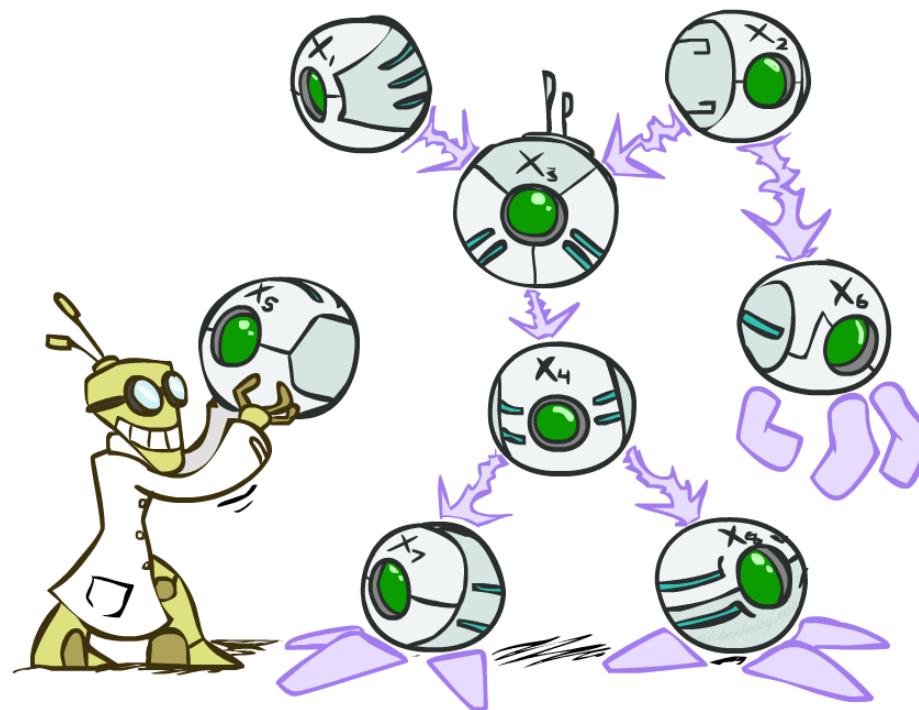


Announcements

- Midterm 1: Almost done grading. Will be returned by the end of the week.
- Probability Review Sessions
 - See Piazza post
- Project 3: Reinforcement Learning
 - Due Friday 10/14 at 5pm
- Contest 2: Due Sunday 10/16 at 11:59pm
- Homework 6: Probability and Bayes' Nets
 - Due Monday 10/17 at 11:59pm

CS 188: Artificial Intelligence

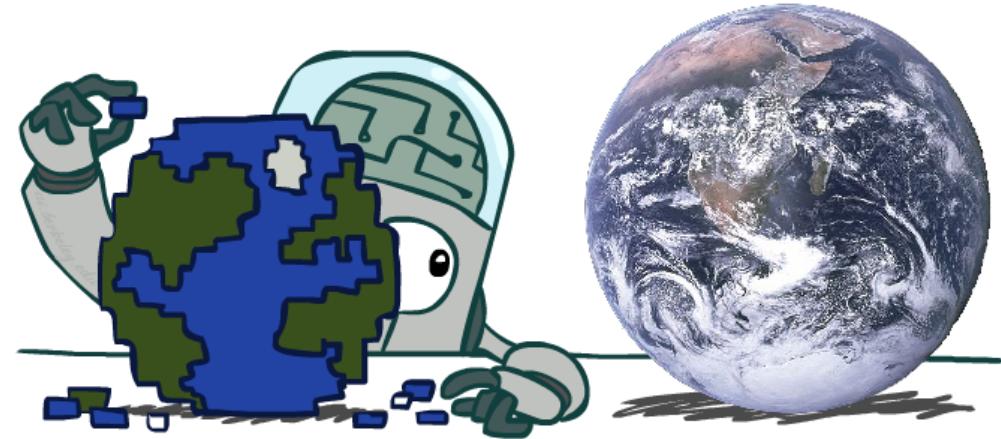
Bayes' Nets



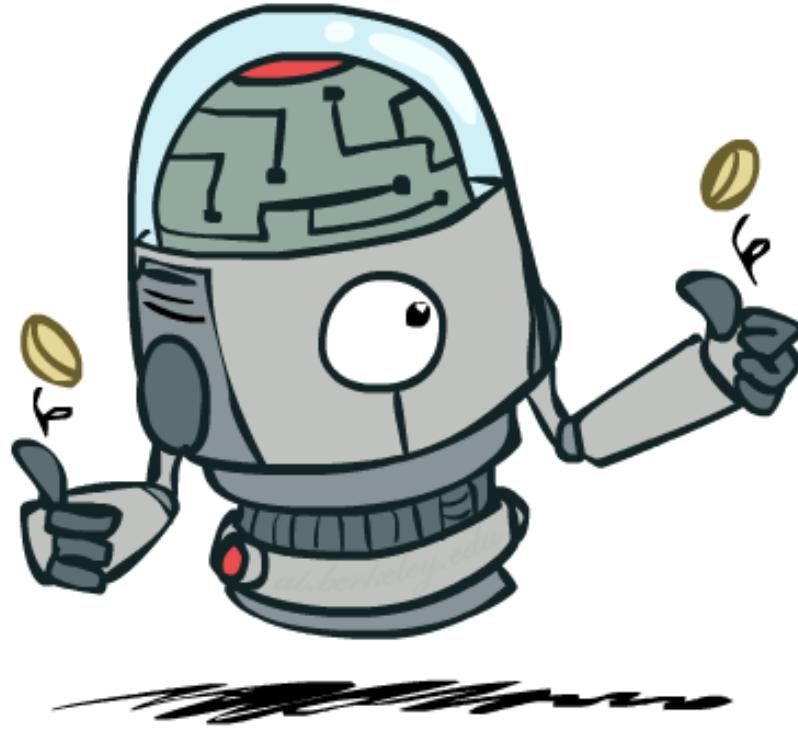
Instructors: Josh Hug & Adam Janin --- University of California, Berkeley

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



Independence

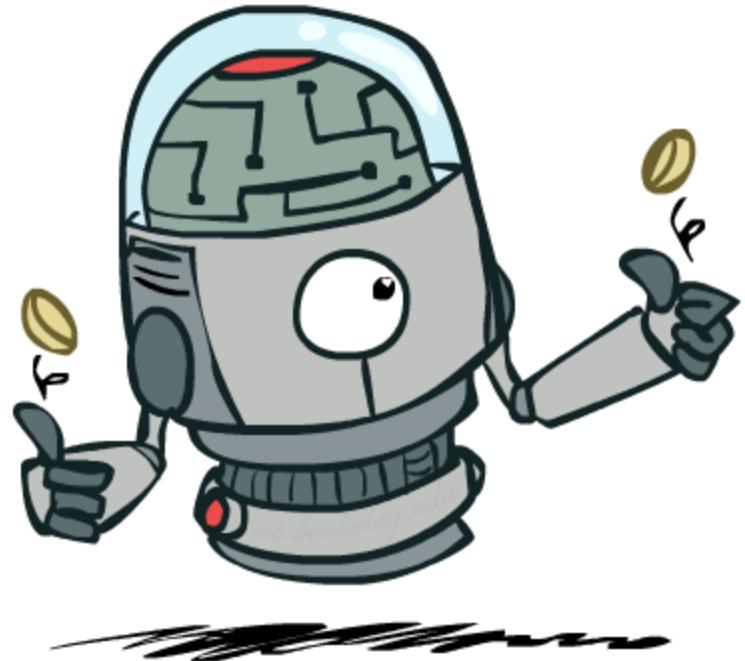
- Two variables are *independent* if:

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

- This says that their joint distribution *factors* into a product two simpler distributions
- Another form:

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

- We write: $X \perp\!\!\!\perp Y$
- Independence is a simplifying *modeling assumption*
 - *Empirical* joint distributions: at best “close” to independent



Independence?



What could we assume for {Weather, Traffic, Cavity, Toothache}?

Weather \perp Cavity

Weather \perp Toothache

Traffic \perp Cavity

Traffic \perp Toothache

Testing Independence

$P_1(T, W)$

T	W	P
hot	sun	0.4
hot	rain	0.1
cold	sun	0.2
cold	rain	0.3

$P(T)$

T	P
hot	0.5
cold	0.5

$P_2(T, W) = P(T) \cdot P(W)$

$P(W)$

W	P
sun	0.6
rain	0.4

T	W	P
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

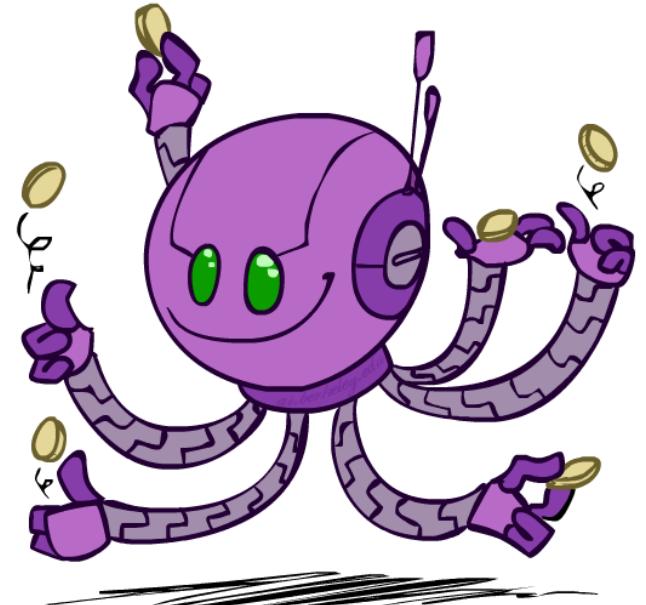
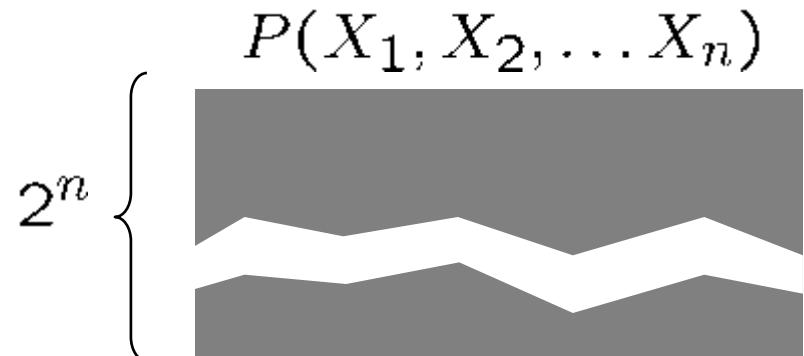
$$P(X_2)$$

H	0.5
T	0.5

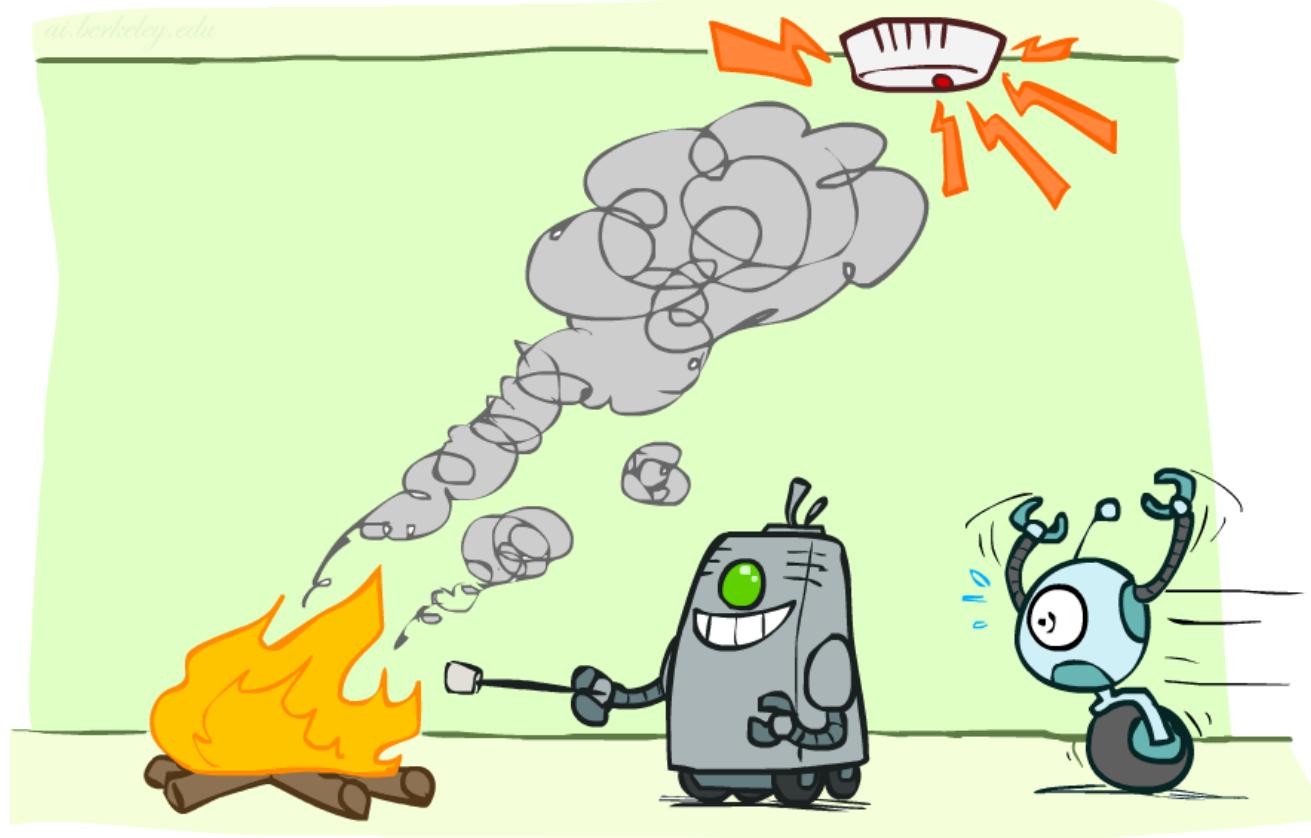
...

$$P(X_n)$$

H	0.5
T	0.5

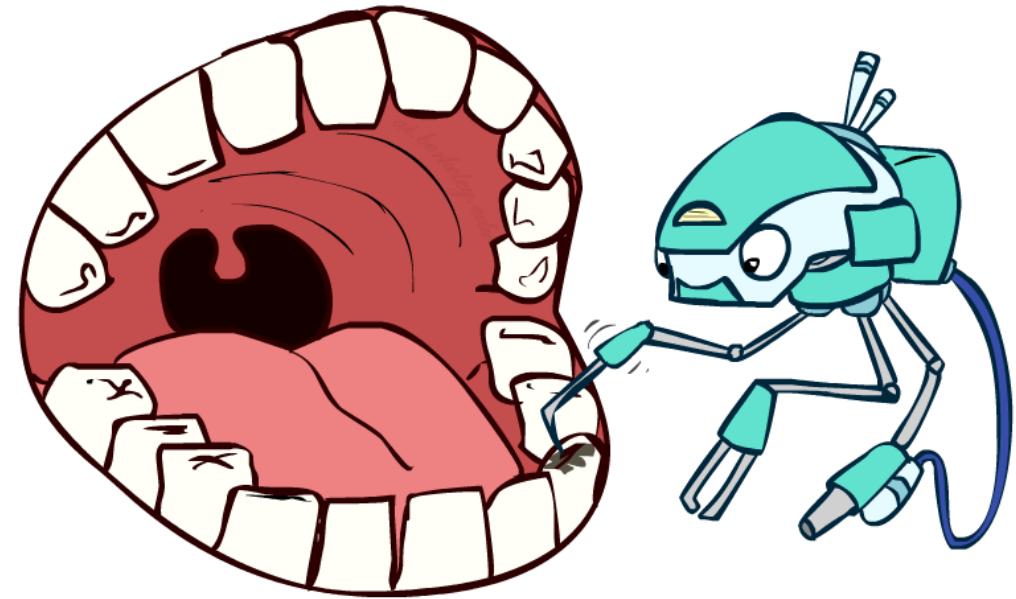


Conditional Independence



Conditional Independence

- $P(\text{Toothache}, \text{Cavity}, \text{Catch})$
- If I have a cavity, the probability that the probe catches it doesn't depend on whether I have a toothache:
 - $P(+\text{catch} | +\text{toothache}, +\text{cavity}) = P(+\text{catch} | +\text{cavity})$
- The same independence holds if I don't have a cavity:
 - $P(+\text{catch} | +\text{toothache}, -\text{cavity}) = P(+\text{catch} | -\text{cavity})$
- Catch is *conditionally independent* of Toothache given Cavity:
 - $P(\text{Catch} | \text{Toothache}, \text{Cavity}) = P(\text{Catch} | \text{Cavity})$
- Equivalent statements:
 - $P(\text{Toothache} | \text{Catch}, \text{Cavity}) = P(\text{Toothache} | \text{Cavity})$
 - $P(\text{Toothache}, \text{Catch} | \text{Cavity}) = P(\text{Toothache} | \text{Cavity}) P(\text{Catch} | \text{Cavity})$
 - 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 is conditionally independent of Y given Z $X \perp\!\!\!\perp Y | Z$

if and only if:

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

or, equivalently, if and only if

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

Conditional Independence



If X is conditionally independent of Y given Z

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

What can you say about $P(Y|Z, X)$?

Conditional Independence



If X is conditionally independent of Y given Z

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

What can you say about $P(Y|Z, X) = P(Y|Z)$

Conditional Independence

- What about this domain:

- Traffic
- Umbrella
- Raining



Conditional Independence

- What about this domain:

- Traffic
- Umbrella
- Raining

$U \perp T ?$



Conditional Independence

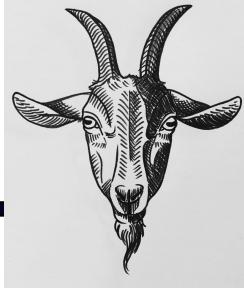
- What about this domain:

- Traffic
- Umbrella
- Raining

$$U \perp T \mid R$$

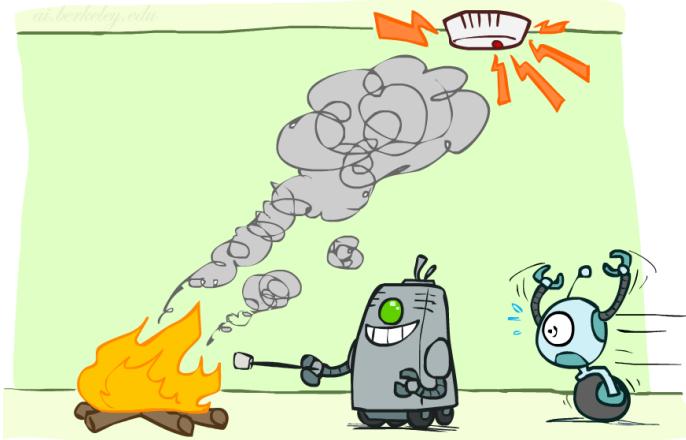


Conditional Independence



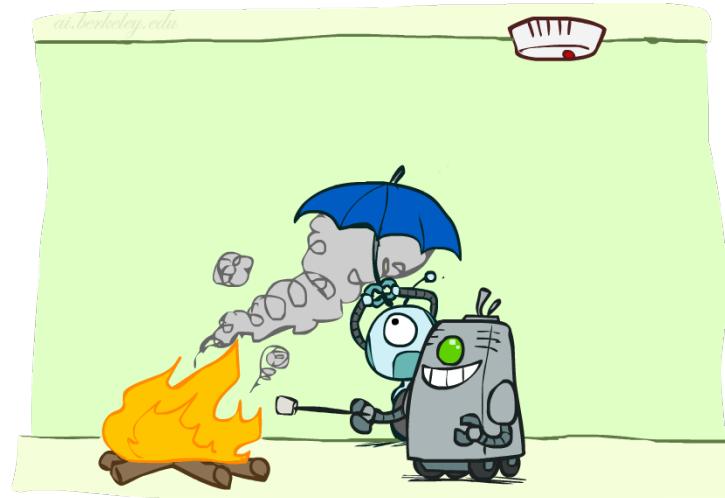
- What about this domain:

- Fire
- Smoke
- Alarm



$$F \perp A | S$$

*



Conditional Independence and the Chain Rule

- Chain rule:

$$P(X_1, X_2, \dots, X_n) = P(X_1)P(X_2|X_1)P(X_3|X_1, X_2) \dots$$

- Trivial decomposition:

$$P(\text{Traffic, Rain, Umbrella}) =$$

$$P(\text{Rain})P(\text{Traffic}|\text{Rain})P(\text{Umbrella}|\text{Rain, Traffic})$$

- With assumption of conditional independence:

$$P(\text{Traffic, Rain, Umbrella}) =$$

$$P(\text{Rain})P(\text{Traffic}|\text{Rain})P(\text{Umbrella}|\text{Rain})$$



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

Conditional Independence and the Chain Rule

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

$P(\text{Traffic, Rain, Umbrella}) =$

$P(\text{Traffic}) P(\text{Rain}|\text{Traffic}) P(\text{Umbrella}|\text{Traffic, Rain})$

$P(\text{Traffic, Rain, Umbrella}) =$

$P(\text{Umbrella}) P(\text{Traffic}|\text{Umbrella}) P(\text{Rain}|\text{Umbrella, Traffic})$

$P(\text{Traffic, Rain, Umbrella}) =$

$P(\text{Rain}) P(\text{Traffic}|\text{Rain}) P(\text{Umbrella}|\text{Rain, Traffic})$

But we said $U \perp T | R$!



Conditional Independence and the Chain Rule

$$P(T, R, U) = P(R)P(T|R)P(U|R, T)$$

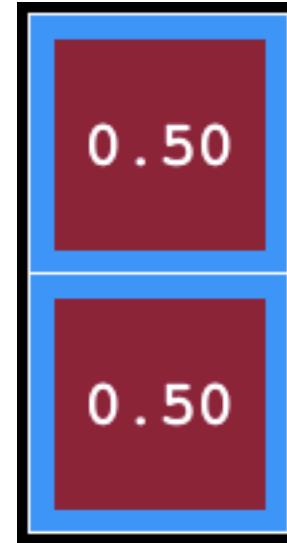
$$U \perp T \mid R \Leftrightarrow P(U|R, T) = P(U|R)$$

$$P(T, R, U) = P(R)P(T|R)P(U|R)$$



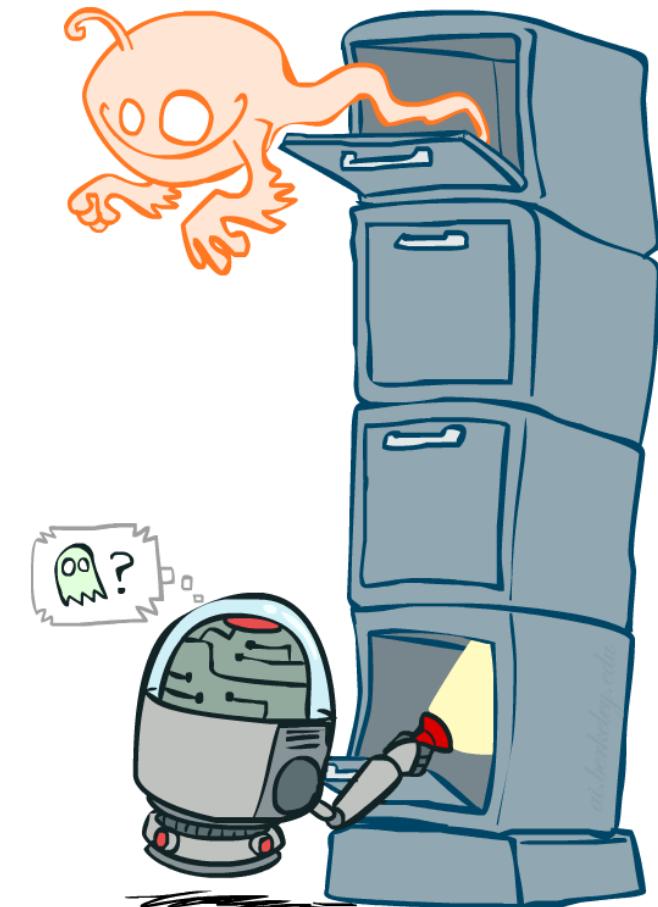
Ghostbusters Chain Rule

- 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
- Givens:
 $P(+g) = 0.5$
 $P(-g) = 0.5$
 $P(+t | +g) = 0.8$
 $P(+t | -g) = 0.4$
 $P(+b | +g) = 0.4$
 $P(+b | -g) = 0.8$

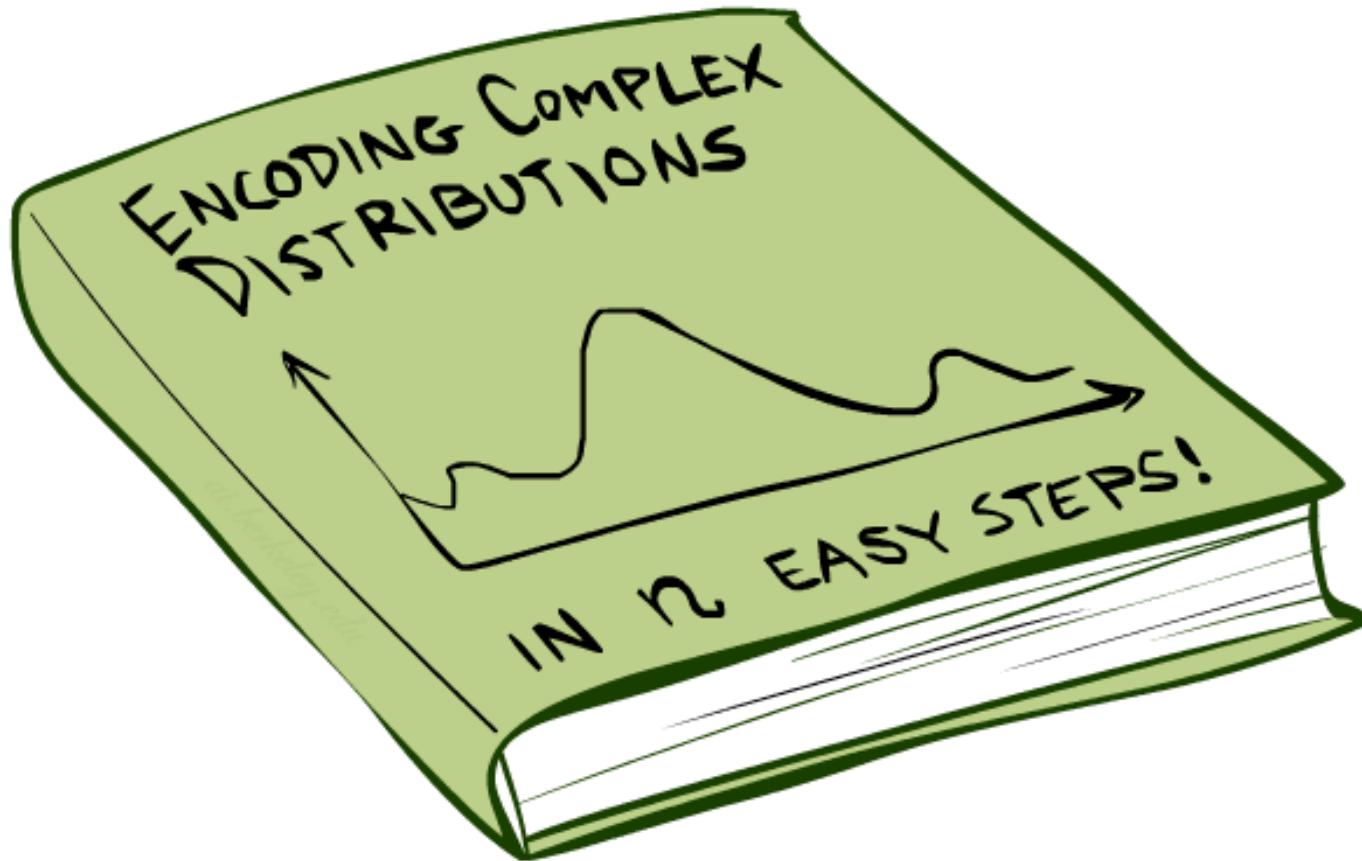


$$P(T, B, G) = P(G) P(T|G) P(B|G)$$

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

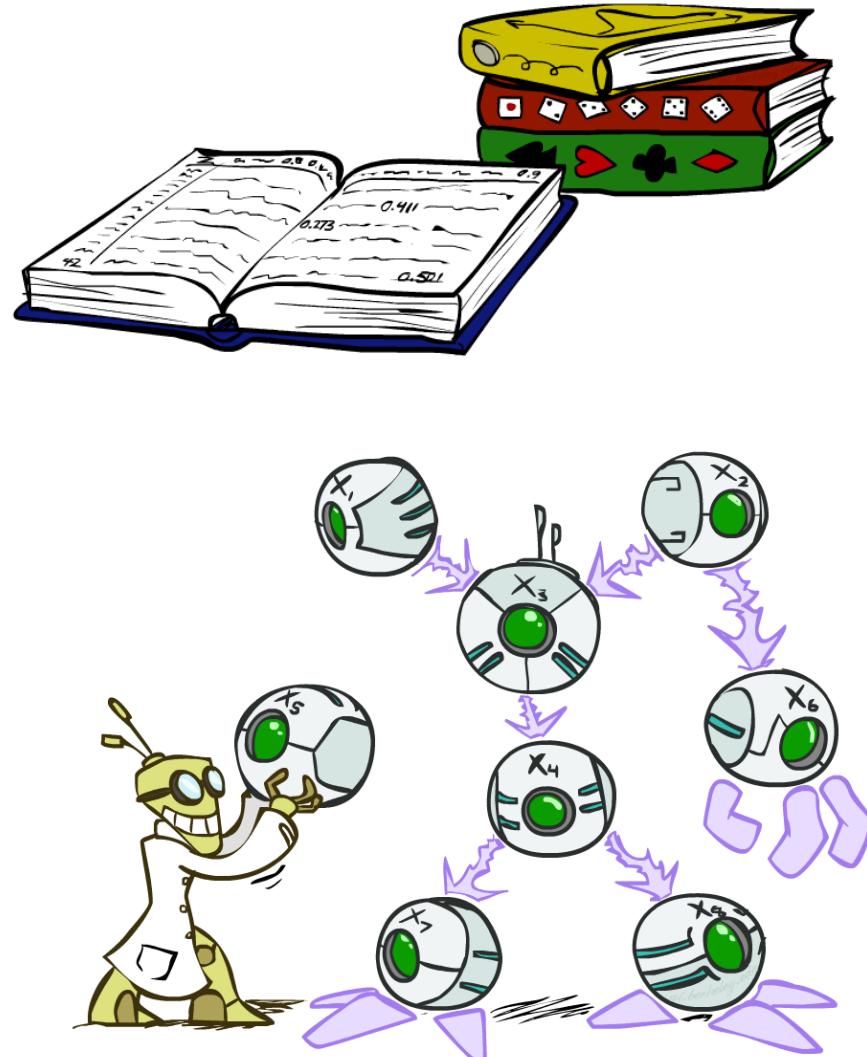


Bayes'Nets: Big Picture

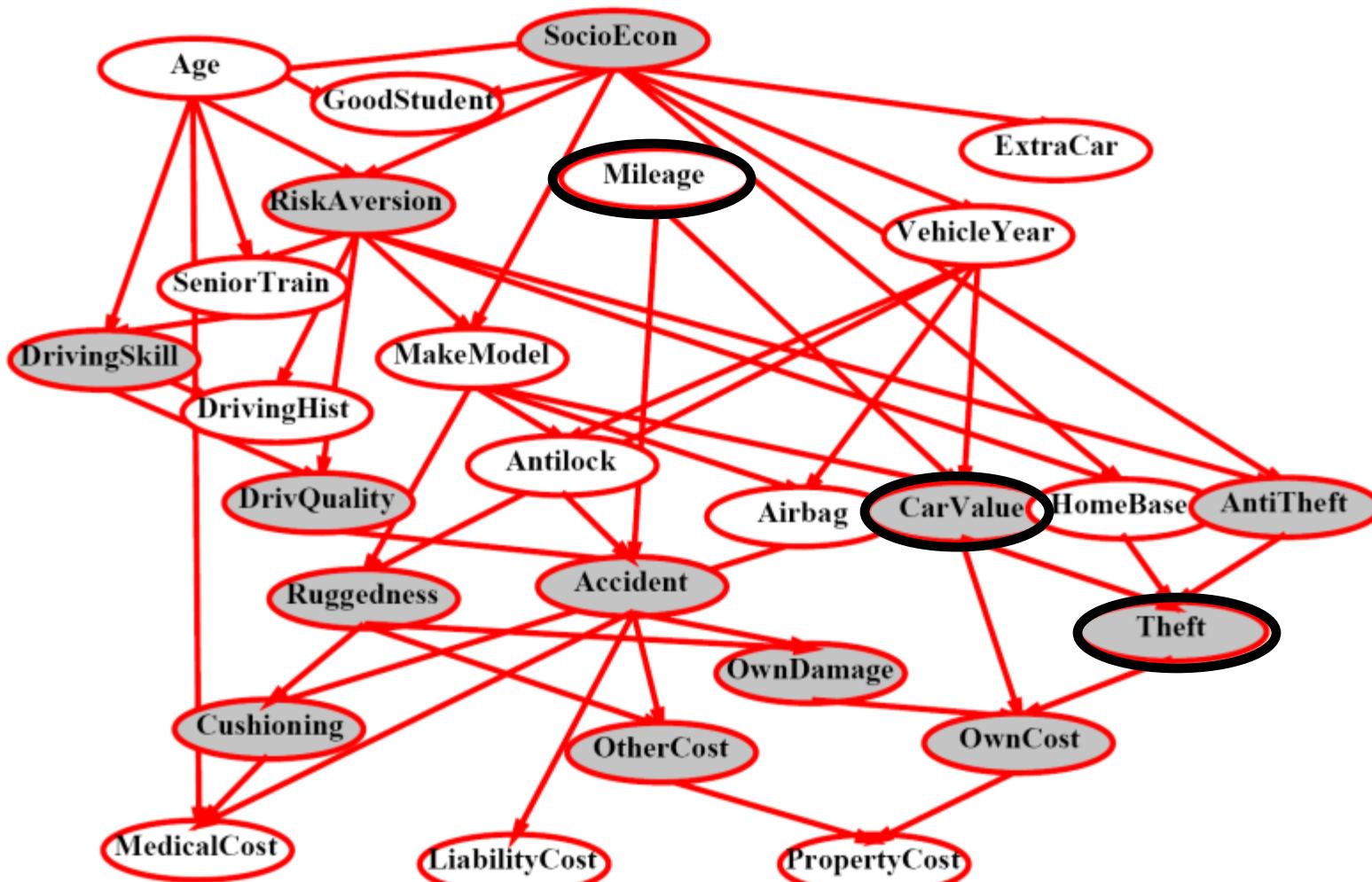


Bayes' Nets: Big Picture

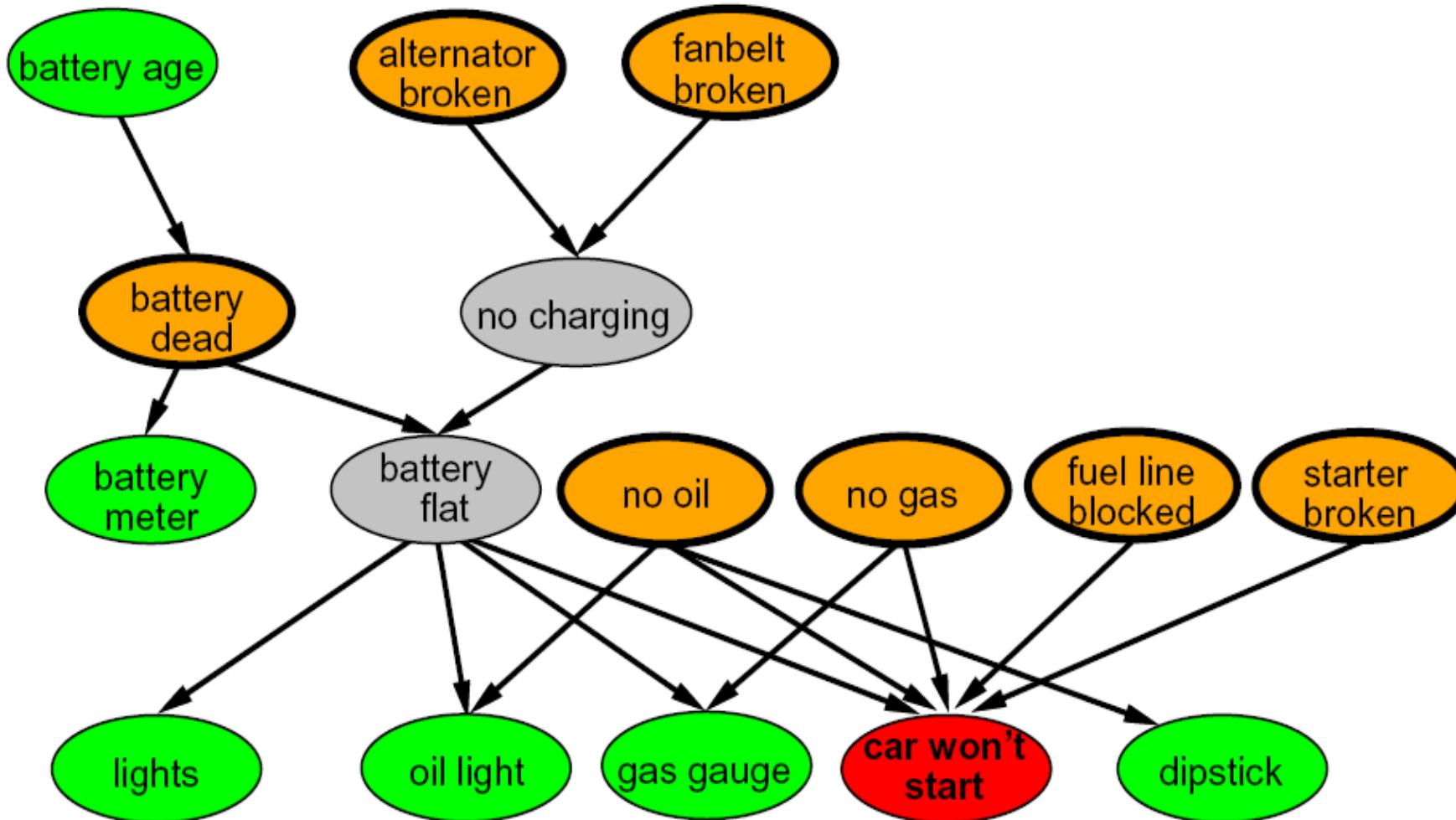
- Two problems with using full joint distribution tables as our probabilistic models:
 - Unless there are only a few variables, the joint is WAY too big to represent explicitly
 - Hard to learn (estimate) anything empirically about more than a few variables at a time
- Bayes' nets: a technique for describing 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
 - For about 10 min, we'll be vague about how these interactions are specified



Example Bayes' Net: Insurance

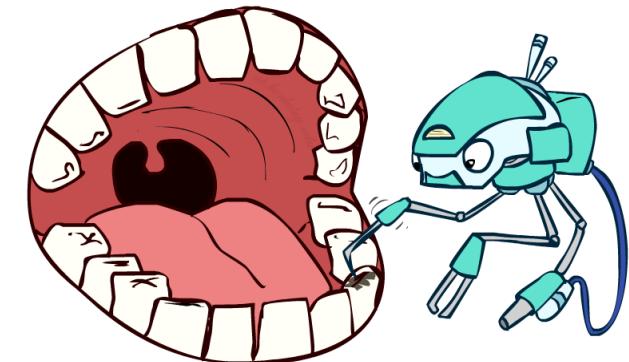
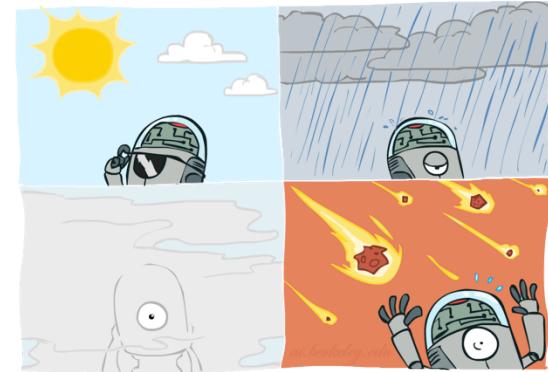
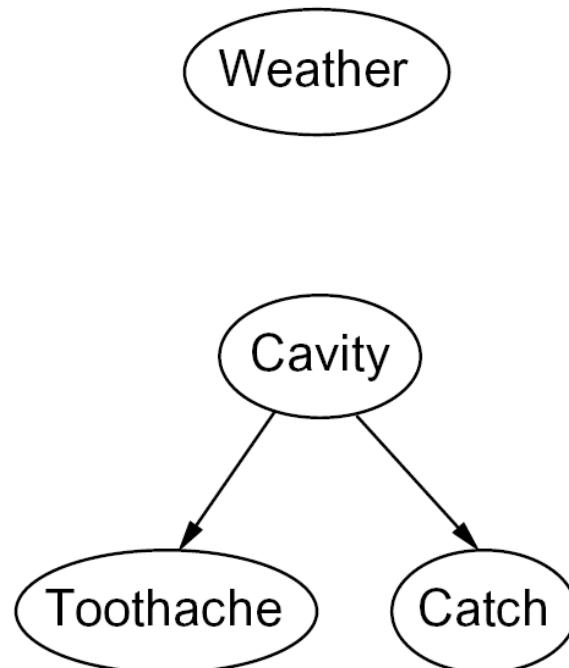


Example Bayes' Net: Car



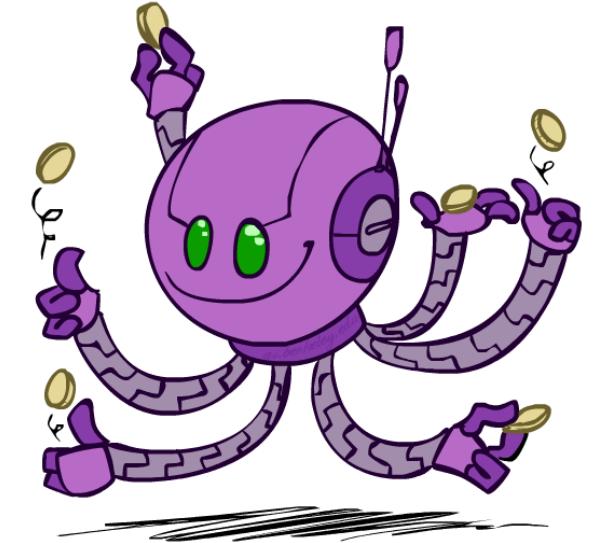
Graphical Model Notation

- 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



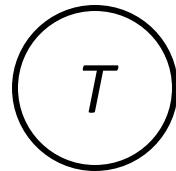
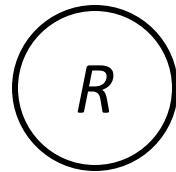
- No interactions between variables: **absolute independence**

Example: Traffic

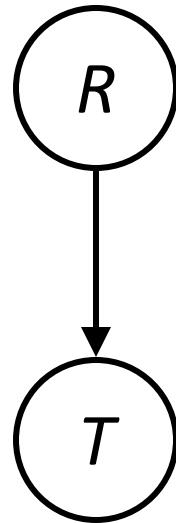
- Variables:
 - R : It rains
 - T : There is traffic



- Model 1: independence



- Model 2: rain causes traffic



- When is an agent using model 2 better?

Example: Traffic II

- Let's build a causal graphical model!

- Variables

- T: Traffic
- R: It rains
- L: Low pressure
- D: Roof drips
- B: Ballgame
- C: Cavity



Example: Traffic II

- Let's build a causal graphical model!

- Variables:

- T: Traffic

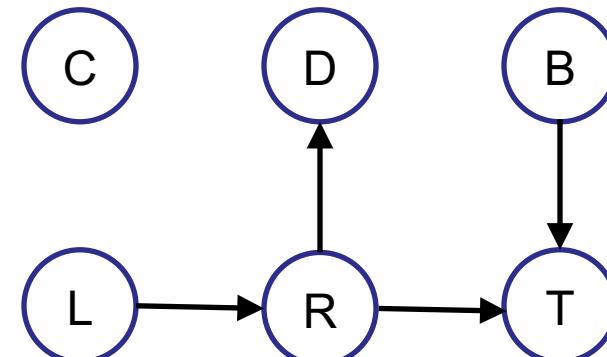
- R: It Rains

- L: Low Pressure

- D: Roof Drips

- B: Ball Game

- C: Cavity



Bayes' Network for Alarm Network?

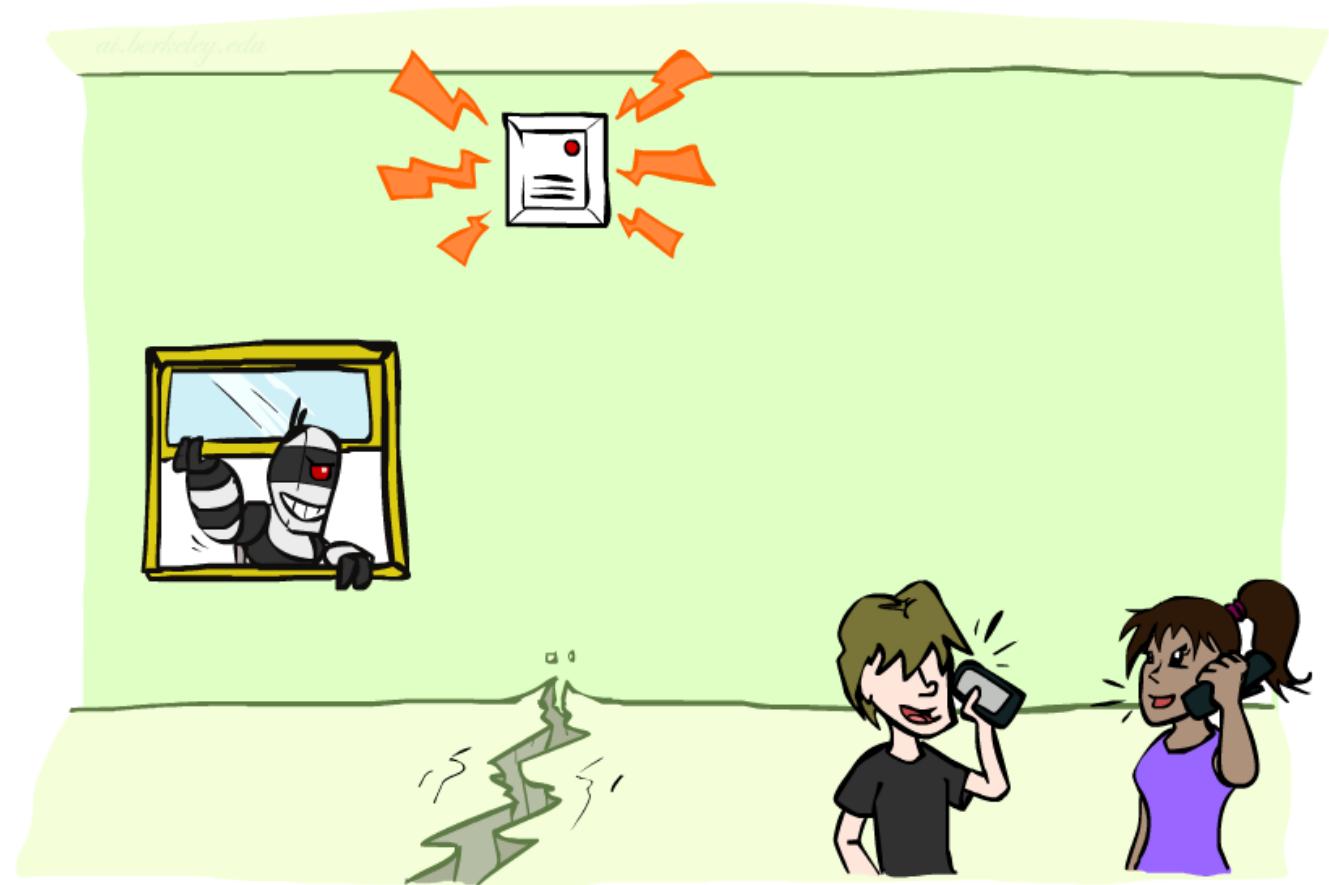


■ Variables

- B: Burglary
- A: Alarm goes off
- M: Mary calls
- J: John calls
- E: Earthquake!

■ Model Assumptions

- Alarm can go off if Burglary or Earthquake
- Mary and John call if they hear the alarm



Bayes' Network for Alarm Network?

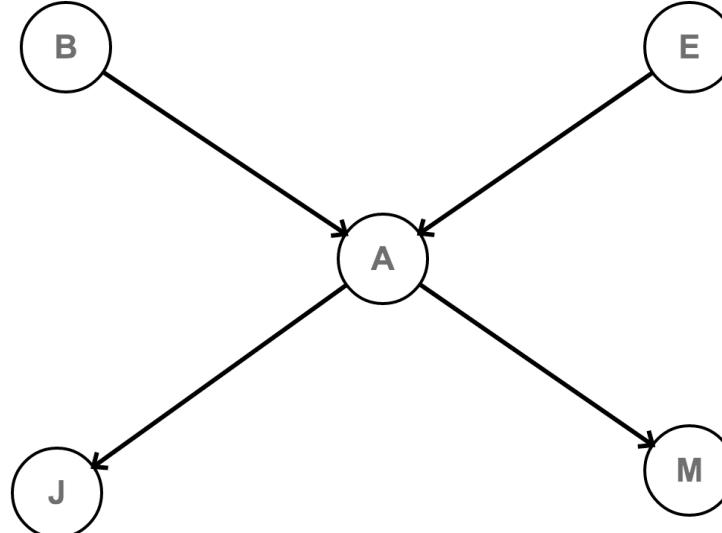


■ Variables

- B: Burglary
- A: Alarm goes off
- M: Mary calls
- J: John calls
- E: Earthquake!

■ Model Assumptions

- Alarm can go off if Burglary or Earthquake
- Mary and John call if they hear the alarm



Bayes' Net Semantics



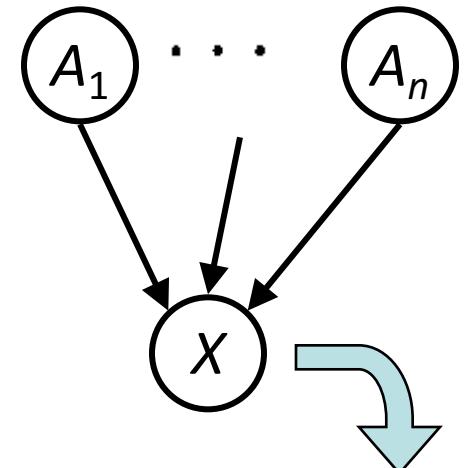


Bayes' Net Semantics

- 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
- Description of a noisy “causal” process



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

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

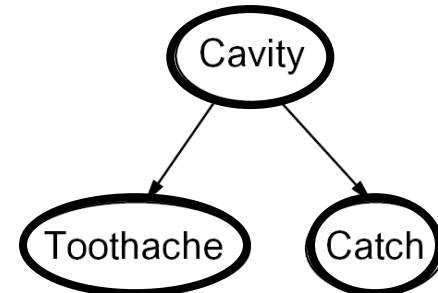
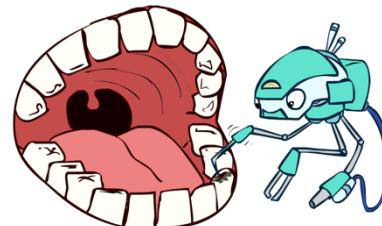
Probabilities in BNs



- 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:

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

- Example:



$$P(+\text{cavity}, +\text{catch}, -\text{toothache}) = P(+\text{cavity}) \ P(-\text{toothache} | +\text{cavity}) \ P(+\text{catch} | +\text{cavity})$$

Probabilities in BNs



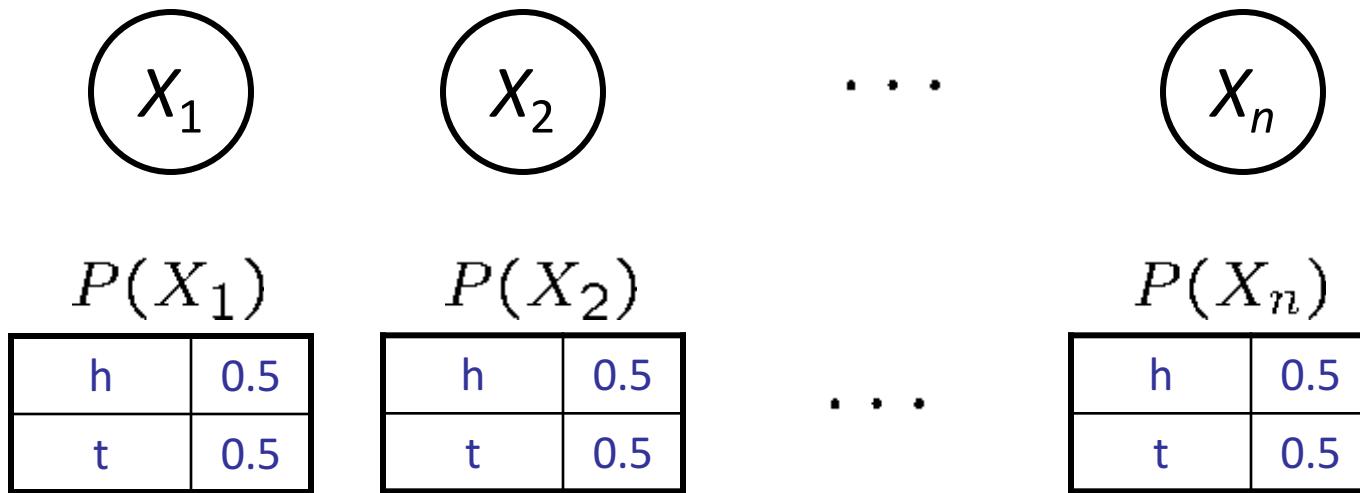
- Why are we guaranteed that setting

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

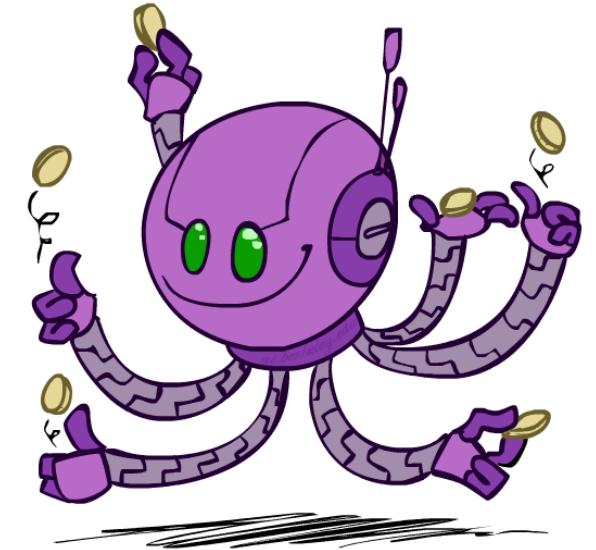
results in a proper joint distribution?

- Chain rule (valid for all distributions): $P(x_1, x_2, \dots, x_n) = \prod_{i=1}^n P(x_i | x_1 \dots x_{i-1})$
- Assume conditional independences: $P(x_i | x_1, \dots, x_{i-1}) = P(x_i | \text{parents}(X_i))$
→ Consequence: $P(x_1, x_2, \dots, x_n) = \prod_{i=1}^n P(x_i | \text{parents}(X_i))$
- Not every BN can represent every joint distribution
 - The topology enforces certain conditional independencies

Example: Coin Flips

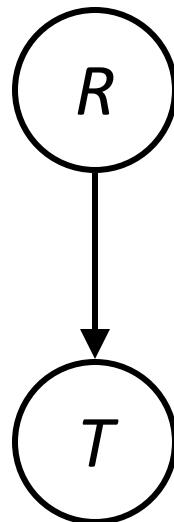
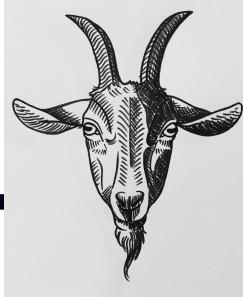


$$P(h, h, t, h) = P(X_1 = h) P(X_2 = h) P(X_3 = t) P(X_4 = h)$$



Only distributions whose variables are absolutely independent can be represented by a Bayes' net with no arcs.

Probability of Rain and No Traffic?



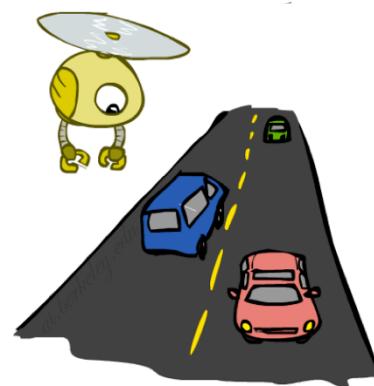
$P(R)$

+r	1/4
-r	3/4

$$\mathbf{P(+r, -t) = P(+r) P(-t | +r) = \frac{1}{4} * \frac{1}{4}}$$

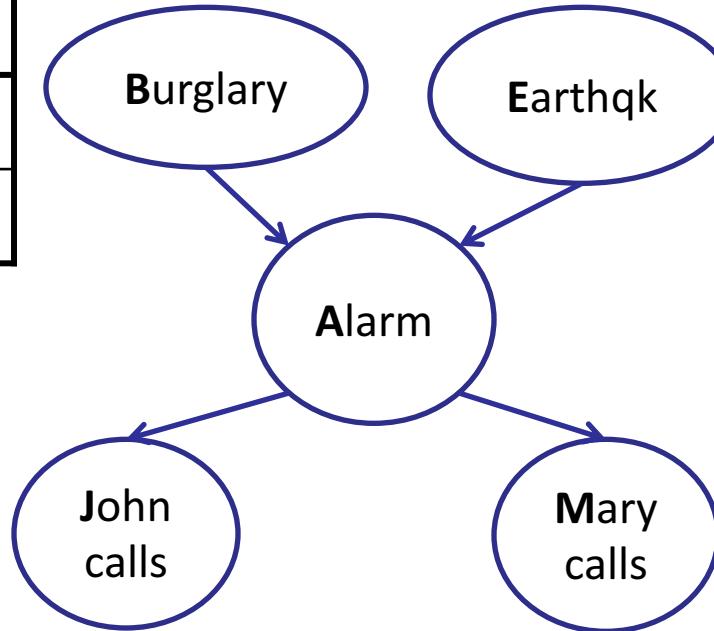
$P(T|R)$

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



Example: Alarm Network

B	P(B)
+b	0.001
-b	0.999



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

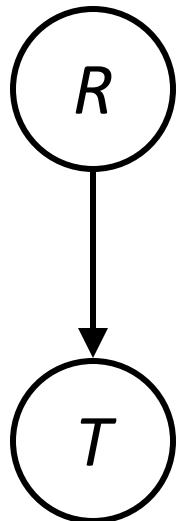
E	P(E)
+e	0.002
-e	0.998



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

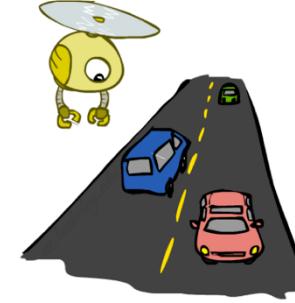
- Causal direction

 $P(R)$

+r	1/4
-r	3/4

 $P(T|R)$

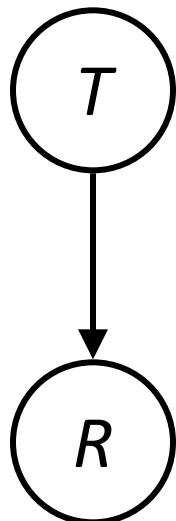
+r	+t	3/4
	-t	1/4
-r	+t	1/2
	-t	1/2

 $P(T, R)$

+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



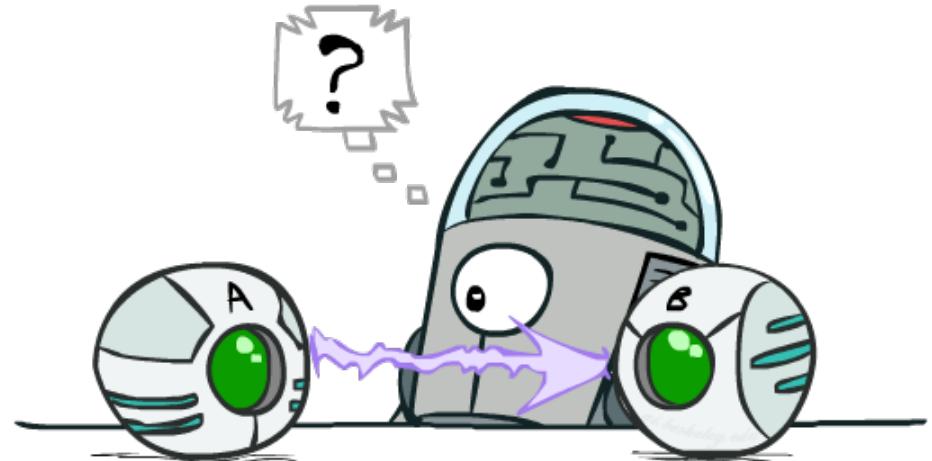
$P(T, R)$

+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**

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



Bayes' Nets

- So far: How a Bayes' net encodes a joint distribution
 - First assembled BNs using an intuitive notion of conditional independence as causality
 - Then saw that key property is conditional independence
- Next lecture: How to answer numerical queries (inference)

