Intelligent Systems (AI-2)

Computer Science cpsc422, Lecture 2

Jan, 7, 2015

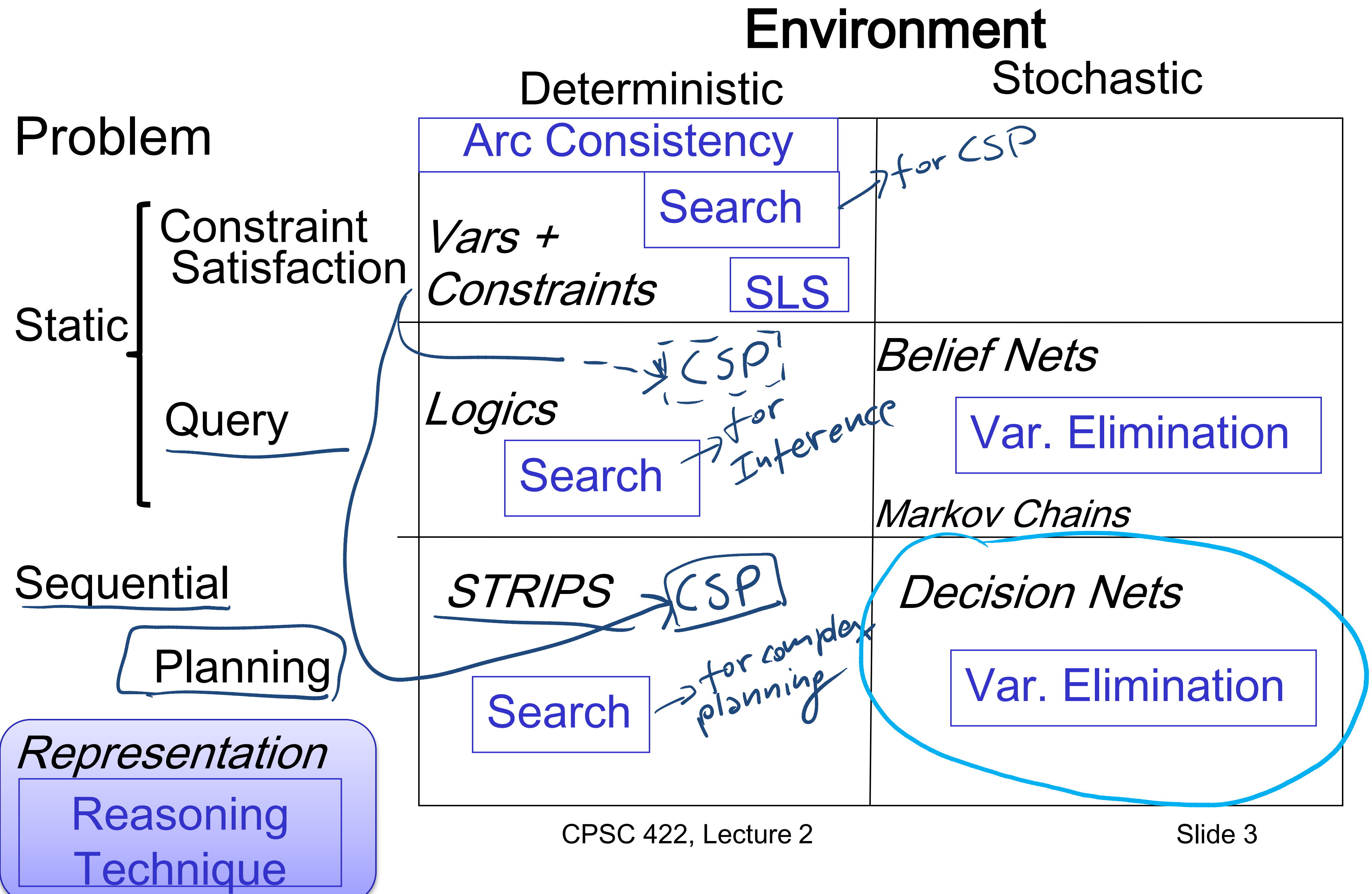
Lecture Overview

Value of Information and Value of Control

Markov Decision Processes (MDPs)

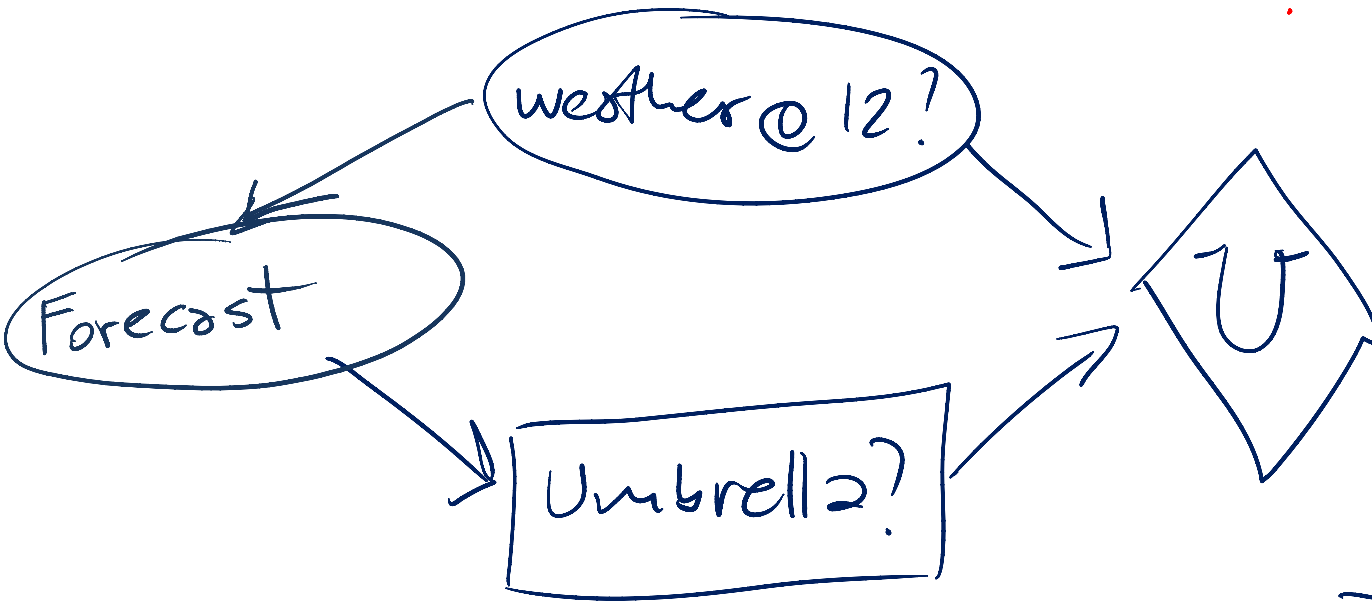
- Formal Specification and example
- Define a policy for an MDP

Cpsc 322 Big Picture



Simple Decision Net

- Early in the morning. Shall I take my **umbrella** today? (I’ll have to go for a long walk at noon)
- Relevant Random Variables?



Polices for Umbrella Problem

- A **policy** specifies what an agent should do under each circumstance (for each decision, consider the parents of the decision node)

In the *Umbrella* case:

D_1 ? T F

pD_1 Rainy Cloudy Sunny

One possible Policy

→ R	T	F	T …
→ C	T	F	T …
→ S	F	F	T …

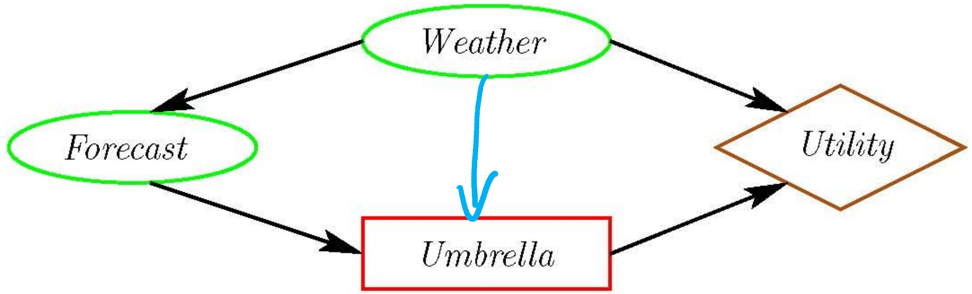
3 policies

How many policies?

2^3

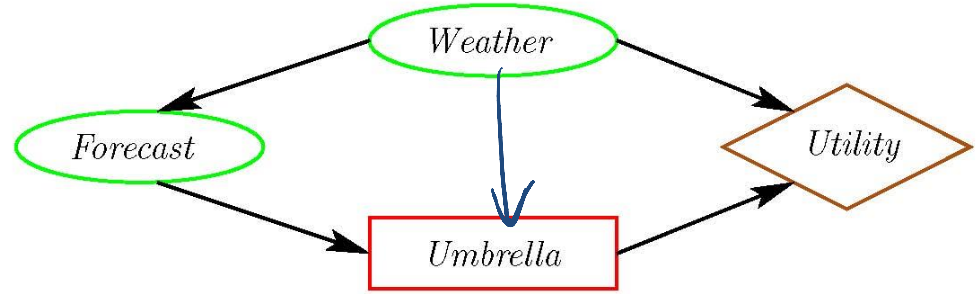
$|D|^{|pD|}$

Value of Information



- Early in the morning. I listen to the **weather forecast**, shall I take my **umbrella** today? (I'll have to go for a long walk **at noon**)
- What would help the agent make a better *Umbrella* decision?

Value of Information



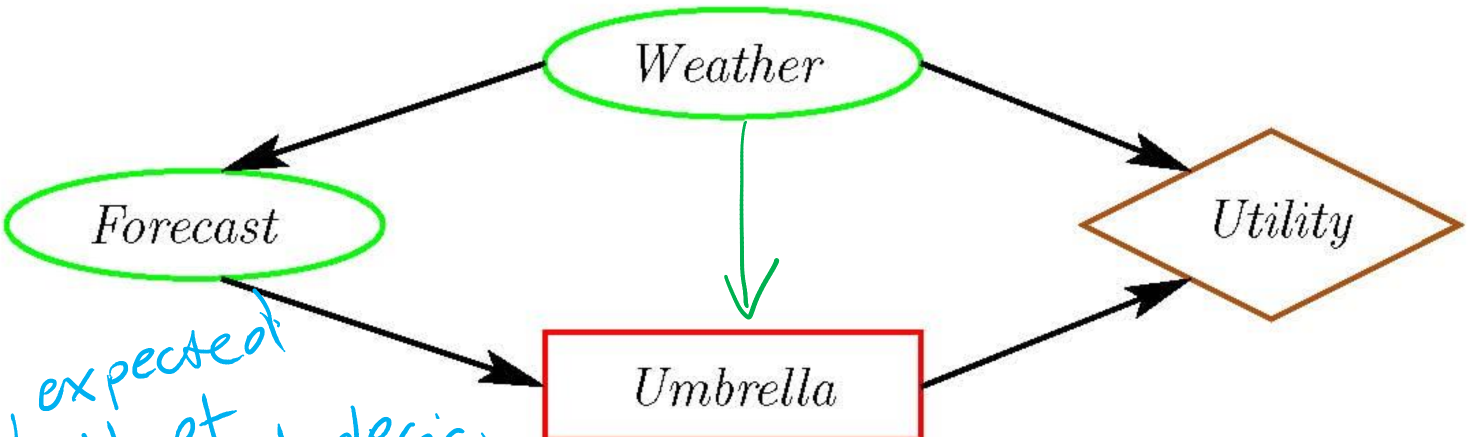
- The value of information of a random variable X for decision D is: EU(knowing X) − EU(not knowing X)

the utility of the network with an arc from X to D **minus** the utility of the network without the arc.

- Intuitively:
  - The value of information is always ≥ 0
  - It is positive only if the agent changes its policy

Value of Information (cont.)

- The value of information provides a bound on how much you should be prepared to pay for a sensor. How much is a **perfect** weather forecast worth?

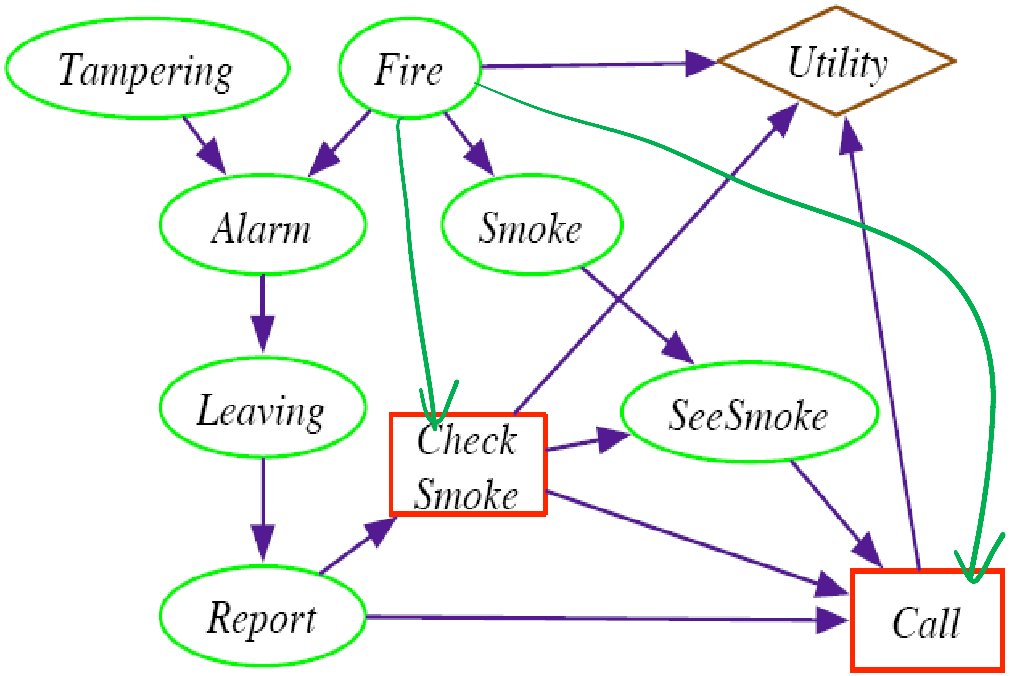


i.e., expected utility of optimal decision

- Original maximum expected utility: 77
- Maximum expected utility when we know Weather: 91
- Better forecast is worth at most: 14

Value of Information

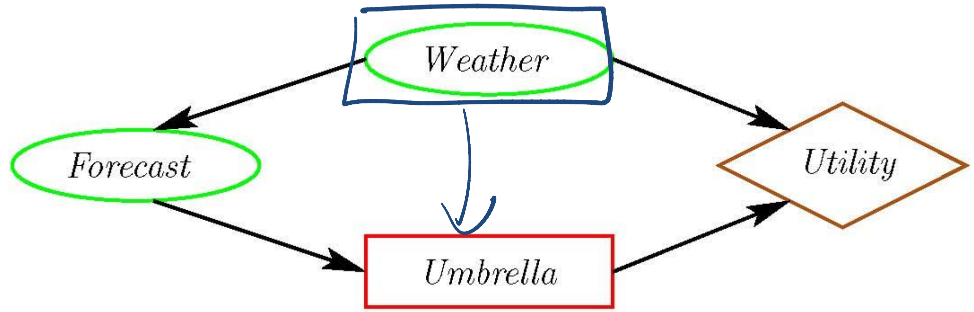
- The value of information provides a bound on how much you should be prepared to pay for a sensor. How much is a **perfect** fire sensor worth?



- Original maximum expected utility: -22.6
- Maximum expected utility when we know Fire: -2
- Perfect fire sensor is worth: 20.6

AIspace

Value of Control



- What would help the agent to make an even better *Umbrella* decision? To maximize its utility.

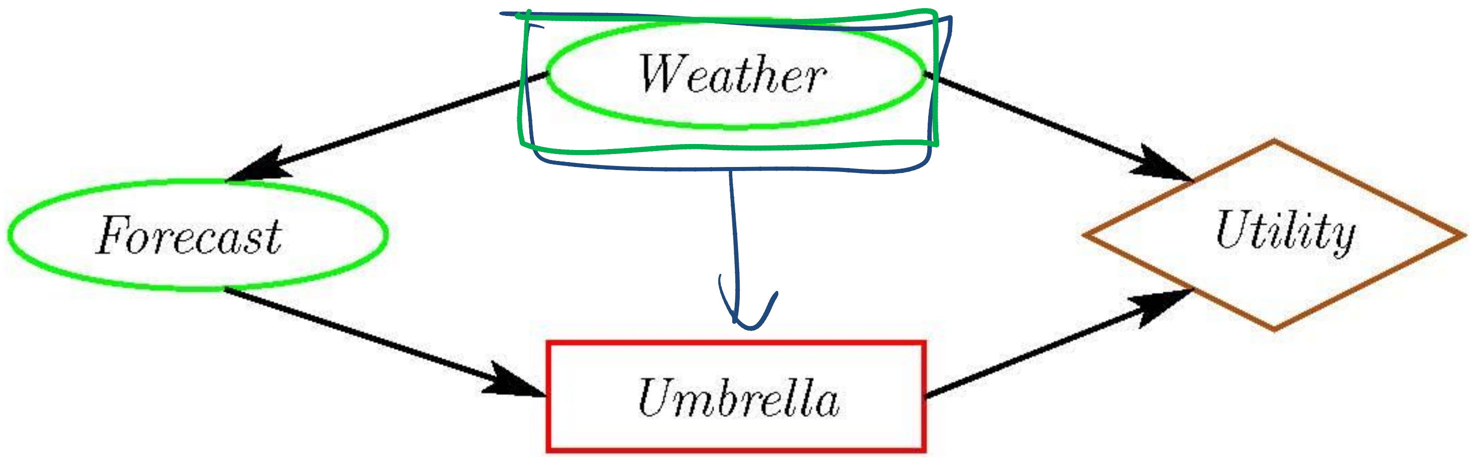
Weather	Umbrella	Value
Rain	true	70
Rain	false	0
noRain	true	20
noRain	false	100

- The value of control of a variable *X* is :

the utility of the network when you make X a decision variable **minus** the utility of the network when X is a random variable.

Value of Control

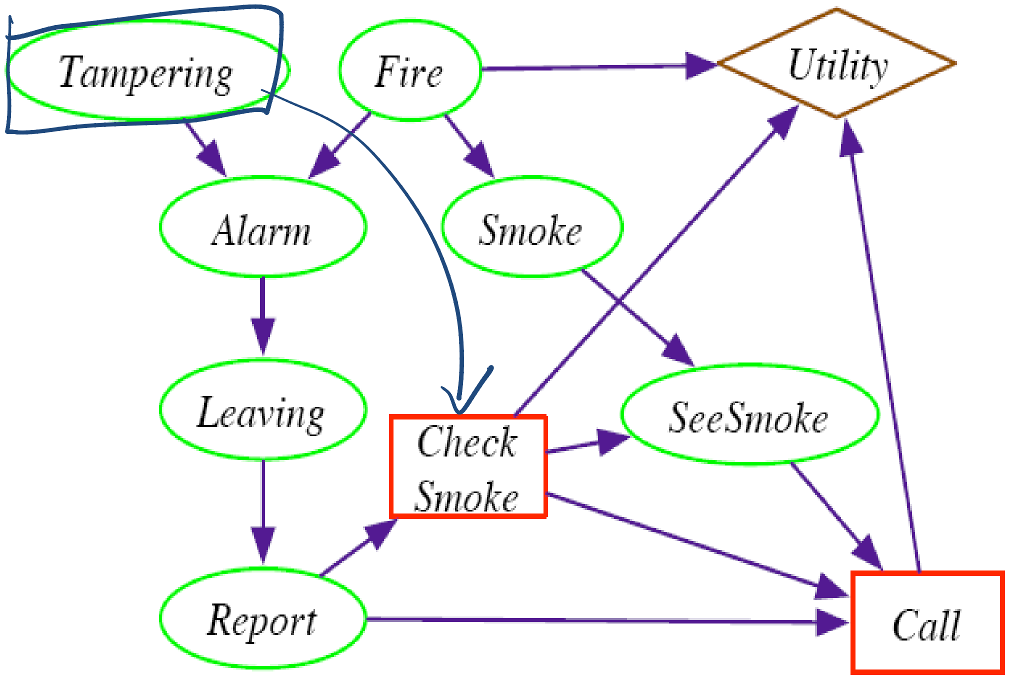
- What if we could control the weather?



- Original maximum expected utility: 77
- Maximum expected utility when we control the weather: 100
- Value of control of the weather: 23

Value of Control

- What if we control Tampering?



- Original maximum expected utility: -22.6
- Maximum expected utility when we control the Tampering: -20.7
- Value of control of Tampering: 1.9
- Let's take a look at the policy
- Conclusion: do not tamper with fire alarms!

Lecture Overview

Value of Information and Value of Control

Markov Decision Processes (MDPs)

- Formal Specification and example
- Define a policy for an MDP

Combining ideas for Stochastic planning

- What is a key limitation of decision networks?

Represent (and optimize) only a fixed number of decisions

- What is an advantage of Markov models?

The network can extend indefinitely

Goal: represent (and optimize) an indefinite sequence of decisions

Decision Processes

Often an agent needs to go beyond a fixed set of decisions – Examples?

- Would like to have an **ongoing decision process**

Infinite horizon problems: process does not stop

Robot surviving on planet, Monitoring Nuc. Plant, . . .

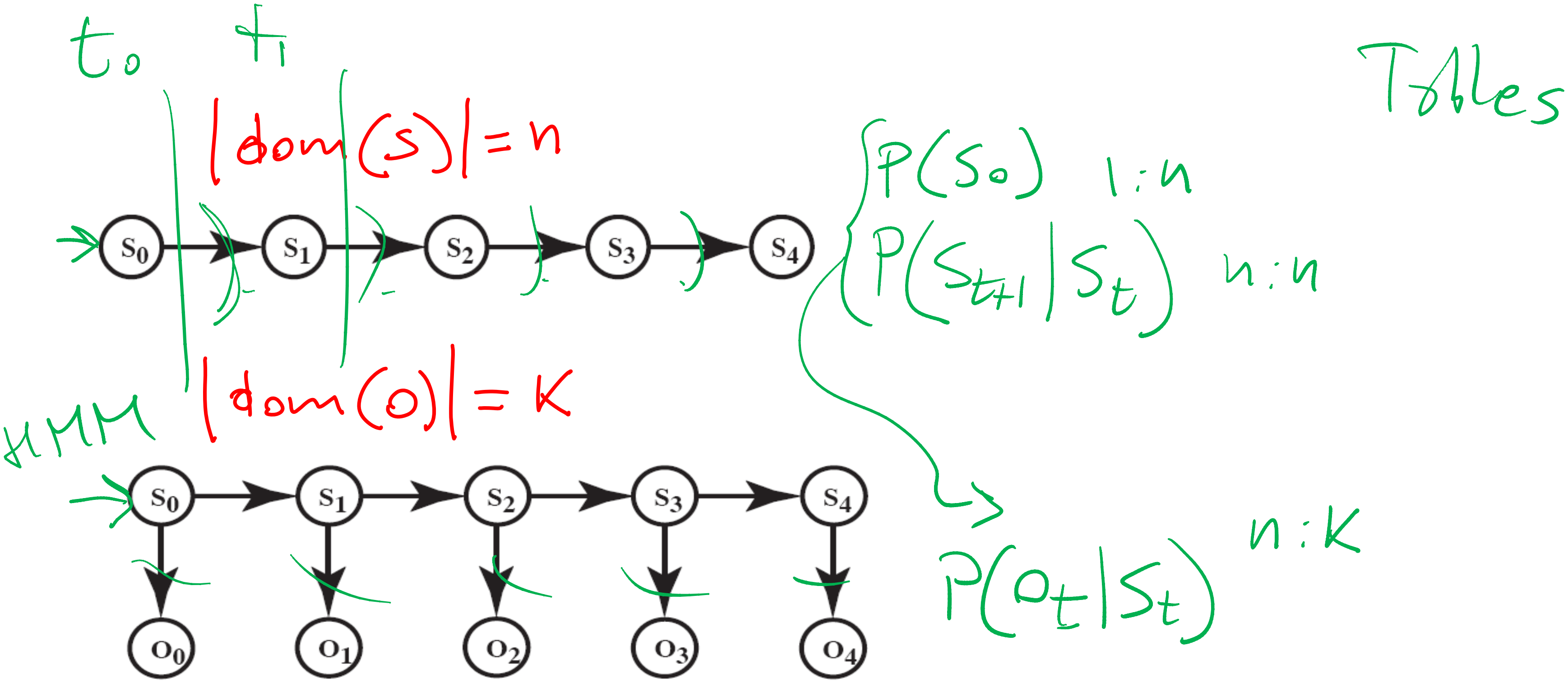
Indefinite horizon problem: the agent does not know when the process may stop

reaching location

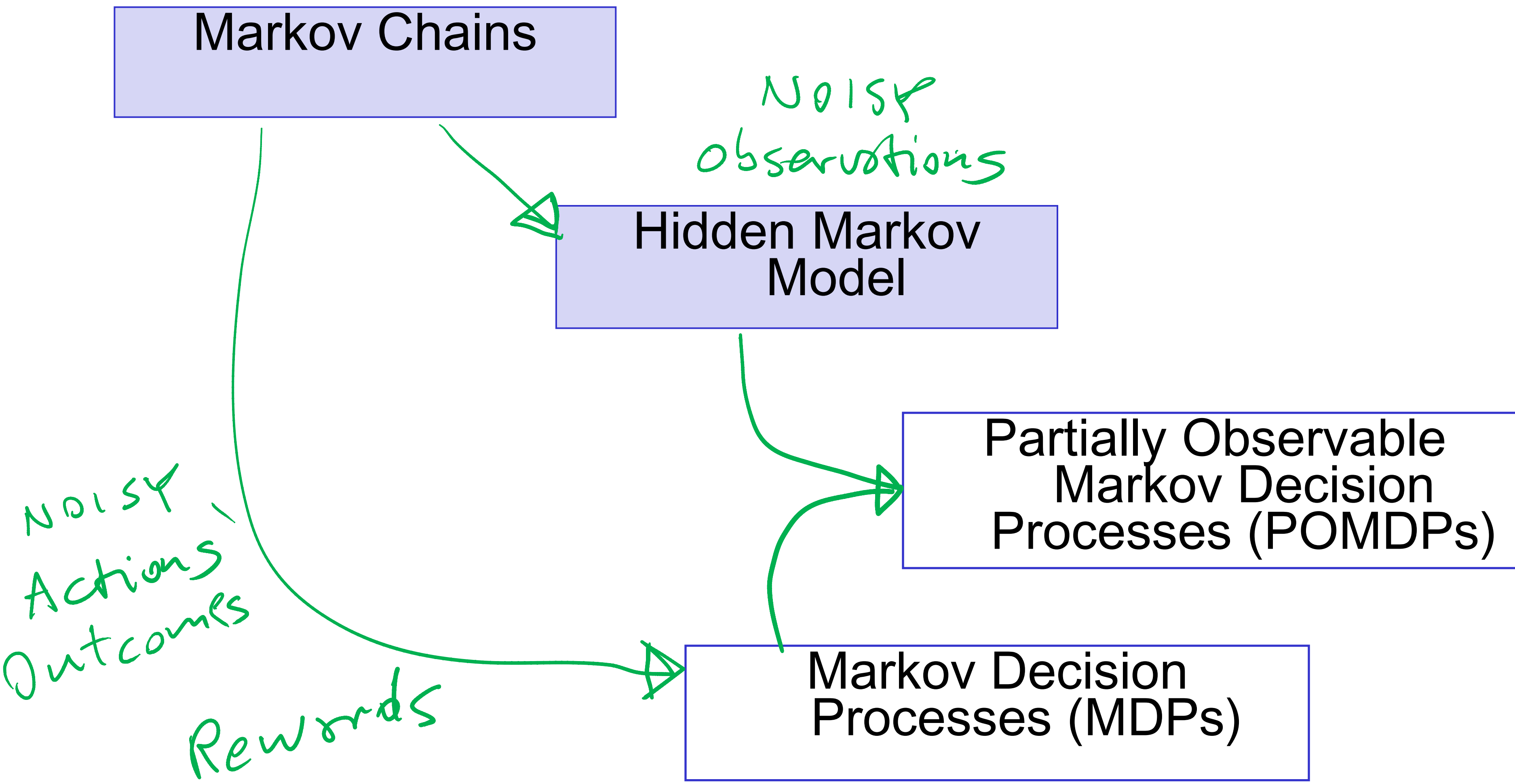
Finite horizon: the process must end at a give time N

in N steps

Recap: Markov Models



Markov Models



How can we deal with indefinite/infinite Decision processes?

We make the same two assumptions we made for….

The action outcome depends only on the current state — Markov

Let S_t be the state at time t …

$P(S_{t+1} \mid S_t, A_t, S_{t-1}, A_{t-1}, \ldots)$

$P(S_{t+1} \mid S_t, A_t)$

The process is *stationary*… — the same for all t

We also need a more flexible specification for the utility. How?

- Defined based on a reward/punishment *R(s)* that the agent receives in each state *s*

$s_0 \; s_1 \; \ldots \ldots \; s_n$

e.g. $\sum \; r_0 \; r_1 \; \ldots \ldots \; r_n$

MDP: formal specification

For an MDP you specify:

- set S of states and set A of actions
- the process' dynamics (or ***transition model***)

  $P(S_{t+1}|S_t, A_t)$
- The **reward function**

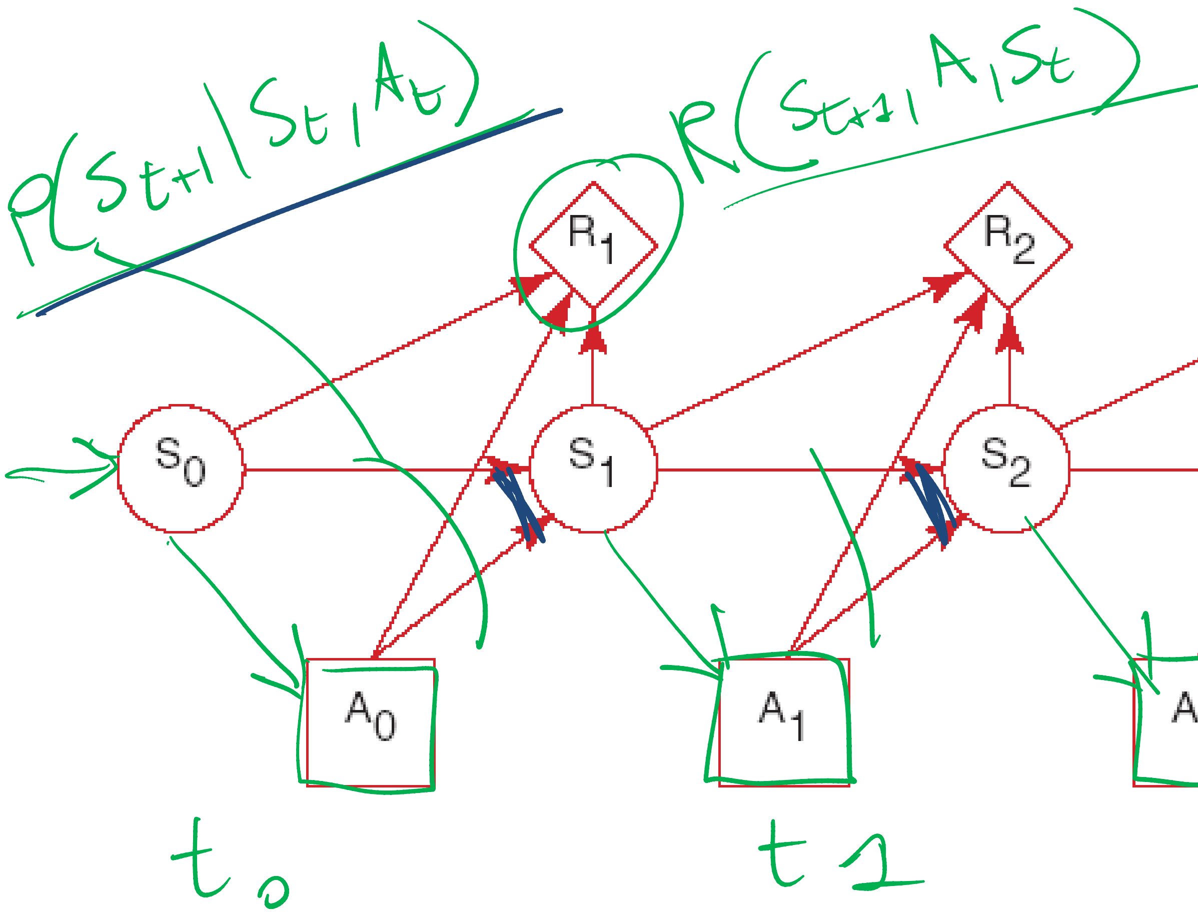
  $R(s, a, s')$

  describing the reward that the agent receives when it performs action *a* in state *s* and ends up in state *s'*
- $R(s)$ is used when the reward depends only on the state s and not on how the agent got there
- **Absorbing/stopping/terminal state** ← S_{ab}

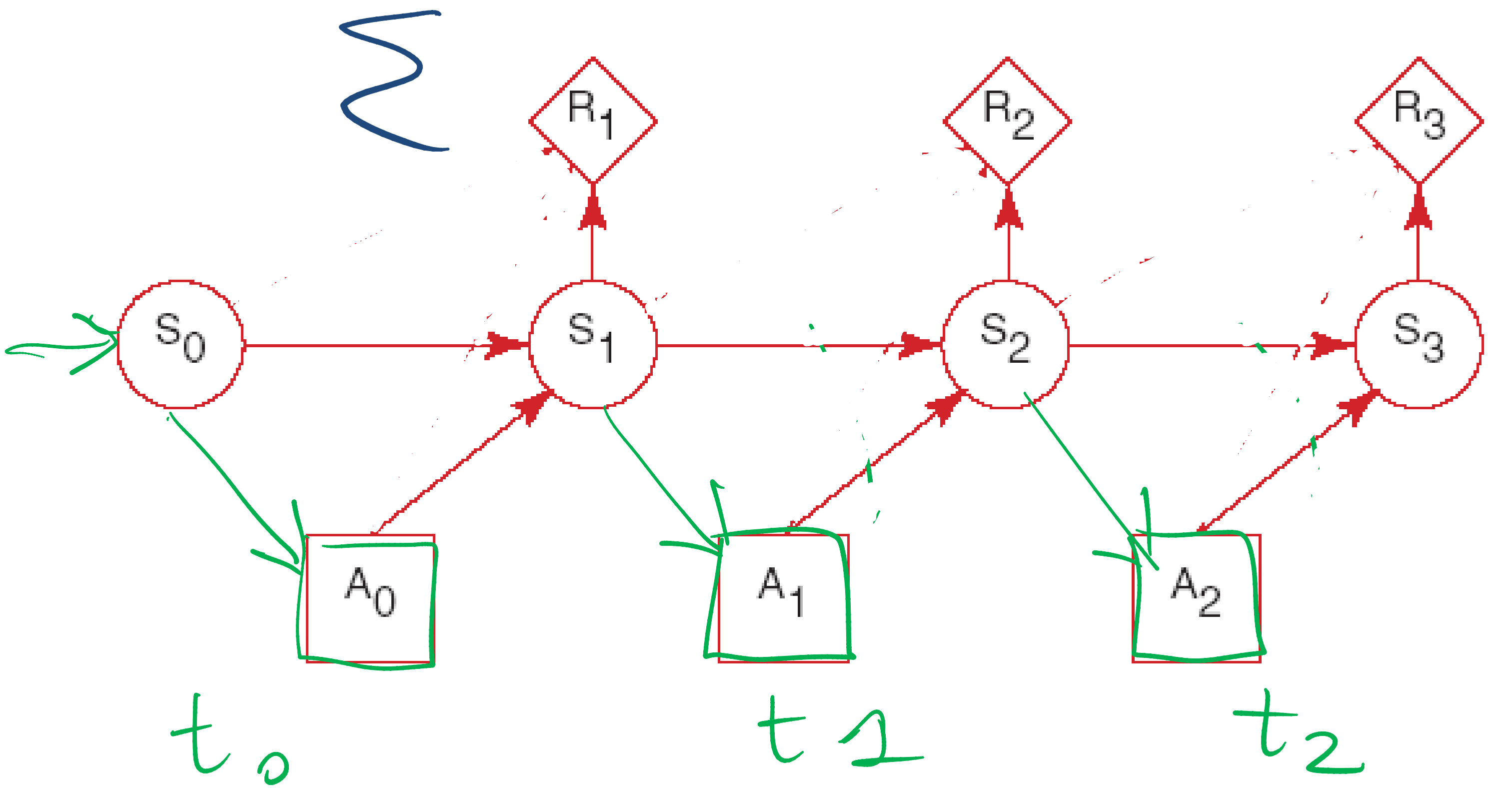
for all action $P(S_{ab} | a, S_{ab}) = 1$ $R(S_{ab}, a, S_{ab}) = 0$

MDP graphical specification

Basically a MDP is a Markov Chain augmented with actions and rewards/values



When Rewards only depend on the state



Learning Goals for today's class

You can:

- Define and compute Value of Information and Value of Control in a decision network
- Effectively represent indefinite/infinite decision processes with a Markov Decision Process (MDP)

TODO for this Fri

- assignment0 – Google Form
- Read textbook 9.5
  - 9.5.1 Value of a Policy
  - 9.5.2 Value of an Optimal Policy
  - 9.5.3 Value Iteration