

**Silvia Rossi**

## Decisioni in caso di incertezza

# Lezione n. 5


# Corso di Laurea: Informatica

# Insegnamento: Sistemi multi-agente

**Email:**  
silrossi@unina.it

# A.A. 2014-2015

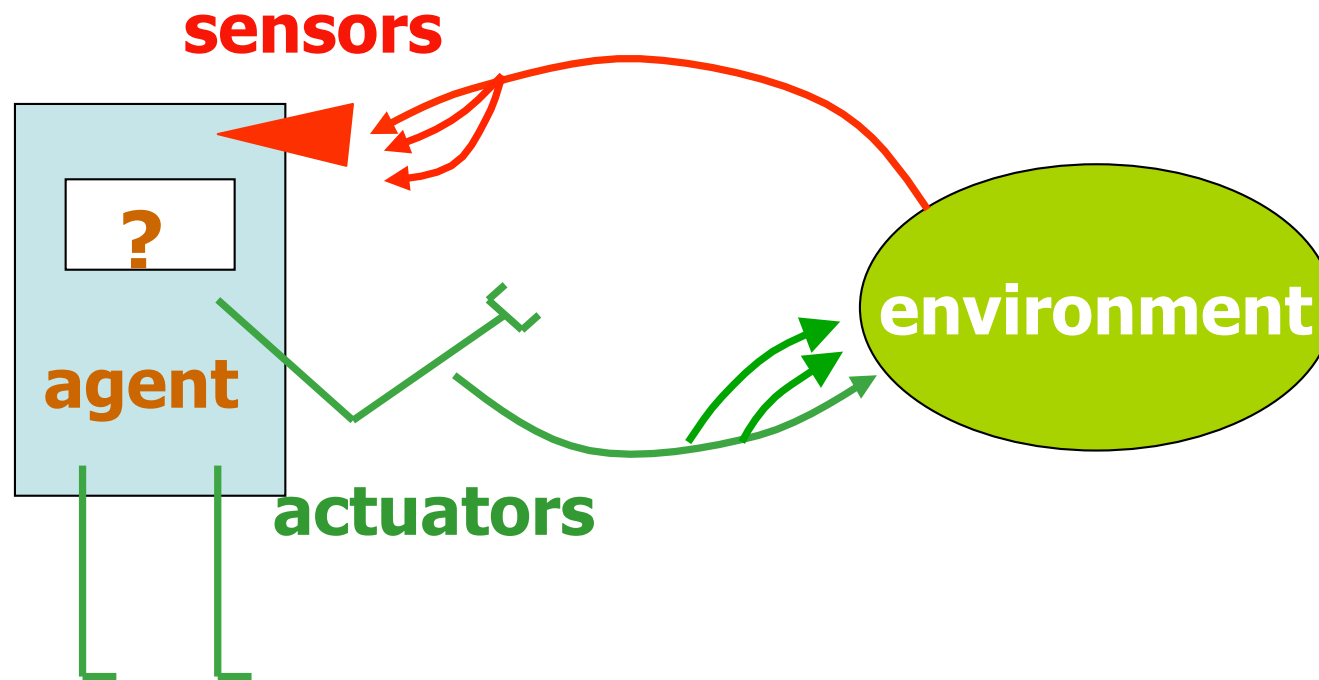




# **Decision-Making with Probabilistic Uncertainty**

*(R&N: 16.1, 16.5, 16.6)*

# Utility-Based Agent

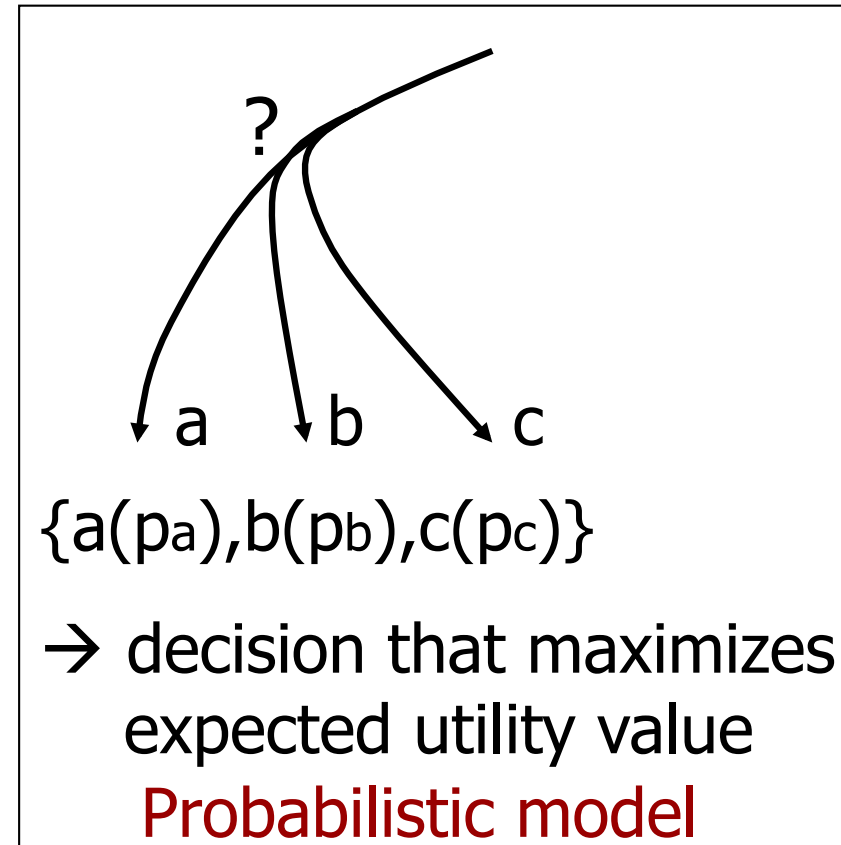
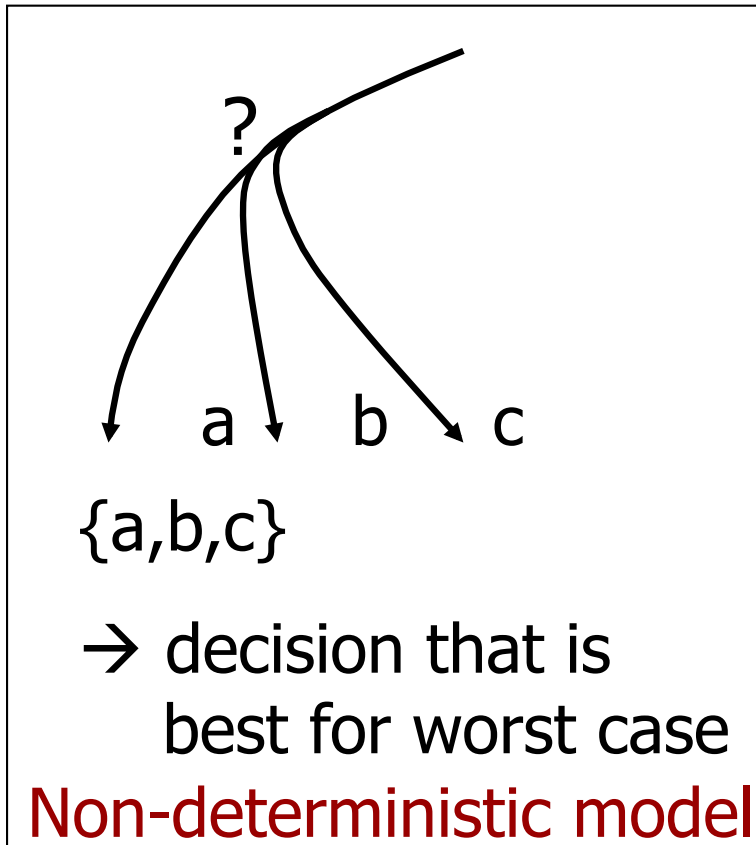


## General Framework

- An agent operates in some given finite state space
- **No goal state**; instead, states provide **rewards** (positive, negative, or null) that quantify in a single unit system what the agent gets when it visits this state (e.g., a bag of gold, a sunny afternoon on the beach, a speeding ticket, etc...)
- Each action has several possible outcomes, each with some probability; sensing may also be imperfect
- The agent's goal is to plan a strategy (here, it is called a **policy**) to maximize the **expected** amount of rewards collected

- Uncertainty in action only  
[The world is fully observable]
- Uncertainty in both action and sensing  
[The world is partially observable]

# Non-deterministic vs. Probabilistic Uncertainty



~ **Adversarial search**

Action a:

$$s \in S \rightarrow a(s) = \underbrace{\{s_1 (p_1), s_2 (p_2), \dots, s_n (p_n)\}}_{\sum_{i=1}^n p_i = 1}$$

**Markov assumption:** The action model  $a(s)$  does not depend on what happened prior to reaching  $s$

Random variable  $X$  with  $n$  values  $x_1, \dots, x_n$  and distribution  $(p_1, \dots, p_n)$

E.g.:  $X$  is the state reached after doing an action  $A$  under uncertainty

Function  $U$  of  $X$

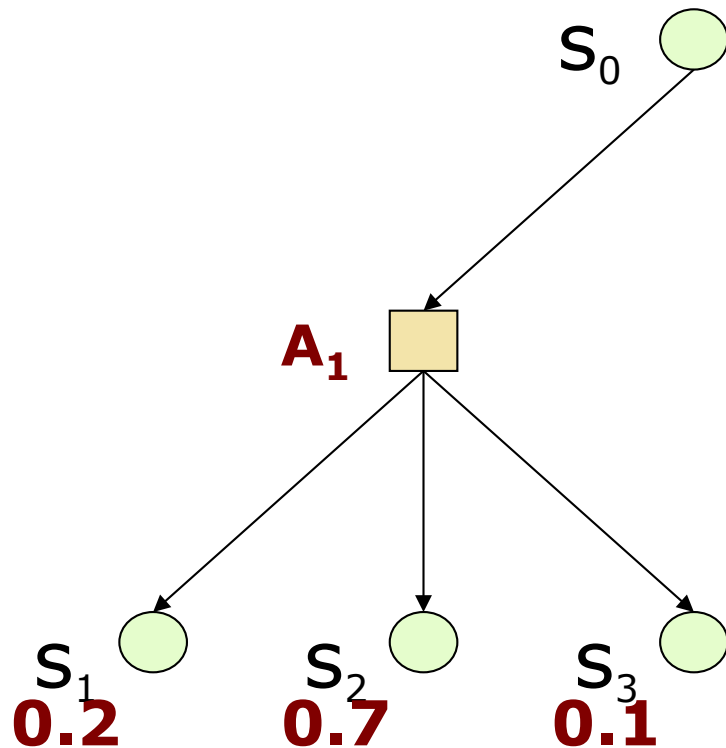
E.g.,  $U$  is the utility of a state

The **expected utility** of  $A$  is

$$EU[A] = \sum_{i=1, \dots, n} p(x_i|A)U(x_i)$$

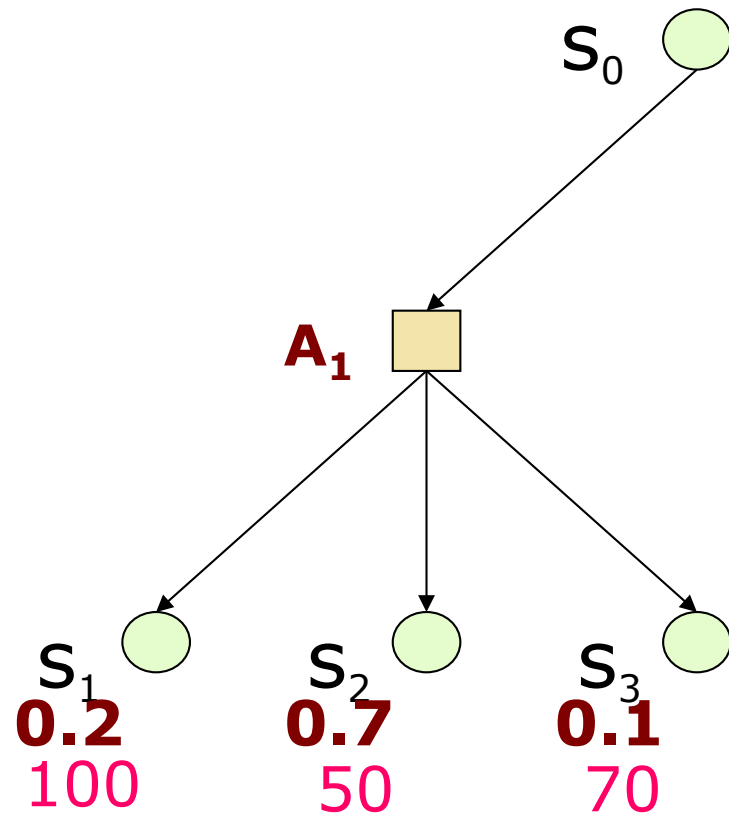


## Starting very simple ...



- $S_0$  describes many actual states of the real world.  $A_1$  reaches  $s_1$  in some states,  $s_2$  in others, and  $s_3$  in the remaining ones
- If the agent could return to  $S_0$  many times in independent ways and if at each time it executed  $A_1$ , then it would reach  $s_1$  20% of the times,  $s_2$  70% of the times, and  $s_3$  10% of the times

# Introducing rewards ...

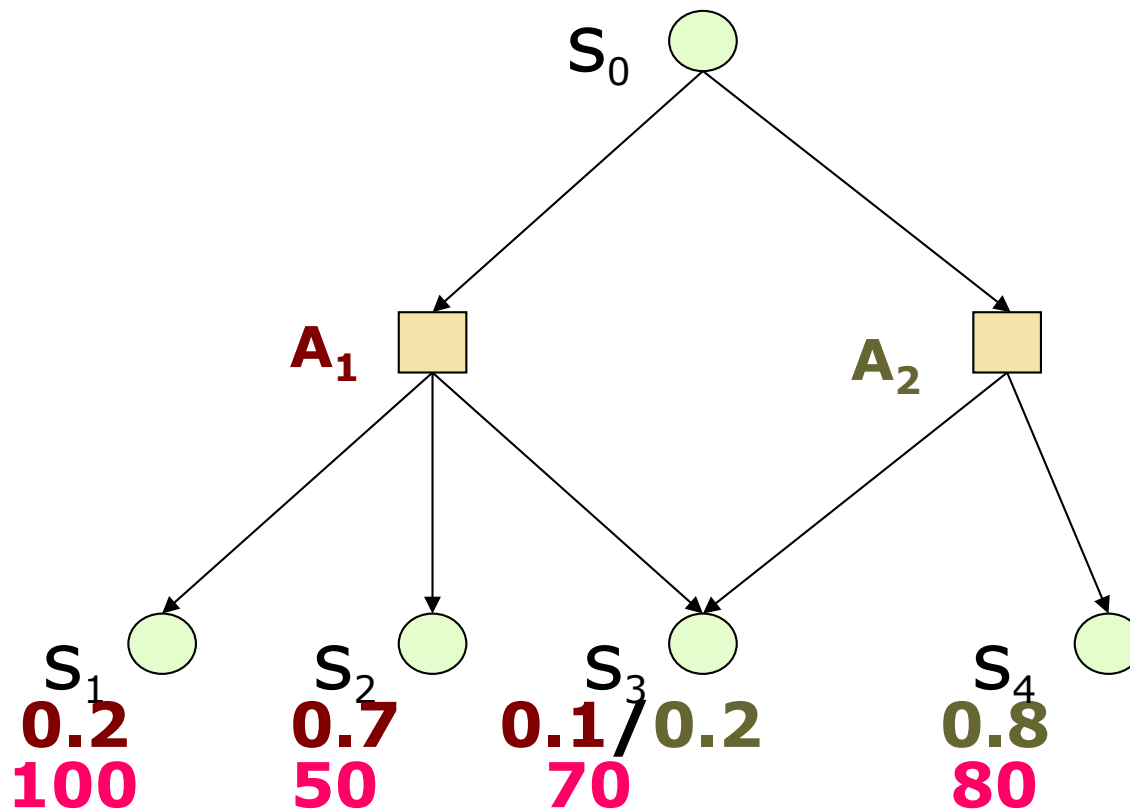


- Assume that the agent receives **rewards** in some states (rewards can be positive or negative)
- If the agent could execute  $A_1$  in  $S_0$  many times, the **average (expected) reward** that it would get is:

$$\begin{aligned} U_1(S_0) &= 100 \times 0.2 + 50 \times 0.7 + 70 \times 0.1 \\ &= 20 + 35 + 7 \\ &= 62 \end{aligned}$$

← rewards associated with states  $s_1$ ,  $s_2$ , and  $s_3$

## ... and a second action ...



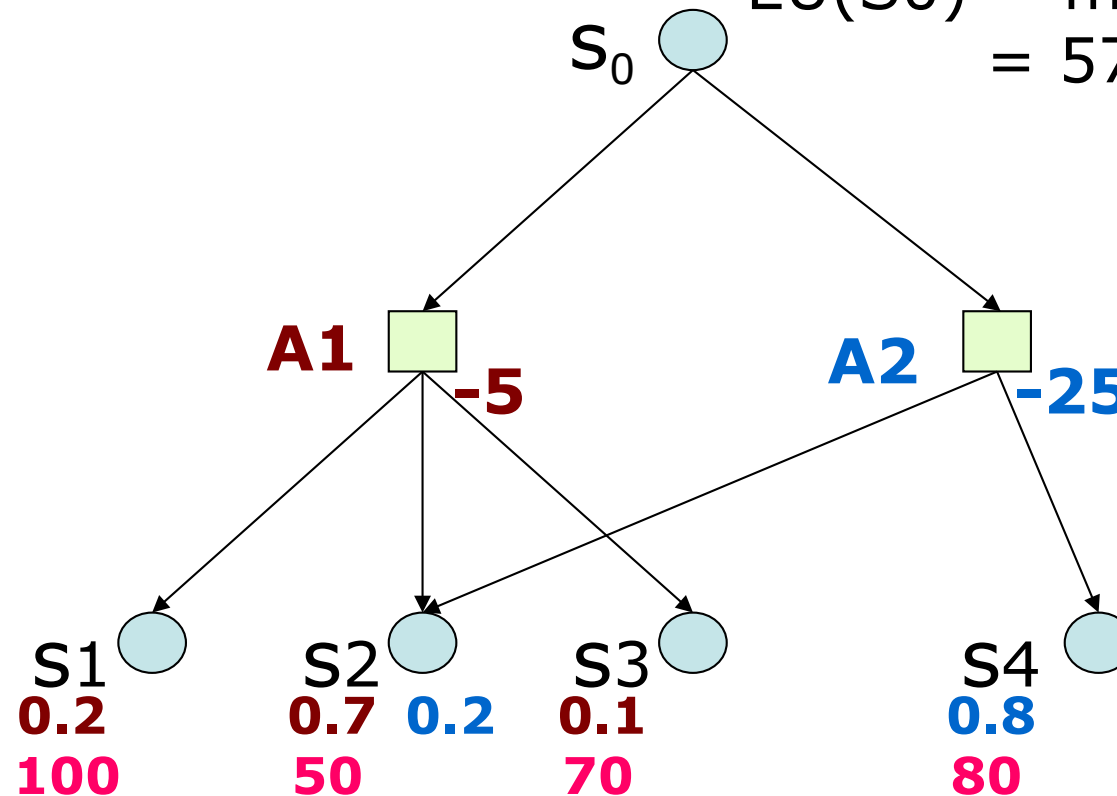
- $U_1(S_0) = 62$
- $U_2(S_0) = 78$
- If the agent chooses to execute  $A_2$ , it will maximize the average collected rewards

# Introducing Action Costs

$$EU(A1) = 62 - 5 = 57$$

$$EU(A2) = 74 - 25 = 49$$

$$EU(S0) = \max\{EU(A1), EU(A2)\} \\ = 57$$



rational agent should choose the action that maximizes agent's expected utility

this is the basis of the field of decision theory

normative criterion for rational choice of action

IT'S ALL SOLVED!!!

Must have **complete** model of:

Actions

Utilities

States

Even if you have a complete model, will be computationally **intractable**

In fact, a truly rational agent takes into account the utility of reasoning as well---**bounded rationality**

Nevertheless, great progress has been made in this area recently, and we are able to solve much more complex decision theoretic problems than ever before

## Decision Theoretic Planning

Simple decision making (ch. 16)

Sequential decision making (ch. 17)

Extend BNs to handle actions and utilities

Also called Influence diagrams

Make use of BN inference

Can do Value of Information calculations



# How do we represent Uncertainty?

We need to answer several questions:

What do we represent & how we represent it?

What language do we use to represent our uncertainty? What are the semantics of our representation?

What can we do with the representations?

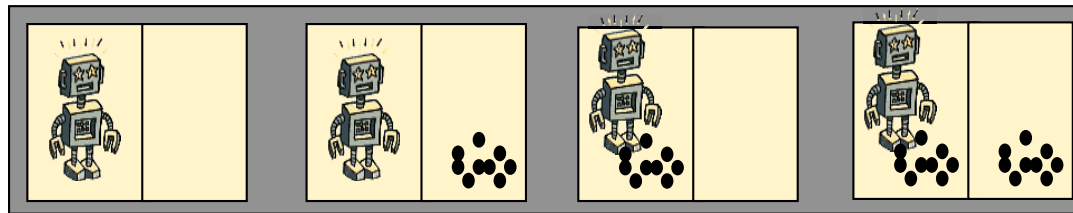
What queries can be answered? How do we answer them?

How do we construct a representation?

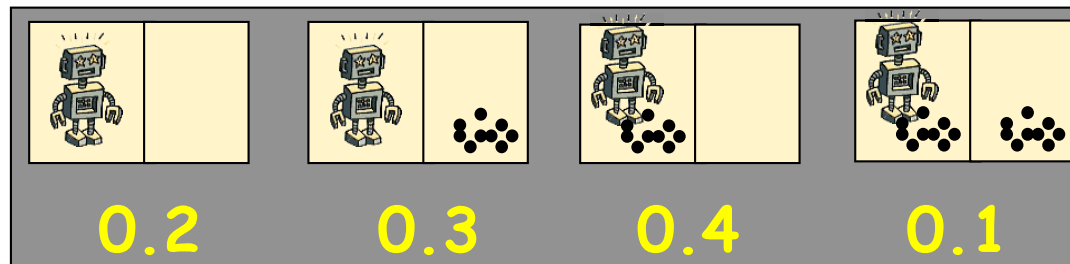
Can we ask an expert? Can we learn from data?

## Example: Belief State

- In the presence of non-deterministic sensory uncertainty, an agent **belief state** represents all the states of the world that it thinks are possible at a given time or at a given stage of reasoning



- In the probabilistic model of uncertainty, a probability is associated with each state to measure its likelihood to be the actual state



## Making decisions under uncertainty

Suppose I believe the following:

$P(A_{25} \text{ gets me there on time} \mid \dots)$	$= 0.04$
$P(A_{90} \text{ gets me there on time} \mid \dots)$	$= 0.70$
$P(A_{120} \text{ gets me there on time} \mid \dots)$	$= 0.95$
$P(A_{1440} \text{ gets me there on time} \mid \dots)$	$= 0.9999$

Which action to choose?

Depends on my **preferences** for missing flight vs. time spent waiting, etc.

**Utility theory** is used to represent and infer preferences

**Decision theory** = probability theory + utility theory

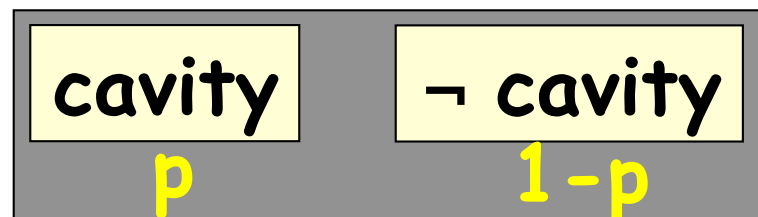
*Decision Theory: An **agent is rational exactly** when it chooses the **action with the maximum expected utility taken over all results of actions.***

```
function DT-AGENT(percept) returns an action
  static: a set probabilistic beliefs about the state of the world

  calculate updated probabilities for current state based on
    available evidence including current percept and previous action
  calculate outcome probabilities for actions,
    given action descriptions and probabilities of current states
  select action with highest expected utility
    given probabilities of outcomes and utility information
  return action
```

## Example

- Consider a world where a dentist agent D meets a new patient P
- D is interested in only one thing: whether P has a cavity, which D models using the proposition Cavity
- Before making any observation, D's belief state is:



- This means that D believes that a fraction  $p$  of patients have cavities

## Where do probabilities come from?

- Frequencies observed in the past, e.g., by the agent, its designer, or others
- Symmetries, e.g.:
  - If I roll a dice, each of the 6 outcomes has probability  $1/6$
- Subjectivism, e.g.:
  - If I drive on Highway 280 at 120mph, I will get a speeding ticket with probability 0.6
  - Principle of indifference: If there is no knowledge to consider one possibility more probable than another, give them the same probability

Basic element: **random variable**

Similar to propositional logic: possible worlds defined by assignment of values to random variables.

**Boolean** random variables

e.g., *Cavity* (do I have a cavity?)

**Discrete** random variables

e.g., *Weather* is one of  $\langle \text{sunny}, \text{rainy}, \text{cloudy}, \text{snow} \rangle$

Domain values must be exhaustive and mutually exclusive

Elementary proposition constructed by assignment of a value to a random variable:

e.g., *Weather = sunny, Cavity = false*  
(abbreviated as  $\neg cavity$ )

Complex propositions formed from elementary propositions and standard logical connectives

e.g., *Weather = sunny  $\vee$  Cavity = false*



**Atomic event:** A **complete** specification of the state of the world about which the agent is uncertain

E.g., if the world consists of only two Boolean variables *Cavity* and *Toothache*, then there are 4 distinct atomic events:

*Cavity* = *false*  $\wedge$  *Toothache* = *false*

*Cavity* = *false*  $\wedge$  *Toothache* = *true*

*Cavity* = *true*  $\wedge$  *Toothache* = *false*

*Cavity* = *true*  $\wedge$  *Toothache* = *true*

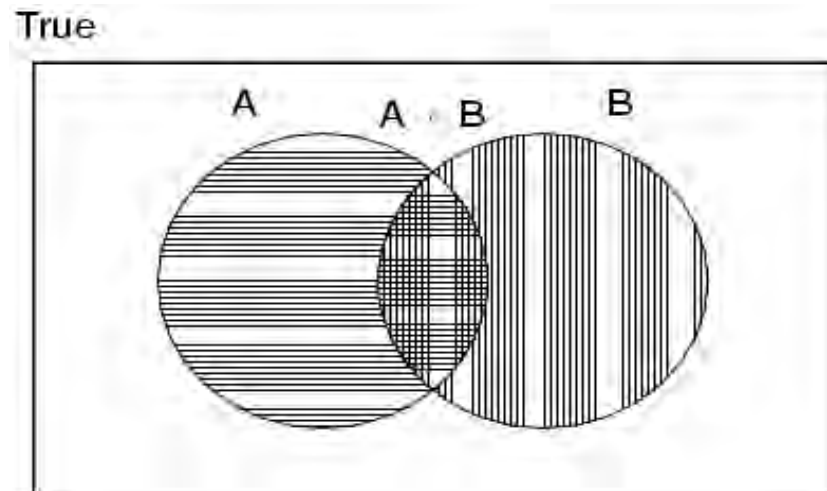
Atomic events are mutually exclusive and exhaustive

For any propositions  $A, B$

$$0 \leq P(A) \leq 1$$

$$P(\text{true}) = 1 \text{ and } P(\text{false}) = 0$$

$$P(A \vee B) = P(A) + P(B) - P(A \wedge B)$$



## Prior or unconditional probabilities of propositions

e.g.,  $P(\text{Cavity} = \text{true}) = 0.1$  and  $P(\text{Weather} = \text{sunny}) = 0.72$  correspond to belief prior to arrival of any (new) evidence

Probability distribution gives values for all possible assignments:

$$\mathbf{P}(\text{Weather}) = \langle 0.72, 0.1, 0.08, 0.1 \rangle$$

(normalized, i.e., sums to 1)

**Joint probability distribution** for a set of random variables gives the probability of every atomic event on those random variables

$\mathbf{P}(\textit{Weather}, \textit{Cavity})$  = a  $4 \times 2$  matrix of values:

<i>Weather</i> =	sunny	rainy	cloudy	snow
<i>Cavity</i> = true	0.144	0.02	0.016	0.02
<i>Cavity</i> = false	0.576	0.08	0.064	0.08

Every question about a domain can be answered by the joint distribution

Start with the joint probability distribution:

	<i>toothache</i>		$\neg$ <i>toothache</i>	
	<i>catch</i>	$\neg$ <i>catch</i>	<i>catch</i>	$\neg$ <i>catch</i>
<i>cavity</i>	<b>.108</b>	<b>.012</b>	<b>.072</b>	<b>.008</b>
$\neg$ <i>cavity</i>	<b>.016</b>	<b>.064</b>	<b>.144</b>	<b>.576</b>

*The probability of a proposition is equal to the sum of the probabilities of the atomic events in which it holds;*

For any proposition  $\phi$ , sum the atomic events where it is true:  $P(\phi) = \sum_{\omega: \omega \models \phi} P(\omega)$

Start with the joint probability distribution:

	<i>toothache</i>		$\neg$ <i>toothache</i>	
	<i>catch</i>	$\neg$ <i>catch</i>	<i>catch</i>	$\neg$ <i>catch</i>
<i>cavity</i>	<b>.108</b>	<b>.012</b>	<b>.072</b>	<b>.008</b>
$\neg$ <i>cavity</i>	<b>.016</b>	<b>.064</b>	<b>.144</b>	<b>.576</b>

For any proposition  $\phi$ , sum the atomic events where it is true:  $P(\phi) = \sum_{\omega: \omega \models \phi} P(\omega)$

$P(\text{toothache}) = ?$

Start with the joint probability distribution:

	<i>toothache</i>		$\neg$ <i>toothache</i>	
	<i>catch</i>	$\neg$ <i>catch</i>	<i>catch</i>	$\neg$ <i>catch</i>
<i>cavity</i>	<b>.108</b>	<b>.012</b>	<b>.072</b>	<b>.008</b>
$\neg$ <i>cavity</i>	<b>.016</b>	<b>.064</b>	<b>.144</b>	<b>.576</b>

For any proposition  $\phi$ , sum the atomic events where it is true:  $P(\phi) = \sum_{\omega: \omega \models \phi} P(\omega)$

$$P(\text{toothache}) = 0.108 + 0.012 + 0.016 + 0.064 = 0.2$$

$$P(\text{cavity} \vee \text{toothache}) = ?$$

Start with the joint probability distribution:

	<i>toothache</i>		$\neg$ <i>toothache</i>	
	<i>catch</i>	$\neg$ <i>catch</i>	<i>catch</i>	$\neg$ <i>catch</i>
<i>cavity</i>	<b>.108</b>	<b>.012</b>	<b>.072</b>	<b>.008</b>
$\neg$ <i>cavity</i>	<b>.016</b>	<b>.064</b>	<b>.144</b>	<b>.576</b>

For any proposition  $\phi$ , sum the atomic events where it is true:  $P(\phi) = \sum_{\omega: \omega \models \phi} P(\omega)$

$$P(\text{toothache}) = 0.108 + 0.012 + 0.016 + 0.064 = 0.2$$

$$P(\text{cavity} \vee \text{toothache}) =$$



New information can change the probability.

Example: The probability of a cavity increases if we know the patient has a toothache.

If additional information is available, we can no longer use the **prior probabilities!**

$P(A|B)$  is the **conditional or posterior probability** of A given that *all we know is B*:

$$P(\text{Cavity} \mid \text{Toothache}) = 0.8$$

**$P(X|Y)$**  is the **table of all conditional probabilities** over all values of X and Y.

## Conditional Probability

Definition:

$$P(A|B) = P(A \cap B) / P(B)$$

Read  $P(A|B)$ : probability of A given B

can also write this as:

$$P(A \cap B) = P(A|B) P(B)$$

called the **product rule**

Conditional or posterior probabilities

e.g.,  $P(\text{cavity} \mid \text{toothache}) = 0.8$

i.e., given that *toothache* is all I know

Notation for conditional distributions:

$\mathbf{P}(\text{Cavity} \mid \text{Toothache}) = (2\text{-element vector of } 2\text{-element vectors})$

If we know more, e.g., *cavity* is also given, then we have

$P(\text{cavity} \mid \text{toothache}, \text{cavity}) = 1$

New evidence may be irrelevant, allowing simplification, e.g.,

$P(\text{cavity} \mid \text{toothache}, \text{sunny}) = P(\text{cavity} \mid \text{toothache}) = 0.8$

This kind of inference, sanctioned by domain knowledge, is crucial

Definition of conditional probability:

$$P(a \mid b) = P(a \wedge b) / P(b) \text{ if } P(b) > 0$$

**Product rule** gives an alternative formulation:

$$P(a \wedge b) = P(a \mid b) P(b) = P(b \mid a) P(a)$$

A general version holds for whole distributions, e.g.,

$$\mathbf{P}(\textit{Weather}, \textit{Cavity}) = \mathbf{P}(\textit{Weather} \mid \textit{Cavity}) \mathbf{P}(\textit{Cavity})$$

(View as a set of  $4 \times 2$  equations, **not** matrix mult.)

**Chain rule** is derived by successive application of product rule:

$$\begin{aligned} \mathbf{P}(X_1, \dots, X_n) &= \mathbf{P}(X_2, \dots, X_n) \mathbf{P}(X_1 \mid X_2, \dots, X_n) \\ &= \mathbf{P}(X_3, \dots, X_n) \mathbf{P}(X_2 \mid X_3, \dots, X_n) \mathbf{P}(X_1 \mid X_2, \dots, X_n) \\ &= \dots \\ &= \prod_{i=1}^n \mathbf{P}(X_i \mid X_{i+1}, \dots, X_n) \end{aligned}$$

$$P(A \wedge B \wedge C) = P(A|B,C) P(B|C) P(C)$$

Start with the joint probability distribution:

	<i>toothache</i>		$\neg$ <i>toothache</i>	
	<i>catch</i>	$\neg$ <i>catch</i>	<i>catch</i>	$\neg$ <i>catch</i>
<i>cavity</i>	<b>.108</b>	<b>.012</b>	<b>.072</b>	<b>.008</b>
$\neg$ <i>cavity</i>	<b>.016</b>	<b>.064</b>	<b>.144</b>	<b>.576</b>

Can also compute conditional probabilities:

$$\begin{aligned}
 P(\neg \text{cavity} \mid \text{toothache}) &= \frac{P(\neg \text{cavity} \wedge \text{toothache})}{P(\text{toothache})} \\
 &= \frac{0.016 + 0.064}{0.108 + 0.012 + 0.016 + 0.064} \\
 &= 0.4
 \end{aligned}$$

Denominator can be viewed as a **normalization constant**  $\alpha$

	<i>toothache</i>		$\neg$ <i>toothache</i>	
	<i>catch</i>	$\neg$ <i>catch</i>	<i>catch</i>	$\neg$ <i>catch</i>
<i>cavity</i>	<b>.108</b>	<b>.012</b>	.072	.008
$\neg$ <i>cavity</i>	<b>.016</b>	<b>.064</b>	.144	.576

$$\begin{aligned}
 \mathbf{P}(\text{Cavity} \mid \text{toothache}) &= \alpha, \mathbf{P}(\text{Cavity}, \text{toothache}) \\
 &= \alpha, [\mathbf{P}(\text{Cavity}, \text{toothache}, \text{catch}) + \\
 &\quad \mathbf{P}(\text{Cavity}, \text{toothache}, \neg \text{catch})] \\
 &= \alpha, [<0.108, 0.016> + <0.012, 0.064>] \\
 &= \alpha, <0.12, 0.08> = <0.6, 0.4>
 \end{aligned}$$

General idea: compute distribution on query variable by fixing **evidence variables** and summing over **hidden variables**

## Inference by enumeration, contd.

Typically, we are interested in  
the posterior joint distribution of the **query variables  $\mathbf{Y}$**   
given specific values  **$\mathbf{e}$**  for the **evidence variables  $\mathbf{E}$**

Let the **hidden variables** be  **$\mathbf{H} = \mathbf{X} - \mathbf{Y} - \mathbf{E}$**

Then the required summation of joint entries is done by summing out the hidden variables:

$$\mathbf{P}(\mathbf{Y} \mid \mathbf{E} = \mathbf{e}) = \alpha \mathbf{P}(\mathbf{Y}, \mathbf{E} = \mathbf{e}) = \alpha \sum_{\mathbf{h}} \mathbf{P}(\mathbf{Y}, \mathbf{E} = \mathbf{e}, \mathbf{H} = \mathbf{h})$$

The terms in the summation are joint entries because  **$\mathbf{Y}$** ,  **$\mathbf{E}$**  and  **$\mathbf{H}$**  together exhaust the set of random variables

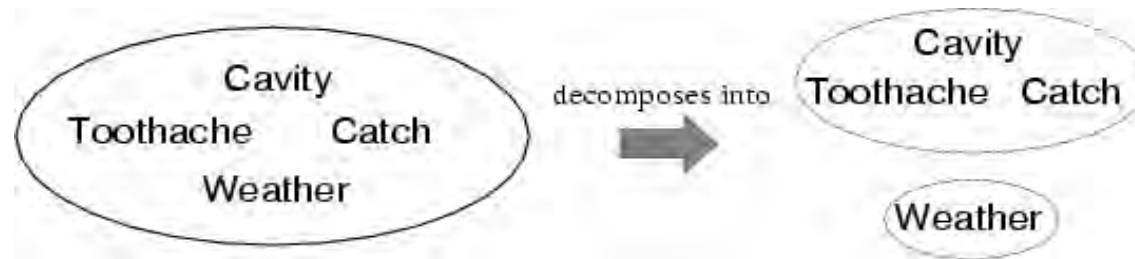
Obvious problems:

1. Worst-case time complexity  $O(d^n)$  where  $d$  is the largest arity
2. Space complexity  $O(d^n)$  to store the joint distribution
3. How to find the numbers for  $O(d^n)$  entries?



$A$  and  $B$  are independent iff

$$\mathbf{P}(A|B) = \mathbf{P}(A) \quad \text{or} \quad \mathbf{P}(B|A) = \mathbf{P}(B) \quad \text{or} \quad \mathbf{P}(A, B) = \mathbf{P}(A) \mathbf{P}(B)$$



$$\begin{aligned} &\mathbf{P}(\textit{Toothache}, \textit{Catch}, \textit{Cavity}, \textit{Weather}) \\ &= \mathbf{P}(\textit{Toothache}, \textit{Catch}, \textit{Cavity}) \mathbf{P}(\textit{Weather}) \end{aligned}$$

32 entries reduced to 12; for  $n$  independent biased coins,  $O(2^n) \rightarrow O(n)$

Absolute independence powerful but rare

Dentistry is a large field with hundreds of variables, none of which are independent. What to do?

Unfortunately, random variables of interest are not independent of each other

A more suitable notion is that of **conditional independence**

Two variables  $X$  and  $Y$  are **conditionally independent** given  $Z$  if

$$P(X = x | Y = y, Z = z) = P(X = x | Z = z) \text{ for all values } x, y, z$$

That is, learning the values of  $Y$  does not change prediction of  $X$  once we know the value of  $Z$

$$I(X, Y | Z)$$

Three propositions:

Gas

Battery

Starts

$P(\text{Battery}|\text{Gas}) = P(\text{Battery})$   
Gas and Battery are independent

$P(\text{Battery}|\text{Gas}, \text{Starts}) \neq P(\text{Battery}|\text{Starts})$   
Gas and Battery are not independent given Starts

## Conditional independence

$\mathbf{P}(\textit{Toothache}, \textit{Cavity}, \textit{Catch})$  has  $2^3 - 1 = 7$  independent entries

If I have a cavity, the probability that the probe catches in it doesn't depend on whether I have a toothache:

$$(1) \mathbf{P}(\textit{catch} \mid \textit{toothache}, \textit{cavity}) = \mathbf{P}(\textit{catch} \mid \textit{cavity})$$

The same independence holds if I haven't got a cavity:

$$(2) \mathbf{P}(\textit{catch} \mid \textit{toothache}, \neg \textit{cavity}) = \mathbf{P}(\textit{catch} \mid \neg \textit{cavity})$$

*Catch* is **conditionally independent** of *Toothache* given *Cavity*:

$$\mathbf{P}(\textit{Catch} \mid \textit{Toothache}, \textit{Cavity}) = \mathbf{P}(\textit{Catch} \mid \textit{Cavity})$$

## Conditional independence

*Catch* is **conditionally independent** of *Toothache* given *Cavity*:

$$\mathbf{P}(\textit{Catch} \mid \textit{Toothache}, \textit{Cavity}) = \mathbf{P}(\textit{Catch} \mid \textit{Cavity})$$

Equivalent statements:

$$\mathbf{P}(\textit{Toothache} \mid \textit{Catch}, \textit{Cavity}) = \mathbf{P}(\textit{Toothache} \mid \textit{Cavity})$$

$$\mathbf{P}(\textit{Toothache}, \textit{Catch} \mid \textit{Cavity}) = \mathbf{P}(\textit{Toothache} \mid \textit{Cavity}) \mathbf{P}(\textit{Catch} \mid \textit{Cavity})$$

$$\mathbf{P}(\textit{Catch}, \textit{Toothache} \mid \textit{Cavity}) = \mathbf{P}(\textit{Toothache} \mid \textit{Cavity}) \mathbf{P}(\textit{Catch} \mid \textit{Cavity})$$

$$\mathbf{P}(A, B \mid C) = \mathbf{P}(A \mid C) \mathbf{P}(B \mid C)$$

$$\mathbf{P}(A, B \mid C) = \mathbf{P}(A \mid B, C) \mathbf{P}(B \mid C)$$

$$P(X, Y | Z) = P(X | Z)P(Y | Z)$$

$$P(X | Y, Z) = P(X | Z) \text{ and } P(Y | X, Z) = P(Y | Z)$$

The decomposition of large probabilistic domains into weakly connected subsets via conditional independence is one of the most important developments in the recent history of AI.

## Conditional independence contd.

Write out full joint distribution using chain rule:

$$\begin{aligned} & \mathbf{P}(\textit{Toothache}, \textit{Catch}, \textit{Cavity}) \\ &= \mathbf{P}(\textit{Toothache} \mid \textit{Catch}, \textit{Cavity}) \mathbf{P}(\textit{Catch}, \textit{Cavity}) \\ &= \mathbf{P}(\textit{Toothache} \mid \textit{Catch}, \textit{Cavity}) \mathbf{P}(\textit{Catch} \mid \textit{Cavity}) \mathbf{P}(\textit{Cavity}) \\ &= \mathbf{P}(\textit{Toothache} \mid \textit{Cavity}) \mathbf{P}(\textit{Catch} \mid \textit{Cavity}) \mathbf{P}(\textit{Cavity}) \end{aligned}$$

I.e.,  $2 + 2 + 1 = 5$  independent numbers

In most cases, the use of conditional independence reduces the size of the representation of the joint distribution from exponential in  $n$  to linear in  $n$ .

Conditional independence is our most basic and robust form of knowledge about uncertain environments.

Example 1:  $I(X, Y | \emptyset)$  and not  $I(X, Y | Z)$

Example 2:  $I(X, Y | Z)$  and not  $I(X, Y | \emptyset)$

conclusion: independence does not imply conditional independence!



Product rule  $P(a \wedge b) = P(a \mid b) P(b) = P(b \mid a) P(a)$

The order is unimportant

$\Rightarrow$  **Bayes' rule:**  $P(a \mid b) = P(b \mid a) P(a) / P(b)$

or in distribution form

$$\mathbf{P}(Y|X) = \mathbf{P}(X|Y) \mathbf{P}(Y) / \mathbf{P}(X) = \alpha \mathbf{P}(X|Y) \mathbf{P}(Y)$$

Useful for assessing **diagnostic** probability from **causal** probability:

$$P(\text{Cause}|\text{Effect}) = P(\text{Effect}|\text{Cause}) P(\text{Cause}) / P(\text{Effect})$$

E.g., let  $M$  be meningitis,  $S$  be stiff neck:

$$P(m|s) = P(s|m) P(m) / P(s) = 0.8 \times 0.0001 / 0.1 = 0.0008$$

Note: posterior probability of meningitis still very small!

If we wish to determine the absolute probability of  $P(C \mid T)$  and we do not know  $P(T)$ , we can also carry out a complete case analysis (e.g. for  $C$  and  $\neg C$ ) and use the fact that  $P(C \mid T) + P(\neg C \mid T) = 1$  (here boolean variables):

$$P(C|T) = \frac{P(T|C) P(C)}{P(T)}$$

$$P(\neg C|T) = \frac{P(T|\neg C) P(\neg C)}{P(T)}$$

$$P(C|T) + P(\neg C|T) = \frac{P(T|C) P(C)}{P(T)} + \frac{P(T|\neg C) P(\neg C)}{P(T)}$$

$$P(T) = P(T|C) P(C) + P(T|\neg C) P(\neg C)$$

By substituting into the first equation:

$$P(C|T) = \frac{P(T|C) P(C)}{P(T|C) P(C) + P(T|\neg C) P(\neg C)}$$

For random variables with multiple values:

$$\mathbf{P(Y | X)} = \alpha \mathbf{P(X | Y)P(Y)}$$

where  $\alpha$  is the normalization constant needed to make the entries in  **$\mathbf{P(Y | X)}$**  sum to 1.

Your doctor tells you that you have tested positive for a serious but rare ( $1/10000$ ) disease. This test (T) is correct to 99% (1% false positive & 1% false negative results). What does this mean for you?

$$P(D | T) = \frac{P(T | D) P(D)}{P(T)} \times \frac{P(T|D) P(D)}{P(T|D) P(D) + P(T|\neg D) P(\neg D)}$$

$$P(D) = 0.0001 \quad P(T | D) = 0.99 \quad P(T | \neg D) = 0.01$$

$$P(D | T) = \frac{0.99 \times 0.0001}{0.99 \times 0.0001 + 0.01 \times 0.9999} \times \frac{0.000099}{0.000099 + 0.009999} = \frac{0.000099}{0.010088} \approx 0.01$$

**Moral:** If the test imprecision is much greater than the rate of occurrence of the disease, then a positive result is not as threatening as you might think.

A common model in early diagnosis:

Symptoms are conditionally independent given the disease (or fault)

Thus, if

$X_1, \dots, X_n$  denote whether the symptoms exhibited by the patient (headache, high-fever, etc.) and

$H$  denotes the hypothesis about the patients health

then,  $P(X_1, \dots, X_n, H) = P(H)P(X_1|H) \dots P(X_n|H)$ ,

This **naïve Bayesian** model allows compact representation

It does embody strong independence assumptions

## Bayes' Rule and conditional independence

$$\begin{aligned} \mathbf{P}(\text{Cavity} \mid \text{toothache} \wedge \text{catch}) \\ &= \alpha \mathbf{P}(\text{toothache} \wedge \text{catch} \mid \text{Cavity}) \mathbf{P}(\text{Cavity}) \\ &= \alpha \mathbf{P}(\text{toothache} \mid \text{Cavity}) \mathbf{P}(\text{catch} \mid \text{Cavity}) \mathbf{P}(\text{Cavity}) \end{aligned}$$

This is an example of a **naïve Bayes** model:

$$\mathbf{P}(\text{Cause}, \text{Effect}_1, \dots, \text{Effect}_n) = \mathbf{P}(\text{Cause}) \prod_i \mathbf{P}(\text{Effect}_i \mid \text{Cause})$$



Total number of parameters is **linear** in  $n$

Product rule  $P(a \wedge b) = P(a \mid b) P(b) = P(b \mid a) P(a)$

The order is unimportant

$\Rightarrow$  **Bayes' rule:**  $P(a \mid b) = P(b \mid a) P(a) / P(b)$

or in distribution form

$$\mathbf{P}(Y|X) = \mathbf{P}(X|Y) \mathbf{P}(Y) / \mathbf{P}(X) = \alpha \mathbf{P}(X|Y) \mathbf{P}(Y)$$

Useful for assessing **diagnostic** probability from **causal** probability:

$$P(\text{Cause}|\text{Effect}) = P(\text{Effect}|\text{Cause}) P(\text{Cause}) / P(\text{Effect})$$

E.g., let  $M$  be meningitis,  $S$  be stiff neck:

$$P(m|s) = P(s|m) P(m) / P(s) = 0.8 \times 0.0001 / 0.1 = 0.0008$$

Note: posterior probability of meningitis still very small!



A common model in early diagnosis:

Symptoms are conditionally independent given the disease (or fault)

Thus, if


$X_1, \dots, X_n$  denote whether the symptoms exhibited by the patient (headache, high-fever, etc.) and

$H$  denotes the hypothesis about the patients health

then,  $P(X_1, \dots, X_n, H) = P(H)P(X_1|H) \dots P(X_n|H)$ ,

This **naïve Bayesian** model allows compact representation

It does embody strong independence assumptions



Three prisoners, *A, B, and C*, are locked in their cells. It is common knowledge that one of them will be executed the next day and the others pardoned. Only the governor knows which one will be executed. Prisoner A asks the guard a favor:

"Please ask the governor who will be executed, and then take a message to one of my friends B or C to let him know that he will be pardoned in the morning."

The guard agrees, and comes back later and tells A that he gave the pardon message to B. What are A's chances of being executed, given this information?

$Fx = \text{"x will be freed"}$

$Ex = \text{"x will be executed"}$

$$P(Ea \mid Fb) = (P(Fb \mid Ea) * P(Ea)) / P(Fb) = (1 * 1/3) / 2/3 = 1/2 !$$

$F'b = \text{"The guard said that Fb"}$

$$P(Ea \mid F'b) = (P(F'b \mid Ea) * P(Ea)) / P(F'b) = (1/2 * 1/3) / 1/3 = 1/3$$

The guard has a choice of whom to inform in the case where A will be executed



All'inizio, è ovvio che:

$$P(A_1) = P(A_2) = P(A_3) = \frac{1}{3}$$

Supponiamo che la porta scelta è la numero 1.

**B** = evento "il presentatore apre la porta 3".  $P(B) = 0.50$

Nel caso in cui la macchina sia dietro la porta 1, il presentatore sarà libero di scegliere la porta 2 o 3 casualmente.

$$\text{Pertanto, } P(B | A_1) = 1 / 2$$

Nel caso in cui la macchina sia dietro la porta 2, il presentatore sarà obbligato ad aprire la porta 3.

$$\text{Pertanto } P(B | A_2) = 1$$

Nel caso in cui la macchina sia dietro la porta 3, il presentatore sarà obbligato ad aprire la porta 2.

$$\text{Pertanto } P(B | A_3) = 0$$

$$P(A_1|B) = \frac{P(B|A_1)P(A_1)}{P(B)} = \frac{\frac{1}{2} \cdot \frac{1}{3}}{\frac{1}{2}} = \frac{1}{3}$$

$$P(A_2|B) = \frac{P(B|A_2)P(A_2)}{P(B)} = \frac{1 \cdot \frac{1}{3}}{\frac{1}{2}} = \frac{2}{3}$$

$$P(A_3|B) = \frac{P(B|A_3)P(A_3)}{P(B)} = \frac{0 \cdot \frac{1}{3}}{\frac{1}{2}} = 0.$$



# Bayesian networks

*(R&N: 14.1, 14.2)*

## Purpose of Bayesian Networks

Facilitate the description of a collection of beliefs by making explicit causality relations and conditional independence among beliefs

Provide a more efficient way (than by using joint distribution tables) to update belief strengths when new evidence is observed



Belief networks

Probabilistic networks

Causal networks



A simple, graphical notation for conditional independence assertions and hence for compact specification of full joint distributions

Syntax:

- a set of nodes, one per variable

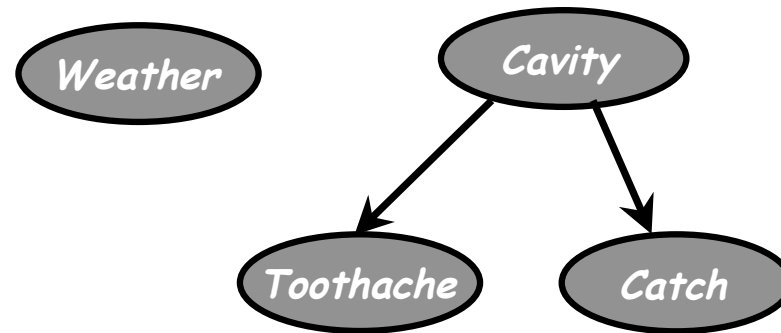
- a directed, acyclic graph (link  $\approx$  "directly influences")

- a conditional distribution for each node given its parents:

$$\mathbf{P}(X_i \mid \text{Parents}(X_i))$$

In the simplest case, conditional distribution represented as a **conditional probability table** (CPT) giving the distribution over  $X_i$  for each combination of parent values

Topology of network encodes conditional independence assertions:



Weather is independent of other variables  
Toothache and Catch are independent given Cavity

I'm at work, neighbor John calls to say my alarm is ringing, but neighbor Mary doesn't call. Sometime it's set off by a minor earthquake. Is there a burglar?

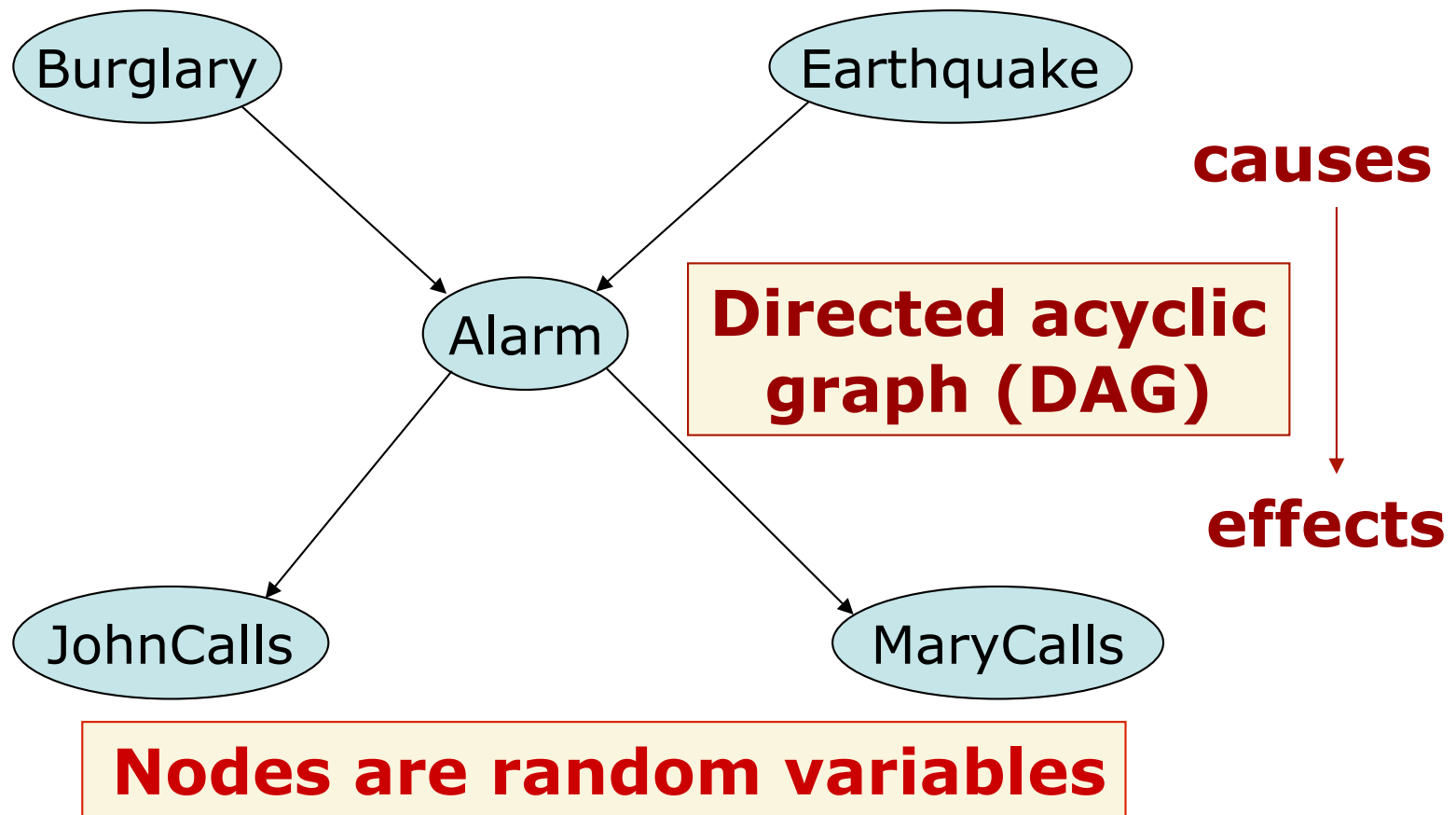
**Variables: Burglar, Earthquake, Alarm, JohnCalls, MaryCalls**

Network topology reflects "causal" knowledge:

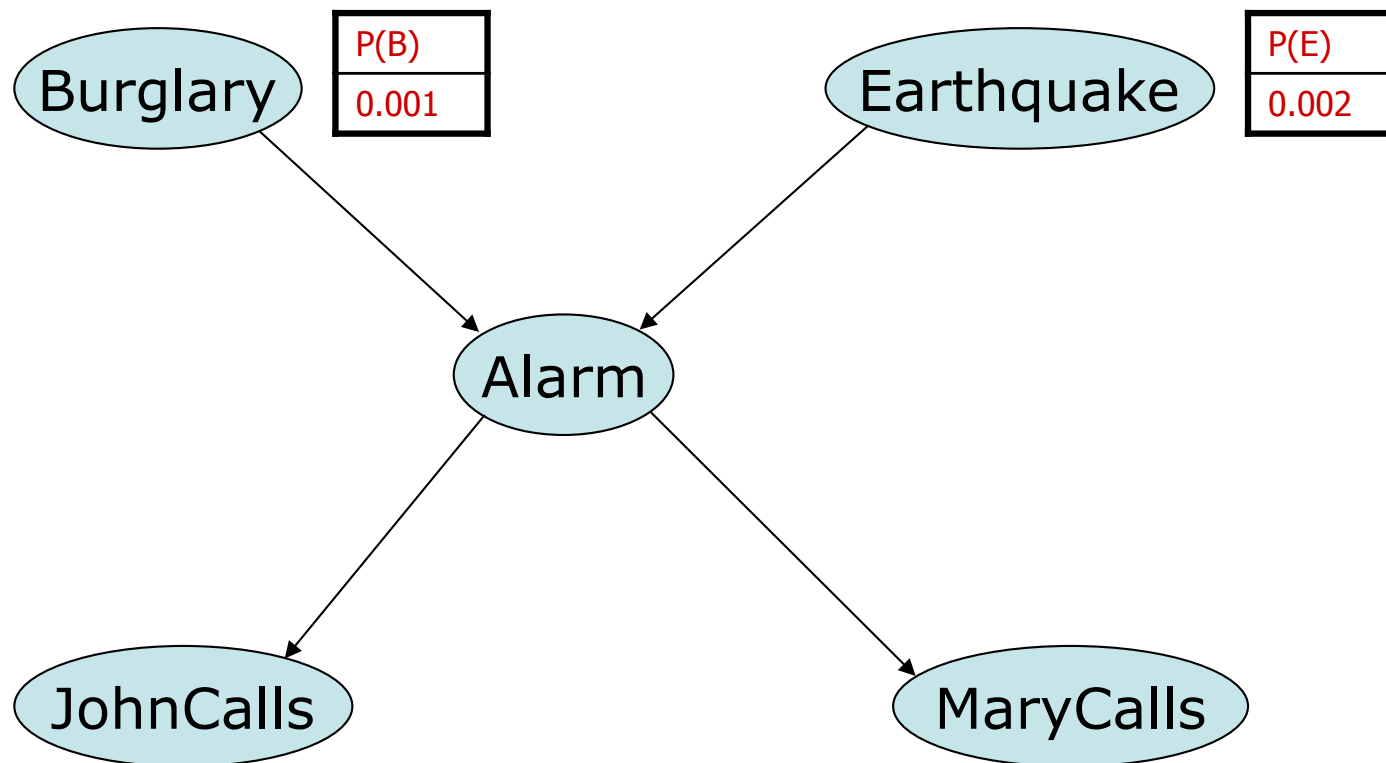
- A burglar can set the alarm off
- An earthquake can set the alarm off
- The alarm can cause Mary to call
- The alarm can cause John to call

# A Simple Belief Network

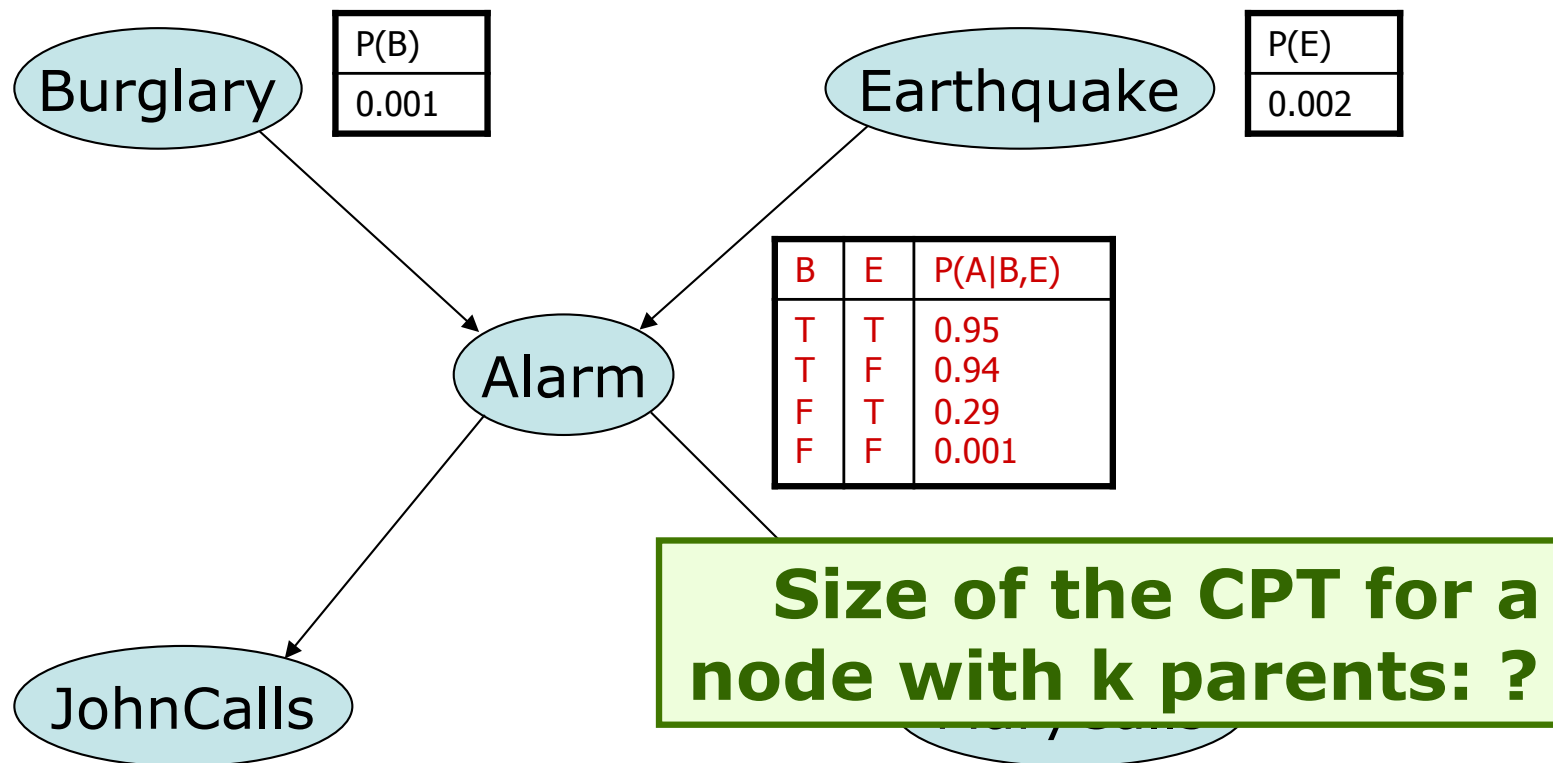
Intuitive meaning of arrow  
from  $x$  to  $y$ : “ $x$  has direct  
influence on  $y$ ”



# Assigning Probabilities to Roots

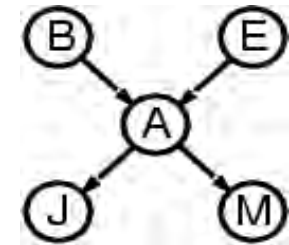


# Conditional Probability Tables



A CPT for Boolean  $X_i$  with  $k$  Boolean parents has  $2^k$  rows for the combinations of parent values

Each row requires one number  $p$  for  $X_i = \text{true}$  (the number for  $X_i = \text{false}$  is just  $1-p$ )

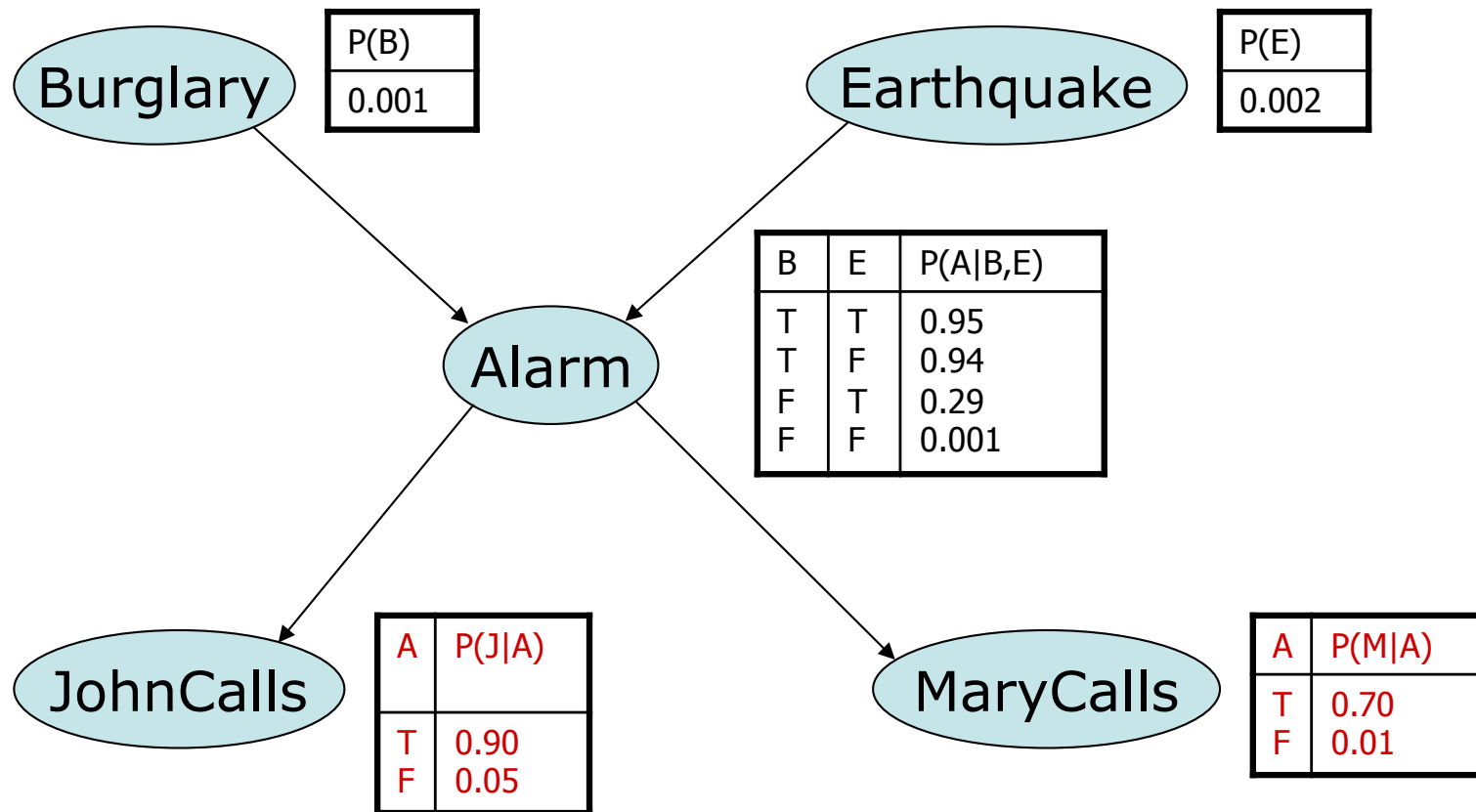


If each variable has no more than  $k$  parents, the complete network requires  $O(n \cdot 2^k)$  numbers

I.e., grows linearly with  $n$ , vs.  $O(2^n)$  for the full joint distribution

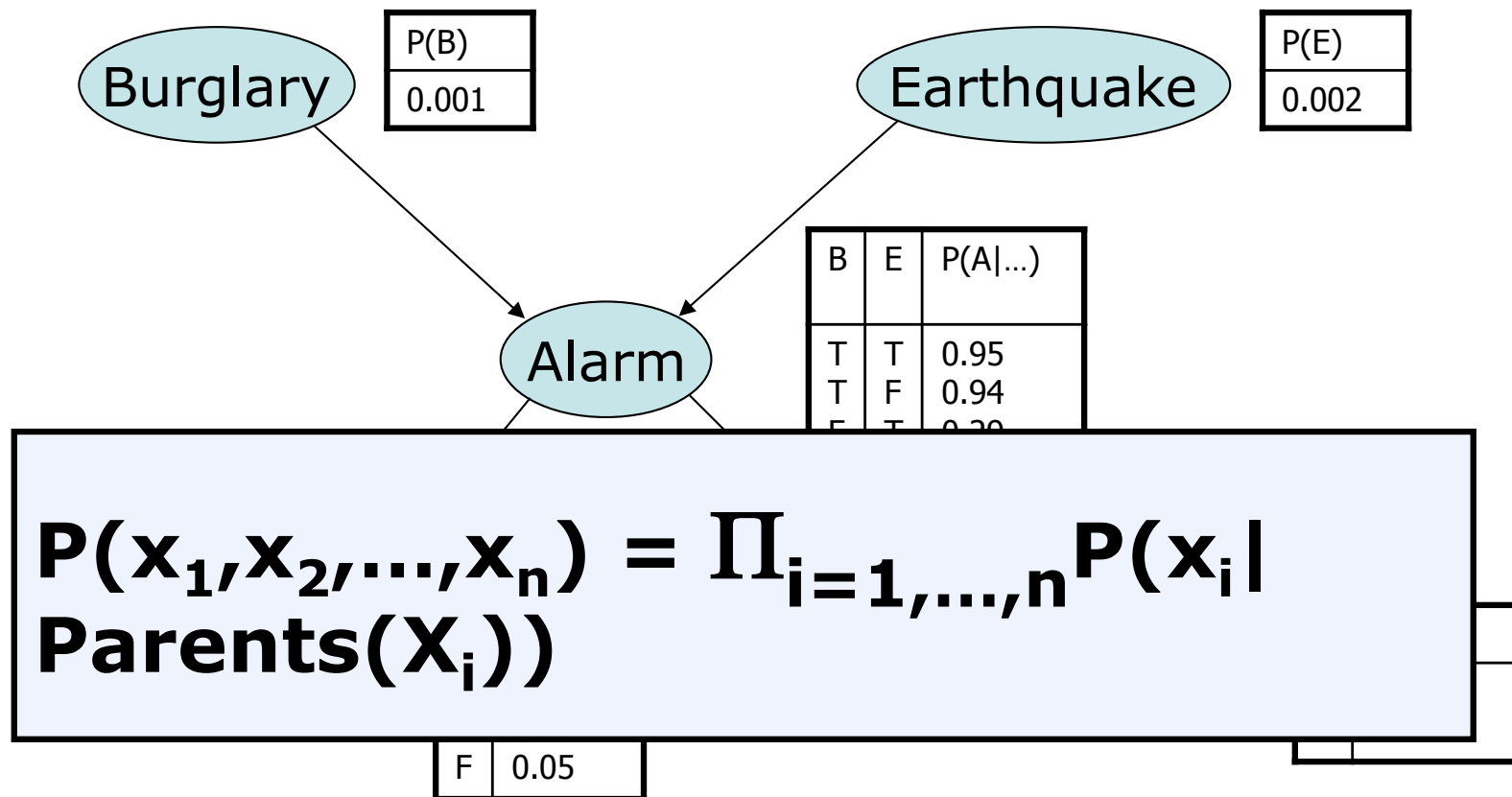
For burglary net,  $1 + 1 + 4 + 2 + 2 = 10$  numbers (vs.  $2^5 - 1 = 31$ )

# Conditional Probability Tables



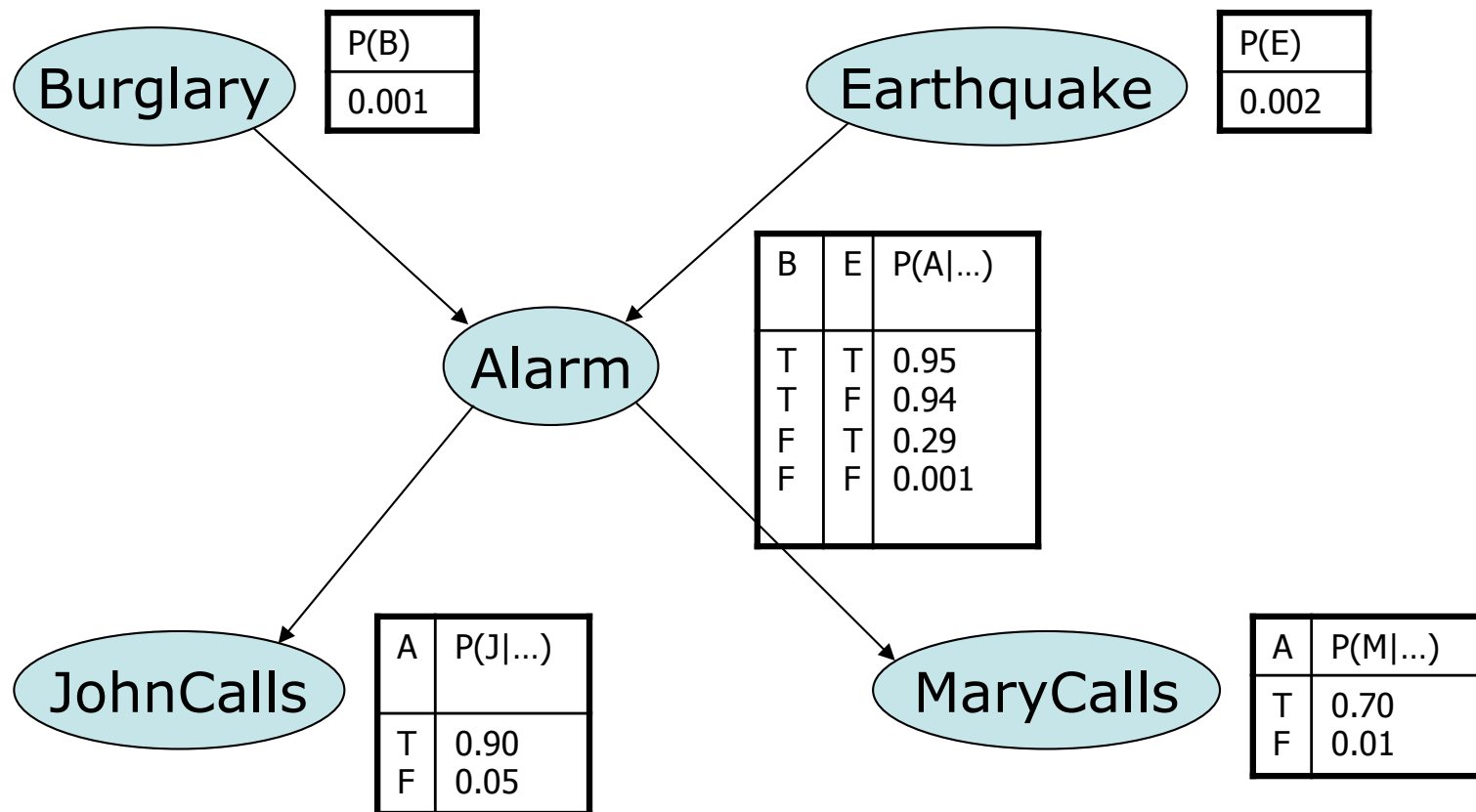


# What the BN Means



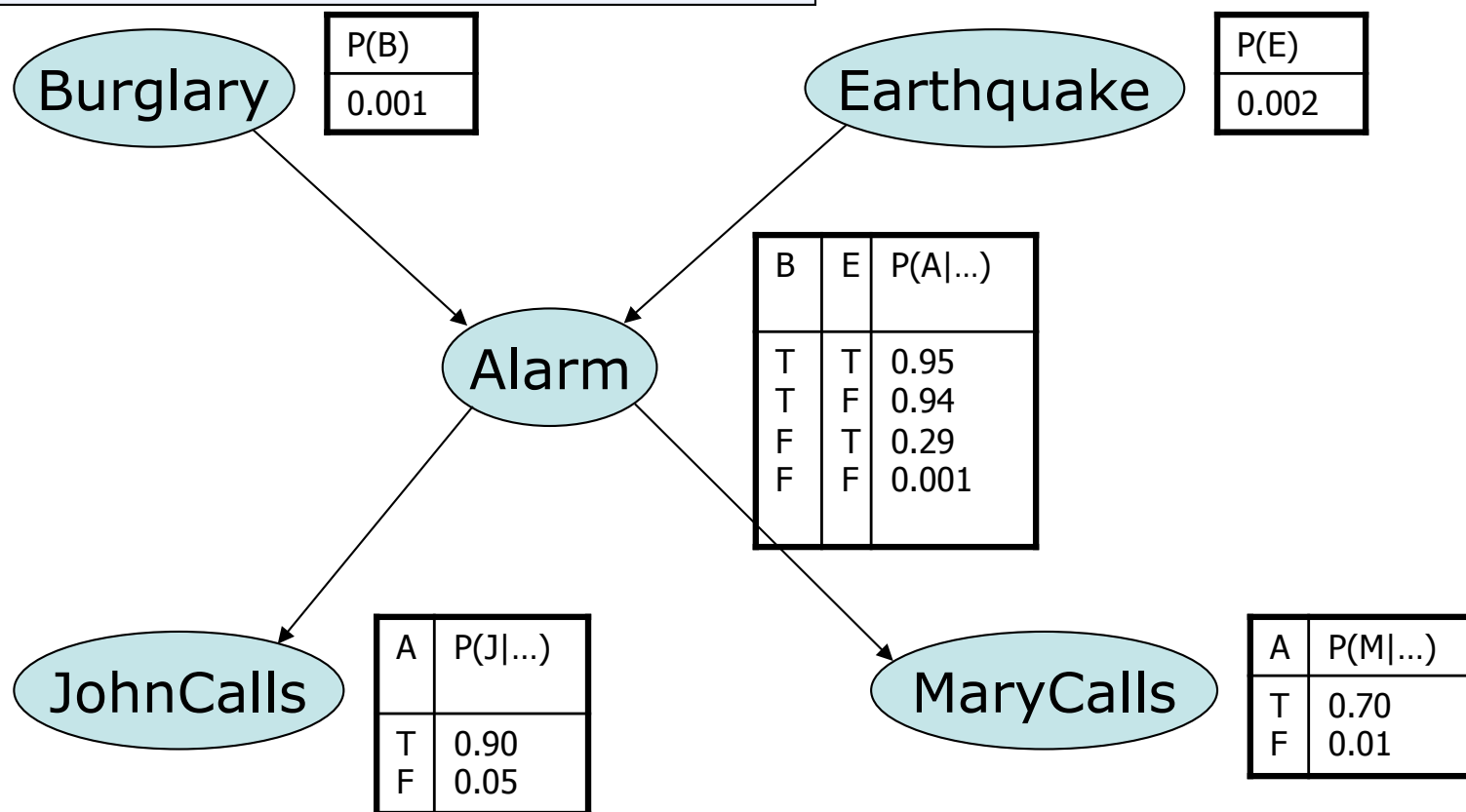
# Calculation of Joint Probability

$P(J \wedge M \wedge A \wedge \neg B \wedge \neg E)?$

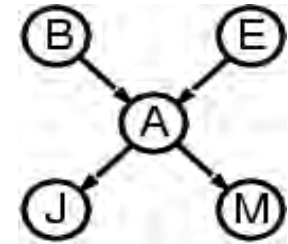


# Calculation of Joint Probability

$$\begin{aligned} &P(J \wedge M \wedge A \wedge \neg B \wedge \neg E) \\ &= P(J|A)P(M|A)P(A|\neg B, \neg E)P(\neg B)P(\neg E) \\ &= 0.9 \times 0.7 \times 0.001 \times 0.999 \times 0.998 \\ &= 0.00062 \end{aligned}$$



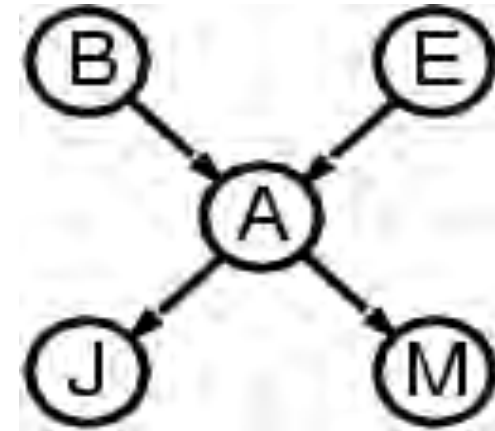
# Calculation of Joint Probability



- $P(J \wedge M \wedge A \wedge \neg B \wedge \neg E)$   
 $= P(J \wedge M | A, \neg B, \neg E) \times P(A \wedge \neg B \wedge \neg E)$   
 $= P(J | A, \neg B, \neg E) \times P(M | A, \neg B, \neg E) \times P(A \wedge \neg B \wedge \neg E)$   
(J and M are independent given A)
- $P(J | A, \neg B, \neg E) = P(J | A)$   
(J and  $\neg B \wedge \neg E$  are independent given A)
- $P(M | A, \neg B, \neg E) = P(M | A)$
- $P(A \wedge \neg B \wedge \neg E) = P(A | \neg B, \neg E) \times P(\neg B | \neg E) \times P(\neg E)$   
 $= P(A | \neg B, \neg E) \times P(\neg B) \times P(\neg E)$   
( $\neg B$  and  $\neg E$  are independent)
- $P(J \wedge M \wedge A \wedge \neg B \wedge \neg E) = P(J | A)P(M | A)P(A | \neg B, \neg E)P(\neg B)P(\neg E)$

# Calculation of Joint Probability

- $P(J \wedge M \wedge A \wedge \neg B \wedge \neg E) = P(J|A)P(M|A)P(A|\neg B, \neg E)P(\neg B)P(\neg E)$

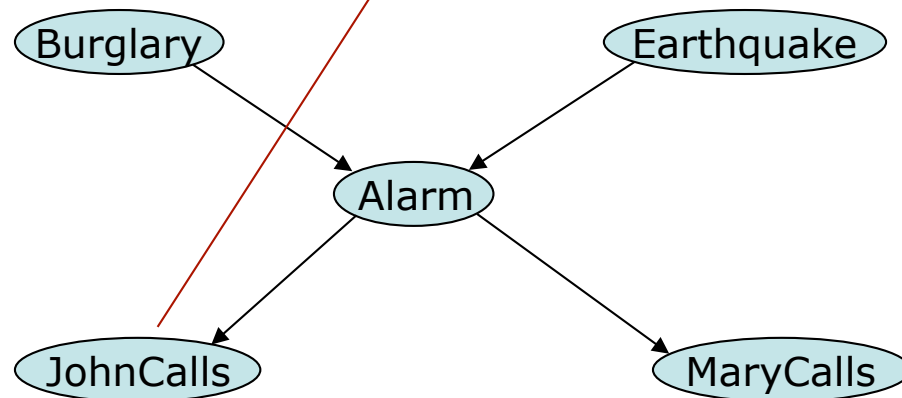


$$P(X_1, \dots, X_n) = \prod_{i=1}^n P(x_i \mid \text{parents}(X_i))$$

## What The BN Encodes

Each of the beliefs JohnCalls and MaryCalls is independent of Burglary and Earthquake given Alarm or  $\neg$ Alarm

For example, John does not observe any burglaries directly



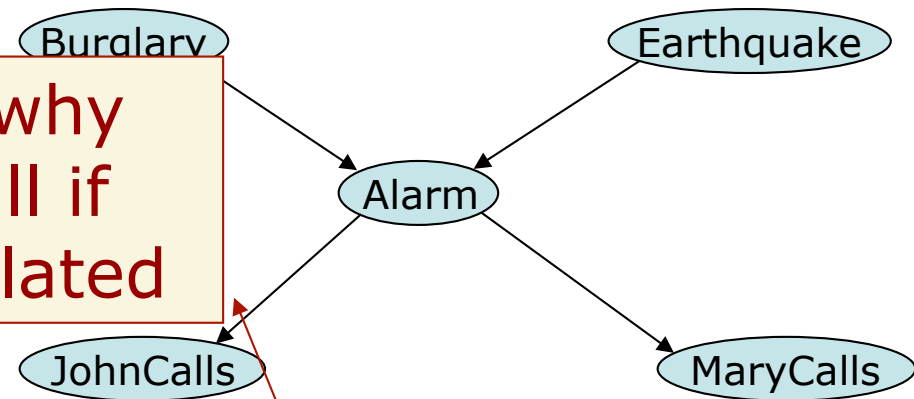
$$P(b \wedge j) \neq P(b) P(j)$$
$$P(b \wedge j | a) = P(b | a) P(j | a)$$

- The beliefs JohnCalls and MaryCalls are independent given Alarm or  $\neg$ Alarm

# What The BN Encodes

Each of the beliefs JohnCalls and MaryCalls is independent of Burglary and Earthquake given Alarm or  $\neg$ Alarm

For instance, the reasons why John and Mary may not call if there is an alarm are unrelated



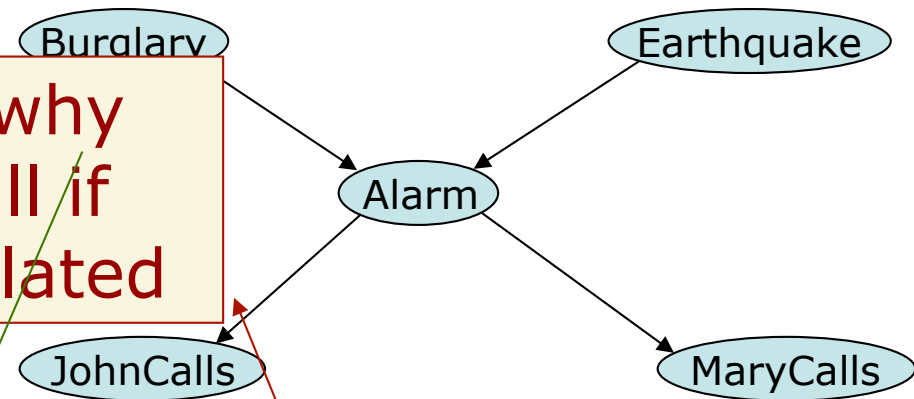
$$P(b \wedge j | a) = P(b | a) P(j | a)$$
$$P(j \wedge m | a) = P(j | a) P(m | a)$$

- The beliefs JohnCalls and MaryCalls are independent given Alarm or  $\neg$ Alarm

# What The BN Encodes

Each of the beliefs JohnCalls and MaryCalls is independent of Burglary and Earthquake given Alarm or  $\neg$ Alarm

For instance, the reasons why John and Mary may not call if there is an alarm are unrelated



Note that these reasons could be other beliefs in the network. The probabilities summarize these non-explicit beliefs

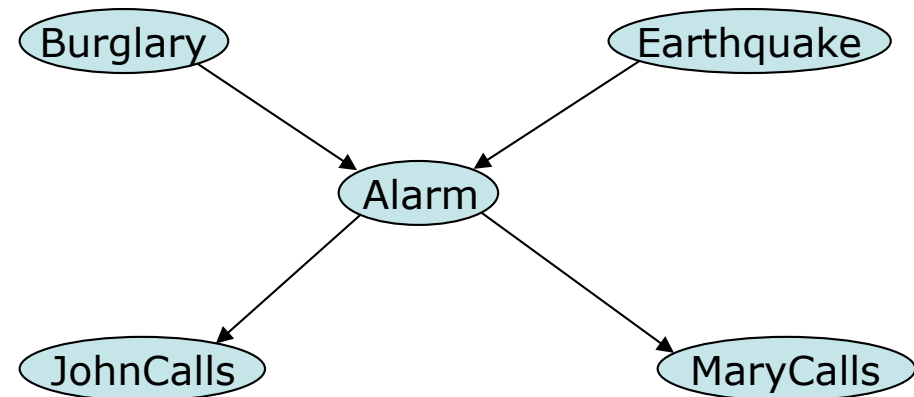
The beliefs JohnCalls and MaryCalls are independent given Alarm or  $\neg$ Alarm



## What The BN Encodes

Each of the beliefs JohnCalls and MaryCalls is independent of Burglary and Earthquake given Alarm or  $\neg$ Alarm

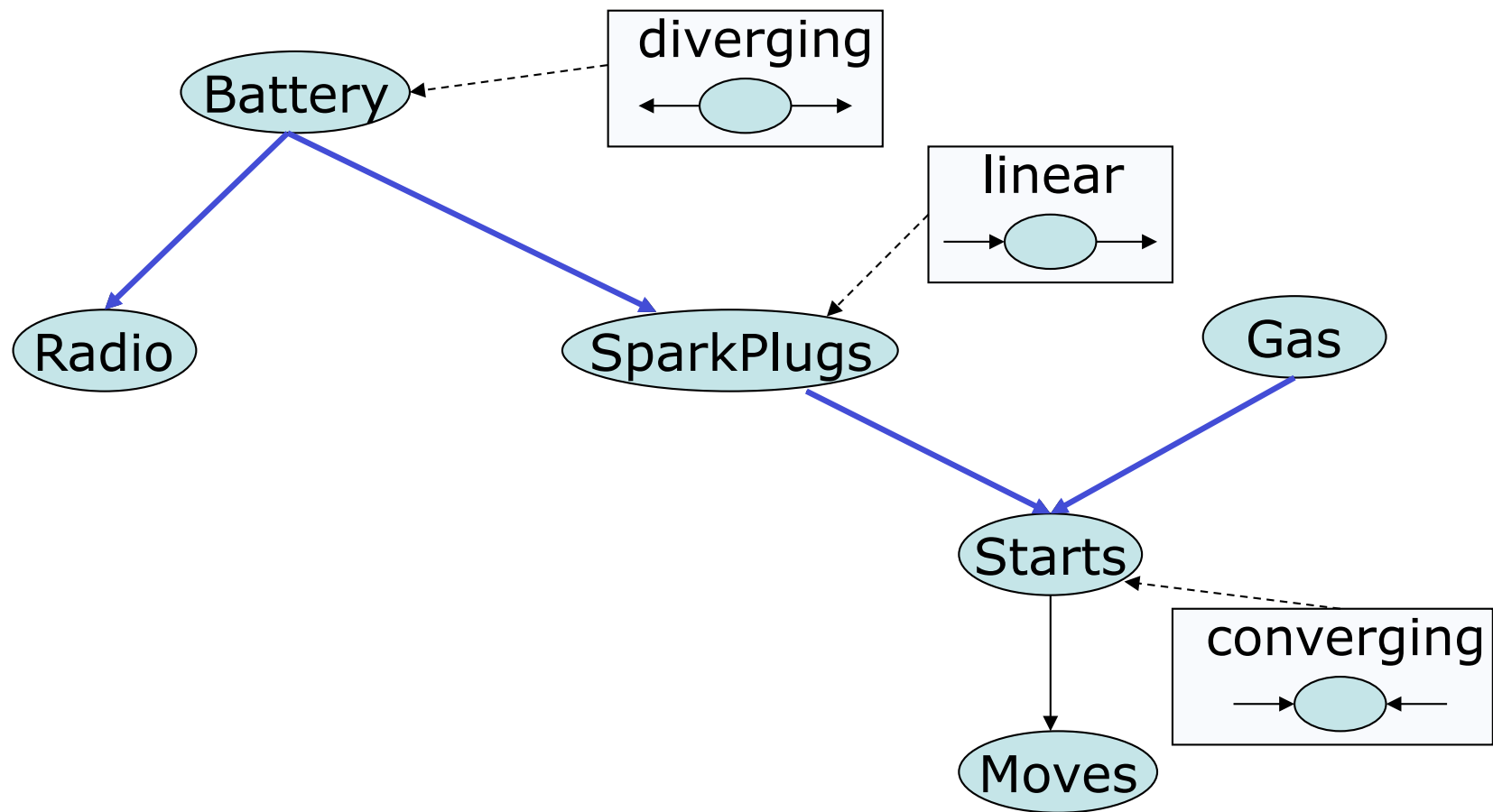
Burglary and Earthquake are independent



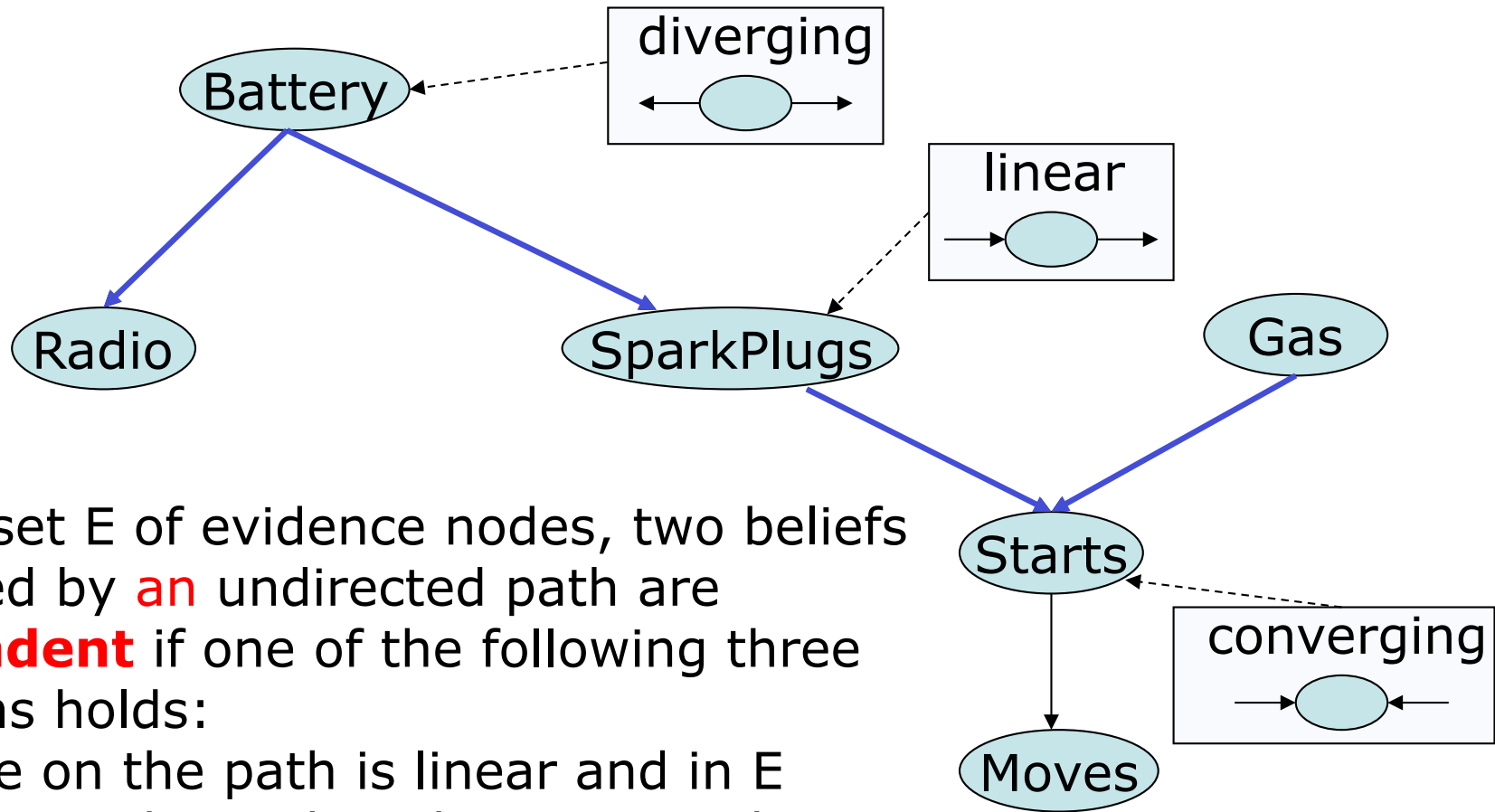
A node is independent of its non-descendants given its parents

- The beliefs JohnCalls and MaryCalls are independent given Alarm or  $\neg$ Alarm

# Types Of Nodes On A Path



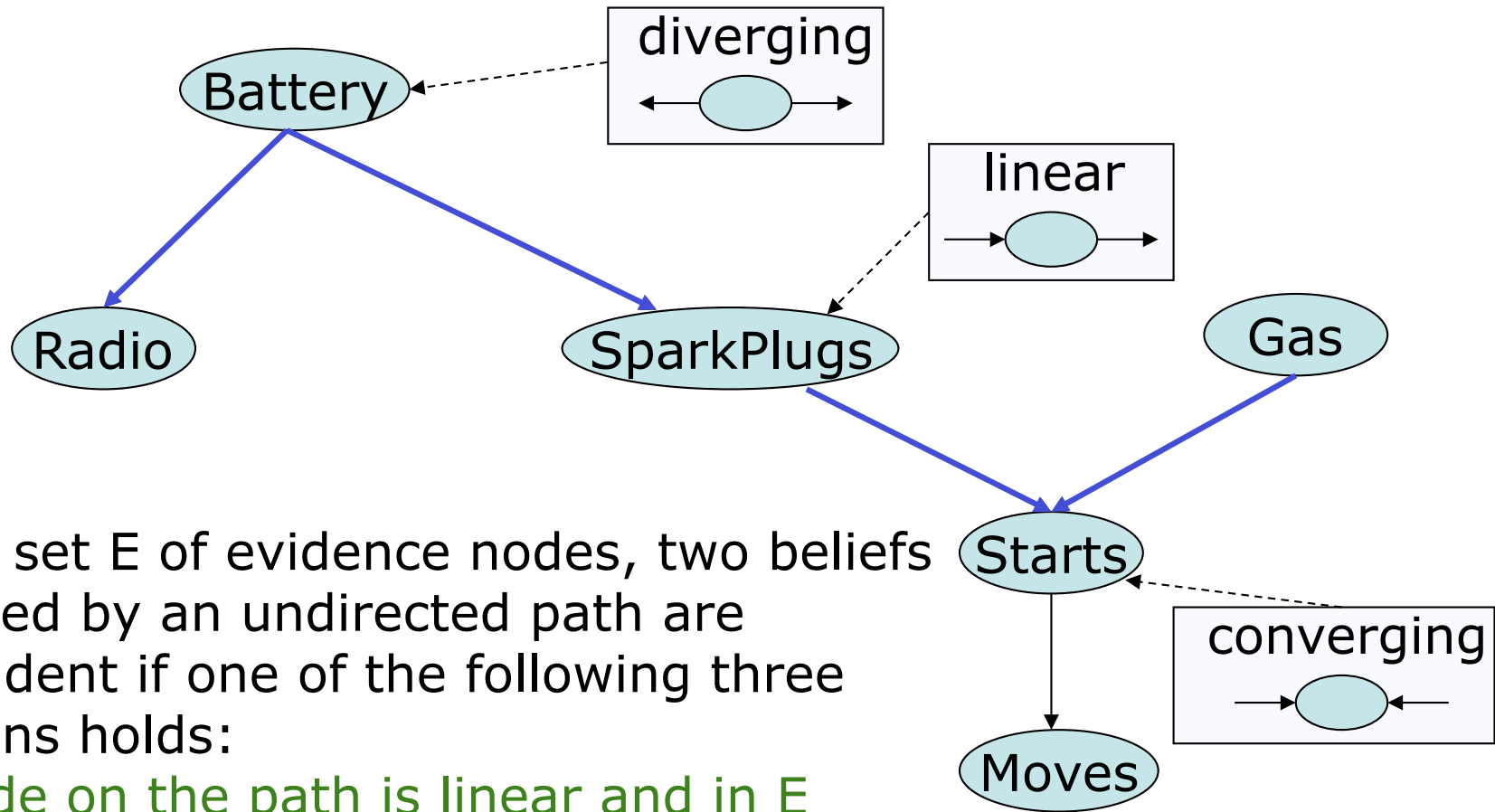
# Independence Relations in BN



Given a set  $E$  of evidence nodes, two beliefs connected by **an** undirected path are **independent** if one of the following three conditions holds:

1. A node on the path is linear and in  $E$
2. A node on the path is diverging and in  $E$
3. A node on the path is converging and neither this node, nor any descendant is in  $E$

# Independence Relations In BN



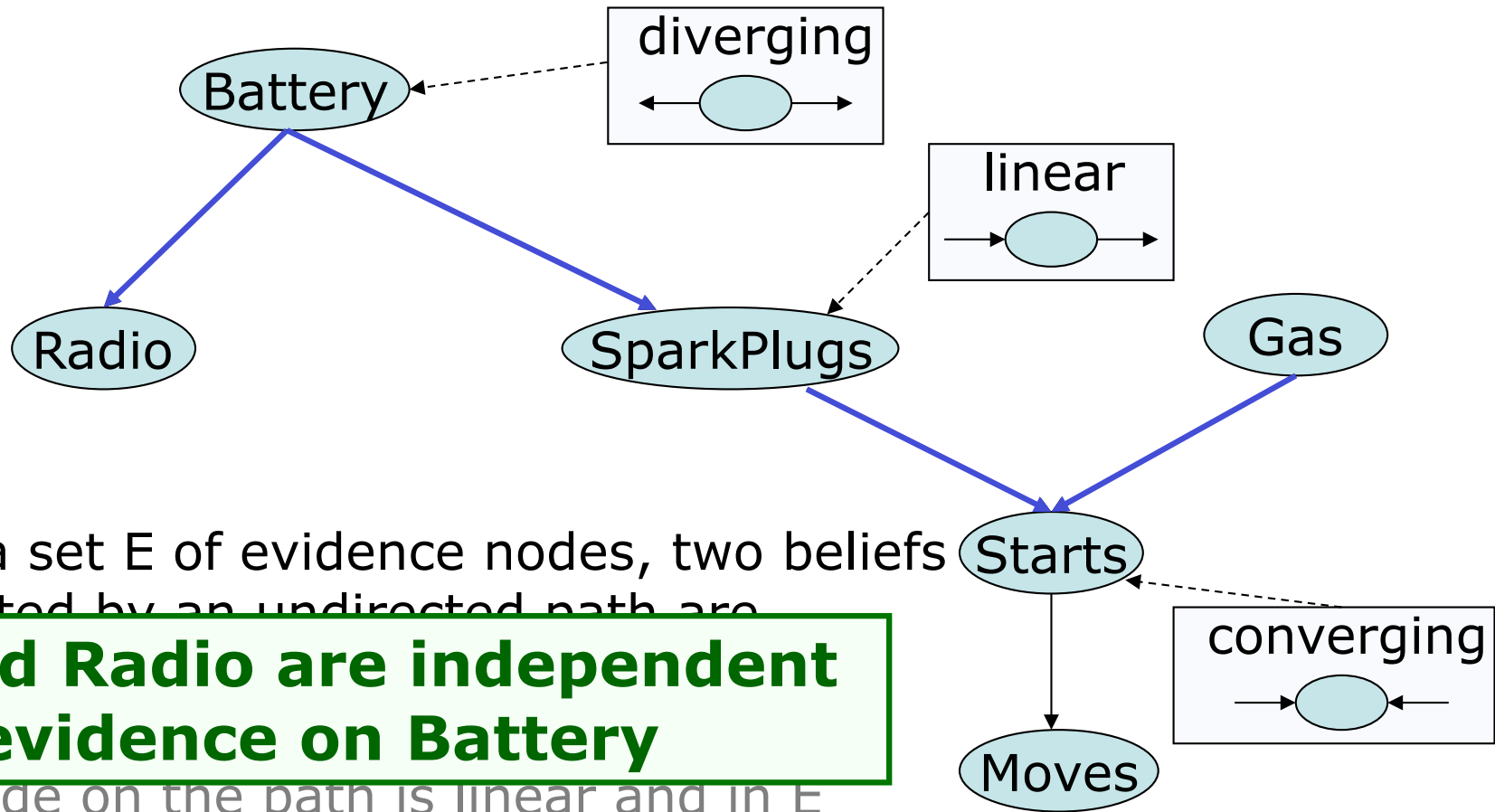
Given a set  $E$  of evidence nodes, two beliefs connected by an undirected path are independent if one of the following three conditions holds:

1. A node on the path is linear and in  $E$
2. A node on the path is diverging and in  $E$

**Gas and Radio are independent given evidence on SparkPlugs**

in  $E$

# Independence Relations In BN

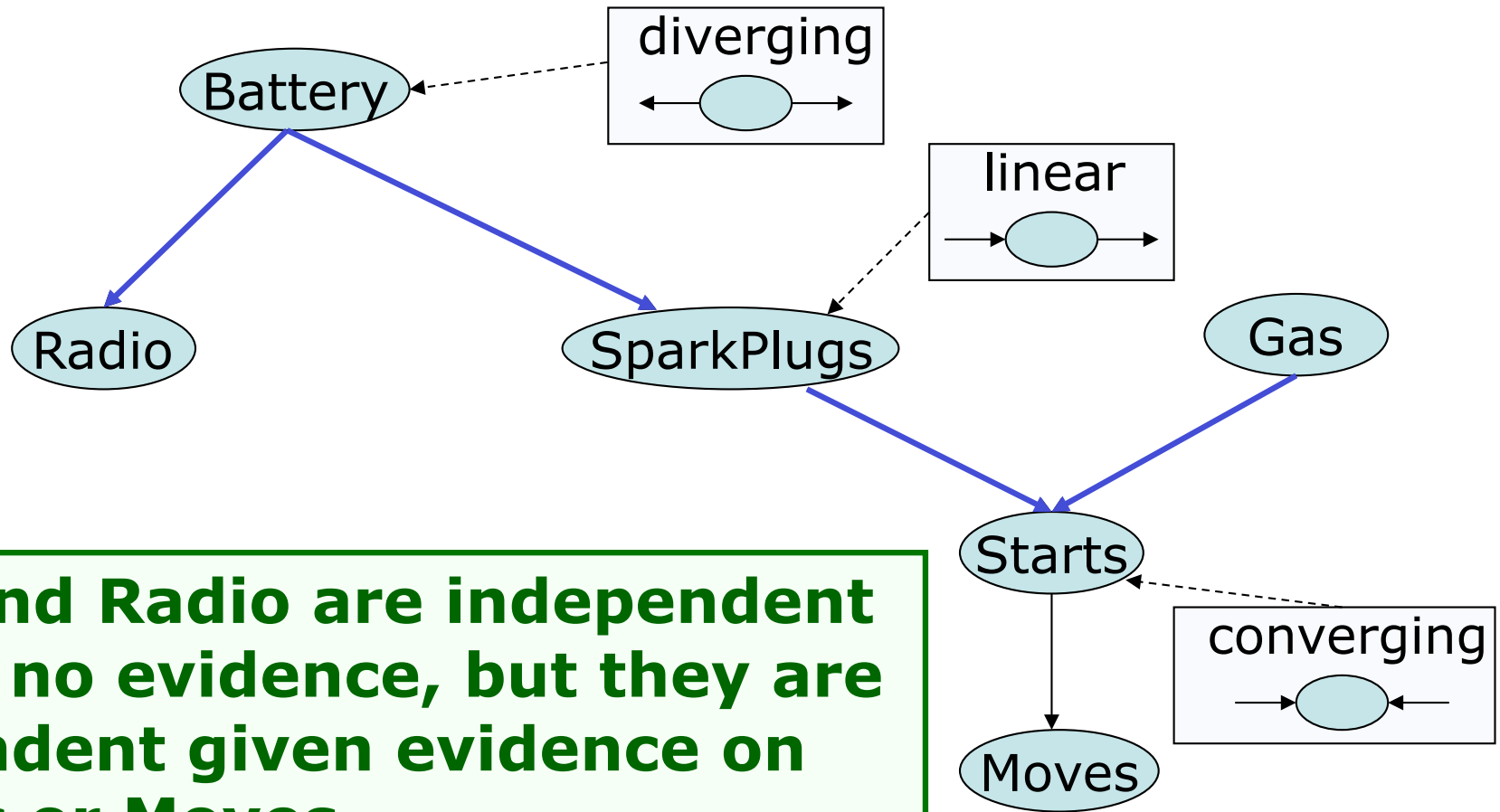


Given a set E of evidence nodes, two beliefs connected by an undirected path are

**Gas and Radio are independent given evidence on Battery**

1. A node on the path is linear and in E
2. A node on the path is diverging and in E
3. A node on the path is converging and neither this node, nor any descendant is in E

# Independence Relations In BN

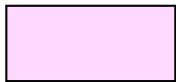


**Gas and Radio are independent given no evidence, but they are dependent given evidence on Starts or Moves**

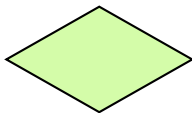
2. A node on the path is diverging and in E
3. A node on the path is converging and neither this node, nor any descendant is in E



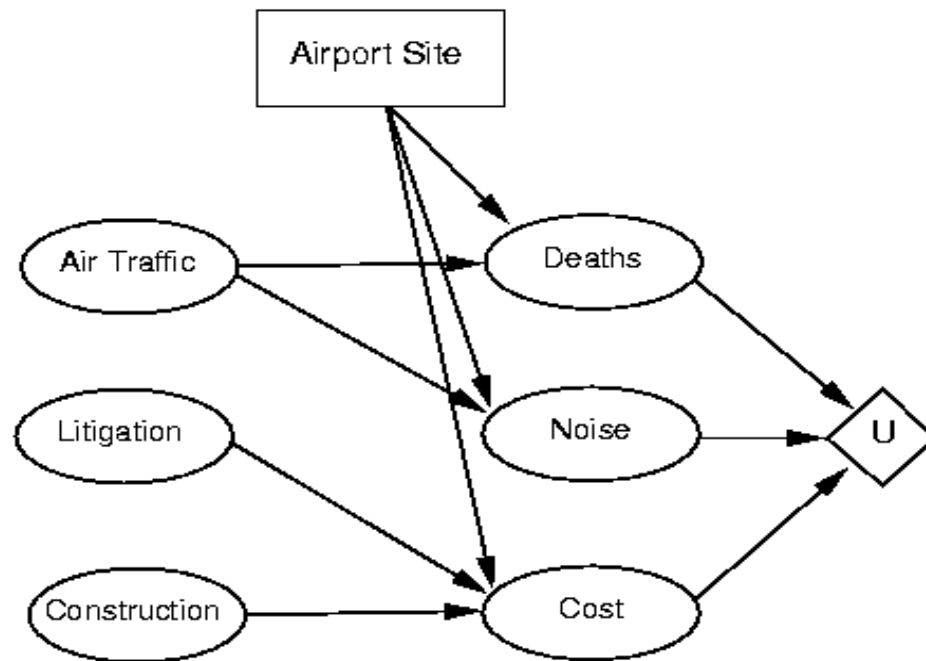
Chance nodes: random variables, as in BNs



Decision nodes: actions that decision maker can take

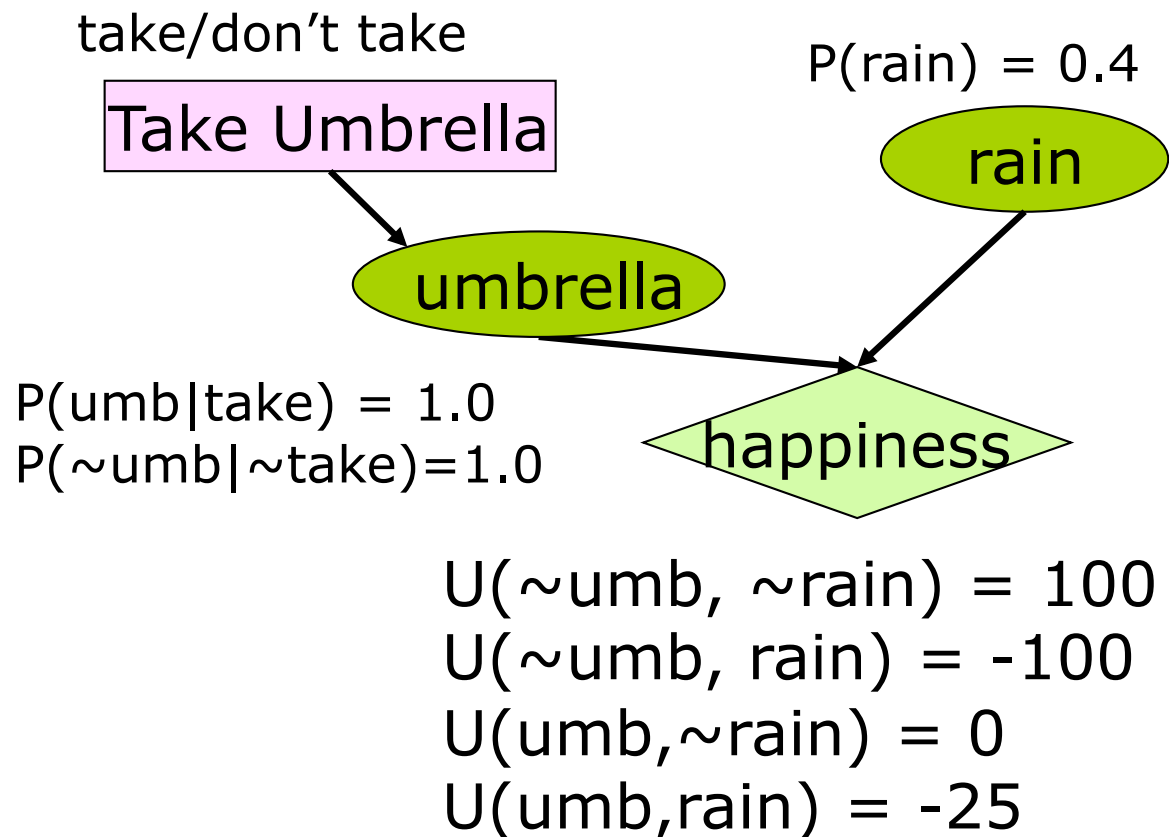


Utility/value nodes: the utility of the outcome state.





# Umbrella Network



Set the evidence variables for current state

For each possible value of the decision node:

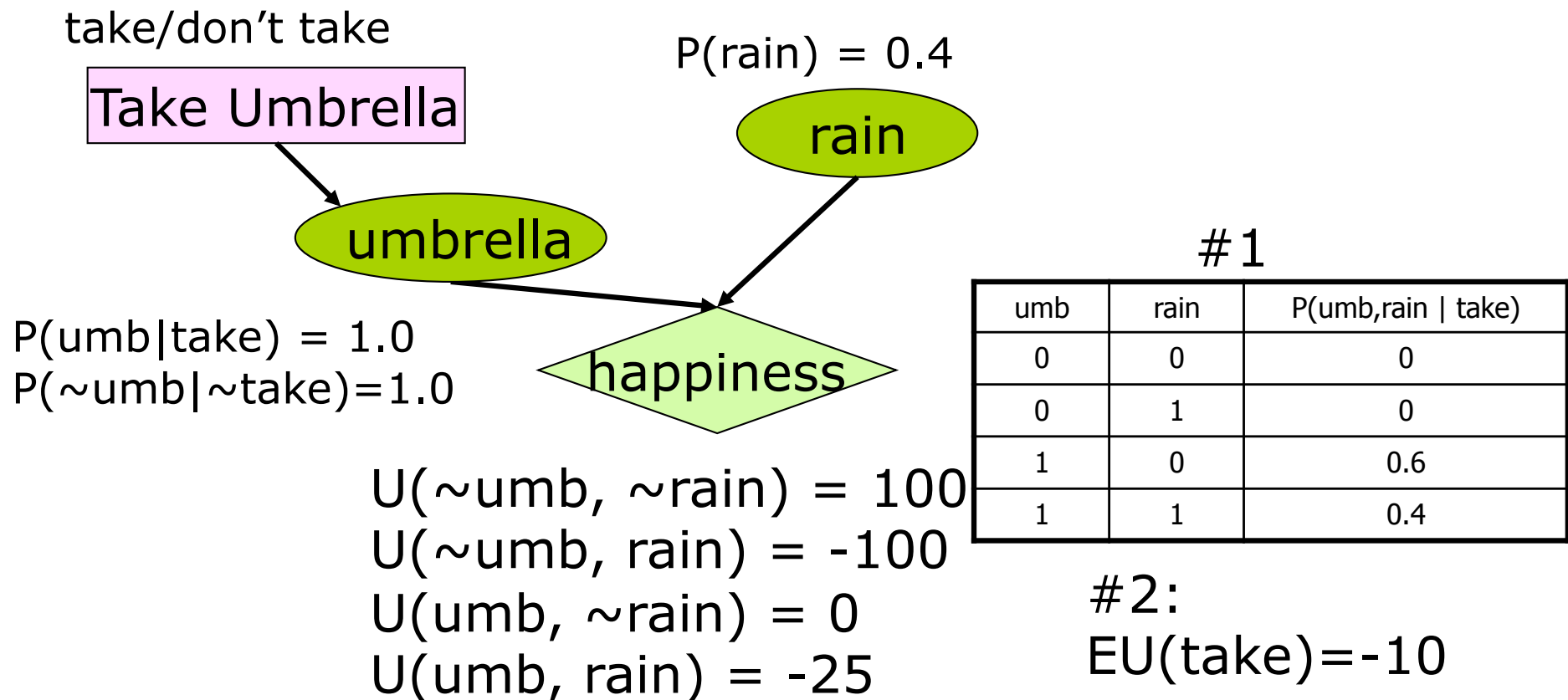
- Set decision node to that value

- Calculate the posterior probability of the parent nodes of the utility node, using BN inference

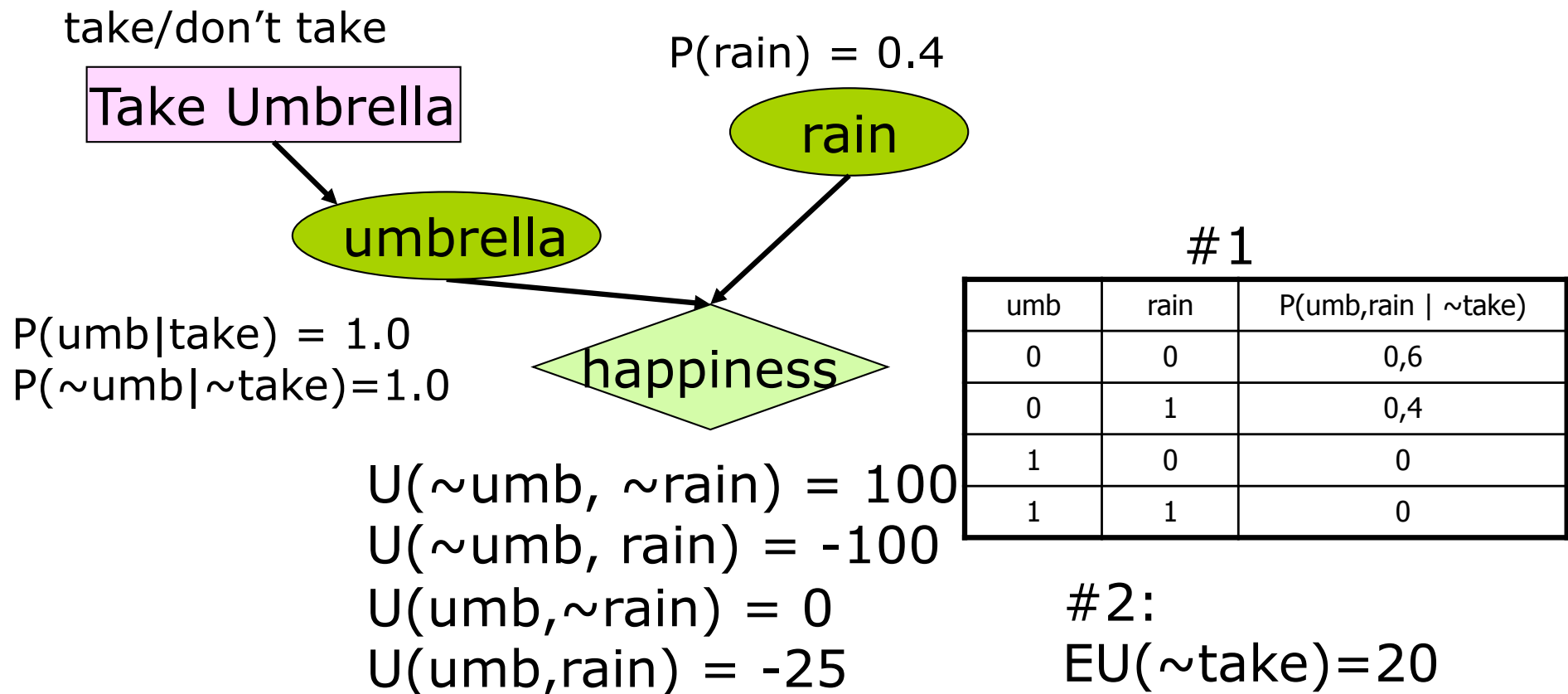
- Calculate the resulting utility for action

return the action with the highest utility

# Umbrella Network



# Umbrella Network



Suppose agent's current knowledge is  $E$ . The value of the current best action  $\alpha$  is

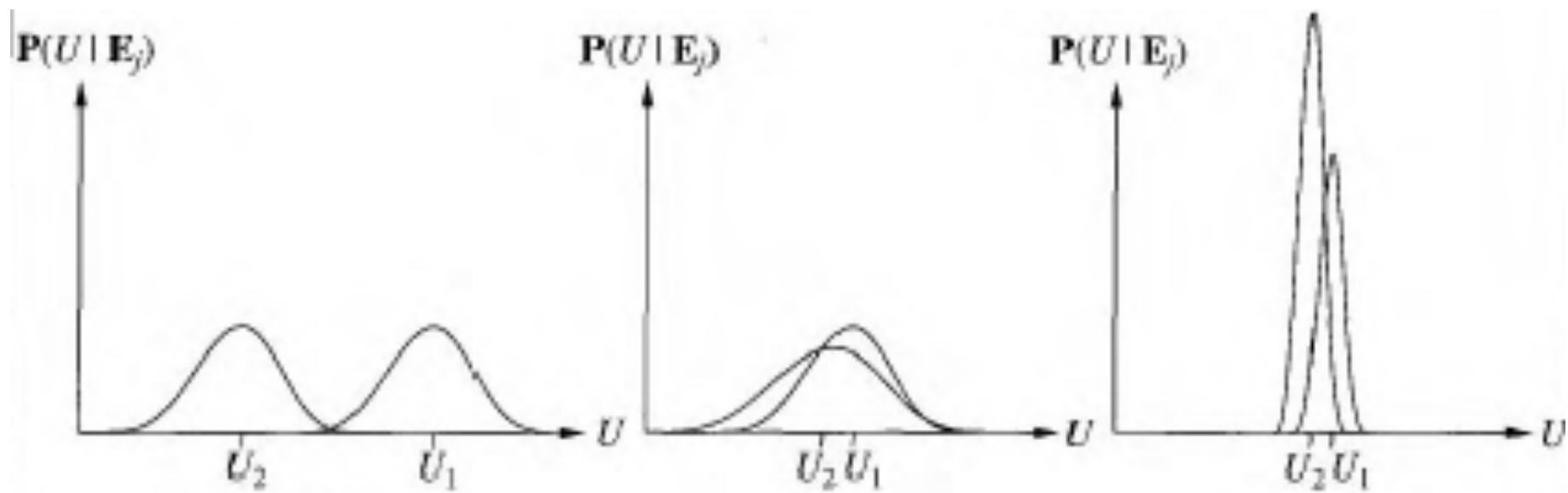
$$EU(\alpha | E) = \max_A \sum_i U(\text{Result}_i(A))P(\text{Result}_i(A) | E, \text{Do}(A))$$

the value of the new best action (after new evidence  $E'$  is obtained):

$$EU(\alpha' | E, E') = \max_A \sum_i U(\text{Result}_i(A))P(\text{Result}_i(A) | E, E', \text{Do}(A))$$

the value of information for  $E'$  is:

$$\text{VOI}(E') = \sum_k P(e_k | E) EU(\alpha_{ek} | e_k, E) - EU(\alpha | E)$$



$$\forall j, E \quad VPI_E(E_j) \geq 0$$

VPI is not additive

$$VPI_E(E_j, E_k) \neq VPI_E(E_j) + VPI_E(E_k)$$

VPI is order-independent

$$\begin{aligned} VPI_E(E_j, E_k) &= VPI_E(E_j) + VPI_{E, E_j}(E_k) \\ &= VPI_E(E_k) + VPI_{E, E_k}(E_j) \end{aligned}$$

**function INFORMATION-GATHERING-AGENT (percept)**  
**return an *action***

**static:  $D$ , a decision network**

integrate *percept* into  $D$

$j \leftarrow$  the value that maximizes  $VPI(E_j) - Cost(E_j)$

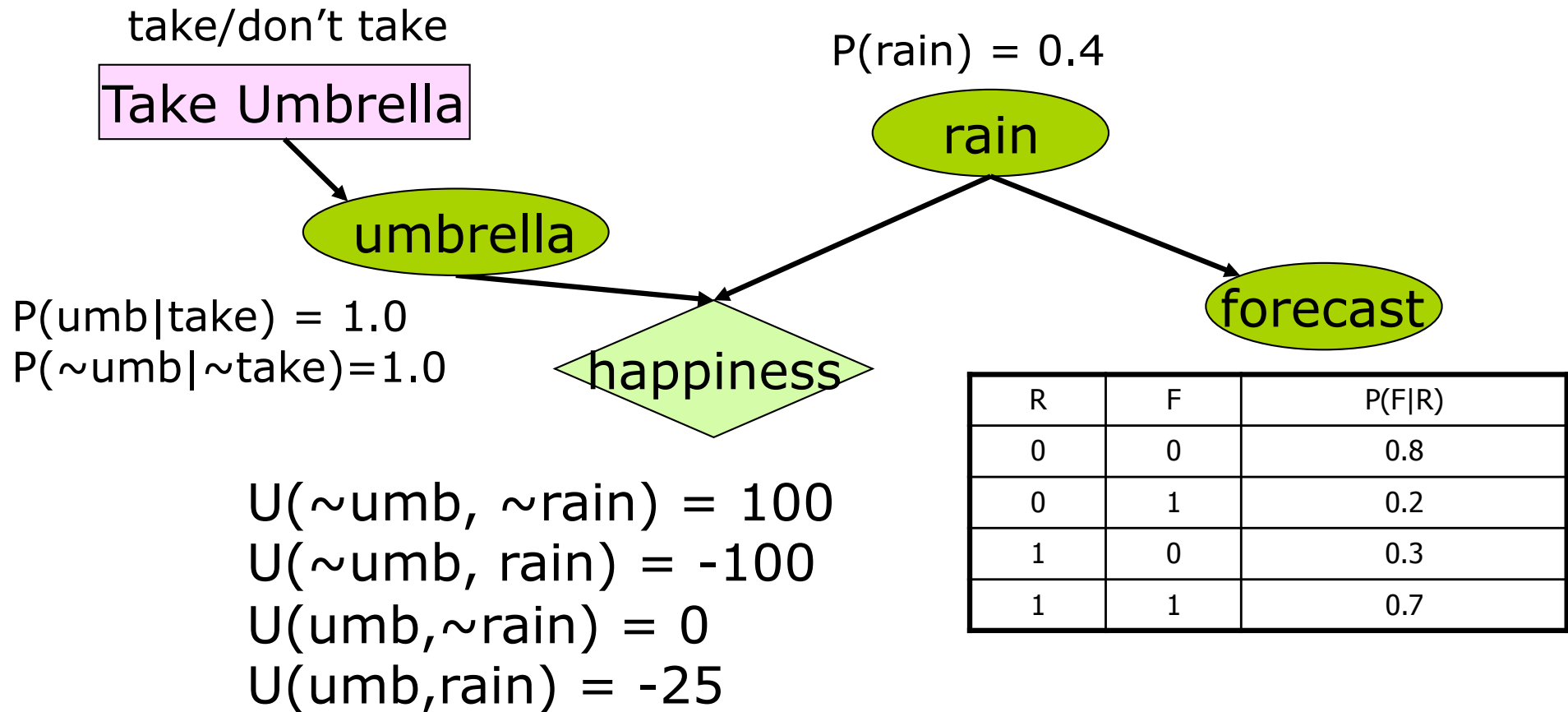
if  $VPI(E_j) > Cost(E_j)$

    then return  $REQUEST(E_j)$

else return the best action from  $D$



# Umbrella Network



VOI(forecast)=

$$P(\text{rainy})EU(\alpha_{\text{rainy}}) + \\ P(\sim\text{rainy})EU(\alpha_{\sim\text{rainy}}) - \\ EU(\alpha)$$

the value of information for  $E'$  is:

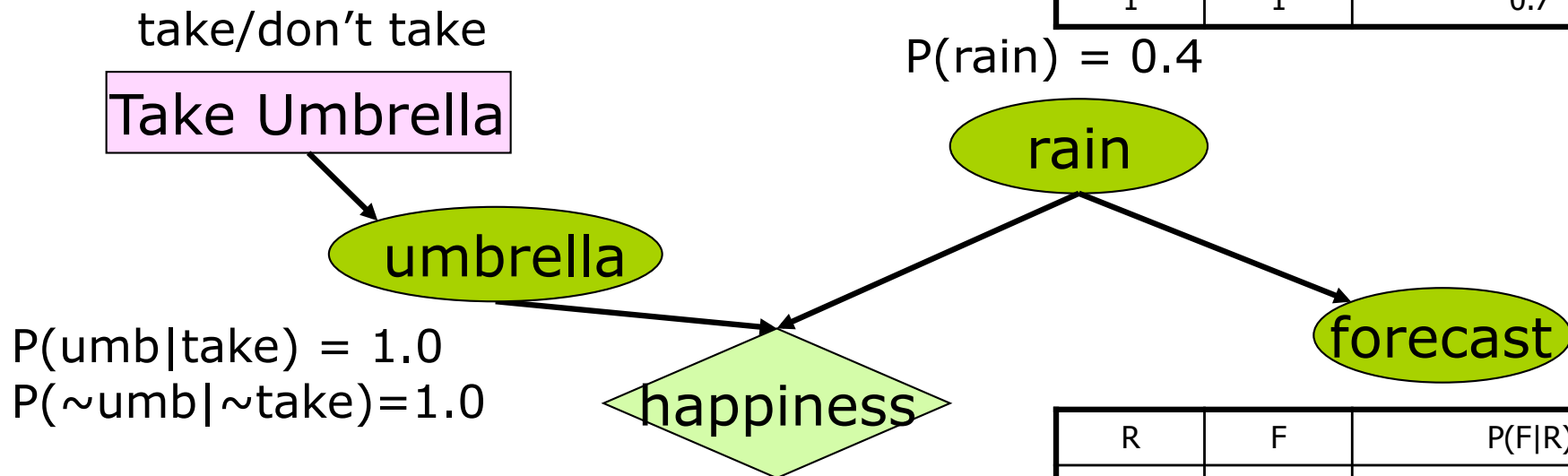
$$VOI(E') = \sum_k P(e_k | E)EU(\alpha_{e_k} | e_k, E) - EU(\alpha | E)$$

# Umbrella Network

$$P(R|F=\text{rainy}) = 0.4$$

F	R	P(R F)
0	0	0.8
0	1	0.2
1	0	0.3
1	1	0.7

$$P(\text{rain}) = 0.4$$



$$\begin{aligned}
 U(\sim\text{umb}, \sim\text{rain}) &= 100 \\
 U(\sim\text{umb}, \text{rain}) &= -100 \\
 U(\text{umb}, \sim\text{rain}) &= 0 \\
 U(\text{umb}, \text{rain}) &= -25
 \end{aligned}$$

R	F	P(F R)
0	0	0.8
0	1	0.2
1	0	0.3
1	1	0.7

umb	rain	P(umb,rain   take, rainy)
0	0	0
0	1	0
1	0	0.3
1	1	0.7

#1: EU(take|  
rainy) = -17,5  
**(-7+36= 29)**

umb	rain	P(umb,rain   ~take, rainy)
0	0	0.3
0	1	0.7
1	0	0
1	1	0


#2: EU(~take|  
rainy) = -40

umb	rain	P(umb,rain   take, ~rainy)
0	0	0
0	1	0
1	0	0.8
1	1	0.2

#3: EU(take|  
~rainy) = -5

umb	rain	P(umb,rain   ~take, ~rainy)
0	0	0.8
0	1	0.2
1	0	0
1	1	0

#4: EU(~take|  
~rainy) = 60

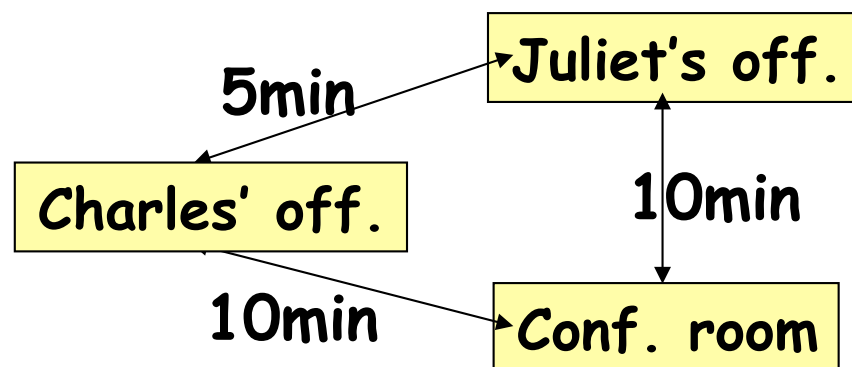


# **Decision-Making with Probabilistic Uncertainty**

*(R&N: 17.1, 17.2, 17.3)*

## A complete (but still very simple) example: Finding Juliet

- A robot, Romeo, is in Charles' office and must deliver a letter to Juliet
- Juliet is either in her office, or in the conference room. Each possibility has probability 0.5
- Traveling takes 5 minutes between Charles' and Juliet's office, 10 minutes between Charles' or Juliet's office and the conference room



- To perform his duties well and save battery, the robot wants to deliver the letter while minimizing the time spent in transit

## States and Actions in Finding-Juliet Problem

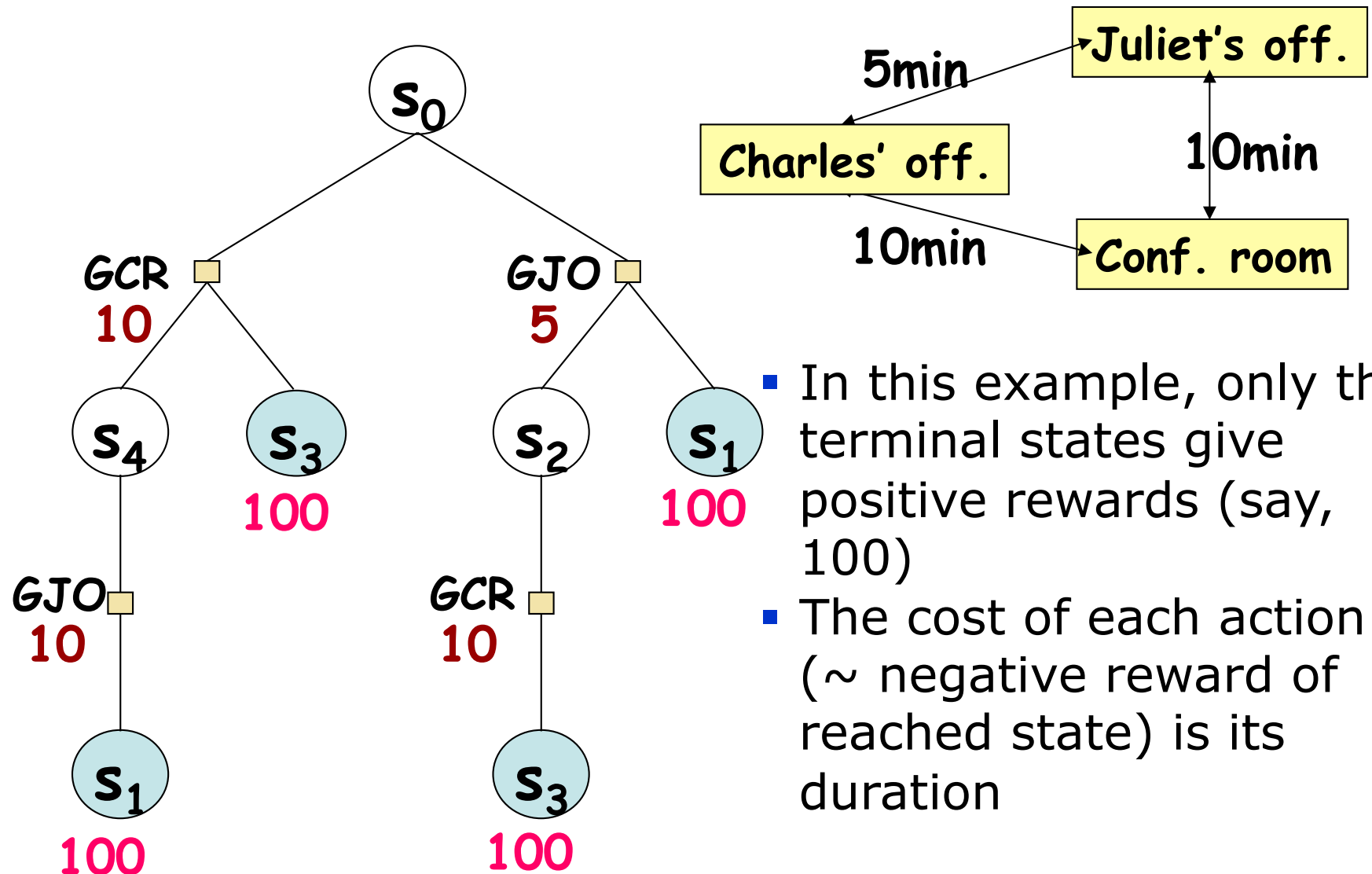
### ■ States:

- $S_0$ : Romeo in Charles' office
- $S_1$ : Romeo in Juliet's office and Juliet here
- $S_2$ : Romeo in Juliet's office and Juliet not here
- $S_3$ : Romeo in conference room and Juliet here
- $S_4$ : Romeo in conference room and Juliet not here
- In this example,  $S_1$  and  $S_3$  are **terminal** states

### ■ Actions:

- GJO (go to Juliet's office)
- GCR (go to conference room)
- The uncertainty in an action is directly linked to the uncertainty in Juliet's location

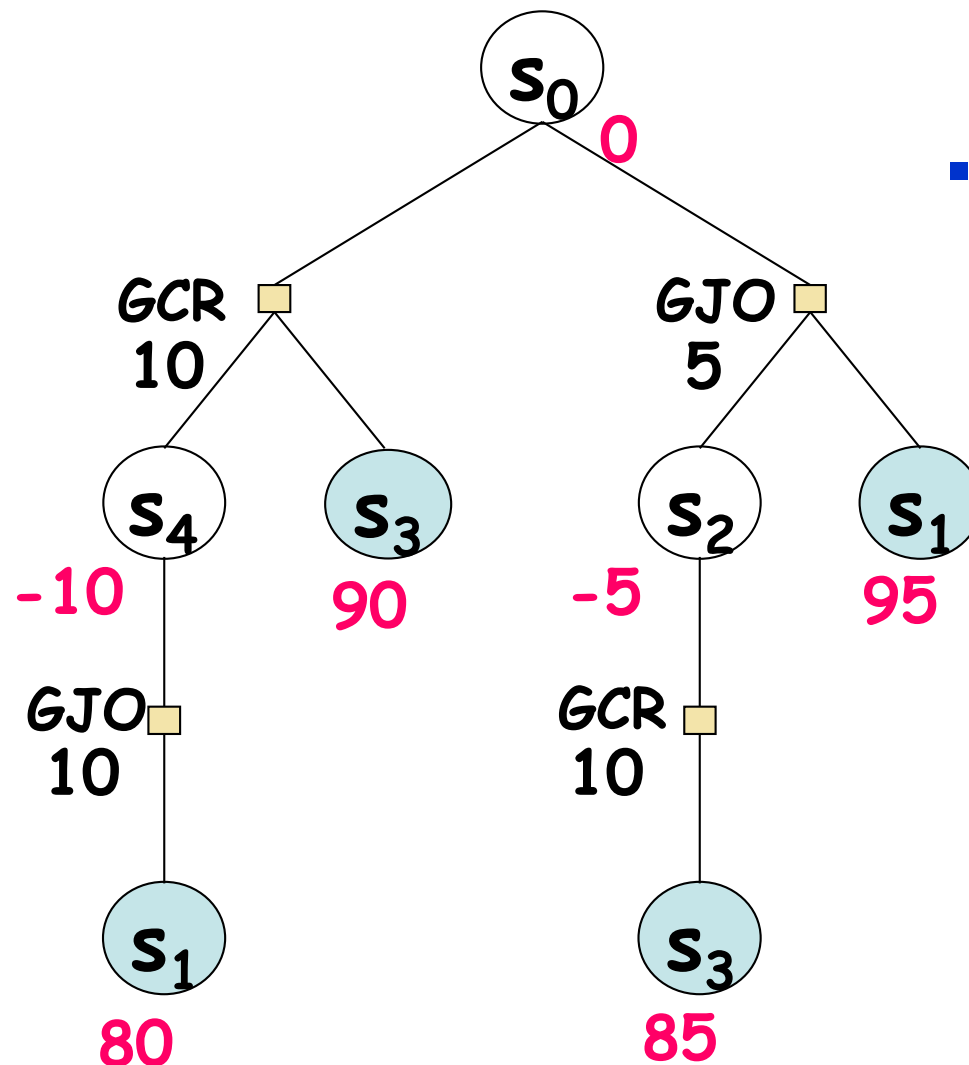
# State/Action Tree



- In this example, only the terminal states give positive rewards (say, 100)
- The cost of each action (~ negative reward of reached state) is its duration



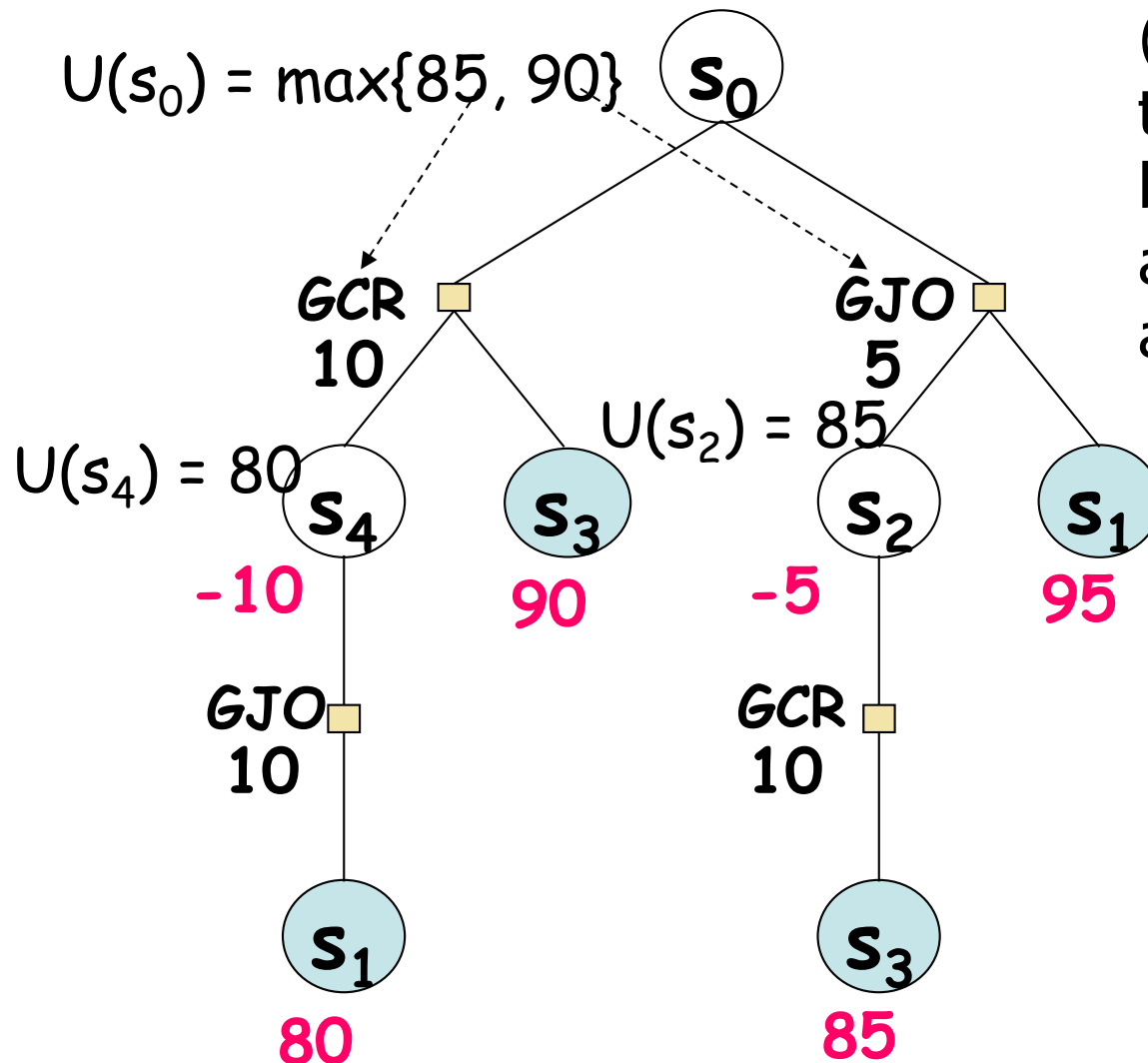
# State/Action Tree



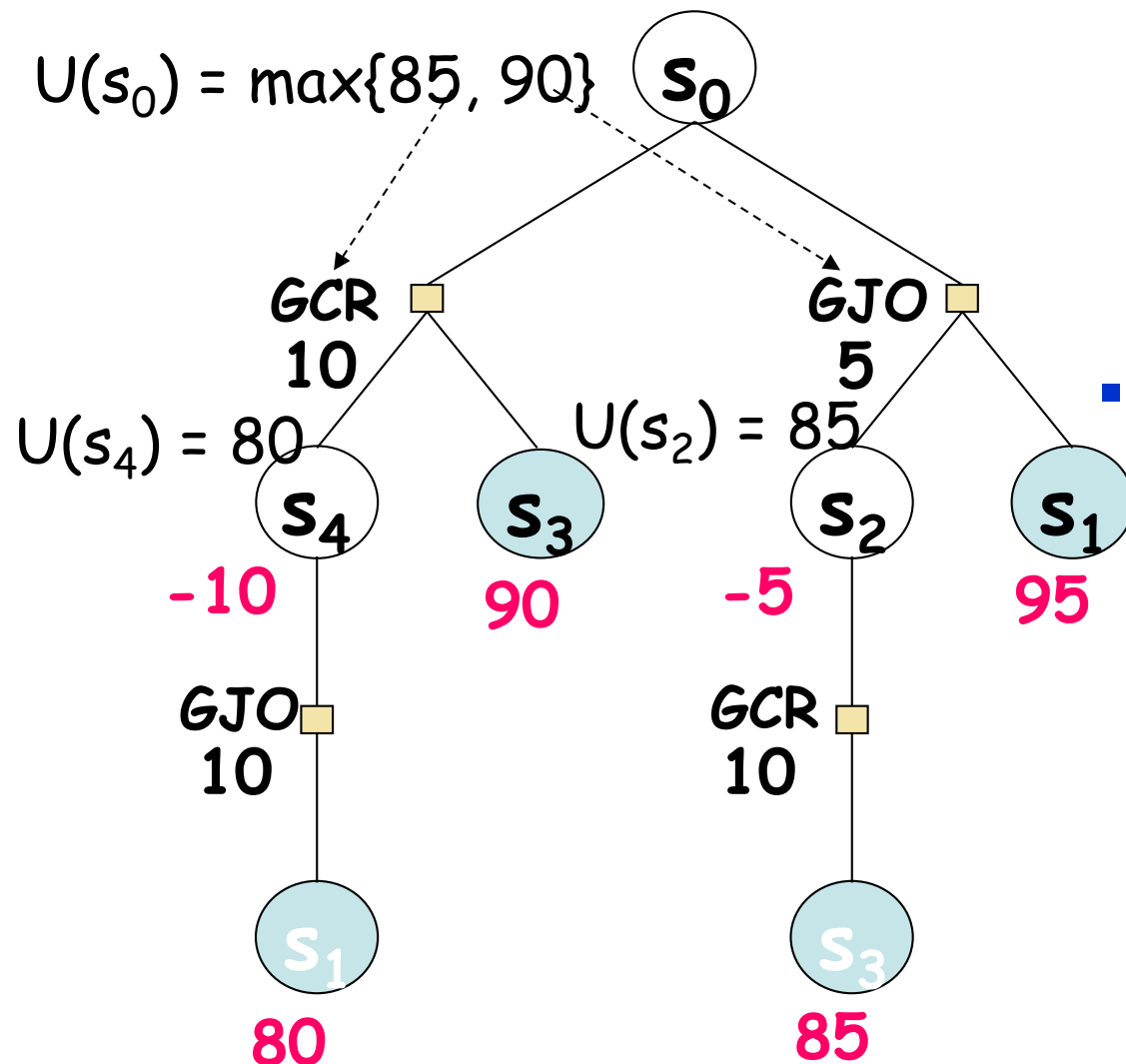
- In this example, only the terminal states give positive rewards (say, 100)
- The cost of each action (~ negative reward of reached state) is its duration

# State/Action Tree

- The computation of  $U$  (utility) at each non-terminal node resembles backing up the values of an evaluation function in adversarial search



# State/Action Tree



- The computation of  $U$  (utility) at each non-terminal node resembles backing up the values of an evaluation function in adversarial search
- The best choice in  $s_0$  is to select GJO

- Inputs:
  - Initial state  $s_0$
  - Action (transition) model
  - Reward  $R(s)$  collected in each state  $s$
- A state is **terminal** if it has no successor
- Starting at  $s_0$ , the agent keeps executing actions until it reaches a terminal state
- Its goal is to maximize the expected sum of rewards collected (**additive** rewards)
- Assume for a while that the same state can't be reached twice (**no cycles**)  
[finite state space  $\rightarrow$  finite state/action tree]

## Utility of a State

The **utility** of a state  $s$  measures its desirability:


- If  $s$  is terminal:

$$U(s) = R(s)$$

- If  $s$  is non-terminal,

$$U(s) = R(s) + \max_{a \in \text{Appl}(s)} \sum_{s' \in \text{Succ}(s,a)} P(s'|a.s) U(s')$$

[the reward of  $s$  augmented by the expected sum of rewards collected in future states]


$$U(s) = R(s) + \max_{a \in \text{Appl}(s)} \sum_{s' \in \text{Succ}(s,a)} P(s'|a.s) U(s')$$

Appl(s) is the set of  
all actions applicable  
to state s

Succ(s,a) is the set of all  
possible states after  
applying a to s

$P(s'|a.s)$  is the probability  
of being in  $s'$  after  
executing a in s

[the reward of s augmented by the expected  
sum of rewards collected in future states]



## Utility with Action Costs

$$U(s) = R(s) + \max_{a \in \text{Appl}(s)} [-\text{cost}(a) + \sum_{s' \in \text{Succ}(s,a)} P(s'|a.s) U(s')]$$

**Bellman equation**




- A **policy** is a function that maps each state  $s$  into the action to execute if  $s$  is reached
- The **optimal policy**  $P^*$  is the policy that always lead to maximizing the expected sum of rewards collected in future states (Maximum Expected Utility principle)

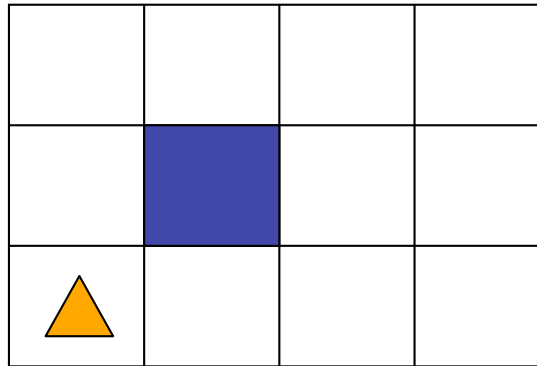
$$P^*(s) = \text{arg max}_{a \in \text{Appl}(s)} [-\text{cost}(a) + \sum_{s' \in \text{Succ}(s,a)} P(s'|a.s)U(s')]$$

$$U(s) = R(s) + \max_{a \in \text{Appl}(s)} [-\text{cost}(a) + \sum_{s' \in \text{Succ}(s,a)} P(s'|a.s)U(s')]$$

- 1) What if the set of states reachable from the initial state is too large to be entirely generated (e.g., there is a time limit)?
- 2) How to deal with cycles (states that can be reached multiple times)?

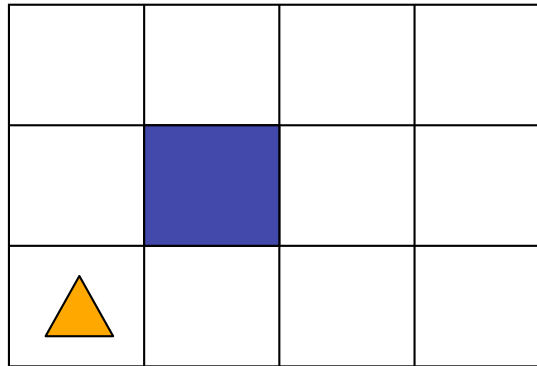
- 
- 1) What if the set of states reachable from the initial state is too large to be entirely generated (e.g., there is a time limit)?
- Expand the state/action tree to some depth  $h$
  - Estimate the utilities of leaf nodes  
[Reminiscent of evaluation function in game trees]
  - Back-up utilities as described before (using estimated utilities at leaf nodes)

# Simple Robot Navigation Problem



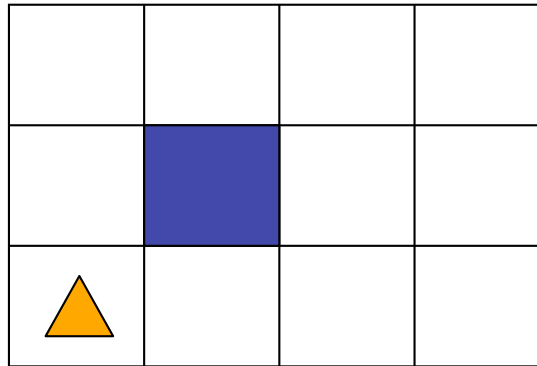
- In each state, the possible actions are U, D, R, and L

# Probabilistic Transition Model



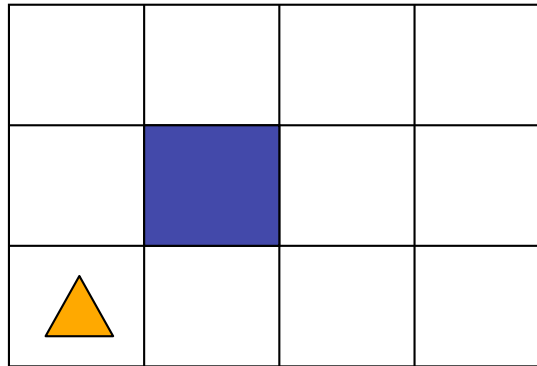
- In each state, the possible actions are U, D, R, and L
- The effect of U is as follows (**transition model**):
  - With probability 0.8 the robot does the right thing (U, D, R, L)

# Probabilistic Transition Model



- In each state, the possible actions are U, D, R, and L
- The effect of U is as follows (**transition model**):
  - With probability 0.8 the robot does the right thing (U, D, R, L)
  - With probability 0.1 it moves in a direction perpendicular to the intended one

# Probabilistic Transition Model

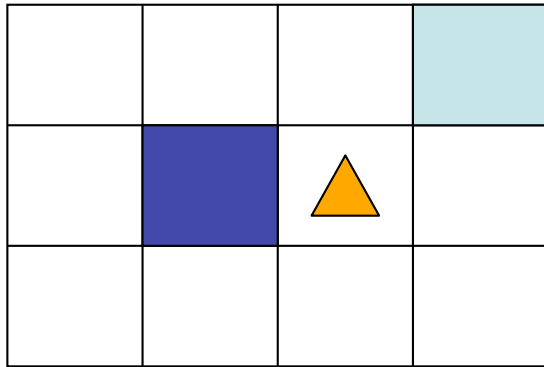


- In each state, the possible actions are U, D, R, and L
- The effect of U is as follows (**transition model**):
  - With probability 0.8 the robot does the right thing (U, D, R, L)
  - With probability 0.1 it moves in a direction perpendicular to the intended one
  - If the robot can't move, it stays in the same square

The transition properties depend only on the current state, not on previous history (how that state was reached)



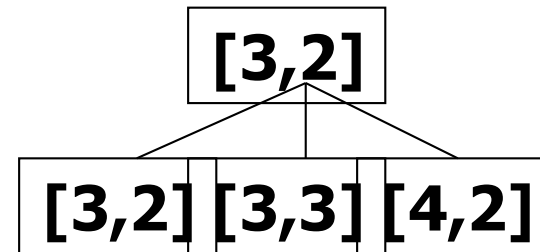
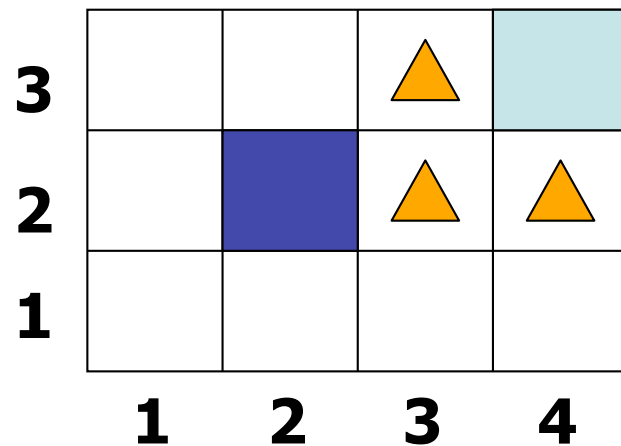
# Sequence of Actions



[3,2]

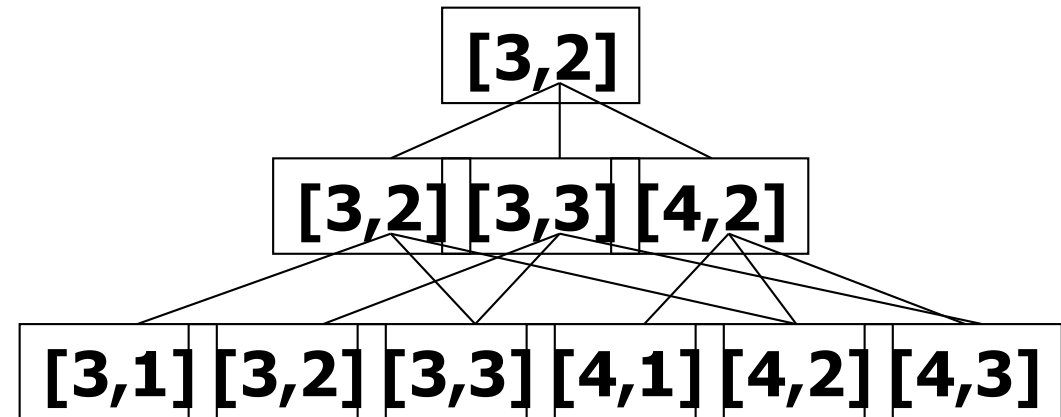
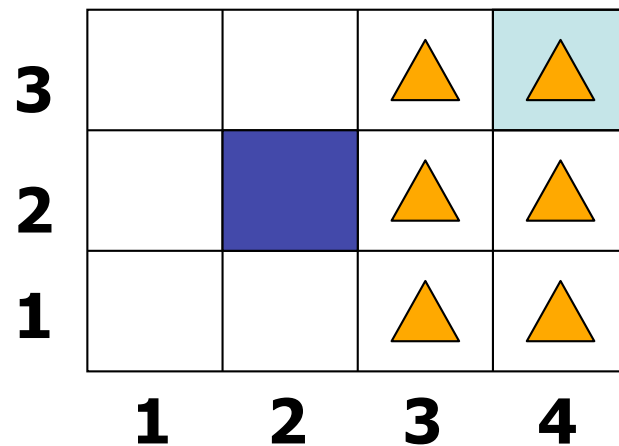
- Planned sequence of actions: (U, R)

# Sequence of Actions



- Planned sequence of actions: (U, R)
- U is executed

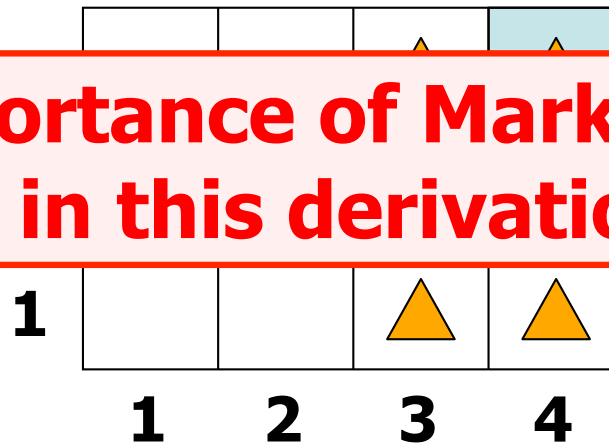
# Histories



- Planned sequence of actions: (U, R)
- U has been executed
- R is executed
- There are 9 possible sequences of states
  - called histories – and 6 possible final states for the robot!

# Probability of Reaching the Goal

**Note importance of Markov property  
in this derivation**



- $P([4,3] \mid (U,R).[3,2]) =$   

$$P([4,3] \mid R.[3,3]) \times P([3,3] \mid U.[3,2])$$

$$+ P([4,3] \mid R.[4,2]) \times P([4,2] \mid U.[3,2])$$
- $P([4,3] \mid R.[3,3]) = 0.8$
- $P([3,3] \mid U.[3,2]) = 0.8$
- $P([4,3] \mid R.[4,2]) = 0.1$
- $P([4,2] \mid U.[3,2]) = 0.1$
- $P([4,3] \mid (U,R).[3,2]) = 0.65$

# Utility Function

<b>3</b>				<b>+1</b>
<b>2</b>				<b>-1</b>
<b>1</b>				
	<b>1</b>	<b>2</b>	<b>3</b>	<b>4</b>

- $[4,3]$  provides power supply
- $[4,2]$  is a sand area from which the robot cannot escape

# Utility Function

<b>3</b>				<b>+1</b>
<b>2</b>				<b>-1</b>
<b>1</b>				
	<b>1</b>	<b>2</b>	<b>3</b>	<b>4</b>

- $[4,3]$  provides power supply
- $[4,2]$  is a sand area from which the robot cannot escape
- The robot needs to recharge its batteries

# Utility Function

<b>3</b>				<b>+1</b>
<b>2</b>				<b>-1</b>
<b>1</b>				
	<b>1</b>	<b>2</b>	<b>3</b>	<b>4</b>

- $[4,3]$  provides power supply
- $[4,2]$  is a sand area from which the robot cannot escape
- The robot needs to recharge its batteries
- $[4,3]$  or  $[4,2]$  are terminal states

# Utility of a History

<b>3</b>				<b>+1</b>
<b>2</b>				<b>-1</b>
<b>1</b>				
	<b>1</b>	<b>2</b>	<b>3</b>	<b>4</b>

- $[4,3]$  provides power supply
- $[4,2]$  is a sand area from which the robot cannot escape
- The robot needs to recharge its batteries
- $[4,3]$  or  $[4,2]$  are terminal states
- The utility of a history is defined by the utility of the last state (+1 or -1) minus  $n/25$ , where  $n$  is the number of moves



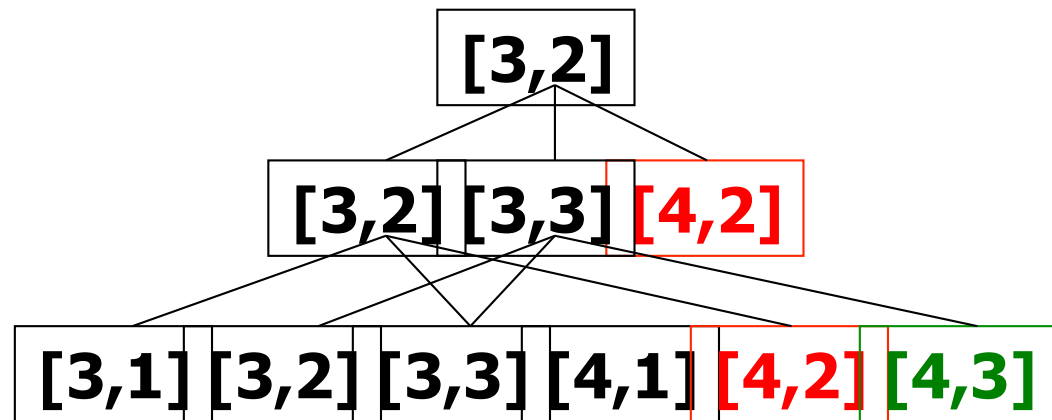
# Utility of an Action Sequence

<b>3</b>				<b>+1</b>
<b>2</b>				<b>-1</b>
<b>1</b>				
	<b>1</b>	<b>2</b>	<b>3</b>	<b>4</b>

- Consider the action sequence (U,R) from [3,2]

# Utility of an Action Sequence

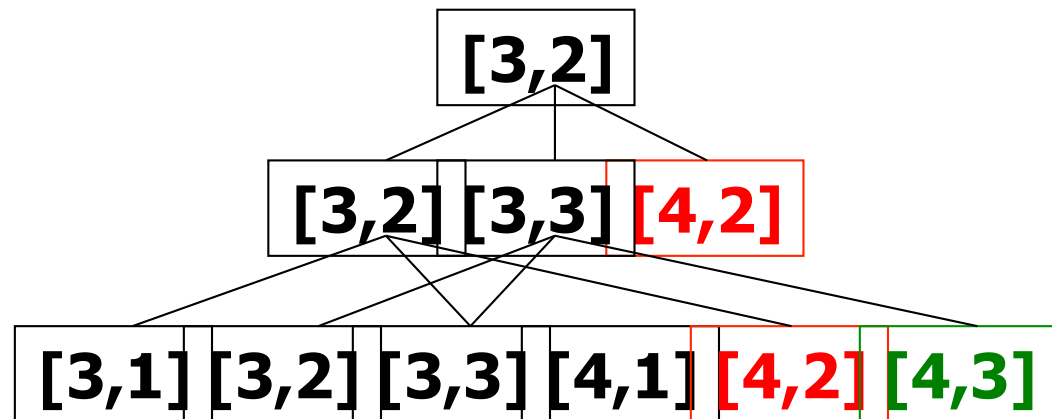
3				+1
2				-1
1				
	1	2	3	4



- Consider the action sequence (U,R) from [3,2]
- A run produces one among 7 possible histories, each with some probability

# Utility of an Action Sequence

			+1
			-1

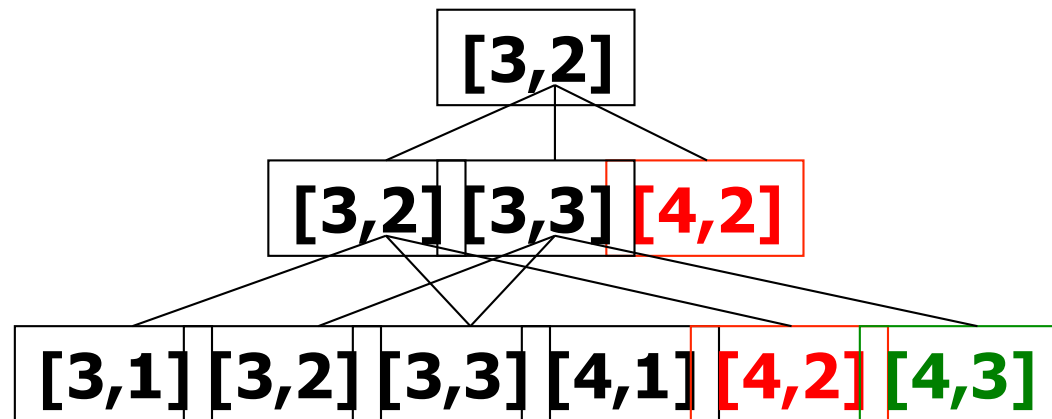


- Consider the action sequence (U,R) from [3,2]
- A run produces one among 7 possible histories, each with some probability
- The utility of the sequence is the expected utility of the histories:

$$U = \sum_h U_h P(h)$$

# Optimal Action Sequence

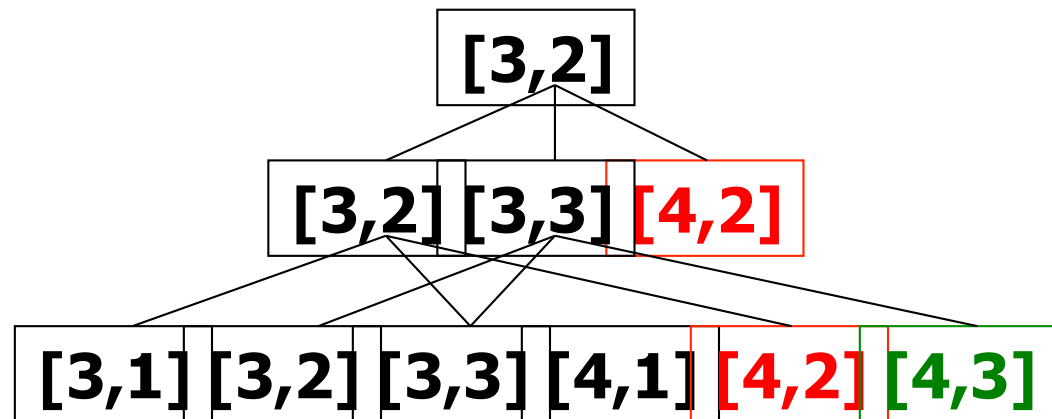
			+1
			-1



- Consider the action sequence (U,R) from [3,2]
- A run produces one among 7 possible histories, each with some probability
- The utility of the sequence is the expected utility of the histories
- The optimal sequence is the one with maximal utility


# Optimal Action Sequence

			+1
			-1



- Consider the action sequence (U,R) from [3,2]
- A run produces probability
- The utility of the sequence is the expected utility of the histories
- The optimal sequence is the one with maximal utility
- But is the optimal action sequence what we want to compute?

**only if the sequence is executed blindly!**



Therefore, a solution must specify what the agent should do for ***any state*** that the agent might reach.

A solution of this kind is called a policy.

# Terminal States, Rewards, and Costs

3	-.04	-.04	-.04	+1
2	-.04		-.04	-1
1	-.04	-.04	-.04	-.04
	1	2	3	4

- Two terminal states: (4,2) and (4,3)
- Rewards:
  - $R(4,3) = +1$  [The robot finds gold]
  - $R(4,2) = -1$  [The robot gets trapped in quick sands]
  - $R(s) = -0.04$  in all other states
- Actions have zero cost  
[actually, they are encoded in the negative rewards of non-terminal states]

## Utility of a State

The **utility** of a state  $s$  measures its desirability:

- If  $s$  is terminal:

$$U(s) = R(s)$$

- If  $s$  is non-terminal,

$$U(s) = R(s) + \max_{a \in \text{Appl}(s)} \sum_{s' \in \text{Succ}(s,a)} P(s'|a.s) U(s')$$

[the reward of  $s$  augmented by the expected sum of rewards collected in future states]



# State Utilities

3	0.81	0.87	0.92	+1
2	0.76		0.66	-1
1	0.71	0.66	0.61	0.39
	1	2	3	4

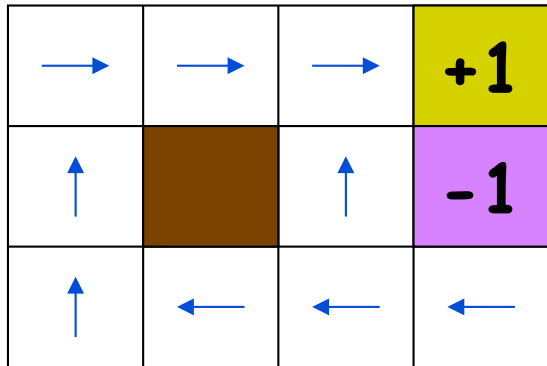
- The utility of a state  $s$  is the maximal expected amount of reward that the robot will collect from  $s$  and future states by executing some action in each encountered state, until it reaches a terminal state (infinite horizon)
- Under the Markov and infinite horizon assumptions, the utility of  $s$  is independent of when and how  $s$  is reached  
[It only depends on the possible sequences of states after  $s$ , not on the possible sequences before  $s$ ]

# Policy (Reactive/Closed-Loop Strategy)

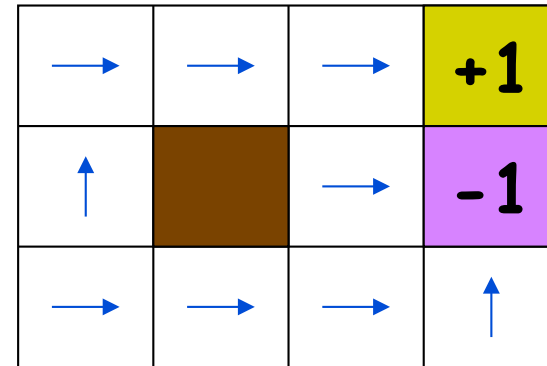
<b>3</b>	→	→	→	<b>+1</b>
<b>2</b>	↑		↑	<b>-1</b>
<b>1</b>	↑	←	←	←
	<b>1</b>	<b>2</b>	<b>3</b>	<b>4</b>

- A **policy**  $\Pi$  is a complete mapping from states to actions

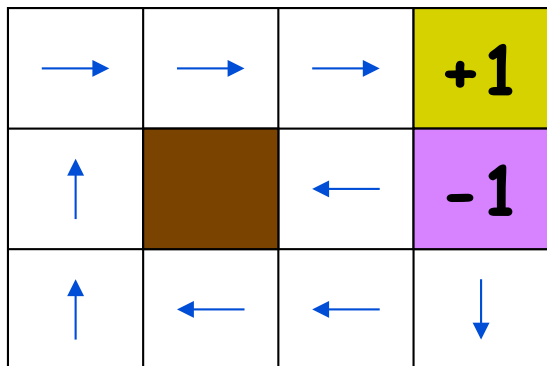
# Optimal Policies for Various $R(s)$



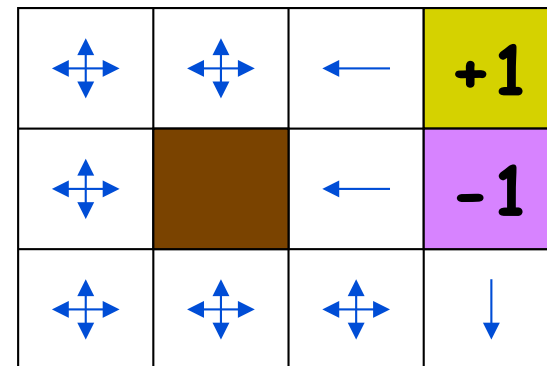
$$R(s) = -0.04$$



$$R(s) = -2$$



$$R(s) = -0.01$$



$$R(s) > 0$$

**Repeat:**

- ◆  **$s \leftarrow$  sensed state**
- ◆ **If  $s$  is terminal then exit**
- ◆  **$a \leftarrow \Pi(s)$**
- ◆ **Perform  $a$**

# Optimal Policy

3	→	→	→	+1
2	↑		↑	-1
1	↑	←	←	←
1				

**Note that [3,2] is a “dangerous” state that the optimal policy tries to avoid**

- A policy  $\Pi$  is a complete mapping from states to actions.
- The optimal policy  $\Pi^*$  is the one that always yields a history (ending at a terminal state) with maximal expected utility.

Makes sense because of Markov property

# Optimal Policy

<b>3</b>	→	→	→	<b>+1</b>
<b>2</b>	↑		↑	<b>-1</b>
<b>1</b>	↑	←	←	←
	<b>1</b>	<b>2</b>	<b>3</b>	<b>4</b>

- A policy  $\Pi$  is a complete rule for choosing actions
- The optimal policy  $\Pi^*$  is the policy that maximizes the expected utility over all possible histories with maximal expected utility

**This problem is called a Markov Decision Problem (MDP)**

**How to compute  $\Pi^*$ ?**

History  $H = (s_0, s_1, \dots, s_n)$

The utility of  $H$  is **additive** iff:

$$U(s_0, s_1, \dots, s_n) = R(0) + U(s_1, \dots, s_n) = \sum R(i)$$



**Reward**

History  $H = (s_0, s_1, \dots, s_n)$

The utility of  $H$  is **additive** iff:

$$U(s_0, s_1, \dots, s_n) = R(0) + U(s_1, \dots, s_n) = \sum R(i)$$

Robot navigation example:

$$R(n) = +1 \text{ if } s_n = [4, 3]$$

$$R(n) = -1 \text{ if } s_n = [4, 2]$$

$$R(i) = -1/25 \text{ if } i = 0, \dots, n-1$$



# Defining Equations

3	→	→	→	+1
2	↑		↑	-1
1	↑	←	←	←
	1	2	3	4

- If  $s$  is terminal:  
$$U(s) = R(s)$$
- If  $s$  is non-terminal:  
$$U(s) = R(s) + \max_{a \in \text{Appl}(s)} \sum_{s' \in \text{Succ}(s,a)} P(s'|a.s) U(s')$$

[Bellman equation]
- $P^*(s) = \arg \max_{a \in \text{Appl}(s)} \sum_{s' \in \text{Succ}(s,a)} P(s'|a.s) U(s')$

# Defining Equations

3	→	→	→	+1
2	↑		↑	-1
1	↑	←	←	←
	1	2	3	4

The utility of  $s$  depends on the utility of other states  $s'$  (possibly, including  $s$ ), and vice versa

The equations are non-linear

- If  $s$  is terminal:

$$U(s) = R(s)$$

- If  $s$  is non-terminal:

$$U(s) = R(s) + \max_{a \in \text{Appl}(s)} \sum_{s' \in \text{Succ}(s,a)} P(s'|a.s) U(s')$$

[Bellman equation]

- $P^*(s) = \arg \max_{a \in \text{Appl}(s)} \sum_{s' \in \text{Succ}(s,a)} P(s'|a.s) U(s')$

# Value Iteration Algorithm

3	0	0	0	+1
2	0		0	-1
1	0	0	0	0
	1	2	3	4

→

3	0.81	0.87	0.92	+1
2	0.76		0.66	-1
1	0.71	0.66	0.61	0.39
	1	2	3	4

1. Initialize the utility of each non-terminal states to  $U_0(s) = 0$
2. For  $t = 0, 1, 2, \dots$  Do

$$U_{t+1}(s) = R(s) + \max_{a \in \text{Appl}(s)} \sum_{s' \in \text{Succ}(s,a)} P(s'|a.s) U_t(s')$$

for each non-terminal state  $s$

# Value Iteration Algorithm

3	0.81	0.87	0.92	+1
2	0.76		0.66	-1
1	0.71	0.66	0.61	0.39
	1	2	3	4

→

3	→	→	→	+1
2	↑		↑	-1
1	↑	←	←	←
	1	2	3	4

1. Initialize the utility of each non-terminal states to  $U_0(s) = 0$
2. For  $t = 0, 1, 2, \dots$  do

$$U_{t+1}(s) = R(s) + \max_{a \in \text{Appl}(s)} \sum_{s' \in \text{Succ}(s,a)} P(s'|a.s) U_t(s')$$

for each non-terminal state  $s$

3. For each non-terminal state  $s$  do

$$P^*(s) = \arg \max_{a \in \text{Appl}(s)} \sum_{s' \in \text{Succ}(s,a)} P(s'|a.s) U(s')$$

# Value Iteration Algorithm

3	0.81	0.87	0.92	+1
0	0.76		0.66	1

→

3	→	→	→	+1
0	↑		↑	1

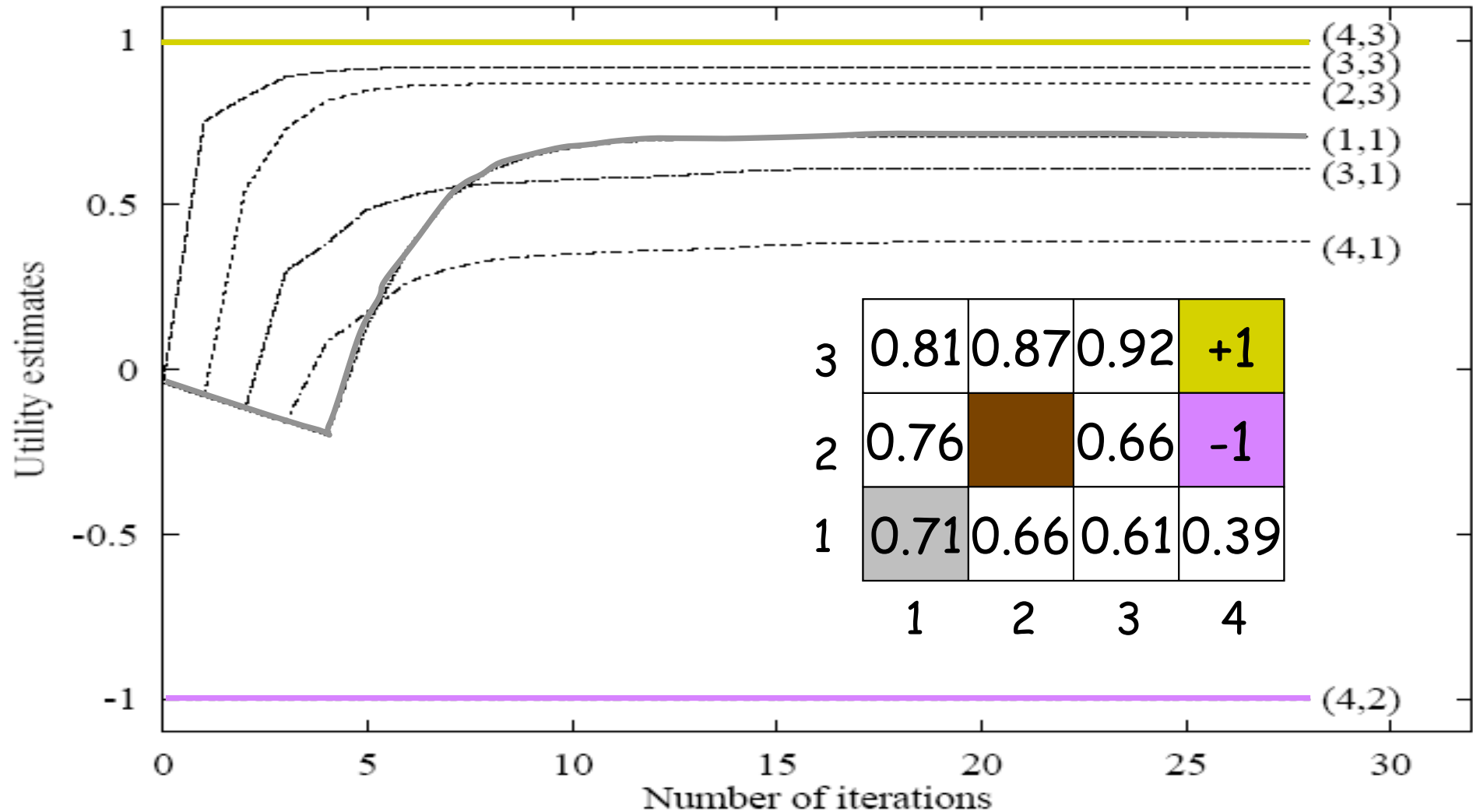
Value iteration is essentially the same as computing the best move from each state using a state/action tree expanded to a large depth  $h$  (with estimated utilities of the non-terminal leaf nodes set to 0)

By doing the computation for all states simultaneously, it avoids much redundant computation

3. For each non-terminal state  $s$  do

$$P^*(s) = \arg \max_{a \in \text{Appl}(s)} \sum_{s' \in \text{Succ}(s,a)} P(s'|a.s) U(s')$$

# Convergence of Value Iteration



# Convergence of Value Iteration

- If:
  - The number of states is finite
  - There exists at least one terminal state that gives a positive reward and is reachable with non-zero probability from every other non-terminal state (connectivity of the state space)
  - $R(s) \leq 0$  at every non-terminal state
  - The cost of every action is  $\geq 0$
- Then value iteration converges toward an optimal policy *if we wait long enough*
- But what if the above conditions are not verified?