Most materials in this chapter are based on materials from http://aima.cs.berkeley.edu/ and http://ai.berkeley.edu

Chapter 6
Bayes' Nets

COMP 3270
Artificial Intelligence

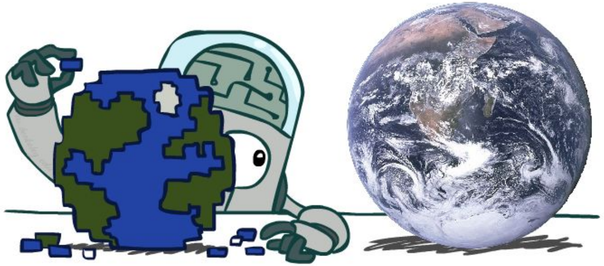
Dirk Schnieders

Outline

- Part A: Representation
- Part B: Independence
- Part C: Inference

Probabilistic Models

- Models describe how (a portion of) the world works
- Models are always simplifications
  - May not account for every variable
  - May not account for all interactions between variables
  - “All models are wrong; but some are useful.”
    – George E. P. Box
- What do we do with probabilistic models?
  - We (or our agents) need to reason about unknown variables, given evidence
  - Example: explanation (diagnostic reasoning)
  - Example: prediction (causal reasoning)
  - Example: value of information



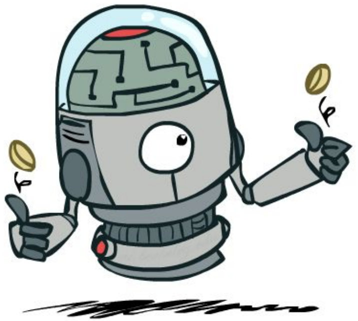
Independence

- Two variables are independent in a joint distribution if

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

$$X \perp\!\!\!\perp Y$$

$$\forall x, y \, P(x, y) = P(x)P(y)$$



  - Says the joint distribution factors into a product of two simple ones
  - Note: Usually, in practice, variables are not independent
- Can use independence as a modeling assumption
  - Simplifying assumption
  - Empirical joint distributions: at best "close" to independent
  - What could we assume for {Weather, Traffic, Cavity}?

Example: Independence check

- Given: Distribution
- Task: Check for independence

T and W are not independent

T	W	$P_1(T,W)$
hot	sun	0.4
hot	rain	0.1
cold	sun	0.2
cold	rain	0.3

T	P(T)
hot	0.5
cold	0.5

W	P(W)
sun	0.6
rain	0.4

T and W are Independent

T	W	$P_2(T,W)$
hot	sun	0.3
hot	rain	0.2
cold	sun	0.3
cold	rain	0.2

Example: Independence

- N fair, independent coin flips

$P(X_1)$

H	0.5
T	0.5

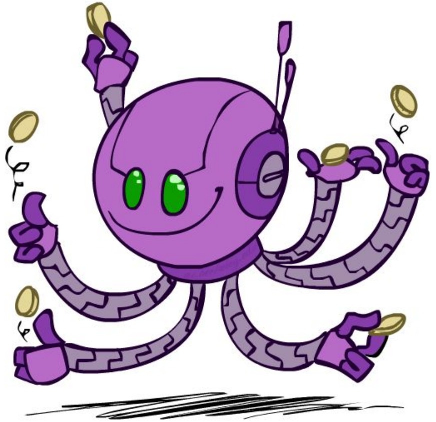
$P(X_2)$

H	0.5
T	0.5

$\dots$

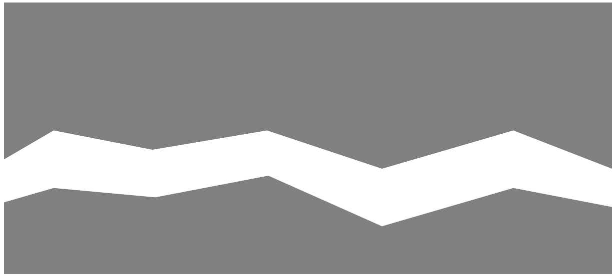
$P(X_n)$

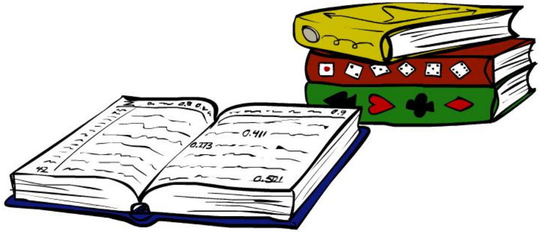
H	0.5
T	0.5



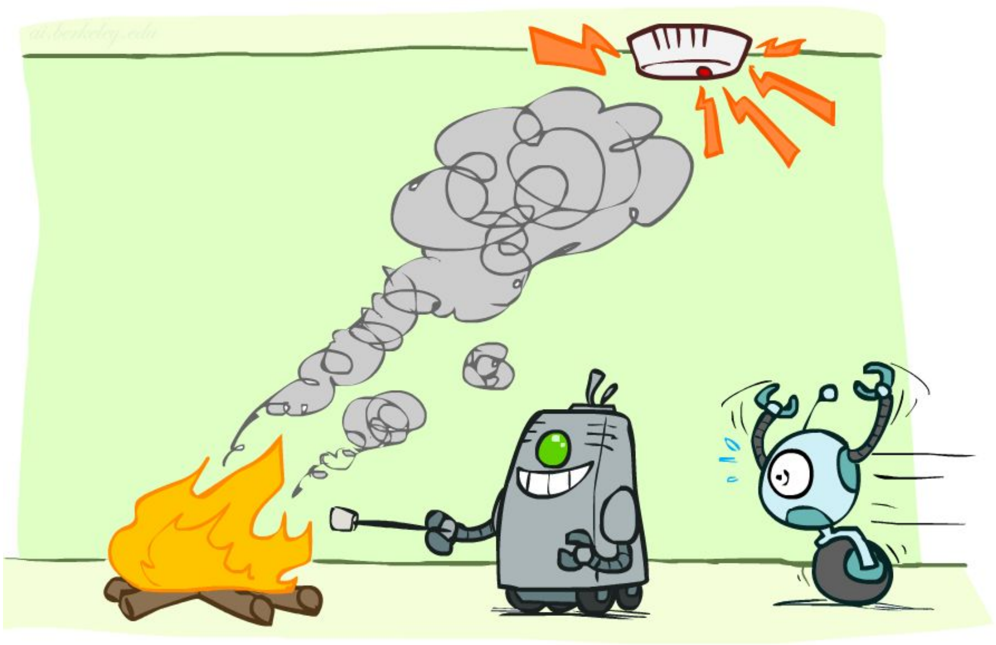
$P(X_1, X_2, \dots X_n)$

2^n

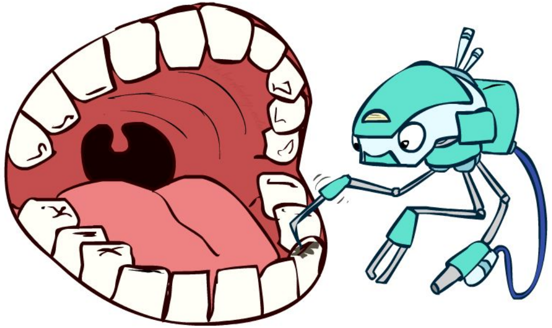




Conditional Independence



Conditional Independence



- P(Toothache, Cavity, Catch)
- If I have a cavity, the probability that the probe catches in it doesn't depend on whether I have a toothache:
  - P(+catch | +toothache, +cavity) = P(+catch | +cavity)
- The same independence holds if I don't have a cavity:
  - P(+catch | +toothache, -cavity) = P(+catch| -cavity)
- Catch is conditionally independent of Toothache given Cavity:
  - P(Catch | Toothache, Cavity) = P(Catch | Cavity)
- Equivalent statements:
  - P(Toothache | Catch , Cavity) = P(Toothache | Cavity)
  - P(Toothache, Catch | Cavity) = P(Toothache | Cavity) P(Catch | Cavity)
  - One can be derived from the other easily

Conditional Independence

- Unconditional (absolute) independence very rare (why?)
- Conditional independence is our most basic and robust form of knowledge about uncertain environments

$$X \perp\!\!\!\perp Y | Z$$

- X is conditionally independent of Y given Z iff

$$\forall x, y, z : P(x, y|z) = P(x|z)P(y|z)$$

- or, equivalently, iff

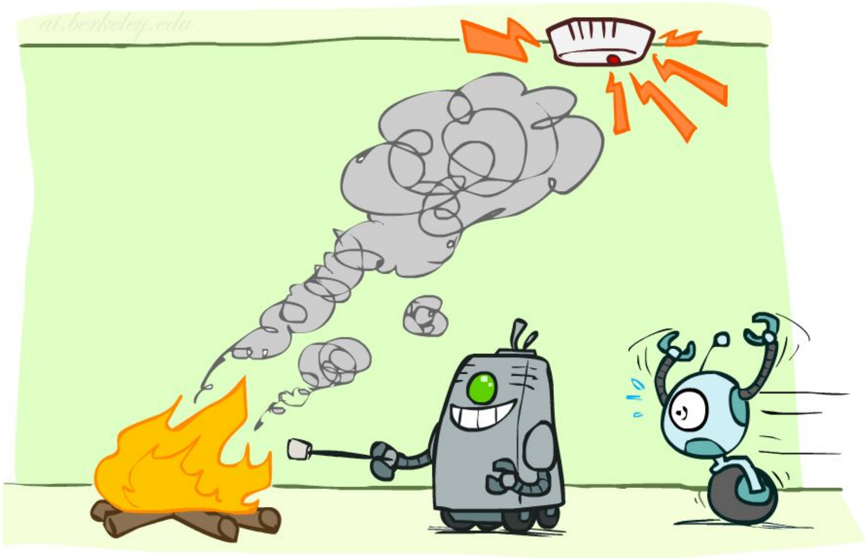
$$\forall x, y, z : P(x|z, y) = P(x|z)$$

Quiz: Conditional Independence

- What about this domain
  - Traffic (T)
  - Umbrella (U)
  - Raining (R)

Quiz: Conditional Independence

- What about this domain
  - Fire (F)
  - Smoke (S)
  - Alarm (A)



Conditional Independence and the Chain Rule

- Chain rule: $P(X_1, X_2, \ldots X_n) = P(X_1)P(X_2|X_1)P(X_3|X_1, X_2)\ldots$

- Trivial decomposition:
$$P(\text{Traffic}, \text{Rain}, \text{Umbrella}) =$$
$$P(\text{Rain})P(\text{Traffic}|\text{Rain})P(\text{Umbrella}|\text{Rain}, \text{Traffic})$$

- With assumption of conditional independence:
$$P(\text{Traffic}, \text{Rain}, \text{Umbrella}) =$$
$$P(\text{Rain})P(\text{Traffic}|\text{Rain})P(\text{Umbrella}|\text{Rain})$$

- Bayes' nets / graphical models help us express conditional independence assumptions

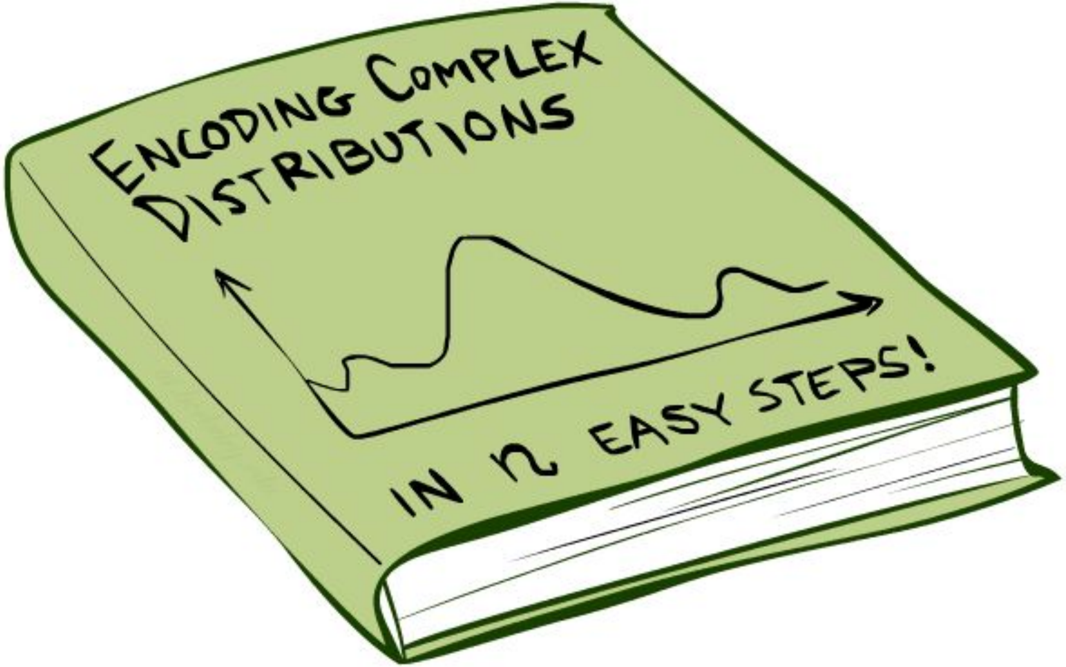
Ghostbusters Chain Rule

- Each sensor depends only on where the ghost is
- That means, the two sensors are conditionally independent, given the ghost position
  - T: Top square is red
  - B: Bottom square is red
  - G: Ghost is in the top
- Given:
  - P(+g) = 0.5
  - P(-g) = 0.5
  - P(+t |+g) = 0.8
  - P(+t |-g) = 0.4
  - P(+b|+g) = 0.4
  - P(+b|-g) = 0.8

P(T,B,G) = P(G) P(T|G) P(B|G)

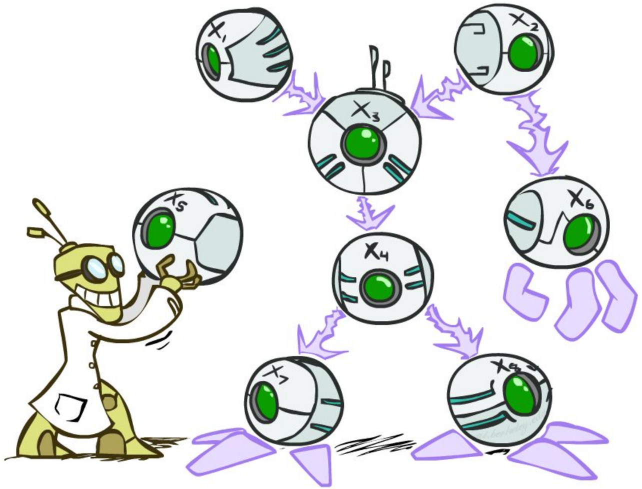
T	B	G	P(T,B,G)
+t	+b	+g	0.16
+t	+b	-g	0.16
+t	-b	+g	0.24
+t	-b	-g	0.04
-t	+b	+g	0.04
-t	+b	-g	0.24
-t	-b	+g	0.06
-t	-b	-g	0.06

Bayes'Nets: Big Picture

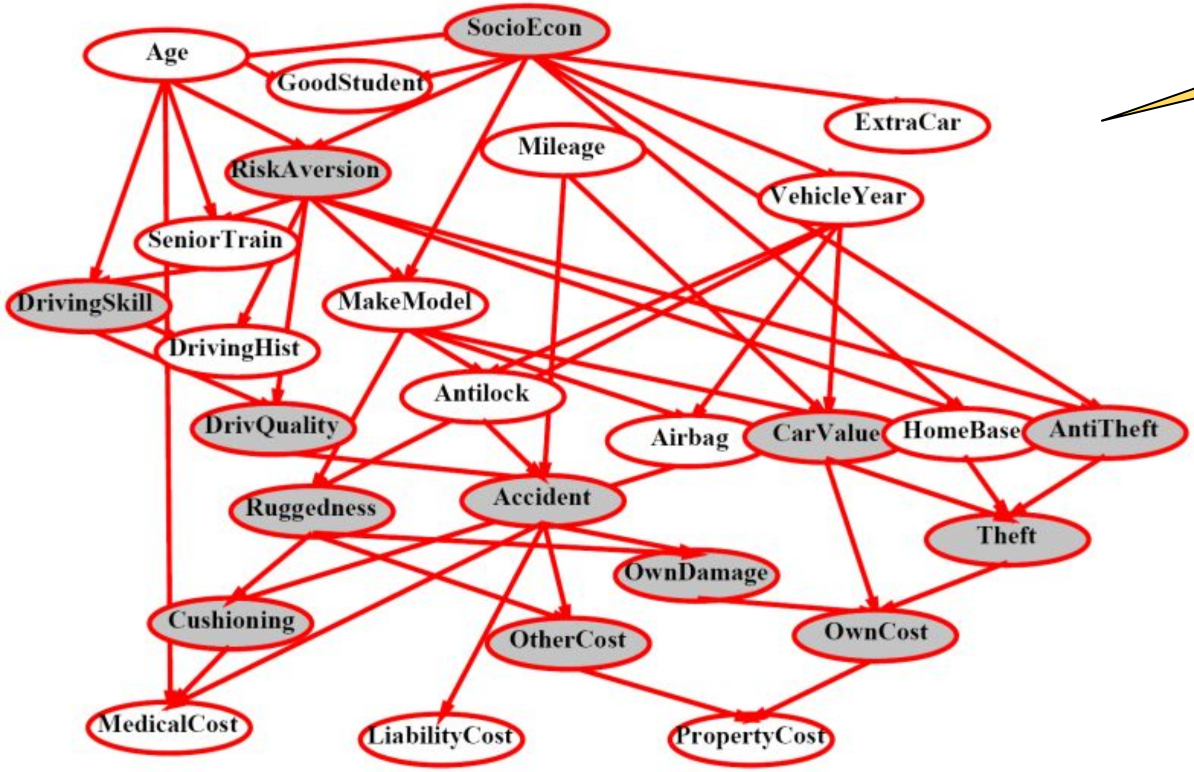


Bayes' Nets: Big Picture

- Two problems with using full joint distribution tables as our probabilistic models:
  - Unless there are only a few variables, the joint is WAY too big to represent explicitly
  - Hard to learn (estimate) anything empirically about more than a few variables at a time
- Bayes' nets: a technique for describing complex joint distributions (models) using simple, local distributions (conditional probabilities)
  - More properly called graphical models
  - We describe how variables locally interact
  - Local interactions chain together to give global, indirect interactions



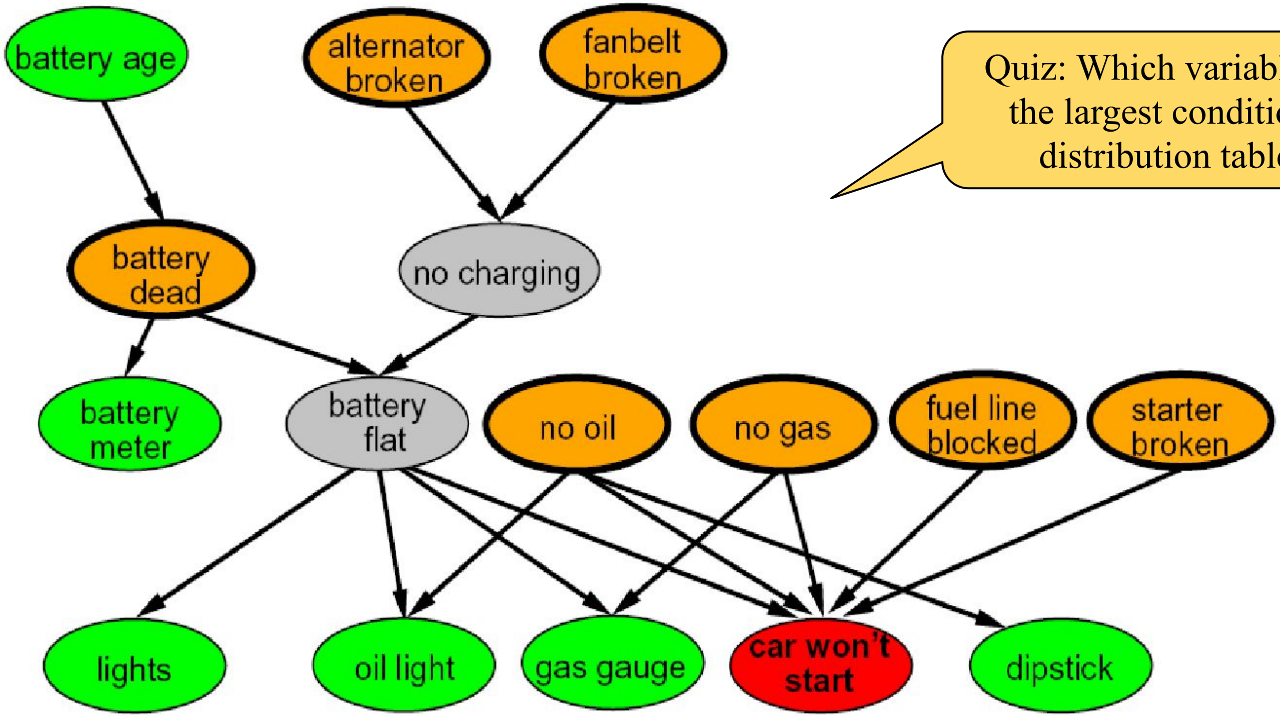
Example Bayes' Net: Insurance



27 variables

Quiz: How many entries in the joint distribution?

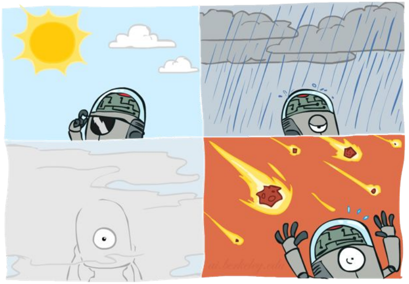
Example Bayes' Net: Car

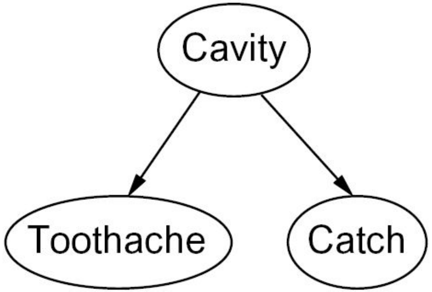


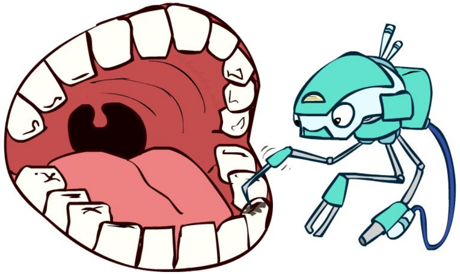
Quiz: Which variable has the largest conditional distribution table?

Graphical Model Notation

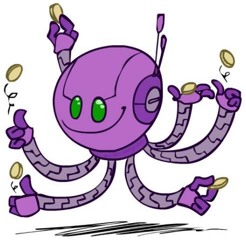
- Nodes: variables (with domains)
  - Can be assigned (observed) or unassigned (unobserved)
- Arcs: interactions
  - Similar to CSP constraints
  - Indicate "direct influence" between variables
  - Formally: encode conditional independence (more later)
- For now: imagine that arrows mean direct causation (in general, they don't!)







Example: Coin Flips

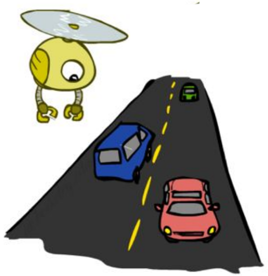


- N independent coin flips

$$X_1 \qquad X_2 \qquad \cdots \qquad X_n$$

- No interactions between variables: absolute independence

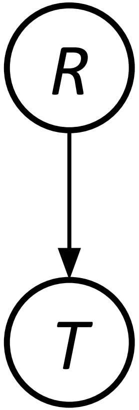
Example: Traffic



- Variables:
    - R: It rains
    - T: There is traffic

- Model 1: independence

- Model 2: rain causes traffic



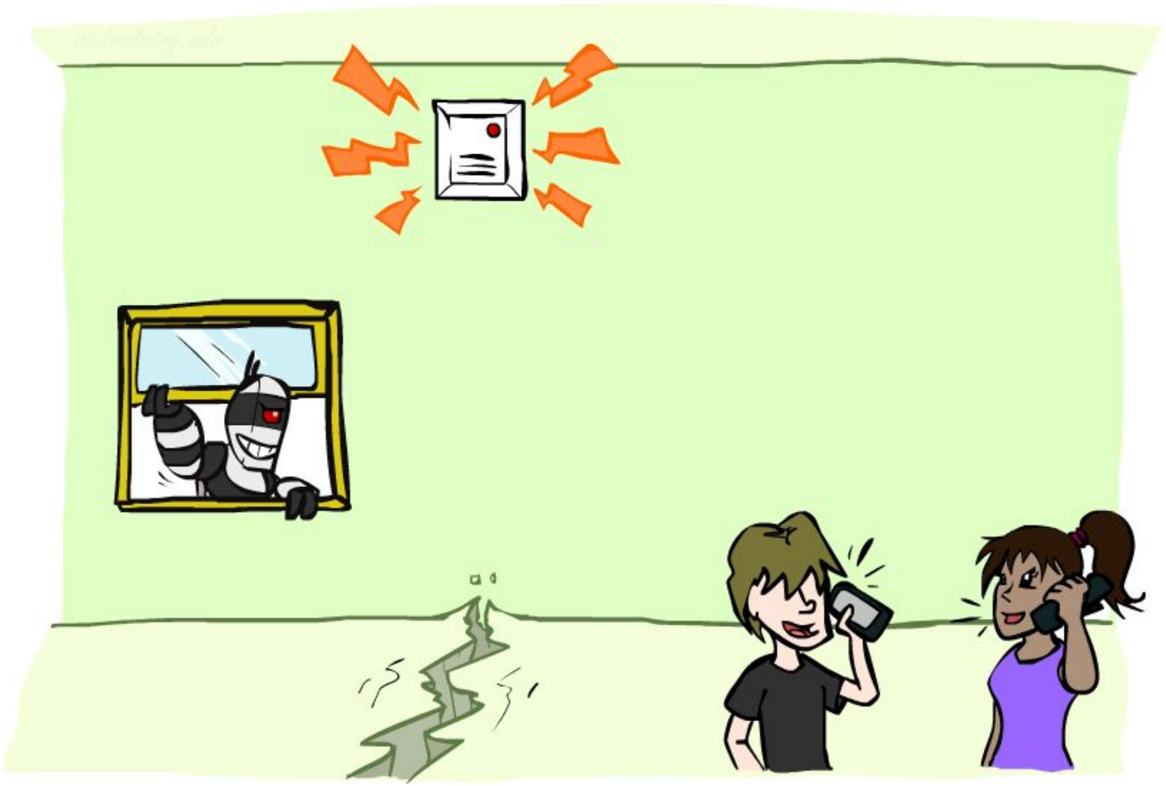
- Quiz: Why is an agent using model 2 better?

Quiz: Traffic II

- Build a causal graphical model!
- Variables
  - T: Traffic
  - R: It rains
  - L: Low pressure
  - D: Roof drips
  - B: Ballgame
  - C: Cavity

Quiz: Alarm Network

- Build a causal graphical model!
- Variables
  - B: Burglary
  - A: Alarm goes off
  - M: Mary calls
  - J: John calls
  - E: Earthquake!



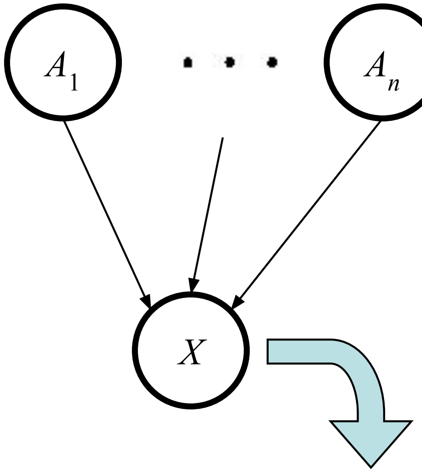
Bayes' Net Semantics

Bayes' Net Semantics

- A set of nodes, one per variable X
- A directed, acyclic graph
- A conditional distribution for each node
  - A collection of distributions over X, one for each combination of parents' values

$$P(X|a_1 \dots a_n)$$

  - CPT: conditional probability table



$$P(X|A_1 \dots A_n)$$

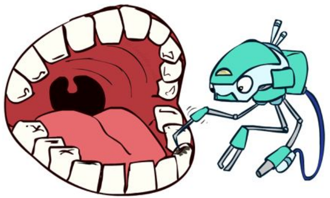
A Bayes net = Topology (graph) + Local Conditional Probabilities

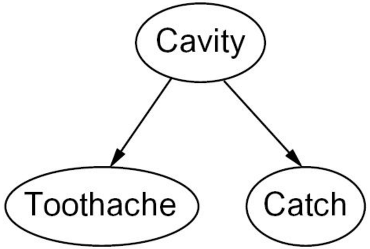
Probabilities in BNs

- Bayes' nets implicitly encode joint distributions
  - As a product of local conditional distributions
  - To see what probability a BN gives to a full assignment, multiply all the relevant conditionals together

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

  - Quiz:





$P(+cavity,\ +catch,\ -toothache)$

Probabilities in BNs

- Why are we guaranteed that setting

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

  results in a proper joint distribution?

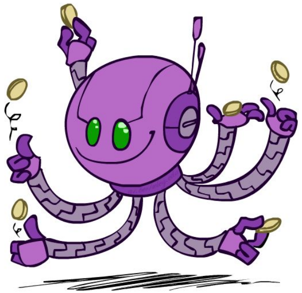
- Chain rule (valid for all distributions): $P(x_1, x_2, \ldots x_n) = \prod_{i=1}^{n} P(x_i | x_1 \ldots x_{i-1})$

- Assume conditional independencies: $P(x_i | x_1, \ldots x_{i-1}) = P(x_i | \mathit{parents}(X_i))$

  - Consequence: $P(x_1, x_2, \ldots x_n) = \prod_{i=1}^{n} P(x_i | \mathit{parents}(X_i))$

- Not every BN can represent every joint distribution
  - The topology enforces certain conditional independencies

Quiz: Coin Flips



X_1 X_2 $\cdots$ X_n

$P(X_1)$

h	0.5
t	0.5

$P(X_2)$

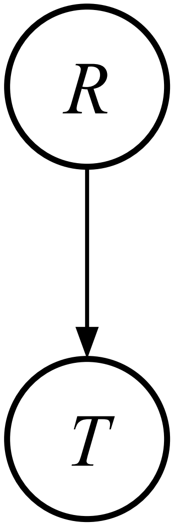
h	0.5
t	0.5

$P(X_n)$

h	0.5
t	0.5

$$P(h, h, t, h) =$$

Quiz: Traffic



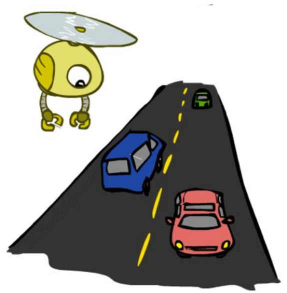
$P(R)$

+r	1/4
-r	3/4

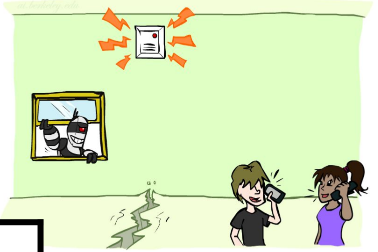
$P(T|R)$

+r	+t	3/4
	-t	1/4
-r	+t	1/2
	-t	1/2

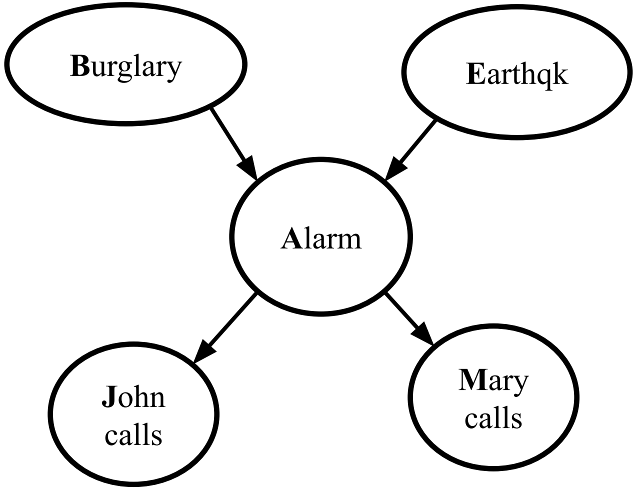
$$P(+r, -t) =$$



Example: Alarm Network



B	P(B)
+b	0.001
-b	0.999



E	P(E)
+e	0.002
-e	0.998

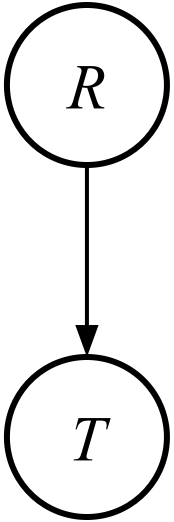
A	J	P(J|A)
+a	+j	0.9
+a	-j	0.1
-a	+j	0.05
-a	-j	0.95

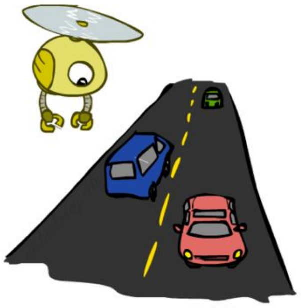
A	M	P(M|A)
+a	+m	0.7
+a	-m	0.3
-a	+m	0.01
-a	-m	0.99

B	E	A	P(A|B,E)
+b	+e	+a	0.95
+b	+e	-a	0.05
+b	-e	+a	0.94
+b	-e	-a	0.06
-b	+e	+a	0.29
-b	+e	-a	0.71
-b	-e	+a	0.001
-b	-e	-a	0.999

Example: Traffic

- Causal direction





$P(R)$

+r	1/4
-r	3/4

$P(T|R)$

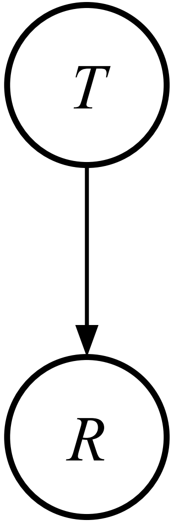
+r	+t	3/4
	-t	1/4
-r	+t	1/2
	-t	1/2

$P(T, R)$

+r	+t	3/16
+r	-t	1/16
-r	+t	6/16
-r	-t	6/16

Example: Reverse Traffic

- Reverse causality?



$P(T)$

+t	9/16
-t	7/16

$P(R|T)$

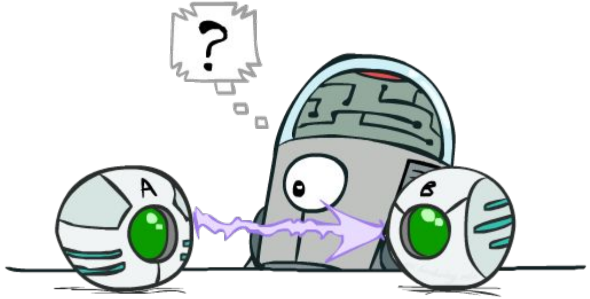
+t	+r	1/3
	-r	2/3

-t	+r	1/7
	-r	6/7

$P(T,R)$

+r	+t	3/16
+r	-t	1/16
-r	+t	6/16
-r	-t	6/16

Causality?



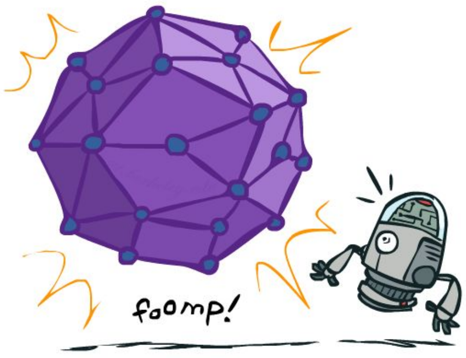
- When Bayes' nets reflect the true causal patterns
  - Often simpler (nodes have fewer parents)
  - Often easier to think about
  - Often easier to elicit from experts
- BNs need not actually be causal
  - Sometimes no causal net exists over the domain (especially if variables are missing)
  - E.g. consider the variables Traffic and Drips
  - End up with arrows that reflect correlation, not causation
- What do the arrows really mean?
  - Topology may happen to encode causal structure
  - Topology really encodes conditional independence

$$P(x_i|x_1,\ldots x_{i-1}) = P(x_i|\mathit{parents}(X_i))$$

Size of a Bayes' Net

- A: How many entries (rows) are in a joint distribution table for a BN with N Boolean variables?
  - 2^N
- B: How many entries (rows) in all (C)PT tables for a BN with N Boolean variables where each variable has at most k parents?
  - $< N2^{k+1}$
- Both A and B give you the power to calculate

  $P(X_1, X_2, \ldots, X_n)$
- BNs
  - Huge space savings!
  - Easier to elicit local CPTs
  - Faster to answer queries

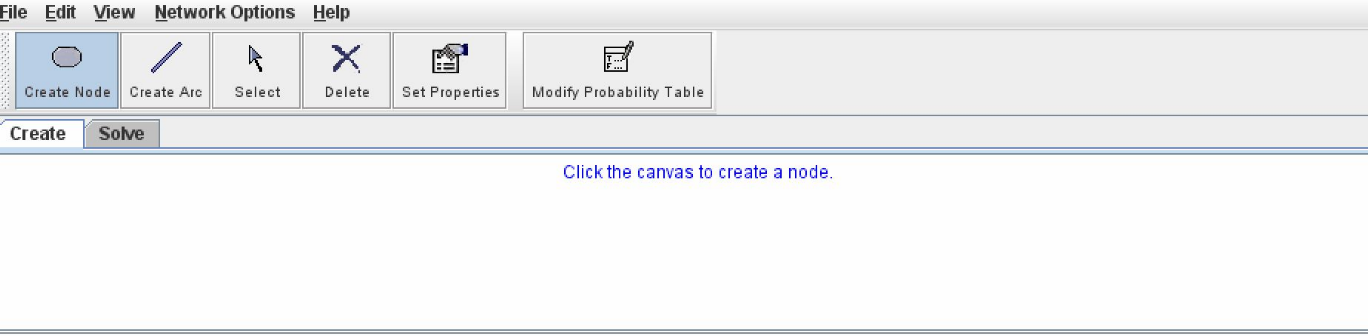


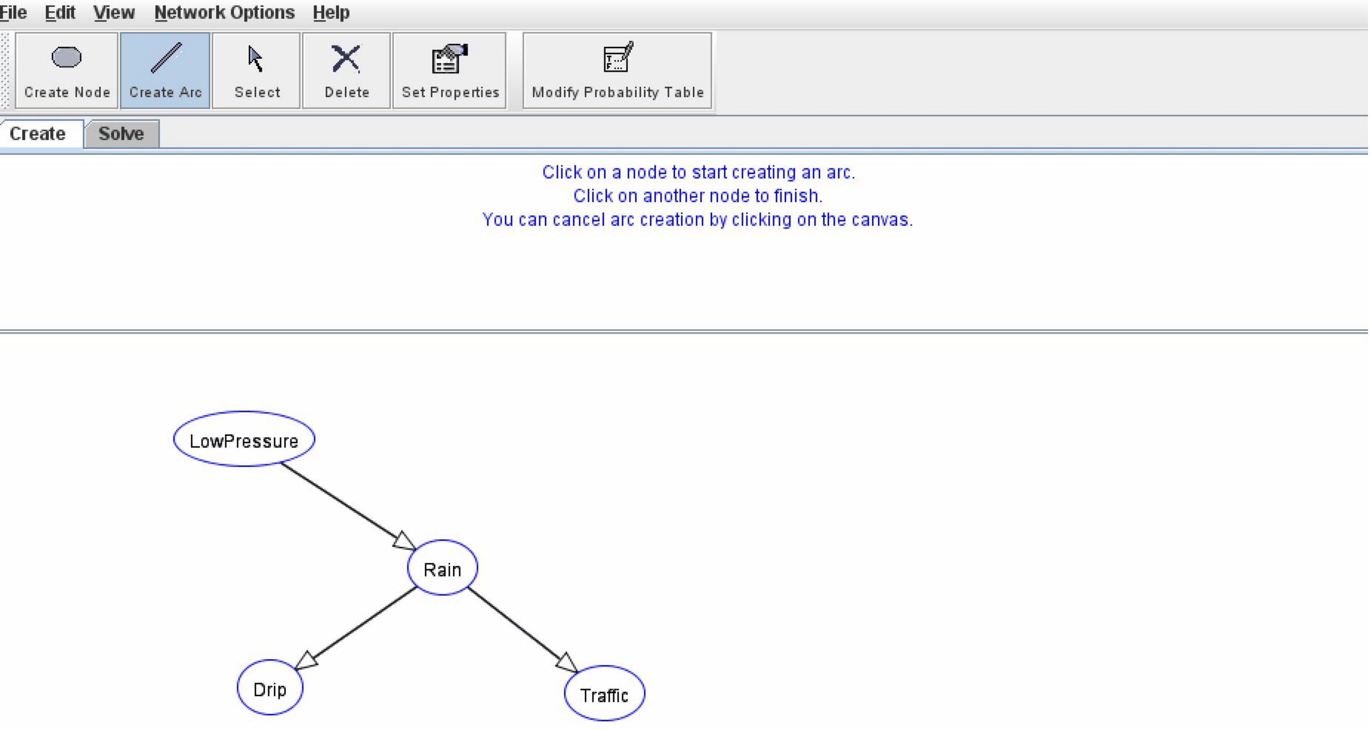
Quiz: Suppose we have 30 nodes each with 5 parents.

- How many rows for A?
- How many rows for B?

Demo

Belief and Decision Networks
http://aispace.org/bayes/


File
Edit
View
Network Options
Help
Create Node
Create Arc
Select
Delete
Set Properties
Modify Probability Table
Create
Solve
Click the canvas to create a node.

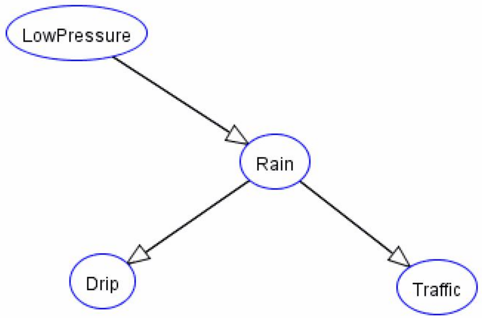

File
Edit
View
Network Options
Help
Create Node
Create Arc
Select
Delete
Set Properties
Modify Probability Table
Create
Solve
Click on a node to start creating an arc.
Click on another node to finish.
You can cancel arc creation by clicking on the canvas.
LowPressure
Rain
Drip
Traffic

File Edit View Network Options Help

Create Node | Create Arc | Select | Delete | Set Properties | Modify Probability Table

Create | Solve

Click on an entity to select or drag the mouse to select multiple entities.



Outline

- Part A: Representation
- Part B: Independence
- Part C: Inference

Probability Recap

- Conditional probability $P(x|y) = \dfrac{P(x,y)}{P(y)}$

- Product rule $P(x,y) = P(x|y)P(y)$

- Chain rule
$$\begin{aligned} P(X_1, X_2, \ldots X_n) &= P(X_1)P(X_2|X_1)P(X_3|X_1,X_2)\ldots \\ &= \prod_{i=1}^{n} P(X_i|X_1,\ldots,X_{i-1}) \end{aligned}$$

- X, Y independent if and only if: $\forall x,y : P(x,y) = P(x)P(y)$

- X and Y are conditionally independent given Z if and only if:

$$\forall x,y,z : P(x,y|z) = P(x|z)P(y|z)$$

$$X \perp\!\!\!\perp Y | Z$$

Conditional Independence Recap

- X and Y are independent if

$$\forall x, y \;\; P(x, y) = P(x)P(y) \longrightarrow X \perp\!\!\!\perp Y$$

- X and Y are conditionally independent given Z

$$\forall x, y, z \;\; P(x, y|z) = P(x|z)P(y|z) \longrightarrow X \perp\!\!\!\perp Y|Z$$

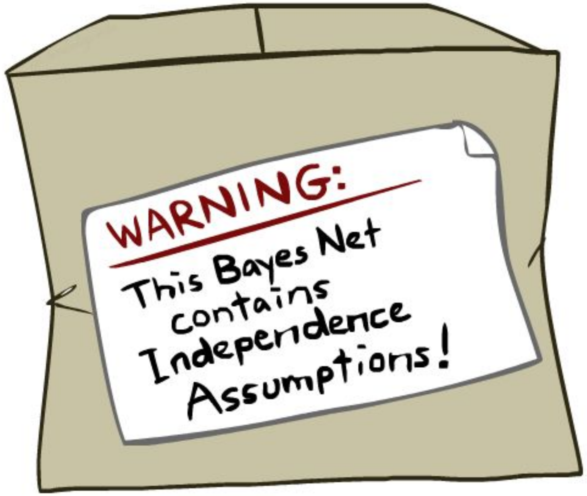
- (Conditional) independence is a property of a distribution

Bayes Nets: Assumptions

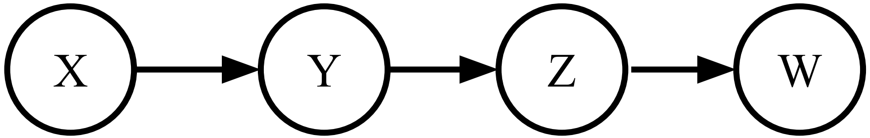
- Assumptions we are required to make to define the Bayes net when given the graph

$$P(x_i|x_1 \cdots x_{i-1}) = P(x_i|parents(X_i))$$

- Beyond above "chain rule → Bayes net" conditional independence assumptions
  - Often additional conditional independencies
  - They can be read off the graph
- Important for modeling: understand assumptions made when choosing a Bayes net graph



Quiz - Conditional Independence Assumptions

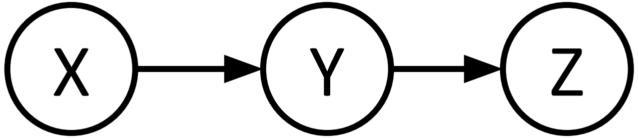


- a) Write down the conditional independence assumptions directly from simplifications in chain rule

- b) Write down additional implied conditional independence assumptions, if any

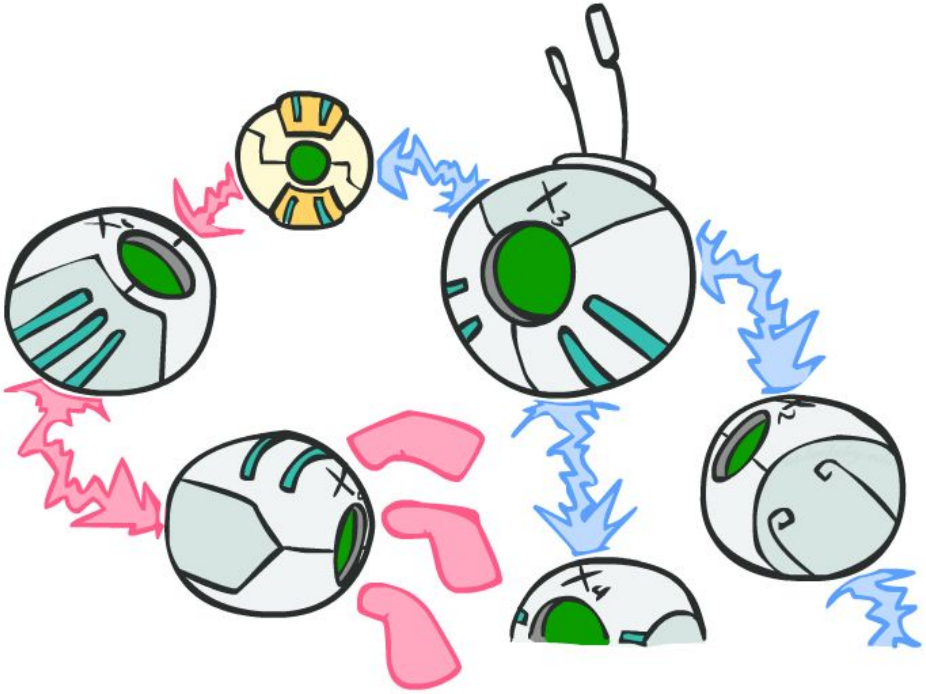
Independence in a BN

- Important question about a BN:
  - Are two nodes independent given certain evidence?
  - If yes, can prove using algebra (tedious in general)
  - If no, can prove with a counter example
  - Example:



- Question: are X and Z necessarily independent?
  - Answer: no. Example: low pressure causes rain, which causes traffic.
  - X can influence Z, Z can influence X (via Y)
  - Quiz: Could they be independent?

D-separation: Outline



D-separation: Outline

- Study independence properties for triples
- Analyze complex cases in terms of member triples
- D-separation: a condition / algorithm for answering such queries

Causal Chain

- This configuration is a “causal chain”

X: Low pressure

Y: Rain

Z: Traffic

$$P(x, y, z) = P(x)P(y|x)P(z|y)$$

- Guaranteed X independent of Z ?
  - No!
  - One example set of CPTs for which X is not independent of Z is sufficient to show this independence is not guaranteed.
  - Example:
    - Low pressure causes rain causes traffic, high pressure causes no rain causes no traffic
    - In numbers:
    - P(+y | +x) = 1, P(-y | - x) = 1, P(+z | +y) = 1, P(-z | -y) = 1

Causal Chain

- This configuration is a "causal chain"

X: Low pressure

Y: Rain

Z: Traffic

$$P(x, y, z) = P(x)P(y|x)P(z|y)$$

- Guaranteed X independent of Z given Y?

$$P(z|x, y) = \frac{P(x, y, z)}{P(x, y)}$$

$$= \frac{P(x)P(y|x)P(z|y)}{P(x)P(y|x)}$$

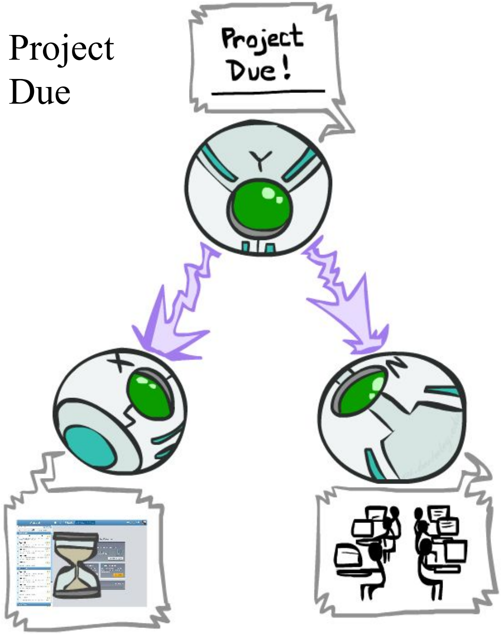
$$= P(z|y)$$

Yes!
Evidence along the chain "blocks" the influence

Common Cause

- This configuration is a "common cause"

Y: Project Due



X: Forum Busy

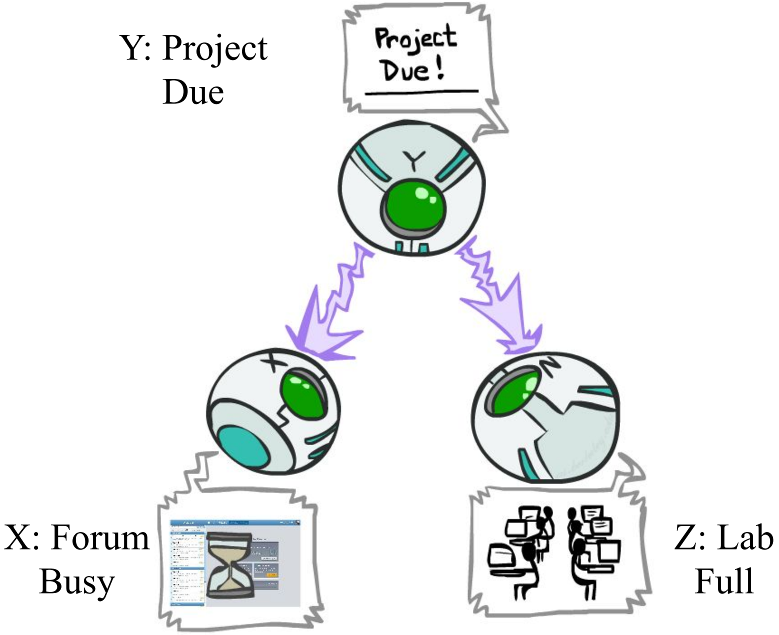
Z: Lab Full

$$P(x, y, z) = P(y)P(x|y)P(z|y)$$

- Guaranteed X independent of Z ?
  - No!
  - One example set of CPTs for which X is not independent of Z is sufficient to show this independence is not guaranteed
  - Example:
    - Project due causes both forums busy and lab full
    - In numbers:
      - P(+x |+y) = 1, P(-x|-y) = 1, P(+z|+y) = 1, P(-z|-y) = 1

Common Cause

- This configuration is a "common cause"



$$P(x, y, z) = P(y)P(x|y)P(z|y)$$

- Guaranteed X and Z independent given Y?

$$P(z|x, y) = \frac{P(x, y, z)}{P(x, y)}$$

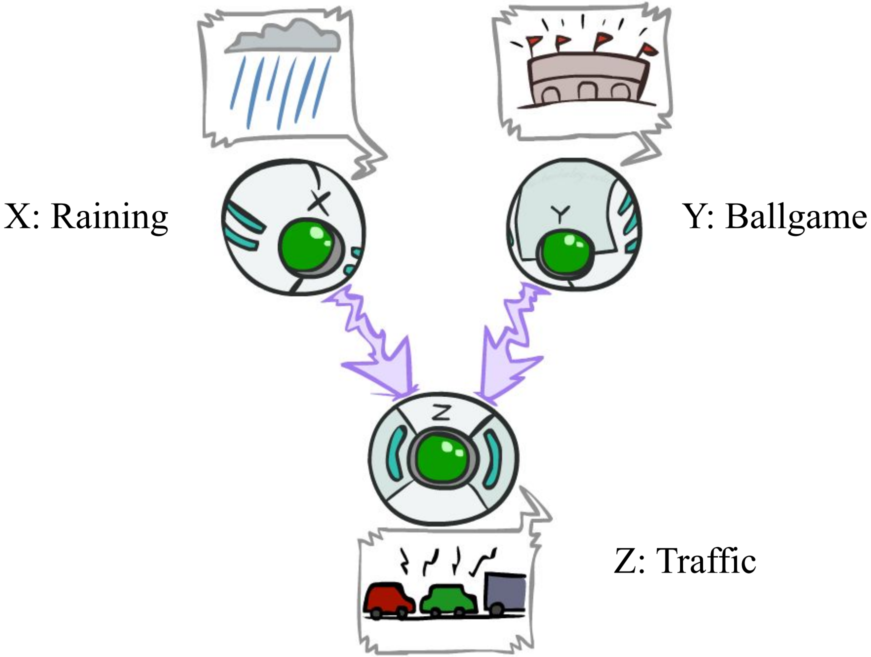
$$= \frac{P(y)P(x|y)P(z|y)}{P(y)P(x|y)}$$

$$= P(z|y)$$

Yes!
Observing the cause blocks influence between effects

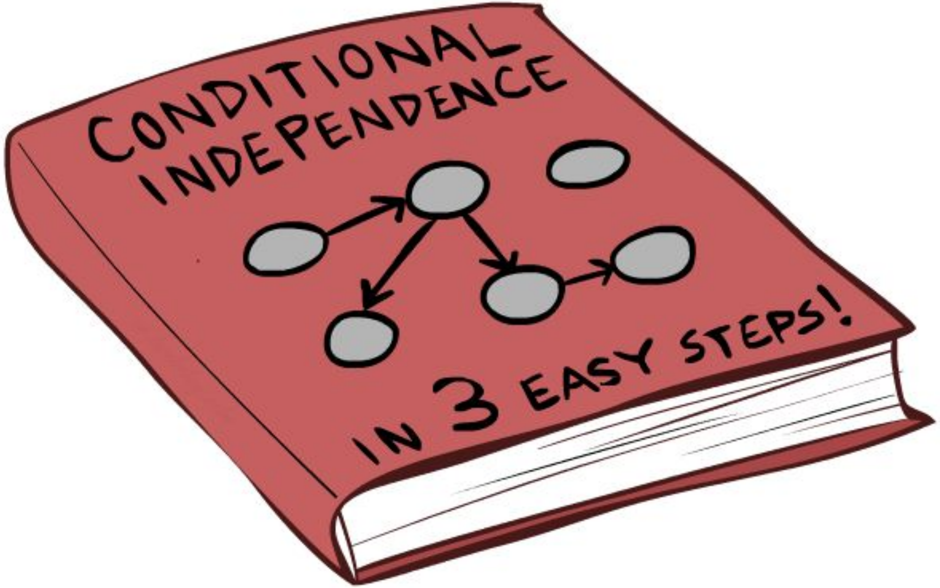
Common Effect

- Last configuration: two causes of one effect (v-structures)



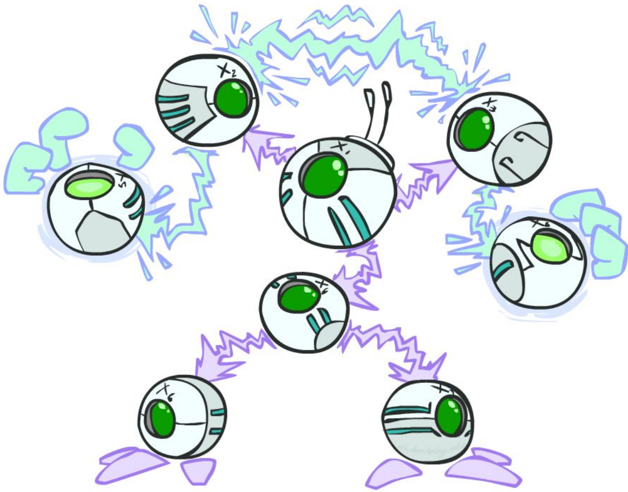
- Are X and Y independent?
  - Yes: the ballgame and the rain cause traffic, but they are not correlated
  - Quiz: Prove they must be
- Are X and Y independent given Z?
  - No: seeing traffic puts the rain and the ballgame in competition
- This is backwards from the other cases
  - Observing an effect activates influence between possible causes

The General Case



The General Case

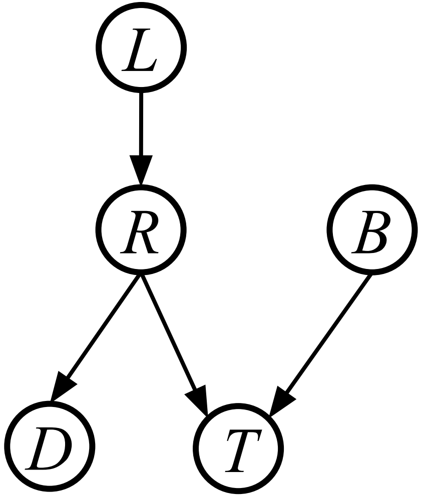
- General question: in a given BN, are two variables independent (given evidence)?
- Solution: analyze the graph
- Any complex example can be broken into repetitions of the three canonical cases

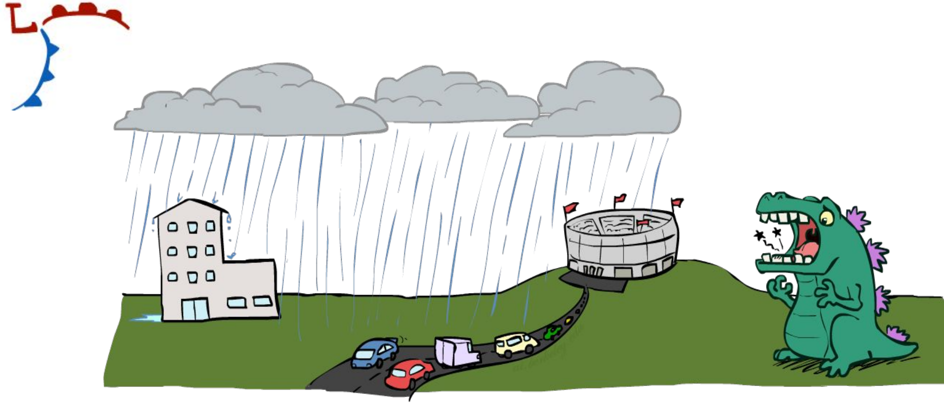


Reachability

- Recipe
  - Shade evidence nodes, look for paths in the resulting graph
- Attempt 1
  - If two nodes are connected by an undirected path not blocked by a shaded node, they are conditionally independent

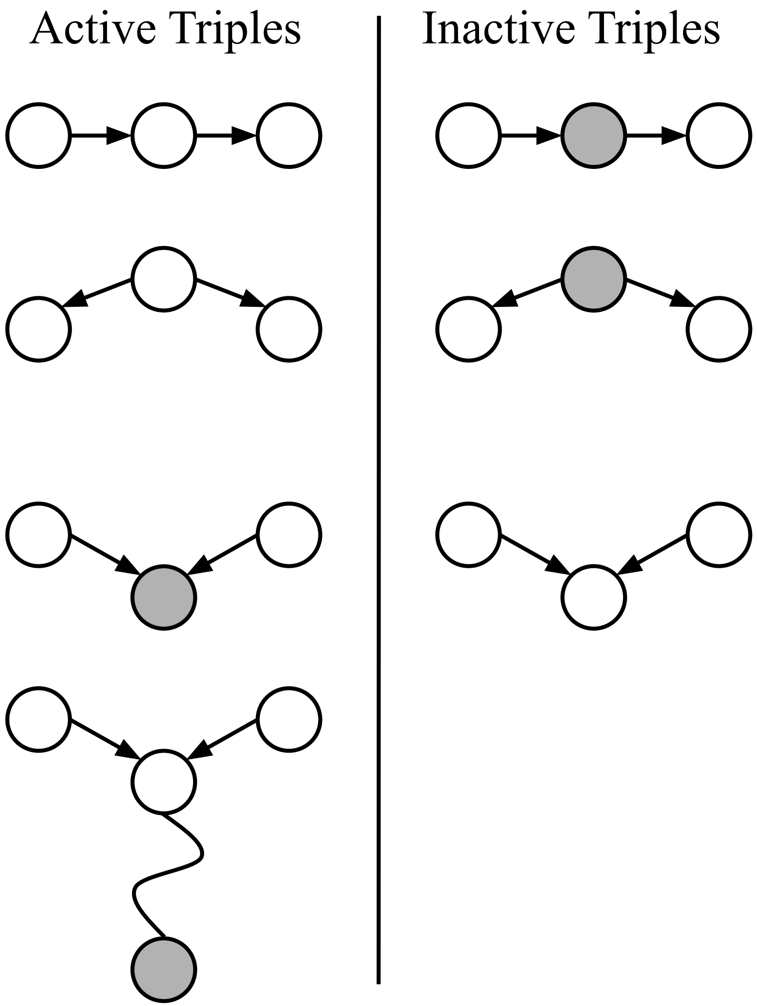
- Almost works, but not quite





Active / Inactive Paths

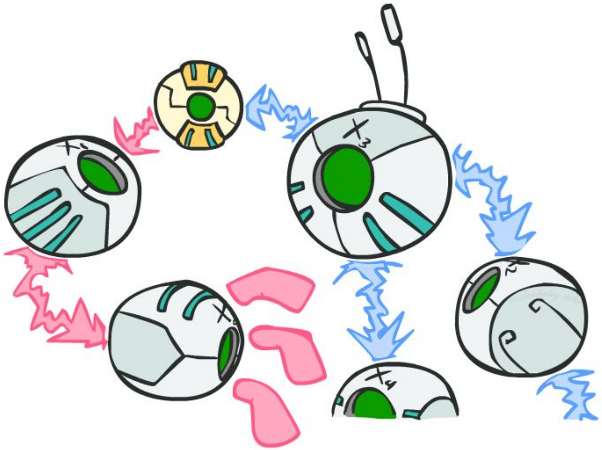
- Question: Are X and Y conditionally independent given evidence variables {Z}?
  - Yes, if X and Y "d-separated" by Z
  - Consider all (undirected) paths from X to Y
  - No active paths = independence!
- A path is active if each triple is active
  - Causal chain A → B → C where B is unobserved (either direction)
  - Common cause A ← B → C where B is unobserved
  - Common effect (aka v-structure) A → B ← C where B or one of its descendents is observed
- All it takes to block a path is a single inactive segment



D-Separation

- Query: $X_i \perp\!\!\!\perp X_j | \{X_{k_1}, ..., X_{k_n}\}$

- Check all (undirected!) paths between X_i and X_j
- If one or more active, then independence not guaranteed
- Otherwise (i.e. if all paths are inactive), then independence is guaranteed

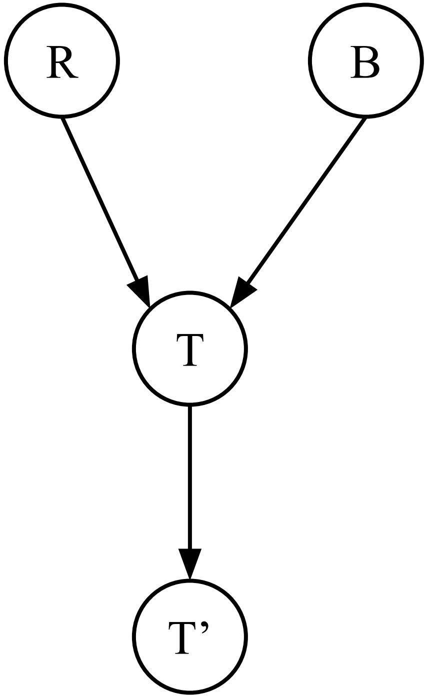


Example 1

$R \perp\!\!\!\perp B$ *Yes*

$R \perp\!\!\!\perp B | T$

$R \perp\!\!\!\perp B | T'$



Example 2

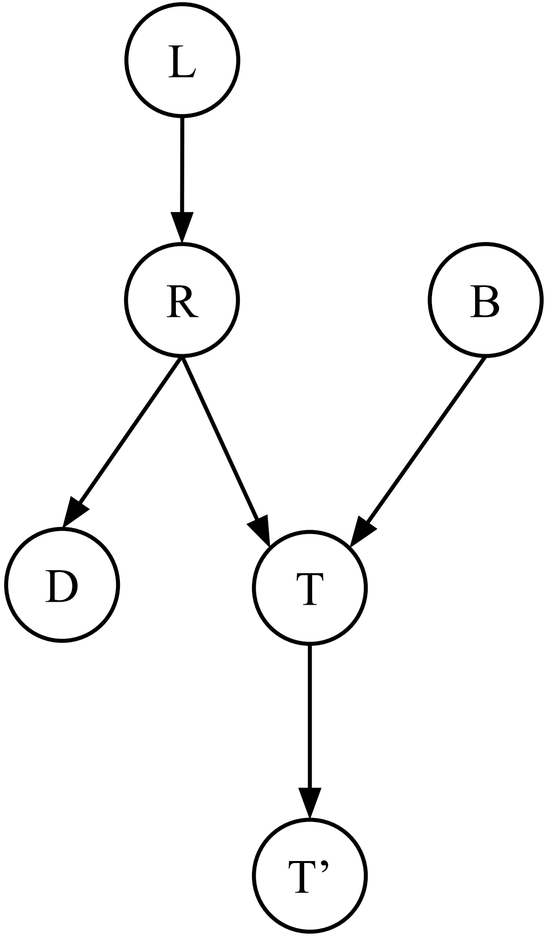
$L \perp\!\!\!\perp T' | T$ Yes

$L \perp\!\!\!\perp B$ Yes

$L \perp\!\!\!\perp B | T$

$L \perp\!\!\!\perp B | T'$

$L \perp\!\!\!\perp B | T, R$ Yes

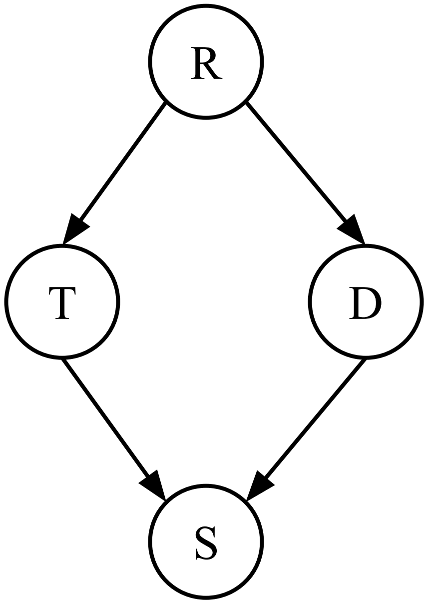


Example 3

$$T \perp\!\!\!\perp D$$

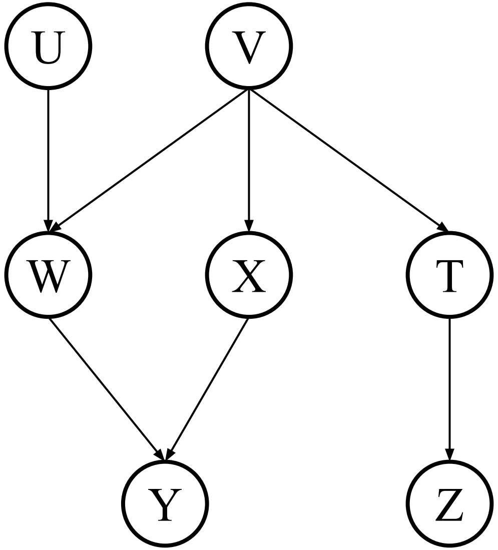
$$T \perp\!\!\!\perp D | R$$ Yes

$$T \perp\!\!\!\perp D | R, S$$



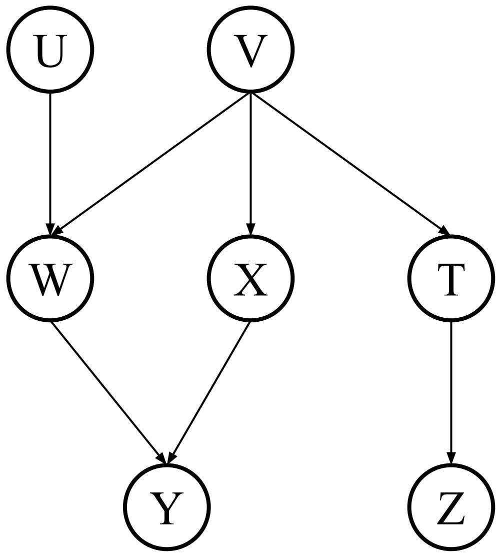
Quiz - 1

$V \perp Z$?



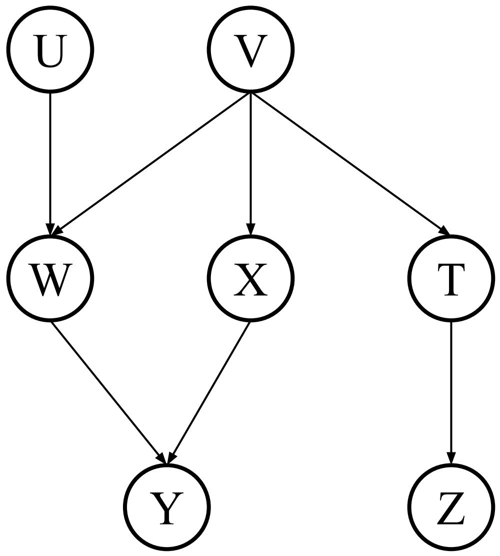
Quiz - 2

$V \perp Z \mid T$?



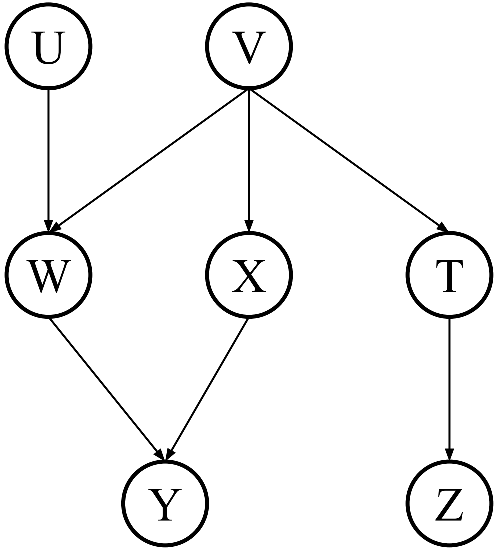
Quiz - 3

$U \perp V$?



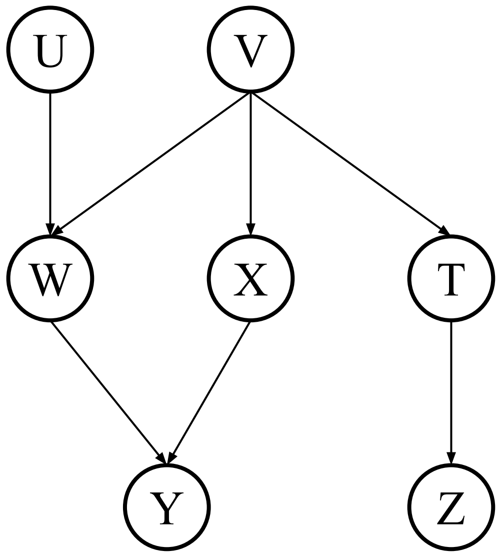
Quiz - 4

$U \perp V \mid W$?



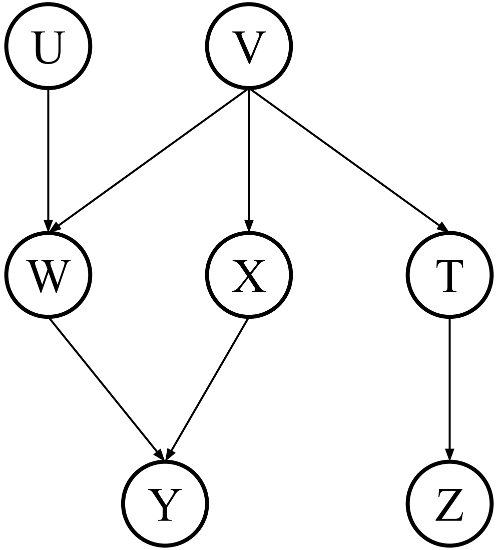
Quiz - 5

$U \perp V \mid X$?



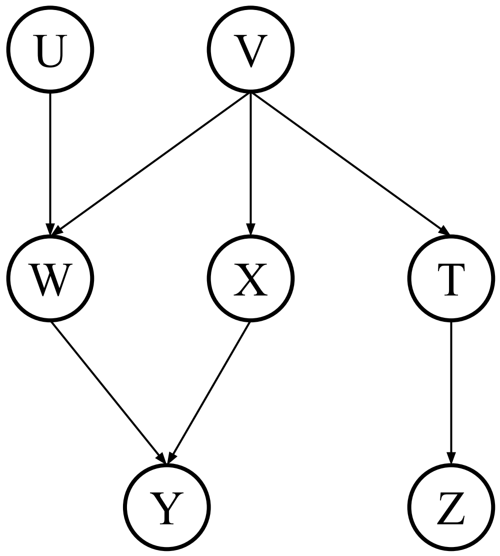
Quiz - 6

$U \perp V \mid Y$?



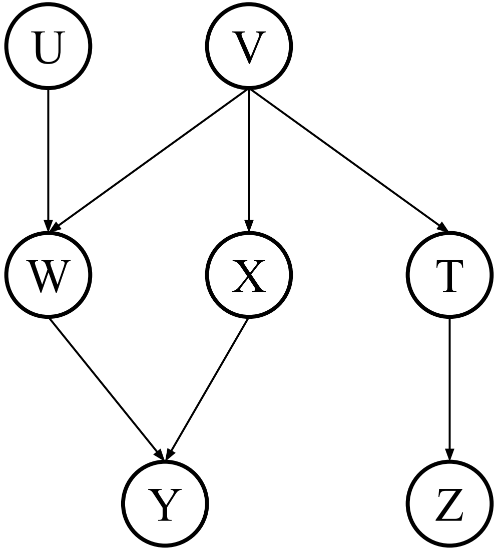
Quiz - 7

$U \perp V \mid Z$?



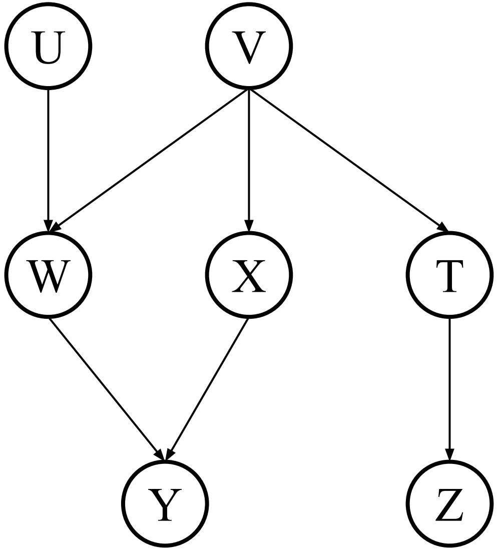
Quiz - 8

W ⊥ X ?



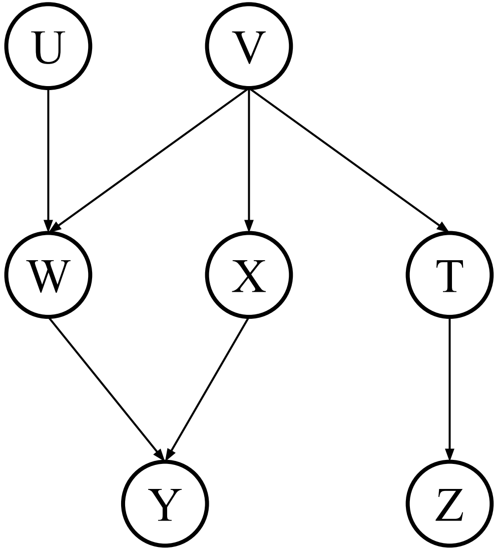
Quiz - 9

$X \perp T \mid V$?



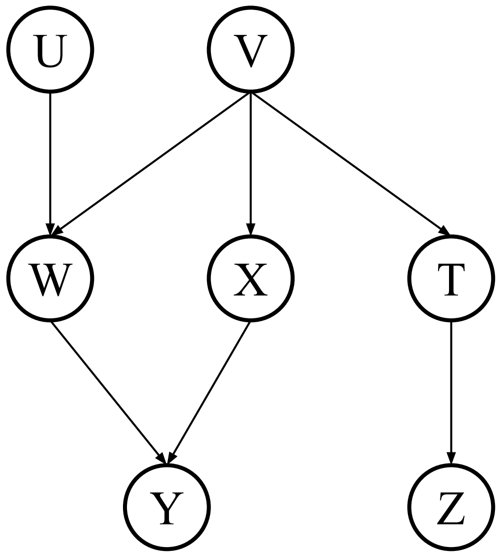
Quiz - 10

$X \perp W \mid U$?



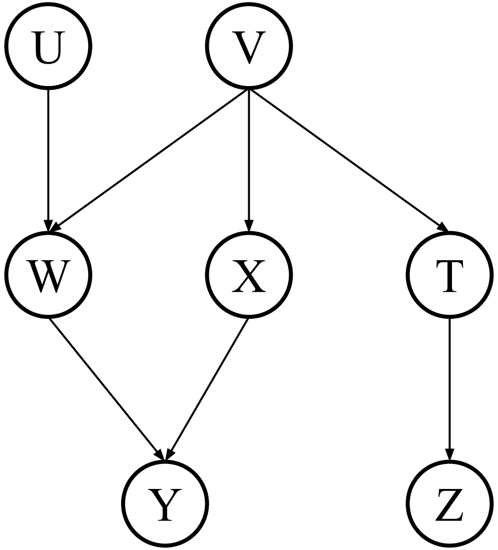
Quiz - 11

$Y \perp Z$?



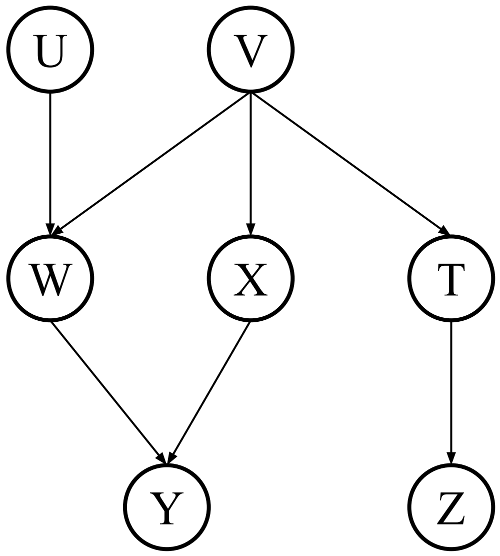
Quiz - 12

$Y \perp Z \mid T$?



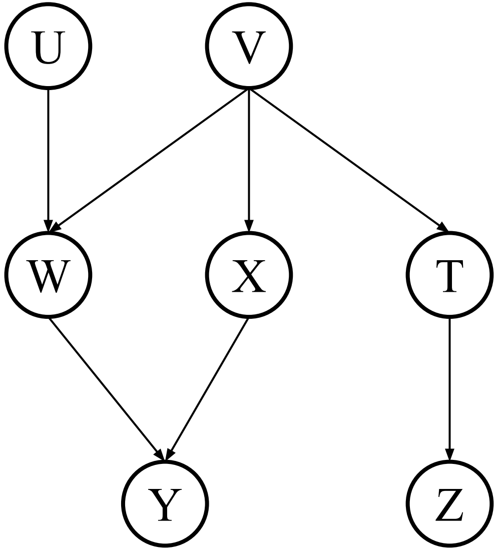
Quiz - 13

$Y \perp Z \mid X$?



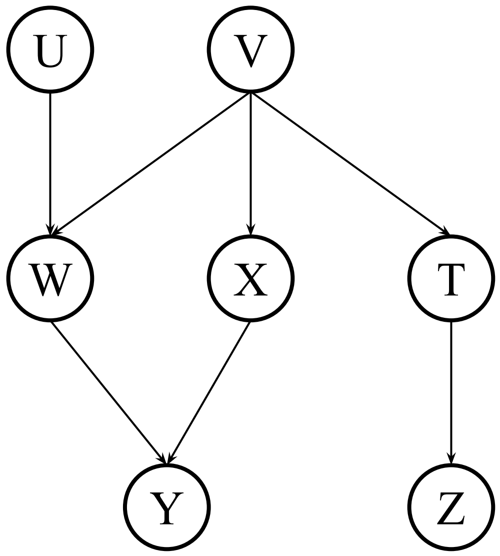
Quiz - 14

$Y \perp Z \mid V$?



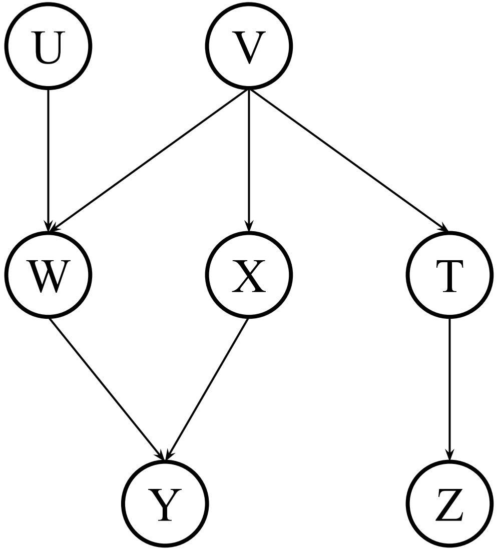
Quiz - 15

$W \perp Z \mid V$?



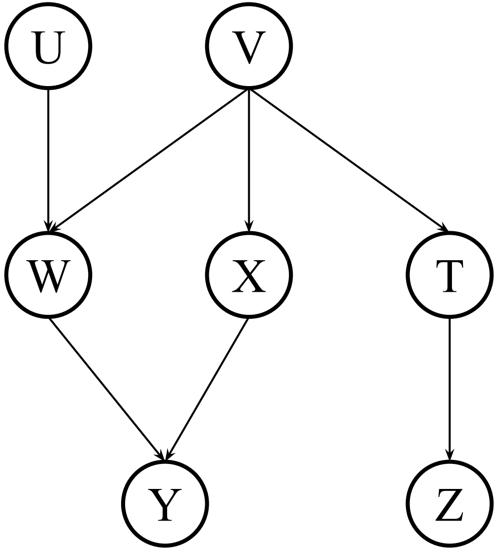
Quiz - 16

$U \perp Z$?



Quiz - 17

$U \perp Z \mid Y$?

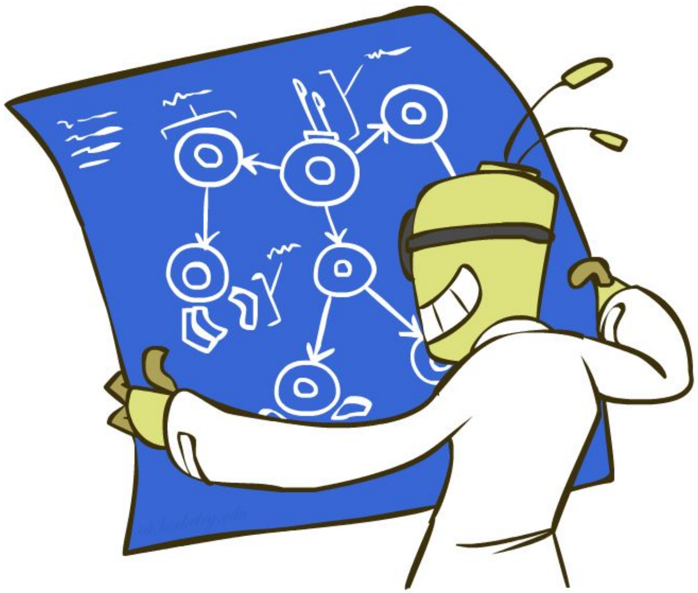


Structure Implications

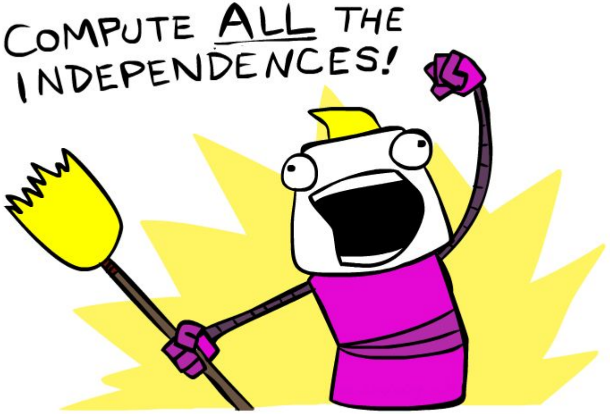
- Given a Bayes net structure, can run d-separation algorithm to build a complete list of conditional independencies that are necessarily true of the form

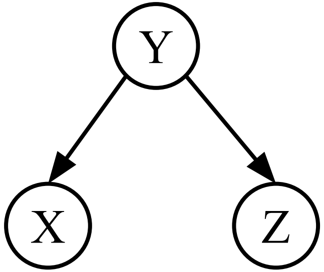
$$X_i \perp\!\!\!\perp X_j | \{X_{k_1}, ..., X_{k_n}\}$$

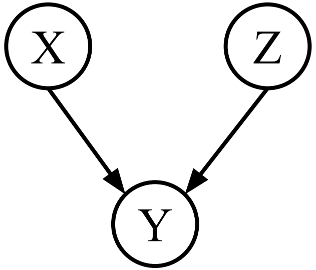
- This list determines the set of probability distributions that can be represented

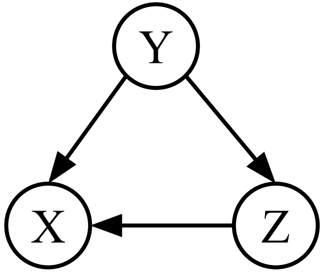


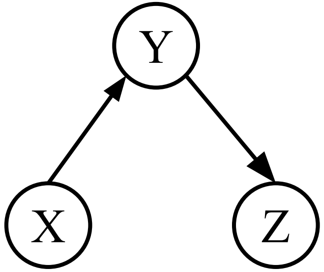
Quiz: Computing All Independences





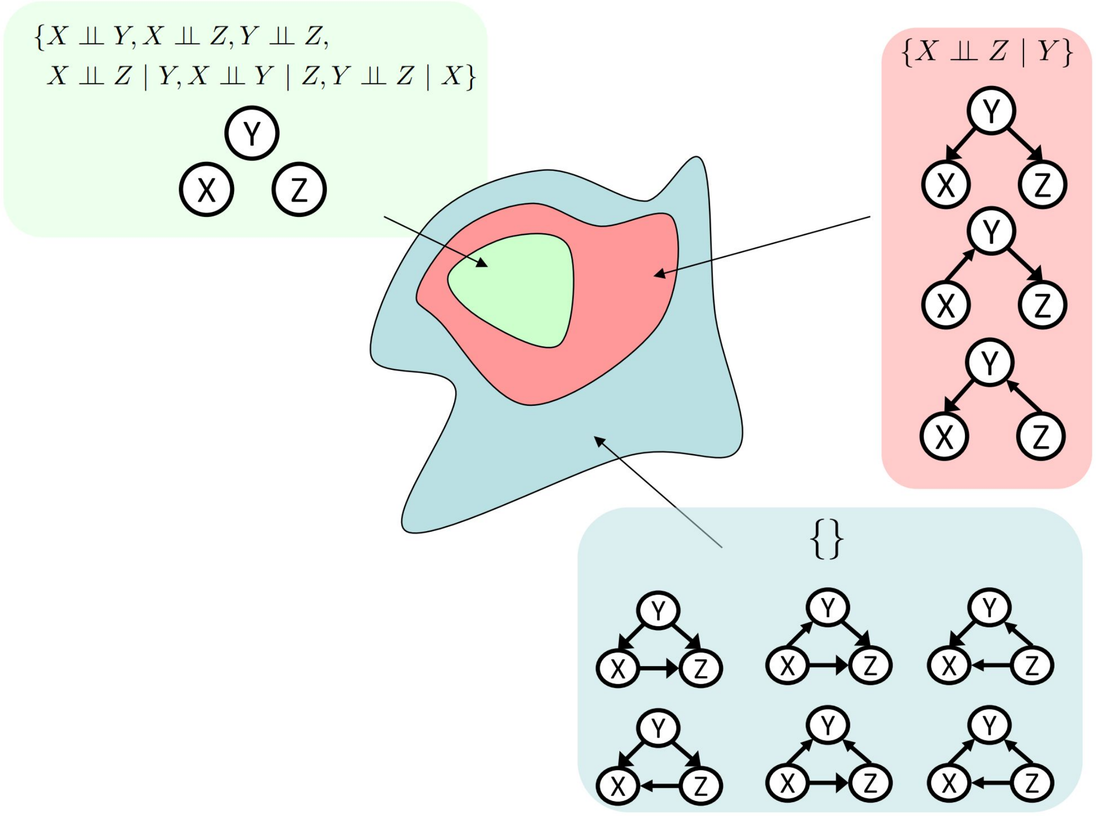






Topology Limits Distributions

- Given some graph topology G, only certain joint distributions can be encoded
- The graph structure guarantees certain (conditional) independencies
- (There might be more independence)
- Adding arcs increases the set of distributions, but has several costs
- Full conditioning can encode any distribution

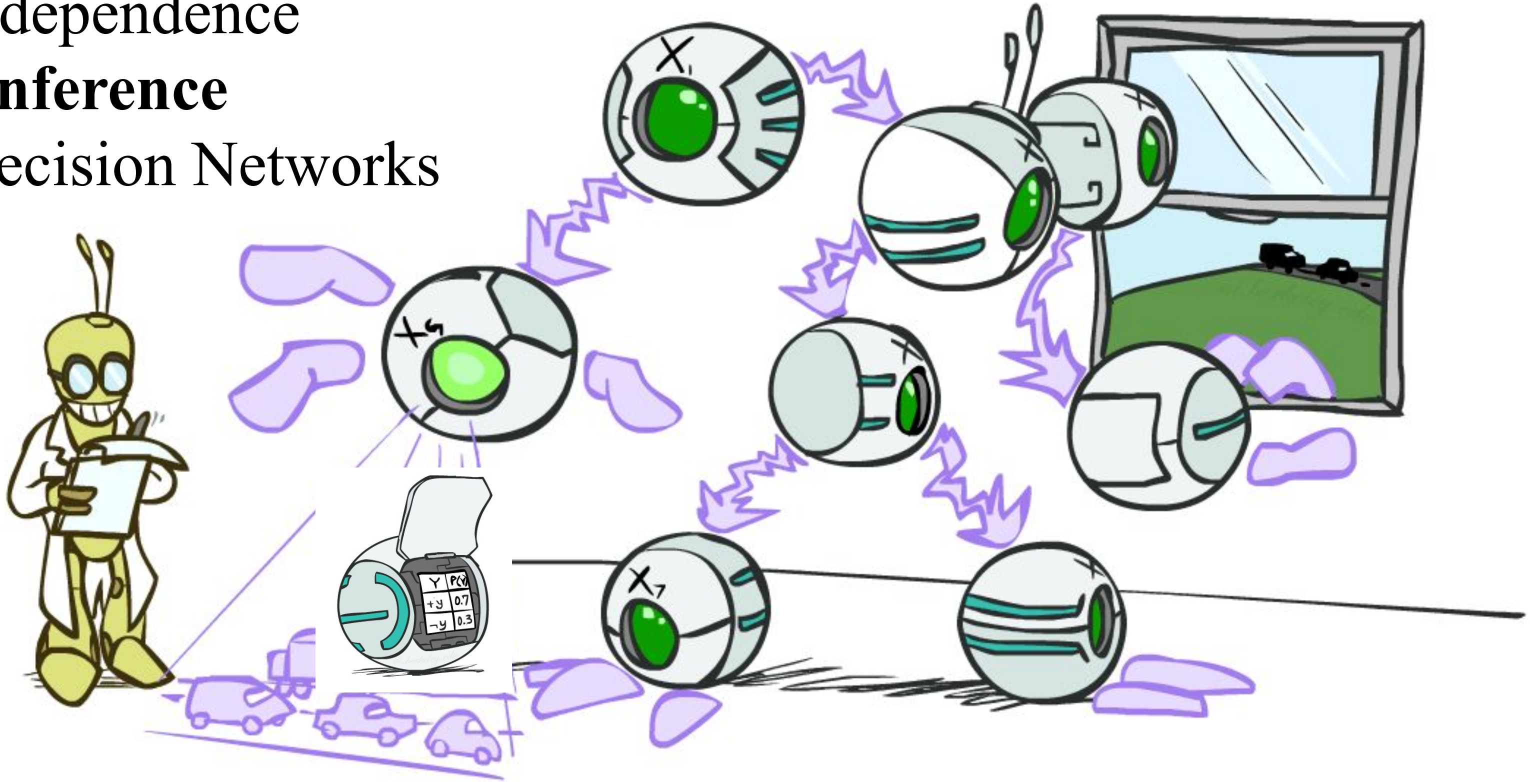


Bayes Nets Representation Summary

- Bayes nets compactly encode joint distributions
- Guaranteed independencies of distributions can be deduced from BN graph structure
- D-separation gives precise conditional independence guarantees from graph alone
- A Bayes' net's joint distribution may have further (conditional) independence that is not detectable until you inspect its specific distribution

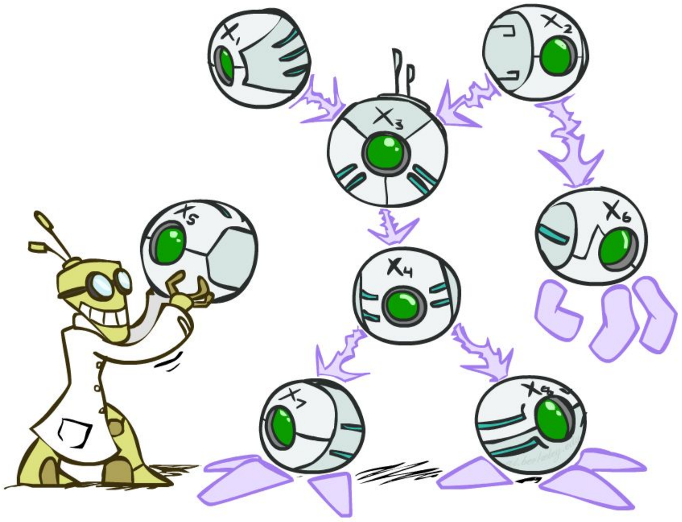
Outline

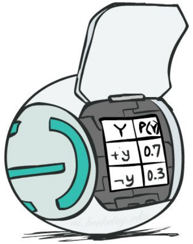
- Part A: Representation
- Part B: Independence
- **Part C: Inference**
- Part D: Decision Networks



Bayes' Net Representation

- A directed, acyclic graph, one node per random variable
- A conditional probability table (CPT) for each node
- Bayes' nets implicitly encode joint distributions

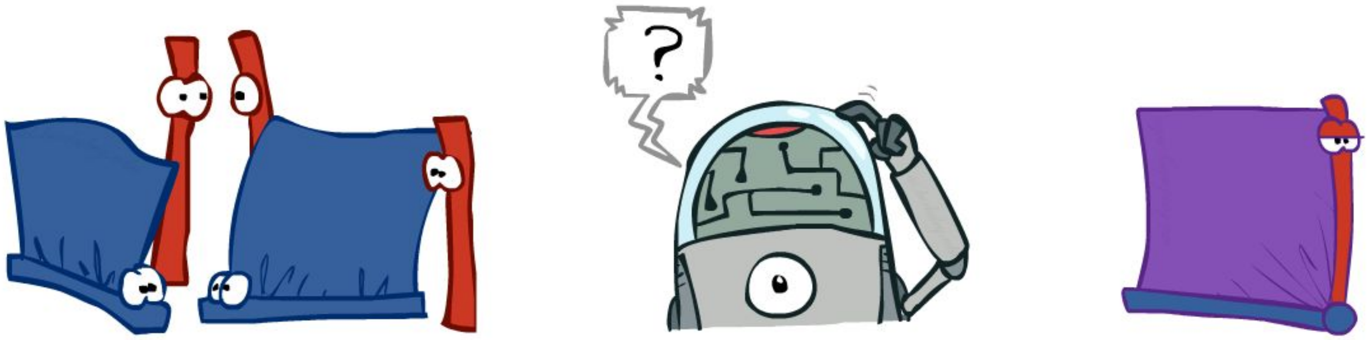




Inference

- Inference: calculating some useful quantity from a joint probability distribution
- Example: posterior probability

$$P(Q|E_1 = e_1, \ldots E_k = e_k)$$



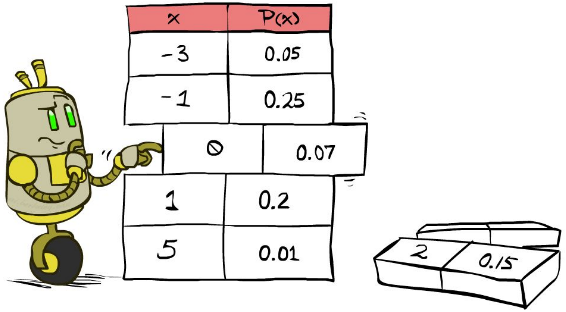
Inference by Enumeration

- General case
  - Evidence variables: $E_1 \ldots E_k = e_1 \ldots e_k$
  - Query variable: Q
  - Hidden variables: $H_1 \ldots H_r$

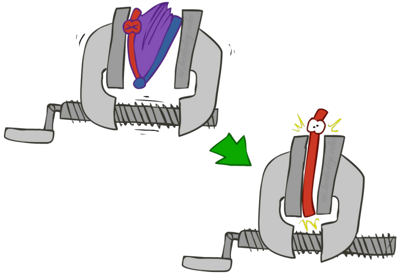
$\left.\begin{array}{l} E_1 \ldots E_k = e_1 \ldots e_k \\ Q \\ H_1 \ldots H_r \end{array}\right\} \begin{array}{c} X_1, X_2, \ldots X_n \\ \textit{All variables} \end{array}$

We want: $P(Q|e_1 \ldots e_k)$

Step 1: Select the entries consistent with the evidence



Step 2: Sum out H to get joint of Query and evidence



$$P(Q, e_1 \ldots e_k) = \sum_{h_1 \ldots h_r} P(Q, h_1 \ldots h_r, e_1 \ldots e_k)$$

$X_1, X_2, \ldots X_n$

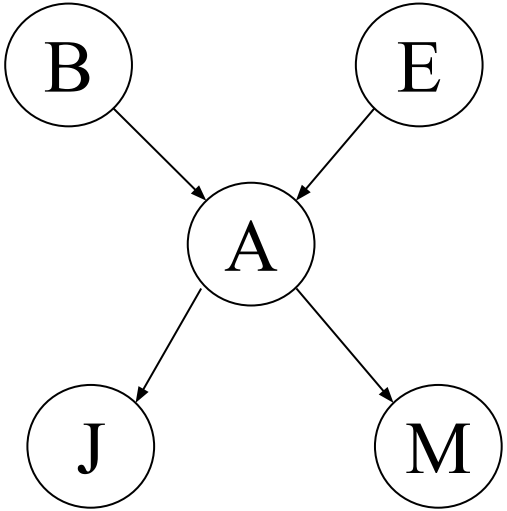
Step 3: Normalize

$$\times \frac{1}{Z}$$

$$Z = \sum_q P(Q, e_1 \cdots e_k)$$

$$P(Q|e_1 \cdots e_k) = \frac{1}{Z} P(Q, e_1 \cdots e_k)$$

Example: Inference by Enumeration



$$P(B \mid +j, +m) \propto_B P(B, +j, +m)$$

$$= \sum_{e,a} P(B, e, a, +j, +m)$$

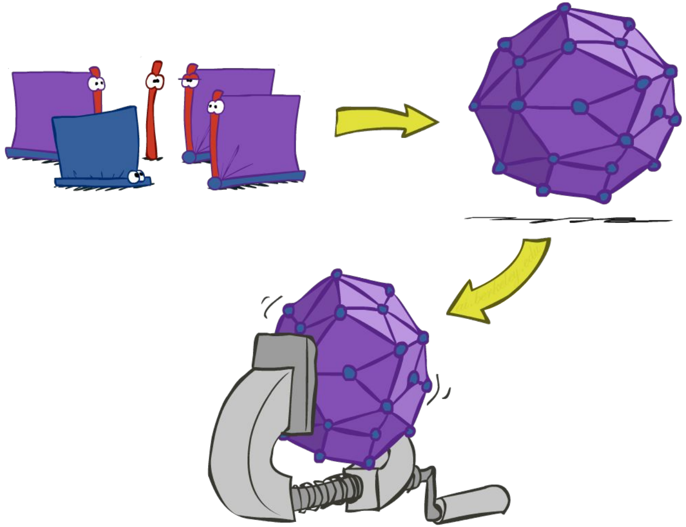
$$= \sum_{e,a} P(B)P(e)P(a|B,e)P(+j|a)P(+m|a)$$

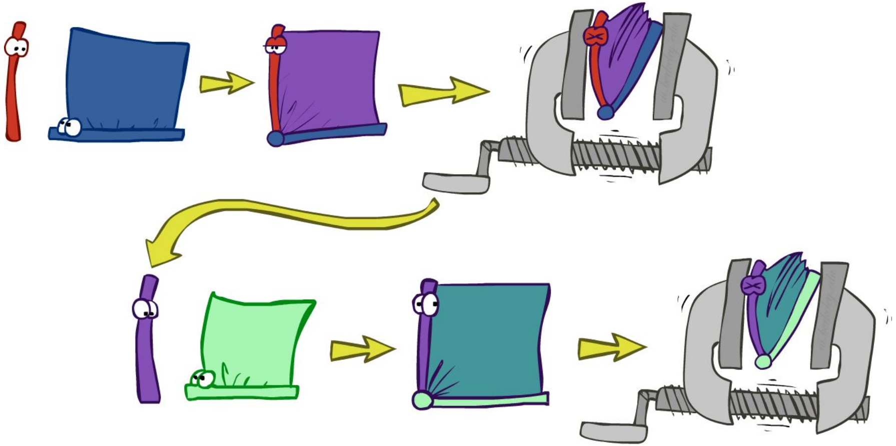
$$= P(B)P(+e)P(+a|B,+e)P(+j|+a)P(+m|+a) + P(B)P(+e)P(-a|B,+e)P(+j|-a)P(+m|-a)$$
$$P(B)P(-e)P(+a|B,-e)P(+j|+a)P(+m|+a) + P(B)P(-e)P(-a|B,-e)P(+j|-a)P(+m|-a)$$

Inference by Enumeration vs. Variable Elimination

- Inference by enumeration is slow
  - You join up the whole joint distribution before you sum out the hidden variables

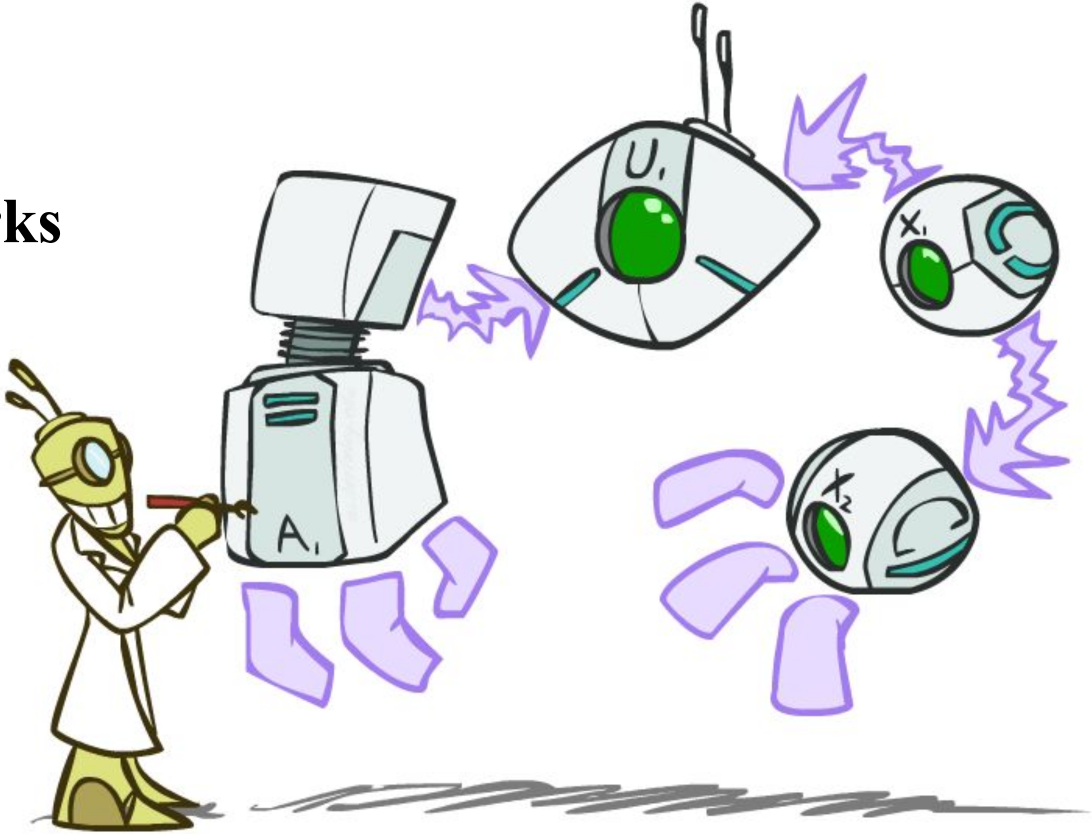
- Idea: interleave joining and marginalizing!
  - Called “Variable Elimination”
  - Usually much faster than inference by enumeration



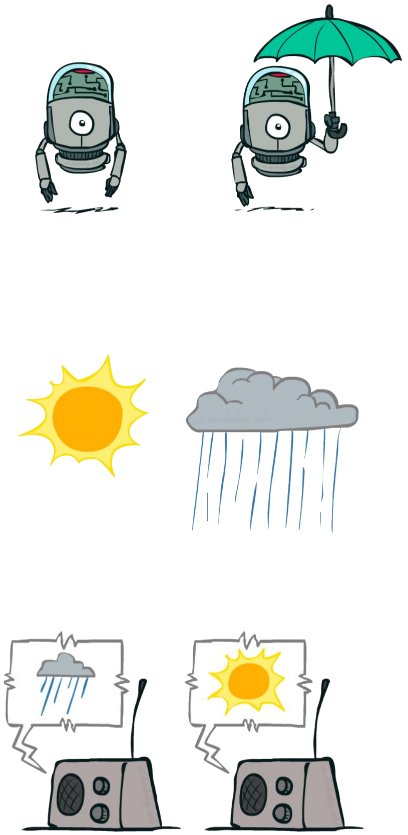


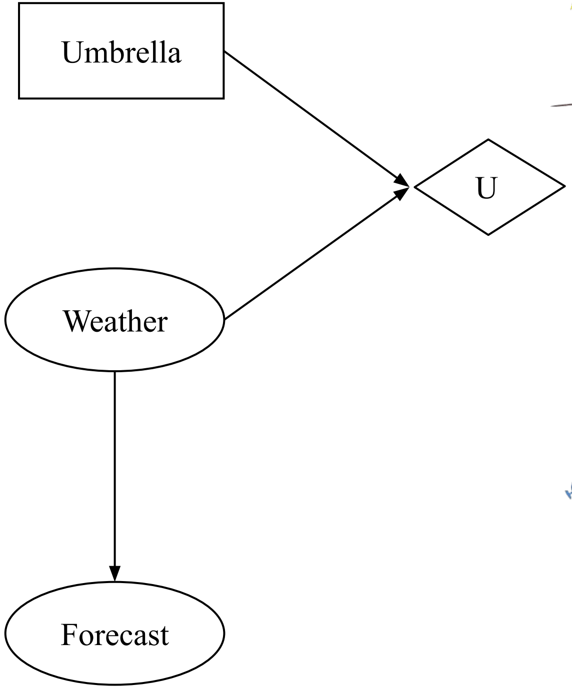
Outline

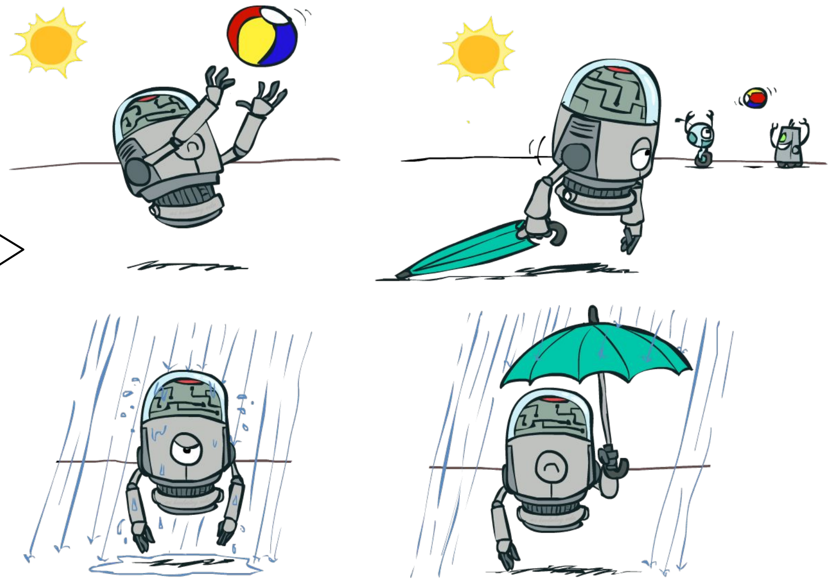
- Part A: Representation
- Part B: Independence
- Part C: Inference
- **Part D: Decision Networks**



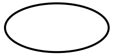
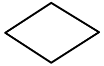
Decision Networks

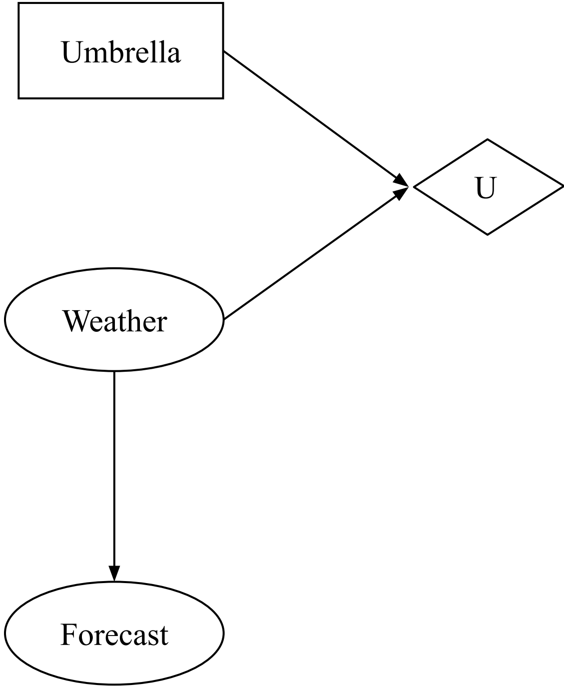






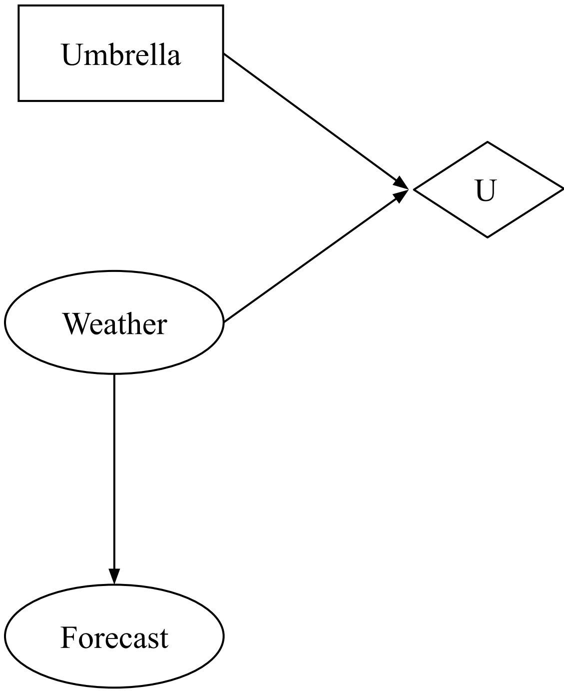
Decision Networks

- MEU: choose the action which maximizes the expected utility given the evidence
- Can directly operationalize this with decision networks
  - Bayes nets with nodes for utility and actions
  - Lets us calculate the expected utility for each action
- New node types
  - Chance nodes (just like BNs) 
  - Actions (rectangles, cannot have parents, act as observed evidence) 
  - Utility node (diamond, depends on action and chance nodes) 

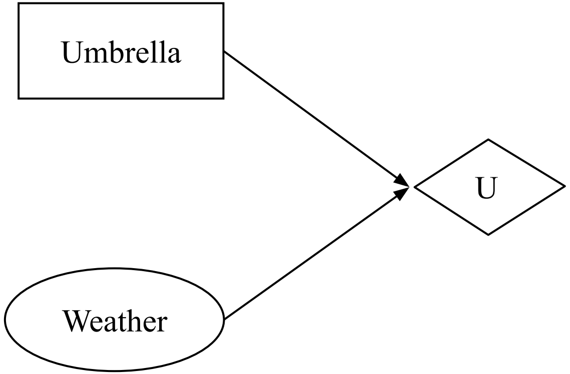


Decision Networks

- Action selection
  - Instantiate all evidence
  - Set action node(s) each possible way
  - Calculate posterior for all parents of utility node, given the evidence
  - Calculate expected utility for each action
  - Choose maximizing action



Decision Networks



W	P(W)
sun	0.7
rain	0.3

A	W	U(A,W)
leave	sun	100
leave	rain	0
take	sun	20
take	rain	70

Umbrella = leave

$$\mathrm{EU}(\mathrm{leave}) = \sum_w P(w)U(\mathrm{leave}, w)$$

$$= 0.7 \cdot 100 + 0.3 \cdot 0 = 70$$

Umbrella = take

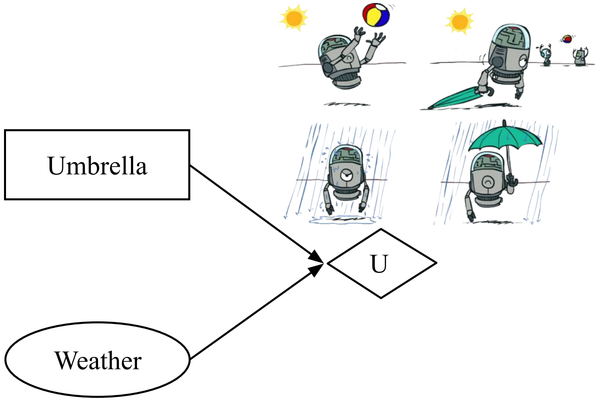
$$\mathrm{EU}(\mathrm{take}) = \sum_w P(w)U(\mathrm{take}, w)$$

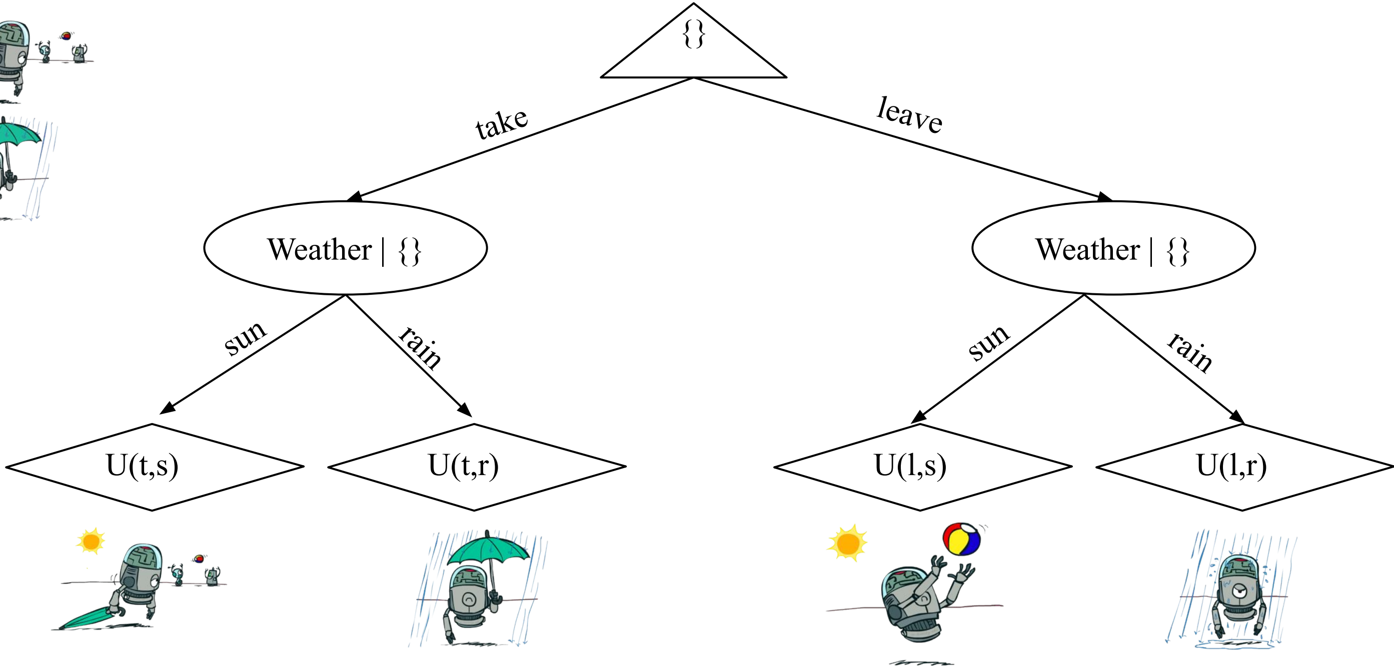
$$= 0.7 \cdot 20 + 0.3 \cdot 70 = 35$$

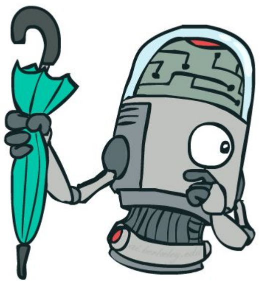
Optimal decision = leave

$$\mathrm{MEU}(\emptyset) = \max_a \mathrm{EU}(a) = 70$$

Decisions as Outcome Trees







- Almost exactly like expectimax / MDPs
- What’s changed?

Example: Decision Networks

Umbrella = leave

$$\mathrm{EU}(\mathrm{leave}|\mathrm{bad}) = \sum_{w} P(w|\mathrm{bad})U(\mathrm{leave}, w)$$

$$= 0.34 \cdot 100 + 0.66 \cdot 0 = 34$$

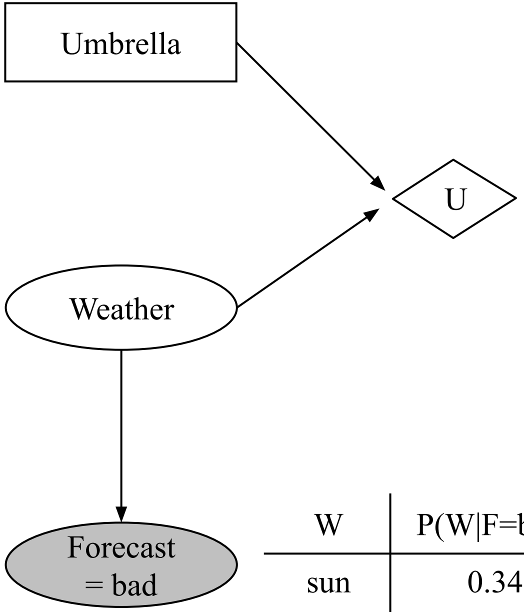
Umbrella = take

$$\mathrm{EU}(\mathrm{take}|\mathrm{bad}) = \sum_{w} P(w|\mathrm{bad})U(\mathrm{take}, w)$$

$$= 0.34 \cdot 20 + 0.66 \cdot 70 = 53$$

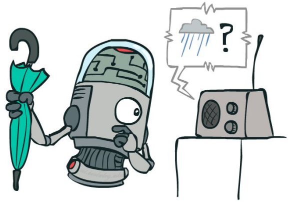
Optimal decision = take

$$\mathrm{MEU}(F = \mathrm{bad}) = \max_{a} \mathrm{EU}(a|\mathrm{bad}) = 53$$



A	W	U(A,W)
leave	sun	100
leave	rain	0
take	sun	20
take	rain	70

W	P(W|F=bad)
sun	0.34
rain	0.66



Decisions as Outcome Trees