

Last Time:

□ game theory

This Time:

□ world modeling

lecture 12

421, FALL '26

Andrea Bajcsy

What is a world model? Why do we care?

↓
The idea that organisms - humans, animals, maybe robots? - carry "mental models" of the world so that they can try out alternatives before acting.

Technically the transition function $T(s' | s, a)$ is a world model! But so far, we have assumed a known or hand-designed representation of state, $s \in S$. But this is never true in practice! We need to deal with real obs, $o_t \in \mathcal{O}$, coming from sensor like our eyes, touch, taste, smell, and sound!

ex. POMDPs!

ok, well the traditional thing to do is to design a state estimator: $p(s_t | o_{0:t}, a_{0:t})$ which gives us a (distribution) over the current state given a history of the robot's / agent's observations and actions.

⚠ BUT what *is* the representation of s ? 🍅 🍷 🧄
ex: complex interactions like cutting an onion -- what is s ?

ex. or when you are cooking in a cluttered kitchen, do you pay attention to every object? its color/appearance? its temperature?

⚠ on top of this, how do you model the consequences of your actions for such a complicated interaction!?

The idea of world modeling is to study how an organism learns a state representation and a transition function directly from $\mathcal{D} = \{(\mathbf{o}_{0:t}^i, \mathbf{a}_{0:t}^i)\}_{i=1}^N$ observation-action data:

ENCODER $\mathcal{E}(\mathbf{o}) \rightarrow \mathbf{z}$
 observation $\uparrow$ latent state or "representation" $\nwarrow$

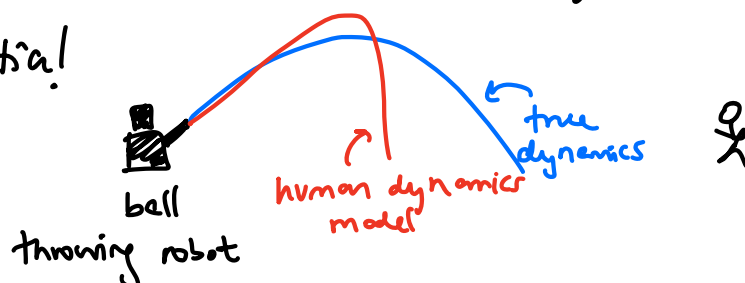
DYNAMICS / TRANSITION $\mathbf{z}' = f(\mathbf{z}, \mathbf{a})$

ex. $\mathcal{D} = \left\{ \left[\begin{array}{c} \text{☺} \quad \text{☹} \quad \text{☹} \\ \mathbf{o}_{0:t} \end{array} \right], \left[\begin{array}{c} \text{down-down-down} \\ \mathbf{a}_{0:t} \end{array} \right] \right\} \rightarrow \left\{ \mathbf{z} \in \mathcal{Z}, f(\mathbf{z}_t, \mathbf{a}_t) \right\}$
 "images of smile" "muscle twitches" "face state" "latent" state "model of how you make expressions"

BUT FIRST, let's talk about:

WORLD MODELS IN RL Reddy, Dragon, Levine (Neurips, 2018)

So far, whenever we have learned from human data (ex. RL, ^{intent} inference) or modeled human behavior, we assumed they were operating / planning under a "perfect world model." But people may deviate from seemingly optimal decisions — this could be because they have 'incorrect' understanding of the world
 ex. classic cog sci work (Caramazza, Cognition 1981) shows that human judgements about physics of projectiles underestimate the effects of inertia!



If robots can understand human "internal world models" then they can coordinate better! ex. better shared controllers/assistance in [Reddy et. al, 2018].

⚠️ For this part, let's assume the state representation is known!



H's policy is a function of $\hat{T}$: $\pi_{\hat{T}}(a|s) \propto e^{Q_{\hat{T}}(s,a)}$

Assume we know $r(s,a) \rightarrow Q_{\hat{T}}(s,a) = r(s,a) + \gamma \mathbb{E}_{s' \sim \hat{T}(s'|s,a)}[V(s')]$
 $\uparrow$ Boltzmann ration-e wrt. $\hat{T}$!

Given a dataset of H behaviors $\mathcal{D} = \{(s,a)\}$, learn:

$$\hat{T}^* := \arg \max_{\hat{T}} \sum_{(s,a) \in \mathcal{D}} \log \pi_{\hat{T}}(a|s) \quad \oplus \text{ This is hard opt! see paper}$$

Once the robot learns $\hat{T}^*$, it can learn to, ex., better assist H!

ex. shared autonomy via internal-to-real dynamics transfer:

IDBA: assist someone by just helping them get to the next state they want to get to $\rightarrow$ intent $\equiv$ next state!

$$a_{\text{assist}}^R = \arg \min_{a^R} D_{KL}[\hat{T}(s'|s, a^R) \parallel T^{\text{real}}(s'|s, a^R)]$$

intuition: robot wants next state distribution in reality to match next state distrib. the person is predicting in their mind.

ex.
imagine the ball is actuated (ex. small drone!)

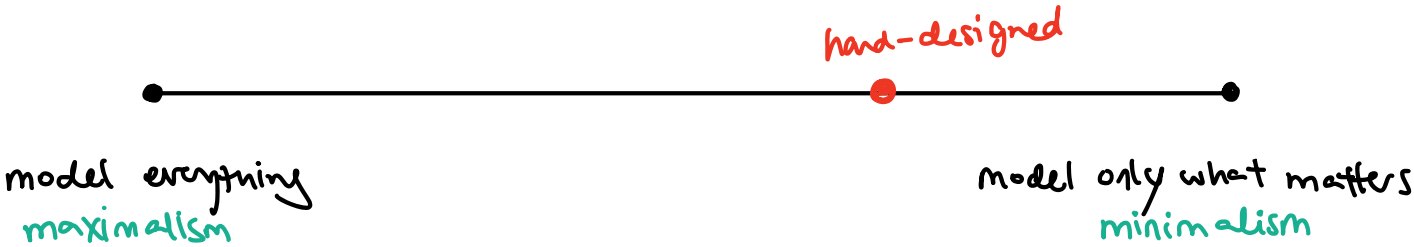


⊕ see paper for real/useful examples!

the mini drone can "brake" after being thrown to resemble the H's prediction

WORLD MODELS IN ROBOTICS

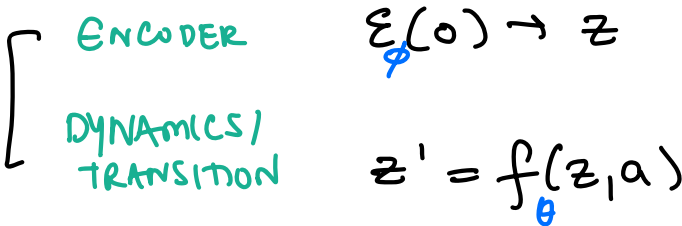
ok but wait , robots don't perfectly understand the world either? And what about the big assumption that we know how to write down the state of the world??
 Let's try to place different approaches to world modeling on a spectrum:



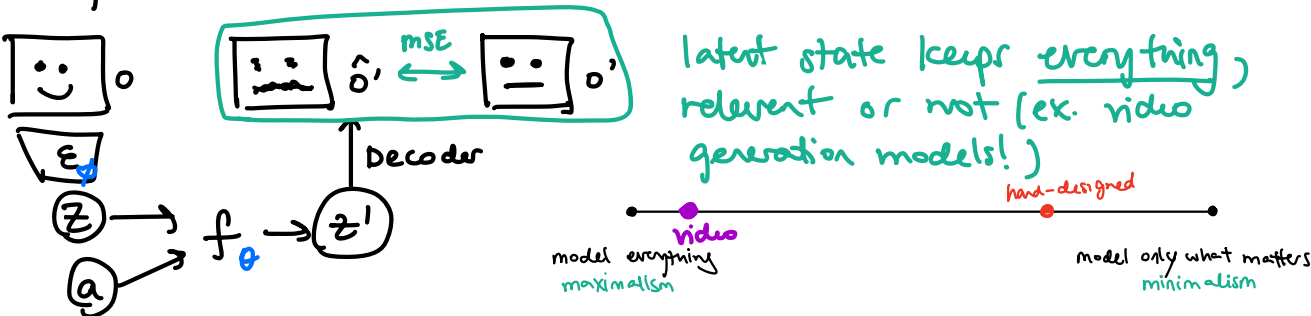
Q when we wrote down a state space + transition model, $T(s'|s,a)$, where on spectrum are we? **Probably here!**

Q now let's say I give you a dataset $\mathcal{D} = \{(o, a, o')\}$

You get to **learn** the state rep + transitions — what should the learning objective be?

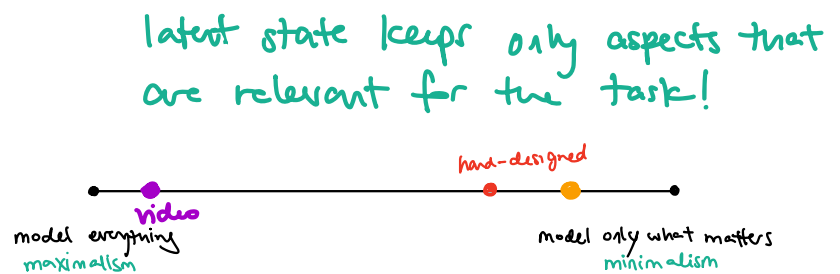
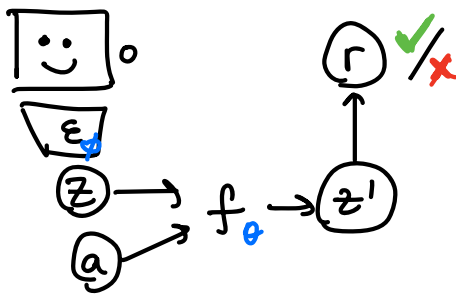


A1 Reconstruction! Your latent z must be able to predict/generate every pixel in your observation!



Q What are some potential problems with this world model?

A2 Predict only what "matters"! State abstractions (and transition models) should be task-driven & minimal. You can discard stuff that doesn't matter in the high-D observation (ex. TV, textures, bg etc.) and focus only on stuff that does (ex. stop light, pothole, color of your fav. mug, etc.). The reward function is a pretty good signal for what "matters"! So now we have:



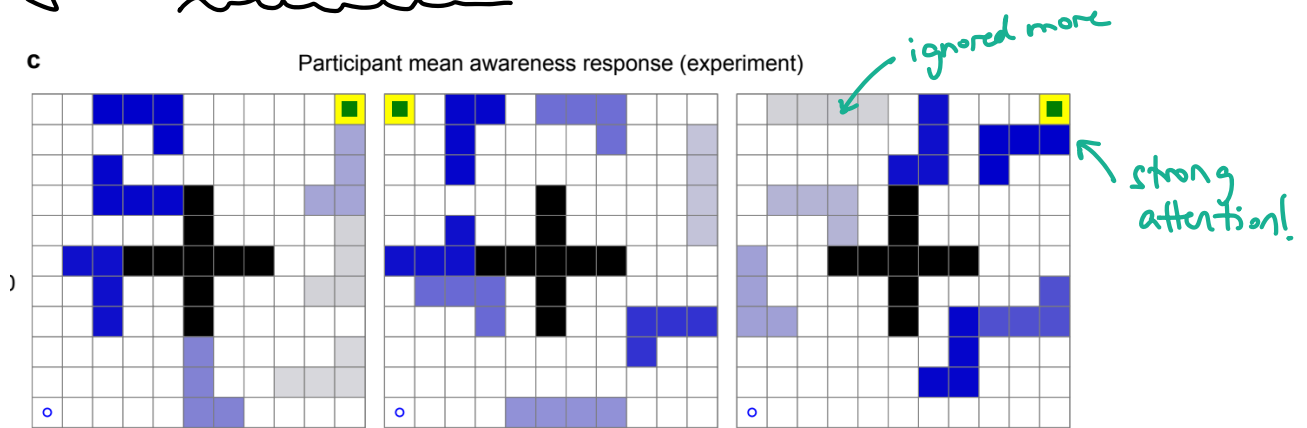
In fact, cognitive science supports the idea that "People construct simplified mental representations to plan" (Ho, Nature 2022)

This paper argues that people are not restricted to a single, full, or fixed representation of the world.

This is especially important b/c humans are resource-rational meaning we try our best to make rational decisions with bounded computation.

The process of adapting the representation & tailoring it to specific tasks/needs is referred to in this work as controlling a task construal:

ex.



H's have to solve maze & go from $\circ \rightarrow \square$. These heatmaps show what objects they focused on/remembered during each task.

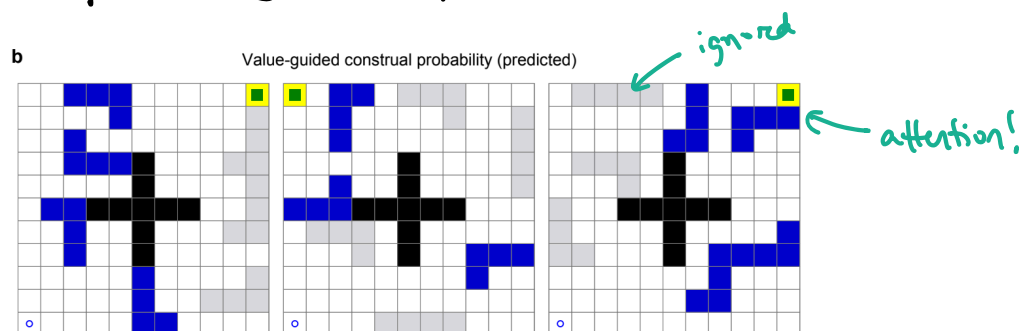
Let's formalize this idea:

- construal $c \subseteq \{\text{obstacles}\}$
- policy under c : $\pi_c = \arg \max_{\pi} R_c(\pi)$
- value of representation:

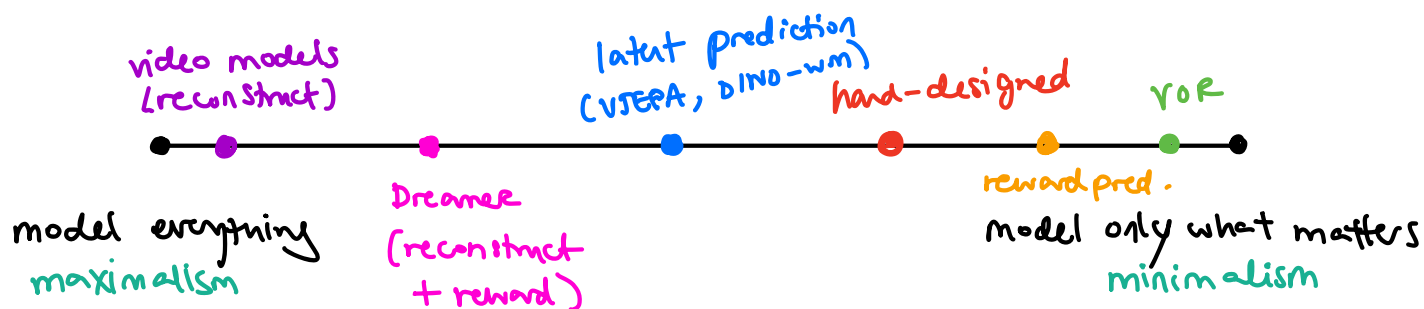
$$VOR(c) = \underbrace{R^*(\pi_c)}_{\text{Reward of policy executed in the REAL WORLD (ex. real maze)}} - \underbrace{C(c)}_{\text{Cognitive cost (ex. } |c| \text{ like num obst.)}}$$

← cumulative reward under c

If you compute VOR for the maze, the model does a very good job predicting the participant data!



There are a lot of unknowns on understanding how humans and robots build world models.



All of this matters b/c

- 1) many HRI failures occur b/c mismatch b/w what robot's model encodes + what human cares about
- 2) we can potentially learn to build better robots — better in the sense of more generalizable + comp. efficient, like ppl! — if we understand humans.