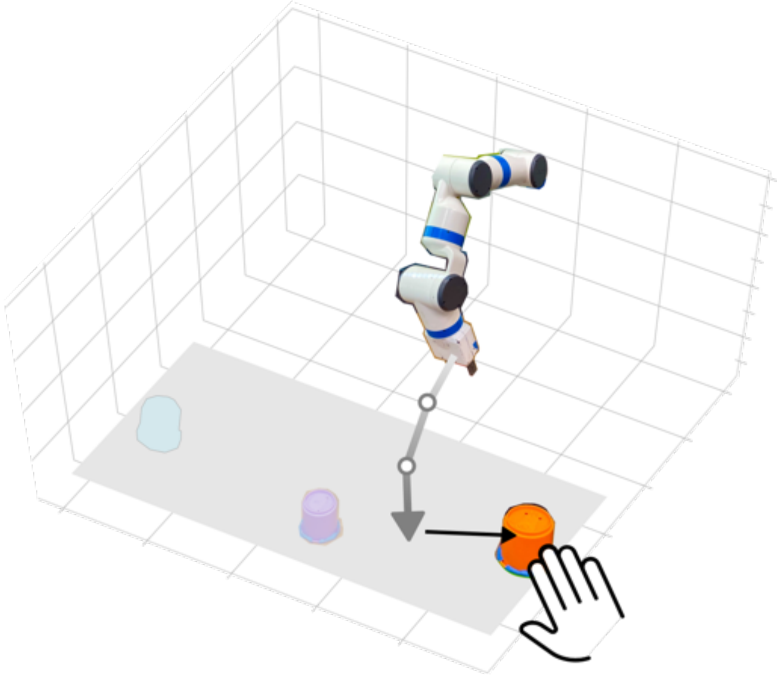


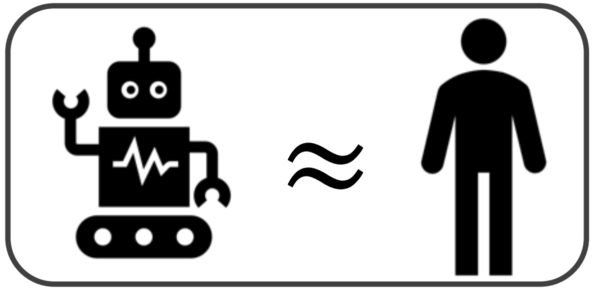


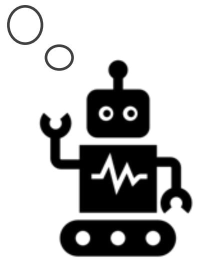


Noisily rational robot

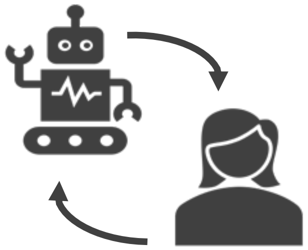


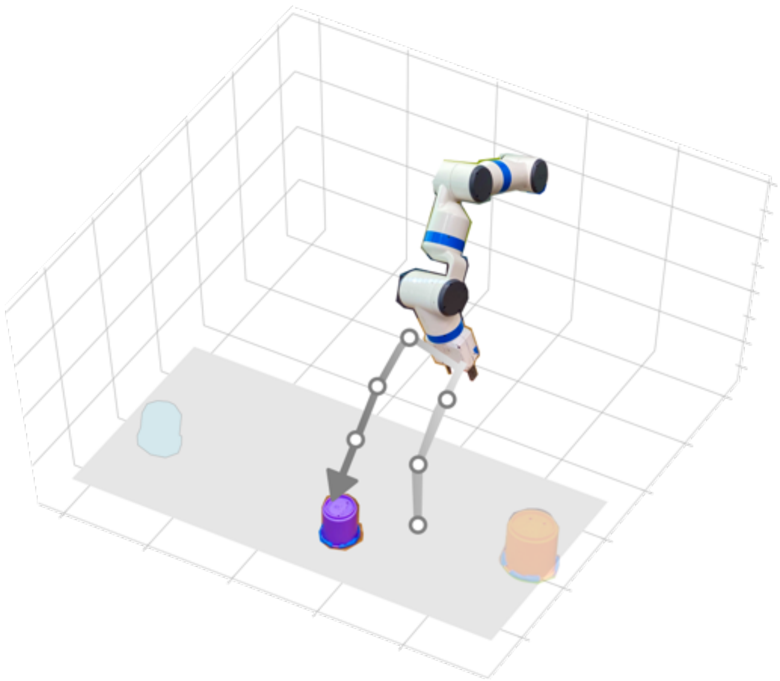


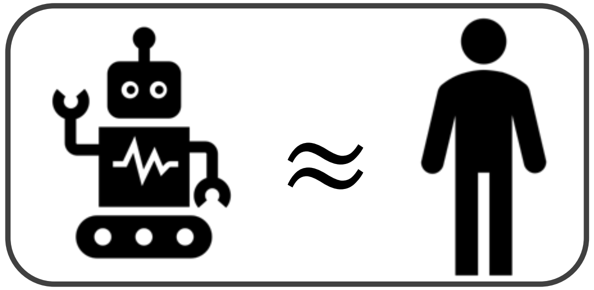


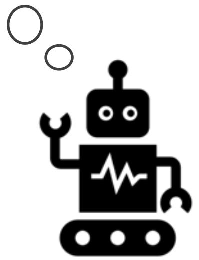


Noisily rational robot

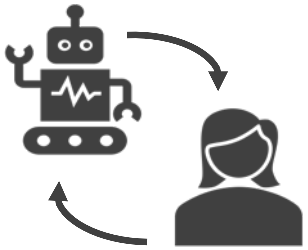


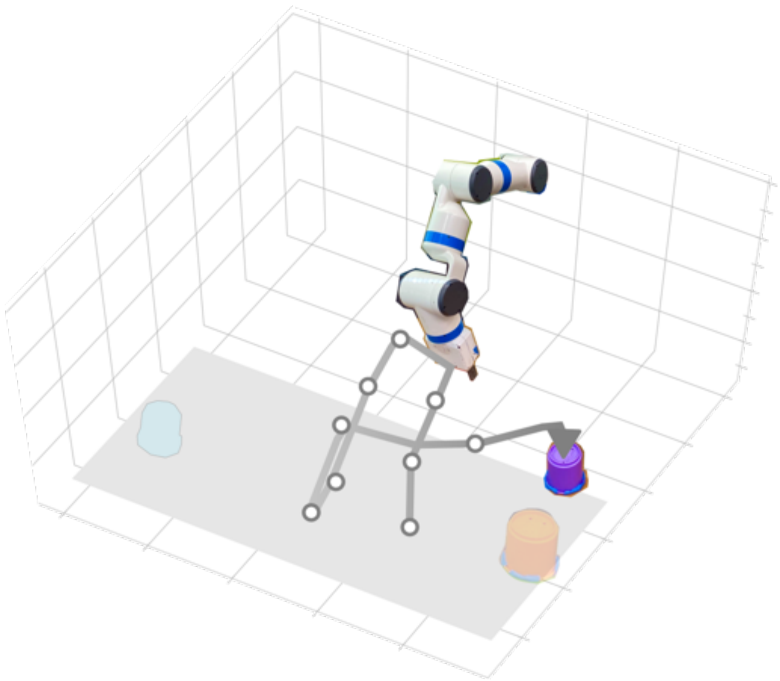


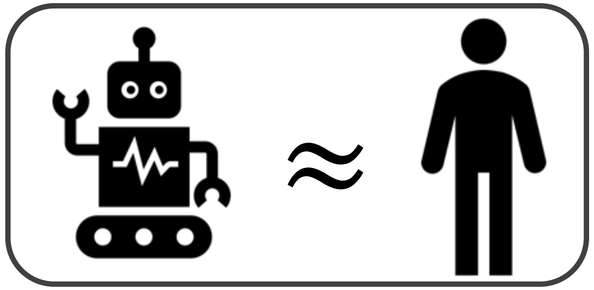


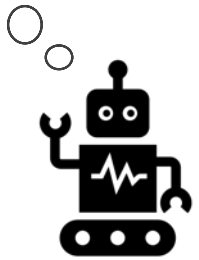


Noisily rational robot

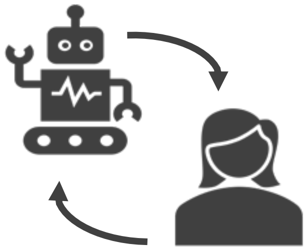


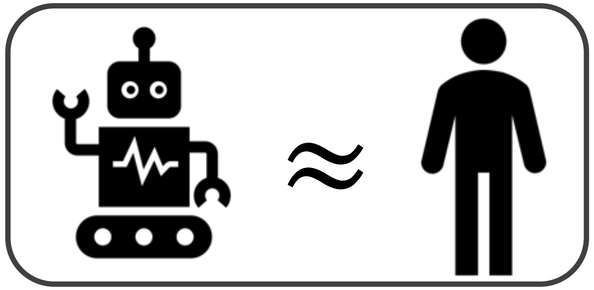


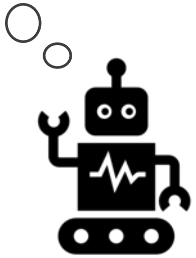




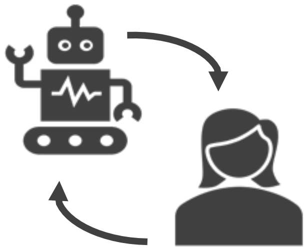
Noisily rational robot

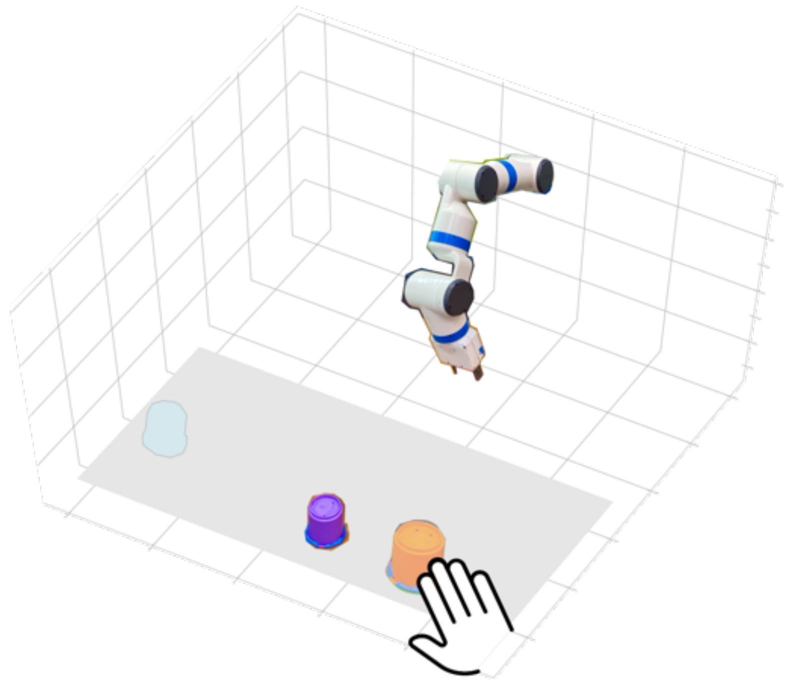


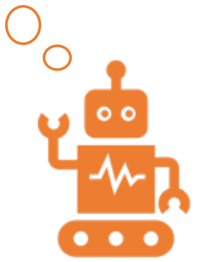


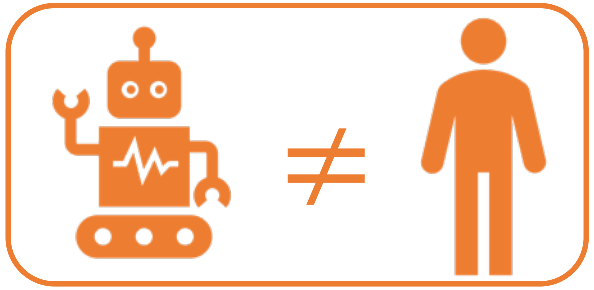


Risk-aware robot

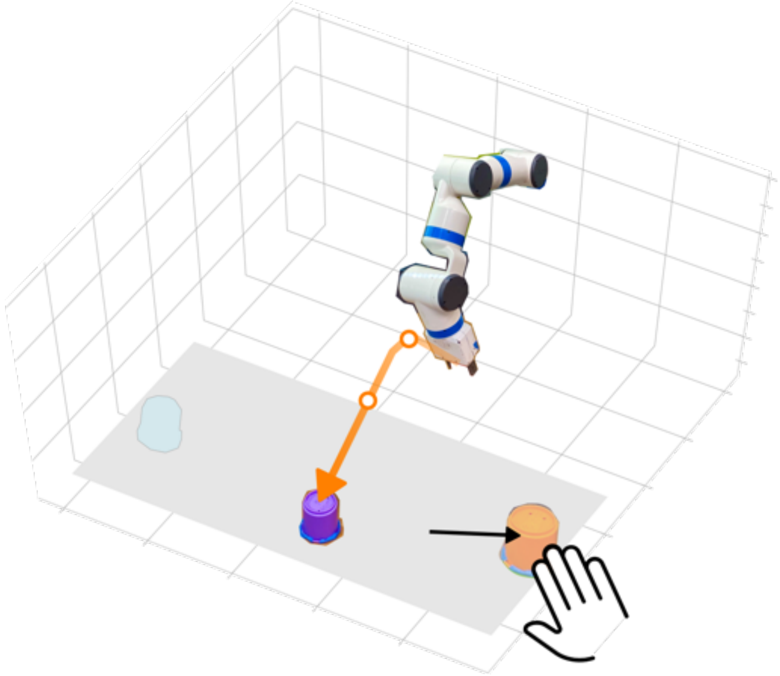


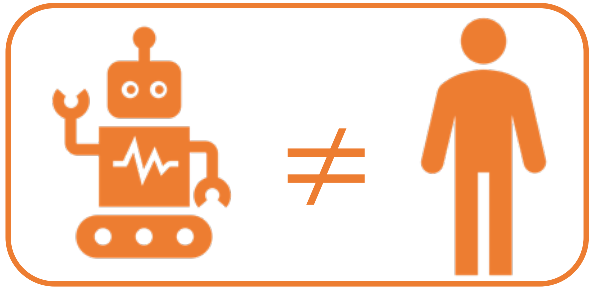




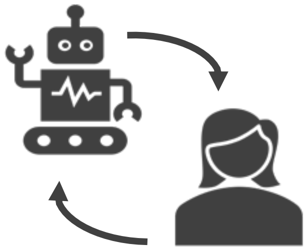


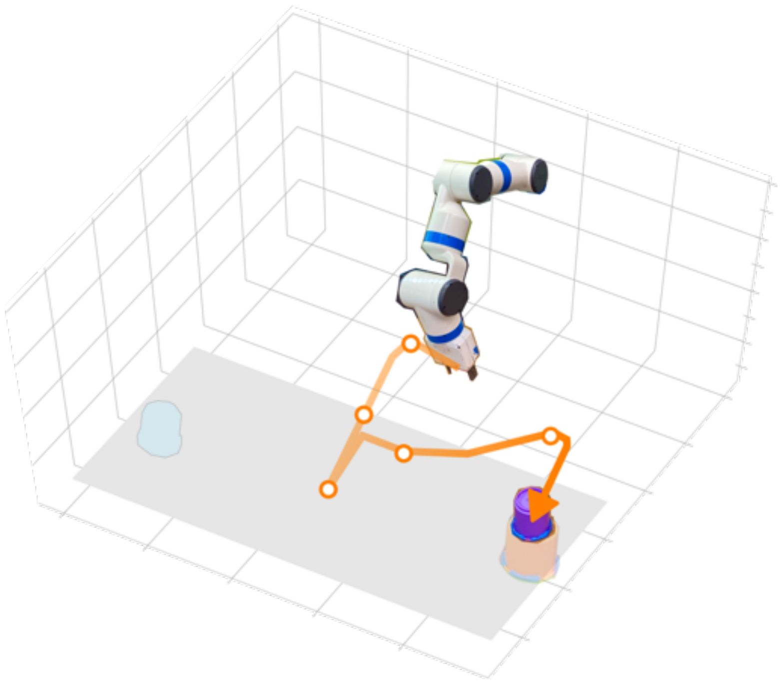
Risk-aware robot

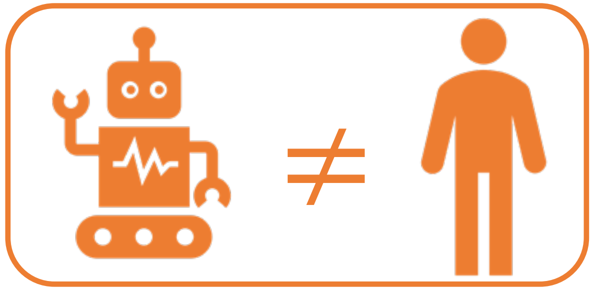


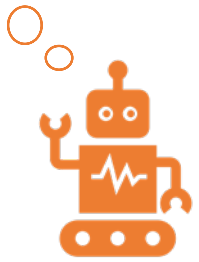


Risk-aware robot

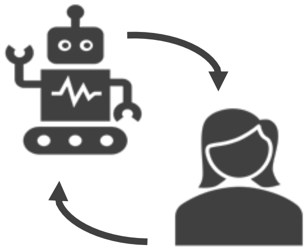


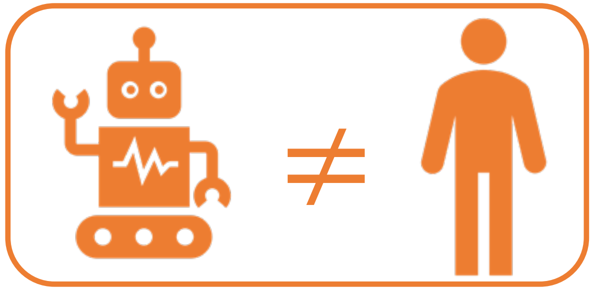


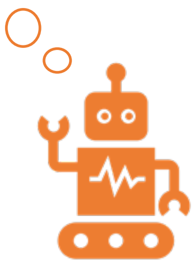




Risk-aware robot



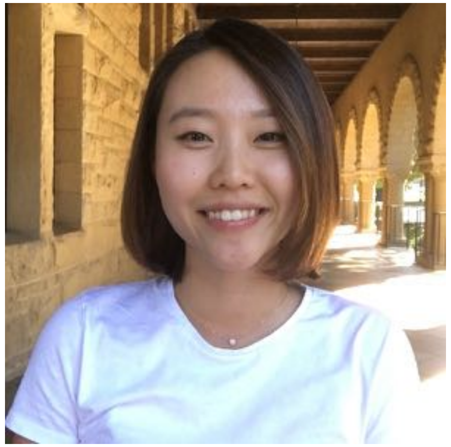




Key Idea:

We capture ***suboptimal*** human behavior using risk-aware human models from cumulative prospect theory.

Erdem Biyik



Minae Kwon

[Kwon, et al. HRI20]

Human Models

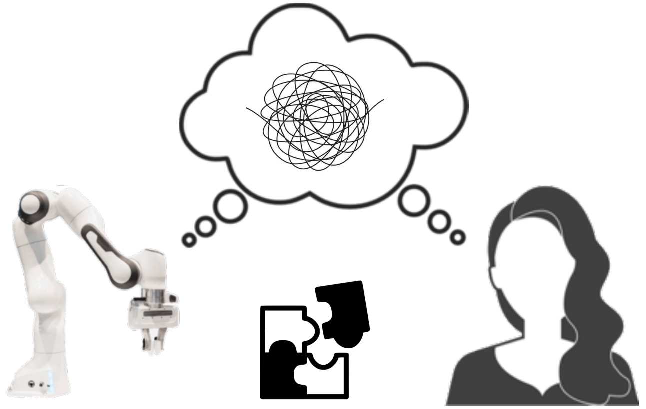
- Data-efficient learning of reward functions with different sources of data
- What happens on the ends of the risk spectrum?

Human Models

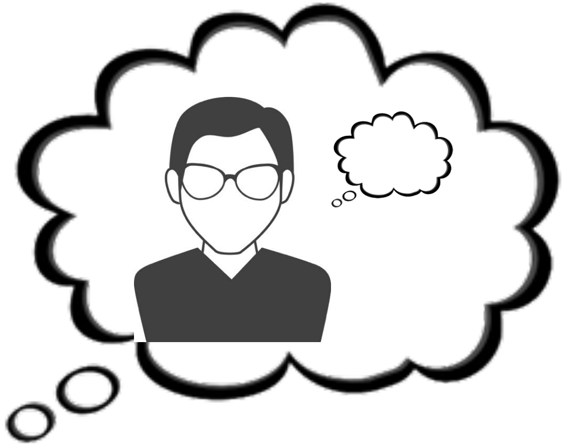
- Data-efficient learning of reward functions with different sources of data
- What happens on the ends of the risk spectrum?

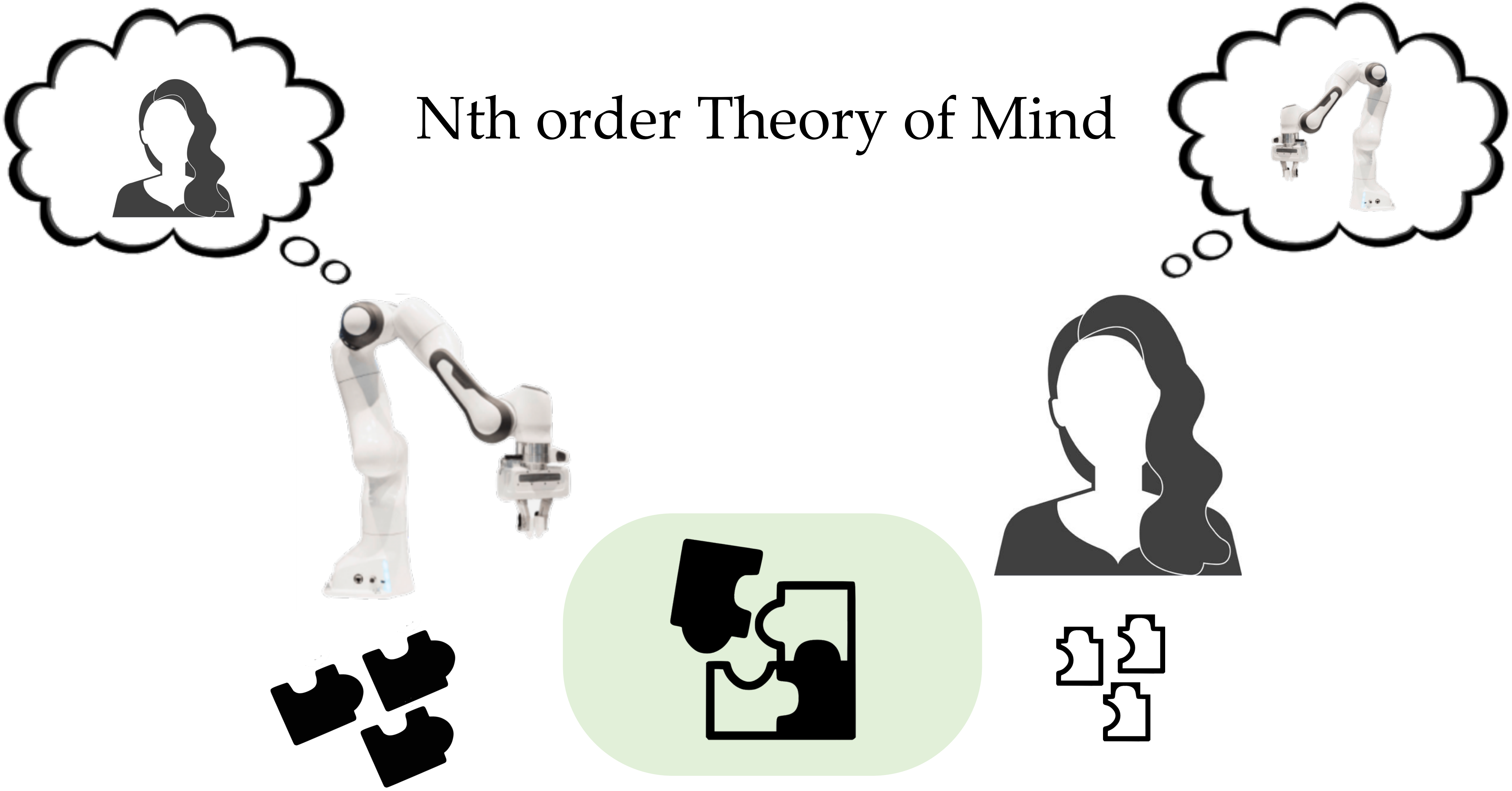
Conventions

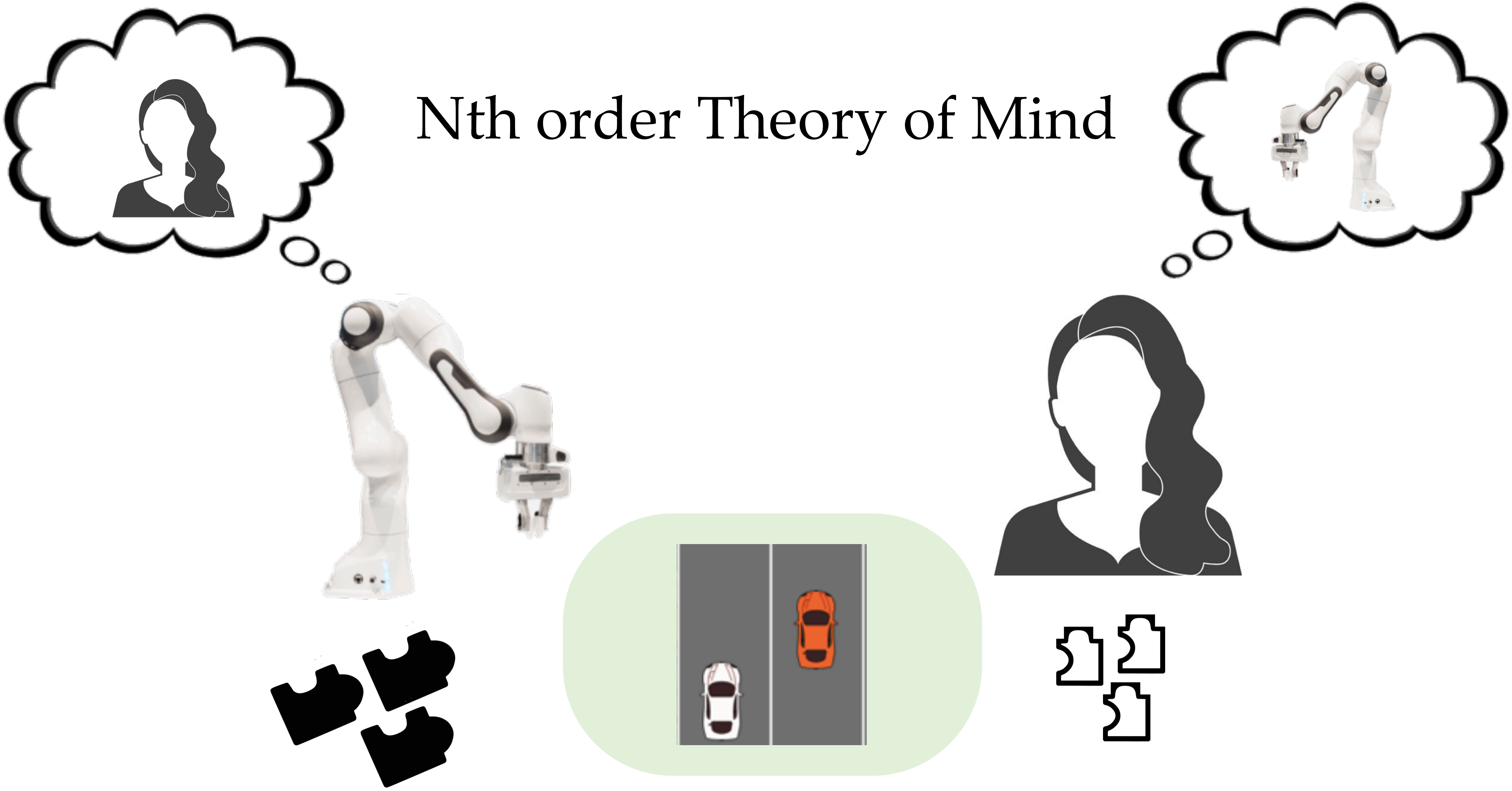
- What low dimensional representations are necessary when collaborating with humans?



Nth order Theory of Mind




Nth order Theory of Mind


Nth order Theory of Mind

Interaction as a Dynamical System

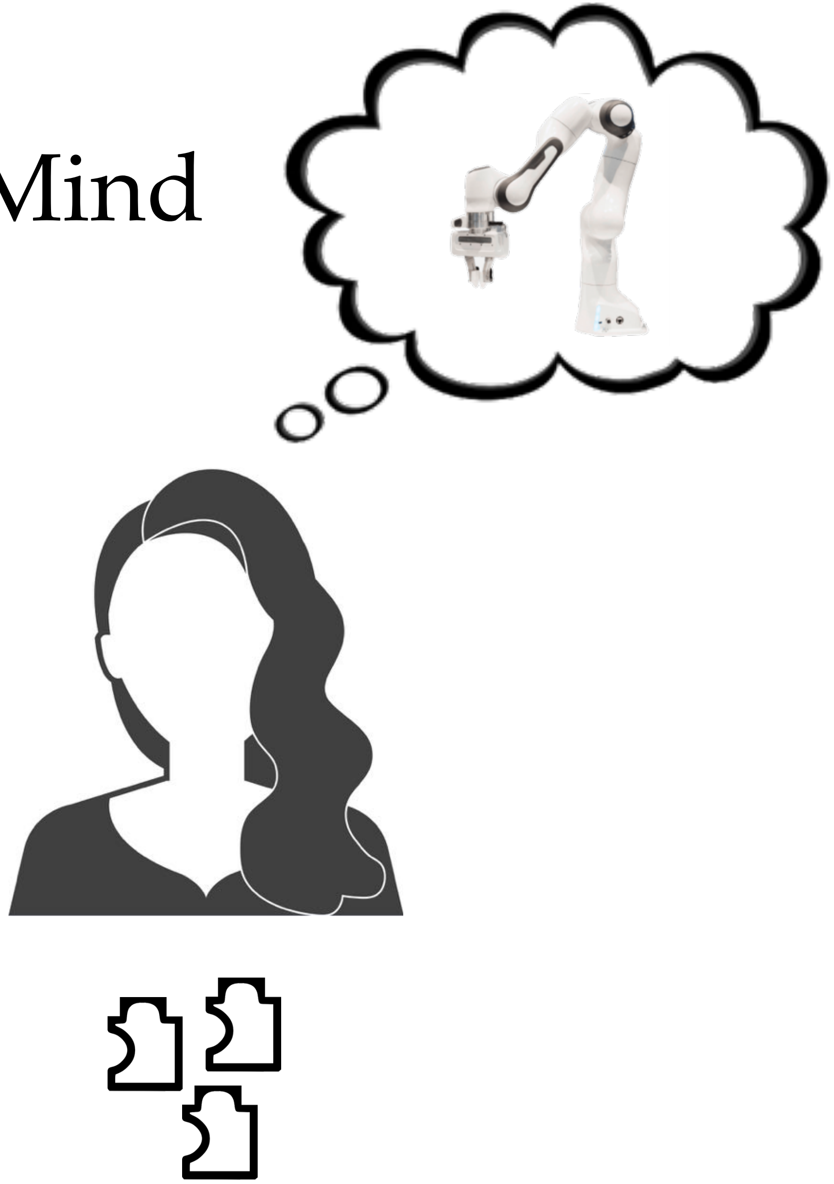
$$a^*_{\mathcal{R}} = \operatorname*{argmax}_{u_{\mathcal{R}}} R_{\mathcal{R}}(s, a_{\mathcal{R}}, a^*_{\mathcal{H}}(s, a_{\mathcal{R}}))$$

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

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

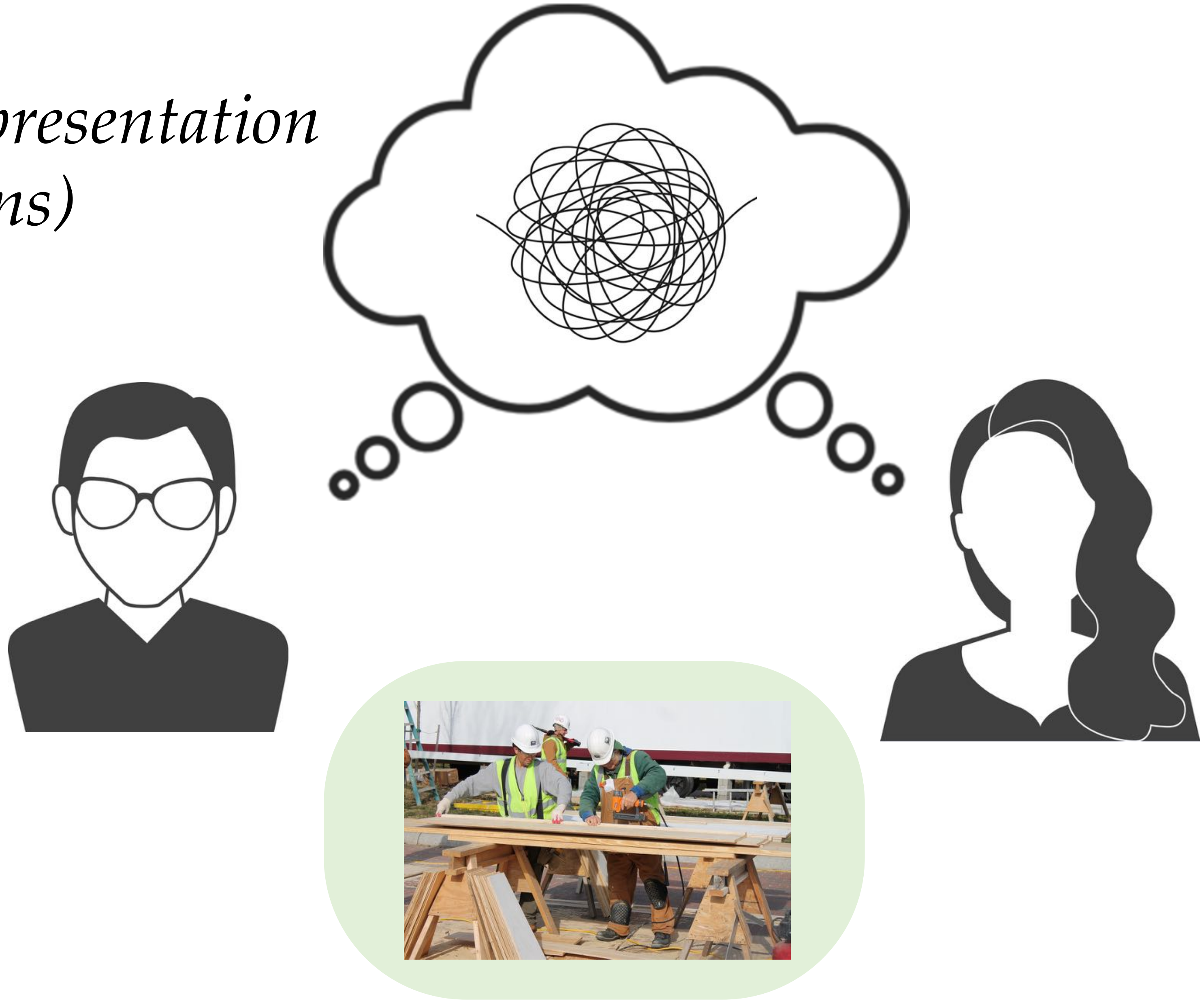
$$a^*_{\mathcal{H}}(s, a_{\mathcal{R}}) \approx \operatorname*{argmax}_{u_{\mathcal{H}}} R_{\mathcal{H}}(s, a_{\mathcal{R}}, a_H)$$

Nth order Theory of Mind

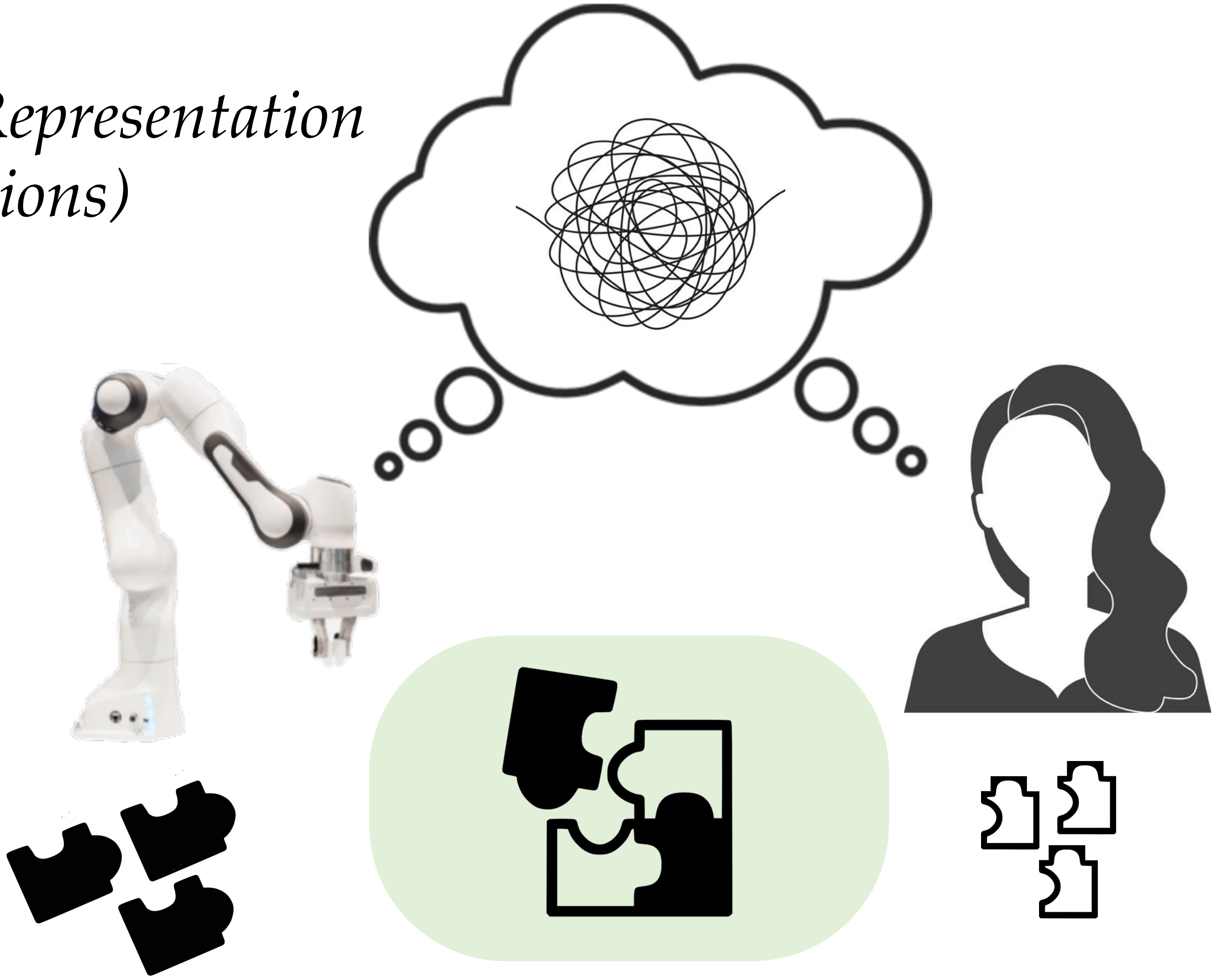


Interactive tasks are usually not the same as playing chess!

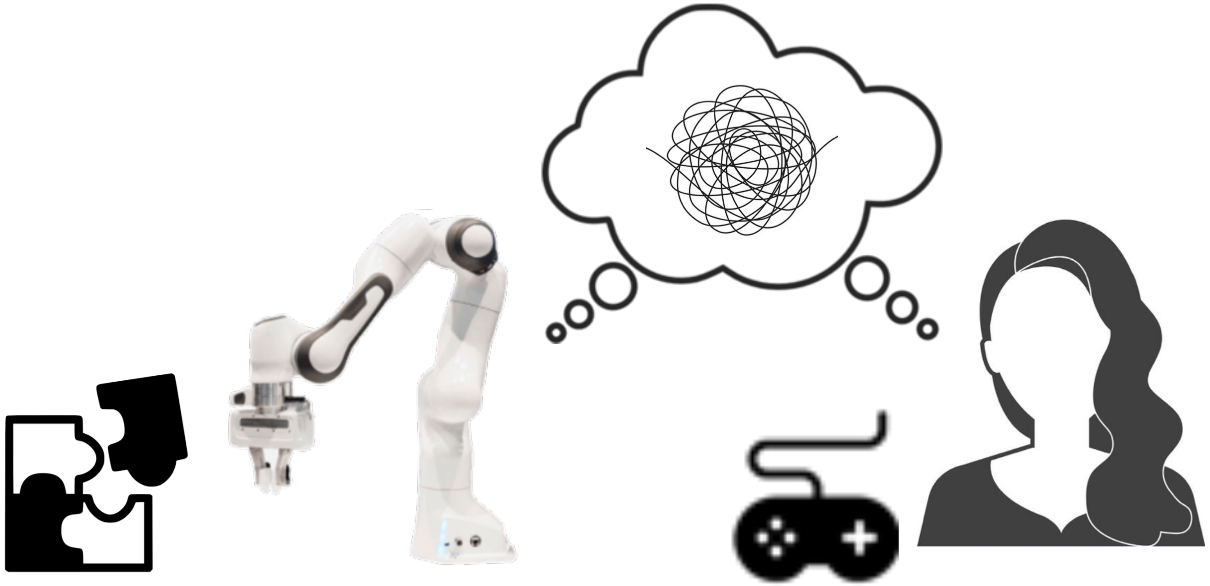
Shared Representation (conventions)



Shared Representation (conventions)



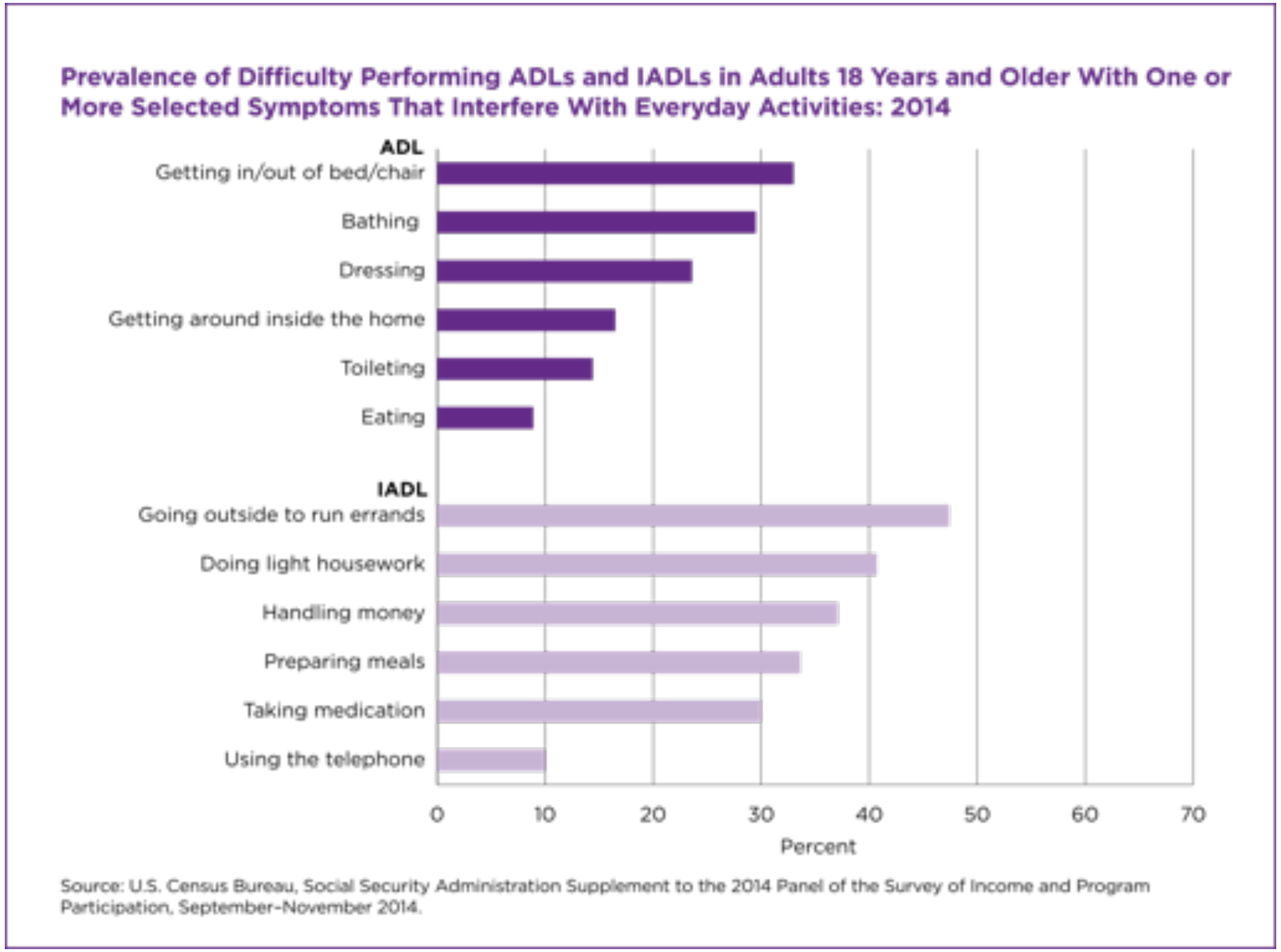
Conventions are low-dimensional shared representations that capture the interaction and can change over time.

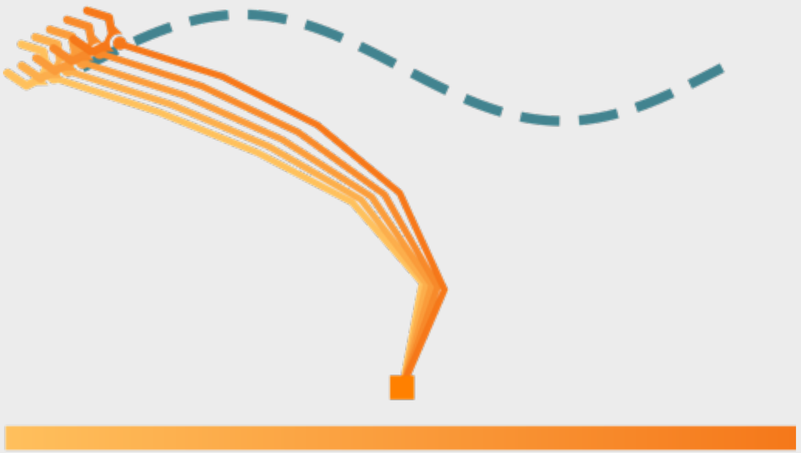


What are *conventions*?

Can robots directly *learn* conventions from interactions?

Can robots *influence* conventions?


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
ADL
Getting in/out of bed/chair
Bathing
Dressing
Getting around inside the home
Toileting
Eating
IADL
Going outside to run errands
Doing light housework
Handling money
Preparing meals
Taking medication
Using the telephone
0
10
20
30
40
50
60
70
Percent
Source: U.S. Census Bureau, Social Security Administration Supplement to the 2014 Panel of the Survey of Income and Program Participation, September–November 2014.



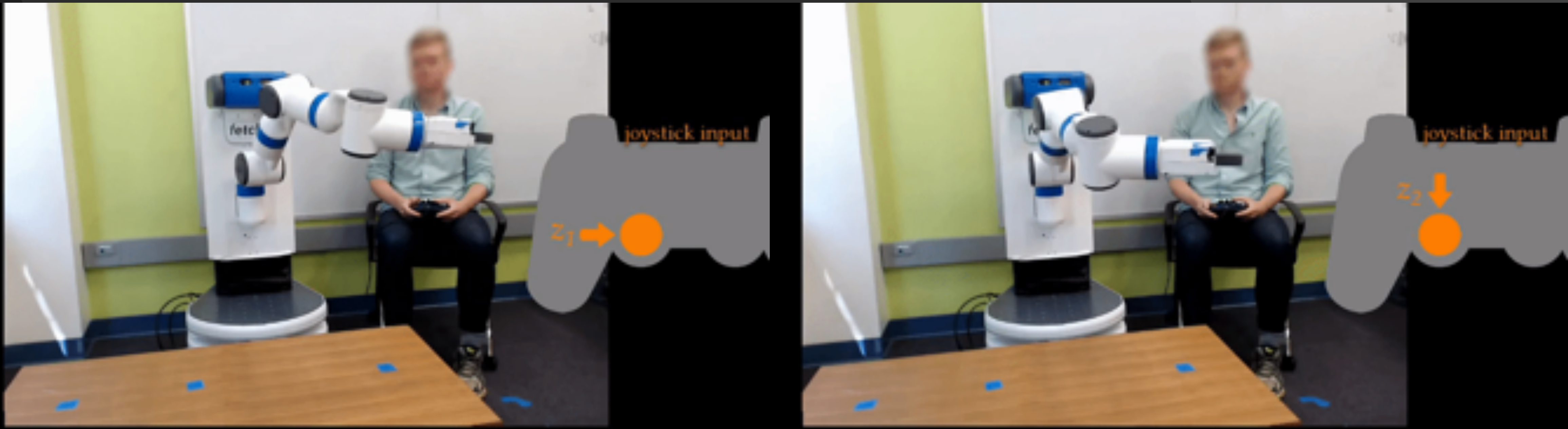
- Assistive robotic arms are ***dexterous***
- This dexterity makes it hard for users to ***control*** the robot

- How can robots ***learn*** low-dimensional representations that make controlling the robot intuitive?

Our Vision

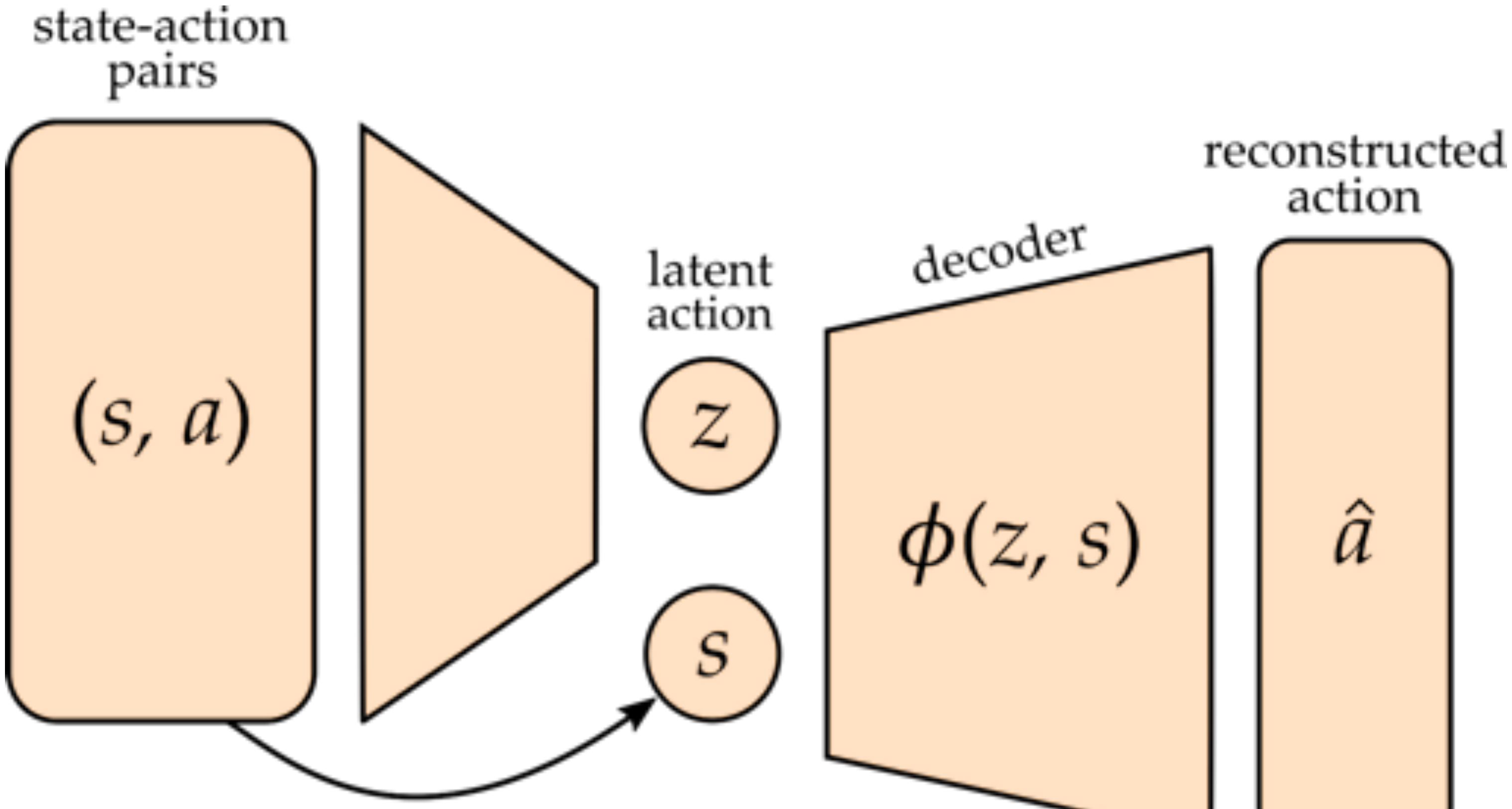
Offline, expert demonstrations of *high-dimensional* motions

Our Vision



Learn *low-dimensional* latent representations for online control

Model Structure (cVAE)



Learning Intuitive Latent Actions

Conditioned. The meaning of the latent action z depends on the current state s.

$$\hat{a} = \phi(z, s)$$

Controllable. The robot can move between states in the dataset.

Consistent. The same z causes the robot to behave similarly nearby.

Scalable. Larger latent actions cause larger changes in the state.

Learning Intuitive Latent Actions

Conditioned. The meaning of the latent action z depends on the current state s.

$$\hat{a} = \phi(z, s)$$

Controllable. The robot can move between states in the dataset.

$$given\ (s, s') \qquad \exists\, z \in \mathcal{Z} \qquad s.t.\ \ s' = \mathcal{T}(s, \phi(z, s))$$

Consistent. The same z causes the robot to behave similarly nearby.

Scalable. Larger latent actions cause larger changes in the state.

Learning Intuitive Latent Actions

Conditioned. The meaning of the latent action z depends on the current state s.

$$\hat{a} = \phi(z, s)$$

Controllable. The robot can move between states in the dataset.

$$given\ (s, s') \qquad \exists\, z \in \mathcal{Z} \qquad s.t.\ \ s' = \mathcal{T}(s, \phi(z, s))$$

Consistent. The same z causes the robot to behave similarly nearby.

$$d_M(\mathcal{T}(s_1, \phi(z, s_1)), \mathcal{T}(s_2, \phi(z, s_2))) < \epsilon \quad when \quad \|s_1 - s_2\| < \delta$$

Scalable. Larger latent actions cause larger changes in the state.

Learning Intuitive Latent Actions

Conditioned. The meaning of the latent action z depends on the current state s.

$$\hat{a} = \phi(z, s)$$

Controllable. The robot can move between states in the dataset.

$$given\ (s, s') \qquad \exists\, z \in \mathcal{Z} \qquad s.t.\ \ s' = \mathcal{T}(s, \phi(z, s))$$

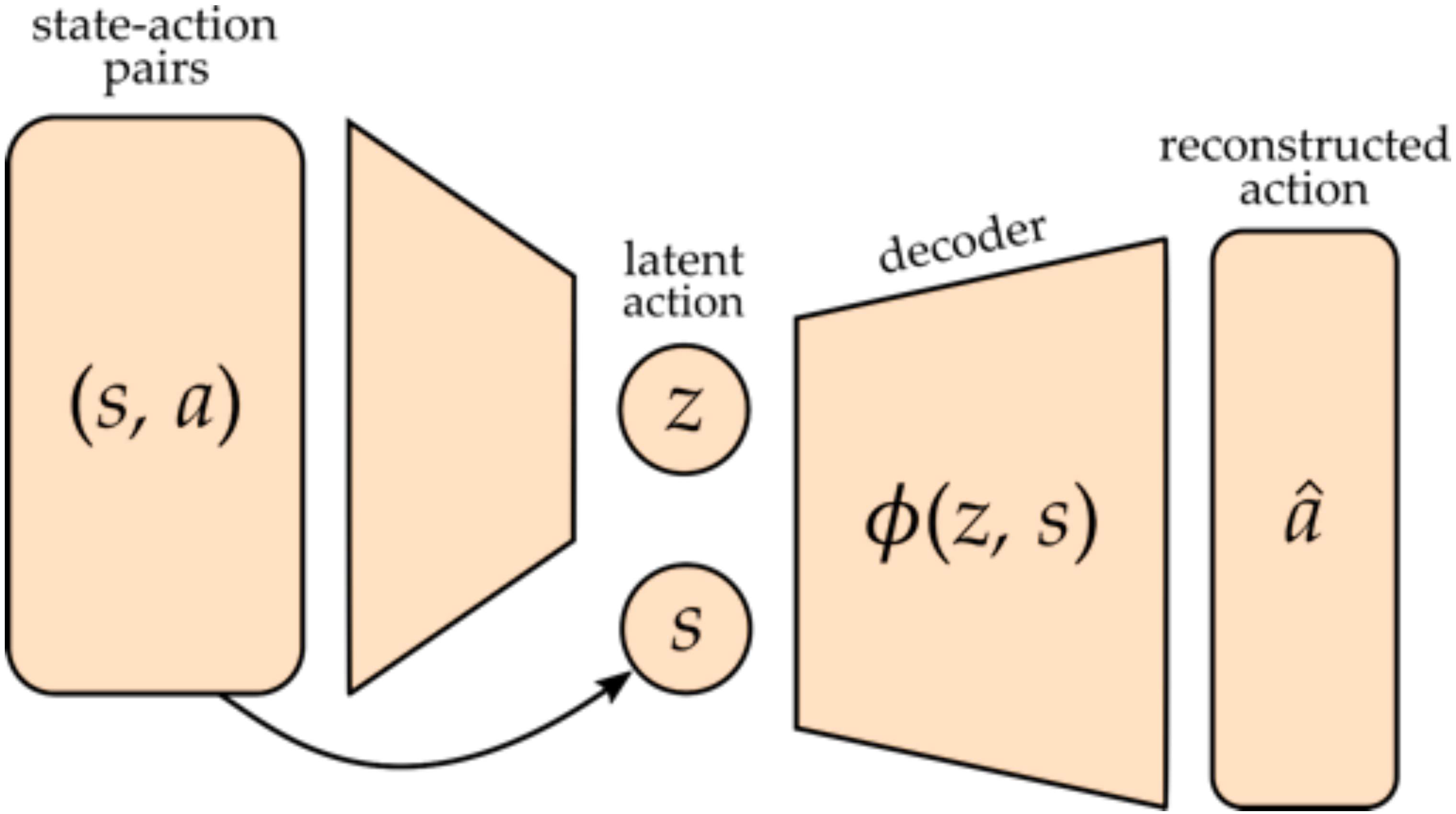
Consistent. The same z causes the robot to behave similarly nearby.

$$d_M(\mathcal{T}(s_1, \phi(z, s_1)), \mathcal{T}(s_2, \phi(z, s_2))) < \epsilon \quad when \quad \|s_1 - s_2\| < \delta$$

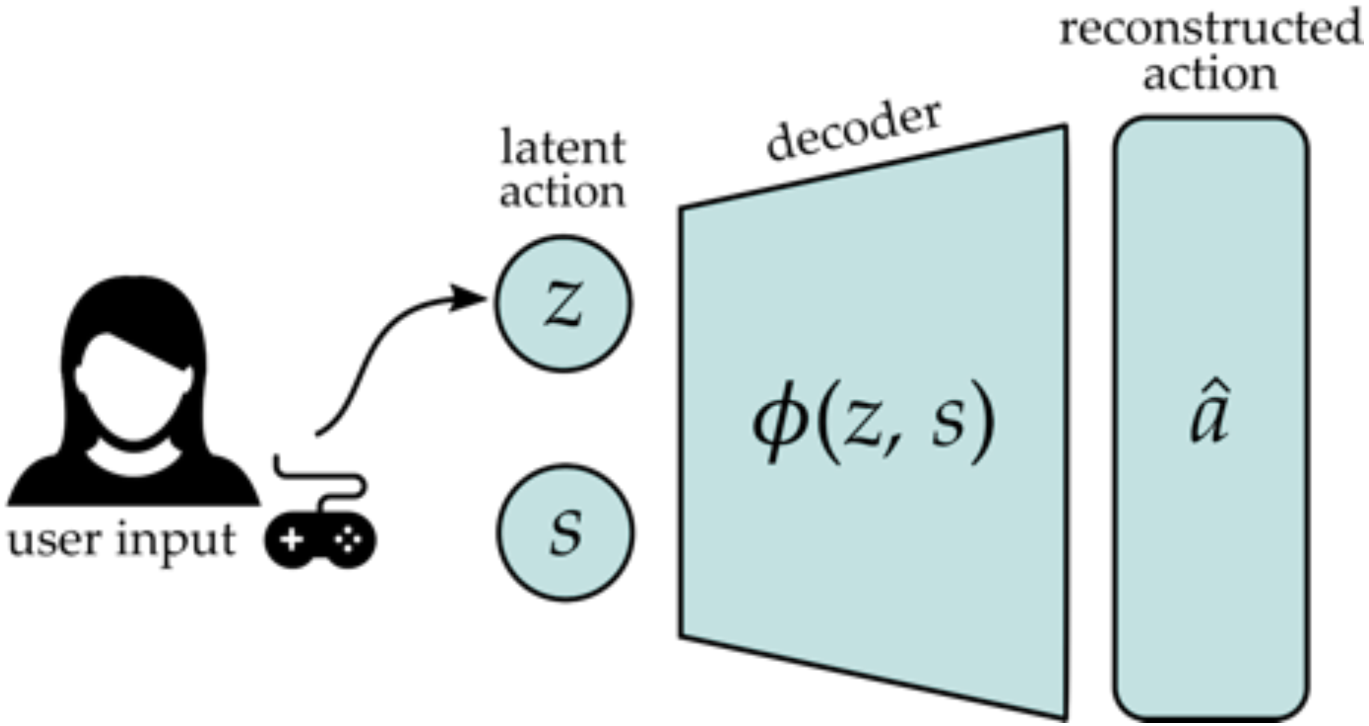
Scalable. Larger latent actions cause larger changes in the state.

$$\|s - \mathcal{T}(s, \phi(z, s))\| \to \infty \quad as \quad \|z\| \to \infty$$

Model Structure (cVAE)



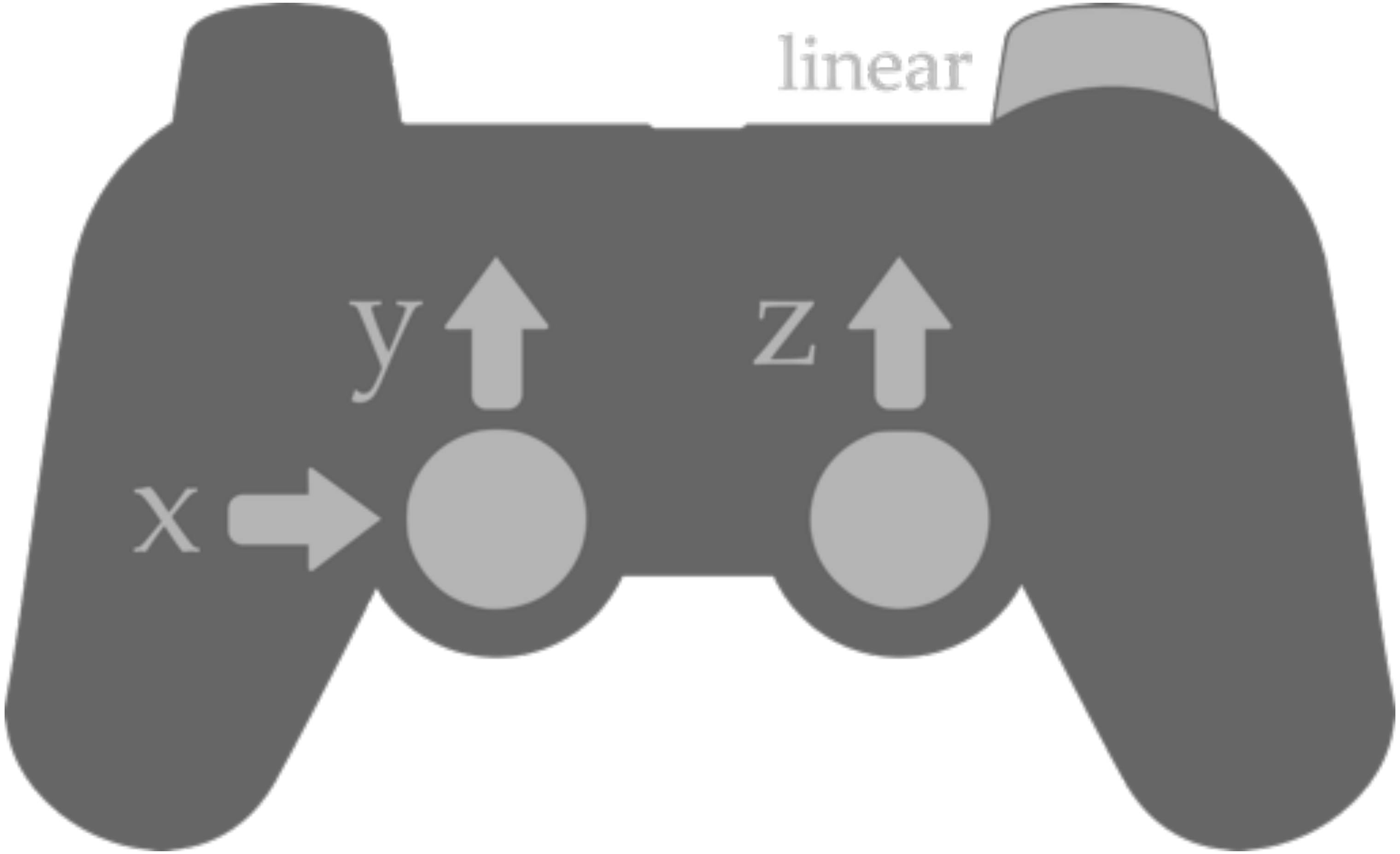
Model Structure (cVAE)



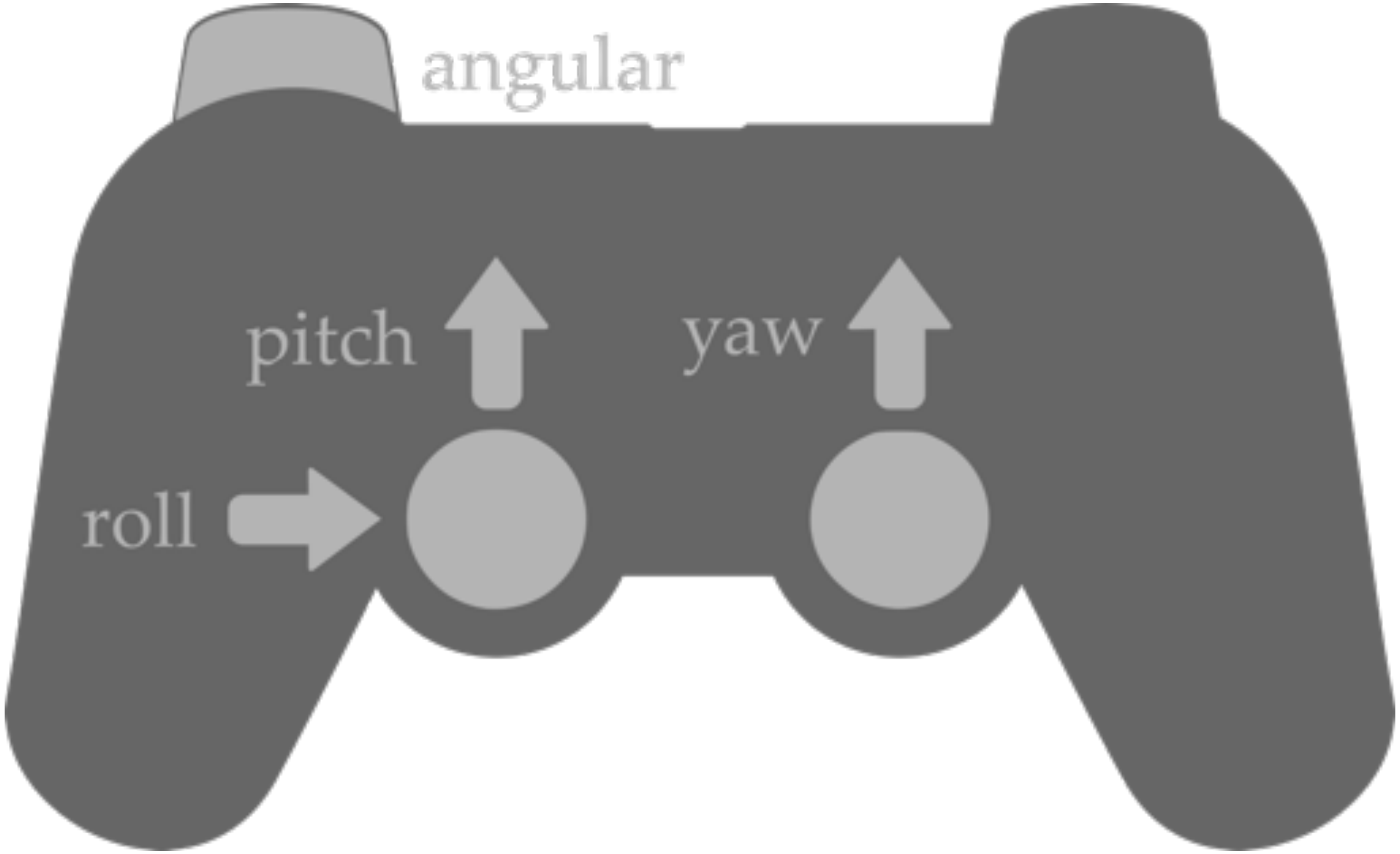
User Study

- 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*)

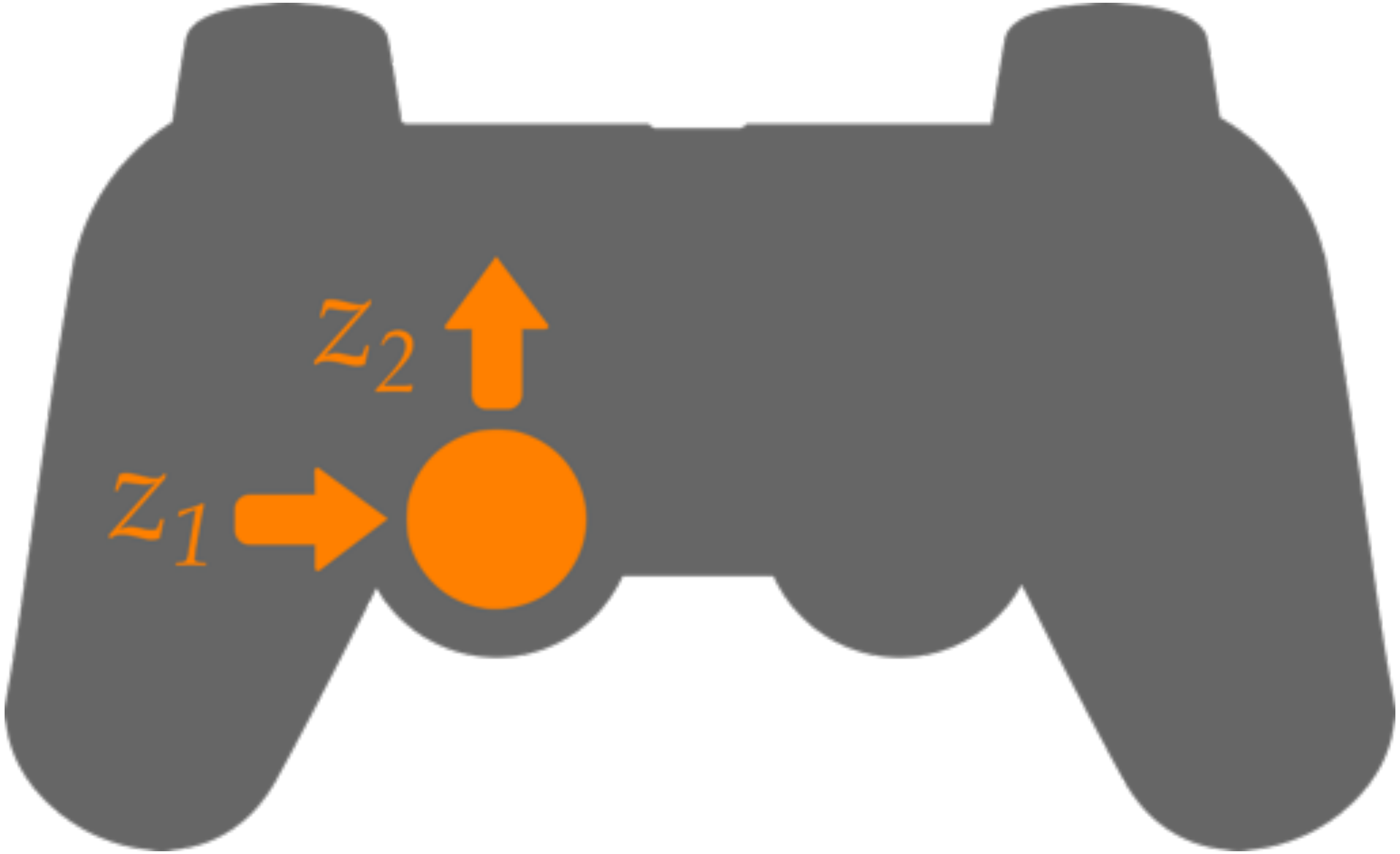
End-Effector



End-Effector



Latent Actions



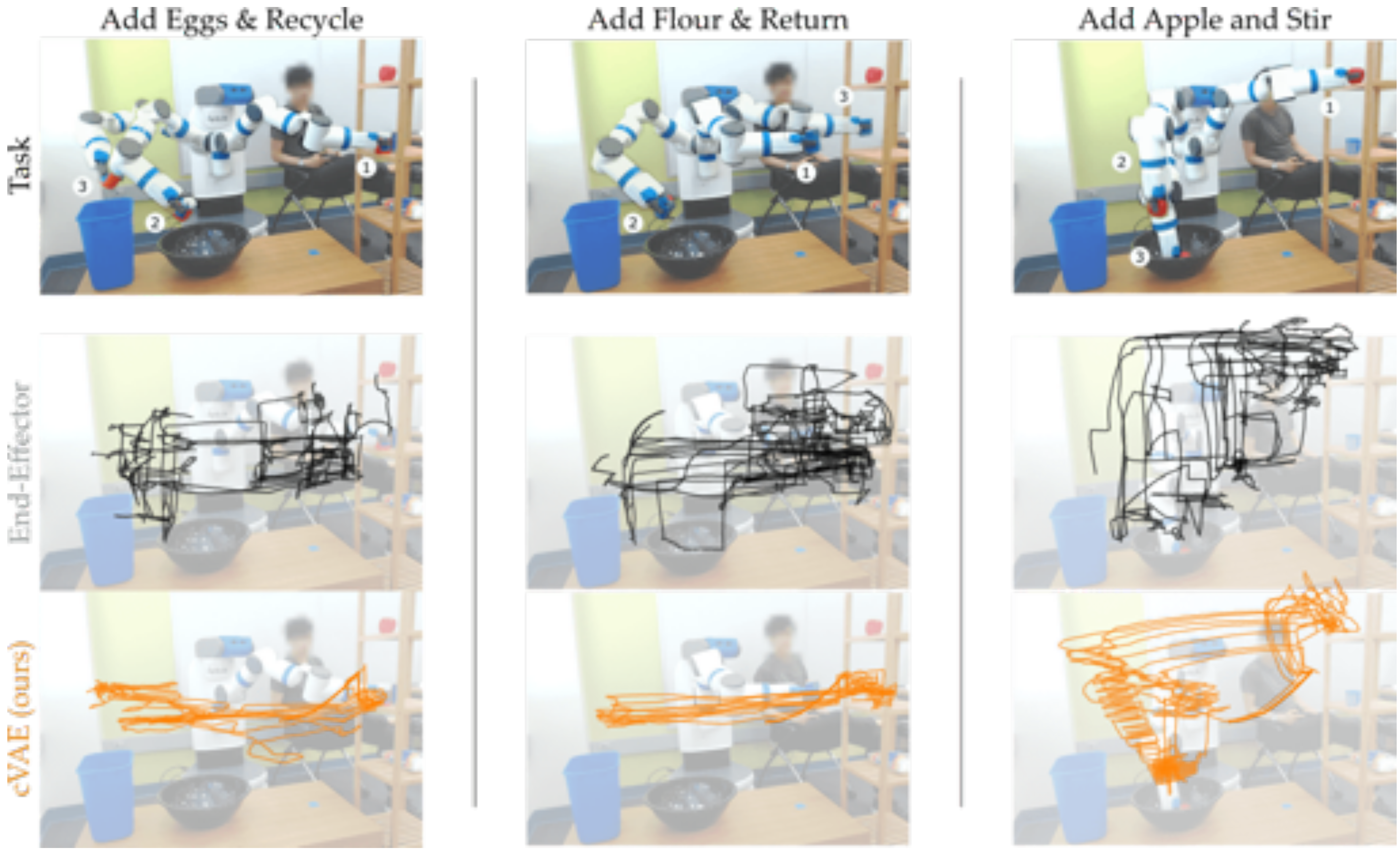
4x Speed

(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

Summary so far…

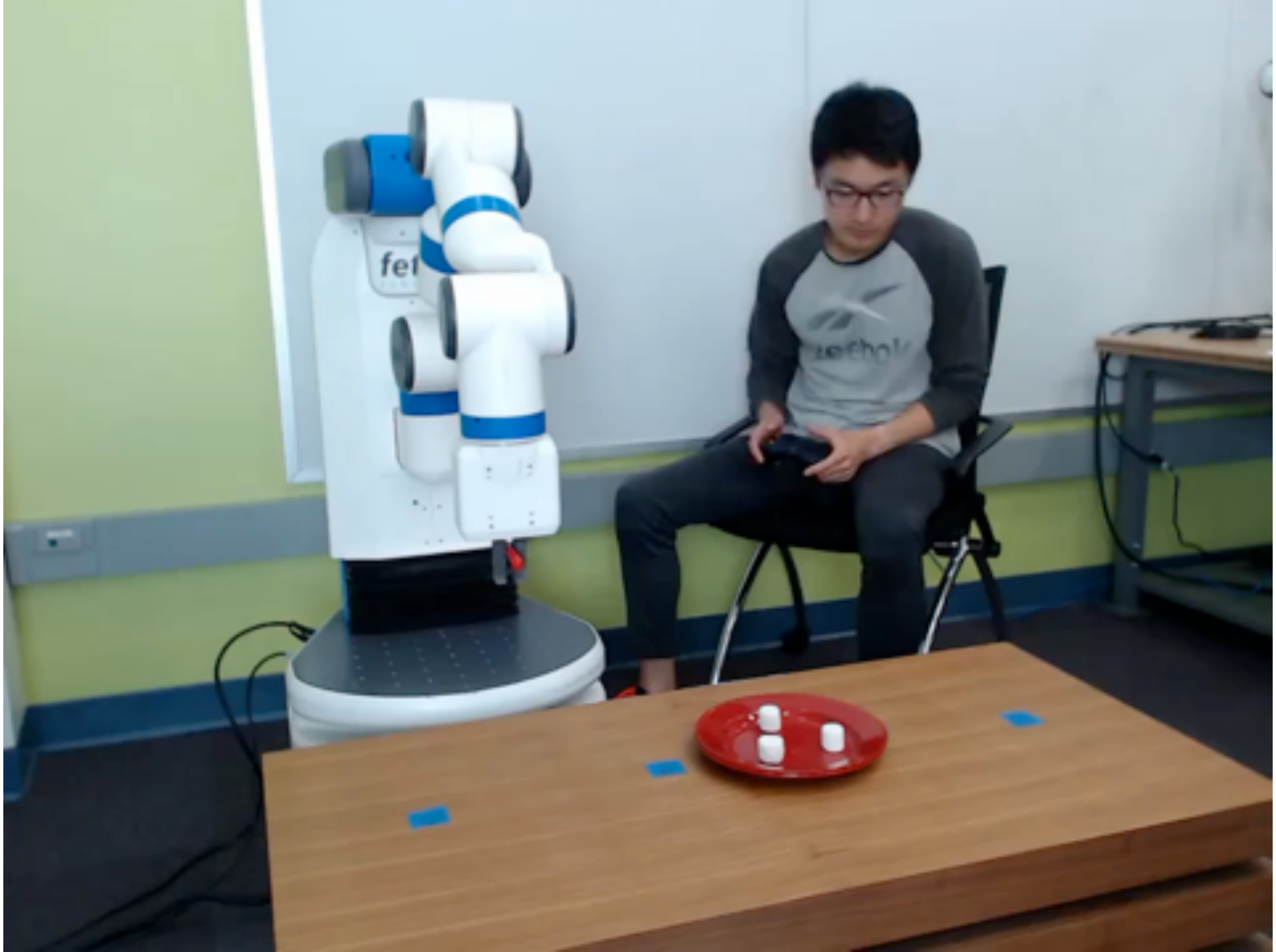
- We *embedded* personalized behaviors to latent spaces
- *Formalized* the properties these latent spaces should satisfy
- Learned from *efficient* amounts of data

Dylan Losey

[Losey, et al., ICRA 2020]

Latent actions enable intuitive
low-dimensional control…

…but is this enough for
precise manipulation tasks?



Precise Manipulation

Cutting

Scooping

Yes

Latent Actions + Shared Autonomy

Start

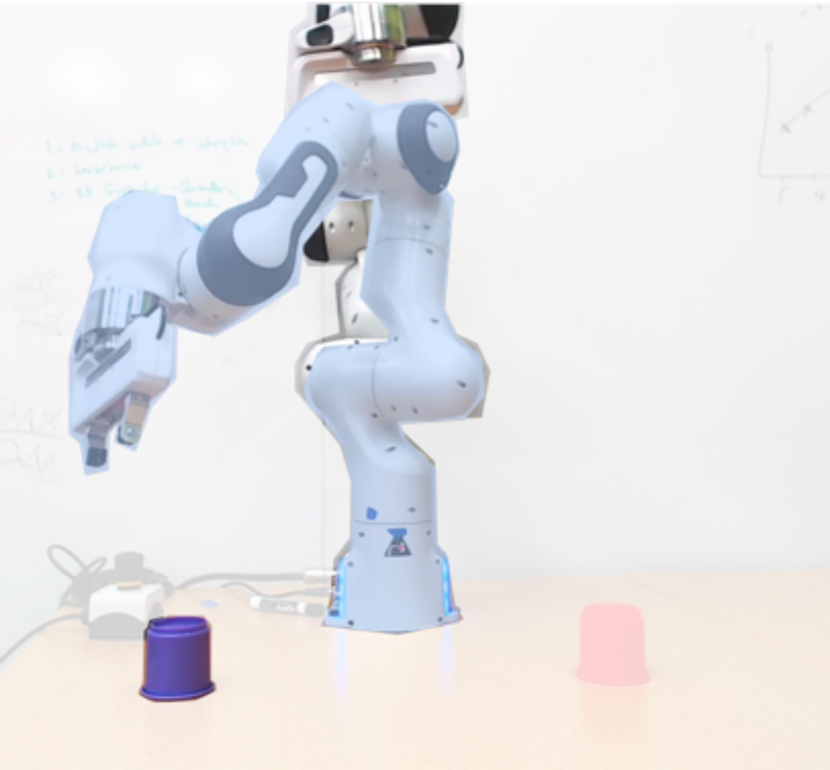
No Assistance

Start

Shared Autonomy

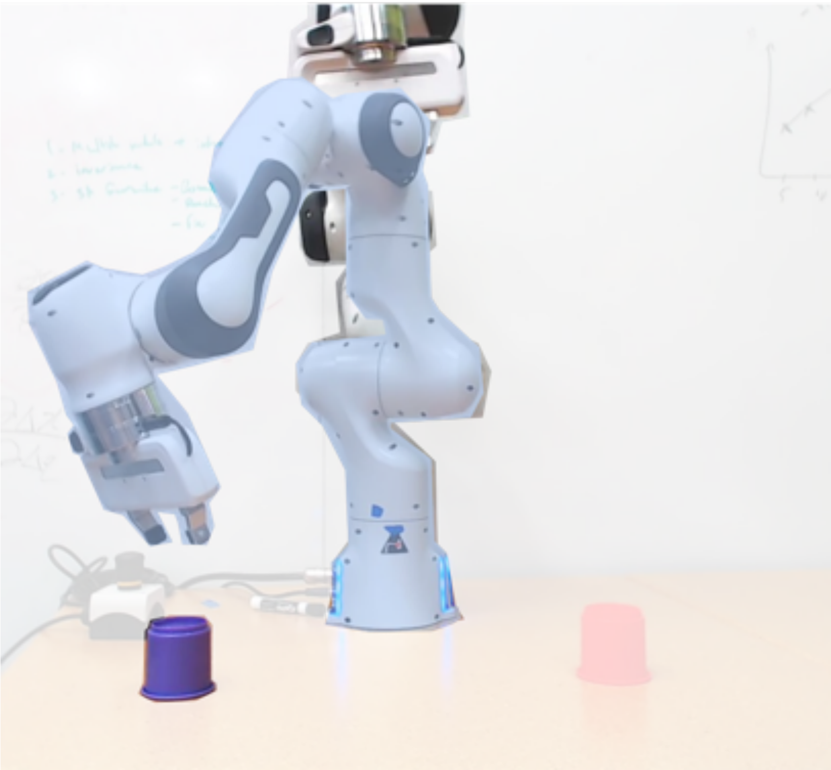
Latent Actions + Shared Autonomy

Control Goal



No Assistance

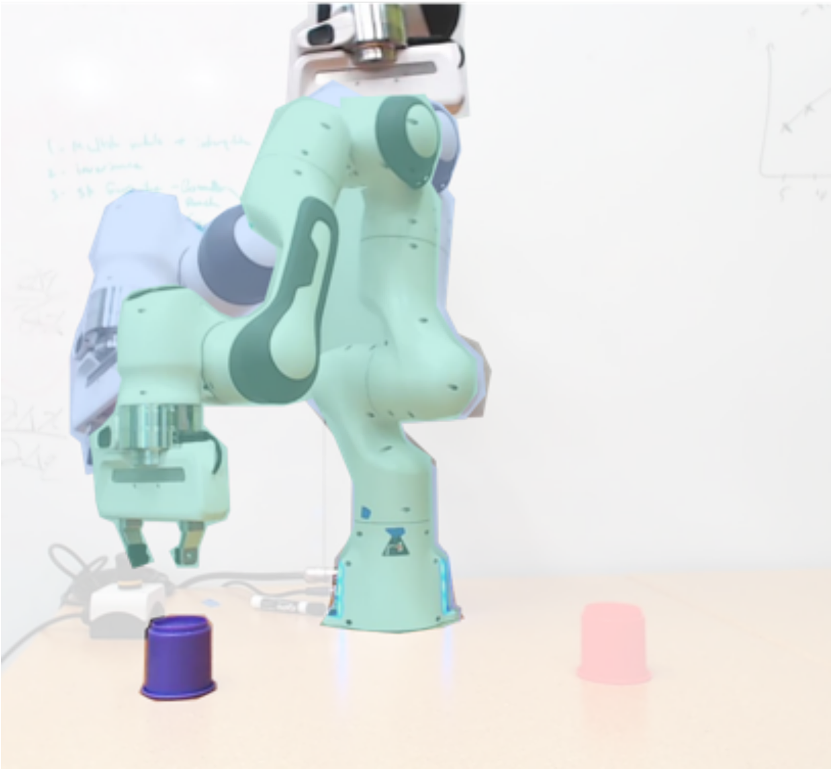
Control Goal



Shared Autonomy

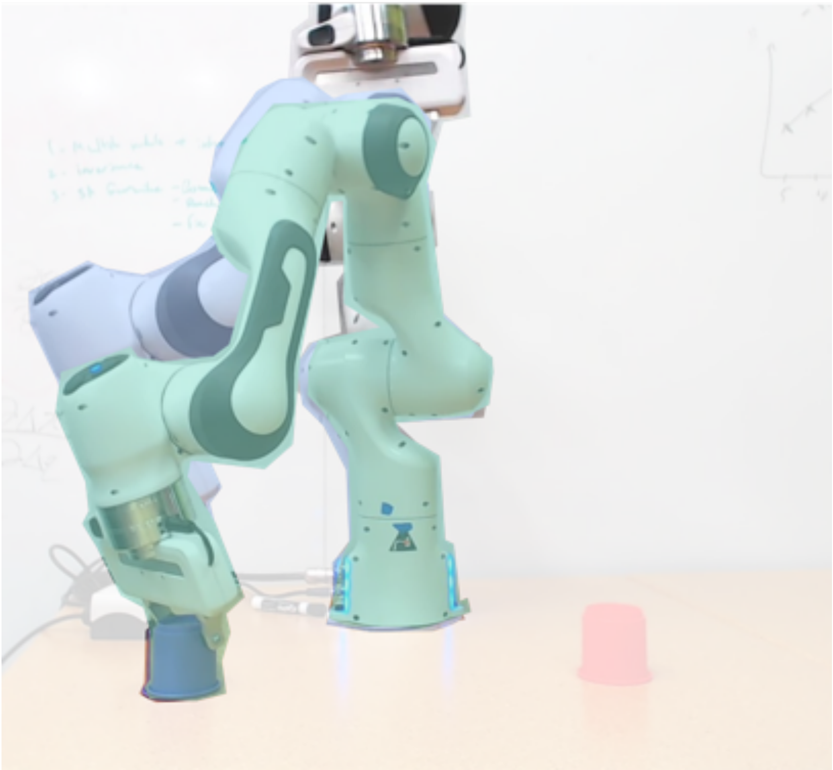
Latent Actions + Shared Autonomy

Control Preference



No Assistance

Control Preference



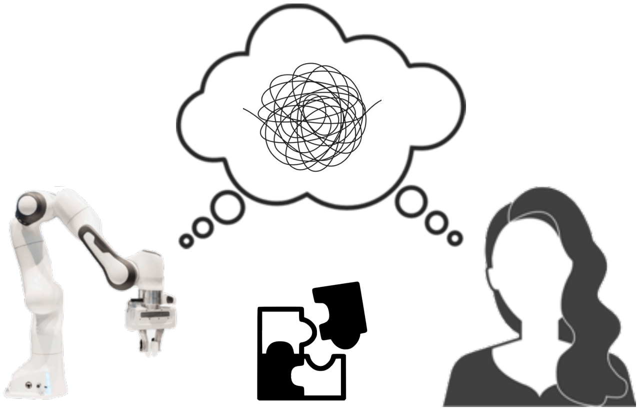
Shared Autonomy

Human Models

- Data-efficient learning of reward functions with different sources of data
- What happens on the ends of the risk spectrum?

Conventions

- What low dimensional representations are necessary when collaborating with humans?



1) There is an *opportunity* for learning and control

… to formalize and solve challenging problems of interaction with humans.

2) We need to design *computational models of human* behavior

Can we rely on low-dimensional statistics that capture high-dimensional interactions?

1) There is an *opportunity* for learning and control

… to formalize and solve challenging problems of interaction with humans.

2) We need to design *computational models of human* behavior

Can we rely on low-dimensional statistics that capture high-dimensional interactions?

3) We spend a lot of effort learning what humans want or do…
… but humans constantly *change*

What can learning and control do?


GETCRUISE.COM

Two different driving equilibria from years of repeated interactions


iliad
intelligent and interactive autonomous systems

1) There is an *opportunity* for learning and control

… to formalize and solve challenging problems of interaction with humans.

2) We need to design *computational models of human* behavior

Can we rely on low-dimensional statistics that capture high-dimensional interactions?

3) We spend a lot of effort learning what humans want or do…
… but humans constantly *change*

What can learning and control do?