Partially Observable Markov Decision Processes

Sanjiban Choudhury

Cornell Bowers CIS
Computer Science

Uncertainty

Types of uncertainty

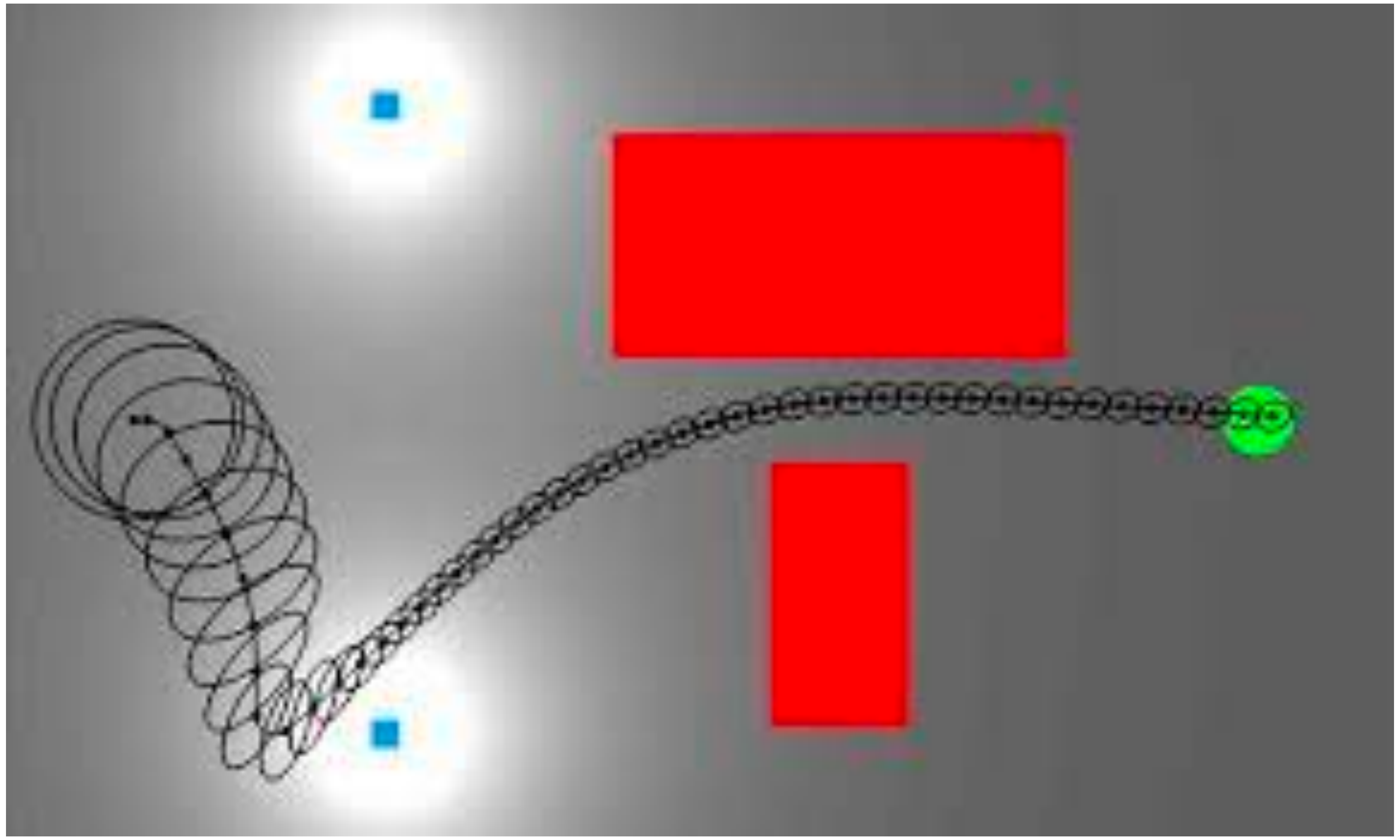
Aleatoric uncertainty

(Inherent randomness that cannot be explained away)

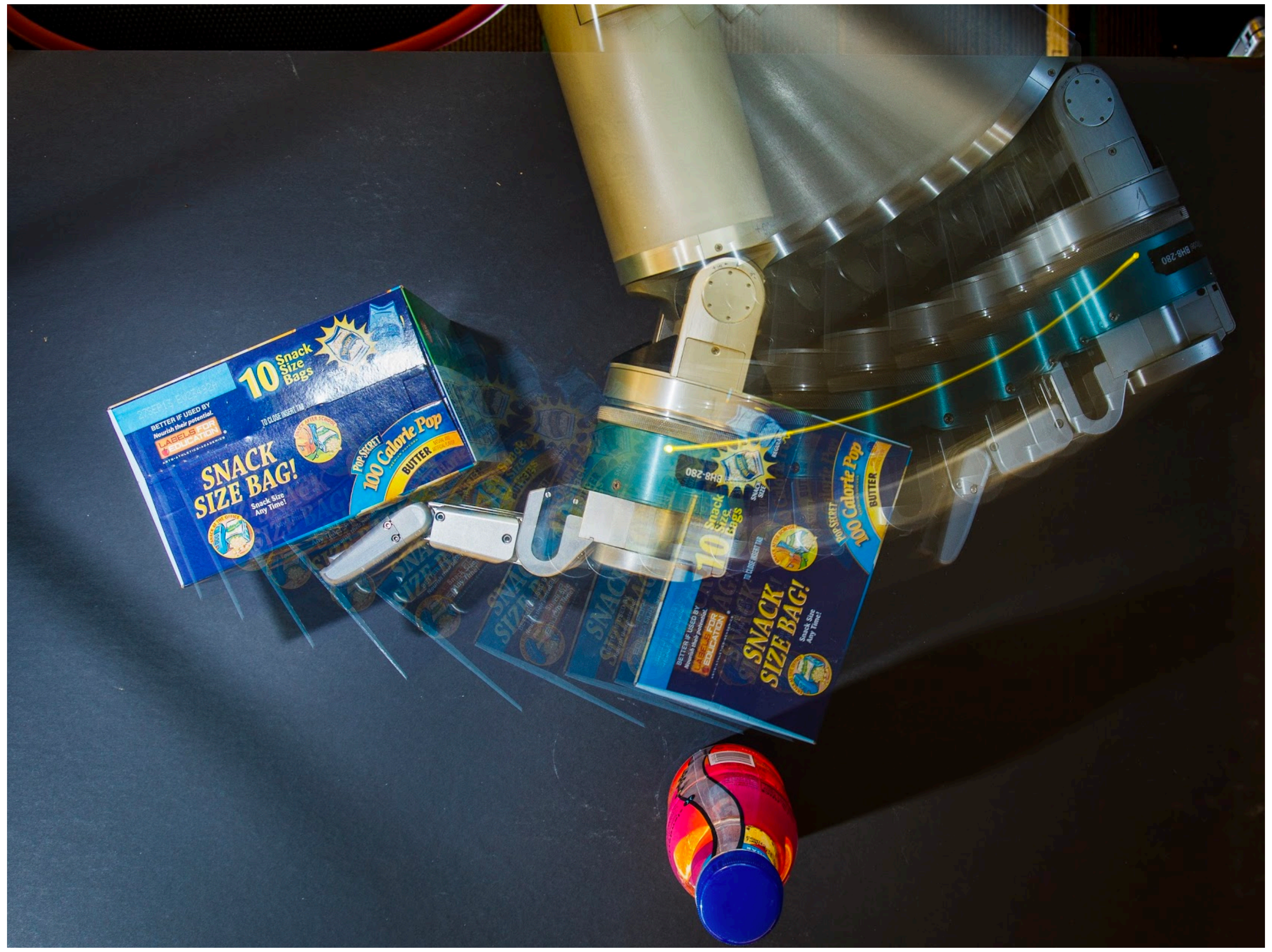
Epistemic uncertainty

(Uncertainty can be reduced through observations)

Epistemic Uncertainty



Uncertain about state

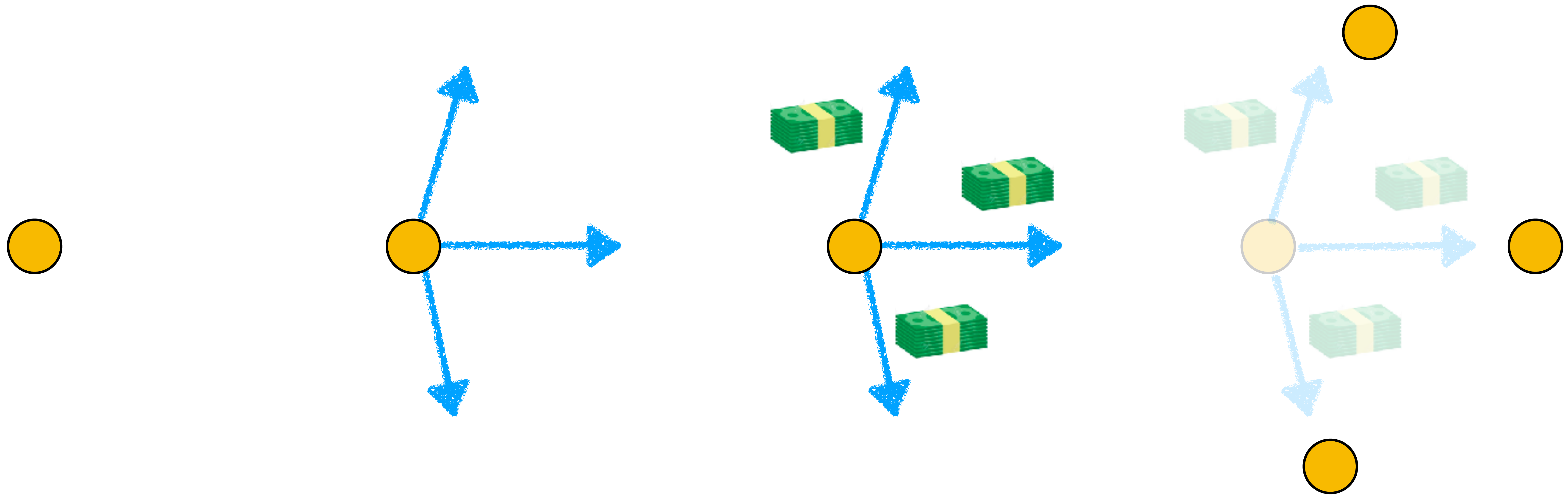


Uncertain about transitions

Markov Decision Process

A mathematical framework for modeling sequential decision making

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



Partially Observable Markov Decision Process

A mathematical framework for modeling sequential decision making

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

State is not observable!

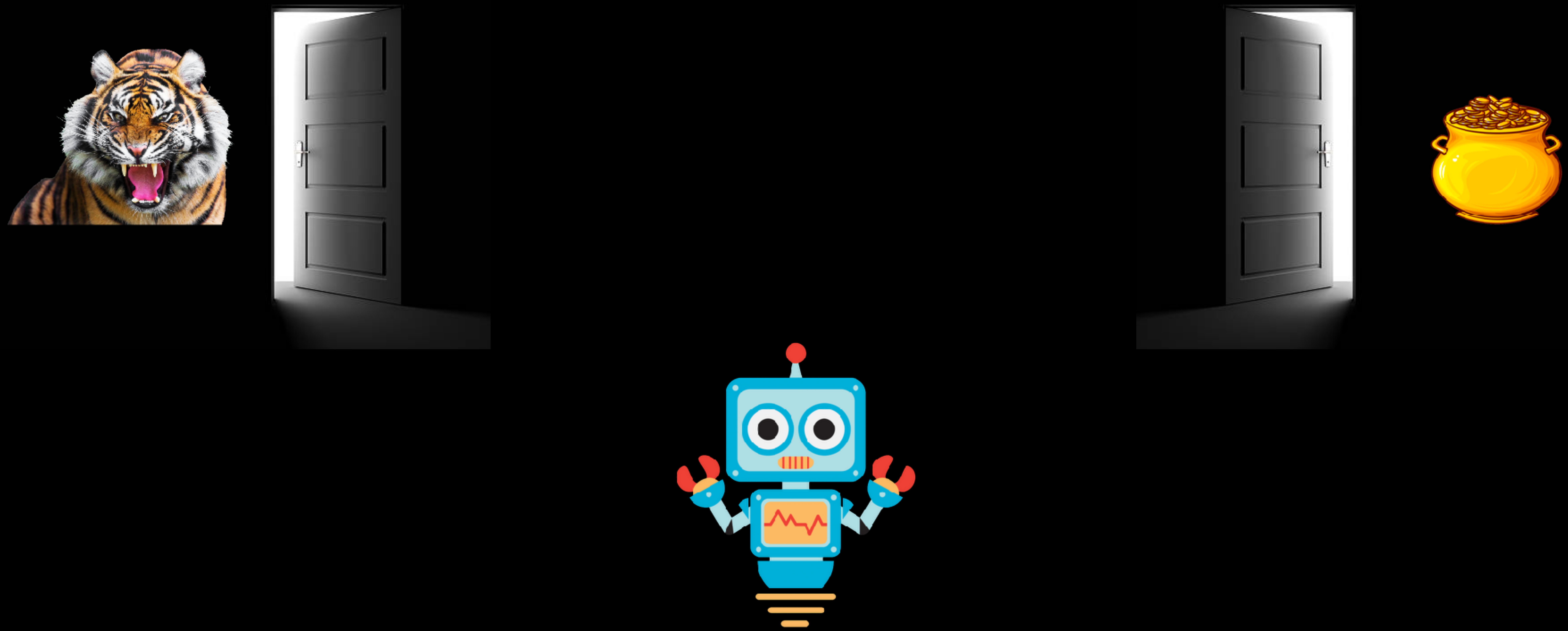
Partially Observable Markov Decision Process

A mathematical framework for modeling sequential decision making

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

How do we solve such MDPs ??

The Tiger Problem



The Tiger Problem

There are two doors, one with a pot of gold, one with a tiger

You don't know where the tiger is

You can either open door left, open door right, or listen

Reward for gold=+10, tiger=-100, listen=-1

Listen tells you with 0.85 prob which door the tiger is in

Let's solve this on the board

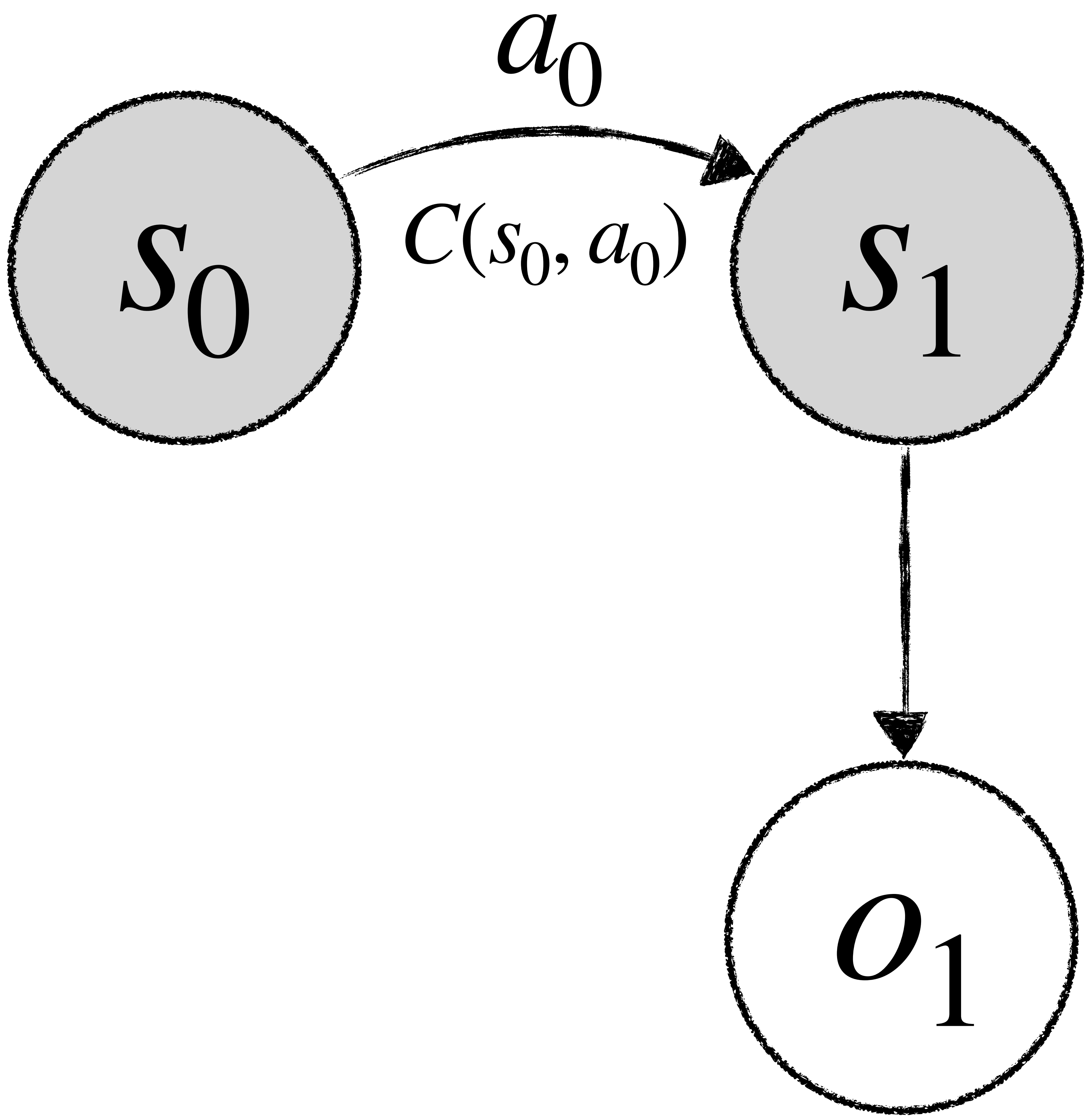
Partially Observable Markov Decision Process

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

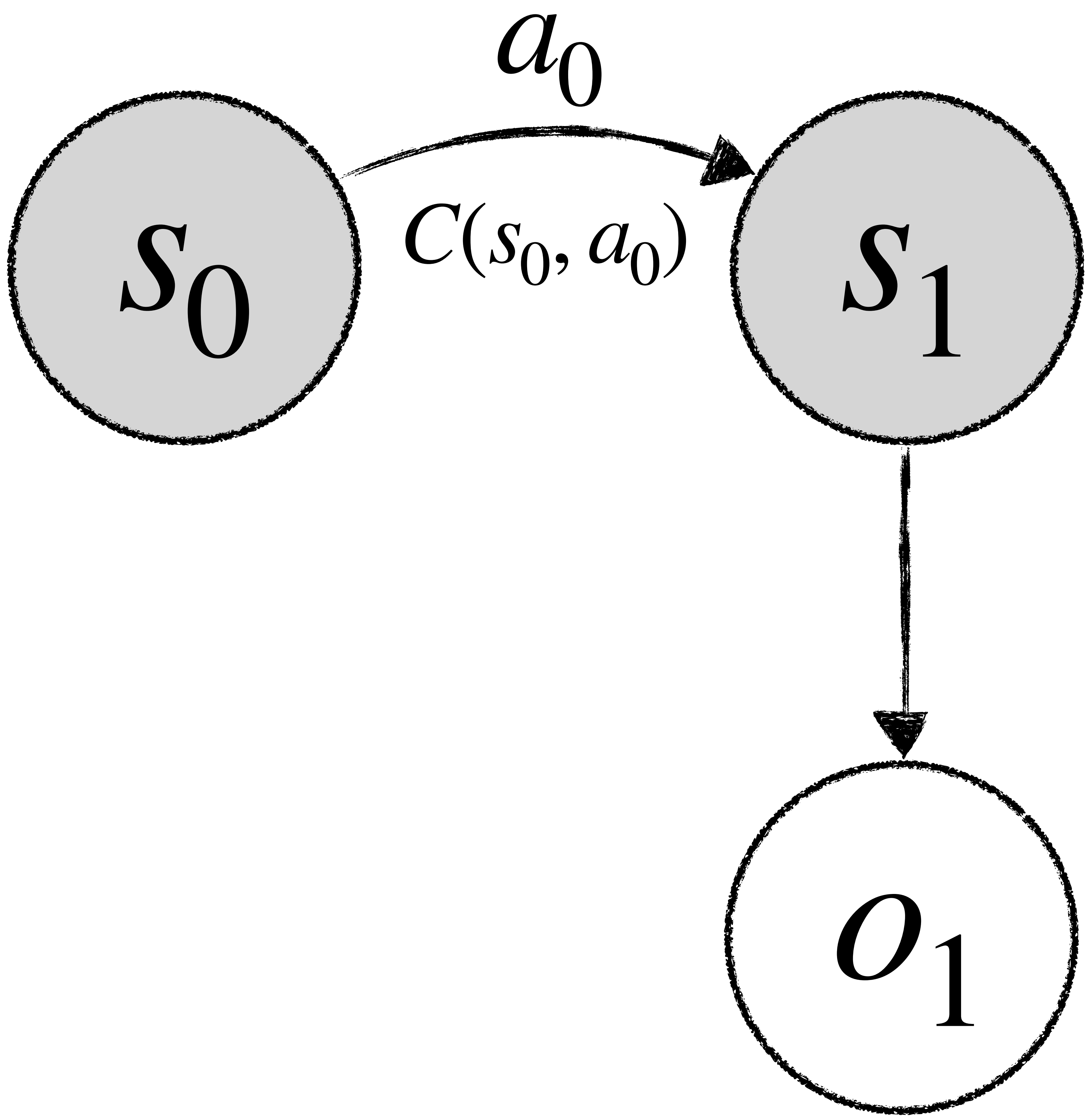
Observations

The Graphical Model

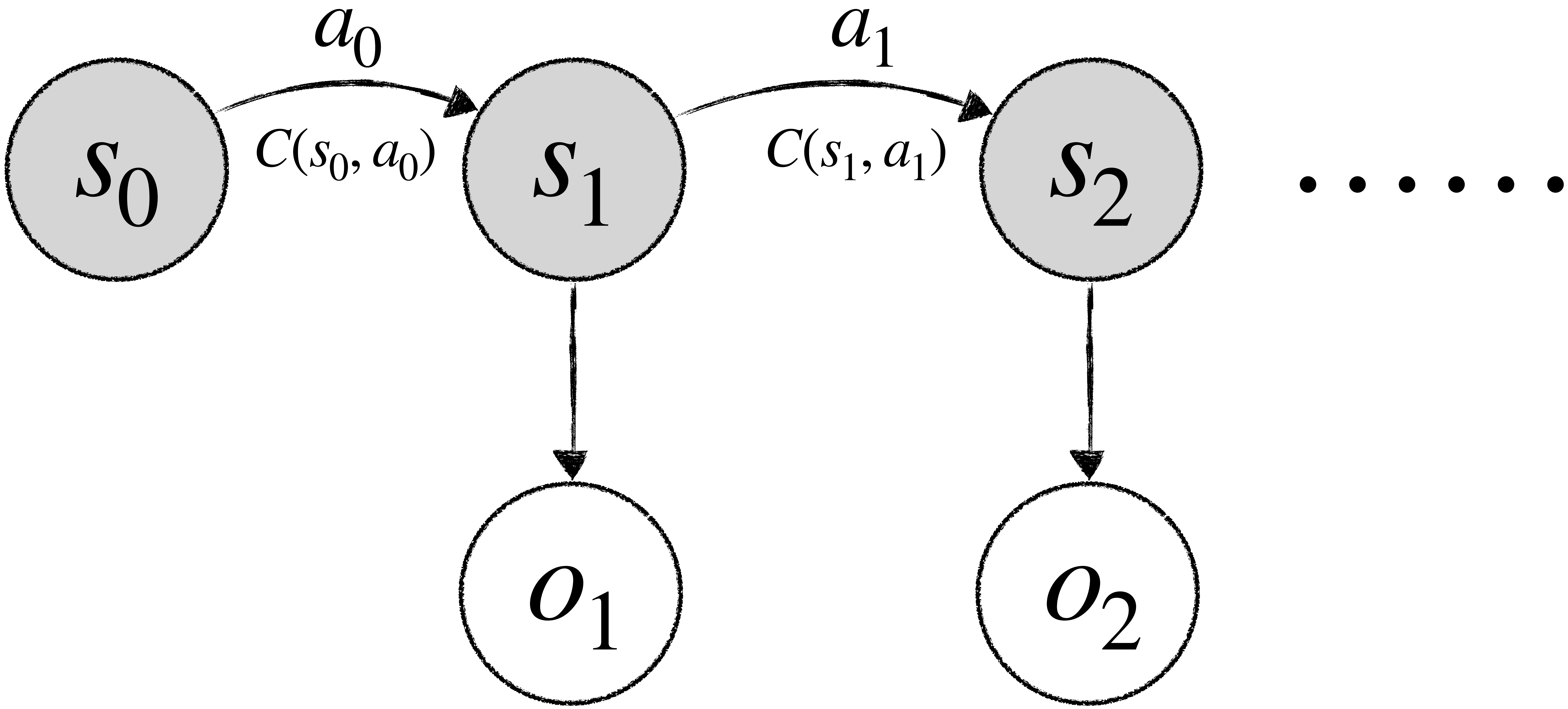
The Graphical Model



The Graphical Model



The Graphical Model



Convert MDP over states
to MDP over *belief*

Belief State

$$b_t$$

Probability over states given history of actions and observations

$$b_t = P(s_t \,|\, o_t, a_{t-1}, \ldots, a_1, o_1, a_0)$$

Belief State is Markovian!

$$b_{t+1} = P(s_{t+1} \,|\, o_{t+1}, a_t, \ldots, a_1, o_1, a_0)$$

Belief State is Markovian!

$$b_{t+1} = P(s_{t+1} \,|\, o_{t+1}, a_t, \ldots, a_1, o_1, a_0)$$

(Bayes Rule) $\propto P(o_{t+1} \,|\, s_{t+1}) P(s_{t+1} \,|\, a_t, o_t, \ldots, a_1, o_1, a_0)$

Belief State is Markovian!

$$b_{t+1} = P(s_{t+1} \,|\, o_{t+1}, a_t, \ldots, a_1, o_1, a_0)$$

$$\text{(Bayes Rule)} \;\propto P(o_{t+1} \,|\, s_{t+1}) P(s_{t+1} \,|\, a_t, o_t, \ldots, a_1, o_1, a_0)$$

$$\text{(Transition Function)} \propto P(o_{t+1} \,|\, s_{t+1}) \sum_{s_t} P(s_{t+1} \,|\, s_t, a_t) P(s_t \,|\, o_t, a_{t-1}, \ldots)$$

Belief State is Markovian!

$$b_{t+1} = P(s_{t+1} \mid o_{t+1}, a_t, \ldots, a_1, o_1, a_0)$$

$$\text{(Bayes Rule)} \;\propto P(o_{t+1} \mid s_{t+1}) P(s_{t+1} \mid a_t, o_t, \ldots, a_1, o_1, a_0)$$

$$\text{(Transition Function)} \propto P(o_{t+1} \mid s_{t+1}) \sum_{s_t} P(s_{t+1} \mid s_t, a_t) P(s_t \mid o_t, a_{t-1}, \ldots)$$

$$\propto P(o_{t+1} \mid s_{t+1}) \sum_{s_t} P(s_{t+1} \mid s_t, a_t) \quad b_t$$

The "Transition Function" of Belief

$$b_{t+1} \propto P(o_{t+1} \mid s_{t+1}) \sum_{s_t} P(s_{t+1} \mid s_t, a_t) \quad b_t$$

New Belief

Observation Prob

Transition Prob

Old Belief

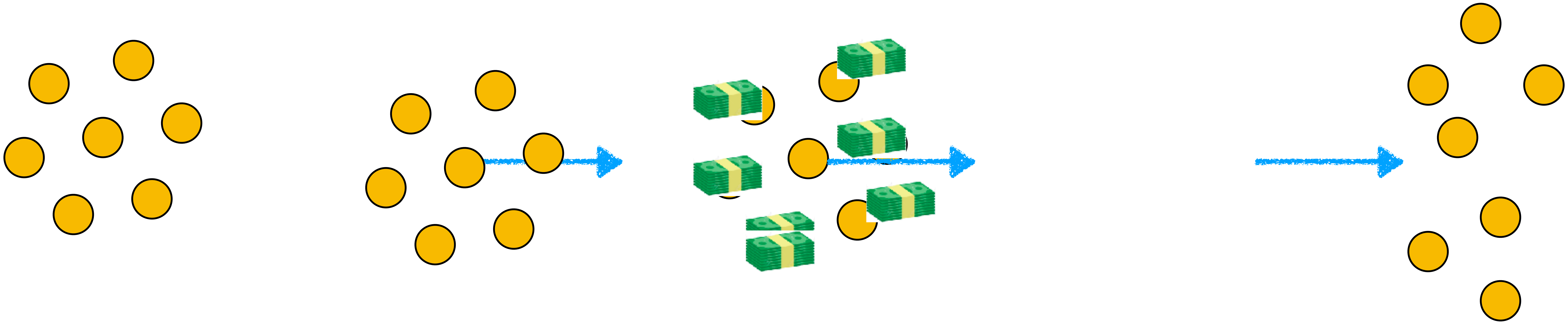
The "Cost Function" in Belief Space

$$c(b_t, a_t) = \sum_s b_t(s) c(s, a_t)$$

Belief Cost is simply the expected cost under my current belief

Belief Markov Decision Process

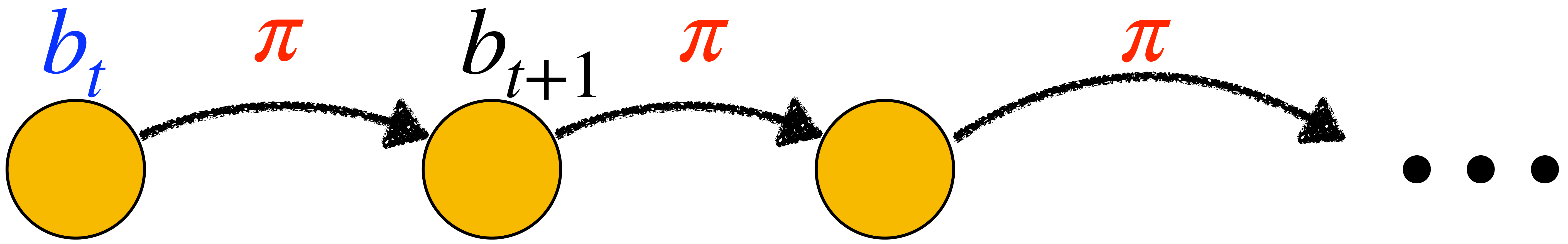
$$< B\,, A\,, C^{B}\,, \mathcal{T}^{B} >$$



The "Value" Function

$$V^{\pi}(b_t)$$

Read this as: Value of a policy at a given belief and time



$$V^{\pi}(b_t) = c_t + \gamma c_{t+1} + \gamma^2 c_{t+2} +$$

The Bellman Equation in Belief Space

$$V^{*}(b_t) = \min_{a_t} \left[c(b_t, a_t) + \gamma \mathbb{E}_{b_{t+1}} V^{*}(b_{t+1})) \right]$$

Optimal Value

Cost

Optimal Value of Next State

Are we done?

Seems like everything we learned so far can be "lifted" to belief space!

A slight “wrinkle”

What is the size of the belief space?

Consider the tiger MDP with 2 states.
How many belief states can there be?

Belief space is enormous

For N finite state MDP,
it's continuous with N dimensions

It's infinite dimensional
for continuous MDPs

Belief space is enormous

Working with an explicit belief space is a no-go ...

But is there an "implicit" belief representation?

Belief space is enormous

Working with an explicit belief space is a no-go ...

But is there an “implicit” belief representation?

Idea: What if we directly work with the history of observations and actions?

$$h_t = \{o_t, a_{t-1}, o_{t-1}, a_{t-2}, \ldots\}$$

Idea: What if we directly work with the history of observations and actions?

$$h_t = \{o_t, a_{t-1}, o_{t-1}, a_{t-2}, \ldots\}$$

History seems to have all the information we need to represent belief

What sort of models can represent history?

$$h_t = \{o_t, a_{t-1}, o_{t-1}, a_{t-2}, \dots\}$$

Sequence models like Transformers!

Turn all your models into sequence models!

$$\pi : h_t \to a_t$$

(Sequence of tokens) (Action tokens)

$$Q : h_t, a_t \to \mathbb{R}$$

(Sequence of tokens + action token)

The Bellman Equation in Belief Space

$$V^{*}(h_t) = \min_{a_t} \left[c(h_t, a_t) + \gamma \mathbb{E}_{b_{t+1}} V^{*}(h_{t+1})) \right]$$

Turn all our algorithms to history models

BC

DAGGER

REINFORCE

Q-learning