TÉCNICO LISBOA

Planning, Learning and Intelligent Decision Making

Lecture 4

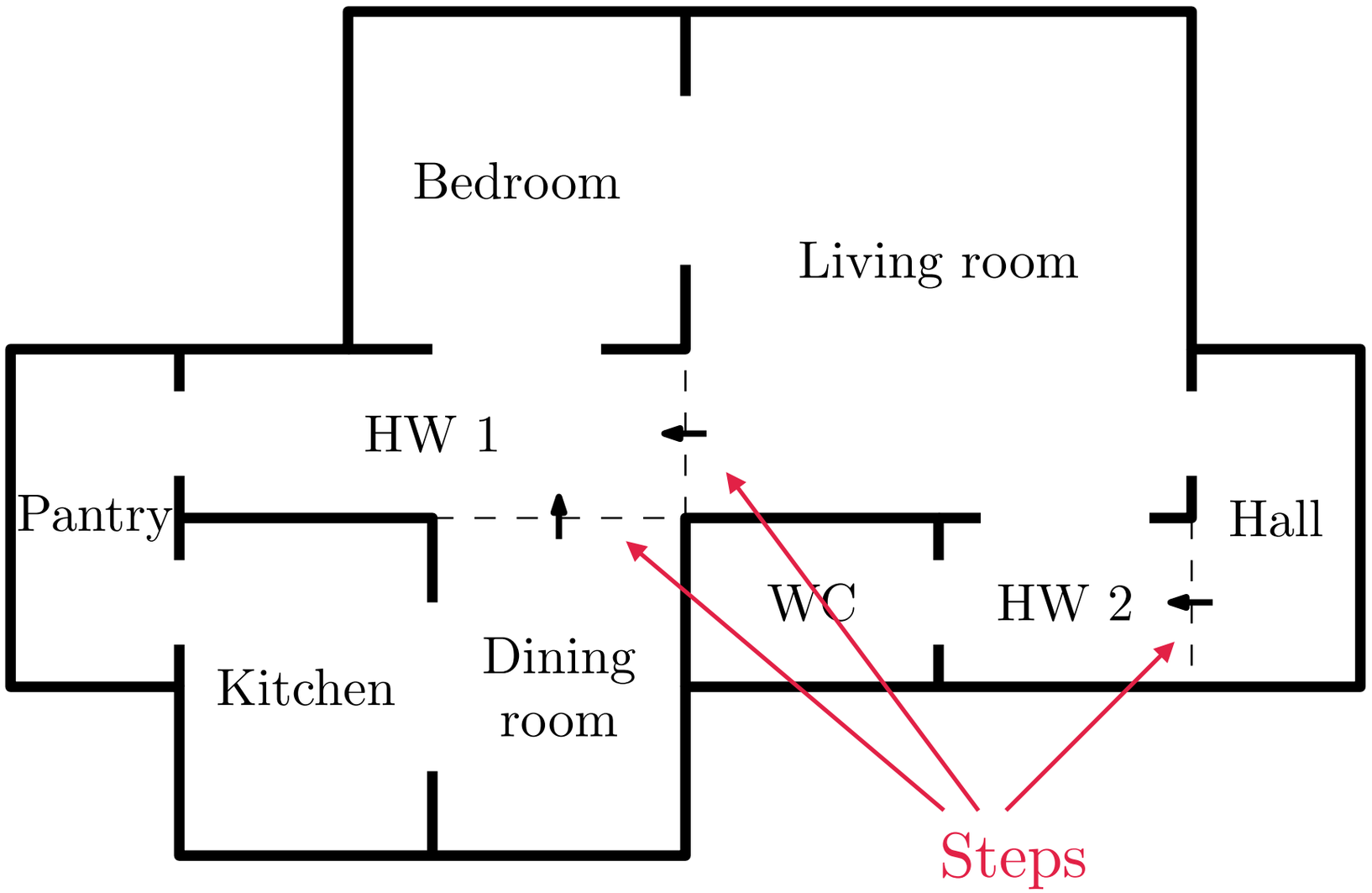
PADInt 2024

Sequential decision problems

The household robot

Household robot

- Consider the household



Household robot

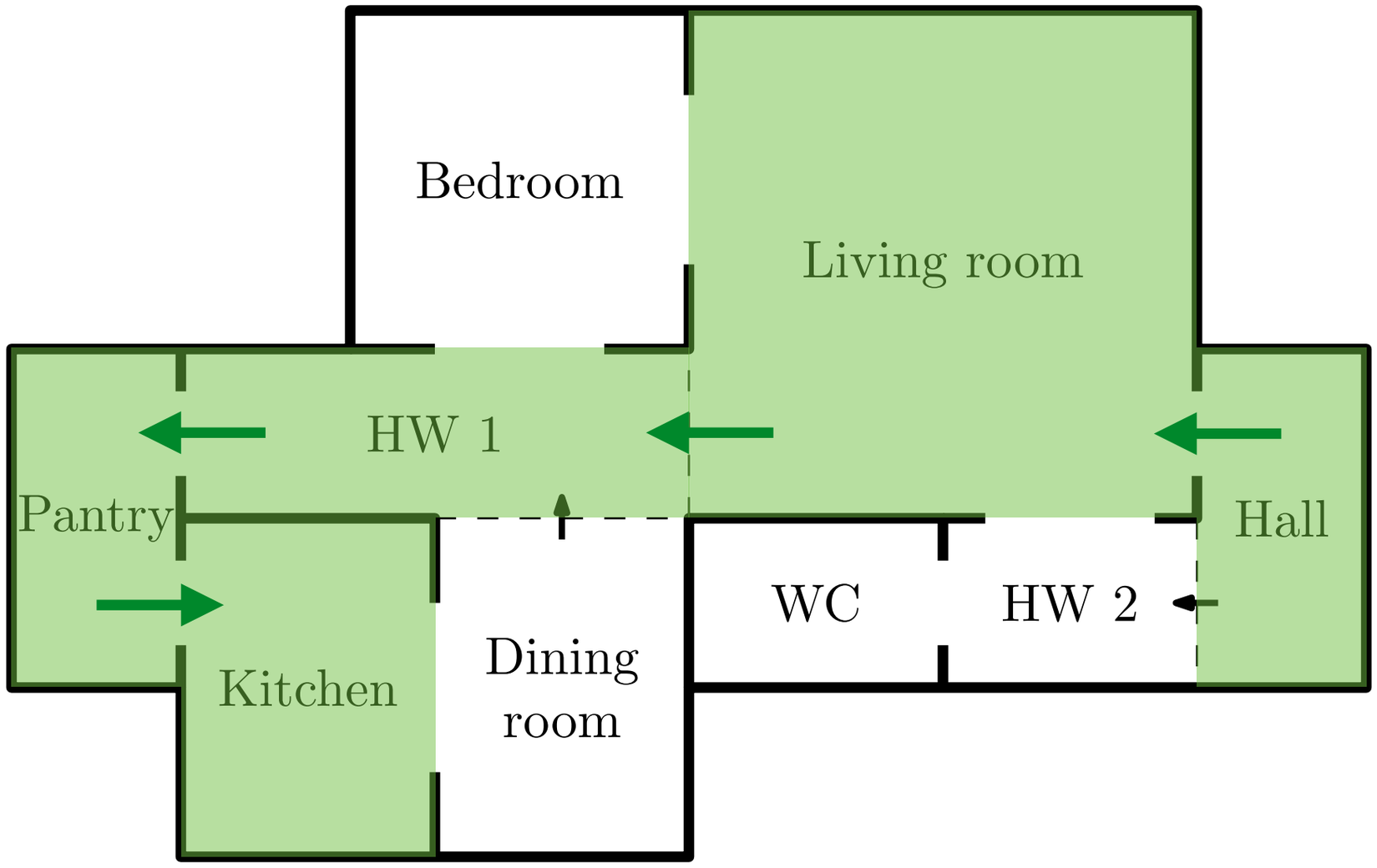
- Robot moves in the environment, assisting human users
- When at the Hall, receives a request from the Kitchen

A single decision

- We can model the problem as a single decision
- Robot must select among several paths

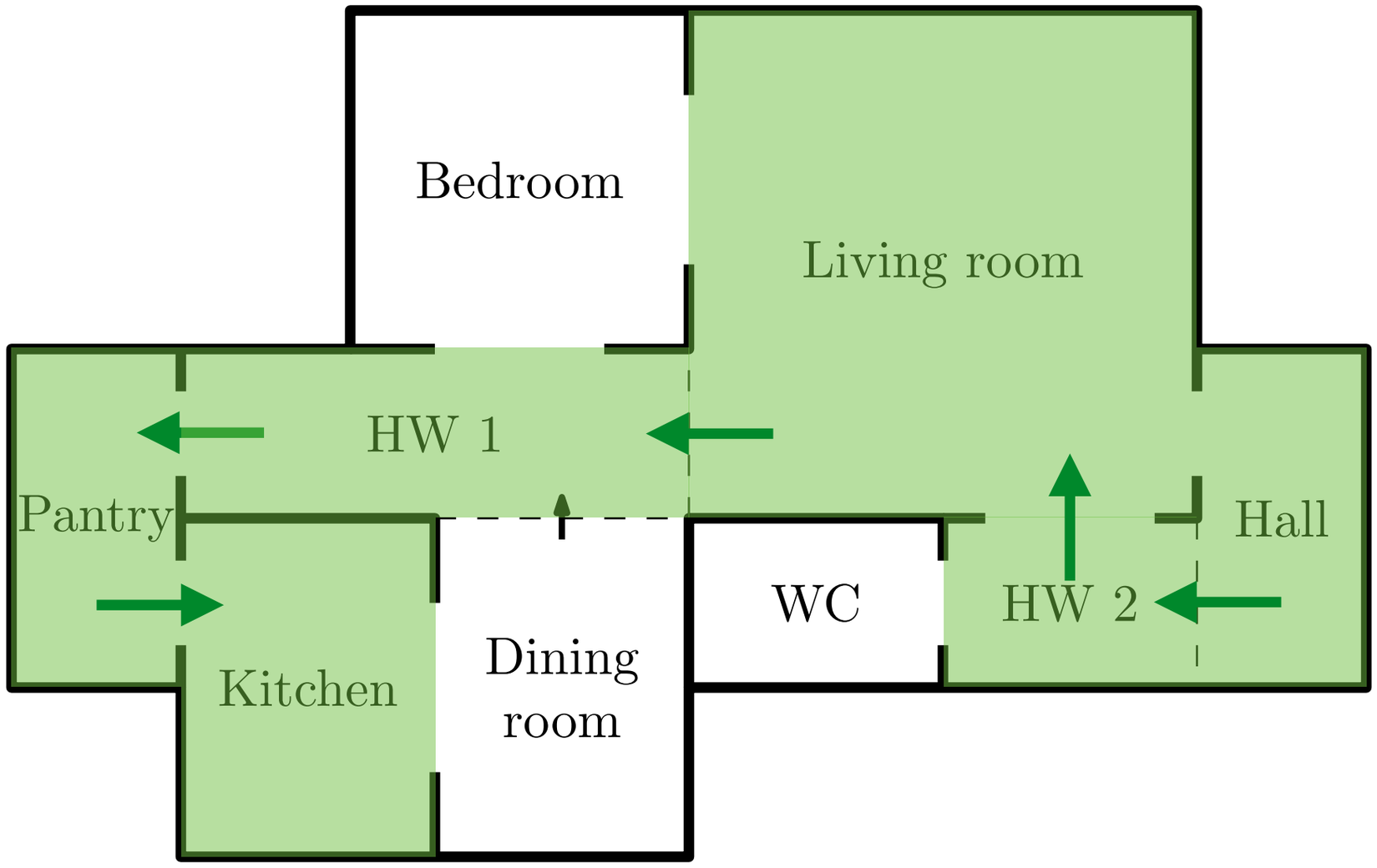
Path A

Hall → Living room → Hallway 1 → Pantry → Kitchen



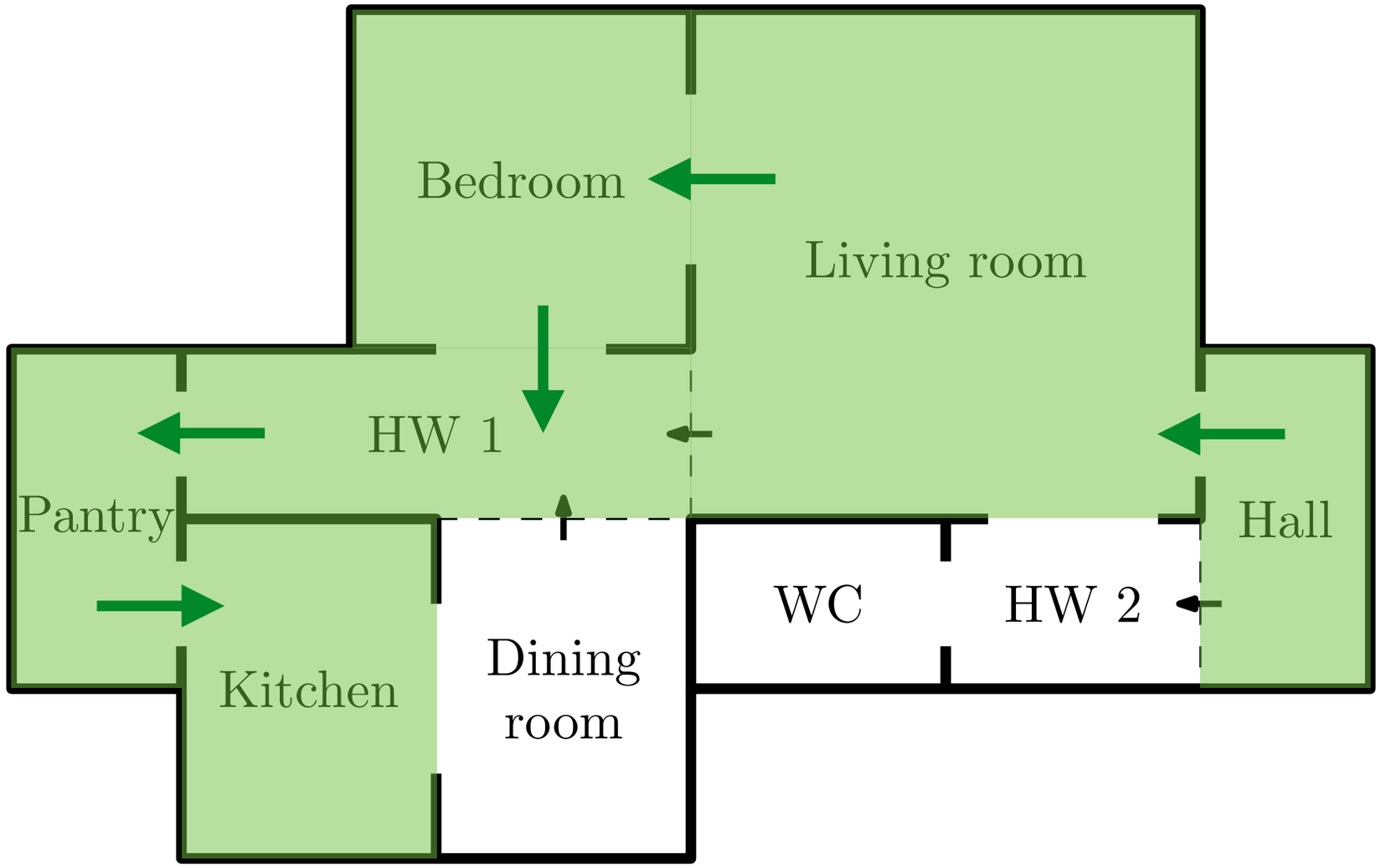
Path B

Hall → Hallway 2 → Living room → Hallway 1 → Pantry → Kitchen



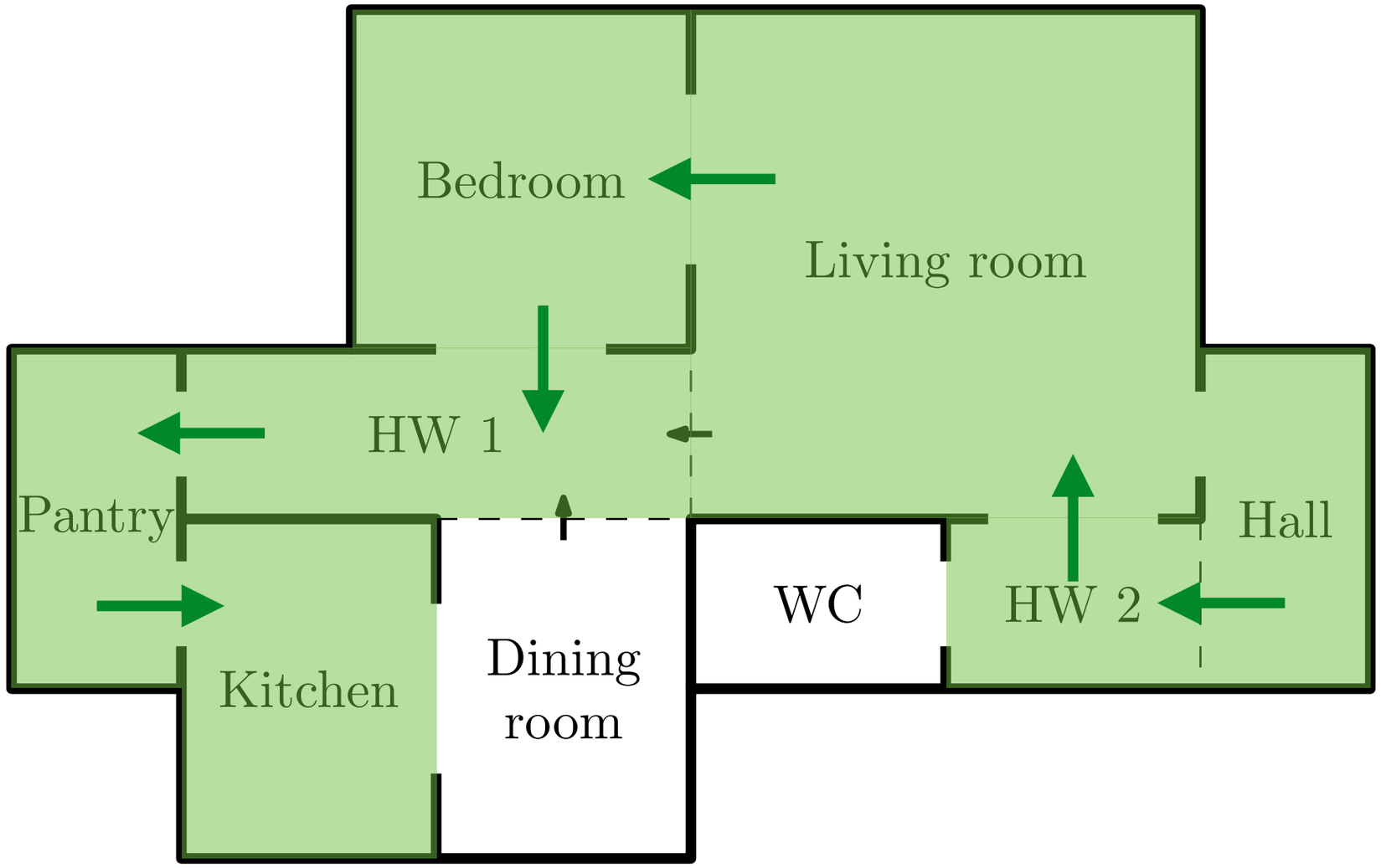
Path C

Hall → Living room → Bedroom → Hallway 1 → Pantry → Kitchen



Path D

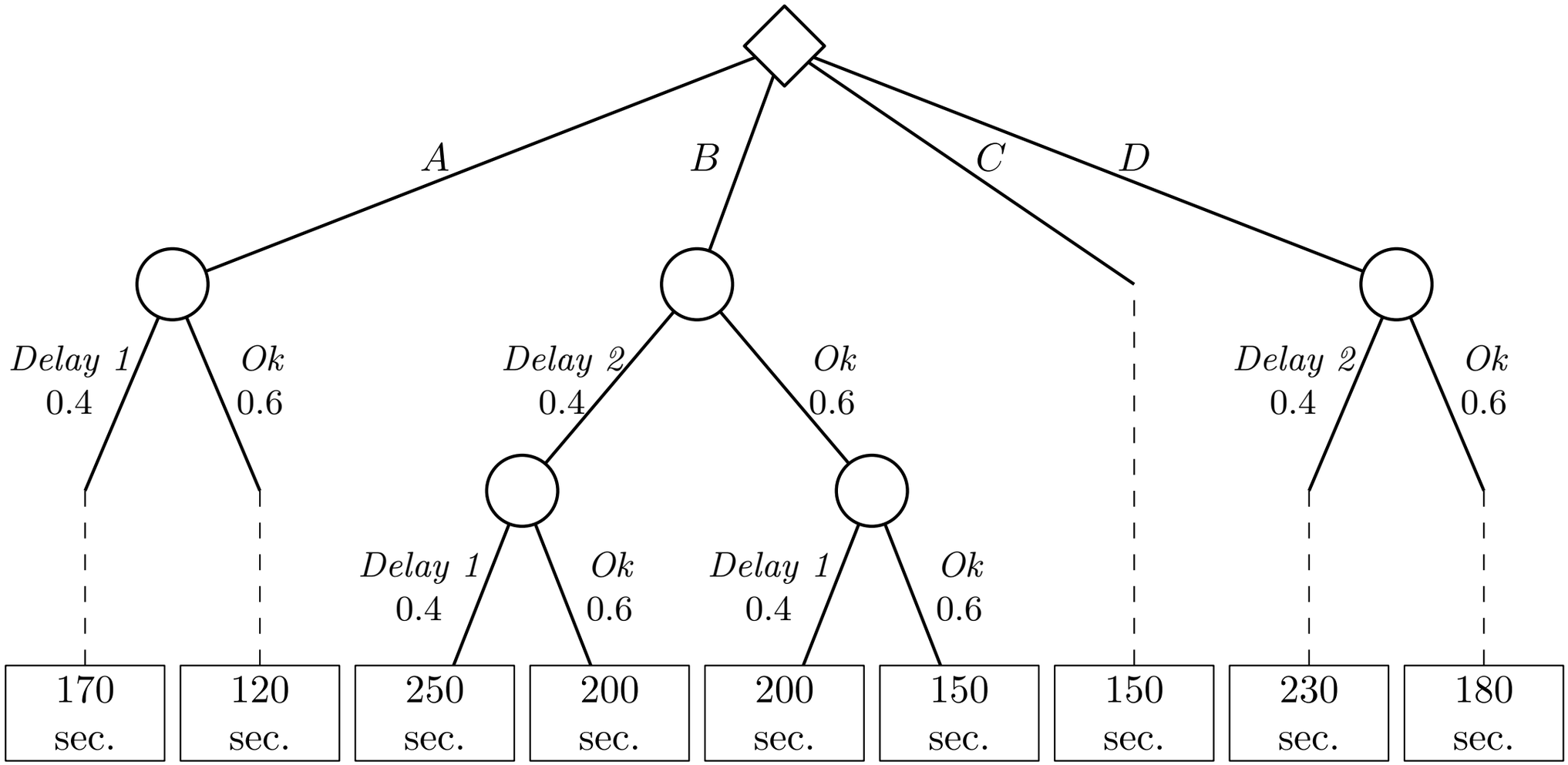
Hall → Hallway 2 → Living room → Bedroom → Hallway 1 → Pantry → Kitchen



A single decision

- Moving between two rooms takes around 30 seconds
- In steps, with a probability 0.4, it takes around 80 seconds

Decision tree



$Q(\text{A}) = \text{-}140$ $Q(\text{B}) = \text{-}190$ $Q(\text{C}) = \text{-}150$ $Q(\text{D}) = \text{-}200$

Observation n. 1

Costs vs. utility

- In many problems, we use **negative utilities**
- E.g., the student problem:
  - We used negative utilities to express loss in grades
- E.g., the robot problem:
  - We used negative utilities to express loss in time

Negative utility = cost

The notion of "goal"

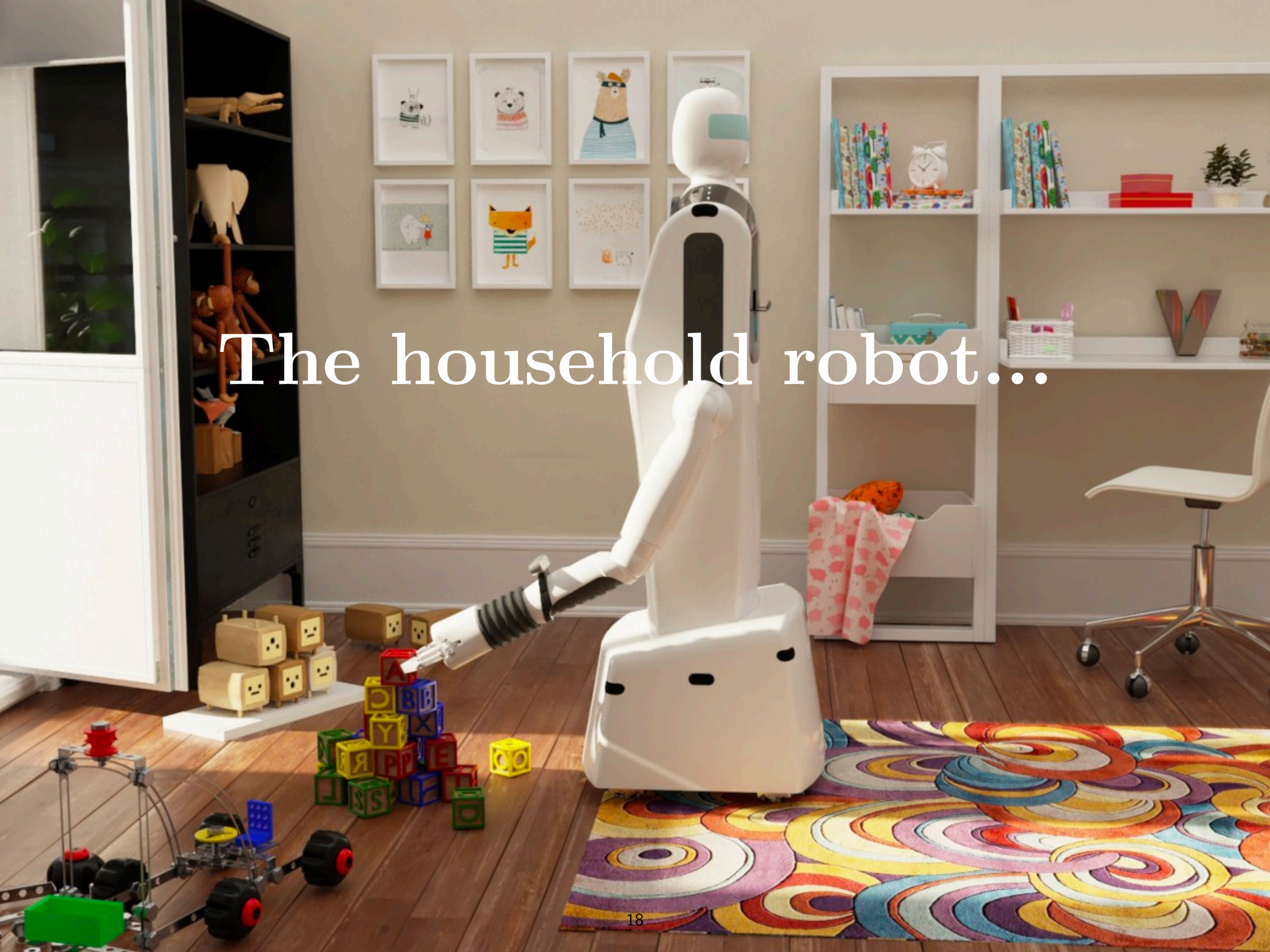
- Cost (or utility) implicitly express the **goal** of the decision maker
- We are the **designers** of such goal: we provide the decision-maker/ agent with a cost (or utility)
- The cost expresses **our own preferences** (as designers) regarding the behavior of the agent

Observation n. 2

Sequential problems

- Sequential problems (like the household robot) are poorly modeled by listing all sequences of actions

Sequence of decisions


The household robot...

Household robot

- Robot moves in the environment, assisting human users
- When at the Hall, receives a request from the Kitchen

One "movement", one decision

Sequence of decisions

- We use expected utility (only tool so far)

- At each step, what are the possible outcomes?

$$\mathcal{X} = \{\text{B, D, H, H1, H2, K, L, P, W}\}$$

Same symbol as before

Sequence of decisions

- We use expected utility (only tool so far)

- At each step, the robot has available a set of actions:

$$\mathcal{A} = \{\text{U(p), D(own), L(eft), R(ight), S(tay)}\}$$

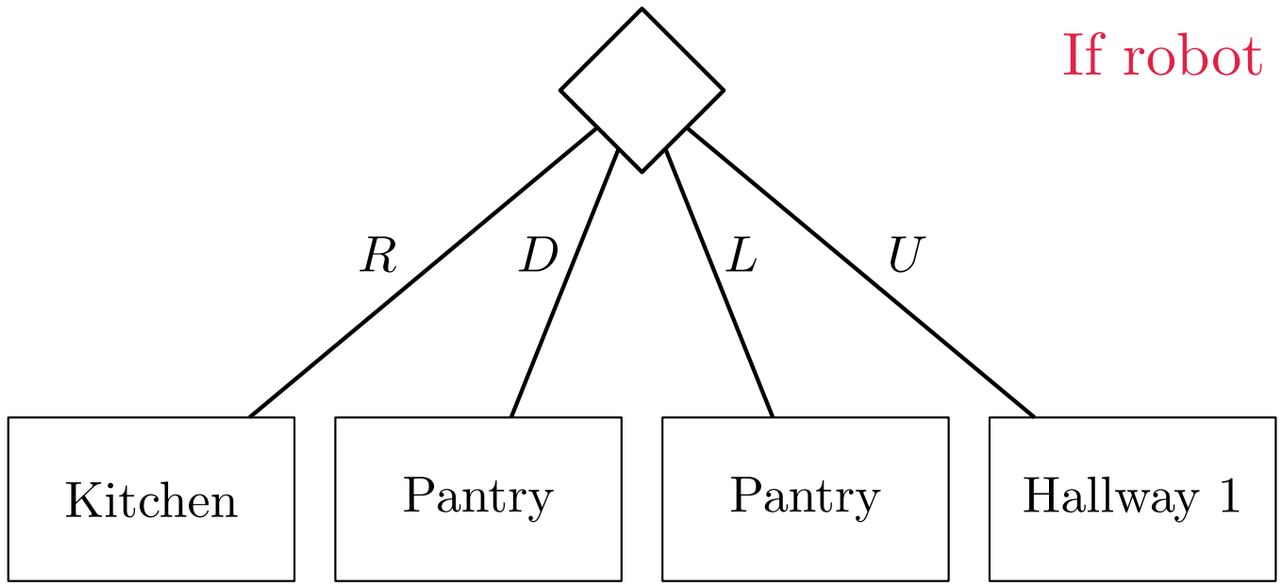
Same symbol as before

Sequence of decisions

- Motions across a step fail with probability 0.4
  - The probability of different outcomes depends on the position of the robot

Sequence of decisions

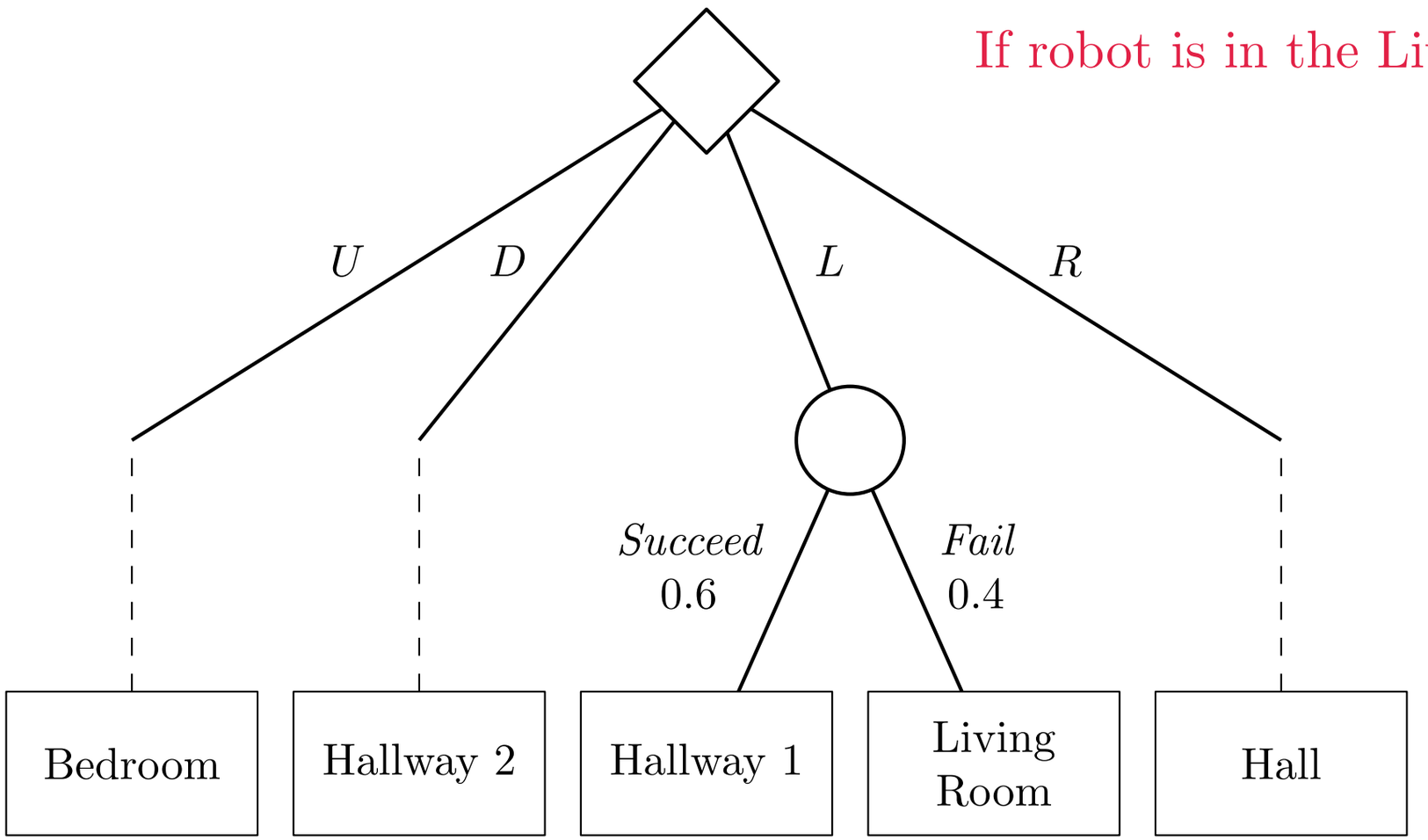
- The probability of different outcomes depends on the position of the robot



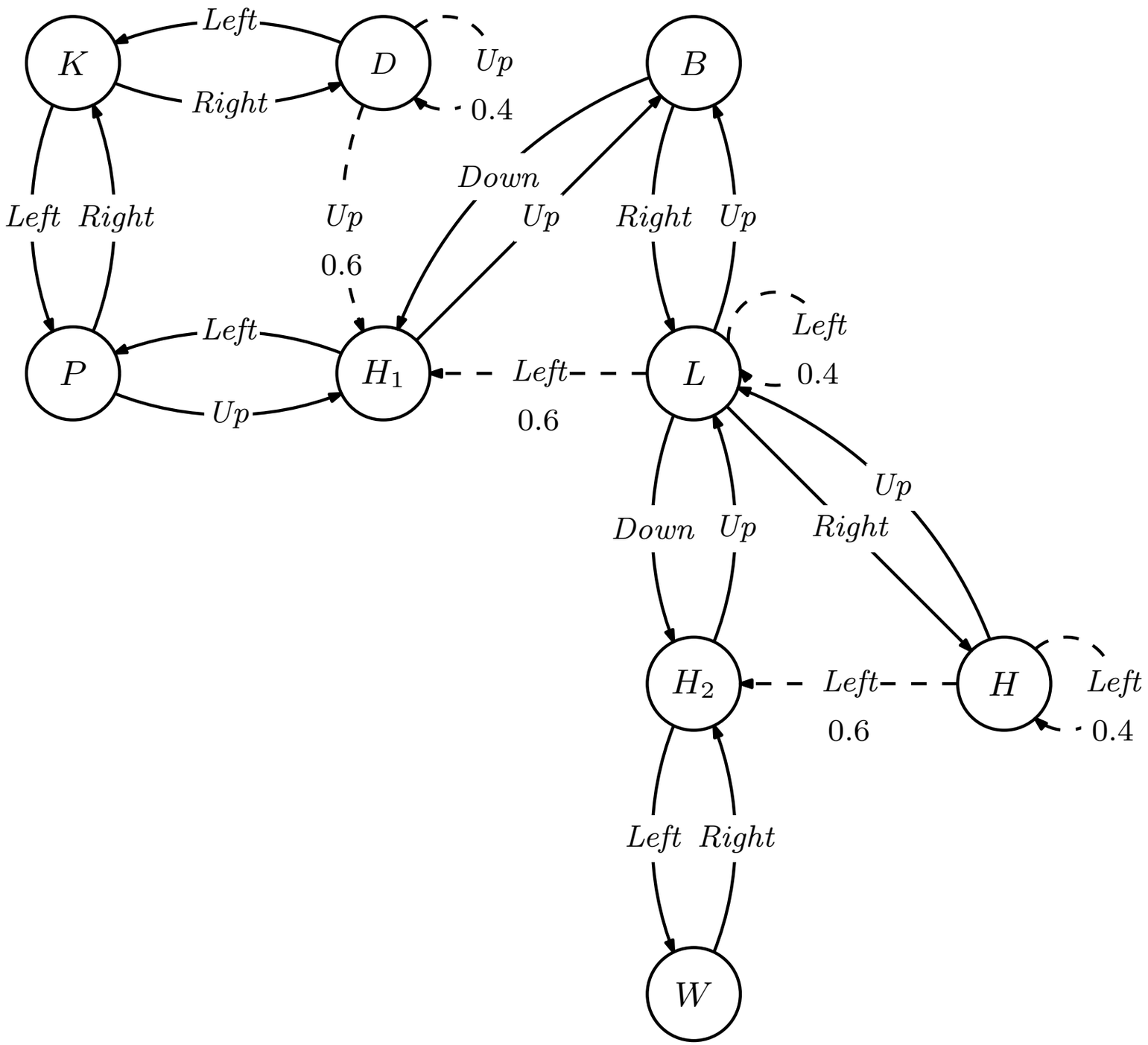
If robot is in the Pantry...

Sequence of decisions

- The probability of different outcomes depends on the position of the robot



Movement of the robot



Sequence of decisions

- The probability of different outcomes depends on the position of the robot
  - We write $\mathbf{P}(x' \mid x, a)$ to denote the probability of outcome x' when robot is in position x and makes action a

What about cost?

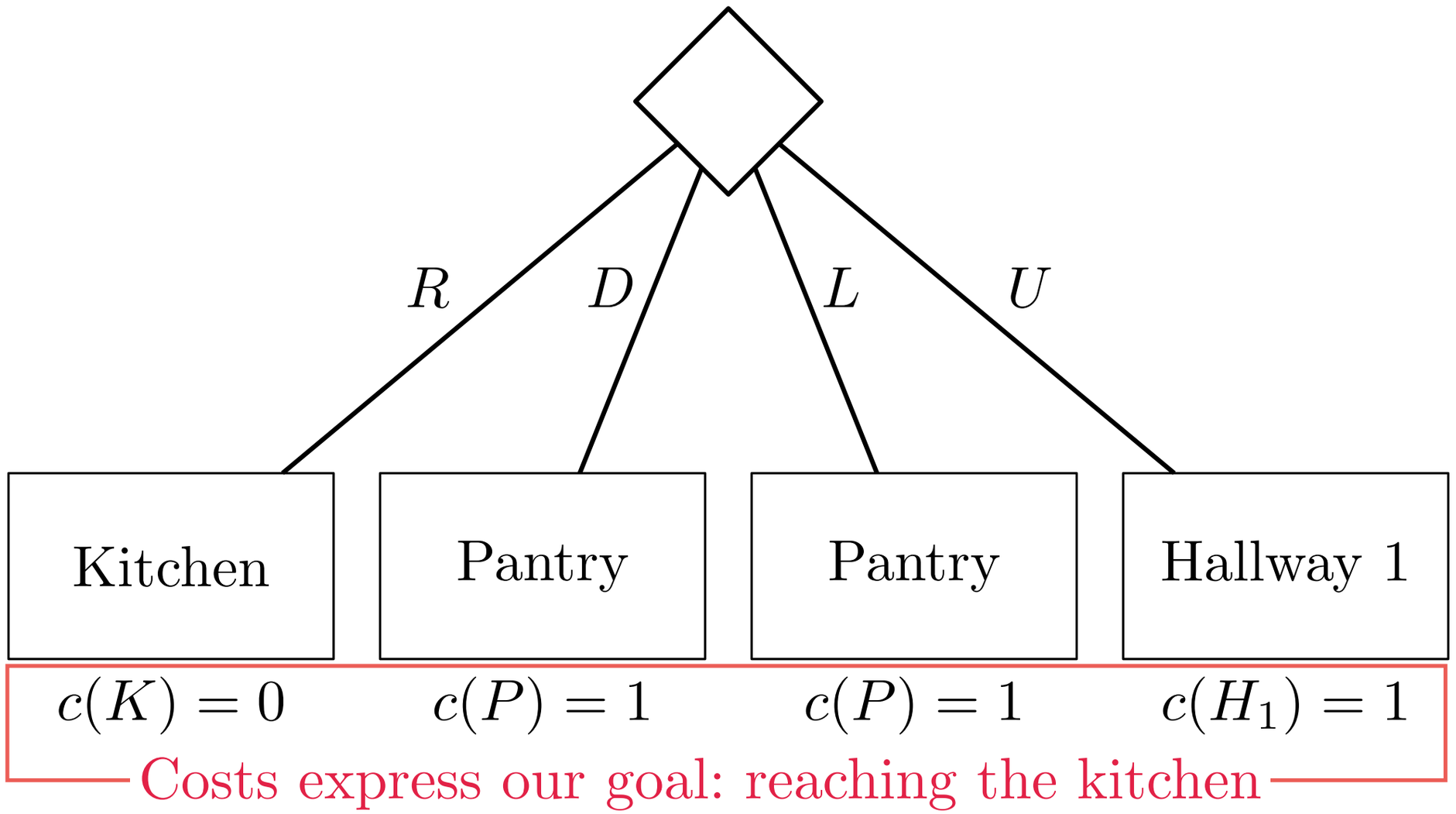
- What should the cost be?
  - Should depend on the position of the robot
  - Suggestion:
    - 1 whenever not in kitchen

Why?

... however...

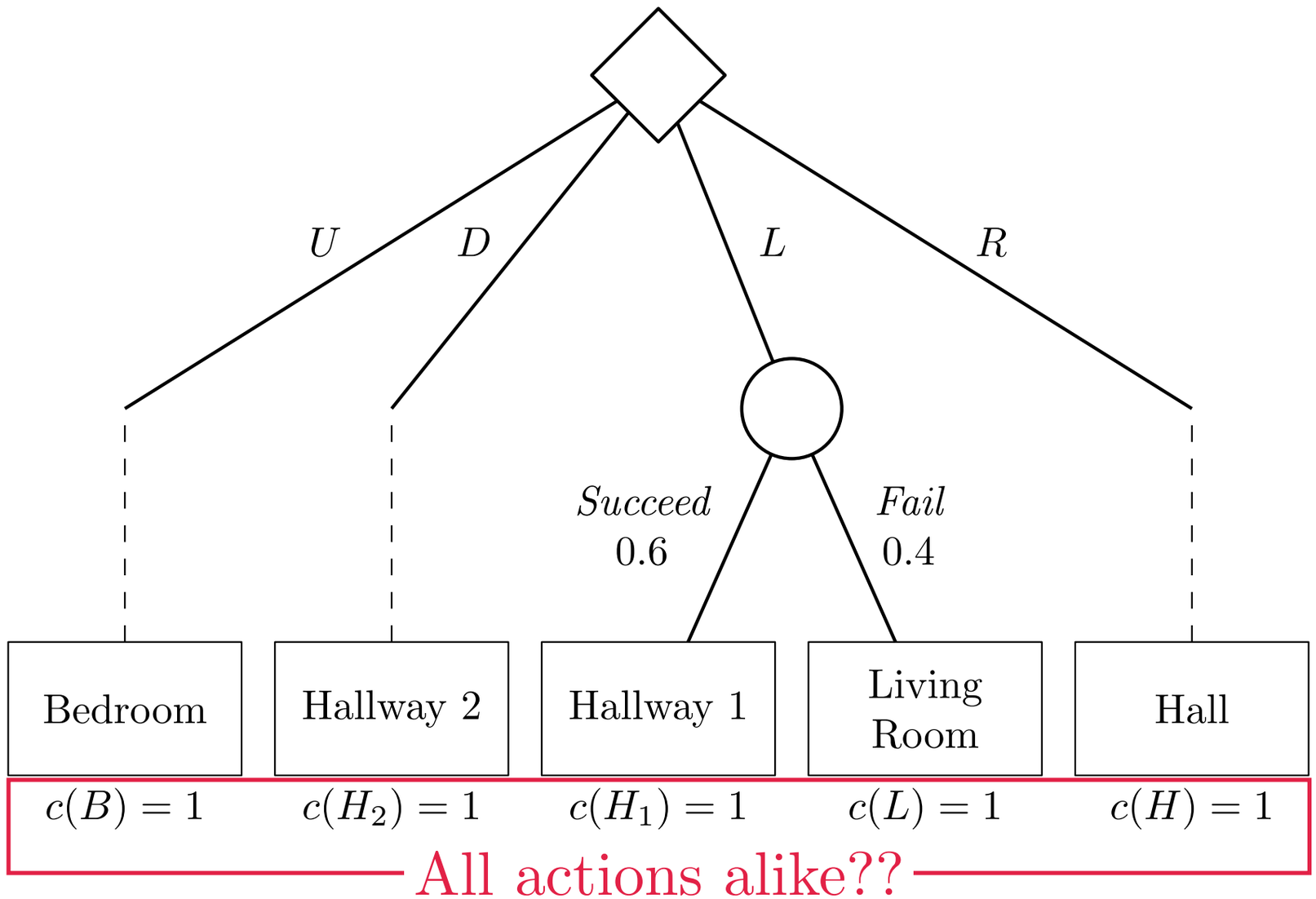
...

- If the robot is at the Pantry...



...

- If the robot is in the Living room...



Immediate cost

- The cost used evaluates **instantaneously** the position/action of the robot
- We will call it the **immediate cost**

How do we choose?

- The current choice influences **future choices**
- Cost does not provide **long-term information** on the actions

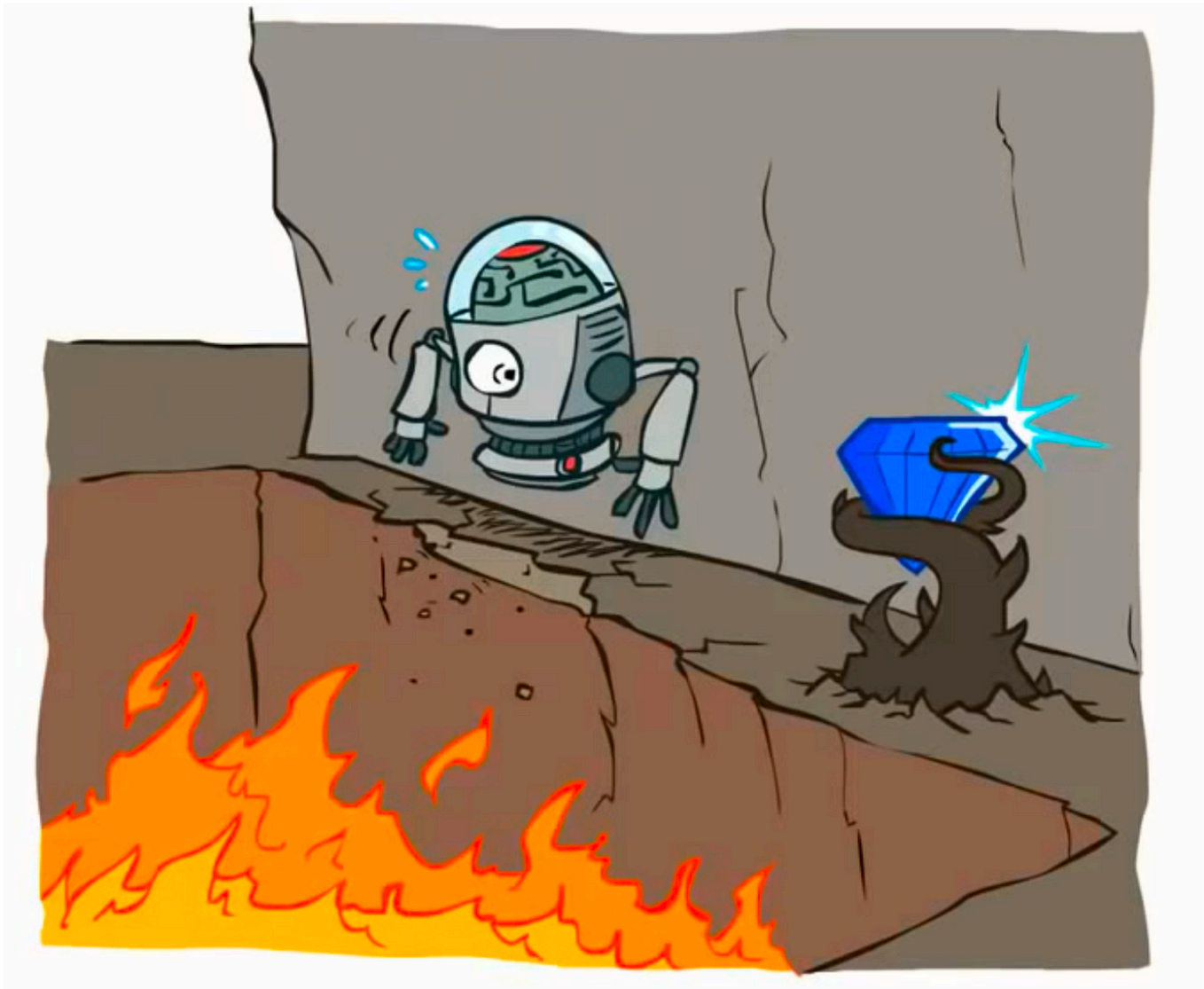
Expected utility <u>cannot</u>
be "naively" used

We need a richer "toolkit"

Two difficulties:

1. How to describe/model such a problem (in general)?

2. How to solve it (in general)?



Markov decision processes

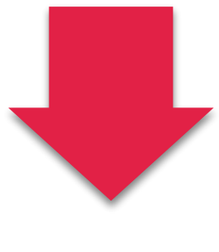
What does the model need?

- Identify the **information** that the decision depends on

States

What does the model need?

- Identify the **actions** that the agent can take



Actions

What does the model need?

- Describe the action **outcomes**

Dynamics

What does the model need?

- Describe the **goal** of the agent

Costs

States

States

- Relevant information for decision making
- We represent the state at time t as x_t
- Set of possible states is $\mathcal{X}$ (finite, most of the time)
- Each step, the agent makes a decision (**decision epoch**)

Actions

Action

- Means by which the agent influences the "environment"
- We represent the action at time t as a_t
- Set of possible actions is $\mathcal{A}$ (finite)

Dynamics

Dynamics

- Describe how the state evolves as a consequence of the agent's actions
- We assume that it verifies the **Markov property**

Markov property

Key Property: Markov property

The state at instant $t+1$ depends only on the state and action at time step t, i.e.,

$$\mathbb{P}\left[\mathrm{x}_{t+1} = y \mid \mathbf{x}_{0:t} = \boldsymbol{x}_{0:t}, \mathbf{a}_{0:t} = \boldsymbol{a}_{0:t}\right] = \mathbb{P}\left[\mathrm{x}_{t+1} = y \mid \mathrm{x}_t = x_t, \mathrm{a}_t = a_t\right]$$

Additional assumptions:

- The probabilities $\boxed{\mathbb{P}[\mathrm{x}_{t+1} = y \mid \mathrm{x}_t = x, \mathrm{a}_t = a]}$ do not depend on t

Transition probability from x
to y given a

- For each action $a \in \mathcal{A}$, we store the transition probabilities in a **matrix**

$$[\mathbf{P}_a]_{xy} = \mathbb{P}\left[\mathrm{x}_{t+1} = y \mid \mathrm{x}_t = x, \mathrm{a}_t = a\right]$$

Costs

Immediate costs

- Instantaneously evaluates **state and action**
- Represented as a function $c : \mathcal{X} \times \mathcal{A} \to \mathbb{R}$
- For simplicity, we assume that $c(x, a) \in [0, 1]$

Markov decision process

- **Model** for sequential decision processes
- Described by:
  - State space, $\mathcal{X}$
  - Action space, $\mathcal{A}$
  - Transition probabilities, $\{\mathbf{P}_a, a \in \mathcal{A}\}$
  - Immediate cost function, $\mathbf{c}$

Useful notation

- Sometimes we write:
  - $\mathbf{P}(y \mid x, a)$ to denote $[\mathbf{P}_a]_{xy}$

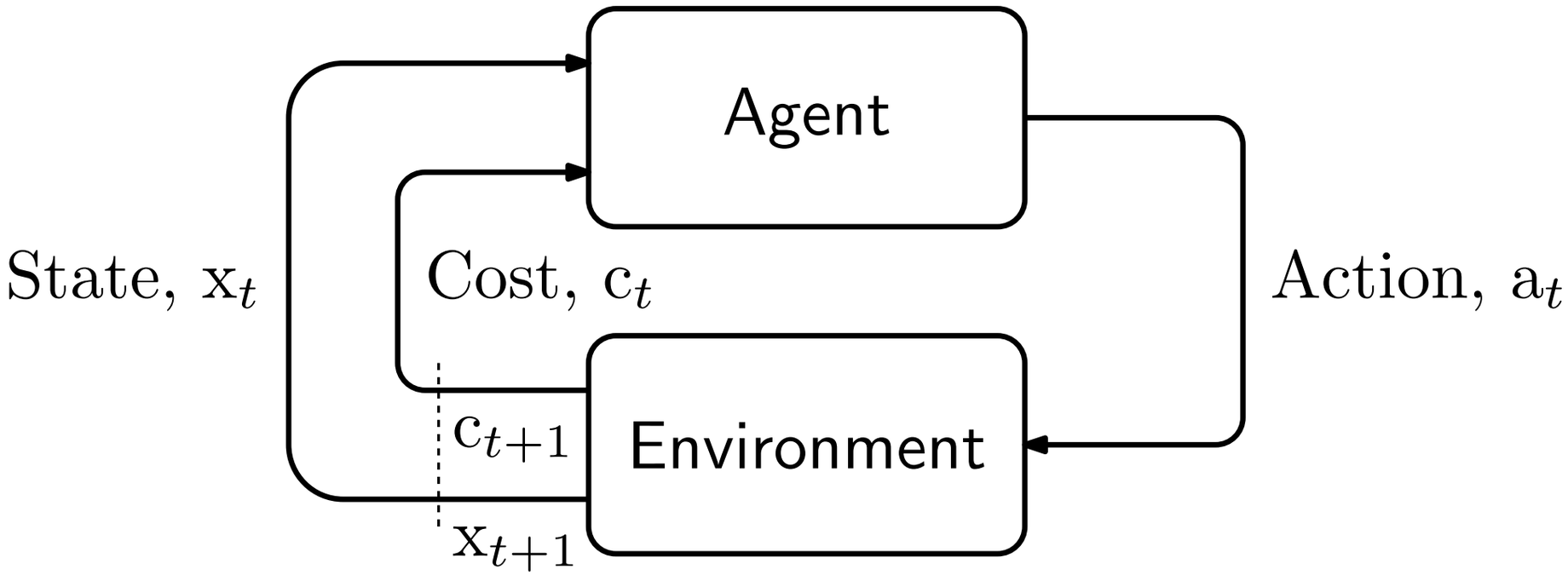
$$\mathbf{P}_a = \begin{bmatrix} \mathbf{P}_a(x_1 \mid x_1) & \mathbf{P}_a(x_2 \mid x_1) & \dots & \mathbf{P}_a(x_N \mid x_1) \\ \mathbf{P}_a(x_1 \mid x_2) & \mathbf{P}_a(x_2 \mid x_2) & \dots & \mathbf{P}_a(x_N \mid x_2) \\ \vdots & & \ddots & \vdots \\ \mathbf{P}_a(x_1 \mid x_N) & \mathbf{P}_a(x_2 \mid x_N) & \dots & \mathbf{P}_a(x_N \mid x_N) \end{bmatrix}$$

Useful notation

- Sometimes we write:
  - $\mathbf{C}$ to denote the cost matrix, with $[\mathbf{C}]_{xa} = c(x, a)$

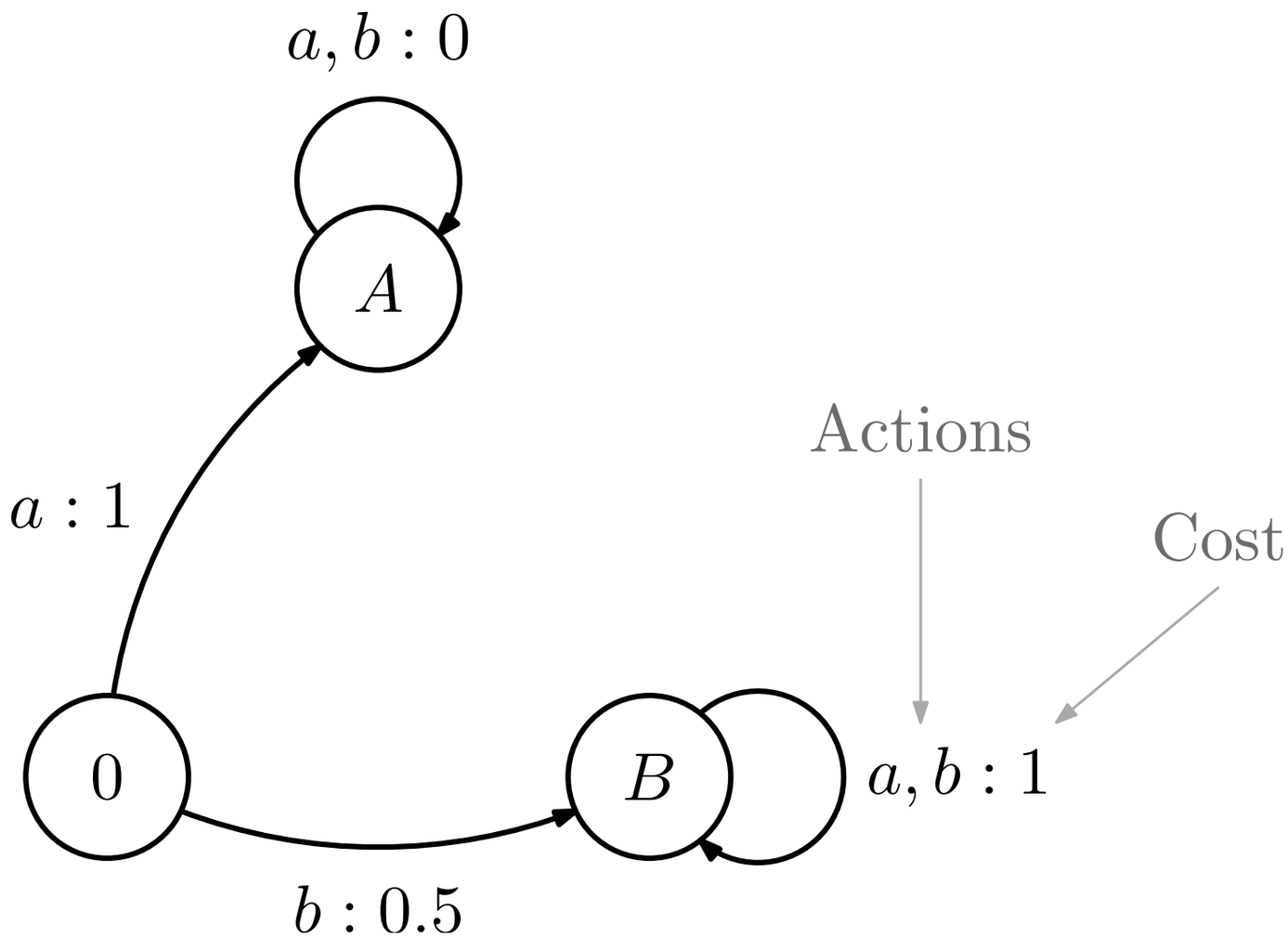
$$\mathbf{C} = \begin{bmatrix} c(x_1,\ a_1) & c(x_1,\ a_2) & \dots & c(x_1,\ a_M) \\ c(x_2,\ a_1) & c(x_2,\ a_2) & \dots & c(x_2,\ a_M) \\ \vdots & & \ddots & \vdots \\ c(x_N,\ a_1) & c(x_N,\ a_2) & \dots & c(x_N,\ a_M) \end{bmatrix}$$

Markov decision process



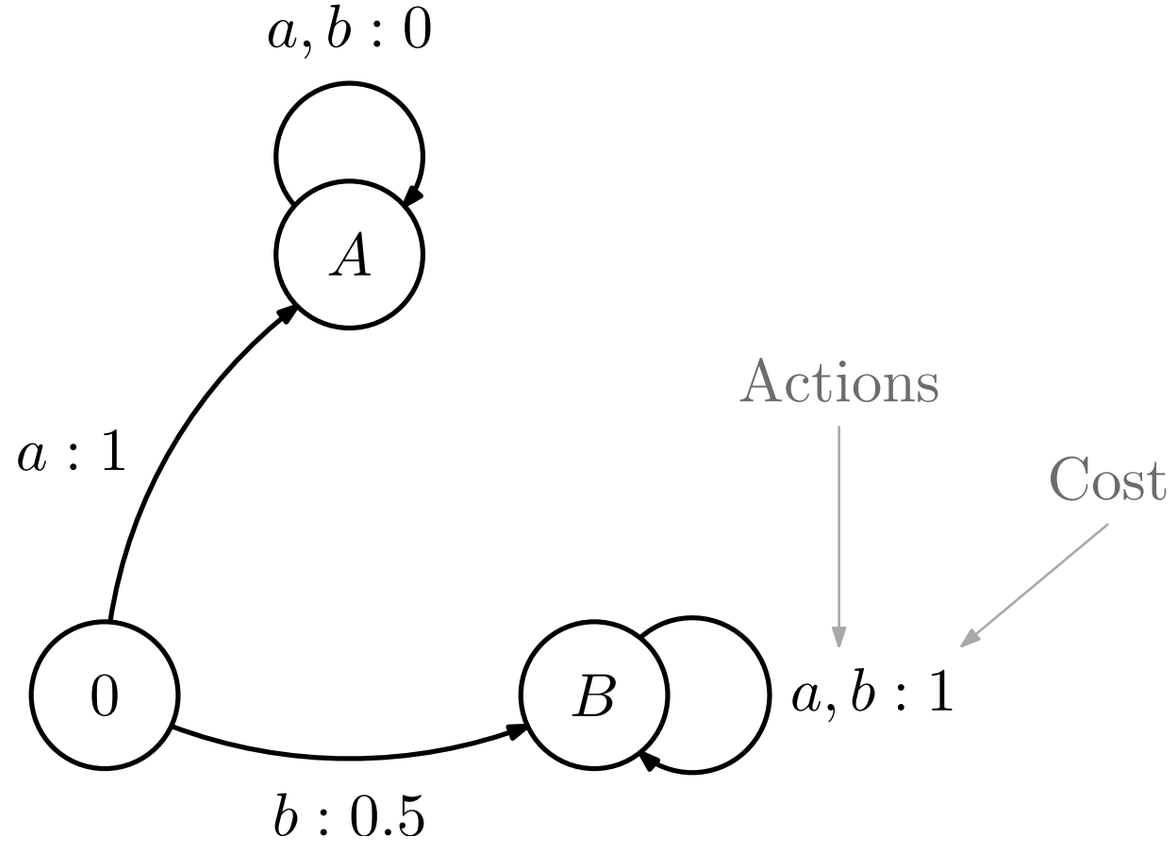
Examples

Example 1



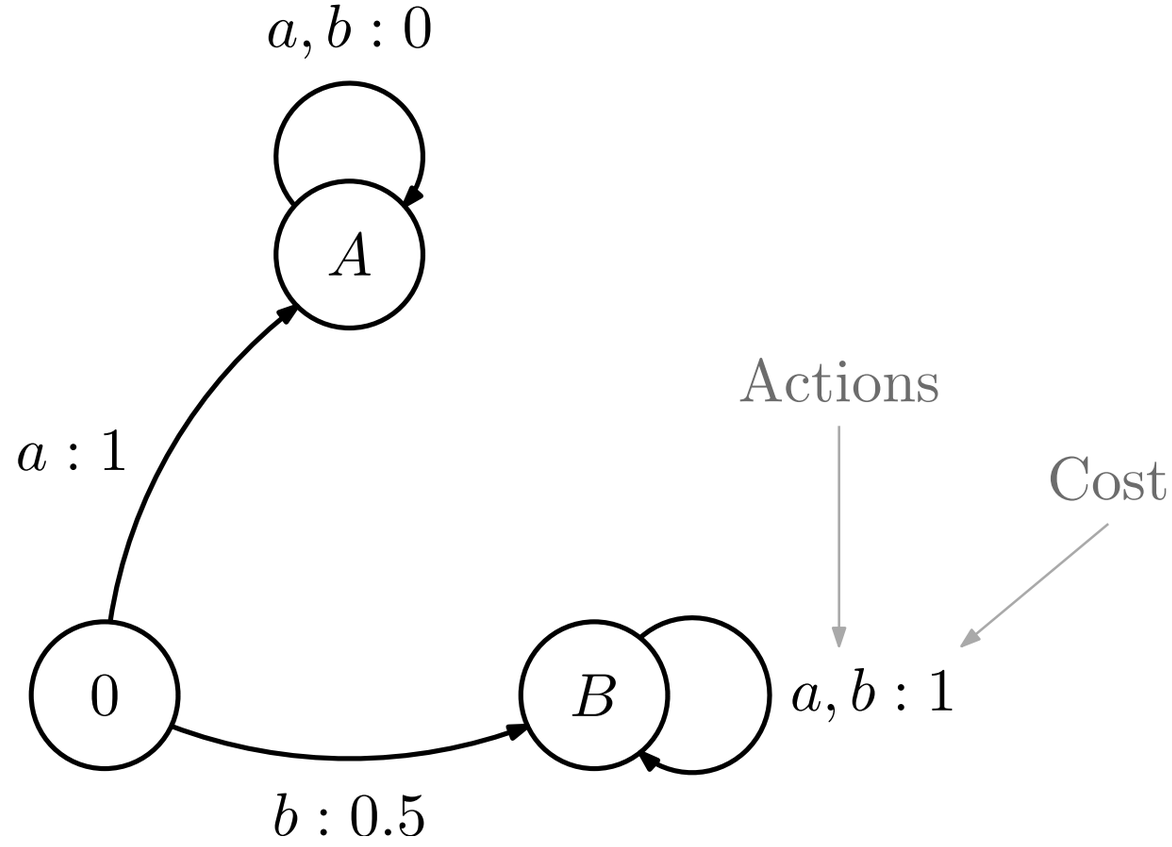
Model definition

- States:
  - $\mathcal{X} = \{0, A, B\}$



Model definition

- Actions:
  - $\mathcal{A} = \{a, b\}$

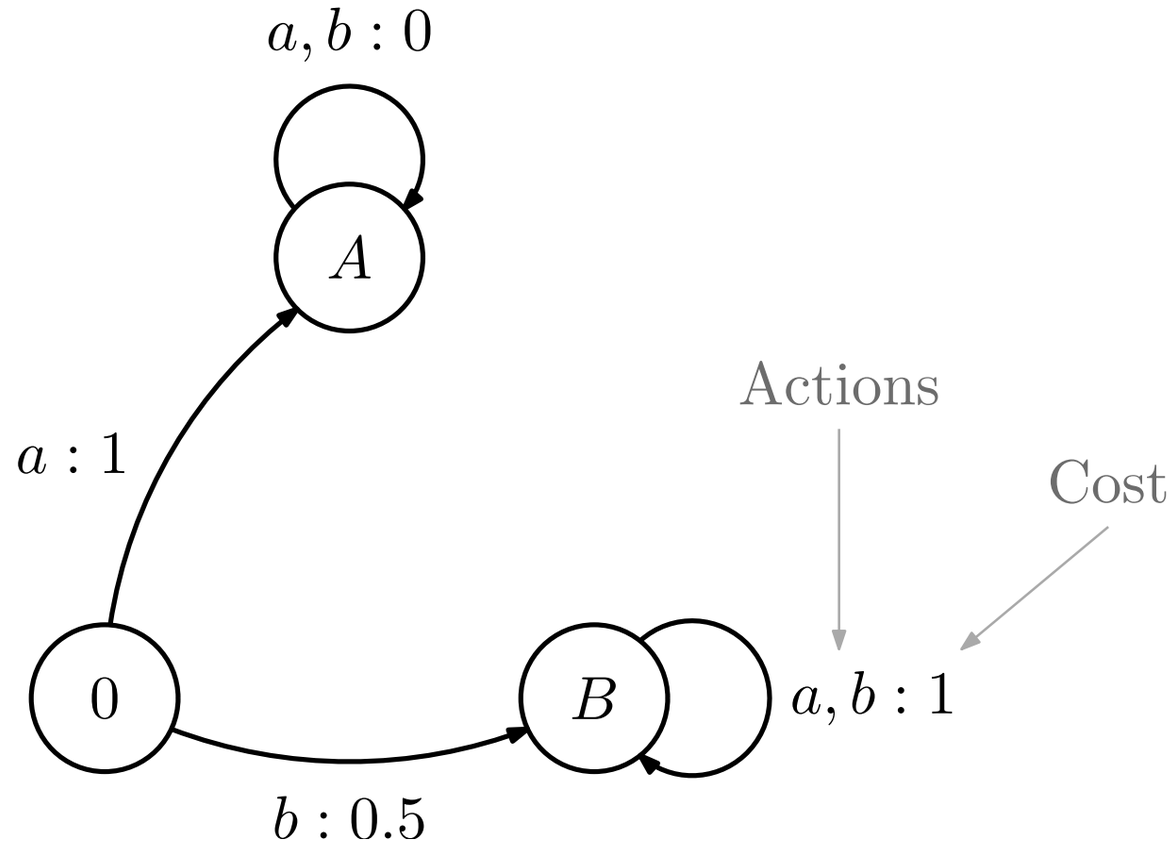


Model definition

- Transition probabilities:

$$\mathbf{P}_a = \begin{bmatrix} 0 & 1 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix}$$

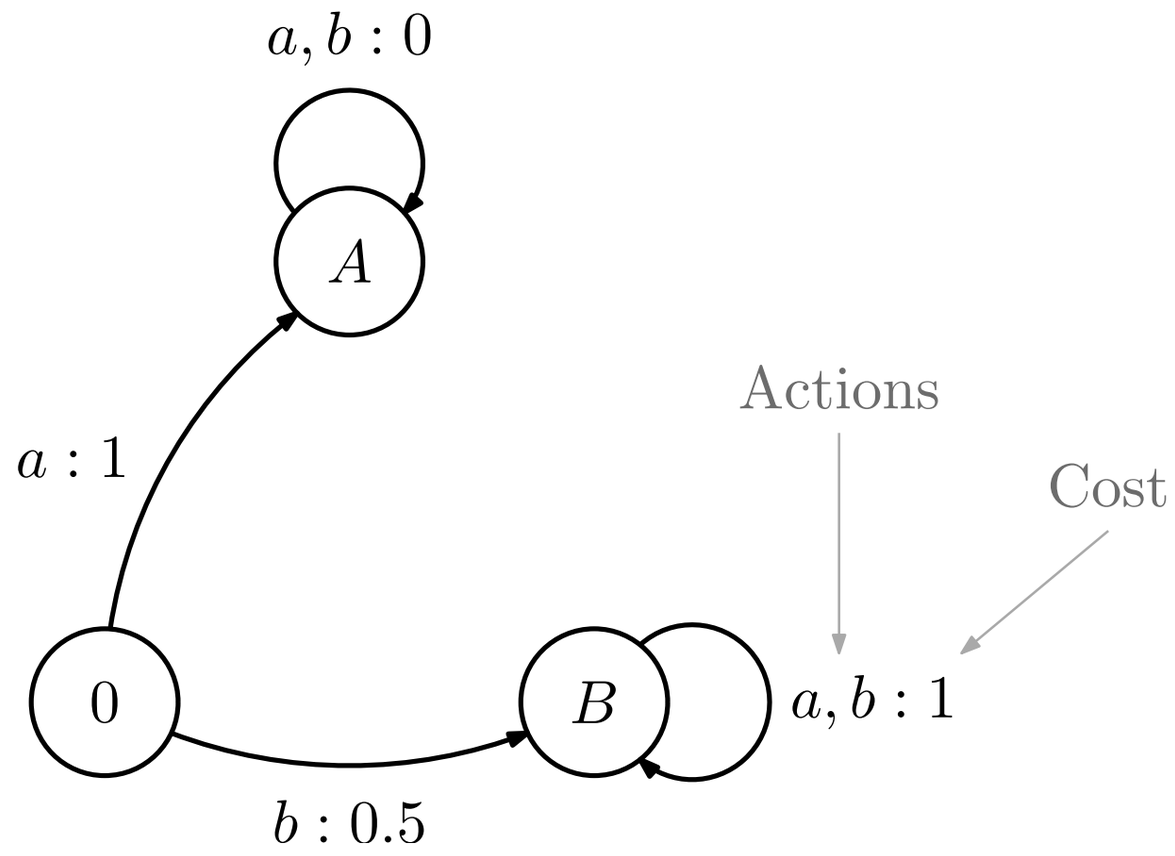
$$\mathbf{P}_b = \begin{bmatrix} 0 & 0 & 1 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix}$$



Model definition

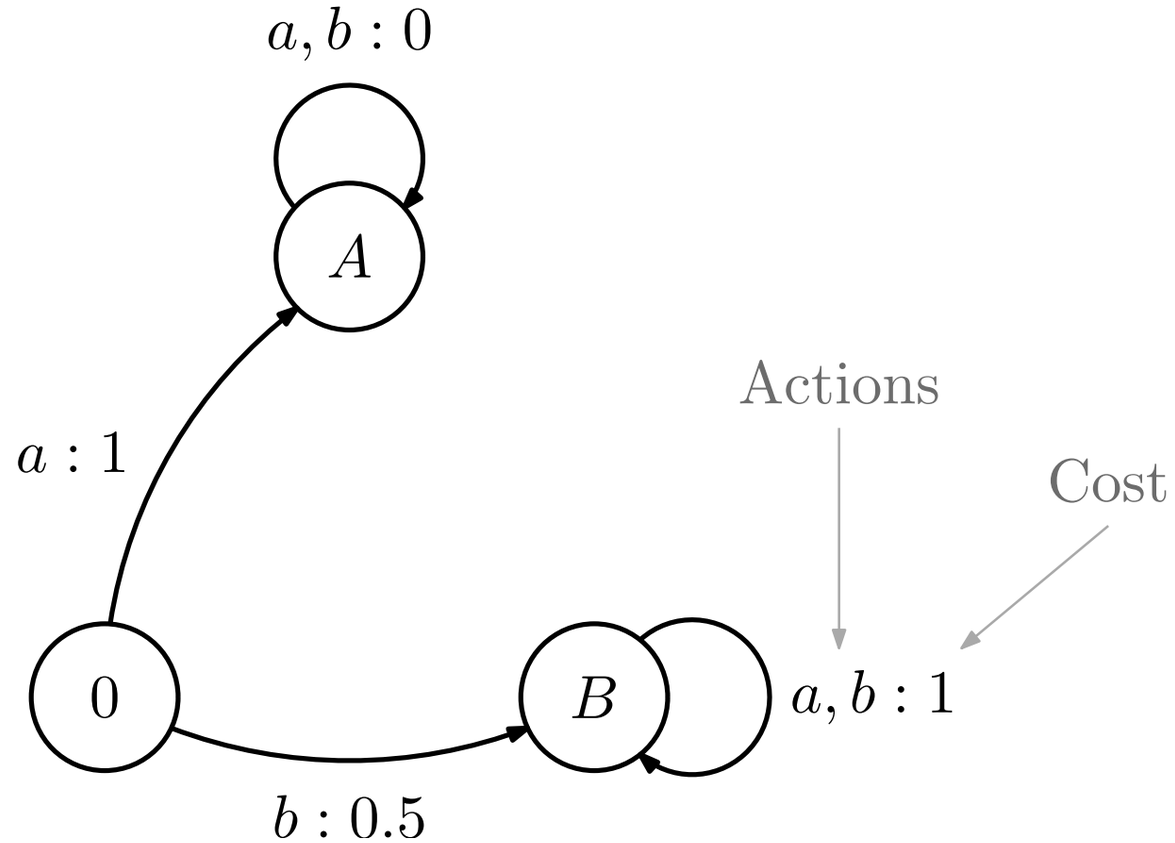
- Cost:

$$\mathbf{C} = \begin{bmatrix} 1 & \frac{1}{2} \\ 0 & 0 \\ 1 & 1 \end{bmatrix}$$



What is the best decision?

- Depends on what "best" means
  - If single decision, then best is b
  - If multiple decisions, then best is a



Example 2

- A company wants to hire a computer engineer
- After initial trial, N candidates are selected for interview

Example 2

- Candidates are interviewed sequentially
- Order of the candidates for interview was selected randomly

Example 2

- Manager must decide, after each interview, whether to hire or not (no second chances)
- Manager knows whether an interviewed candidate is the best so far
- If no candidate has been hired in the meantime, candidate N is necessarily hired

How to model this?

- What are the states?
  - What is relevant for the manager's decision?
    - Current candidate best so far or not
    - How many candidates have been interviewed/are missing
- State-space:
  - $\mathcal{X} = \{(B,1), (B,2), (\neg B,2), \ldots, (B,N), (\neg B,N), H\}$

Not best so far

Best so far

Hired

How to model this?

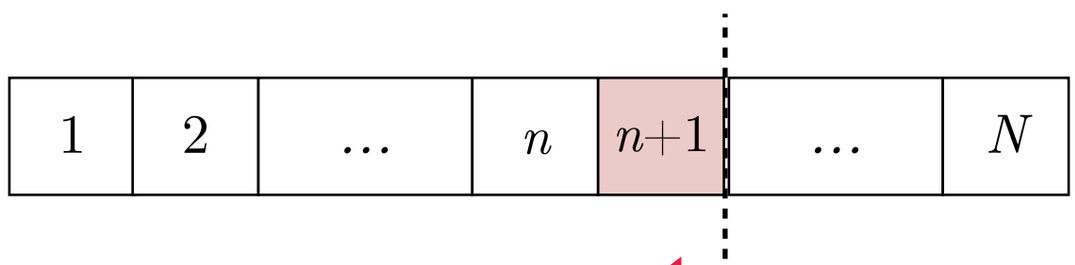
- What are the actions?
  - $\mathcal{A} = \{H, \neg H\}$

How to model this?

- Transition probabilities:
  - ... tough!

How to model this?

- Transition probabilities:
  - What is the probability that the $(n+1)$th candidate is the best so far?



Probability that the best among first $n+1$ candidates is candidate $n+1$?

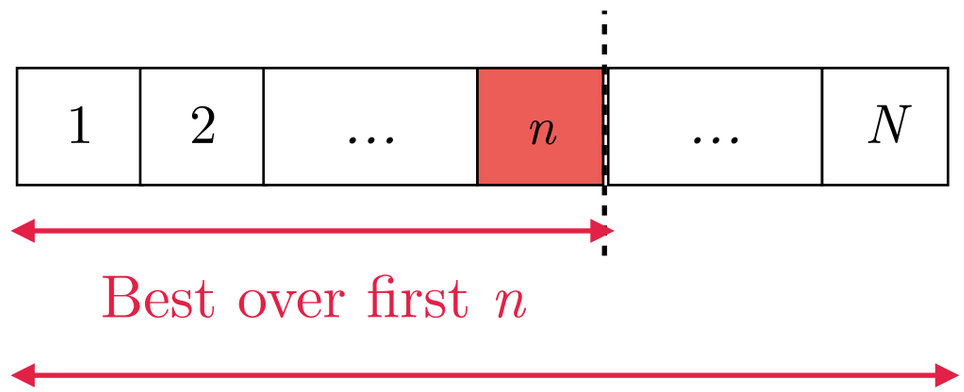
$$1 \;/\; (n+1)$$

How to model this?

- Transition probabilities:
  - What is the probability that the $(n+1)$th candidate is the best so far?
    - $1 \,/\, (n+1)$
  - What's the probability that the $(n+1)$th candidate is **not** the best so far?
    - $n \,/\, (n+1)$

How to model this?

- Cost:
  - ... hiring a guy who is not the best so far incurs maximum cost (clearly, that guy is not the best)
  - ... what about hiring a guy who is the best so far after n interviews?
    - How likely is it that it is not the best overall?

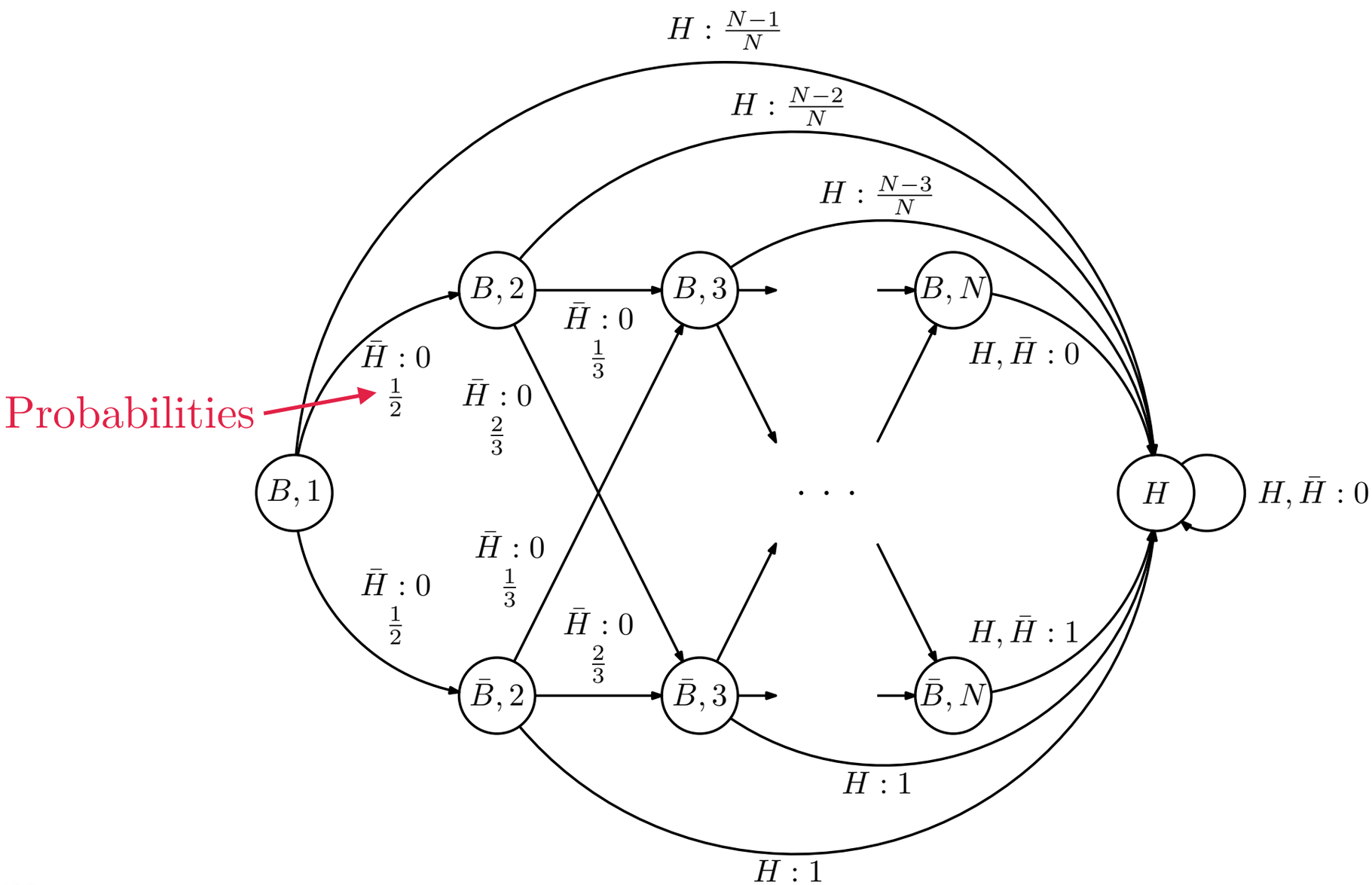


How to model this?

- Cost:
  - ... hiring a guy who is not the best so far incurs maximum cost (clearly, that guy is not the best)
  - ... what about hiring a guy who is the best so far after n interviews?
    - How likely is it that it is not the best overall?
      - $(N - n)/N$

How to model this?

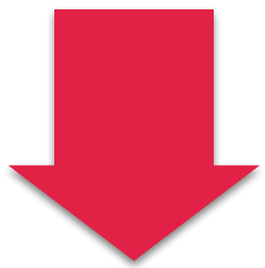
- Putting everything together:



Decisions with Markov decision processes

Optimality?

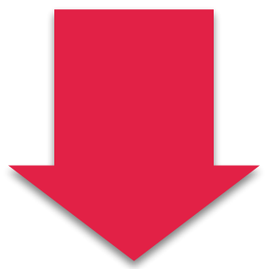
- Given a Markov decision process, $(\mathcal{X}, \mathcal{A}, \{\mathbf{P}_a\}, c)$...
  - ... what do we want to do?



Select the
"best" actions

Optimality

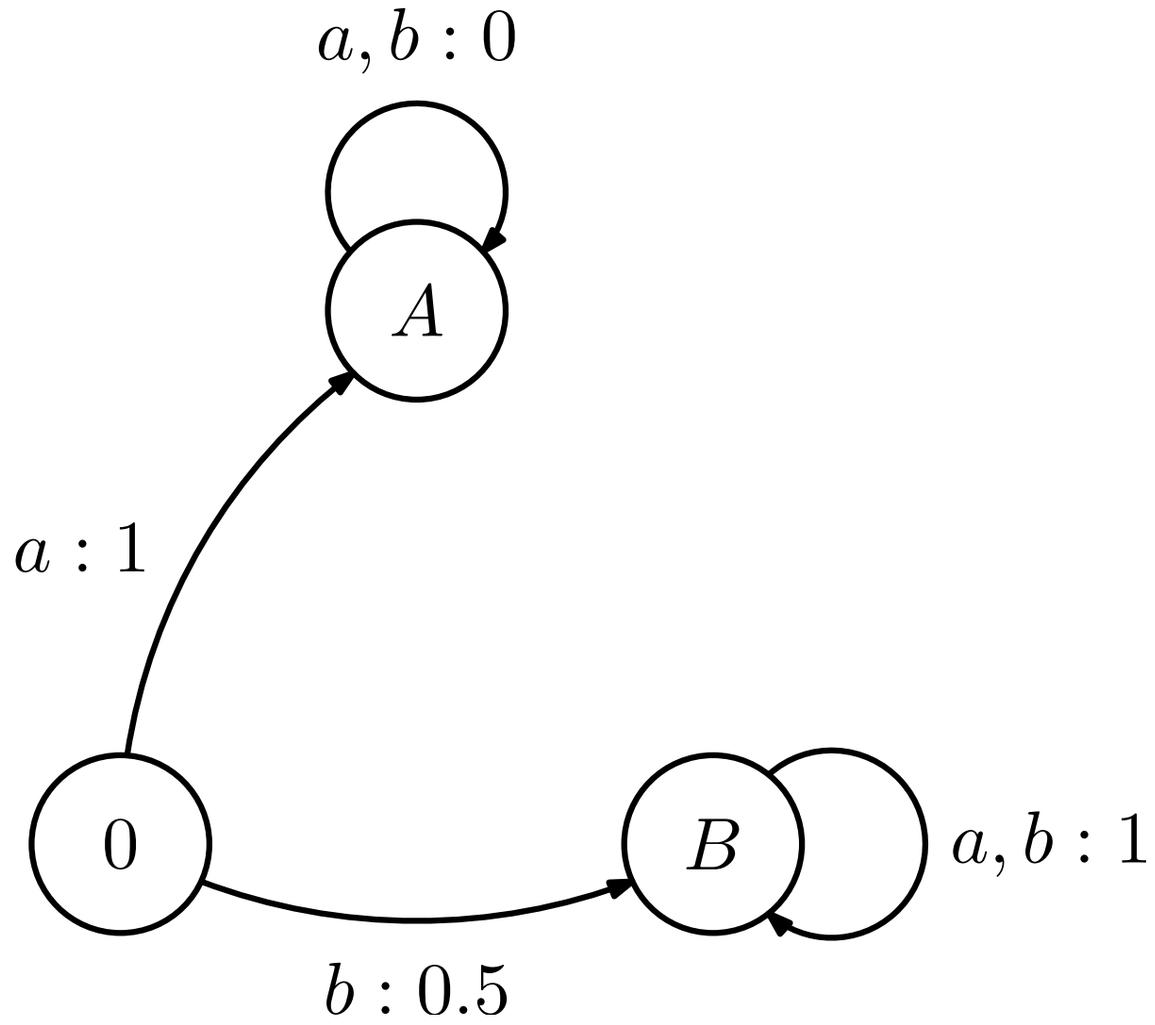
- What are the "best" actions?
- We need a criterion to compare different **ways of selecting actions**



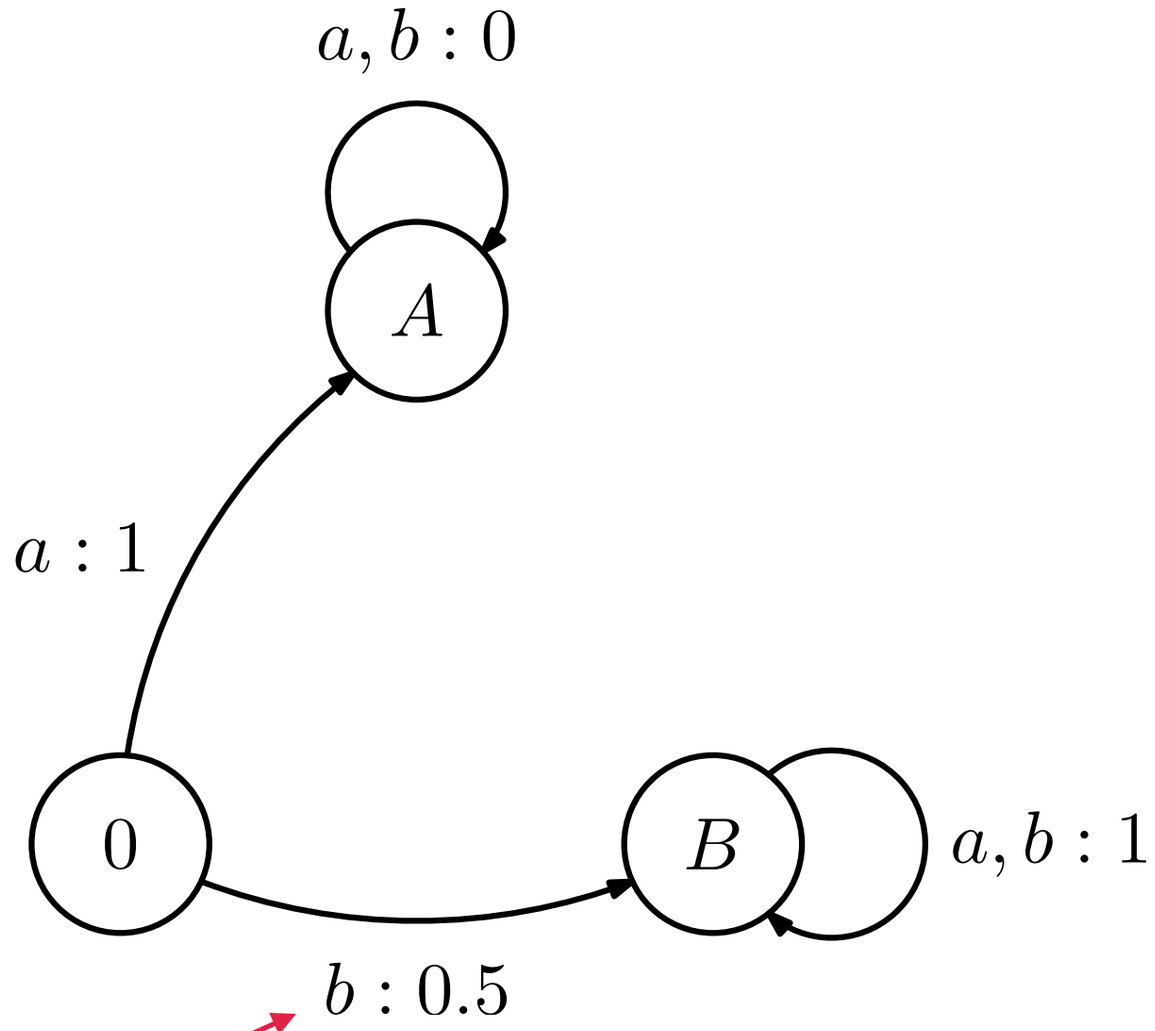
Optimality criterion

Example

What is the best action?



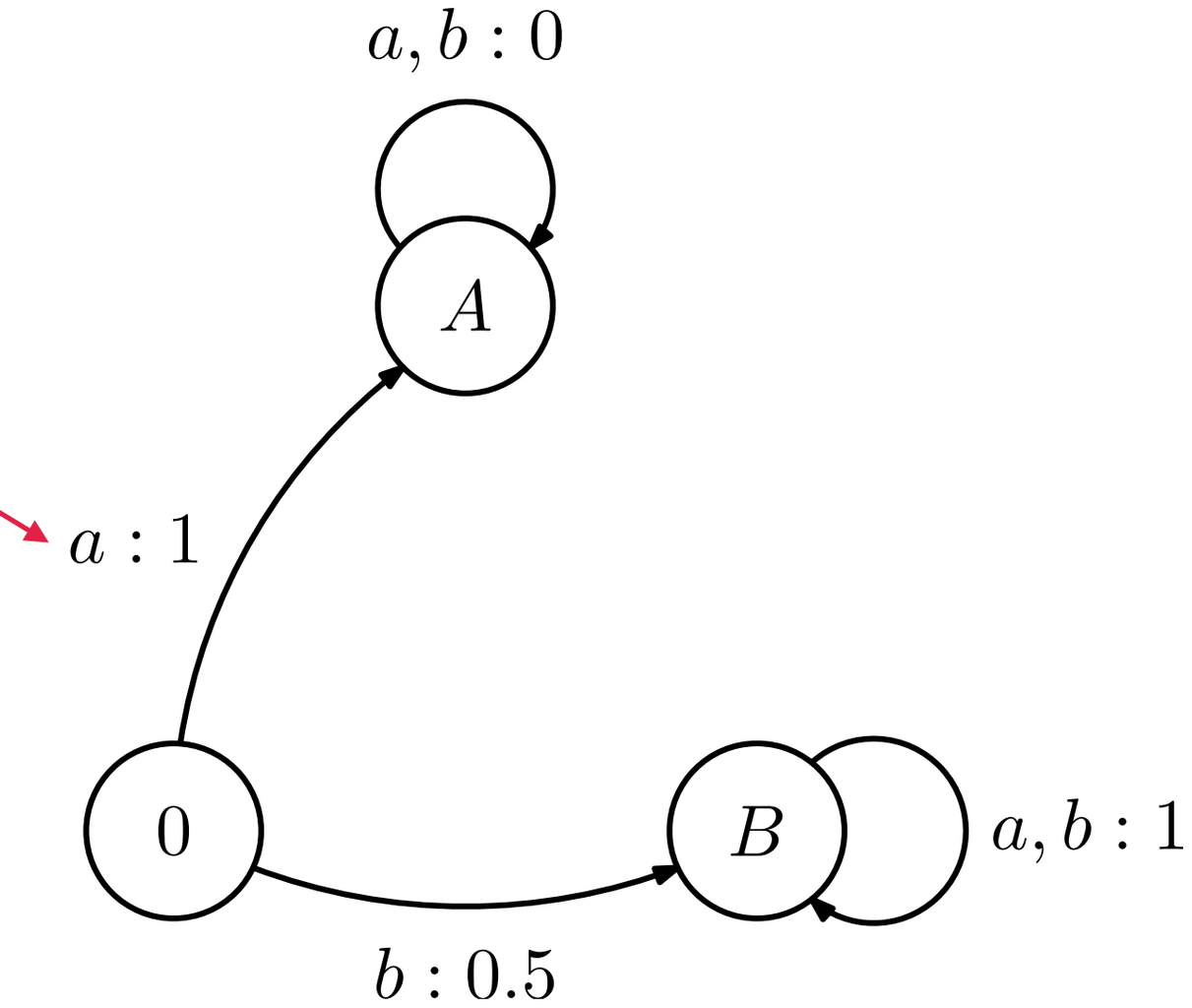
Example



If there is a
single decision,
b is the best!

Example

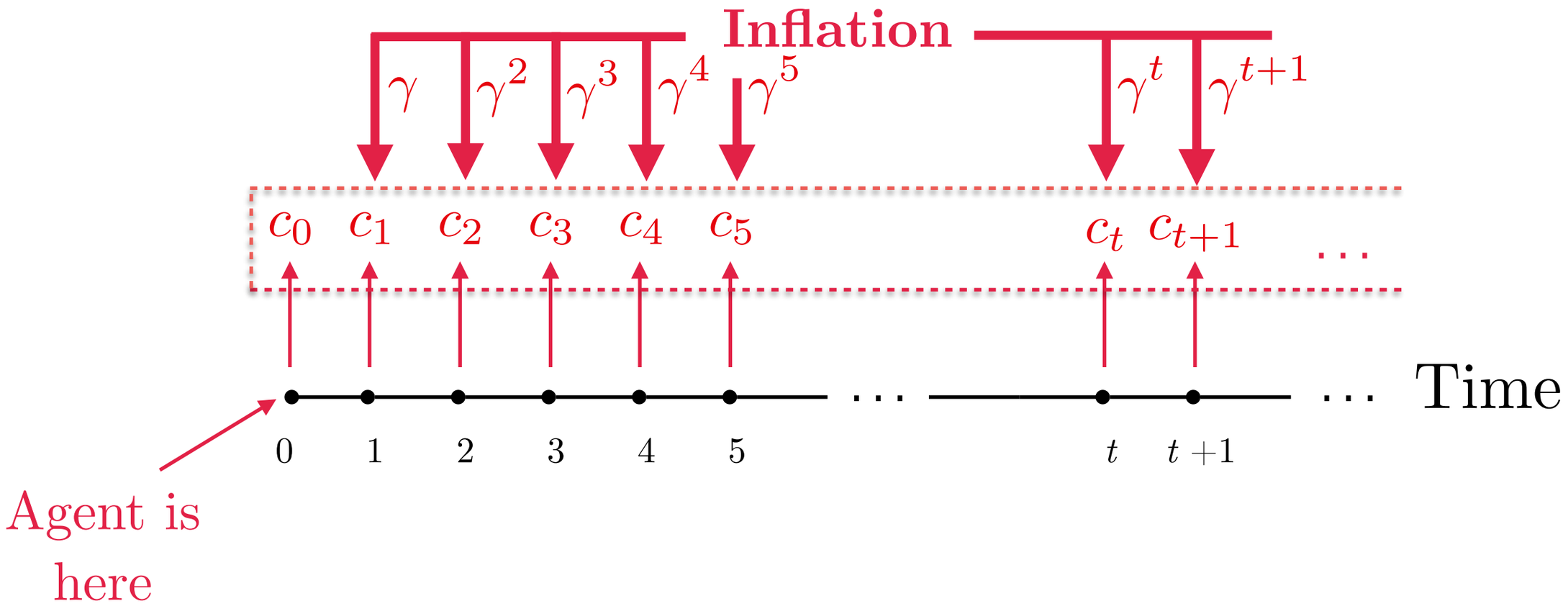
If there is more than one decision, a is the best!



Discounted cost-to-go

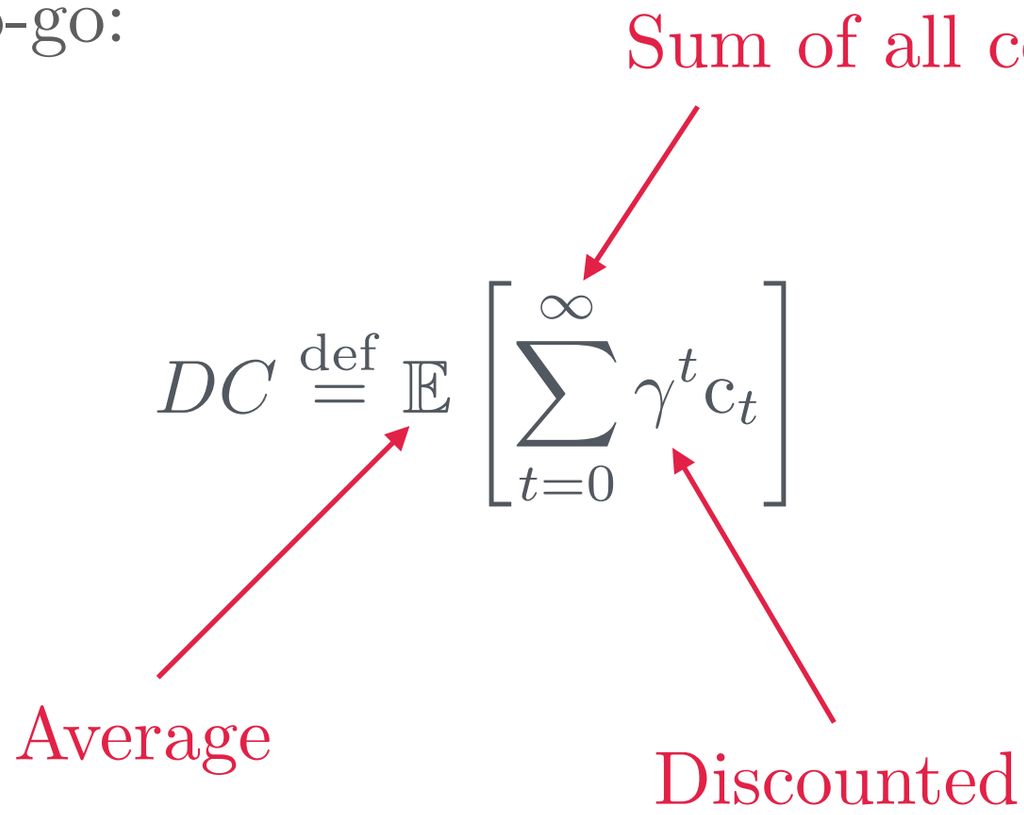
- Assumptions:
  - The agent lives forever (we don't know n. of decisions)
  - There is an inflation rate (costs in the future are not as bad as costs now)
  - Agent wants to pay as little as possible

Discounted cost-to-go



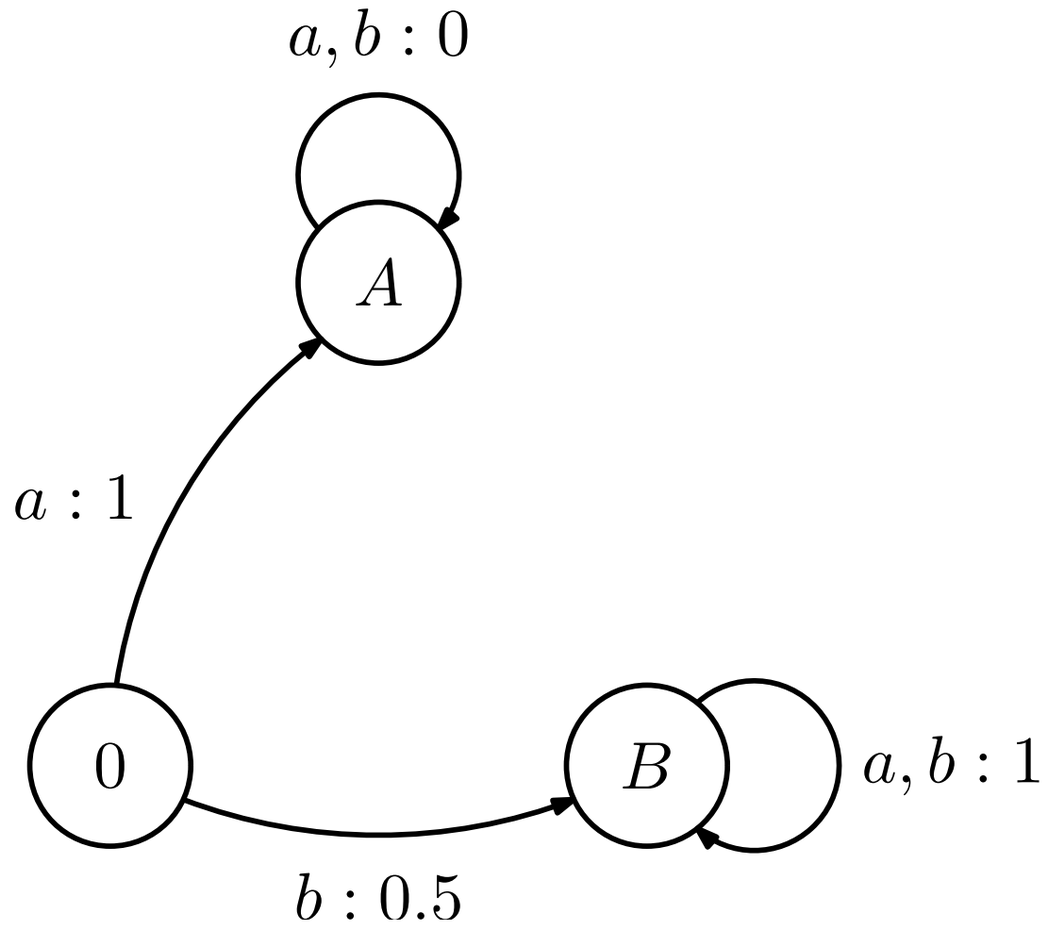
Discounted cost-to-go

- Discounted cost-to-go:



Example

- What is the discounted cost-to-go if we always select b?
  - **It depends on where we start!**

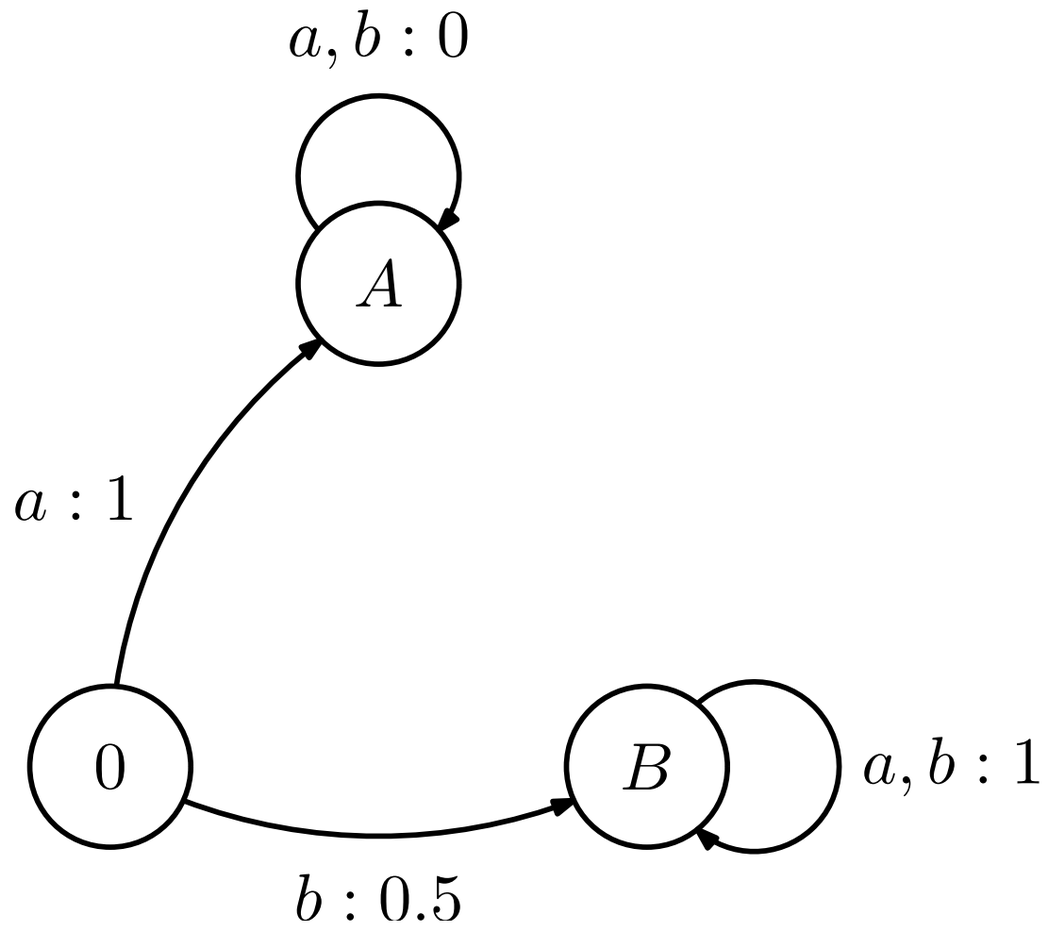


Example

- What if we start in A?

$$J(A) = 0 + \gamma 0 + \ldots = 0$$

Cost-to-go if we start in A

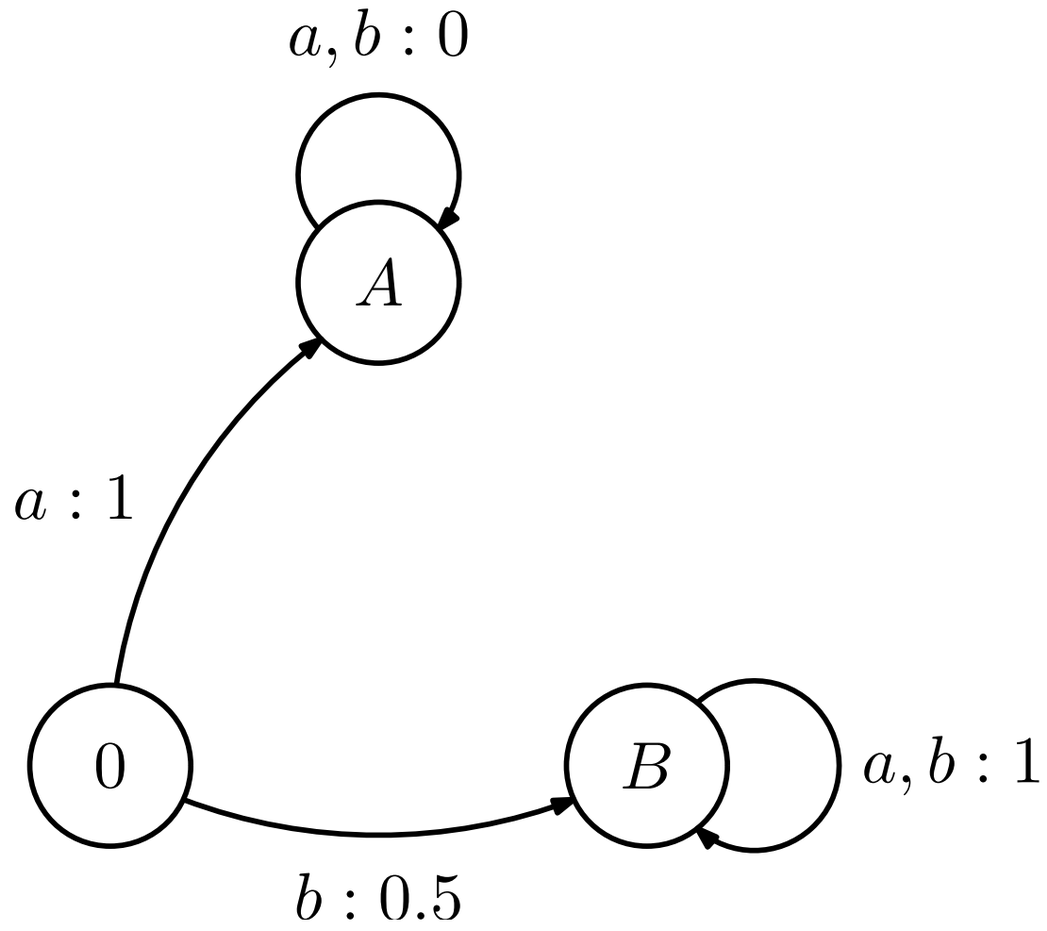


Example

- What if we start in B?

$$J(B) = 1 + \gamma 1 + \ldots$$

$$= \frac{1}{1-\gamma}$$



Example

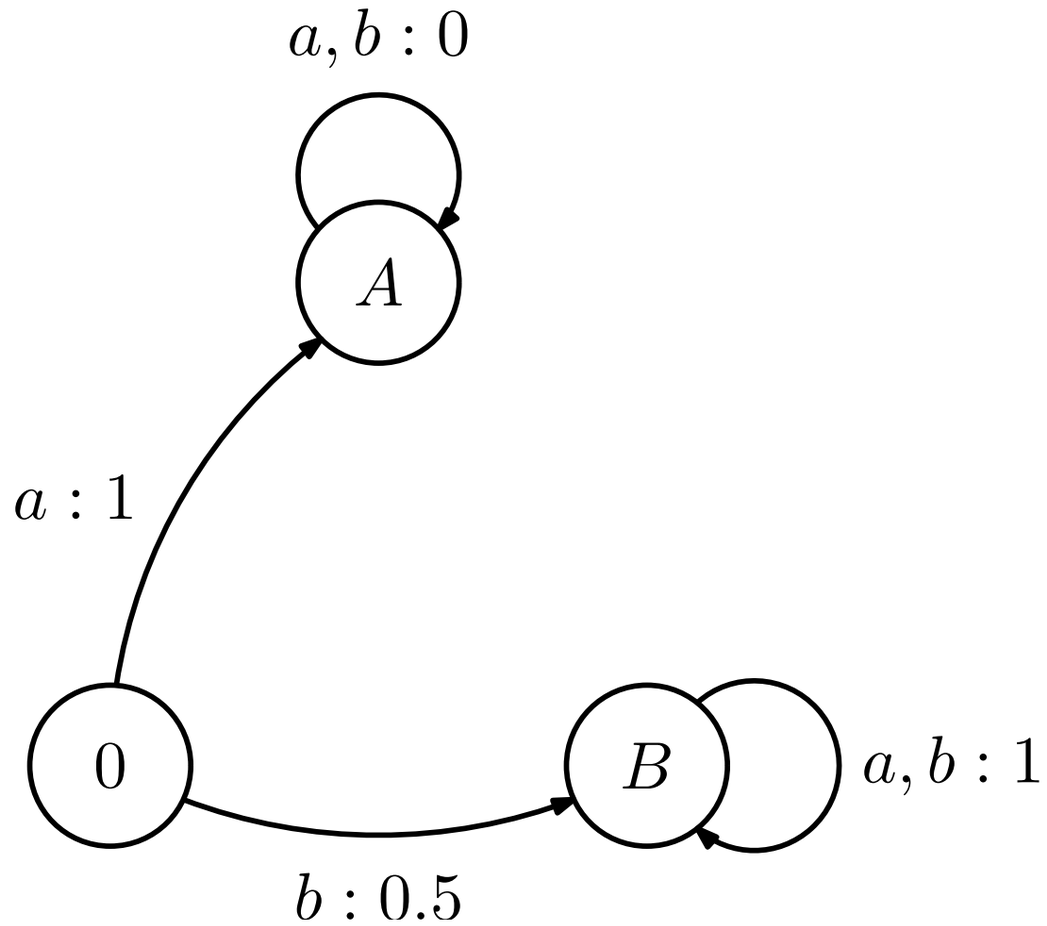
- What if we start in 0?

$$J(0) = 0.5 + \gamma 1 + \gamma^2 1 + \ldots$$

$$= 0.5 + \gamma \underbrace{(1 + \gamma 1 + \ldots)}_{J(B)}$$

$$= 0.5 + \gamma J(B)$$

$$= \frac{1}{2} \cdot \frac{1+\gamma}{1-\gamma}$$



Example

- What is the discounted cost-to-go if we always select b?

$$\boldsymbol{J} = \begin{bmatrix} \frac{1}{2} \cdot \frac{1+\gamma}{1-\gamma} \\ 0 \\ \frac{1}{1-\gamma} \end{bmatrix}$$