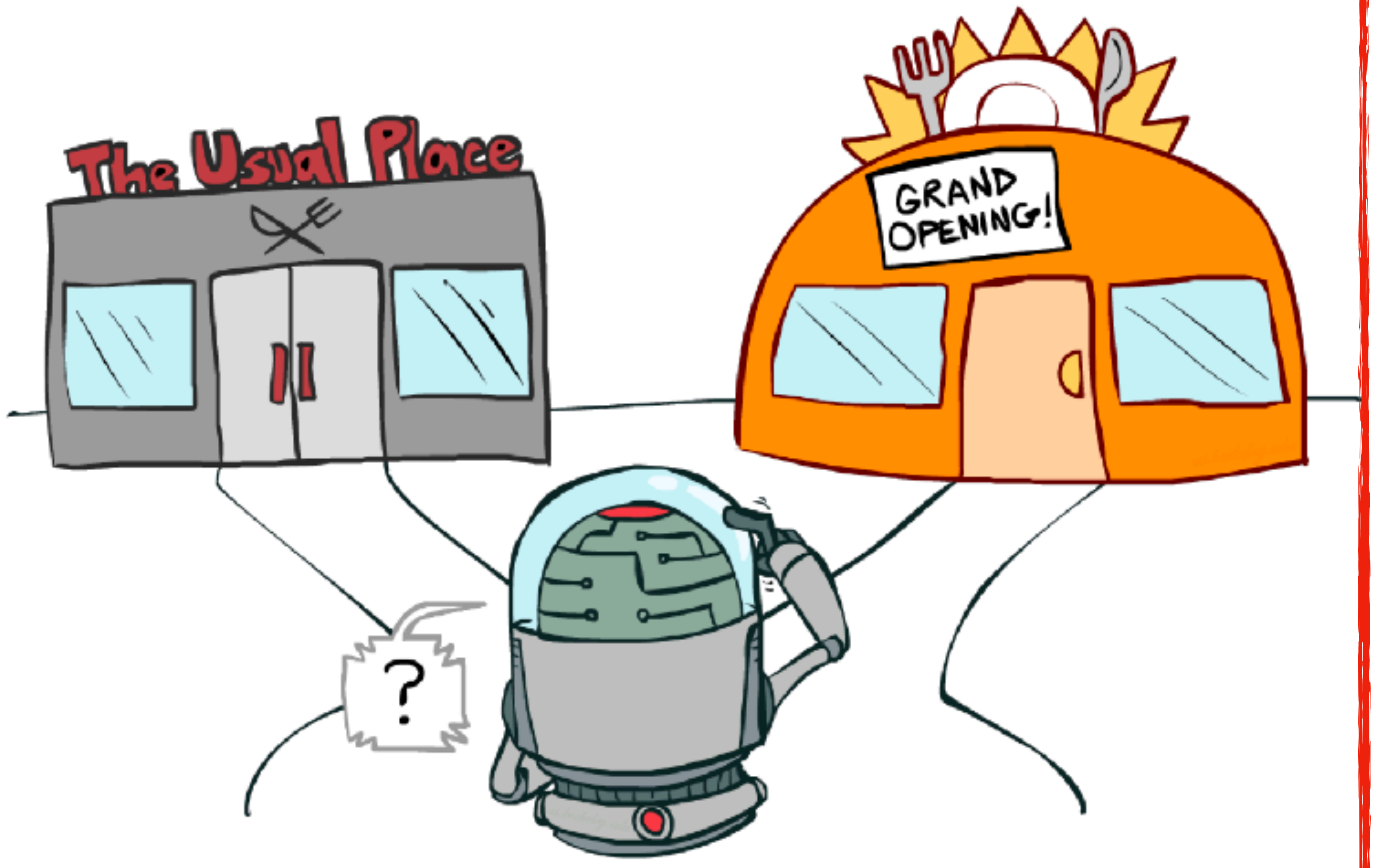
Dealing with Uncertainty

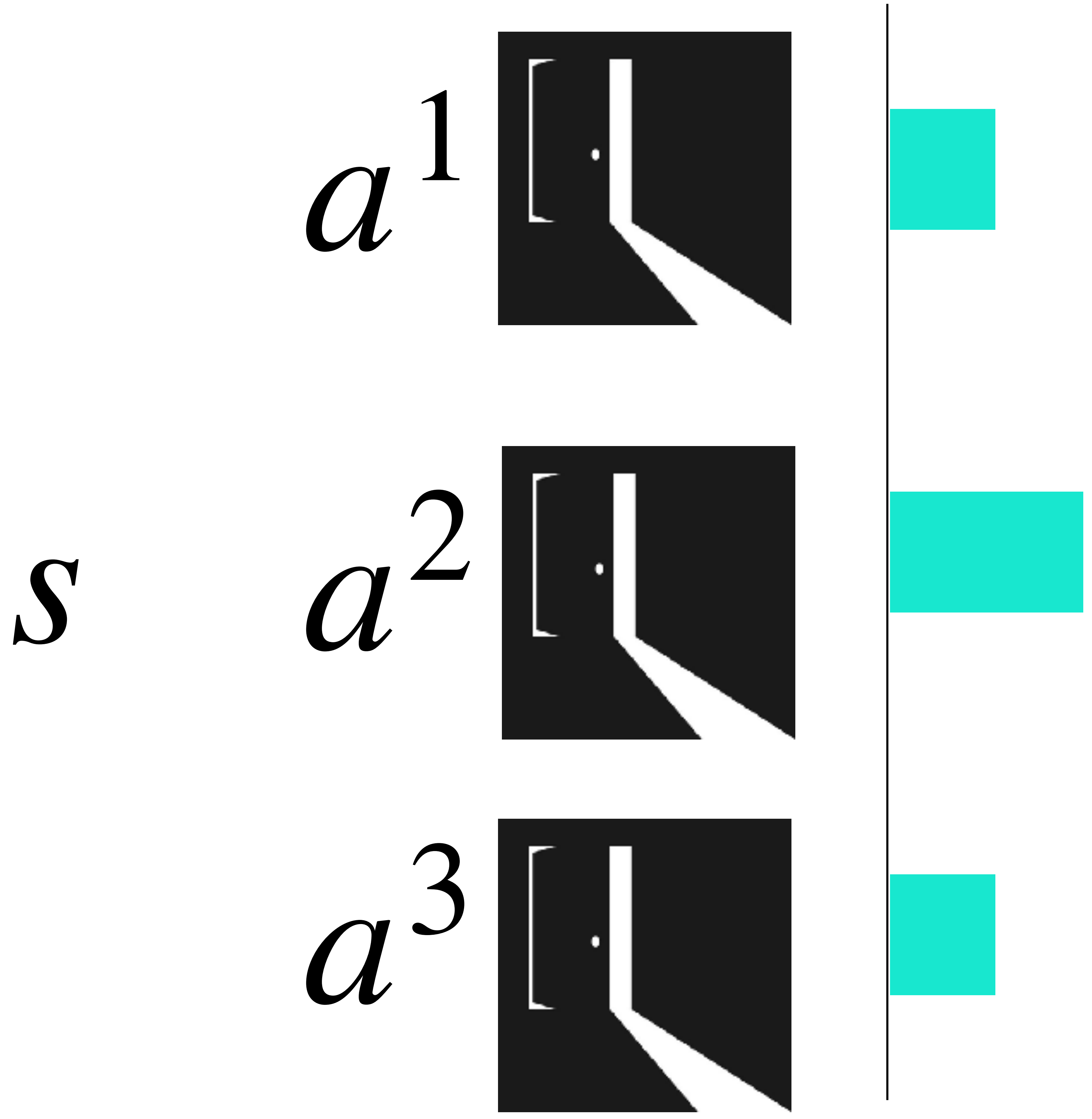
Sanjiban Choudhury

Cornell Bowers CIS
Computer Science

Two Ingredients of RL



Exploration Exploitation



Estimate Values $Q(s,a)$

Uncertainty

Types of uncertainty

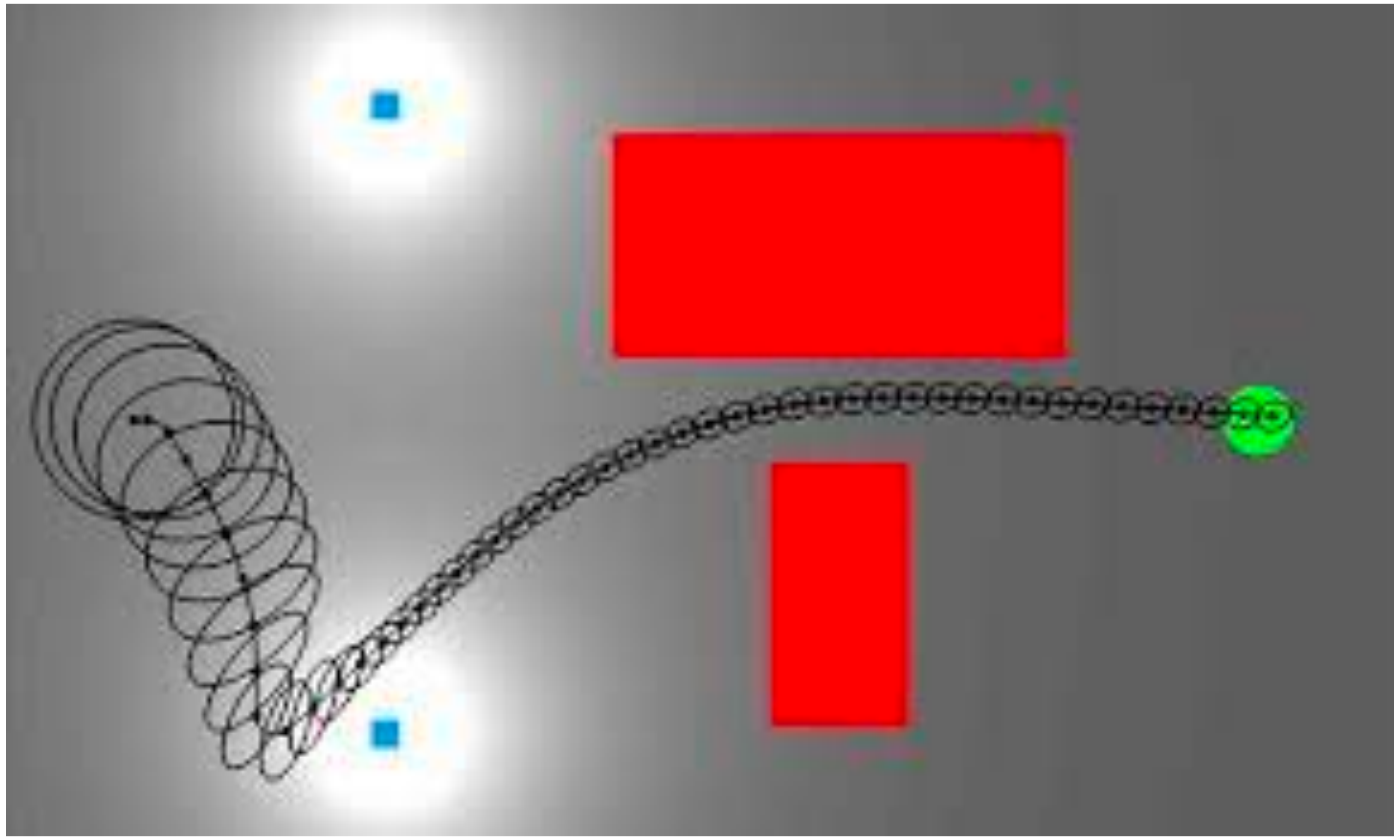
Aleatoric uncertainty

(Inherent randomness that cannot be explained away)

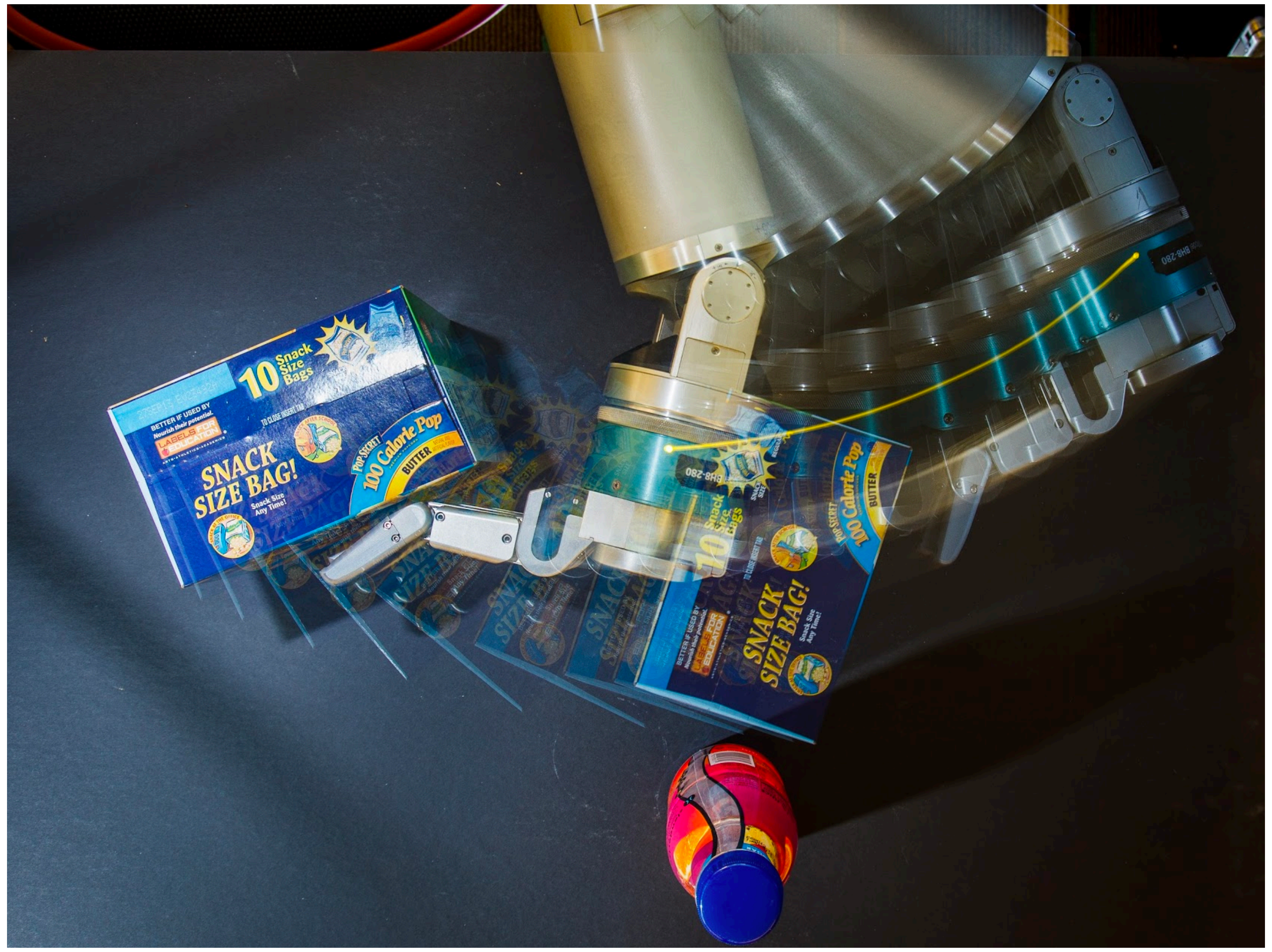
Epistemic uncertainty

(Acquire knowledge!)

Epistemic Uncertainty



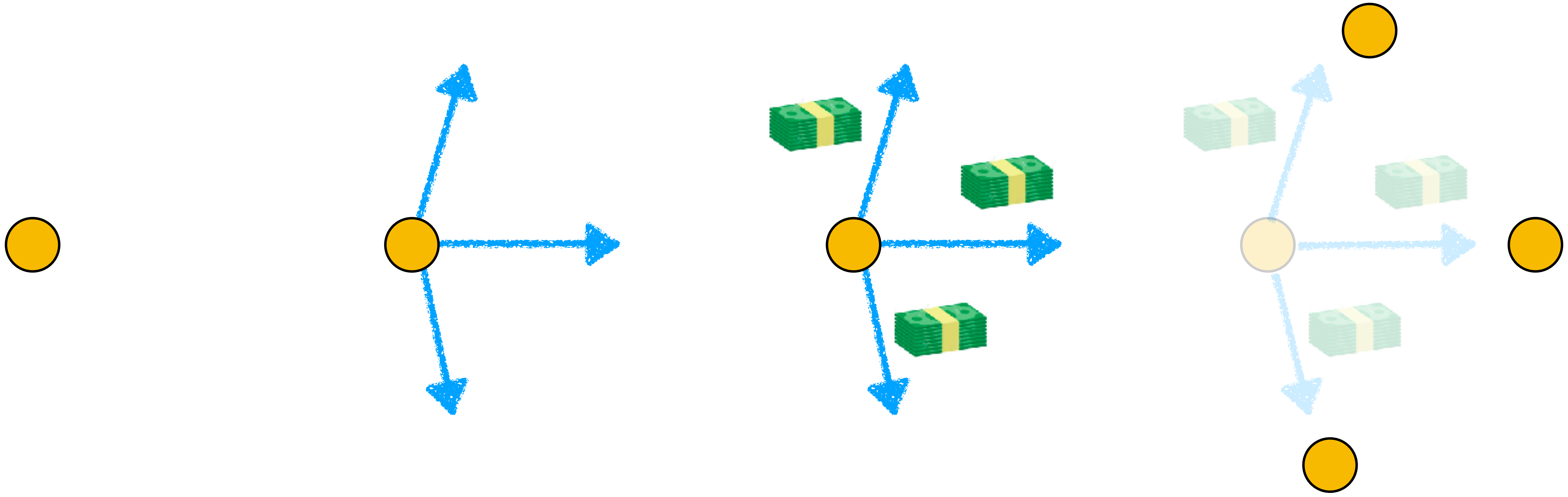
Uncertain about state



Uncertain about transitions

Can be uncertain about any of these things!

$$< S , A , C , \mathcal{T} >$$



What do we want to do about uncertainty?

Pure Exploration

Collapse uncertainty as quickly as possible

20 questions

Optimally explore / exploit

Take information gathering steps, but be robust along the way

Life!

Pure Exploitation

Be robust against uncertainty

UAV flying in wind

Activity!

Rank the following robotics applications based on pure exploration (highest) to pure exploitation (lowest)

When poll is active respond at **PollEv.com/sc2582**

Self-driving through an intersection

Human-robot shared autonomy

UAV autonomously mapping a building

Grasping an occluded object on the top-shelf

Fast off-road driving over terrain

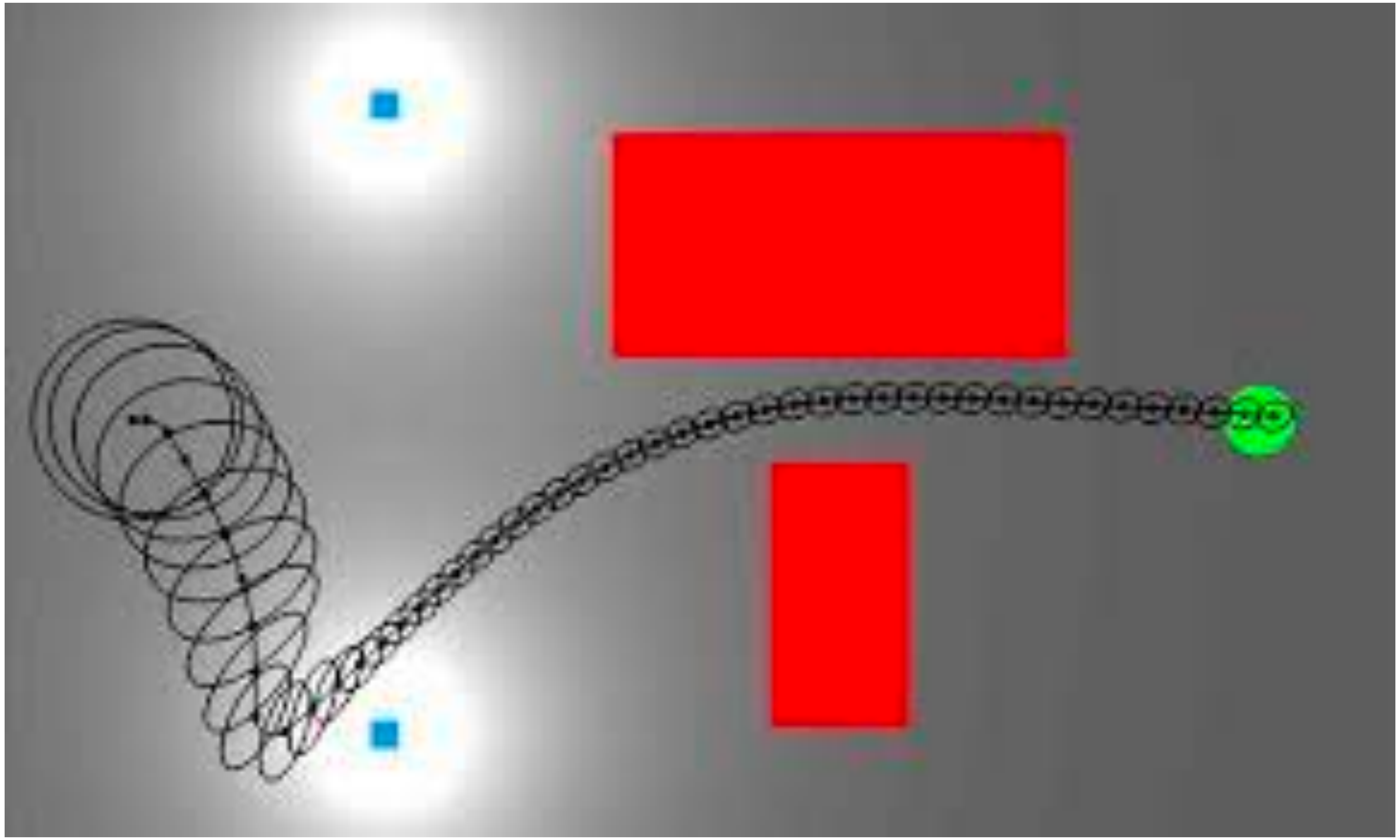
But what is the *optimal* exploration-exploitation algorithm?

Bayes Optimality:

The Holy Grail

POMDPs: The Siren's Call

Let's work through an example: Uncertain about the robot pose

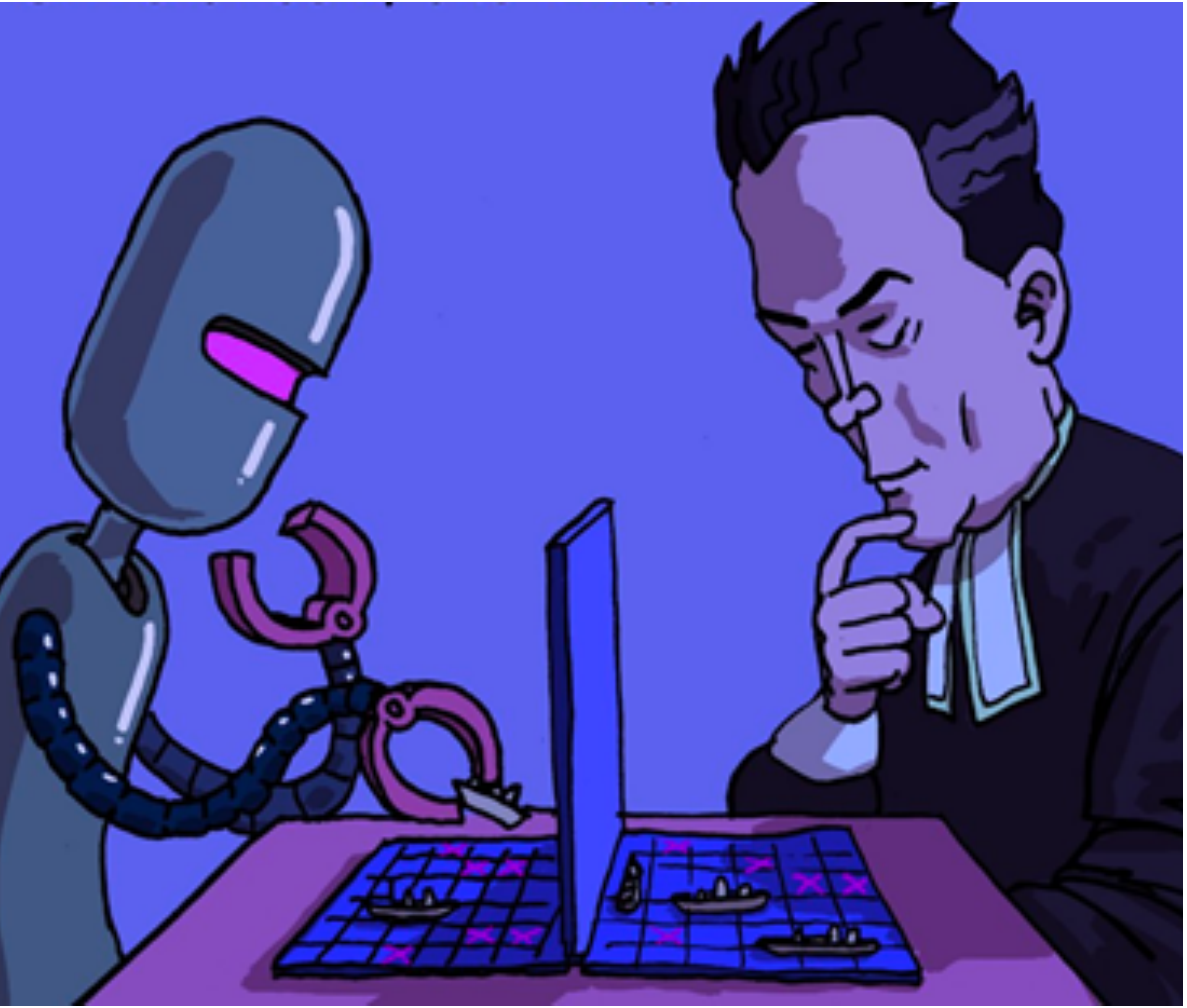


GAME
OVER!

Belief Space Planning is NP-Hard at best, undecidable at worst

Need to relax our problem!

What if we wanted to explore as optimally as possible using prior information?



Information Gain

20 Questions

Let's say you have a set of hypotheses

$$\{\theta_1, \theta_2, \ldots, \theta_n\}$$

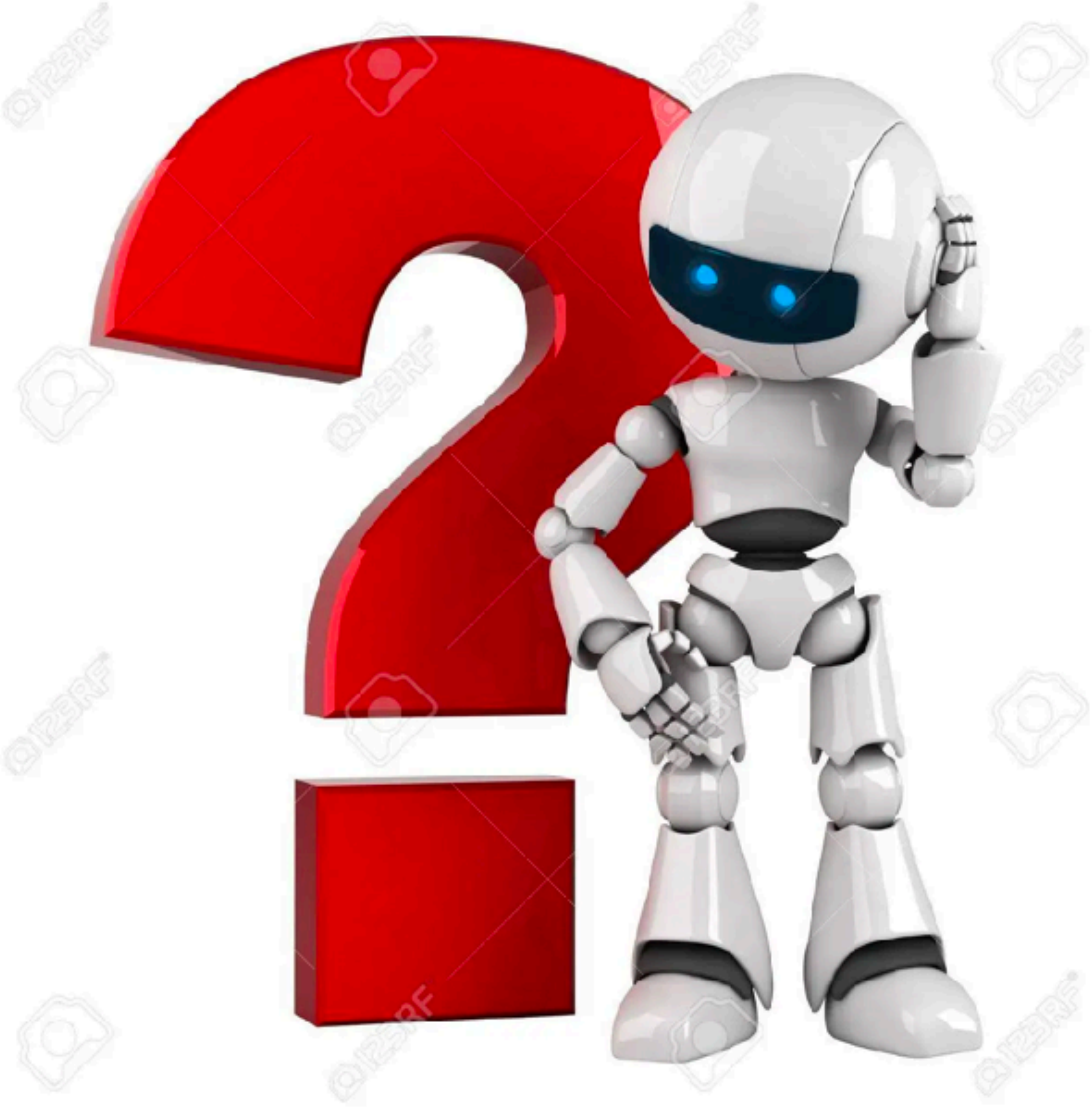
and a set of tests

$$\{t_1, t_2, \ldots, t_n\}$$

Given a prior over hypotheses $P(\theta)$

Find the minimal number of tests to identify hypothesis

20 Questions



Let’s say you have a set of hypotheses

$$\{\theta_1, \theta_2, \ldots, \theta_n\}$$

and a set of tests

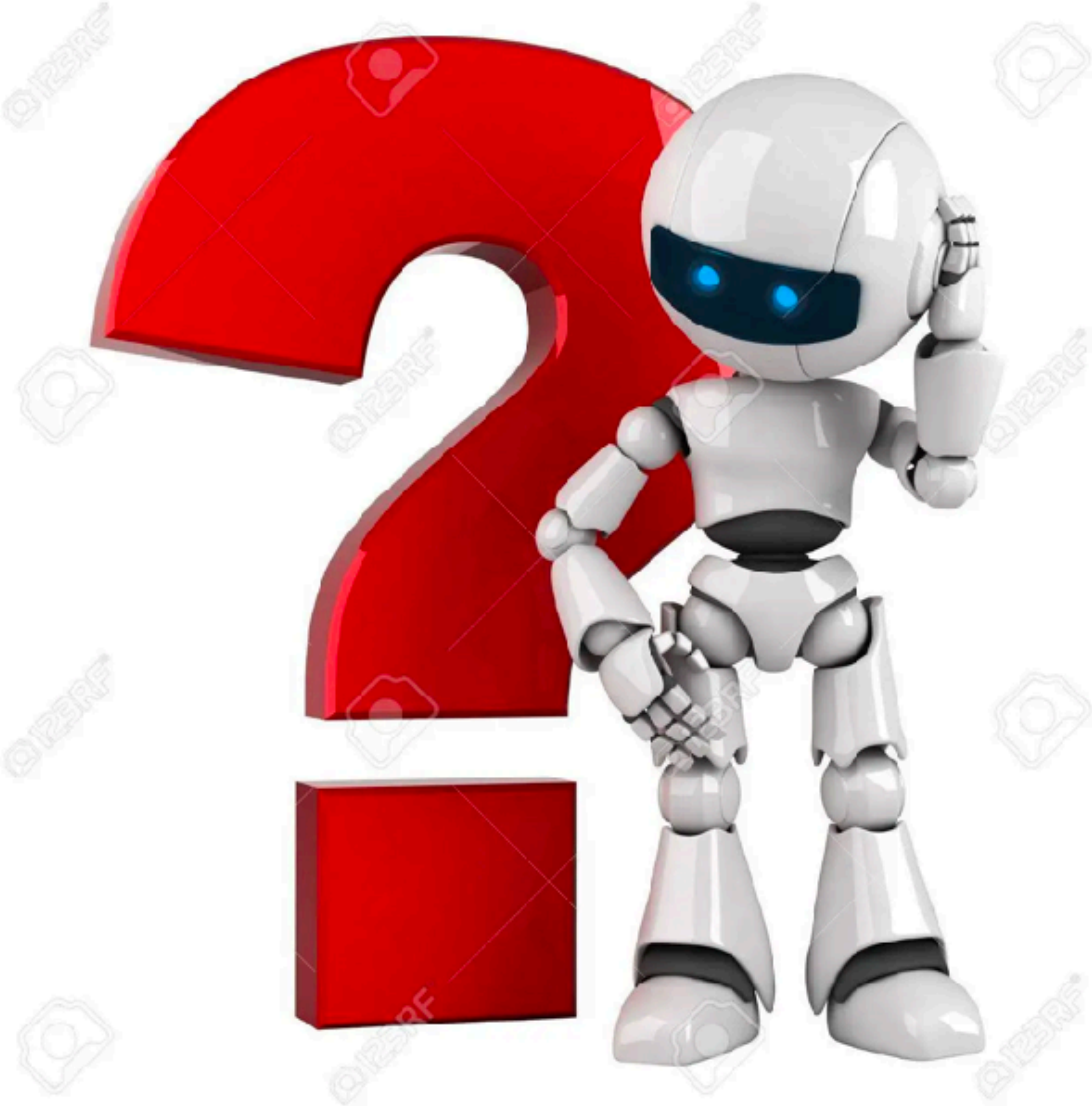
$$\{t_1, t_2, \ldots, t_n\}$$

NP-HARD

Given a prior over hypotheses $P(\theta)$

Find the minimal number of tests to identify hypothesis

A simple algorithm



Greedily pick the test that maximizes information gain

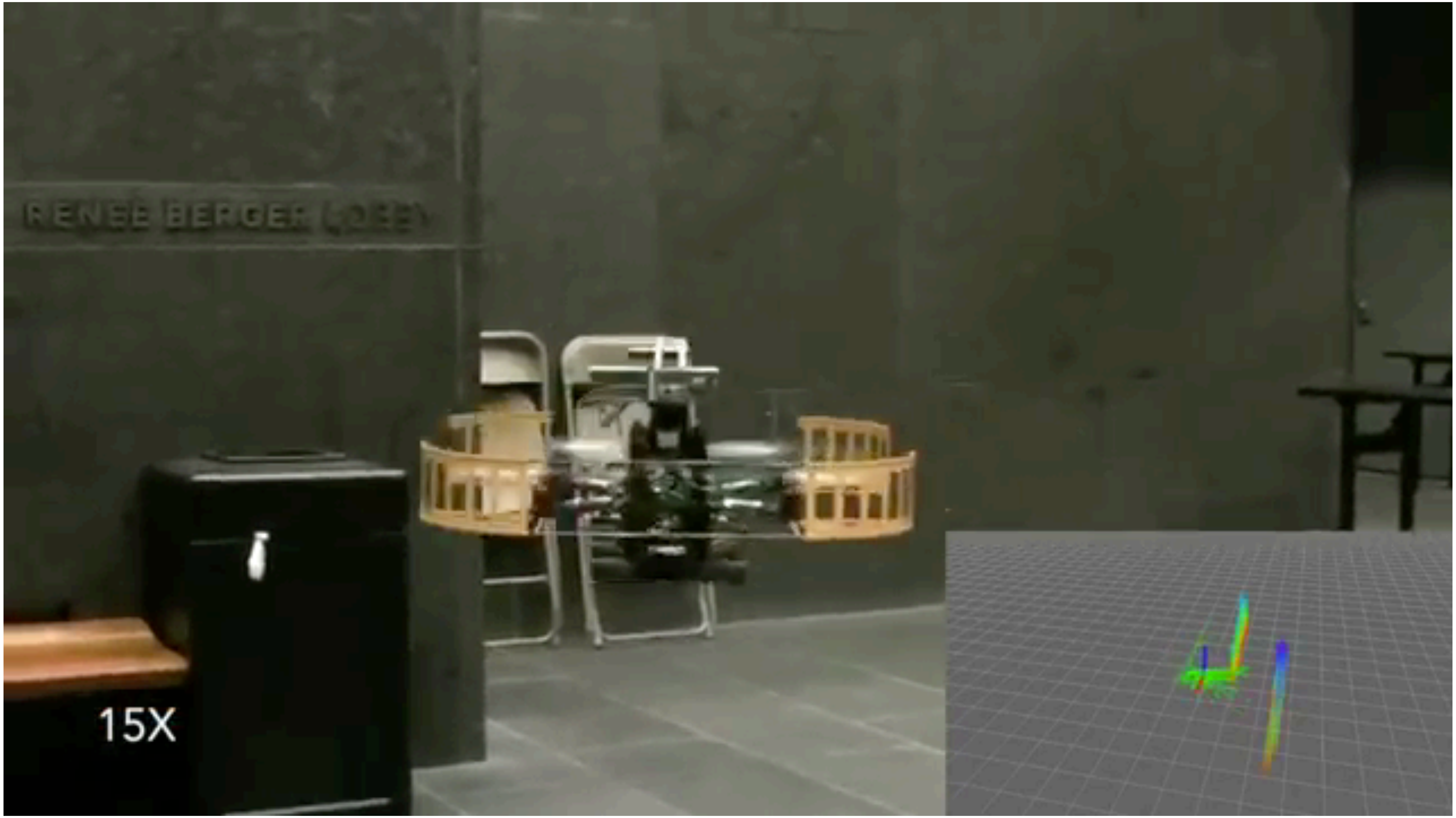
$$\max_t H(\theta) - \mathbb{E}_o H(\theta \mid t, o)$$

Entropy Posterior entropy

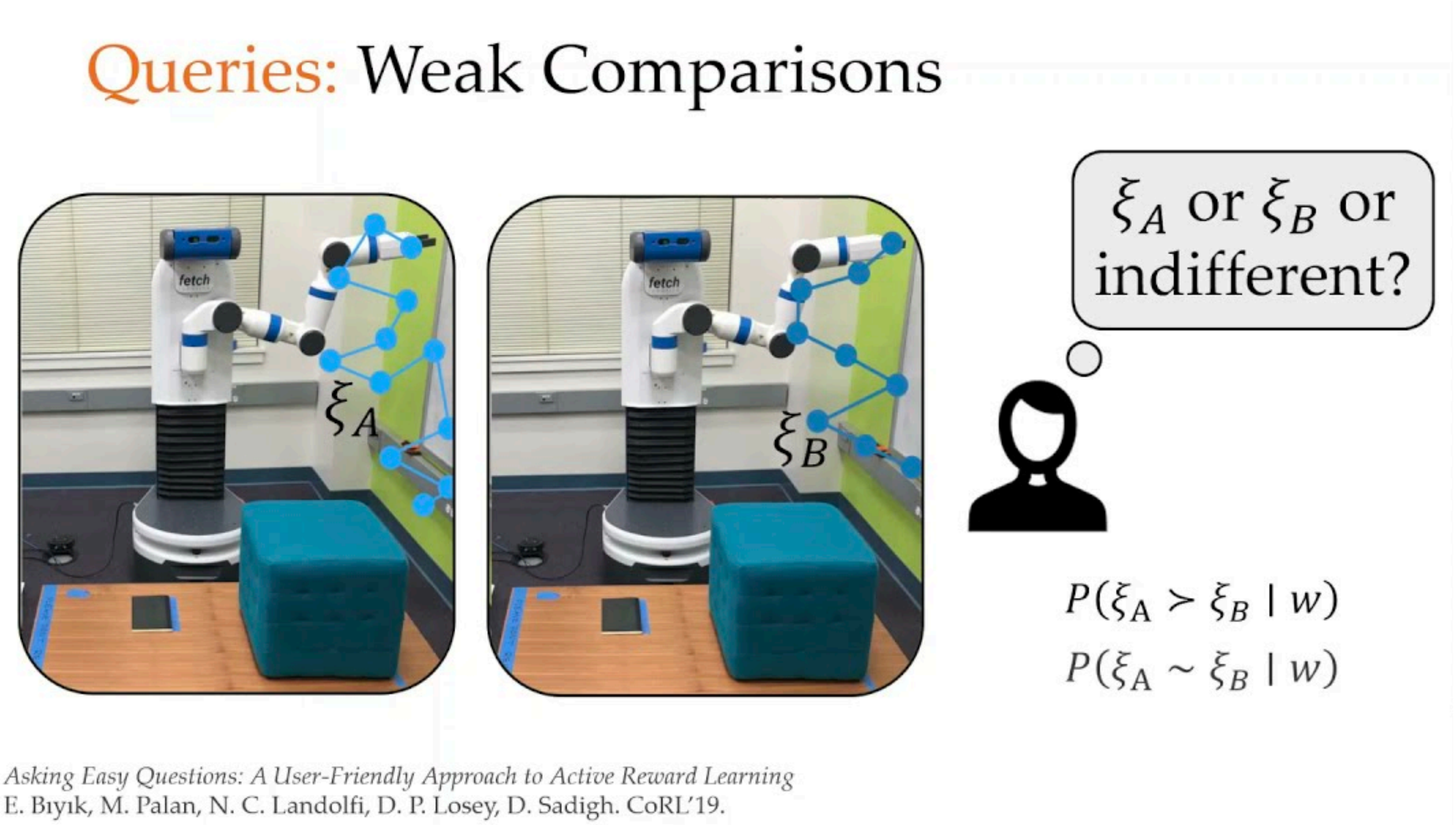
Entropy is adaptive sub modular => Greedy is near-optimal

Applications

Autonomous mapping



Active Preference Learning



Queries: Weak Comparisons

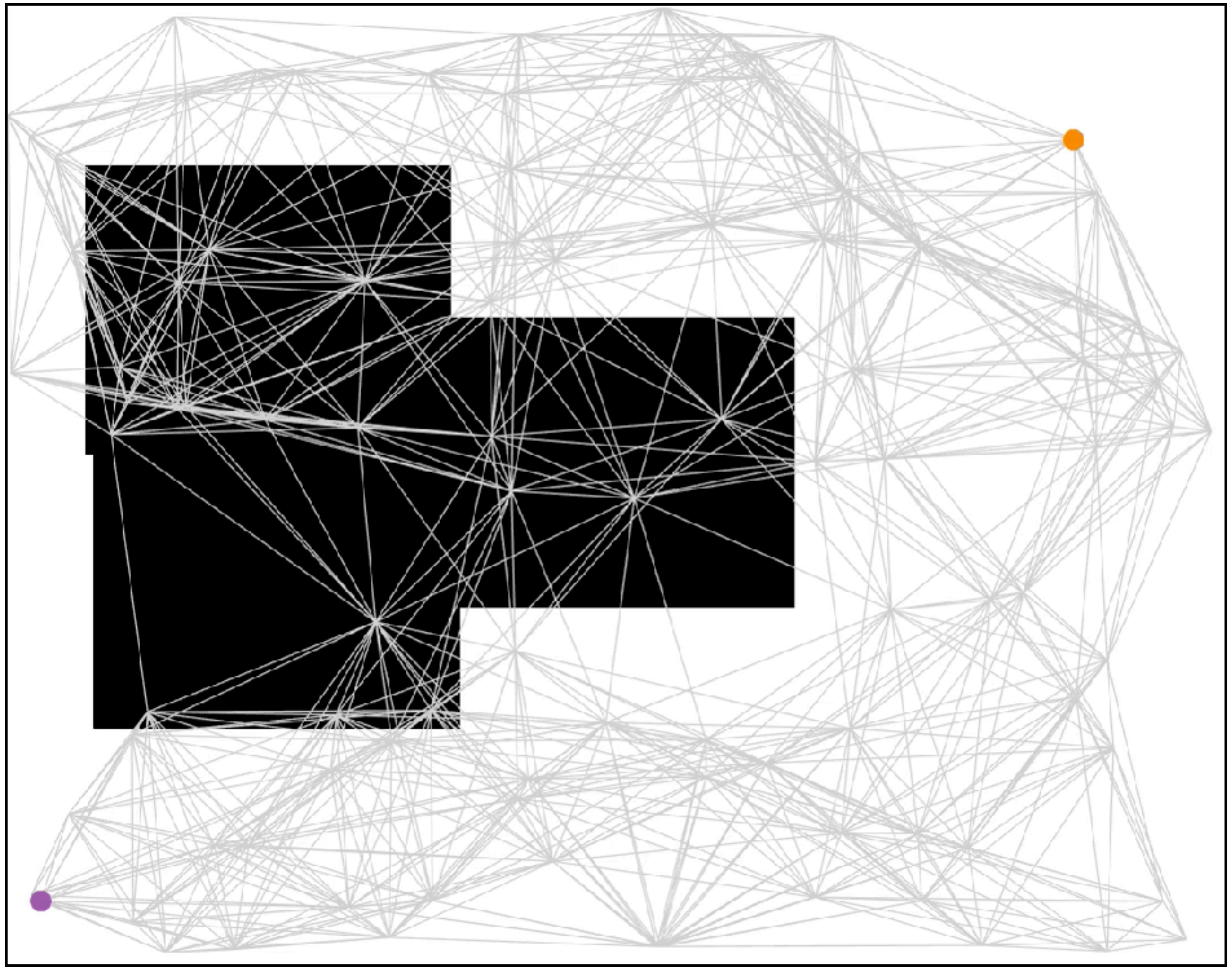
$P(\xi_\mathrm{A} \succ \xi_B \mid w)$

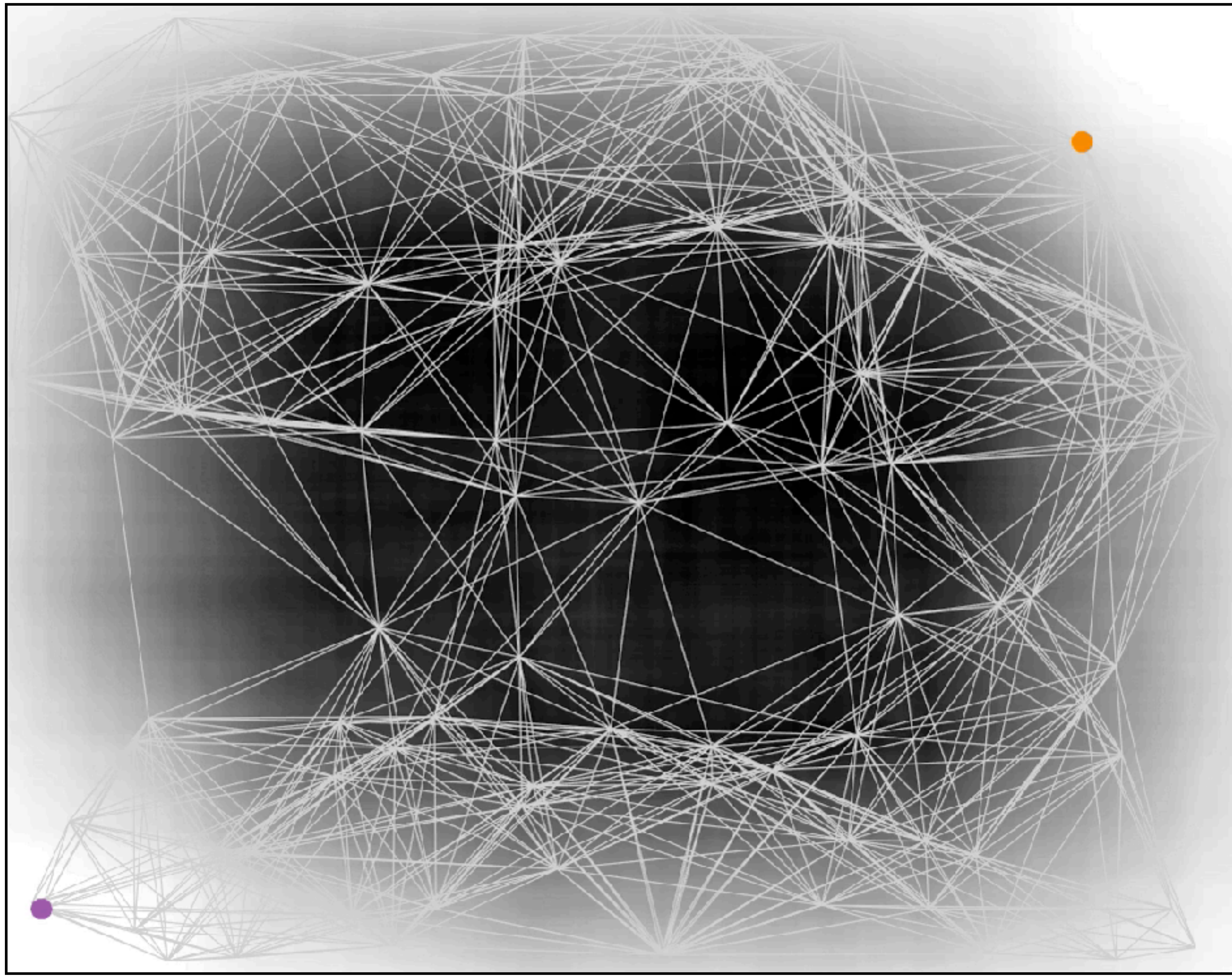
$P(\xi_\mathrm{A} \sim \xi_B \mid w)$

Asking Easy Questions: A User-Friendly Approach to Active Reward Learning
E. Bıyık, M. Palan, N. C. Landolfi, D. P. Losey, D. Sadigh. CoRL'19.

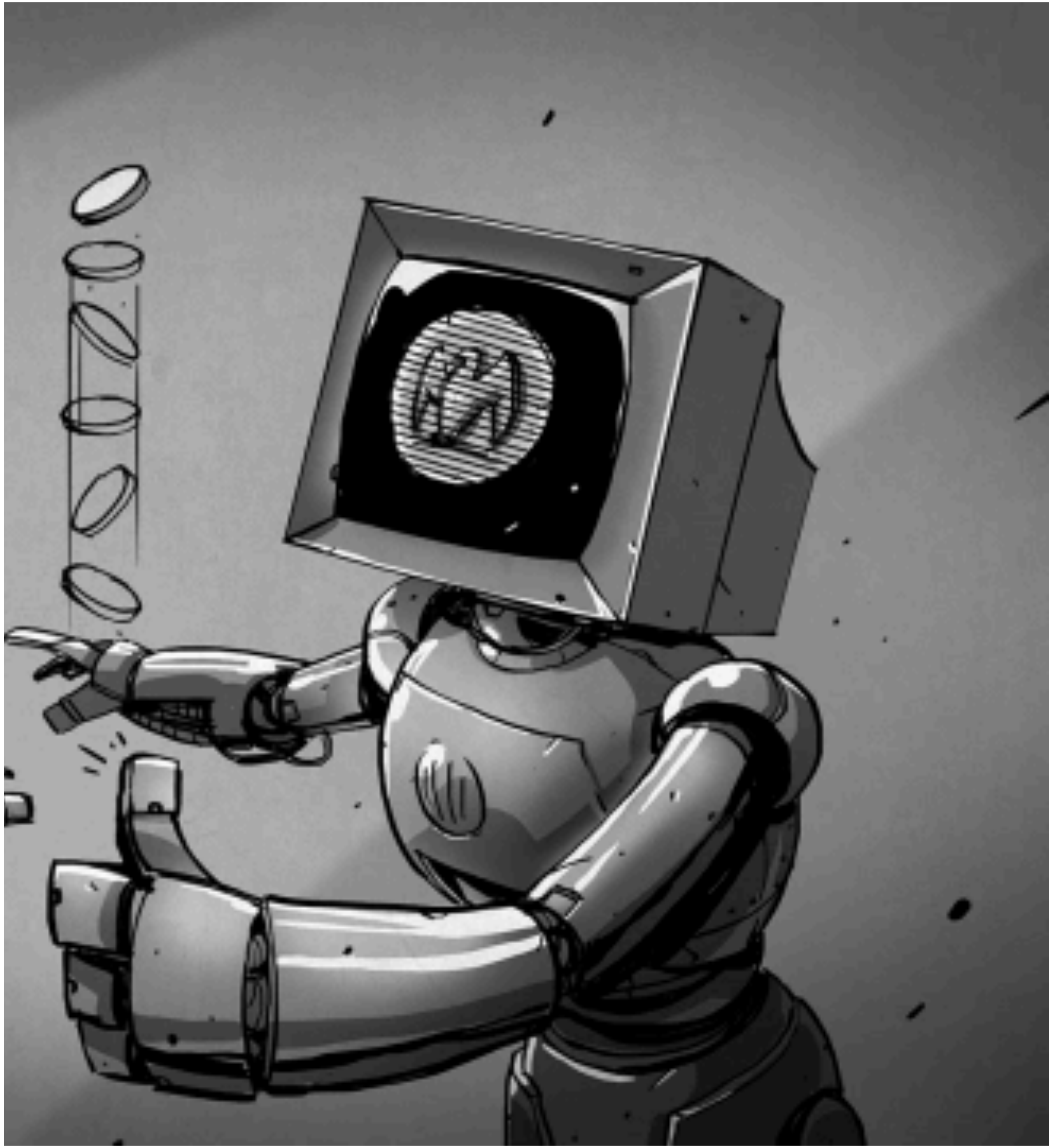
Optimal edge evaluation for shortest path

[CJS+ NeurIPS'17] [CSS IJCAI'18]





Can we find a better exploration / exploitation algorithm?



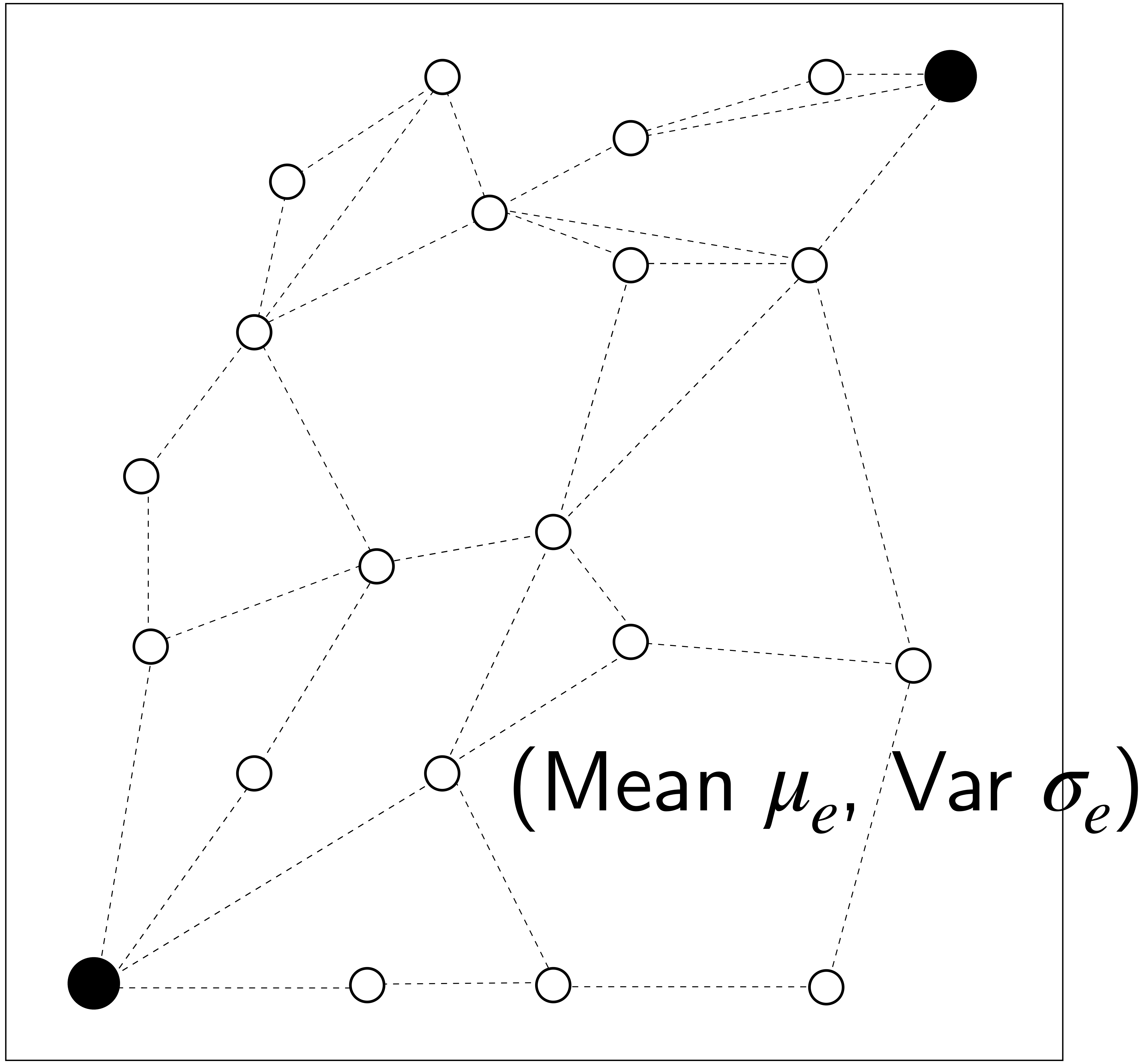
Posterior Sampling

The Online Shortest Path Problem

You just moved to Cornell and are traveling from office to home.

You would like to get home quickly but you are uncertain about travel times along each edge

Suppose we had a prior on travel time for each edge (Mean μ_e, Var σ_e)



What if …

… we just sampled travel times from our prior and solved the shortest path?

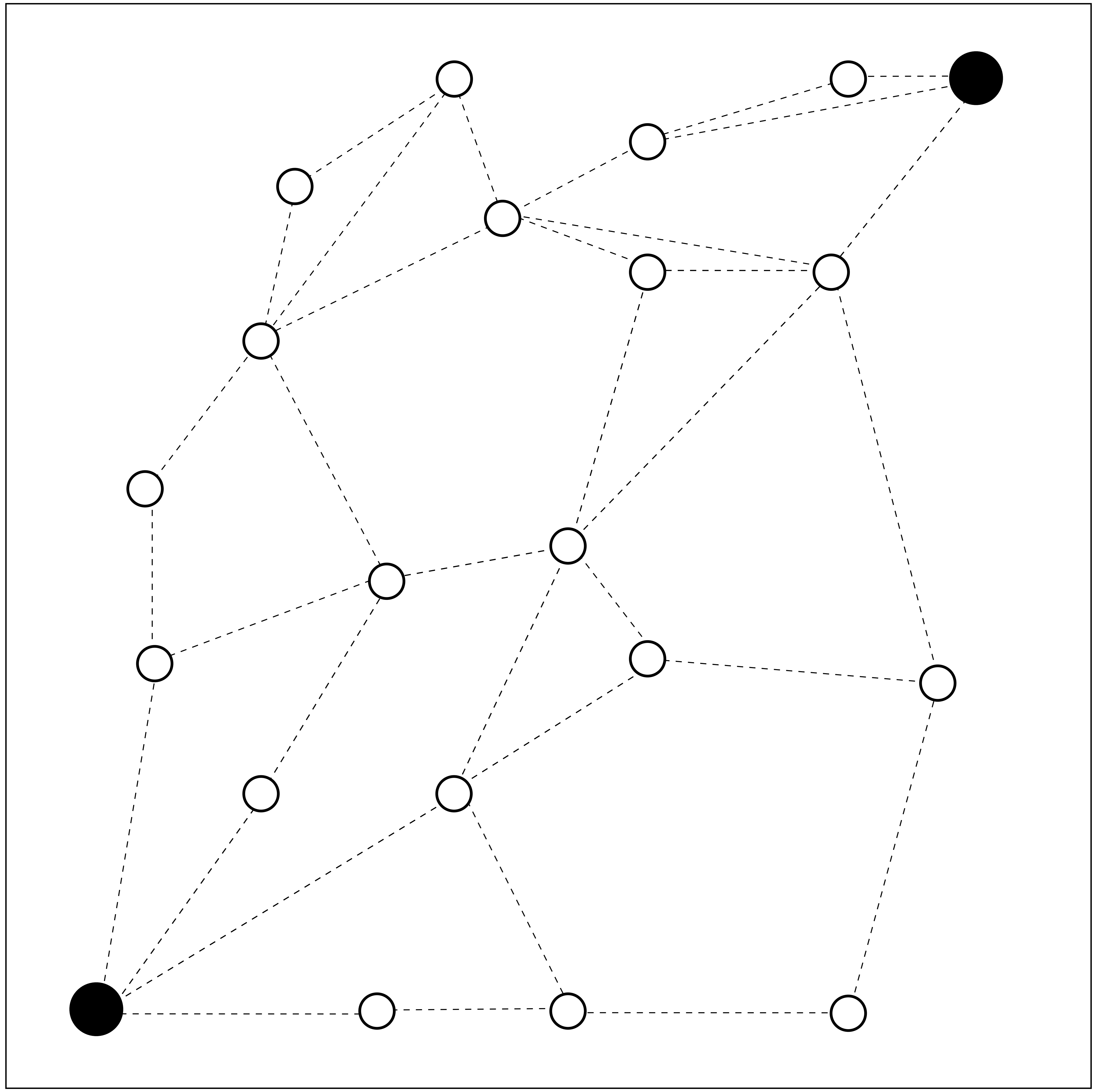
A suspicious simple algorithm

Repeat forever:

Sample edge times from posterior

Compute shortest path

Travel along path, and update posterior



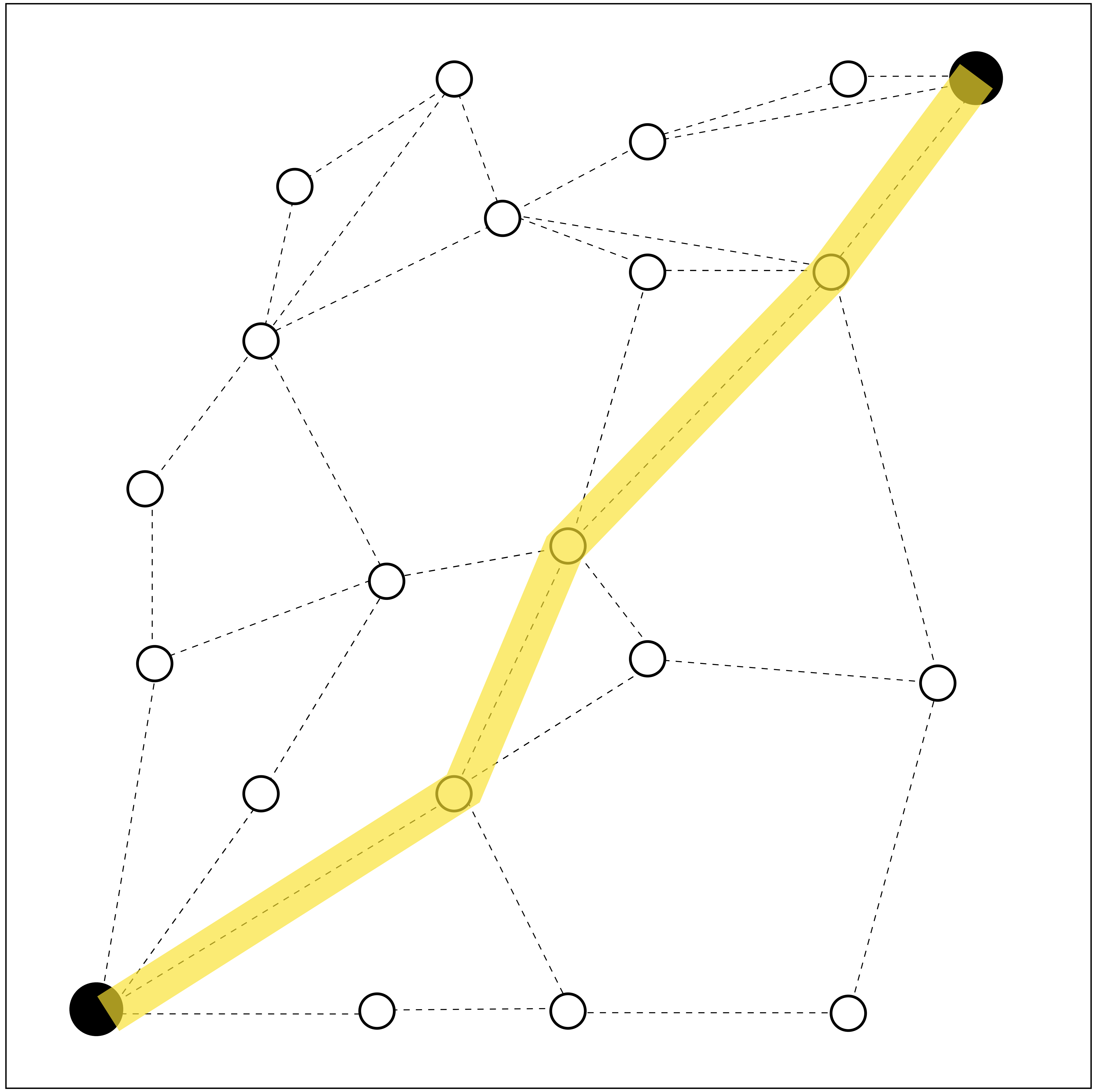
A suspiciously simple algorithm

Repeat forever:

Sample edge times from posterior

Compute shortest path

Travel along path, and update posterior



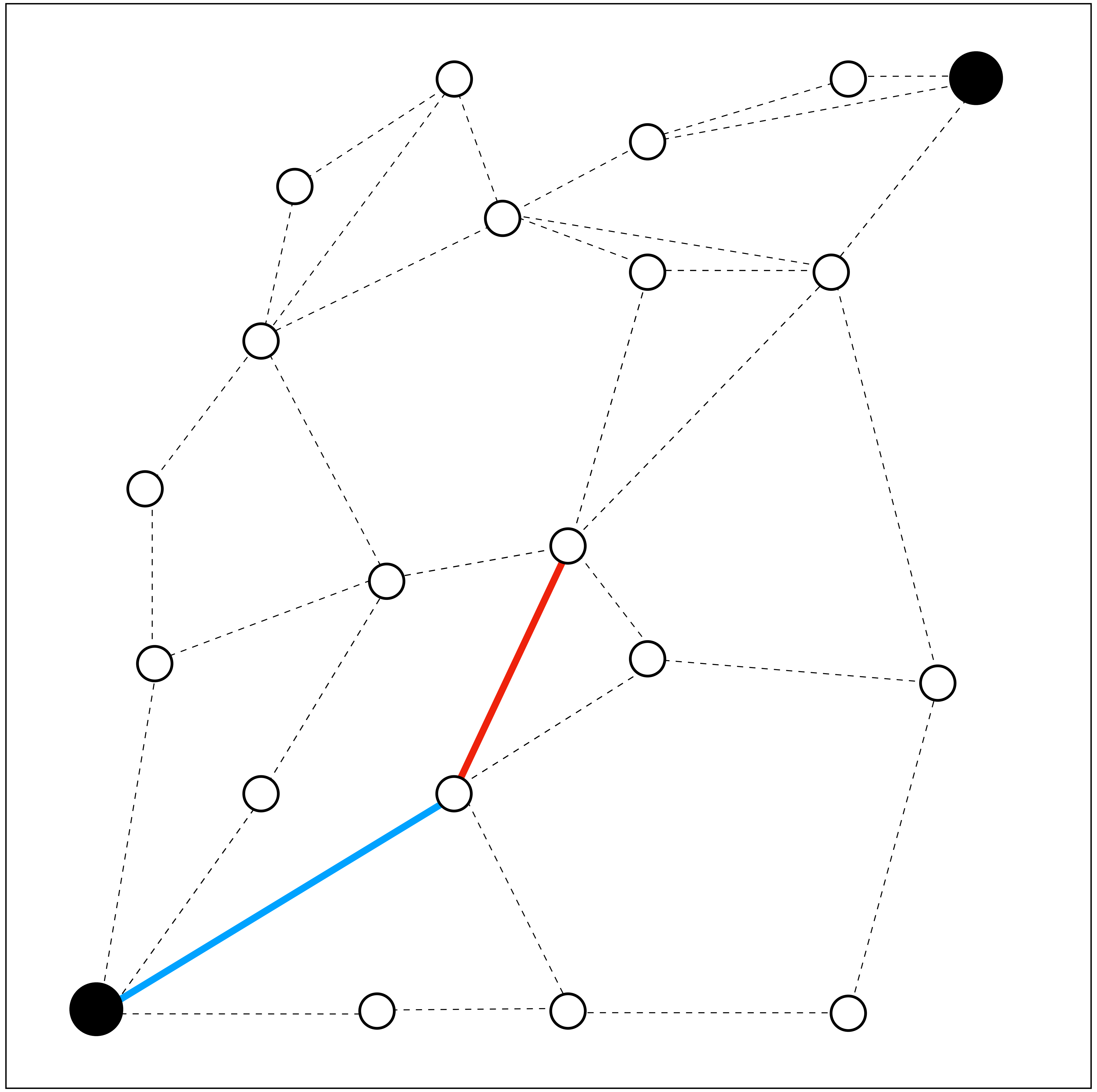
A suspiciously simple algorithm

Repeat forever:

Sample edge times from posterior

Compute shortest path

Travel along path, and update posterior



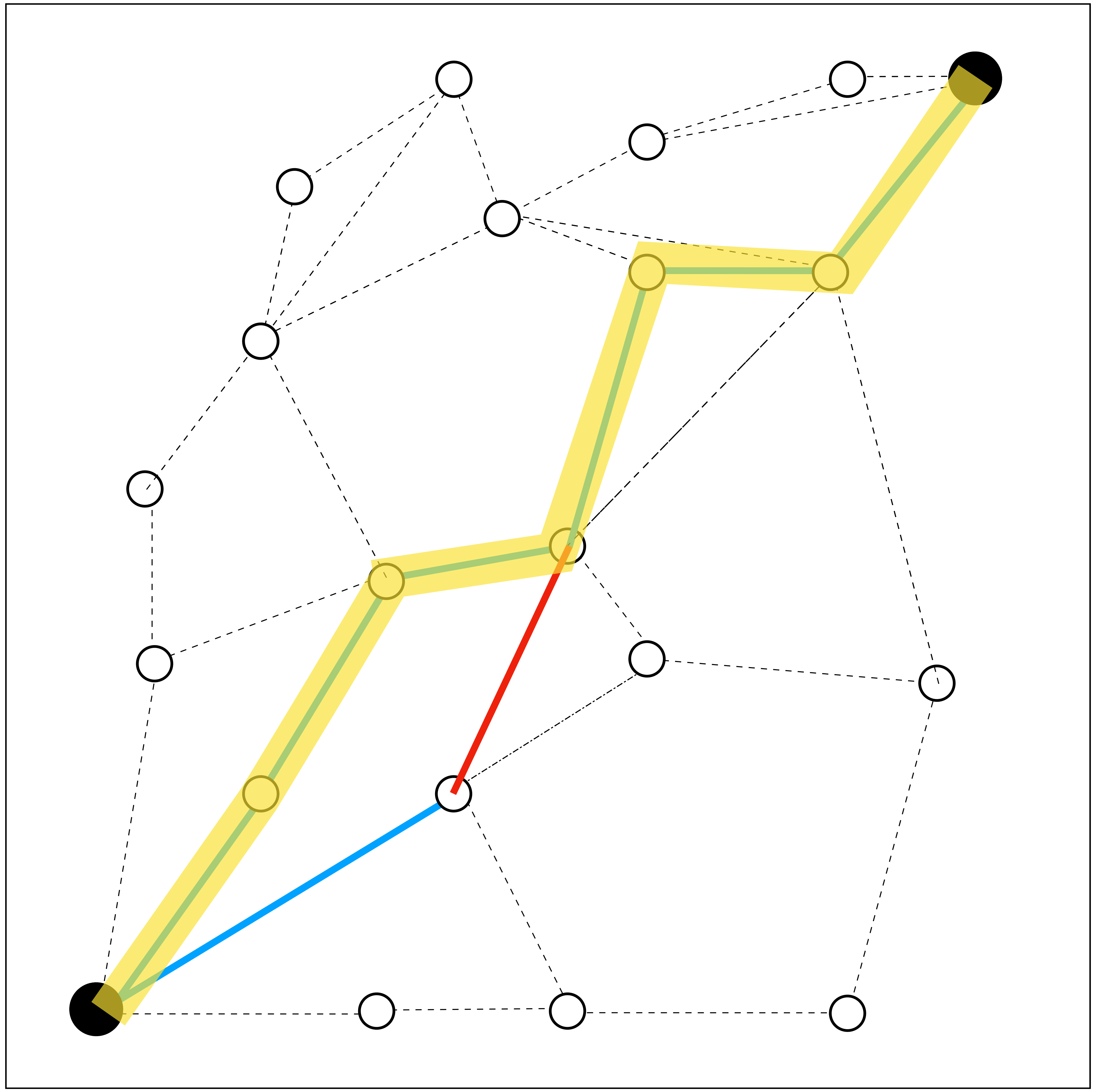
A suspiciously simple algorithm

Repeat forever:

Sample edge times from posterior

Compute shortest path

Travel along path, and update posterior



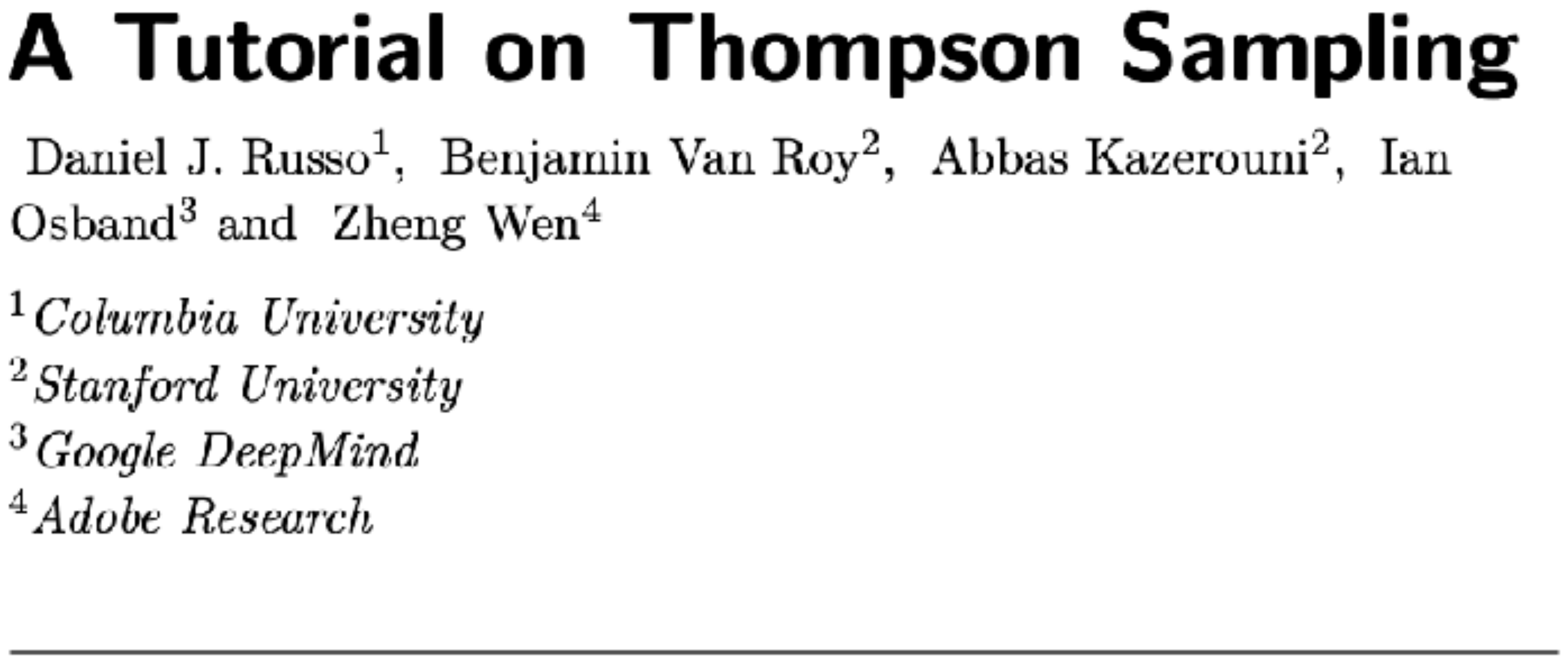
Can we lift this idea to general MDP

Repeat forever:

```
    Sample model from posterior

    Compute optimal policy

    Execute policy, observe s,a,s',
    Update model
```

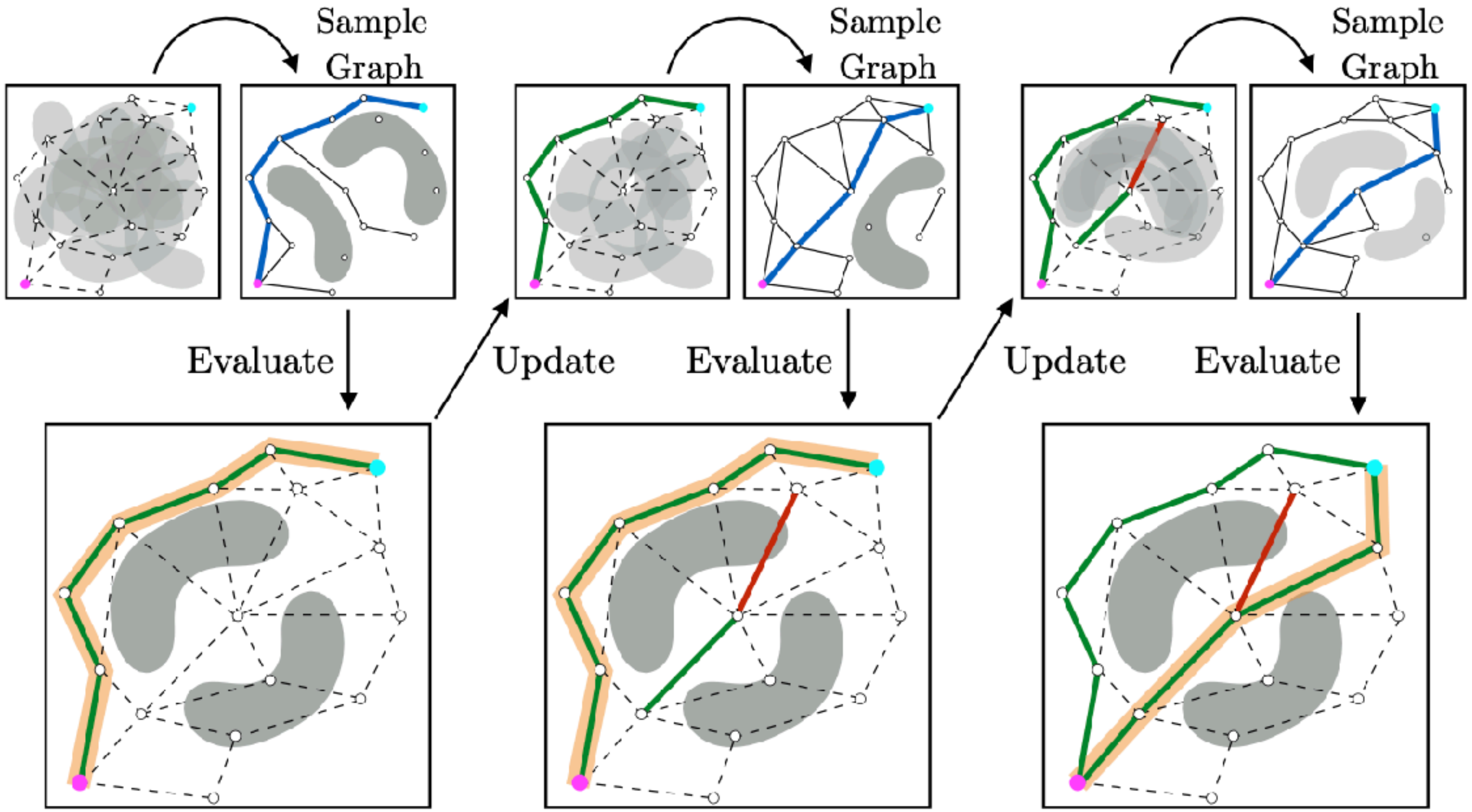


A Tutorial on Thompson Sampling

Daniel J. Russo[1], Benjamin Van Roy[2], Abbas Kazerouni[2], Ian Osband[3] and Zheng Wen[4]

[1] *Columbia University*
[2] *Stanford University*
[3] *Google DeepMind*
[4] *Adobe Research*

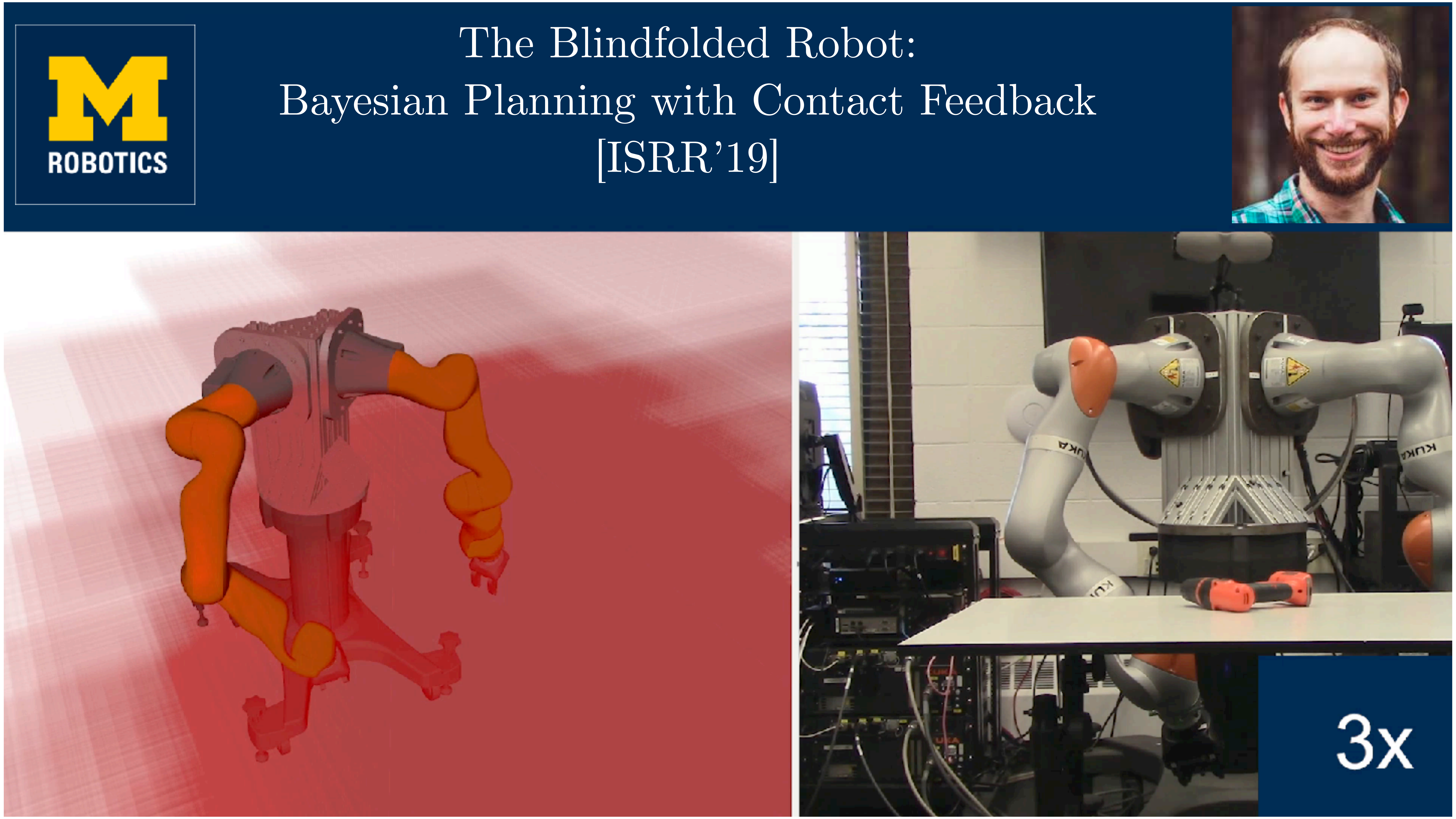
Posterior Sampling for Motion Planning



Posterior Sampling for Anytime Motion Planning on Graphs with Expensive-to-Evaluate Edges

Brian Hou, Sanjiban Choudhury, Gilwoo Lee, Aditya Mandalika, and Siddhartha S. Srinivasa

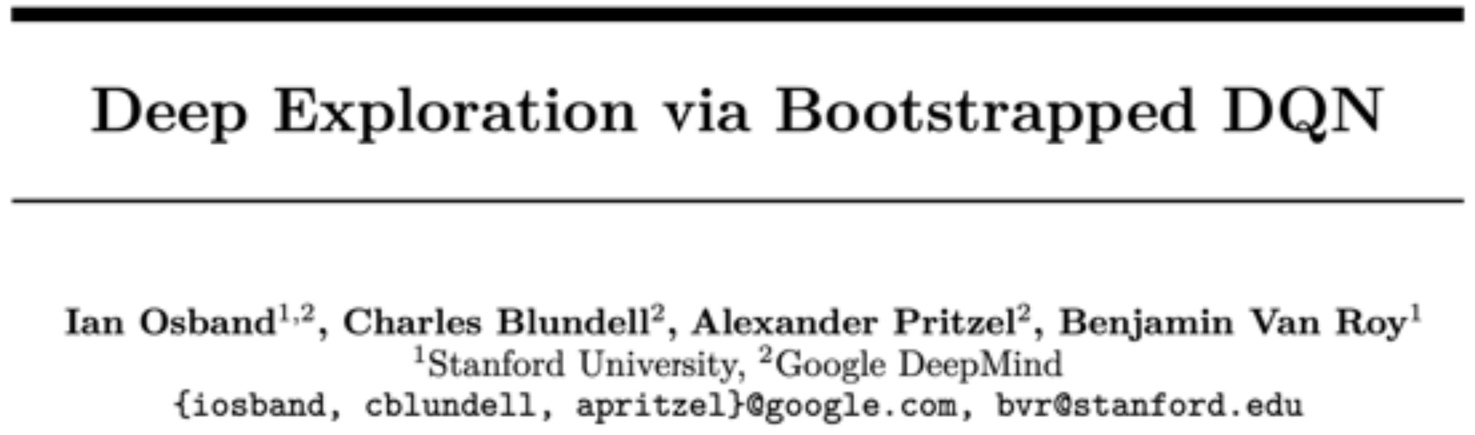
Real Robot Problems!



The Blindfolded Robot:
Bayesian Planning with Contact Feedback
[ISRR'19]

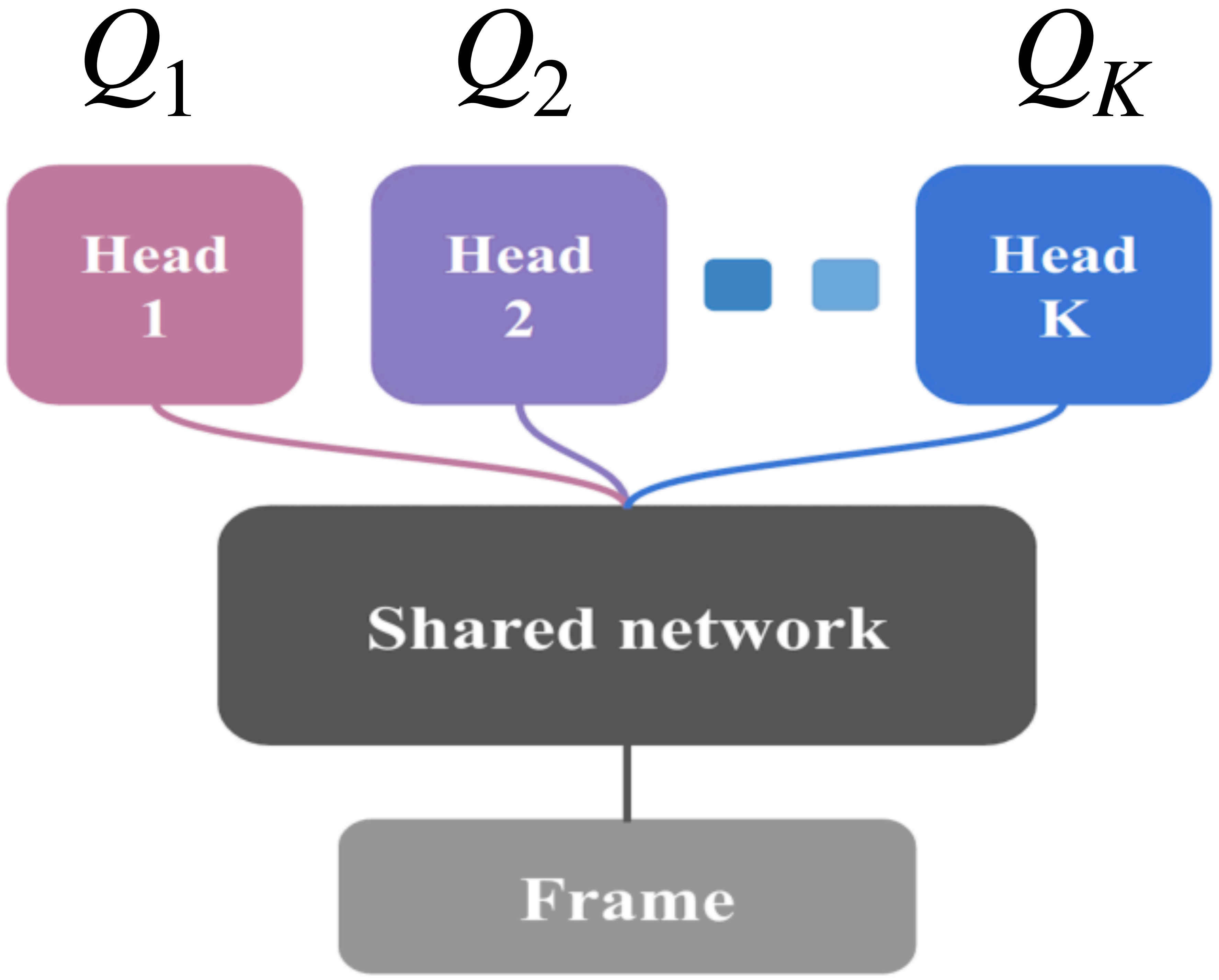
Posterior Sampling for Reinforcement Learning

1. sample Q-function Q from $p(Q)$
2. act according to Q for one episode
3. update $p(Q)$



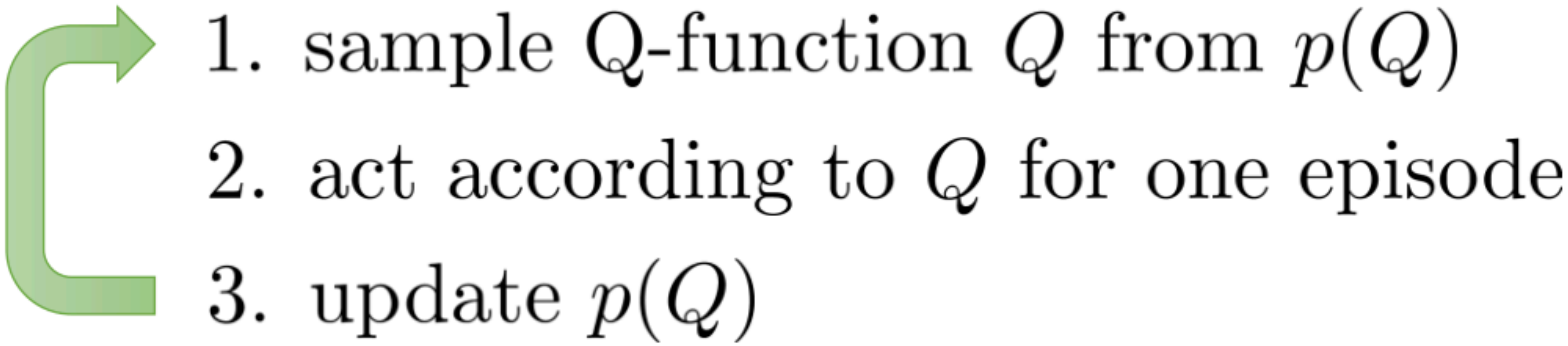
Deep Exploration via Bootstrapped DQN

Ian Osband[1,2], Charles Blundell[2], Alexander Pritzel[2], Benjamin Van Roy[1]
[1]Stanford University, [2]Google DeepMind
{iosband, cblundell, apritzel}@google.com, bvr@stanford.edu



Bootstrapped Q Network

Posterior Sampling for Reinforcement Learning



1. sample Q-function Q from $p(Q)$
2. act according to Q for one episode
3. update $p(Q)$

Atari

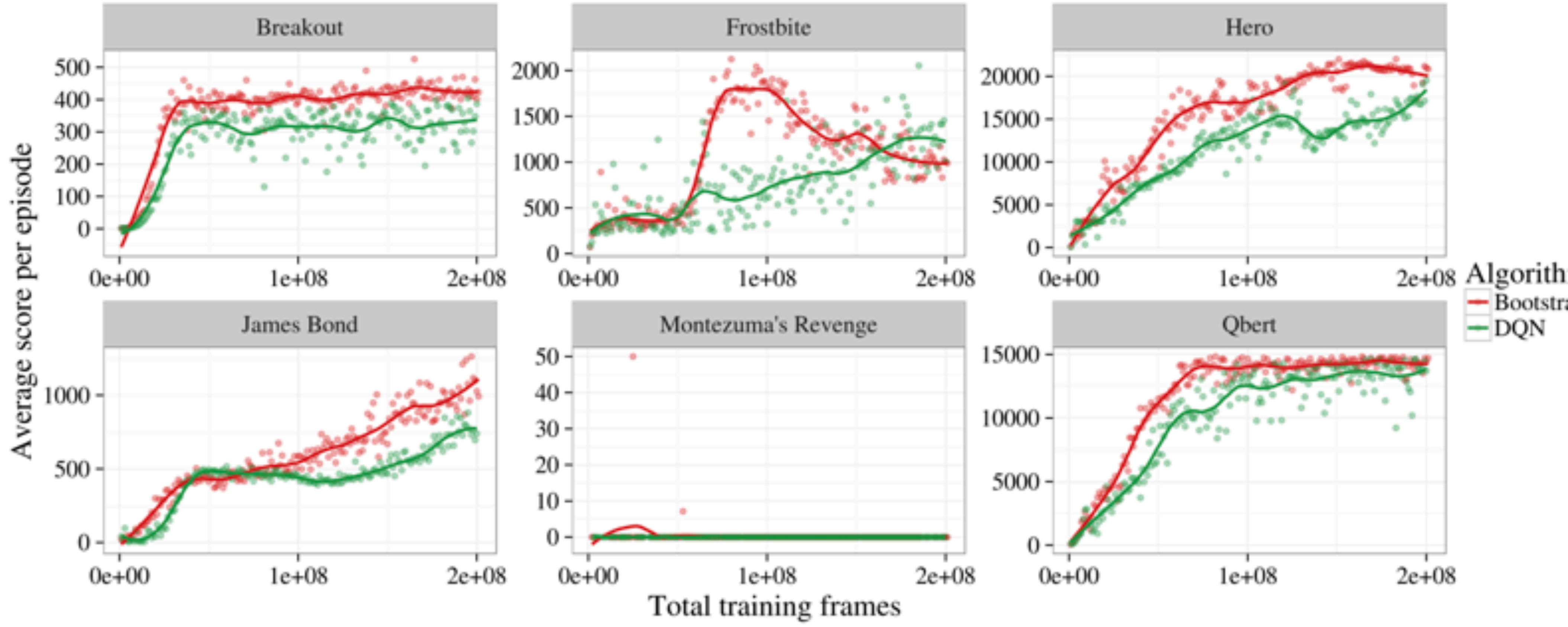


Figure 6: Bootstrapped DQN drives more efficient exploration.

Why does work better than taking random actions?