Intelligent Systems (AI-2)

Computer Science cpsc422, Lecture 2

Jan, 13, 2021

Lecture Overview

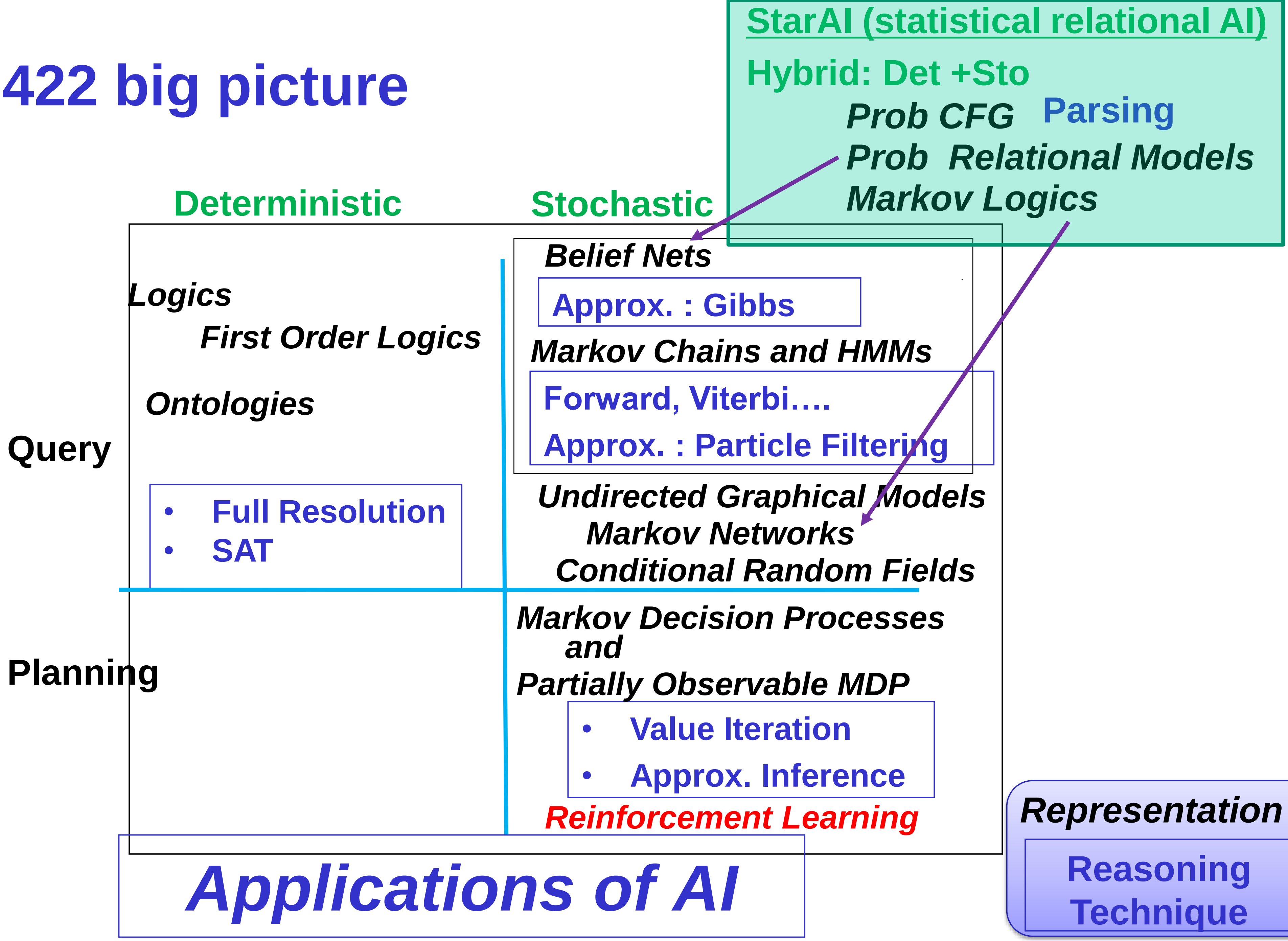
Value of Information and Value of Control

Recap Markov Chain

Markov Decision Processes (MDPs)

- Formal Specification and example

422 big picture



StarAI (statistical relational AI)

Hybrid: Det +Sto

Prob CFG **Parsing**

Prob Relational Models

Markov Logics

Deterministic **Stochastic**

Belief Nets

Approx. : Gibbs

Markov Chains and HMMs

Forward, Viterbi….

Approx. : Particle Filtering

Undirected Graphical Models

Markov Networks

Conditional Random Fields

Logics

First Order Logics

Ontologies

Query

- **Full Resolution**
- **SAT**

Planning

Markov Decision Processes and Partially Observable MDP

- **Value Iteration**
- **Approx. Inference**

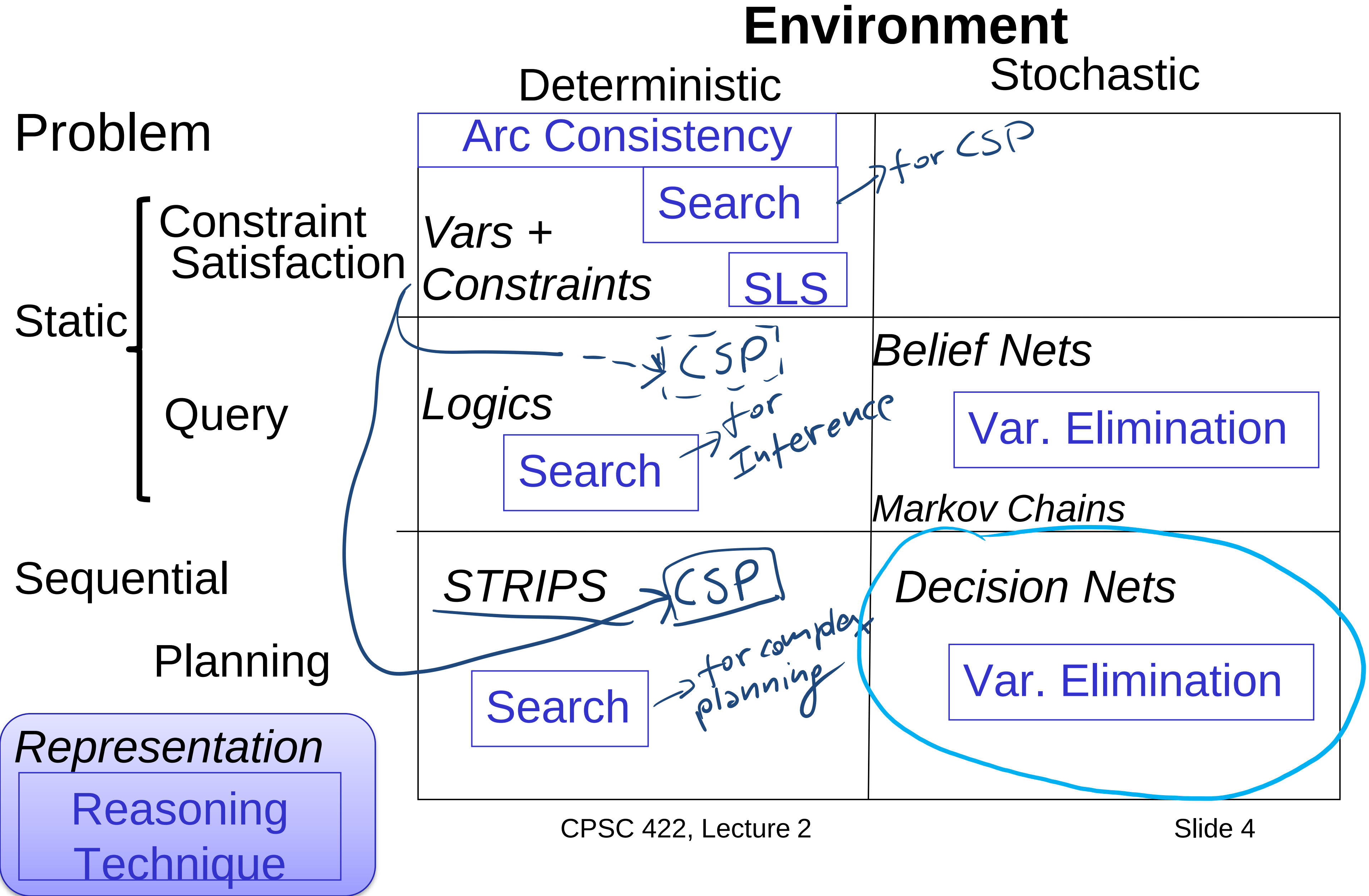
Reinforcement Learning

Applications of AI

Representation

Reasoning Technique

Cpsc 322 Big Picture



Problem		Environment: Deterministic	Environment: Stochastic
Static	Constraint Satisfaction	*Vars + Constraints* Arc Consistency Search SLS	
	Query	*Logics* Search	*Belief Nets* Var. Elimination *Markov Chains*
Sequential	Planning	*STRIPS* Search	*Decision Nets* Var. Elimination

Representation

Reasoning Technique

→ for CSP

CSP

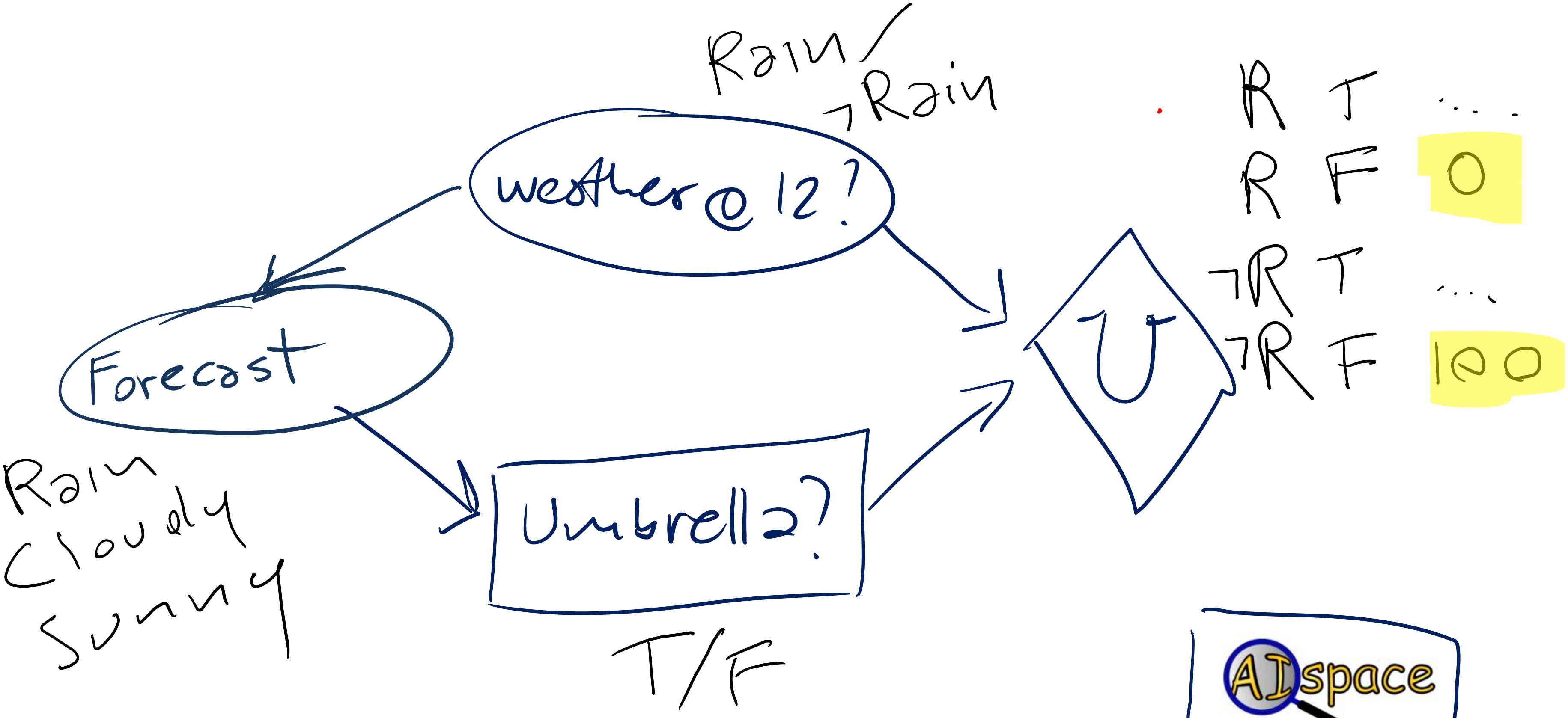
→ for Inference

CSP

→ for complex planning

Simple Decision Net

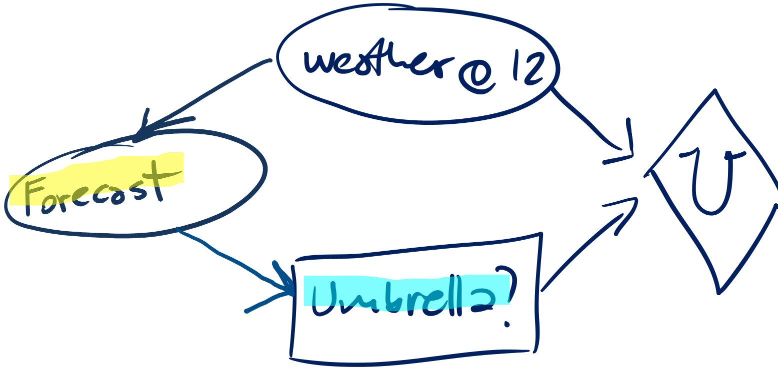
- Early in the morning. Shall I take my **umbrella** today? (I'll have to go for a long walk at noon)
- Relevant Random Variables… with reasonable values…



R	T	....
R	F	0
¬R	T	....
¬R	F	100

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?

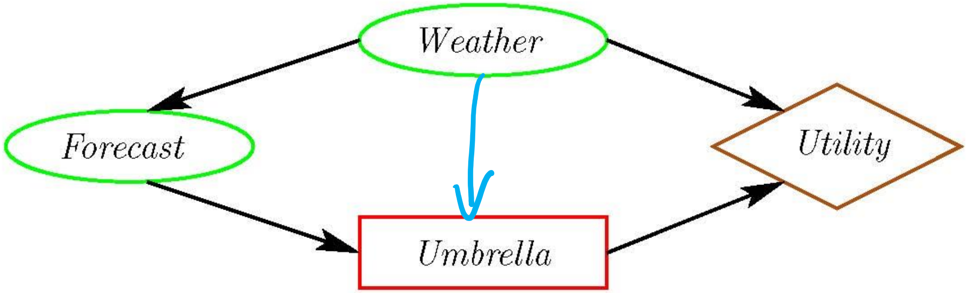
2^3

For this example

$|D|^{|pD|}$

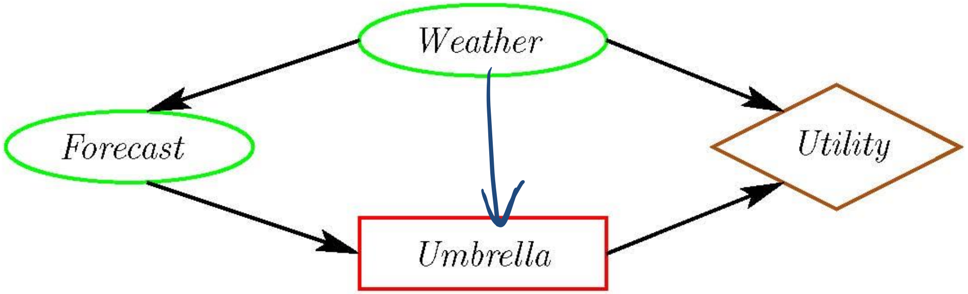
In general

Value of Information



- Early in the morning. I listen to the **weather forecas**t, 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



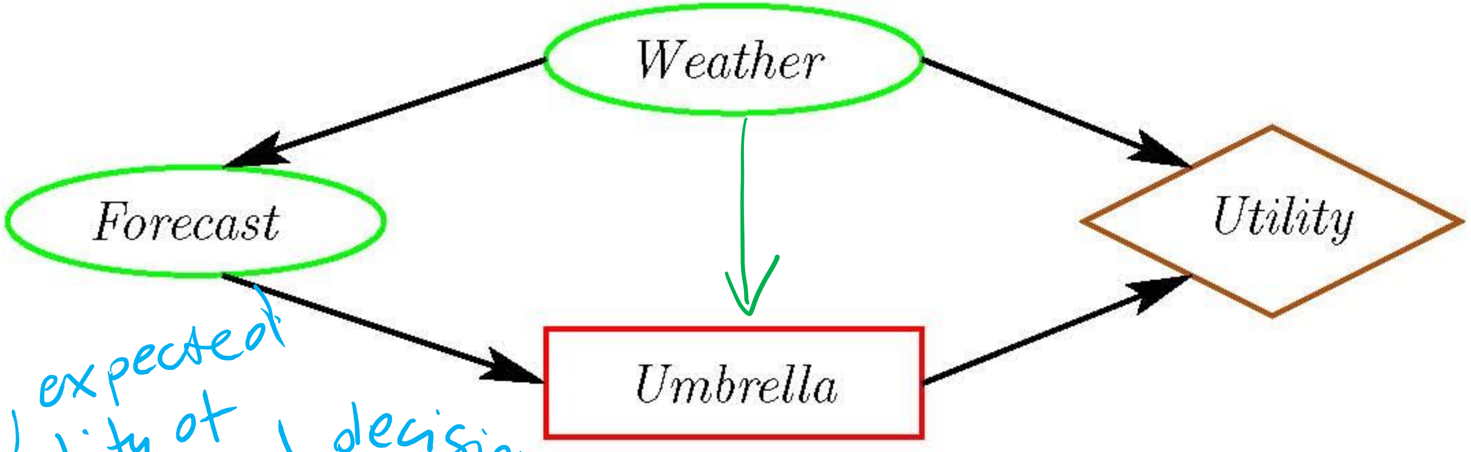
- The value of information of a random variable *X* for decision *D* is: EU(knowing X) − EU(not knowing X)

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?

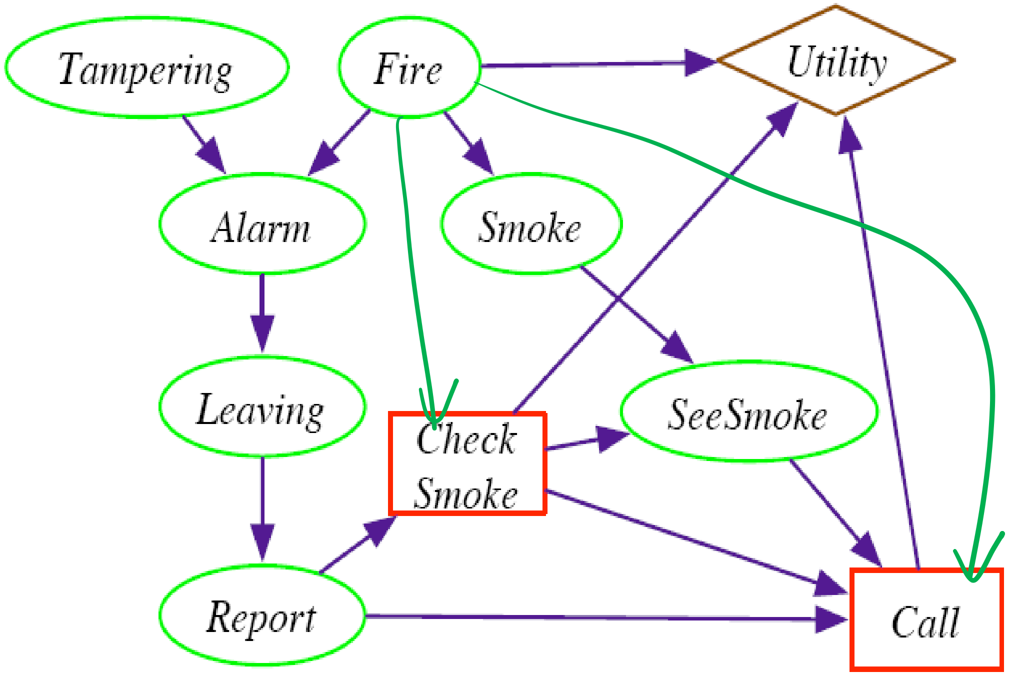


i.e., expected utility of optimal decision

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

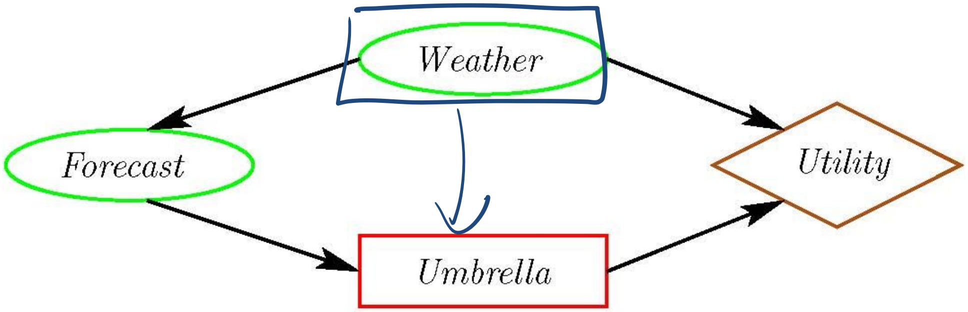
Value of Information

- 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

Value of Control



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

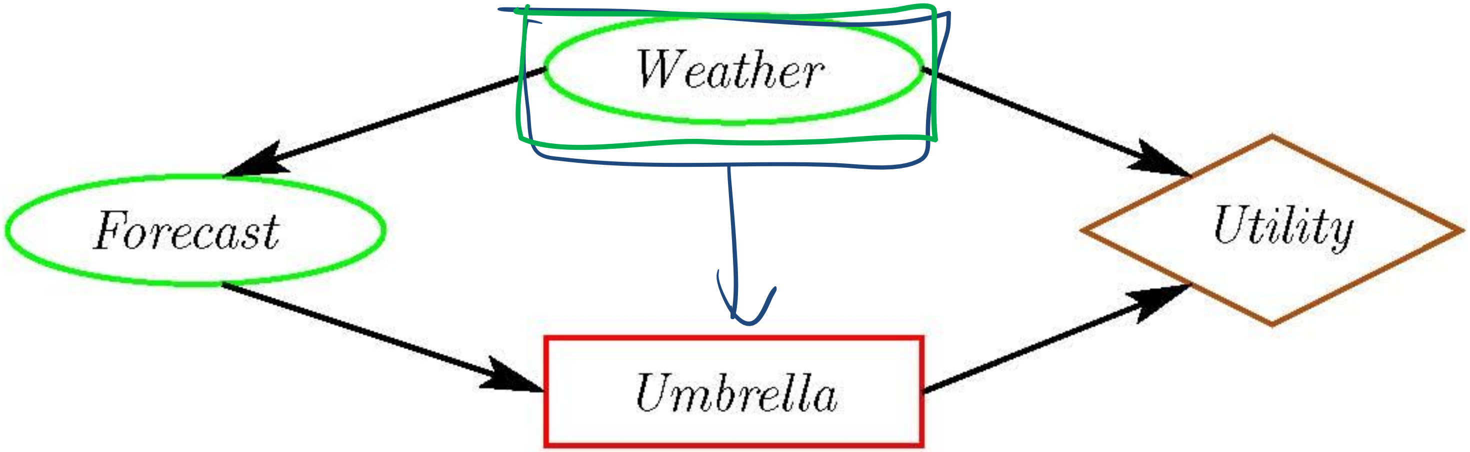
Weather	Umbrella	Value
Rain	true	70
Rain	false	0
noRain	true	20
noRain	false	100

- 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

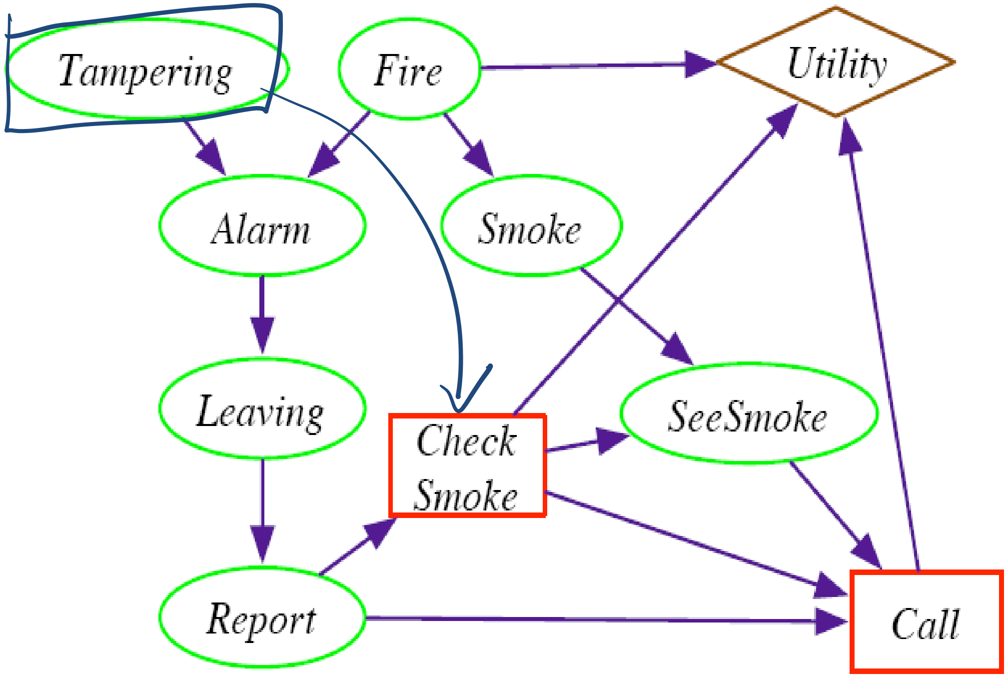
- What if we could control the weather?



- 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 optimal policy
- Conclusion: do not tamper with fire alarms!

Lecture Overview

Value of Information and Value of Control

Recap Markov Chain

Markov Decision Processes (MDPs)

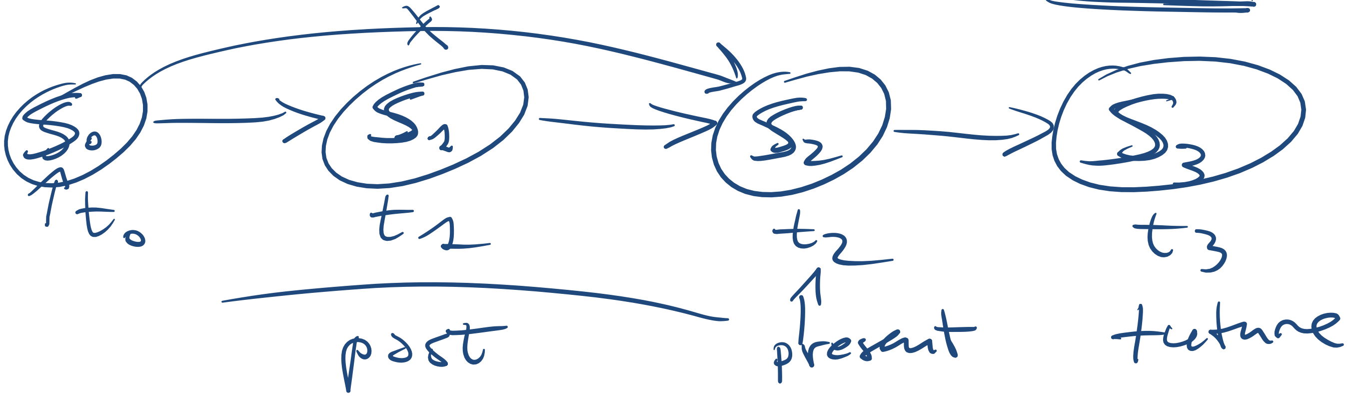
- Formal Specification and example

Lecture Overview (from my 322)

- **Recap**
- **Temporal Probabilistic Models**
- Start Markov Models
    - **Markov Chain**
    - Markov Chains in Natural Language Processing

Simplest Possible DBN

- One random variable for each time slice: let's assume S_t represents the state at time t. with domain $\{\underline{\boxed{v_1 \ldots v_n}}\}$

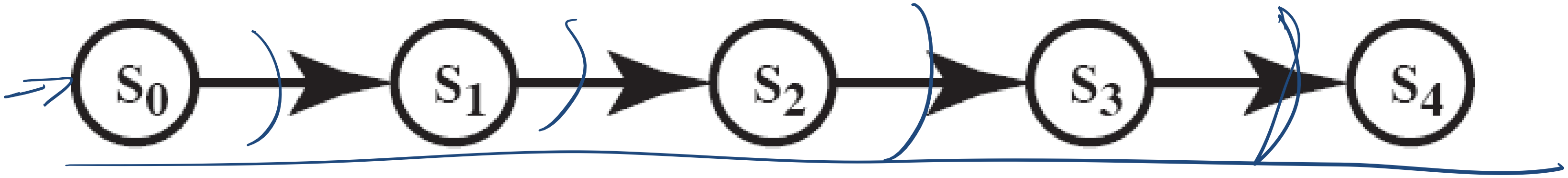


- Each random variable depends only on the previous one
- Thus $P(S_{t+1} \mid S_0 \cdots S_t) = P(\underline{S_{t+1}} \mid \underline{S_t})$

- Intuitively S_t conveys all of the information about the history that can affect the future states.
- "The future is independent of the past given the present."

Simplest Possible DBN (cont')

$P(S_4 | S_3)$



- How many CPTs do we need to specify?

4 $P(S_1 | S_0)$ $P(S_2 | S_1)$ etc.

A. 1

C . 2

D. 3

[✓ **B. 4**]

- *Stationary process assumption:* the mechanism that regulates how state variables change overtime is stationary, that is it can be described by a single transition model

- $P(S_t|S_{t-1})$ is the same for all t

Stationary Markov Chain (SMC)

A stationary Markov Chain : for all t >0

- $P(S_{t+1}| S_0,\dots,S_t) = P(S_{t+1}|S_t)$ and Markov assumption
- $P(S_{t+1}|S_t)$ is the same stationary

So we only need to specify?

[✓ **A.** $P(S_{t+1}|S_t)$ and $P(S_0)$]

B. $P(S_0)$

C . $P(S_{t+1}|S_t)$

D. $P(S_t|S_{t+1})$

Stationary Markov Chain (SMC)

A stationary Markov Chain : for all t >0

- $P(S_{t+1} | S_0, \ldots, S_t) = P(S_{t+1} | S_t)$ and Markov assumption
- $P(S_{t+1} | S_t)$ is the same stationary

We only need to specify $P(S_0)$ and $P(S_{t+2} | S_t)$

- Simple Model, easy to specify
- Often the natural model
- The network can extend indefinitely

Stationary Markov-Chain: Example

Domain of variable S_i is {t , q, p, a, h, e}

t	.6
q	.4
p	0
a	0
h	0
e	

Probability of initial state $P(S_0)$

Stochastic Transition Matrix $P(S_{t+1}|S_t)$

Which of these two is a possible STM?

$$\begin{array}{cc|cccccc} & & & & S_{t+1} & & & \\ & & t & q & p & a & h & e \\ \hline & t & 0 & .3 & 0 & .3 & .4 & 0 \\ & q & .4 & 0 & .6 & 0 & 0 & 0 \\ S_t & p & 0 & 0 & 1 & 0 & 0 & 0 \\ & a & 0 & 0 & .4 & .6 & 0 & 0 \\ & h & 0 & 0 & 0 & 0 & 0 & 1 \\ & e & 1 & 0 & 0 & 0 & 0 & 0 \end{array}$$

$$\begin{array}{cc|cccccc} & & & & S_{t+1} & & & \\ & & t & q & p & a & h & e \\ \hline & t & 1 & 0 & 0 & 0 & 0 & 0 \\ & q & 0 & 1 & 0 & 0 & 0 & 0 \\ S_t & p & .3 & 0 & 1 & 0 & 0 & 0 \\ & a & 0 & 0 & 0 & 1 & 0 & 0 \\ & h & 0 & 0 & 0 & 0 & 0 & 1 \\ & e & 0 & 0 & 0 & .2 & 0 & 1 \end{array}$$

$\Sigma > 1$

[✓ **A.** Left one only]

B. *Right one only*

C . *Both*

D. *None*

E. *I want to sleep*

Stationary Markov-Chain: Example

Domain of variable S_i is {t , q, p, a, h, e}

six possible values

We only need to specify…

$P(S_0)$

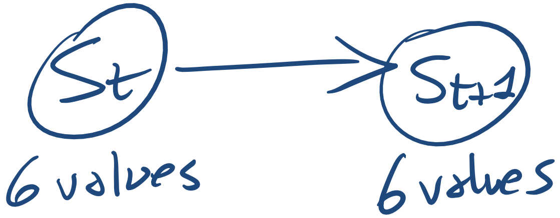
Probability of initial state

t	.6
q	.4
p	0
a	0
h	0
e	0

Stochastic Transition Matrix

$P(S_{t+1}|S_t)$

Q



S_{t+1}

S_t	t	q	p	a	h	e
t	0	.3	0	.3	.4	0
q	.4	0	.6	0	0	0
p	0	0	1	0	0	0
a	0	0	.4	.6	0	0
h	0	0	0	0	0	1
e	1	0	0	0	0	0

$\leftarrow P(S_{t+1}|S_t=q)$

$\leftarrow P(S_{t+1}|S_t=p)$

Markov-Chain: Inference

Probability of a sequence of states $S_0 \dots S_T$

$$\underbrace{P(S_0,\dots,S_T)} = P(S_0)\, P(S_1 | S_0)\, P(S_2 | S_1) \dots\dots$$

$P(u,e,e) \rightarrow$

$P(S_0)$

t	.6
q	.4
p	0
a	0
h	0
e	0

$P(S_{t+1}|S_t)$

	t	q	p	a	h	e
t	0	.3	0	.3	.4	0
q	.4	0	.6	0	0	0
p	0	0	1	0	0	0
a	0	0	.4	.6	0	0
h	0	0	0	0	0	1
e	1	0	0	0	0	0

Example:

$$\underline{P(t, q, p)} =$$

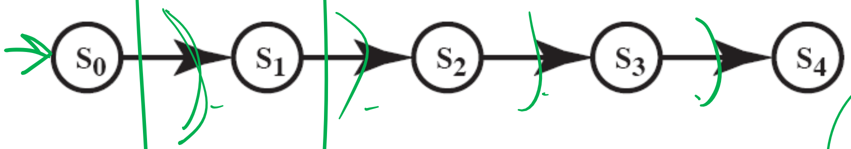
$$P(t) * P(q|t) * P(p|q)$$

$$.6 * .3 * .6 = .108$$

Recap: Markov Models

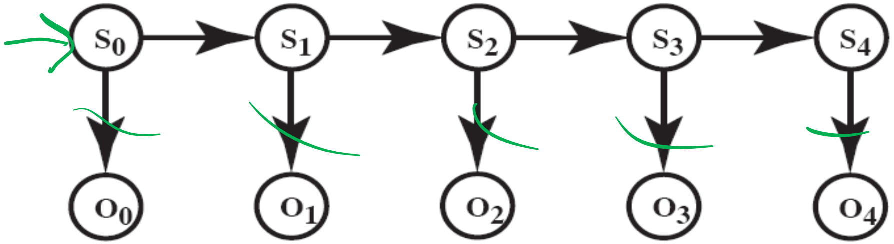
t_0 t_1

$|dom(S)| = n$



$|dom(O)| = K$

HMM



Tables

$P(S_0)$ $1:n$

$P(S_{t+1}|S_t)$ $n:n$

$P(O_t|S_t)$ $n:K$

Lecture Overview

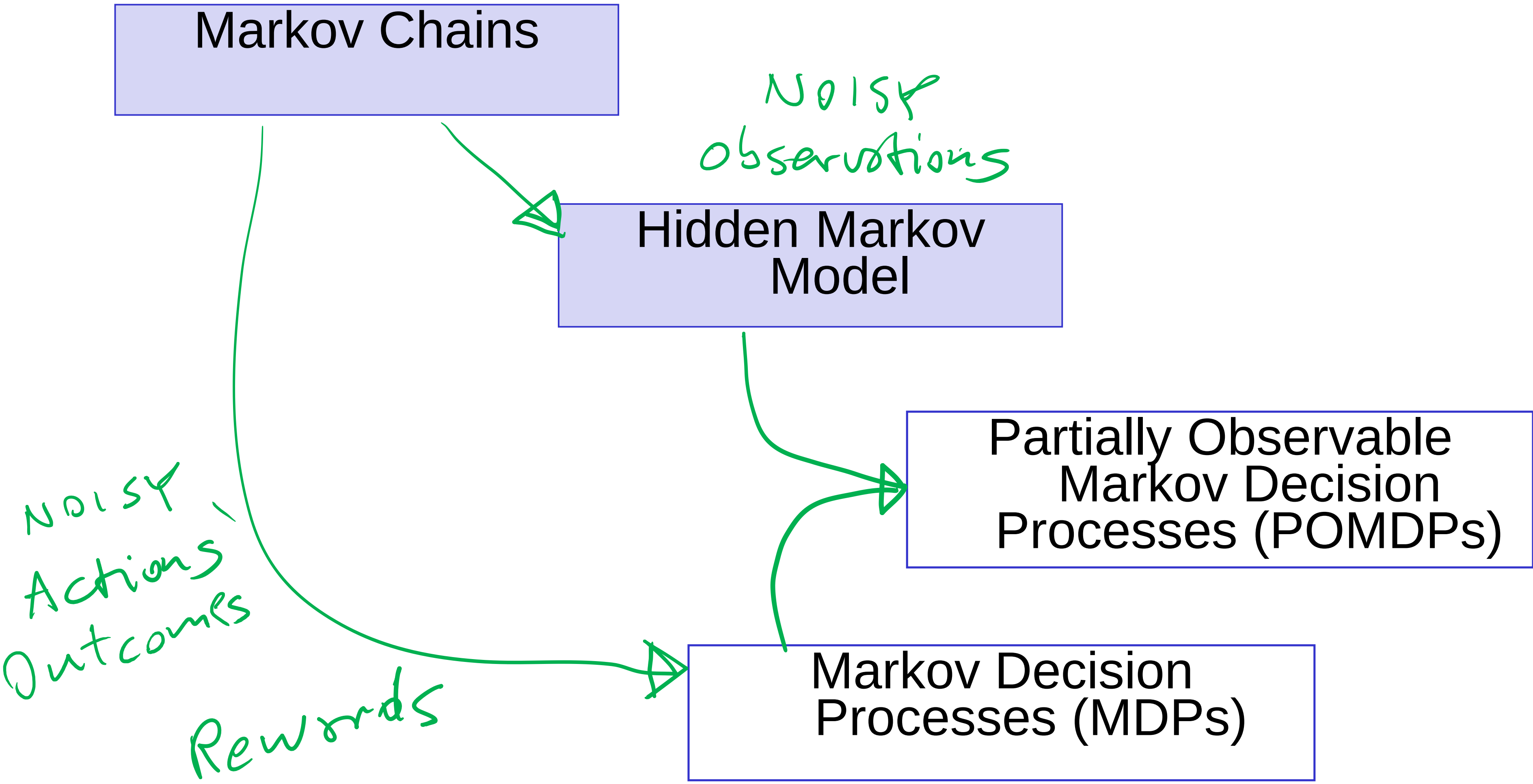
Value of Information and Value of Control

Recap Markov Chain

Markov Decision Processes (MDPs)

- Formal Specification and example

Markov Models



Markov Chains

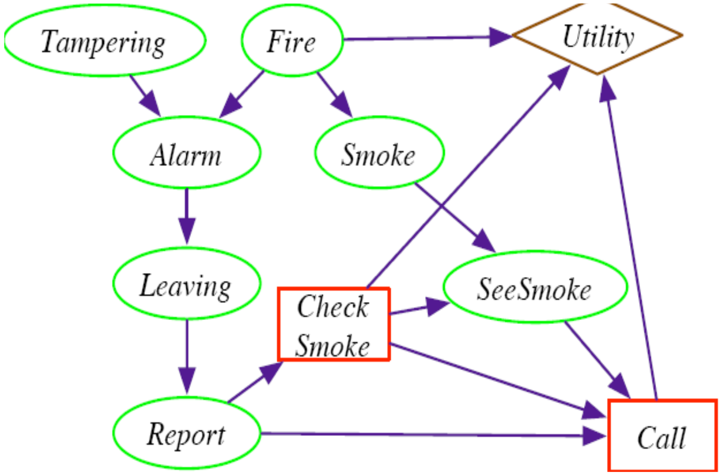
Hidden Markov Model

Partially Observable Markov Decision Processes (POMDPs)

Markov Decision Processes (MDPs)

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

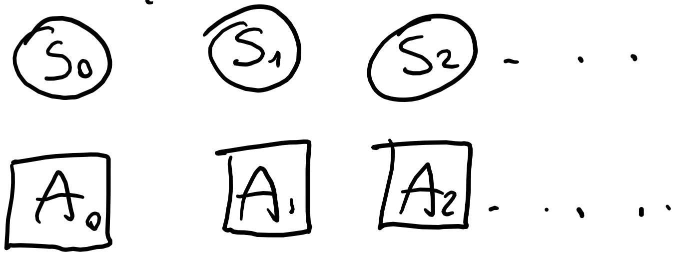
Finite horizon: the process must end at a give time N

in N steps

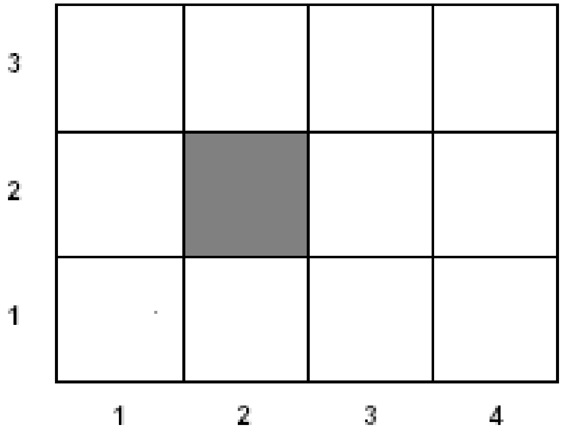
How can we deal with indefinite/infinite Decision processes?

Like in a Markov Chain one random variable S_t for each time slice represents the state at time *t*.

And A_t be the action/decision at time *t*



Example



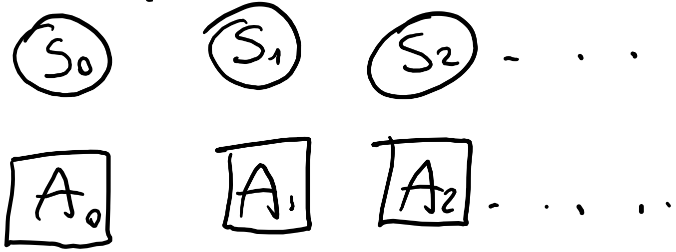
Agent moves in the above grid :

- States are the cells of the grid
- Actions *Up, Down, Left, Right*

How can we deal with indefinite/infinite Decision processes?

Like in a Markov Chain one random variable S_t for each time slice represents the state at time *t*.

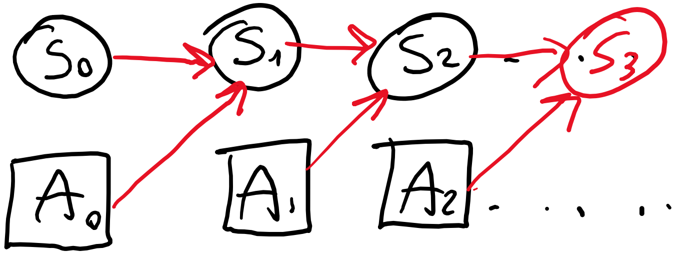
And A_t be the action/decision at time *t*



Make the same two assumptions we made for Markov Chains

(a) The action outcome (the state S_{t+1} at time *t+1*) only depends on S_t and A_t

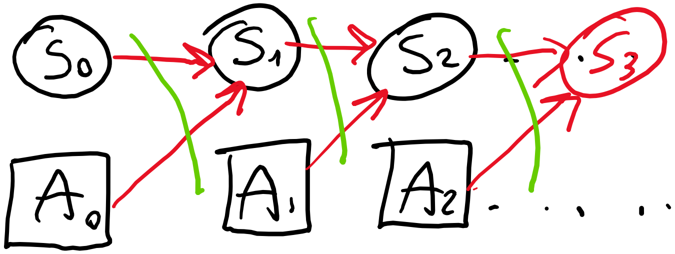
$$P(S_{t+1}|S_t,A_t,S_{t-1},A_{t-1}\ldots)=$$
$$= P(S_{t+1}|S_t,A_t)$$



(b) The process is *stationary…*

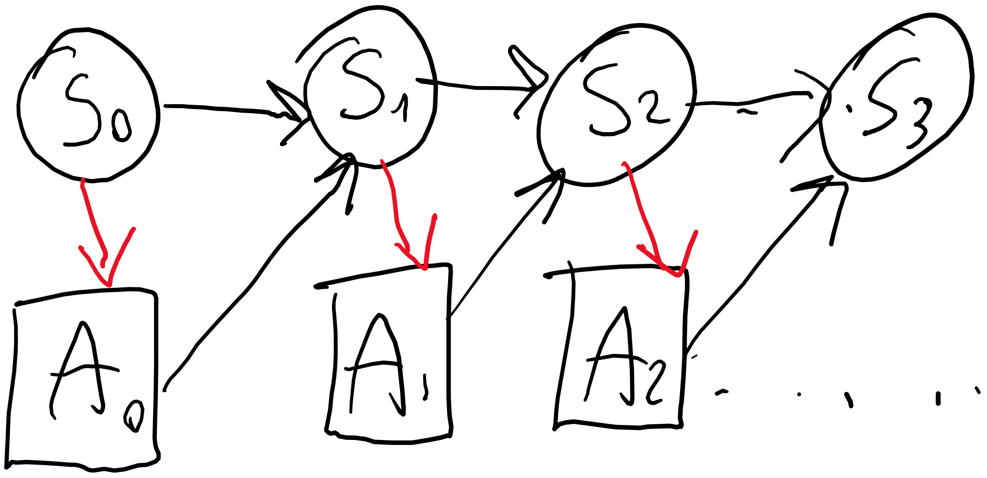
$P(S_{t+1}|S_t,A_t)$

the same for all t

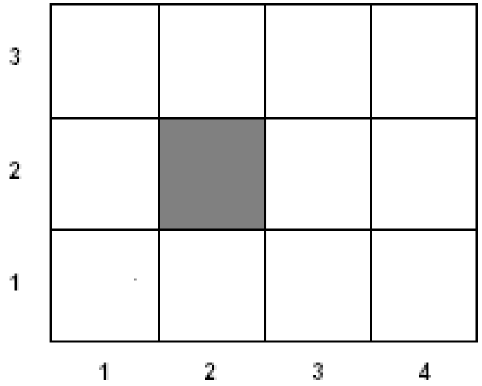


same conditional probabilities

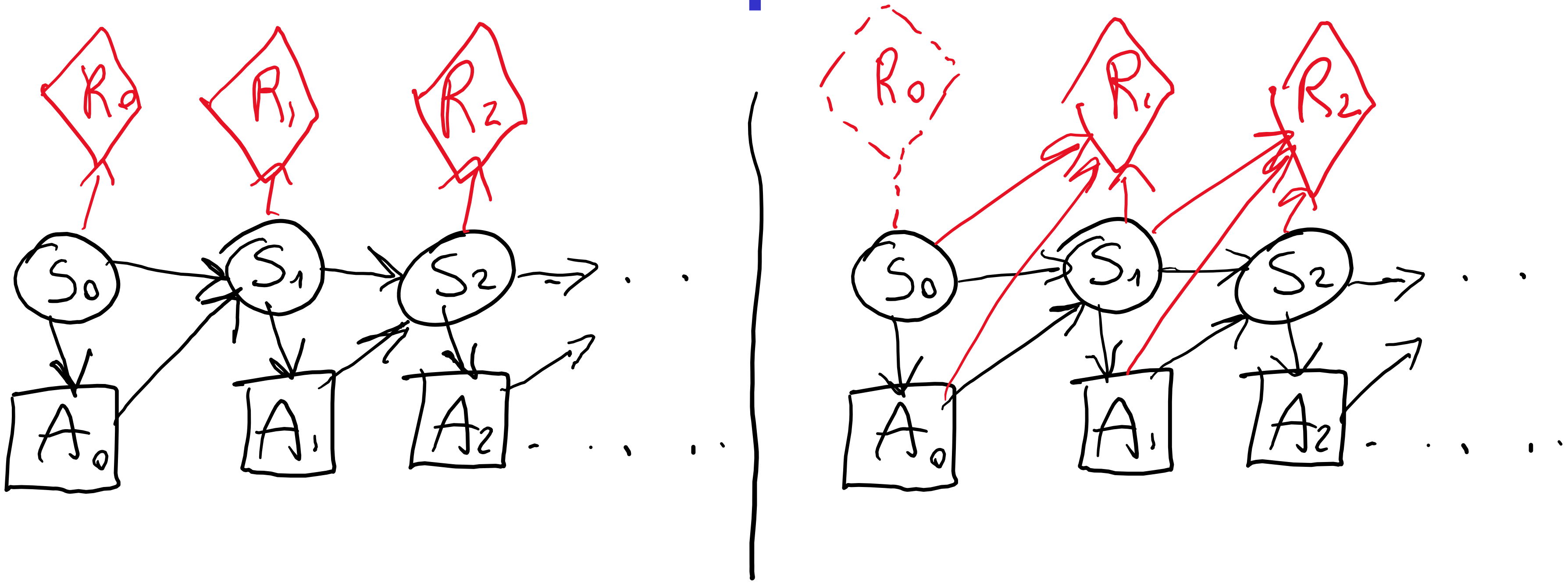
Agent knows in which state it is at time **t** when it selects the action at time **t**



The robot always knows in
Which cell it is…



How can we deal with indefinite/infinite Decision processes?

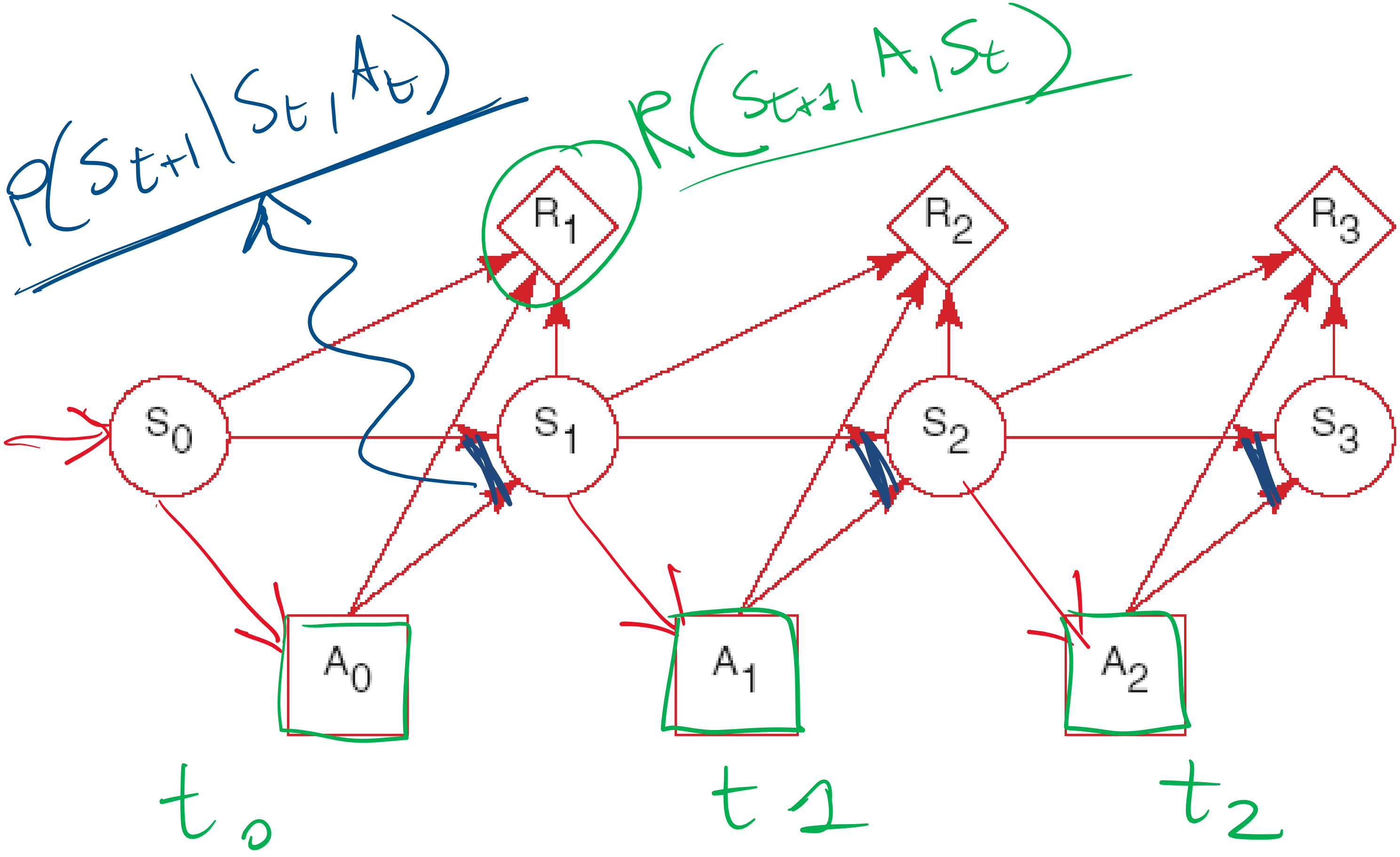


We also need a more **flexible specification for the utility**. How?

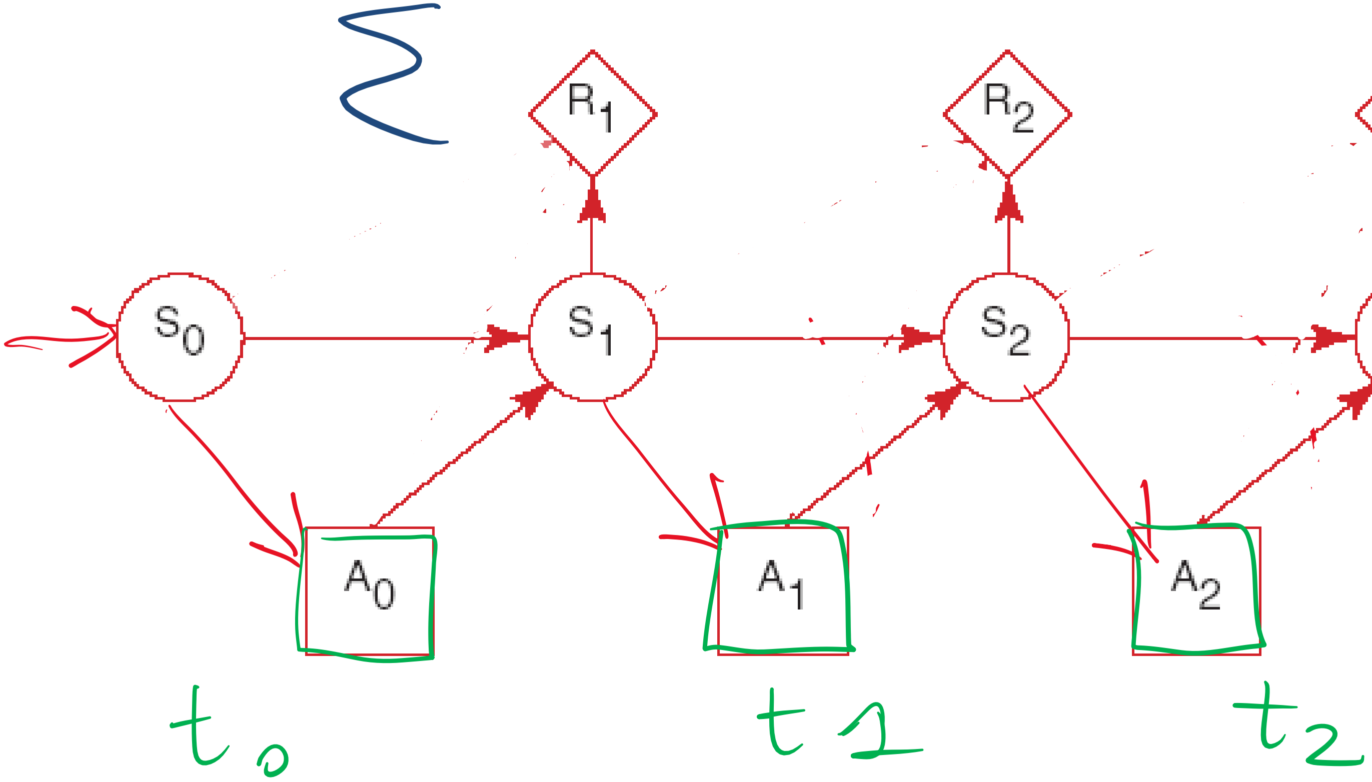
- Defined based on a reward/punishment that the agent receives in each time slice
- Typically summing them up

MDP graphical specification

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

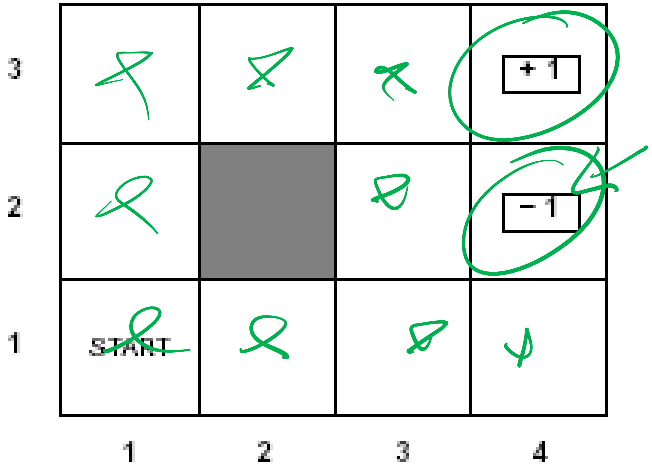


When Rewards only depend on the state



CPSC 422, Lecture 2

Example MDP: Rewards



$$R(s) = \begin{cases} -0.04 & \text{(small penalty) for nonterminal states} \\ \pm 1 & \text{for terminal states} \end{cases}$$

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

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

Summary Decision Processes: MDPs

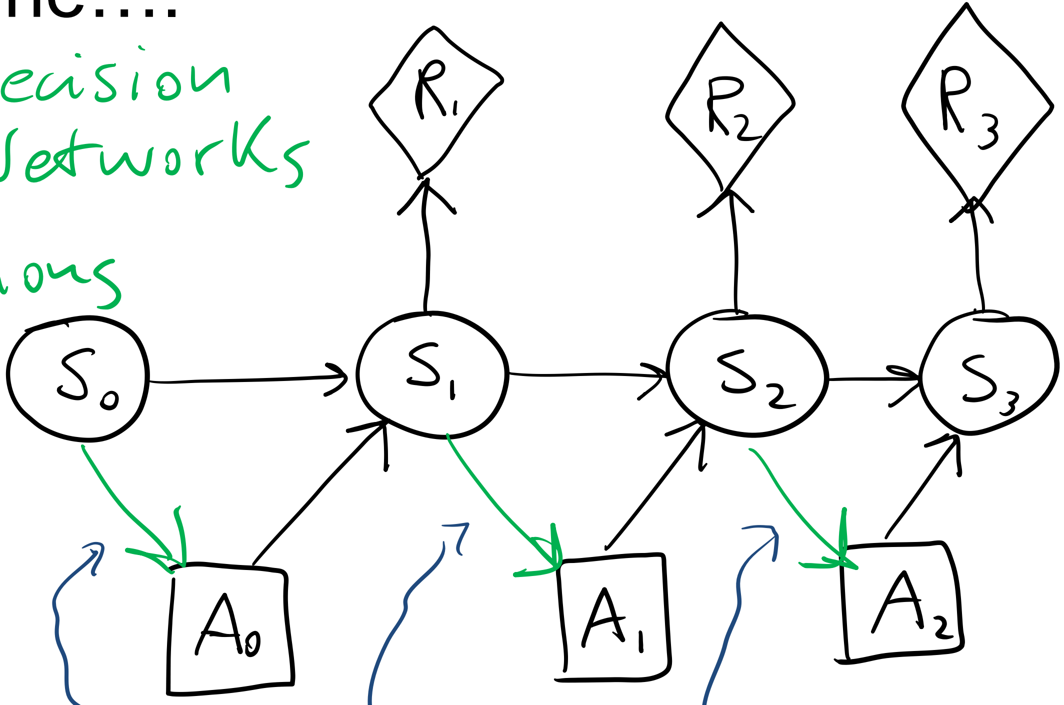
To manage an ongoing (indefinite… infinite) decision process, we combine….

Markov Chains & Decision Networks

Markovian

Stationary

Assumptions



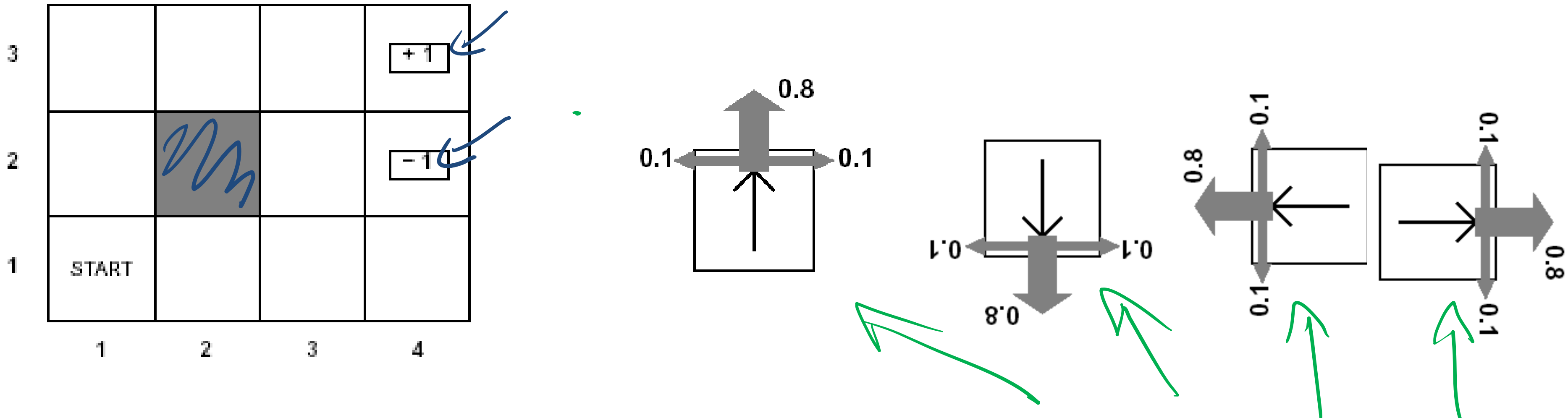
Utility not just at the end

BUT

Sequence of rewards

Fully Observable

Example MDP: Scenario and Actions



Agent moves in the above grid via actions *Up, Down, Left, Right*

Each action has:

- 0.8 probability to reach its intended effect
- 0.1 probability to move at right angles of the intended direction
- If the agents bumps into a wall, it stays there

How many states? 11 11 12

There are two terminal states (3,4) and (2,4)

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)
- Compute the probability distribution on states given a sequence of actions in an MDP
- Define a policy for an MDP

TODO for Fri

- **Read textbook 9.4**
- **Read textbook 9.5**
  - **9.5.1 Value of a Policy**

  - **9.5.2 Value of an Optimal Policy**
  - **9.5.3 Value Iteration**

CPSC 322 Review “Exam”

https://forms.gle/SpQwrXfonTZrVf4P7

Based on CPSC 322 material

- Logic
- Uncertainty
- Decision Theory

Review material (e.g., 322 slides from 2017):

https://www.cs.ubc.ca/~carenini/TEACHING/CPSC322-17S/index.html