Reasoning Under Uncertainty

Russell & Norvig chapter 12

T-622-ARTI

Spring 2023

Introducing Uncertainty (Ch. 12)

It is not the world that is imperfect, it is our knowledge of it.

So far we have assumed that:

- World states are perfectly observable:
  - The current state is exactly known
- Action representations are perfect:
  - States are exactly predicted

- We will now investigate how an agent can cope with imperfect information

Intelligence

Pei Wang – Temple University

“Intelligence” … is about whether, or how much, a system can improve its problem-solving capability by adaptation, that is, learning from its past experience.

https://cis.temple.edu/~pwang/GTI-book/

Intelligent vs. Instinctive

In an **instinctive system**, all major components and their relations are determined when the system is formed, and remain unchanged afterwards.

In an **intelligent system**, all major components and their relations are adaptive to the environment. The system learns new beliefs, organize actions into skills, establish new goals, all as attempts to improve its goal-achieving capability, under the assumption that in general the future will be similar to the past.

Intelligence is the ability to adapt to the environment with insufficient knowledge and resources.

https://cis.temple.edu/~pwang/GTI-book/

Sources of Uncertainty

- **The Representation Language**
- **Imperfect Observation of the World**
- **Ignorance, Laziness, Efficiency**

Representing the Real World

Agent's conceptualization
- Representation language

3x3 matrix filled
with 1, 2, .., 8, and
'empty'

Real world

8-puzzle

Representing the Real World

Agent's conceptualization
- Representation language

Logic sentences using propositions like
Block(A), On(A,B), Handempty, ...
and connectives

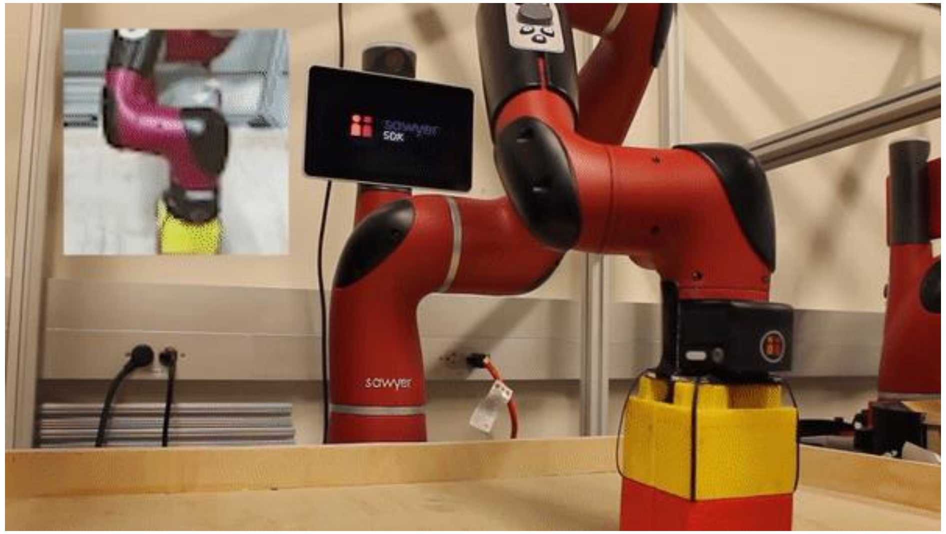
Real world

Blocks world

Who provides the representation language?

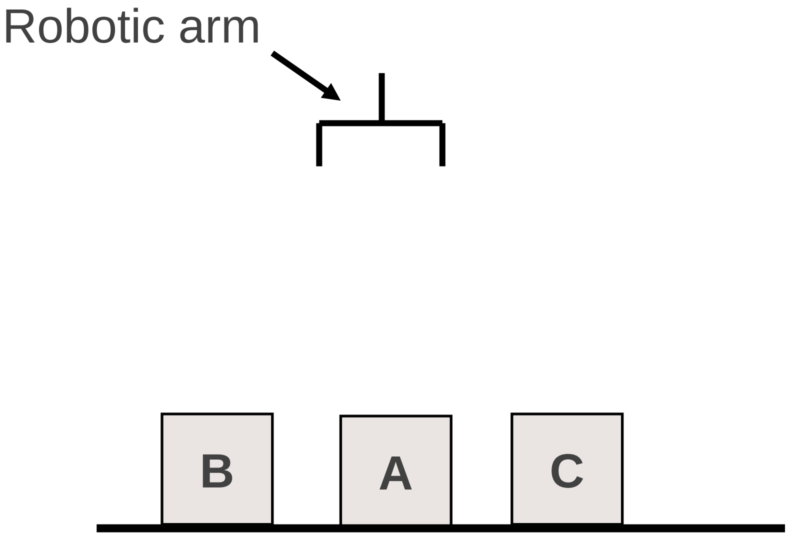
- The agent’s designer
- Few practical techniques exist to allow an agent to autonomously abstract features of the real world into useful concepts
  - e.g. transformer embeddings, Thing2Vec
- Techniques allowing an agent to develop its own representation language using these concepts?
  - Inductive learning techniques are steps in this direction
- In the following slides:
  - Representation language is provided by the agent’s designer
  - or developed over time by the agent

Robots stacking blocks



https://bair.berkeley.edu/blog/2019/05/20/solar/

Stacking blocks
Simplified representation

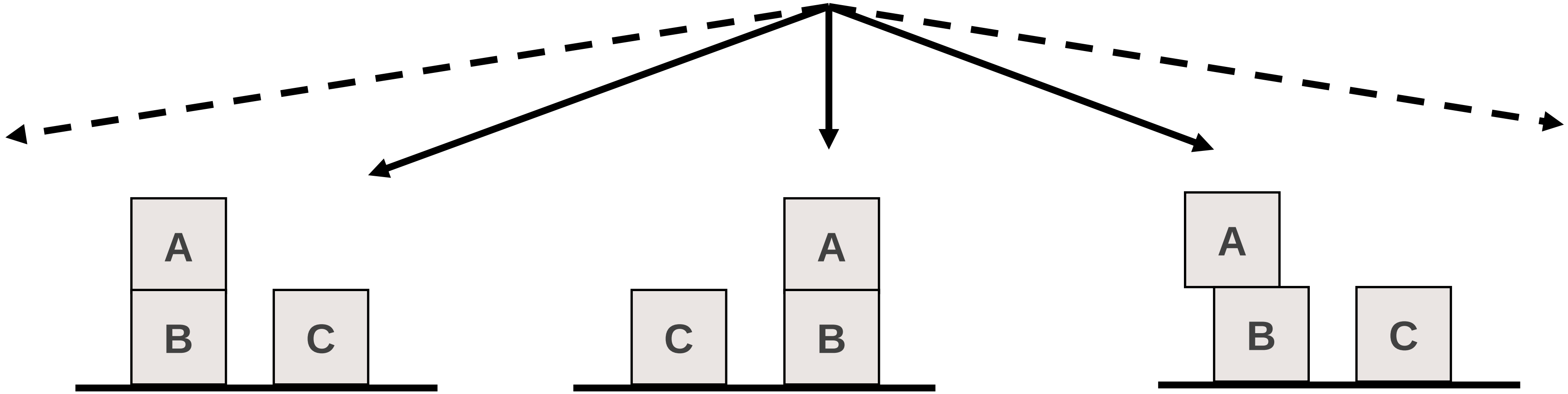


$$On(A, Table) On(B, Table) \wedge On(C, Table) \wedge \cdots$$

First Source of Uncertainty:
The Representation Language

- There are many more states of the real world than can be expressed in the representation language
- So, any state represented in the language may correspond to many different states of the real world, which the agent can't represent distinguishably

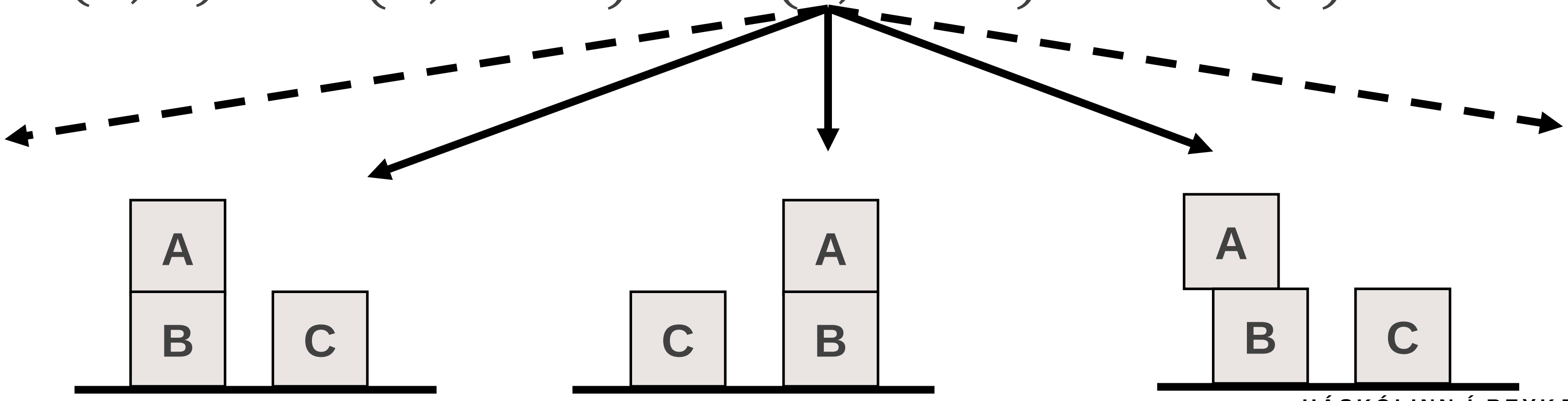
$$On(A,B) \wedge On(B,Table) \wedge On(C,Table) \wedge Clear(A) \wedge Clear(C)$$



First Source of Uncertainty:
The Representation Language

- 6 propositions On(x,y), where x, y = A, B, C and x ≠ y
- 3 propositions On(x,Table), where x = A, B, C
- 3 propositions Clear(x), where x = A, B, C
- At most 2^{12} states can be distinguished in the language
- But there are infinitely many states of the real world

$$On(A,B) \land On(B,Table) \land On(C,Table) \land Clear(A) \land Clear(C)$$



An action representation may be incorrect

$\boldsymbol{Stack(C, A)}$:

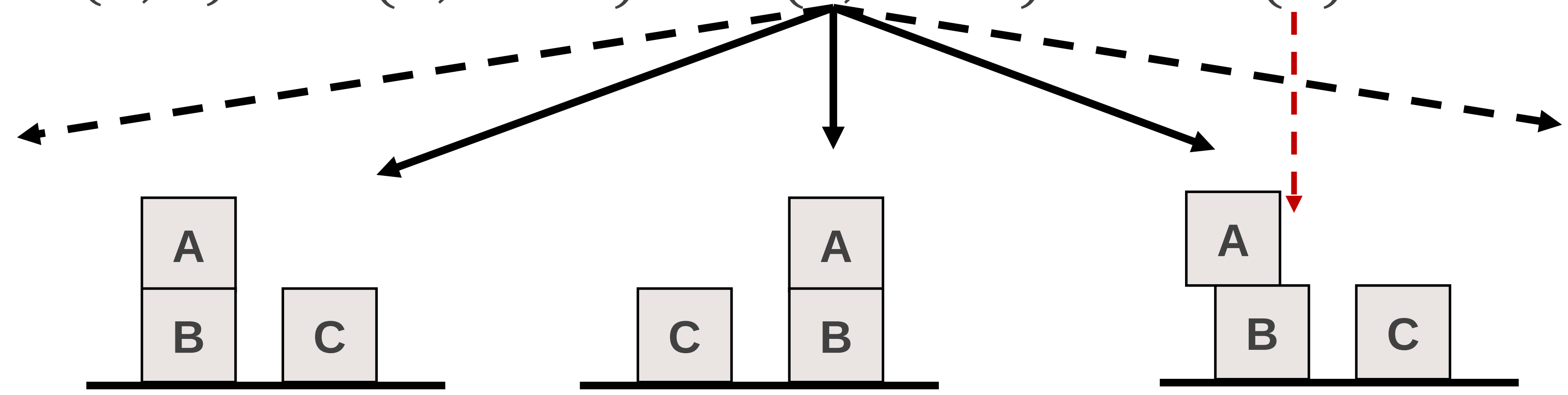
$Precondition = Holding(C) \land Block(A) \land Block(C) \land Clear(A)$

$Delete = Clear(A), Holding(C)$

$Add = On(C, A), Clear(C), Handempty$

is likely not to have the described effects in case 3 because the precondition is “incomplete”

$$On(A, B) \land On(B, Table) \land On(C, Table) \land Clear(A) \land Clear(C)$$



An action may describe alternative effects

$\boldsymbol{Stack(C,A)}$:

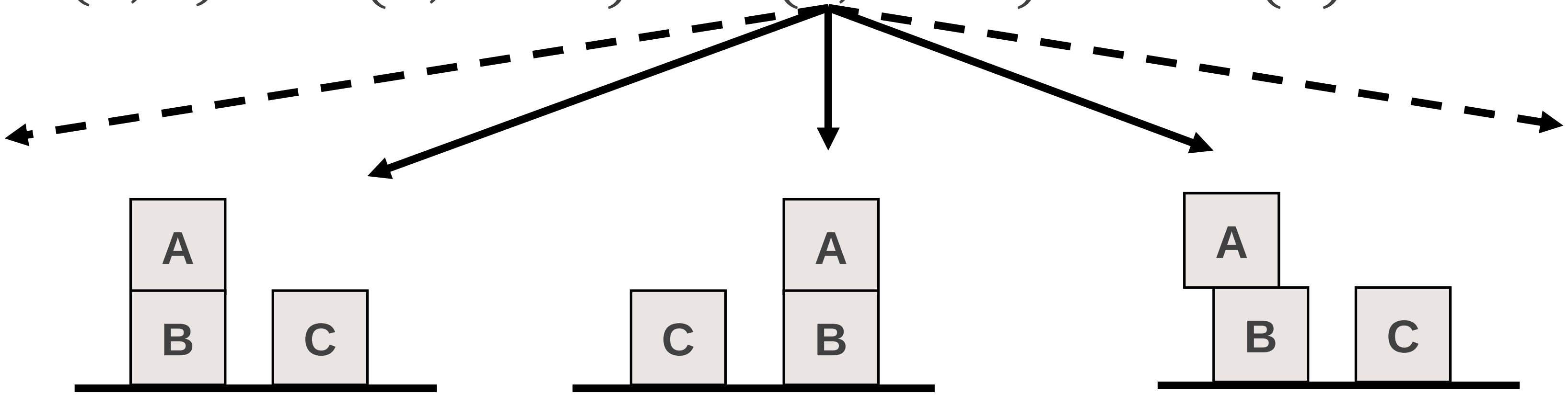
$$Precondition = Holding(C) \land Block(A) \land Block(C) \land Clear(A)$$
$$If\ On(A,x) \land (x \neq Table)$$

$$E_1 \begin{cases} Delete = Clear(A), Holding(C) \\ Add = On(C,A), Clear(C), Handempty \end{cases}$$

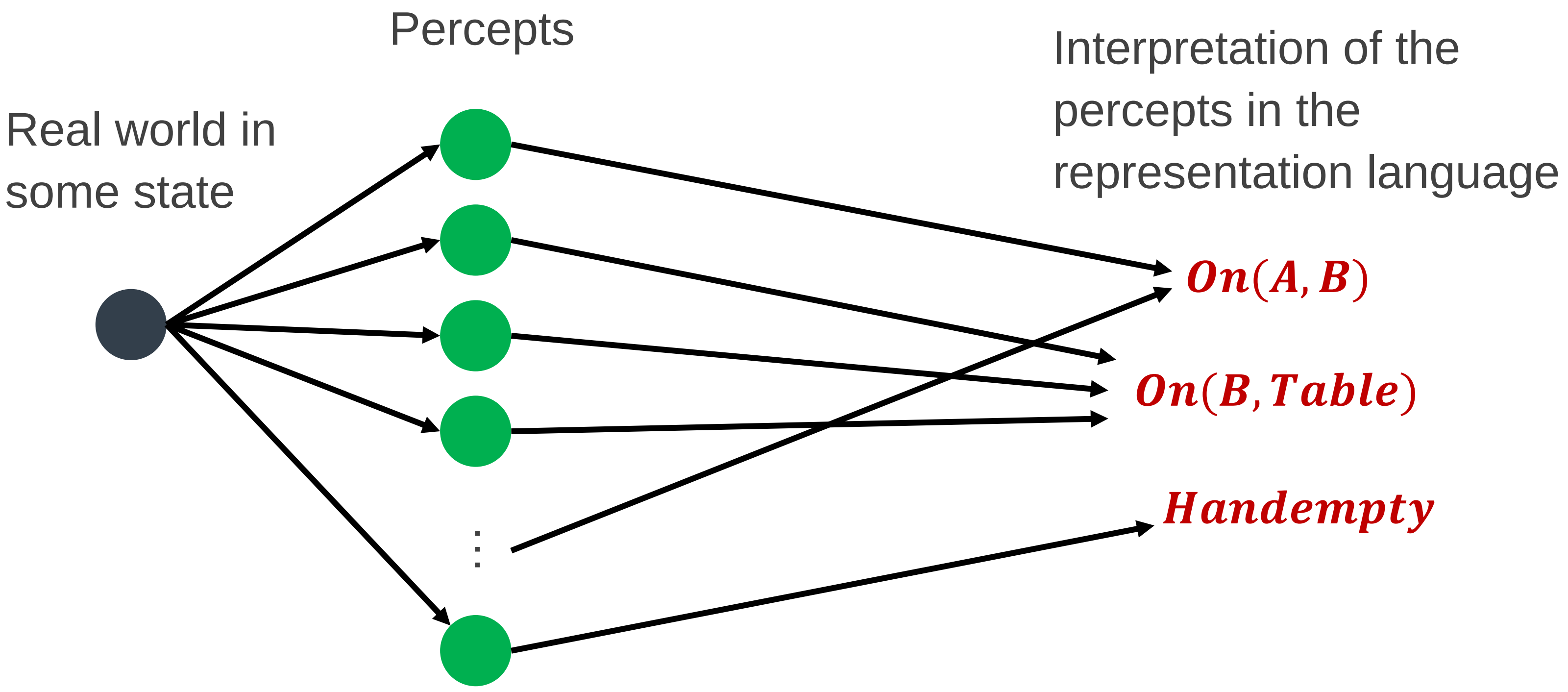
OR

$$E_2 \begin{cases} Delete = On(A,x), Holding(C) \\ Add = On(C,Table), Clear(C), Handempty, On(A,Table), Clear(A), Clear(x) \end{cases}$$

$$On(A,B) \land On(B,Table) \land On(C,Table) \land Clear(A) \land Clear(C)$$



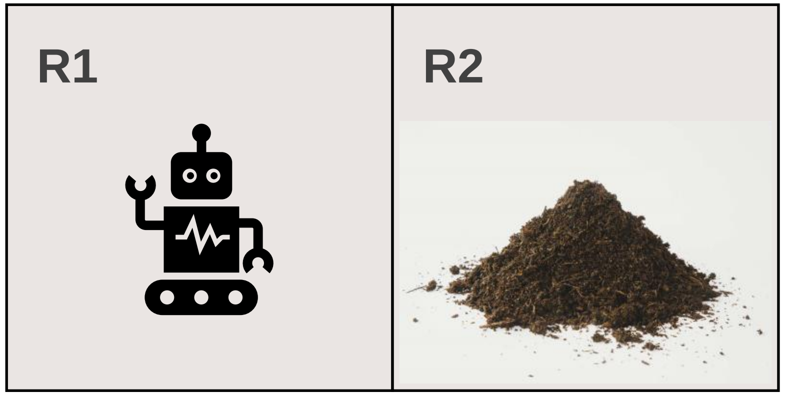
Observation of the real world



Percepts can be user's inputs, sensory data (e.g., image pixels), information received from other agents, ...

Second source of Uncertainty:
Imperfect Observation of the World

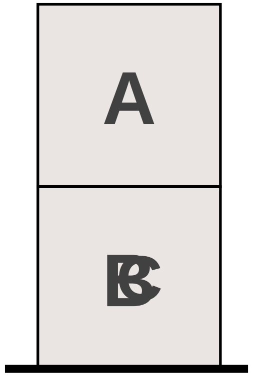
- **Observation of the world can be:**
  - *Partial* - a vision sensor can’t see through obstacles (lack of percepts)



The robot may not know whether there is dirt in room R2

Second source of Uncertainty:
Imperfect Observation of the World

- **Observation of the world can be:**
  - *Partial* - a vision sensor can’t see through obstacles (lack of percepts)
  - *Ambiguous* - percepts have multiple possible interpretations



$$\longrightarrow \quad On(A,B) \vee On(A,C)$$

Second source of Uncertainty:
Imperfect Observation of the World

- **Observation of the world can be:**
  - *Partial* - a vision sensor can't see through obstacles (lack of percepts)
  - *Ambiguous* - percepts have multiple possible interpretations
  - *Incorrect*

Third Source of Uncertainty:
Ignorance, Laziness, Efficiency

- An action may have a long list of preconditions:
  - **Drive-Car:**

$$P = Have(Keys) \land \neg Empty(GasTank) \land BatteryOk \land IgnitionOk \land \neg FlatTires \land \neg Stolen(Car) \ldots$$

- The agent's designer may **ignore** some preconditions,
  or by **laziness** or **efficiency**, may not include all in the action representation
- Results in representation that is incorrect:
  - Executing the action may not have the described effects
- Or has several alternative effects

Third Source of Uncertainty:
Ignorance, Laziness, Efficiency

- Example of uncertain reasoning: diagnosing a dental patient’s toothache
- Using logic to cope with this domain fails:
  - **Laziness:** Too much work to list complete set of antecedents and consequences and too hard to use those rules
  - **Theoretical ignorance:** Medical science has no complete theory for domain
  - **Practical ignorance:** Even if rules are known, individual patients introduce uncertainty because not all test may have been run

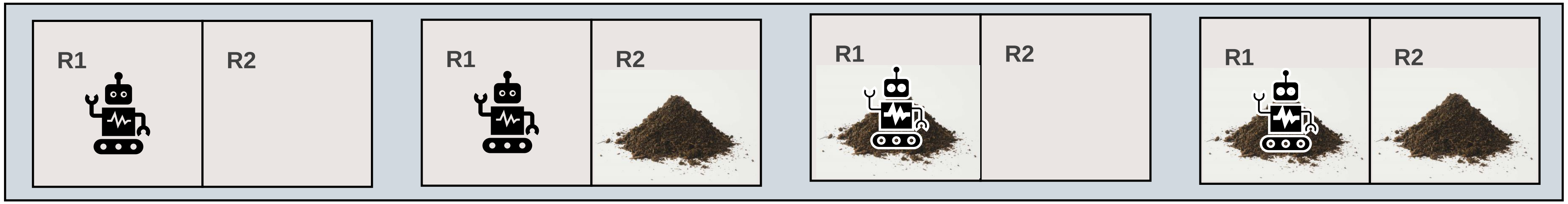
Representation of Uncertainty

- Many models of uncertainty
- We will consider two important models:

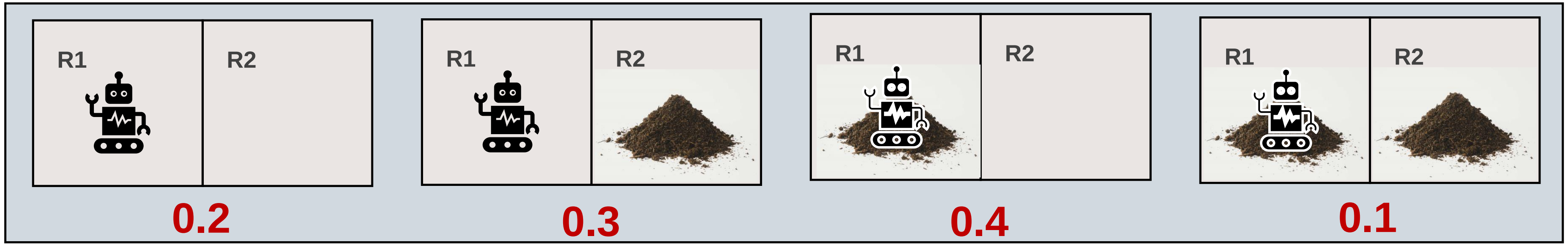
1. **Non-deterministic model:**
   - Uncertainty is represented by a set of possible values
     - a set of possible worlds, a set of possible effects, ...
2. **Probabilistic model:**
   - Uncertainty is represented by a probabilistic distribution over a set of possible values

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

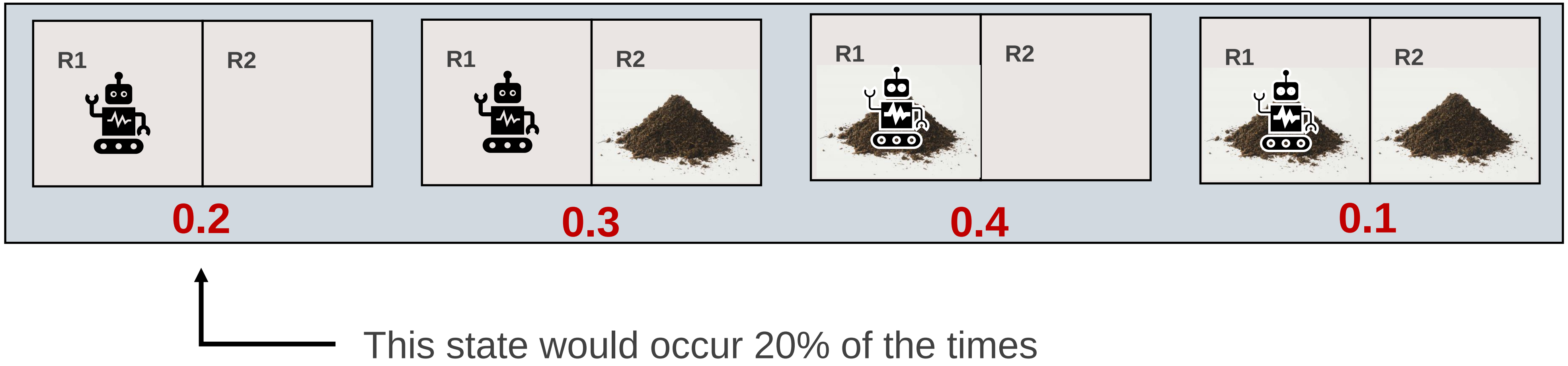


In the probabilistic model of uncertainty, a probability is associated with each state to measure its likelihood to be the actual state



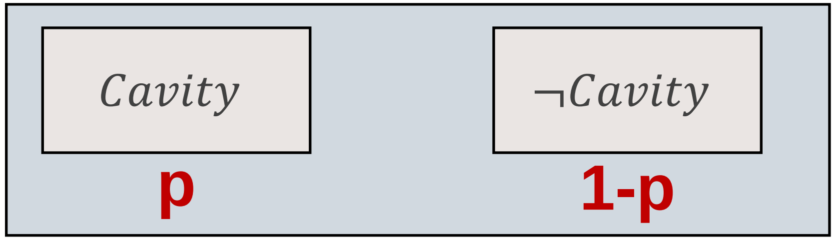
What do probabilities mean?

- Probabilities have a natural frequency interpretation
- The agent believes:
  - If it returned many times to a situation where it has the same belief state → the actual states in this situation would occur at a relative frequency defined by the probabilistic distribution



Example: Dentist Agent

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



- D believes that a fraction p of patients have cavities

Where do probabilities come from?

- **Frequencies observed in the past:**
  - by the agent, its designer, or others
- **Symmetries:**
  - Rolling dice, each of the 6 outcomes has probability 1/6
- **Subjectivism:**
  - Driving on Highway 280 at 120mph, will get speeding ticket with probability 0.6
  - Principle of indifference:
    - No knowledge if one possibility more probable than another, give them the same probability

Decision Theory

General theory of rational decisions

The theory of probability provides a way of summarizing the uncertainty that comes from our laziness and ignorance

Agents can have preferences for different possible outcomes

Utility Theory: Every state (or sequence) has a degree of usefulness to an agent

General theory of rational decisions:

$$Decision\ Theory = probability\ theory + utility\ theory$$

An agent is rational if and only if it chooses the action that yields the highest expected utility, averaged over all the possible outcomes of the action

→ Maximum Expected Utility

A decision-theoretic agent that selects rational actions

```
function DT-Agent(percept) returns an action

        persistent:  belief_state probabilistic beliefs about the current state
                     of the world
                     action the agent's action

        update belief_state based on action and percept

        calculate outcome probabilities for actions
              given action descriptions and current belief_state

        select action with highest expected utility
              given probabilities of outcomes and utility information

        return action
```

Probabilistic Belief States and Bayesian Networks

Russell & Norvig chapter 13

T-622-ARTI

Spring 2023

Probabilistic Belief

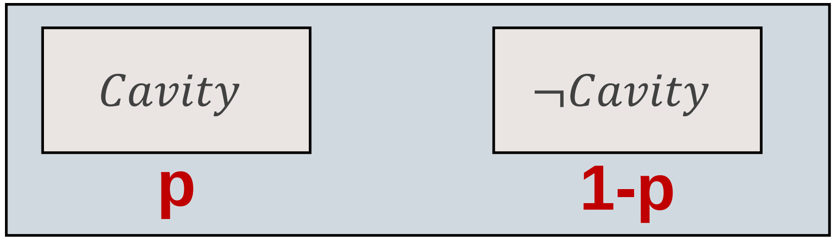
- Consider a world where a dentist agent D meets with a new patient P
- D is interested in only whether P has a cavity:
  - a state is described with a single proposition: $Cavity$
- Before observing P, D does not know if P has a cavity, but from years of practice, he believes $Cavity$ with some probability p and $\neg Cavity$ with probability $1 - p$
- The proposition is now a boolean random variable and $(Cavity, p)$ is a probabilistic belief

Probabilistic Belief State

The world has only two possible states, which are respectively described by $Cavity$ and $\neg Cavity$

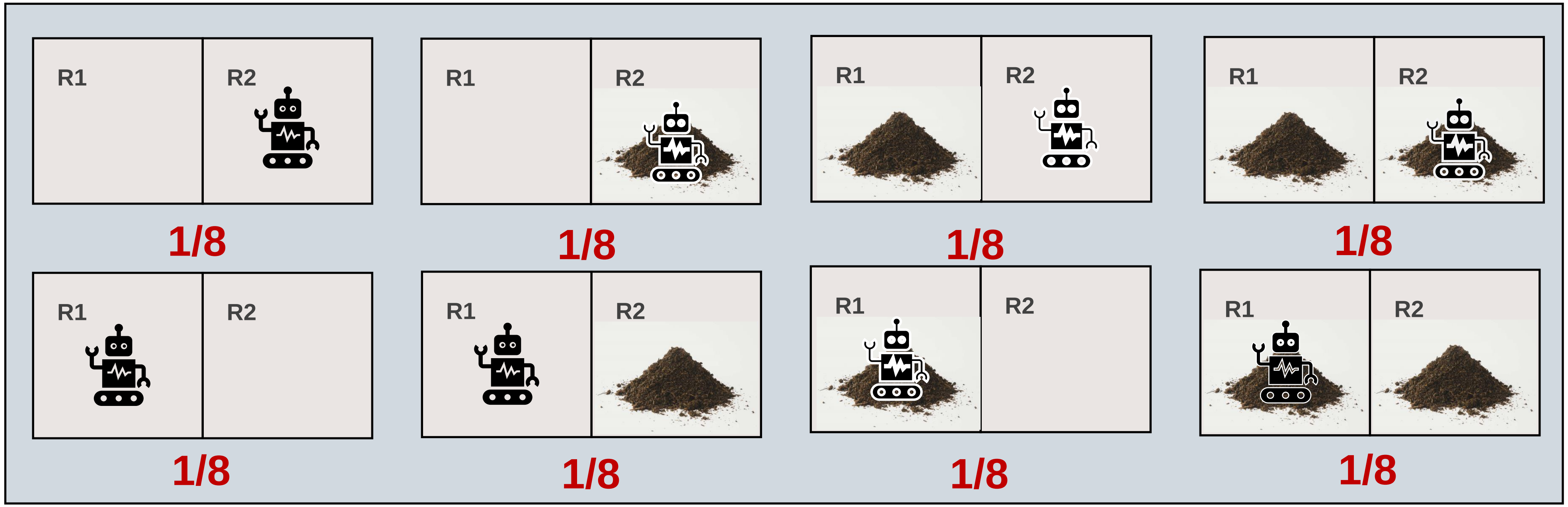
The probabilistic belief state of an agent is a probabilistic distribution over all the states that the agent thinks possible

The dentist agent's belief state is:

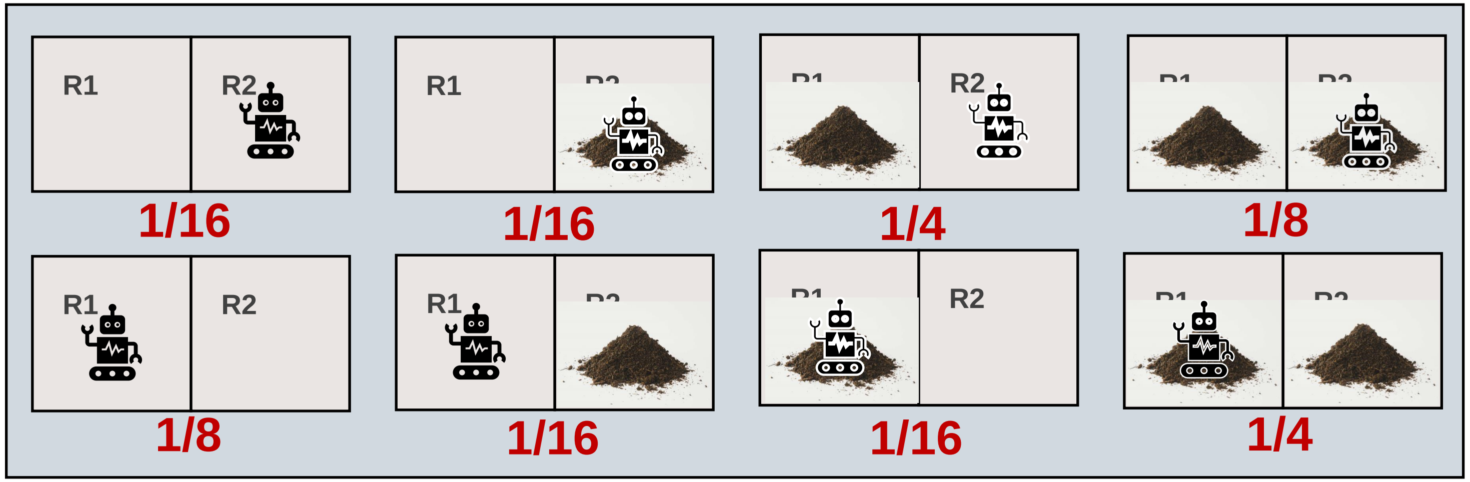


Vacuum Robot

- The Principle of Indifference:
  - If the robot has no idea what the state of the world is, and thinks that all states are equally probable, its belief state is:



How are beliefs and belief states related?



How does a belief affect the entire belief state and the other beliefs?

$(Clean(R1), 5/16)$
$(Clean(R2), 0.5)$
$(In(Robot, R1), 0.5)$
$(In(Robot, R2), 0.5)$

It is usually more convenient to deal with individual beliefs than with entire belief states:

- The robot may choose to execute $Suck(R2)$ only if $Clean(R2)$ has low probability

- The robot may directly observe whether $Clean(R1)$ or $Clean(R2)$

Back to the dentist's …

- Let's represent the world of the dentist D using three propositions:
  - $Cavity$, $Toothache$, and $Catch$
- D's belief state consists of $2^3 = 8$ states, each with some probability:

$$\{Cavity \wedge Toothache \wedge Catch,$$
$$\neg Cavity \wedge Toothache \wedge Catch,$$
$$Cavity \wedge \neg Toothache \wedge PCatch, \ldots\}$$

The belief state is defined by the full joint probability of the propositions

	$toothache$		$\neg toothache$	
	$Catch$	$\neg Catch$	$Catch$	$\neg Catch$
$cavity$	0.108	0.012	0.072	0.008
$\neg cavity$	0.016	0.064	0.144	0.576

Probabilistic Inference

	$toothache$		$\neg toothache$	
	$Catch$	$\neg Catch$	$Catch$	$\neg Catch$
$cavity$	0.108	0.012	0.072	0.008
$\neg cavity$	0.016	0.064	0.144	0.576

$$P(cavity \vee toothache) = 0.108 + 0.012 + \cdots$$
$$= 0.28$$

Probabilistic Inference

	$toothache$		$\neg toothache$	
	$Catch$	$\neg Catch$	$Catch$	$\neg Catch$
$cavity$	0.108	0.012	0.072	0.008
$\neg cavity$	0.016	0.064	0.144	0.576

$$P(cavity) = 0.108 + 0.012 + 0.072 + 0.008$$
$$= 0.2$$

Marginal probability

Probabilistic Inference

	$toothache$		$\neg toothache$	
	$catch$	$\neg catch$	$catch$	$\neg catch$
$cavity$	0.108	0.012	0.072	0.008
$\neg cavity$	0.016	0.064	0.144	0.576

Marginalization: Summing out – summing up the probabilities for each possible value of the other variables, taking them out of the equation

$$\begin{aligned} P(Cavity) &= P(Cavity, toothache, catch) + P(Cavity, tootache, \neg catch) \\ &\quad + P(Cavity, \neg toothache, catch) + P(Cavity, \neg tootache, \neg catch) \\ &= \langle 0.108, 0.016\rangle + \langle 0.012\ 0.064\rangle + \langle 0.072, 0.144\rangle + \langle 0.008, 0.576\rangle \\ &= \langle 0.2, 0.8\rangle \end{aligned}$$

$$P(Cavety) = \sum_{Toothache} \sum_{Catch} P(Cavity \wedge Toothache \wedge Catch)$$

Conditional Probability

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

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

$P(A|B)$ is the posterior probability of A given B

$$P(A|B) = \frac{P(A \wedge B)}{P(B)}$$

	toothache		*¬toothache*	
	catch	*¬catch*	*catch*	*¬catch*
cavity	0.108	0.012	0.072	0.008
¬cavity	0.016	0.064	0.144	0.576

$$P(cavity|toothache) = \frac{P(cavity \wedge toothache)}{P(toothache)}$$

$$= \frac{0.108 + 0.012}{0.108 + 0.012 + 0.016 + 0.064} = 0.6$$

Interpretation: After observing $toothache$, the patient is no longer an “average” one, and the prior probability (0.2) of $cavity$ is no longer valid.

$P(Cavity|toothache)$ is calculated by keeping the ratios of the probabilities of the 4 cases unchanged, and normalizing their sum to 1

	toothache		*¬toothache*	
	catch	*¬catch*	*catch*	*¬catch*
cavity	0.108	0.012	0.072	0.008
¬cavity	0.016	0.064	0.144	0.576

$$P(cavity|toothache) = \frac{P(cavity \land toothache)}{P(toothache)}$$

$$= \frac{0.108 + 0.012}{0.108 + 0.012 + 0.016 + 0.064} = 0.6$$

$$P(\neg cavity|toothache) = \frac{P(\neg cavity \land toothache)}{P(toothache)}$$

$$= \frac{0.016 + 0.064}{0.108 + 0.012 + 0.016 + 0.064} = 0.4$$

	toothache		*¬toothache*	
	catch	*¬catch*	*catch*	*¬catch*
cavity	0.108	0.012	0.072	0.008
¬cavity	0.016	0.064	0.144	0.576

$$P(cavity|toothache) = \frac{0.108 + 0.012}{0.108 + 0.012 + 0.016 + 0.064} = 0.6$$

$$P(\neg cavity|toothache) = \frac{0.016 + 0.064}{0.108 + 0.012 + 0.016 + 0.064} = 0.4$$

$$P(Cavity|Toothache) = \big(P(Cavity|Toothache), P(Cavity|Tootache)\big)$$

$$= \alpha P(Cavity \wedge Toothache)$$

$$= \alpha \sum_{Catch} P(Cavity \wedge Toothache \wedge Catch)$$

$$= \alpha[(0.108, 0.016) + (0.012, 0.064)]$$

$$= \alpha(0.12, 0.08) = (0.6, 0.4)$$

Normalization constant

Conditional Probability

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

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

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

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

$$P(Cavety) = \sum_{Toothache}\sum_{Catch} P(Cavity \wedge Toothache \wedge Catch)$$

$$= \sum_{Toothache}\sum_{Catch} P(Cavity|Toothache, Catch)P(Toothache \wedge Catch)$$

Independence

Two random variables A and B are independent if

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

hence $P(A|B) = P(A)$

Two random variables A and B are independent given C, if

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

hence $P(A|B,C) = P(A|C)$

Updating the Belief State

	$toothache$		$\neg toothache$	
	$catch$	$\neg catch$	$catch$	$\neg catch$
$cavity$	0.108	0.012	0.072	0.008
$\neg cavity$	0.016	0.064	0.144	0.576

- Let E be the evidence such that $P(Toothache|E) \ = \ 0.8$
- We want to compute $P(CA \wedge TO \wedge CH|E) = P(CA \wedge CH|TO, E)\ P(TO|E)$
- Since E is not directly related to the cavity or the catch, we consider that CA and CH are independent of E given T

  hence: $P(CA \wedge CH|T, E) \ = \ P(CA \wedge CH|T)$

Updating the Belief State

	$toothache$		$\neg toothache$	
	$catch$	$\neg catch$	$catch$	$\neg catch$
$cavity$	0.108	0.012	0.072	0.008
$\neg cavity$	0.016	0.064	0.144	0.576

- Let E be the evidence such that $P(Toothache|E) = 0.8$
- We want to compute $P(CA \wedge TO \wedge CH|E) = P(CA \wedge CH|TO, E)\, P(TO|E)$
- Since E is not directly related to the cavity or the catch, we consider that CA and CH are independent of E given T

  hence: $P(CA \wedge CH|T, E) = P(CA \wedge CH|T)$

Updating the Belief State

	$toothache$		$\neg toothache$	
	$catch$	$\neg catch$	$catch$	$\neg catch$
$cavity$	~~0.108~~ 0.432	~~0.012~~ 0.048	~~0.072~~ 0.018	~~0.008~~ 0.002
$\neg cavity$	~~0.016~~ 0.064	~~0.064~~ 0.256	~~0.144~~ 0.036	~~0.576~~ 0.114

To get these 4 probabilities
we normalize their sum to 0.8

ch that $P(Toothache|E) = 0.8$

$A \wedge TO \wedge CH|E) = P(C$

To get these 4 probabilities
we normalize their sum to 0.2

- Since E is not directly related to the cavity or th d CH are independent of E given T

  hence: $P(CA \wedge CH|T,E) = P(CA \wedge CH|T)$

Issues

- If a state is described by n propositions, then a belief state contains 2^n states (possibly, some have probability 0)
- **Modeling difficulty:** many numbers must be entered in the first place
- **Computational issue:** memory size and time

Enter Bayesian Networks

	$toothache$		$\neg toothache$	
	$catch$	$\neg catch$	$catch$	$\neg catch$
$cavity$	0.108	0.012	0.072	0.008
$\neg cavity$	0.016	0.064	0.144	0.576

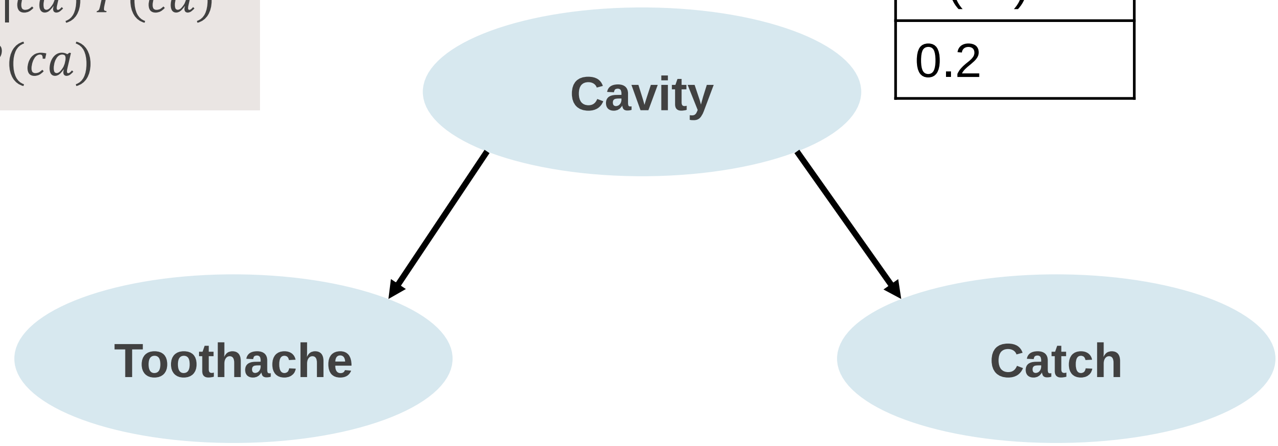
- Toothache and Catch are independent given Cavity (or $\neg Cavity$), but this relation is hidden in the numbers!
- Bayesian networks explicitly represent independence among propositions to reduce the number of probabilities defining a belief state

Bayesian Network

- Notice that Cavity is the “cause” of both Toothache and Catch, and represent the causality links explicitly
- Give the prior probability distribution of Cavity
- Give the conditional probability tables of Toothache and Catch

$$
\begin{aligned}
P(ca \wedge to \wedge ch) &= P(to \wedge ch|ca)\,P(ca) \\
&= P(to|ca)\,P(ch|ca)\,P(ca)
\end{aligned}
$$

P(ca)
0.2



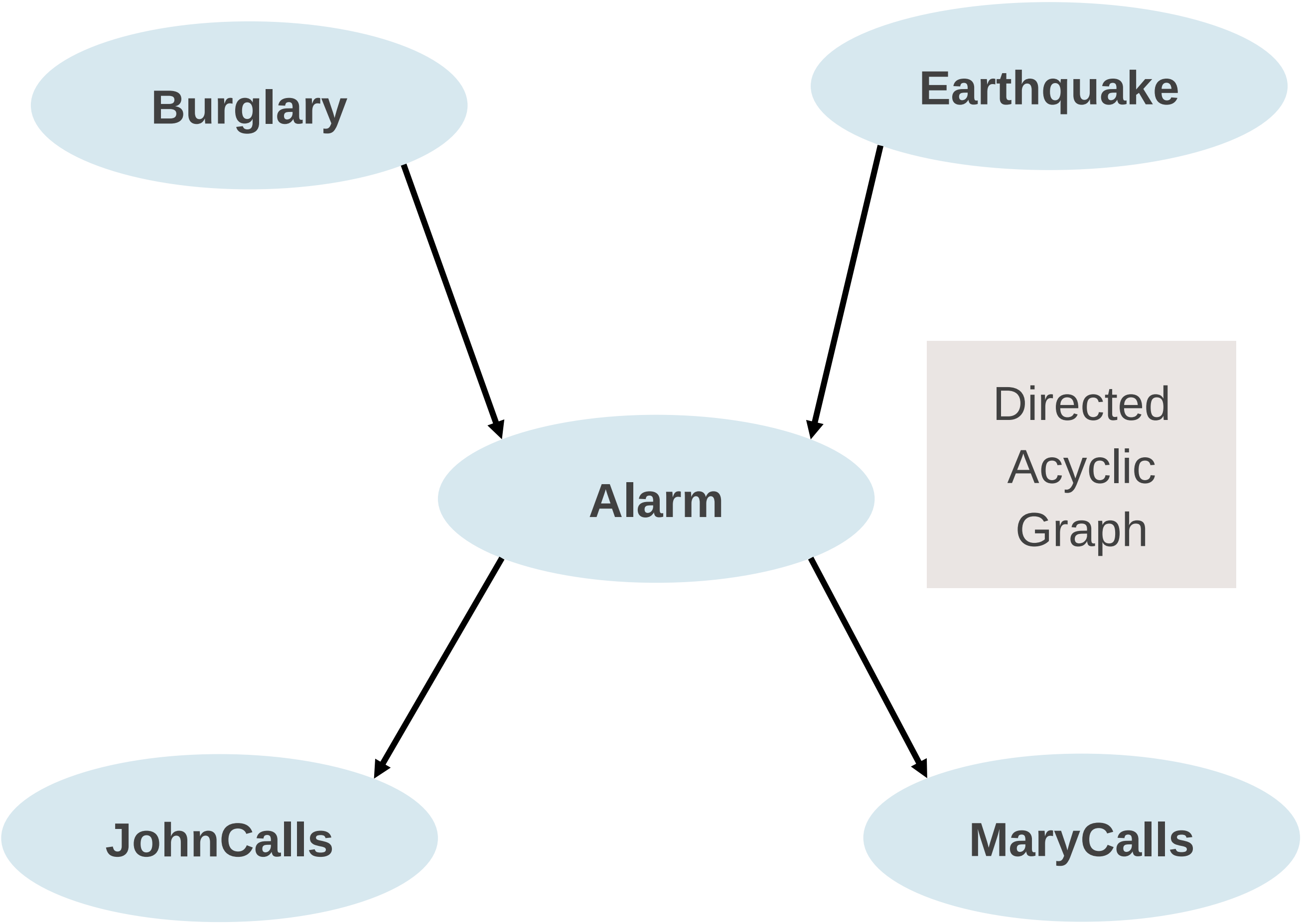
	P(to|ca)
Cavity	0.6
¬Cavity	0.1

	P(ch|ca)
Cavity	0.9
¬Cavity	0.02

5 probabilities instead of 7

A More Complex BN

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



Directed
Acyclic
Graph

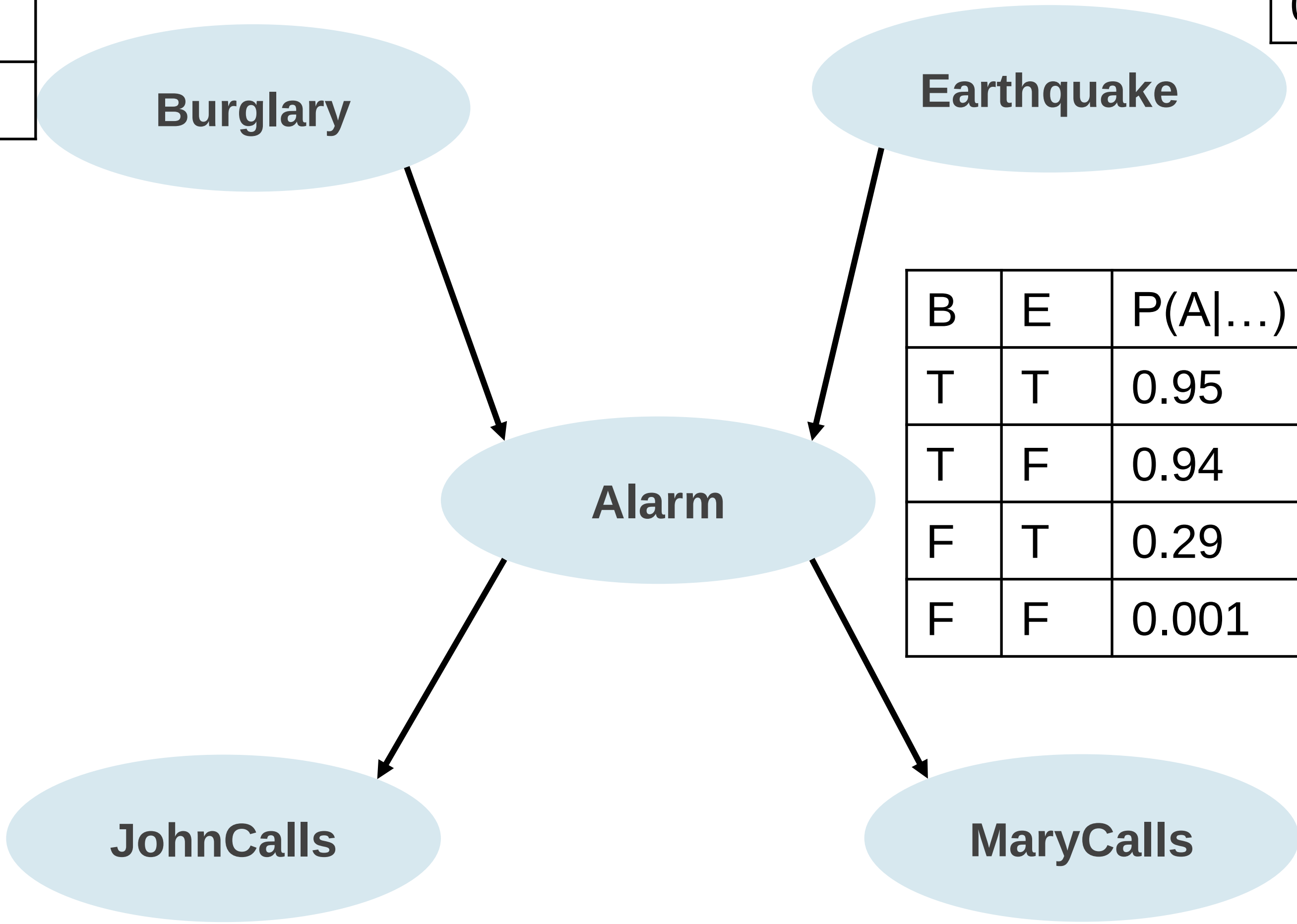
Causes

Effects

A More Complex BN

P(B)
0.001

P(E)
0.002



Size of the CPT for a node with k parents: 2^k

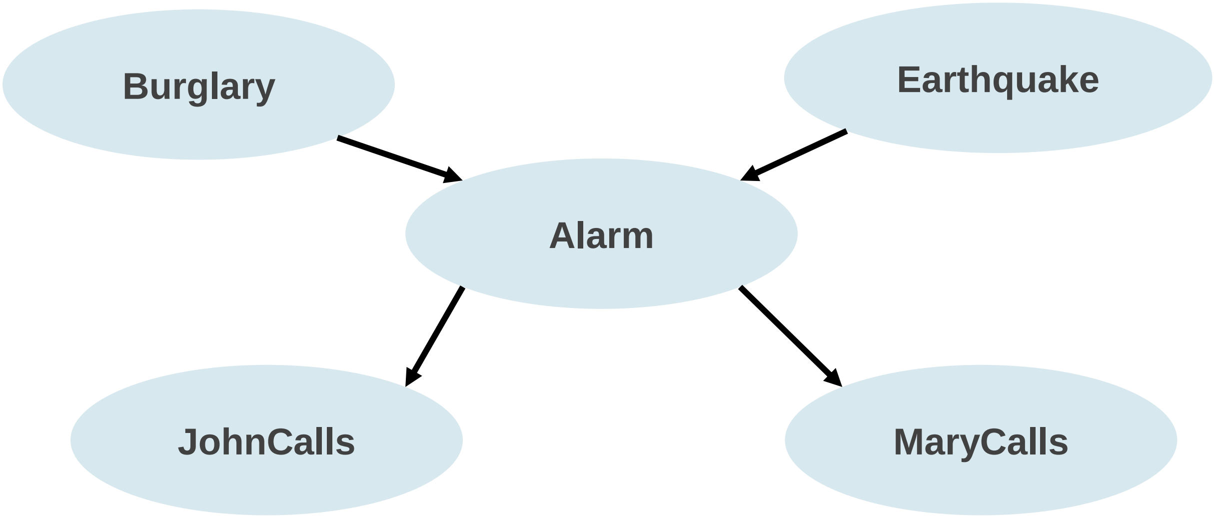
B	E	P(A|…)
T	T	0.95
T	F	0.94
F	T	0.29
F	F	0.001

A	P(J|…)
T	0.9
F	0.05

A	P(M|…)
T	0.7
F	0.01

10 probabilities, instead of 31

What does the BN encode?

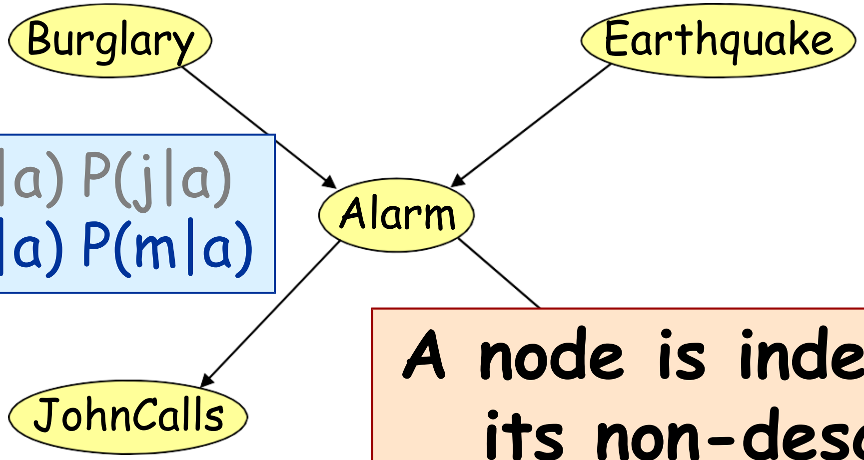


$$P(b \wedge j) \neq P(b)\, P(j)$$
$$P(b \wedge j|a) = P(b|a)\, P(j|a)$$

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

What does the BN encode?



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

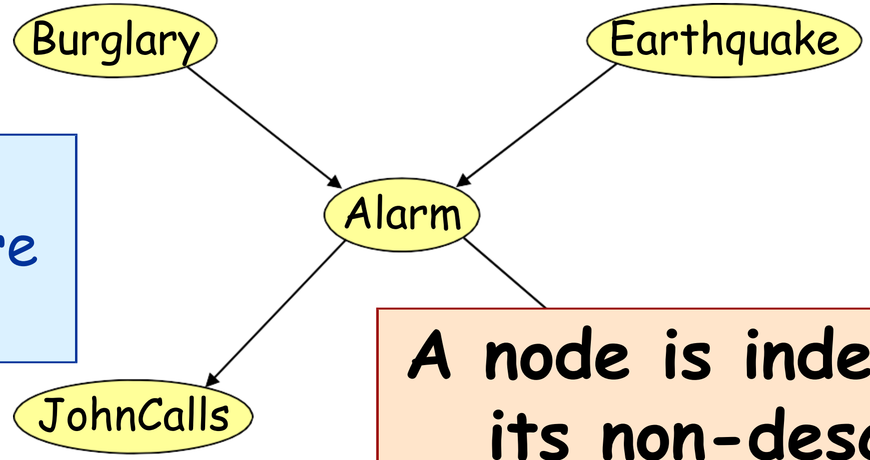
$P(j \wedge m \mid a) = P(j \mid a)\, P(m \mid a)$

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

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

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

What does the BN encode?



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

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

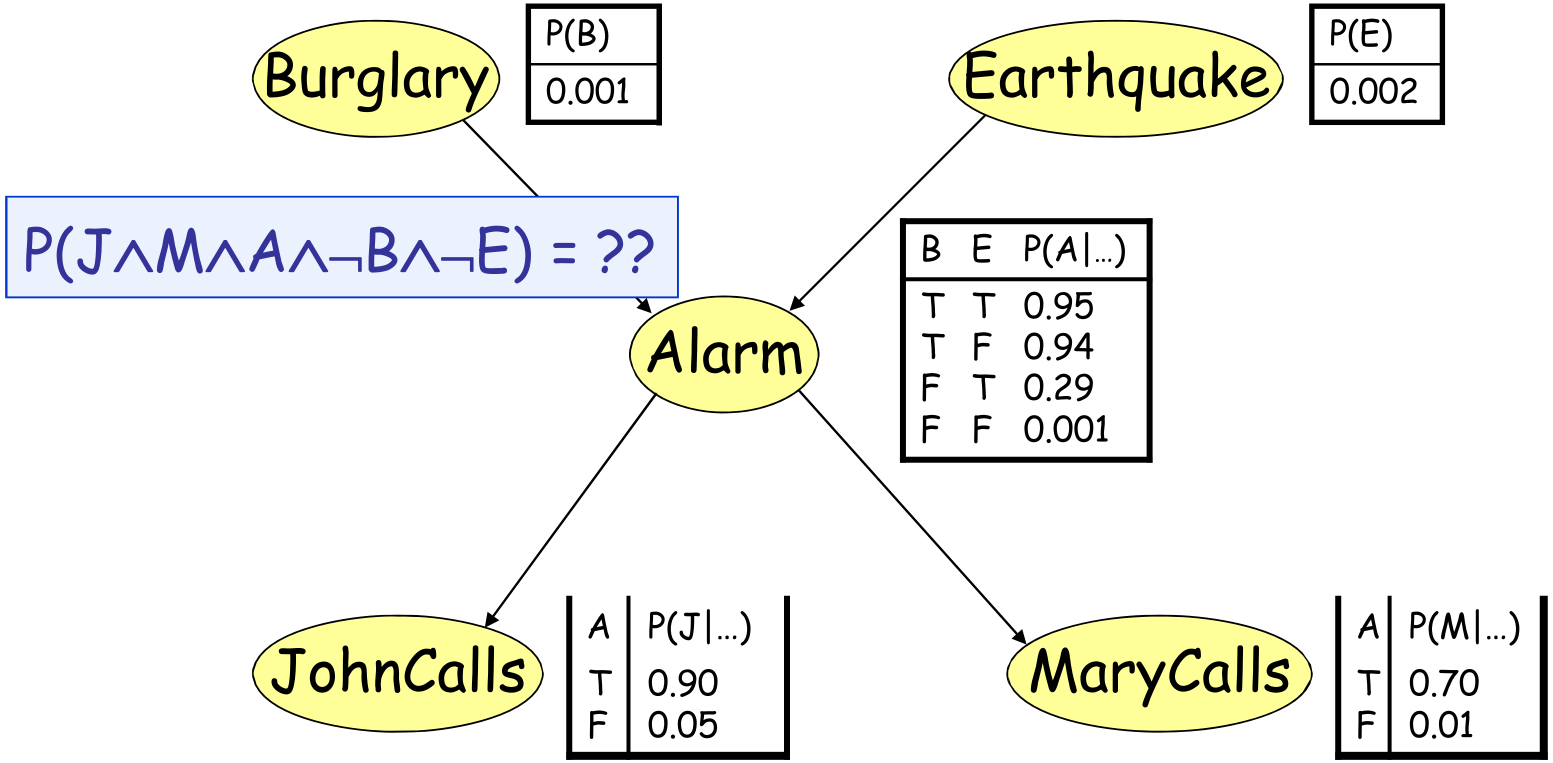
Locally Structured World

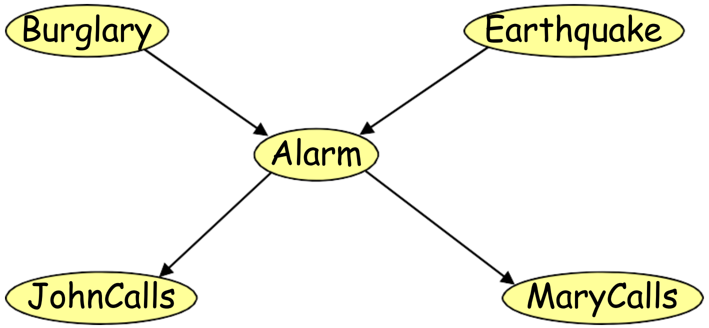
- A world is locally structured (or sparse) if each of its components interacts directly with relatively few other components
- In a sparse world, the CPTs are small and the BN contains much fewer probabilities than the full joint distribution
- If the # of entries in each CPT is bounded by a constant, i.e., O(1), then the # of probabilities in a BN is linear in n - the # of propositions - instead of 2^n for the joint distribution

But does a BN represent a belief state?

In other words, can we compute the full joint distribution of the propositions from it?

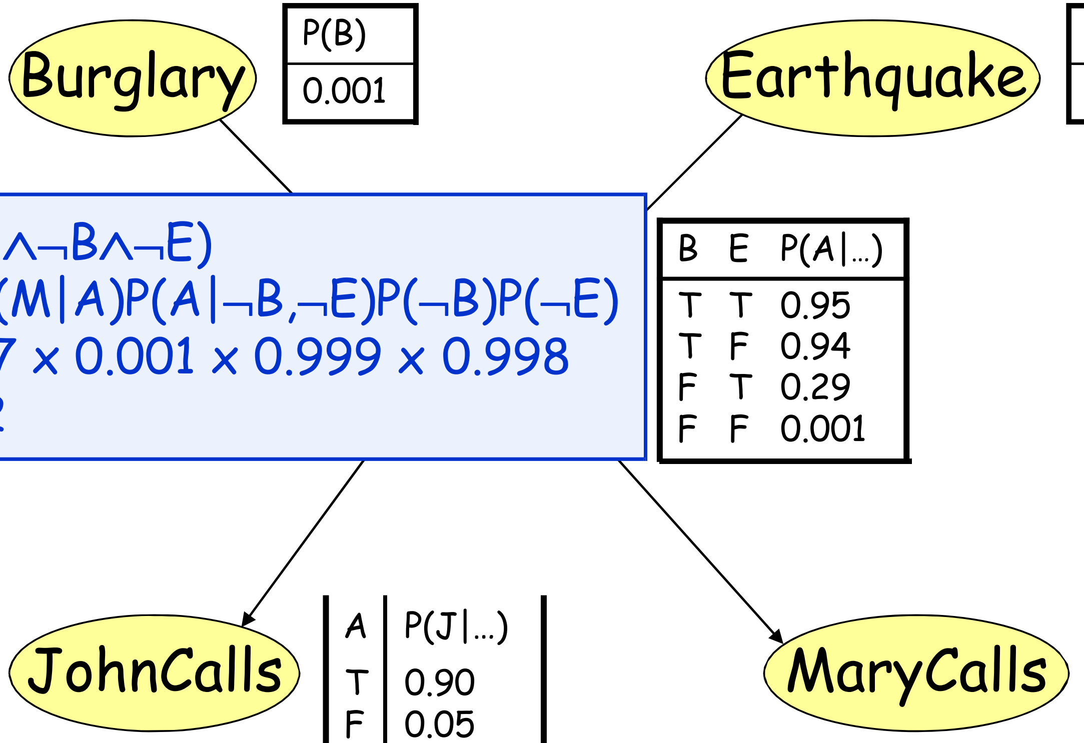
Calculation of Joint Probability





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



$$P(J \land M \land A \land \neg B \land \neg E)$$
$$= P(J|A)P(M|A)P(A|\neg B,\neg E)P(\neg B)P(\neg E)$$
$$= 0.9 \times 0.7 \times 0.001 \times 0.999 \times 0.998$$
$$= 0.00062$$

P(B)
0.001

P(E)
0.002

B	E	P(A|…)
T	T	0.95
T	F	0.94
F	T	0.29
F	F	0.001

A	P(J|…)
T	0.90
F	0.05

A	P(M|…)
T	0.70
F	0.01

Calculation of Joint Probability

Burglary

P(B)
0.001

Earthquake

P(E)
0.002

$P(J \wedge M \wedge A \wedge \neg B \wedge \neg E)$
$= P(J|A)P(M|A)P(A|\neg B,\neg E)P(\neg B)P(\neg E)$
$= 0.9 \times 0.7 \times 0.001 \times 0.999 \times 0.998$
$= 0.00062$

B	E	P(A|…)
T	T	0.95
T	F	0.94
F	T	0.29
F	F	0.001

$$P(x_1 \wedge x_2 \wedge \ldots \wedge x_n) = \prod_{i=1,\ldots,n} P(x_i|\text{parents}(X_i))$$

0.70

F 0.05

F 0.01

→ full joint distribution table

Calculation of Joint Probability

Burglary

P(B)
0.001

Since a BN defines the full joint distribution of a set of propositions, it represents a belief state

$P(J \land M \land A \land \neg B \land \neg E)$
$= P(J|A)P(M|A)P(A|\neg B,\neg E)$
$= 0.9 \times 0.7 \times 0.001 \times 0.999 \times 0.998$
$= 0.00062$

T	F	0.94
F	T	0.29
F	F	0.001

$$P(x_1 \land x_2 \land \ldots \land x_n) = \prod_{i=1,\ldots,n} P(x_i|parents(X_i))$$

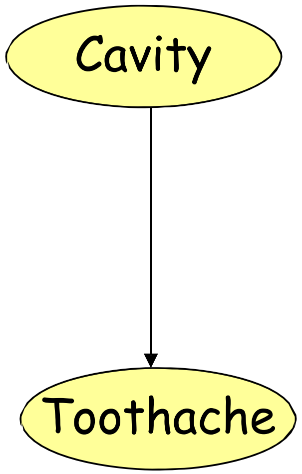
0.70

F 0.05

F 0.01

→ full joint distribution table

Querying the BN



P(C)
0.1

C	P(T|c)
T	0.4
F	0.01111

- The BN gives P(t|c)
- What about P(c|t)?
- P(Cavity|t)
  $= P(Cavity \wedge t)/P(t)$
  $= P(t|Cavity)\ P(Cavity)\ /\ P(t)$
  [Bayes' rule]
- $P(c|t) = \alpha\ P(t|c)\ P(c)$
- Querying a BN is just applying the trivial Bayes' rule on a larger scale

Querying the BN

- New evidence E indicates that JohnCalls with some probability p
- We would like to know the posterior probability of the other beliefs, e.g. P(Burglary|E)
- $P(B|E) = P(B \wedge J|E) + P(B \wedge \neg J|E)$
  $= P(B|J,E)\, P(J|E) + P(B|\neg J,E)\, P(\neg J|E)$
  $= P(B|J)\, P(J|E) + P(B|\neg J)\, P(\neg J|E)$
  $= p\, P(B|J) + (1-p)\, P(B|\neg J)$
- We need to compute P(B|J) and $P(B|\neg J)$

Querying the BN

- $P(b|J) = \alpha\, P(b \wedge J)$

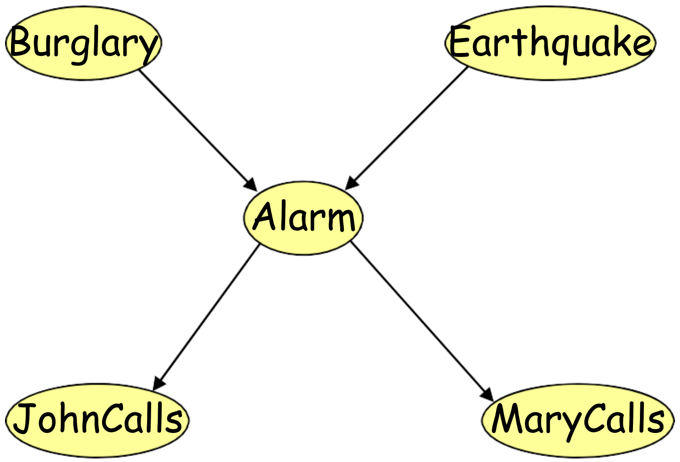
  $= \alpha \sum_m \sum_a \sum_e P(b \wedge J \wedge m \wedge a \wedge e)$ [marginalization]

  $= \alpha \sum_m \sum_a \sum_e P(b)P(e)P(a|b,e)P(J|a)P(m|a)$ [BN]

  $= \alpha\, P(b) \sum_e P(e) \sum_a P(a|b,e)P(J|a) \sum_m P(m|a)$ [re-ordering]

- Depth-first evaluation of $P(b|J)$ leads to computing each of the 4 following products twice:
  $P(J|A)\,P(M|A)$, $P(J|A)\,P(\neg M|A)$, $P(J|\neg A)\,P(M|\neg A)$, $P(J|\neg A)\,P(\neg M|\neg A)$
- Bottom-up (right-to-left) computation + caching – e.g., variable elimination algorithm (see R&N) – avoids such repetition
- For singly connected BN, the computation takes time linear in the total number of CPT entries ($\rightarrow$ time linear in the # propositions if CPT's size is bounded)

Comparison to Classical Logic

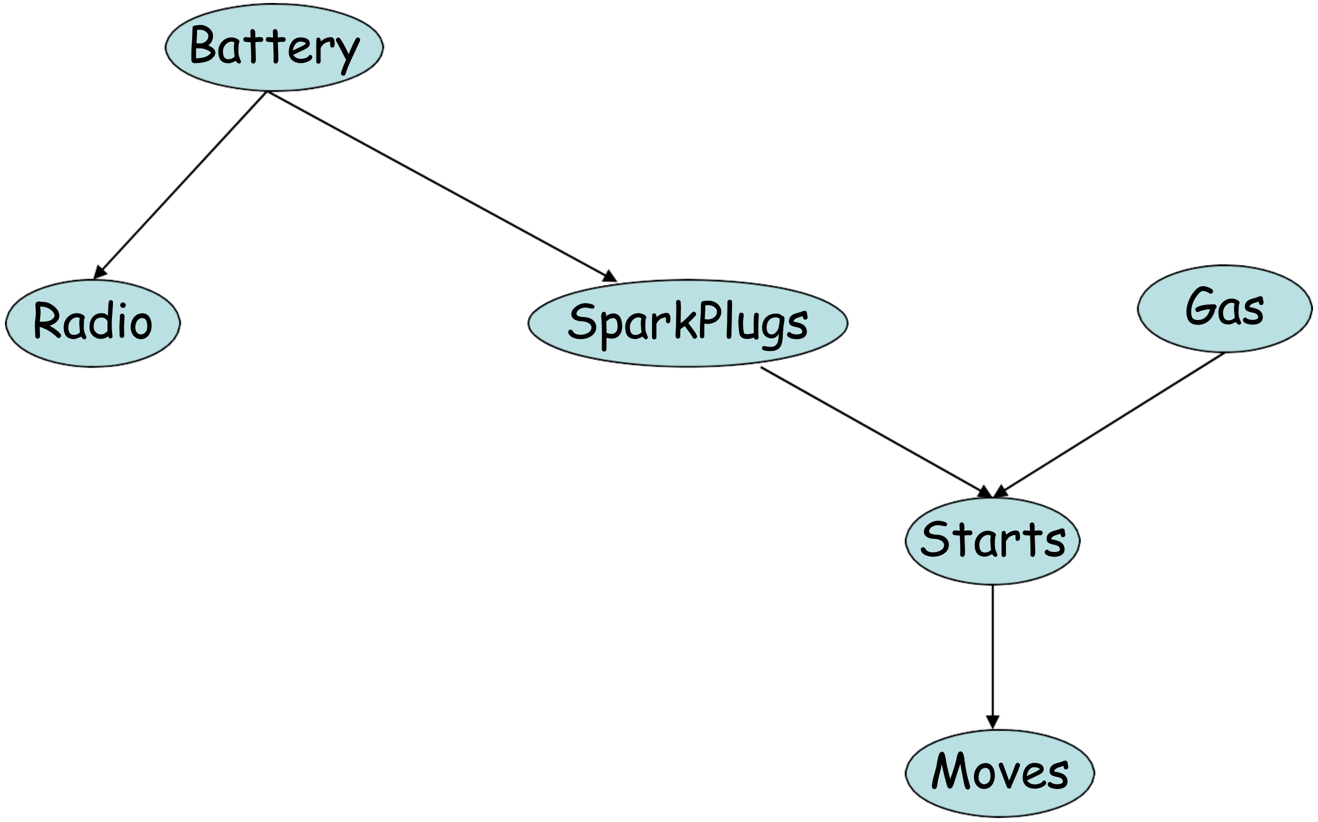


Burglary $\rightarrow$ Alarm
Earthquake $\rightarrow$ Alarm
Alarm $\rightarrow$ JohnCalls
Alarm $\rightarrow$ MaryCalls

If the agent observes $\neg$JohnCalls,
it infers $\neg$Alarm, $\neg$MaryCalls, $\neg$Burglary, and $\neg$Earthquake

If it observes JohnCalls, then it infers nothing

More Complicated
Singly-Connected Belief Net



Some Applications of BN

- Medical diagnosis
- Troubleshooting of hardware/software systems
- Fraud/uncollectible debt detection
- Data mining
- Analysis of genetic sequences
- Data interpretation, computer vision, image understanding