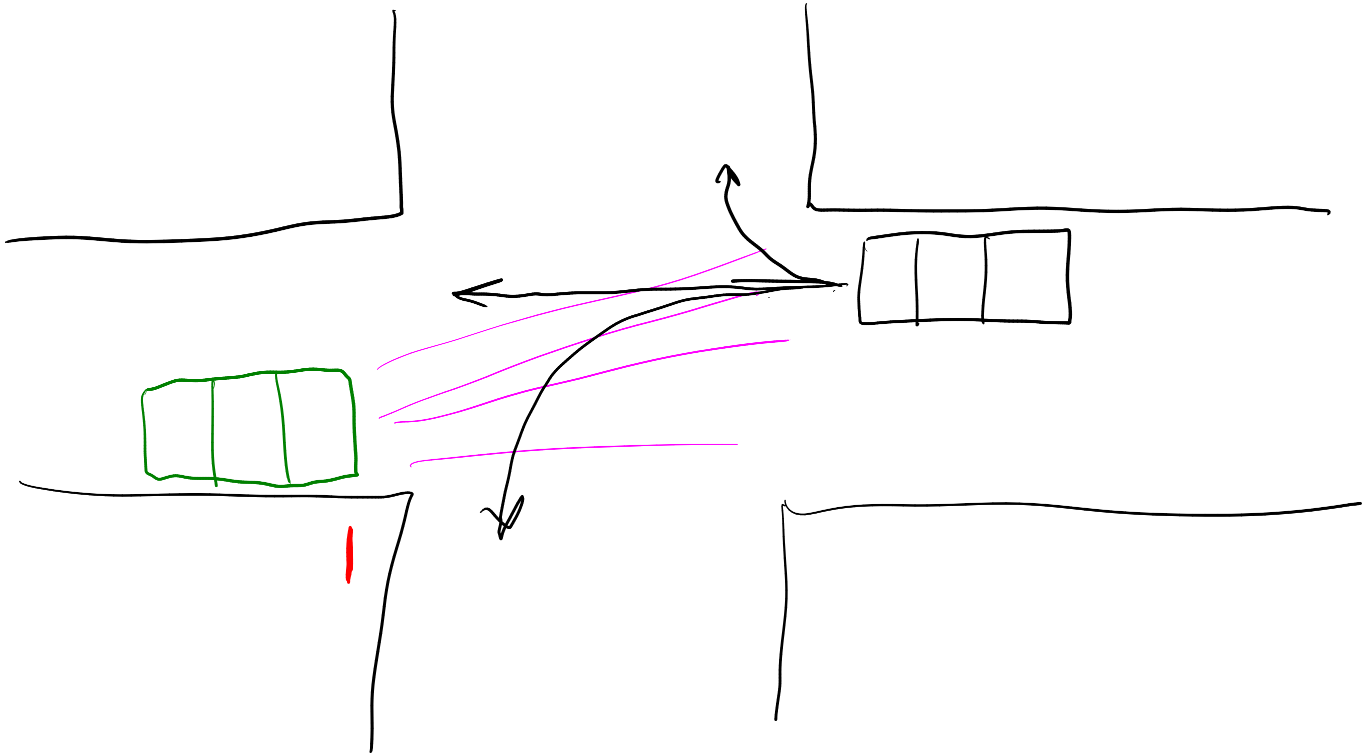
POMDPs

POMDPs

- We've been living a lie:

```
s = observe(env)
```



Types of Uncertainty

Types of Uncertainty

Alleatory

Types of Uncertainty

Alleatory

Types of Uncertainty

Alleatory

Epistemic (Static)

Types of Uncertainty

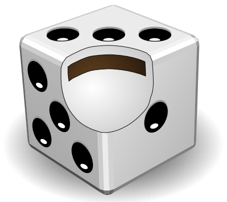
Alleatory

Epistemic (Static)

Types of Uncertainty

Alleatory

Epistemic (Static)



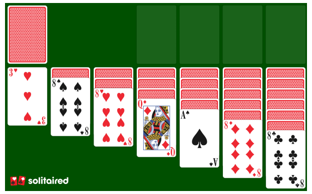
Epistemic (Dynamic)

Types of Uncertainty

Alleatory

Epistemic (Static)

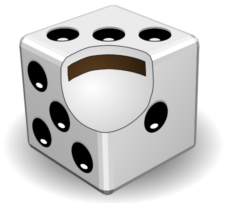
Epistemic (Dynamic)



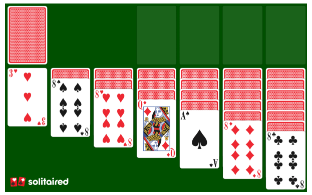
Types of Uncertainty

Alleatory

Epistemic (Static)



Epistemic (Dynamic)

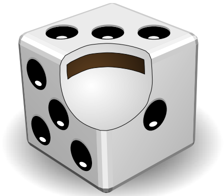


Interaction

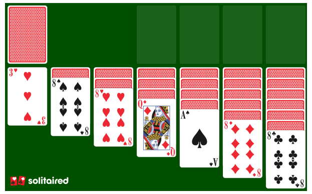
Types of Uncertainty

Alleatory

Epistemic (Static)



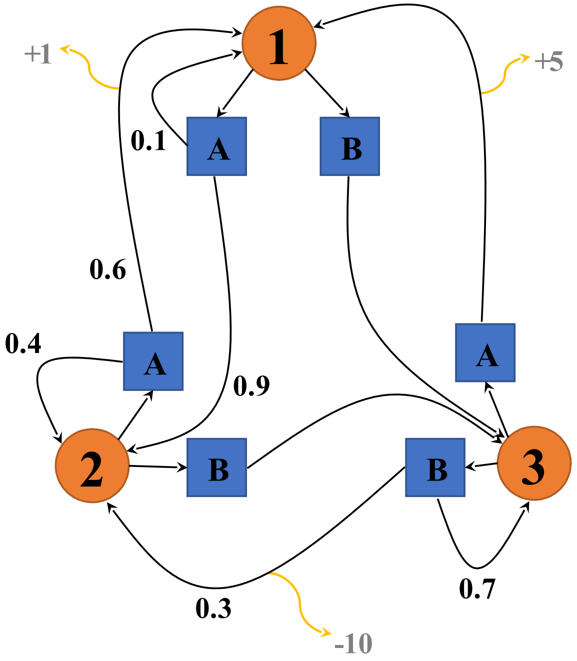
Epistemic (Dynamic)



POMDPs

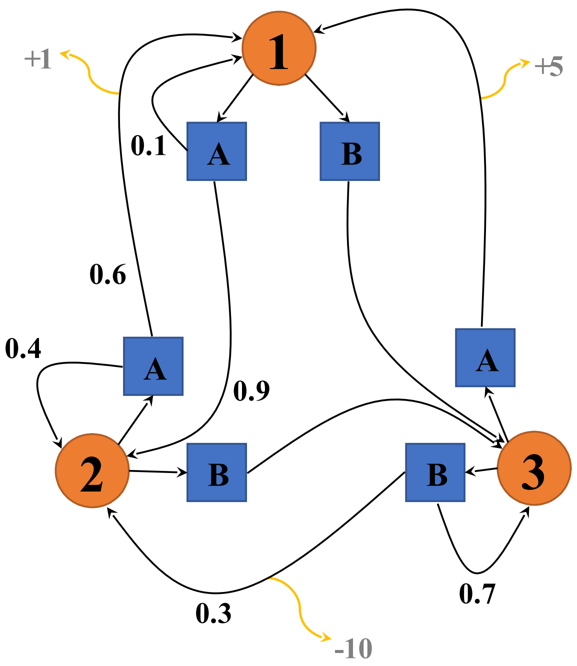
Interaction

Markov Decision Process (MDP)



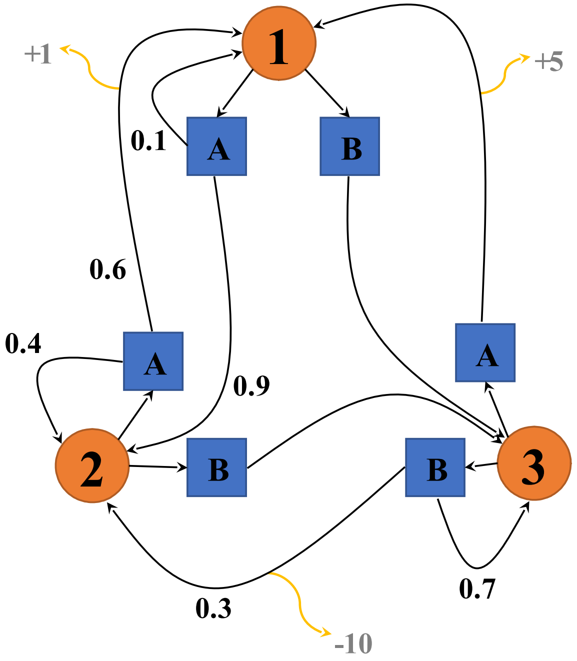
- $\mathcal{S}$ - State space
- $T : \mathcal{S} \times \mathcal{A} \times \mathcal{S} \to \mathbb{R}$ - Transition probability distribution

Markov Decision Process (MDP)



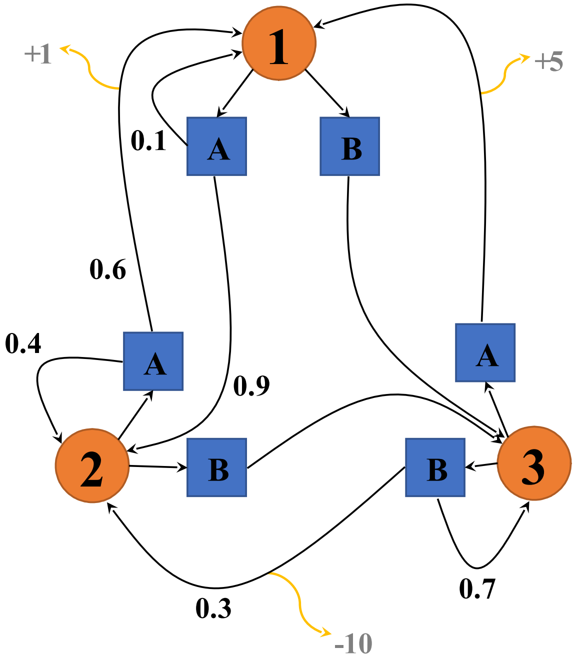
- $\mathcal{S}$ - State space
- $T : \mathcal{S} \times \mathcal{A} \times \mathcal{S} \to \mathbb{R}$ - Transition probability distribution
- $\mathcal{A}$ - Action space

Markov Decision Process (MDP)



- $\mathcal{S}$ - State space
- $T : \mathcal{S} \times \mathcal{A} \times \mathcal{S} \to \mathbb{R}$ - Transition probability distribution
- $\mathcal{A}$ - Action space
- $R : \mathcal{S} \times \mathcal{A} \to \mathbb{R}$ - Reward

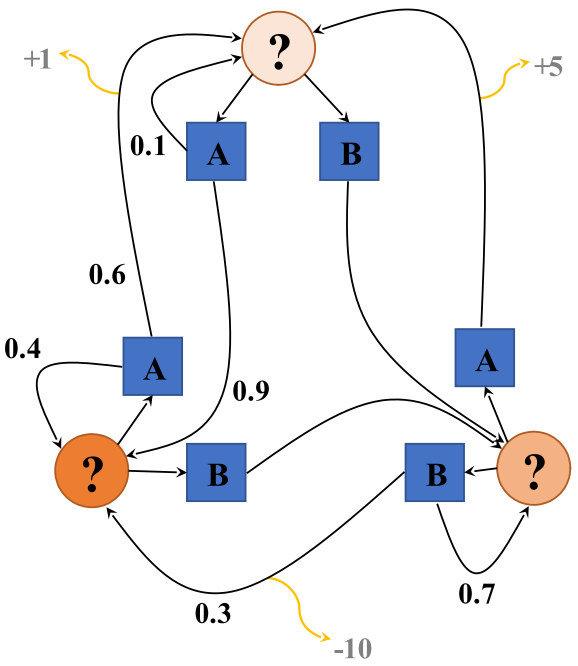
Markov Decision Process (MDP)



- $\mathcal{S}$ - State space
- $T : \mathcal{S} \times \mathcal{A} \times \mathcal{S} \to \mathbb{R}$ - Transition probability distribution
- $\mathcal{A}$ - Action space
- $R : \mathcal{S} \times \mathcal{A} \to \mathbb{R}$ - Reward

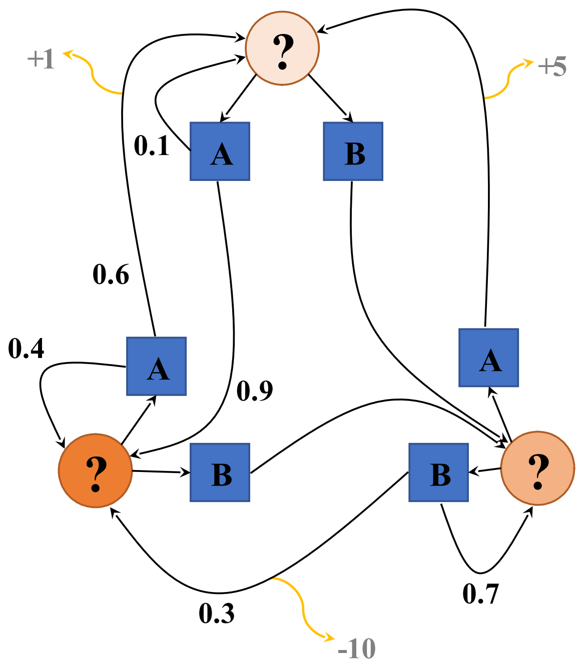
Alleatory

Partially Observable Markov Decision Process (POMDP)



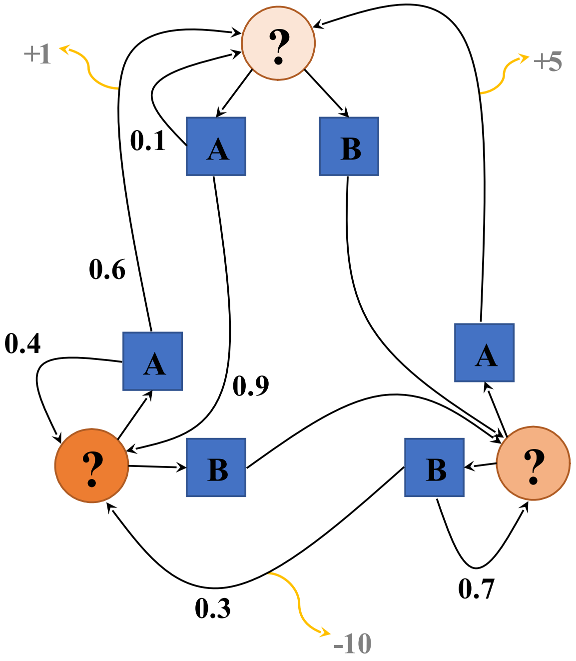
- $\mathcal{S}$ - State space
- $T : \mathcal{S} \times \mathcal{A} \times \mathcal{S} \to \mathbb{R}$ - Transition probability distribution
- $\mathcal{A}$ - Action space
- $R : \mathcal{S} \times \mathcal{A} \to \mathbb{R}$ - Reward

Partially Observable Markov Decision Process (POMDP)



- $\mathcal{S}$ - State space
- $T : \mathcal{S} \times \mathcal{A} \times \mathcal{S} \rightarrow \mathbb{R}$ - Transition probability distribution
- $\mathcal{A}$ - Action space
- $R : \mathcal{S} \times \mathcal{A} \rightarrow \mathbb{R}$ - Reward
- $\mathcal{O}$ - Observation space

Partially Observable Markov Decision Process (POMDP)



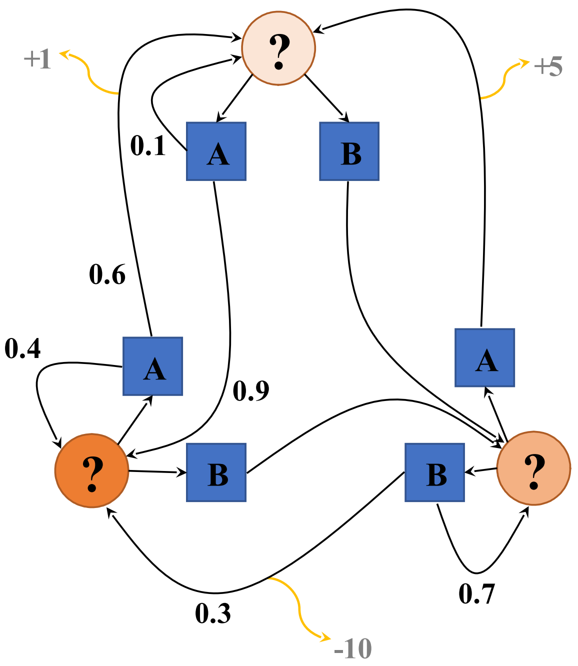
- $\mathcal{S}$ - State space
- $T : \mathcal{S} \times \mathcal{A} \times \mathcal{S} \rightarrow \mathbb{R}$ - Transition probability distribution
- $\mathcal{A}$ - Action space
- $R : \mathcal{S} \times \mathcal{A} \rightarrow \mathbb{R}$ - Reward
- $\mathcal{O}$ - Observation space
- $Z : \mathcal{S} \times \mathcal{A} \times \mathcal{S} \times \mathcal{O} \rightarrow \mathbb{R}$ - Observation probability distribution

a'

$Z(o \mid a, s')$

$Z(o \mid s, a, s')$

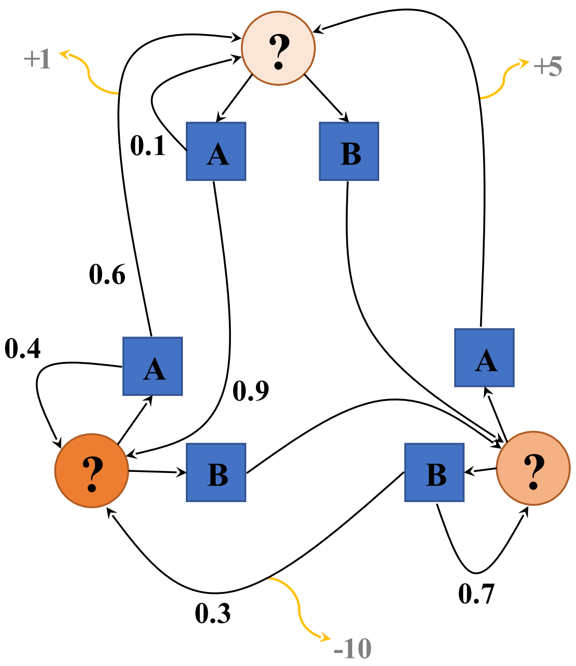
Partially Observable Markov Decision Process (POMDP)



- $\mathcal{S}$ - State space
- $T : \mathcal{S} \times \mathcal{A} \times \mathcal{S} \rightarrow \mathbb{R}$ - Transition probability distribution
- $\mathcal{A}$ - Action space
- $R : \mathcal{S} \times \mathcal{A} \rightarrow \mathbb{R}$ - Reward
- $\mathcal{O}$ - Observation space
- $Z : \mathcal{S} \times \mathcal{A} \times \mathcal{S} \times \mathcal{O} \rightarrow \mathbb{R}$ - Observation probability distribution

Alleatory

Partially Observable Markov Decision Process (POMDP)



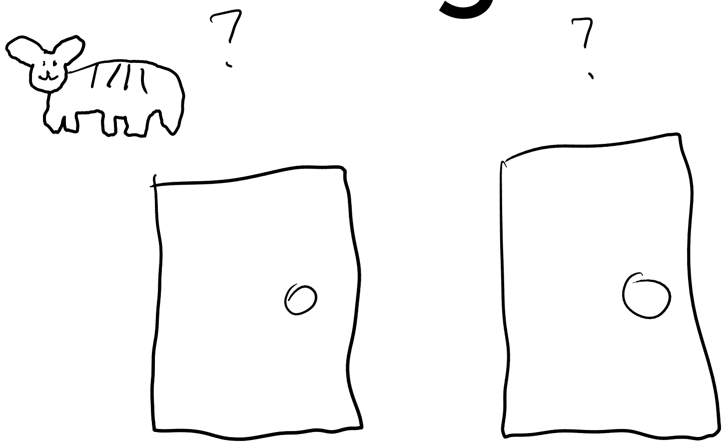
- $\mathcal{S}$ - State space
- $T : \mathcal{S} \times \mathcal{A} \times \mathcal{S} \to \mathbb{R}$ - Transition probability distribution
- $\mathcal{A}$ - Action space
- $R : \mathcal{S} \times \mathcal{A} \to \mathbb{R}$ - Reward
- $\mathcal{O}$ - Observation space
- $Z : \mathcal{S} \times \mathcal{A} \times \mathcal{S} \times \mathcal{O} \to \mathbb{R}$ - Observation probability distribution

Alleatory

Epistemic (Static)

Epistemic (Dynamic)

Tiger POMDP Definition



$-100 \leftarrow$ open, tiger behind

$+10 \leftarrow$ open, tiger not behind

$$R(s,a) = \begin{cases} -100 & \text{if } a = s \\ -1 & \text{if } a = \text{Listen} \\ 10 & \text{o.w.} \end{cases}$$

Listen: 85% correct observation

$$S = \{\overset{TL}{L}, \overset{TR}{R}\}$$

$$A = \{\overset{OL}{L}, \overset{OR}{R}, \overset{L}{\text{Listen}}\}$$

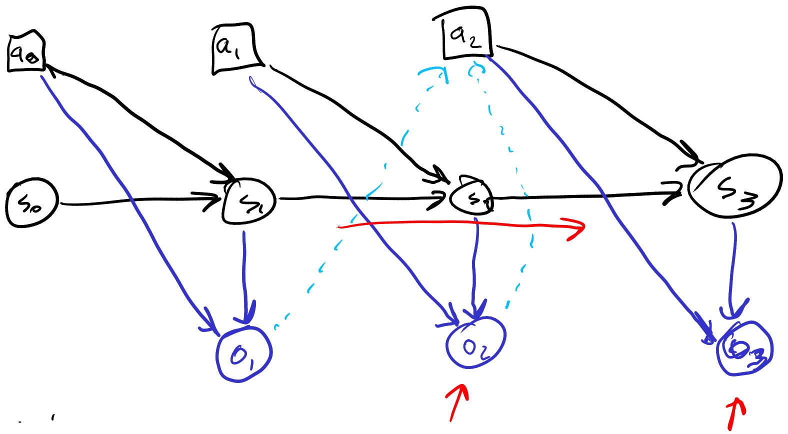
$$\gamma = 0.95$$

$$\mathcal{O} = S$$

$$T^{\text{listen}} = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix} \qquad T^{L} = T^{R} = \begin{bmatrix} 0.5 & 0.5 \\ 0.5 & 0.5 \end{bmatrix}$$

$$Z(o \mid a, s') = \begin{cases} 0.85 & \text{if } a = \text{Listen and } s' = o \\ 0.15 & \text{if } a = \text{Listen and } s' \neq o \\ 0.5 & \text{if } a \neq \text{Listen} \end{cases}$$

Hidden Markov Models and Beliefs



For an MDP

$$P(s_{t+1} \mid s_0, a_0 \ldots s_t, a_t) = T(s_{t+1} \mid s_t, a_t)$$

$$\cancel{P(o_{t+1} \mid o_0, a_0 \ldots o_t, a_t) = P(o_{t+1} \mid a_t, o_{t-1})}$$

$$b_0(s) \equiv P(s_0 = s)$$

$$h_t \equiv (b_0, a_0, o_1, a_1 \ldots a_{t-1}, o_t)$$

$$b_t(s) \equiv P(s_t = s \mid h_t)$$

$$\underline{P(b_{t+1} \mid b_0, a_0 \ldots b_t, a_t) = P(b_{t+1} \mid b_t, a_t)}$$

$$S = \{L, R\}$$

$$B = \Delta(S) = [0,1]^{|S|-1}$$

$$b = [0.2, 0.8]$$

Bayesian Belief Updates

$h_t = (b_0, a_0, o_1, \dots, a_{t-1}, o_t)$

$$b_t \equiv P(s_t \mid h_t) = P(s_t \mid h_{t-1}, a_{t-1}, o_t)$$

$$= \frac{P(o_t \mid s_t, h_{t-1}, a_{t-1})\, P(s_t \mid h_{t-1}, a_{t-1})}{P(o_t \mid h_{t-1}, a_{t-1})}$$

$$\propto P(o_t \mid s_t, h_{t-1}, a_{t-1})\, P(s_t \mid h_{t-1}, a_{t-1})$$

$$= P(o_t \mid s_t, a_{t-1}) \sum_{s_{t-1}} \underbrace{P(s_t \mid s_{t-1}, a_{t-1})} \underbrace{P(s_{t-1} \mid h_{t-1}, \cancel{a_{t-1}})}$$

$$= Z(o_t \mid a_{t-1}, s_t) \sum_{s_{t-1}} T(s_t \mid s_{t-1}, a_{t-1})\, b_{t-1}(s_{t-1})$$

measurement update

Prediction

$$\boxed{b'(s') \propto Z(o \mid a, s') \sum_s T(s' \mid s, a)\, b(s)}$$

normalize by $\sum_{s'} b'(s')$

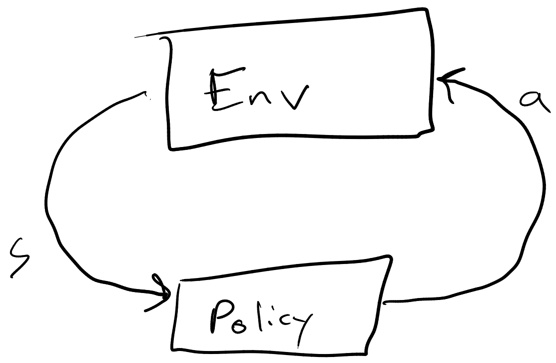
$Z(o \mid a, s')$

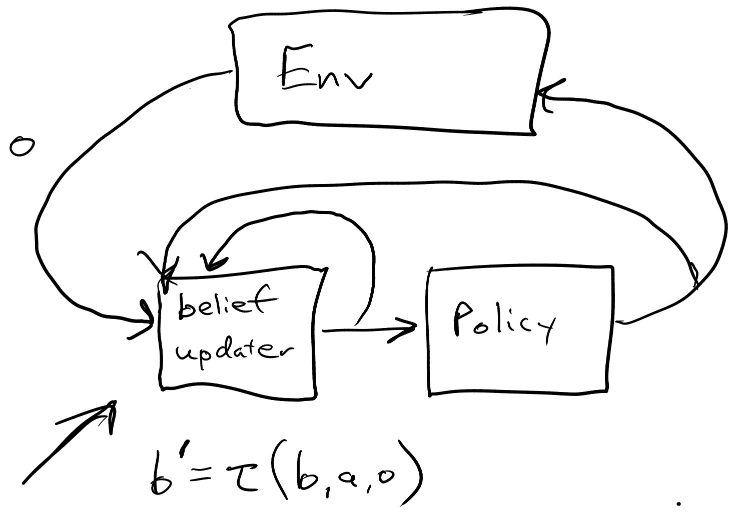
$$P(s_t \mid h_{t-1}, a_{t-1})$$
$$= \sum_{s_{t-1}} P(s_t, s_{t-1} \mid h_{t-1}, a_{t-1})$$
$$= \sum_{s_{t-1}} P(s_t \mid s_{t-1}, h_{t-1}, a_{t-1})\, P(s_{t-1} \mid h_{t-1}, a_{t-1})$$

$s_t \perp h_{t-1} \mid s_{t-1}, a_{t-1}$

$$= \sum P(s_t \mid s_{t-1}, a_{t-1})\, P(s_{t-1} \mid h_{t-1}, a_{t-1})$$

Filtering Loop





Tiger Example

Recap

S, A, O, R, Z, T, γ

$b' = \tau(b, a, o)$

Next Time

Policies