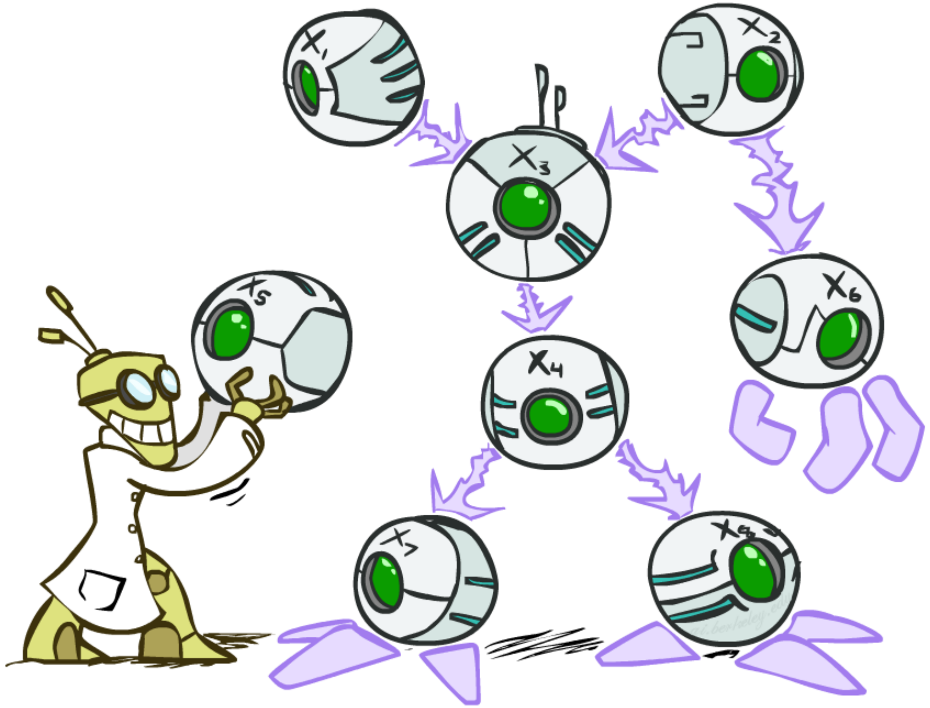
# CS 343: Artificial Intelligence

## Bayes Nets: Representation & Independence




Profs. Peter Stone and Yuke Zhu — The University of Texas at Austin

[These slides based on those of Dan Klein and Pieter Abbeel for CS188 Intro to AI at UC Berkeley. All CS188 materials are available at http://ai.berkeley.edu.]

# Announcements

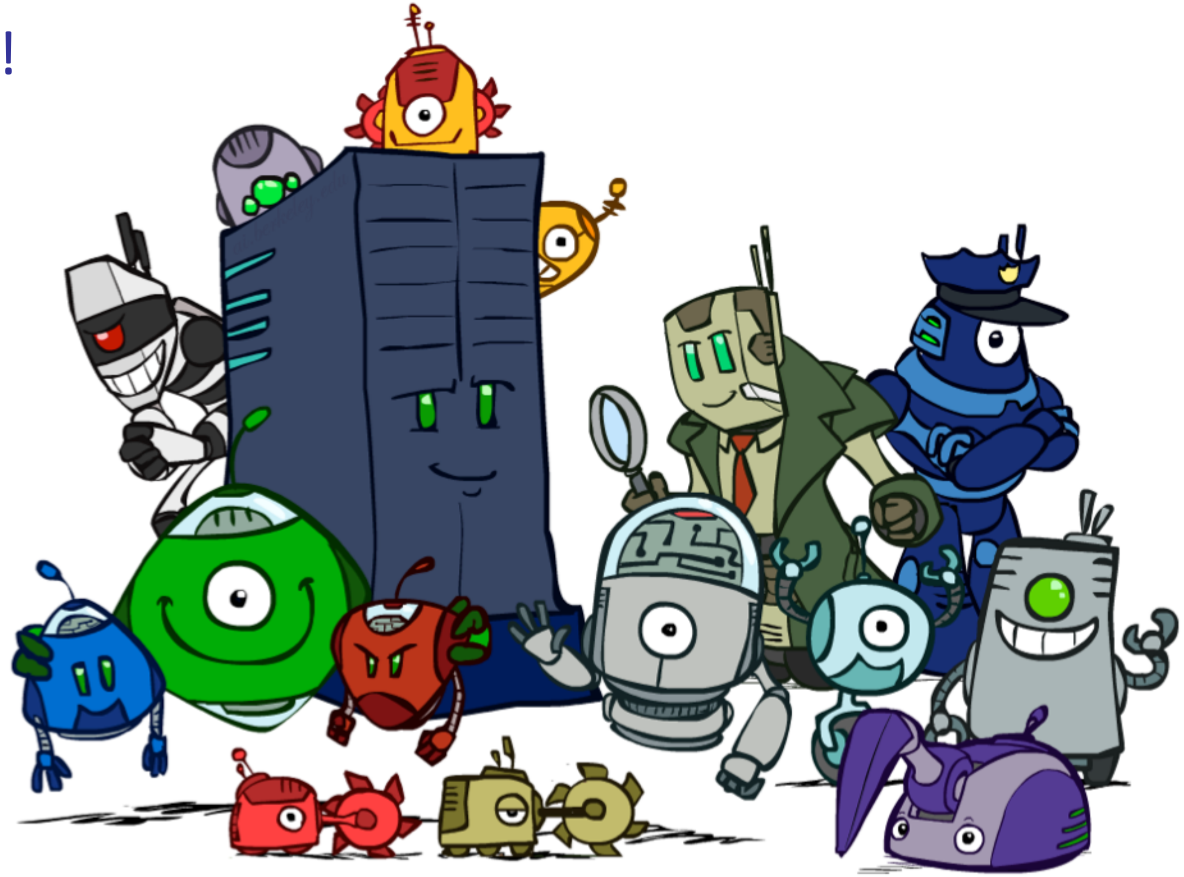
- Past Dues
  - HW1-4: Search, CSPs, Games, MDPs
  - 8 reading responses: AI100 report; 7 Textbook readings
  - P0,1, 2: tutorial, Search, Games
- Upcoming EdX Homeworks
  - HW5: RL – due Monday 3/22 at 11:59 pm
  - HW6: Bayes Nets – due Monday 4/5 at 11:59 pm
- Upcoming Programming Projects
  - P3: RL – due Wednesday 3/31 at 11:59pm
- Midterm Exam
  - Thursday 3/25 or so (Materials up to this week)

# CS343 Outline

- We're done with Part I: Search and Planning!

- We've seen how AI methods can solve problems in:
  - Search
  - Constraint Satisfaction Problems
  - Games
  - Markov Decision Problems
  - Reinforcement Learning

- Next up: Uncertainty and Learning!



# CS343 Outline

- We're done with Part I: Search and Planning!
- Part II: Probabilistic Reasoning
  - Diagnosis
  - Speech recognition
  - Tracking objects
  - Robot mapping
  - Genetics
  - Error correcting codes
  - … lots more!
- Part III: Machine Learning

# Probability Recap

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

- 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: $$\forall x, y : P(x,y) = P(x)P(y) \qquad X \perp\!\!\!\perp Y$$

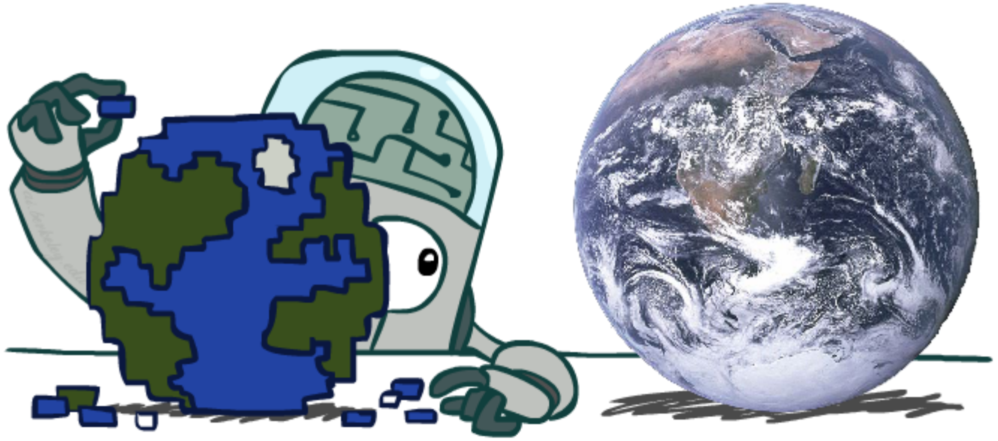
- X and Y are conditionally independent given Z if and only if:

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

# 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



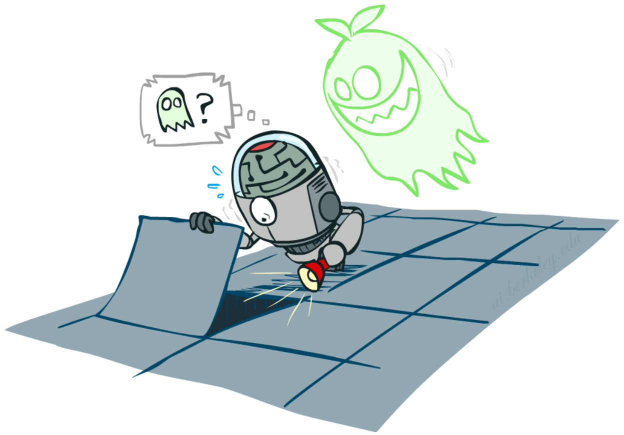
- What do we do with probabilistic models?
  - We (or our agents) need to reason about unknown variables, given evidence
  - Example: explanation (diagnostic reasoning)
  - Example: prediction (causal reasoning)
  - Example: value of information

# Probabilistic Models

- A probabilistic model is a joint distribution over a set of random variables
- Probabilistic models:
  - (Random) variables with domains
  - Assignments are called *outcomes*
  - Joint distributions: say whether assignments (outcomes) are likely
  - *Normalized:* sum to 1.0
  - Ideally: only certain variables directly interact
- Constraint satisfaction problems:
  - Variables with domains
  - Constraints: state whether assignments are possible
  - Ideally: only certain variables directly interact

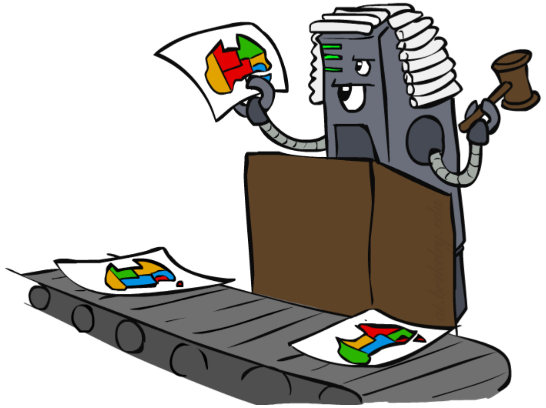
Distribution over T,W

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



Constraint over T,W

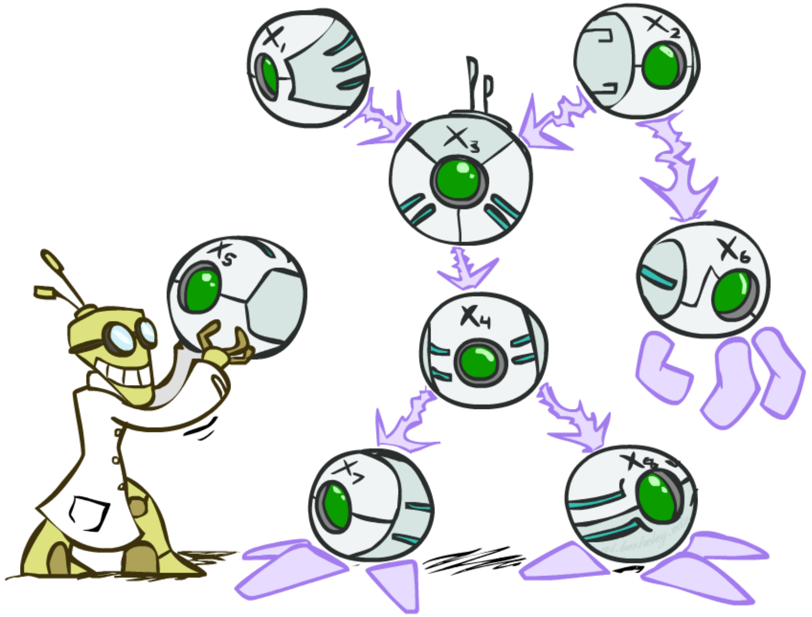
| T | W | P |
|---|---|---|
| hot | sun | T |
| hot | rain | F |
| cold | sun | F |
| cold | rain | T |



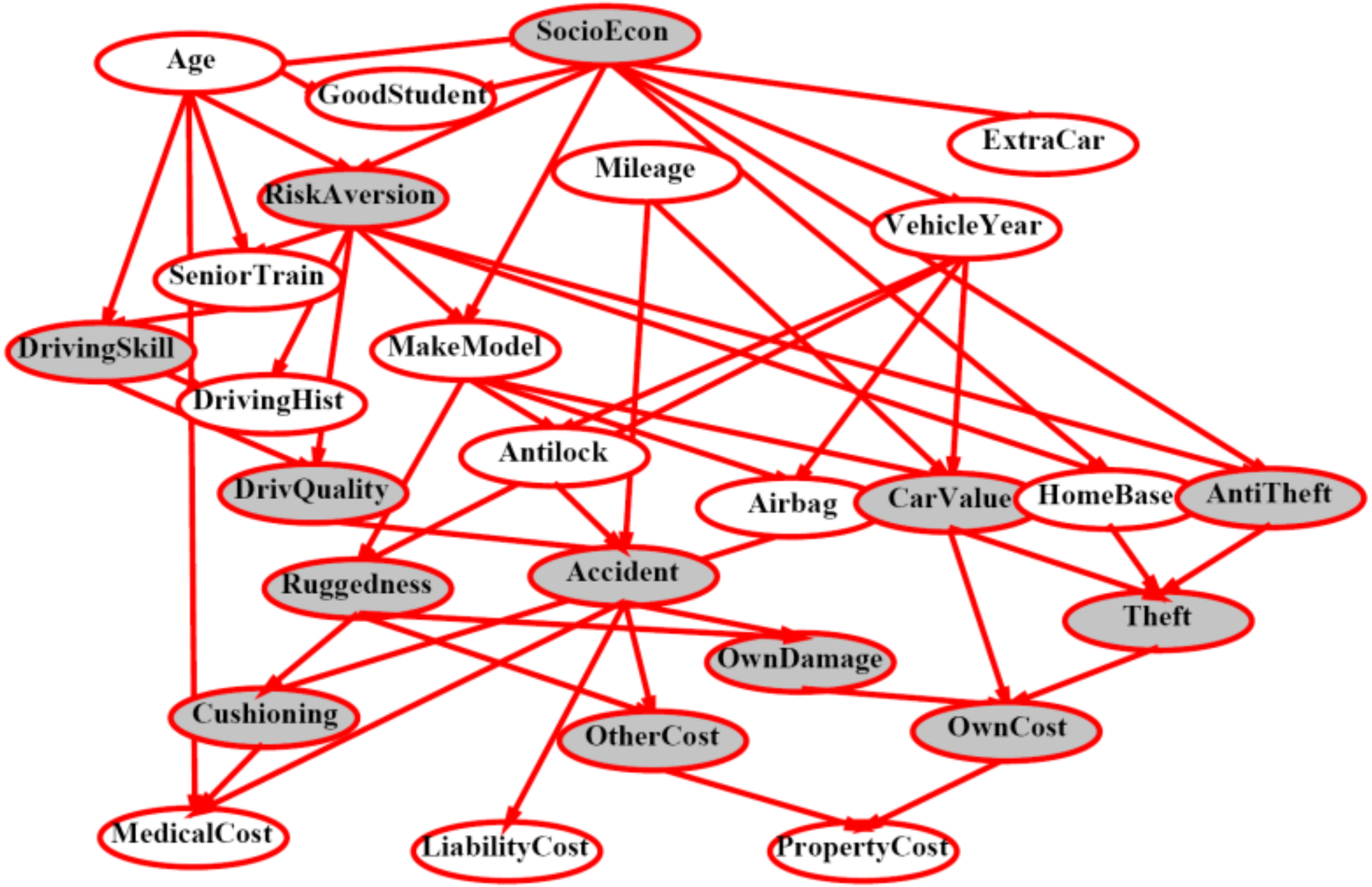
# Bayes Nets: Big Picture

- 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

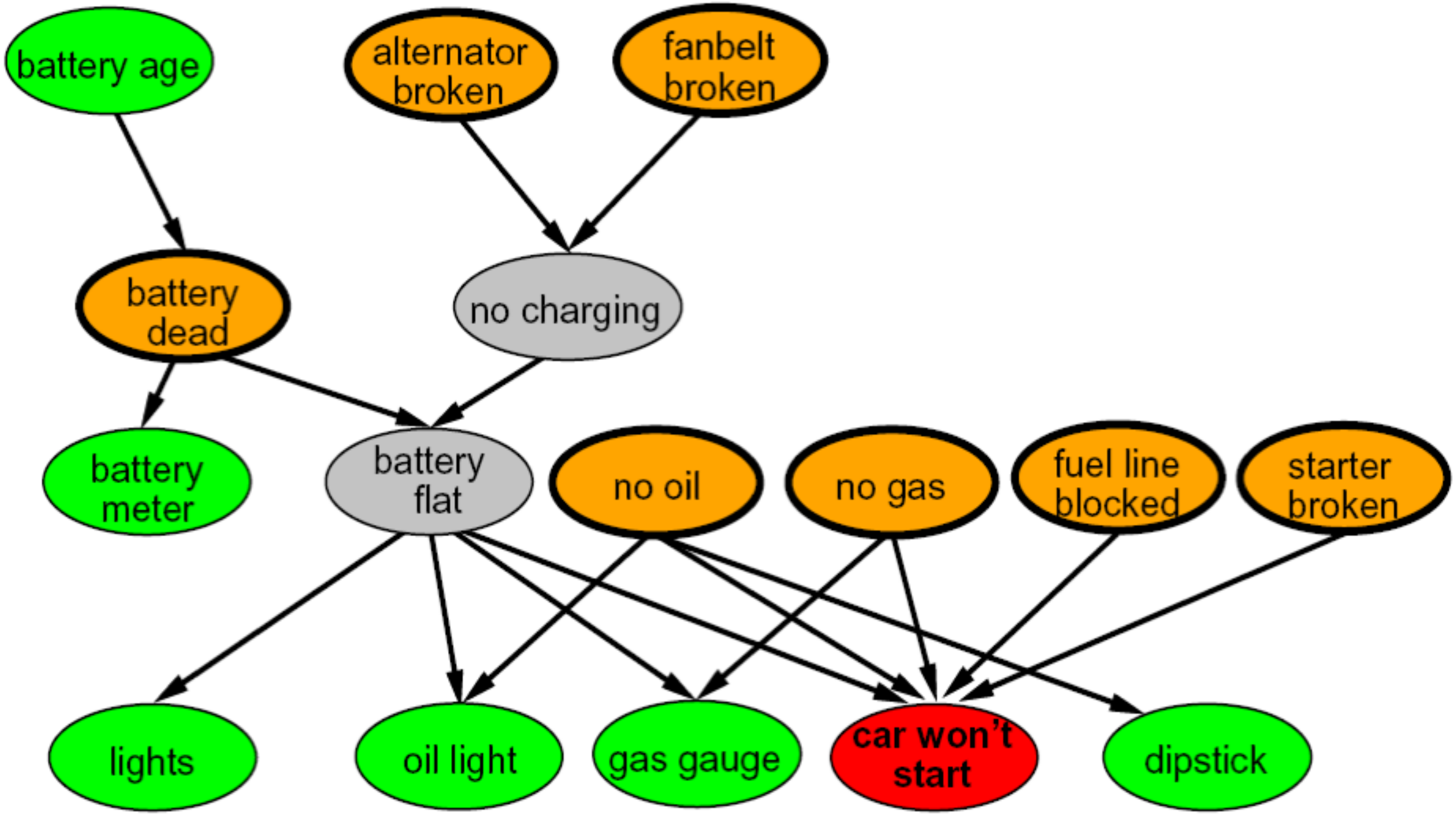
- Bayes nets: a technique for describing complex joint distributions (models) using simple, local distributions (conditional probabilities)
  - More properly called graphical models
  - We describe how variables locally interact
  - Local interactions chain together to give global, indirect interactions
  - For about 10 min, we'll be vague about how these interactions are specified



# Example Bayes Net: Insurance

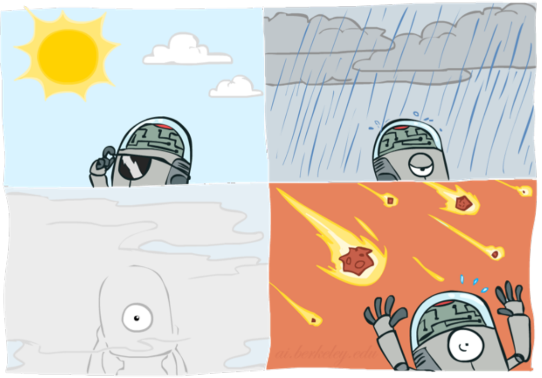


# Example Bayes Net: Diagnosis

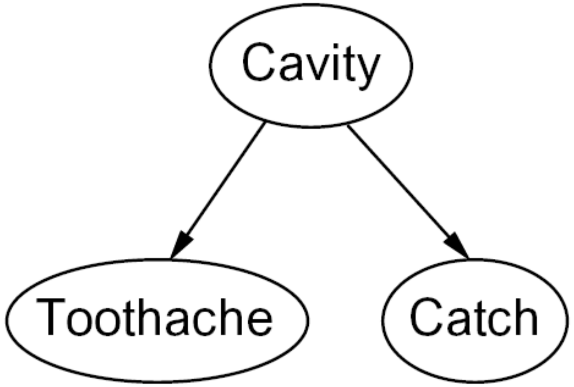


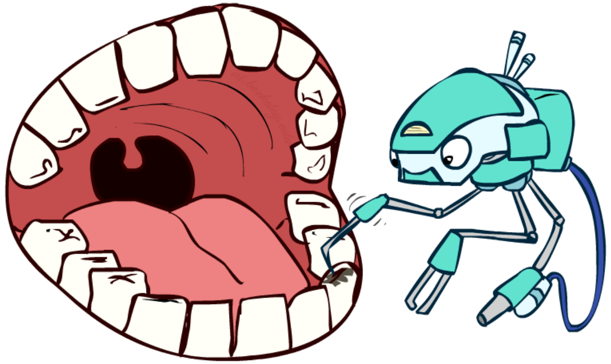
# Graphical Model Notation

- 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)

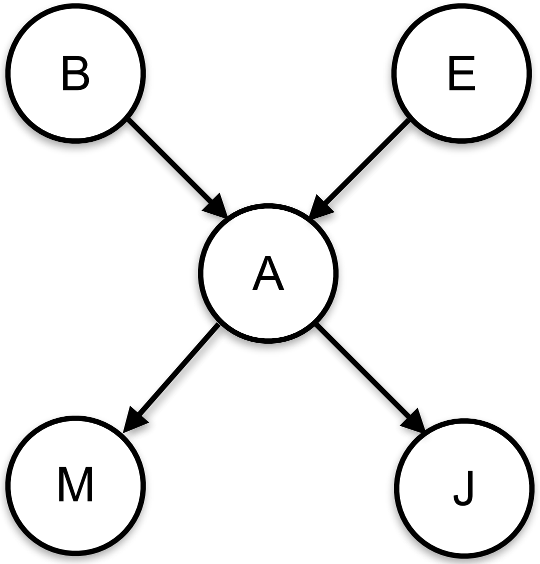


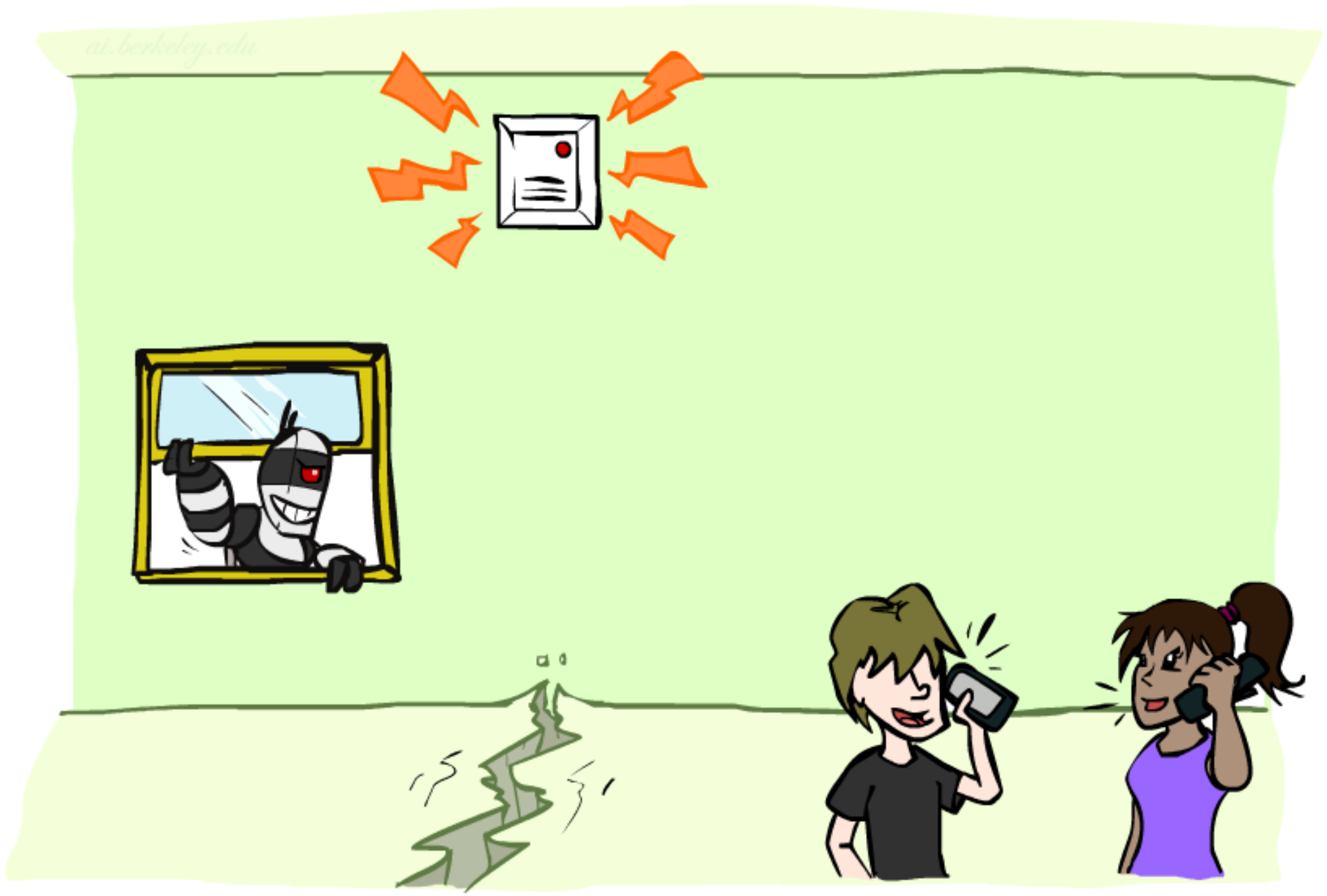



- For now: imagine that arrows mean direct causation (in general, they don't!)

# Example: Alarm Network

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



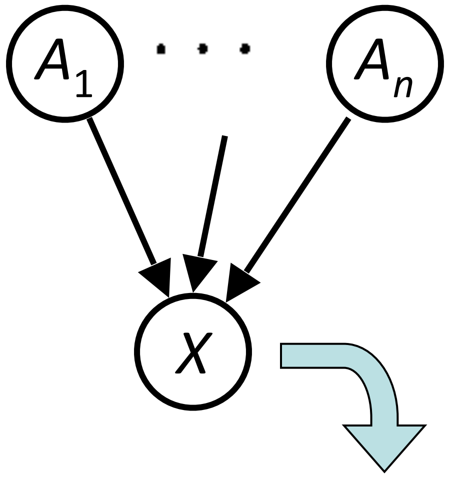
# Bayes Net Semantics

# 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




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

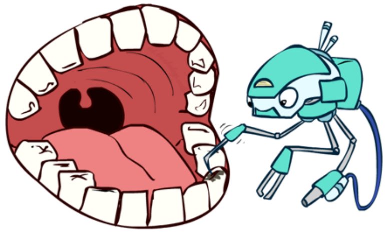
*A Bayes net = Topology (graph) + Local Conditional Probabilities*

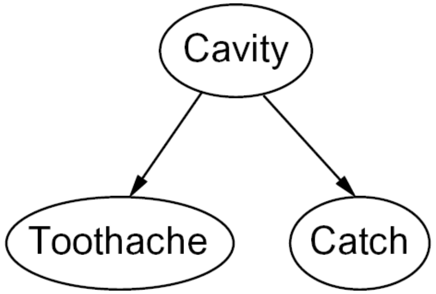
# Probabilities in BNs

- 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:

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

  - Example:






$$P(+\textit{cavity}, +\textit{catch}, \textit{-toothache})$$

# 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): $P(x_1, x_2, \ldots x_n) = \prod_{i=1}^{n} P(x_i | x_1 \ldots x_{i-1})$

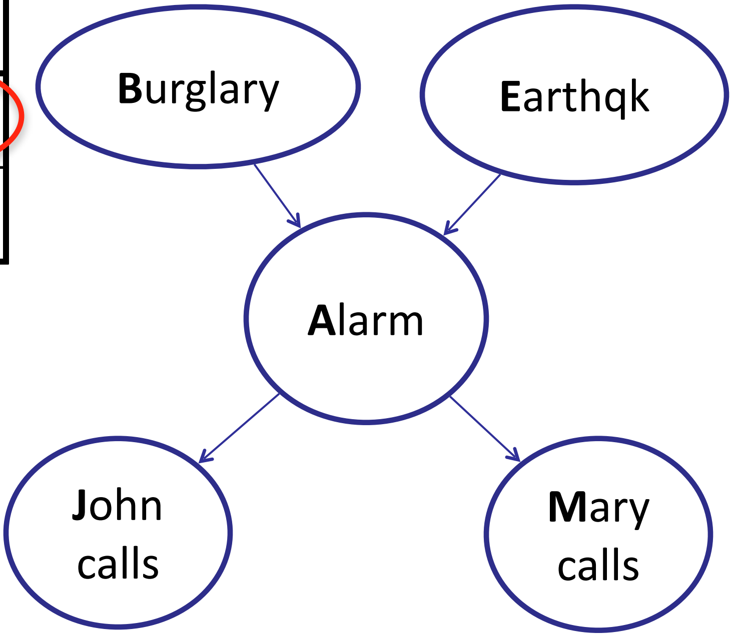
- <u>Assume</u> conditional independences: $P(x_i | x_1, \ldots x_{i-1}) = P(x_i | \mathit{parents}(X_i))$

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

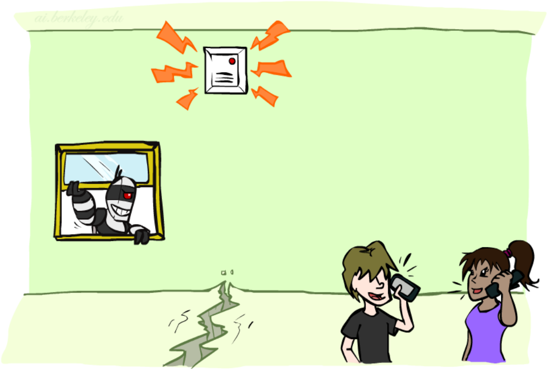
- Not every BN can represent every joint distribution
  - The topology enforces certain conditional independencies

# Example: Alarm Network

| B | P(B) |
|---|---|
| +b | 0.001 |
| -b | 0.999 |




| E | P(E) |
|---|---|
| +e | 0.002 |
| -e | 0.998 |



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

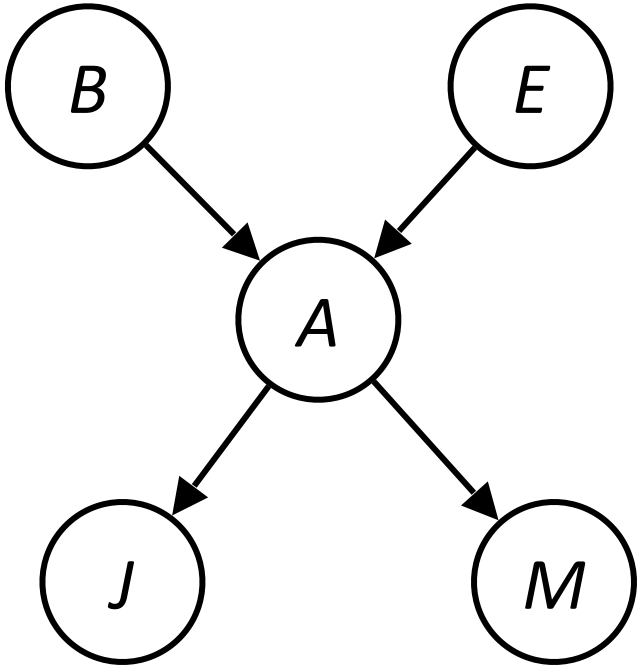
| A | J | P(J\|A) |
|---|---|---|
| +a | +j | 0.9 |
| +a | -j | 0.1 |
| -a | +j | 0.05 |
| -a | -j | 0.95 |

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

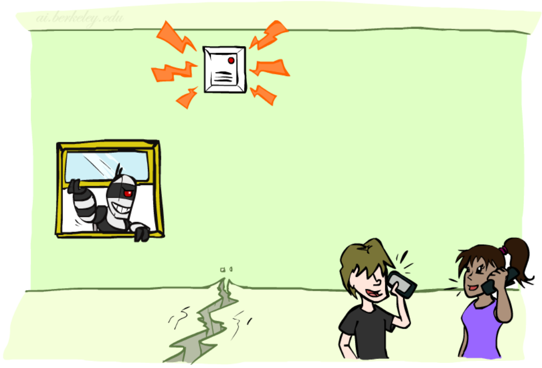
| B | E | A | P(A\|B,E) |
|---|---|---|---|
| +b | +e | +a | 0.95 |
| +b | +e | -a | 0.05 |
| +b | -e | +a | 0.94 |
| +b | -e | -a | 0.06 |
| -b | +e | +a | 0.29 |
| -b | +e | -a | 0.71 |
| -b | -e | +a | 0.001 |
| -b | -e | -a | 0.999 |

# Example: 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 |

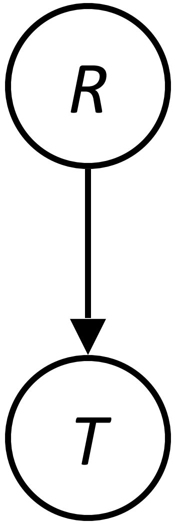
$$P(+b, -e, +a, -j, +m) =$$

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

$$0.001 \times 0.998 \times 0.94 \times 0.1 \times 0.7$$

# Example: Traffic

- Causal direction




$P(R)$

| | |
|---|---|
| +r | 1/4 |
| -r | 3/4 |

$P(T|R)$

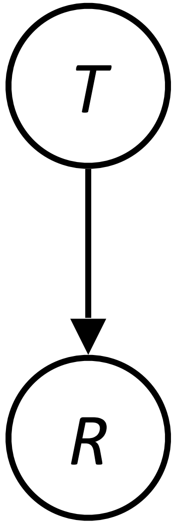
| | | |
|---|---|---|
| +r | +t | 3/4 |
| | -t | 1/4 |
| -r | +t | 1/2 |
| | -t | 1/2 |

$P(T, R)$

| | | |
|---|---|---|
| +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 |

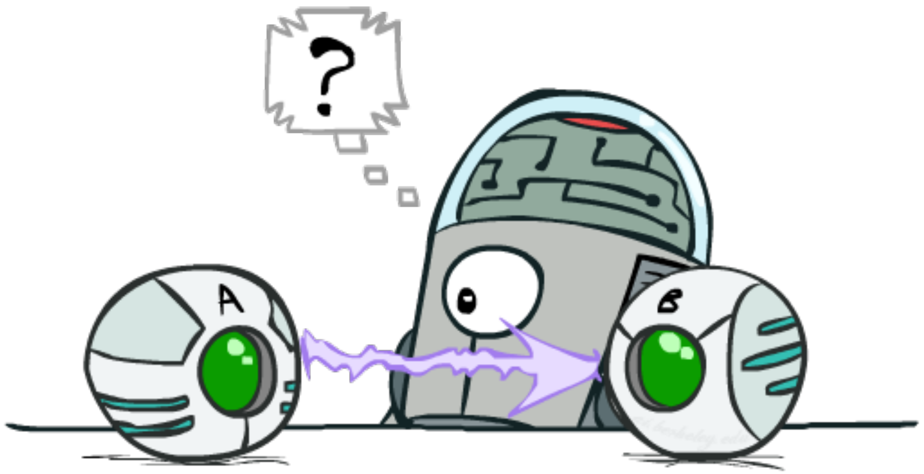
$P(T,R)$

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

# Causality?

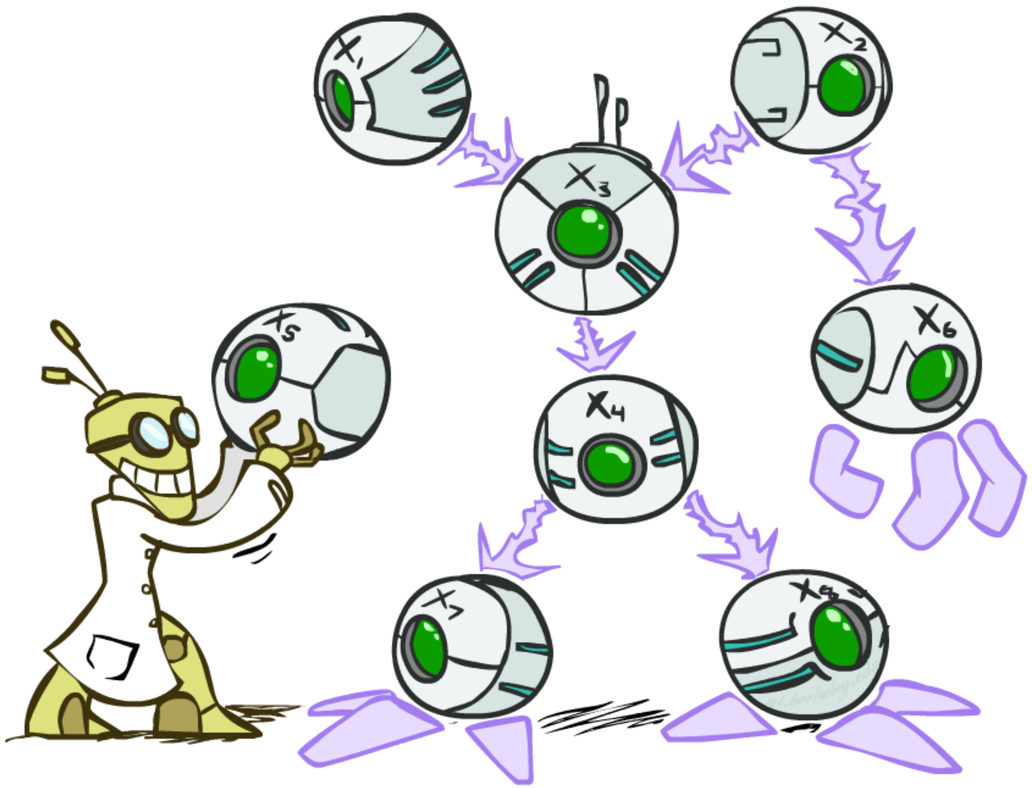
- When Bayes nets reflect the true causal patterns:
  - Often simpler (nodes have fewer parents)
  - Often easier to think about
  - Often easier to elicit from experts
- BNs need not actually be causal
  - Sometimes no causal net exists over the domain (especially if variables are missing)
  - E.g. consider the variables *Traffic* and *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

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



# 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
- Thursday: how to answer numerical queries (inference)



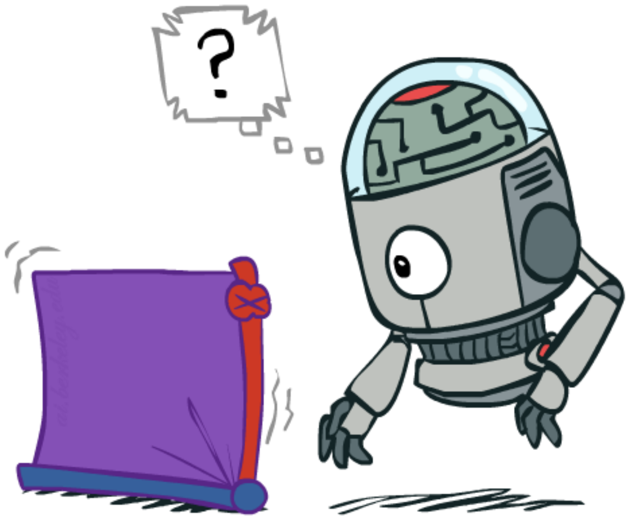
# Size of a Bayes Net

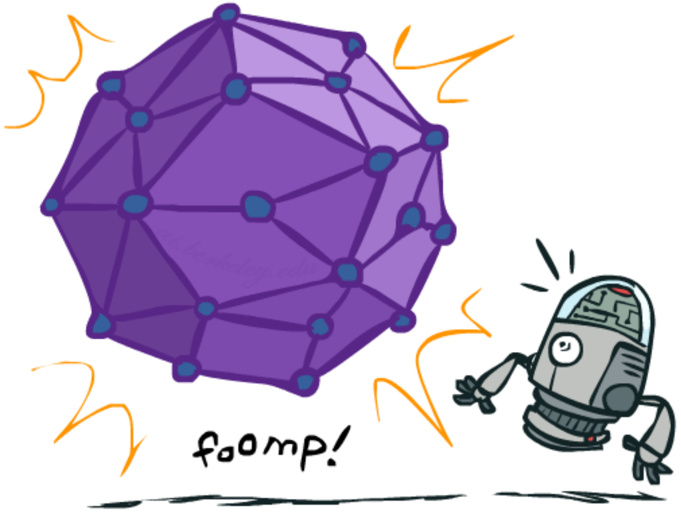
- How big is a joint distribution over N Boolean variables?
  - $2^N$
- How big is an N-node net if nodes have up to k parents?
  - $O(N * 2^{k+1})$

- Both give you the power to calculate

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

- BNs: Huge space savings!
- Also easier to elicit local CPTs
- Also faster to answer queries (coming)



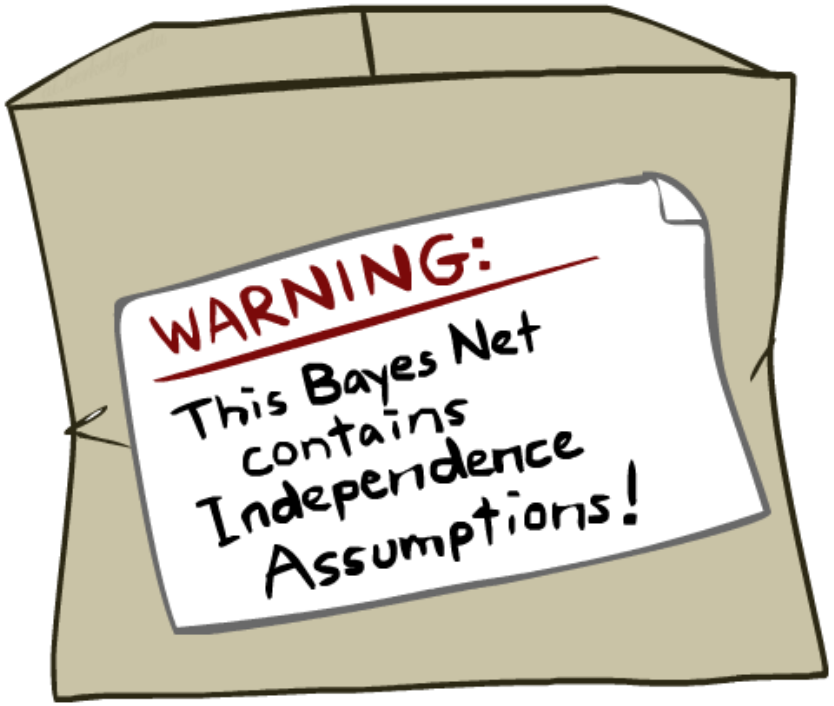



# Bayes Nets: Assumptions

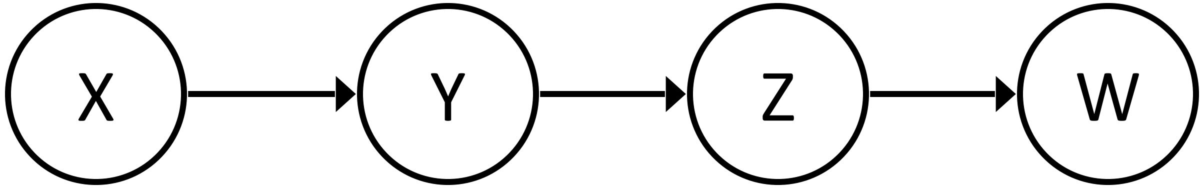
- Assumptions we are required to make to define the Bayes net when given the graph:

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

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



# Example




- Conditional independence assumptions directly from simplifications in chain rule:

Standard chain rule: $p(x, y, z, w) = p(x)p(y|x)p(z|x, y)p(w|x, y, z)$

Bayes net: $p(x, y, z, w) = p(x)p(y|x)p(z|y)p(w|z)$

Since: $z \perp\!\!\!\perp x \mid y$ and $w \perp\!\!\!\perp x, y \mid z$ (cond. indep. given parents)

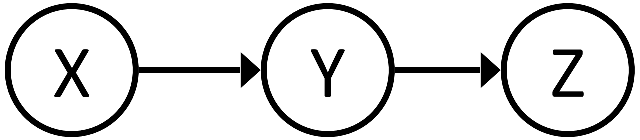
- Additional implied conditional independence assumptions? $w \perp\!\!\!\perp x \mid y$

$$p(w|x, y) = \frac{p(w, x, y)}{p(x, y)} = \frac{\sum_z p(x)p(y|x)p(z|y)p(w|z)}{p(x)p(y|x)} = \sum_z p(z|y)p(w|z) = \sum_z p(z|y)p(w|z, y)$$

$$= \sum_z p(z, w|y) = p(w|y)$$

# 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)
    - Addendum: they *could* be independent: how?

# Bayes Nets

- ✓ Representation
- ✓ Conditional Independences
- Probabilistic Inference
  - Enumeration (exact, exponential complexity)
  - Variable elimination (exact, worst-case exponential complexity, often better)
  - Probabilistic inference is NP-complete
  - Sampling (approximate)
- Learning Bayes' Nets from Data