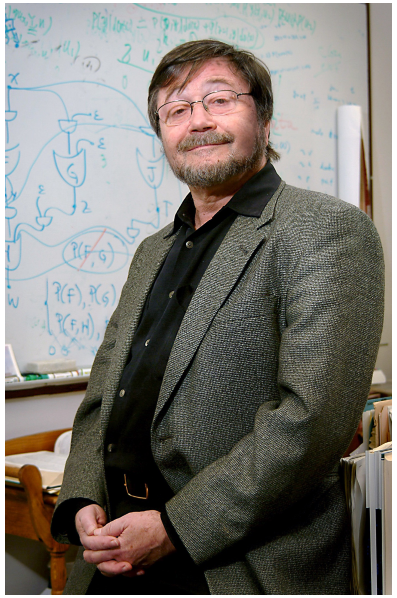
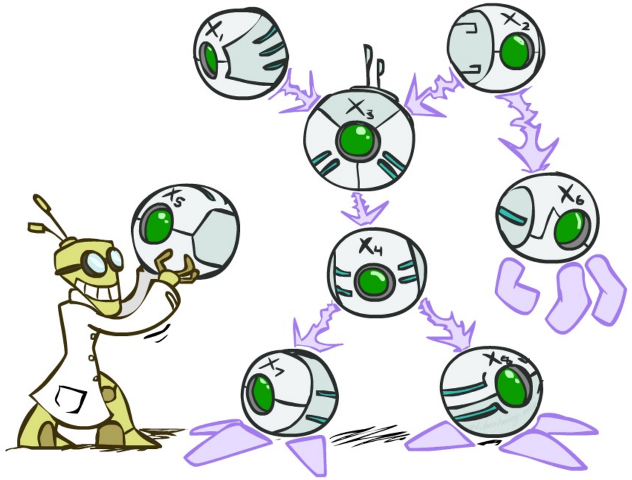


Artificial Intelligence



Probabilistic Reasoning with Bayes Nets

CS 444 – Spring 2021

Dr. Kevin Molloy

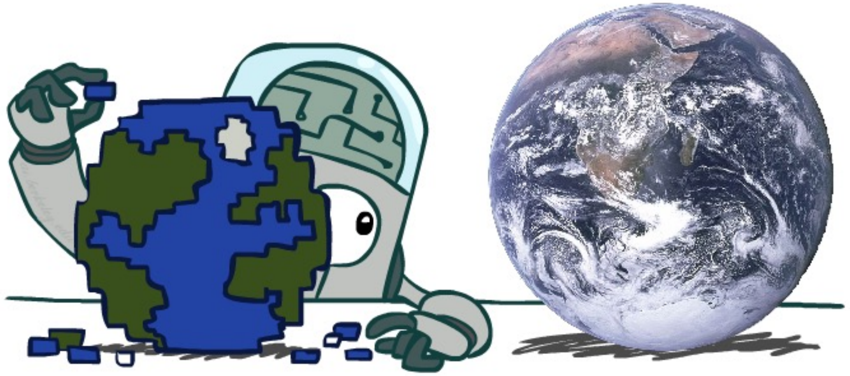
Department of Computer Science

James Madison University

Much of this lecture is taken from AIMA resources

JMU JAMES MADISON UNIVERSITY.

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

Bayesian Networks

A simple, graphical notation for conditional independence assertions and hence for compact specification of full joint distribution

Syntax:

- A set of nodes, one per variable
- A directed, acyclic graph (link is approximately "directly influences")
- a conditional distribution for each node given its parents $P(X_i \mid Parents(X_i))$.

In the simplest case, conditional distribution represented as a conditional probability table (CPT) giving the distribution over X_i for each combination of parents values.

Example of a Bayesian Network

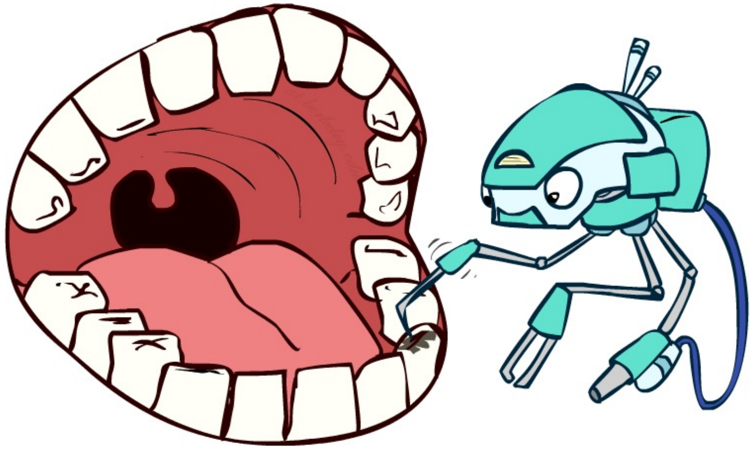
Topology of network encodes conditional independence assertions:

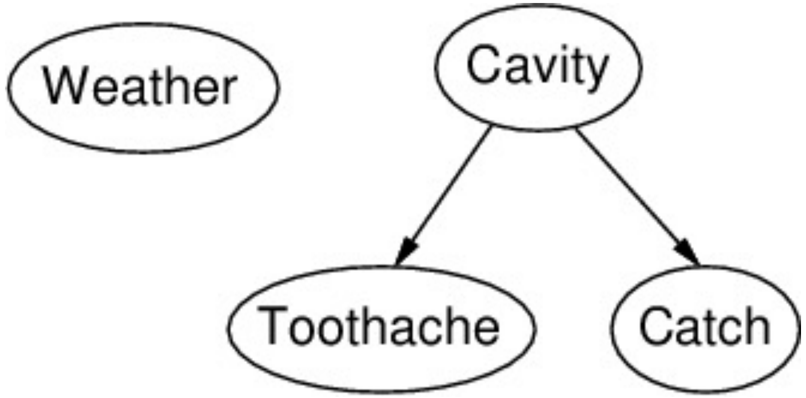
- P(Toothache, Cavity, Catch)

- 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)





Conditional Independence

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

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

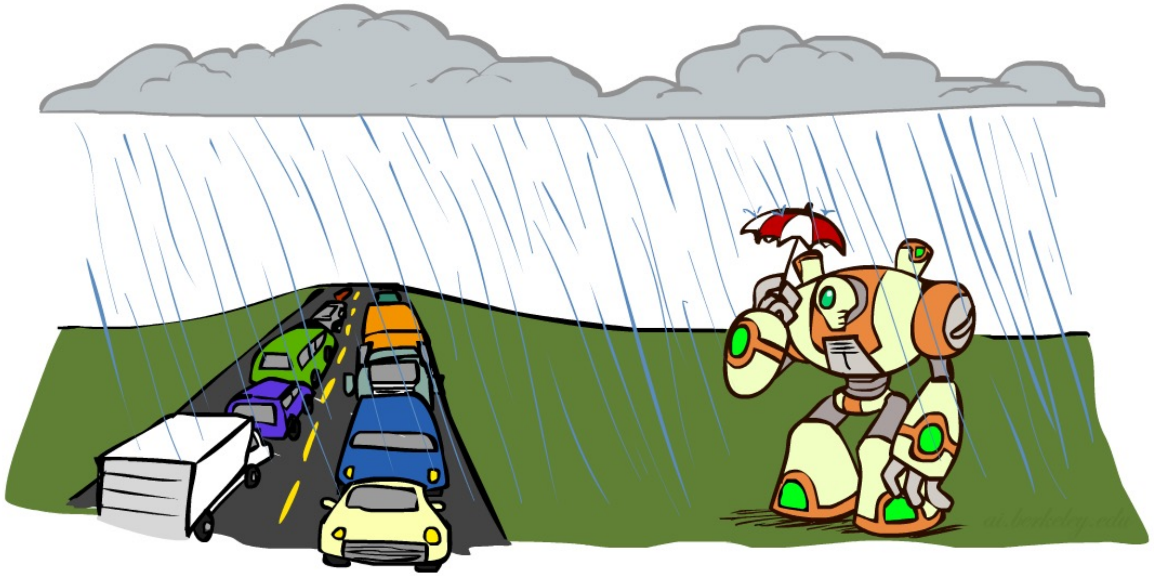
- X is conditionally independent of Y given Z $X \perp\!\!\!\perp Y | Z$

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

  or, equivalently, if and only if $\forall x, y, z : P(x|z, y) = P(x|z)$

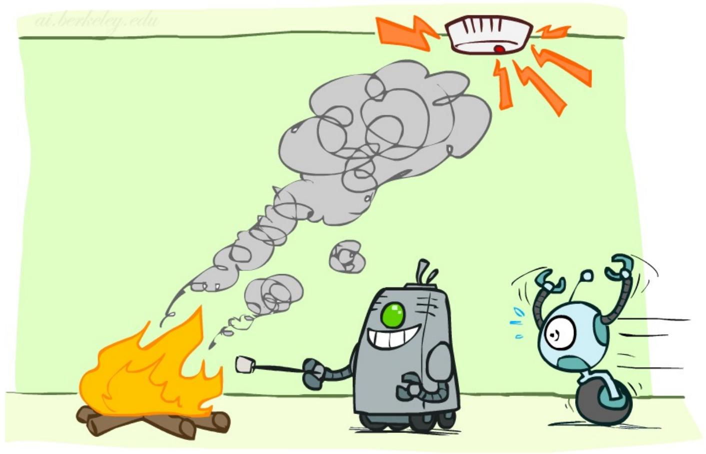
Short Quiz: Conditional Independence

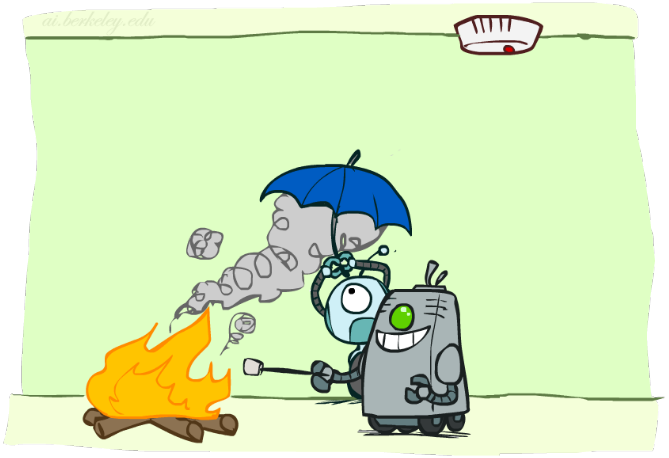
- What about this domain:
  - Traffic
  - Umbrella
  - Raining



Short Quiz: Conditional Independence

- What about this domain:
  - Fire
  - Smoke
  - Alarm



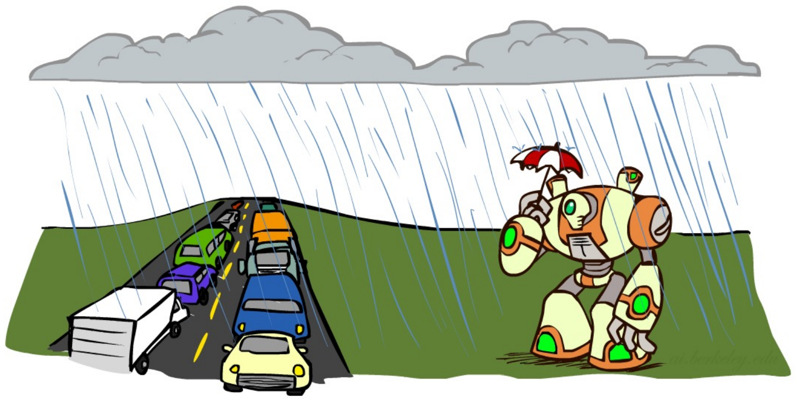


Conditional Ind. And the Chain Rule

- 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

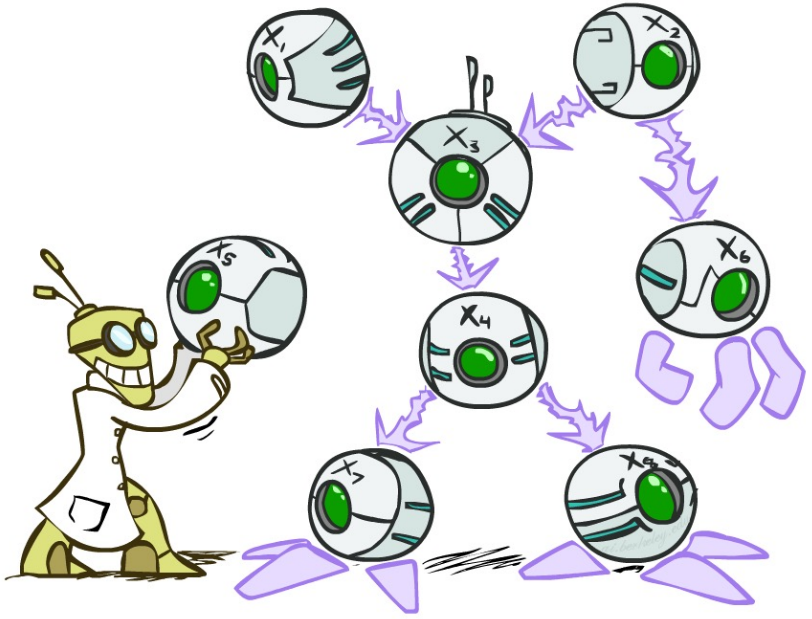
- Each sensor depends only on where the ghost is
- That means, the two sensors are conditionally independent, given the ghost position
- T: Top square is red
  B: Bottom square is red
  G: Ghost is in the top
- Givens:
  $P(+g) = 0.5$
  $P(-g) = 0.5$
  $P(+t \mid +g) = 0.8$
  $P(+t \mid -g) = 0.4$
 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

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

Big Picture of Bayes' Nets

- 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

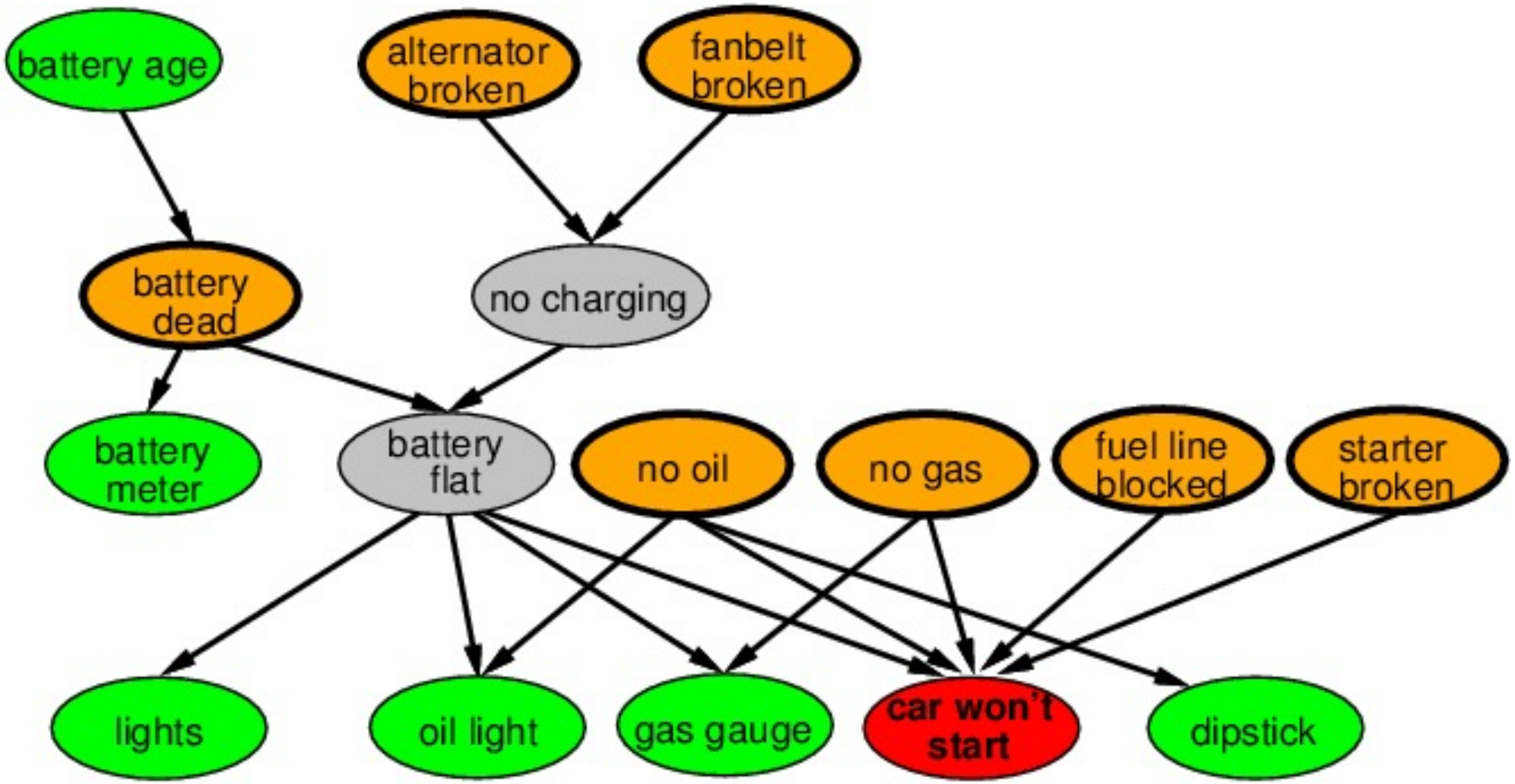


Example Car Diagnosis

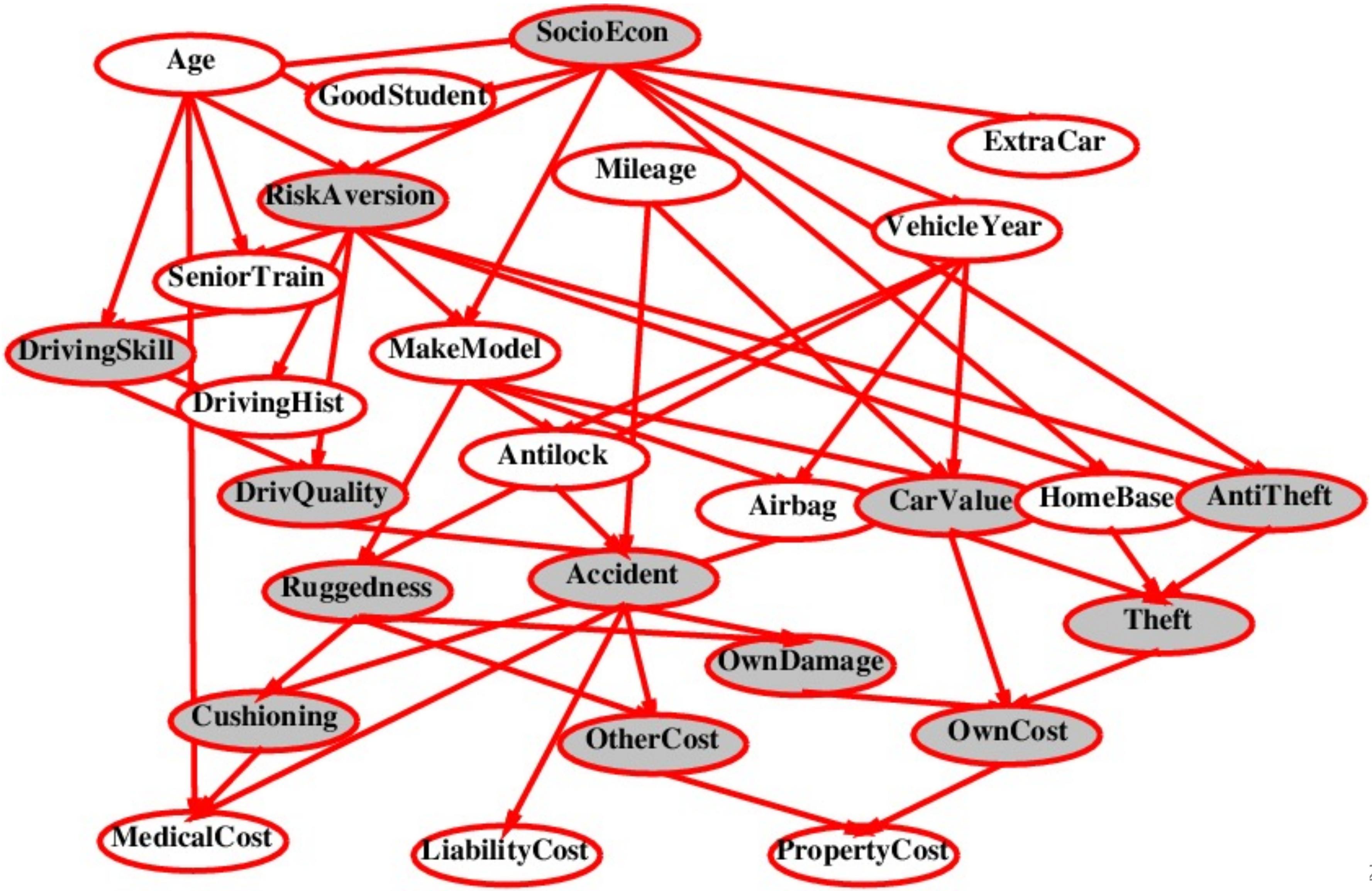
Initial evidence: car won't start

Testable variables (Green), "Broken, so fix it" variables are orange

Hidden variables (gray) ensure sparse structure, reduce parameters



Example: Car Insurance



Global Semantics

Global semantics defines the full joint distribution as the product of the local conditional distributions.

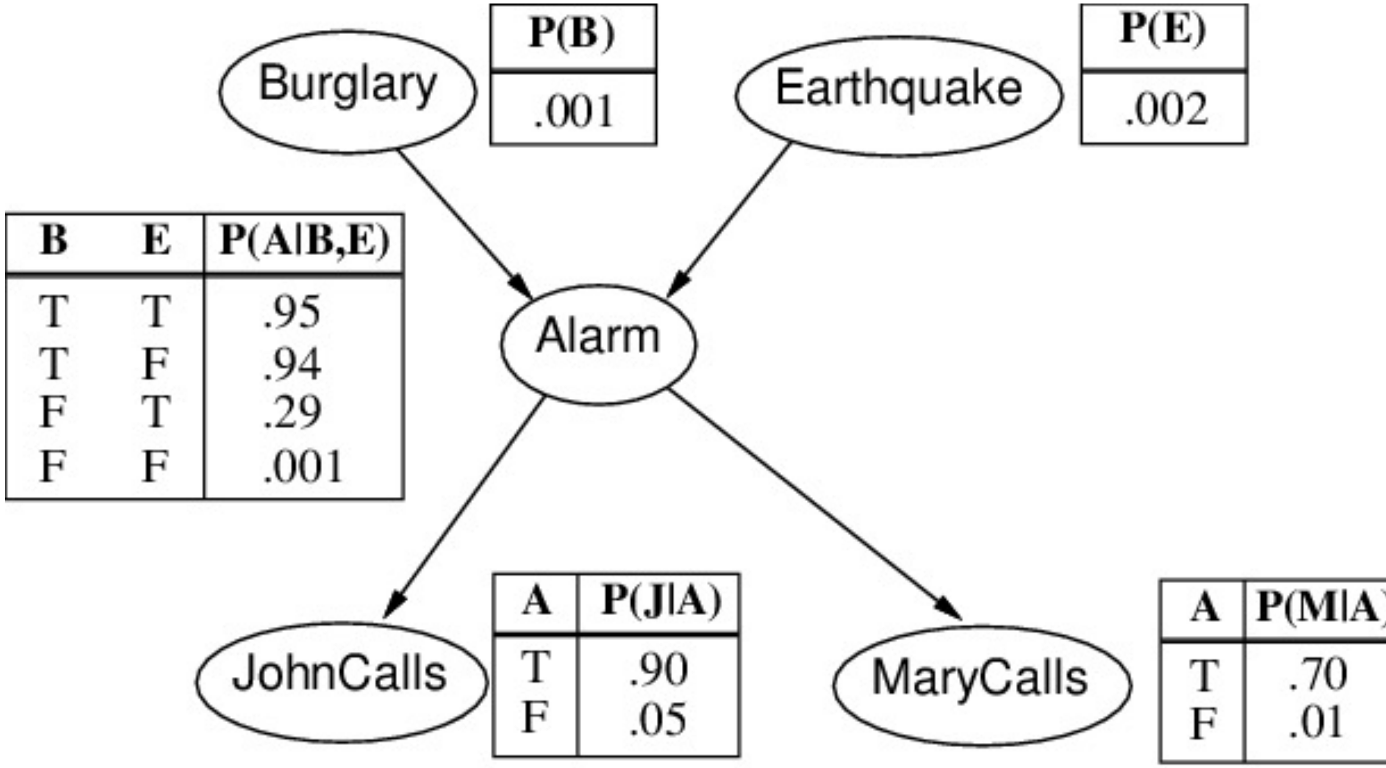
$$P(x_1, \ldots, x_n) = \prod_{i=1}^{n} P(x_i | parents\ (X_i))$$

Quiz: Compute $P(j \land m \land a \land \neg b \land \neg e)$

$= P(j\ |a\ )P(M\ |\ a)\ P\ (a\ |\ \neg b, \neg e)P(\neg b)\ P(\neg e)$

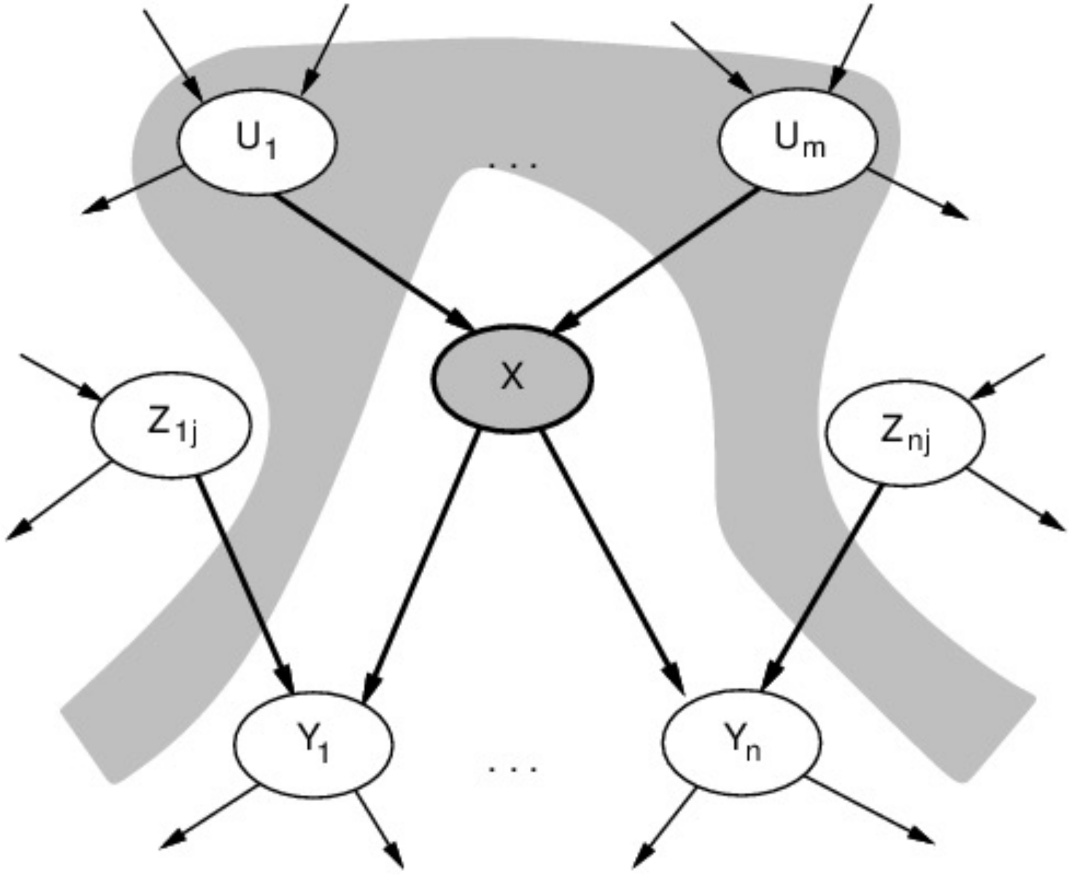
$= 0.9 \times 0.7 \times 0.001 \times 0.999 \times .998$

≈ 0.00063



Local Semantics

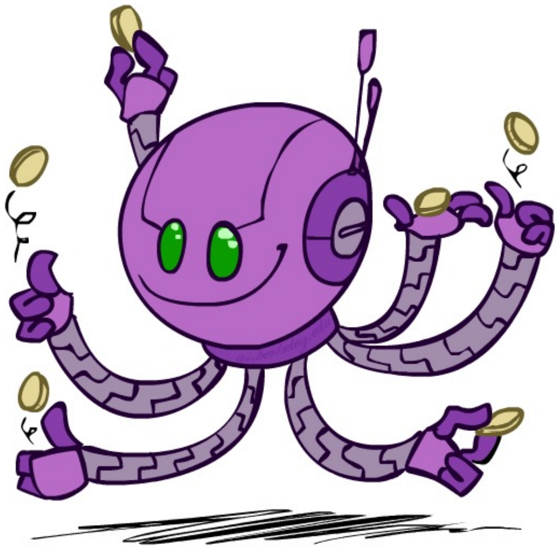
Local semantics: each node is conditionally independent of its nondescenants given its parents



Theorem: Local semantics ⇔ global semantics

Example: Coin Flips

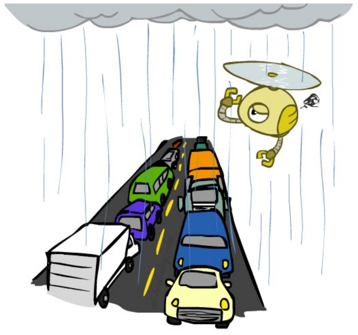
- N independent coin flips



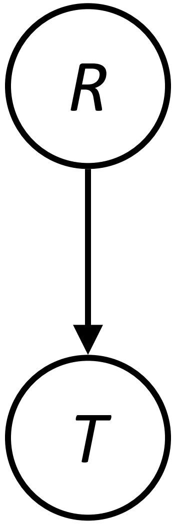
- No interactions between variables: absolute independence

Example: Traffic

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



- Model 1: independence



- Why is an agent using model 2 better?

Example of a Bayesian Network

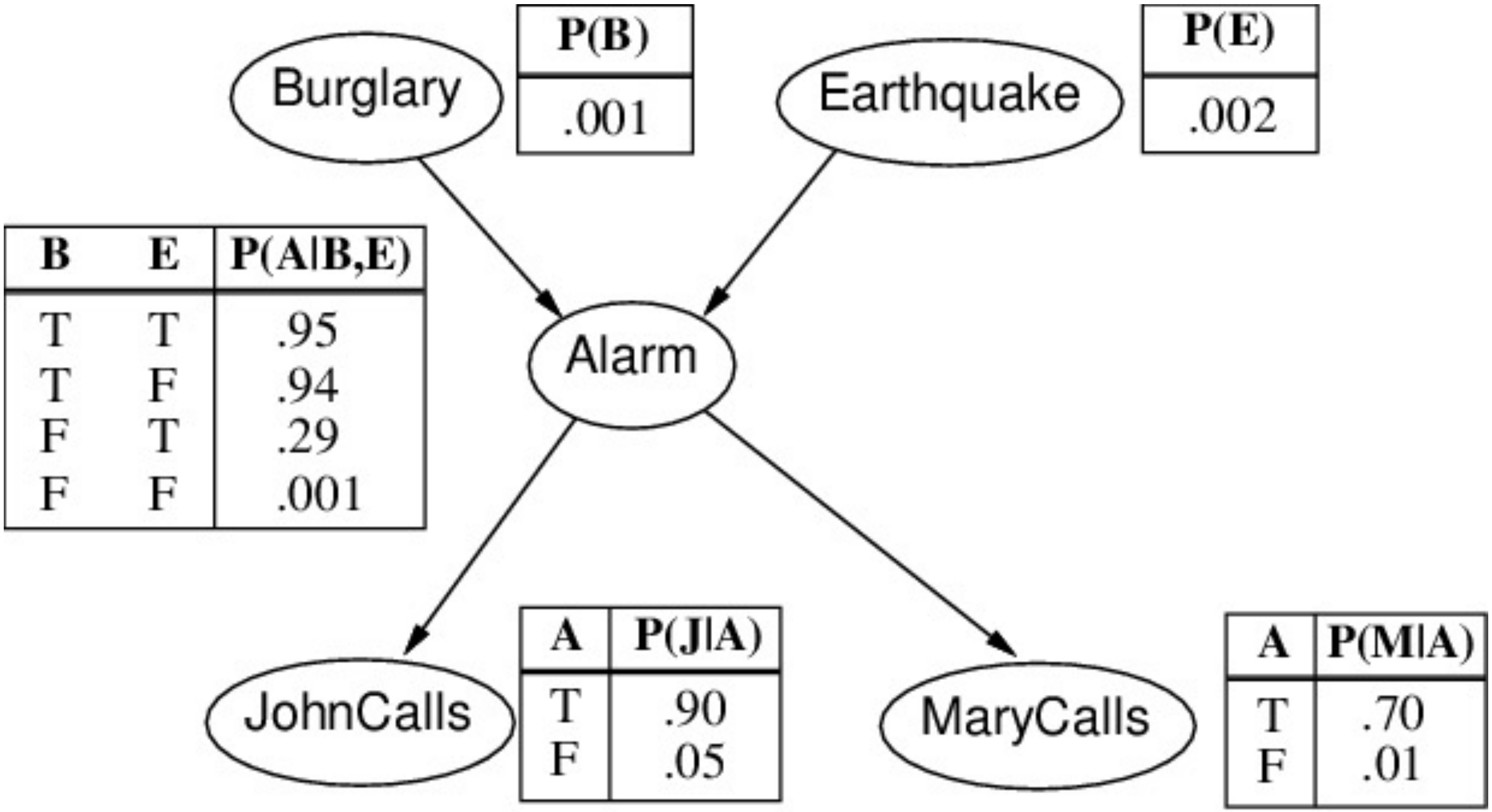
I'm at work, neighbor John calls to say my burglar alarm is ringing, but my neighbor Mary doesn't call. Sometimes the alarm is set off by earthquakes. Is there a burglar?

Variables: Burglar, Earthquake, Alarm, JohnCalls, MaryCalls

Network topology reflects "causal" knowledge:

- *A burglar can set off the alarm*
- *An earthquake can set the alarm off*
- *The alarm can cause Mary to call*
- *The alarm can cause John to Call.*

Example Bayesian Network



P(B)
.001

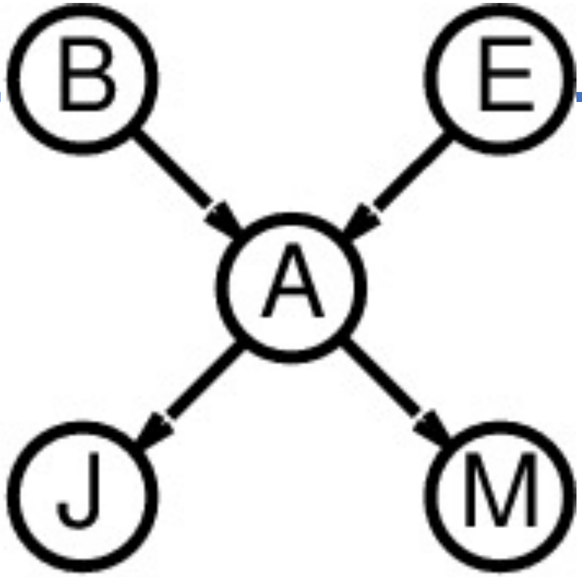
P(E)
.002

B	E	P(A|B,E)
T	T	.95
T	F	.94
F	T	.29
F	F	.001

A	P(J|A)
T	.90
F	.05

A	P(M|A)
T	.70
F	.01

Compactness



A CPT for Boolean X_i with k Boolean parents.

Has:

2^k rows for the combinations of parent values

Each row requires on number *p for* $X_i = true$

(the number for X_i = false is simply 1 – p)

If each variable has no more than k parents, the complete network requires $O(n \cdot 2^k)$ numbers

i.e. **grows linearly with n**, vs $O(2^n)$ for the full joint distribution.

For the burglary net, 1 + 1 + 4 + 2 + 2 = 10 numbers (vs. $2^5 - 1 = 31$).

In-Class Problem

We have a bag of 3 biased coins: a, b, c with probabilities of coming up heads of 20%, 60%, and 80% respectively. One coin is drawn randomly from the bag (with equal likelihood od drawing each of the 3 coins), and then the coin is flipped 3 times to generate the outcomes X_1, X_2, and X_3.

1. Draw the Bayesian network corresponding to this setup and define the necessary CPTs.
2. Calculate which coin was most likely to have been drawn from the bag if the observed flips come out heads twice and tails once.