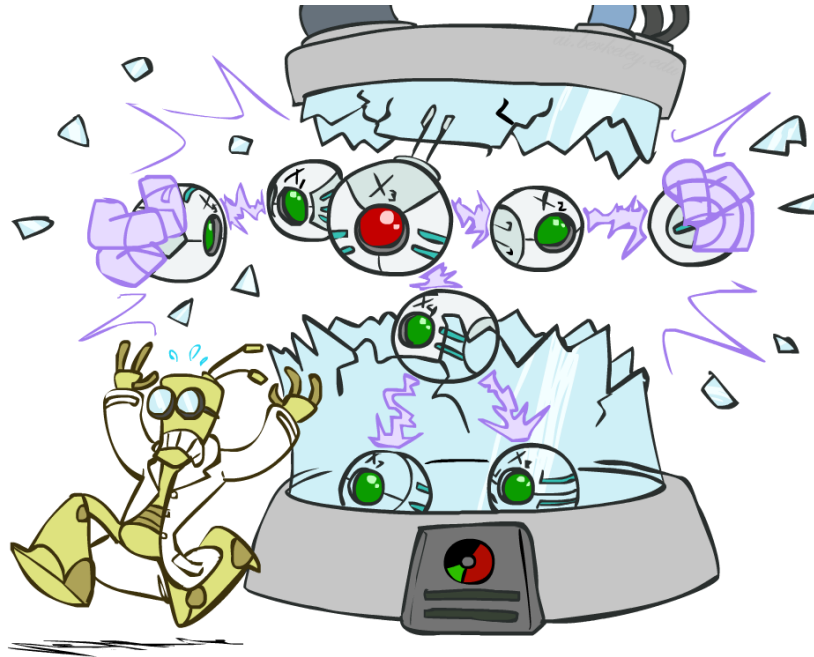


AI: Representation and Problem Solving

Bayes Nets: Independence



Instructor: Pat Virtue

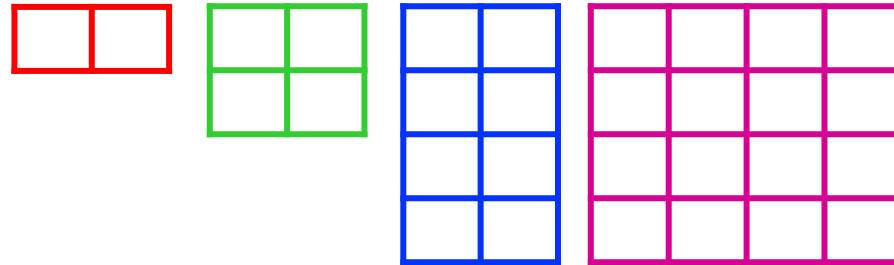
Slide credits: CMU AI and <http://ai.berkeley.edu>

Bayesian Networks

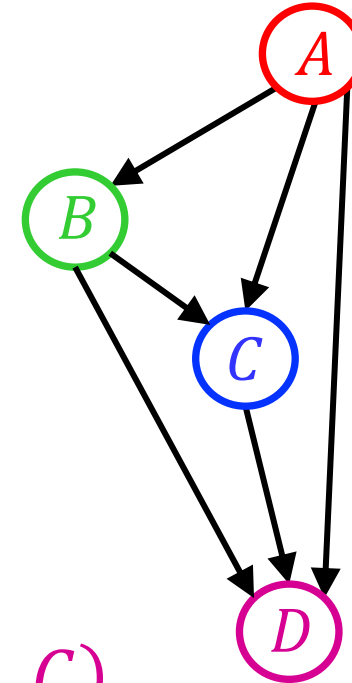
One node per random variable

DAG

One CPT per node: $P(\text{node} \mid \text{Parents}(\text{node}))$



Bayes net



$$P(A, B, C, D) = P(A) P(B|A) P(C|A, B) P(D|A, B, C)$$

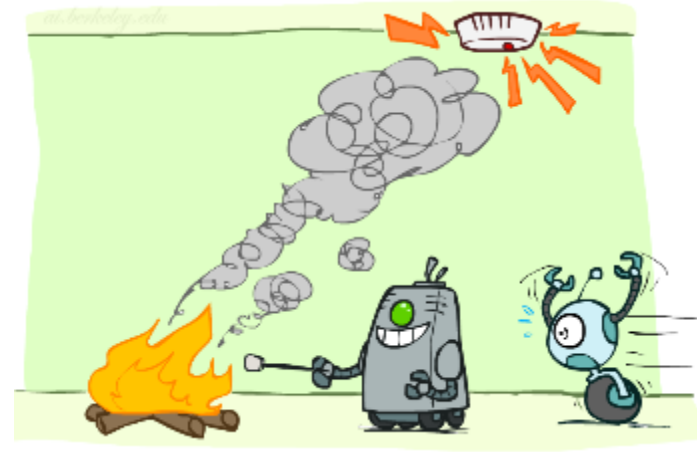
Encode joint distributions as product of conditional distributions on each variable

$$P(X_1, \dots, X_N) = \prod_i P(X_i \mid \text{Parents}(X_i))$$

Build Bayes Net Using Chain Rule

Binary random variables

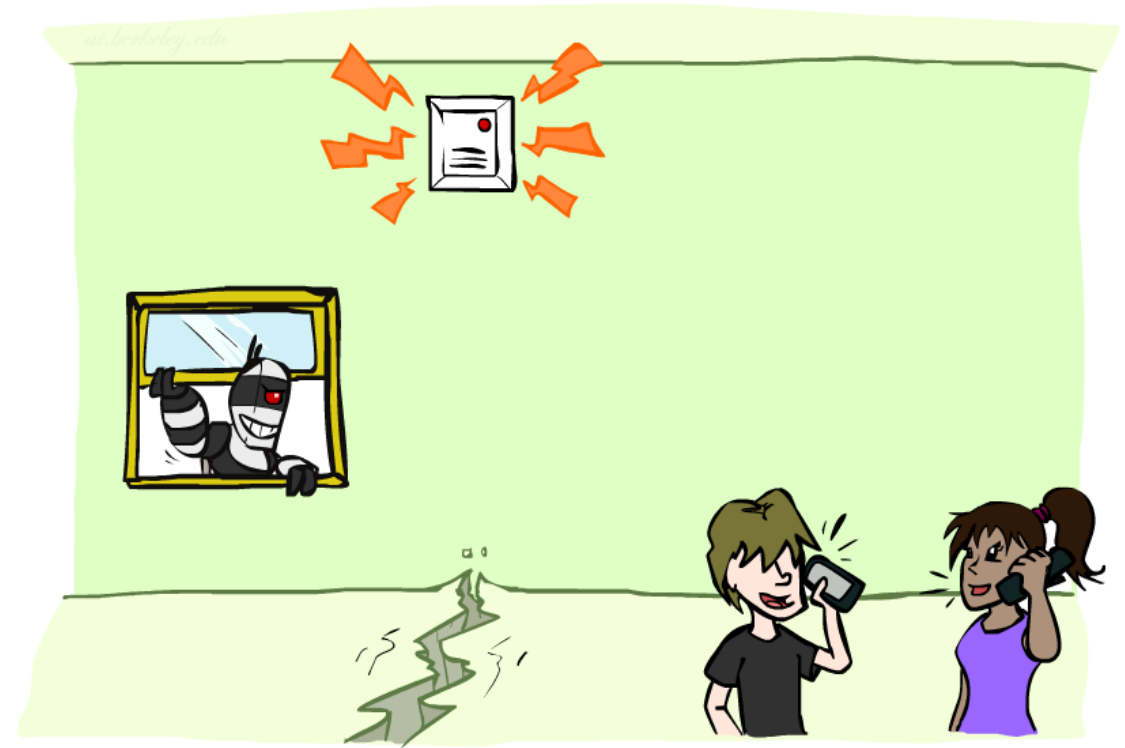
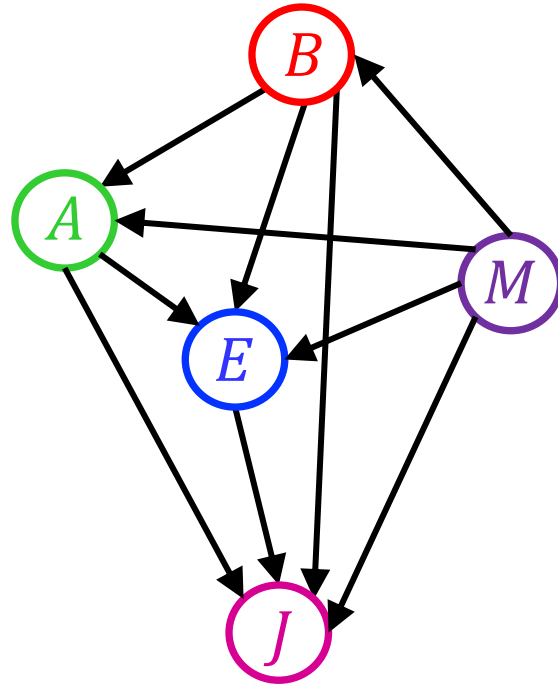
- Fire
- Smoke
- Alarm



Question

Variables

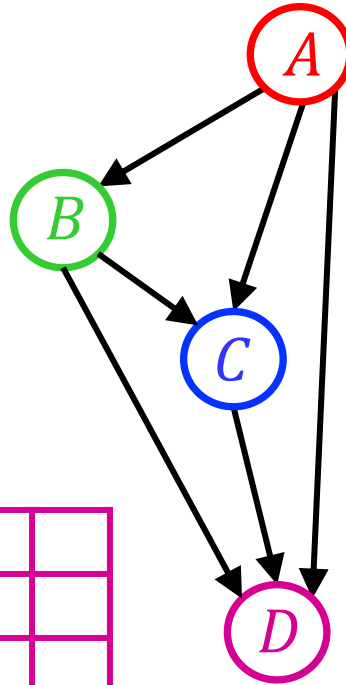
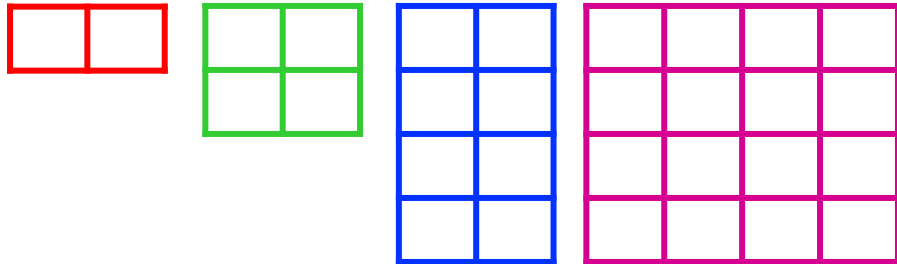
- B: Burglary
- A: Alarm goes off
- M: Mary calls
- J: John calls
- E: Earthquake!



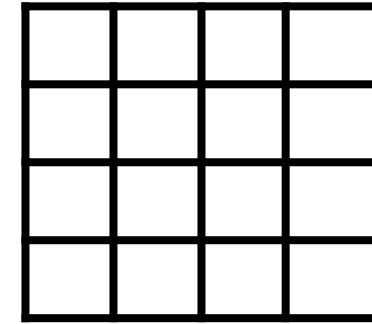
Given the Bayes net, write the joint distribution?

Answer Any Query from Bayes Net

Bayes Net and
Conditional
Probability Tables



Joint

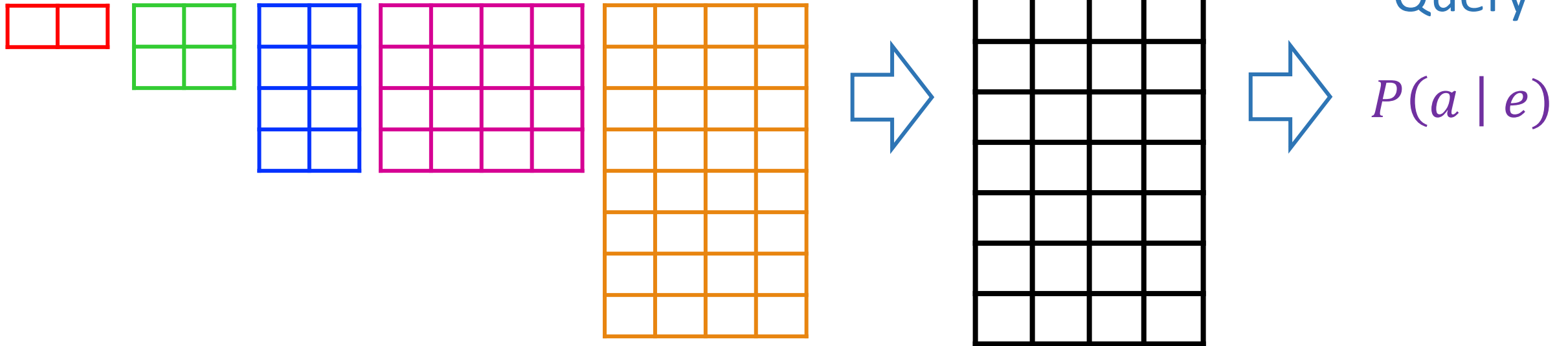


Query

$$P(a | e)$$

Answer Any Query from Condition Probability Tables

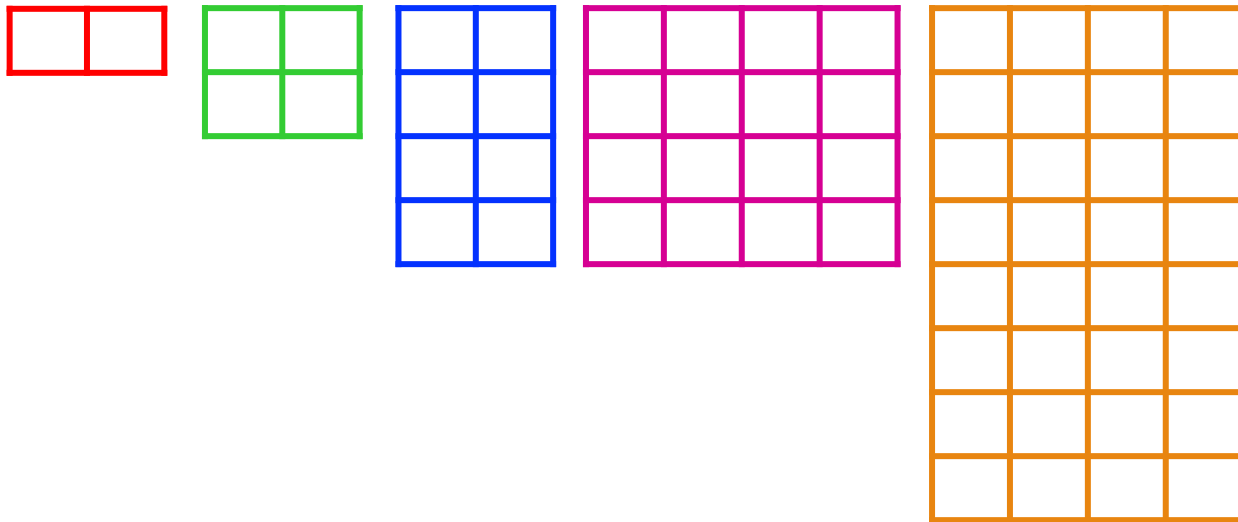
Conditional Probability Tables
and Chain Rule



$$P(A) \ P(B|A) \ P(C|A, B) \ P(D|A, B, C) \ P(E|A, B, C, D)$$

Answer Any Query from Condition Probability Tables

Conditional Probability Tables and Chain Rule



$$P(A) \ P(B|A) \ P(C|A, B) \ P(D|A, B, C) \ P(E|A, B, C, D)$$

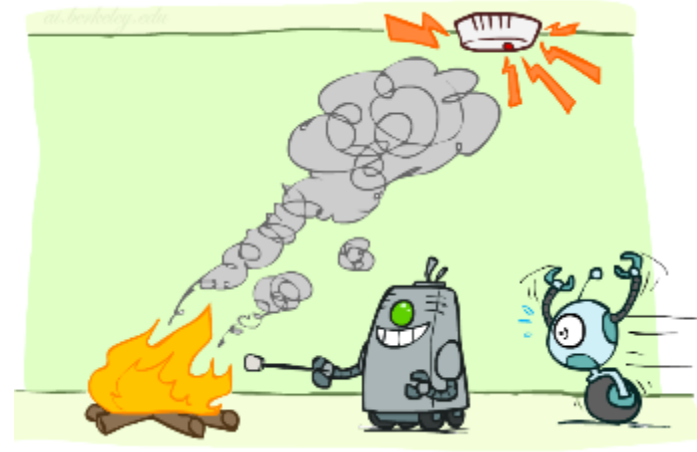
Problems

- Huge
 - n variables with d values
 - d^n entries
- We aren't given the right tables

Do We Need the Full Chain Rule?

Binary random variables

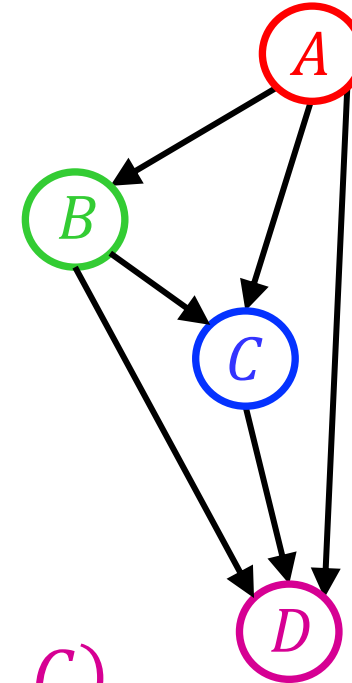
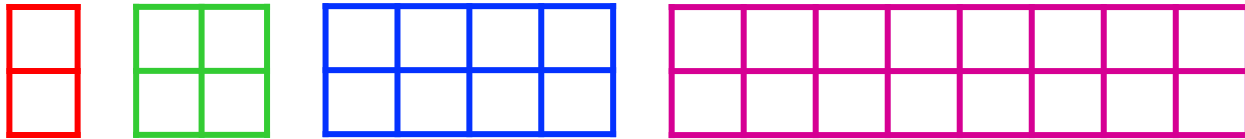
- Fire
- Smoke
- Alarm



Bayesian Networks

Bayes net

Example of full chain rule Bayes net



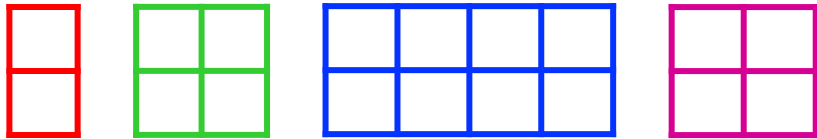
$$P(A, B, C, D) = P(A) P(B|A) P(C|A, B) P(D|A, B, C)$$

Encode joint distributions as product of conditional distributions on each variable

$$P(X_1, \dots, X_N) = \prod_i P(X_i | \text{Parents}(X_i))$$

Bayesian Networks

Missing edges (compared to full chain rule)
encode our independence assumptions

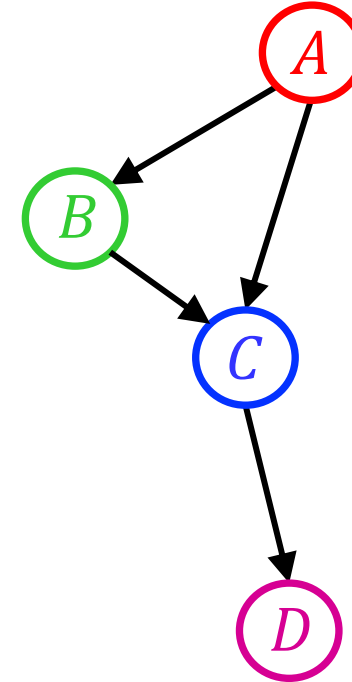


$$P(A, B, C, D) = P(A) P(B|A) P(C|A, B) P(D|C)$$

Encode joint distributions as product of conditional distributions on each variable

$$P(X_1, \dots, X_N) = \prod_i P(X_i | \text{Parents}(X_i))$$

Bayes net



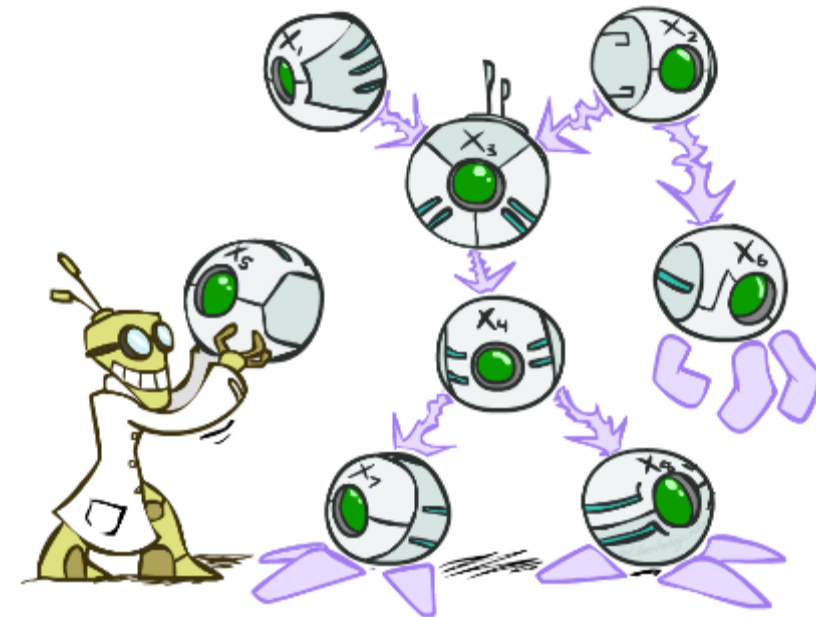
Bayes' Nets: Big Picture

Two problems with using full joint distribution tables as our probabilistic models:

- Usually, 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)

- A type of **probabilistic graphical models**
- We describe how variables locally interact
- Local interactions chain together to give global, indirect interactions



Graphical Model Notation

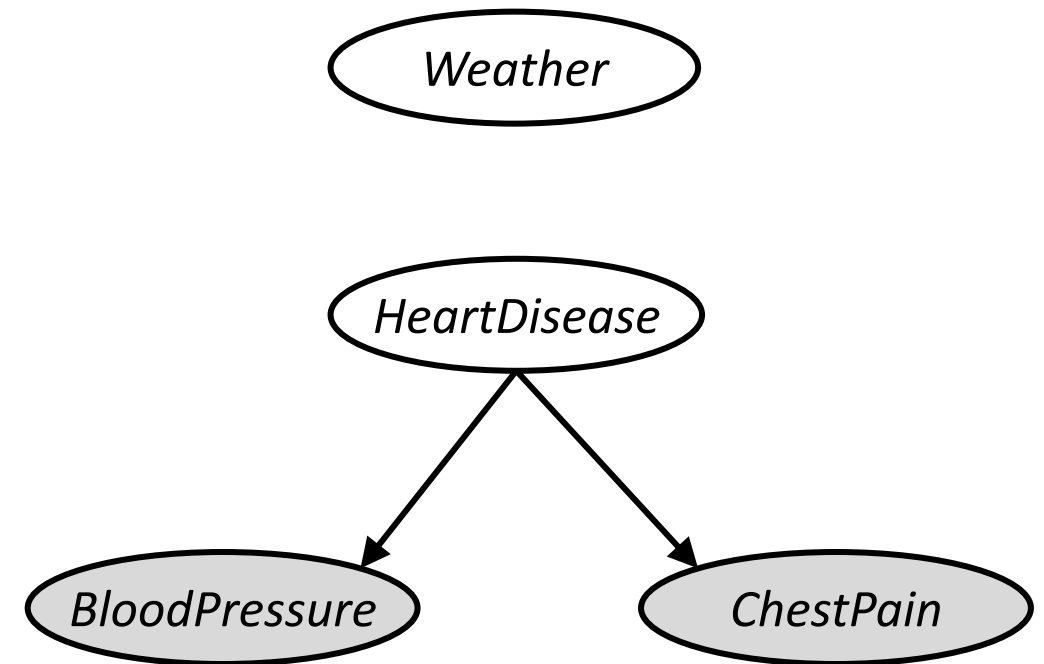
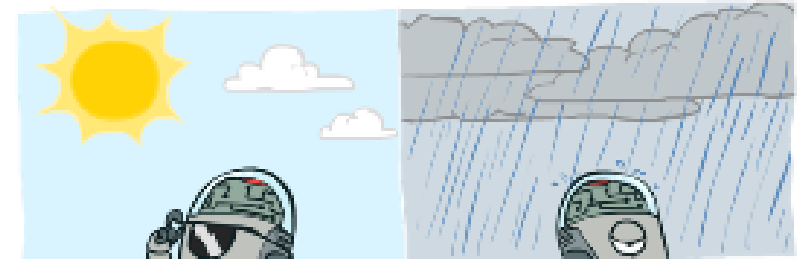
Nodes: variables (with domains)

- Can be assigned (observed) or unassigned (unobserved)
- We'll shade nodes to indicate observed variables
- Observed **does not** mean *Variable = true*
Observed just means that we will have the value for that variable

Edges

- Indicate “direct influence” between variables
- Absence of edges: encode conditional independence

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



Maggie Makar, University of Michigan



Assistant Professor
Computer Science and Engineering
University of Michigan

<https://mymakar.github.io/>

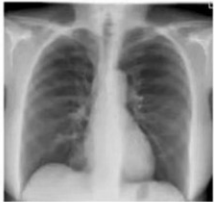
Causally-motivated shortcut removal using auxiliary labels

M. Makar, B. Packer, D. Moldovan, D. Blalock, Y. Halpern, A. D'Amour

AIStats 2022 [[paper](#)]

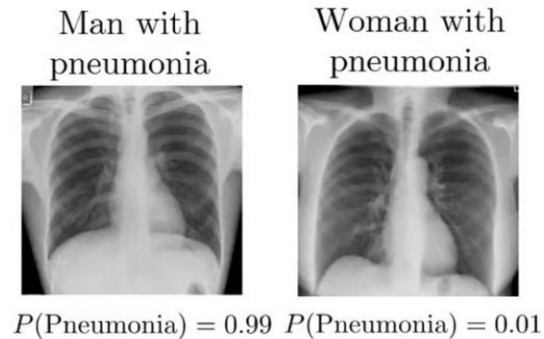
Pneumonia detection under biased sampling

Training data

Target label:	Healthy	Pneumonia
	$Y = 0$	$Y = 1$
Men $V = 0$		
Women $V = 1$		

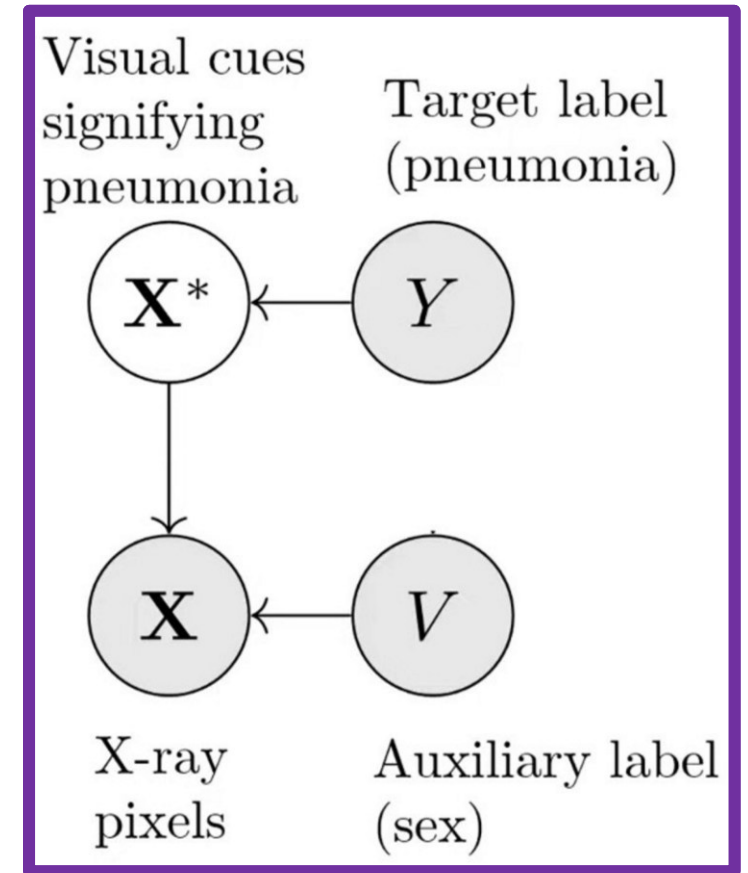
Auxiliary label:

Testing data



Shortcut learning

Cases courtesy of Dr. Andrew Dixon, Dr Henry Knip, Dr. Usman Bashir, and Dr. Ian Bickle , Radiopaedia.org, rID: 48366, 31388, 18394 and rID: 50318



Causally-motivated shortcut removal using auxiliary labels

M. Makar, B. Packer, D. Moldovan, D. Blalock, Y. Halpern, A. D'Amour

AIStats 2022 [[paper](#)]

Danielle Belgrave, Microsoft Research



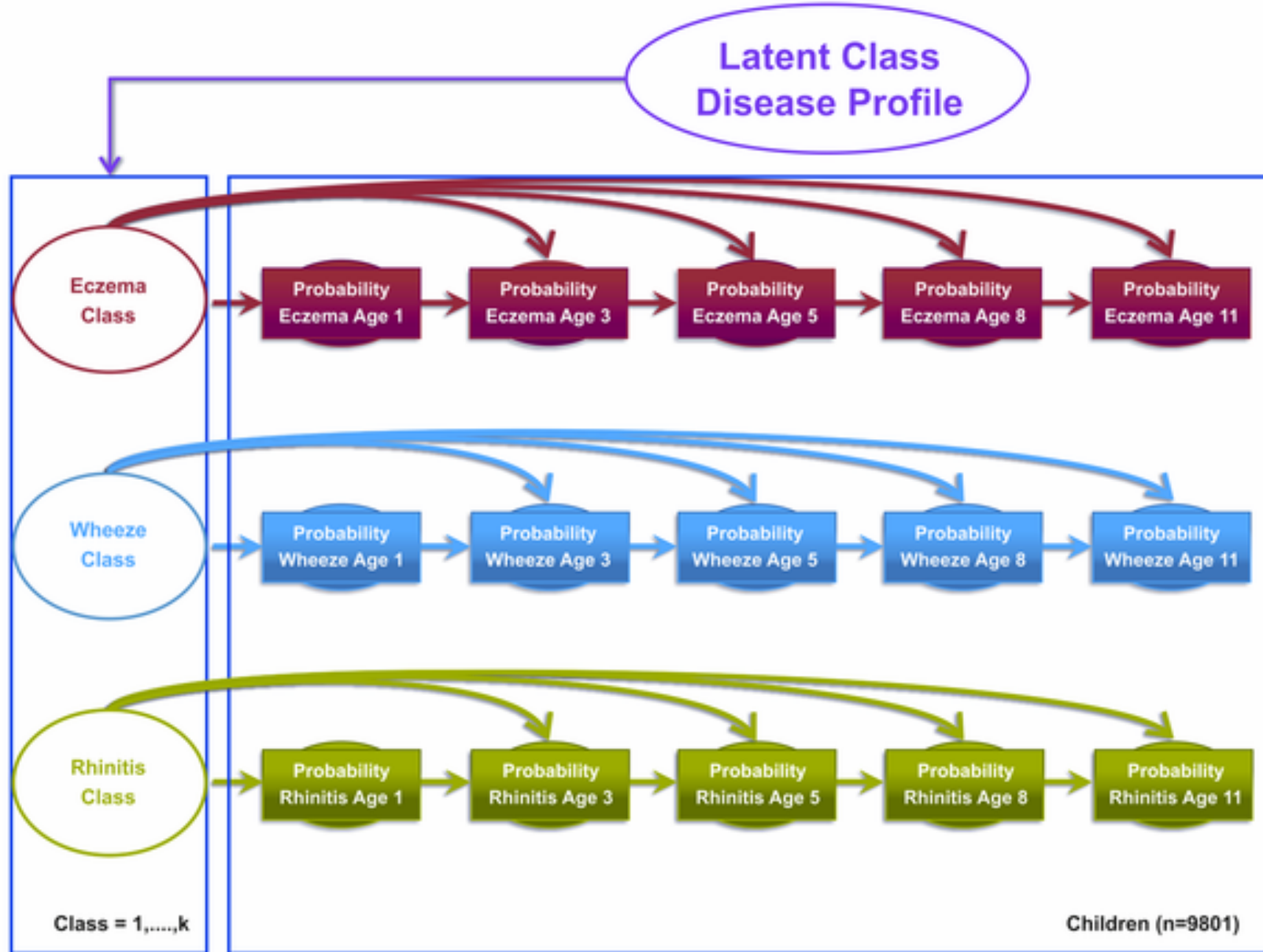
Vice President
AI/ML
GSK



<https://www.daniellebelgrave.com>

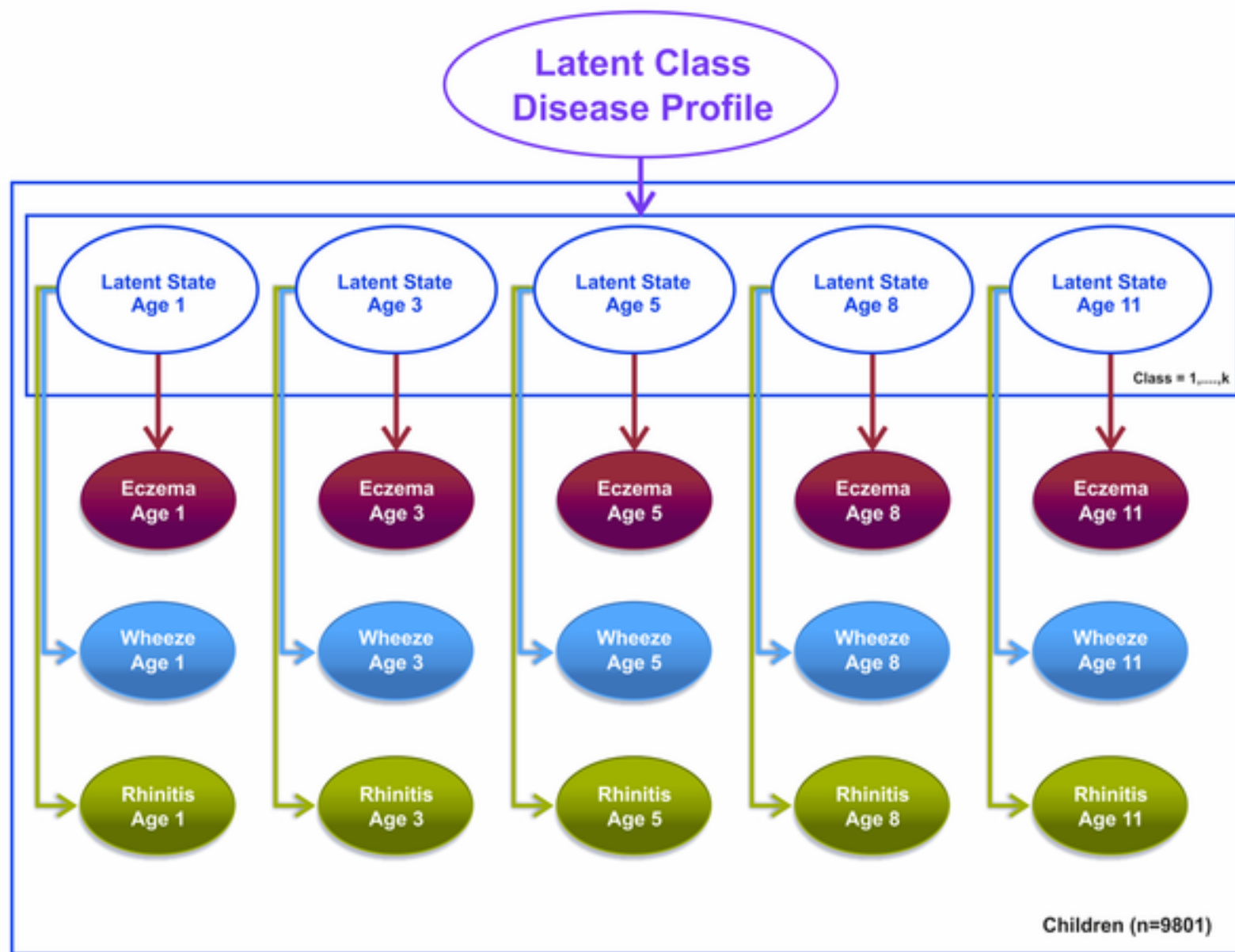
Developmental Profiles of Eczema, Wheeze, and Rhinitis:
Two Population-Based Birth Cohort Studies
Danielle Belgrave, et al. *PLOS Medicine*, 2014

<https://journals.plos.org/plosmedicine/article?id=10.1371/journal.pmed.1001748>



Danielle Belgrave, et al. *PLOS Medicine*, 2014

<https://journals.plos.org/plosmedicine/article?id=10.1371/journal.pmed.1001748>



Danielle Belgrave, et al. *PLOS Medicine*, 2014

<https://journals.plos.org/plosmedicine/article?id=10.1371/journal.pmed.1001748>

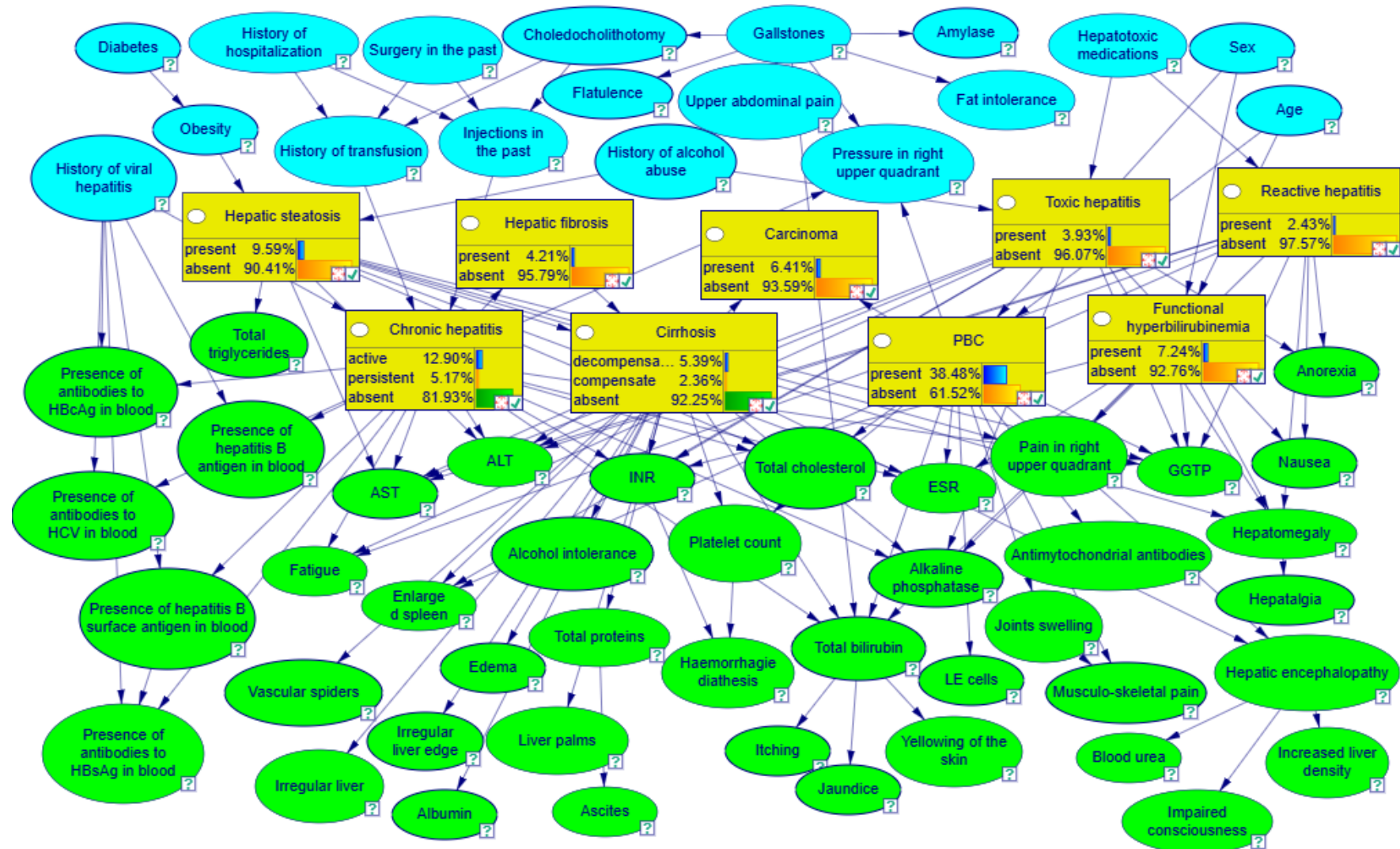
Characteristic	MAAS Cohort		ALSPAC Cohort		Joint MAAS and ALSPAC Cohort	
	<i>n</i> /Total	Percent	<i>n</i> /Total	Percent	<i>n</i> /Total	Percent
Gender (Female)	617/1,136	54.3	4,212/8,665	48.61	4,829/9,801	49.3
Eczema						
Age 1 y	383/1,077					
Age 3 y	355/1,061					
Age 5 y	340/1,050					
Age 8 y	285/1,027					
Age 11 y	216/924					
Wheeze						
Age 1 y	300/1,087					
Age 3 y	257/1,095					
Age 5 y	238/1,056					
Age 8 y	185/1,024					
Age 11 y	173/916					
Rhinitis						
Age 1 y	8/943					
Age 3 y	49/1,075					
Age 5 y	292/1,039					
Age 8 y	297/1,027					
Age 11 y	321/927					

	<i>n</i> /Total	Percent
Gender (Female)	617/1,136	54.3
Eczema		
Age 1 y	383/1,077	35.6
Age 3 y	355/1,061	33.5
Age 5 y	340/1,050	32.4
Age 8 y	285/1,027	27.8
Age 11 y	216/924	23.4

Danielle Belgrave, et al. *PLOS Medicine*, 2014

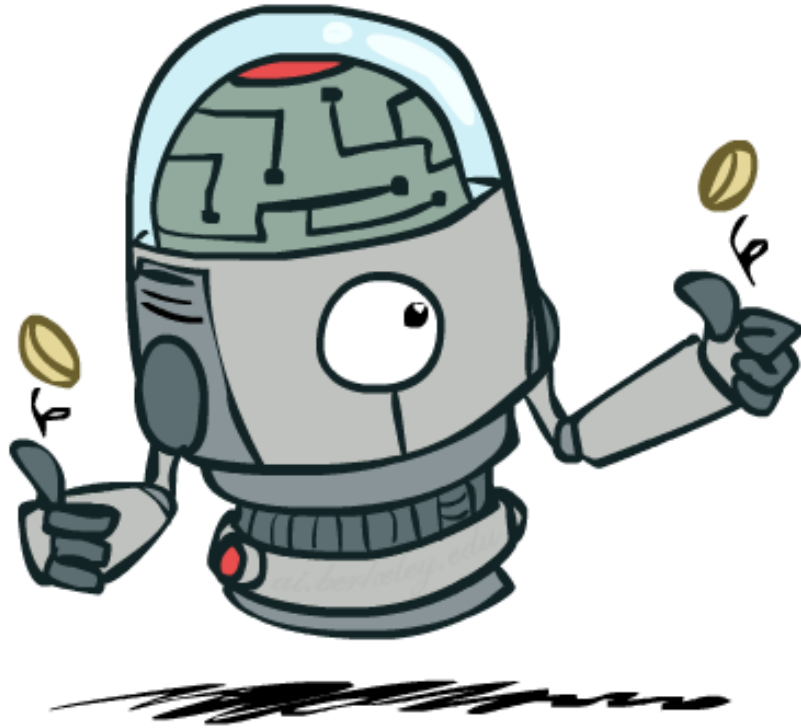
<https://journals.plos.org/plosmedicine/article?id=10.1371/journal.pmed.1001748>

Example: Liver Disorders



<https://demo.bayesfusion.com/bayesbox.html>

Independence



Independence

Two variables X and Y are *independent* if

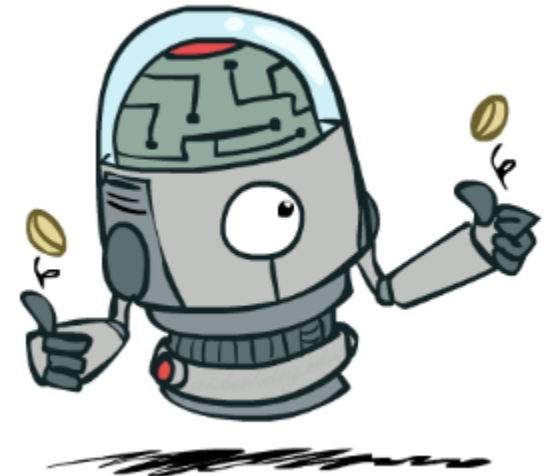
$$\forall x, y \quad P(x, y) = P(x) P(y)$$

- This says that their joint distribution *factors* into a product of two simpler distributions
- Combine with product rule $P(x, y) = P(x | y) P(y)$ we obtain another form:
$$\forall x, y \quad P(x | y) = P(x) \quad \text{or} \quad \forall x, y \quad P(y | x) = P(y)$$

Example: two dice rolls R_1 and R_2

$$P(R_1=5, R_2=5) = P(R_1=5) P(R_2=5) = 1/6 \times 1/6 = 1/36$$

$$P(R_2=5 | R_1=5) = P(R_2=5)$$



Example: Independence

n fair, independent coin flips:

$P(X_1)$

H	0.5
T	0.5

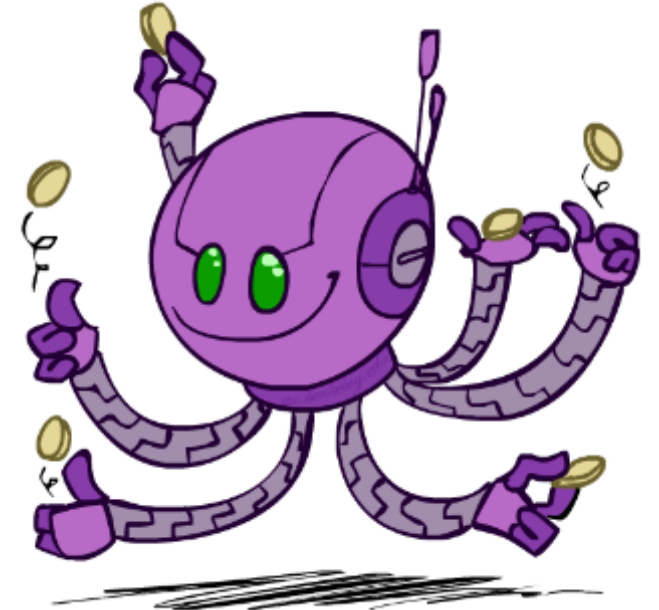
$P(X_2)$

H	0.5
T	0.5

...

$P(X_n)$

H	0.5
T	0.5



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

2^n



Poll 1

Are T and W independent?

$P_1(T, W)$

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

$P(W)$

W	P
sun	0.6
rain	0.4

Poll 1

Are T and W independent?

No

$P(T, W)$

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

$P(T)P(W)$

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

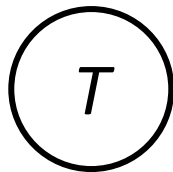
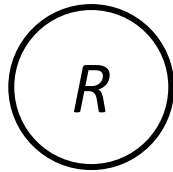
Example: Traffic

Variables:

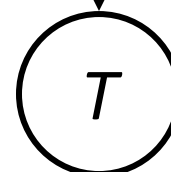
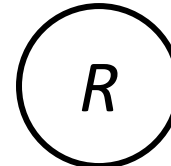
- R : rain or not
- T : traffic or not



Model 1: independence



Model 2: rain affects traffic



Why is an agent using model 2 better?

Conditional Independence

Absolute (unconditional) 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

if and only if:

$$\forall x, y, z \quad P(x \mid y, z) = P(x \mid z)$$

or, equivalently, if and only if

$$\forall x, y, z \quad P(x, y \mid z) = P(x \mid z) P(y \mid z)$$

Independence Rules

Independence

If A and B are independent, then:

$$P(A, B) = P(A)P(B)$$

$$P(A \mid B) = P(A)$$

$$P(B \mid A) = P(B)$$

Conditional independence

If A and B are conditionally independent given C, then:

$$P(A, B \mid C) = P(A \mid C)P(B \mid C)$$

$$P(A \mid B, C) = P(A \mid C)$$

$$P(B \mid A, C) = P(B \mid C)$$

Conditional Independence

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

If it's raining, the probability that there is traffic doesn't depend on whether see an umbrella:

- $P(+\text{traffic} \mid +\text{umbrella}, +\text{rain}) = P(+\text{traffic} \mid +\text{rain})$

The same independence holds if it's not raining:

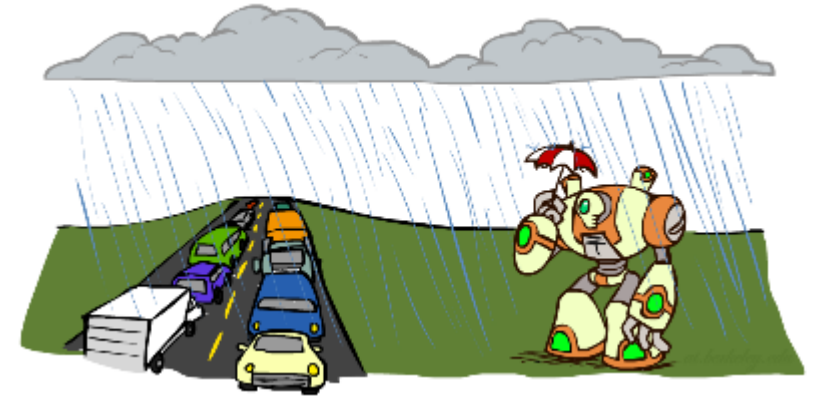
- $P(+\text{traffic} \mid +\text{umbrella}, -\text{rain}) = P(+\text{traffic} \mid -\text{rain})$

Traffic is *conditionally independent* of Umbrella given Rain:

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

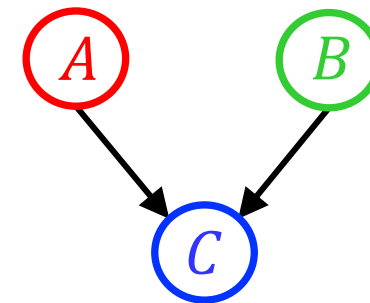
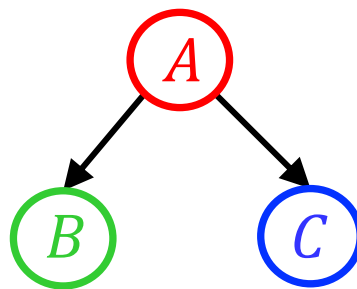
Equivalent statements:

- $P(\text{Umbrella} \mid \text{Traffic}, \text{Rain}) = P(\text{Umbrella} \mid \text{Rain})$
- $P(\text{Umbrella}, \text{Traffic} \mid \text{Rain}) = P(\text{Umbrella} \mid \text{Rain}) P(\text{Traffic} \mid \text{Rain})$
- One can be derived from the other easily



Poll 2

Match the product of CPTs to the Bayes net.



I. $P(A) P(B|A) P(C|B)$

$P(A) P(B|A) P(C|A)$

$P(A) P(B) P(C|A, B)$

II. $P(A) P(B) P(C|A, B)$

$P(A) P(B|A) P(C|B)$

$P(A) P(B|A) P(C|A)$

III. $P(A) P(B|A) P(C|A)$

$P(A) P(B) P(C|A, B)$

$P(A) P(B|A) P(C|B)$

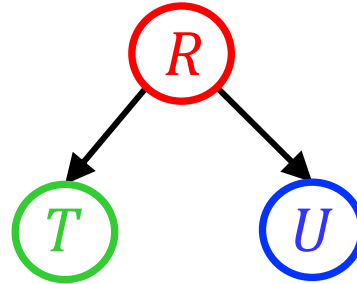
Conditional Independence Semantics

Common local relationships within a Bayes net

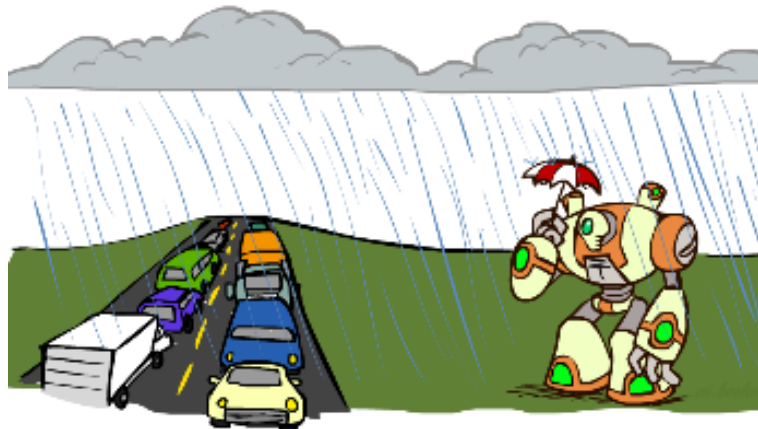
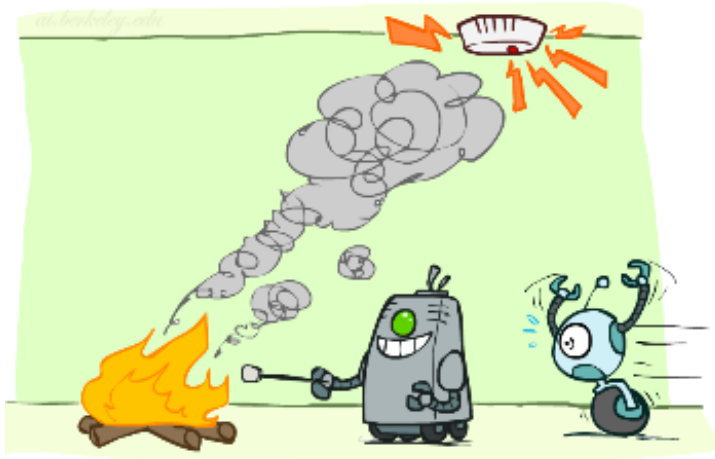
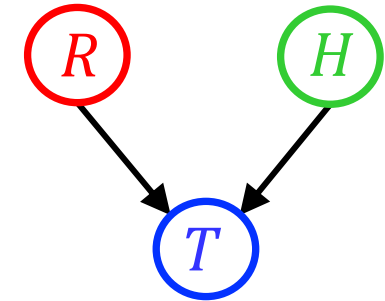
Causal Chain



Common Cause



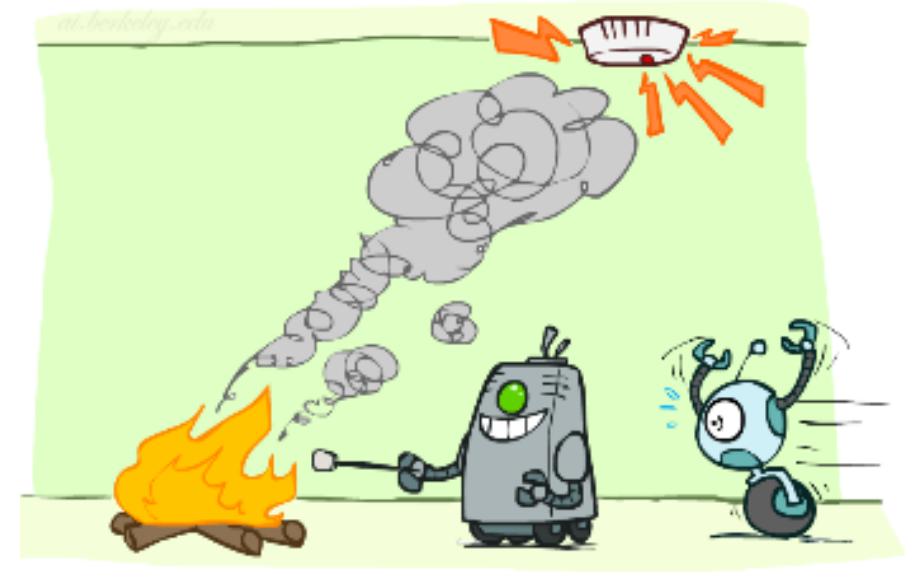
Common Effect



Causal Chain

Fire, Smoke, Alarm

- Causal story to create Bayes net



- Joint distribution

- Assumptions?

Common Cause

Chain rule:

$$P(x_1, x_2, \dots, x_n) = \prod_i P(x_i \mid x_1, \dots, x_{i-1})$$

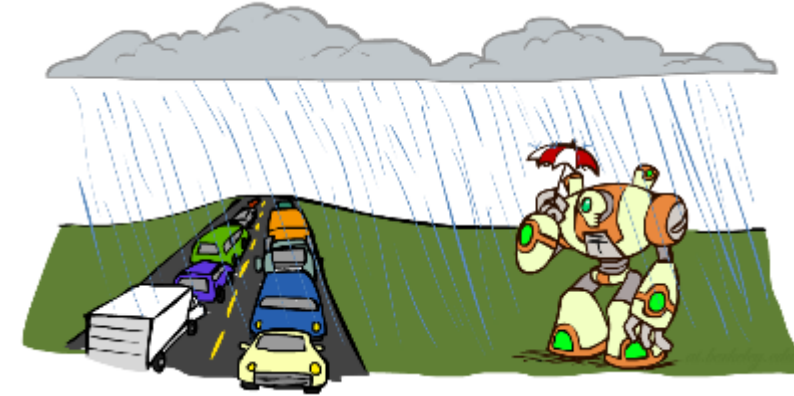
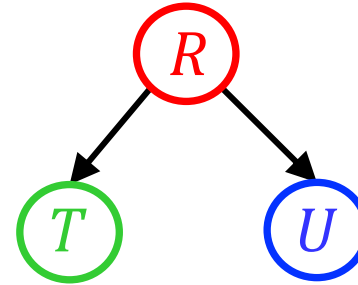
Trivial decomposition:

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

With assumption of conditional independence:

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

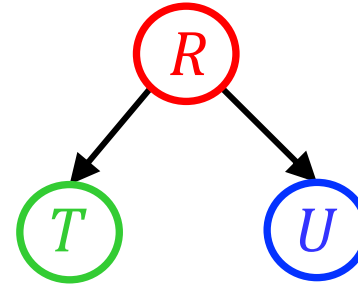
Bayes nets / graphical models help us express conditional independence assumptions



Common Cause

Chain rule:

$$P(x_1, x_2, \dots, x_n) = \prod_i P(x_i \mid x_1, \dots, x_{i-1})$$



Trivial decomposition:

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

With assumption of conditional independence:

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

Bayes nets / graphical models help us express conditional independence assumptions

Common Effect

Chain rule:

$$P(x_1, x_2, \dots, x_n) = \prod_i P(x_i \mid x_1, \dots, x_{i-1})$$

Trivial decomposition:

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

With assumption of conditional independence:

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

Bayes nets / graphical models help us express conditional independence assumptions



Common Effect

Chain rule:

$$P(x_1, x_2, \dots, x_n) = \prod_i P(x_i \mid x_1, \dots, x_{i-1})$$

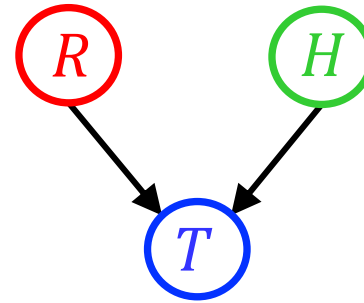
Trivial decomposition:

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

With assumption of conditional independence:

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

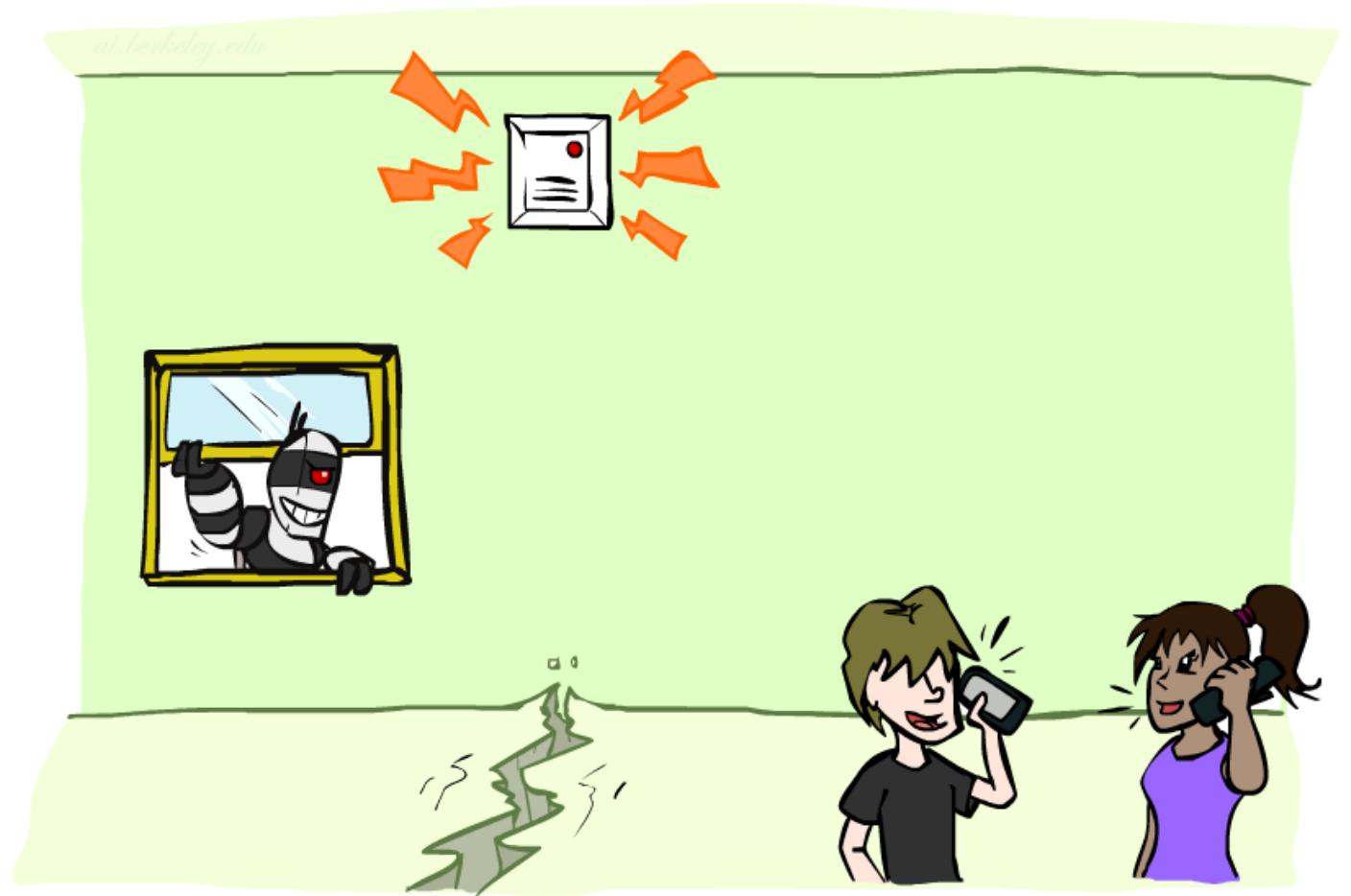
Bayes nets / graphical models help us express conditional independence assumptions



Example: Alarm Network

Variables

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



Example: Alarm Network

Joint distribution factorization example

Generic chain rule

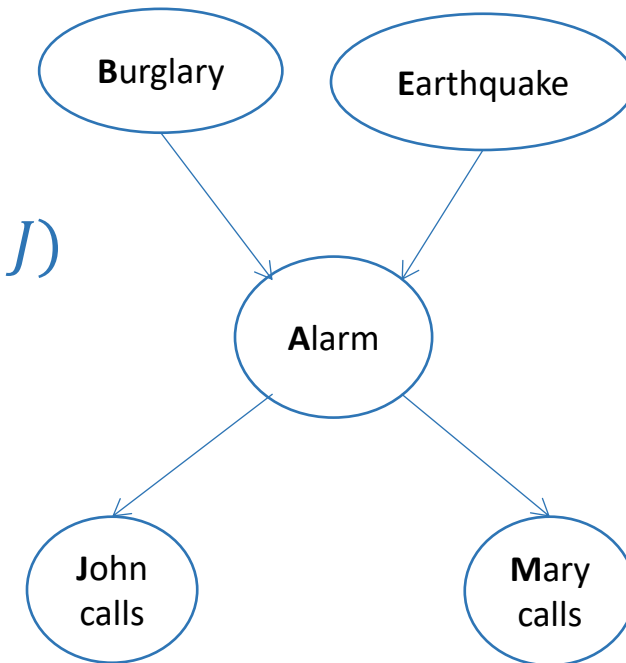
$$\blacksquare P(X_1 \dots X_n) = \prod_i P(X_i | X_1 \dots X_{i-1})$$

$$P(B, E, A, J, M) = P(B) P(E|B) P(A|B, E) P(J|B, E, A) P(M|B, E, A, J)$$

$$P(B, E, A, J, M) = P(B) P(E) P(A|B, E) P(J|A) P(M|A)$$

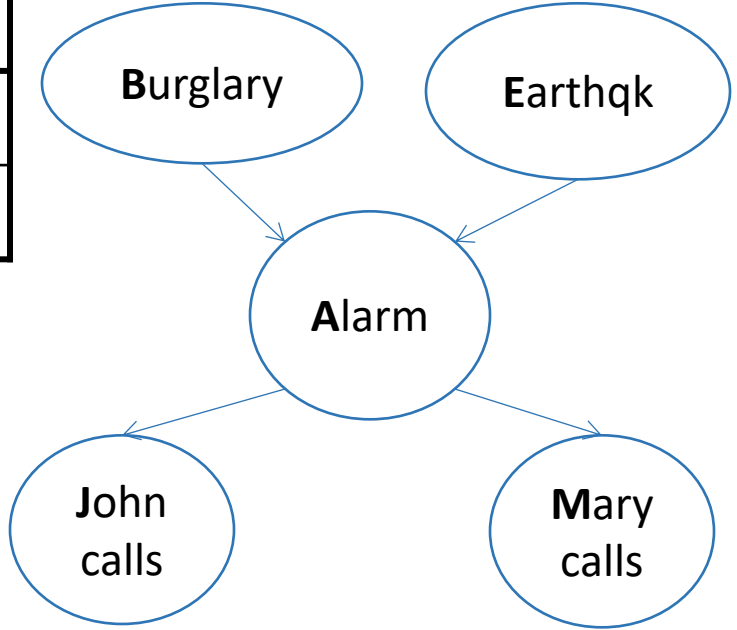
Bayes nets

$$\blacksquare P(X_1 \dots X_n) = \prod_i P(X_i | \text{Parents}(X_i))$$



Example: Alarm Network

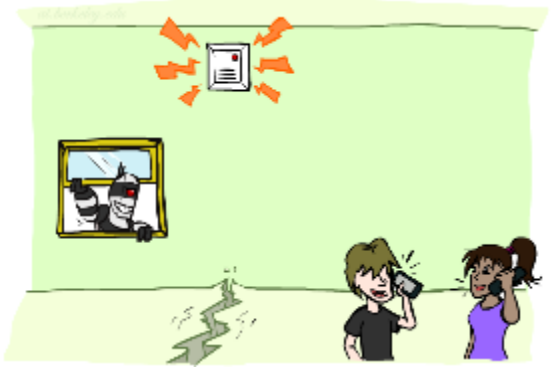
B	P(B)
+b	0.001
-b	0.999



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

Conditional Independence Semantics

For the following Bayes nets, write the joint $P(A, B, C)$

1. Using the chain rule (with top-down order A,B,C)
2. Using Bayes net semantics (product of CPTs)



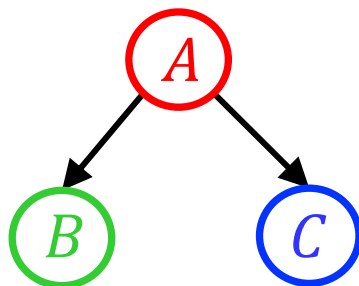
$$P(A) P(B|A) P(C|A, B)$$

$$P(A) P(B|A) P(C|B)$$

Assumption:

$$P(C|A, B) = P(C|B)$$

C is independent from A given B



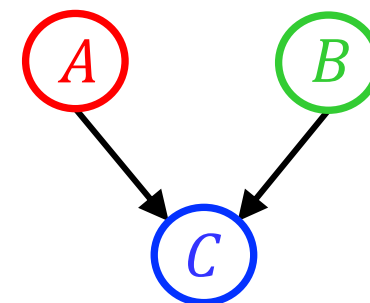
$$P(A) P(B|A) P(C|A, B)$$

$$P(A) P(B|A) P(C|A)$$

Assumption:

$$P(C|A, B) = P(C|A)$$

C is independent from B given A



$$P(A) P(B|A) P(C|A, B)$$

$$P(A) P(B) P(C|A, B)$$

Assumption:

$$P(B|A) = P(B)$$

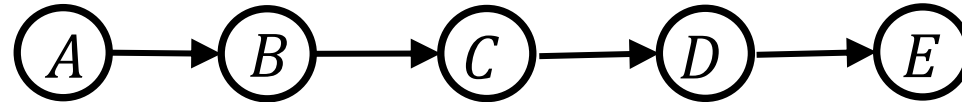
A is independent from B given { }

Bayes Net Independence

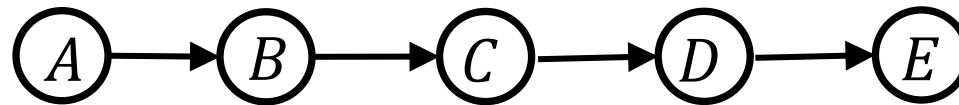


Answering Independence Questions

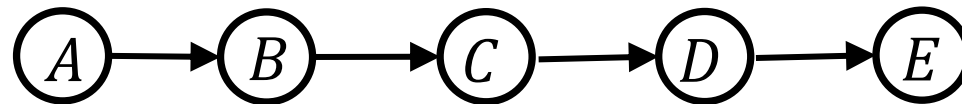
- Is A independent from E?



- Is A independent from E given C?



- Is A independent from C given E?



Active / Inactive Paths

Question: Are X and Y conditionally independent given evidence variables {Z}?

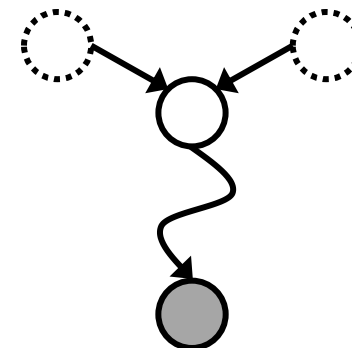
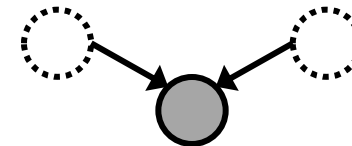
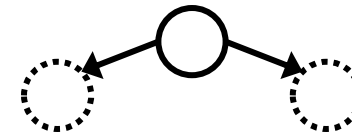
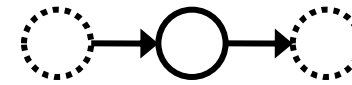
- Yes, if X and Y “d-separated” by Z
- Consider all (undirected) paths from X to Y
- No active paths = independence!

A path is active if each triple is active:

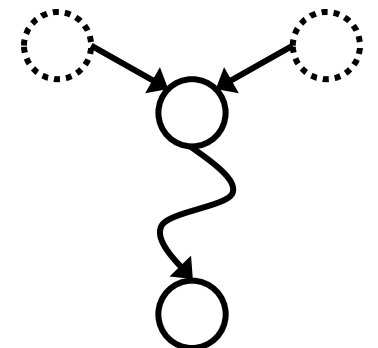
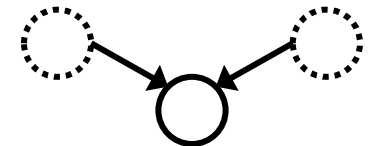
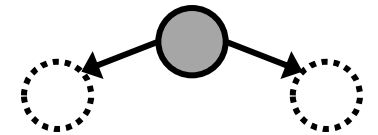
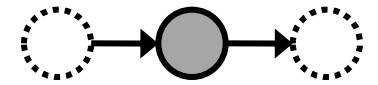
- Causal chain $A \rightarrow B \rightarrow C$ where B is unobserved (either direction)
- Common cause $A \leftarrow B \rightarrow C$ where B is unobserved
- Common effect (aka v-structure)
 $A \rightarrow B \leftarrow C$ where B or one of its descendents is observed

All it takes to block a path is a single inactive segment

Active Paths



Inactive Paths



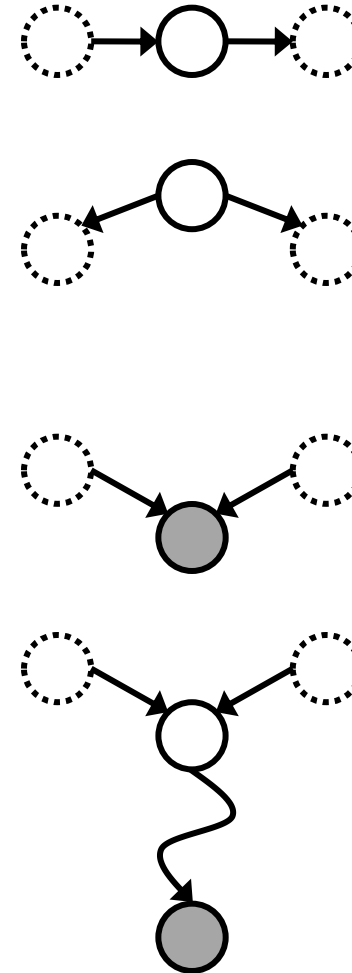
Bayes Ball

Question: Are X and Y conditionally independent given evidence variables {Z}?

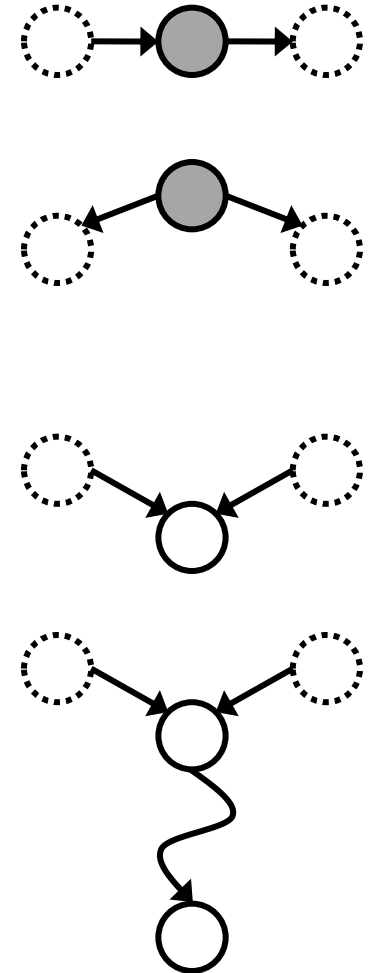


Shachter, Ross D. "Bayes-Ball: Rational Pastime (for Determining Irrelevance and Requisite Information in Belief Networks and Influence Diagrams)." *Proceedings of the Fourteenth conference on Uncertainty in Artificial Intelligence*. 1998.

Active Paths



Inactive Paths

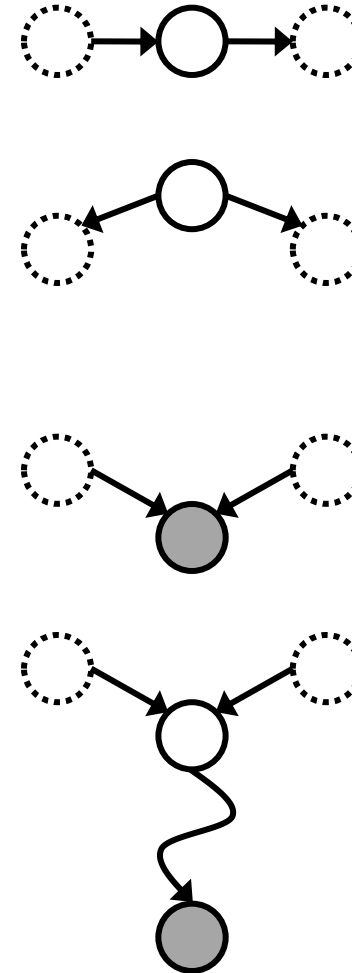


Bayes Ball

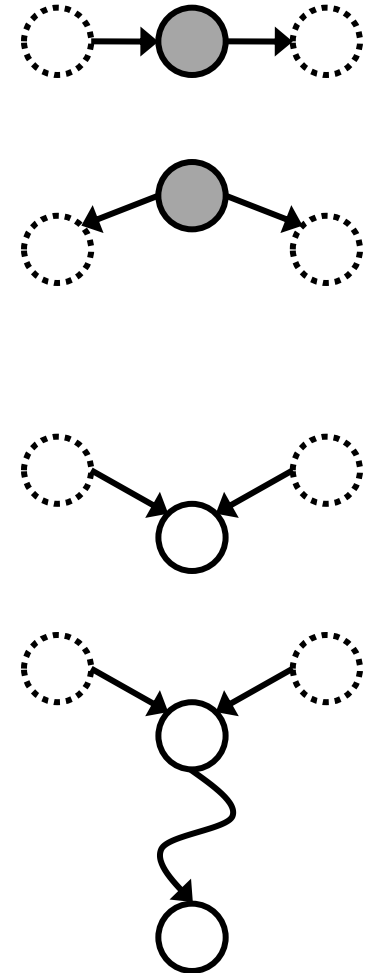
Question: Are X and Y conditionally independent given evidence variables {Z}?

1. Shade in Z
2. Drop a ball at X
3. The ball can pass through any *active* path and is blocked by any *inactive* path (ball can move either direction on an edge)
4. If the ball reaches Y, then X and Y are NOT conditionally independent given Z

Active Paths

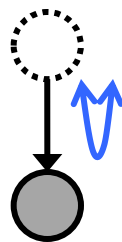
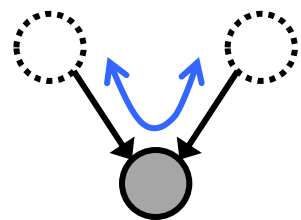
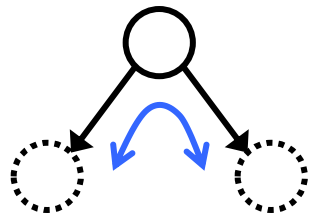
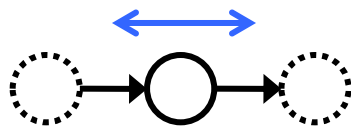


Inactive Paths

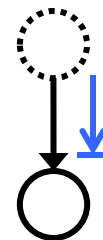
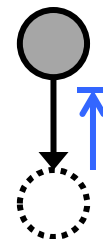
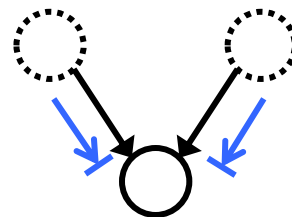
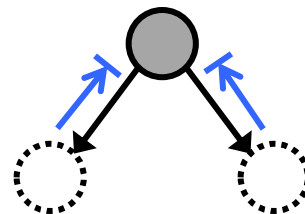
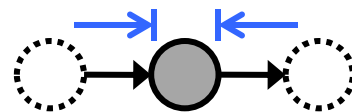


Bayes Ball

Active Paths

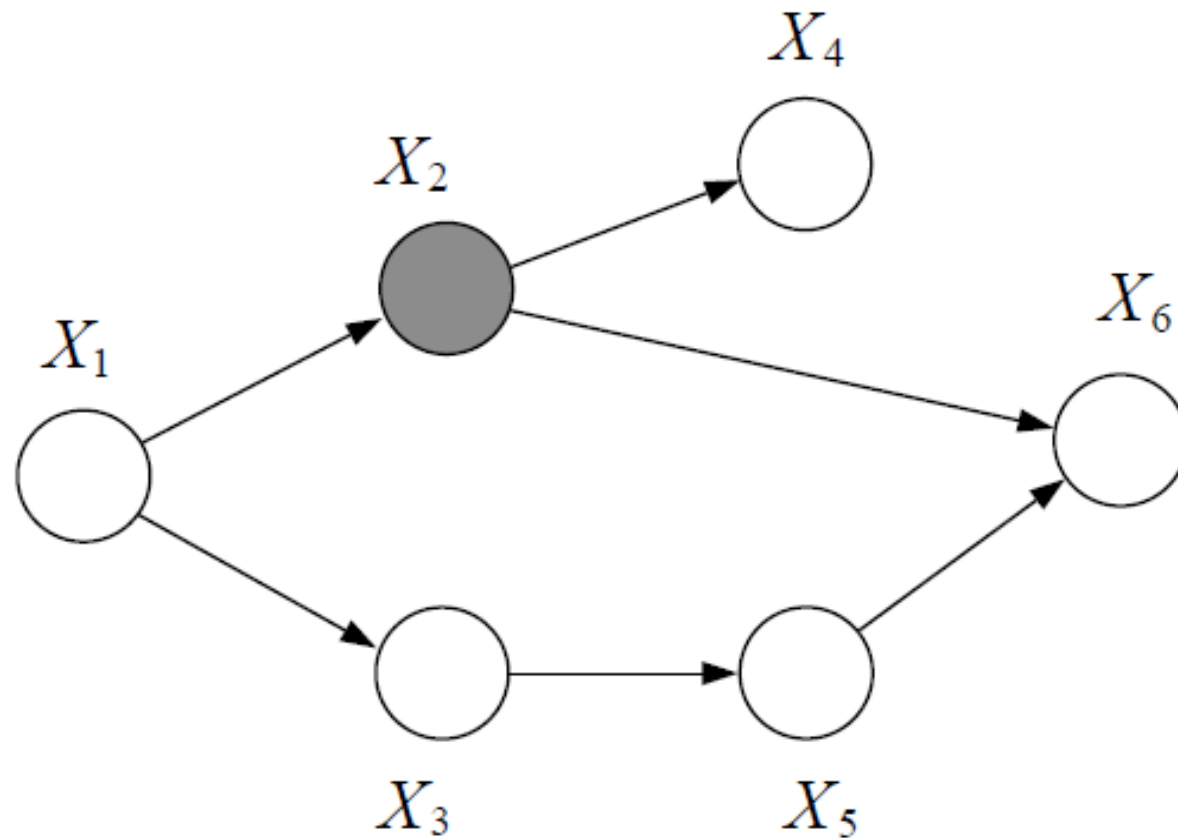


Inactive Paths



Poll 3

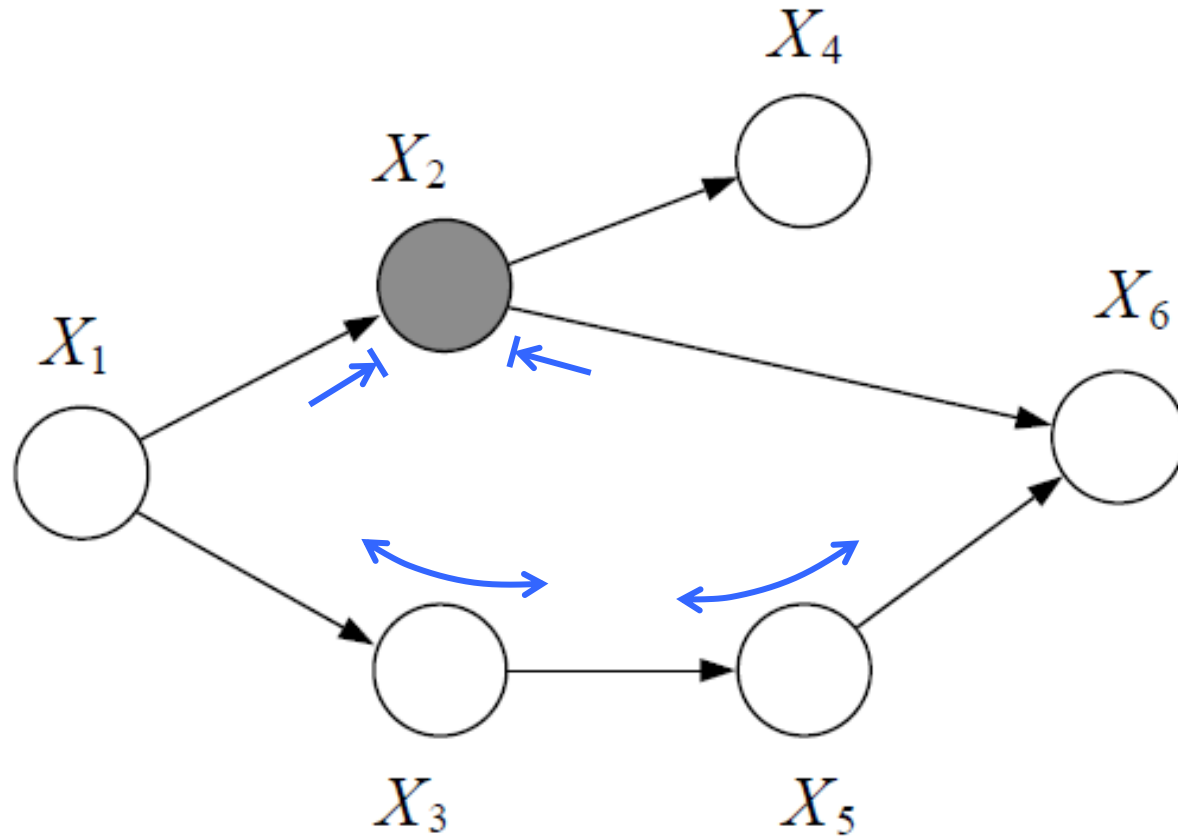
Is X_1 independent from X_6 given X_2 ?



Poll 3

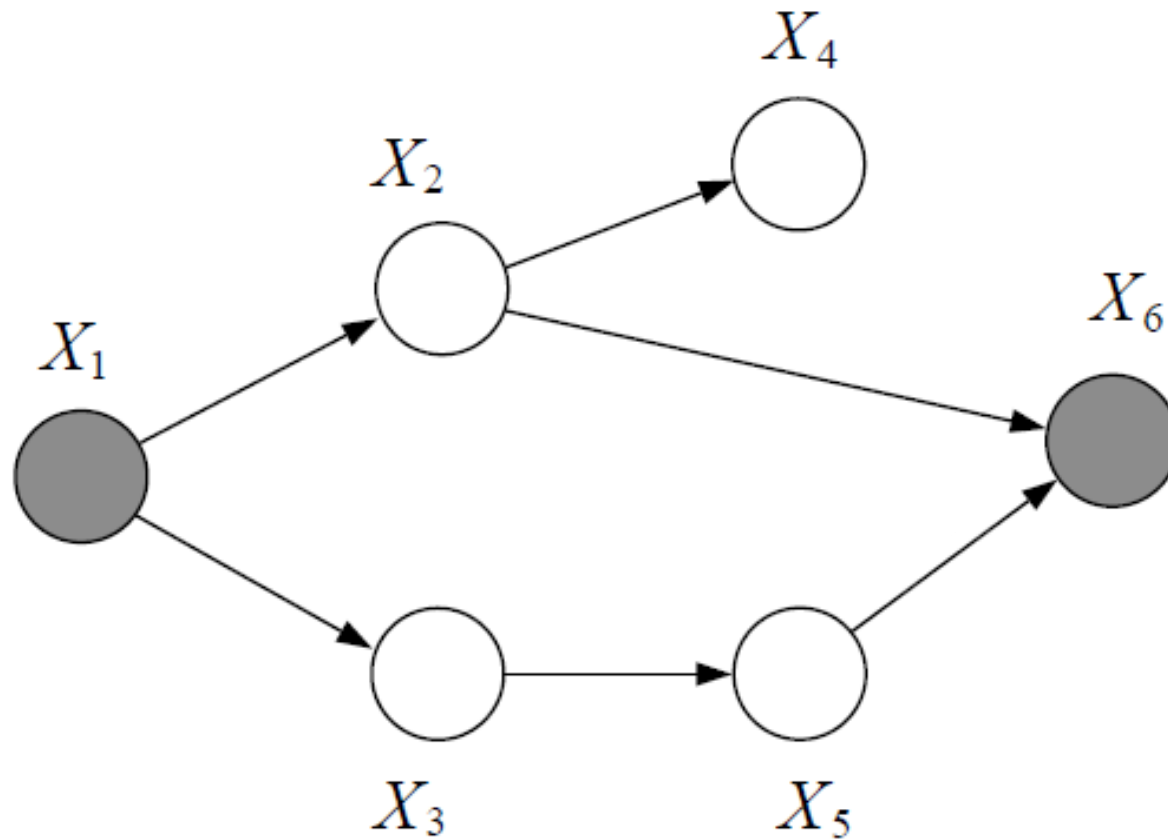
Is X_1 independent from X_6 given X_2 ?

No, the Bayes ball can travel through X_3 and X_5 .



Poll 4

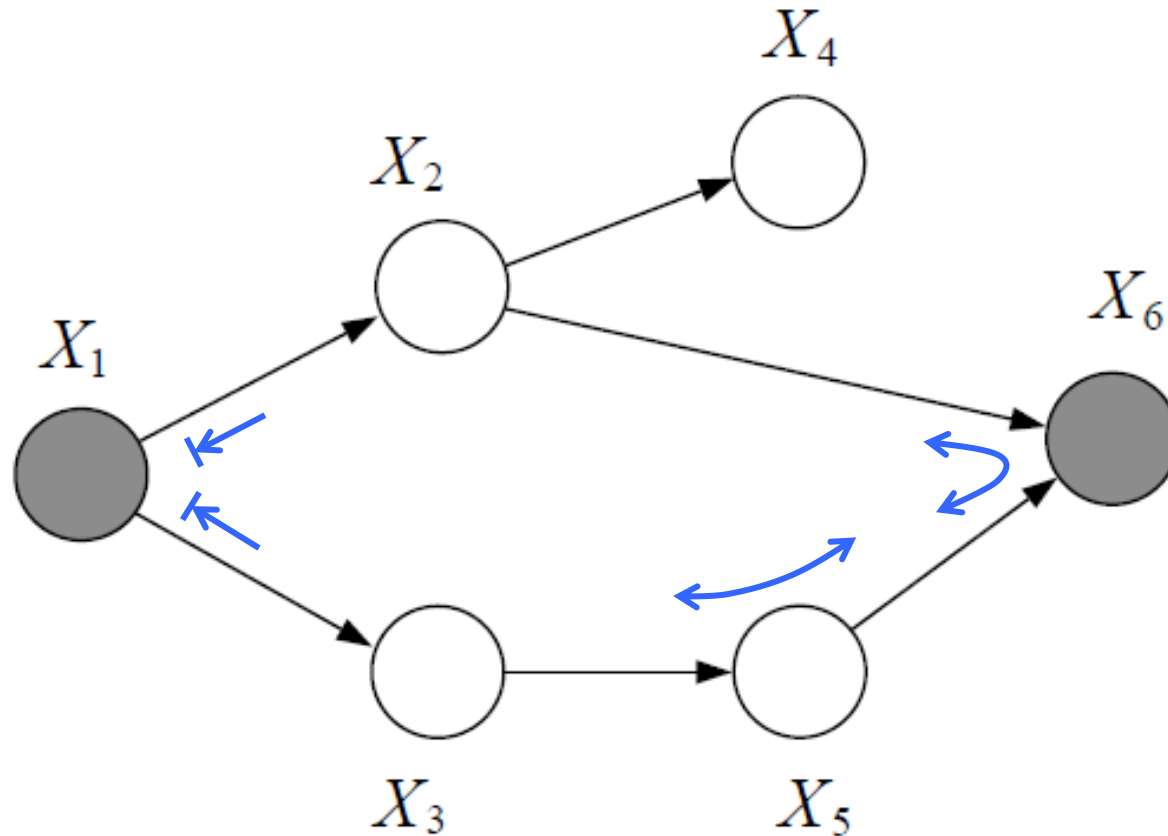
Is X_2 independent from X_3 given X_1 and X_6 ?



Poll 4

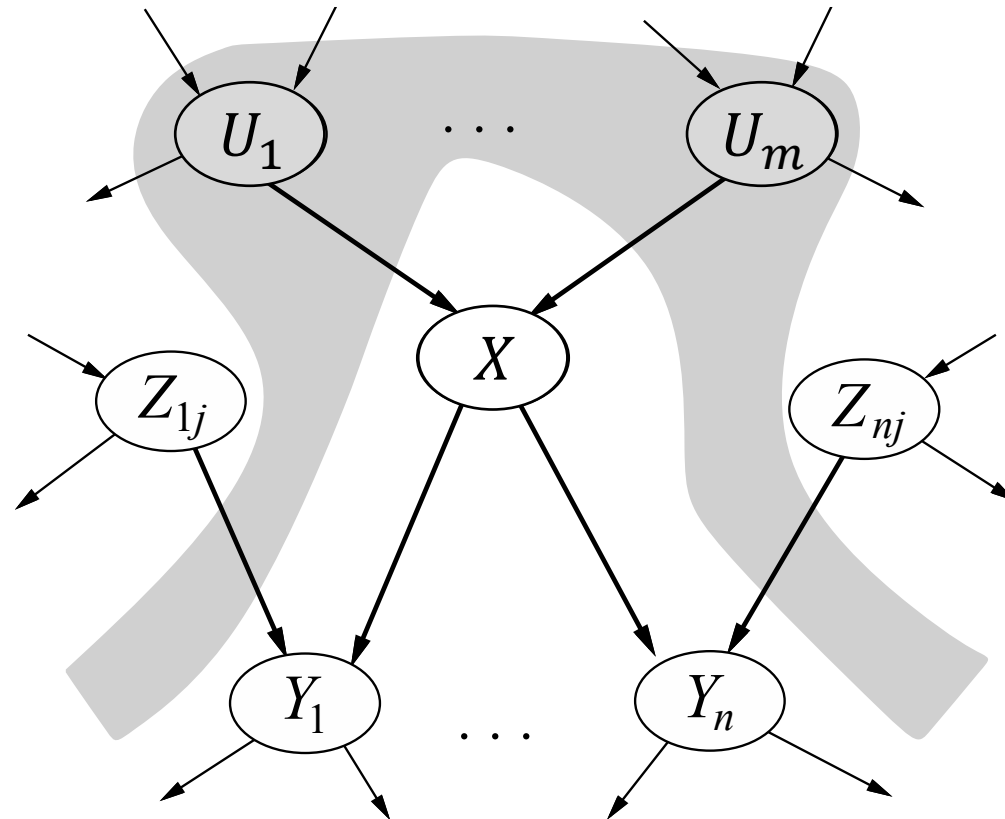
Is X_2 independent from X_3 given X_1 and X_6 ?

No, the Bayes ball can travel through X_5 and X_6 .



Conditional Independence Semantics

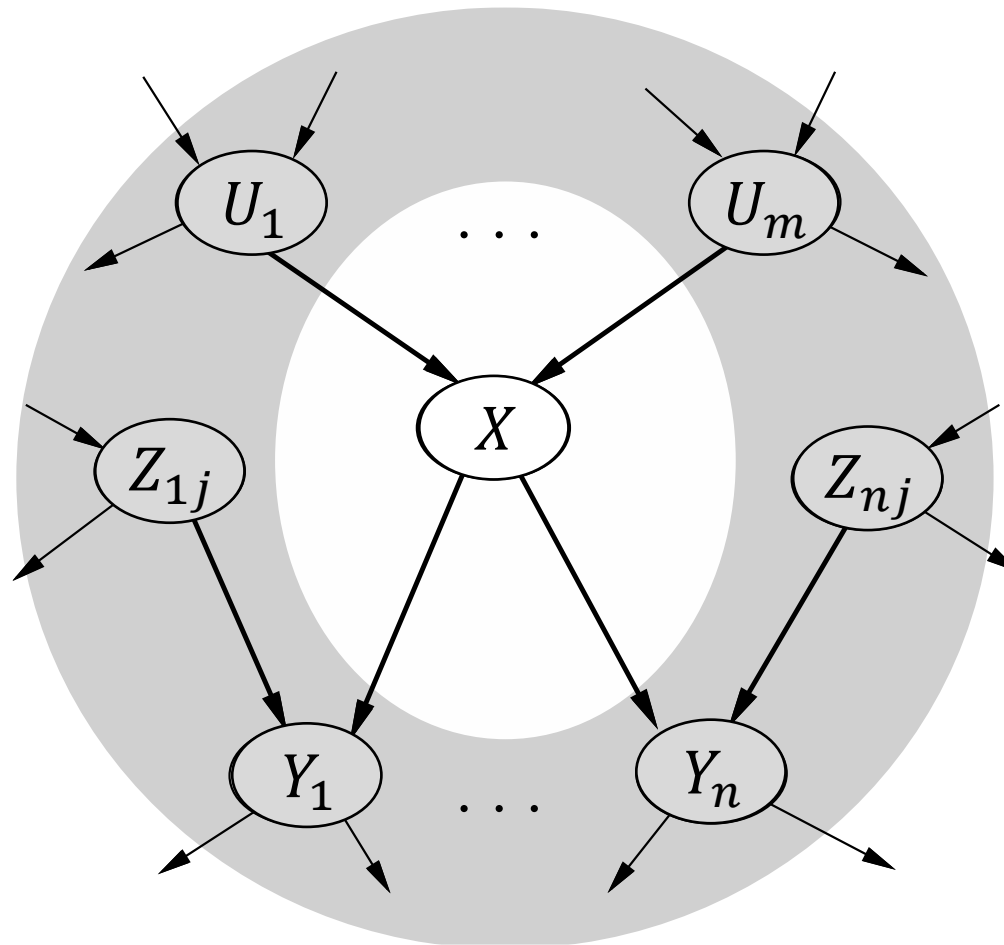
Every variable is conditionally independent of its non-descendants given its parents



Markov blanket

A variable's Markov blanket consists of parents, children, children's other parents

Every variable is conditionally independent of all other variables given its Markov blanket



Answer Any Query from Joint Distribution

Joint distributions are the best!

Problems with joints

- We aren't given the joint table
 - Usually some set of conditional probability tables
- Huge
 - n variables with d values
 - d^n entries

Joint

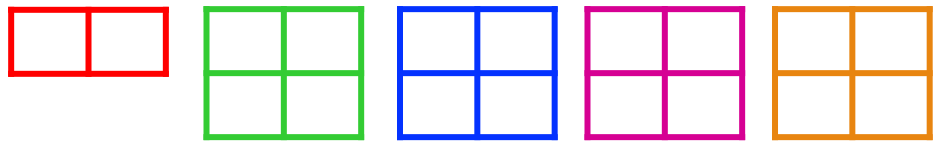
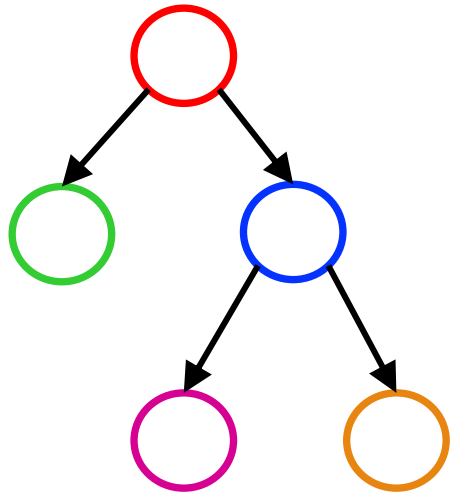


Query

$$P(a | e)$$

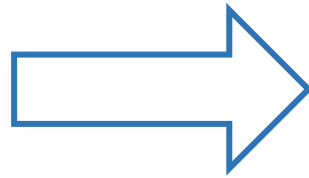
Answer Any Query from Bayes Net

Bayes Net



$P(A)$ $P(B|A)$ $P(C|A)$ $P(D|C)$ $P(E|C)$

Joint



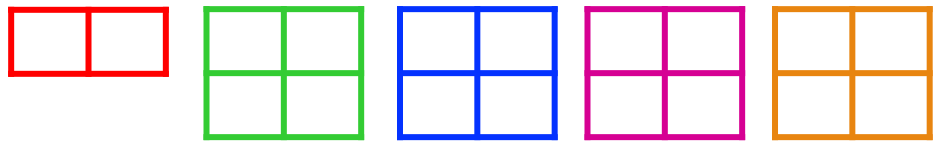
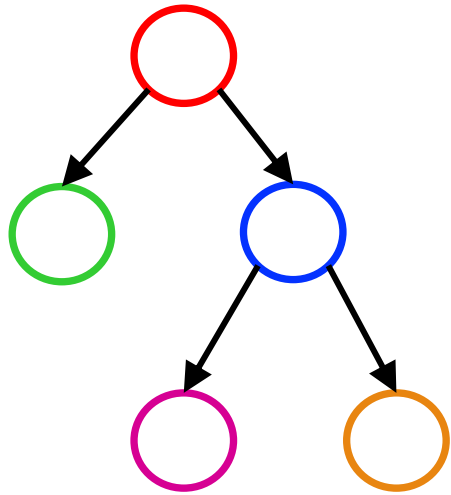


Query

$P(a | e)$

Next: Answer Any Query from Bayes Net

Bayes Net



$P(A)$ $P(B|A)$ $P(C|A)$ $P(D|C)$ $P(E|C)$

Query



$P(a | e)$