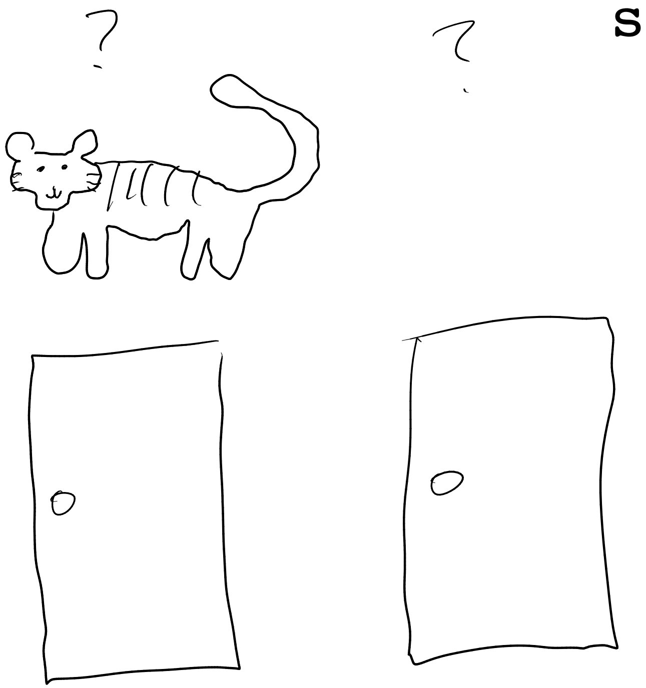
# POMDPs

- We've been living a lie:

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




$A = \{OL, OR, L\}$

85% accurate

-1 Listen

+10 opening ~~good~~ safe door

-100 opening tiger door

Listen until expected value of one of the doors is positive

Use Bayes rule to update tiger probabilities based on observations we get from listening

# Types of Uncertainty

# Types of Uncertainty

**Alleatory**

# Types of Uncertainty

**Alleatory**

# Types of Uncertainty

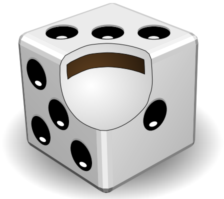
**Alleatory**

**Epistemic (Static)**

# Types of Uncertainty

**Alleatory**

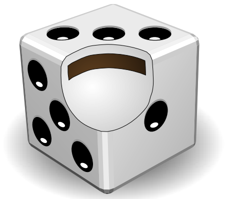

**Epistemic (Static)**



# Types of Uncertainty

**Alleatory**

**Epistemic (Static)**

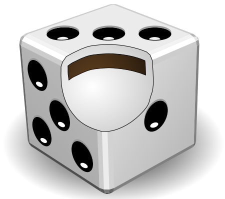


**Epistemic (Dynamic)**

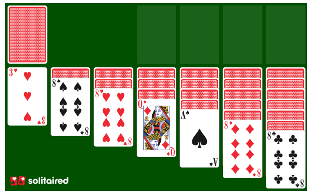
# Types of Uncertainty

**Alleatory**

**Epistemic (Static)**



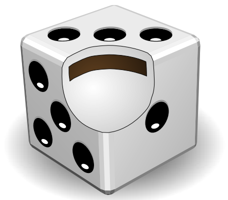
**Epistemic (Dynamic)**



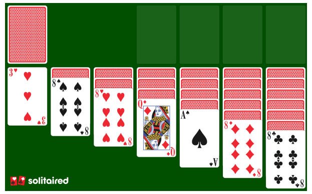
# Types of Uncertainty

**Alleatory**

**Epistemic (Static)**



**Epistemic (Dynamic)**

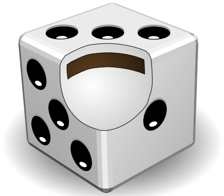



**Interaction**

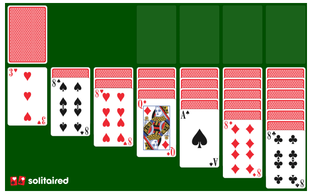
# Types of Uncertainty

**Alleatory**

**Epistemic (Static)**

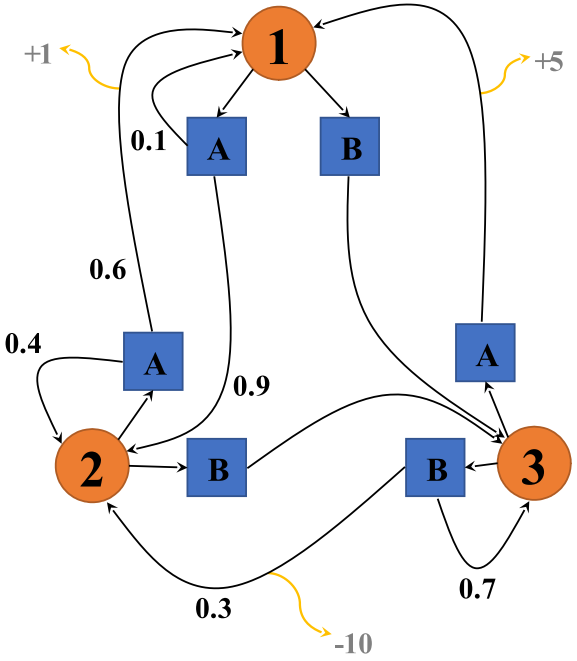


**Epistemic (Dynamic)**



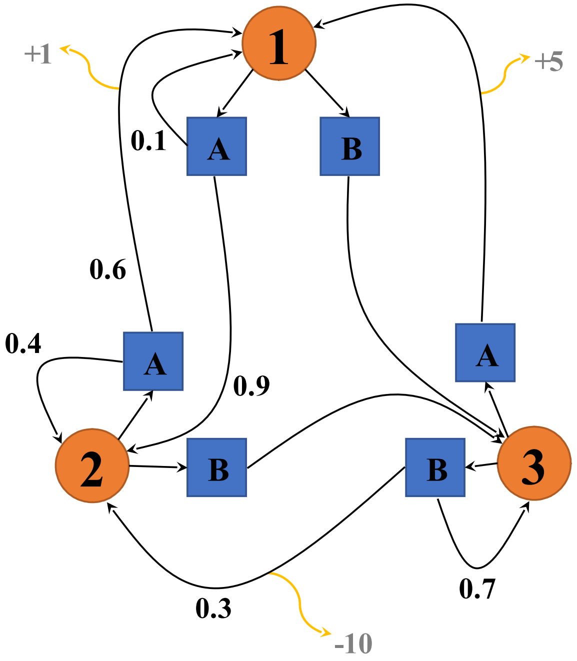

**Interaction**

# Markov Decision Process (MDP)




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

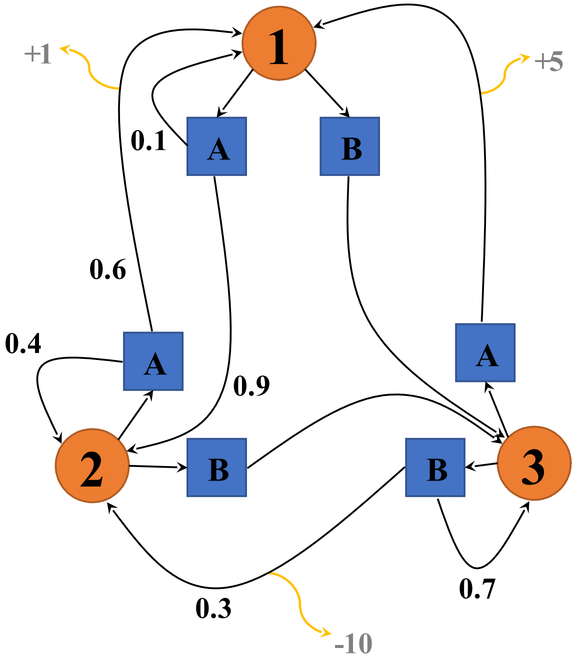
# Markov Decision Process (MDP)




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

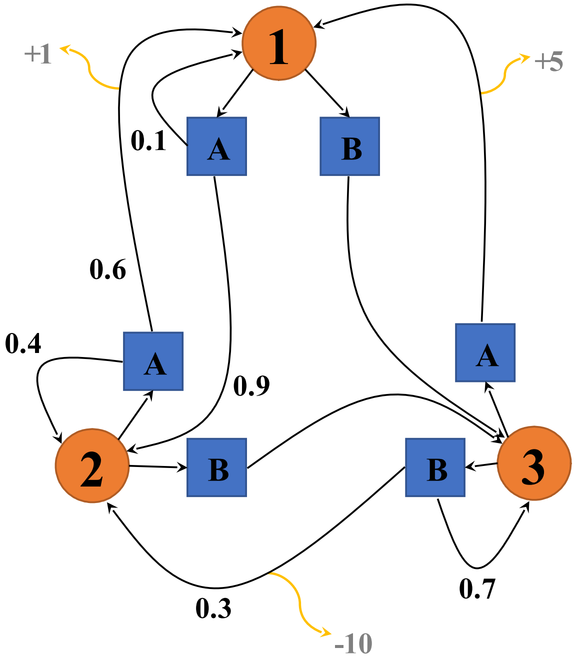
# Markov Decision Process (MDP)

$(S, A, T, R, \gamma)$




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

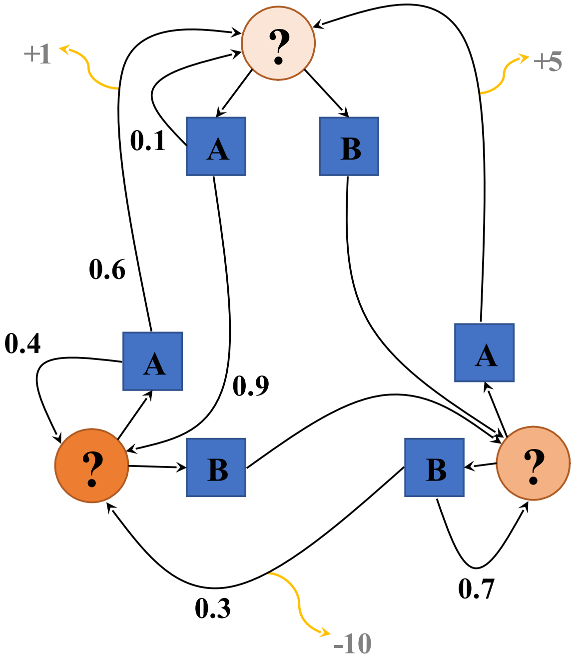
**Alleatory**

RL

Static Epistemic

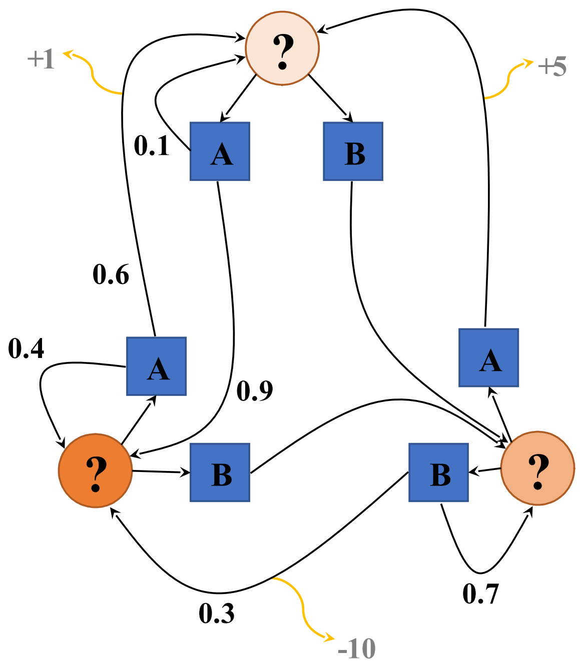
R, T

# Partially Observable Markov Decision Process (POMDP)



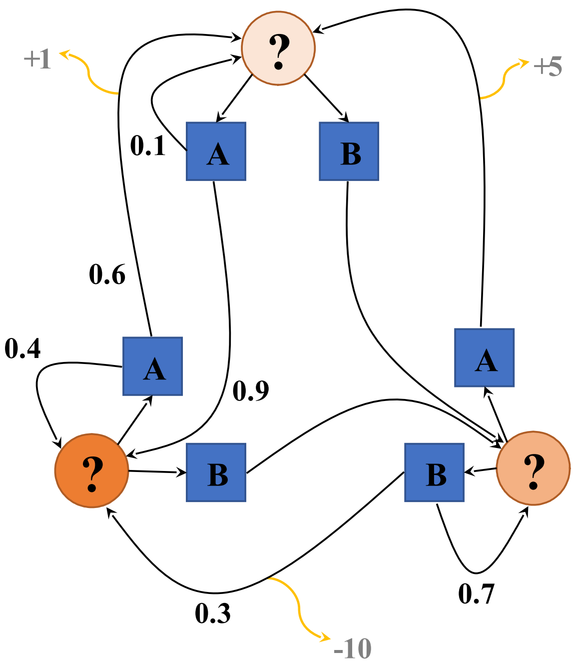

- $\mathcal{S}$ - State space
- $T : \mathcal{S} \times \mathcal{A} \times \mathcal{S} \rightarrow \mathbb{R}$ - Transition probability distribution
- $\mathcal{A}$ - Action space
- $R : \mathcal{S} \times \mathcal{A} \rightarrow \mathbb{R}$ - Reward

# 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
- $T : \mathcal{S} \times \mathcal{A} \times \mathcal{S} \to \mathbb{R}$ - 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

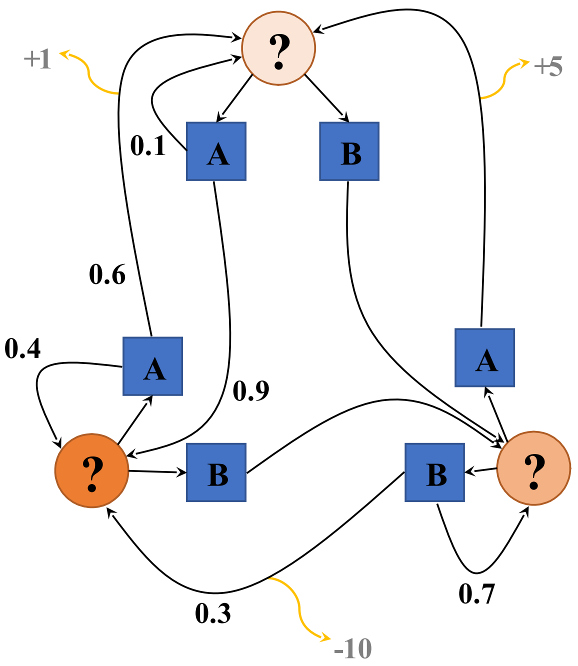
$o \in O$

$O = \{TL, TR\}$

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

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

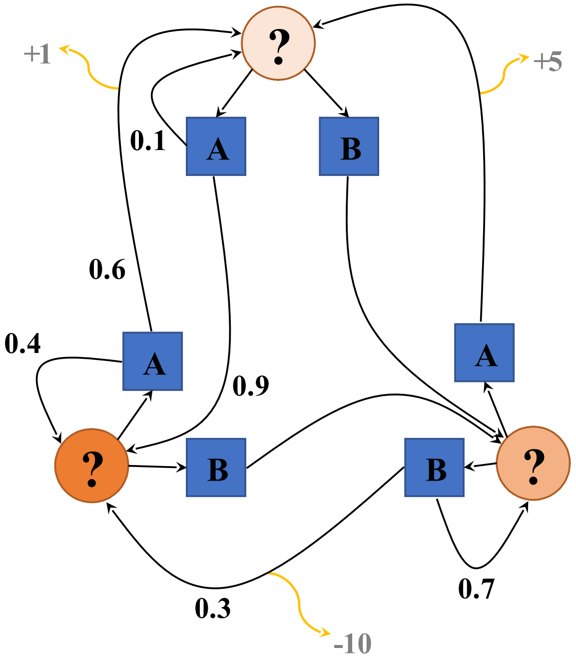
# Partially Observable Markov Decision Process (POMDP)




- $\mathcal{S}$ - 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
- $T : \mathcal{S} \times \mathcal{A} \times \mathcal{S} \rightarrow \mathbb{R}$ - 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

Epistemic (Static)

Epistemic (Dynamic)

# Tiger POMDP Definition

$S = \{TL, TR\}$

$A = \{OL, OR, L\}$

$O = \{TL, TR\}$

$$R(s,a) = \begin{cases} -1 & \text{if } a = L \\ +10 & \text{if } a \in \{OL, OR\} \text{ and } a \neq s \\ -100 & \text{if } a \in \{OL, OR\} \text{ and } a = s \end{cases}$$

$$T(s' \mid s, a) = \begin{cases} 1 & \text{if } a = L,\ s' = s \\ 0 & \text{if } a = L,\ s' \neq s \\ 0.5 & \text{if } a \neq L \end{cases}$$

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

$\gamma = 0.95$

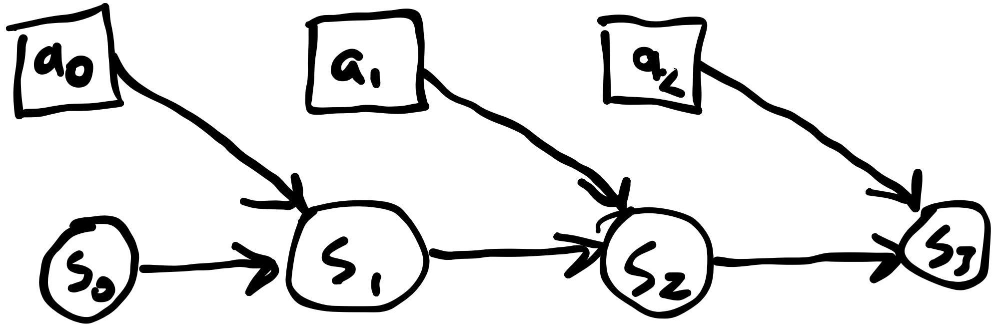
# Hidden Markov Models and Beliefs

Let

# Hidden Markov Models and Beliefs

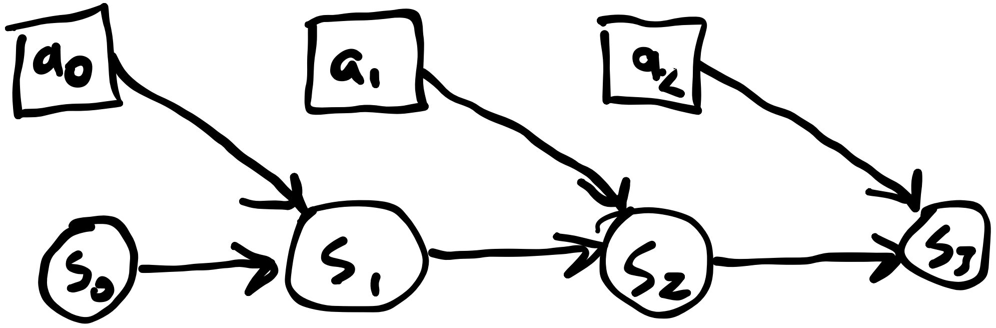

Let

# Hidden Markov Models and Beliefs




Let

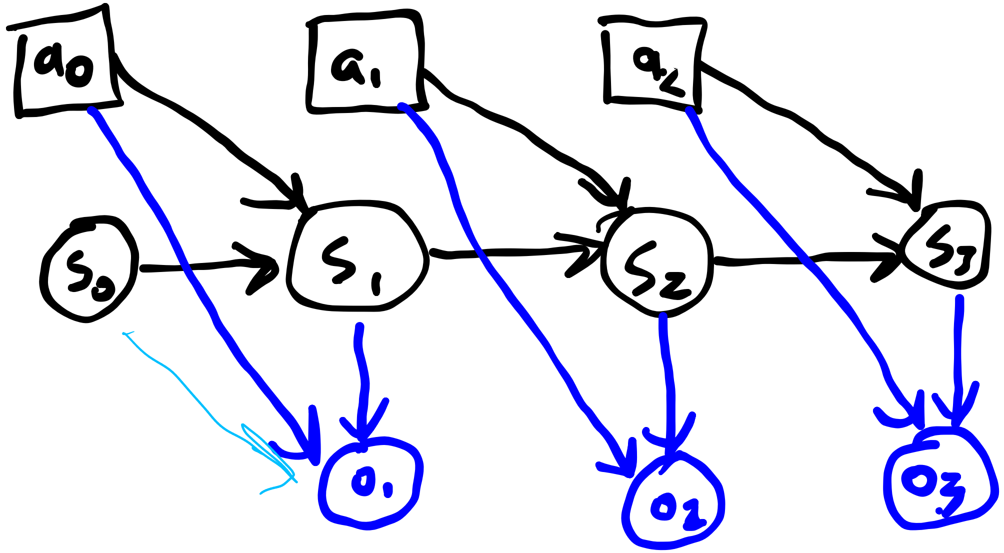
# Hidden Markov Models and Beliefs




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

Let

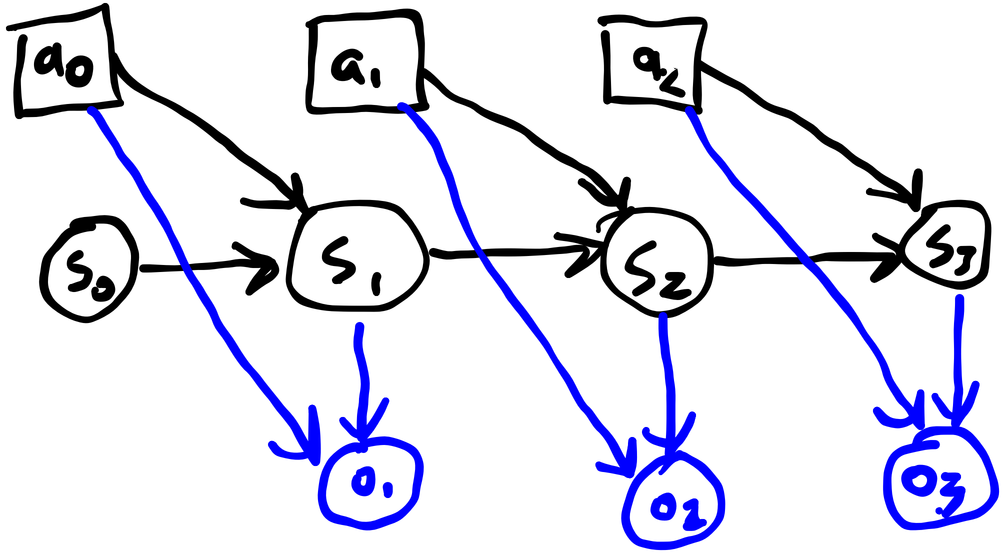
# Hidden Markov Models and Beliefs




$$P(s_{t_1} \mid s_0, a_0, \ldots, s_t, a_t) = T(s_{t+1} \mid s_t, a_t)$$

Let

# Hidden Markov Models and Beliefs

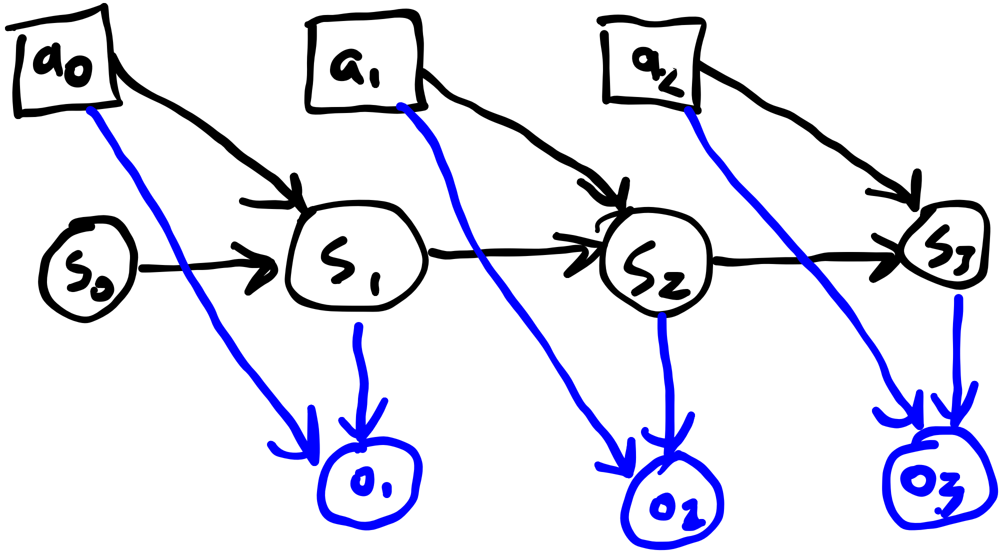



$$P(s_{t_1} \mid s_0, a_0, \ldots, s_t, a_t) = T(s_{t+1} \mid s_t, a_t)$$

$$P(s_{t_1} \mid o_0, a_0, \ldots, o_t, a_t) = P(s_{t+1} \mid a_t, o_{t+t})????$$

Let

# Hidden Markov Models and Beliefs

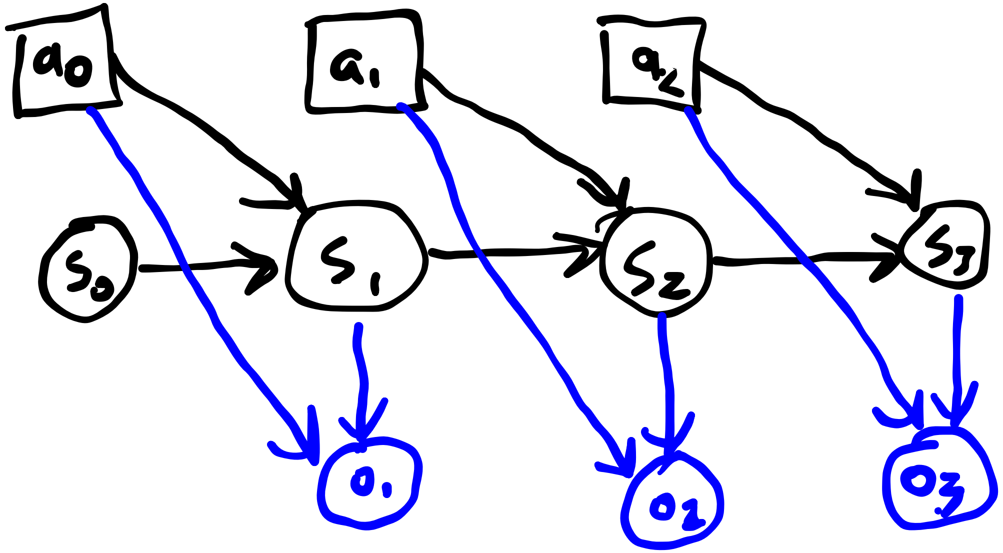



$$P(s_{t_1} \mid s_0, a_0, \ldots, s_t, a_t) = T(s_{t+1} \mid s_t, a_t)$$

$$\cancel{P(s_{t_1} \mid o_0, a_0, \ldots, o_t, a_t) = P(s_{t+1} \mid a_t, o_{t+t})}????$$

Let

# Hidden Markov Models and Beliefs



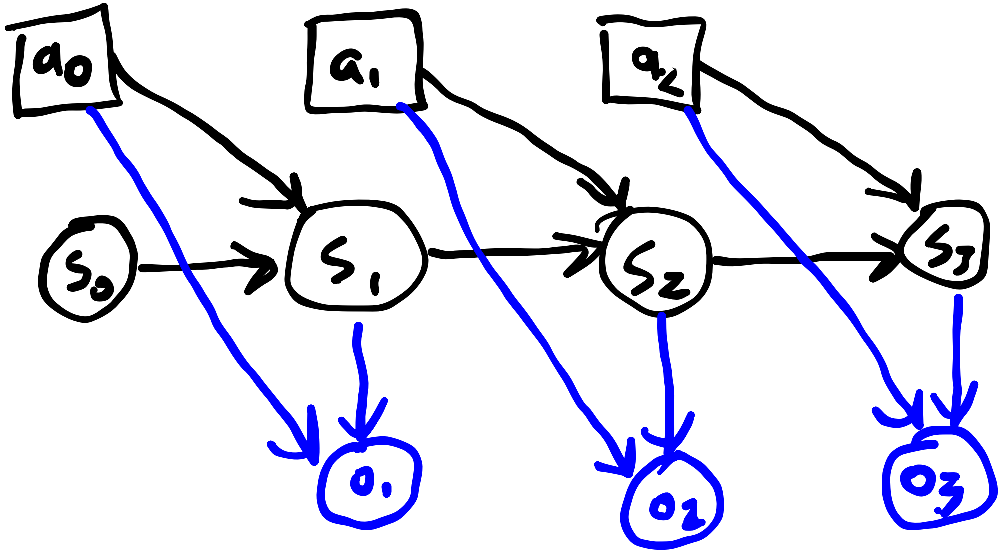

$$P(s_{t_1} \mid s_0, a_0, \ldots, s_t, a_t) = T(s_{t+1} \mid s_t, a_t)$$

$$\cancel{P(s_{t_1} \mid o_0, a_0, \ldots, o_t, a_t) = P(s_{t+1} \mid a_t, o_{t+t})????}$$

Let

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

# Hidden Markov Models and Beliefs



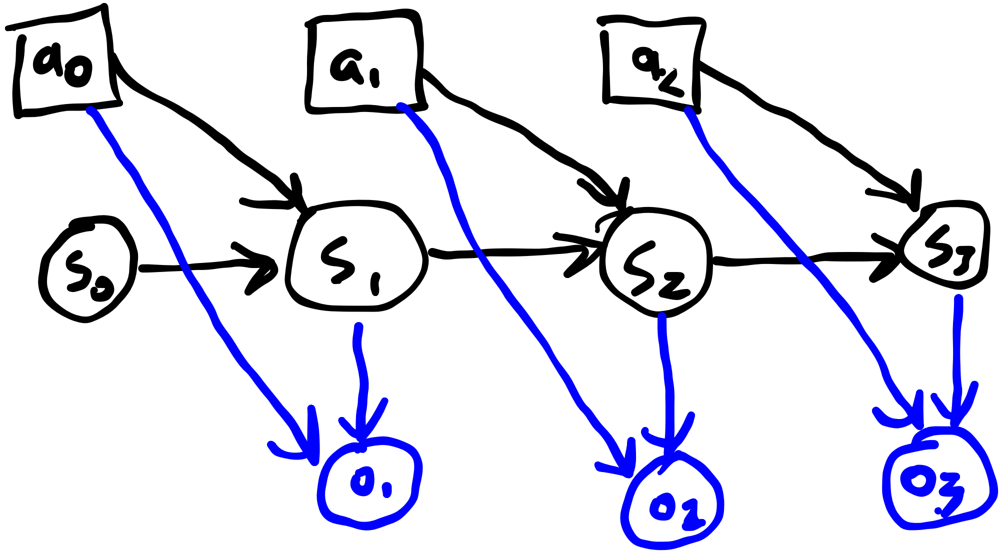

$$P(s_{t_1} \mid s_0, a_0, \ldots, s_t, a_t) = T(s_{t+1} \mid s_t, a_t)$$

$$\cancel{P(s_{t_1} \mid o_0, a_0, \ldots, o_t, a_t) = P(s_{t+1} \mid a_t, o_{t+t})}????$$

Let

- $b_0(s) \equiv P(s_0 = s)$
- $h_t \equiv (b_0, a_0, o_1, a_1, \ldots, a_{t-1}, o_t)$

# Hidden Markov Models and Beliefs




$$P(s_{t_1} \mid s_0, a_0, \ldots, s_t, a_t) = T(s_{t+1} \mid s_t, a_t)$$

$$\cancel{P(s_{t_1} \mid o_0, a_0, \ldots, o_t, a_t) = P(s_{t+1} \mid a_t, o_{t+t})????}$$

Let

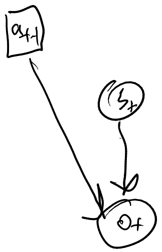
- $b_0(s) \equiv P(s_0 = s)$
- $h_t \equiv (b_0, a_0, o_1, a_1, \ldots, a_{t-1}, o_t)$
- $b_t(s) \equiv P(s_t = s \mid h_t)$

# Bayesian Belief Updates

$$b_t = \tau(b_{t-1}, a_{t-1}, o_t) \qquad h_t = (h_{t-1}, a_{t-1}, o_t)$$

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

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




$$o_t \perp h_t \mid s_t, a_{t-1}$$

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

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

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

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

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

# Filtering Loop

$b = b_0$

loop

receiving $o$

for $s' \in S$

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

$b' \leftarrow b' / \sum_{s'} b'(s')$

$b \leftarrow b'$

# Tiger Example