Efficient Model Learning for Dialog Management

Finale Doshi
Nick Roy

CSAIL

Outline

- What are the benefits and challenges of dialog management?
- How can Partially Observable Markov Decision Processes (POMDPs) help handle dialog uncertainty?
- How can we learn user dialog models online through human-robot interactions?

Why dialog management?

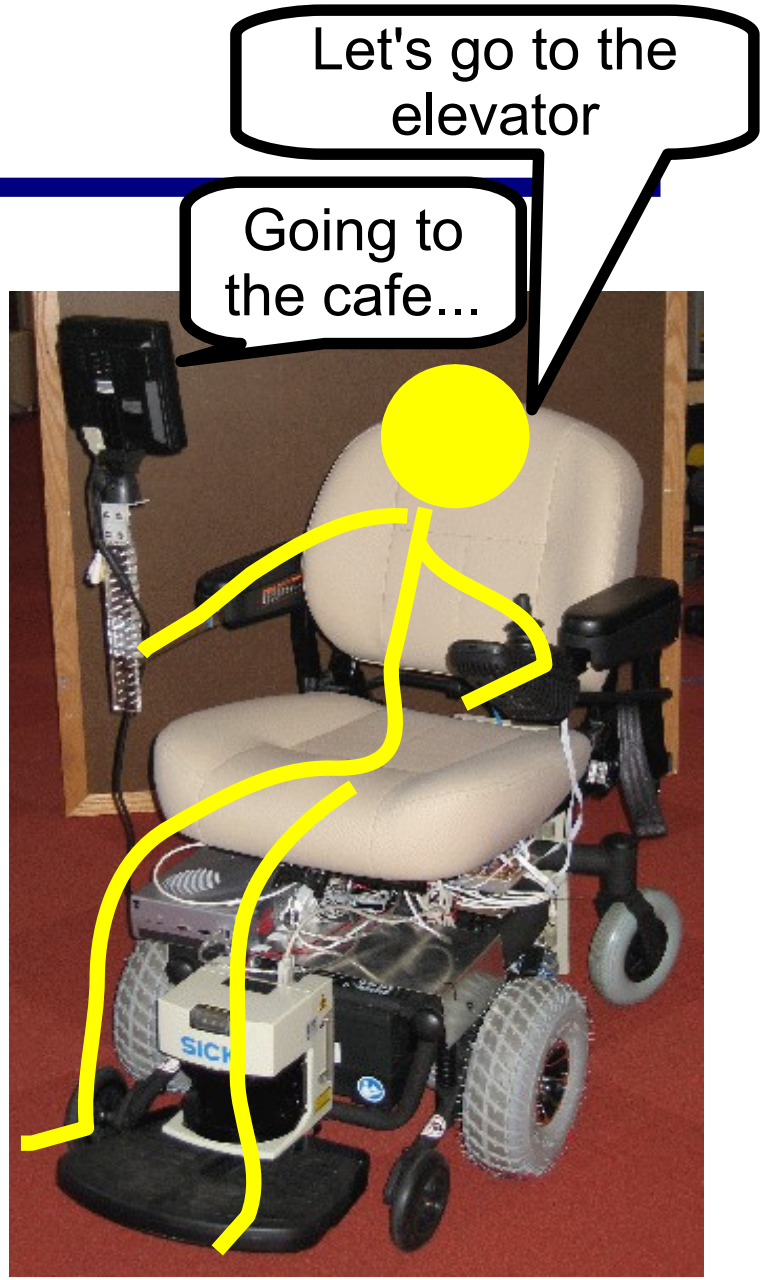
- Wouldn't it be nice if we could simply tell a robot
  - to go to a particular location?
  - to follow alongside someone?
  - to remember the name of a new room?

This is would be particularly useful to wheelchair users with severely limited mobility.

The Problem

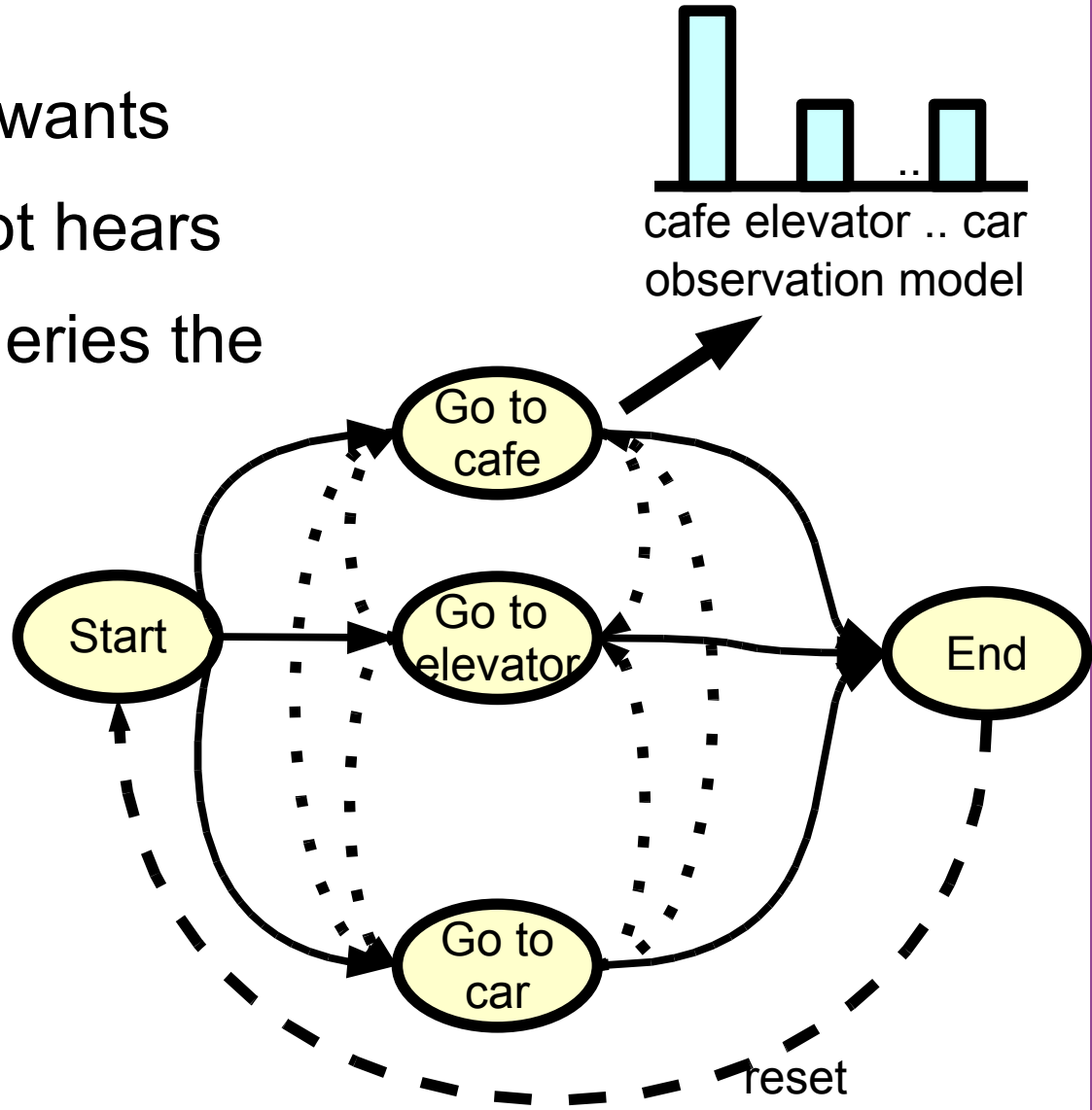
Spoken language allows for natural human robot interaction, but there are several challenges:

- Noisy speech recognition
  - ex. “Gates” becomes “Good”
- Linguistic ambiguities
  - multiple “elevators” may exist
  - the robot must know that an “elevator” is a location



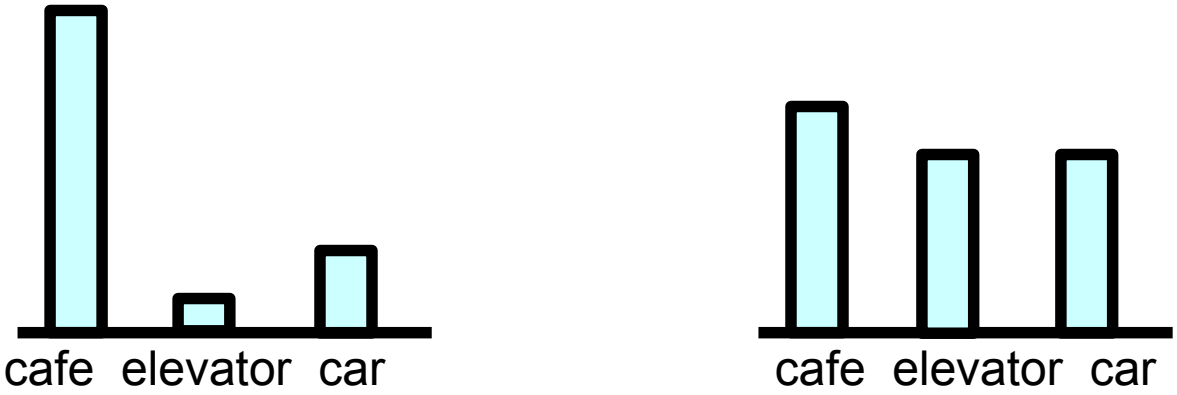
The POMDP Dialog Model

- States (hidden!): the user's wants
- Observations: what the robot hears
- Actions: movements and queries the robot can do
- Reward Model R(s,a)
- Transition Model T(s'|s,a)
- Observation Model O(o|s,a)



Why is the POMDP model useful?

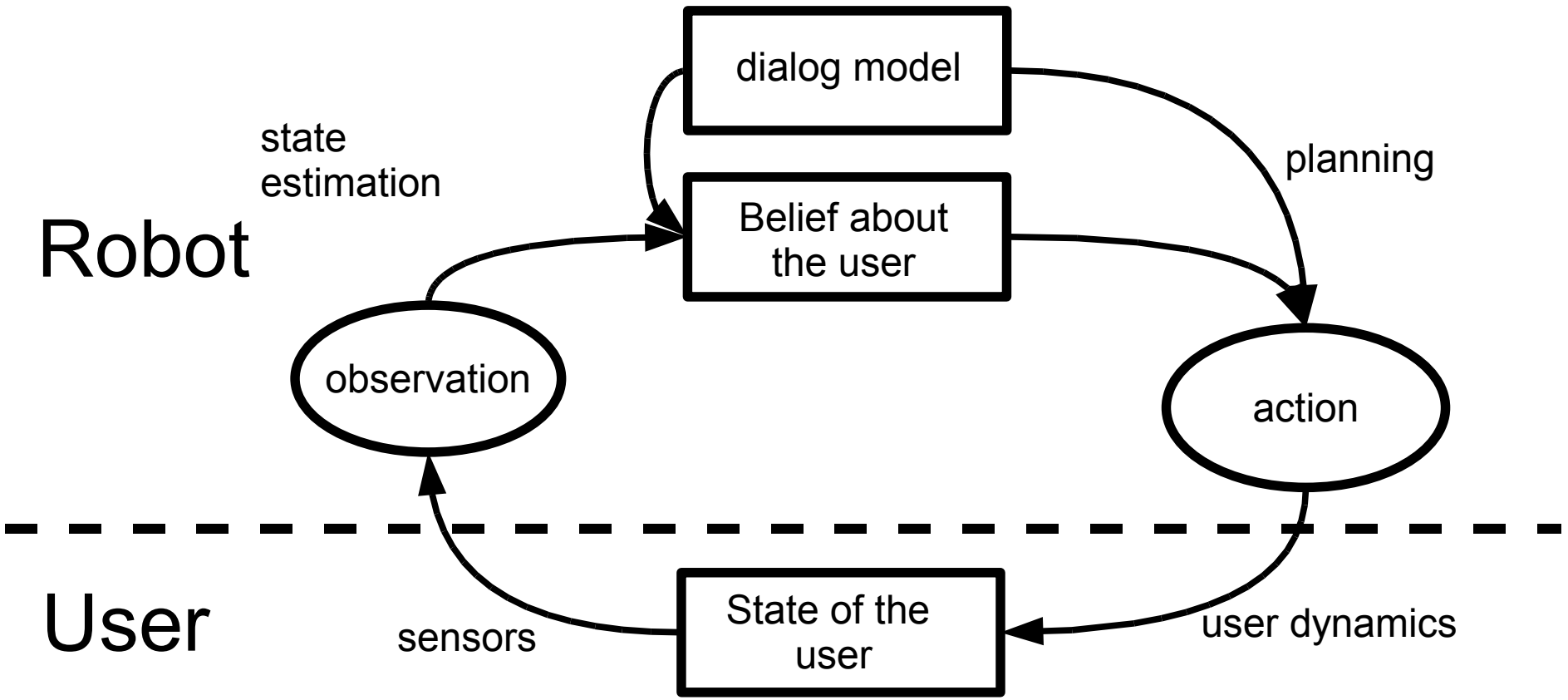
- We track our *belief*, a probability distribution over states.
- We choose an action based on our belief, thus taking into account our uncertainty about what the user really wants.



- POMDPs have been used in several dialog management applications, such as Roy, Pineau, and Thrun (2000) and Williams and Young (2005)

CSAIL

The planning process.



Solving the POMDP Dialog Model

Value of a belief

Value of belief, action pair

$$V_n(b) = max_a Q_n(b,a)$$

$$Q_n(b,a) = R(b,a) + \gamma \sum_{b' \in B} T(b'|b,a) V_{n-1}(b')$$

$$Q_n(b,a) = R(b,a) + \gamma \sum_{o \in O} O(o|b,a) V_{n-1}(b_a^o)$$

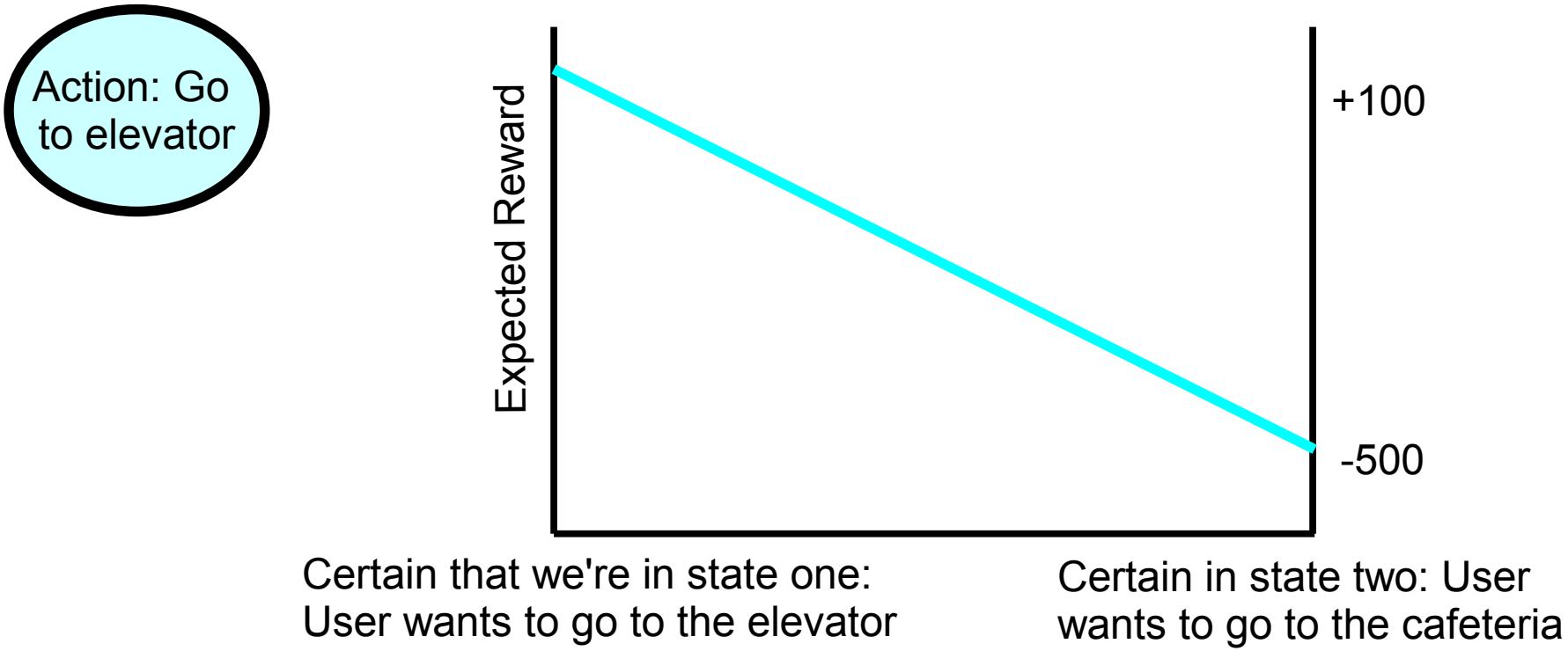
Current reward

Future Reward

CSAIL

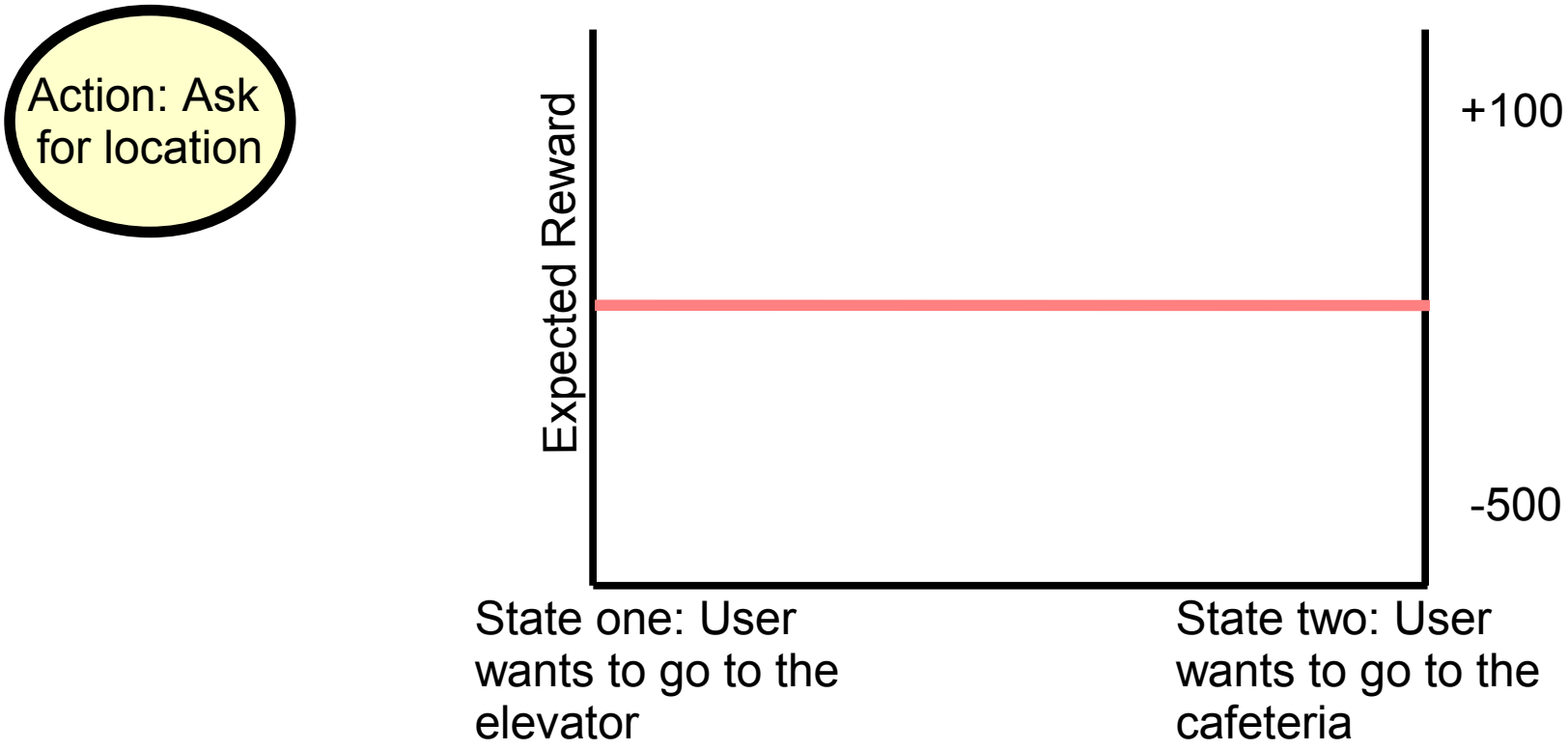
Dialog Model: Solving the POMDP

We think of the previous recursions as building a policy tree...



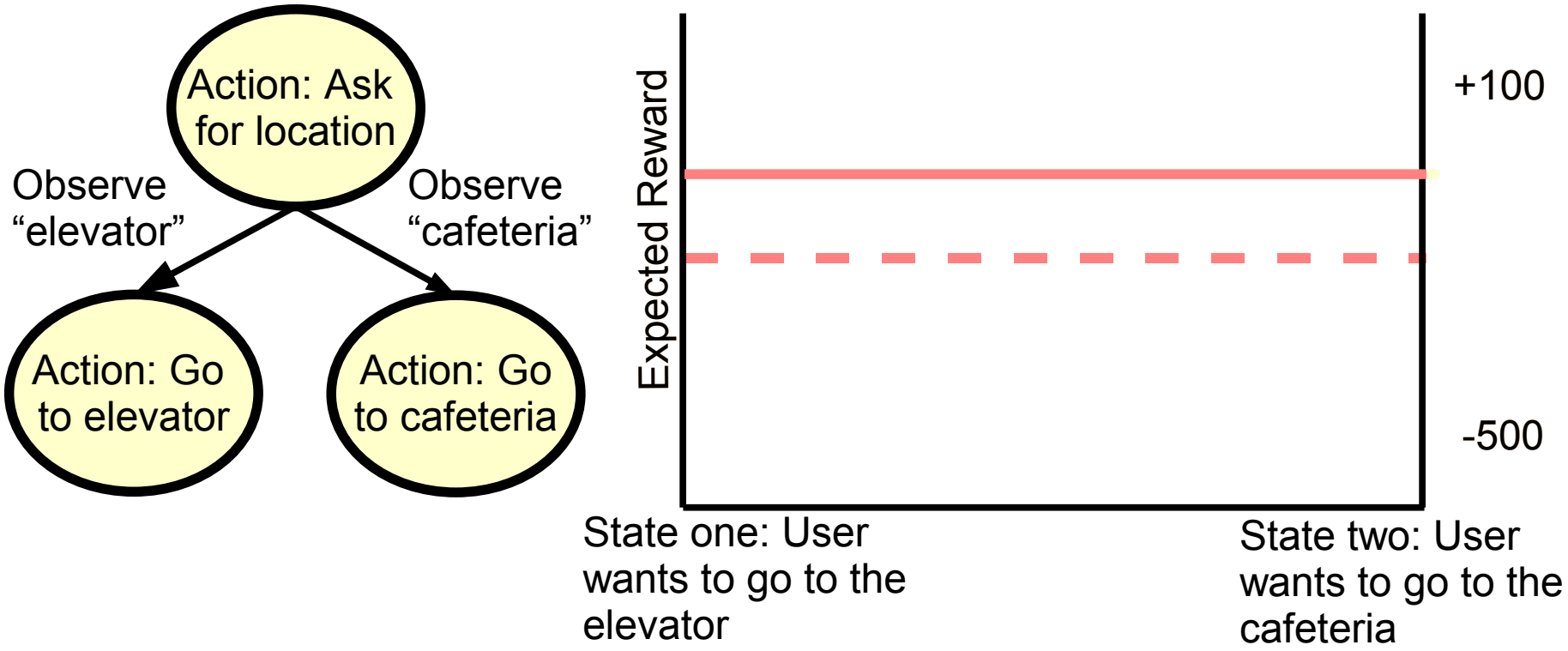
Dialog Model: Solving the POMDP

We think of the previous recursions as building a policy tree...



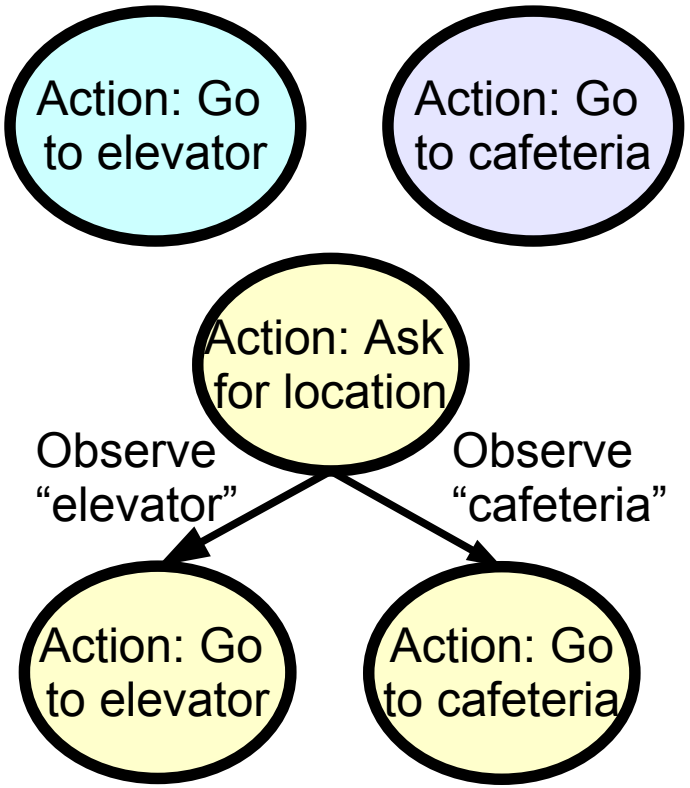
Dialog Model: Solving the POMDP

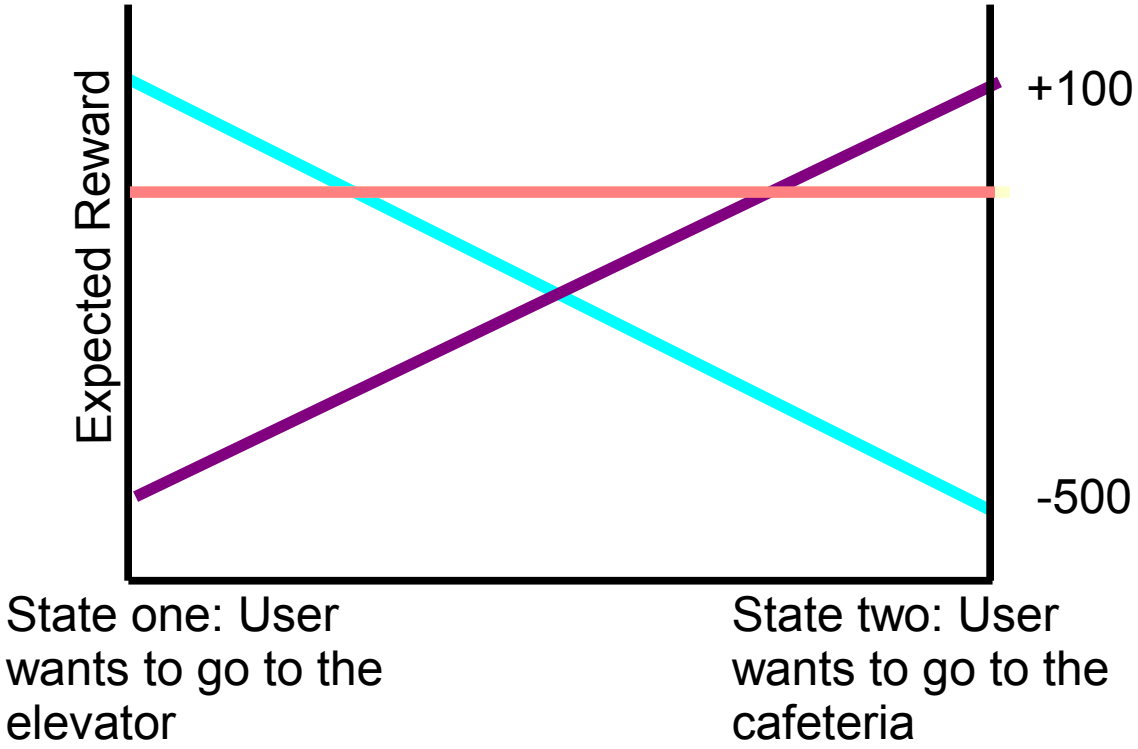
We think of the previous recursions as building a policy tree; planning ahead increases our expected reward.



Dialog Model: Solving the POMDP

Given multiple trees, we can determine the most appropriate action:





Difficulties with the POMDP model

R(user wants Gates, go to Gates) R(user wants Gates, go to Dreyfoos) R(user wants Gates, go to infodesk) R(user wants Gates, go to cafe) R(user wants Gates, go to parking lot) O(hear Gates|user wants Gates, ask where) O(hear Gates|user wants Dreyfoos, ask where) O(hear Gates|user wants cafe, ask where) O(hear Gates|user wants infodesk, ask where) O(hear Gates|user wants parking lot, ask wh m Gates) O(hear Gates|user wants r wants cafe, confirm Gates) O(Gates) O(hear Gates|user wants wants Gates, confirm Dreyfoos) Dreyfoos) O(hear Gates|user wants wants parking lot, confirm Dreyfoos) Dreyfoos) O(hear Gates|user wants wants Dreyfoos, confirm infodesk) O(hear Gates|user wants cafe, confirm infodesk) O(hear Gates|user wants parking lot, confirm infodesk) O(hear Gates|user wants infodesk, confirm infodesk) O(hear Gates|user wants Gates, confirm parking lot) O(hear Gates|user wants Dreyfoos, confirm parking lot) O(hear Gates|user wants cafe, confirm parking lot) O(hear Gates|user wants parking lot, confirm parking lot) O(hear Gates|user wants infodesk, confirm parking lot) O(hear Gates|user wants Gates, confirm cafe) O(hear Gates|user wants Dreyfoos, confirm cafe) O(hear Gates|user wants cafe, confirm cafe) O(hear Gates|user wants parking lot, confir

Even our simple model has **1344** parameters!
(I only listed 35)

CSAIL

Difficulties with the POMDP model

R(user wants Gates, go to Gates) R(user wants Gates, go to Dreyfoos) R(user wants Gates, go to infodesk) R(user wants Gates, go to cafe) R(user wants Gates, go to parking lot) O(hear Gates|user wants Gates, ask where) O(hear Gates|user wants Dreyfoos, ask where) O(hear Gates|user wants cafe, ask where) O(hear Gates|user wants infodes ear Gates| user wants parking lot, ask where) O(hear Gates|u confirm Gates) O(hear Gates|user wants Dreyfoos, confir ates|user wants cafe, confirm Gates) O(hear Gates|user wa firm Gates) O(hear Gates|user wants infodesk, confirm Gates r wants Gates, confirm Dreyfoos) O(hear Gates|user want n Dreyfoos) O(hear Gates|user wants cafe, confirm Dreyfoos) O(hear Gates|user wants parking lot, confirm Dreyfoos) O(hear Gates|user wants infodesk, confirm Dreyfoos) O(hear Gates|user wants Gates, confirm infodesk) O(hear Gates|user wants Dreyfoos, confirm infodesk) O(hear Gates|user wants cafe, confirm infodesk) O(hear Gates|user wants parking lot, confirm infodesk) O(hear Gates| user wants infodesk, confirm infodesk) O(hear Gates|user wants Gates, confirm parking lot) O(hear Gates|user wants Dreyfoos, confirm parking lot) O(hear Gates|user wants cafe, confirm parking lot) O(hear Gates|user wants parking lot, confirm parking lot) O(hear Gates|user wants infodesk, confirm parking lot) O(hear Gates|user wants Gates, confirm cafe) O(hear Gates|user wants Dreyfoos, confirm cafe) O(hear Gates|user wants cafe, confirm cafe) O(hear

This is probably large

CSAIL

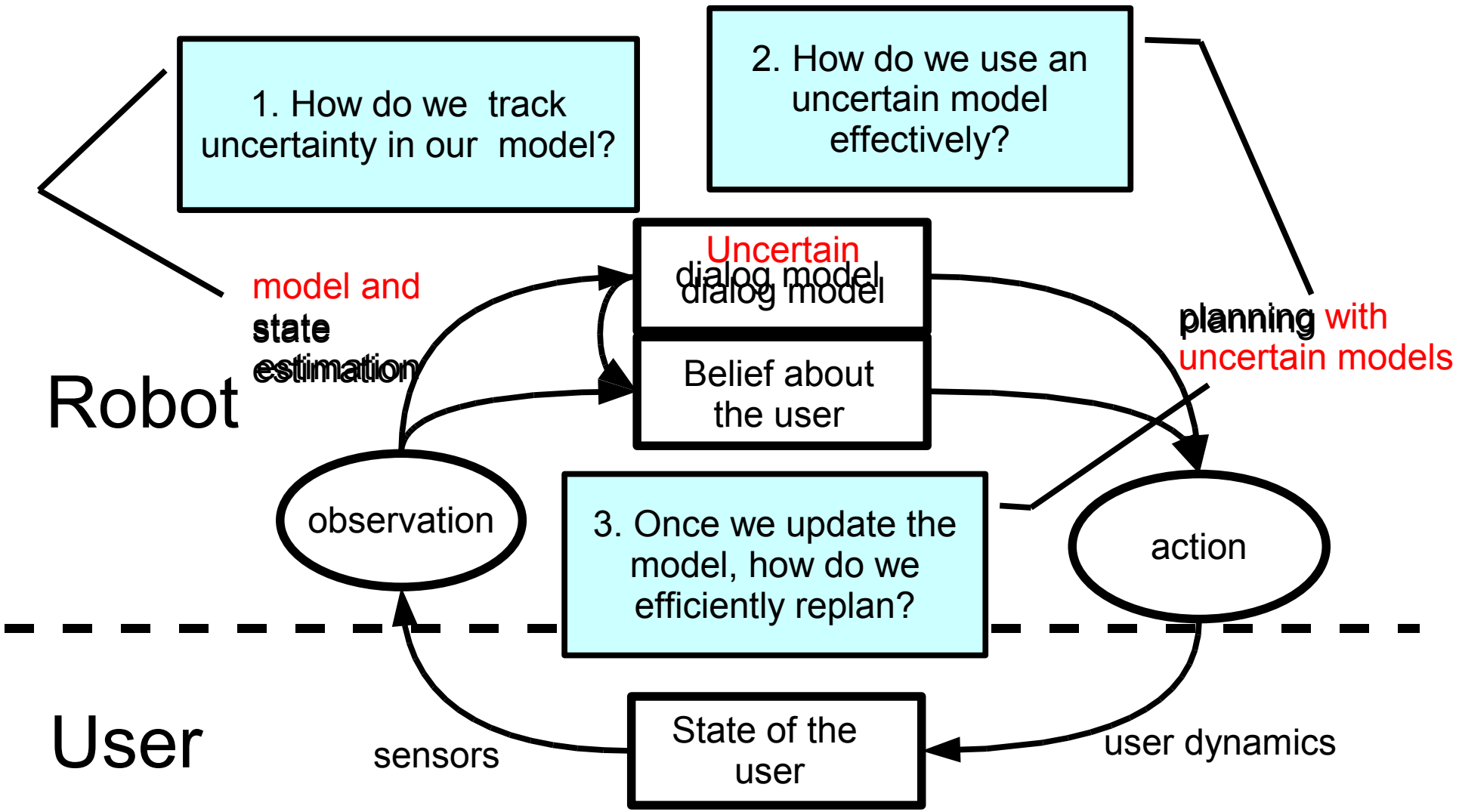
Difficulties with the POMDP model

R(user wants Gates, go to Gates) R(user wants Gates, go to Dreyfoos) R(user wants Gates, go to infodesk) R(user wants Gates, go to cafe) R(user wants Gates, go to parking lot) O(hear Gates|user wants Gates, ask where) O(hear Gates|user wants Dreyfoos, ask where) O(hear Gates|user wants cafe, ask where) O(hear Gates|user wants infodesk, ask where) O(hear Gates|user wants parking lot, ask where) O(hear Gates|user wants Gates, confirm Gates) O(hear Gates|us… confirm Gates) O(hear Gates|user wants cafe, confirm G…s|user wants parking lot, confirm Gates) O(hear Gates|us… confirm Gates) O(hear Gates|user wants Gates, confirm D…ates|user wants Dreyfoos, confirm Dreyfoos) O(hear Gates|us…irm Dreyfoos) O(hear Gates|user wants parking lot, confirm D…ates|user wants infodesk, confirm Dreyfoos) O(hear Gates|us…nfirm infodesk) O(hear Gates|user wants Dreyfoos, confirm infodesk) O(hear Gates|user wants cafe, confirm infodesk) O(hear Gates|user wants parking lot, confirm infodesk) O(hear Gates|user wants infodesk, confirm infodesk) O(hear Gates|user wants Gates, confirm parking lot) O(hear Gates|user wants Dreyfoos, confirm parking lot) O(hear Gates|user wants cafe, confirm parking lot) **O(hear Gates|user wants parking lot, confirm parking lot)** O(hear Gates|user wants infodesk, confirm parking lot) O(hear Gates|user wants Gates, confirm cafe) O(hear Gates|user wants Dreyfoos, confirm cafe) O(hear Gates|user wants cafe, confirm cafe)

This is probably small

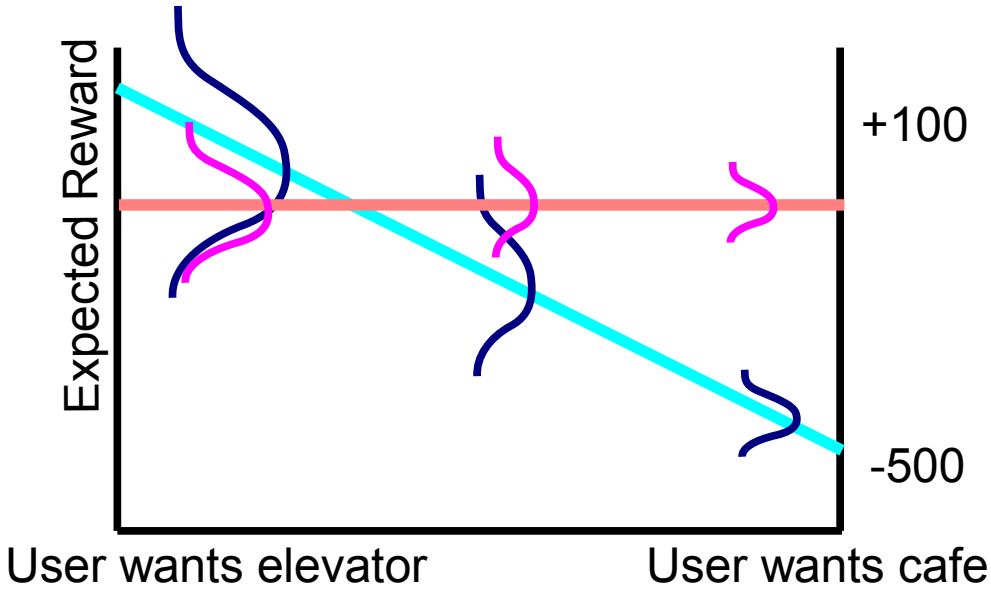
CSAIL

Model uncertainty and the planning process.



1. Modeling Uncertainty: Placing Priors

- An expert provides guesses using “pre-observation counts.” For example: suppose I think P(error) = 10%
  - If I'm pretty certain, guess “saw 100 errors in 1000 tests”
  - If I'm not sure, might guess “saw 1 error in 10 tests”
- Parameter uncertainty induces uncertainty in the value function.



CSAIL

2. Planning with uncertain parameters

If parameters are uncertain, solve POMDP with the expected value of the parameters to optimize reward.

Expectation over states

$$Q_n(b,a) = max_i \, q_{a,i} \cdot b$$

$$q_{a,i}(s) = E[R(s,a)] + \gamma \sum_{o \in O} \sum_{s' \in S} E[T(s'|s,a)] E[O(o|s',a)] V_{n-1,i}(s)$$

Expectations over model parameters

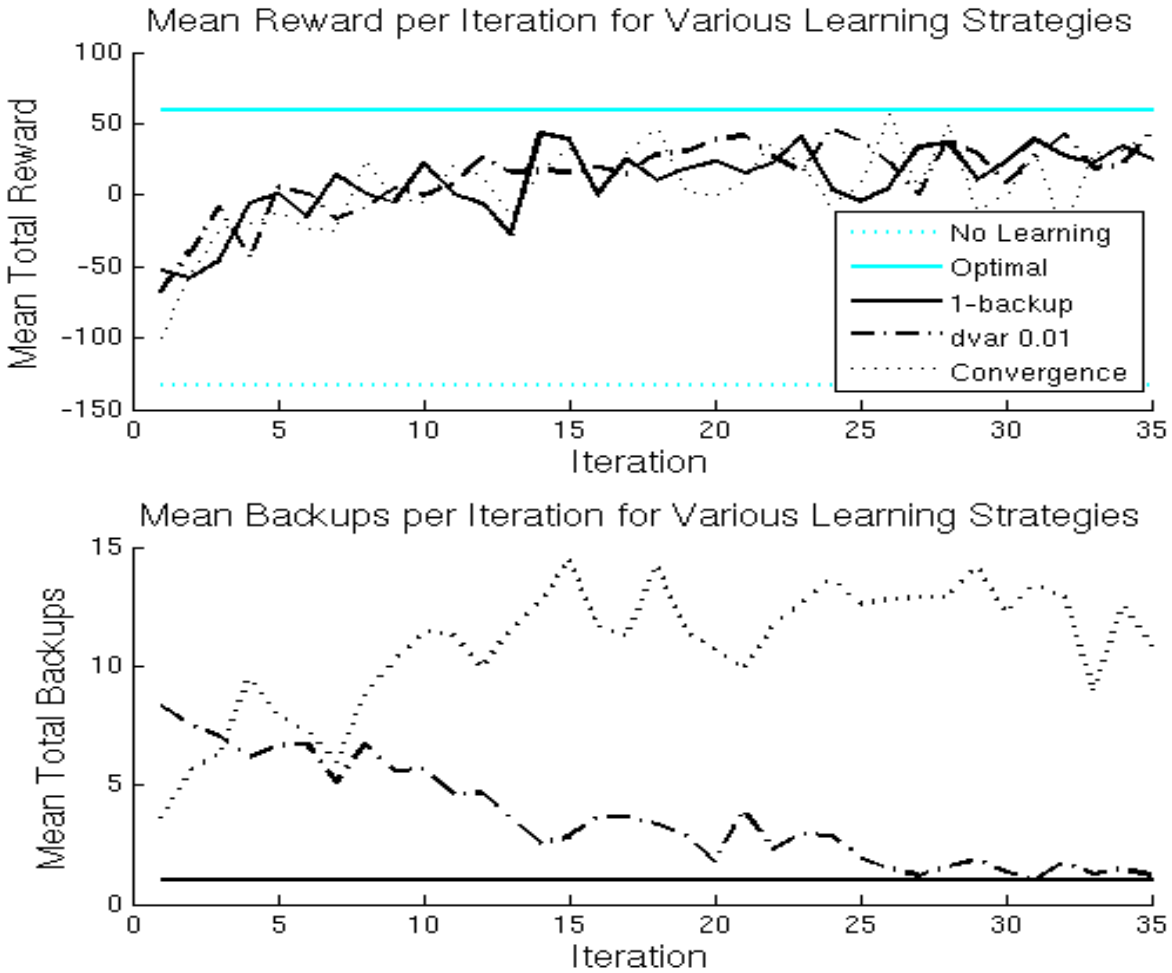
Expectation over model stochasticity

CSAIL

3. Updating the Policy

- How? Create new policy trees initialized with values from the old policy trees.
  - POMDP solution will converge given new expected values for the parameters.
- How deep should the new trees be?
  - Grow to convergence
  - Grow to a fixed depth
  - Grow proportionally to variance reduction in the parameters

Simulation Results



CSAIL

Wheelchair Results

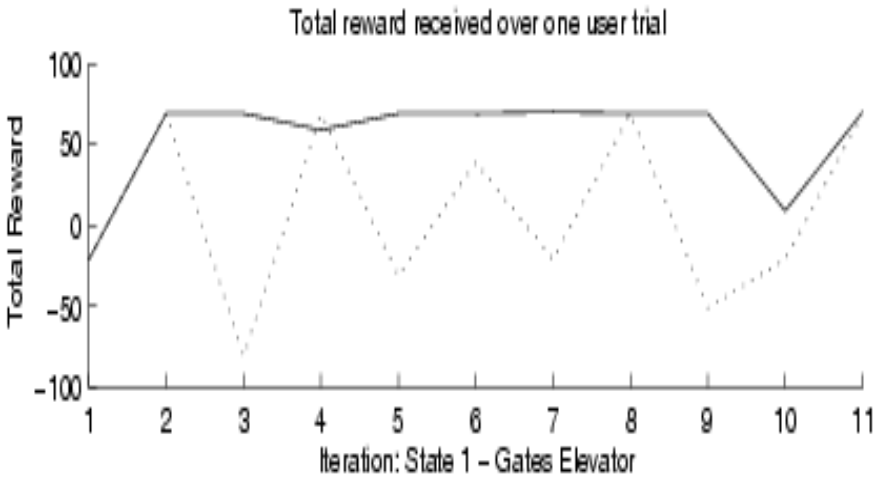
CSAIL

Wheelchair Results: Sample Dialog

Non-learner
User: Take me to the elevator.
Robot: Where did you want to go?
User: The Gates elevator please.
Robot: Do you want to go to the Gates Elevator?
User: Yes.
Robot: Going to Gates.

Learner
User: Take me to the elevator.
Robot: Do you want to go to the Gates Tower?
User: Yes
Robot: Going to Gates

Wheelchair Results by Goal State

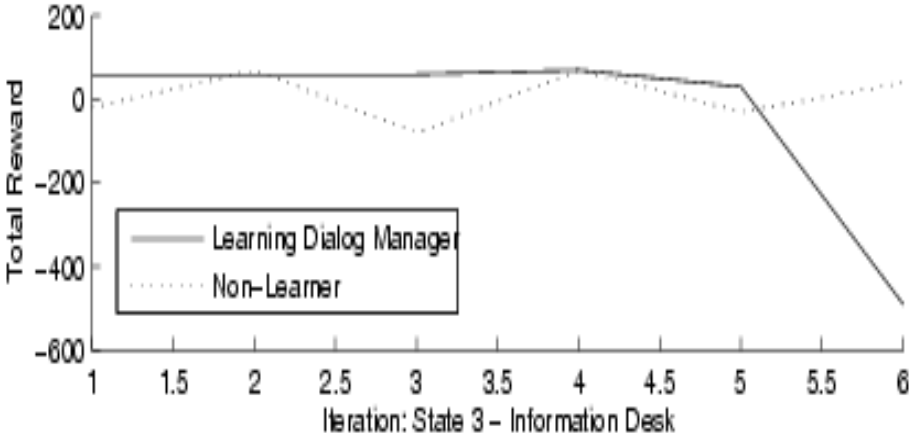


State Property	Performance
Ambiguous vocabulary	✓
New vocabulary	
Very little state-specific vocabulary	

Wheelchair Results by Goal State

State Property	Performance
Ambiguous vocabulary	✓
New vocabulary	✓
Very little state-specific vocabulary	

Wheelchair Results by Goal State



State Property	Performance
Ambiguous vocabulary	✓
New vocabulary	✓
Very little state-specific vocabulary	✗

Wheelchair Video

Continuing Work: Larger Questions

- How can we guarantee the convergence of key parameters?
  - Make POMDP aware of parameter uncertainty by incorporating parameters as additional hidden state.
- How can the robot take exploratory actions *and* act robustly?
  - “Meta-Actions,” queries about what the robot should do, are a low-risk way to discover user preferences.
- What capabilities would be most useful to wheelchair users?

Conclusions

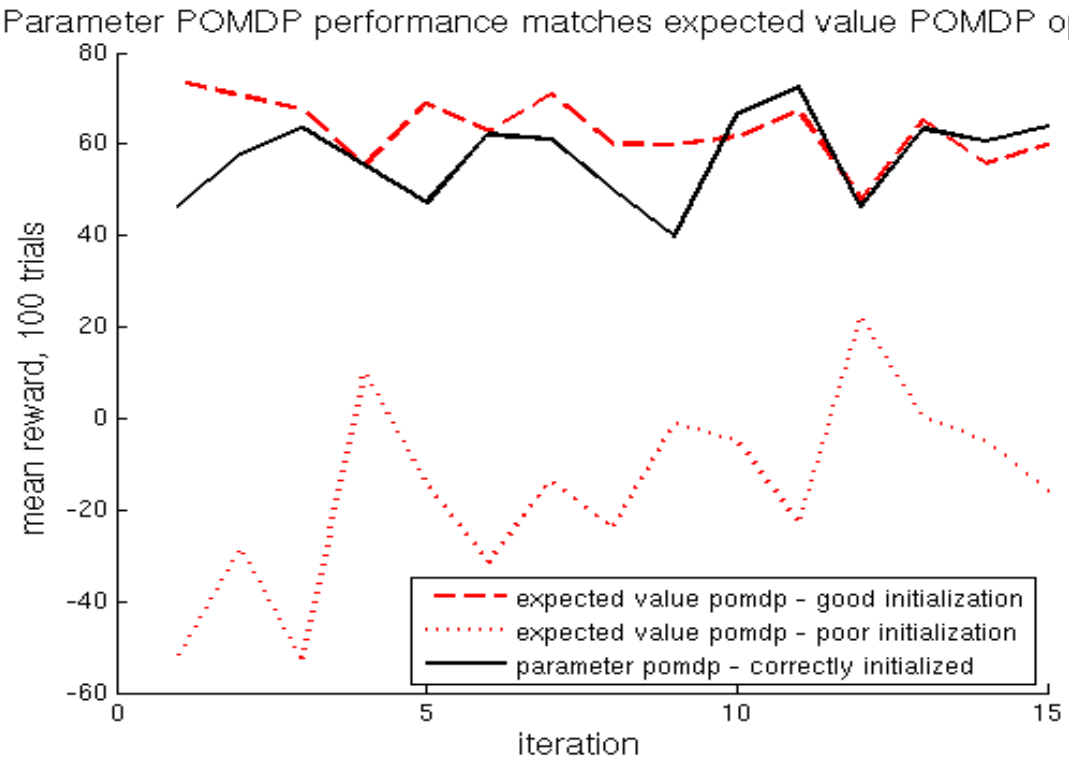
- POMDPs are an effective way to handle uncertainty in human-robot interactions.
- We can refine a rough POMDP dialog model through user interactions.
- With some simple heuristics, this learning process can be completed efficiently online.

Thank-you

Current Work: Larger Questions

How can we guarantee convergence of key parameters?

- Make POMDP aware of parameter uncertainty by incorporating parameters as additional hidden state.

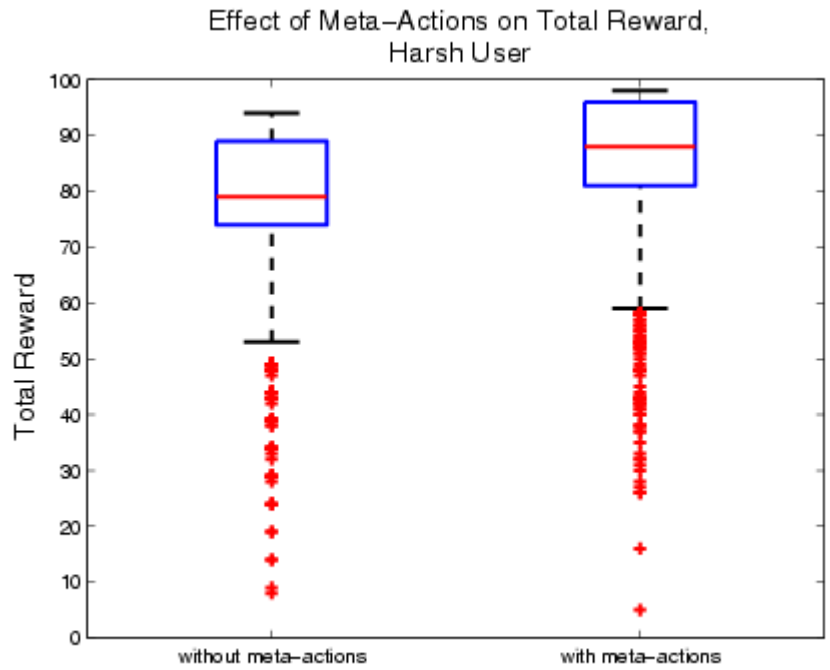


CSAIL

Current Work: Larger Questions

Can the robot take exploratory actions *and* act robustly?

- "Meta-Actions," queries about what the robot should do, are a low-risk way to discover user preferences.



CSAIL

Simulation Results