Review

- Tentamen: Maandag 16 Januari 2012

- Early rule-based (logical) approach to reasoning with uncertainty seemed attractive
- Largely replaced by probability theory
- Bayesian networks and other probabilistic graphical models are the state of the art for reasoning with uncertainty
- Variable elimination is an efficient way of performing probabilistic inference by exploiting the independence structure of a Bayesian network

Decision making

Alice... went on "Would you please tell me, please, which way I ought to go from here?"

"That depends a good deal on where you want to get to," said the Cat.

"I don't much care where" said Alice.

"Then it doesn't matter which way you go," said the Cat.

Lewis Carroll, 1832-1898
Alice's Adventures in Wonderland, 1865

Decision making

- We know how to reason about the state of the world
- Is that enough to implement an intelligent agent?
- No:
  - reasoning without action is void
  - reasoning may require action to gain information
  - action selection requires preferences

Preferences

- Actions result in outcomes
- Agents have preferences over outcomes
- A rational agent will take the action that has the best outcome for them
- Sometimes agents don't know the outcomes of the actions, but they still need to compare actions
- Agents have to act (doing nothing is often an action).

Preferences

If o_1 and o_2 are outcomes

- $o_1 \succeq o_2$ means o_1 is at least as desirable as o_2
- $o_1 \sim o2$ means $o_1 \succeq o_2$ and $o_2 \succeq o_1$
- $o_1 \succ o_2$ means $o_1 \succeq o_2$ but not $o_2 \succeq o_1$

Lotteries

- An agent may not know the outcomes of their actions but only have a probability distribution of the outcomes.
- A lottery is a probability distribution over outcomes. It is written

  $$p_1 : o_1; p_2 : o_2; \ldots ; p_k : o_k$$

  where the o_i are outcomes and $p_i > 0$ such that

  $$\sum_i p_i = 1$$

  The lottery specifies that outcome o_i occurs with probability p_i.
- E.g. $0.1 : cured; 0.9 : uncured$ when receiving treatment

von Neumann-Morgenstern axioms

- Completeness:

$$\forall o_1 \forall o_2 \;\; o_1 \succeq o_2 \text{ or } o_2 \succeq o_1$$

Agents must act; if the actions available to it have outcomes o_1 and o_2 then, by acting, it is explicitly or implicitly preferring one outcome over the other.

- Transitivity:

$$\text{if } o_1 \succeq o_2 \text{ and } o_2 \succeq o_3 \text{ then } o_1 \succeq o_3$$

Suppose it is false, in which case $o_1 \succeq o_2$ and $o_2 \succeq o_3$ and $o_3 \succ o_1$. This would lead to a money pump. Being prepared to pay money to cycle through a set of outcomes is irrational; hence, a rational agent should have transitive preferences.

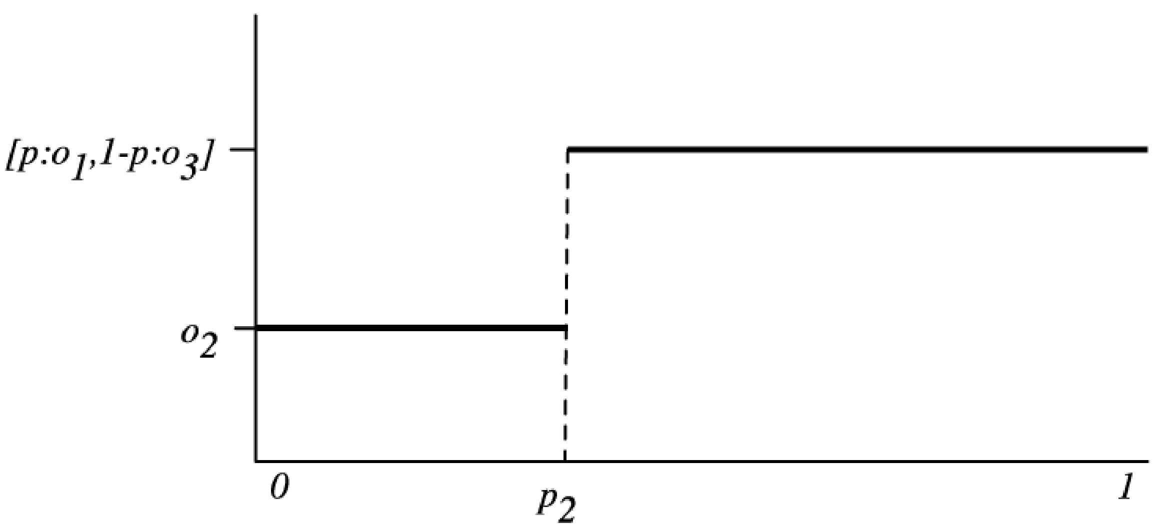
von Neumann-Morgenstern axioms

- Independence:

  if $o_1 \succeq o_2$ then $[p\colon o_1, 1-p\colon o_3] \succeq [p\colon o_2, 1-p\colon o_3]$

  Two gambles mixed with a third one maintain the same preference order as when the two are presented independently of the third one.

- Continuity: suppose $o_1 \succ o_2 \succ o_3$ then there exists a $p \in [0,1]$ such that $o_2 \sim [p\colon o_1, 1-p\colon o_3]$



Rational agents

- If an agent respects the von Neumann-Morgenstern axioms then it is said to be rational
- If an agent is rational, then the preference of an outcome can be quantified using a utility function:

$$U: \text{outcomes} \longrightarrow [0, 1]$$

such that:

  - $o_1 \succeq o_2$ if and only if $U(o_1) \geq U(o_2)$.
  - $U([p_1 : o_1, p_2 : o_2, \ldots, p_k : o_k]) = \sum_{i=1}^{k} p_i \cdot U(o_i)$

Bounded rationality

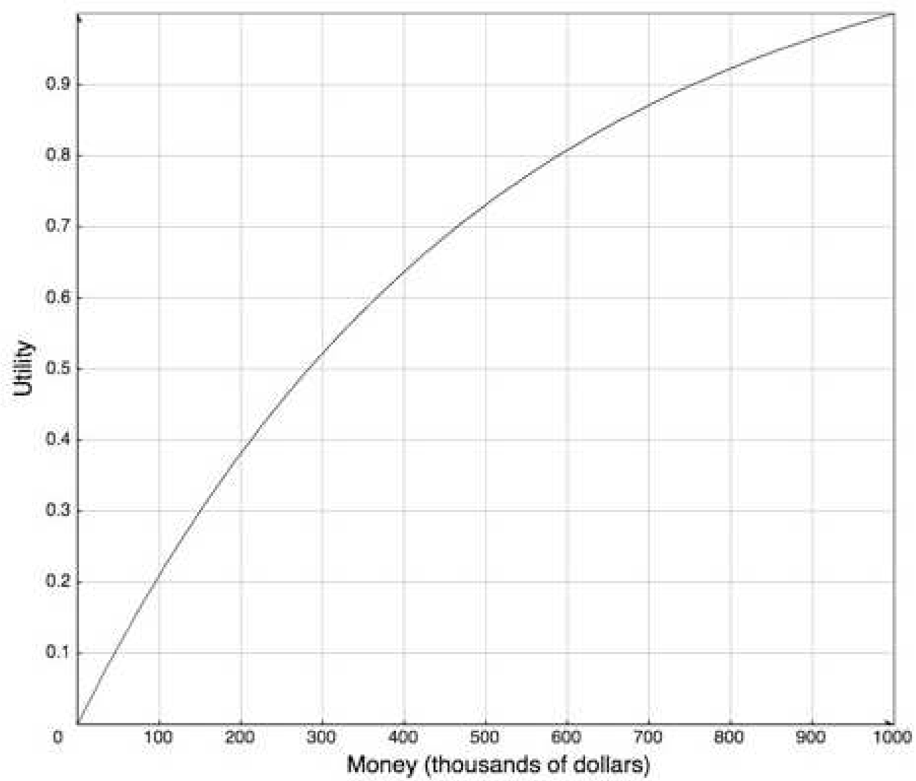
- Proposed by Herbert Simon as an alternative basis for the mathematical modeling of decision making, as used in economics and related disciplines
- Bounded rationality: Rationality of individuals is limited by the information they have, the cognitive limitations of their minds, and the finite amount of time they have to make decisions.
- Bounded rationality searches for satisficing (sufficing and satisfying) solutions

Utilities

$$U: \text{outcomes} \rightarrow [0, 1]$$

- Utility is a measure of desirability of outcomes to an agent.
- Let $u(o)$ be the utility of outcome o to the agent.
- Simple goals can be specified by: outcomes that satisfy the goal have utility 1; other outcomes have utility 0.
- Often utilities are more complicated: for example some function of the amount of damage to a robot, how much energy is left, what goals are achieved, and how much time it has taken.

Utility as a function of money



- People are often risk-aversive
- What do you prefer? 50/50 chance for 0 or 1000000 or 300000 now?
- Risk-aversion is why insurance companies can exist

Framing effects

- A disease is expected to kill 600 people. Two alternative programs have been proposed:
  - Program A: 200 people will be saved
  - Program B: probability 1/3: 600 people will be saved, probability 2/3: no one will be saved
- Which program would you favor?

Framing effects

- A disease is expected to kill 600 people. Two alternative programs have been proposed:
  - Program C: 400 people will die
  - Program D: probability 1/3: no one will die, probability 2/3: 600 will die
- Which program would you favor?

Framing effects

- A disease is expected to kill 600 people. Two alternative programs have been proposed:
  - Program C: 400 people will die
  - Program D: probability 1/3: no one will die, probability 2/3: 600 will die
- Which program would you favor?
- Tversky and Kahneman: 72% chose A over B; 22% chose C over D.

Decision-making under uncertainty

What an agent should do depends on:

- The agent's beliefs: the ways the world could be, given the agent's knowledge.
- The agent's preferences: what the agent wants and tradeoffs when there are risks.
- The agent's ability: what actions are available to it.

Decision theory specifies how to trade off the desirability and probabilities of the possible outcomes for competing actions.

Single decisions

- Decision variables are like random variables that an agent gets to choose a value for.
- For a single decision variable, the agent can choose $D = d$ for any $d \in \mathsf{dom}(D)$.
- The expected utility of decision $D = d$ is

$$E(U \mid d) = \sum_{x_1,\ldots,x_n} P(x_1,\ldots,x_n \mid d) U(x_1,\ldots,x_n,d)$$

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

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

Example

Suppose:

- $P =$ throw party
- $R =$ rain
- $U(p, \neg r) = 500$, $U(p, r) = -100$, $U(\neg p, r) = 0$, $U(\neg p, \neg r) = 50$
- $P(r \mid P) = P(r) = 0.6$

Then:

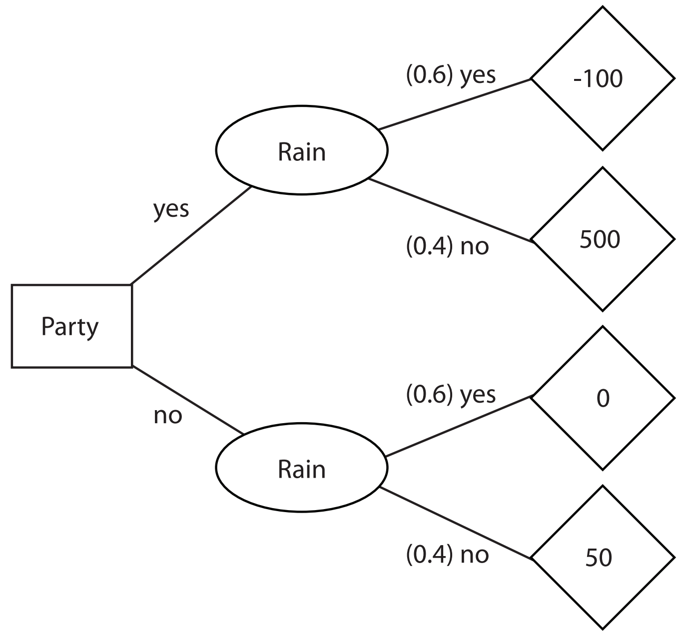
$$\begin{aligned} E(U \mid p) &= 0.6 \cdot -100 + 0.4 \cdot 500 = 140 \\ E(U \mid \neg p) &= 0.6 \cdot 0 + 0.4 \cdot 50 = 20 \end{aligned}$$

Conclusion: Party!

Sequential decisions

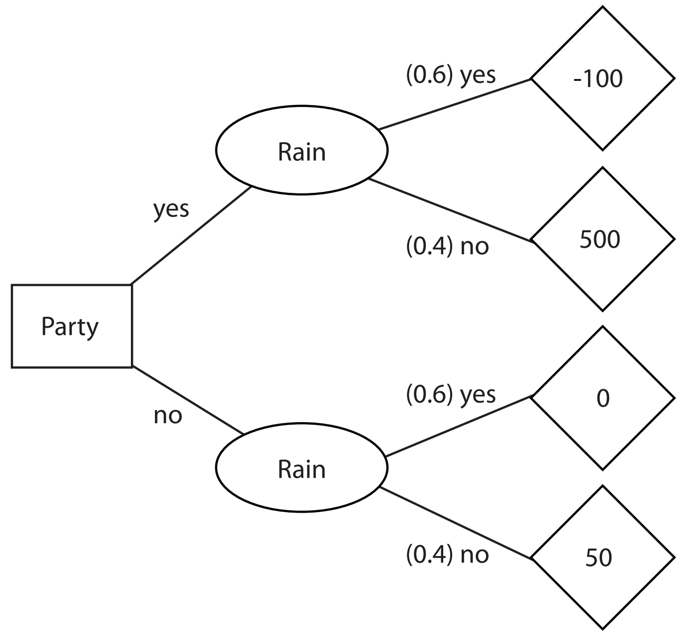
- Multiple decisions made in parallel can be regarded as one big single decision.
- An intelligent agent doesn't carry out just one action or ignore intermediate information
- A more typical scenario is where the agent: observes, acts, observes, acts, . . .
- Subsequent actions can depend on what is observed.
- What is observed depends on previous actions.
- Often the sole reason for carrying out an action is to provide information for future actions. For example: diagnostic tests, range finding, spying.
- Optimal reasoning requires sequential decision making

Decision trees



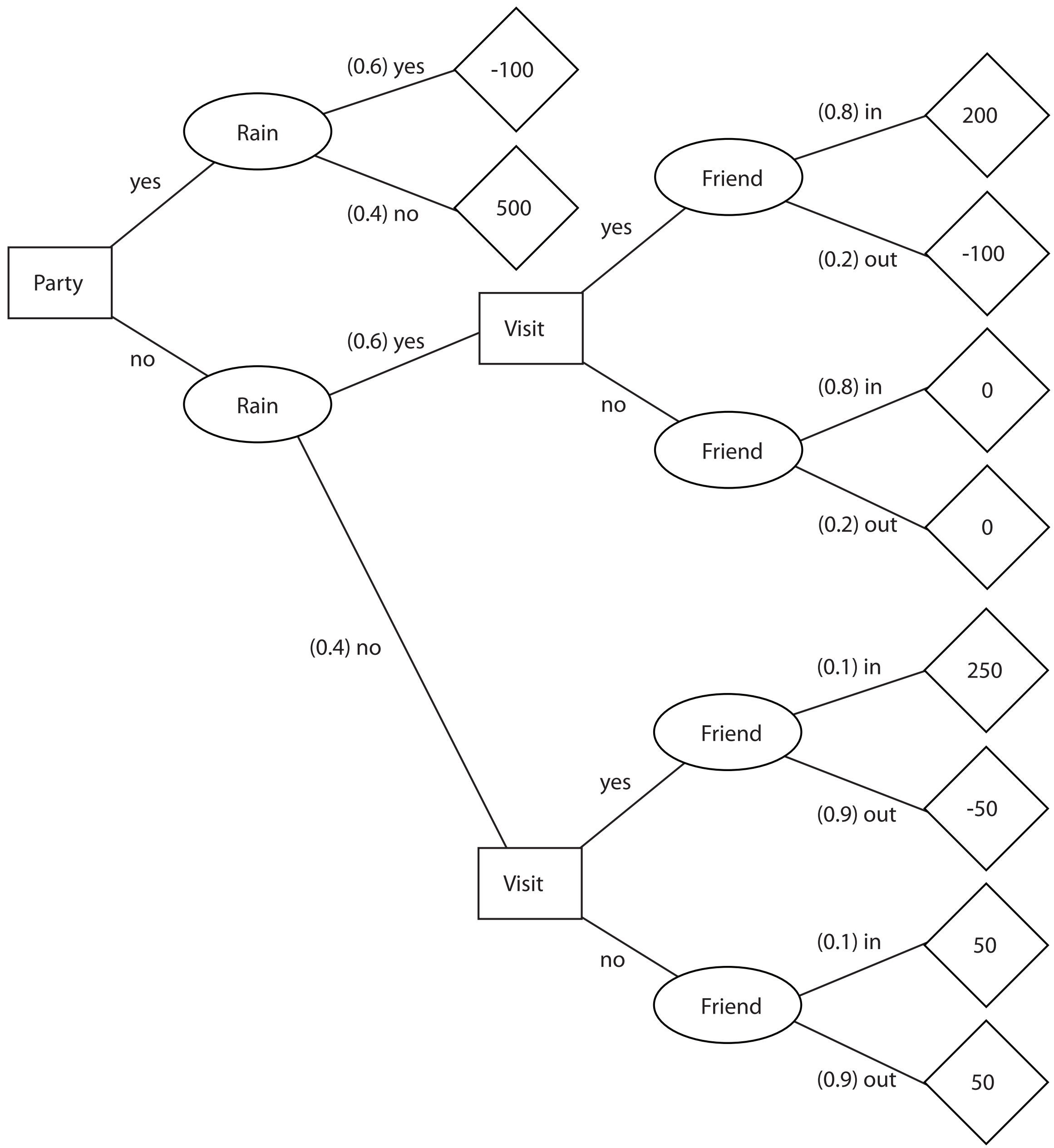
- Decision trees are a way to graphically organise a finite sequential decision process.
- Contains decision nodes, each with branches for each of the alternative decisions.
- Contains chance nodes (random variables)
- Utility at the leaf of each branch

Decision trees



- Expected utility of any decision computed on the basis of the weighted summation of all branches from the decision to all leaves from that branch.

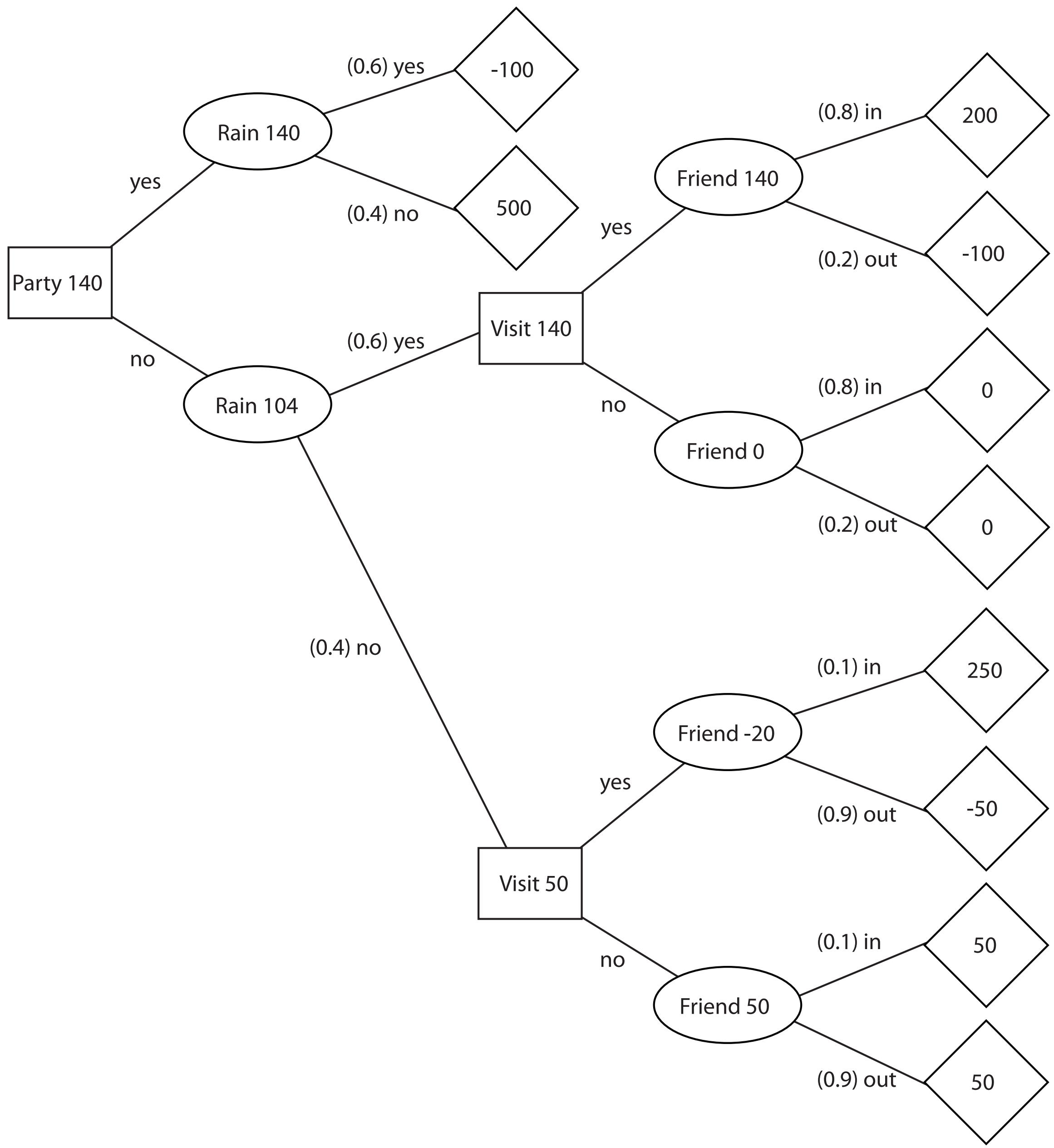
Sequential decision problems



Sequential decision problems

- A sequential decision problem consists of a sequence of decision variables $D_1, \ldots, D_n$.
- Each D_i has a set of information variables $\pi(D_i)$ whose value will be known at the time decision D_i is made.
- Solving a sequence of decisions can be achieved using a decision tree by rolling back the tree:
  - The utility of a chance node is its expected utility
  - For a decision node, the value of the node is the maximum of its child values.
  - The optimal decision sequence is given at each decision node by finding which child node has the maximal value

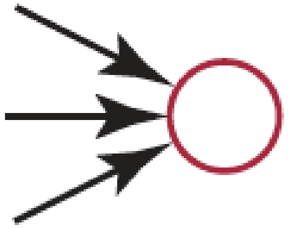
Sequential decision problems



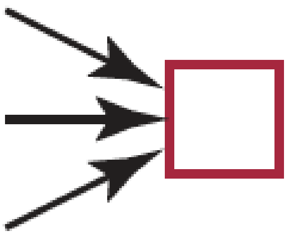
Influence diagrams

- Decision trees are general and explicitly encode the utilities and probabilities associated with each decision and event
- Decision trees grow exponentially fast in the number of variables
- Influence diagrams (IDs) enable a more compact description of a finite sequential decision problem
- IDs extend belief networks to include decision variables and utility.
- Specifies what information is available when the agent has to act.
- Specifies which variables the utility depends on.

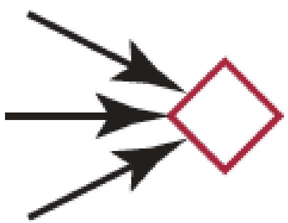
Influence diagrams



- A random variable is drawn as an ellipse. Arcs into the node represent probabilistic dependence.

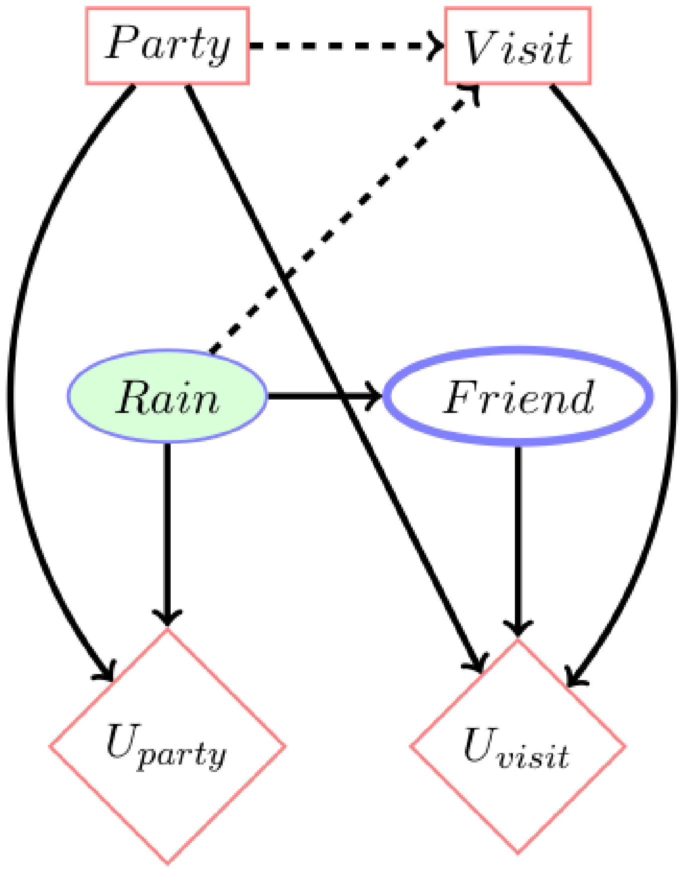


- A decision variable is drawn as an rectangle. Arcs into the node represent information available when the decision is make.



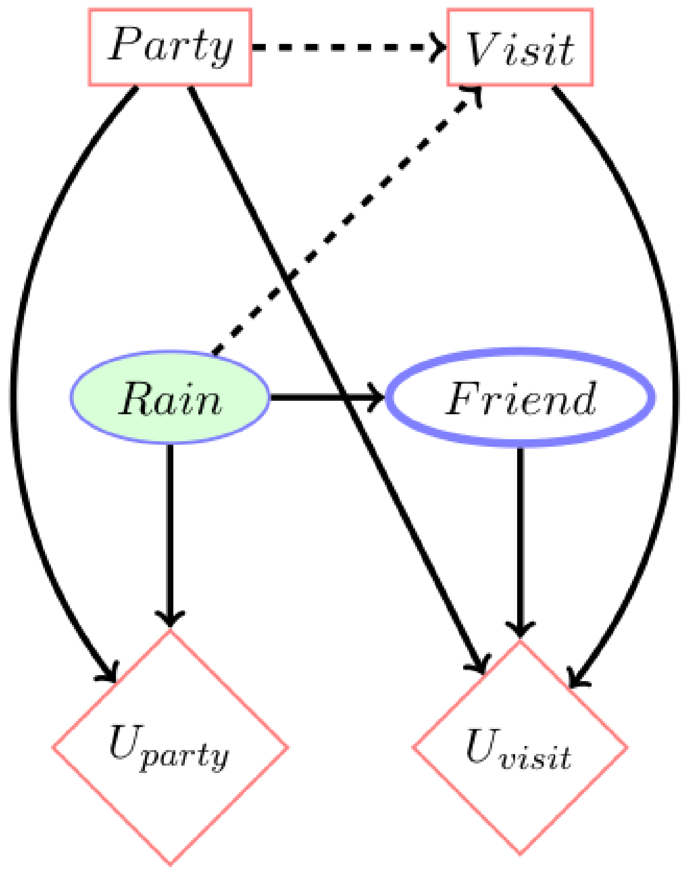
- A utility node is drawn as a diamond. Arcs into the node represent variables that the utility depends on.

Party ID



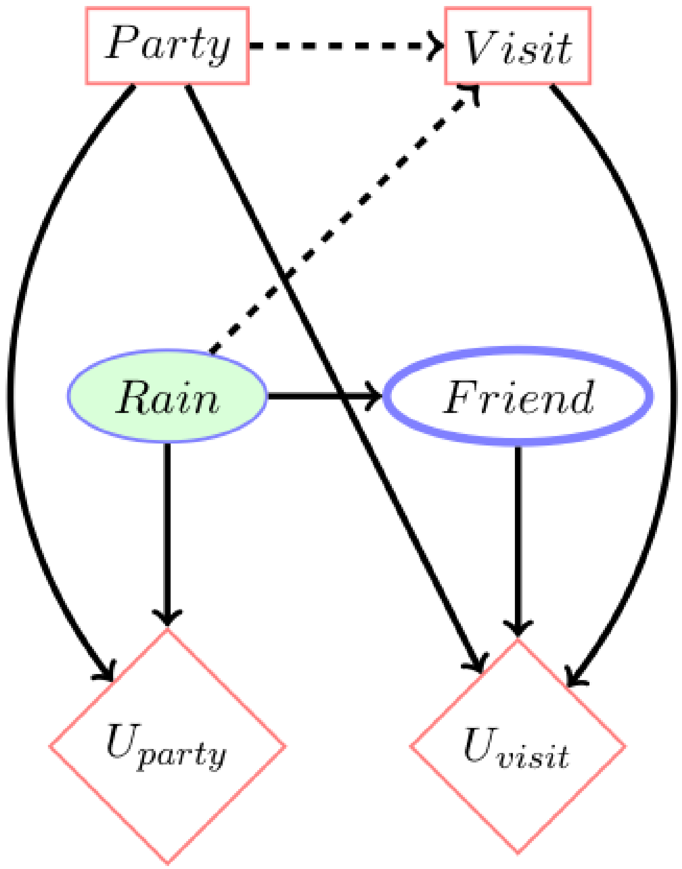
- The decision nodes are totally ordered. This is the order the actions will be taken.

Party ID



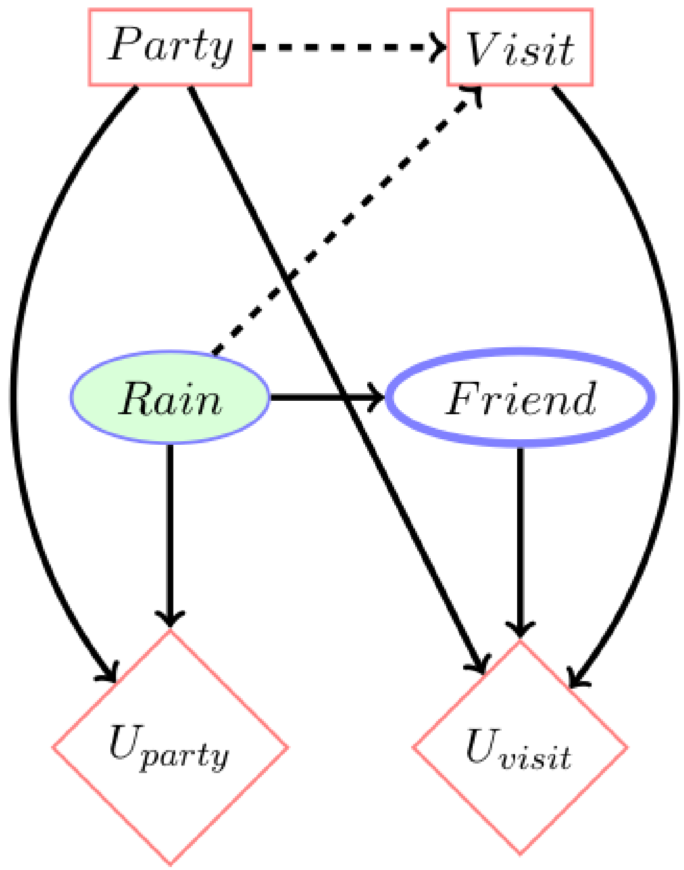
- Informational predecessors indicated by dashed arcs

Party ID



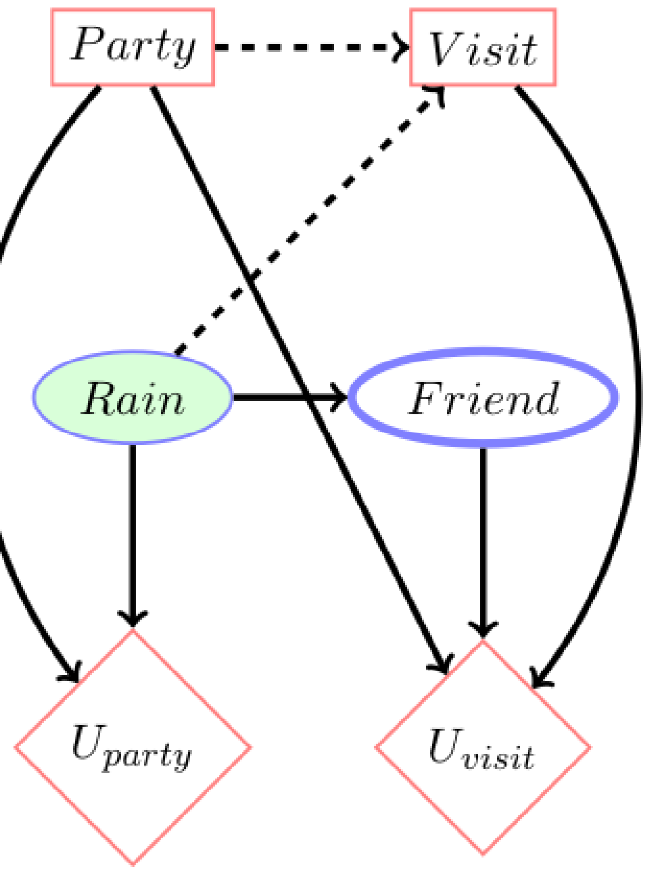
- Colored chance nodes indicate information which will become available

Party ID



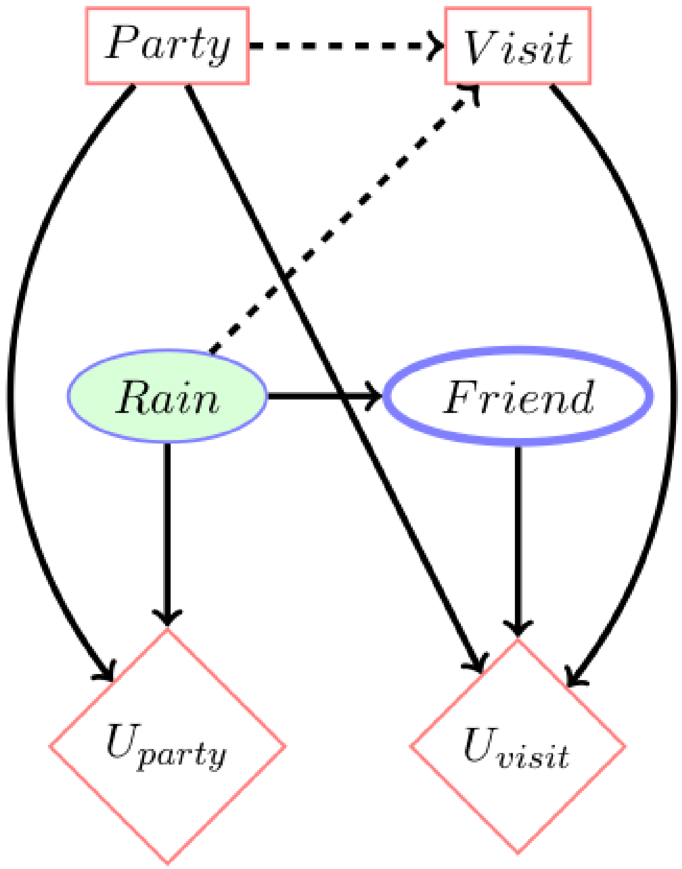
- Any parent of a decision node is a parent of subsequent decision nodes (no forgetting principle).

Party ID



- An ID defines a partial order:
  $\mathcal{X}_1 \prec D_2, \ldots, \mathcal{X}_{n-1} \prec D_n \prec \mathcal{X}_n$

Party ID



- Total utility is given by the sum of the independent utilities (additive utility).

What should an agent do?

- What an agent should do at any time depends on what it will do in the future.
- What an agent does in the future depends on what it did before.
- A policy specifies what an agent should do under each circumstance.
- A policy is a sequence $\Delta = \{\delta_1, \ldots, \delta_N\}$ of decision functions

$$\delta_i : \text{dom}(\pi(D_i)) \longrightarrow \text{dom}(D_i)$$

- This policy means that when the agent has observed $o \in \text{dom}(\pi(D_i))$, it will do $\delta_i(o)$.
- An optimal policy is the one with the highest expected utility.

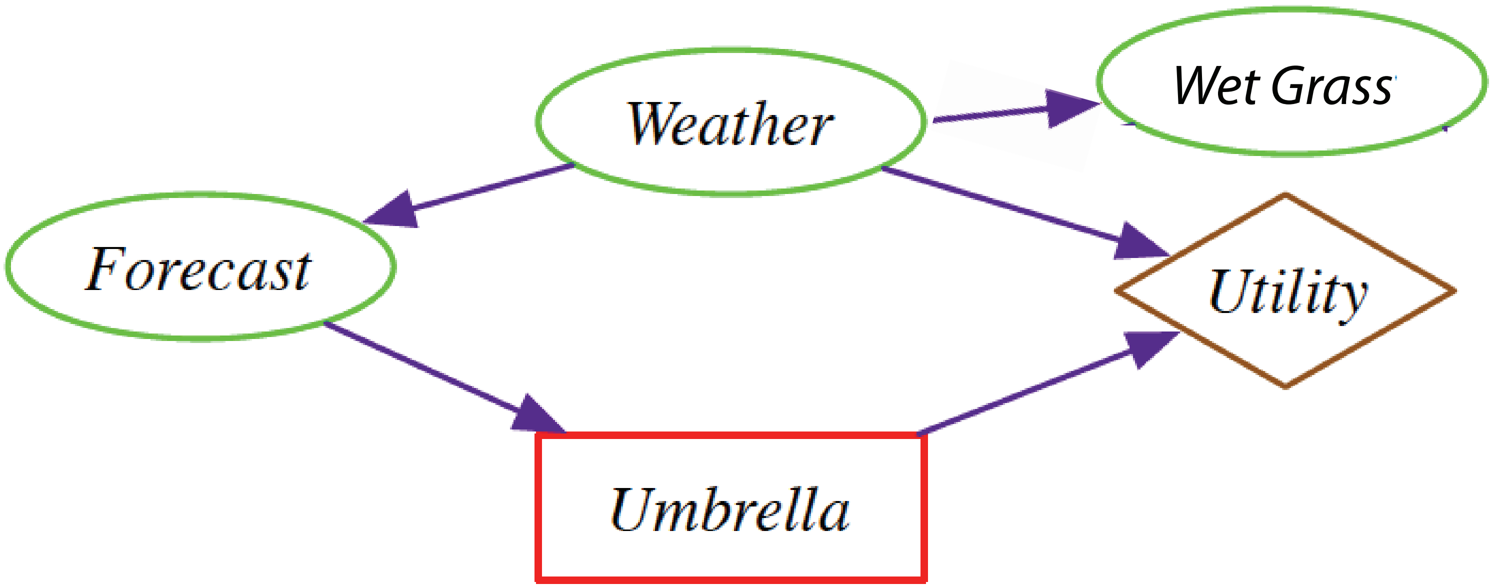
Finding the optimal policy

- Partial order: $\mathcal{X}_1 \prec D_2, \ldots, \mathcal{X}_{n-1} \prec D_n \prec \mathcal{X}_n$
- Recall:
  $E(U \mid d) = \sum_{x_1,\ldots,x_n} P(x_1, \ldots, x_n \mid d) U(x_1, \ldots, x_n, d)$
- The maximal expected utility U^* is given by

$$U^* = \sum_{\mathcal{X}_1} \max_{D_2} \cdots \sum_{\mathcal{X}_{n-1}} \max_{D_n} \sum_{\mathcal{X}_n} \prod_{i \in \mathcal{I}} P(x_i \mid \pi(x_i)) \sum_{j \in \mathcal{J}} U_j(\pi(u_j))$$

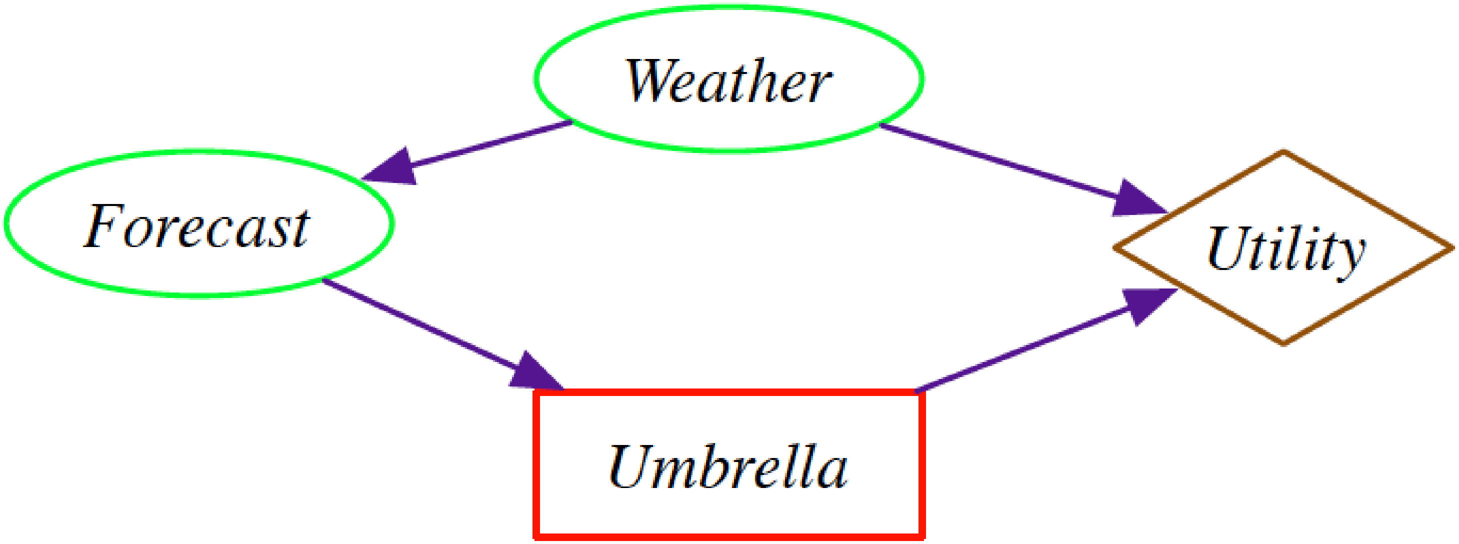
- The optimal policy can be found by variable elimination while maximizing over decisions:
  - first consider the last decision
  - find an optimal decision for each value of its parents and produce a factor of these maximum values.
  - recursively solve for the remaining decisions

Umbrella network



- You don't get to observe the weather when you have to decide whether to take your umbrella. You do get to observe the forecast. Rain will cause wet grass.

Finding the optimal policy



- Remove all variables not ancestors of the utility node

Finding the optimal policy

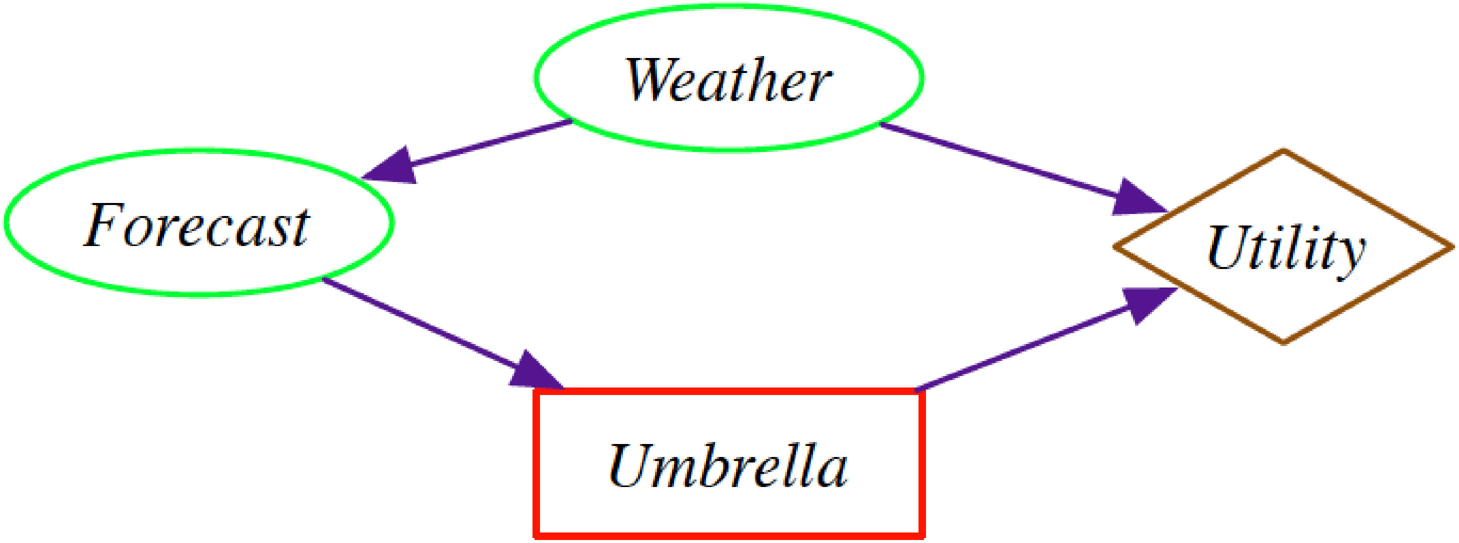
Weather	Value
norain	0.7
rain	0.3

Weather	Fcast	Value
norain	sunny	0.7
norain	cloudy	0.2
norain	rainy	0.1
rain	sunny	0.15
rain	cloudy	0.25
rain	rainy	0.6

Weather	Umb	Value
norain	take	20
norain	leave	100
rain	take	70
rain	leave	0

- Create a factor for each conditional probability table and a factor for the utility.

Finding the optimal policy



$$U^* = \sum_{F,W} \max_{U} f_1(W) f_2(W,F) f_3(W,U)$$

Finding the optimal policy

- Sum out variables not (parents of) a decision node D

$$
\begin{aligned}
U^* &= \sum_F \max_U \sum_W f_1(W) f_2(W,F) f_3(W,U) \\
&= \sum_F \max_U f_4(F,U)
\end{aligned}
$$

Forecast	*Umbrella*	*Value*
sunny	*takeIt*	12.95
sunny	*leaveIt*	49.0
cloudy	*takeIt*	8.05
cloudy	*leaveIt*	14.0
rainy	*takeIt*	14.0
rainy	*leaveIt*	7.0

Finding the optimal policy

$$U^* = \sum_F \max_U f_4(F, U)$$

- Select D that is in a factor f with (some of) 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, $\max_D f$

Forecast	Umbrella
sunny	leaveIt
cloudy	leaveIt
rainy	takeIt

Forecast	Value
sunny	49.0
cloudy	14.0
rainy	14.0

- the final sum returns the maximized expected utility:

$$U^* = \sum_F f_5(F) = 77$$

Value of information

- The amount a decision maker would be willing to pay for information on X prior to making a decision D
- The value of information on X for decision D is the expected utility of the network with an arc from X to D minus the expected utility of the network without the arc.
- The value of information is always non-negative.
- It is positive only if the agent changes its action depending on X.
- The value of information provides a bound on how much an agent should be prepared to pay for a sensor. How much is a better diagnosis worth?

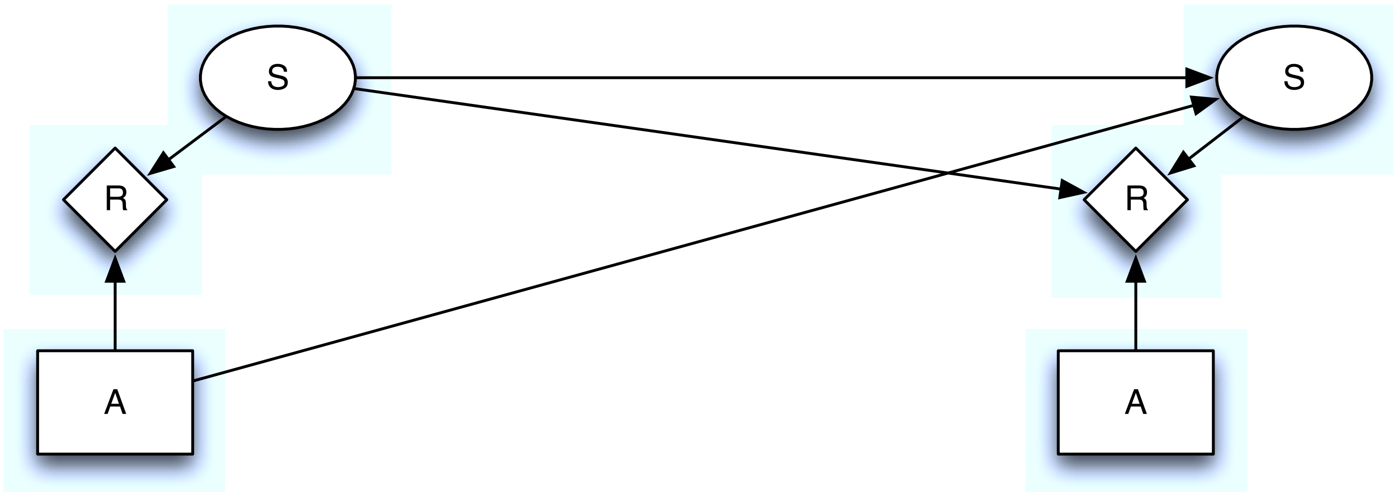
Value of control

- The amount a decision maker would be willing to pay in order to be able to control a random variable X
- The value of control of a variable X is the expected utility of the network when you make X a decision variable minus the expected utility of the network when X is a random variable.
- You need to be explicit about what information is available when you control X.
- Controlling X can be worse than observing X. E.g., controlling a thermometer.
- If you keep the parents the same, the value of control is always non-negative.

Infinite horizon decision problems

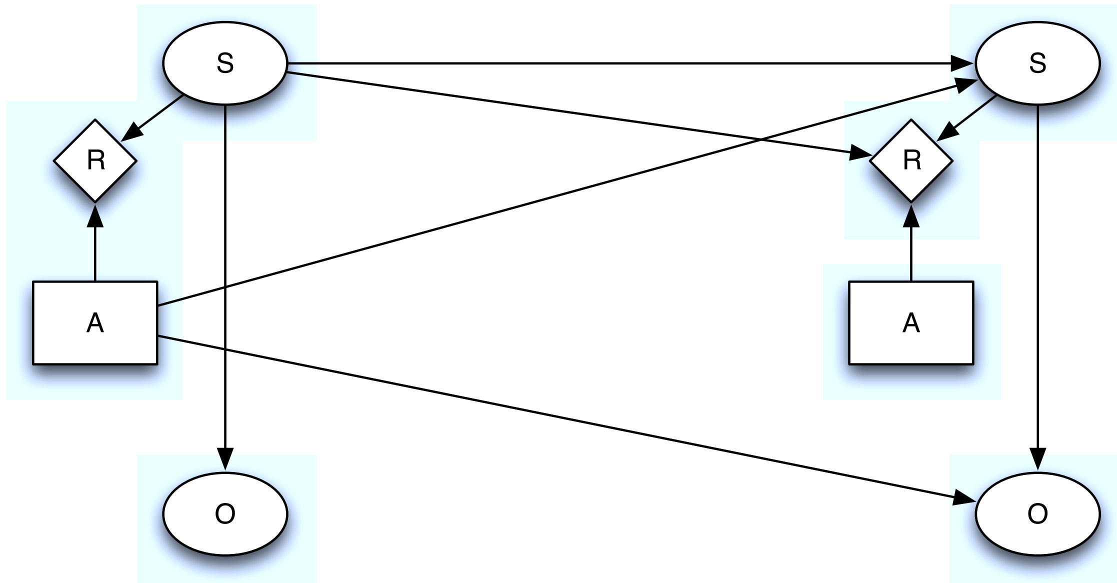
- Often an agent must reason about an ongoing process or it does not know how many actions it will be required to perform
- Called infinite horizon problems when the process may go on forever
- Called indefinite horizon problems when the agent will eventually stop but does not know when
- Discussed methods fail since they start solving from the last decision.
- Modelled as (partially observable) Markov decision problems

MDP



- S, a set of states of the world.
- A, a set of actions.
- $P : S \times S \times A \rightarrow [0, 1]$, written as $P(s'|s, a)$
- $R : S \times A \times S \rightarrow R$, written as $R(s, a, s')$

POMDP



As an MDP but additionally:

- O, a set of possible observations;
- $P(s_0)$, which gives the probability distribution of the starting state
- $P(o \mid s, a)$, which gives the probability of observing o given the state is s and the previous action a.

Solving a POMDP

Goal: maximize

$$E\left[\sum_{t=0}^{h} \gamma^t R_t\right] .$$

- Represent the problem as an influence diagram (expensive, approximate)

Solving a POMDP

Goal: maximize

$$E\left[\sum_{t=0}^{h} \gamma^t R_t\right] .$$

- Represent the problem as an influence diagram (expensive, approximate)
- Find the optimal policy $\pi^*: B \rightarrow A$ over the belief state B using value iteration or policy iteration

Solving a POMDP

Goal: maximize

$$E\left[\sum_{t=0}^{h} \gamma^t R_t\right] .$$

- Represent the problem as an influence diagram (expensive, approximate)
- Find the optimal policy $\pi^*: B \rightarrow A$ over the belief state B using value iteration or policy iteration
- Search through the space of policy graphs

Summary

- Reasoning also presupposes optimal decision making
- Preference capture desirability of outcomes
- Should follow the von Neumann-Morgenstern axioms
- Preference is quantified in terms of utility
- Finite decision problems can be solved using decision trees or influence diagrams
- Infinite decision problems require other solution strategies
- Relevant for robot path planning, adaptive brain-computer interfaces
- Next week: probabilistic logic