Principles of Robot Autonomy II

Human-Robot Interaction

Stanford University

iliad
intelligent and interactive autonomous systems

Recap

- Imitation learning and inverse RL
- Learning from other sources of data – Pairwise Comparisons
- Learning from other sources of data – Foundation Models
- Learning from physical feedback
- Learning from gestures
- Learning from sketches
- Data Quality

Types of Imitation Learning

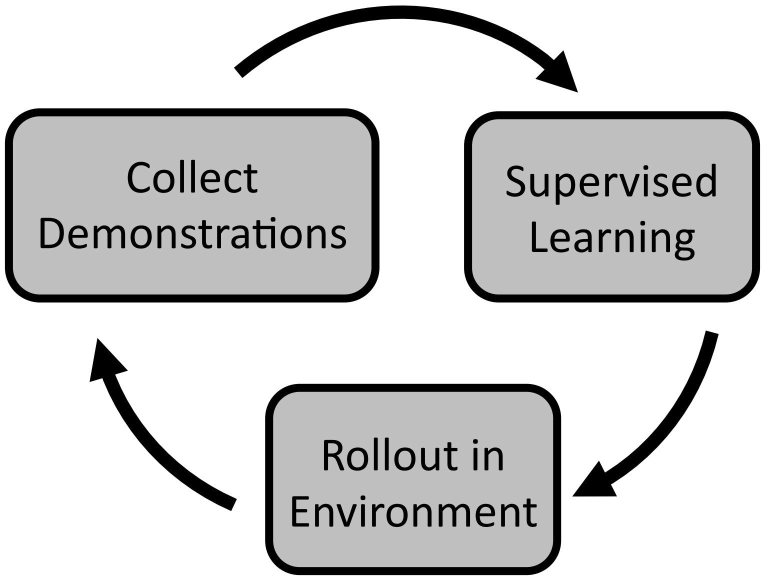
Behavioral Cloning

$$\arg\min_{\theta} \mathbb{E}_{(s,a^*)\sim P^*} L(a^*, \pi_\theta(s))$$

Works well when P^* is close to P_θ

Direct Policy Learning (via Interactive Demonstrator)

Requires Interactive Demonstrator (BC is a 1-step special case)



Inverse RL

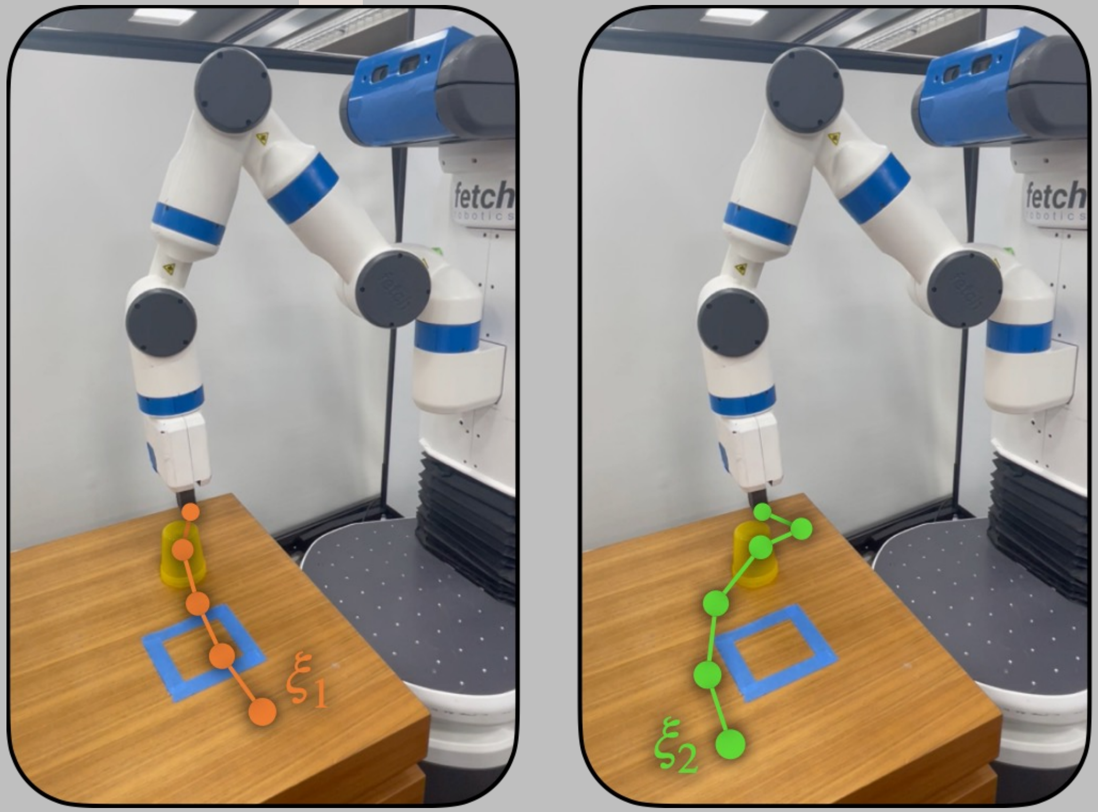
Learn r such that:

$$\pi^* = \overbrace{\arg\max_{\theta} \mathbb{E}_{s\sim P(s|\theta)} r(s, \pi_\theta(s))}^{\text{RL problem}}$$

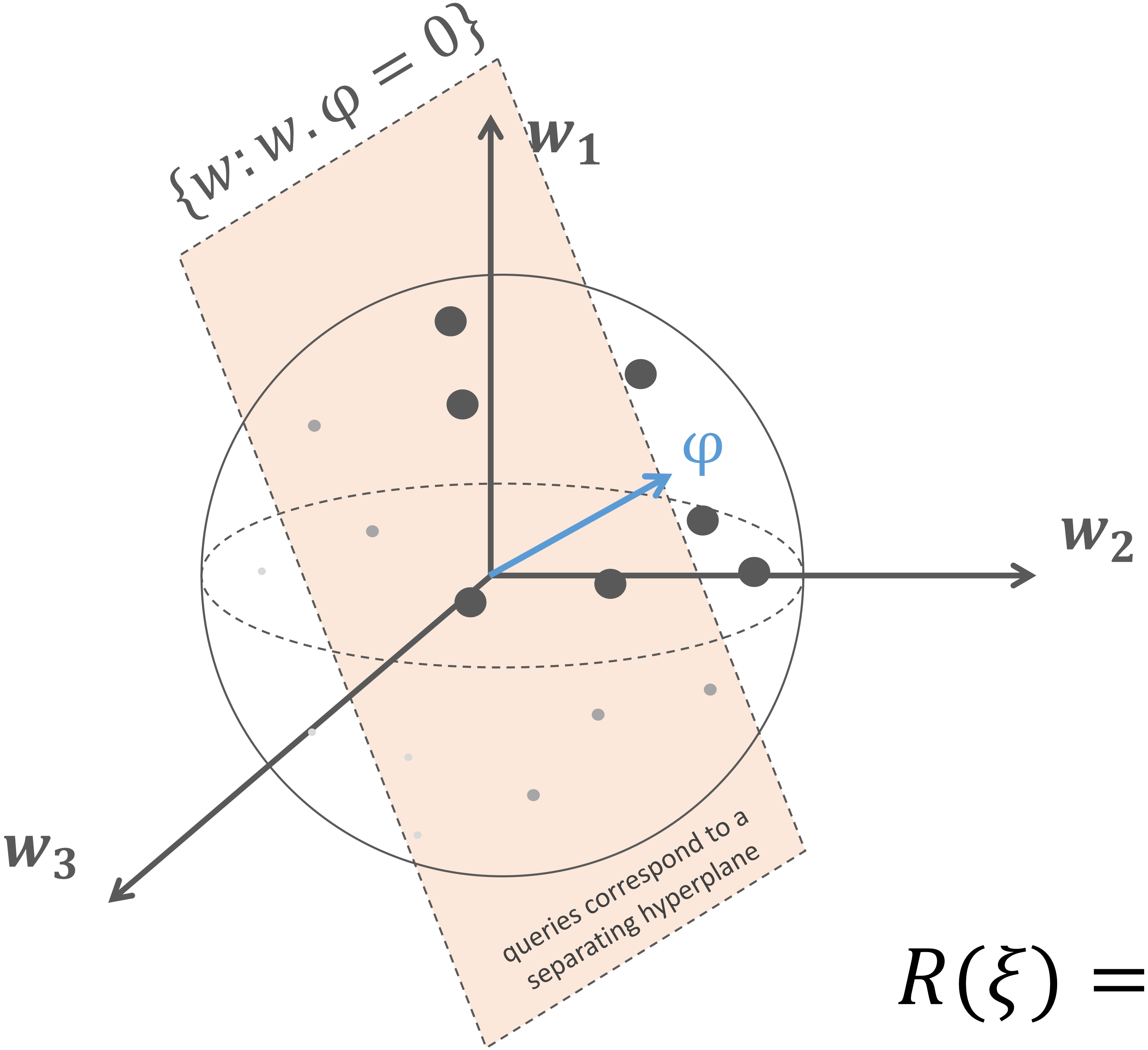
Assume learning r is statistically easier than directly learning π^*

Recap

- Imitation learning and inverse RL
- Learning from other sources of data – Pairwise Comparisons
- Learning from other sources of data – Foundation Models
- Learning from physical feedback
- Learning from gestures
- Learning from sketches
- Data Quality

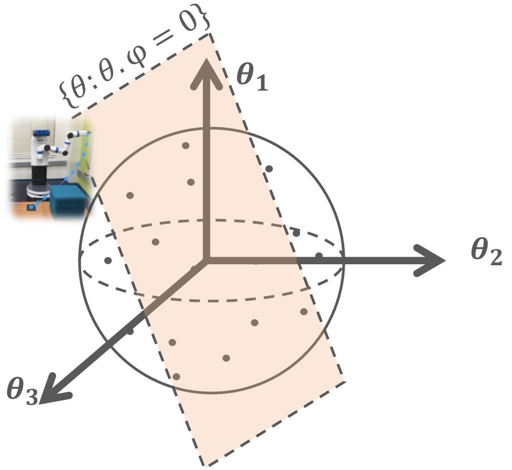


ξ_A or ξ_B?



$$R(\xi) = w \cdot \phi(\xi)$$

Actively *synthesizing* queries



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

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

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

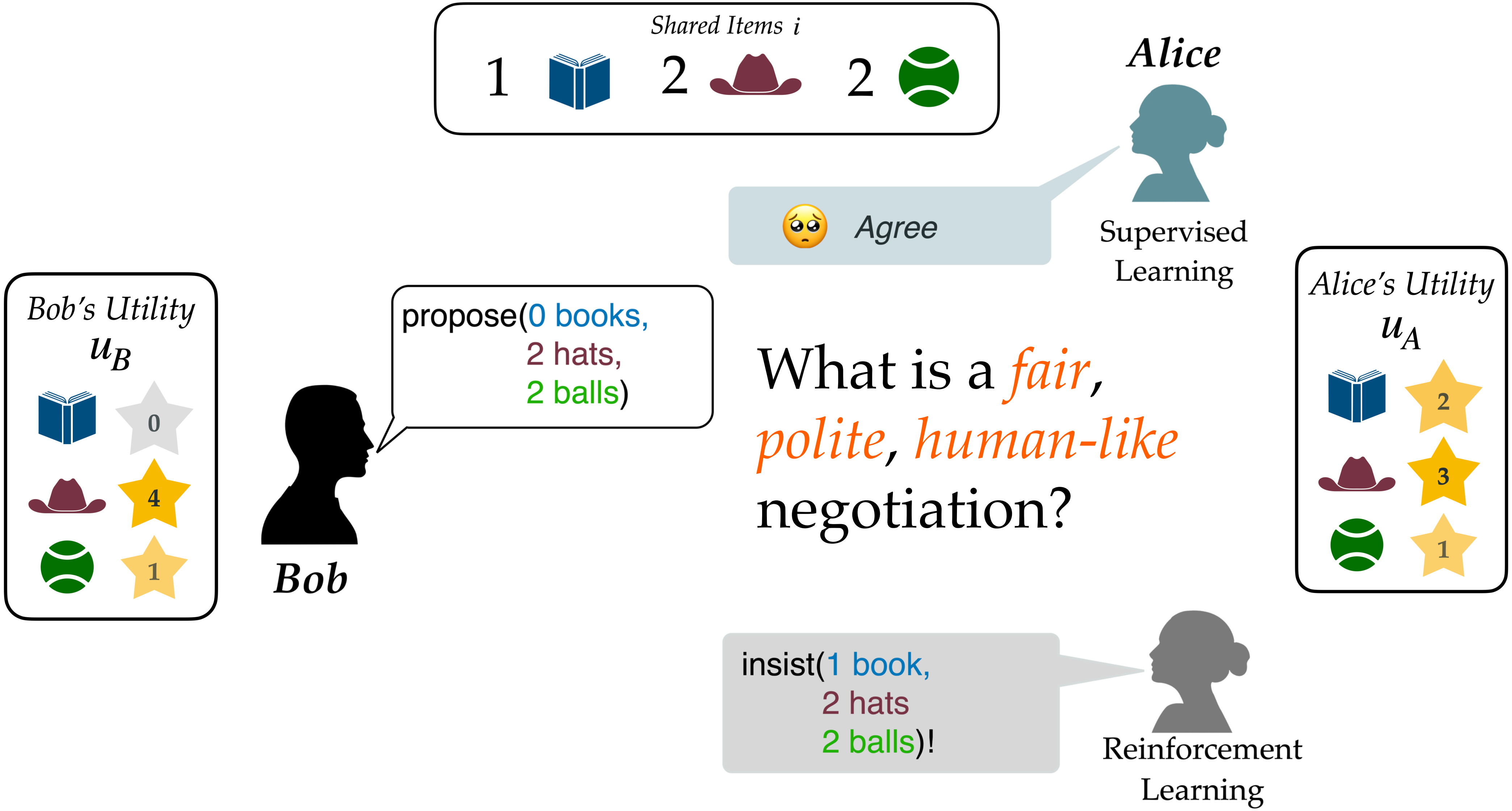
Human update function $\quad f_{\varphi}(\mathbf{w}) = \min(1, \exp(I_t \mathbf{w}^{\top} \varphi))$

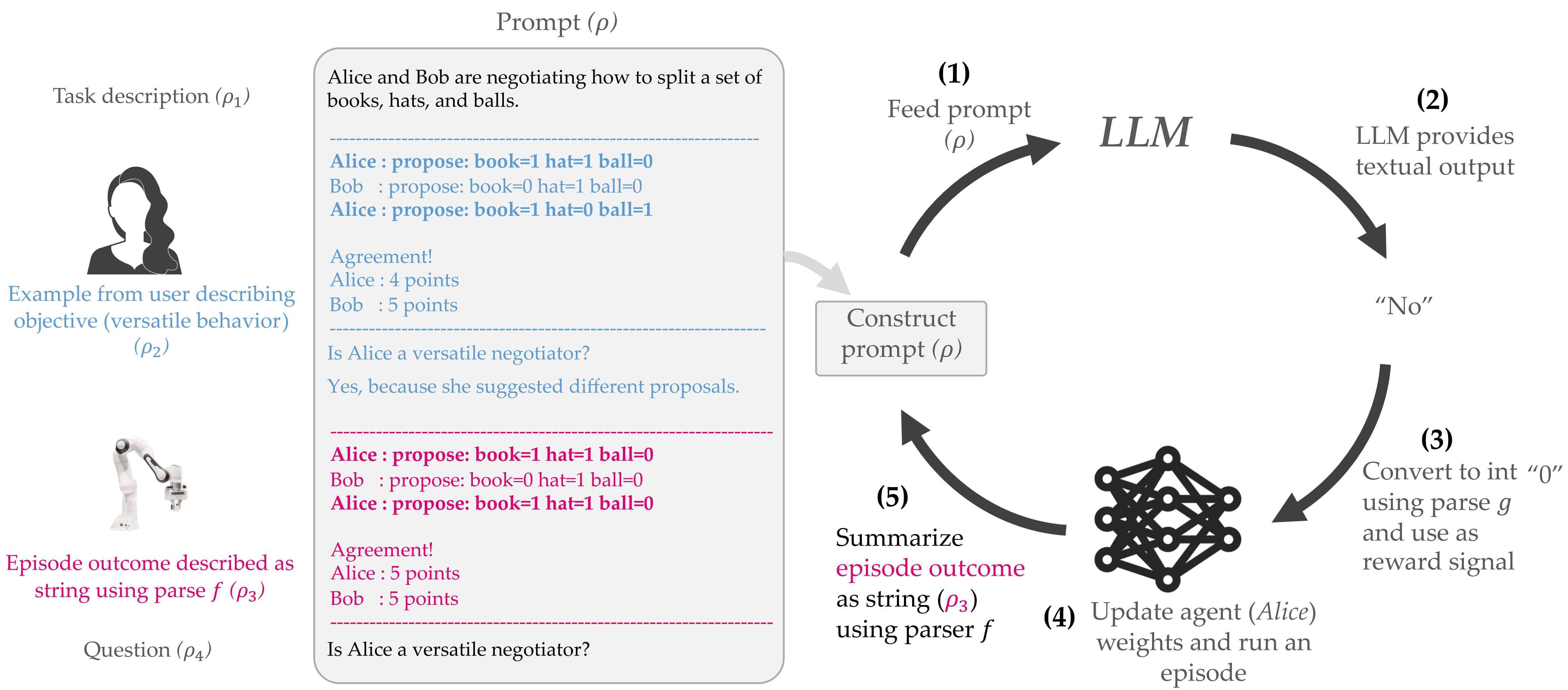
[Sadigh et al. RSS17]
[Biyik et al. CoRL18]
[Biyik et al. CDC19]
[Palan et al. RSS19]
[Biyik et al. CoRL19]
[Basu et al. IROS19]
[Biyik et al. RSS20]
[Myers et al. CoRL21]
[Myers et al. ICRA22]

Recap

- Imitation learning and inverse RL
- Learning from other sources of data – Pairwise Comparisons
- Learning from other sources of data – Foundation Models
- Learning from physical feedback
- Learning from gestures
- Learning from sketches
- Data Quality

Negotiation Domain




Prompt (ρ)
Task description (ρ_1)
Alice and Bob are negotiating how to split a set of books, hats, and balls.
Alice : propose: book=1 hat=1 ball=0
Bob : propose: book=0 hat=1 ball=0
Alice : propose: book=1 hat=0 ball=1
Agreement!
Alice : 4 points
Bob : 5 points
Is Alice a versatile negotiator?
Yes, because she suggested different proposals.
Example from user describing objective (versatile behavior) (ρ_2)
Alice : propose: book=1 hat=1 ball=0
Bob : propose: book=0 hat=1 ball=0
Alice : propose: book=1 hat=1 ball=0
Agreement!
Alice : 5 points
Bob : 5 points
Episode outcome described as string using parse f (ρ_3)
Question (ρ_4)
Is Alice a versatile negotiator?
Construct prompt (ρ)
(1)
Feed prompt (ρ)
LLM
(2)
LLM provides textual output
"No"
(3)
Convert to int "0" using parse g and use as reward signal
(4) Update agent (Alice) weights and run an episode
(5)
Summarize episode outcome as string (ρ_3) using parser f

Recap

- Imitation learning and inverse RL
- Learning from other sources of data – Pairwise Comparisons
- Learning from other sources of data – Foundation Models
- Learning from physical feedback
- Learning from gestures
- Learning from sketches
- Data Quality

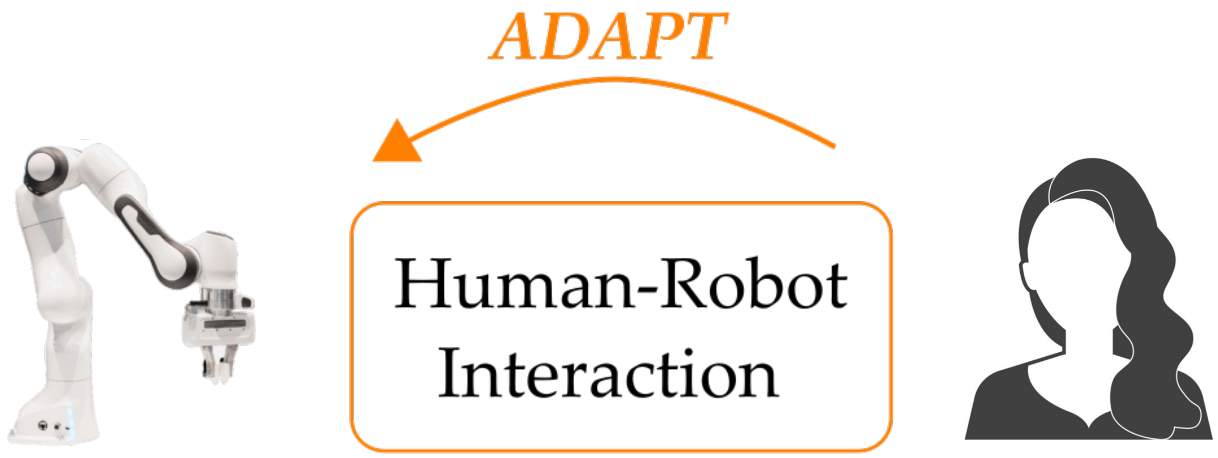
Today's itinerary

- 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

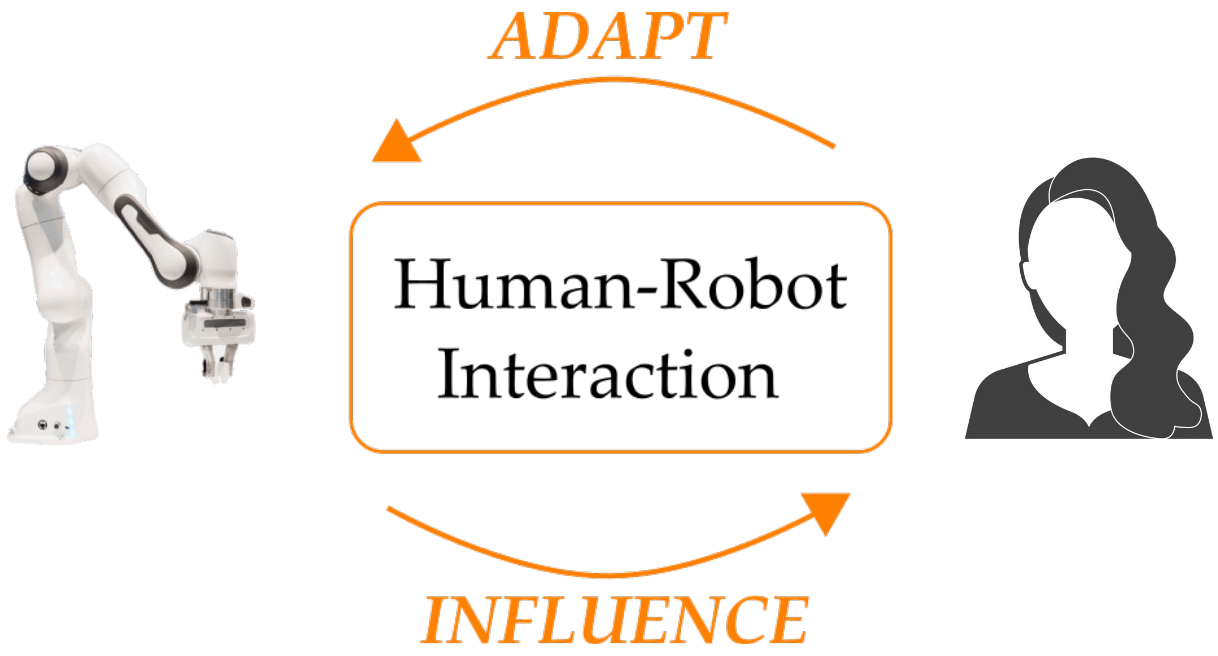
Learning from Humans



Existing research explores how robots ***adapt*** to humans

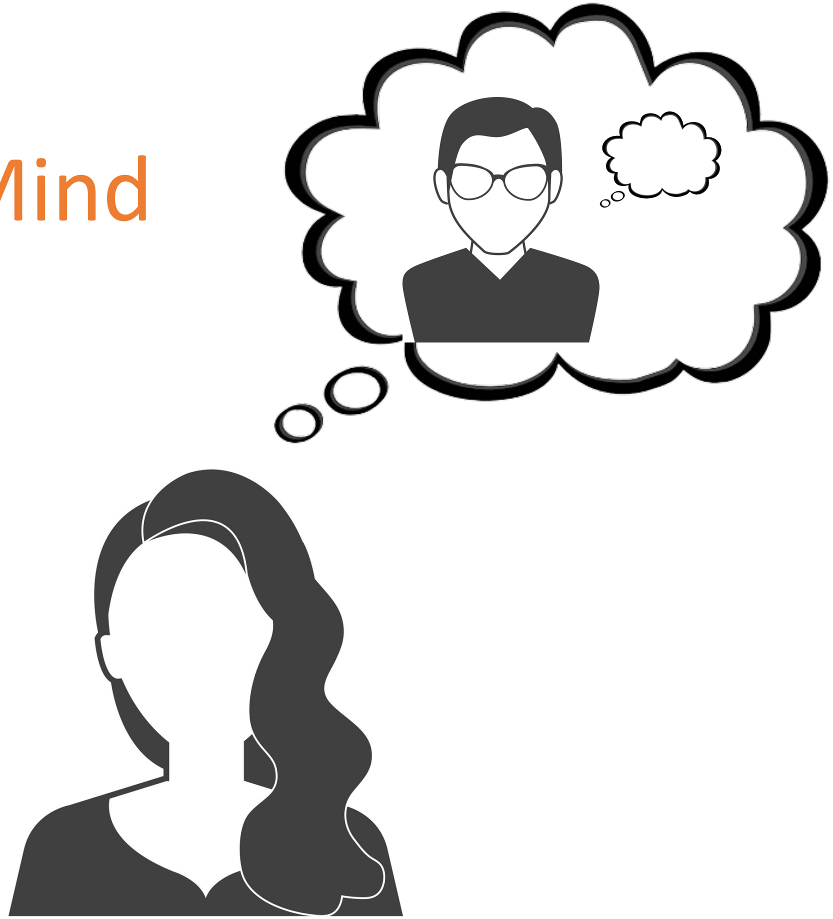
- Imitation learning
- Learning from demonstrations

Influencing Humans

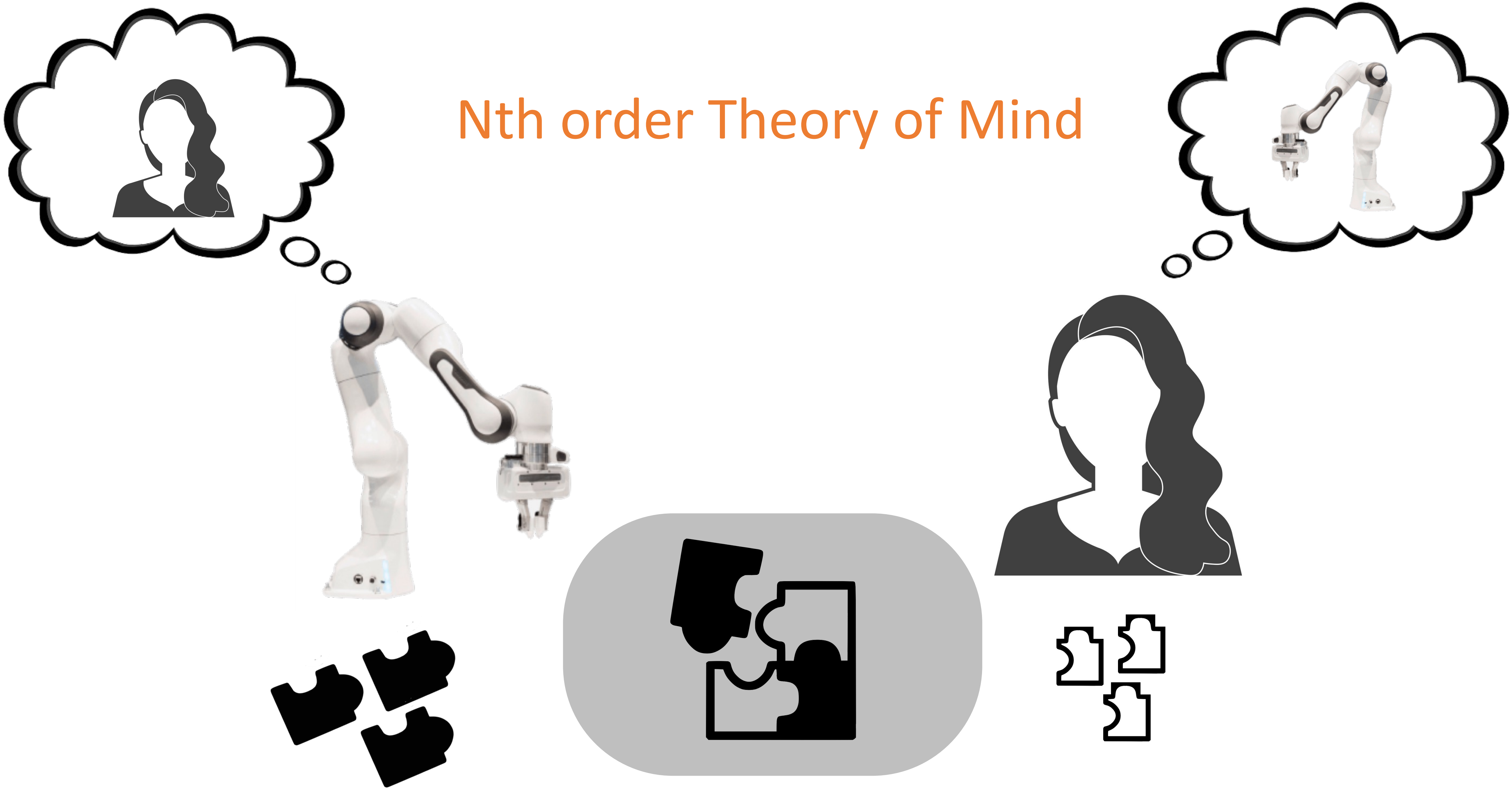


Far less studies how robots ***influence*** humans

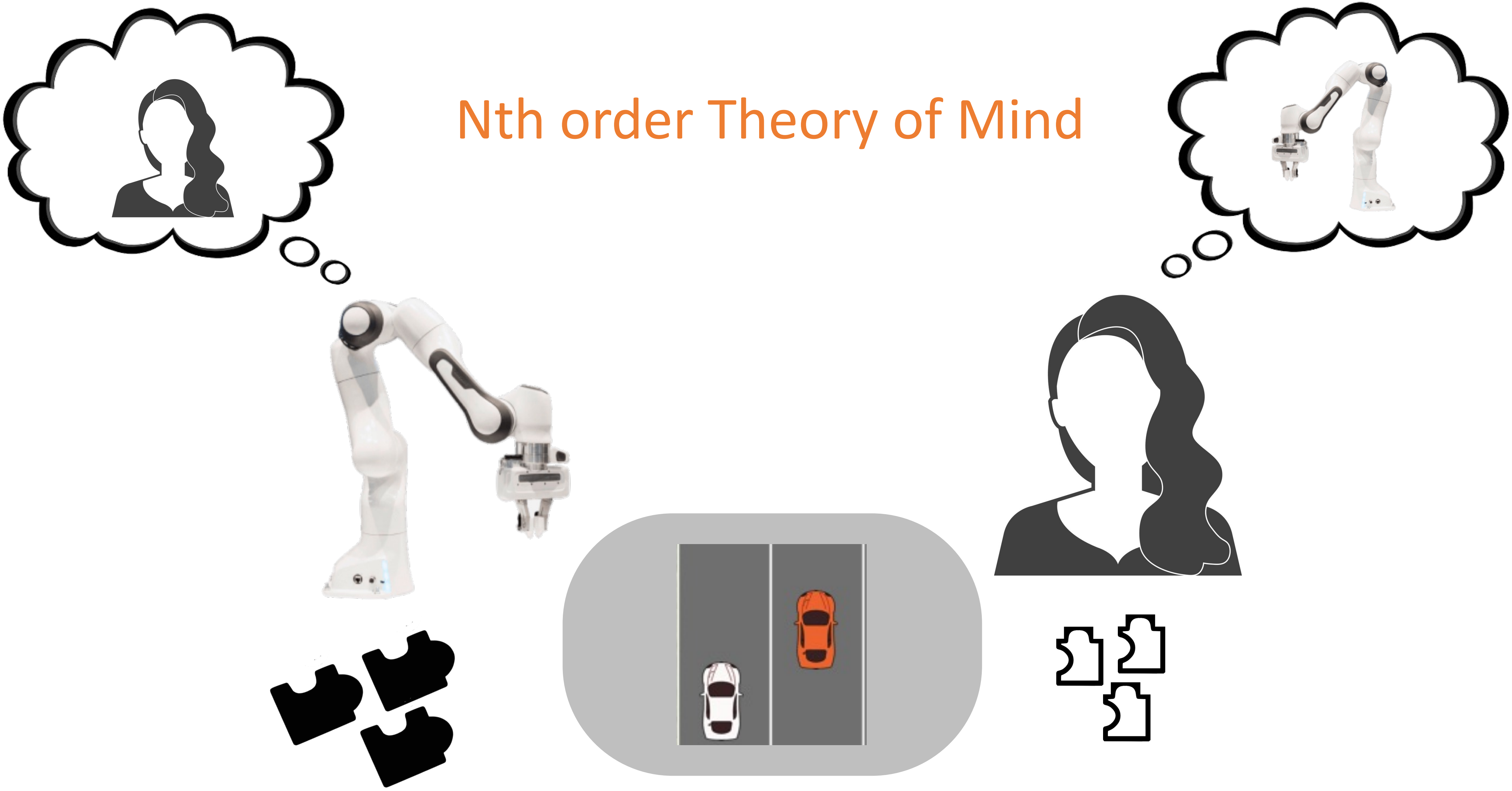
Nth order Theory of Mind



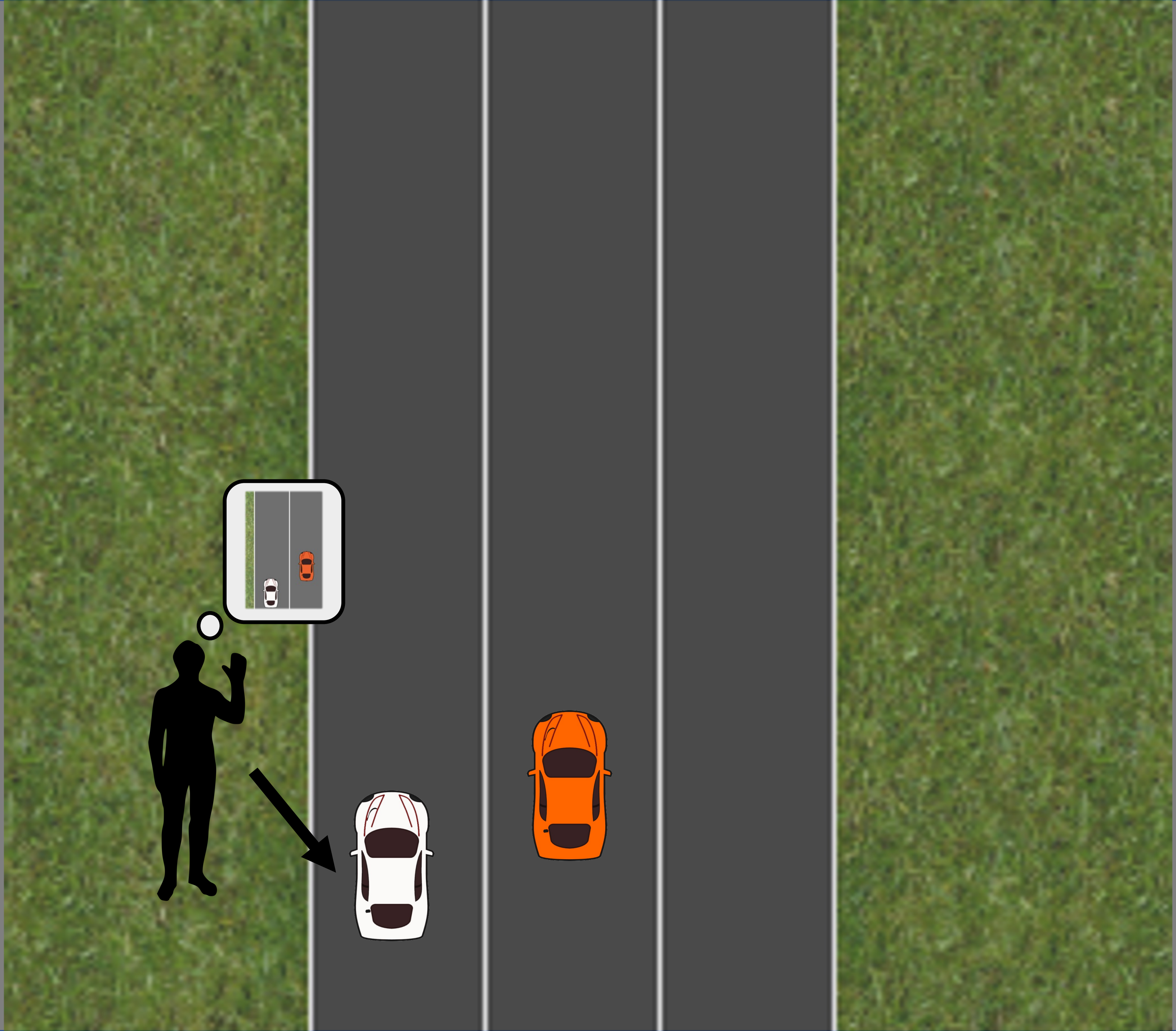
Nth order Theory of Mind



Nth order Theory of Mind



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



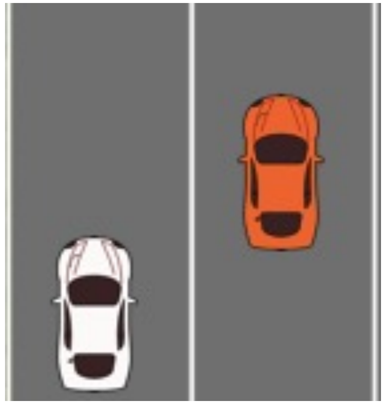
An autonomous car's
actions will *affect* the actions of other
drivers.


FORTUNER
civic

Source: https://twitter.com/nitguptaa/

Interaction as a Dynamical System

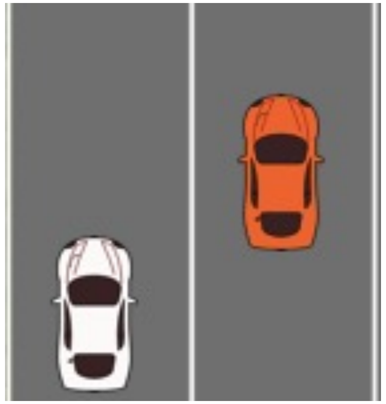
direct control over $u_{\mathcal{R}}$



indirect control over $u_{\mathcal{H}}$

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 $R_{\mathcal{H}}$.

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

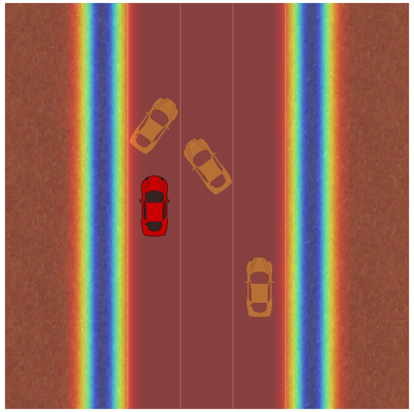
Sadigh et al. RSS 2016, AURO 2018

Learning Driver Models

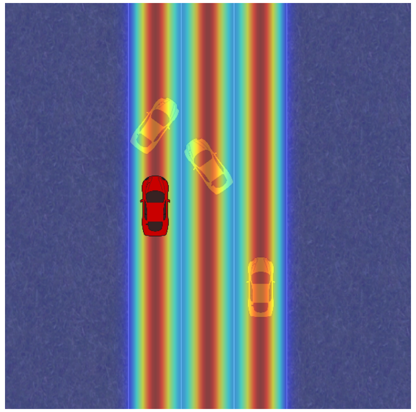
Learn Human's reward function based on Inverse Reinforcement Learning:

$$P(u_{\mathcal{H}} | x, w) = \frac{\exp(R_{\mathcal{H}}(x, u_{\mathcal{R}}, u_{\mathcal{H}}))}{\int \exp\big(R_{\mathcal{H}}(x, u_{\mathcal{R}}, \tilde{u}_{\mathcal{H}})\big)\, d\, \tilde{u}_{\mathcal{H}}}$$

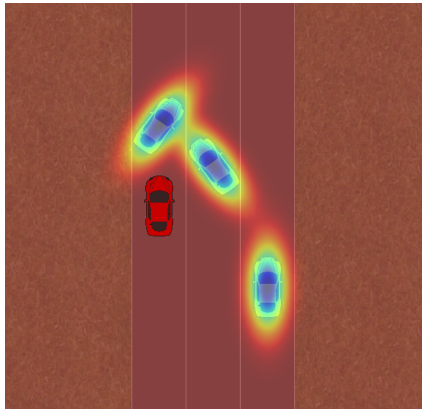
$$R_{\mathcal{H}}(x, u_{\mathcal{R}}, u_{\mathcal{H}}) = w^{\top} \phi(x, u_{\mathcal{R}}, u_{\mathcal{H}})$$



Features for the boundaries of the road.



Features for staying inside the lanes.

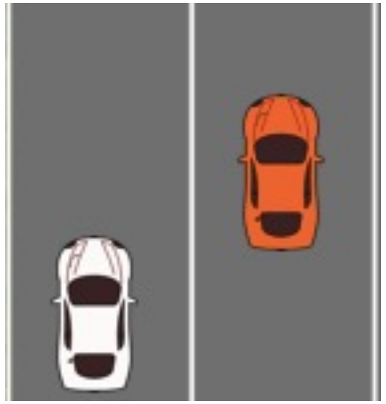


Features for avoiding other vehicles.

[Ziebart' 09] [Levine'10]

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 $R_{\mathcal{H}}$.

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

Approximations for Tractability

- Receding Horizon Control:

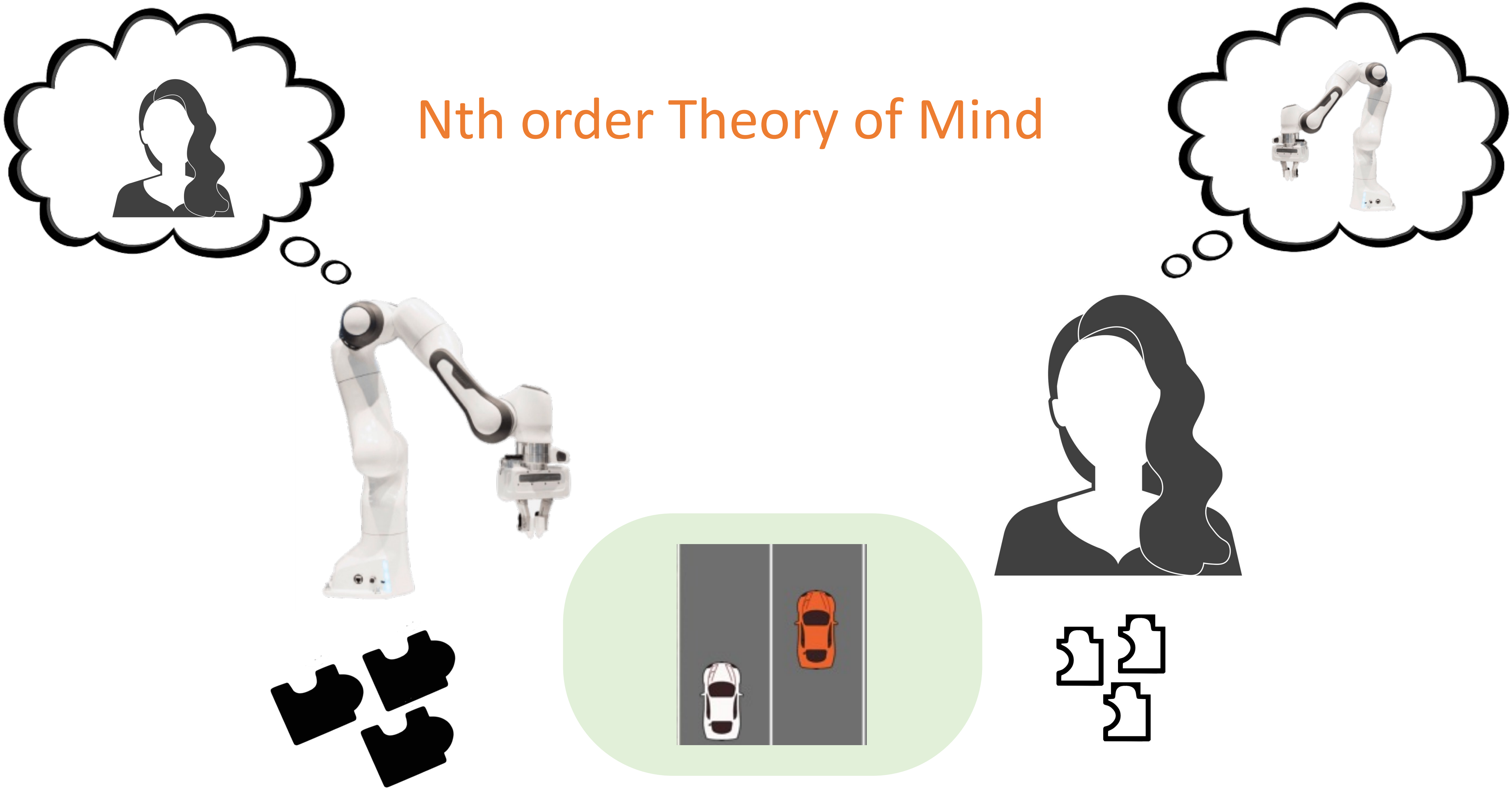
Plan for short time horizon, replan at every step.

- Model the problem as a *Stackelberg game*.

Give the human full access to $u_{\mathcal{R}}$ for the short time horizon.

Nth order Theory of Mind

Nth order Theory of Mind



Approximations for Tractability

- Receding Horizon Control:

Plan for short time horizon, replan at every step.

- Model the problem as a *Stackelberg game*.
Give the human full access to $u_{\mathcal{R}}$ for the short time horizon.

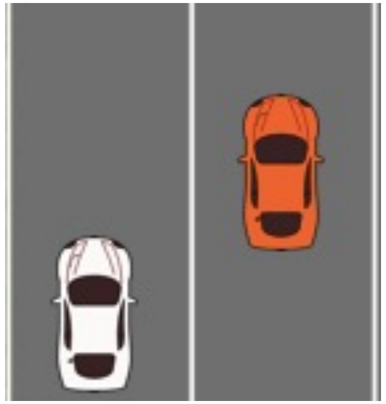
$$u_{\mathcal{H}}^{*}(x, u_{\mathcal{R}}) = \underset{u_{\mathcal{H}}}{\operatorname{argmax}}\, R_{\mathcal{H}}(x, u_{\mathcal{R}}, u_{\mathcal{H}})$$

- Assume deterministic human model.

Solution of Nested Optimization

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

$$R_{\mathcal{R}}(x, u_{\mathcal{R}}, u_{\mathcal{H}}) = \sum_{t=1}^{N} r_{\mathcal{R}}(x^{t}, u_{\mathcal{R}}^{t}, u_{\mathcal{H}}^{t})$$



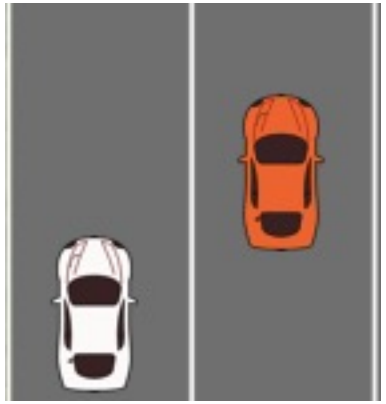
Gradient-Based Method **(Quasi-Newton):**

$$\begin{cases} R_{\mathcal{R}}(x, u_{\mathcal{R}}, u_{\mathcal{H}}^{*}) \\ \dfrac{\partial R_{\mathcal{R}}}{\partial u_{\mathcal{R}}} = \dfrac{\partial R_{\mathcal{R}}}{\partial u_{\mathcal{H}}} \dfrac{\partial u_{\mathcal{H}}^{*}}{\partial u_{\mathcal{R}}} + \dfrac{\partial R_{\mathcal{R}}}{\partial u_{\mathcal{R}}} \end{cases}$$

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

$$R_{\mathcal{H}}(x, u_{\mathcal{R}}, u_{\mathcal{H}}) = \sum_{t=1}^{N} r_{\mathcal{H}}(x^{t}, u_{\mathcal{R}}^{t}, u_{\mathcal{H}}^{t})$$

Solution of Nested Optimization



Given $R_{\mathcal{H}}$ is:,
- smooth,
- its minimum is attained,

for an *unconstrained optimization*, the partial $\frac{\partial R_{\mathcal{H}}}{\partial u_{\mathcal{H}}}$ at the optimum $u_{\mathcal{H}}^*$ evaluates to zero.

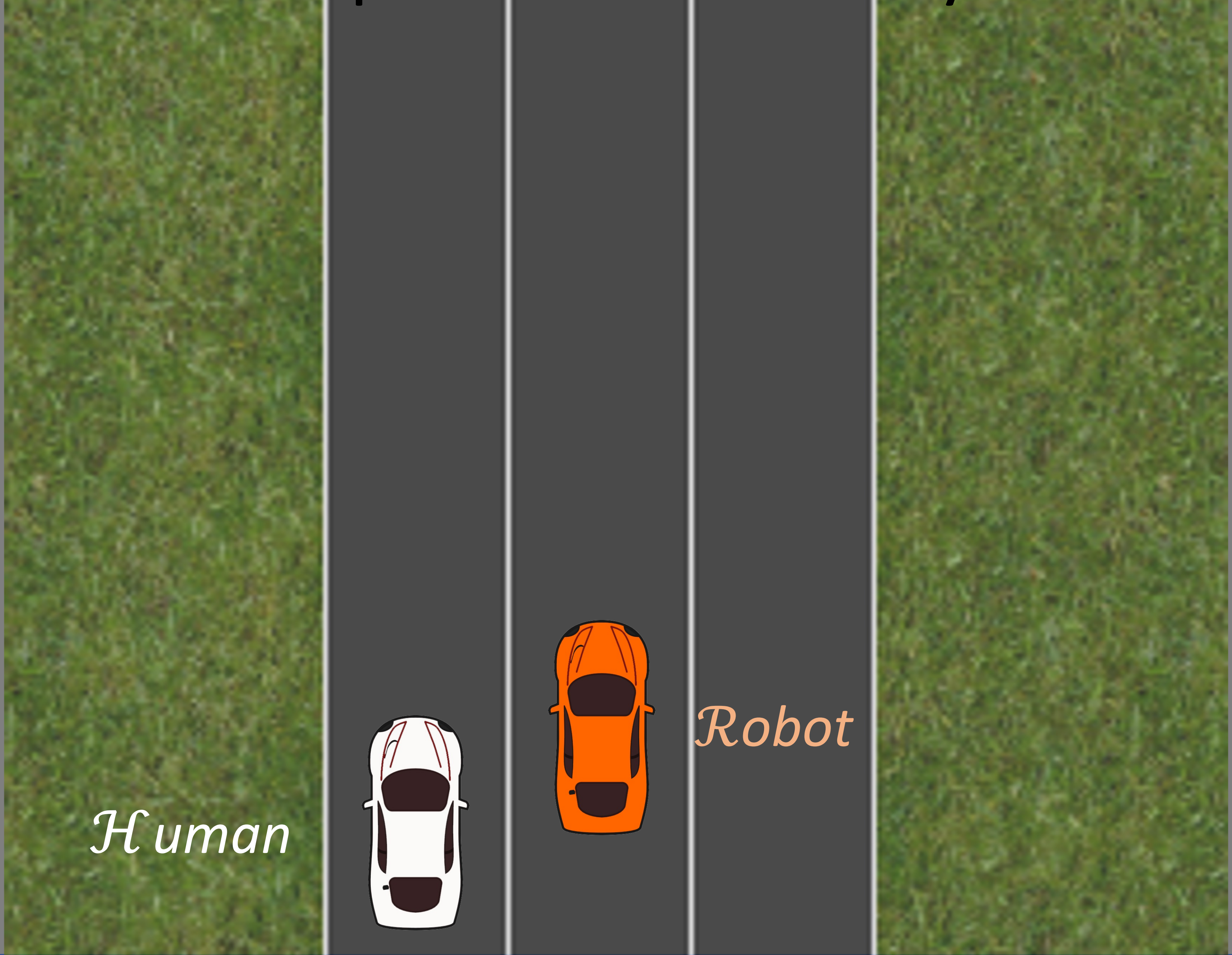
Quasi-Newton method:

$$\frac{\partial R_{\mathcal{R}}}{\partial u_{\mathcal{R}}} = \frac{\partial R_{\mathcal{R}}}{\partial u_{\mathcal{H}}} \cdot \boxed{\frac{\partial u_{\mathcal{H}}^*}{\partial u_{\mathcal{R}}}} + \frac{\partial R_{\mathcal{R}}}{\partial u_{\mathcal{R}}}$$

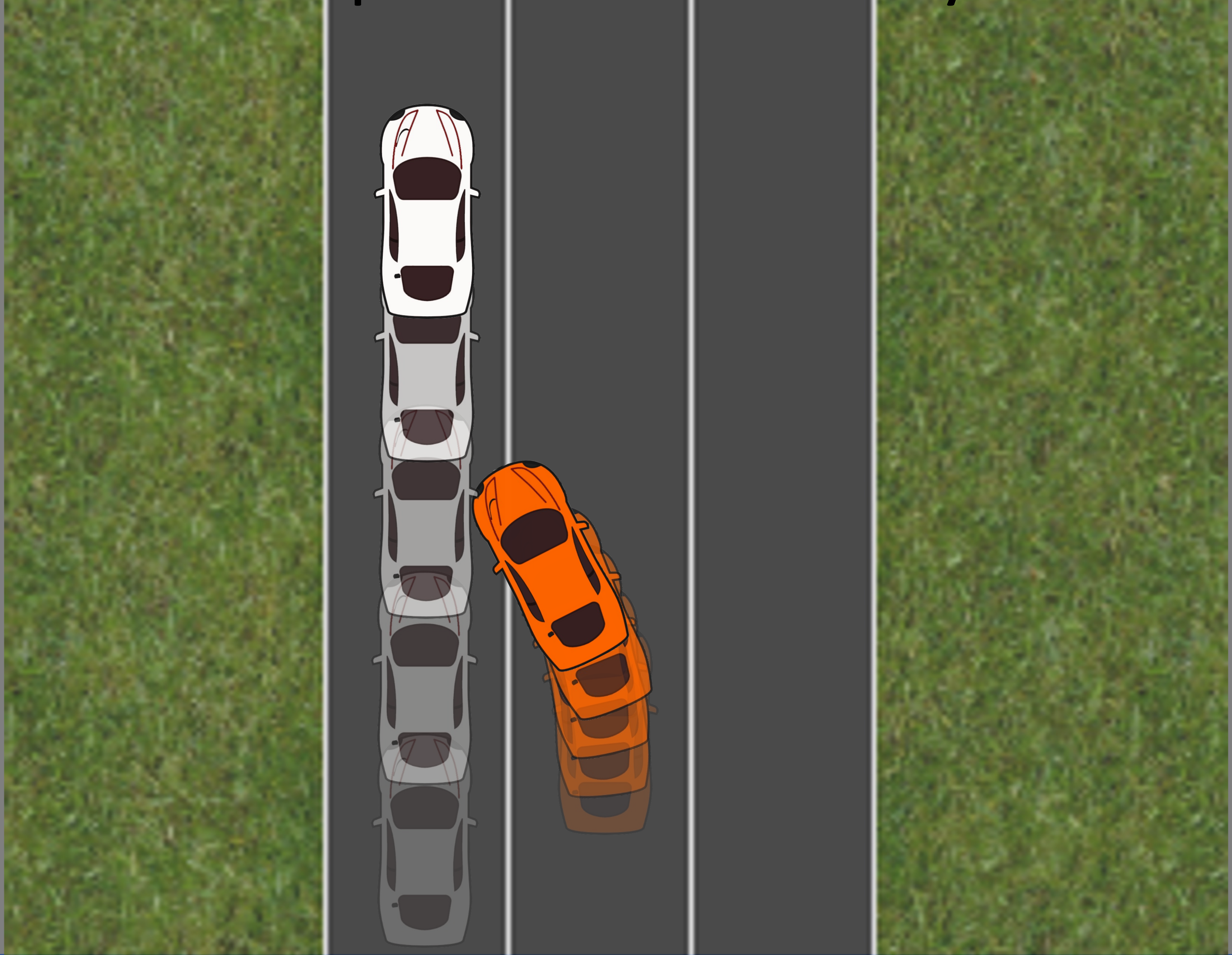
$$\boxed{\frac{\partial R_{\mathcal{H}}}{\partial u_{\mathcal{H}}}\left(x, u_{\mathcal{R}}, u_{\mathcal{H}}^*(x, u_{\mathcal{R}})\right) = 0}$$

$$\frac{\partial^2 R_{\mathcal{H}}}{\partial u_{\mathcal{H}}^2} \cdot \boxed{\frac{\partial u_{\mathcal{H}}^*}{\partial u_{\mathcal{R}}}} + \frac{\partial^2 R_{\mathcal{H}}}{\partial u_{\mathcal{H}} \partial u_{\mathcal{R}}} \cdot \frac{\partial u_{\mathcal{R}}}{\partial u_{\mathcal{R}}} = 0$$

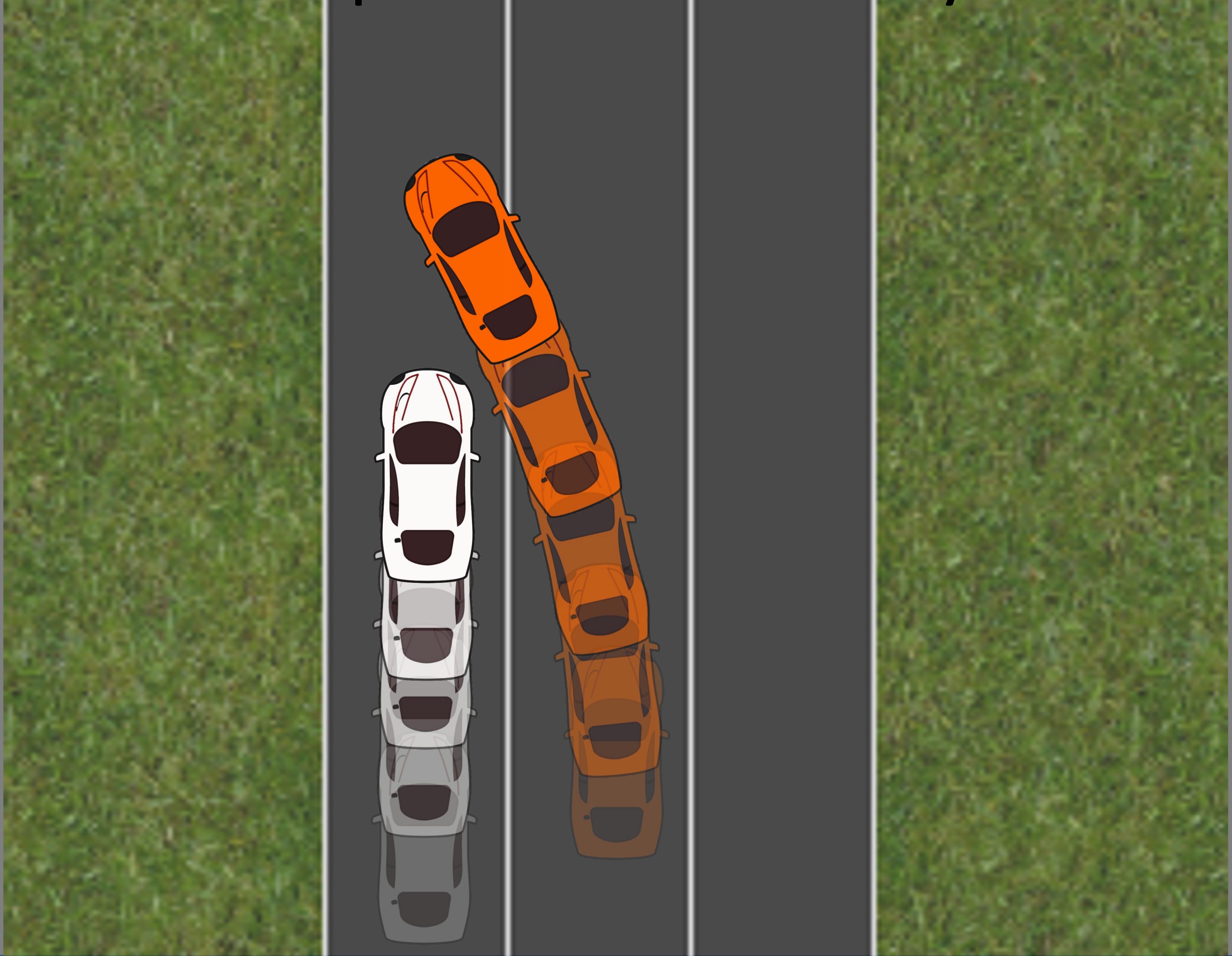
Implication: Efficiency



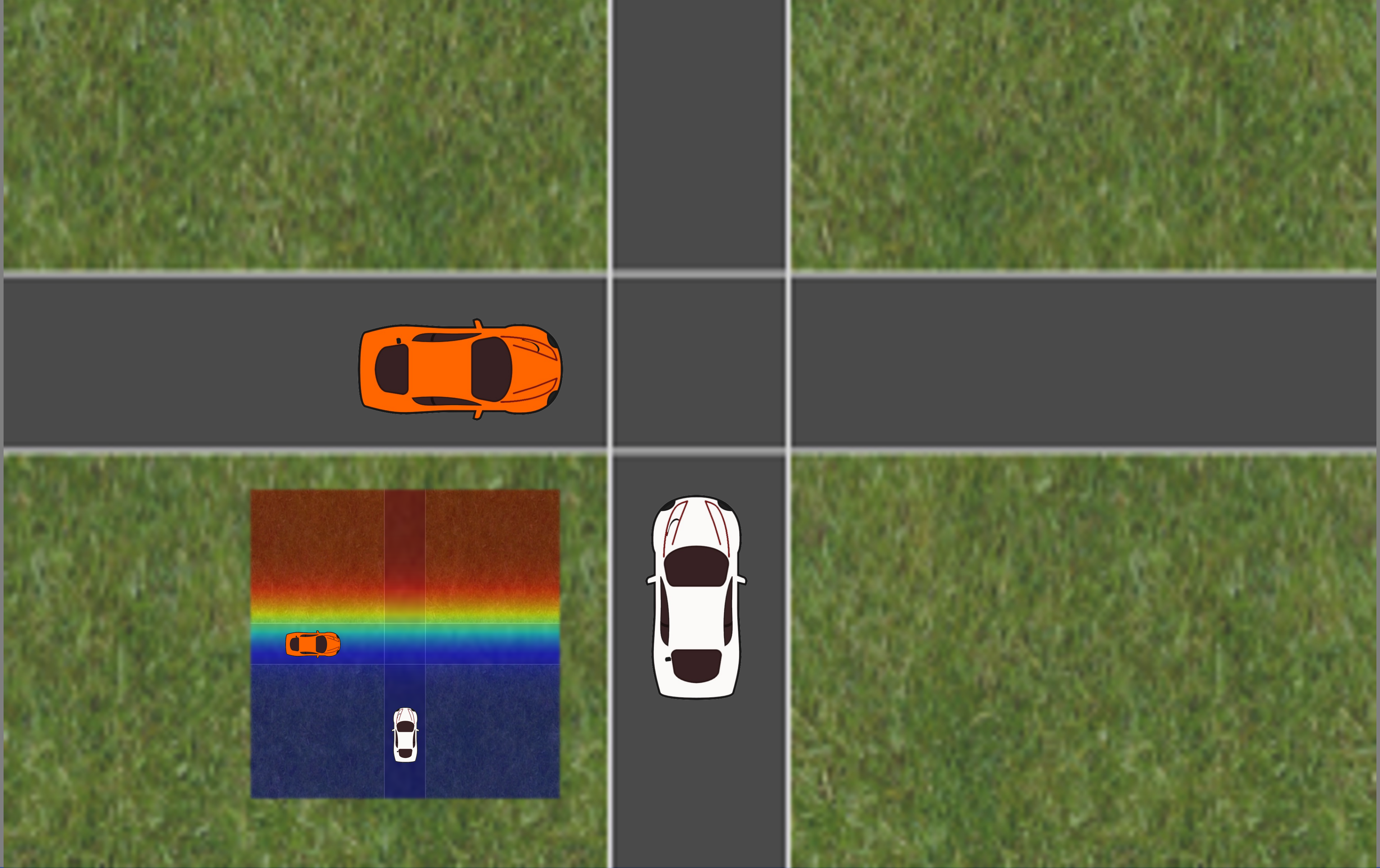
Implication: Efficiency



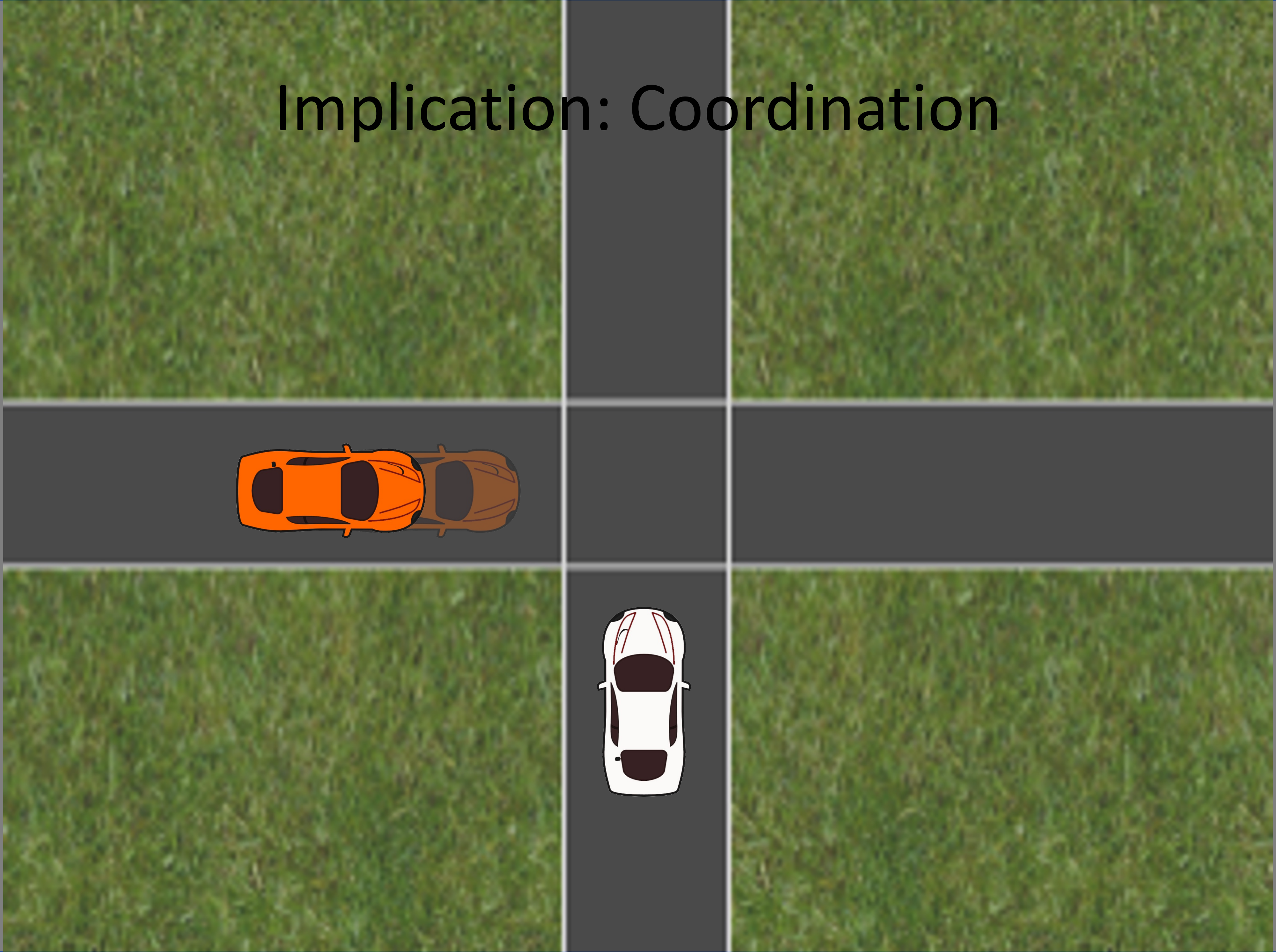
Implication: Efficiency



Implication: Coordination

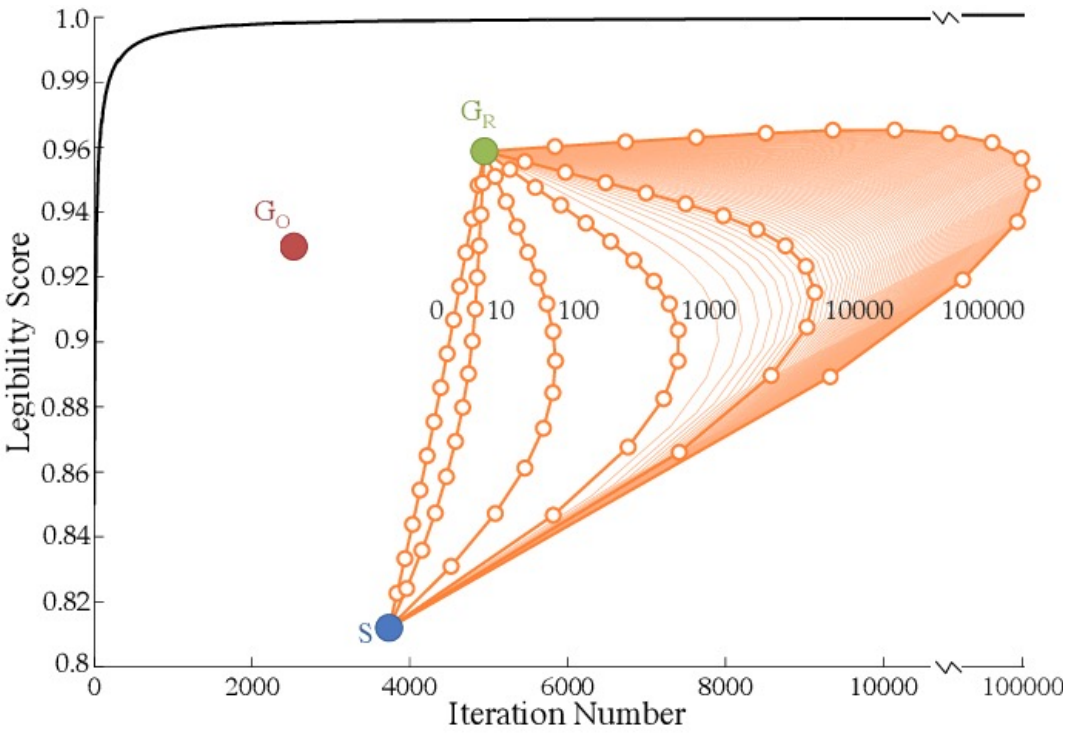


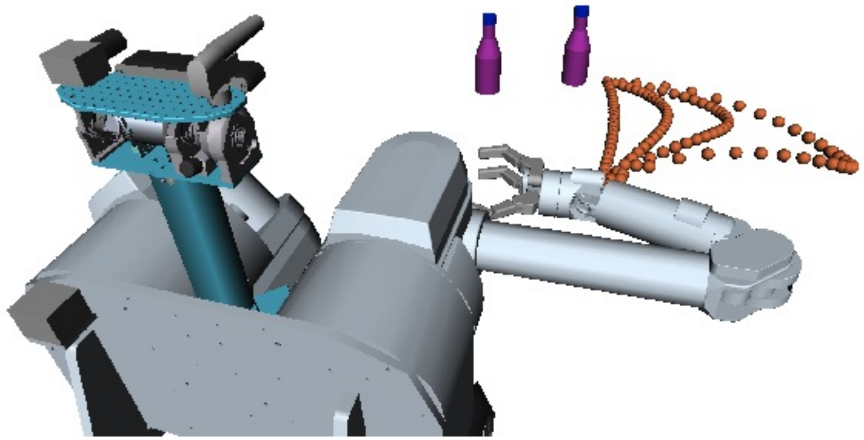
Implication: Coordination

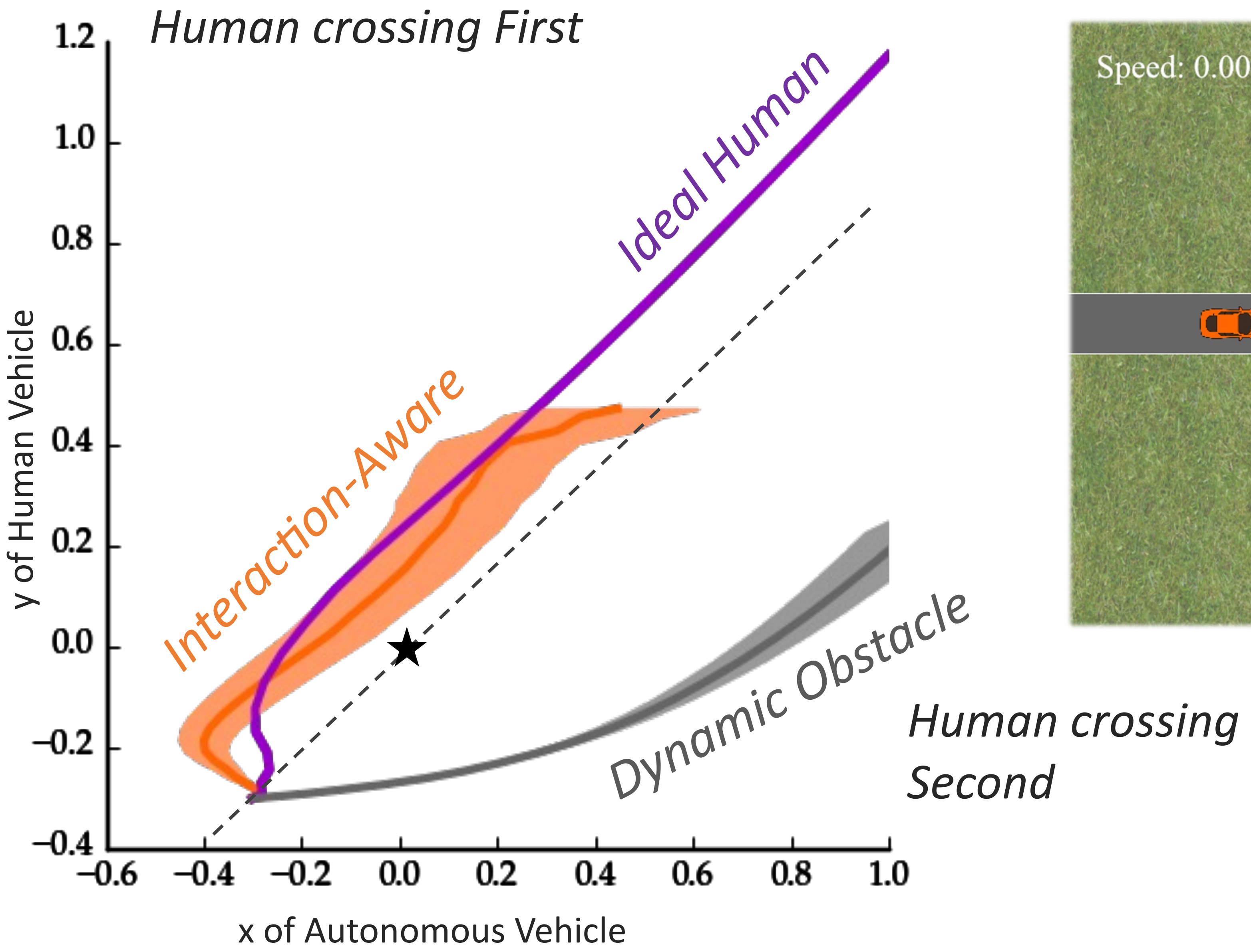


Legible Motion

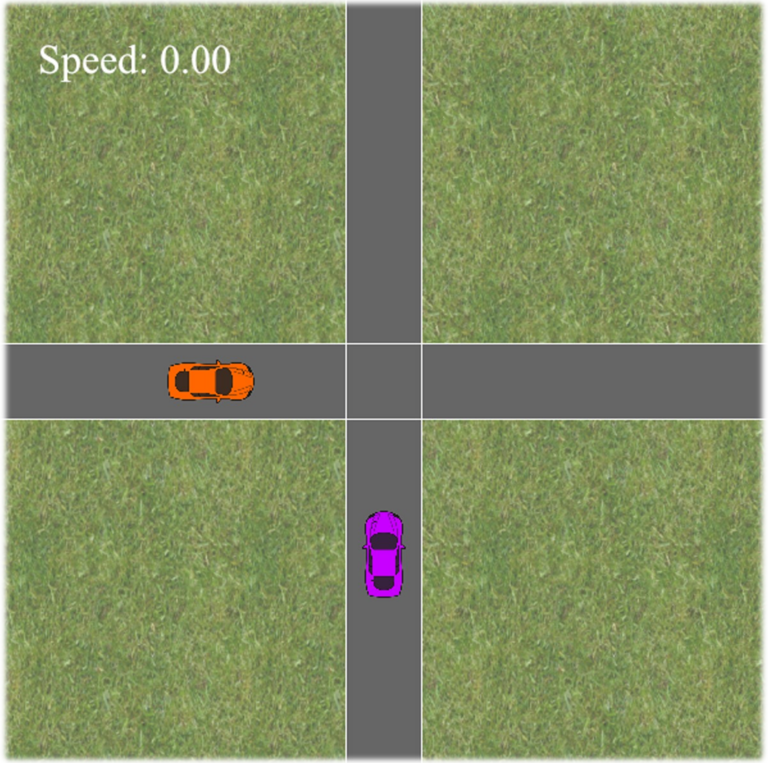
Using robot motion to coordinate with the human better about the robot's goal



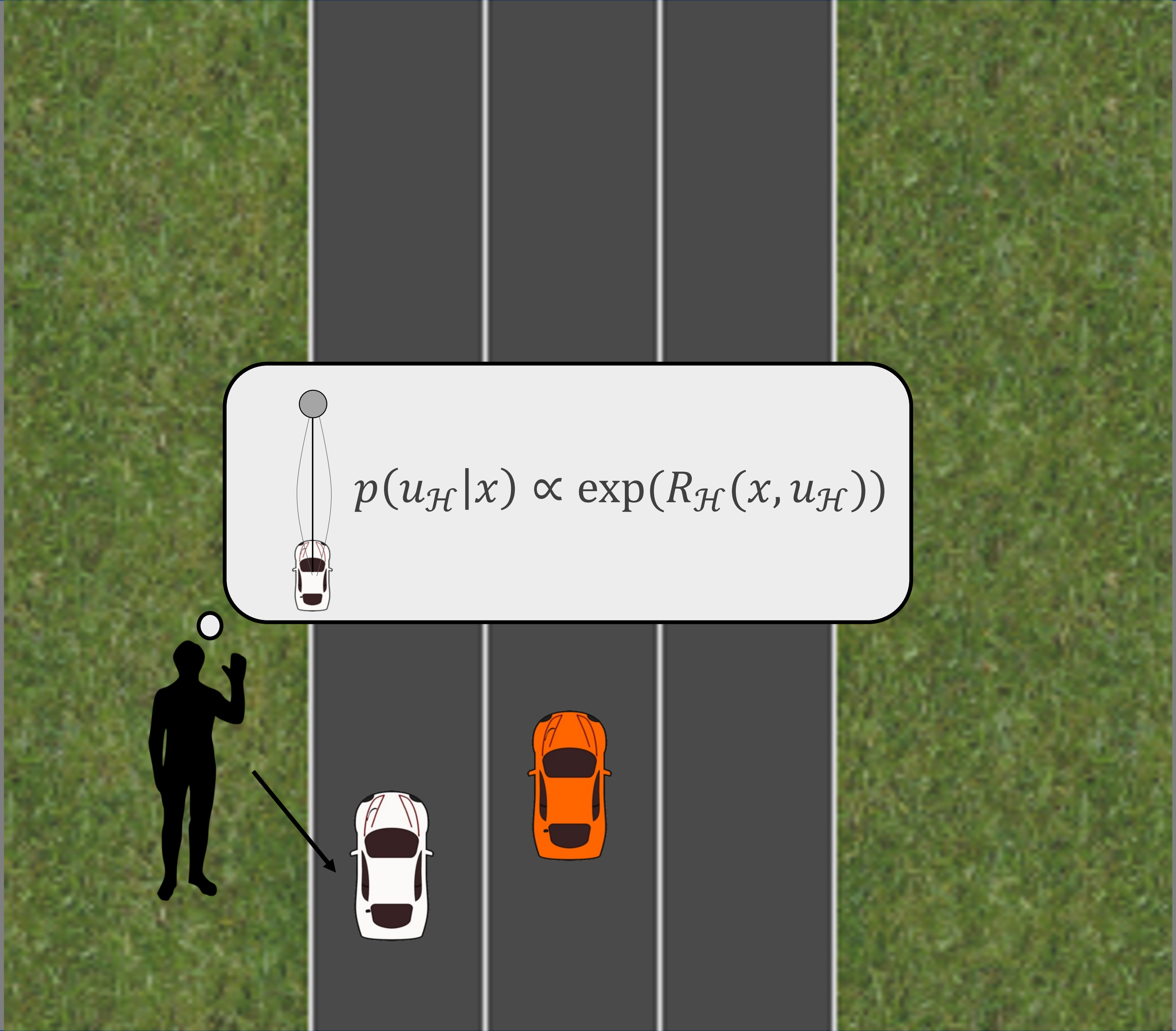




Human crossing First
Ideal Human
Interaction-Aware
Dynamic Obstacle
Human crossing
Second
y of Human Vehicle
x of Autonomous Vehicle
1.2
1.0
0.8
0.6
0.4
0.2
0.0
−0.2
−0.4
−0.6
−0.4
−0.2
0.0
0.2
0.4
0.6
0.8
1.0



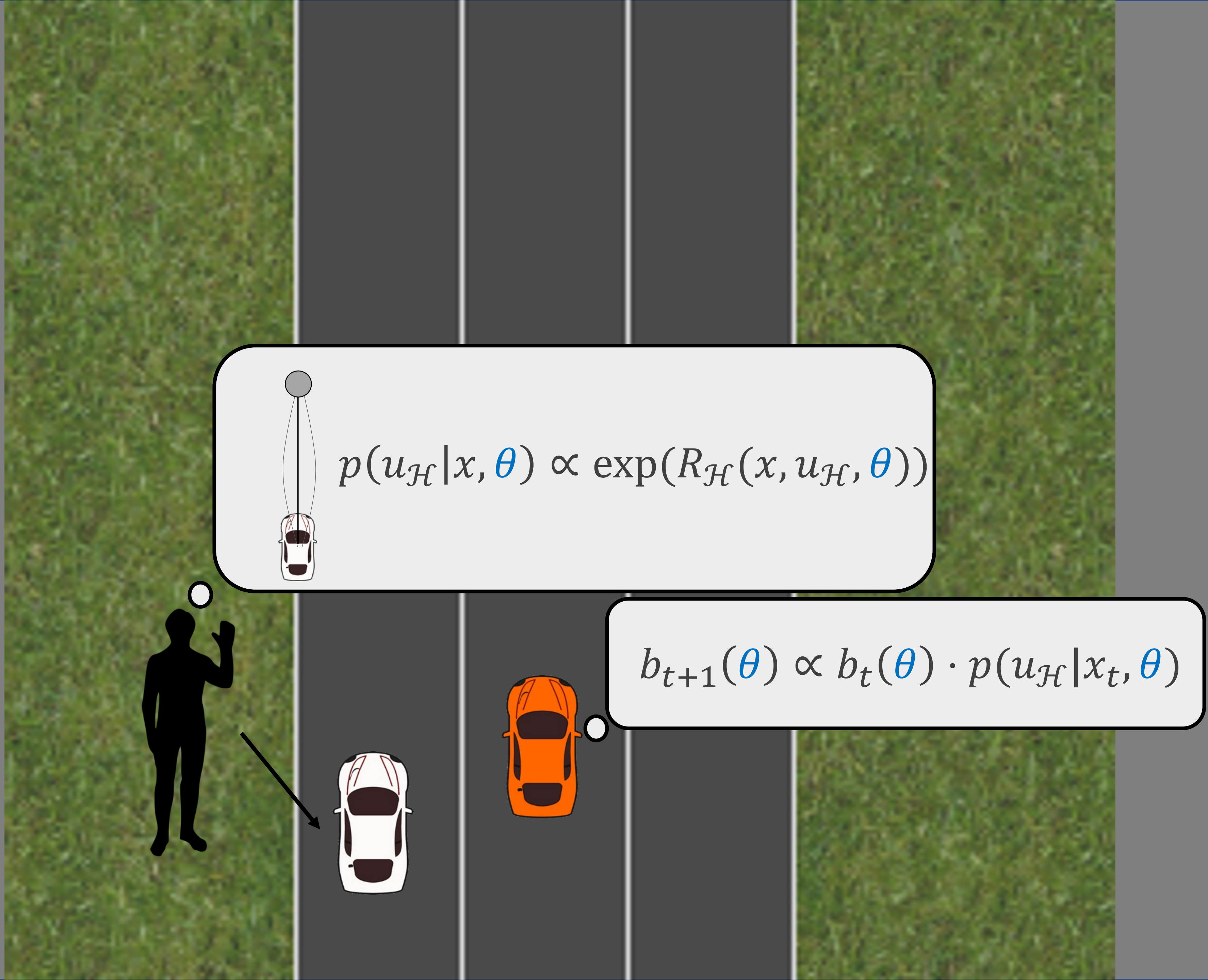
Speed: 0.00


$p(u_{\mathcal{H}}|x) \propto \exp(R_{\mathcal{H}}(x, u_{\mathcal{H}}))$

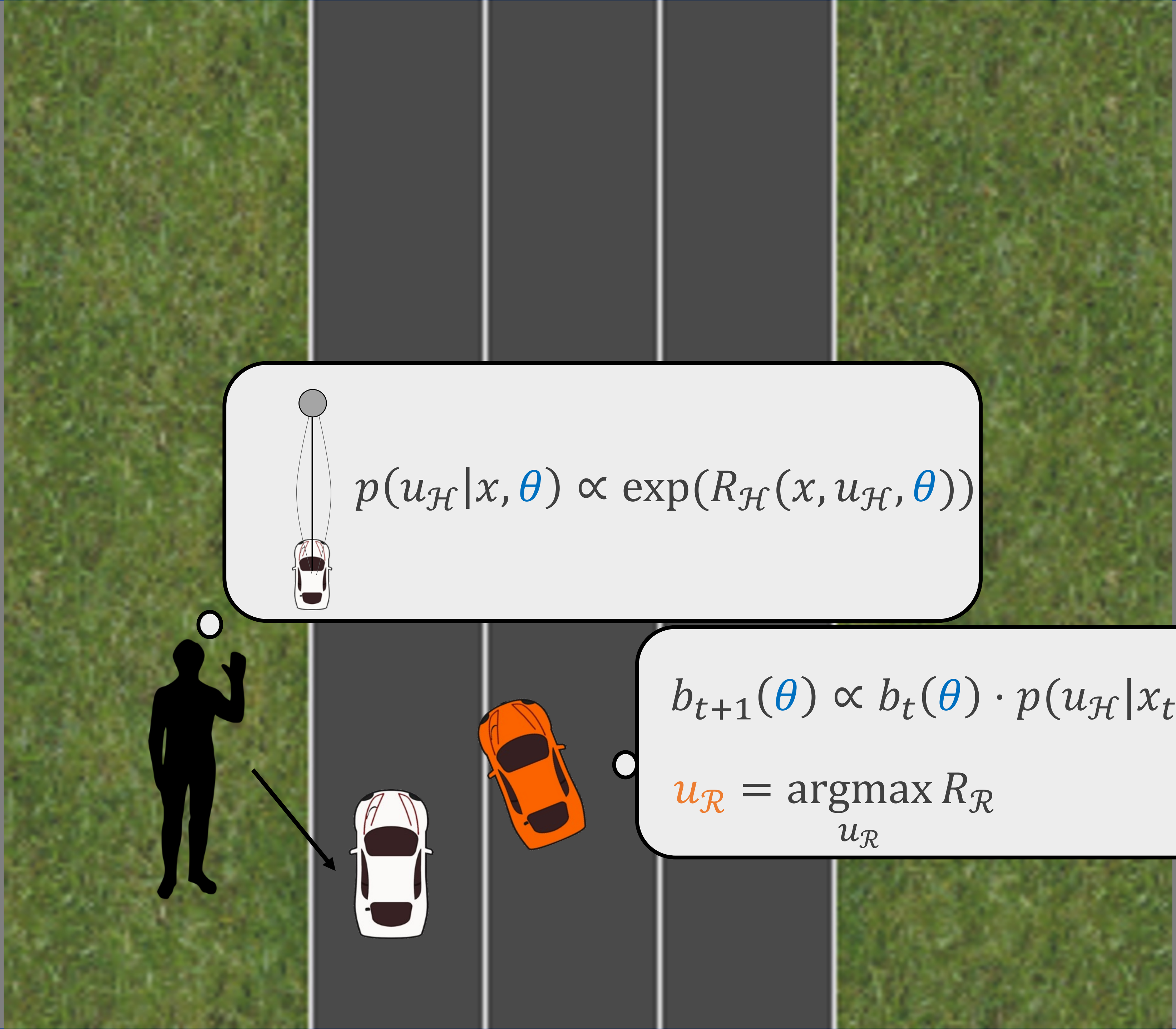

SPEED
SOCIETY
2015/02/06 23:10:05

We can’t rely on a
single driver model.

We need to ***differentiate***
between different drivers.



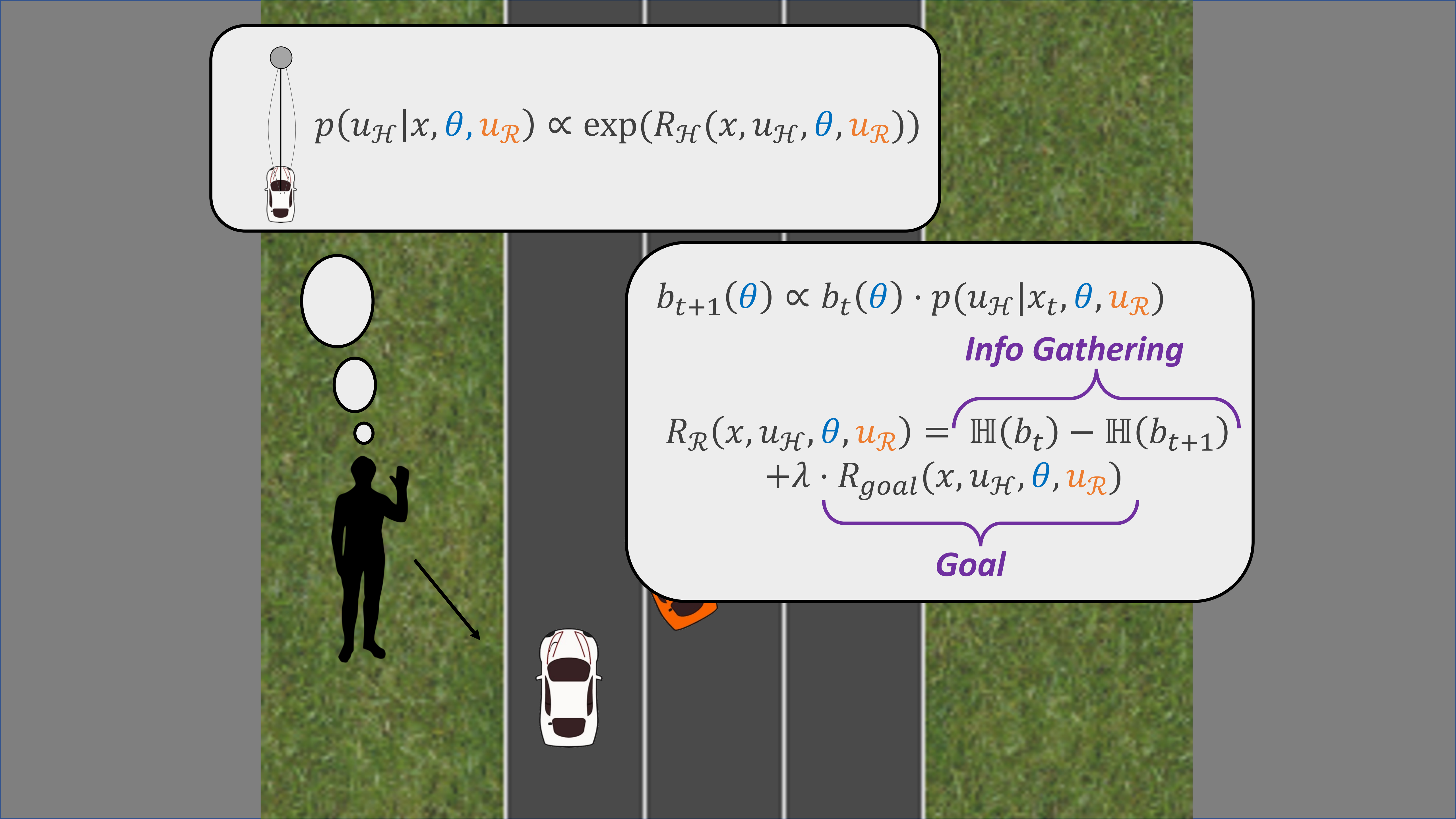
$p(u_{\mathcal{H}}|x,\theta) \propto \exp(R_{\mathcal{H}}(x,u_{\mathcal{H}},\theta))$
$b_{t+1}(\theta) \propto b_t(\theta) \cdot p(u_{\mathcal{H}}|x_t,\theta)$

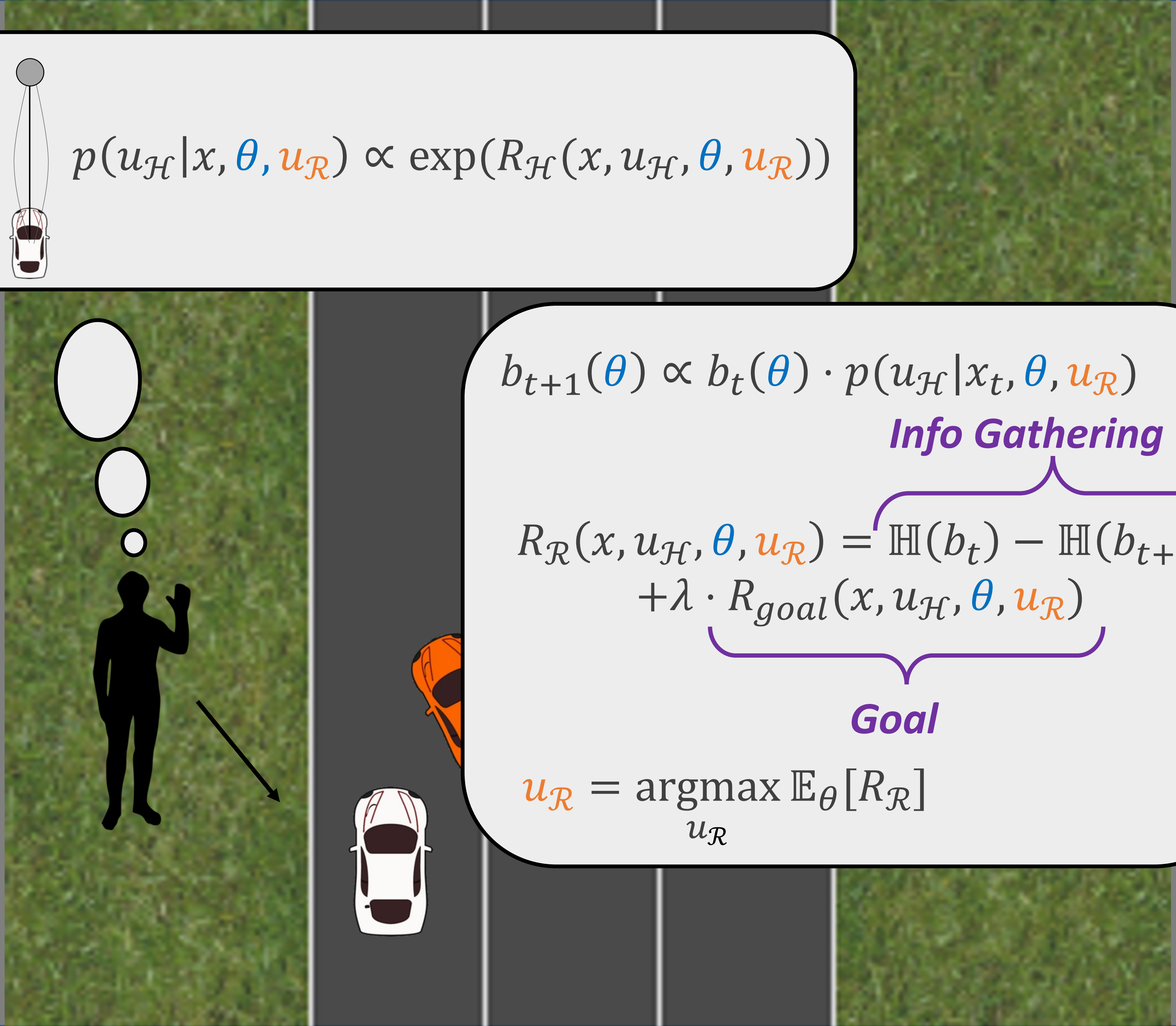

$p(u_{\mathcal{H}} | x, \theta) \propto \exp(R_{\mathcal{H}}(x, u_{\mathcal{H}}, \theta))$
$b_{t+1}(\theta) \propto b_t(\theta) \cdot p(u_{\mathcal{H}} | x_t, \theta)$
$u_{\mathcal{R}} = \underset{u_{\mathcal{R}}}{\operatorname{argmax}} R_{\mathcal{R}}$


FORTUNER
CIVIC

Drivers ***respond*** to
actions of other cars.

…We have an opportunity to
actively gather information.


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

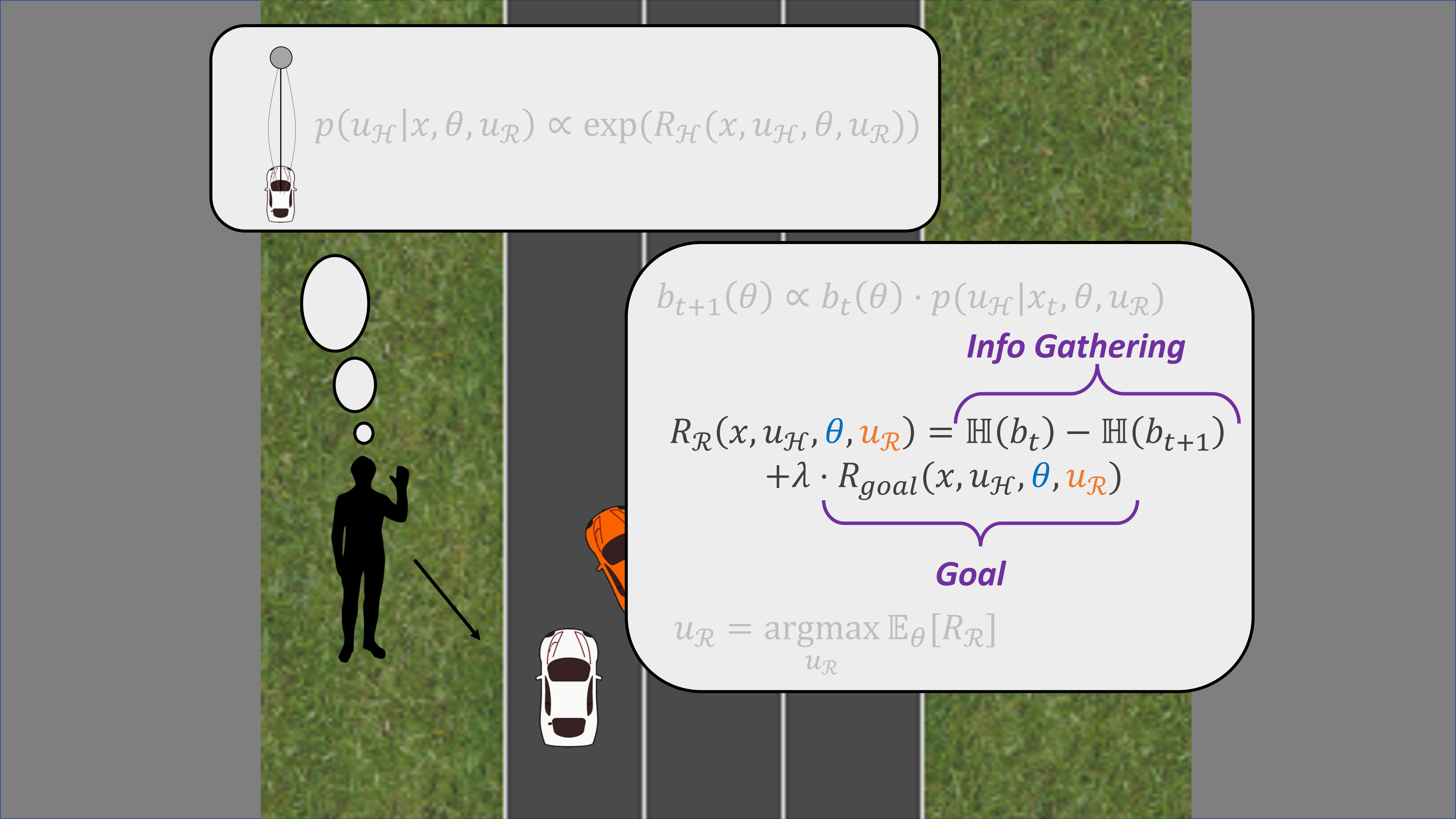


$$p(u_{\mathcal{H}} | x, \theta, u_{\mathcal{R}}) \propto \exp(R_{\mathcal{H}}(x, u_{\mathcal{H}}, \theta, u_{\mathcal{R}}))$$

$$b_{t+1}(\theta) \propto b_t(\theta) \cdot p(u_{\mathcal{H}} | x_t, \theta, u_{\mathcal{R}})$$

$$R_{\mathcal{R}}(x, u_{\mathcal{H}}, \theta, u_{\mathcal{R}}) = \overbrace{\mathbb{H}(b_t) - \mathbb{H}(b_{t+1})}^{\textit{Info Gathering}} + \lambda \cdot \underbrace{R_{goal}(x, u_{\mathcal{H}}, \theta, u_{\mathcal{R}})}_{\textit{Goal}}$$

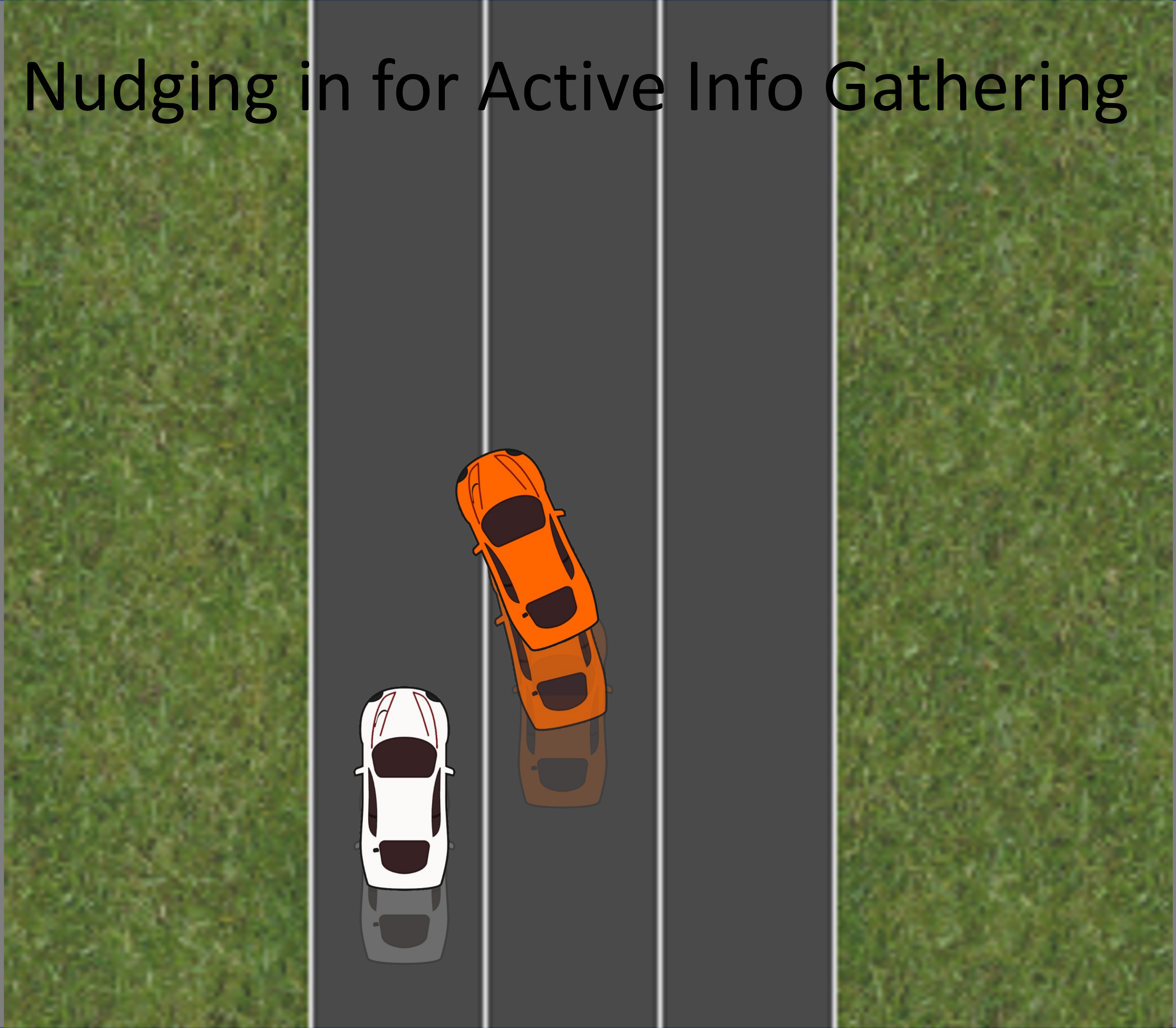
$$u_{\mathcal{R}} = \underset{u_{\mathcal{R}}}{\operatorname{argmax}} \, \mathbb{E}_{\theta}[R_{\mathcal{R}}]$$



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

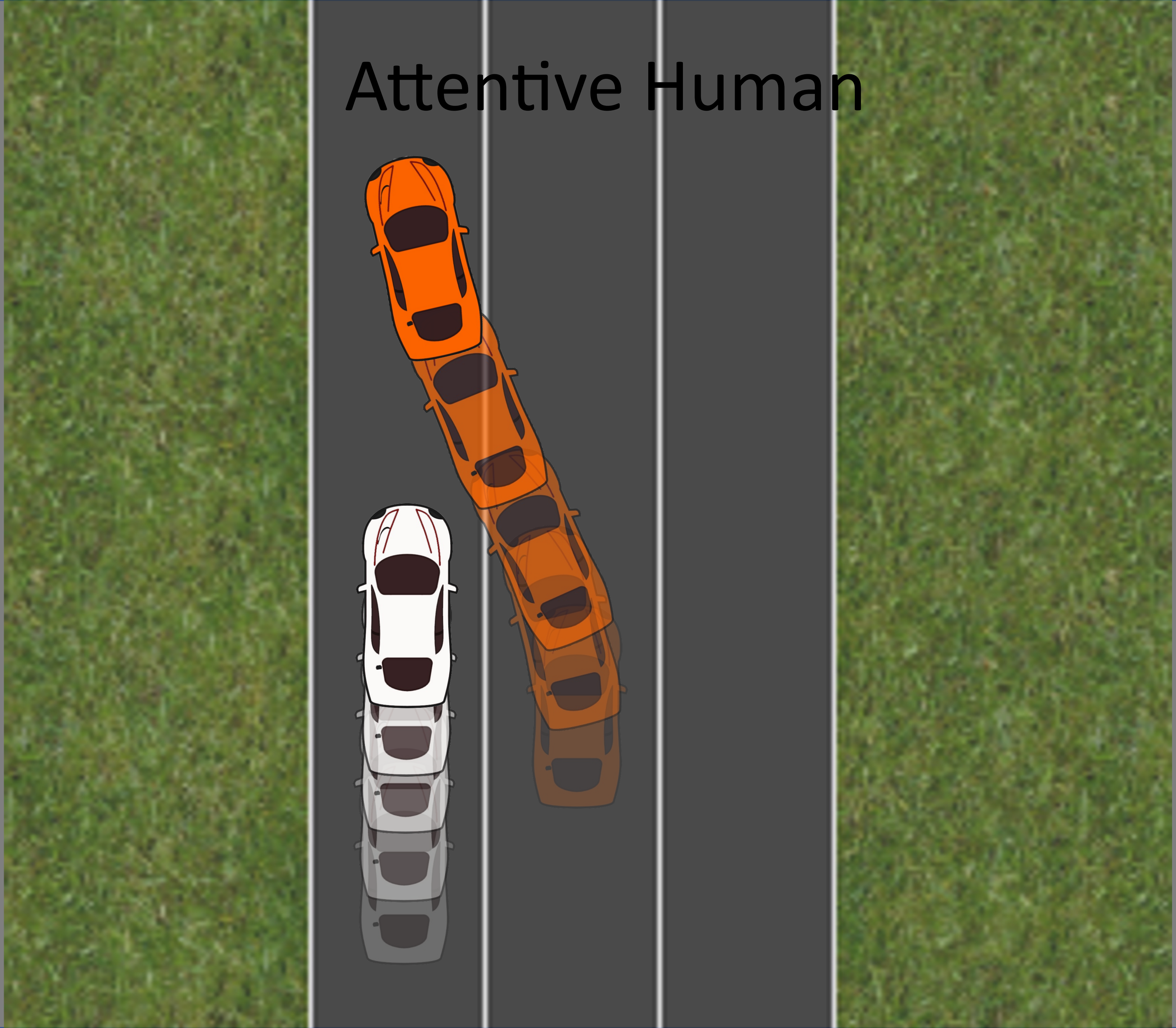
Nudging in for Active Info Gathering

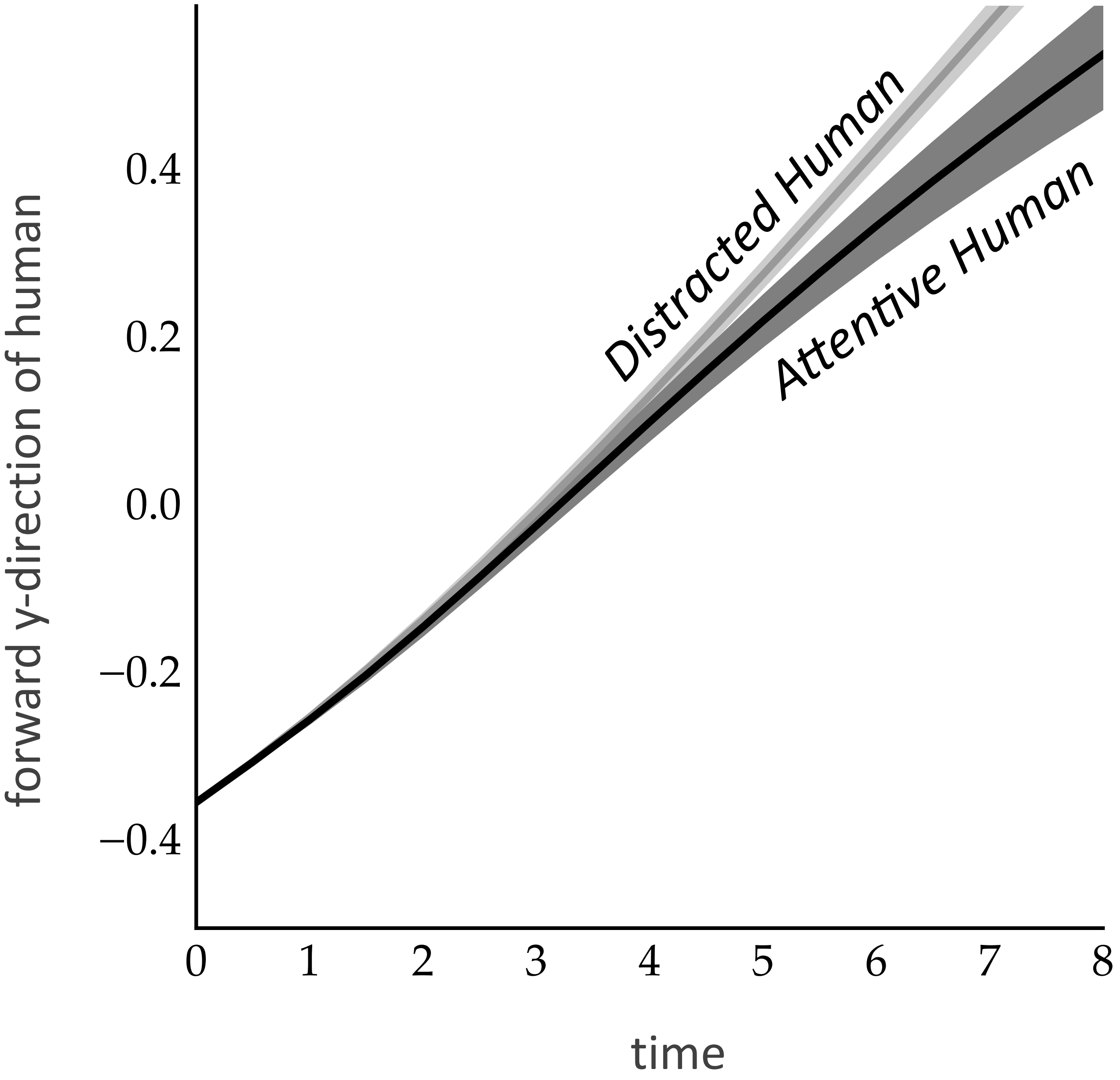
Nudging in for Active Info Gathering



Distracted Human

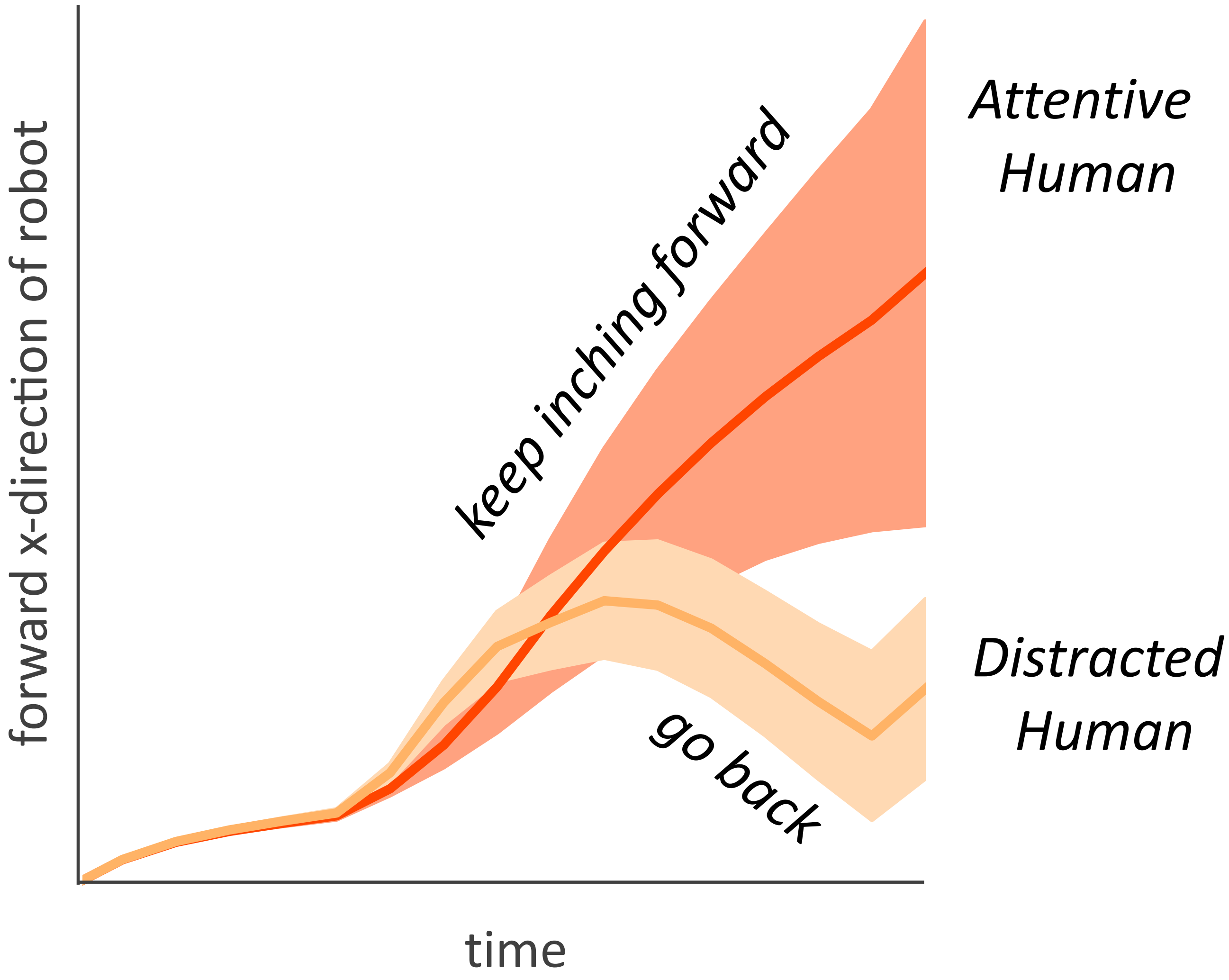
Attentive Human



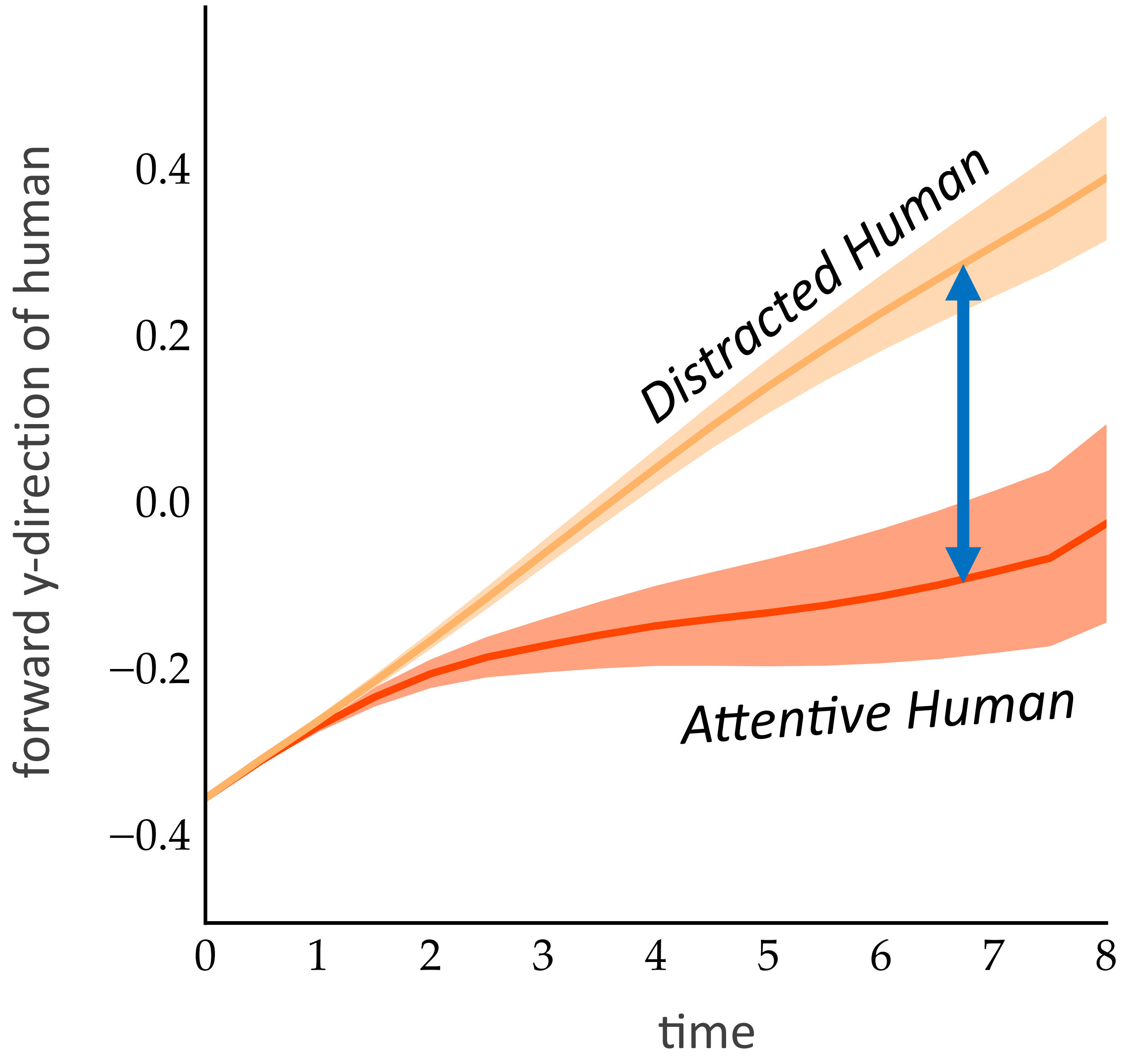


Distracted Human
Attentive Human
0.4
0.2
0.0
−0.2
−0.4
forward y-direction of human
0
1
2
3
4
5
6
7
8
time

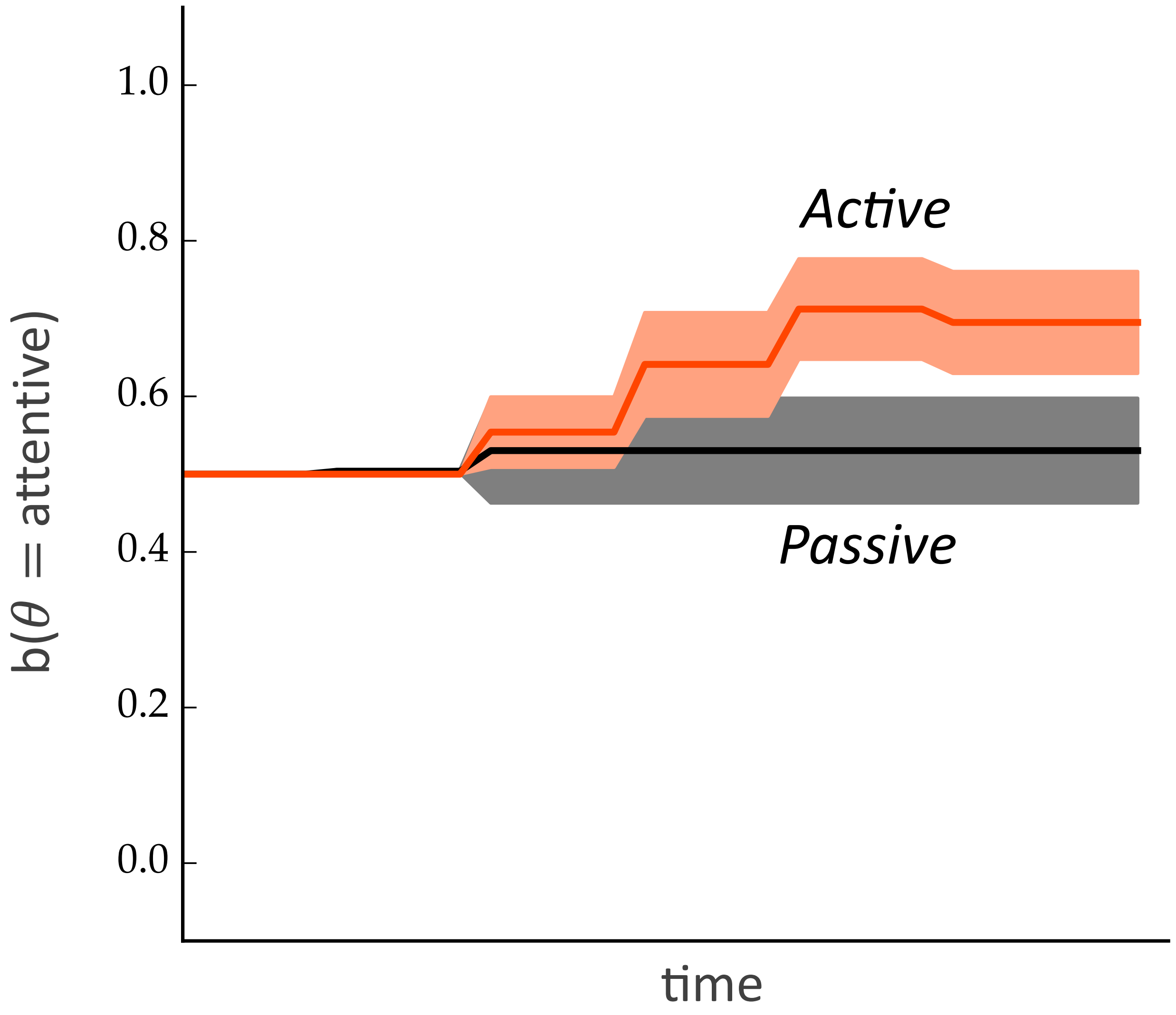
Robot Active Info Gathering



Human Responses



Belief over Driving Style: Active vs Passive



Key Idea:

Robot's actions ***affect*** human's actions. We want to leverage these effects for better safety and efficiency and better estimation.

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