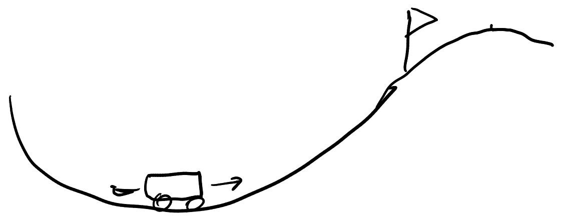
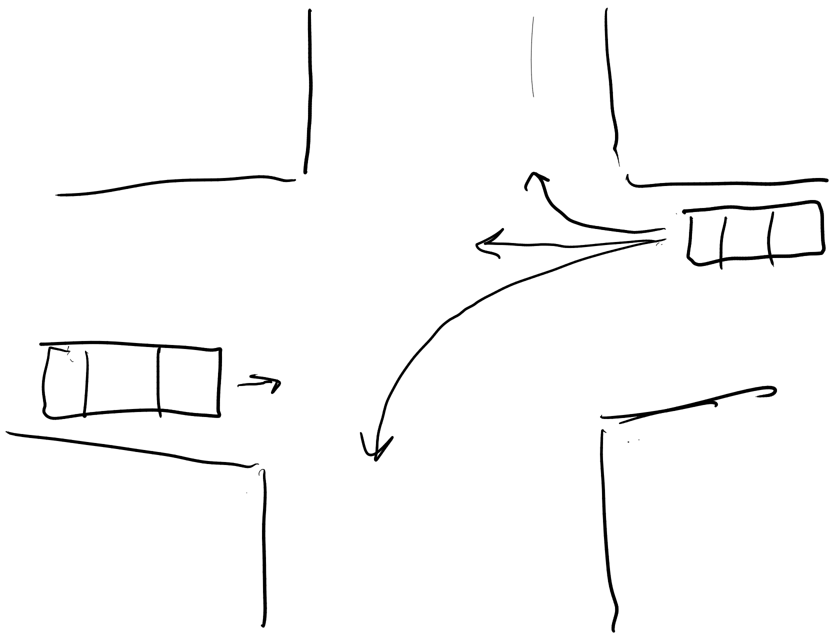


POMDPs

POMDPs

- We've been living a lie:

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



Types of Uncertainty

Types of Uncertainty

Alleatory

Types of Uncertainty

Alleatory

Types of Uncertainty

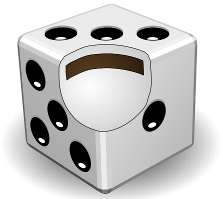
Alleatory

Epistemic (Static)

Types of Uncertainty

Alleatory

Epistemic (Static)



Types of Uncertainty

Alleatory

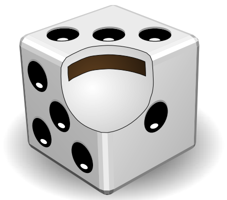
Epistemic (Static)

Epistemic (Dynamic)

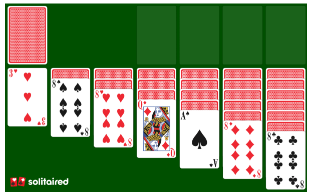
Types of Uncertainty

Alleatory

Epistemic (Static)



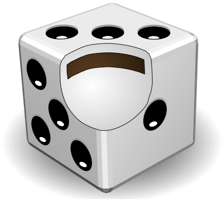
Epistemic (Dynamic)



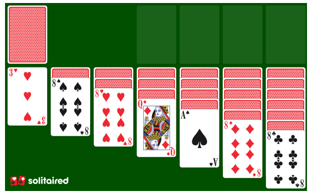
Types of Uncertainty

Alleatory

Epistemic (Static)



Epistemic (Dynamic)

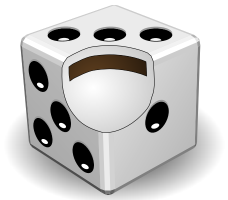


Interaction

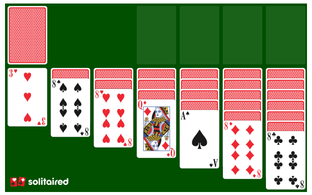
Types of Uncertainty

Alleatory

Epistemic (Static)



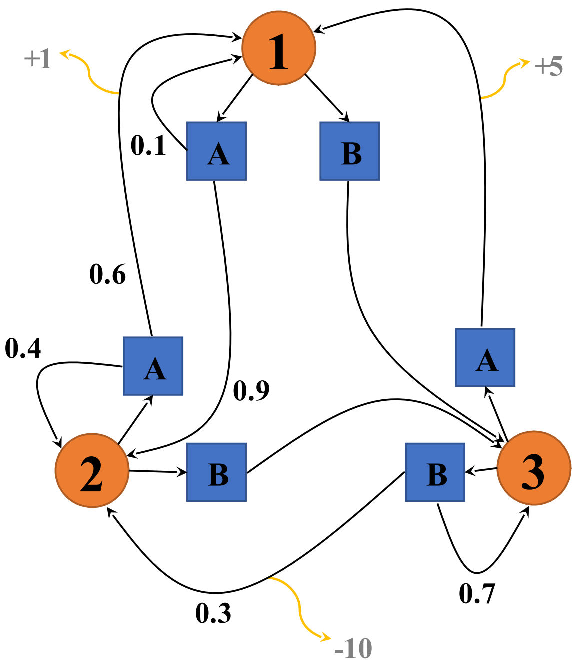
Epistemic (Dynamic)



POMDP

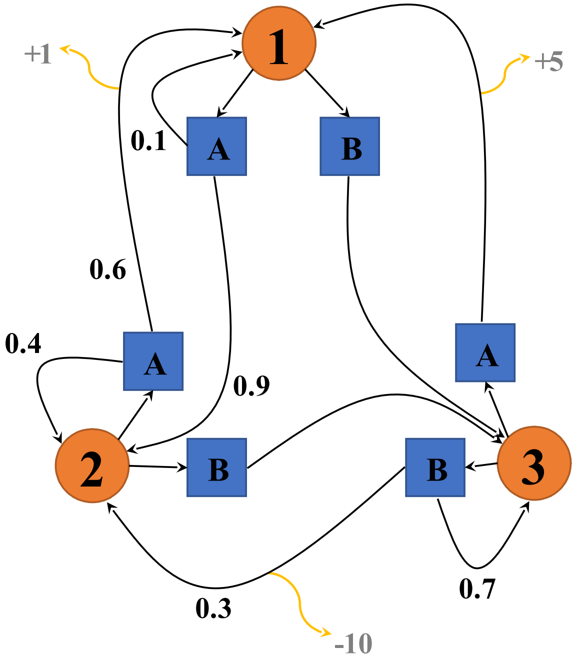
Interaction

Markov Decision Process (MDP)



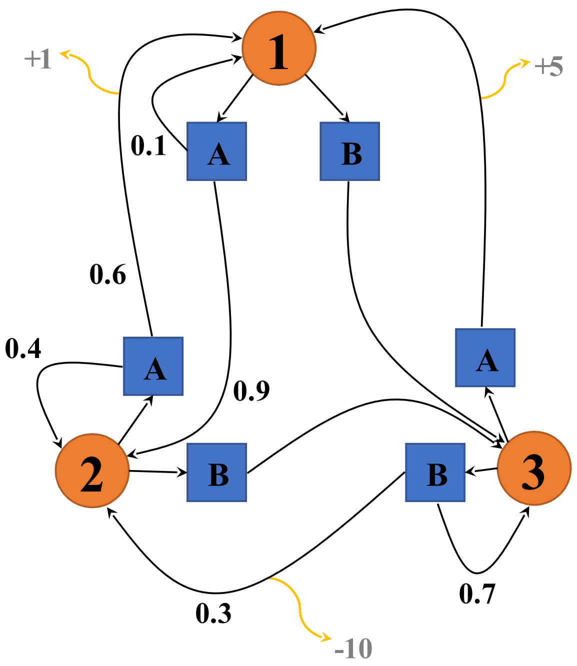
- $\mathcal{S}$ - State space
- $T(s' \mid s, a)$ - Transition probability distribution

Markov Decision Process (MDP)



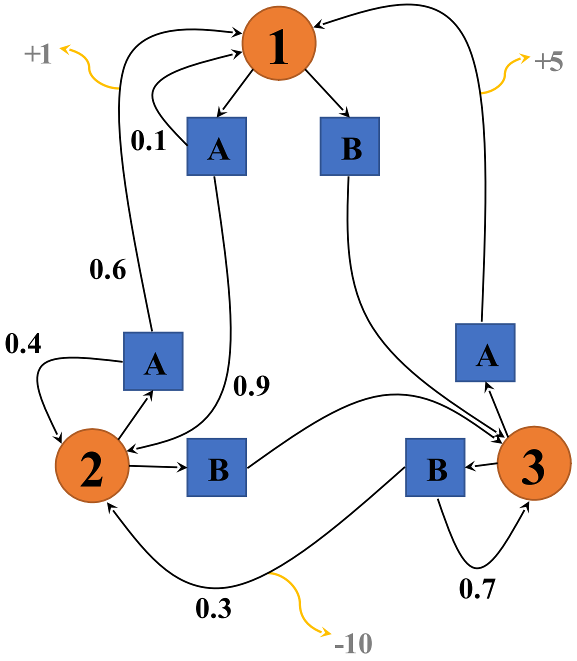
- $\mathcal{S}$ - State space
- $T(s' \mid s, a)$ - Transition probability distribution
- $\mathcal{A}$ - Action space

Markov Decision Process (MDP)



- $\mathcal{S}$ - State space
- $T(s' \mid s, a)$ - Transition probability distribution
- $\mathcal{A}$ - Action space
- $R(s, a)$ - Reward

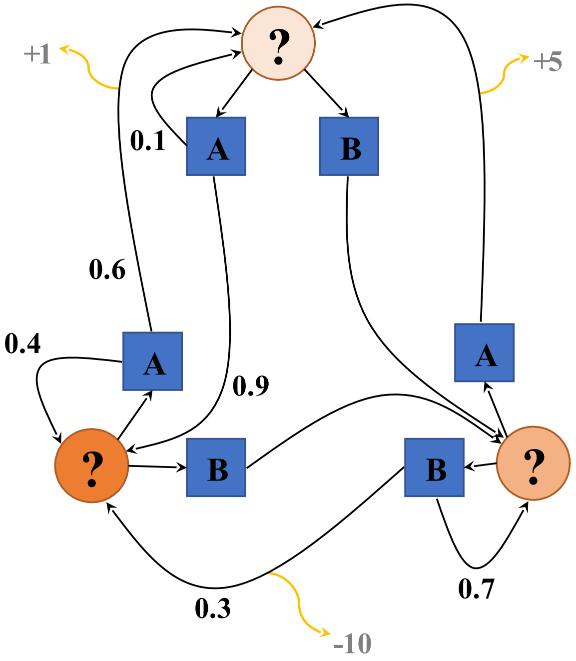
Markov Decision Process (MDP)



- $\mathcal{S}$ - State space
- $T(s' \mid s, a)$ - Transition probability distribution
- $\mathcal{A}$ - Action space
- $R(s, a)$ - Reward

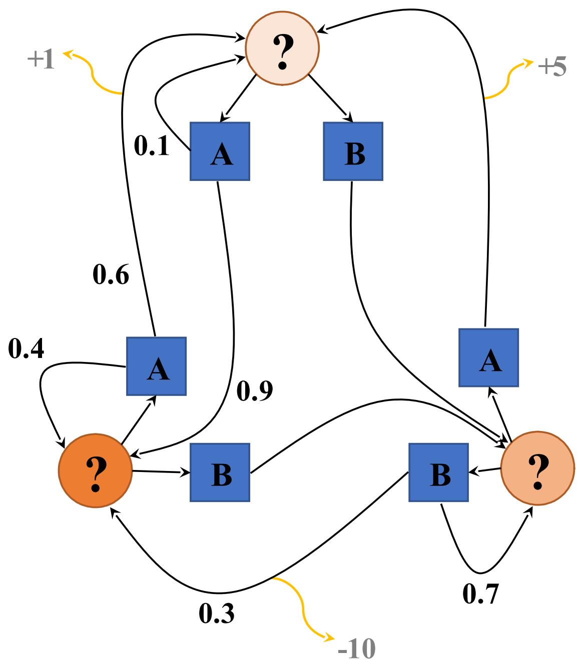
Alleatory

Partially Observable Markov Decision Process (POMDP)



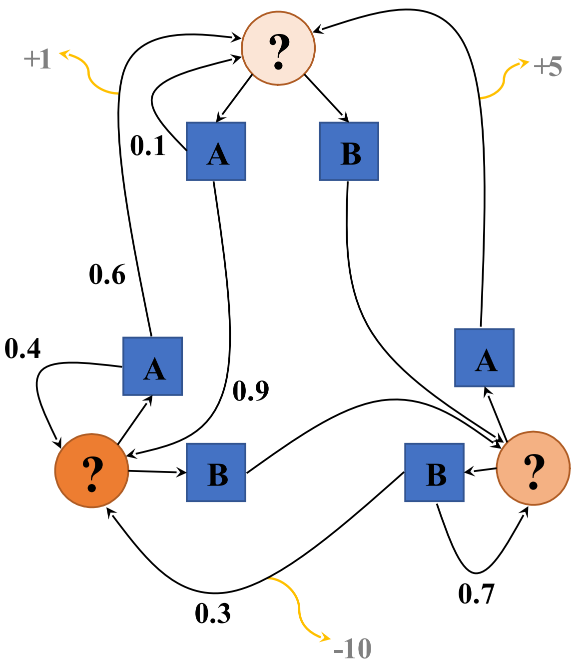
- $\mathcal{S}$ - State space
- $T(s' \mid s, a)$ - Transition probability distribution
- $\mathcal{A}$ - Action space
- $R(s, a)$ - Reward

Partially Observable Markov Decision Process (POMDP)



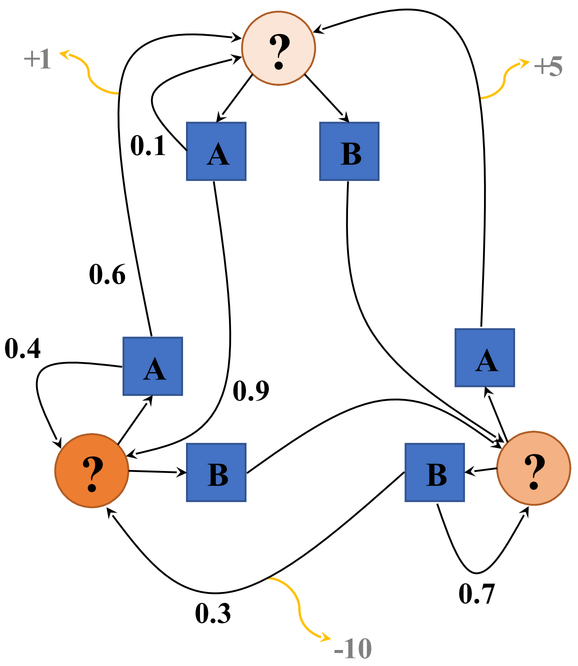
- $\mathcal{S}$ - State space
- $T(s' \mid s, a)$ - Transition probability distribution
- $\mathcal{A}$ - Action space
- $R(s, a)$ - Reward
- $\mathcal{O}$ - Observation space

Partially Observable Markov Decision Process (POMDP)



- $\mathcal{S}$ - State space
- $T(s' \mid s, a)$ - Transition probability distribution
- $\mathcal{A}$ - Action space
- $R(s, a)$ - Reward
- $\mathcal{O}$ - Observation space
- $Z(o \mid a, s')$ - Observation probability distribution

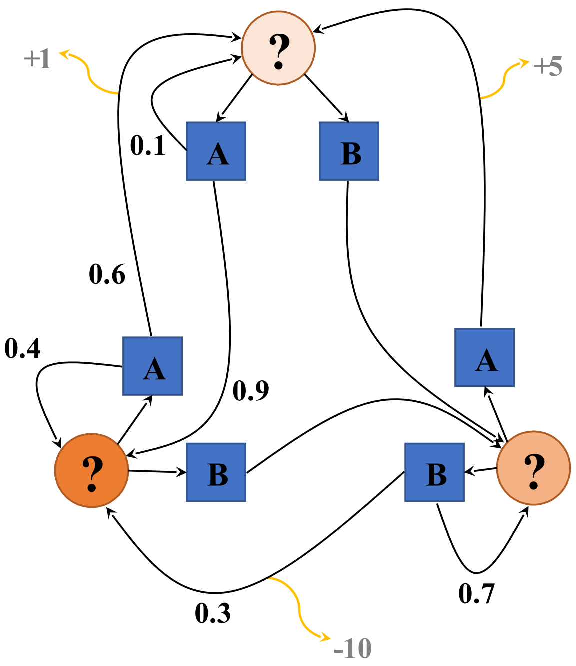
Partially Observable Markov Decision Process (POMDP)



- $\mathcal{S}$ - State space
- $T(s' \mid s, a)$ - Transition probability distribution
- $\mathcal{A}$ - Action space
- $R(s, a)$ - Reward
- $\mathcal{O}$ - Observation space
- $Z(o \mid a, s')$ - Observation probability distribution

Alleatory

Partially Observable Markov Decision Process (POMDP)



- $\mathcal{S}$ - State space
- $T(s' \mid s, a)$ - Transition probability distribution
- $\mathcal{A}$ - Action space
- $R(s, a)$ - Reward
- $\mathcal{O}$ - Observation space
- $Z(o \mid a, s')$ - Observation probability distribution

$\cancel{Z(o \mid s, a)}$

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

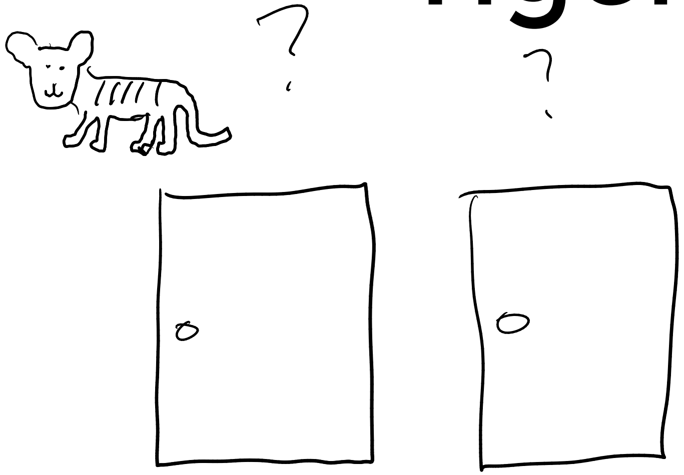
$Z(o \mid a, s')$

Alleatory

Epistemic (Static)

Epistemic (Dynamic)

Tiger POMDP Definition



−100

+10

Listen: 85% correct observation

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

$$\gamma = 0.95$$

$S = \{L, R\}$

$A = \{L, R, \text{Listen}\}$

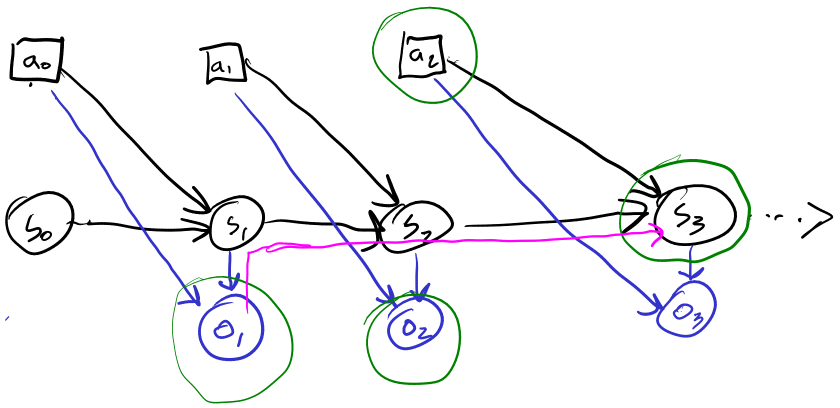
$O = \{L, R\}$

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

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

Hidden Markov Models and Beliefs

$Z(o|a,s')$



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

true

$$P(s_{t+1} | a_0, o_1, a_1, o_2 \ldots o_t, a_t) \stackrel{?}{=} P(s_{t+1} | o_t, a_t)$$

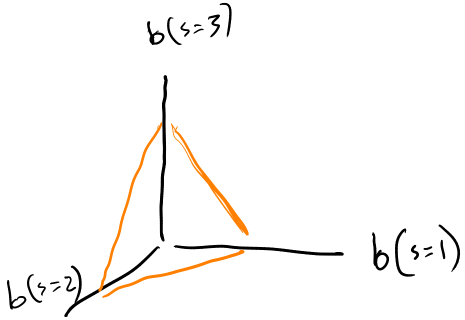
not true

$b_0(s) = b(s)$

initial state distribution

$$h_t \equiv (b_0, a_0, o_1, a_1, o_2 \ldots a_{t-1}, o_t)$$

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



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

POMDPs are MDPs on the belief space

$$\mathcal{B} = \Delta(S)$$

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

Bayesian Belief Updates

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

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

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

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

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

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

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

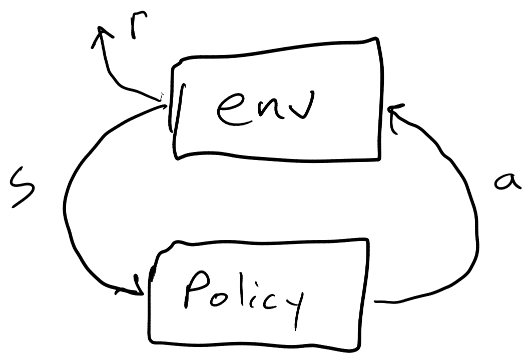
$T(s' \mid s, a)$

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

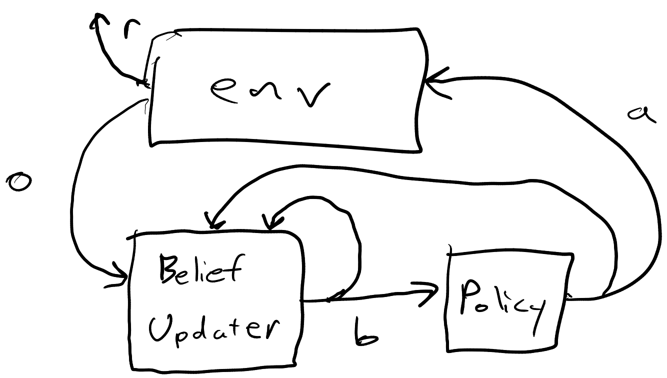
$b' \leftarrow \tau(b, a, o)$

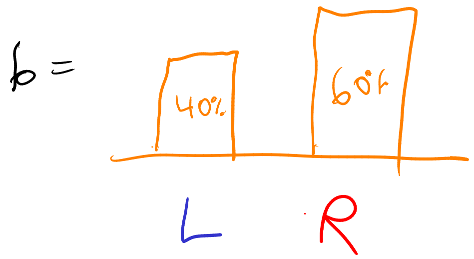
Filtering Loop

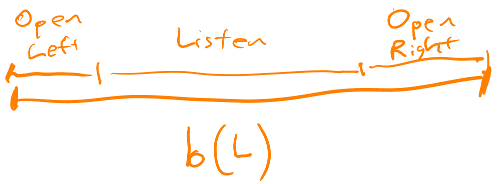
MDP



POMDP







Tiger Example

Recap