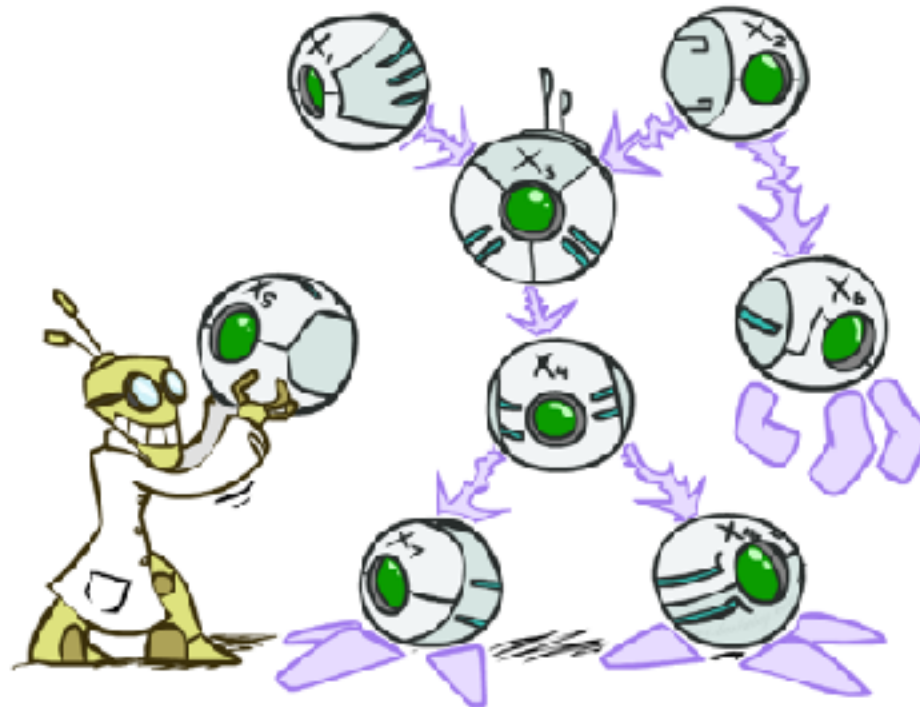# CS 383: Artificial Intelligence

Bayes Nets: Representation

Prof. Scott Niekum — UMass Amherst

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
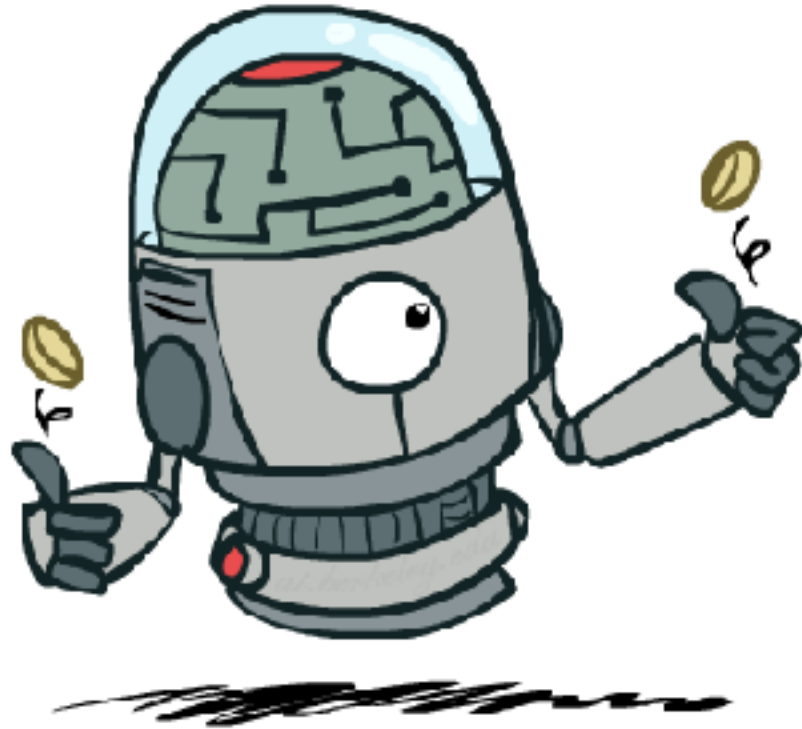
Independence

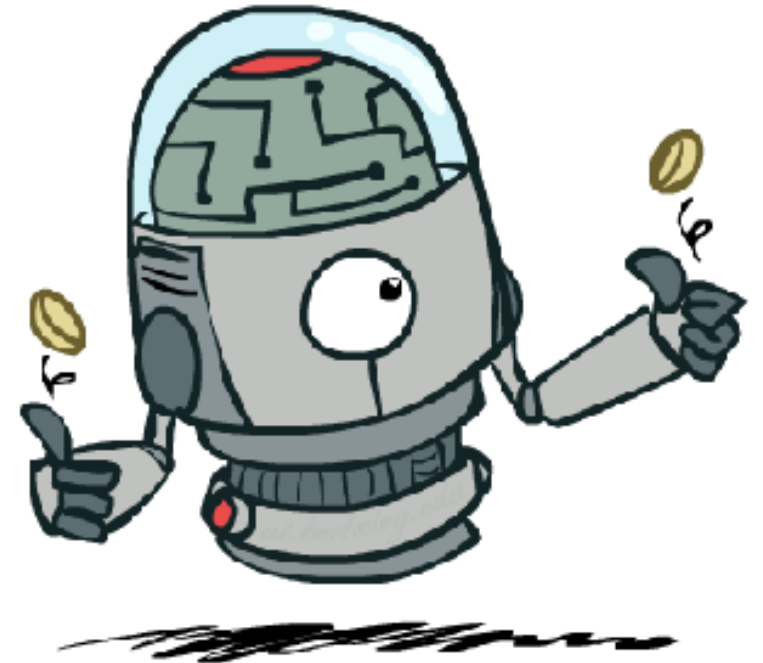- Two variables are *independent* if:

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

  - 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

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

Example: Independence?

$P(T)$

T	P
hot	0.5
cold	0.5

$P_1(T, W)$

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

$P_2(T, W)$

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

$P(W)$

W	P
sun	0.6
rain	0.4

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

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

2^n
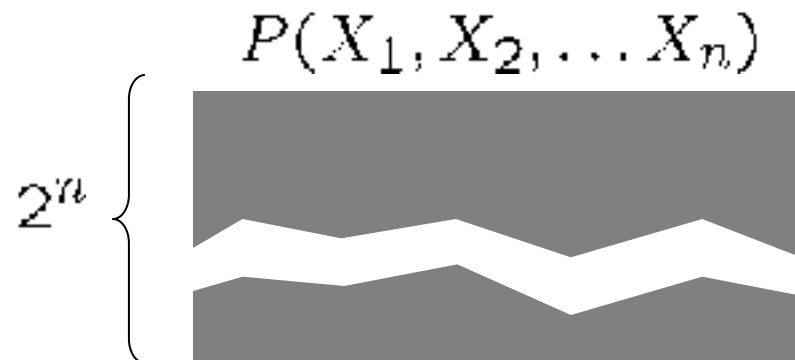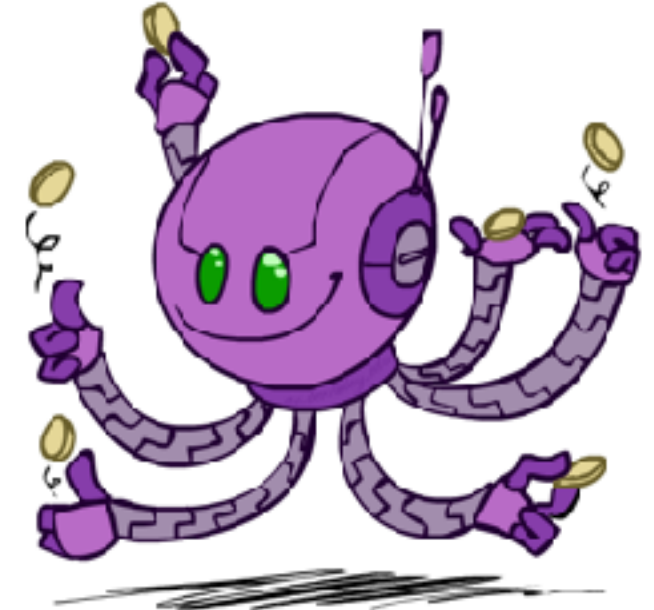
Conditional Independence

- 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
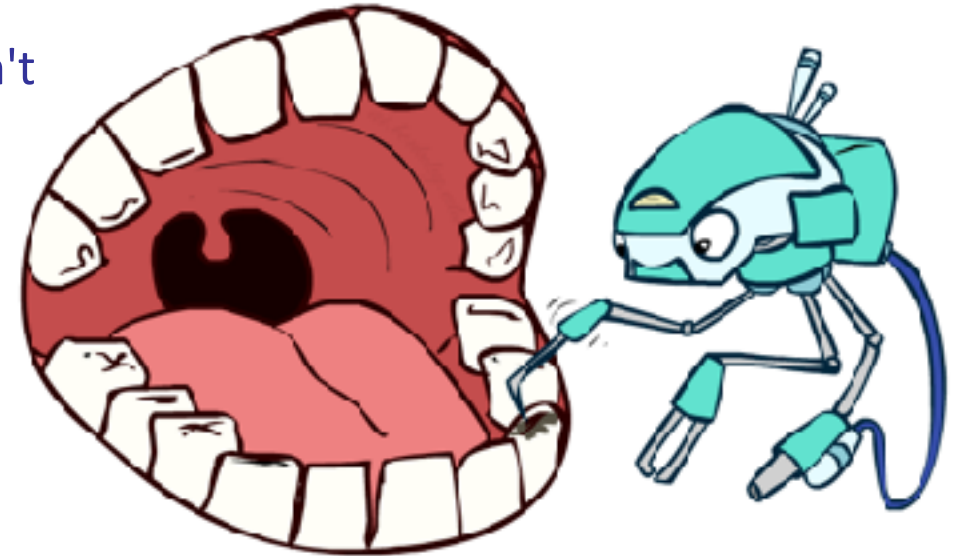
Conditional Independence

- Unconditional (absolute) independence very rare (why?)

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

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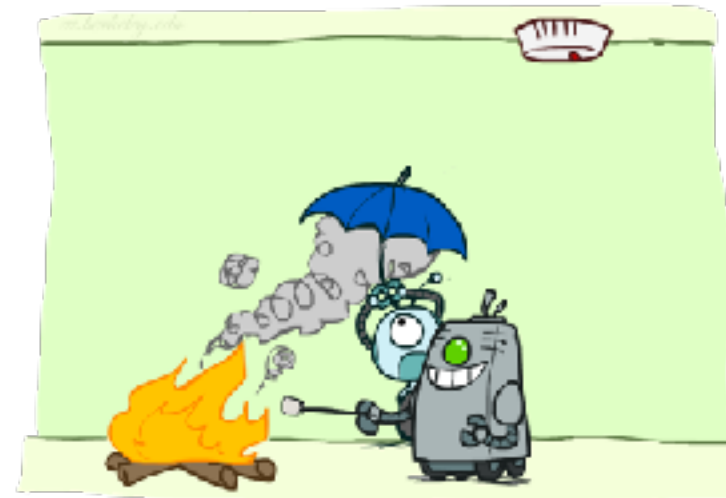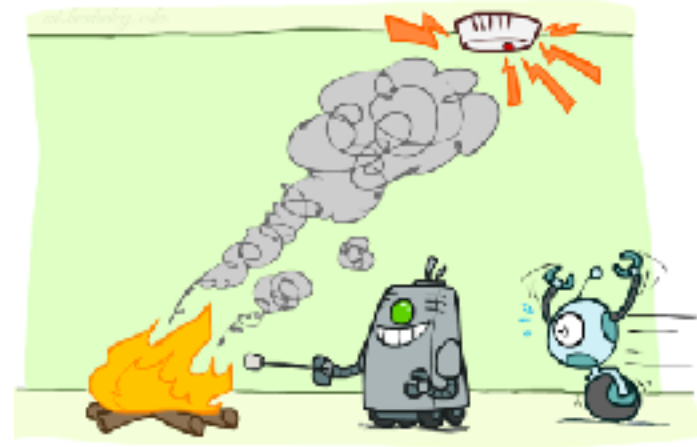$$\forall x, y, z : P(x | z, y) = P(x | z)$$

Conditional Independence

- What about this domain:

    - Traffic
    - Umbrella
    - Raining

Conditional Independence

- What about this domain:

  - Fire
  - Smoke
  - Alarm

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:

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

- 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

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

0.50

0.50

Bayes Nets: Big Picture
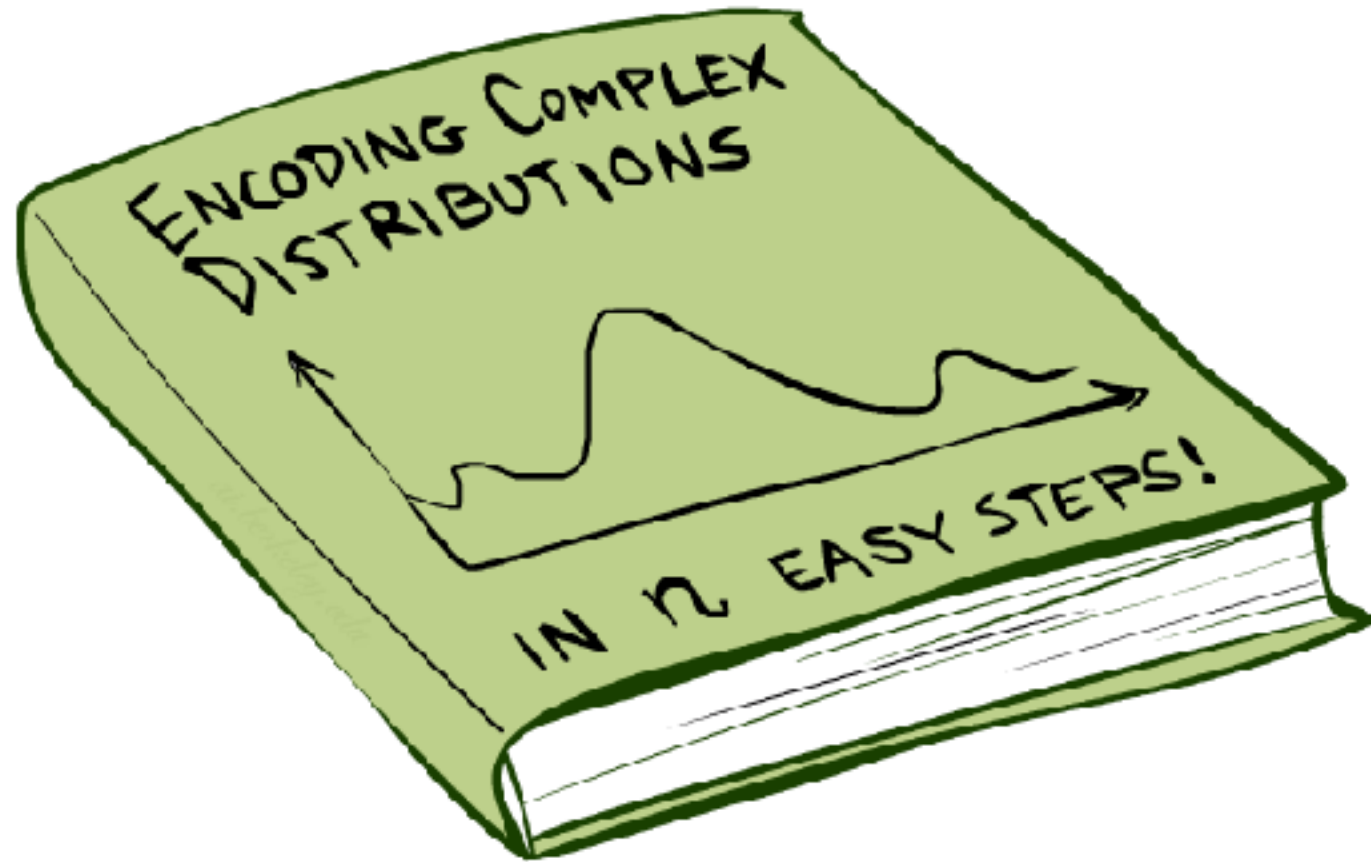
Bayes Nets: Big Picture

- 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
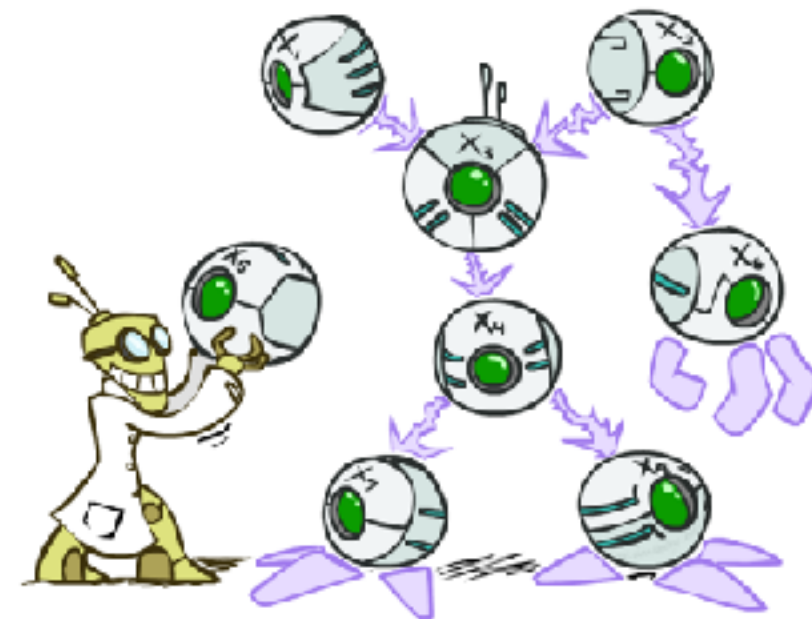
Example Bayes Net: Insurance
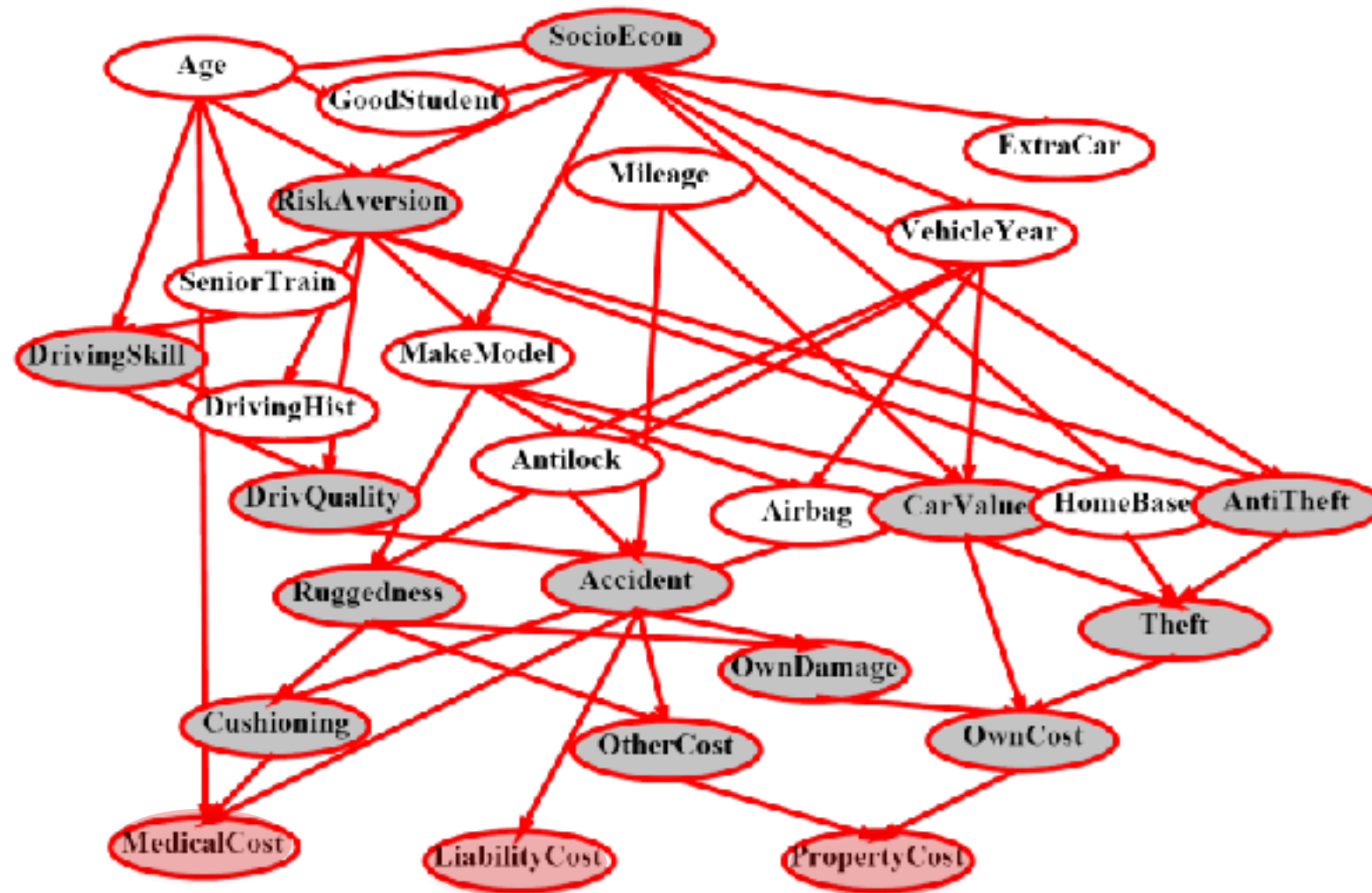
Example Bayes Net: Car
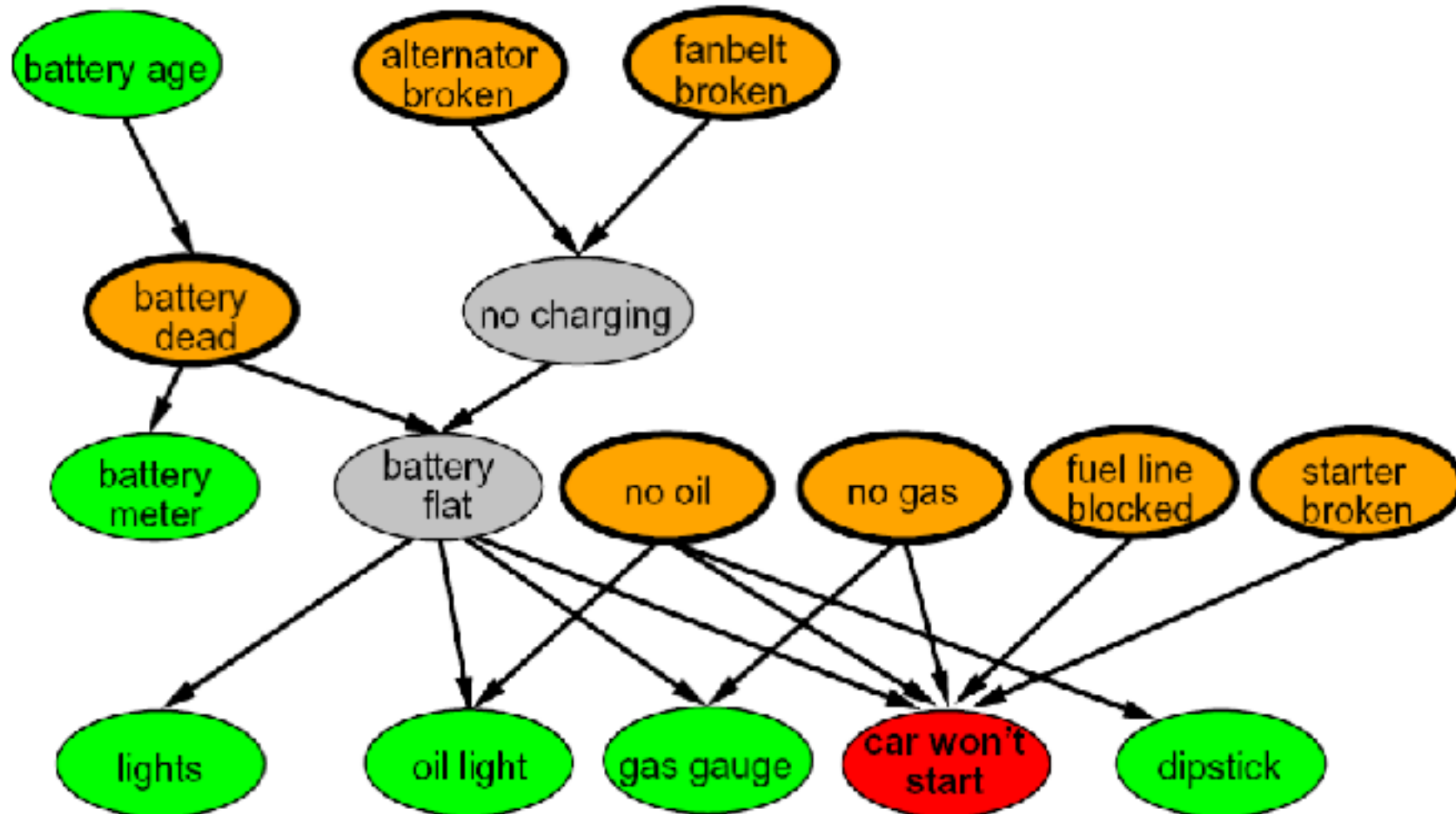
Graphical Model Notation

- **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!)**
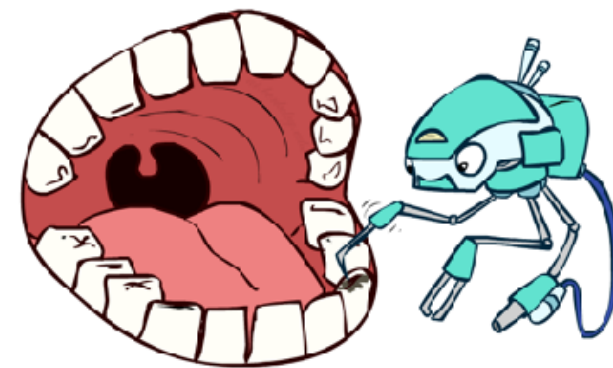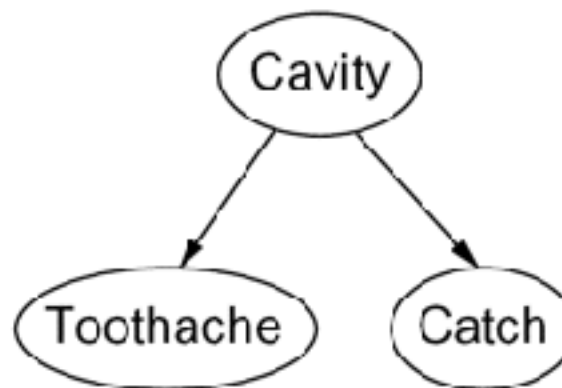
Example: Coin Flips
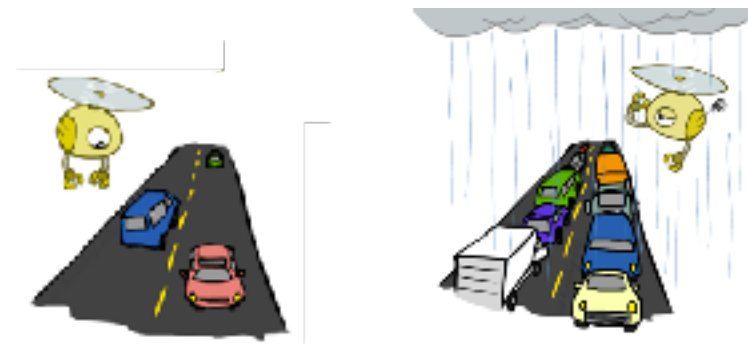
- N independent coin flips

X_1 X_2 . . . X_n
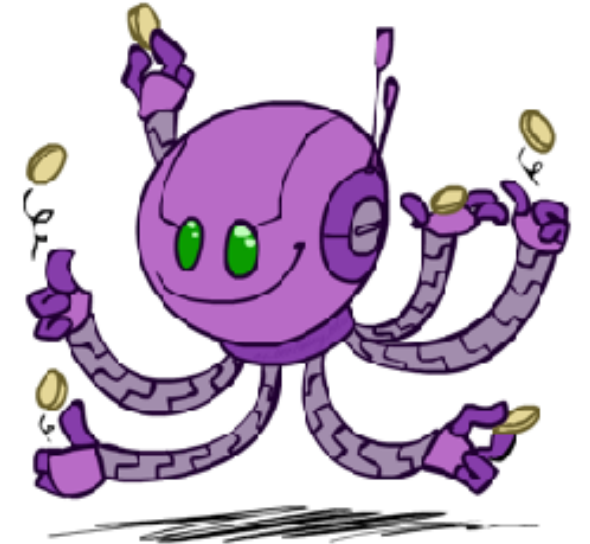
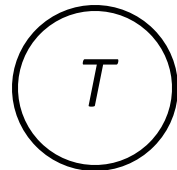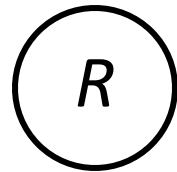- No interactions between variables: absolute independence
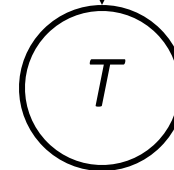
Example: Traffic
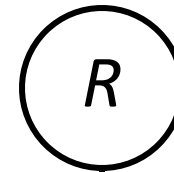
- Variables:
  - R: It rains
  - T: There is traffic

- Model 1: independence

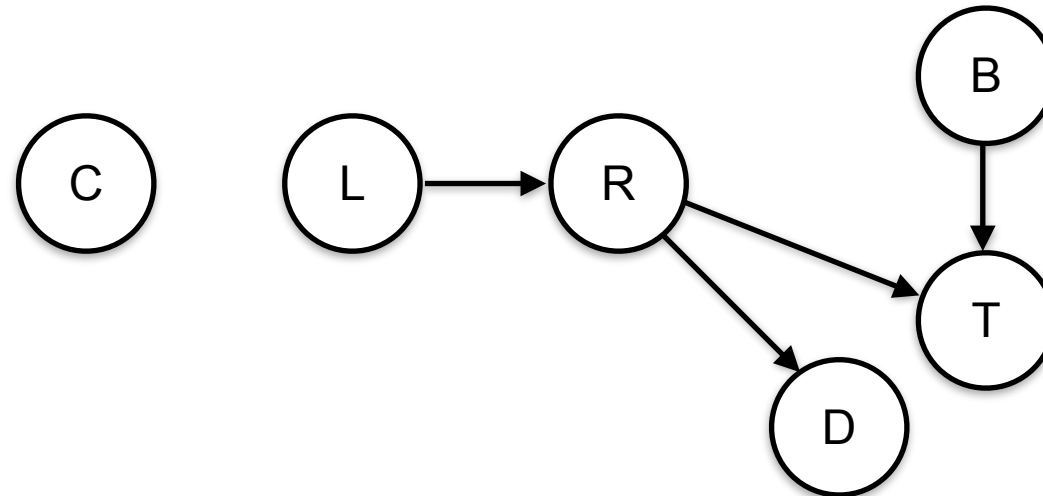- Model 2: rain causes traffic

- Why is an agent using model 2 better?
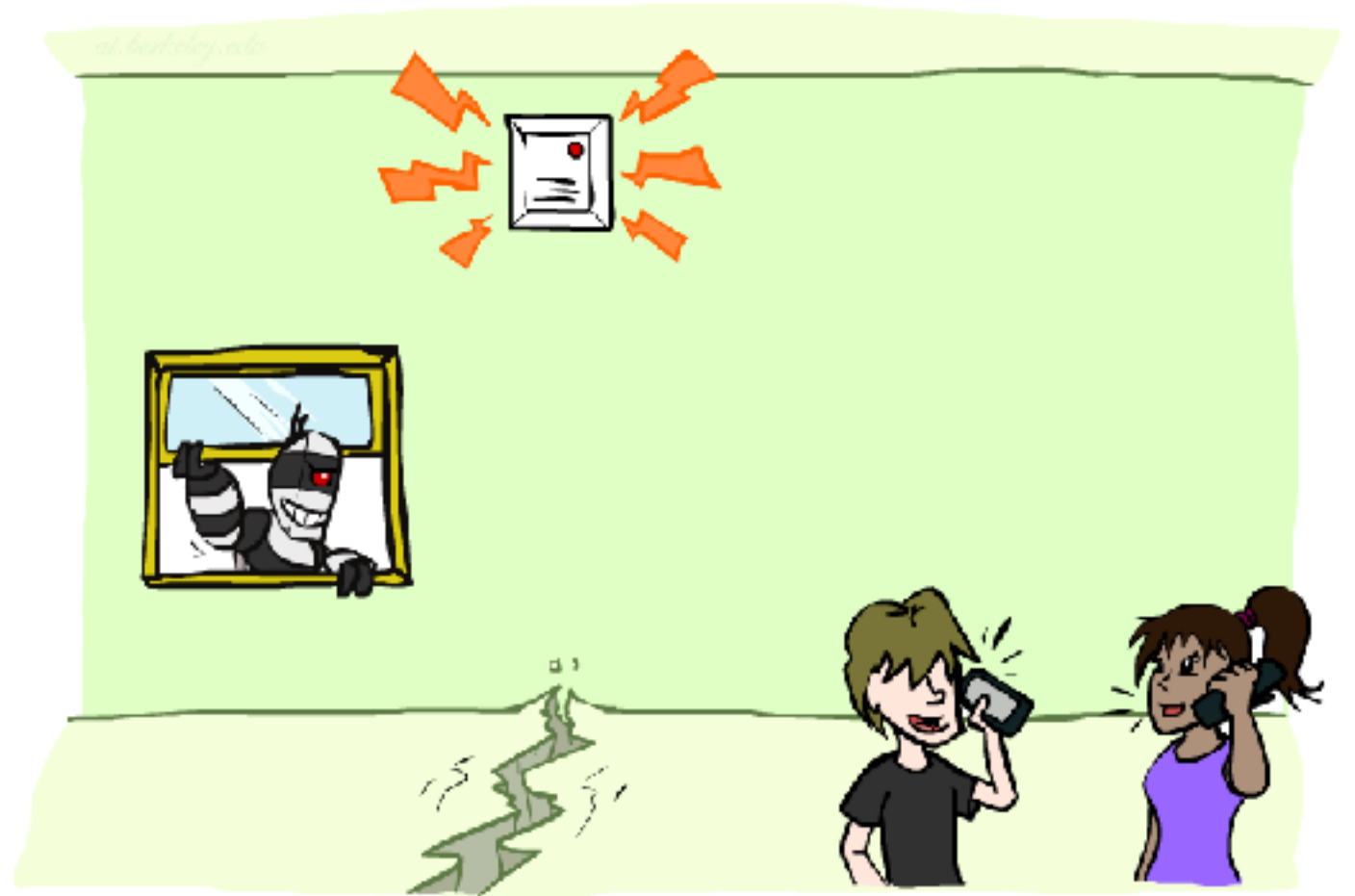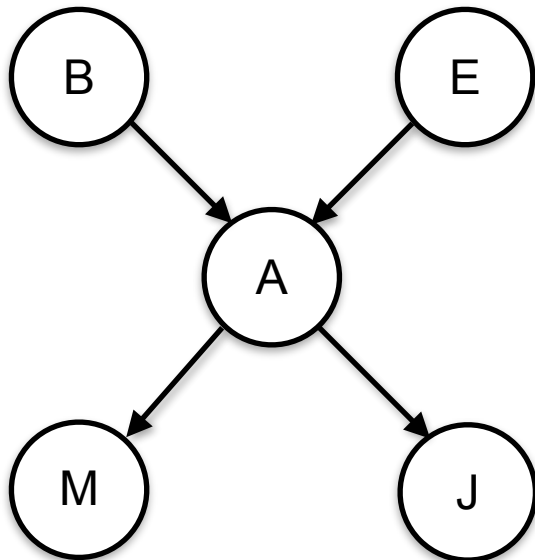
Example: Traffic II

- Let's build a causal graphical model!
- Variables
  - T: Traffic
  - R: It rains
  - L: Low pressure
  - D: Roof drips
  - B: Ballgame
  - C: Cavity

Example: Alarm Network

- 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

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

  - CPT: conditional probability table

  - Description of a noisy "causal" process

$A_1 \cdots A_n$

X

$$P(X|A_1 \ldots A_n)$$
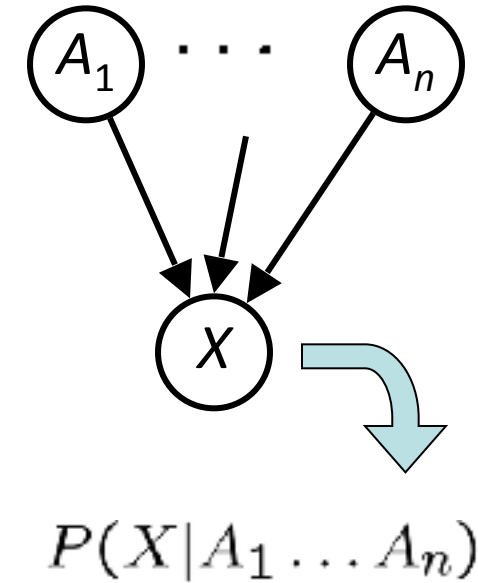
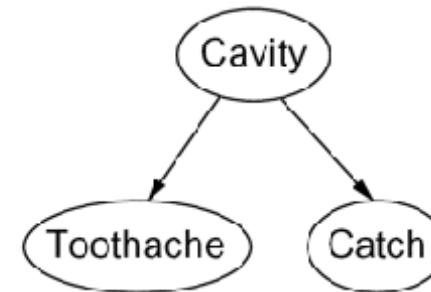*A Bayes net = Topology (graph) + Local Conditional Probabilities*

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:

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

  - Example:

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

Probabilities in BNs

- Why are we guaranteed that setting

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

results in a proper joint distribution?

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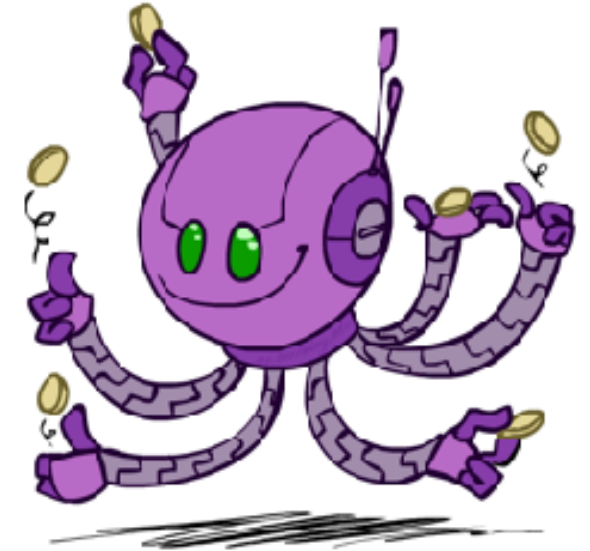
$$P(x_1, x_2, \ldots x_n) = \prod_{i=1}^{n} P(x_i | x_1 \ldots x_{i-1})$$

- <u>Assume</u> conditional independences:

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

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

- Not every BN can represent every joint distribution

  - The topology enforces certain conditional independencies

Example: Coin Flips

X_1 X_2 . . . 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
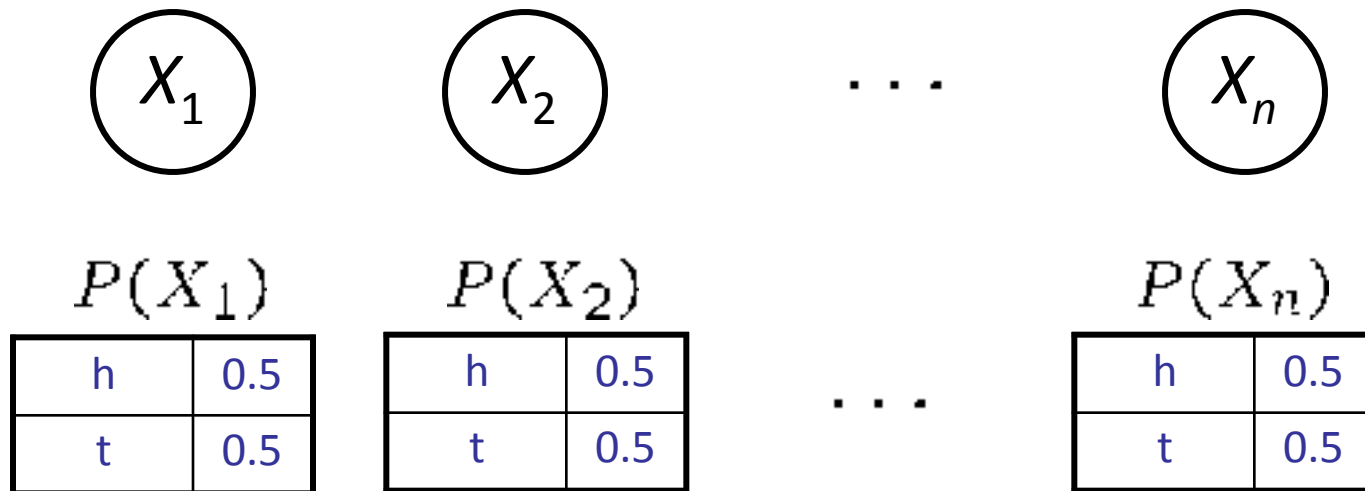
$P(h, h, t, h) =$ 0.5 * 0.5 * 0.5 * 0.5

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

Example: Traffic

$P(R)$
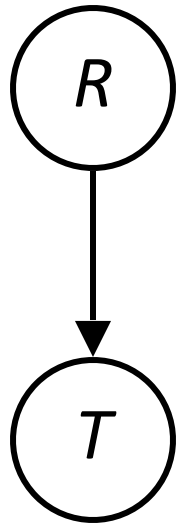
+r	1/4
-r	3/4

$P(T|R)$

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

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

$P(+r, -t) =$ 1/4 * 1/4

Example: Alarm Network

B	P(B)
+b	0.001
-b	0.999

E	P(E)
+e	0.002
-e	0.998

Burglary   **E**arthqk

Alarm

John calls   **M**ary calls

P(+b, +e, -a, +j, -m) = ?

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

R → T

$P(T|R)$

+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?
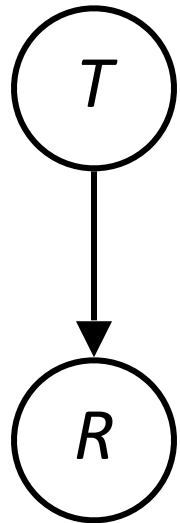
$P(T)$

+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 *Roof 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

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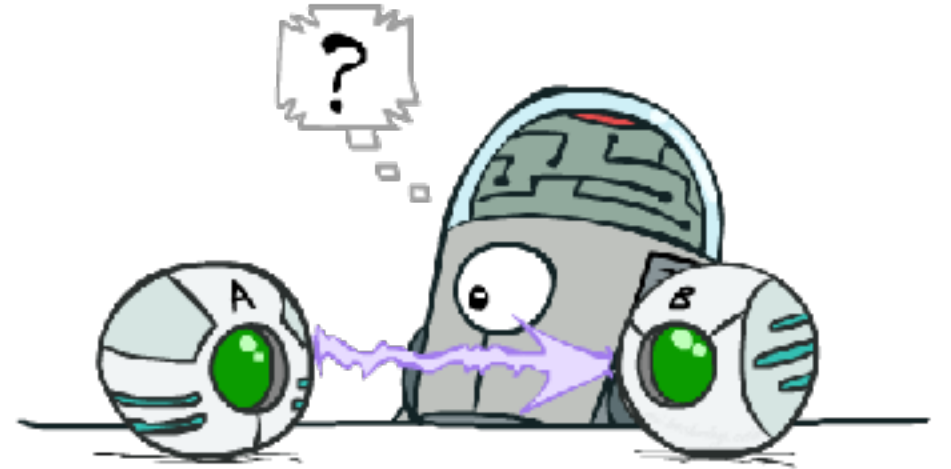
Bayes Nets

- So far: how a Bayes net encodes a joint distribution

- Next: how to answer queries about that distribution
  - Today:
    - First assembled BNs using an intuitive notion of conditional independence as causality
    - Then saw that key property is conditional independence
  - Main goal: answer queries about conditional independence and influence

- After that: how to answer numerical queries (inference)