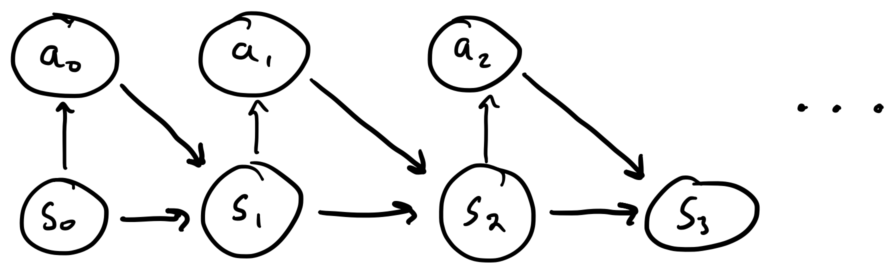
The Markov Property

$$p(s_t \mid s_{t-1}, a_{t-1}, \ldots, s_0, a_0) = p(s_t \mid s_{t-1}, a_{t-1})$$

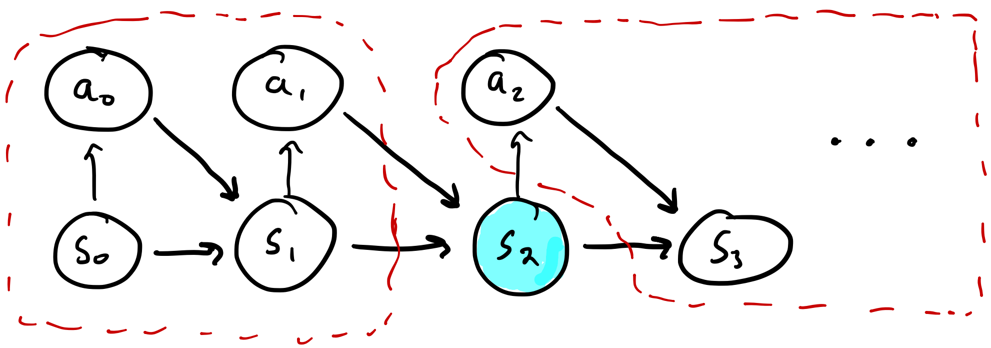
The Markov Property

$$p(s_t \mid s_{t-1}, a_{t-1}, \ldots, s_0, a_0) = p(s_t \mid s_{t-1}, a_{t-1})$$



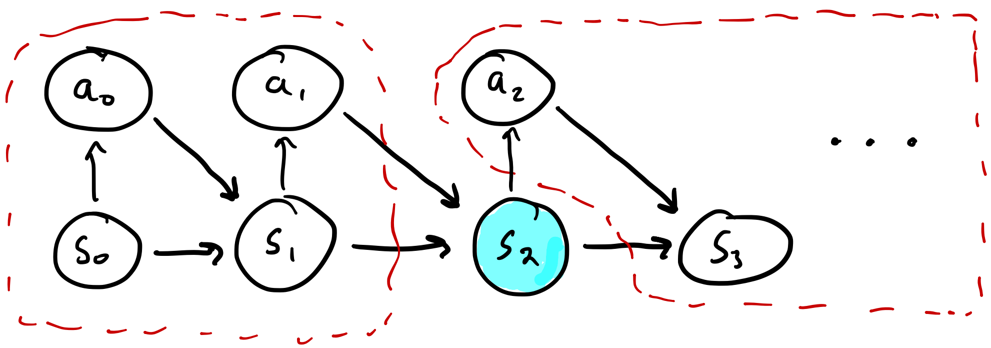
The Markov Property

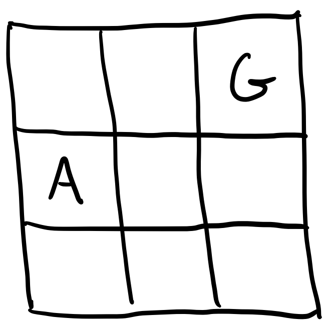
$$p(s_t \mid s_{t-1}, a_{t-1}, \dots, s_0, a_0) = p(s_t \mid s_{t-1}, a_{t-1})$$



The Markov Property

$$p(s_t \mid s_{t-1}, a_{t-1}, \dots, s_0, a_0) = p(s_t \mid s_{t-1}, a_{t-1})$$





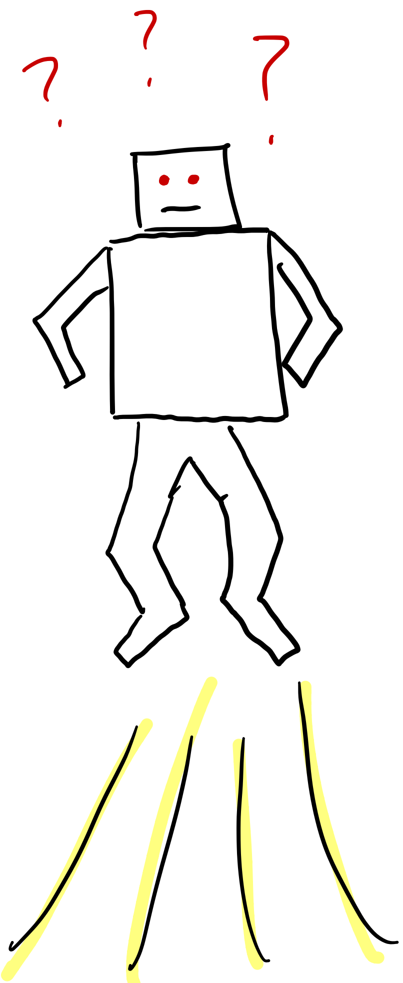
State : $\langle x, y \rangle$

Actions : ↑ ↓ ← →

Once A is visited, all action effects reverse

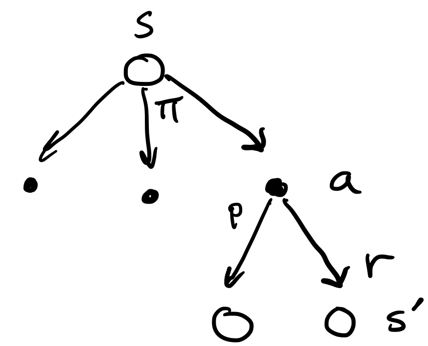
Is this Markovian?

Reward Specification



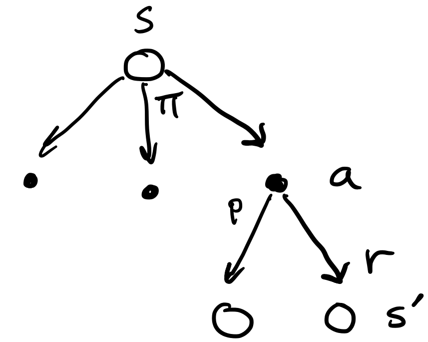
The Bellman Equation

$$V_\pi(s) = \sum_a \pi(a|s) \sum_{s',r} p(s',r|s,a)\left[r + \gamma V_\pi(s')\right]$$



The Bellman Equation

$$V_\pi(s) = \sum_a \pi(a|s) \sum_{s',r} p(s',r|s,a)\left[r + \gamma V_\pi(s')\right]$$



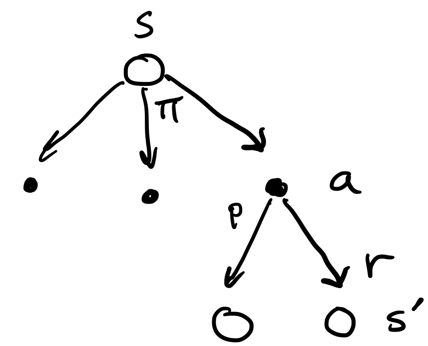
$$V_*(s) = \max_a \sum_{s',r} p(s',r|s,a)\left[r + \gamma V_*(s')\right]$$

$$q_*(s,a) = \sum_{s',r} p(s',r|s,a)\left[r + \gamma \max_{a'} q_*(s',a')\right]$$

$$V_*(s) = \max_a q_*(s,a)$$

The Bellman Equation

$$v_\pi(s) = \sum_a \pi(a|s) \sum_{s',r} p(s',r|s,a)\left[r + \gamma v_\pi(s')\right]$$

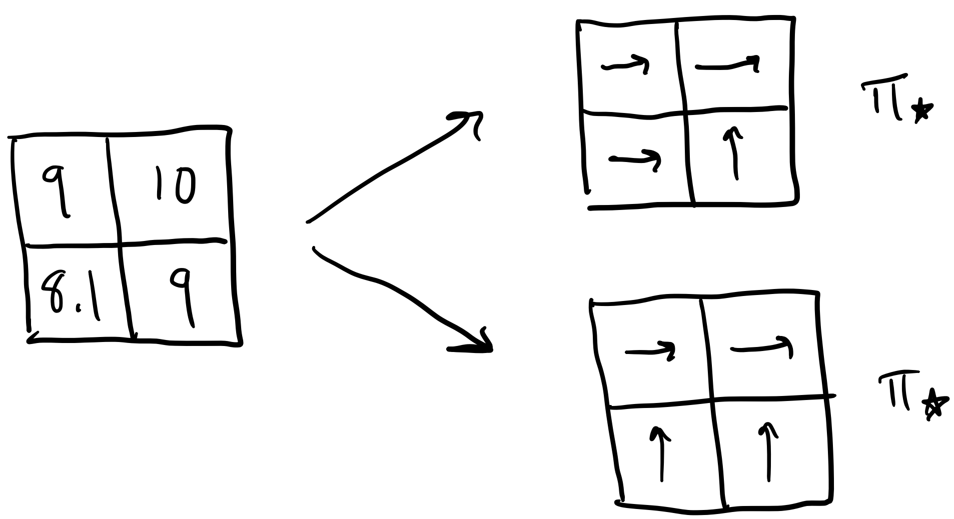


$$v_*(s) = \max_a \sum_{s',r} p(s',r|s,a)\left[r + \gamma v_*(s')\right]$$

$$q_*(s,a) = \sum_{s',r} p(s',r|s,a)\left[r + \gamma \max_{a'} q_*(s',a')\right]$$

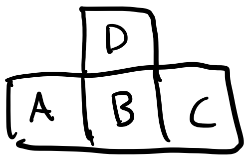
$$v_*(s) = \max_a q_*(s,a)$$

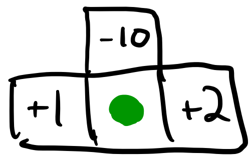
v_*



v_* is unique, but π_* is not!

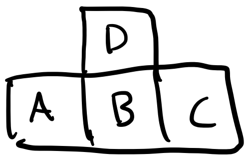
Optimal action selection

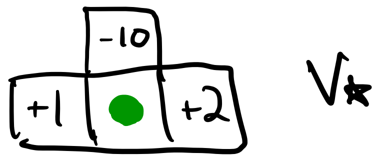




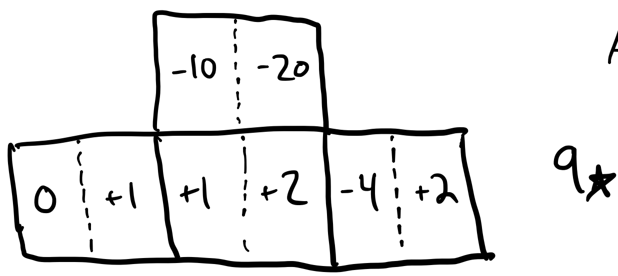
V^*

Optimal action selection





Actions: a_1 + a_2



POMDPs

$O: p(o \mid s, a)$

$b(s) = p(s_t \mid o_1, a_1, \ldots, o_t, a_t)$

Belief update:

$$b'(s') \leftarrow p(o \mid s', a) \sum_s p(s' \mid s, a)\, b(s)$$

POMDPs

$O: p(o | s, a)$

$b(s) = p(s_t | o_1, a_1, \ldots, o_t, a_t)$

Belief update:

$$b'(s') \leftarrow p(o | s', a) \sum_s p(s' | s, a)\, b(s)$$

$|S|$ states, 10 probability buckets. How many belief states?

POMDPs

O: $p(o \mid s, a)$

$b(s) = p(s_t \mid o_1, a_1, \ldots, o_t, a_t)$

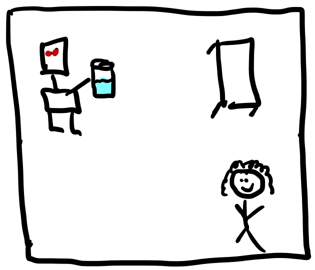
Belief update:

$$b'(s') \leftarrow p(o \mid s', a) \sum_{s} p(s' \mid s, a)\, b(s)$$

$|S|$ states, 10 probability buckets. How many belief states?

$\longrightarrow 10^{|S|}$

Minimal State Representations + Abstraction



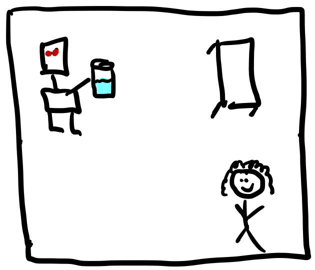
State : ⟨x, y⟩ loc of robot + Jill

⟨x, y⟩ loc of table

⟨x, y⟩ loc of car keys

Goal: Deliver water to Jill

Minimal State Representations + Abstraction

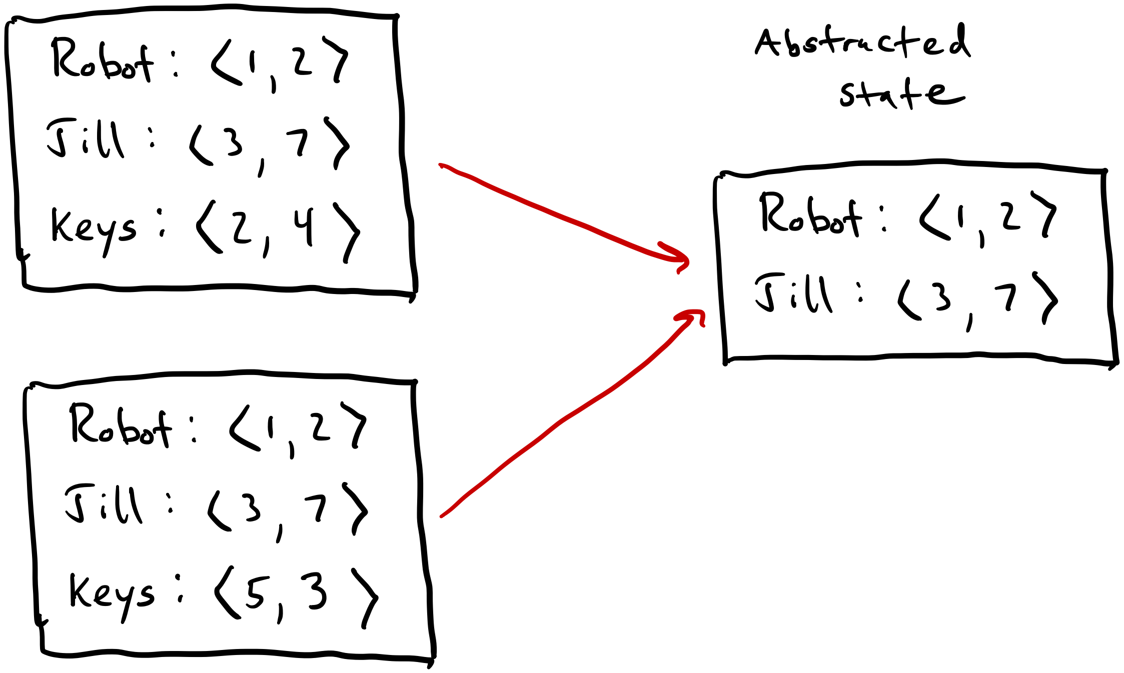


State : ⟨x, y⟩ loc of robot + Jill

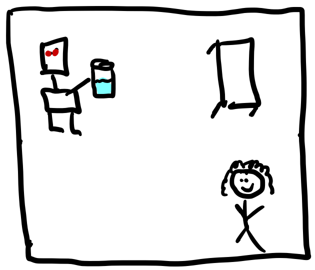
⟨x, y⟩ loc of table

⟨x, y⟩ loc of car keys

Goal: Deliver water to Jill



Minimal State Representations + Abstraction



State: ⟨x, y⟩ loc of robot + Jill
⟨x, y⟩ loc of table
⟨x, y⟩ loc of car keys

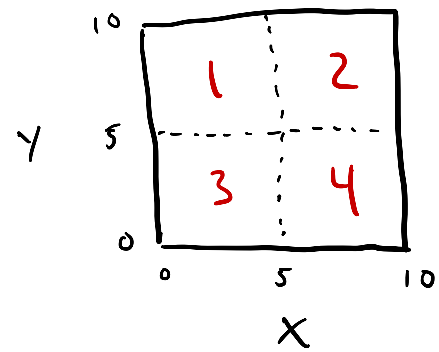
Goal: Deliver water to Jill

Robot: ⟨1, 2⟩
Jill: ⟨3, 7⟩
Keys: ⟨2, 4⟩

Robot: ⟨1, 2⟩
Jill: ⟨3, 7⟩
Keys: ⟨5, 3⟩

→ Abstracted state

Robot: ⟨1, 2⟩
Jill: ⟨3, 7⟩



x < 5, y < 5 → 1

x < 5, y ≥ 5 → 3

⋮