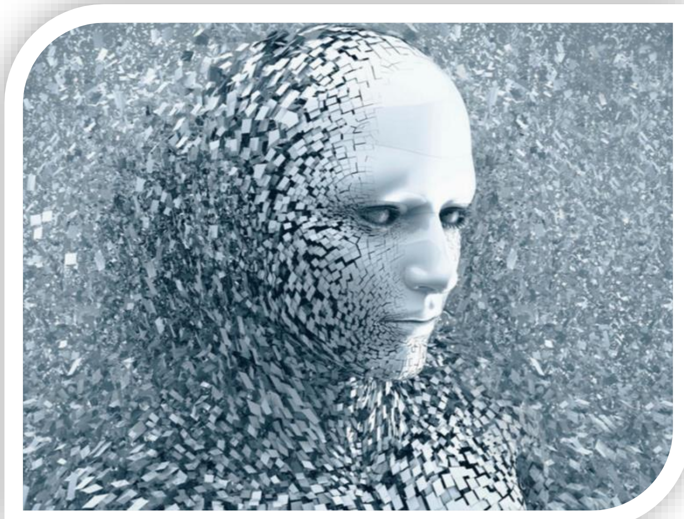


Artificial Intelligence

Session 5:
Probability & Bayes' Nets

School of Computing and Engineering
University of West London, UK

Dr Massoud Zolgharni

Probability

Uncertainty

Assume the following action:
Leave for the airport 90 minutes before my flight departures using the **A25**.

Will that action get me there on time?

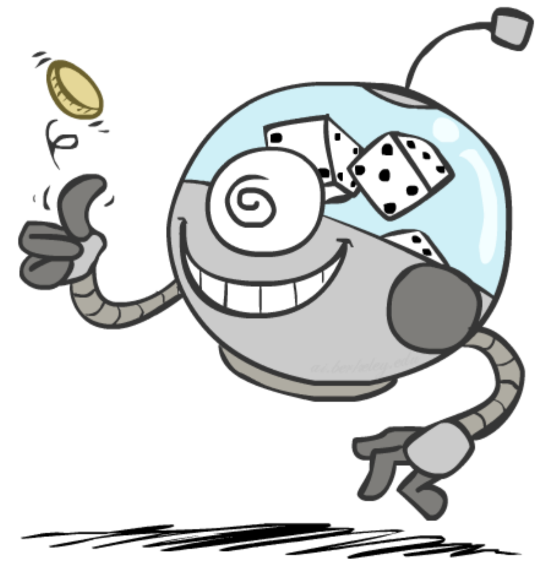
Problems while deciding to take the action or not:
- partial observability (road state, other drivers' plans, etc.)
- noisy sensors (traffic reports)
- uncertainty in action outcomes (flat tire, etc.)
- immense complexity of modelling and predicting traffic

Hence a purely logical approach (without assumption of probabilities) either risks **falsehood**: *"**A25** will get me there on time"*

Or leads to conclusions that are **too weak for decision** making:
*"**A25** will get me there on time if there's no accident on the bridge and it doesn't rain and my tires remain intact etc etc."*

Random Variables

- A random variable is some aspect of the world about which we (may) have uncertainty
  - R = Is it raining?
  - T = Is it hot or cold?
  - D = How long will it take to drive to work?
  - L = Where is the ghost?



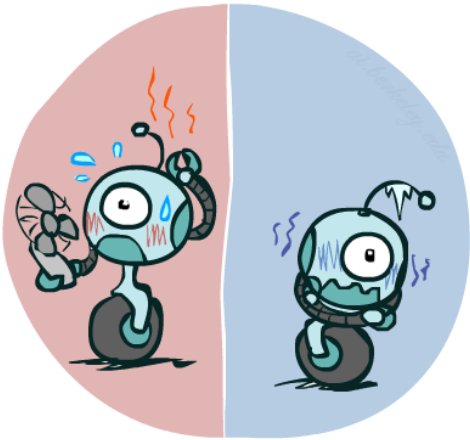
- We denote random variables with capital letters

- Like variables in a CSP, random variables have domains
  - R in {true, false}   (often write as {+r, -r})
  - T in {hot, cold}
  - D in $[0, \infty)$
  - L in possible locations, maybe {(0,0), (0,1), …}

Probability Distributions

- Associate a probability with each value

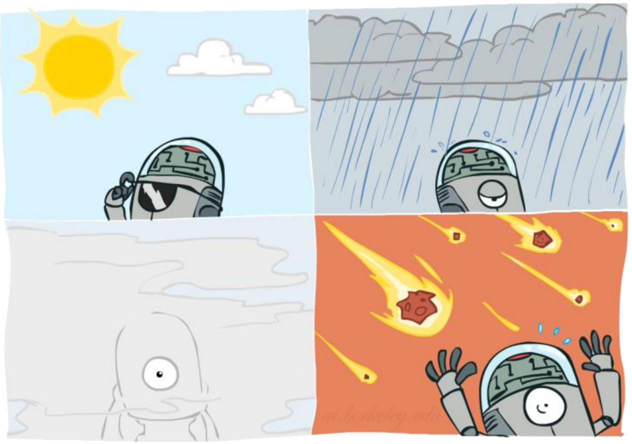
  - Temperature:



$P(T)$

T	P
hot	0.5
cold	0.5

  - Weather:



$P(W)$

W	P
sun	0.6
rain	0.1
fog	0.3
meteor	0.0

Probability Distributions

- Unobserved random variables have distributions

$P(T)$

T	P
hot	0.5
cold	0.5

$P(W)$

W	P
sun	0.6
rain	0.1
fog	0.3
meteor	0.0

Shorthand notation:

$$P(hot) = P(T = hot),$$
$$P(cold) = P(T = cold),$$
$$P(rain) = P(W = rain),$$

- A distribution is a TABLE of probabilities of values

- A probability (lower case value) is a single number

$$P(W = rain) = 0.1$$

- Must have: $\forall x \;\; P(X = x) \geq 0$ and $\sum_x P(X = x) = 1$

Joint Distributions

- A *joint distribution* over a set of random variables: $X_1, X_2, \ldots X_n$ specifies a real number for each assignment (or *outcome*):

$$P(X_1 = x_1, X_2 = x_2, \ldots X_n = x_n)$$

$$P(x_1, x_2, \ldots x_n)$$

  - Must obey:

$$P(x_1, x_2, \ldots x_n) \geq 0$$

$$\sum_{(x_1, x_2, \ldots x_n)} P(x_1, x_2, \ldots x_n) = 1$$

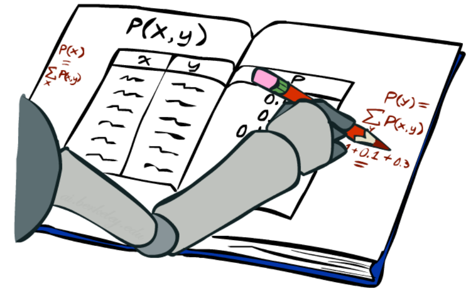
$P(T, W)$

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

- Size of distribution if *n* variables with domain sizes *d*?
  - For all but the smallest distributions, impractical to write out!

Marginal Distributions

- Marginal distributions are sub-tables which eliminate variables
- Marginalization (summing out): Combine collapsed rows by adding



$$P(T, W)$$

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

$$P(W)$$

W	P
sun	0.6
rain	0.4

Events

- An *event* is a set E of outcomes

$$P(E) = \sum_{(x_1 \ldots x_n) \in E} P(x_1 \ldots x_n)$$

- From a joint distribution, we can calculate the probability of any event
  - Probability that it's hot AND sunny? 0.4
  - Probability that it's hot? 0.5
  - Probability that it's hot OR sunny? 0.7

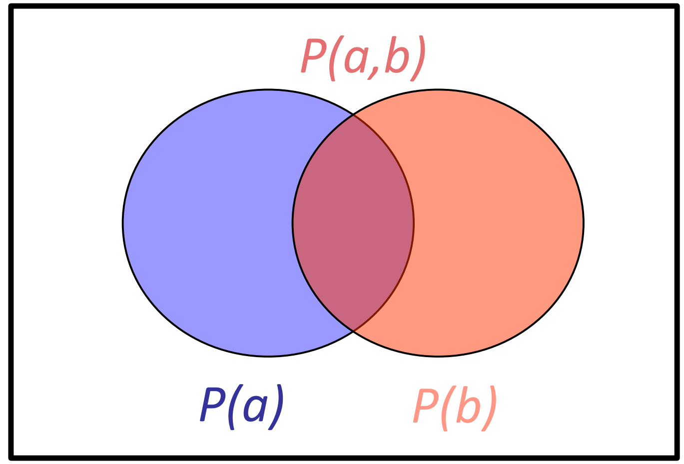
$$P(T, W)$$

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

Conditional Probabilities

P(*a* | *b*) means: Probability for *a* "given" that *b* already happened

$$P(a|b) = \frac{P(a,b)}{P(b)}$$



$$P(T,W)$$

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

$$P(W=s|T=c) = \frac{P(W=s,T=c)}{P(T=c)} = \frac{0.2}{0.5} = 0.4$$

$$= P(W=s,T=c) + P(W=r,T=c)$$

$$= 0.2 + 0.3 \quad = 0.5$$

Probabilistic Inference

- Probabilistic inference: compute a desired probability from a set of other known probabilities

- We generally compute conditional probabilities
  - P(on time | no reported accidents) = 0.90
  - These represent the agent's *beliefs* given the evidence

- Probabilities change with new evidence:
  - P(on time | no accidents, 5 a.m.) = 0.95
  - P(on time | no accidents, 5 a.m., raining) = 0.80
  - Observing new evidence causes *beliefs to be updated*

Product Rule

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

$P(W)$

W	P
sun	0.8
rain	0.2

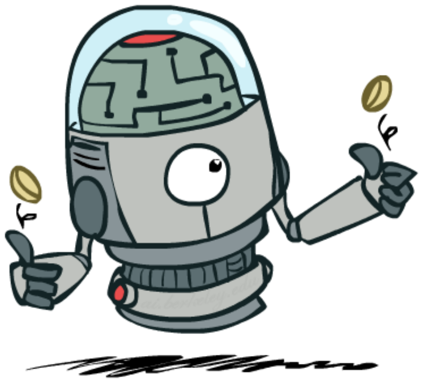
$P(T|W)$

T	W	P
wet	sun	0.1
dry	sun	0.9
wet	rain	0.7
dry	rain	0.3

$P(T,W)$

T	W	P
wet	sun	
dry	sun	
wet	rain	
dry	rain	

Independence



- Two variables are independent if:

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

  - their joint probability distribution *factors* into a product of two simpler distributions (there is no interaction between variables)

- Independence is a simplifying modeling assumption
  - What could we assume for {Coin Toss} or {Weather, Traffic, Cavity, Toothache}?

Factoring large joint distribution into smaller ones

$P(X_1)$

H	0.5
T	0.5

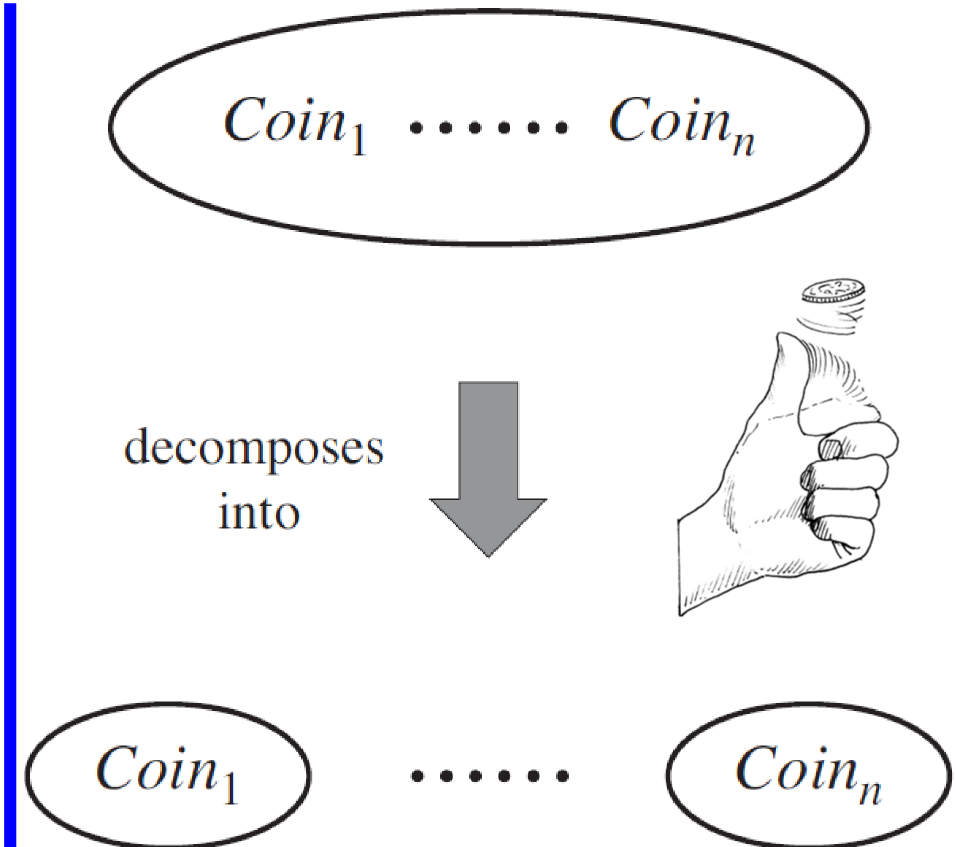
$P(X_2)$

H	0.5
T	0.5

$\cdots$

$P(X_n)$

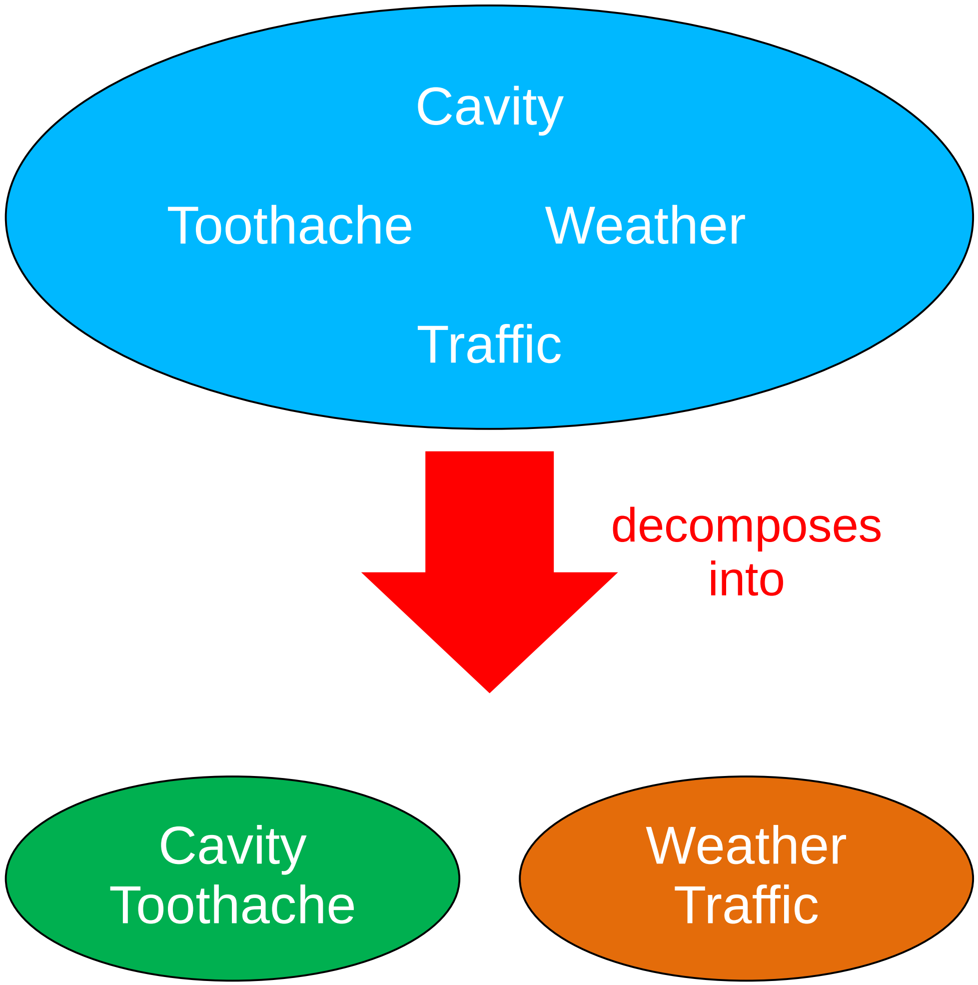
H	0.5
T	0.5

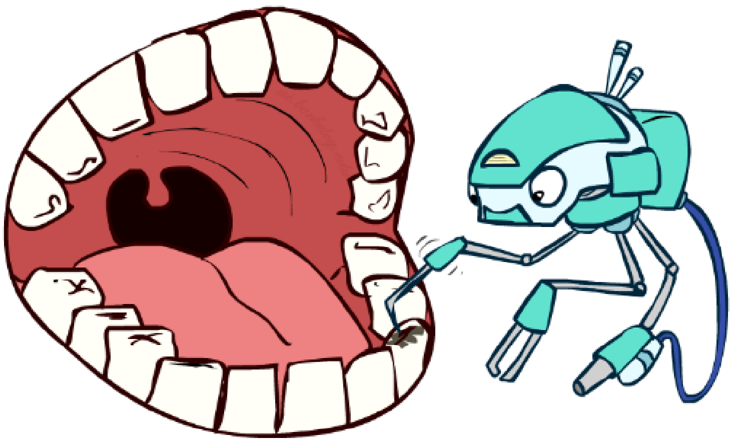


Coin flips are independent

H	0.5
T	0.5

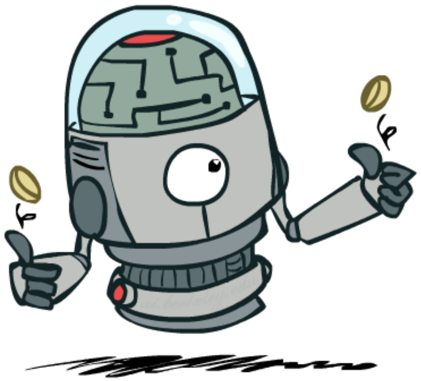
Factoring large joint distribution into smaller ones





$$P(\text{Toothache, Cavity, Weather, Traffic}) = P(\text{Toothache, Cavity}) \times P(\text{Weather, Traffic})$$

Independence



- Two variables are independent if:

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

  - their joint distribution *factors* into a product of two simpler distributions (there is no interaction between variables)

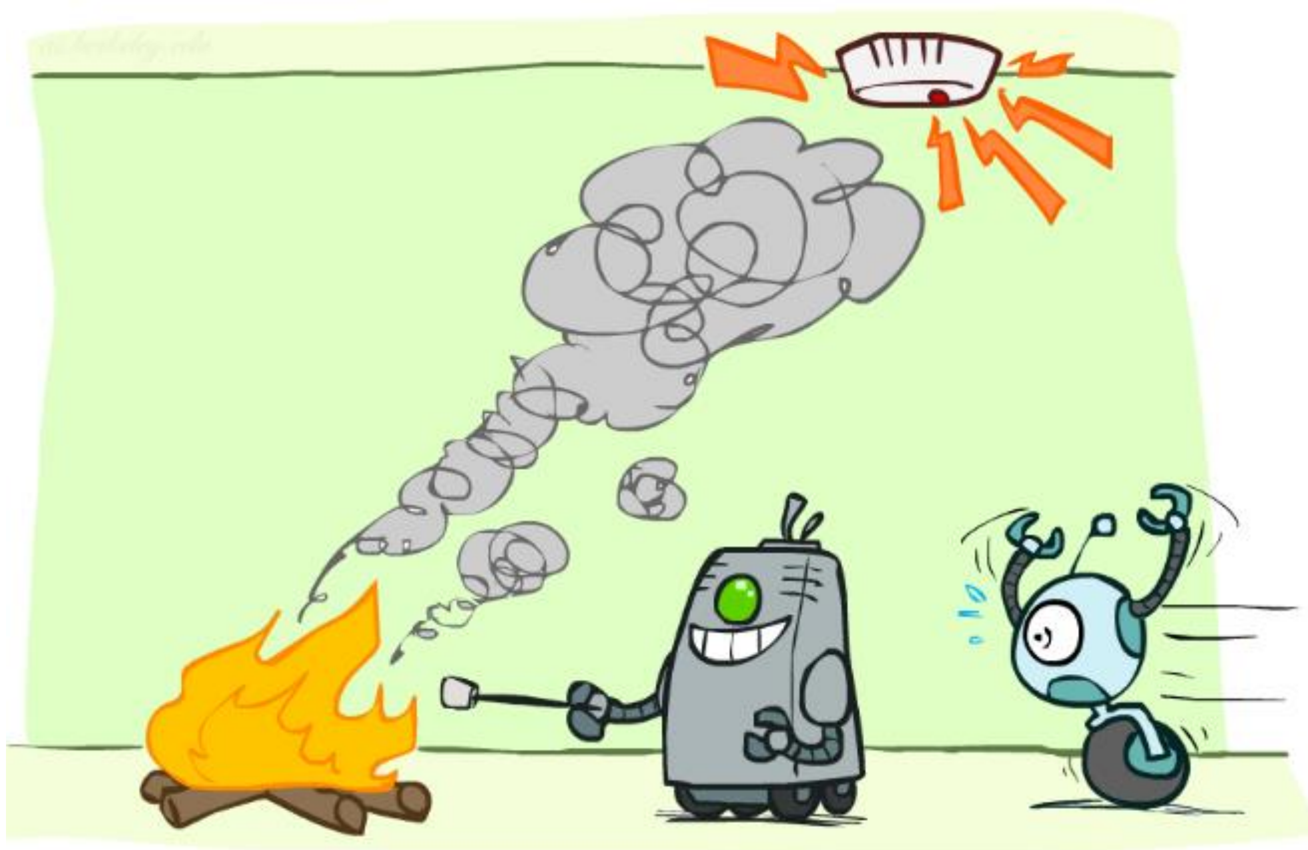
- Independence is a simplifying modeling assumption
  - What could we assume for {Coin Toss} or {Weather, Traffic, Cavity, Toothache}?

  - Another form (learning something about y, does not change my belief about x):

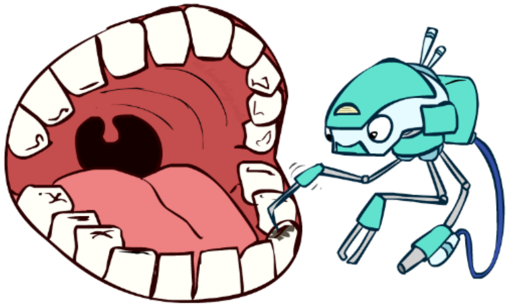
$$P(x|y) = P(x)$$

  - Or: $$P(y|x) = P(y)$$

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 pain:
  - *P*(catch | toothache, cavity) = *P*(catch | cavity)

- Catch is conditionally independent of Toothache given Cavity:
  - Once I know about cavity, knowing about pain does not change (add) anything

- *X* is conditionally independent of *Y* given *Z*

  if and only if:

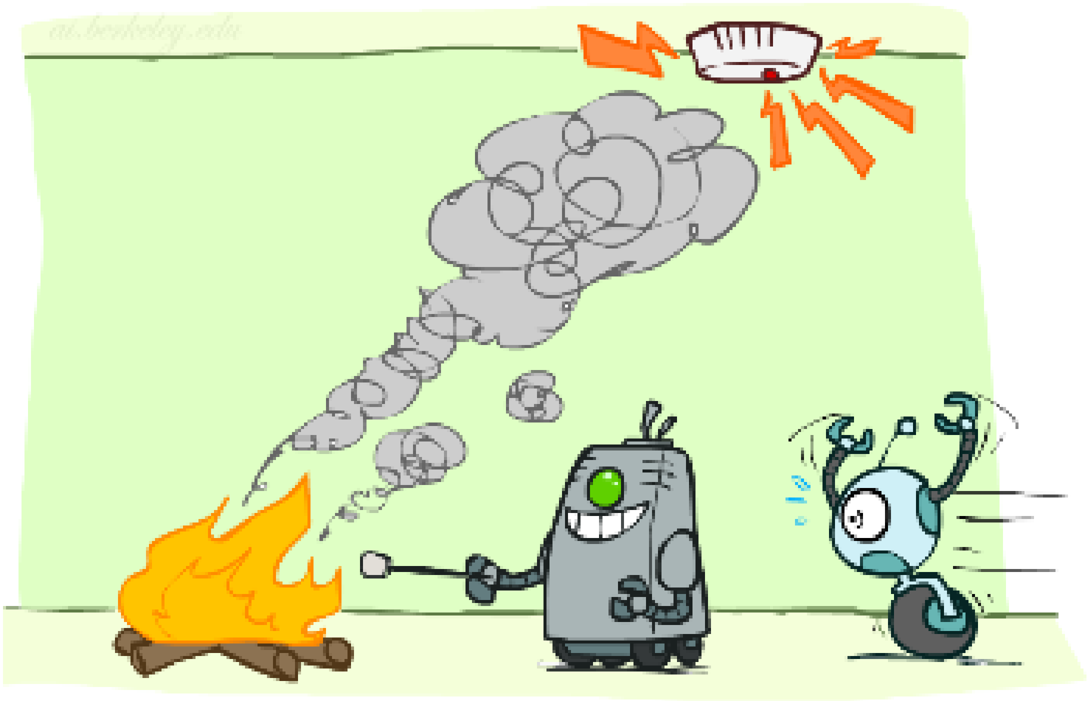
$$P(x|z,y) = P(x|z)$$

  or, equivalently, if and only if

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

Conditional Independence

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



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

$$P(alarm \mid smoke, fire) = P(alarm \mid smoke)$$

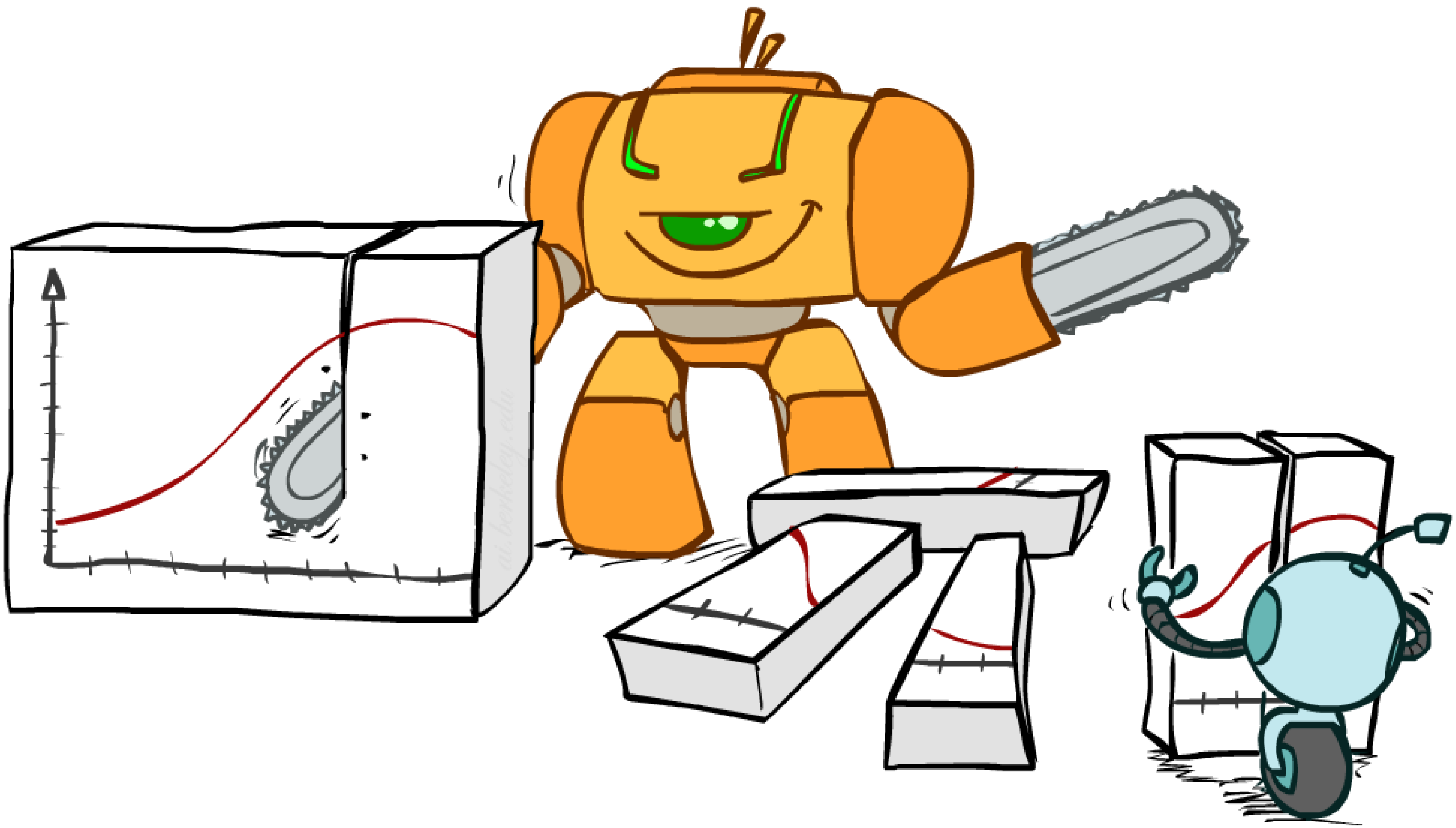
Conditional Independence

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

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

$$P(traffic \mid umbrella, rain) = P(traffic \mid rain)$$

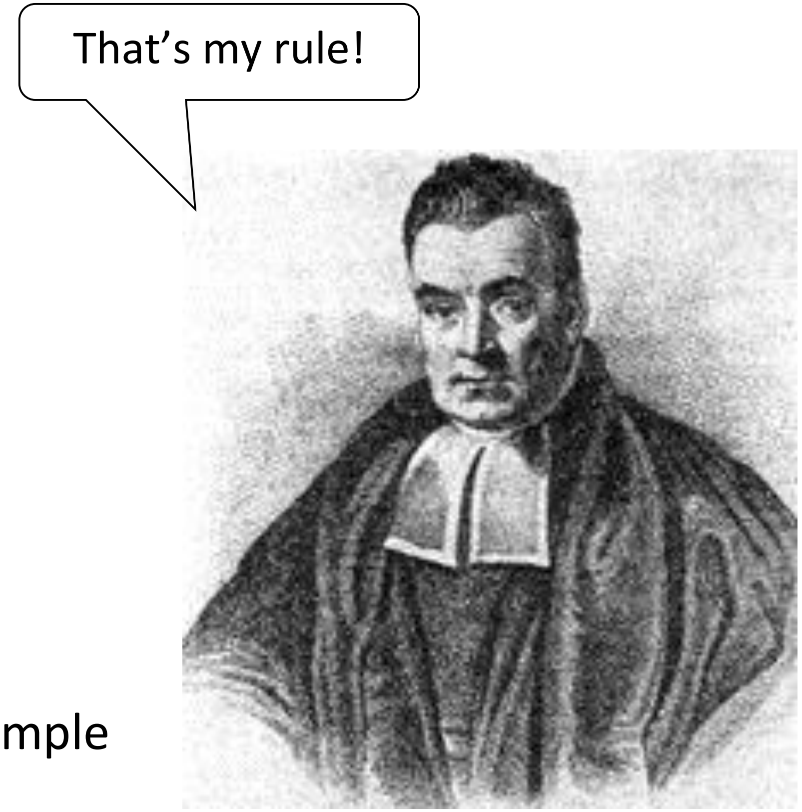
Bayes' Rule



Bayes' Rule

- Two ways to factor a joint distribution over two variables:

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



- Dividing, we get:

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

- Why is this at all helpful?
  - Lets us build one conditional from its reverse
  - Often one conditional is tricky but the other one is simple

- In the running for most important AI equation!

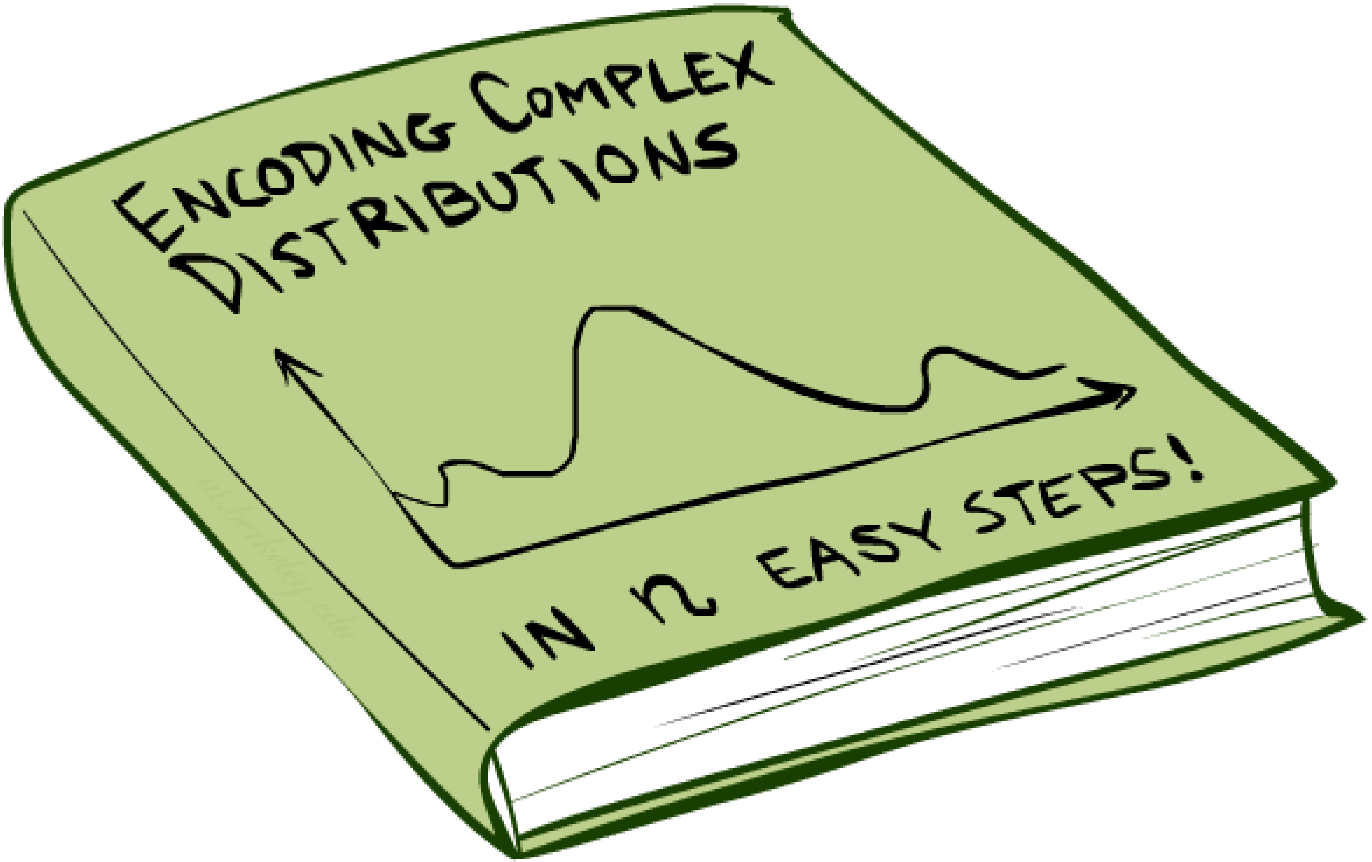
Example: building diagnostic probability from causal probability

$$P(\text{cause}|\text{effect}) = \frac{P(\text{effect}|\text{cause})P(\text{cause})}{P(\text{effect})}$$

- Example:
  - M: meningitis, S: stiff neck

$$P(+m|+s) = \frac{P(+s|+m)P(+m)}{P(+s)}$$

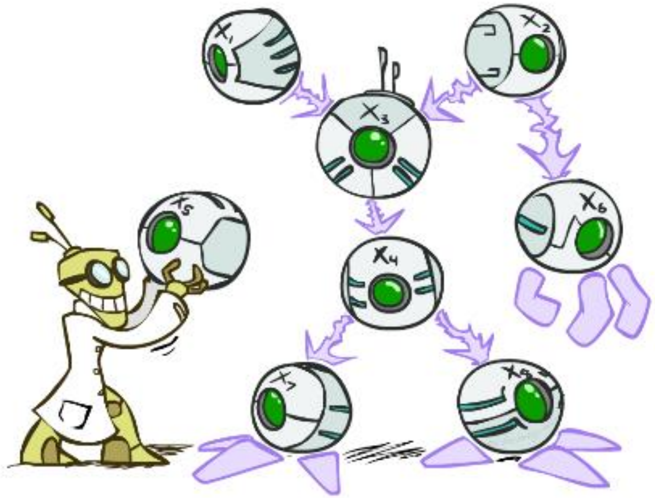
Bayes' Nets



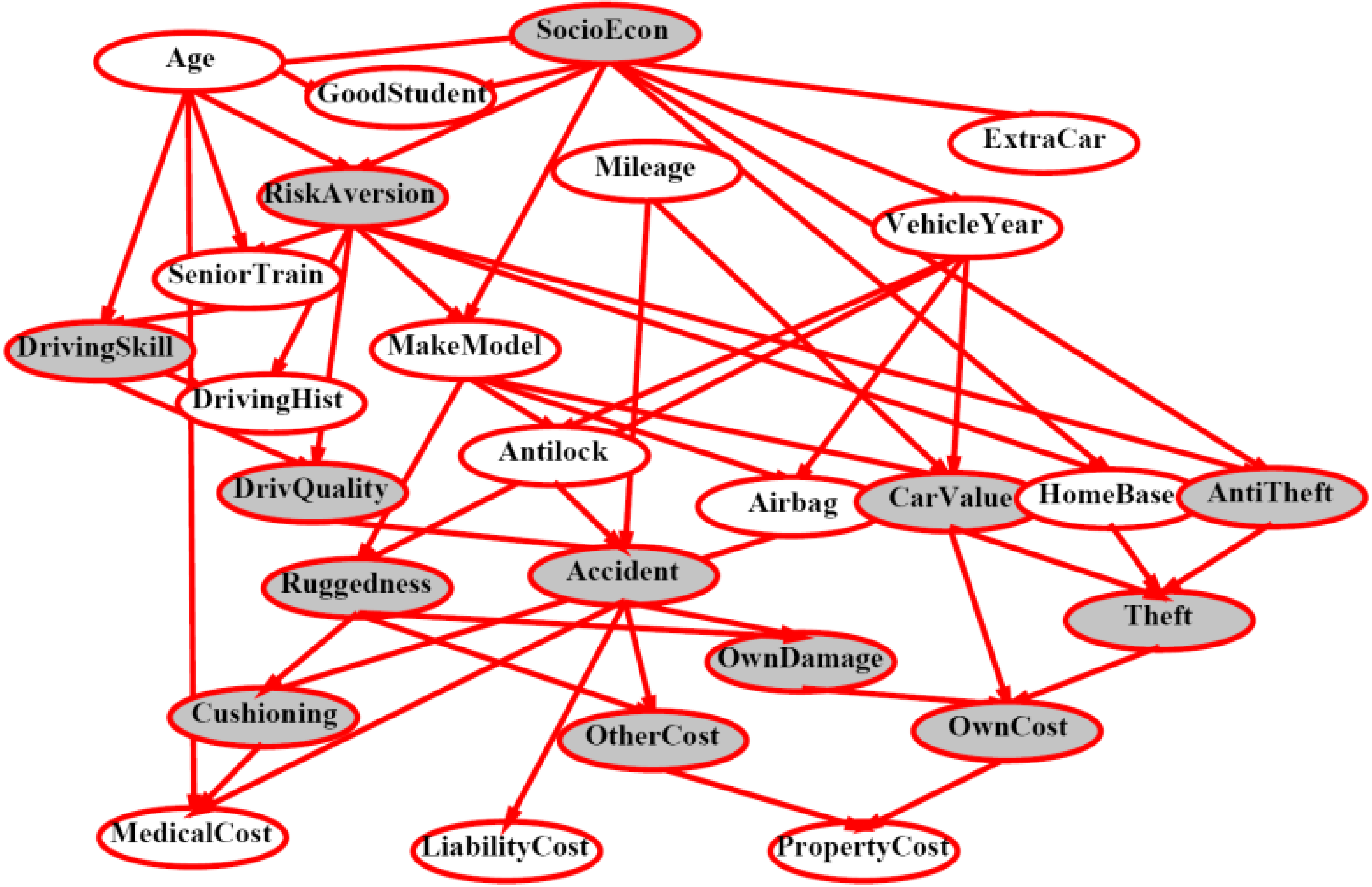
Bayes’ Nets: Big Picture

- Drawbacks 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



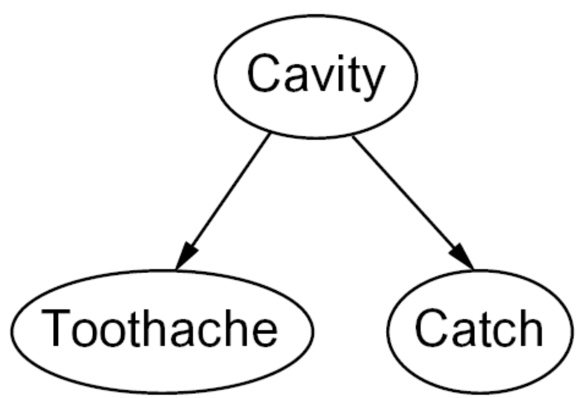
Example Bayes' Net: Insurance

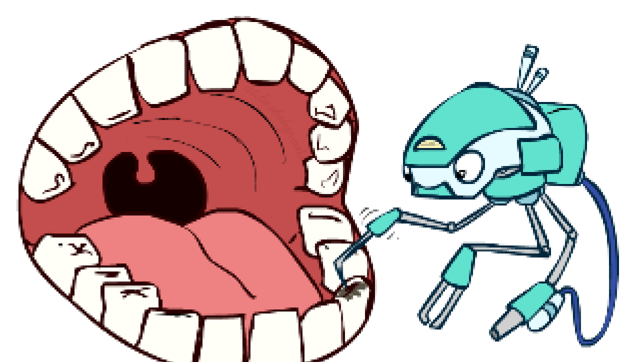


Graphical Model Notation

- Nodes: variables (with domains)
  - Can be assigned (observed) or unassigned (unobserved)

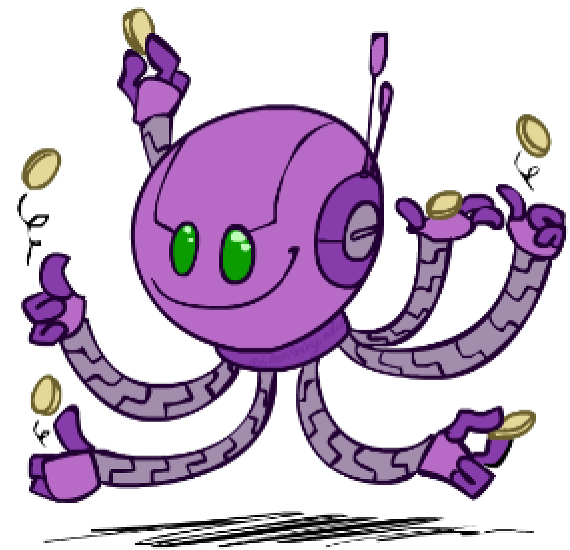
- Arcs: interactions
  - Similar to CSP constraints
  - Indicate "direct influence" between variables





Example: Coin Flips

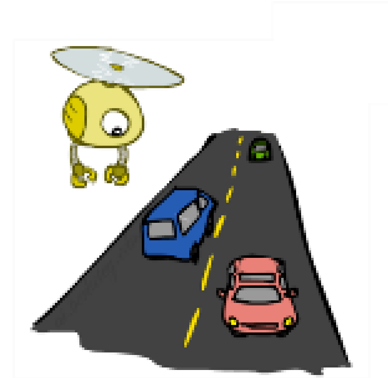
- N independent coin flips

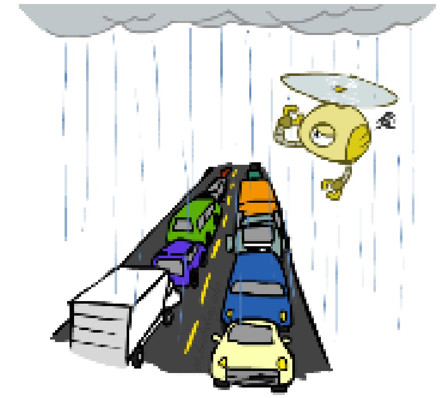


- No interactions between variables:

  absolute independence

Example: Traffic

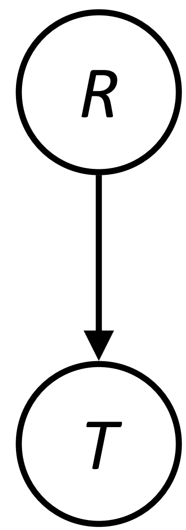




- Variables:
  - R: It rains
  - T: There is traffic

Model 1: independence

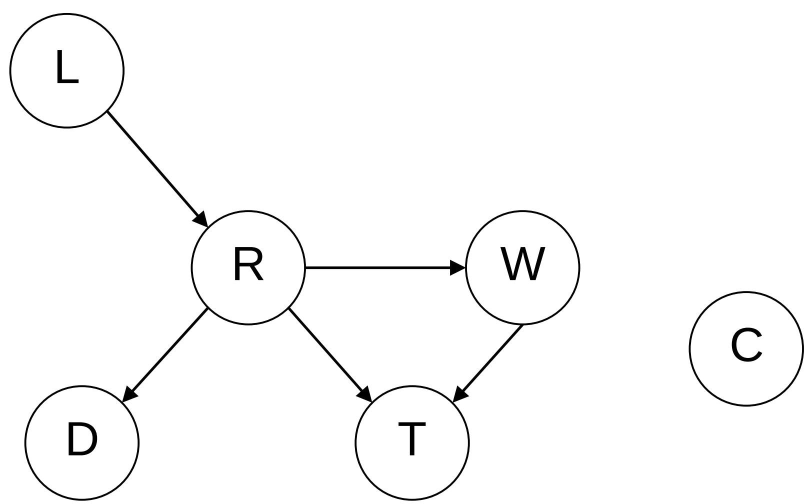
Model 2: rain causes traffic



- Why is an agent using model 2 better?

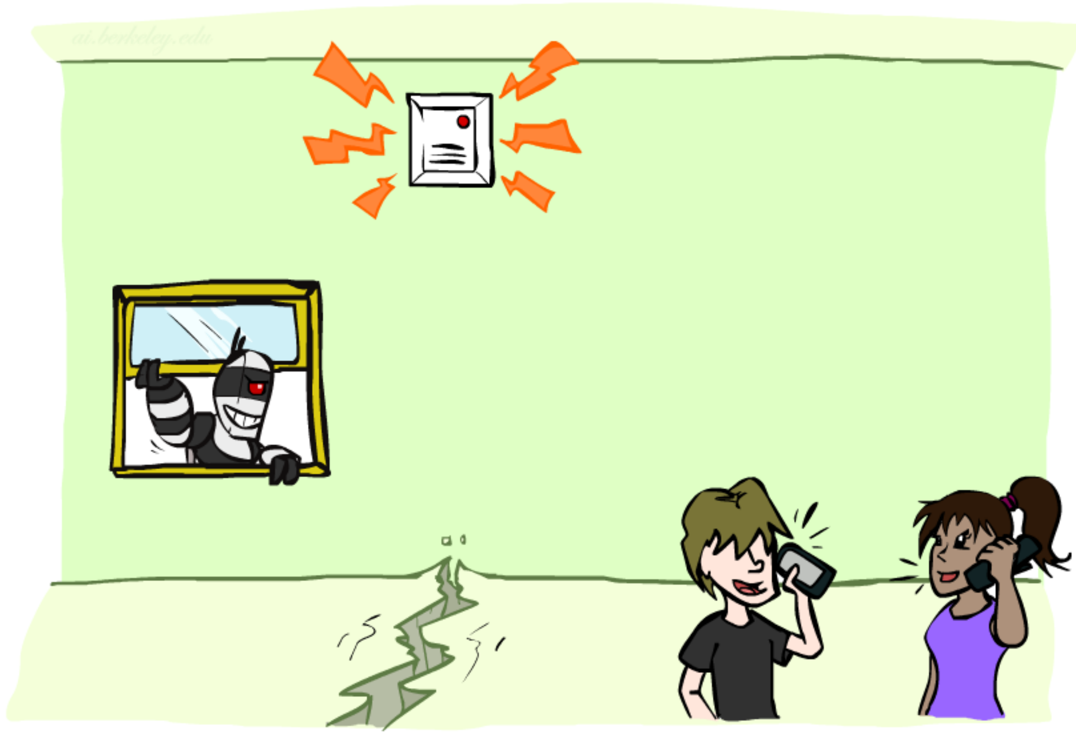
Example: Traffic II

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



Example: Alarm Network

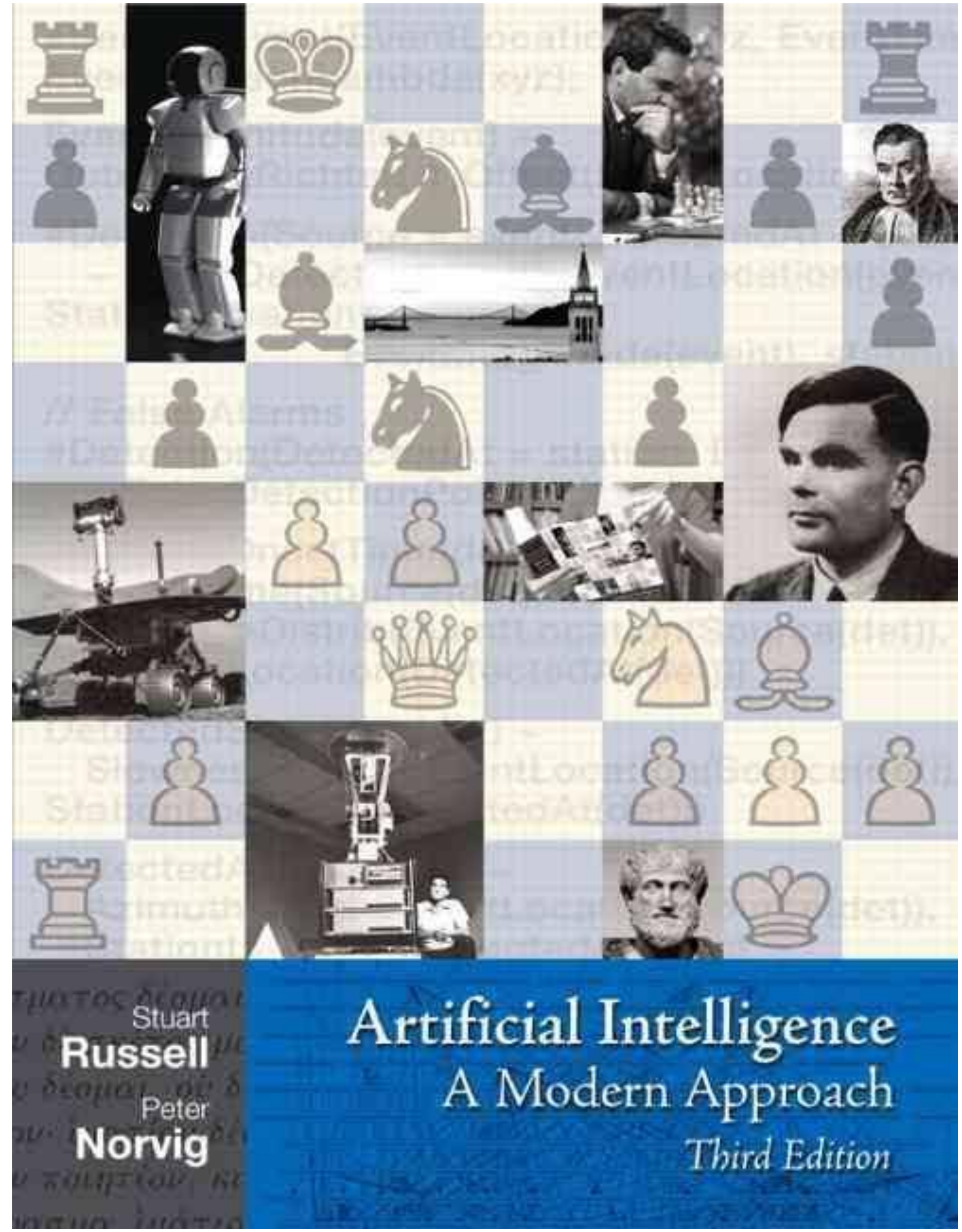
- Variables
  - B: Burglary
  - A: Alarm goes off
  - M: Mary calls
  - J: John calls
  - E: Earthquake!



Recommended reading

Stuart Russell, Peter Norvig: ***Artificial Intelligence A Modern Approach***

Chapter 13



ANY Questions?