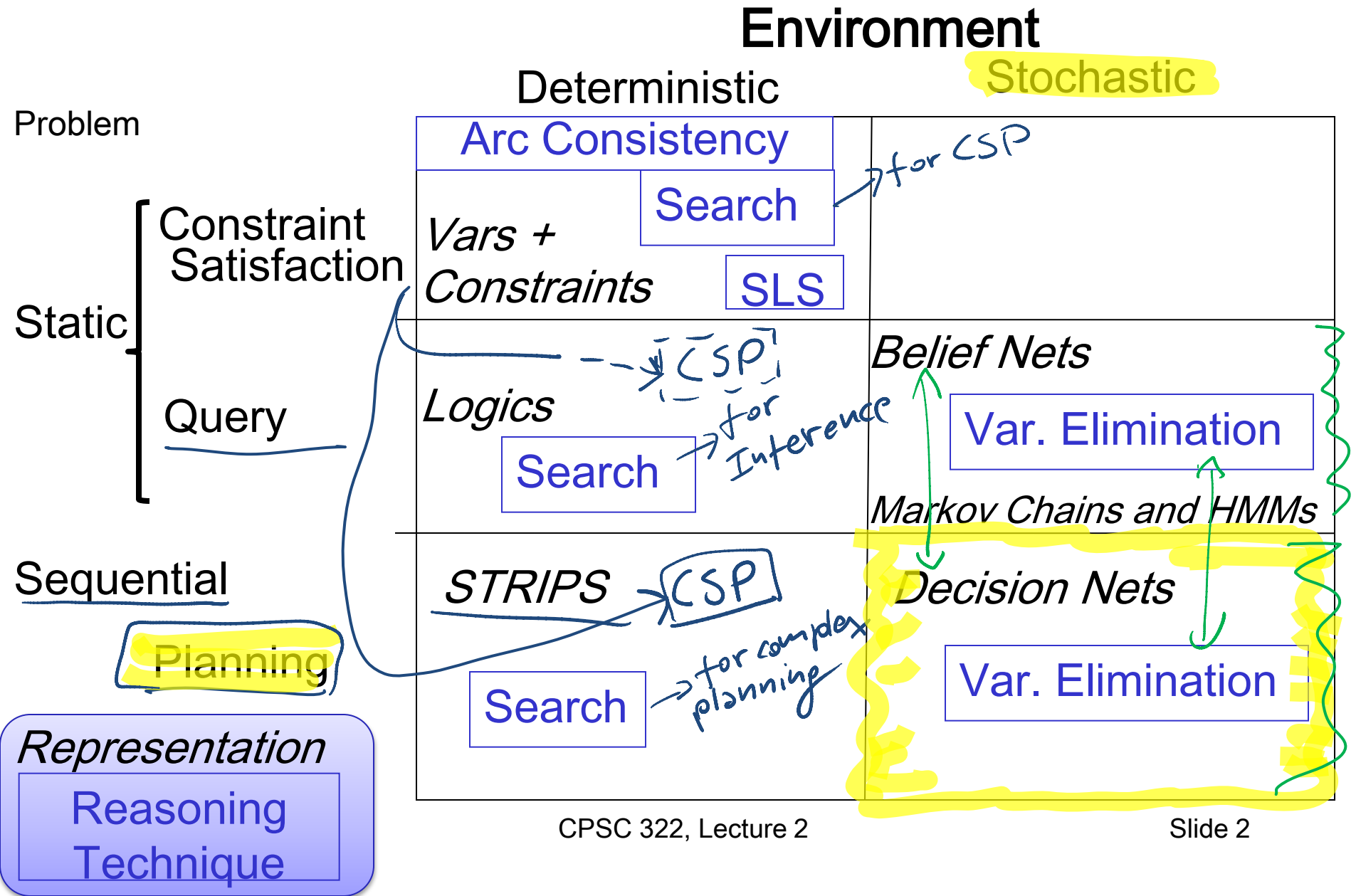
Planning Under Uncertainty

Computer Science cpsc322, Lecture 11

(Textbook Chpt 9.1-3)

June 12, 2012

Planning in Stochastic Environments



Planning Under Uncertainty: Intro

- **Planning** how to select and organize a sequence of actions/decisions to achieve a given goal.
- **Deterministic Goal**: A possible world in which some propositions are true

- **Planning under Uncertainty**: how to select and organize a sequence of actions/decisions to "*maximize the probability*" of *"achieving a given goal"*
- **Goal under Uncertainty**: we'll move from all-or-nothing goals to a richer notion: rating how *happy* the agent is in different possible worlds.

“Single” Action vs. Sequence of Actions

Set of primitive decisions that can be treated as a single macro decision to be made *before acting*

one-off

- Agents makes observations
- Decides on an action
- Carries out the action

Sequential Decisions

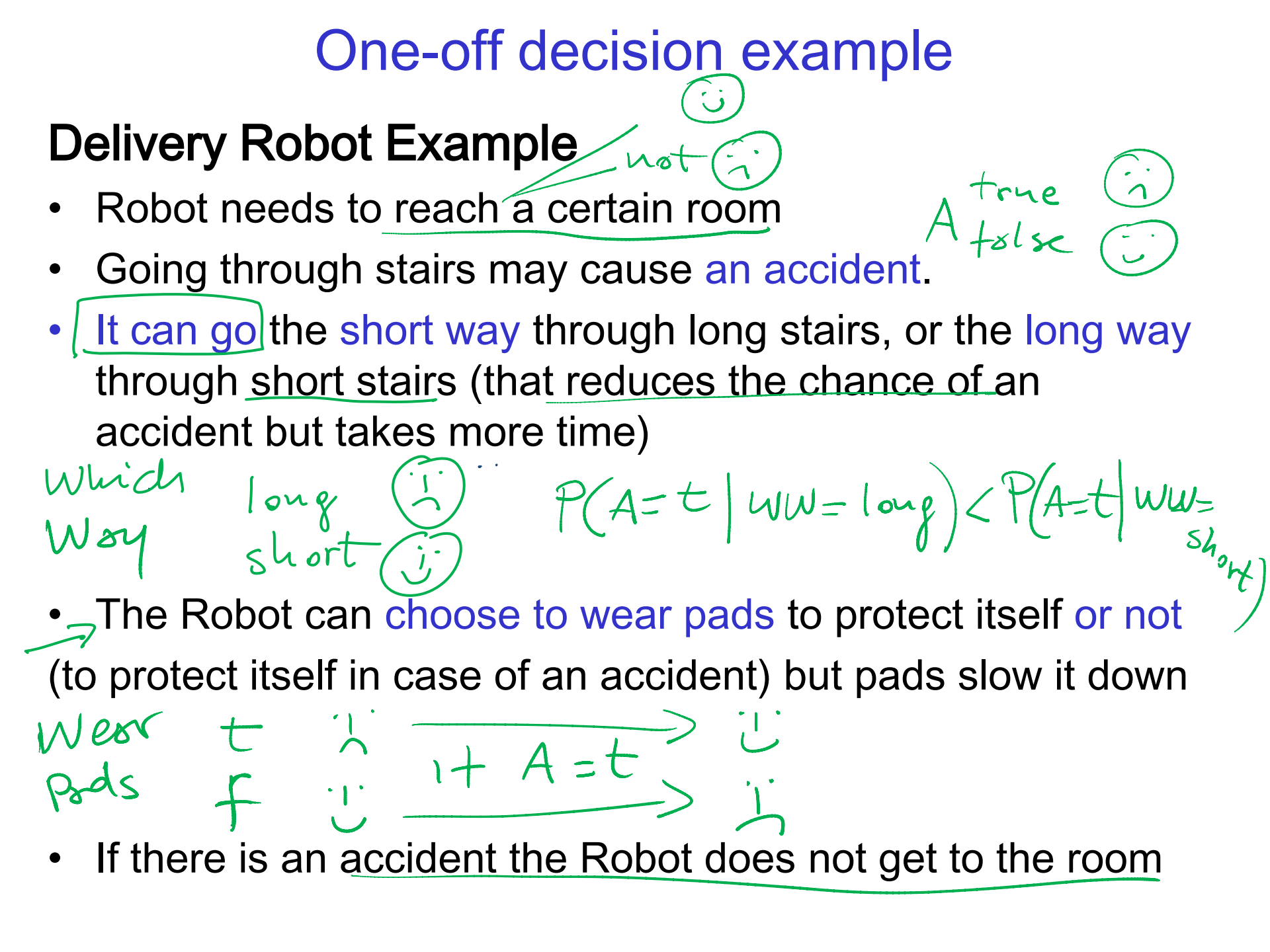
Lecture Overview

- **One-Off Decision**
  - Example
  - Optimal Decision: Utilities / Preferences
  - Single stage Decision Networks
- Sequential Decisions
  - Representation
  - Policies
  - Finding Optimal Policies

One-off decision example

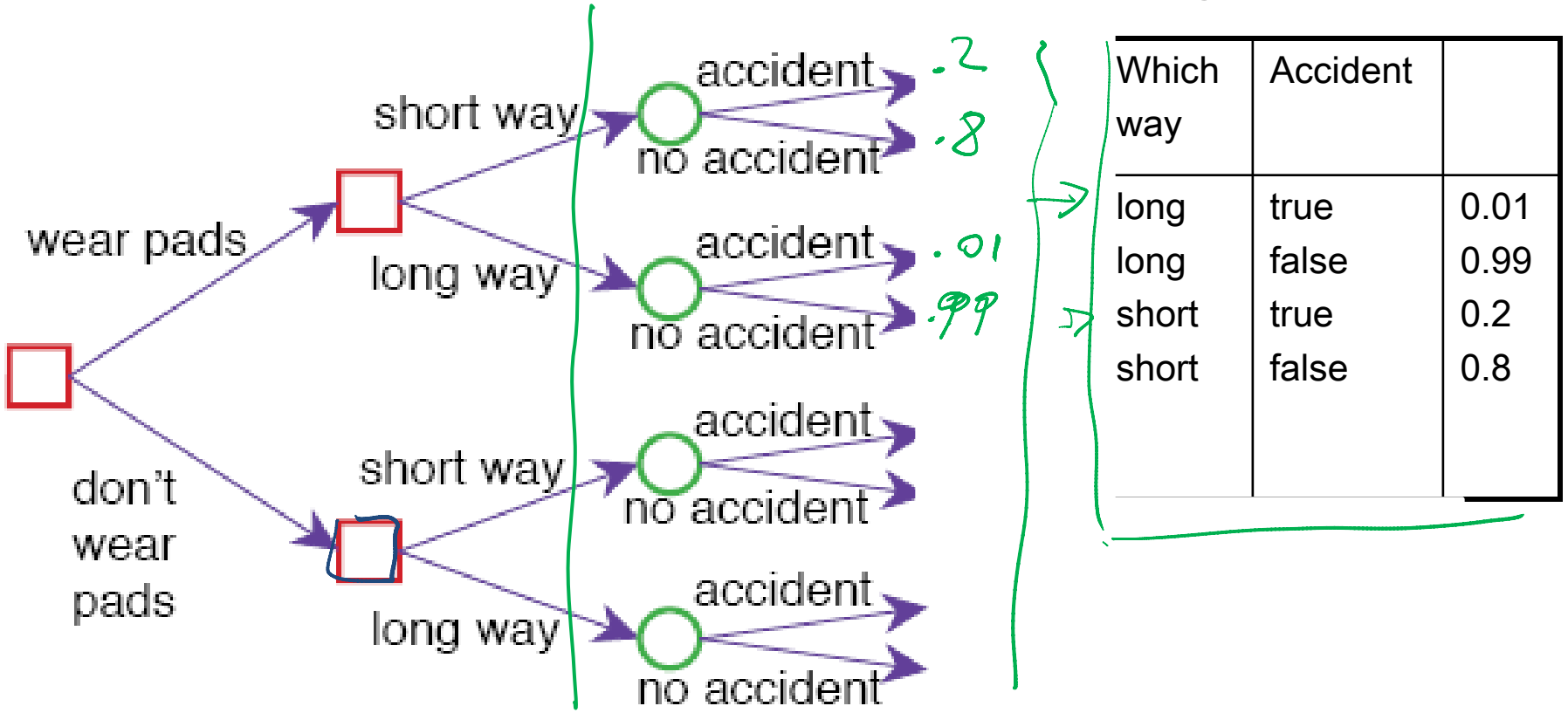
Delivery Robot Example

- Robot needs to reach a certain room
- Going through stairs may cause an accident.
- It can go the short way through long stairs, or the long way through short stairs (that reduces the chance of an accident but takes more time)
- The Robot can choose to wear pads to protect itself or not (to protect itself in case of an accident) but pads slow it down
- If there is an accident the Robot does not get to the room



Decision Tree for Delivery Robot

- This scenario can be represented as the following decision tree

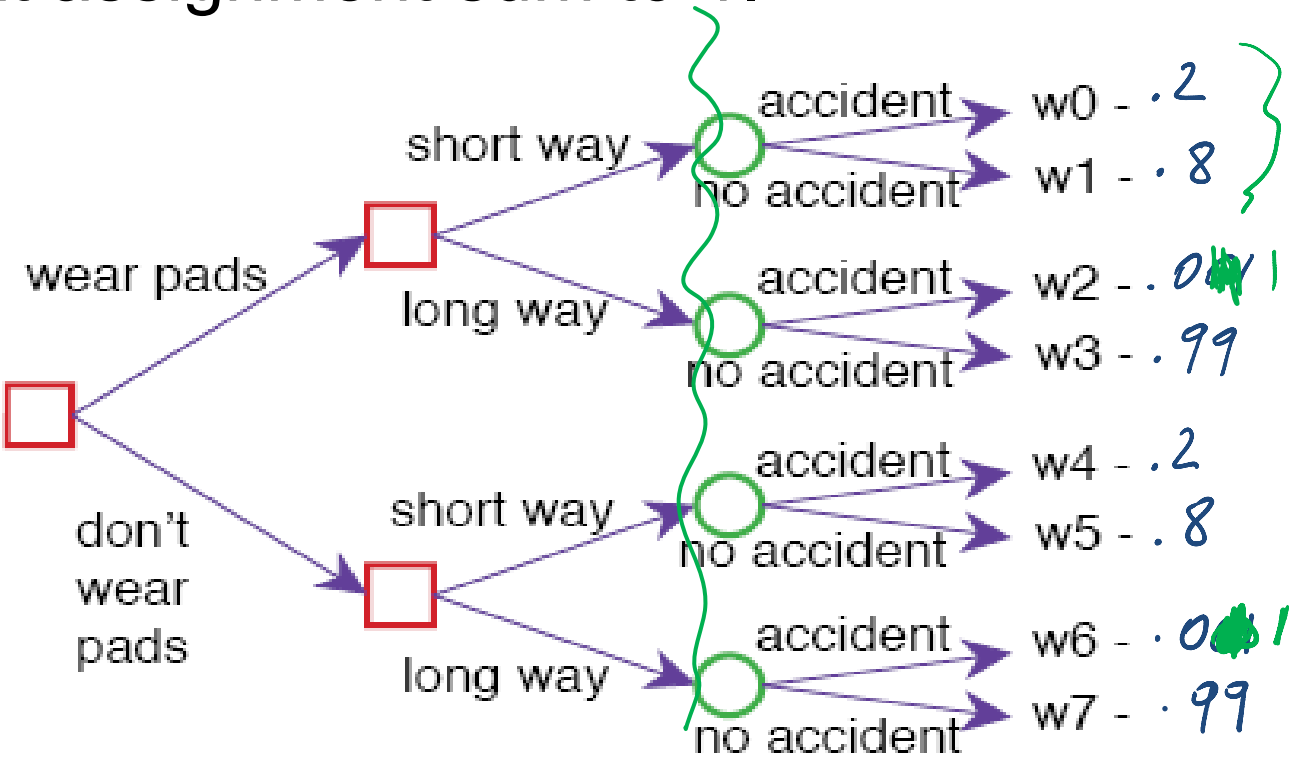


Which way	Accident	
long	true	0.01
long	false	0.99
short	true	0.2
short	false	0.8

- The agent has a set of decisions to make (a macro-action it can perform)
- Decisions can influence random variables
- Decisions have probability distributions over outcomes

Decision Variables: Some general Considerations

- A possible world specifies a value for each random variable and each decision variable.
- For each assignment of values to all decision variables, the probabilities of the worlds satisfying that assignment sum to 1.



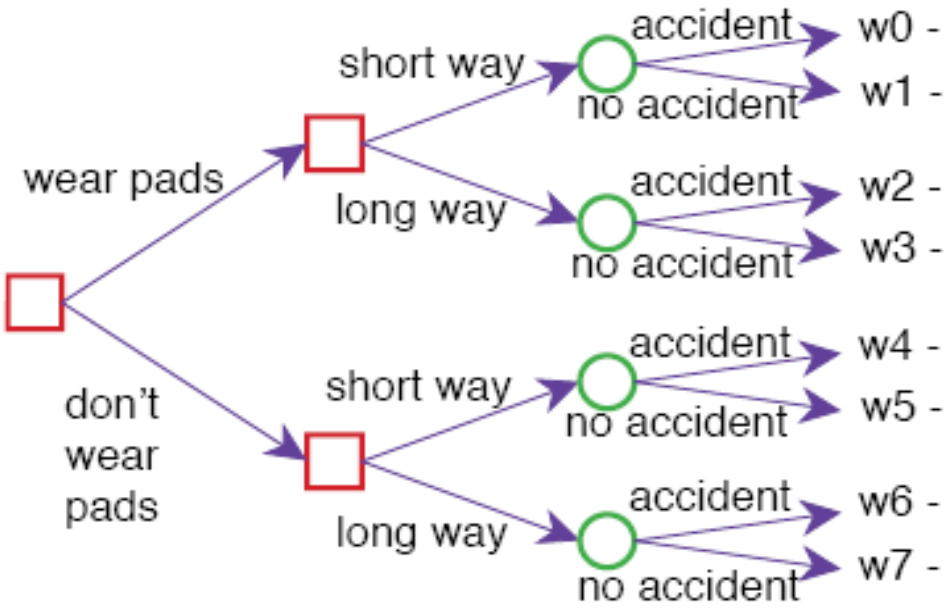
Lecture Overview

- One-Off Decision
  - Example
  - **Optimal Decision: Utilities / Preferences**
  - Single stage Decision Networks
- Sequential Decisions
  - Representation
  - Policies
  - Finding Optimal Policies

What are the optimal decisions for our Robot?

It all depends on how happy the agent is in different situations.

For sure getting to the room is better than not getting there….. but we need to consider other factors..



Utility / Preferences

Utility: a measure of desirability of possible worlds to an agent

- Let U be a real-valued function such that $U(w)$ represents an agent's degree of preference for world w . [0, 100]

Would this be a reasonable utility function for our Robot?

Which way	Accident	Wear Pads	Utility	World
short	true	true	35	w0, moderate damage
short	false	true	95	w1, reaches room, quick, extra weight
long	true	true	30	w2, moderate damage, low energy
long	false	true	75	w3, reaches room, slow, extra weight
short	true	false	3	w4, severe damage
short	false	false	100	w5, reaches room, quick
long	false	false	0	w6, severe damage, low energy
long	true	false	80	w7, reaches room, slow

Best

Worst ?

Utility: Simple Goals

- Can simple (boolean) goals still be specified?

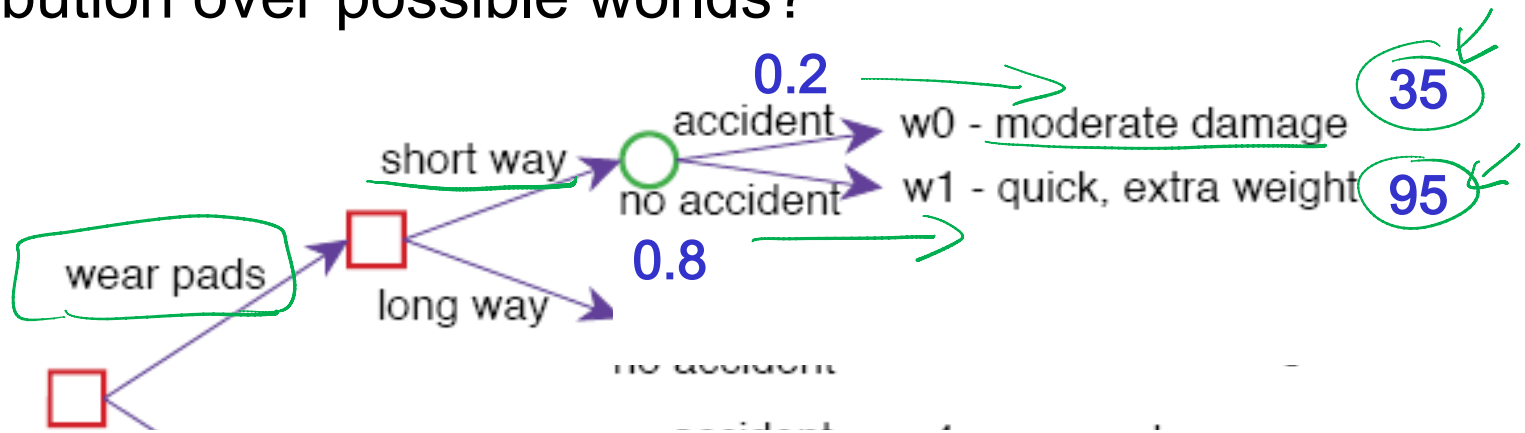
goal: "reaching the room"

Accident must be false

Which way	Accident	Wear Pads	Utility
long	true	true	0
long	true	false	0
long	false	true	100
long	false	false	100
short	true	true	0
short	true	false	0
short	false	true	101
short	false	false	100

Optimal decisions: How to combine Utility with Probability

What is the utility of achieving a certain probability distribution over possible worlds?



- It is its expected utility/value i.e., its average utility, weighting possible worlds by their probability.

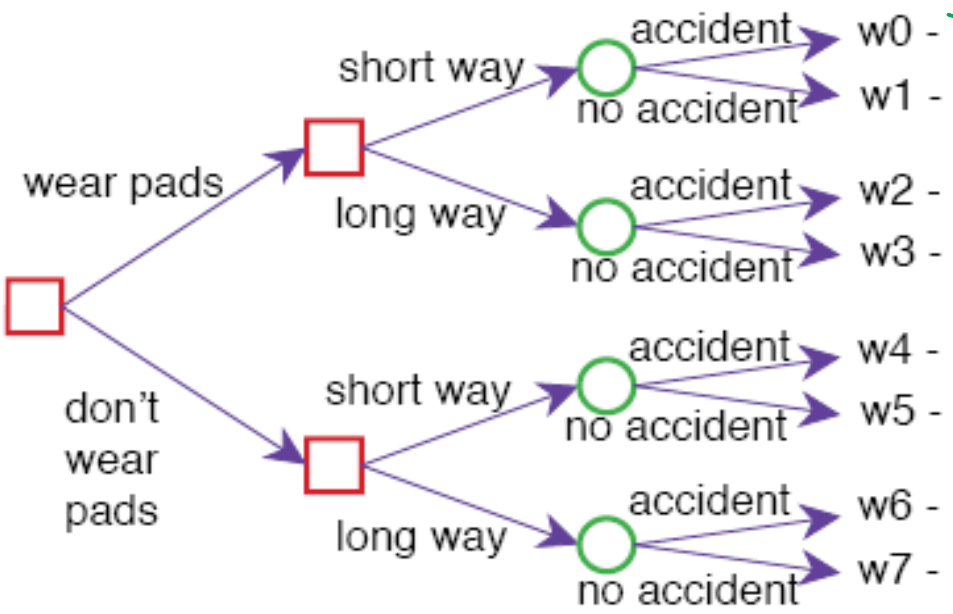
$$EU(wp=t, ww=short) =$$

$$.2 * 35 + .8 * 95$$

Optimal decision in one-off decisions

- Given a set of *n* decision variables var_i (e.g., Wear Pads, Which Way), the agent can choose:
  $D = d_i$ for any $d_i \in \text{dom}(var_1) \times .. \times \text{dom}(var_n)$.

Wear Pads	Which way
true	short
true	long
false	short
false	long

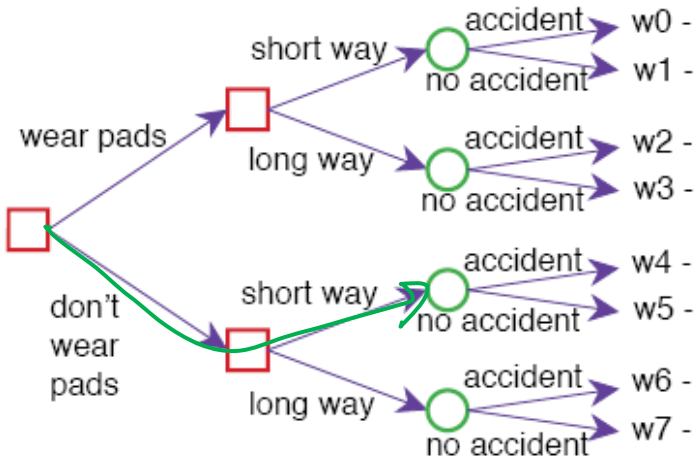


Optimal decision: Maximize Expected Utility

- The expected utility of decision $D = d_i$ is

$$\mathbb{E}(U \mid D = d_i) = \sum_{w \models D = d_i} P(w \mid D = d_i)\, U(w)$$

e.g., $\mathbb{E}(U \mid D = \{WP = \text{false}, WW = \text{short}\}) =$



$P(w_4 \mid D) * U(w_4) +$
$P(w_5 \mid D) * U(w_5)$

- An optimal decision is the decision $D = d_{max}$ whose expected utility is maximal:

$$d_{\max} = \arg\max_{d_i \in dom(D)} \mathbb{E}(U \mid D = d_i)$$

max

Wear Pads	Which way	EU
true	short	—
true	long	—
false	short	—
false	long	—

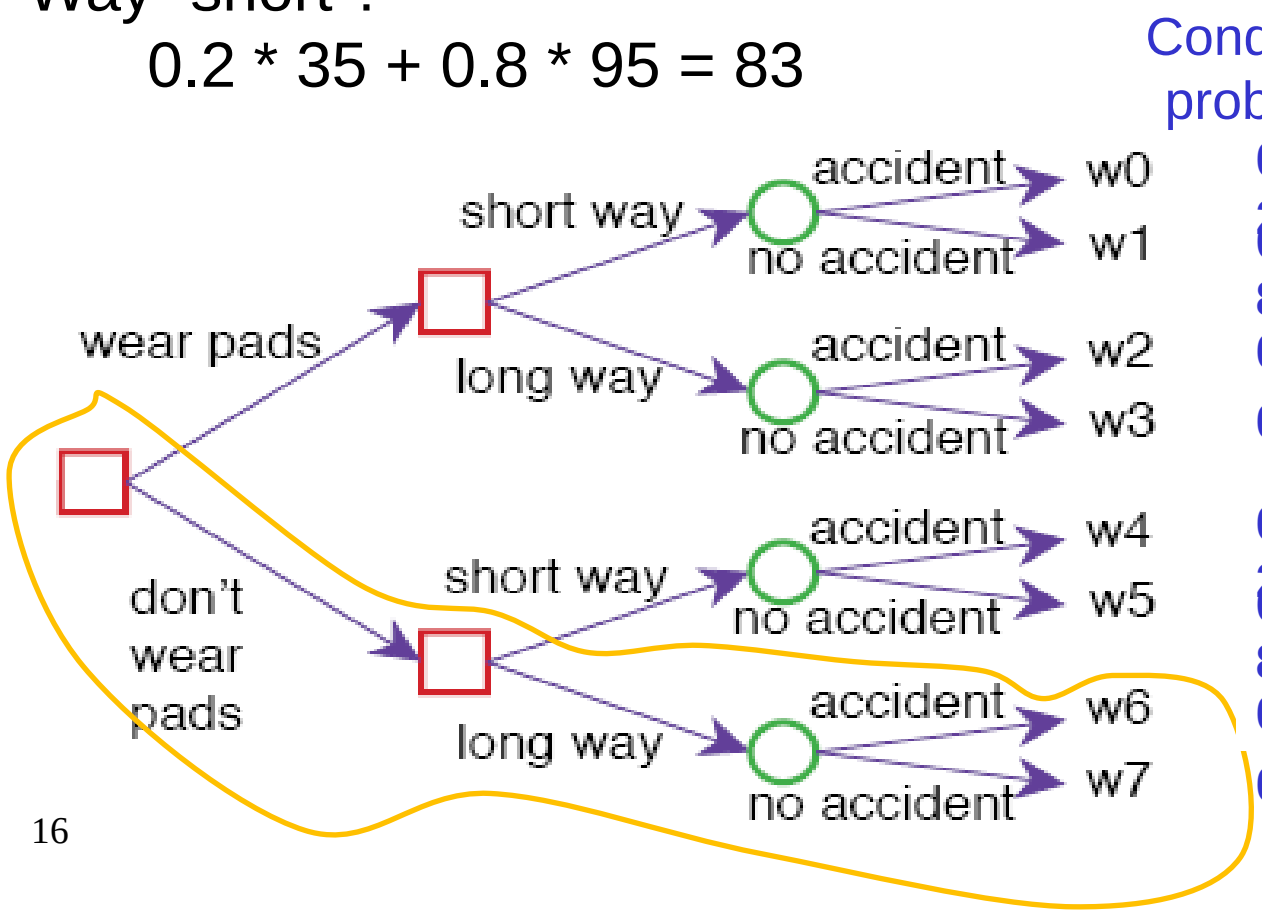
Expected utility of a decision

- The expected utility of decision $D = d_i$ is

$$\mathbb{E}(U \mid D = d_i) = \sum_{w \models (D = d_i)} P(w)\ U(w)$$

- What is the expected utility of Wearpads=yes, Way=short ?

  0.2 * 35 + 0.8 * 95 = 83



	Conditional probability	Utility	E[U|D]
w0	0.2	35	83
w1	0.8	95	
w2	0.01	30	74.55
w3	0.99	75	
w4	0.2	3	80.6
w5	0.8	100	
w6	0.01	0	79.2
w7	0.99	80	

Lecture Overview

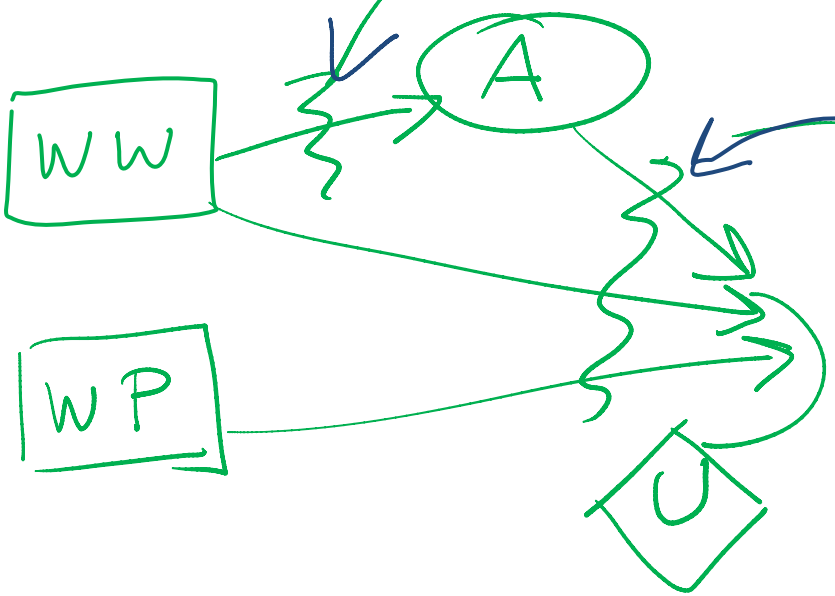
- One-Off Decision
  - Example
  - Optimal Decision: Utilities / Preferences
  - **Single stage Decision Networks**
- Sequential Decisions
  - Representation
  - Policies
  - Finding Optimal Policies

Single-stage decision networks

Extend belief networks with:

- **Decision nodes**, that the agent chooses the value for. Drawn as rectangle.
- **Utility node**, the parents are the variables on which the utility depends. Drawn as a diamond.
- Shows explicitly which decision nodes affect random variables

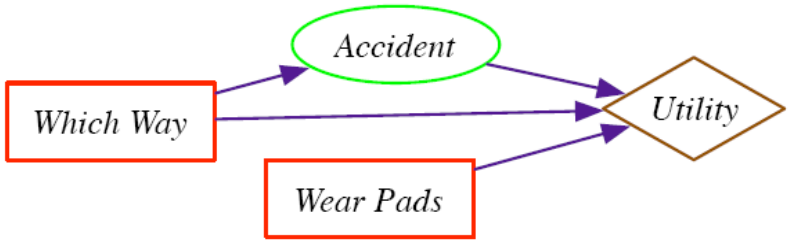
Which way	Accident	
long	true	0.01
long	false	0.99
short	true	0.2
short	false	0.8



Which way	Accident	Wear Pads	Utility
long	true	true	30
long	true	false	0
long	false	true	75
long	false	false	80
short	true	true	35
short	true	false	3
short	false	true	95
short	false	false	100

Finding the optimal decision: We can use VE

Suppose the random variables are $X_1, \ldots, X_n$, the decision variables are the set D, and utility depends on $pU \subseteq \{X_1, \ldots, X_n\} \cup D$



$$\mathbb{E}(U \mid D) = \sum_{X_1,\ldots,X_n} P(X_1,\ldots,X_n \mid D)\, U(pU)$$

$$= \sum \prod P(X_i \mid pX_i)\, U(pU)$$

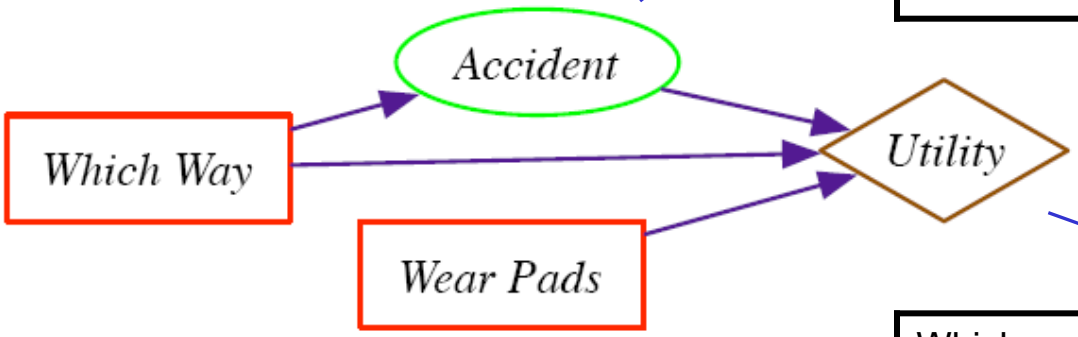
also includes decision vars

To find the optimal decision we can use VE:

1. Create a factor for each conditional probability and for the utility
2. Multiply factors and sum out all of the random variables (This creates a factor on D that gives the expected utility for each d_i)
3. Choose the d_i with the maximum value in the factor.

VE Example: Step 1, create initial factors

Abbreviations:
W = Which Way
P = Wear Pads
A = Accident



$f_1(A,W)$

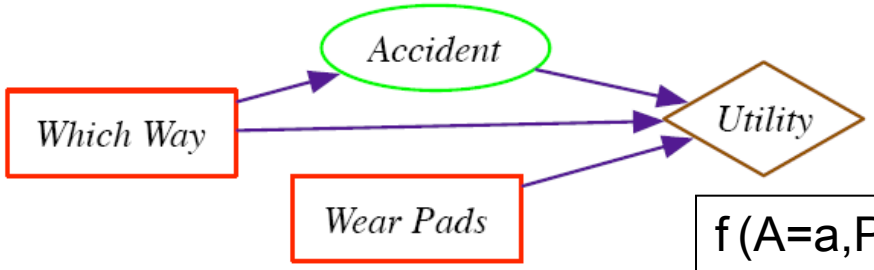
Which Way W	Accident A	P(A|W)
long	true	0.01
long	false	0.99
short	true	0.2
short	false	0.8

$f_2(A,W,P)$

Which way W	Accident A	Pads P	Utility
long	true	true	30
long	true	false	0
long	false	true	75
long	false	false	80
short	true	true	35
short	true	false	3
short	false	true	95
short	false	false	100

$$E(U) = \sum_A P(A|W)\,U(A,W,P)$$
$$= \sum_A f_1(A,W)\; f_2(A,W,P)$$

VE example: step 2, sum out A



Step 2a: compute product
$f(A,W,P) = f_1(A,W) \times f_2(A,W,P)$

$f(A=a,P=p,W=w) = f_1(A=a,W=w) \times f_2(A=a,W=w,P=p)$

Which way W	Accident A	$f_1(A,W)$
long	true	0.01
long	false	0.99
short	true	0.2
short	false	0.8

Which way W	Accident A	Pads P	$f_2(A,W,P)$
long	true	true	30
long	true	false	0
long	false	true	75
long	false	false	80
short	true	true	35
short	true	false	3
short	false	true	95
short	false	false	100

Which way W	Accident A	Pads P	f(A,W,P)
long	true	true	0.01 * 30
long	true	false	
long	false	true	
long	false	false	???
short	true	true	
short	true	false	
short	false	true	
short	false	false	

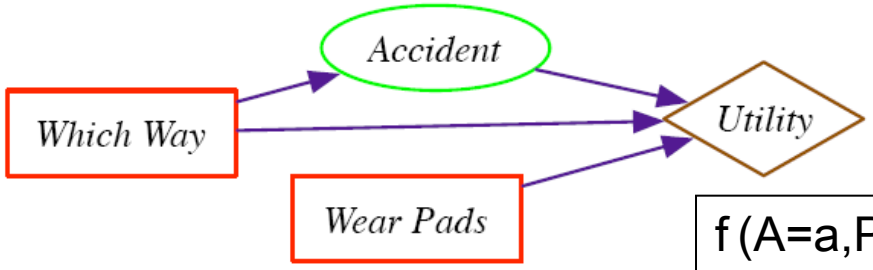
0.99 * 30

0.01 * 80

0.99 * 80

0.8 * 30

VE example: step 2, sum out A



Step 2a: compute product

$f(A,W,P) = f_1(A,W) \times f_2(A,W,P)$

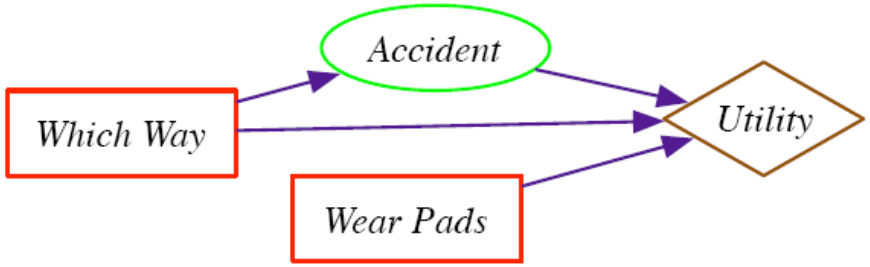
$f(A=a,P=p,W=w) = f_1(A=a,W=w) \times f_2(A=a,W=w,P=p)$

Which way W	Accident A	$f_1(A,W)$
long	true	0.01
long	false	0.99
short	true	0.2
short	false	0.8

Which way W	Accident A	Pads P	$f_2(A,W,P)$
long	true	true	30
long	true	false	0
long	false	true	75
long	false	false	80
short	true	true	35
short	true	false	3
short	false	true	95
short	false	false	100

Which way W	Accident A	Pads P	f(A,W,P)
long	true	true	0.01 * 30
long	true	false	0.01*0
long	false	true	0.99*75
long	false	false	0.99*80
short	true	true	0.2*35
short	true	false	0.2*3
short	false	true	0.8*95
short	false	false	0.8*100

VE example: step 2, sum out A



Step 2b: sum A out of the product f(A,W,P):

$$f_3(W,P) = \sum_A f(A,W,P)$$

Which way W	Pads P	$f_3(W,P)$
long	true	0.01*30+0.99*75=74.55
long	false	
short	true	??
short	false	

Which way W	Accident A	Pads P	f(A,W,P)
long	true	true	0.01 * 30
long	true	false	0.01*0
long	false	true	0.99*75
long	false	false	0.99*80
short	true	true	0.2*35
short	true	false	0.2*3
short	false	true	0.8*95
short	false	false	0.8*100

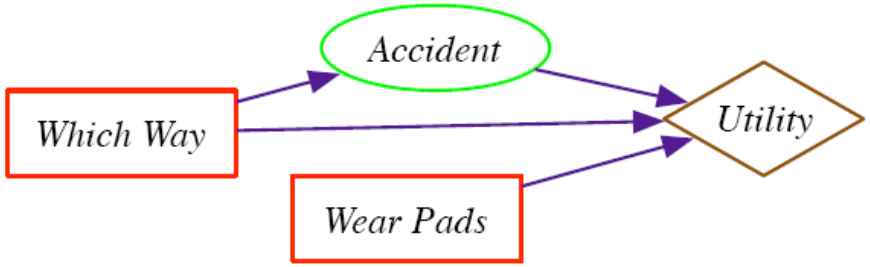
0.2*35 + 0.2*0.3

0.2*35 + 0.8*95

0.99*80 + 0.8*95

0.8 * 95 + 0.8*100

VE example: step 2, sum out A



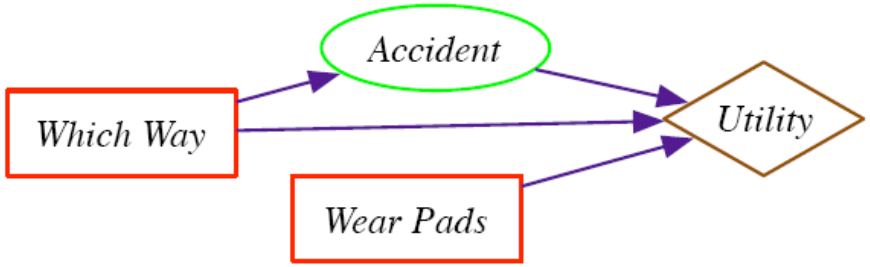
Step 2b: sum A out of the product f(A,W,P):

$$f_3(W,P) = \sum_A f(A,W,P)$$

Which way W	Pads P	f_3(W,P)
long	true	0.01*30+0.99*75=74.55
long	false	
short	true	0.2*35+0.8*95=83
short	false	

Which way W	Accident A	Pads P	f(A,W,P)
long	true	true	0.01 * 30
long	true	false	0.01*0
long	false	true	0.99*75
long	false	false	0.99*80
short	true	true	0.2*35
short	true	false	0.2*3
short	false	true	0.8*95
short	false	false	0.8*100

VE example: step 3, choose decision with max E(U)



Step 2b: sum A out of the product f(A,W,P):

$$f_3(W,P) = \sum_A f(A,W,P)$$

Which way W	Pads P	f_3(W,P)
long	true	0.01*30+0.99*75=74.55
long	false	0.01*0+0.99*80=79.2
short	true	0.2*35+0.8*95=83
short	false	0.2*3+0.8*100=80.6

Which way W	Accident A	Pads P	f(A,W,P)
long	true	true	0.01 * 30
long	true	false	0.01*0
long	false	true	0.99*75
long	false	false	0.99*80
short	true	true	0.2*35
short	true	false	0.2*3
short	false	true	0.8*95
short	false	false	0.8*100

The final factor encodes the expected utility of each decision

- Thus, taking the short way but wearing pads is the best choice, with an expected utility of 83

Learning Goals for today's class – part 1

You can:

- Compare and contrast stochastic **single-stage (one-off)** decisions vs. **multistage** decisions
- Define a **Utility Function** on possible worlds
- Define and compute **optimal one-off decision** (max expected utility)
- Represent one-off decisions as **single stage decision networks** and compute optimal decisions by Variable Elimination

Lecture Overview

- One-Off Decision
  - Example
  - Optimal Decision: Utilities / Preferences
  - Single stage Decision Networks
- **Sequential Decisions**
  - Representation
  - Policies
  - Finding Optimal Policies

“Single” Action vs. Sequence of Actions

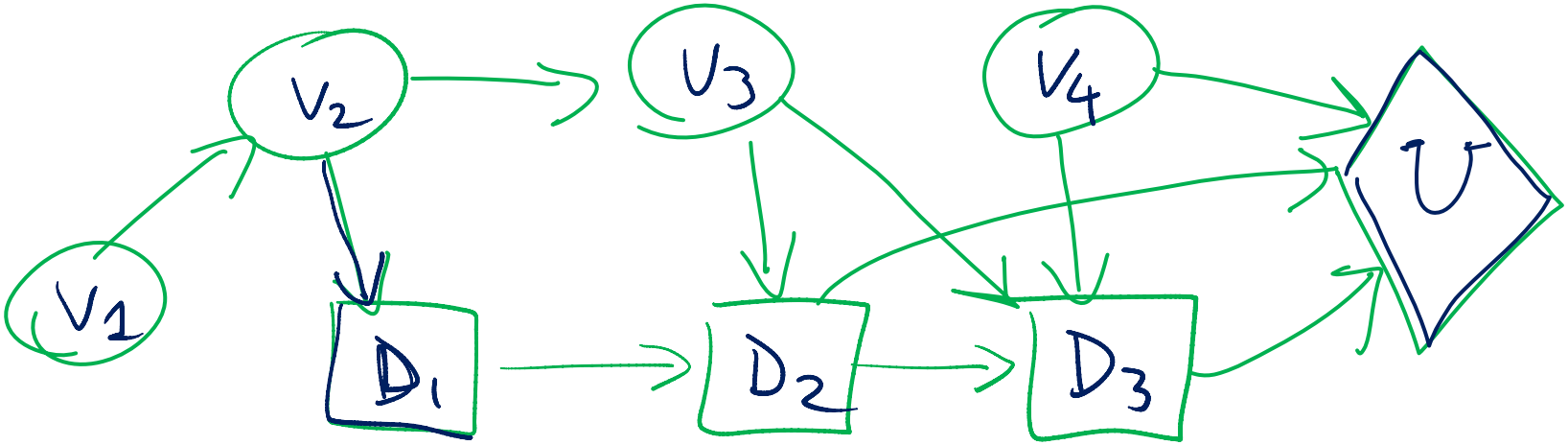
Set of primitive decisions that can be treated as a single macro decision to be made *before acting*

- Agent makes observations
- Decides on an action
- Carries out the action

Sequential decision problems

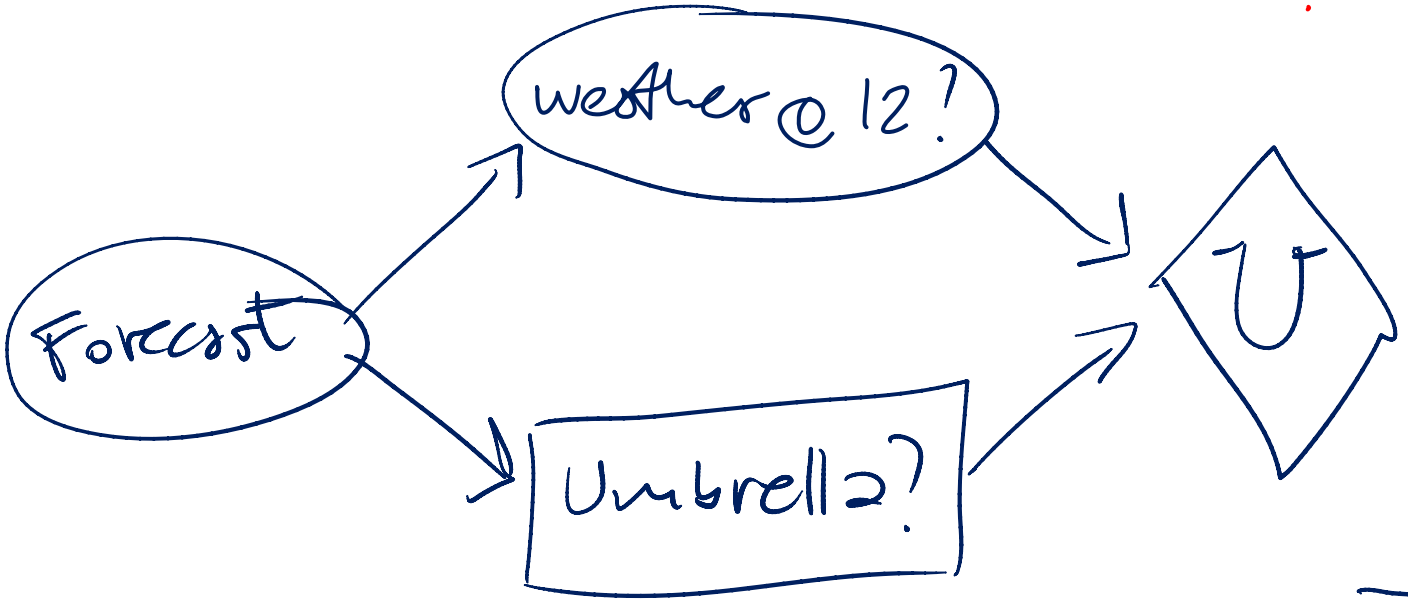
- A sequential decision problem consists of a sequence of decision variables $D_1, \ldots, D_n$.
- Each D_i has an information set of variables pD_i, whose value will be known at the time decision D_i is made.

$pD_3 = \{D_2\ V_3\ V_4\}$



Sequential decisions : Simplest possible

- Only one decision! (but different from one-off decisions)
- Early in the morning. Shall I take my **umbrella** today? (I'll have to go for a long walk at noon)
- Relevant Random Variables?



AIspace

Policies for Sequential Decision Problem: Intro

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

In the *Umbrella* "degenerate" case:

D_1 ? T F

pD_1 Rainy, Cloudy, Sunny

How many policies? 2^3

One possible Policy

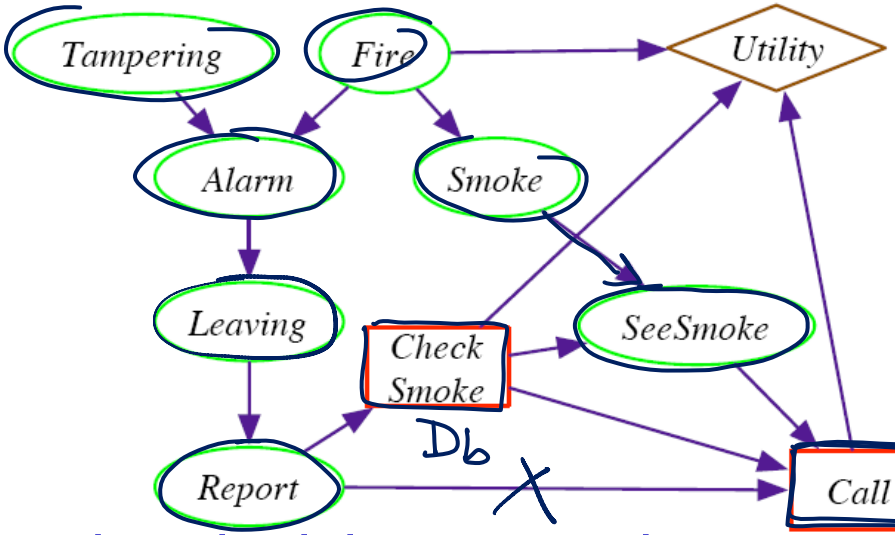
→ R	T	F	T …
→ C	T	F	T …
→ S	F	F	T …

3 policies

$|D|^{|pD|}$

Sequential decision problems: "complete" Example

- A sequential decision problem consists of a sequence of decision variables $D_1, \ldots, D_n$.
- Each D_i has an information set of variables pD_i, whose value will be known at the time decision D_i is made.



$pCS = \{R\}$

$pC = \{R, CS, SS\}$

No-forgetting decision network:

- decisions are totally ordered
- if a decision D_b comes before D_a, then
  - D_b is a parent of D_a
  - any parent of D_b is a parent of D_a

$pCS \subseteq pC$

Lecture Overview

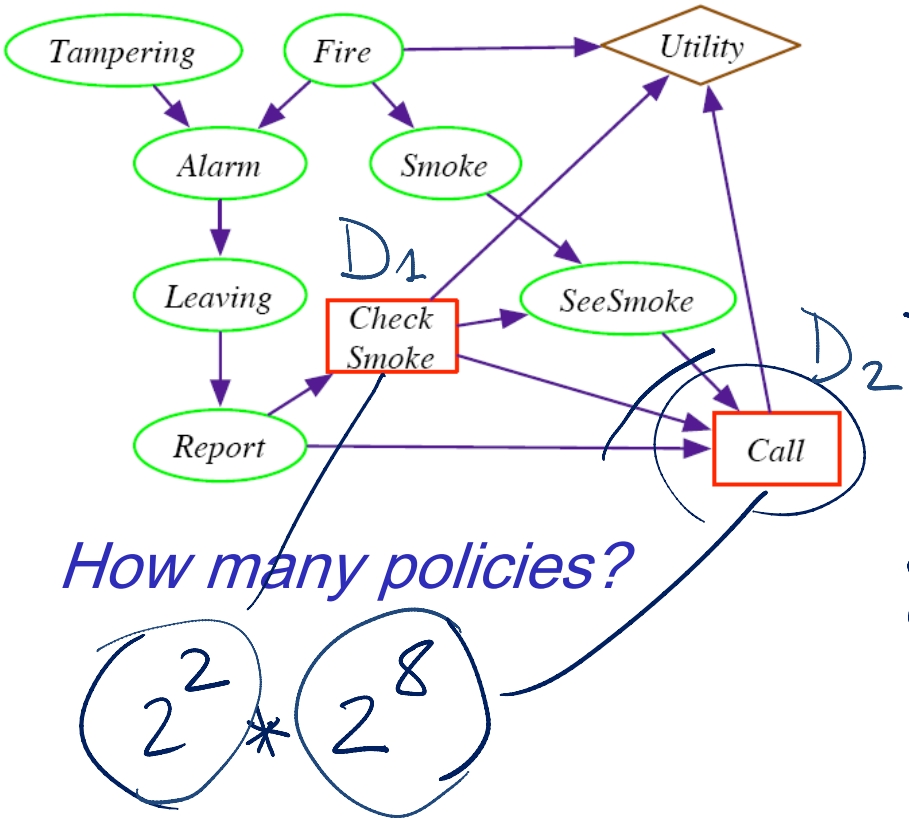
- One-Off Decision
  - Example
  - Optimal Decision: Utilities / Preferences
  - Single stage Decision Networks
- Sequential Decisions
  - Representation
  - **Policies**
  - Finding Optimal Policies

Policies for Sequential Decision Problems

- A **policy** is a sequence of $\delta_1, \ldots, \delta_n$ **decision functions**

$$\delta_i : \text{dom}(pD_i) \rightarrow \text{dom}(D_i)$$

- This policy means that when the agent has observed $O \in \text{dom}(pD_i)$, it will do $\delta_i(O)$



How many policies?

$2^2 * 2^8$

Example:

δ_1

Report	Check Smoke
T	T T F F
F	T F T F

δ_2

Report	CheckSmoke	SeeSmoke	Call
true	true	true	true
true	true	false	false
true	false	true	true
true	false	false	false
false	true	true	true
false	true	false	false
false	false	true	false
false	false	false	false

8

When does a possible world satisfy a policy?

- A possible world specifies a value for each random variable and each decision variable.
- **Possible world** *w* **satisfies policy** δ , written $w \models \delta$ if the value of each decision function in the policy (when applied in *w*).

VARs	
Fire	true
Tampering	false
Alarm	true
Leaving	true
Report	false
Smoke	true
SeeSmoke	true
CheckSmoke	true
Call	true

Decision function for…

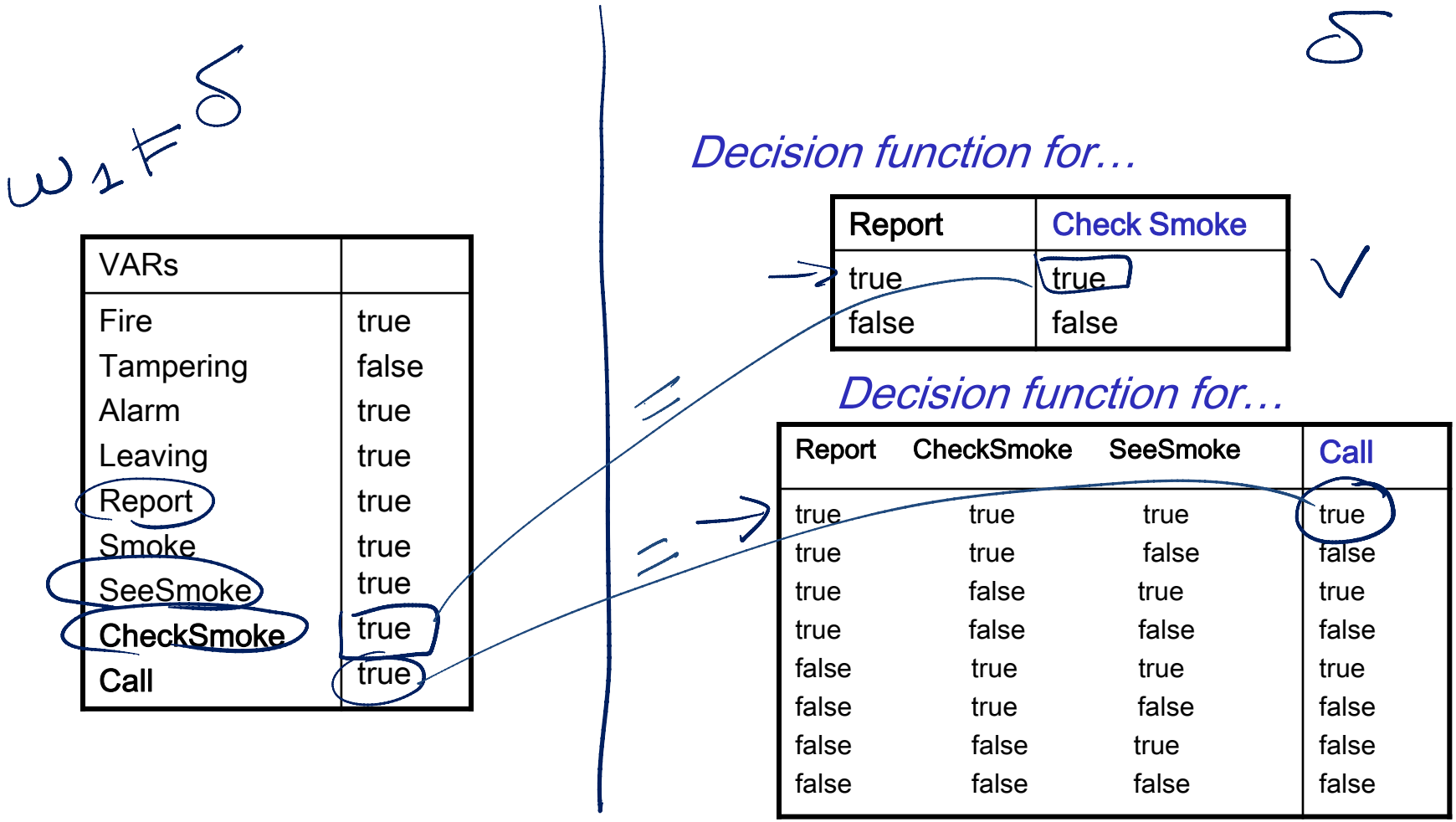
Report	Check Smoke
true	true
false	false

Decision function for…

Report	CheckSmoke	SeeSmoke	Call
true	true	true	true
true	true	false	false
true	false	true	true
true	false	false	false
false	true	true	true
false	true	false	false
false	false	true	false
false	false	false	false

When does a possible world satisfy a policy?

- Possible world *w* satisfies policy *δ* , written *w* ⊨ *δ* if the value of each decision variable is the value selected by its decision function in the policy (when applied in *w*).



VARs	
Fire	true
Tampering	false
Alarm	true
Leaving	true
Report	true
Smoke	true
SeeSmoke	true
CheckSmoke	true
Call	true

Decision function for…

Report	Check Smoke
true	true
false	false

Decision function for…

Report	CheckSmoke	SeeSmoke	Call
true	true	true	true
true	true	false	false
true	false	true	true
true	false	false	false
false	true	true	true
false	true	false	false
false	false	true	false
false	false	false	false

Expected Value of a Policy

- Each possible world *w* has a probability P(*w*) and a utility U(*w*)
- The expected utility of policy δ is

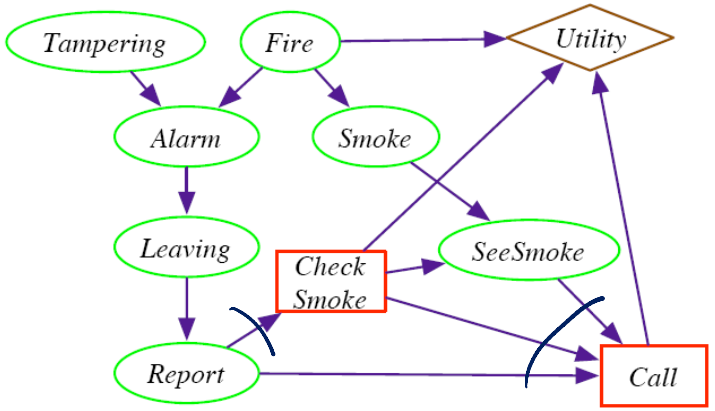
$$\sum_{w \models \delta} P(w) * U(w)$$

- The optimal policy is one with the max expected utility.

Lecture Overview

- One-Off Decision
  - Example
  - Optimal Decision: Utilities / Preferences
  - Single stage Decision Networks
- Sequential Decisions
  - Representation
  - Policies
  - **Finding Optimal Policies**

Complexity of finding the optimal policy: how many policies?



- How many assignments to parents?
- How many decision functions? (binary decisions)
- How many policies?

- If a decision *D* has *k* binary parents, how many assignments of values to the parents are there?

k^2 2^k k Dk

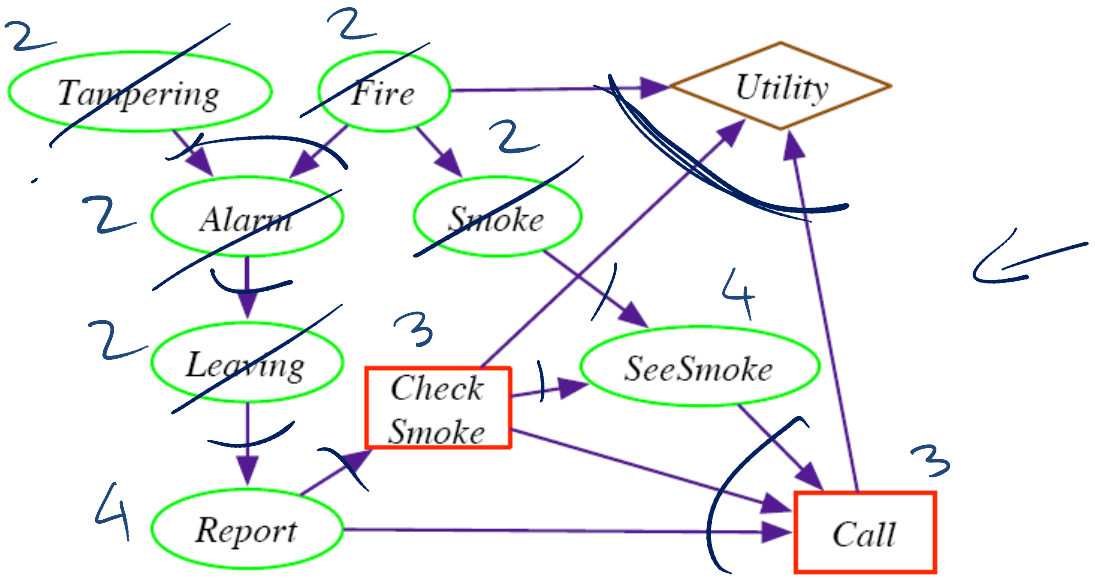
- If there are *b* possible actions (possible values for D), how many different decision functions are there?

$2^{(k+1)}$ $b2^k$ $(b)^{2^k}$

- If there are *d* decisions, each with *k* binary parents and *b* possible actions, how many policies are there?

Finding the optimal policy more efficiently: VE

1. Create a factor for each conditional probability table and a factor for the utility.
2. **Sum out** random variables that are not parents of a decision node.
3. **Eliminate** (aka sum out) the decision variables
4. **Sum out** the remaining random variables.
5. **Multiply the factors**: this is the expected utility of the optimal policy.



AIspace

Eliminate the decision Variables: step3 details

- Select a variable D that corresponds to the latest decision to be made
  - this variable will appear in only one factor with its parents
- Eliminate D by maximizing. This returns:
  - The optimal decision function for D, $\arg\max_D f$
  - A new factor to use in VE, $\max_D f$
- Repeat till there are no more decision nodes.

Example: Eliminate CheckSmoke

Report	CheckSmoke	Value
true	true	-5.0
true	false	-5.6
false	true	-23.7
false	false	-17.5

Report	Value
true	-5.0
false	-17.5

New factor

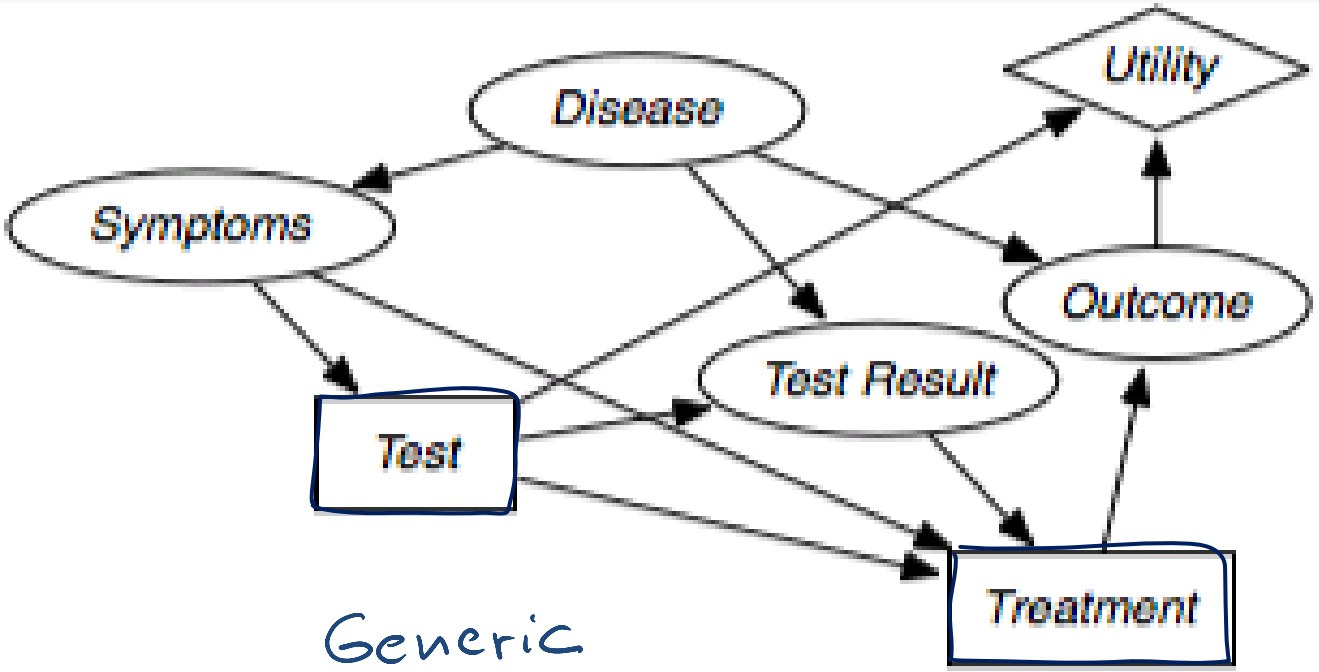
Decision Function

Report	CheckSmoke
true	true
false	false

VE elimination reduces complexity of finding the optimal policy

- We have seen that, if a decision D has k binary parents, there are b possible actions, If there are d decisions,
- Then there are: $(b^{2^k})^d$ *policies*
- Doing variable elimination lets us find the optimal policy after considering only $d \cdot b^{2^k}$ policies (we eliminate one decision at a time)
  - VE **is much more efficient** than searching through policy space.
  - However, this complexity is **still doubly-exponential** we'll only be able to handle relatively small problems.

+ give up nonforgetting assump
+ approx. algorithms

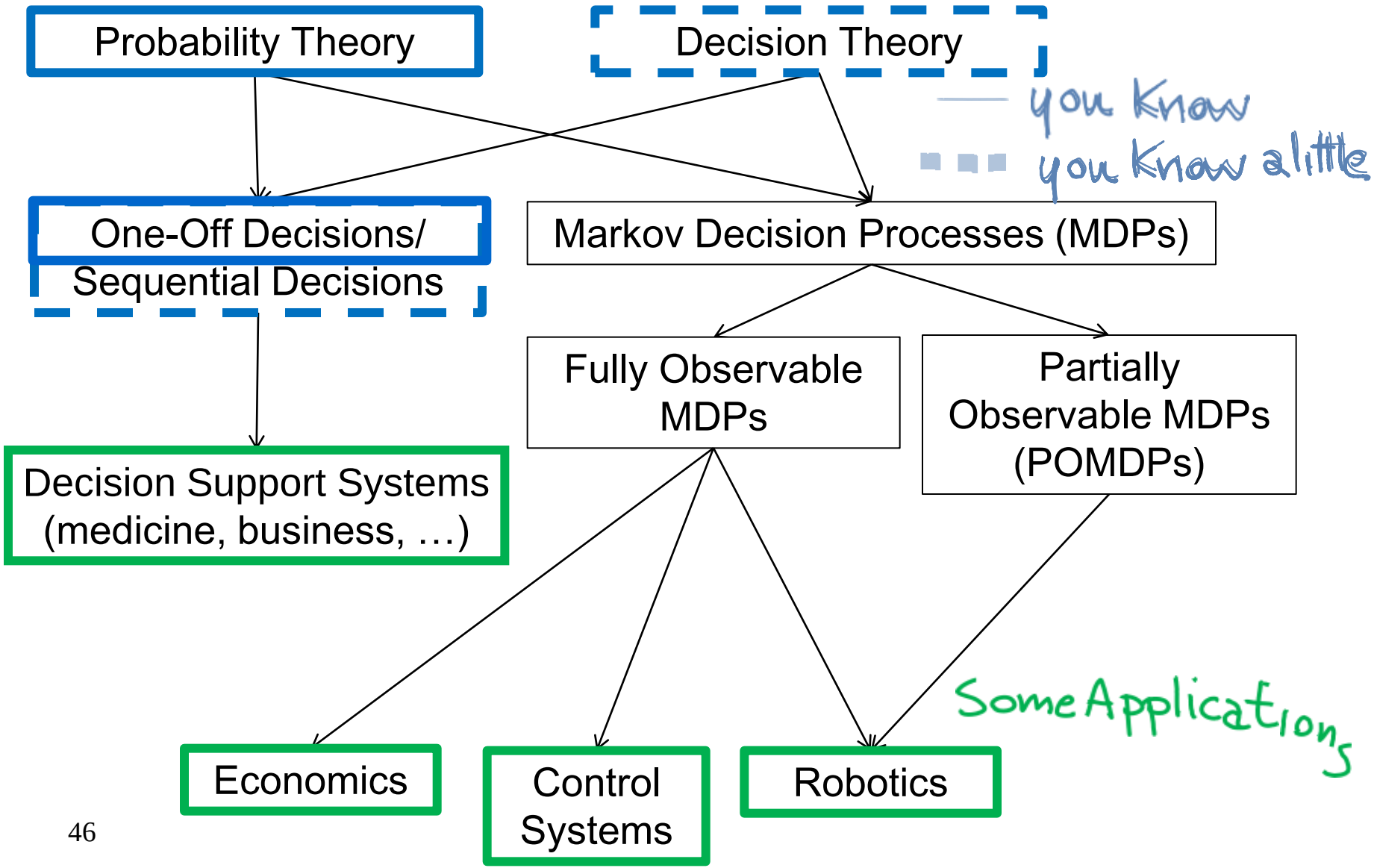


Generic

Figure 9.8: Decision network for diagnosis

to select what test to apply
and then what treatment to prescribe

Big Picture: Planning under Uncertainty

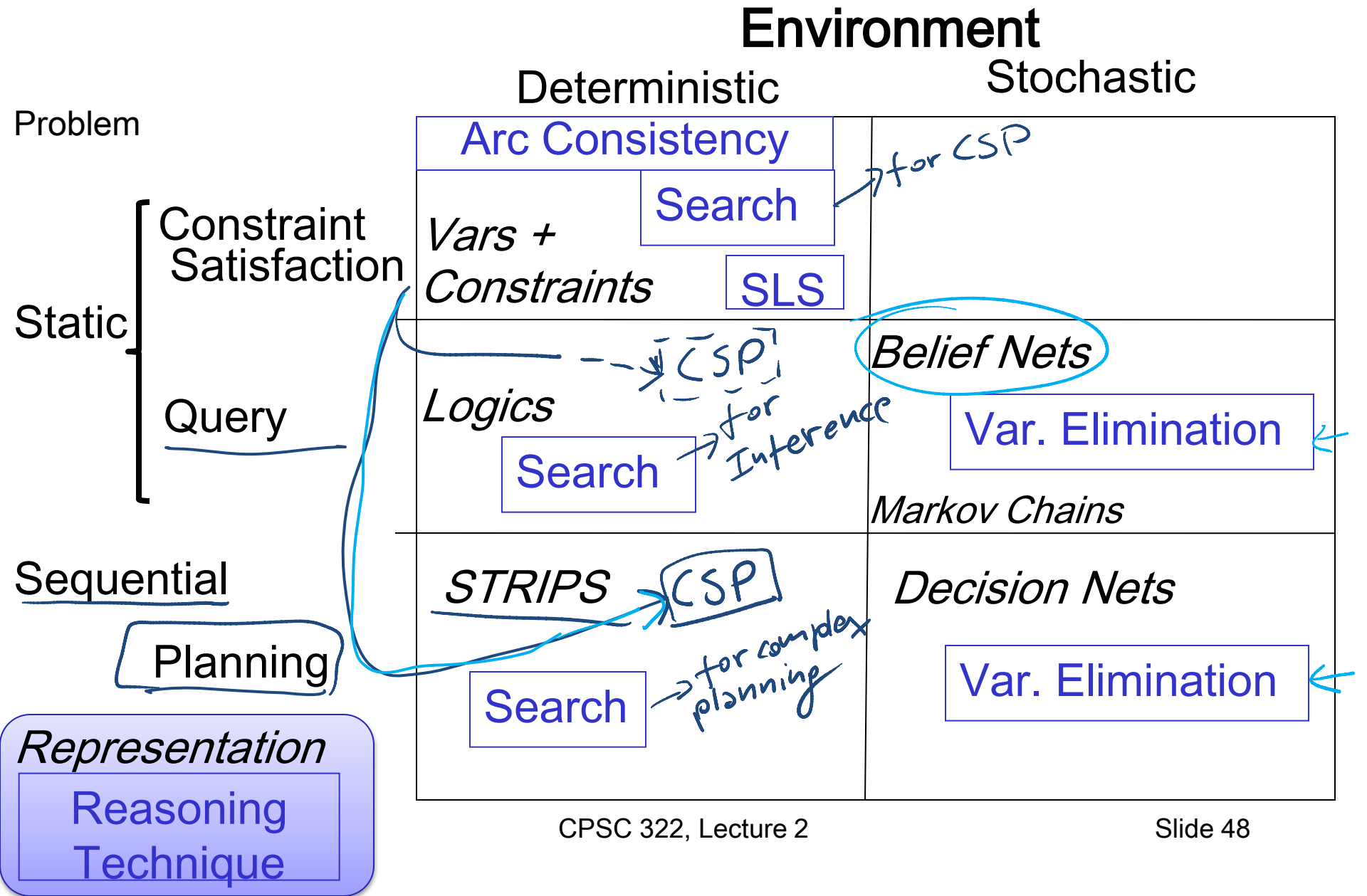


Learning Goals for today's class – part 2

You can:

- Represent **sequential decision problems** as decision networks. And explain the **non forgetting property**
- Verify whether a **possible world satisfies a policy** and define the **expected value of a policy**
- Compute the **number of policies** for a decision problem
- **Compute the optimal policy** by Variable Elimination

Cpsc 322 Big Picture



After 322 .....

340

→ Machine Learning

→ Knowledge Acquisition

→ Preference Elicitation

322 big picture

Mostly

14 422

	Deterministic	Stochastic
CSPs	*Vars + Constraints* Techniques to study SLS Performance	
Query	*Logics* → *First Order Logics* → *Temporal reasoning* → *Description Logics*	*Belief Nets* More sophisticated reasoning *Markov Chains and HMMs*
Planning	*Hierarchical Task Networks* *Partial Order Planning*	*Markov Decision Processes and Partially Observable MDP* More sophisticated reasoning

Where are the components of our representations coming from?

The probabilities?

The utilities?

The logical formulas?

From people and from data!

11

Applications of AI

Announcements

- **Fill out Online Teaching Evaluations Survey.**

- **FINAL EXAM:** Thur June14, 9:00-11:30 pm DMP 310 (NOT regular room)

6-8 pts ~20 pts

Final will comprise: 10 -15 short questions + 3-4 problems

- Work on all **practice exercises** and **sample problems**
- While you revise the **learning goals,** work on **review questions** - I may even reuse some verbatim ☺
- **Come to remaining Office hours!** Today and Tomorrow (2-4 X150 (Learning Center))

STRIKE? http://vpstudents.ubc.ca/news/strike-action/