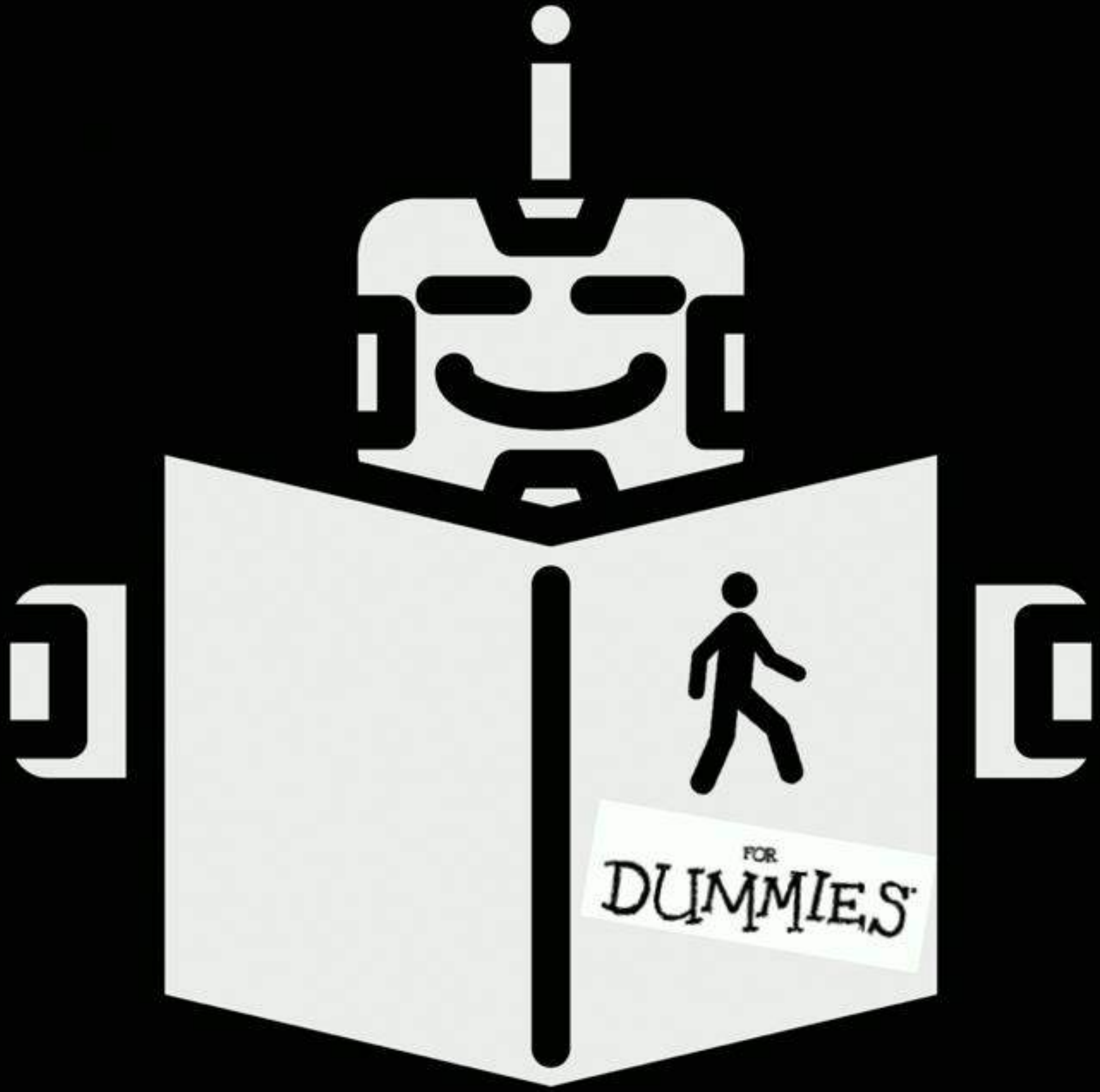
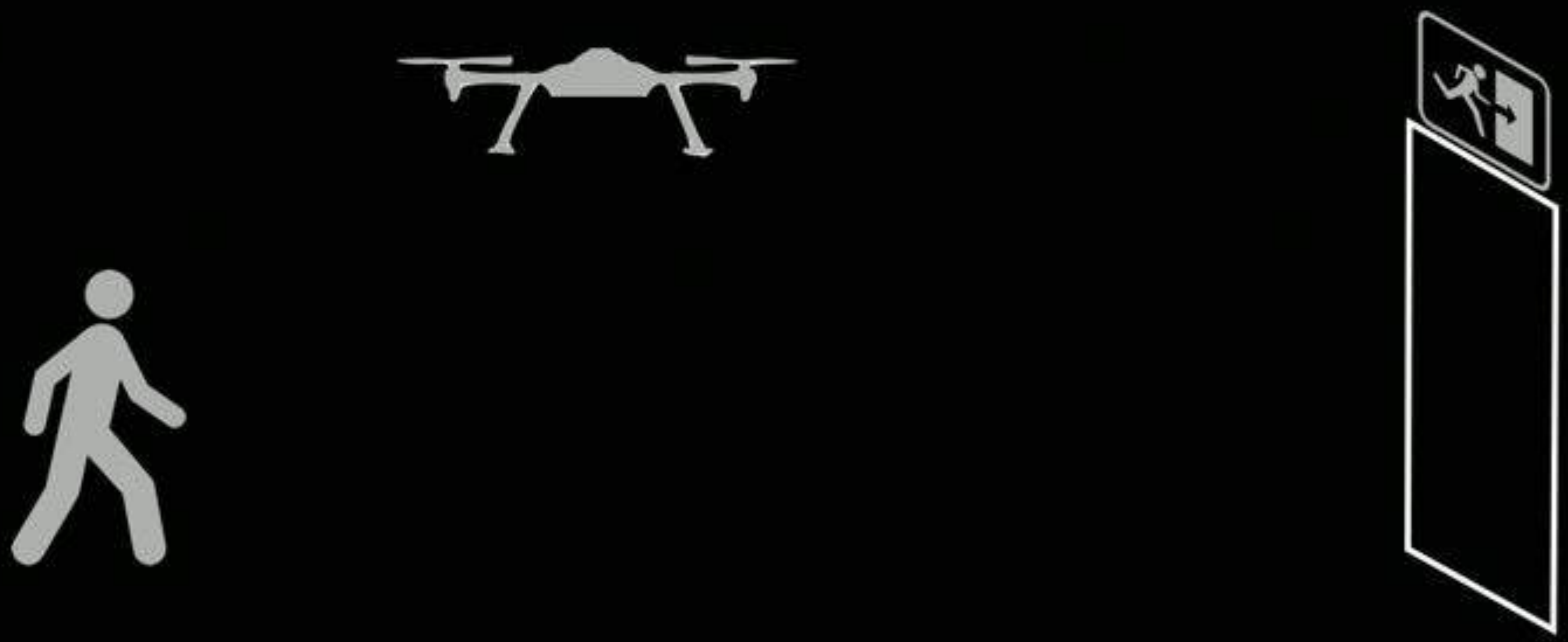
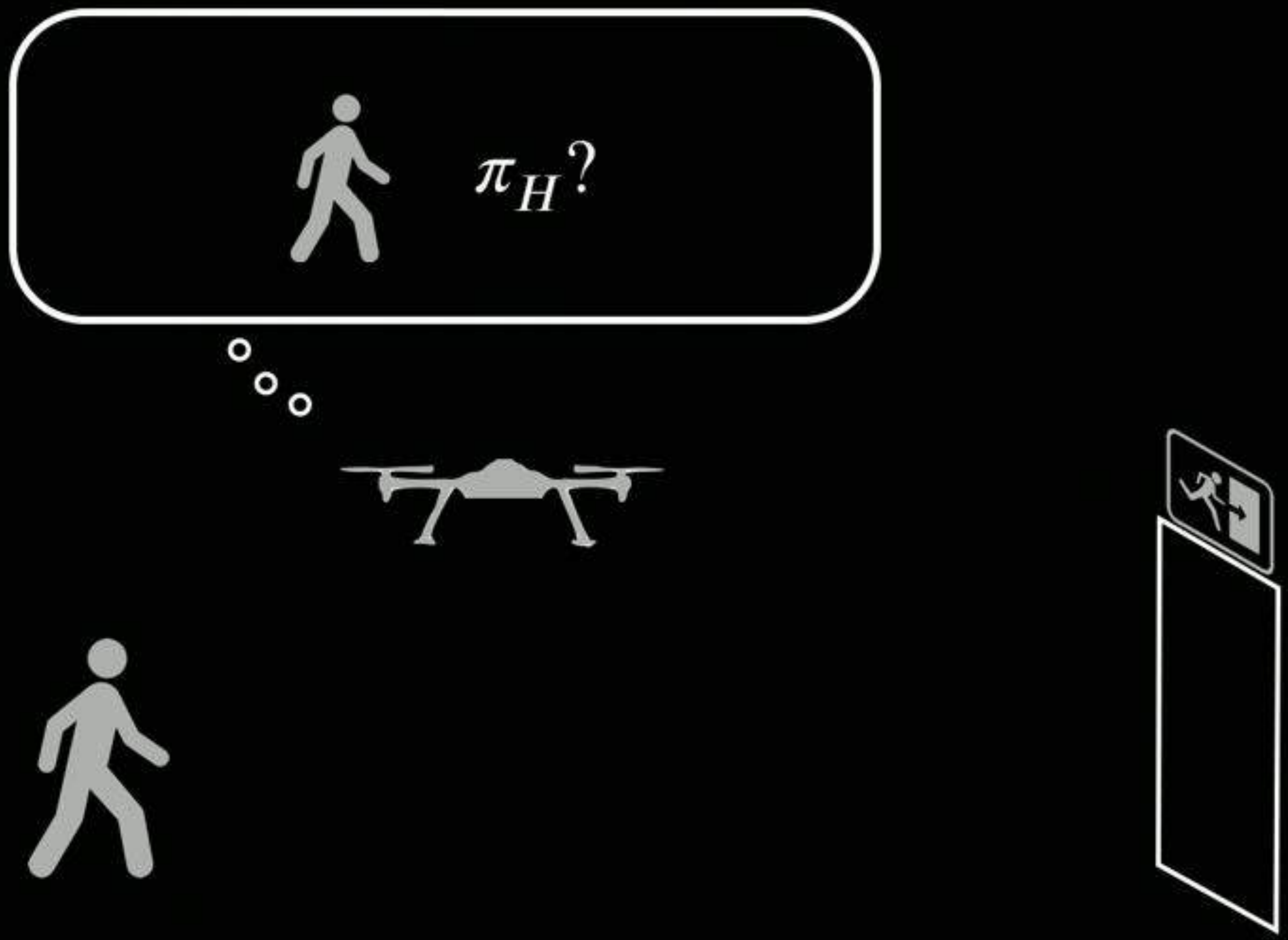
An Optimization-Centric Theory of Mind for Human-Robot Interaction

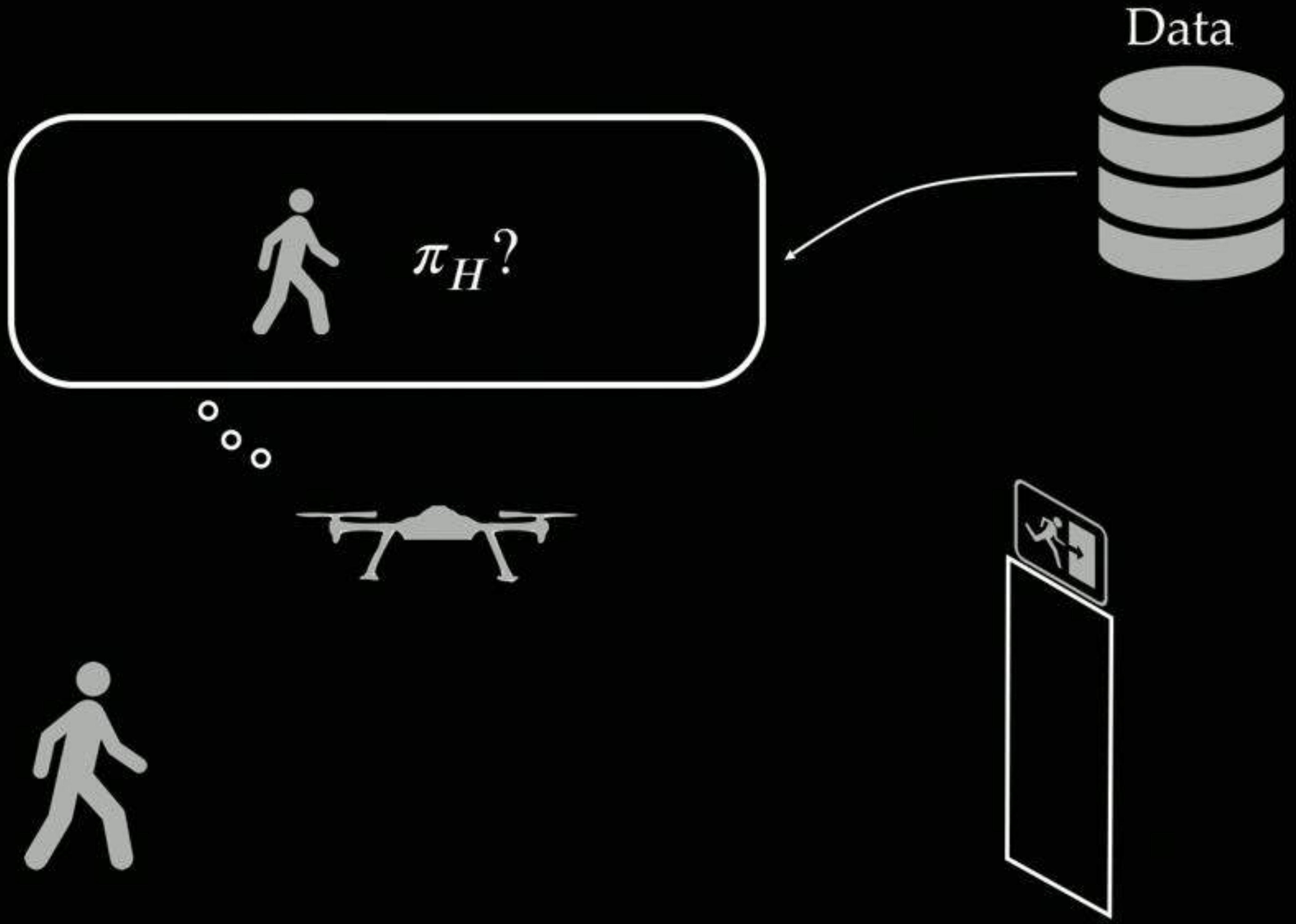
Anca Dragan

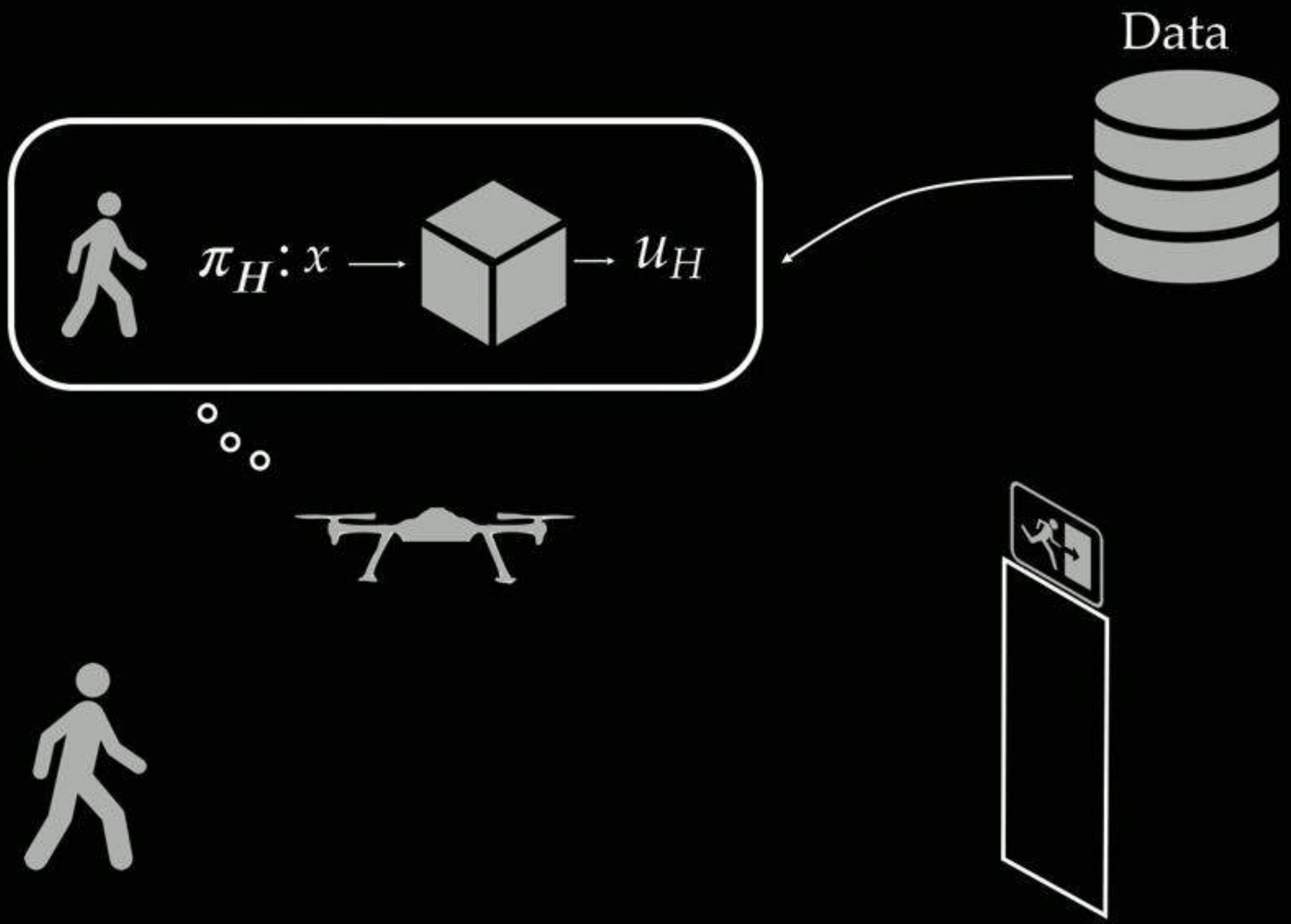

FOR
DUMMIES

$$\max_{\pi_R} \mathbb{E}[U(\pi_R, \pi_H)]$$

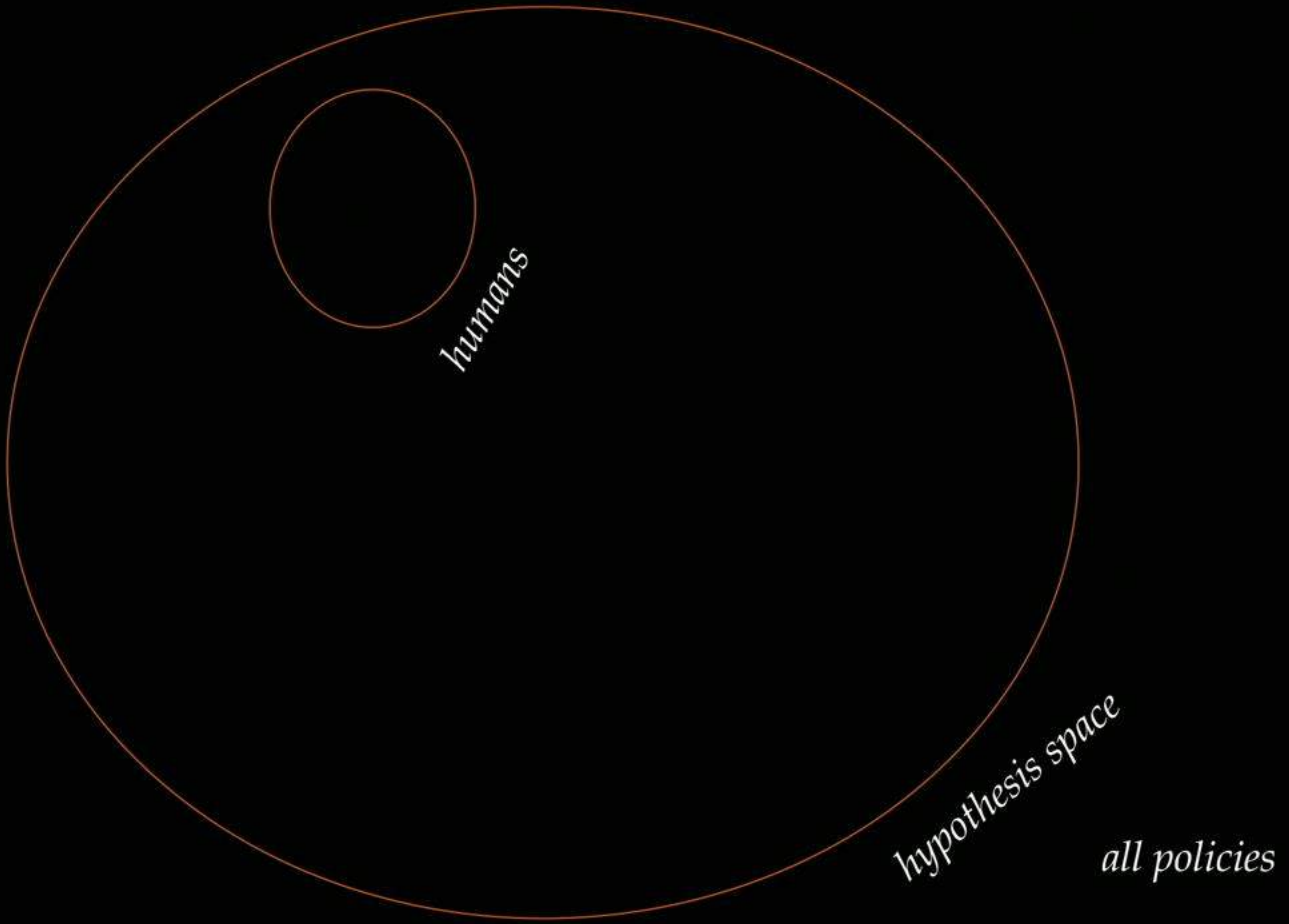


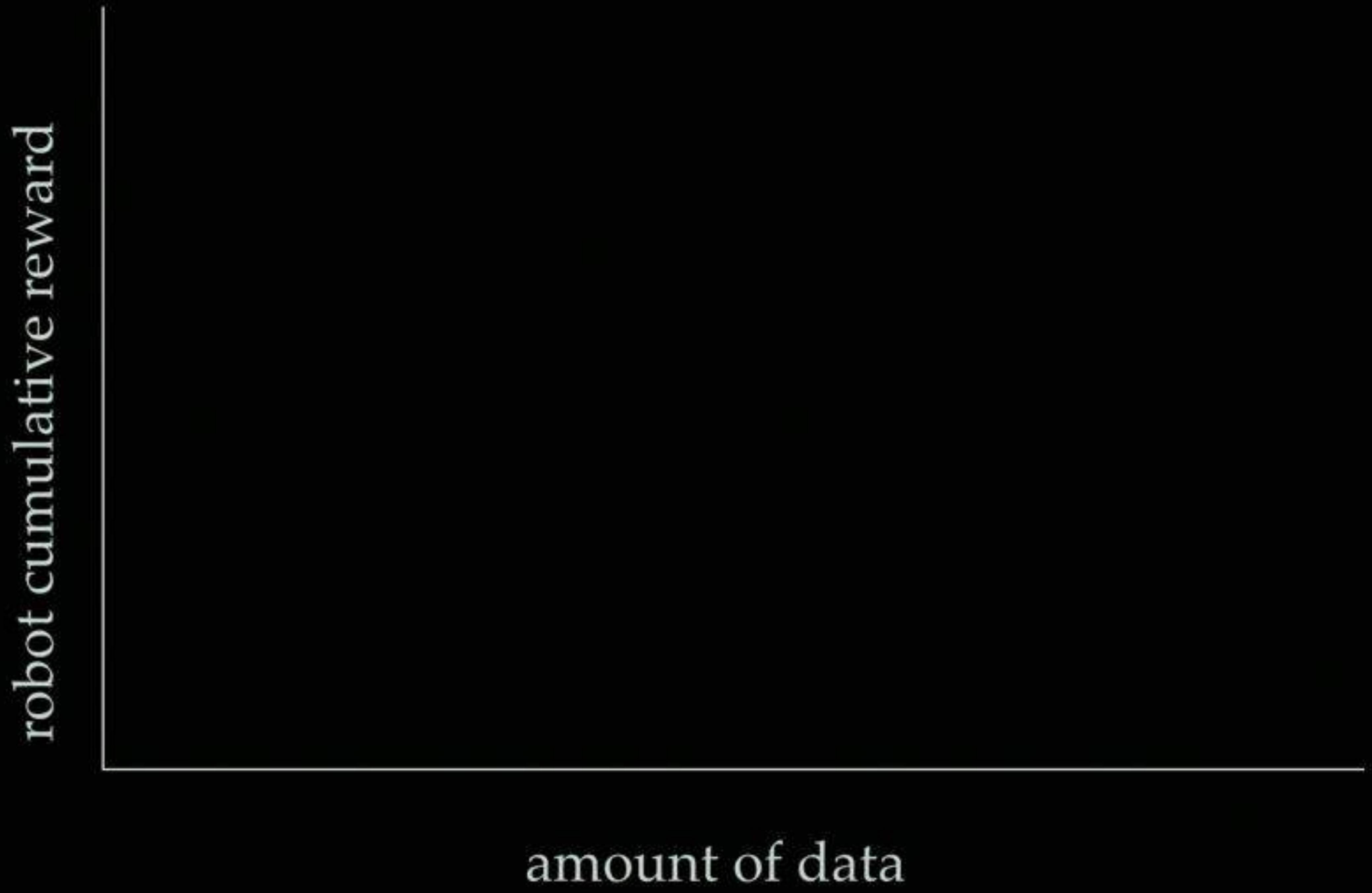

π_H?

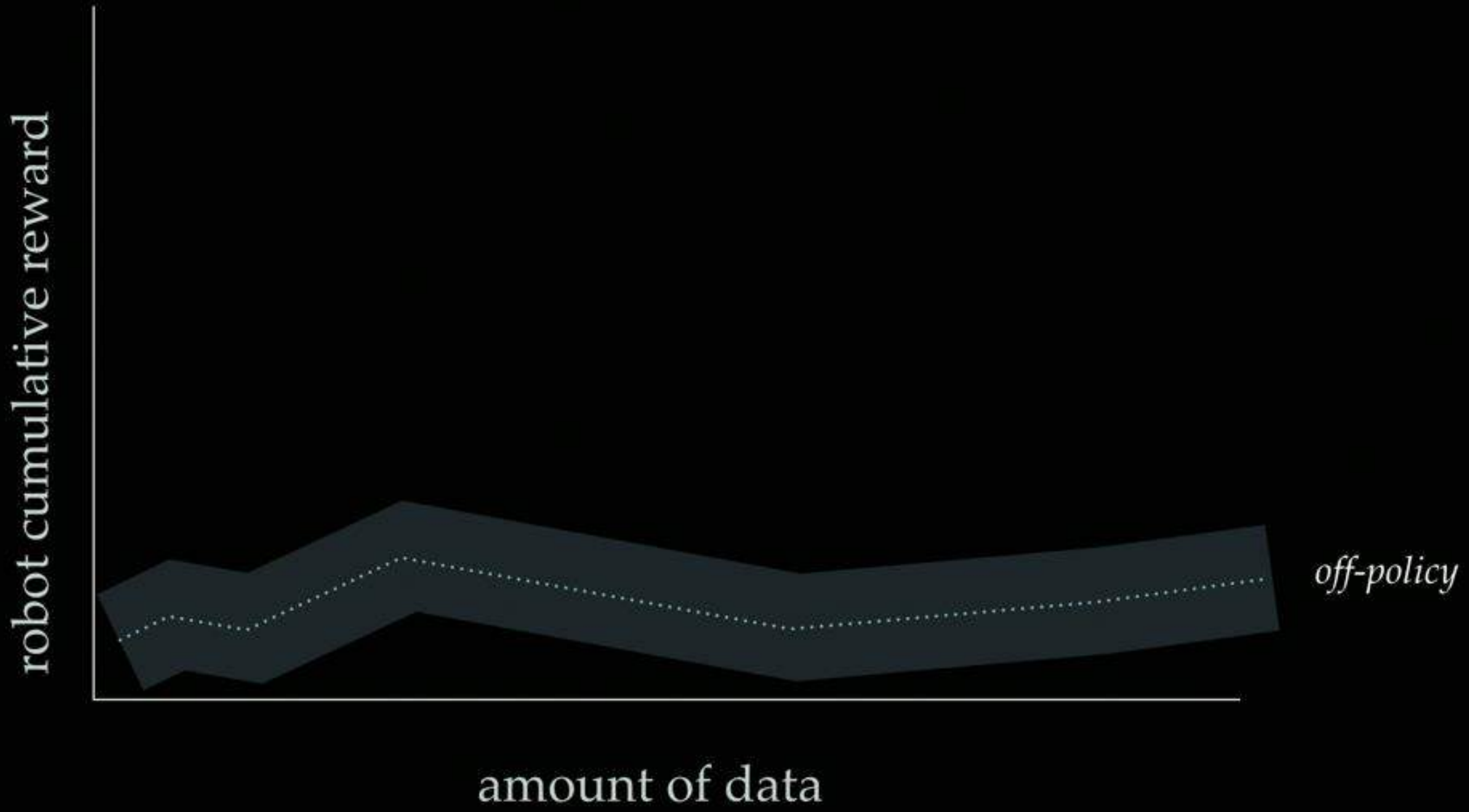

Data
π_H?

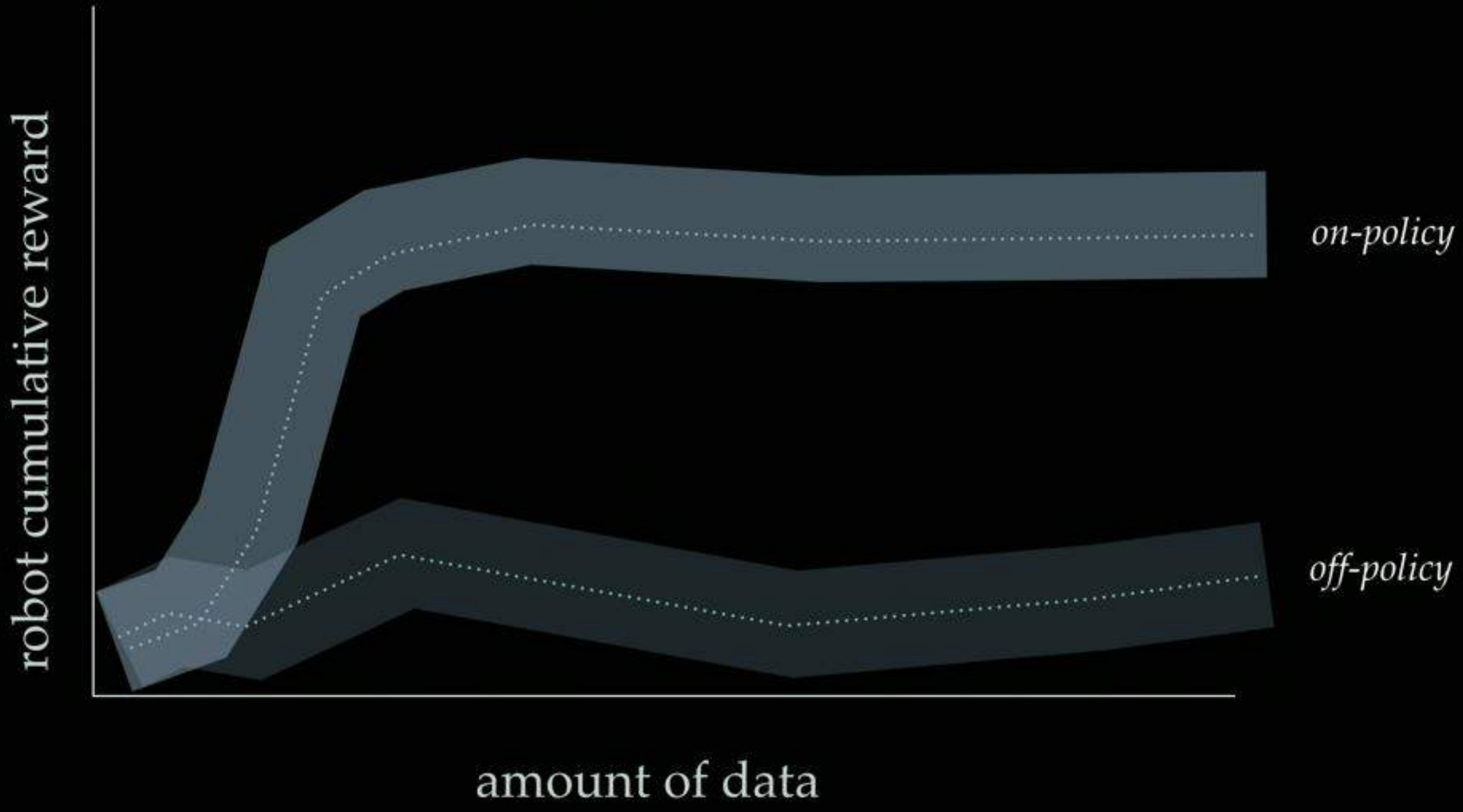


Data
$\pi_H : x \rightarrow \quad \rightarrow u_H$

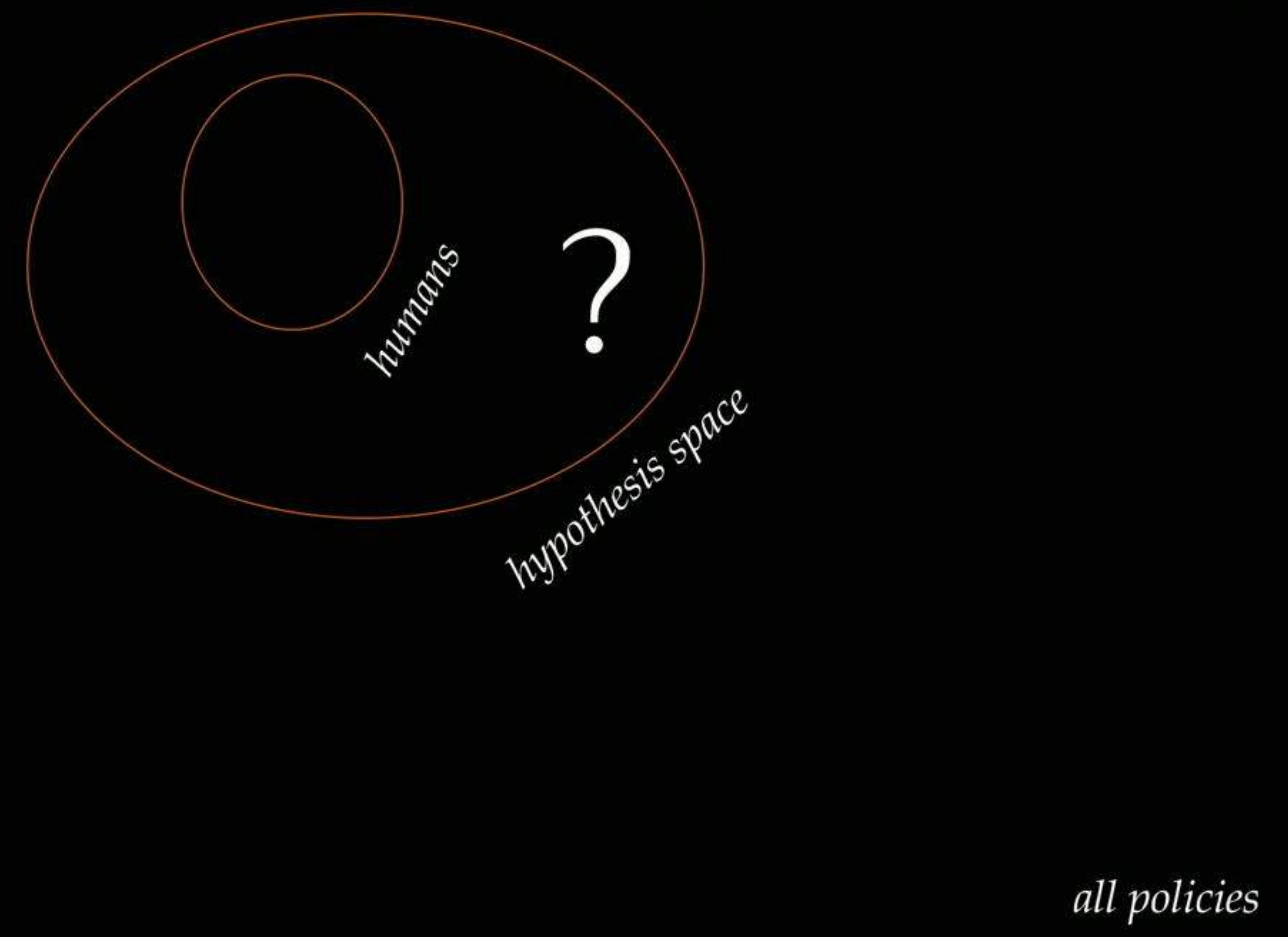

humans
hypothesis space
all policies


robot cumulative reward
amount of data


robot cumulative reward
off-policy
amount of data

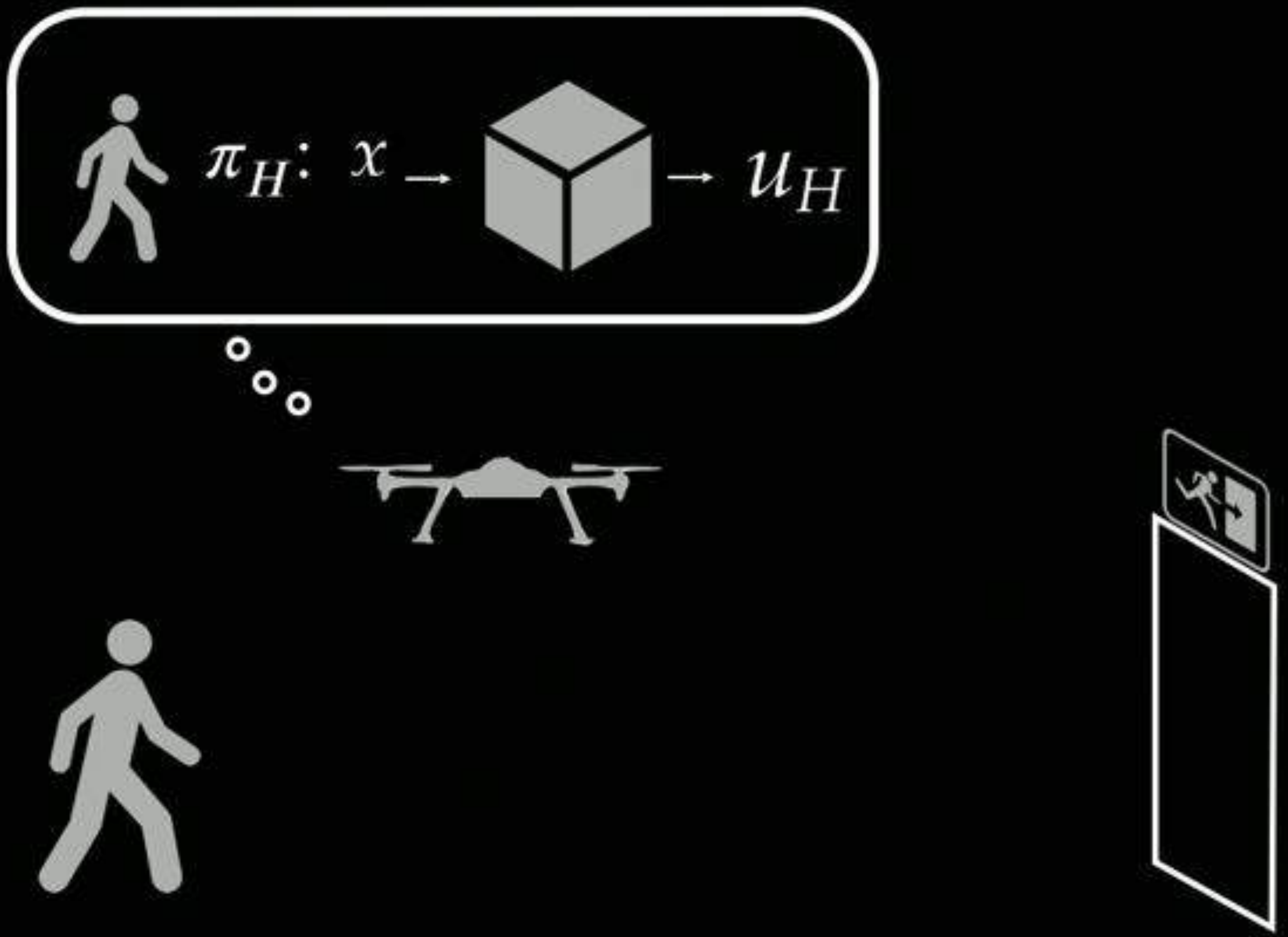


robot cumulative reward
on-policy
off-policy
amount of data

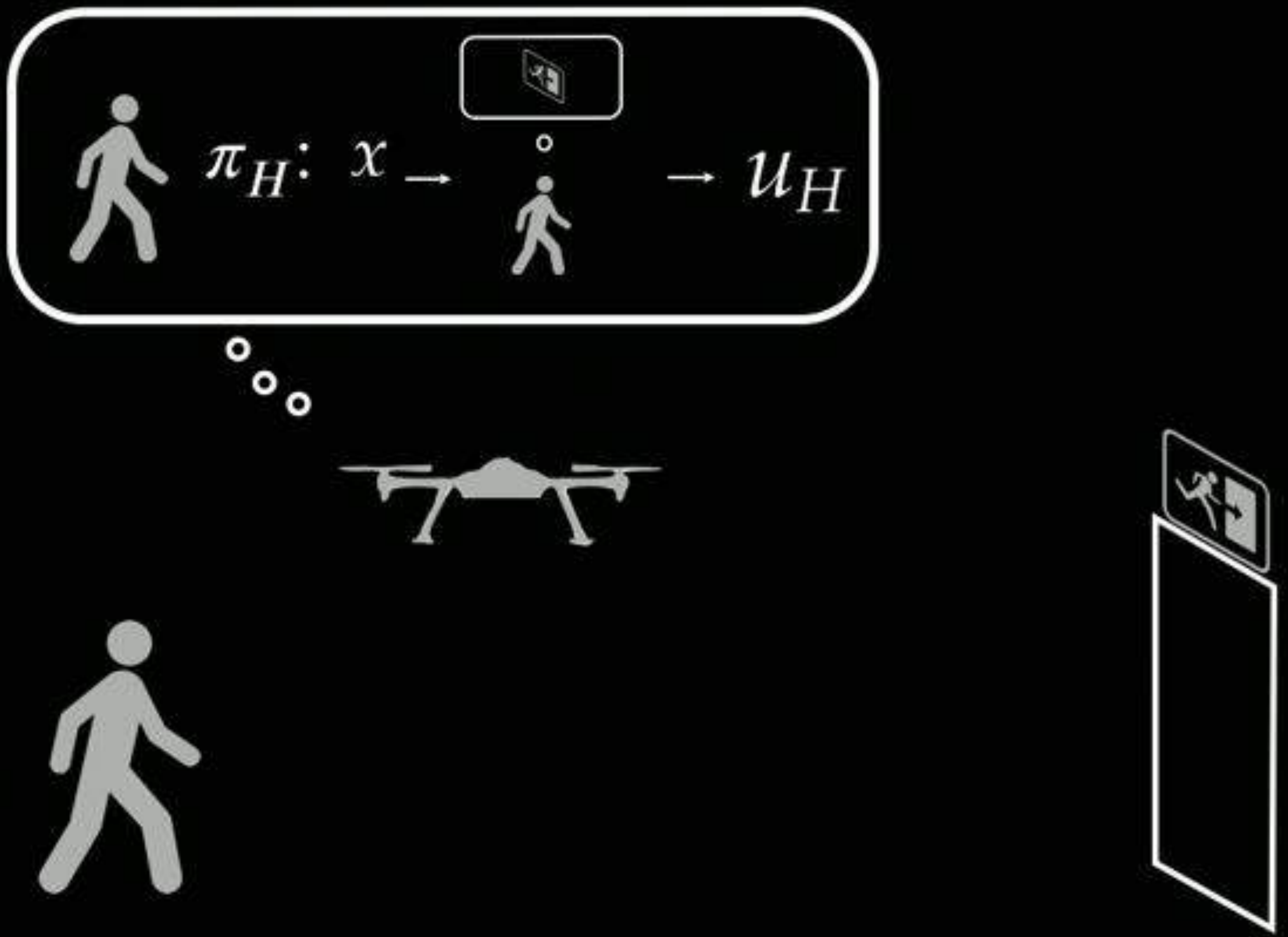

humans
?
hypothesis space
all policies

What is the right inductive bias for HRI?

Humans as black-box policies



Humans as intent-driven agents

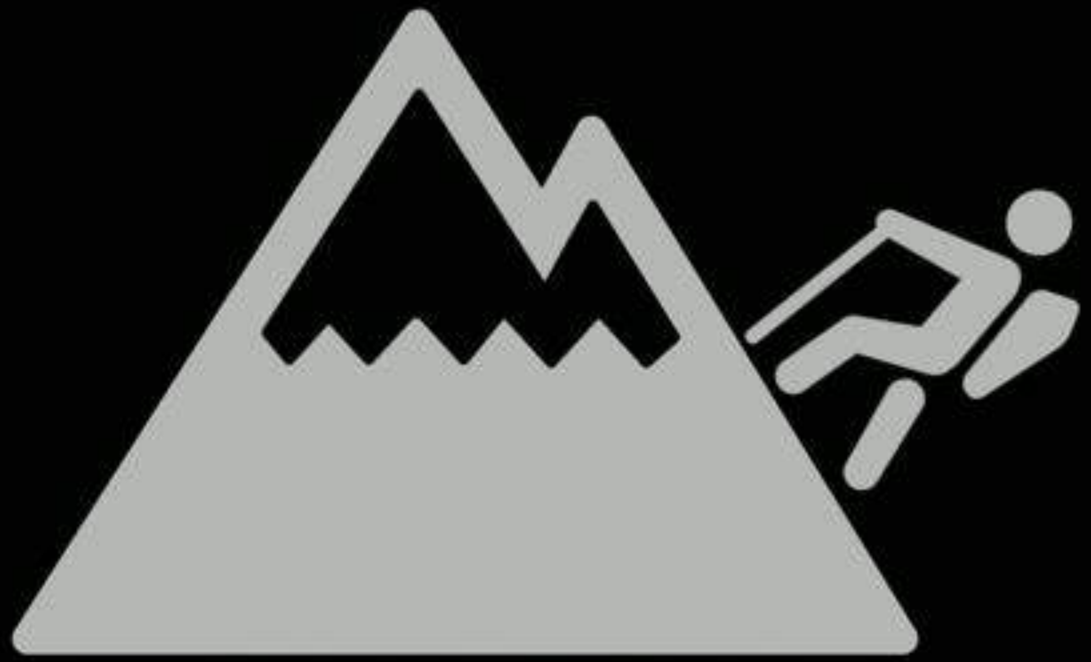


Humans have intent

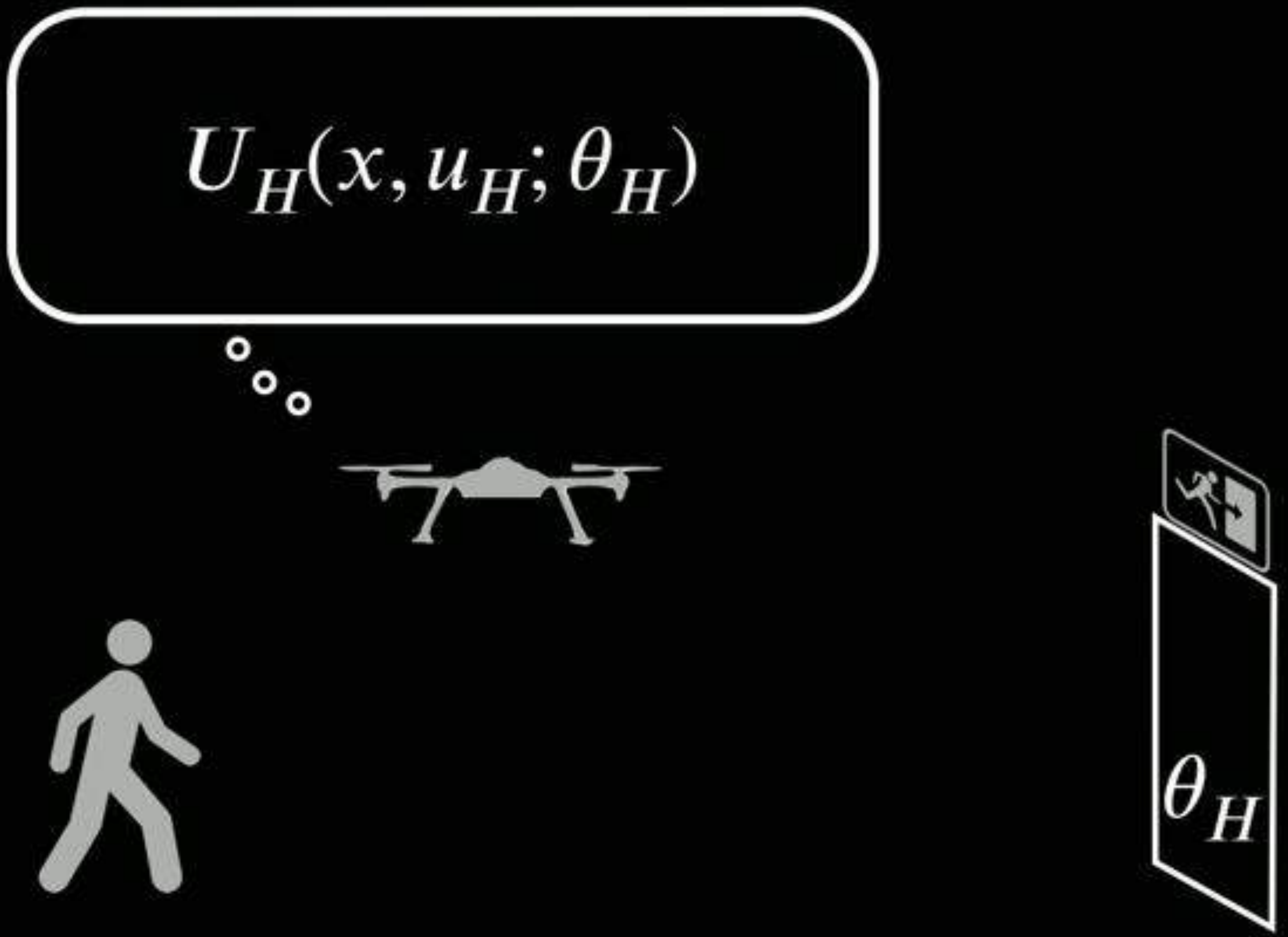
Humans have intent

Humans have intent

inductive bias



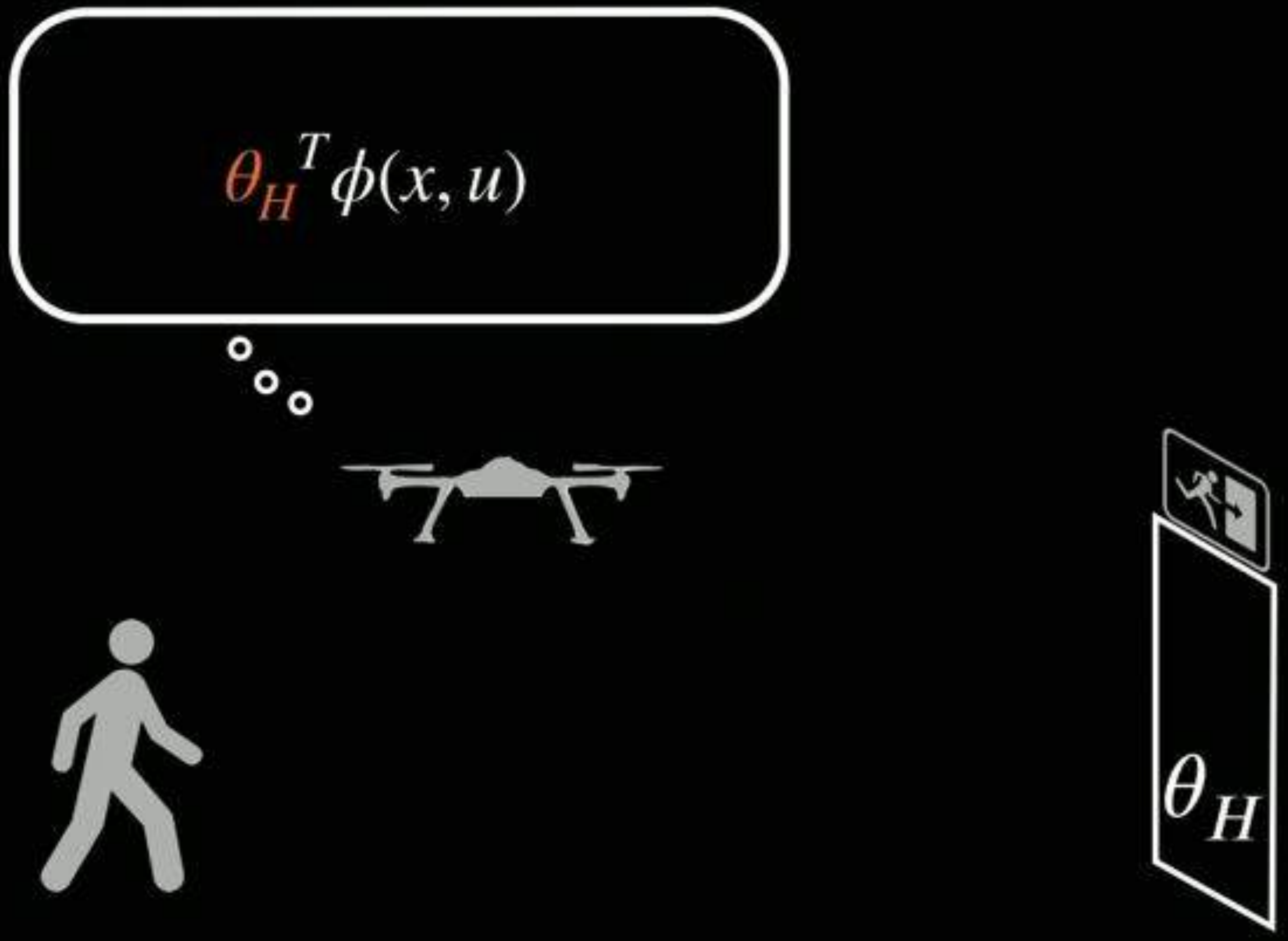
intent via utility



intent via utility

$$w_1 d(x, \theta_H) + w_2 |u_H| + w_3 d(x, x_o)$$

intent via utility



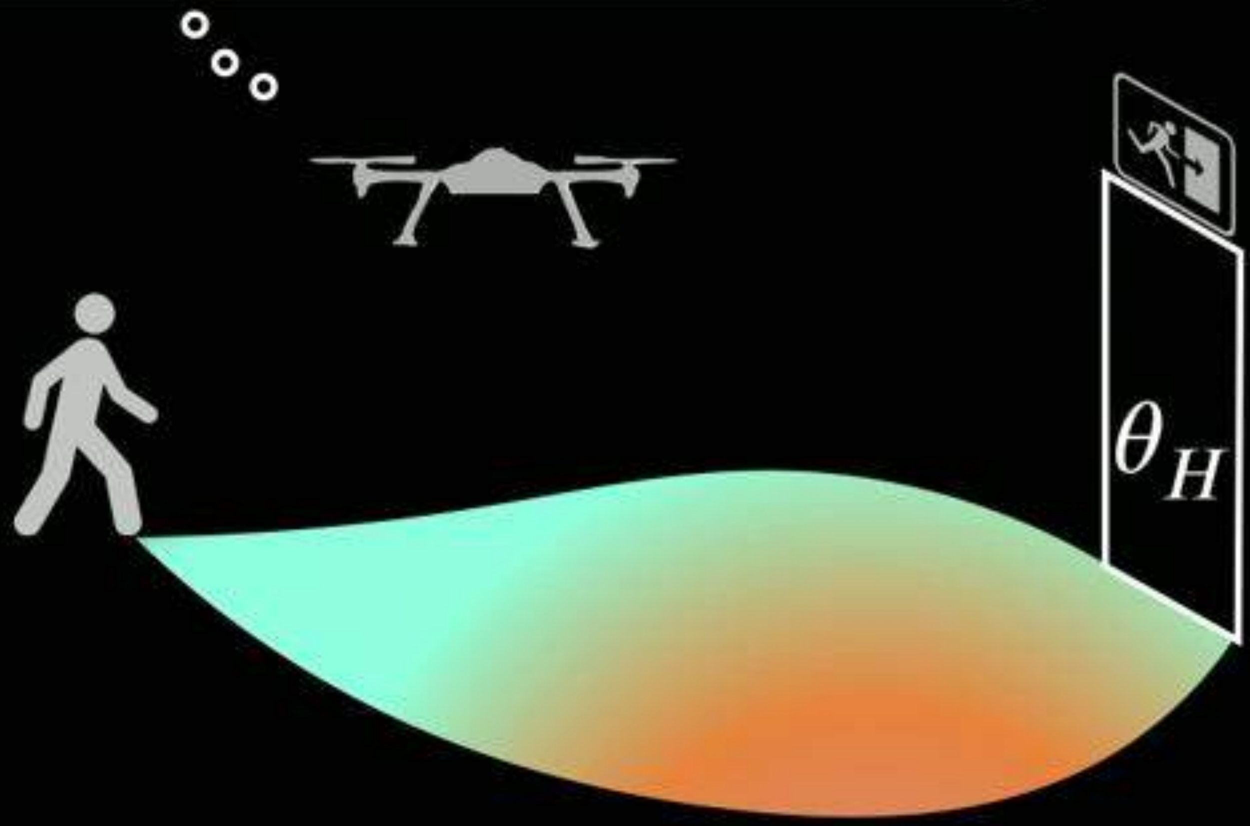
intent via utility

$$\max_{P} \; H(P)$$
$$s.t. \; \mathbb{E}[Q_{\theta_H}] = Q^*_{\theta_H} - \epsilon$$

Humans as noisy-rational agents

don't learn all this

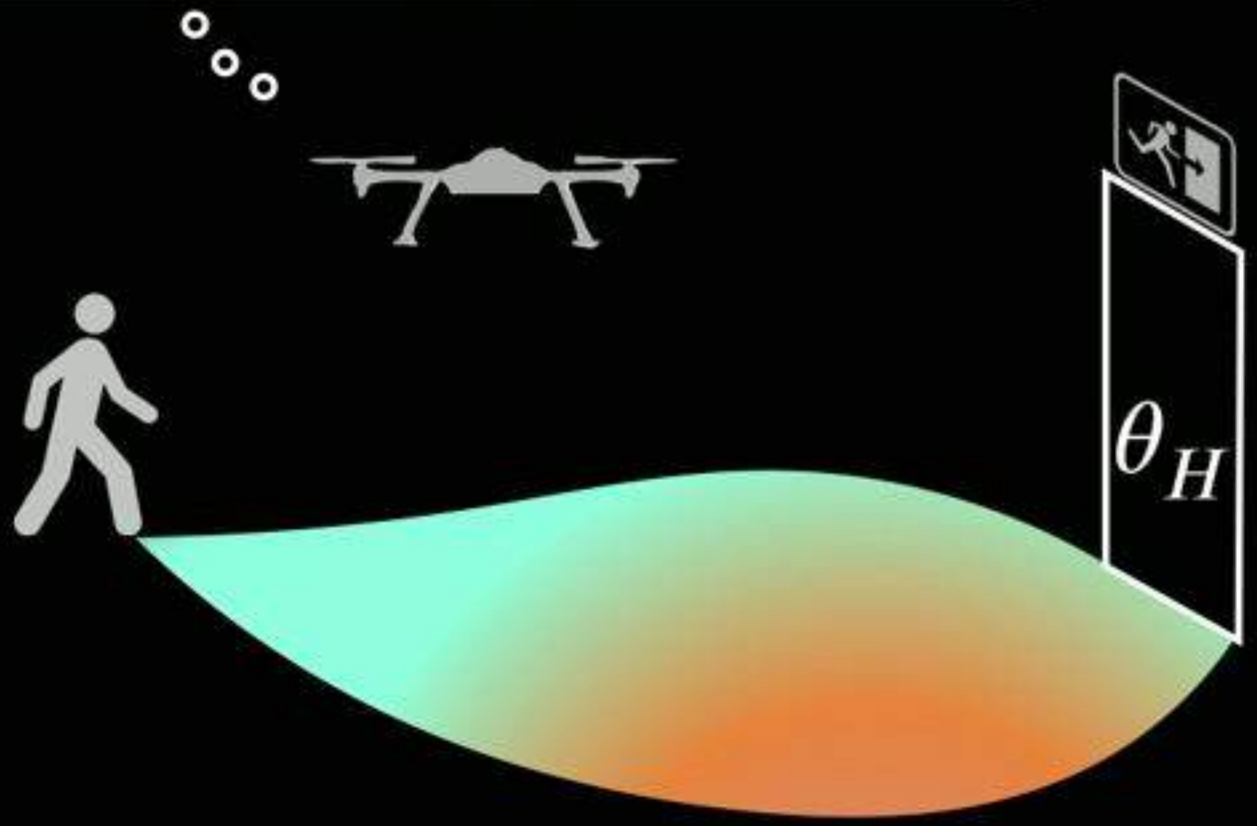
$$P(u_H \mid x, \theta_H) \propto e^{\beta Q(x, u_H; \theta_H)}$$



Humans as noisy-rational agents with unknown utilities

learn this

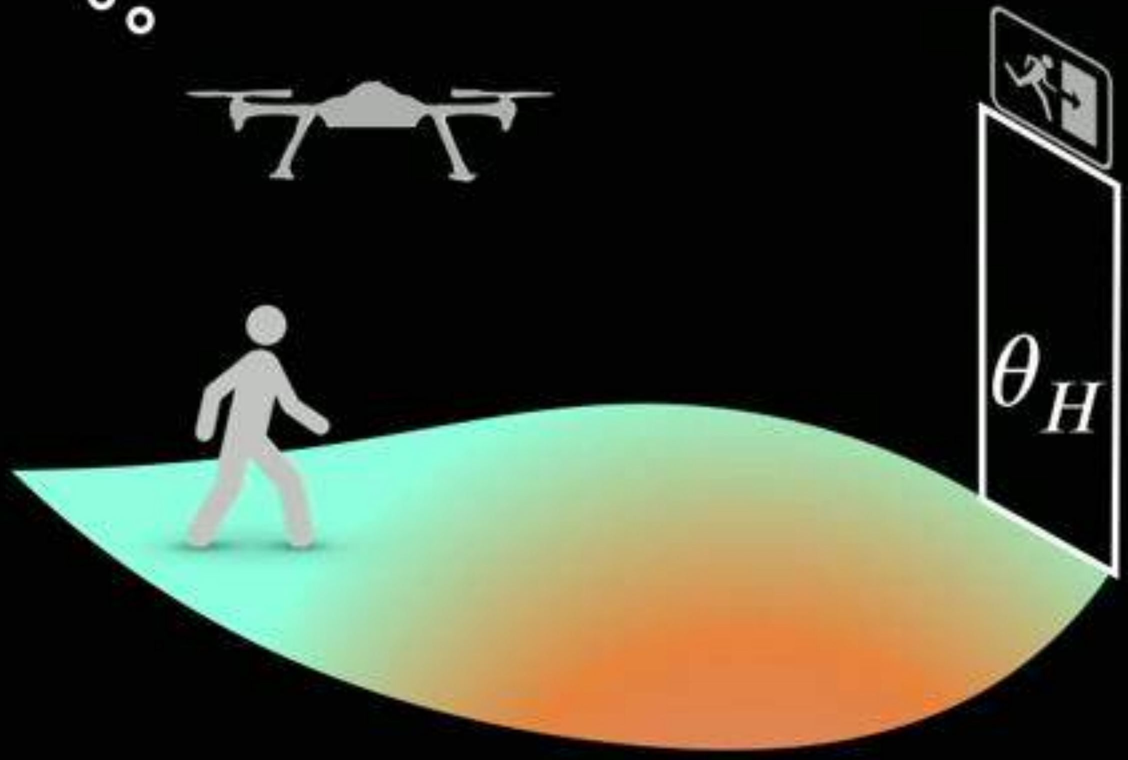
$$P(u_H \mid x, \theta_H) \propto e^{\beta Q(x, u_H; \theta_H)}$$



Humans as noisy-rational agents with unknown utilities

learn this

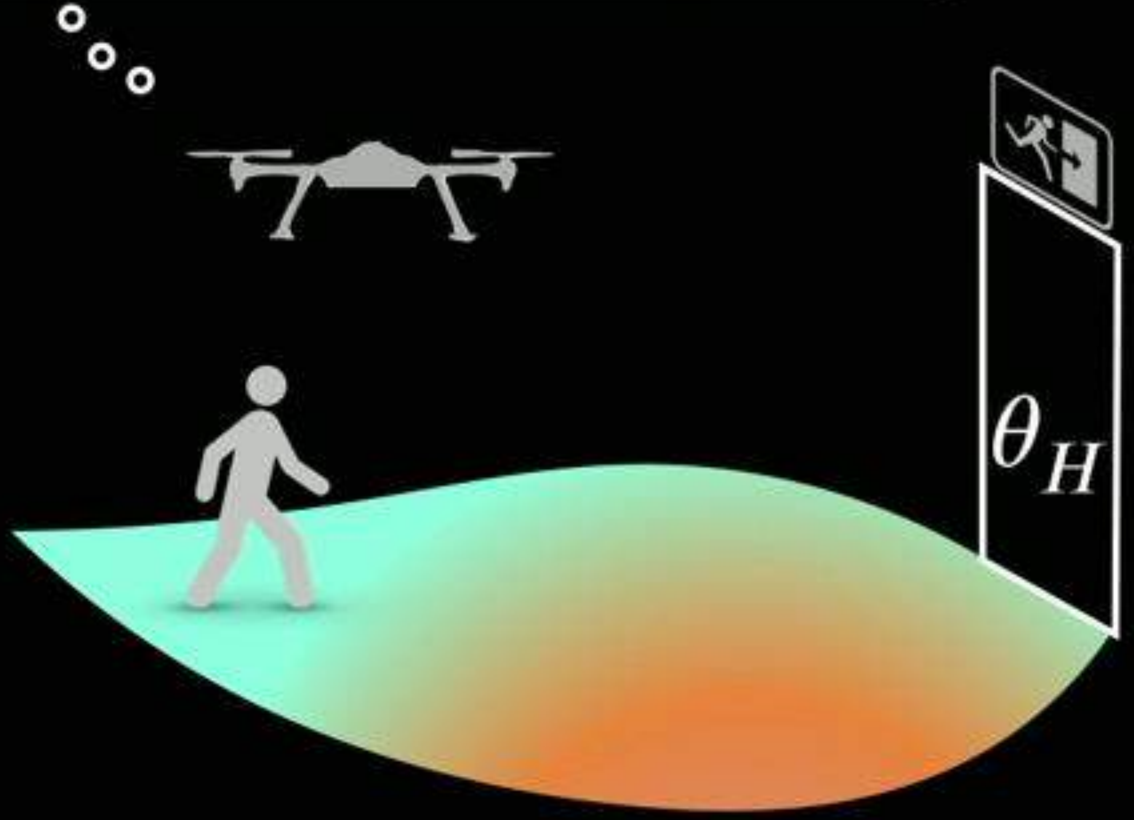
$$P(u_H \mid x, \theta_H) \propto e^{\beta Q(x, u_H; \theta_H)}$$



Humans as noisy-rational agents with unknown utilities

$$P(u_H \mid x, \theta_H) \propto e^{\beta Q(x, u_H; \theta_H)}$$

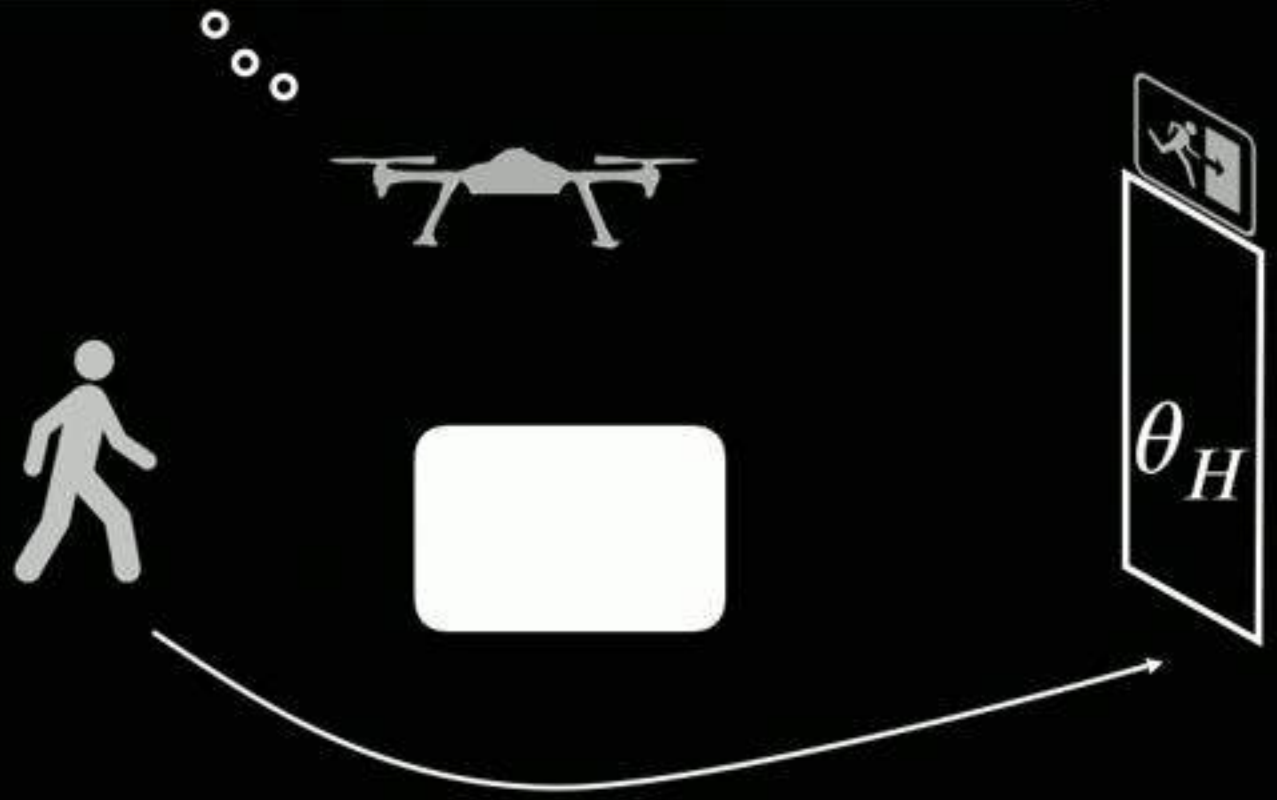
$$b'(\theta_H) \propto b(\theta_H) P(u_H \mid x, \theta_H)$$



Humans as noisy-rational agents with unknown utilities

$$P(u_H \mid x, \theta_H) \propto e^{\beta Q(x, u_H; \theta_H)}$$

$$b'(\theta_H) \propto b(\theta_H) P(u_H \mid x, \theta_H)$$



Humans as noisy-rational agents

action (demonstration) $P(u_H \mid x, \theta_H) \propto e^{\beta Q(x, u_H; \theta_H)}$

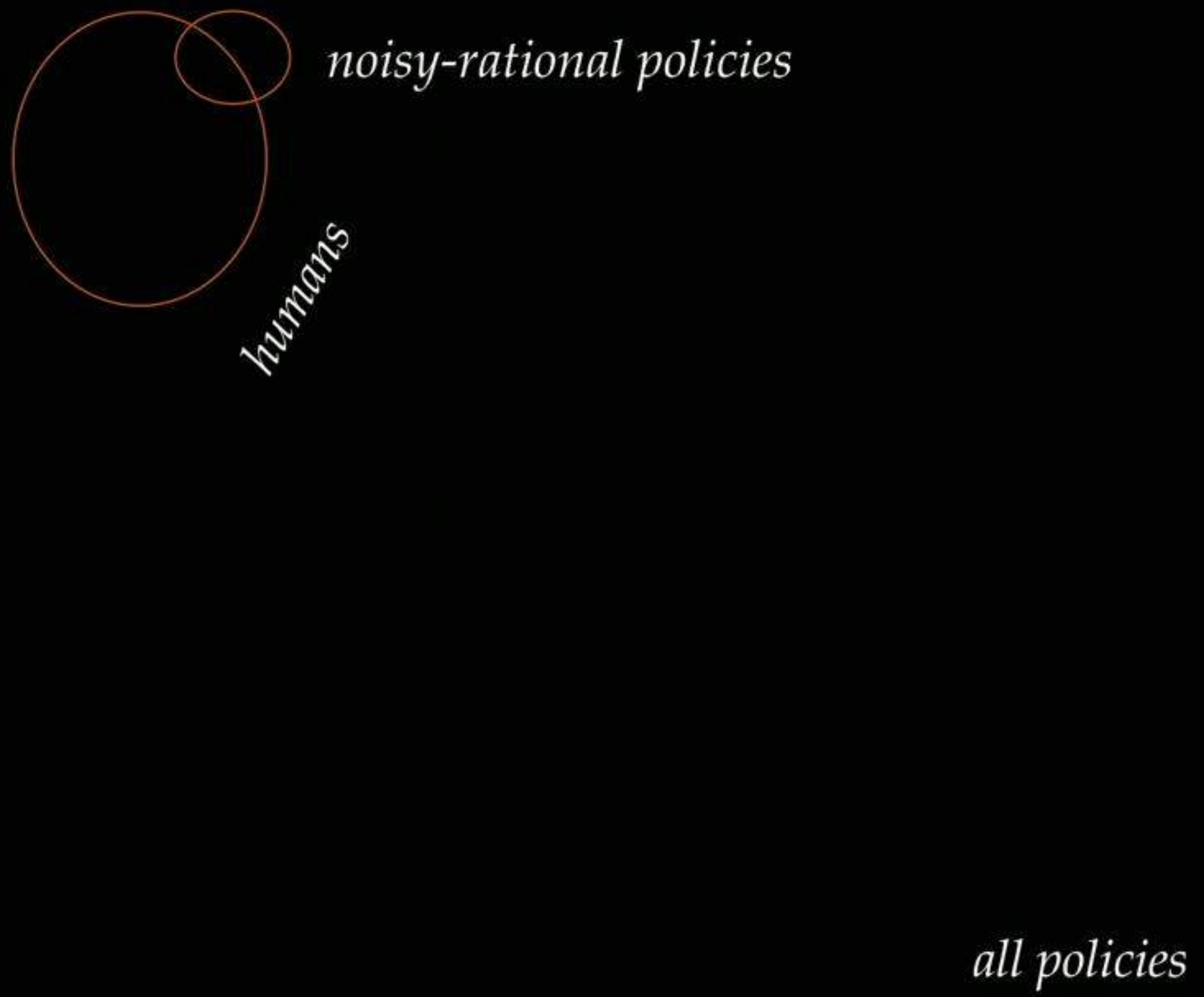
Humans as noisy-rational agents

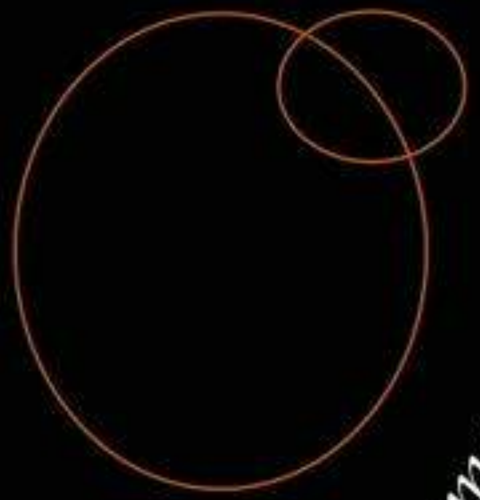
action (demonstration)	$u_H > u \forall u$	$P(u_H \mid x, \theta_H) = \dfrac{e^{\beta Q(x,u_H;\theta_H)}}{\int e^{\beta Q(x,u;\theta_H)} du}$
comparison	$u_A > u_B$	$P(u_A \mid x, u_A, u_B, \theta_H) = \dfrac{e^{\beta Q(x,u_A;\theta_H)}}{e^{\beta Q(x,u_A;\theta_H)} + e^{\beta Q(x,u_B;\theta_H)}}$
correction	$u_H + u_R > u \forall u$	$P(u_H \mid x, u_R, \theta_H) = \dfrac{e^{\beta Q(x,u_H+u_R;\theta_H)}}{\int e^{\beta Q(x,u;\theta_H)} du}$
stop	$u_0 > u_R$	$P(u_0 \mid x, u_R, \theta_H) = \dfrac{e^{\beta Q(x,u_0;\theta_H)}}{e^{\beta Q(x,u_0;\theta_H)} + e^{\beta Q(x,u_R;\theta_H)}}$

Humans as noisy-rational agents

action (demonstration)	$u_H > u \forall u$	$P(u_H \mid x, \theta_H) = \frac{e^{\beta Q(x, u_H; \theta_H)}}{\int e^{\beta Q(x,u;\theta_H)} du}$
comparison	$u_A > u_B$	$P(u_A \mid x, u_A, u_B, \theta_H) = \frac{e^{\beta Q(x, u_A; \theta_H)}}{e^{\beta Q(x,u_A;\theta_H)} + e^{\beta Q(x,u_B;\theta_H)}}$
correction	$u_H + u_R > u \forall u$	$P(u_H \mid x, u_R, \theta_H) = \frac{e^{\beta Q(x, u_H + u_R; \theta_H)}}{\int e^{\beta Q(x,u;\theta_H)} du}$
stop	$u_0 > u_R$	$P(u_0 \mid x, u_R, \theta_H) = \frac{e^{\beta Q(x, u_0; \theta_H)}}{e^{\beta Q(x,u_0;\theta_H)} + e^{\beta Q(x,u_R;\theta_H)}}$

proxy reward, current world state, …


noisy-rational policies
humans
all policies

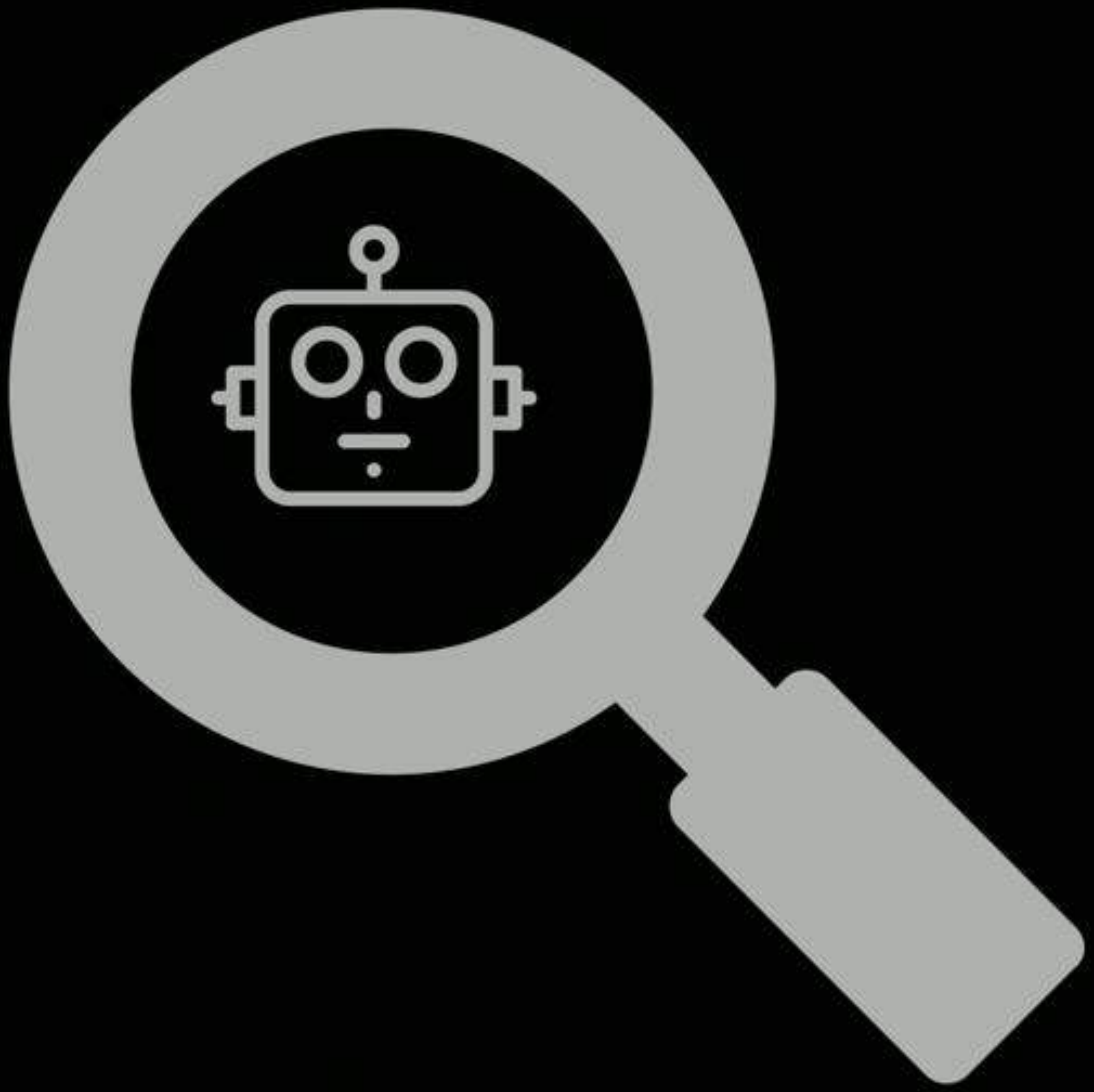


Challenge:

noisy rationality is sometimes too rigid.

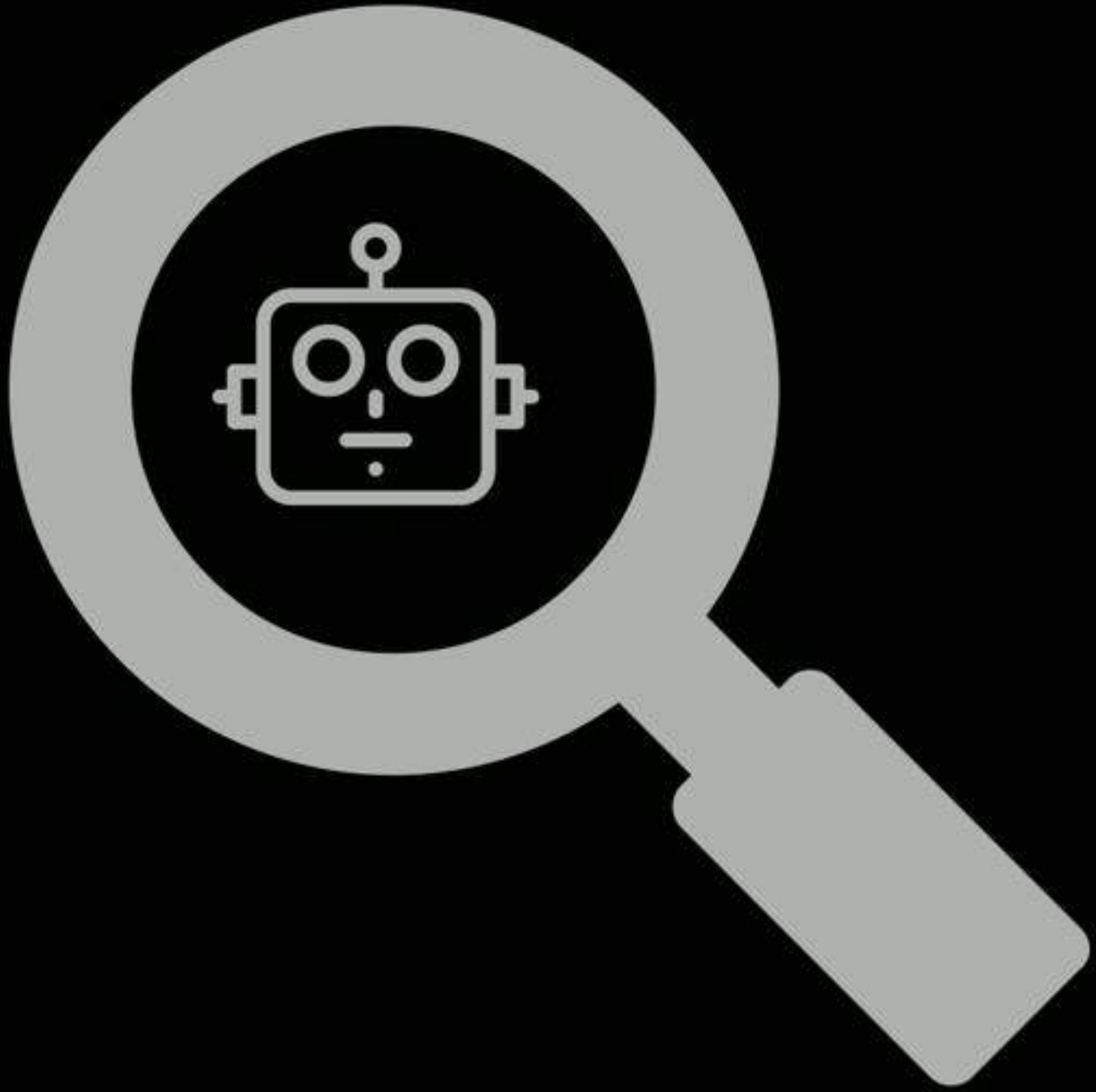
Humans have intent

inductive bias

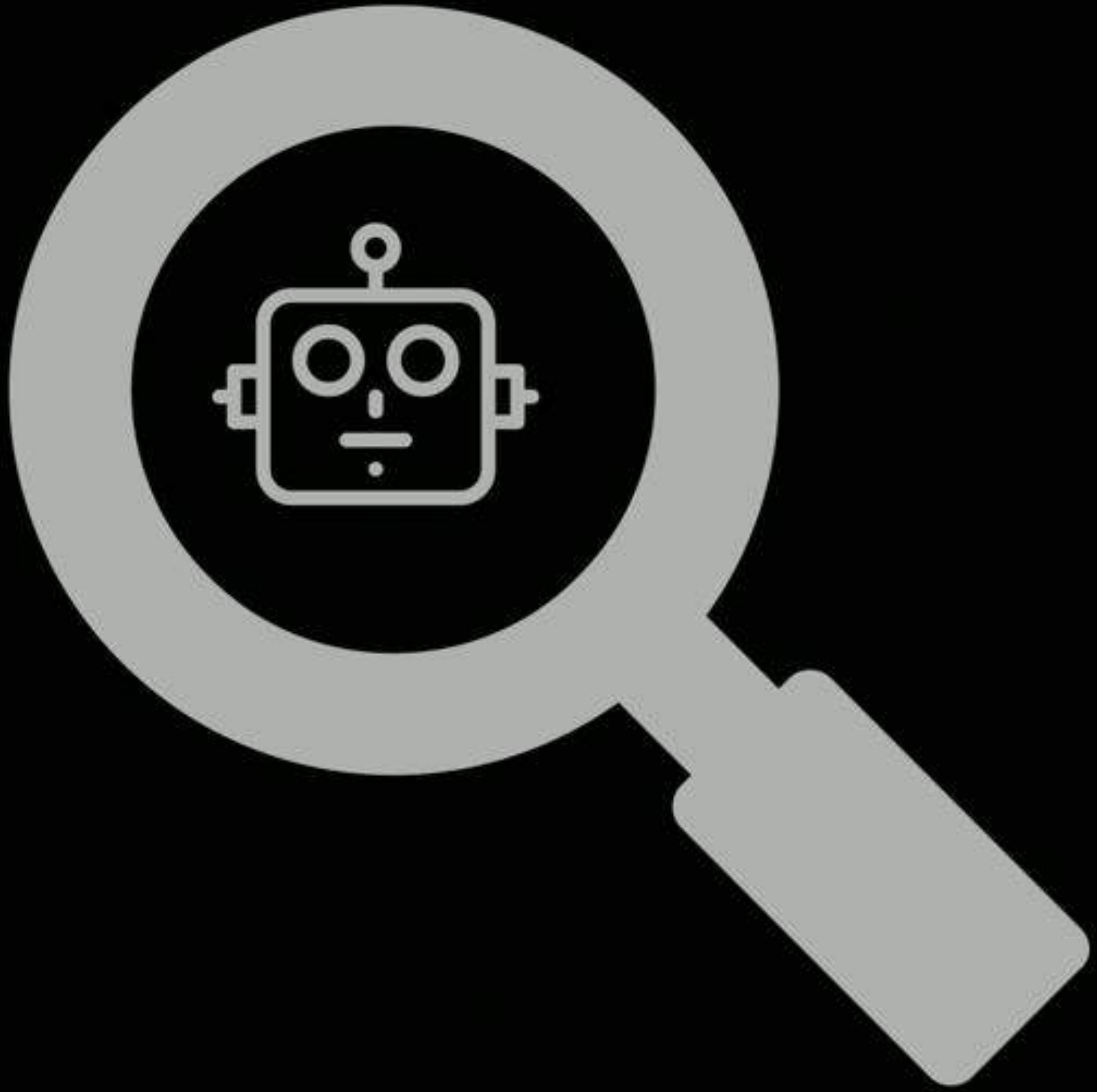


Treating people as robots

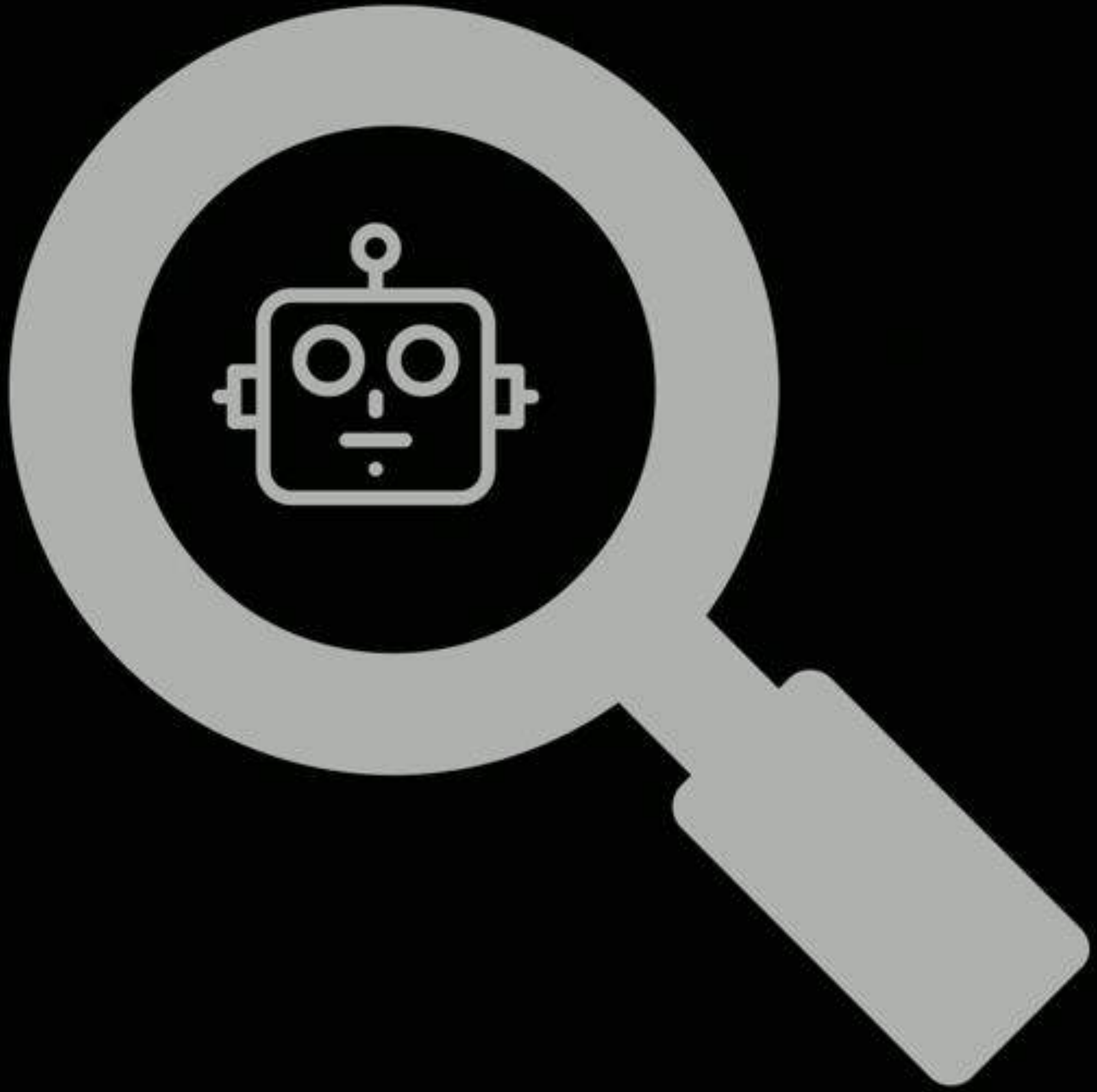
is what makes noisy rationality too rigid.



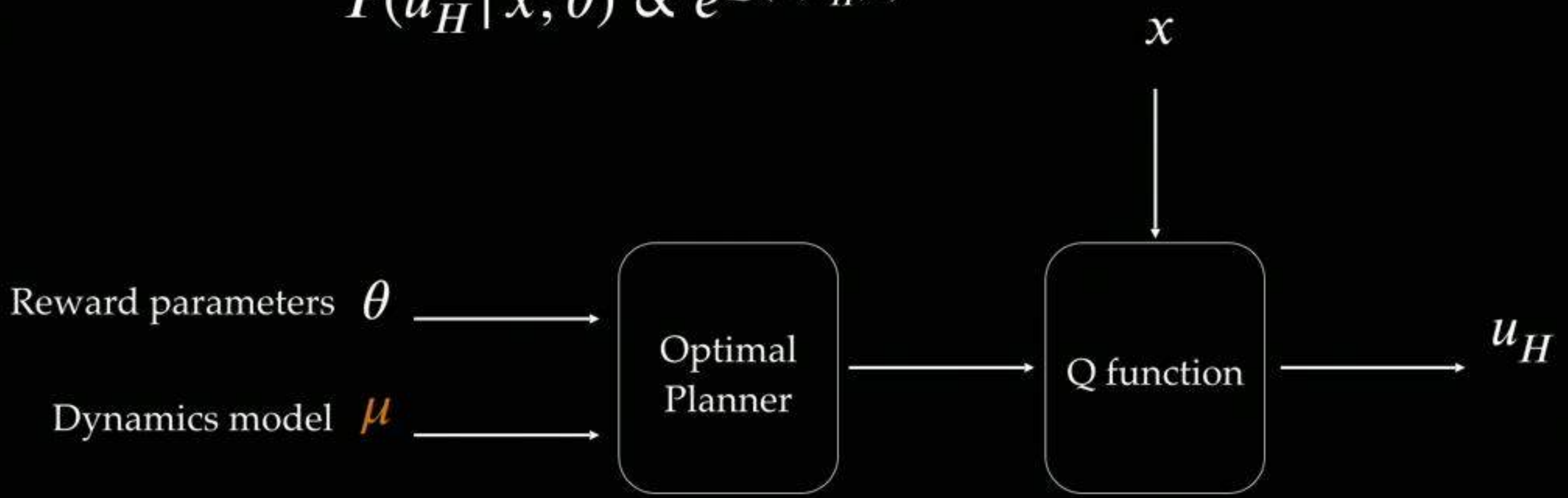
What if it's also the key to fixing it?



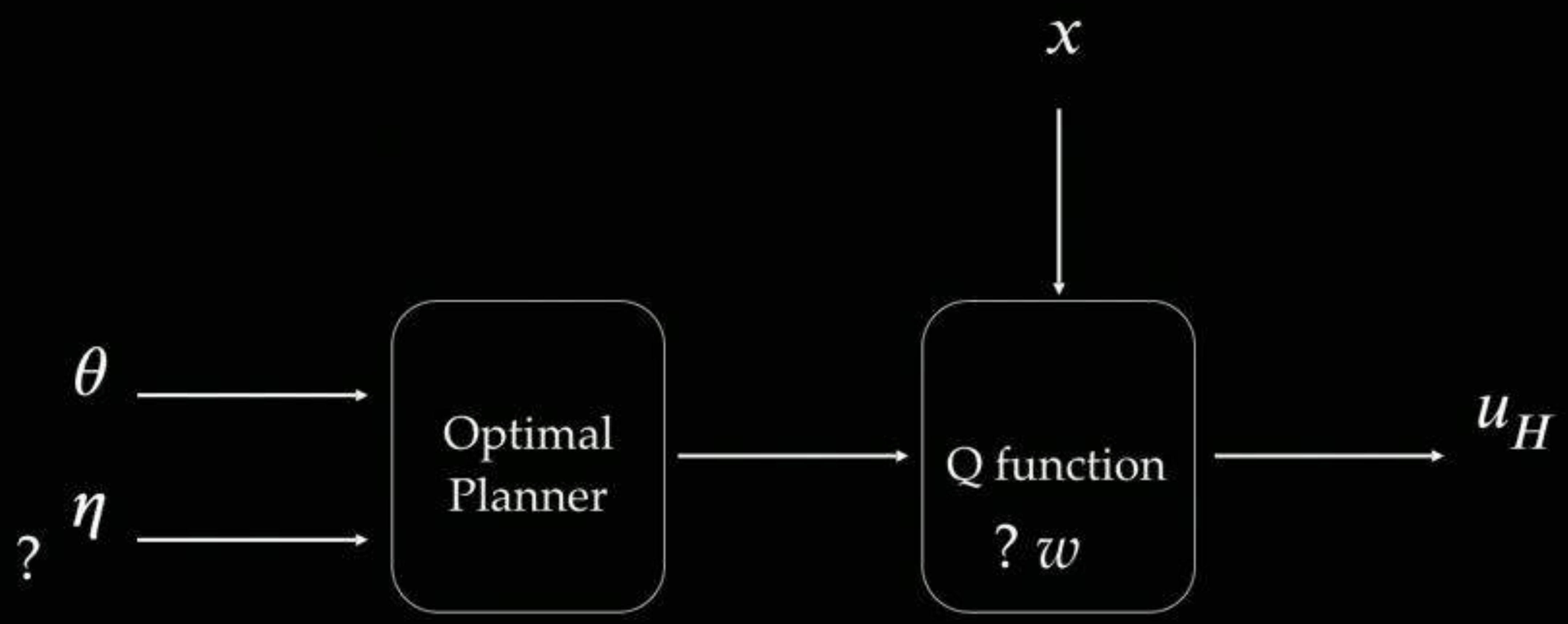
When are robots not rational?



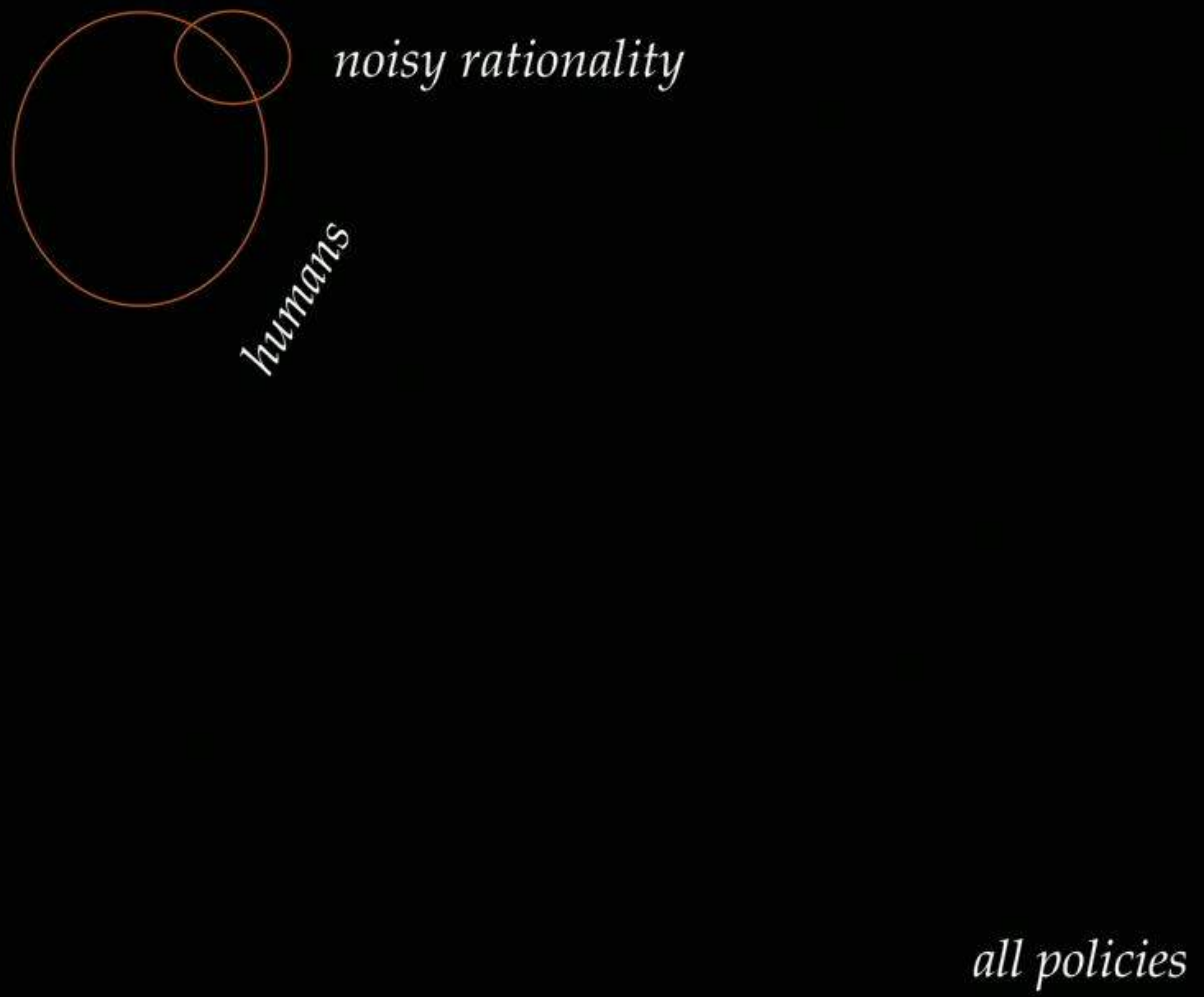
$$P(u_H \mid x; \theta) \propto e^{Q(x, u_H; \theta)}$$

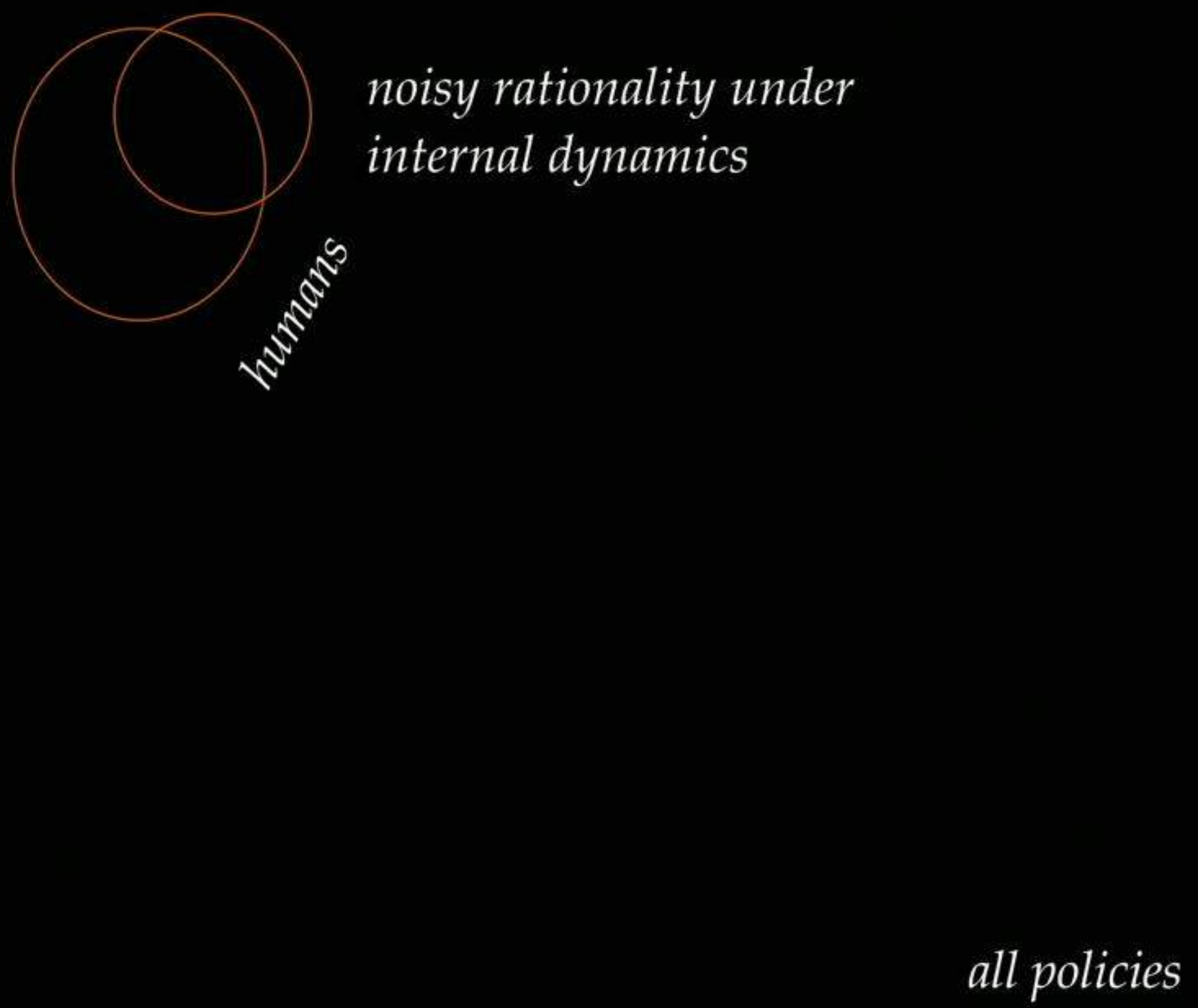


$$P(u_H \mid x; \theta, \eta) \propto e^{Q(x, u_H; \theta, \eta)}$$

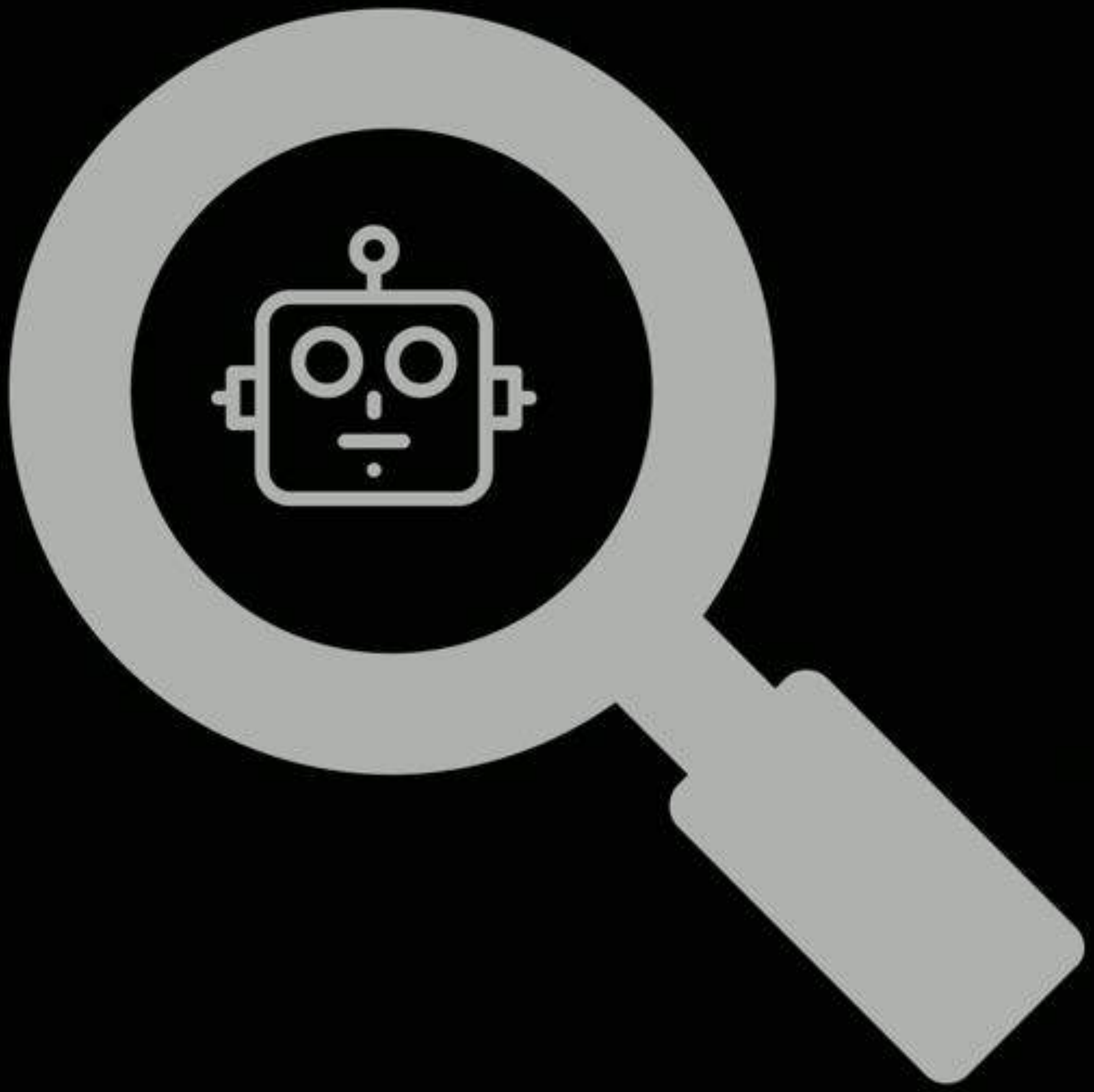


$$\max_{\eta, w} P\left(u_H \middle| x; w\right) - \lambda \delta\left(w, \eta; \theta\right)^2$$


noisy rationality
humans
all policies


noisy rationality under internal dynamics
humans
all policies

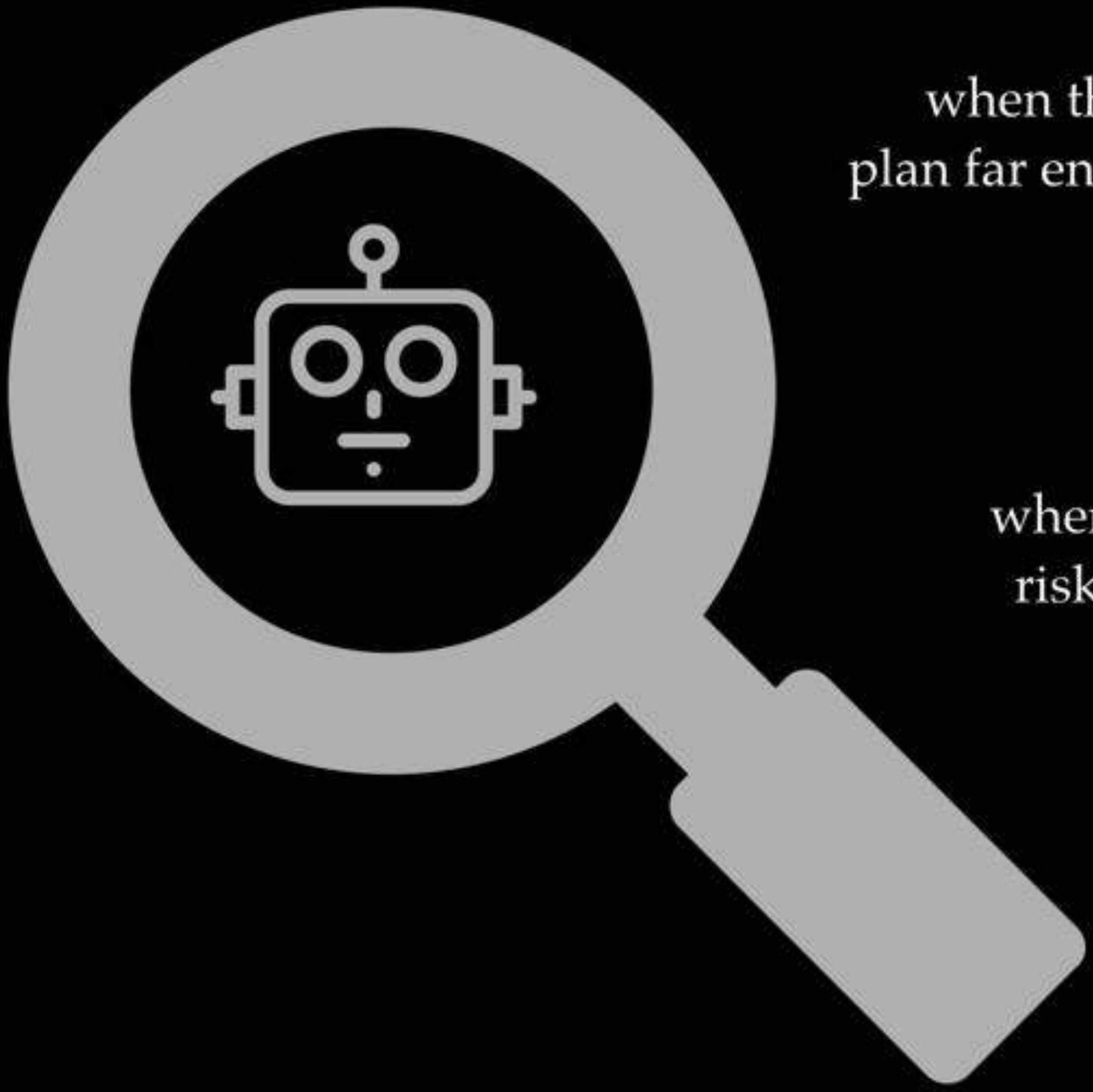
When are robots not rational?



When are robots not rational?

when they don't know the dynamics

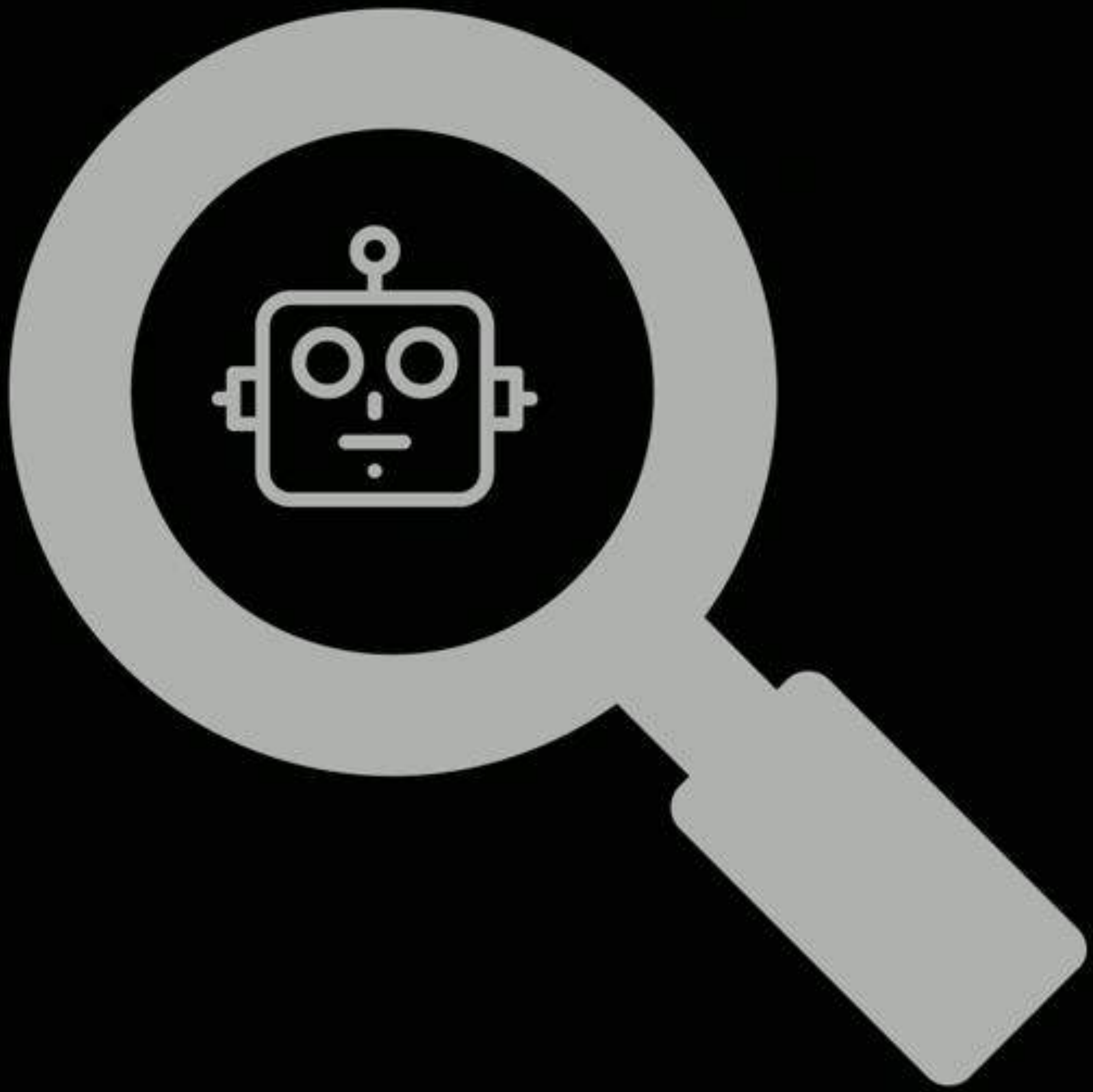
when they can't plan far enough ahead



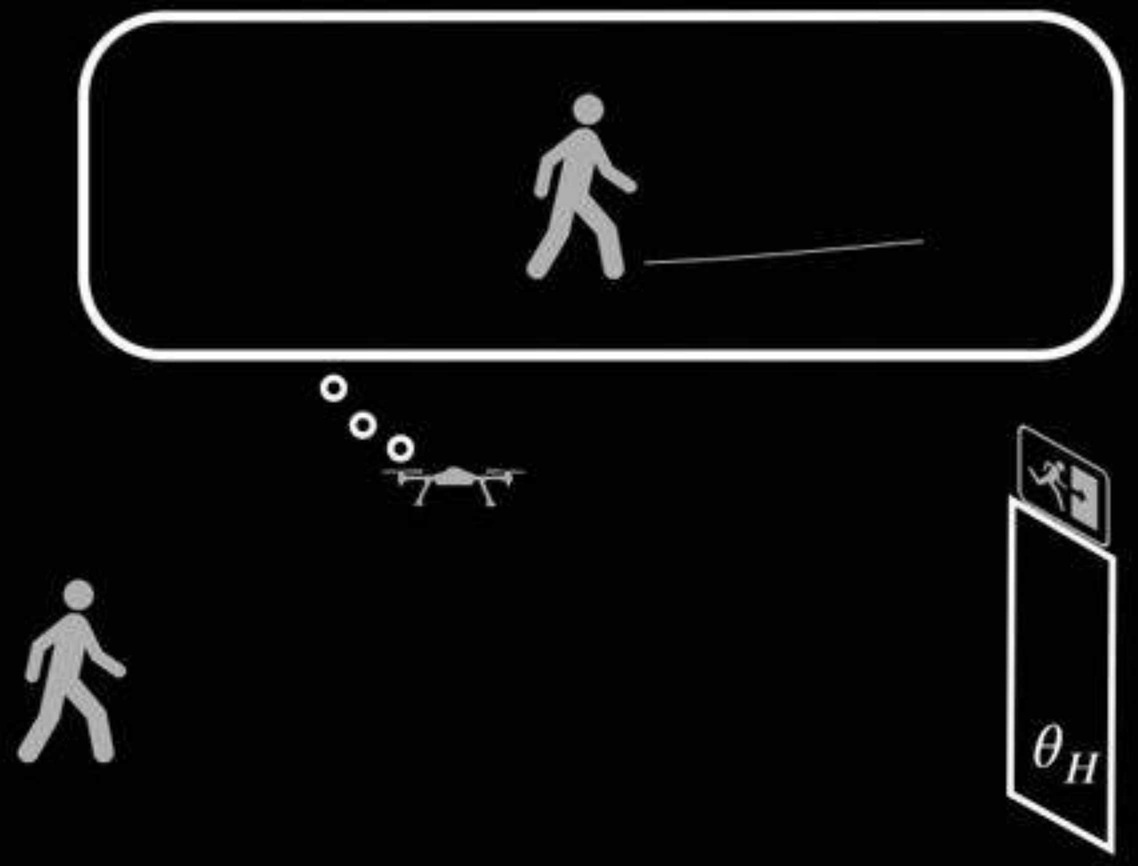
when they're risk-averse

when they're still learning

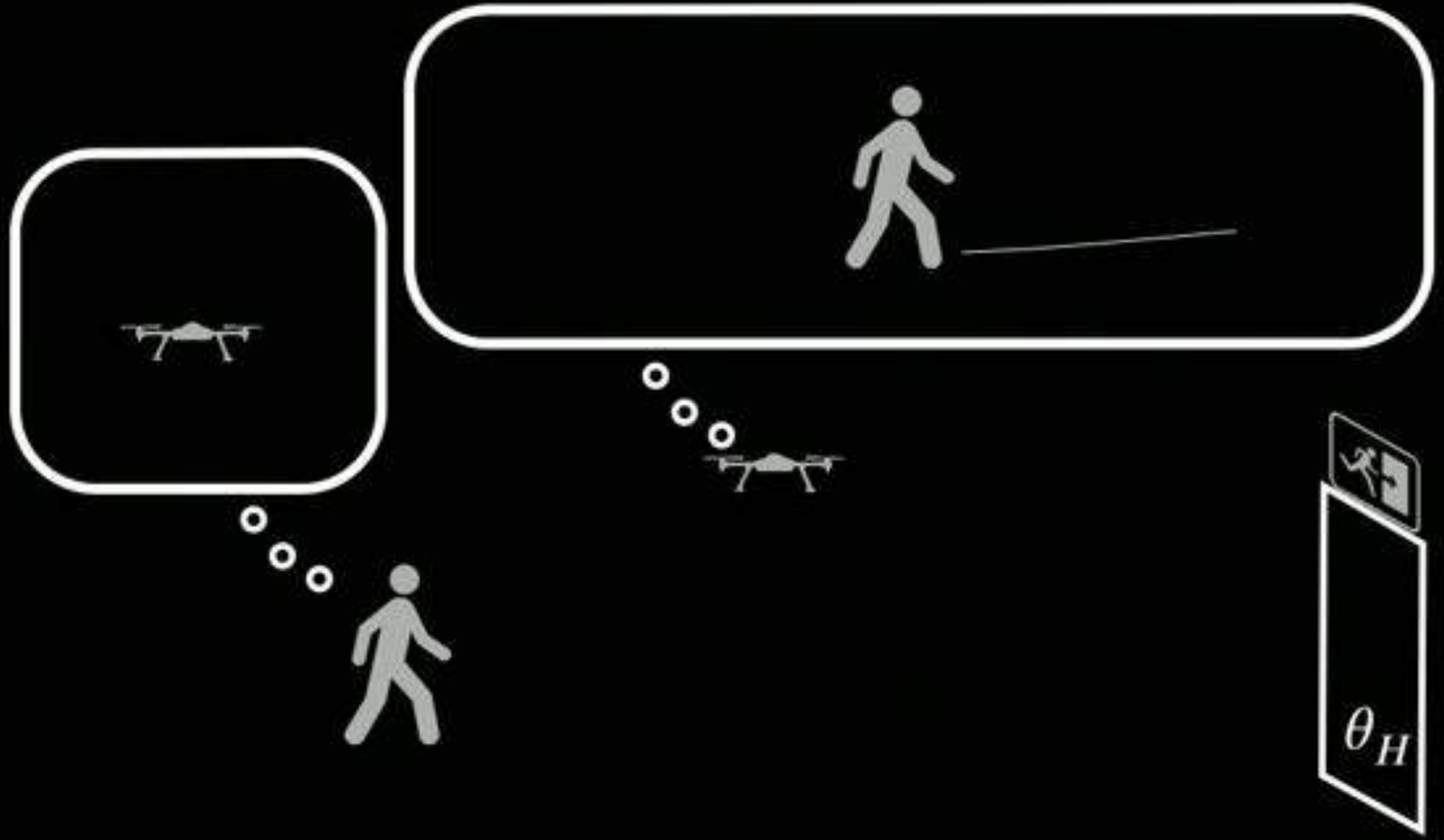
When are humans better than rational robots?



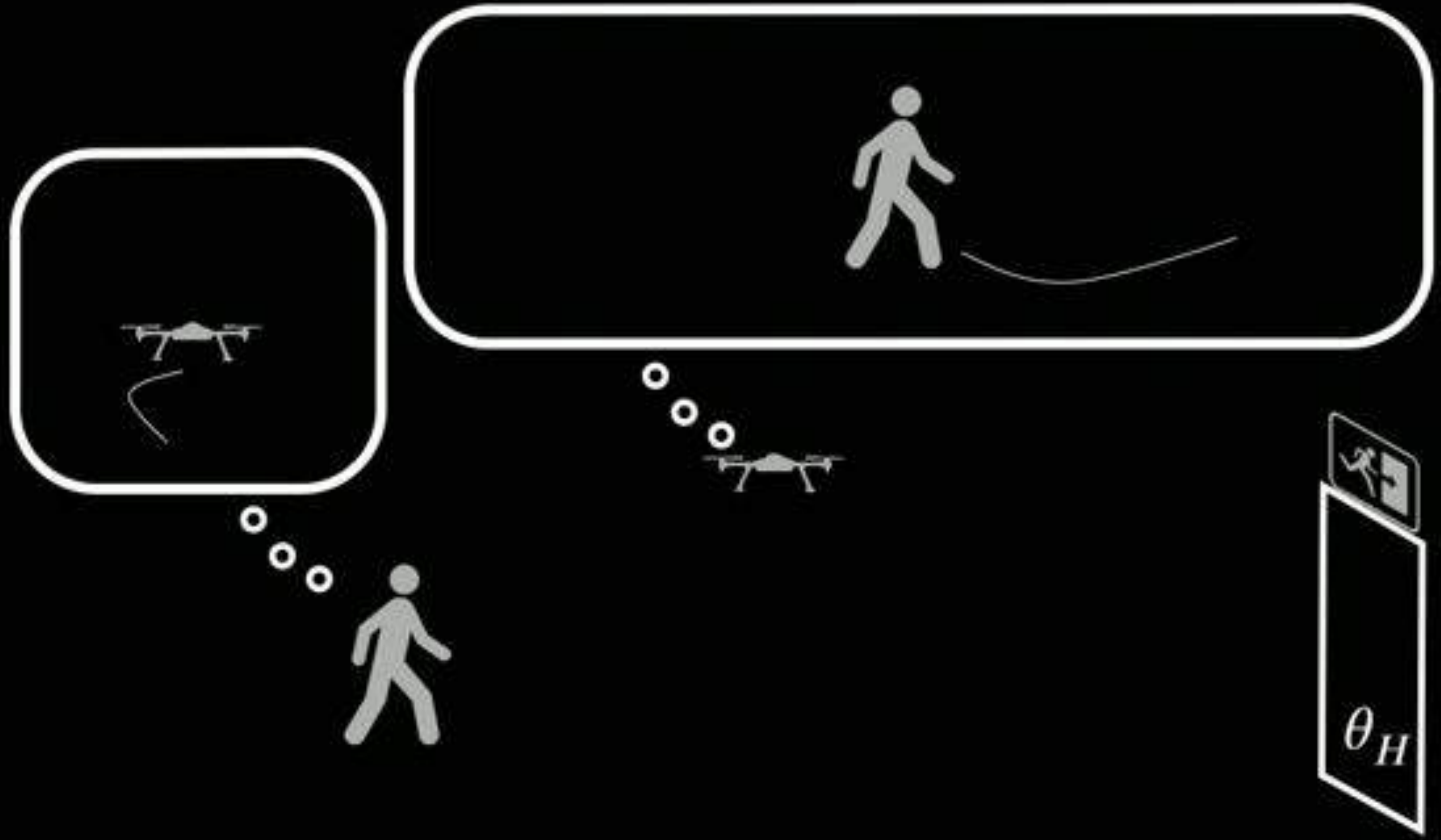

FORTUNER


θ_H

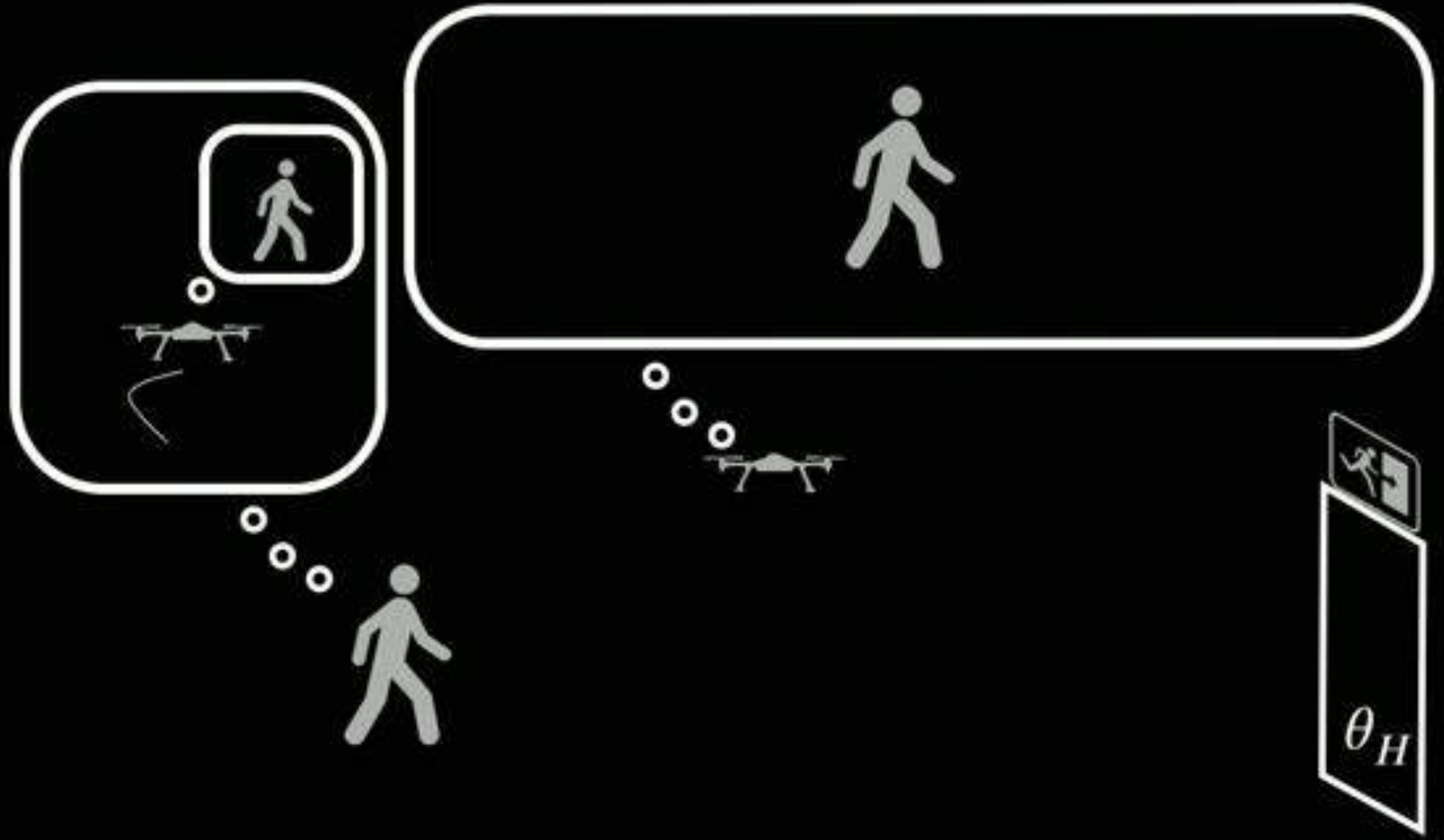
People think about the robot



Implication: the robot's actions influence human actions



People think about what the robot thinks about them

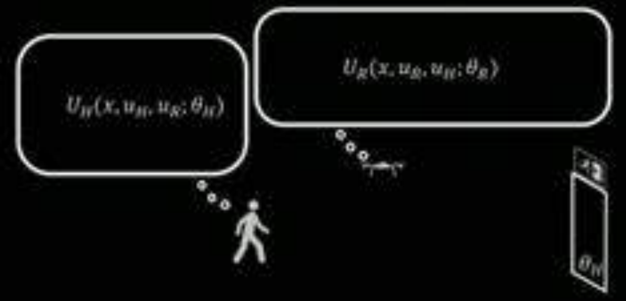


If we could solve this, it'd go like this:

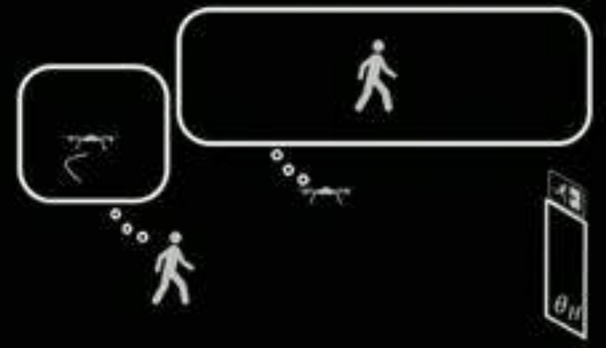
$$V_R(x, T+1) = 0 \quad V_H(x, T+1) = 0$$

Hierarchical approximation

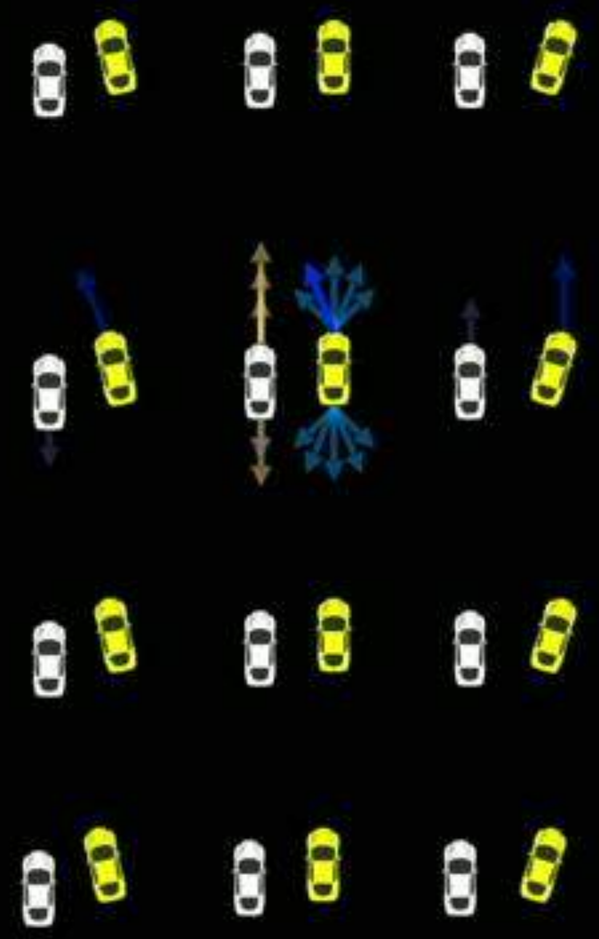
strategic level

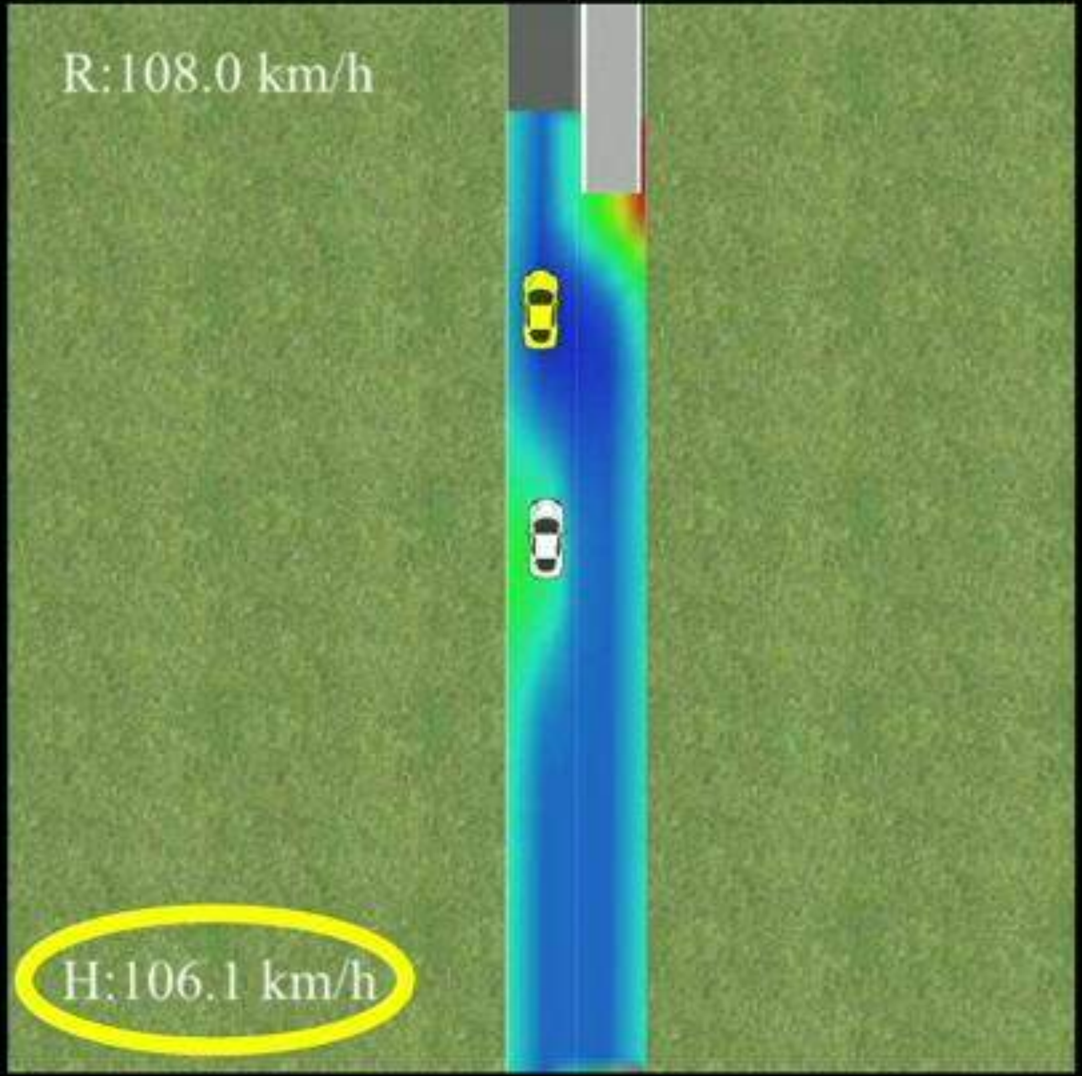


tactical level

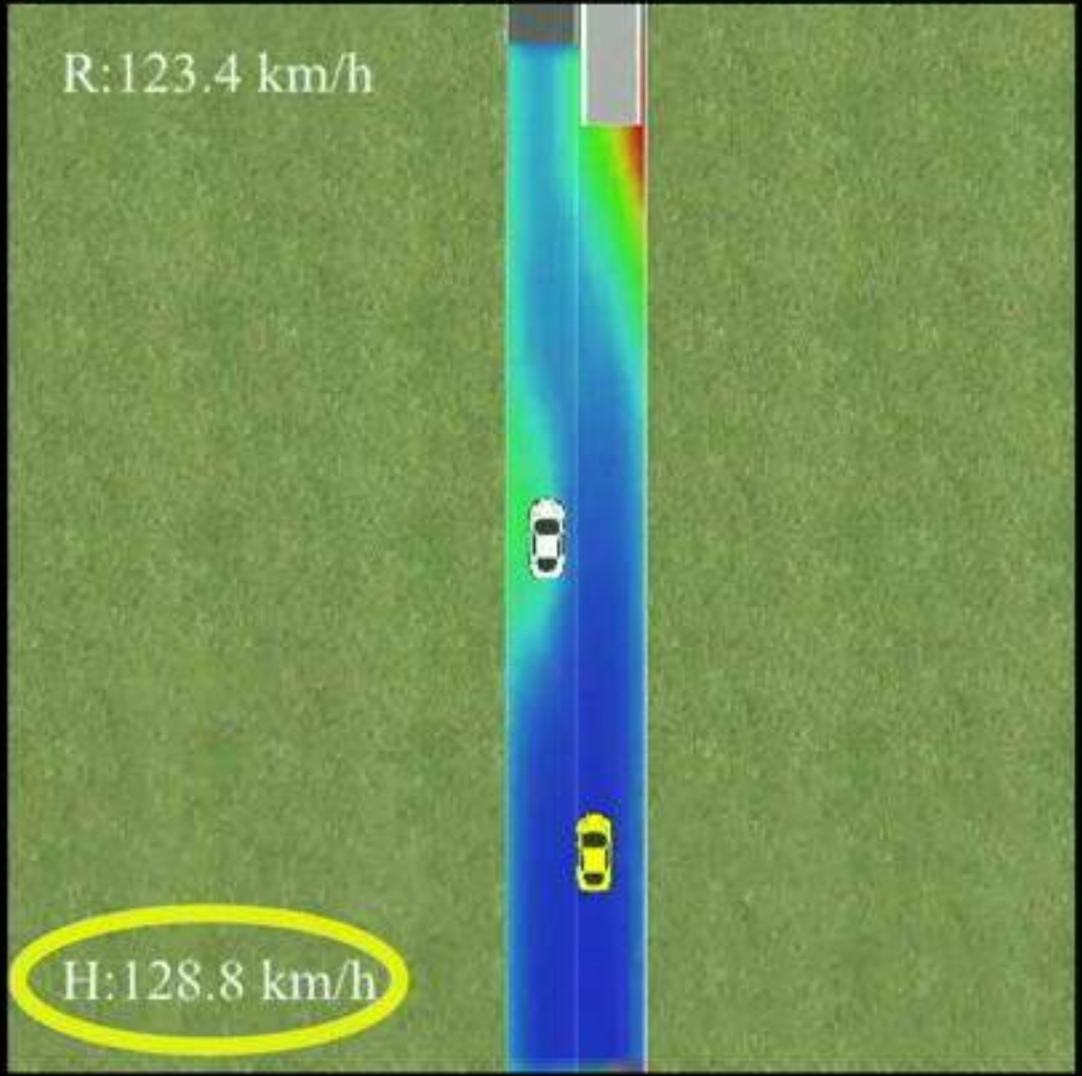


Strategic level: simplified state-action space

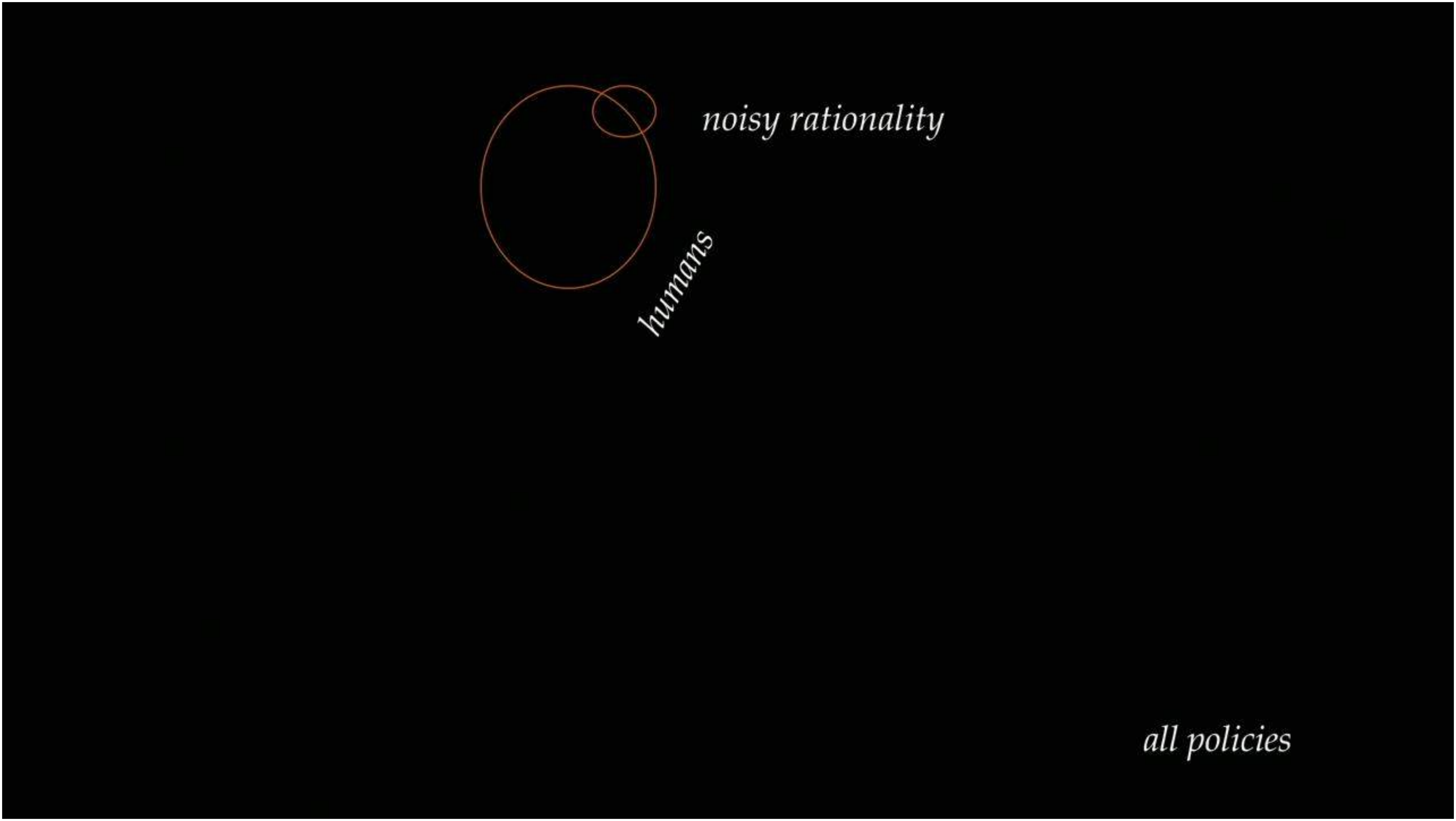


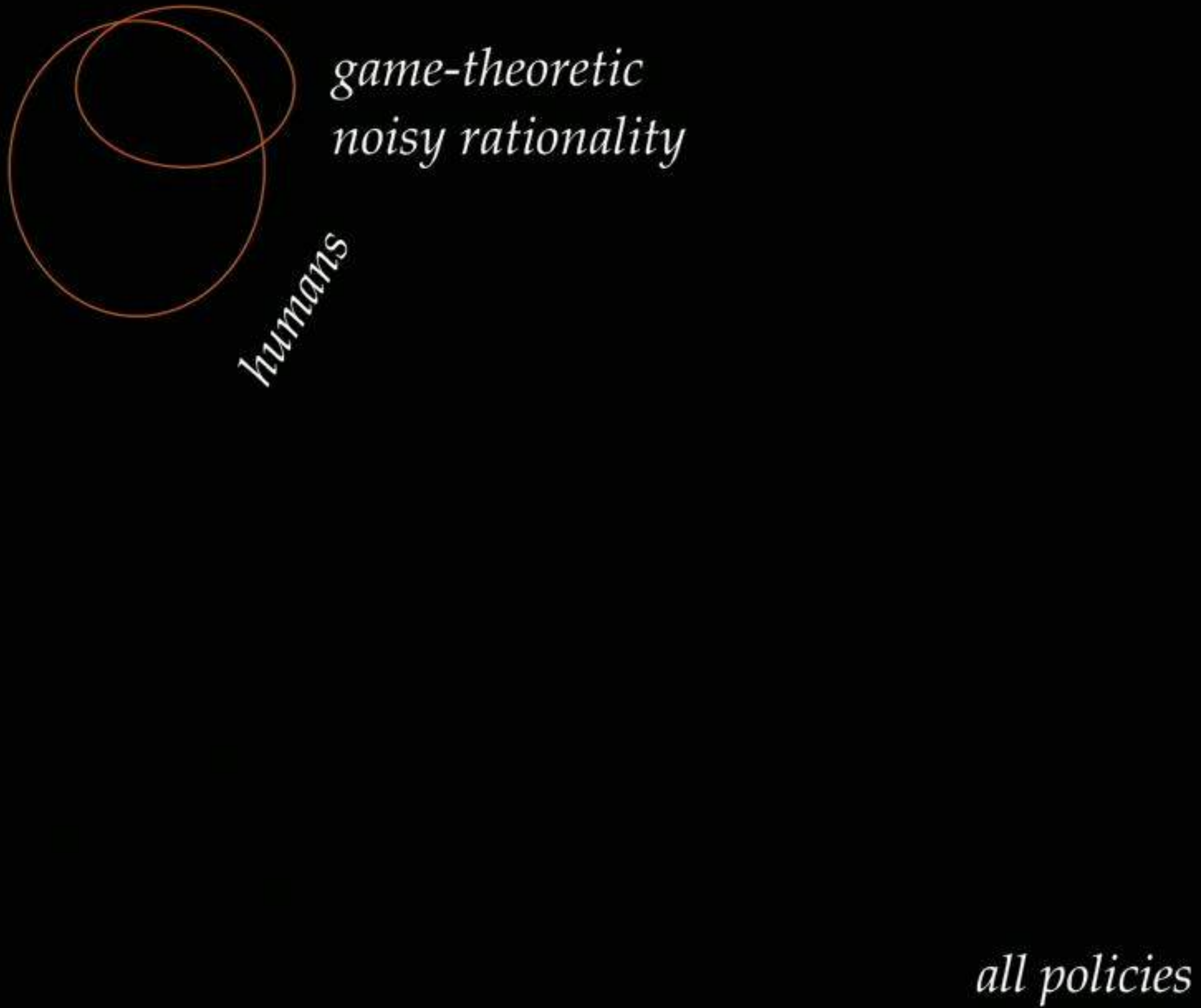


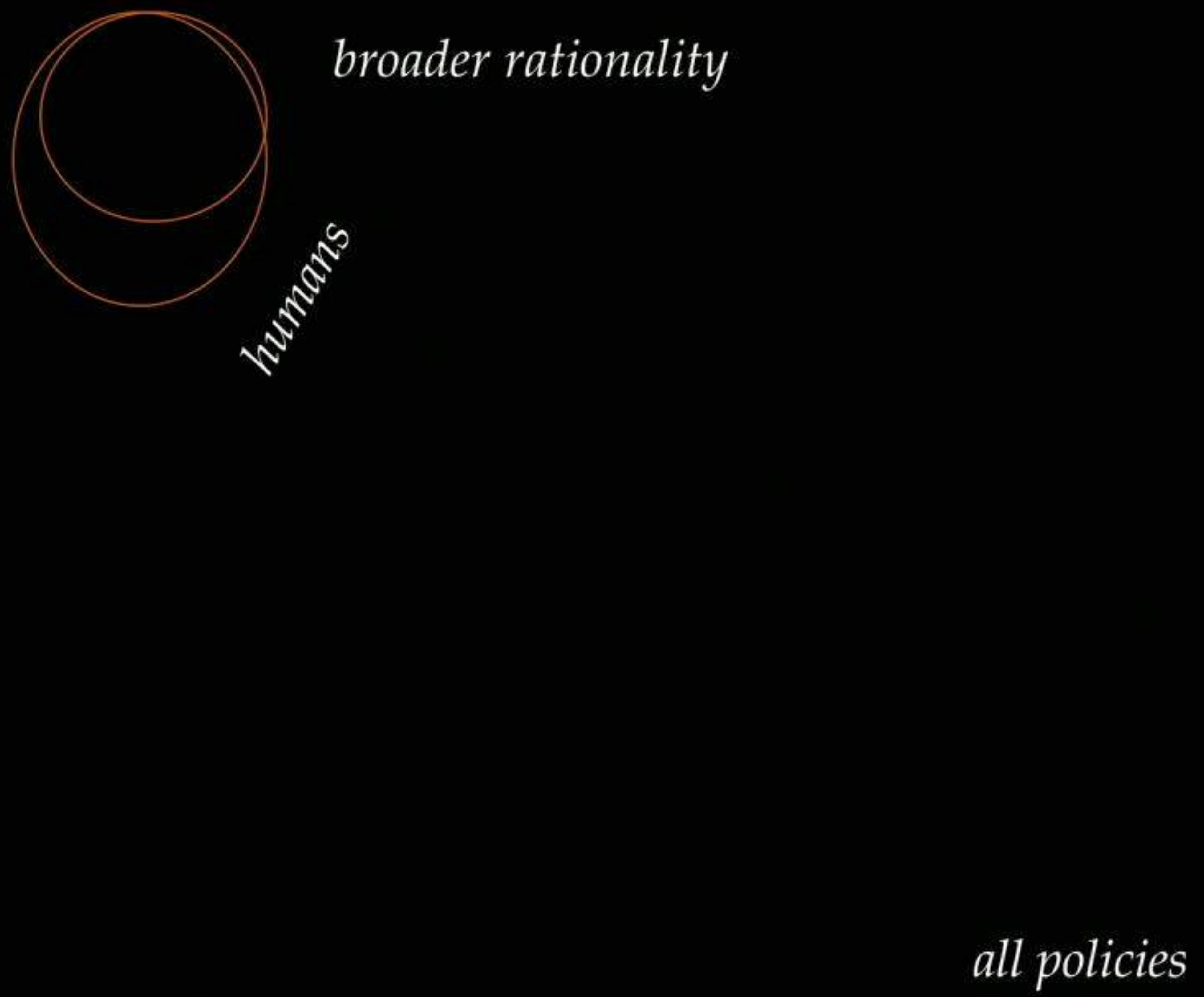
Lead vehicle decelerates



Lead vehicle accelerates


noisy rationality
humans
all policies


game-theoretic
noisy rationality
humans
all policies

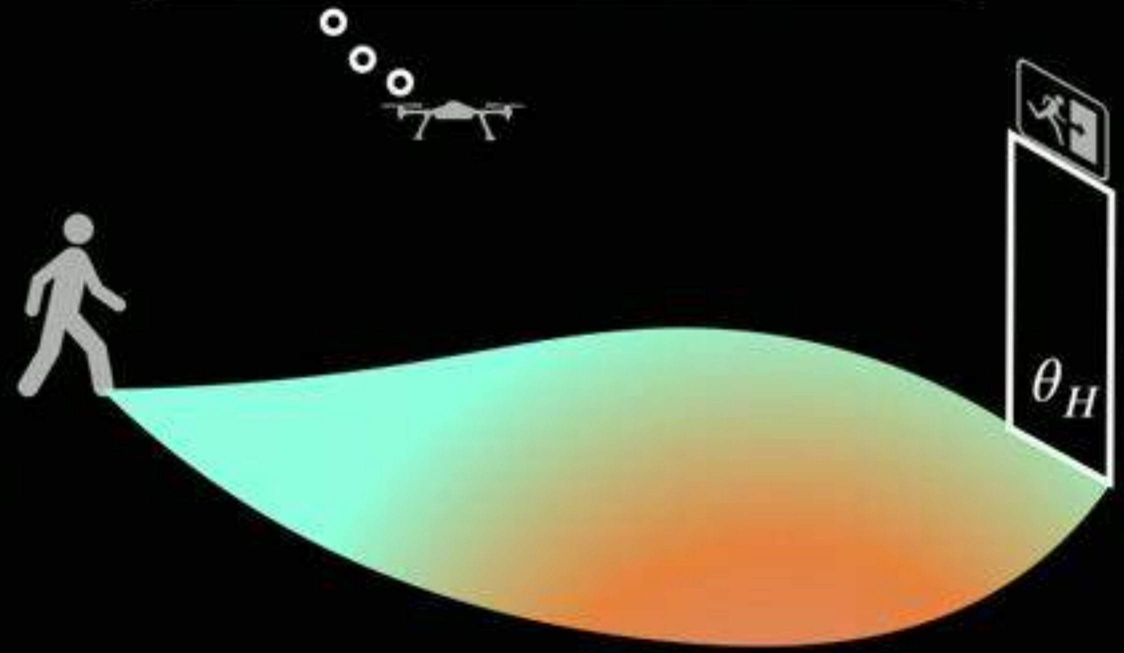

broader rationality
humans
all policies


MIND THE GAP

The rationality coefficient

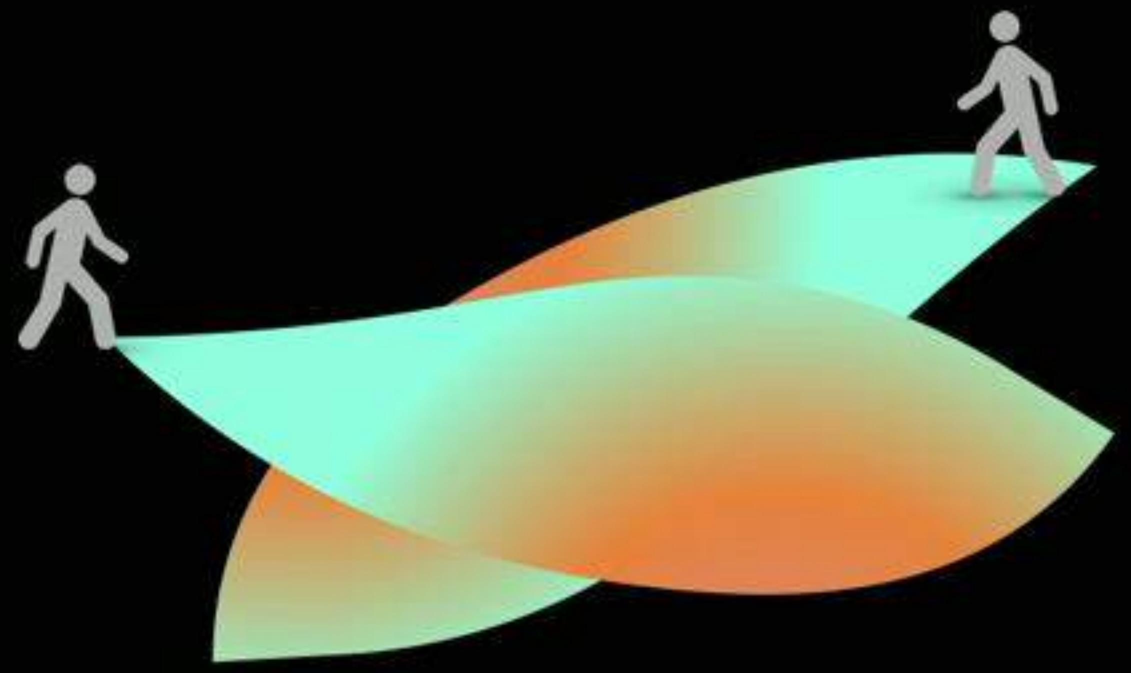
hacky hyperparameter

$$P(u_H \mid x, \theta_H) \propto e^{\beta Q(x, u_H; \theta_H)}$$

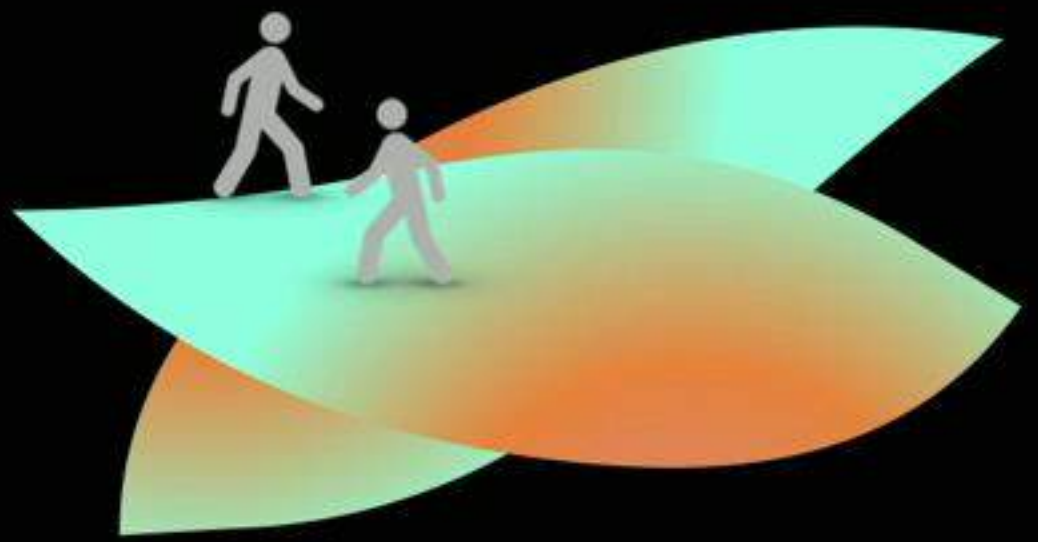


If the human appears
too suboptimal to the model,
be skeptical of the model.

Multiple humans



Multiple humans

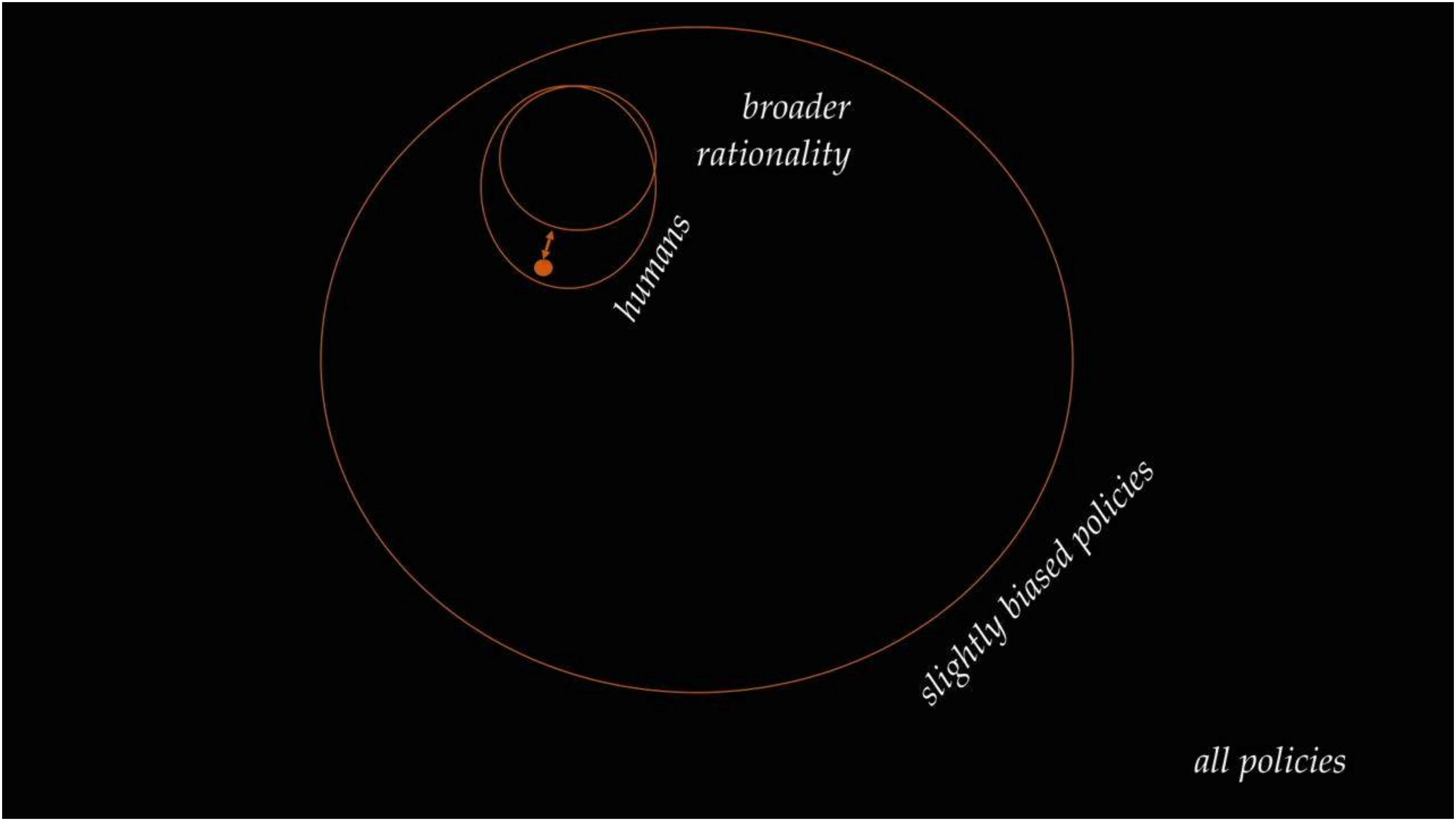


What is the right inductive bias for HRI?

Humans have intent

Humans have intent

inductive bias


broader
rationality
humans
slightly biased policies
all policies

Thanks!