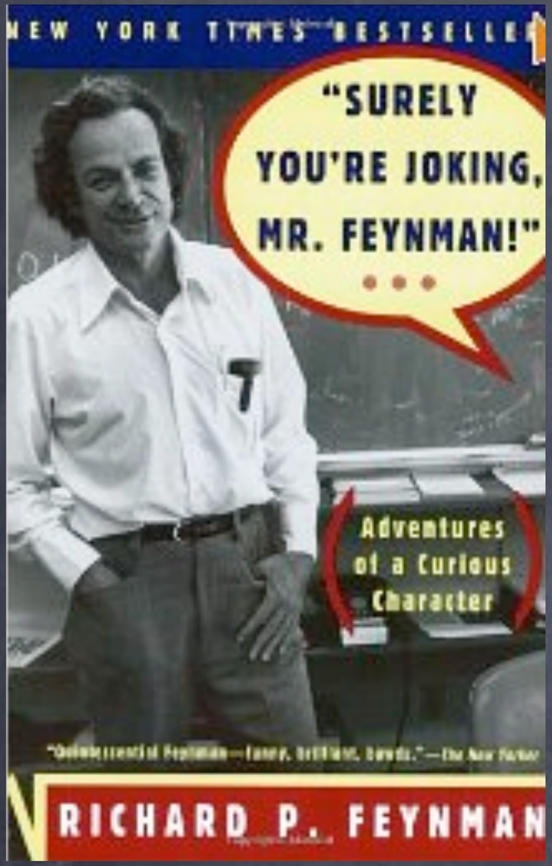
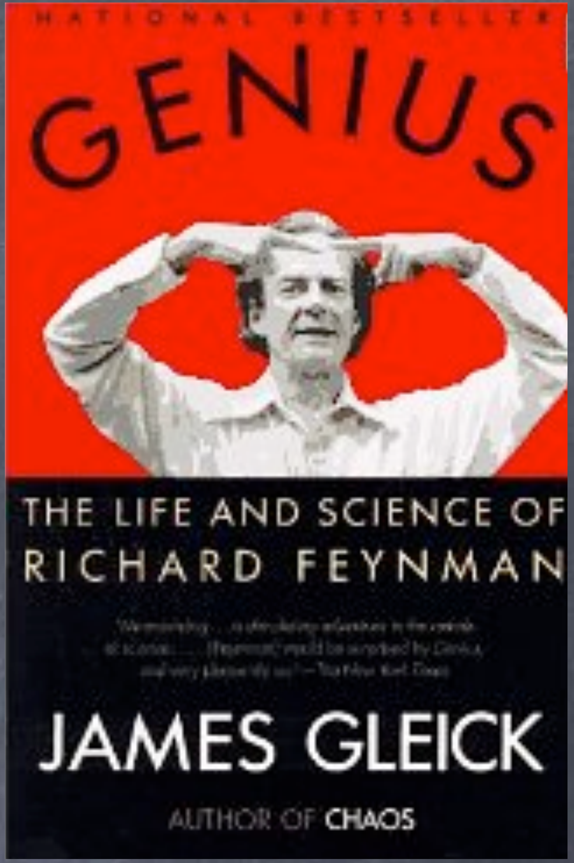
CSC242: Introduction to Artificial Intelligence

Lecture 3.3

Please put away all electronic devices





Richard Feynman

1918–1988

CSC 161-171-172 TAs & Workshop Leaders

https://goo.gl/forms/pZbL3WDyn6YOJwAP2

Representing Uncertainty

- Probability: Sample space, probabilities, events
- Random variables, domains
- Language of probability statements
- Probability distributions, joint distributions, full joint distribution

Probabilistic Inference

- Compute what “follows from” our (uncertain) knowledge
- Make implicit knowledge explicit

Probabilistic Inference

- Computing posterior probabilities for statements given prior probabilities and observed evidence
- Given priors and evidence, compute probabilities given evidence (posteriors)

Probabilistic Inference (Single Variable)

$$\mathbf{P}(X \mid \mathbf{e}) = \alpha\,\mathbf{P}(X, \mathbf{e}) = \alpha \sum_{\mathbf{y}} \mathbf{P}(X, \mathbf{e}, \mathbf{y})$$

Query variable $X : Domain(X) = \{x_1, \ldots, x_m\}$

Evidence variables $\mathbf{E} : \{E_1, \ldots, E_k\}$

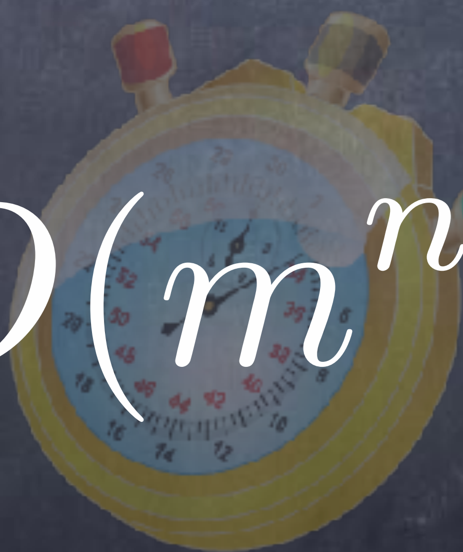
Observations $\mathbf{e} : \{e_1, \ldots, e_k\}$ s.t. $E_i = e_i$

Unobserved variables $\mathbf{Y} : \{Y_1, \ldots, Y_l\}$

$Domain(Y_i) = \{y_{i,1}, \ldots, y_{i,n_i}\}$

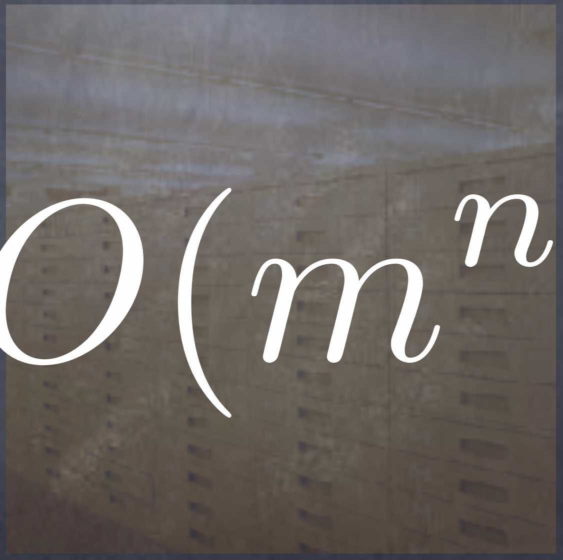
Probabilistic Inference (Single Variable)

$$\mathbf{P}(X \mid \mathbf{e}) = \alpha\, \mathbf{P}(X, \mathbf{e}) = \alpha \sum_{\mathbf{y}} \mathbf{P}(X, \mathbf{e}, \mathbf{y})$$



$O(m^n)$

Time Complexity



$O(m^n)$

Space Complexity

Bayes' Rule

$$P(b \mid a) = \frac{P(a \mid b)P(b)}{P(a)}$$

Thomas Bayes

(c. 1702 – 1761)

Bayesian Diagnosis

$$P(disease \mid symptom) = \frac{P(symptom \mid disease)P(disease)}{P(symptom)}$$

$$toothache \qquad\qquad catch$$

$$(Toothache = true) \qquad\qquad (Catch = true)$$

$$\mathbf{P}(Cavity \mid toothache \wedge catch)$$

$$toothache \qquad\qquad catch$$

$$(Toothache = true) \qquad (Catch = true)$$

$$\mathbf{P}(Cavity \mid toothache \wedge catch)$$

Combining Evidence

$$\mathbf{P}(Cavity \mid toothache \wedge catch)$$

Combining Evidence

$$\mathbf{P}(Cavity \mid toothache \wedge catch)$$
$$= \alpha \, \langle 0.180, 0.016 \rangle \approx \langle 0.871, 0.129 \rangle$$

	$toothache$		$\neg toothache$	
	$catch$	$\neg catch$	$catch$	$\neg catch$
$cavity$	0.108	0.012	0.072	0.008
$\neg cavity$	0.016	0.064	0.144	0.576

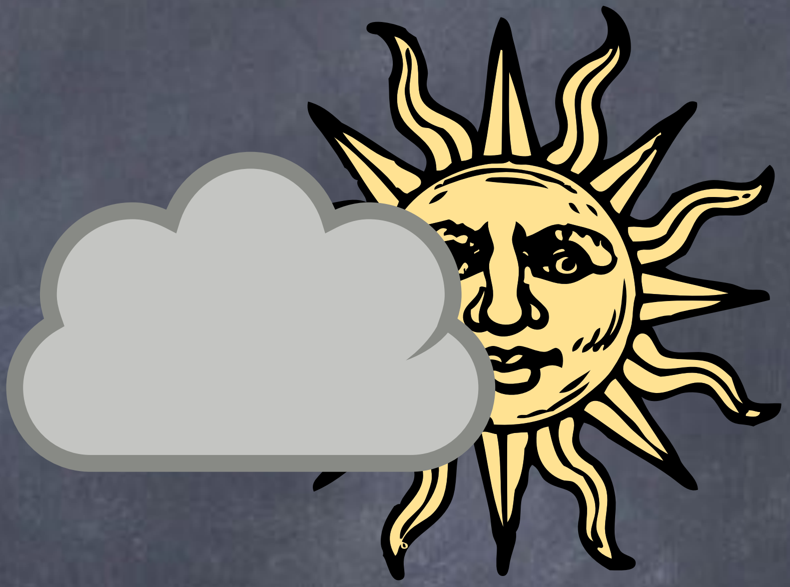
Combining Evidence

$$\mathbf{P}(Cavity \mid toothache \wedge catch)$$
$$= \alpha\, \mathbf{P}(toothache \wedge catch \mid Cavity)\, \mathbf{P}(Cavity)$$

Combining Evidence

- In general, if there are n evidence variables, then there are $O(2^n)$ possible combinations of observed values for which we would need to know the conditional probabilities

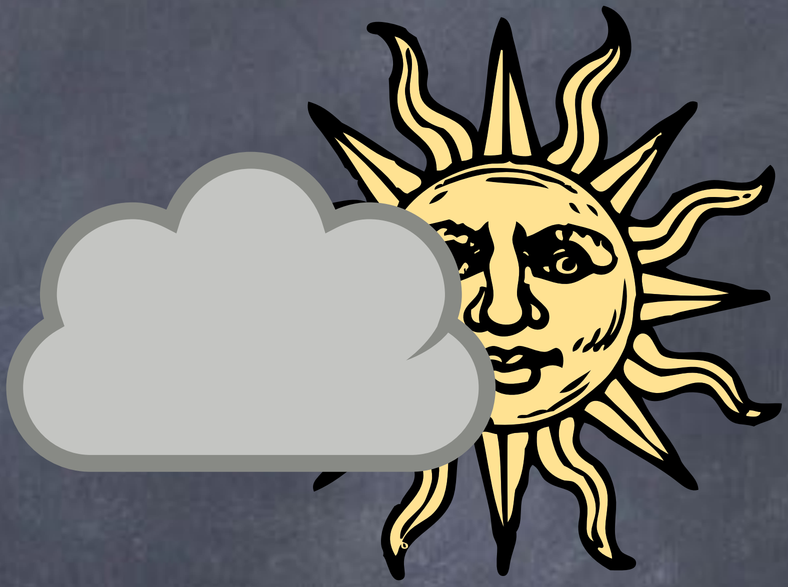
Cavity



Weather

Independence

$Cavity$



$Weather$

Independence

$$P(cavity \mid sunny) = P(cavity)$$

$$P(rain \mid cavity) = P(rain)$$

$$\mathbf{P}(Cavity \mid Weather) = \mathbf{P}(Cavity)$$

$$\mathbf{P}(Weather \mid Cavity) = \mathbf{P}(Weather)$$

Independence

$$P(a \mid b) = P(a)$$
$$P(b \mid a) = P(b)$$
$$P(a \wedge b) = P(a)P(b)$$

Independence

$$\mathbf{P}(X \mid Y) = \mathbf{P}(X)$$
$$\mathbf{P}(Y \mid X) = \mathbf{P}(Y)$$
$$\mathbf{P}(X, Y) = \mathbf{P}(X)\mathbf{P}(Y)$$

Independence

$$\mathbf{P}(X) = \langle p_{x_1}, p_{x_2}, \ldots, p_{x_n} \rangle$$

$$\mathbf{P}(Y) = \langle p_{y_1}, p_{y_2}, \ldots, p_{y_m} \rangle$$

$\mathbf{P}(X, Y)$	x_1	x_2	...	x_n
y_1	$p_{x_1}p_{y_1}$	$p_{x_2}p_{y_1}$	...	$p_{x_n}p_{y_1}$
y_2	$p_{x_1}p_{y_2}$	$p_{x_2}p_{y_2}$	...	$p_{x_n}p_{y_2}$
...	...	...	...	...
y_m	$p_{x_1}p_{y_m}$	$p_{x_2}p_{y_m}$	...	$p_{x_n}p_{y_m}$

Independence

$$\mathbf{P}(X \mid Y) = \mathbf{P}(X)$$

$$\mathbf{P}(Y \mid X) = \mathbf{P}(Y)$$

$$\mathbf{P}(X, Y) = \mathbf{P}(X)\mathbf{P}(Y)$$

Can compute $n \times m$ probabilities

from $n + m$ probabilities

(if random variables are independent)

$$toothache \qquad\qquad catch$$

$$(Toothache = true) \qquad (Catch = true)$$

$$\mathbf{P}(Cavity \mid toothache \wedge catch)$$

$$toothache \qquad\qquad catch$$

$$(Toothache = true) \qquad (Catch = true)$$

$$\mathbf{P}(Cavity \mid toothache \wedge catch)$$
$$= \alpha\, \mathbf{P}(toothache \wedge catch \mid Cavity)\, \mathbf{P}(Cavity)$$

$$toothache \qquad\qquad catch$$

$$(Toothache = true) \qquad (Catch = true)$$

Independent?

$$\mathbf{P}(Cavity \mid toothache \wedge catch)$$

$$= \alpha\, \mathbf{P}(toothache \wedge catch \mid Cavity)\, \mathbf{P}(Cavity)$$

Conditional Independence

- Both $toothache$ and $catch$ are caused by a cavity, but neither has a direct effect on the other
- The variables are independent <u>given the presence or absence of a cavity</u>

Conditional Independence

$$\mathbf{P}(X \mid Y, Z) = \mathbf{P}(X \mid Z)$$
$$\mathbf{P}(Y \mid X, Z) = \mathbf{P}(Y \mid Z)$$
$$\mathbf{P}(X, Y \mid Z) = \mathbf{P}(X \mid Z)\mathbf{P}(Y \mid Z)$$

Conditional Independence

$$\mathbf{P}(Toothache, Catch \mid Cavity) = \mathbf{P}(Toothache \mid Cavity)\mathbf{P}(Catch \mid Cavity)$$

Combining Evidence

$$\mathbf{P}(Cavity \mid toothache \wedge catch)$$

$$= \alpha\, \mathbf{P}(toothache \wedge catch \mid Cavity)\, \mathbf{P}(Cavity)$$

$$= \alpha\, \mathbf{P}(toothache|Cavity)\, \mathbf{P}(catch|Cavity)\, \mathbf{P}(Cavity)$$

Combining Evidence

- For n symptoms (e.g., $Toothache$, $Catch$) that are all conditionally independent given a disease (e.g., $Cavity$), we need $O(n)$ probabilities rather than $O(2^n)$
  - Representation scales to larger problems
  - Conditional probabilities more likely to be available than absolute independence assumptions

Probabilistic Inference

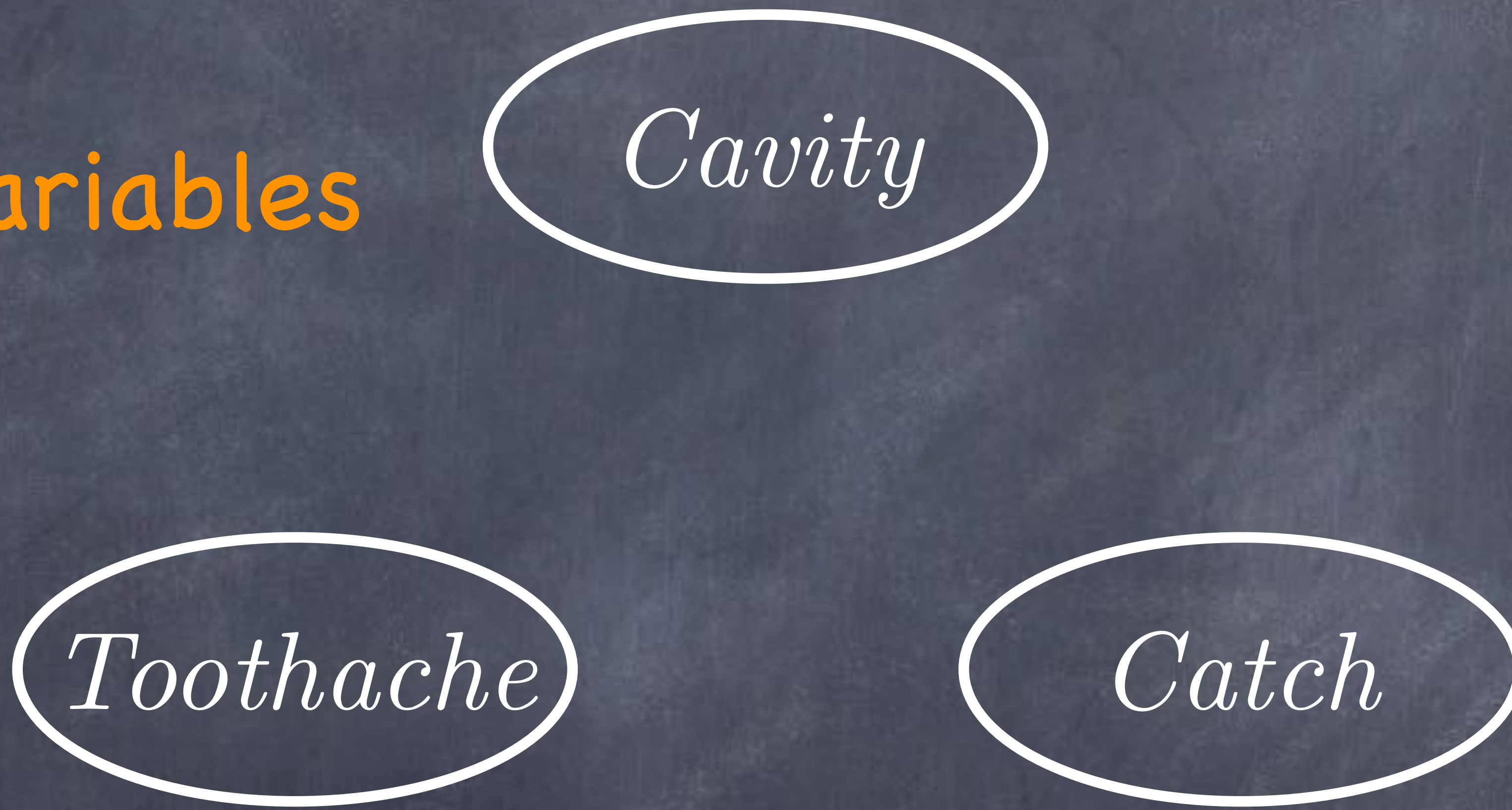
- Full joint distribution: intractable as problem grows
- Independence assumptions reduce number of probabilities required to represent full joint distribution
- Next: Develop a data structure that represents the (in)dependencies among random variables and can be used to compute probabilities full joint distribution

Bayesian Networks

Cavity *Toothache* *Catch*

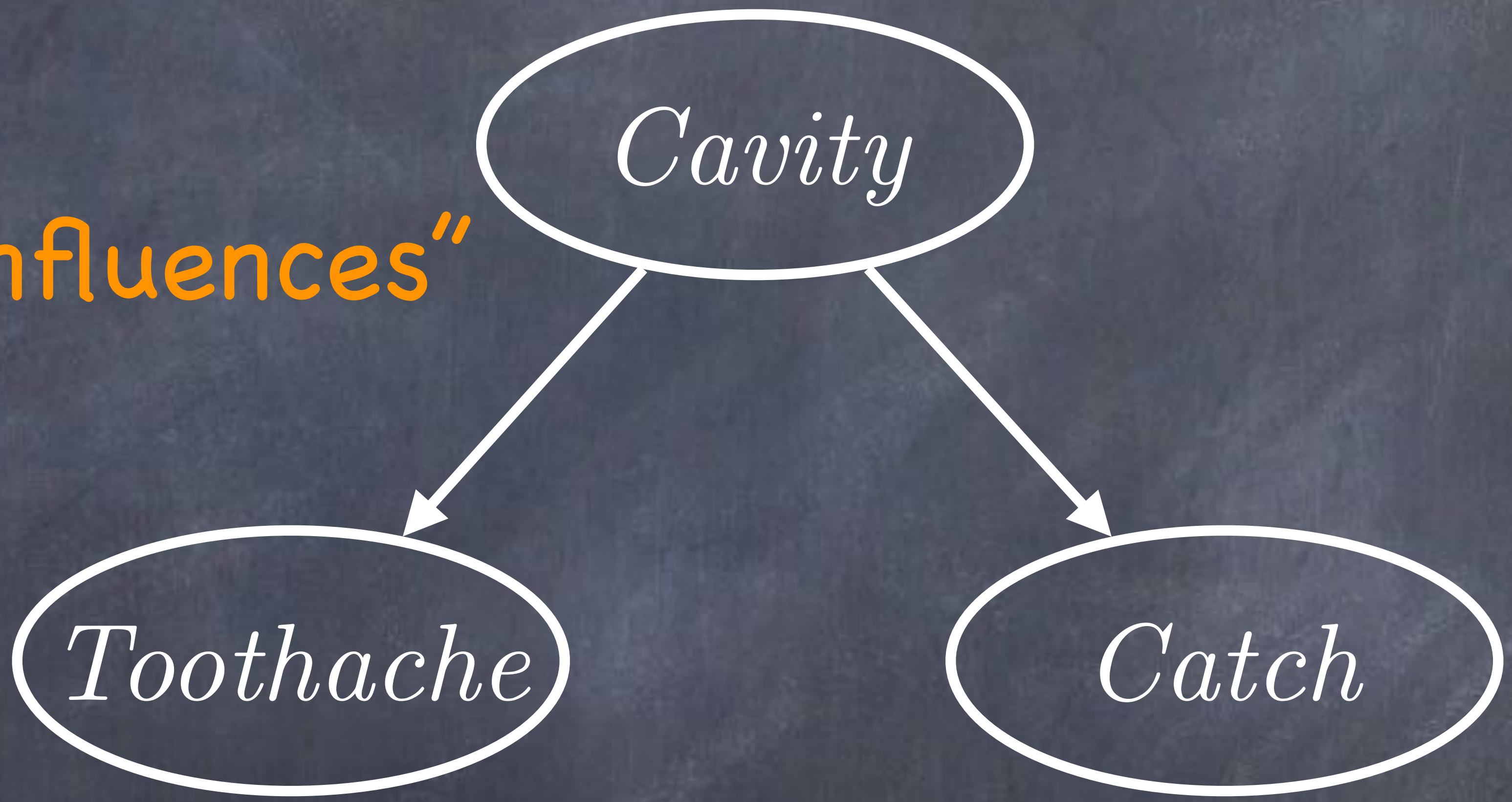
Bayesian Networks

Random Variables

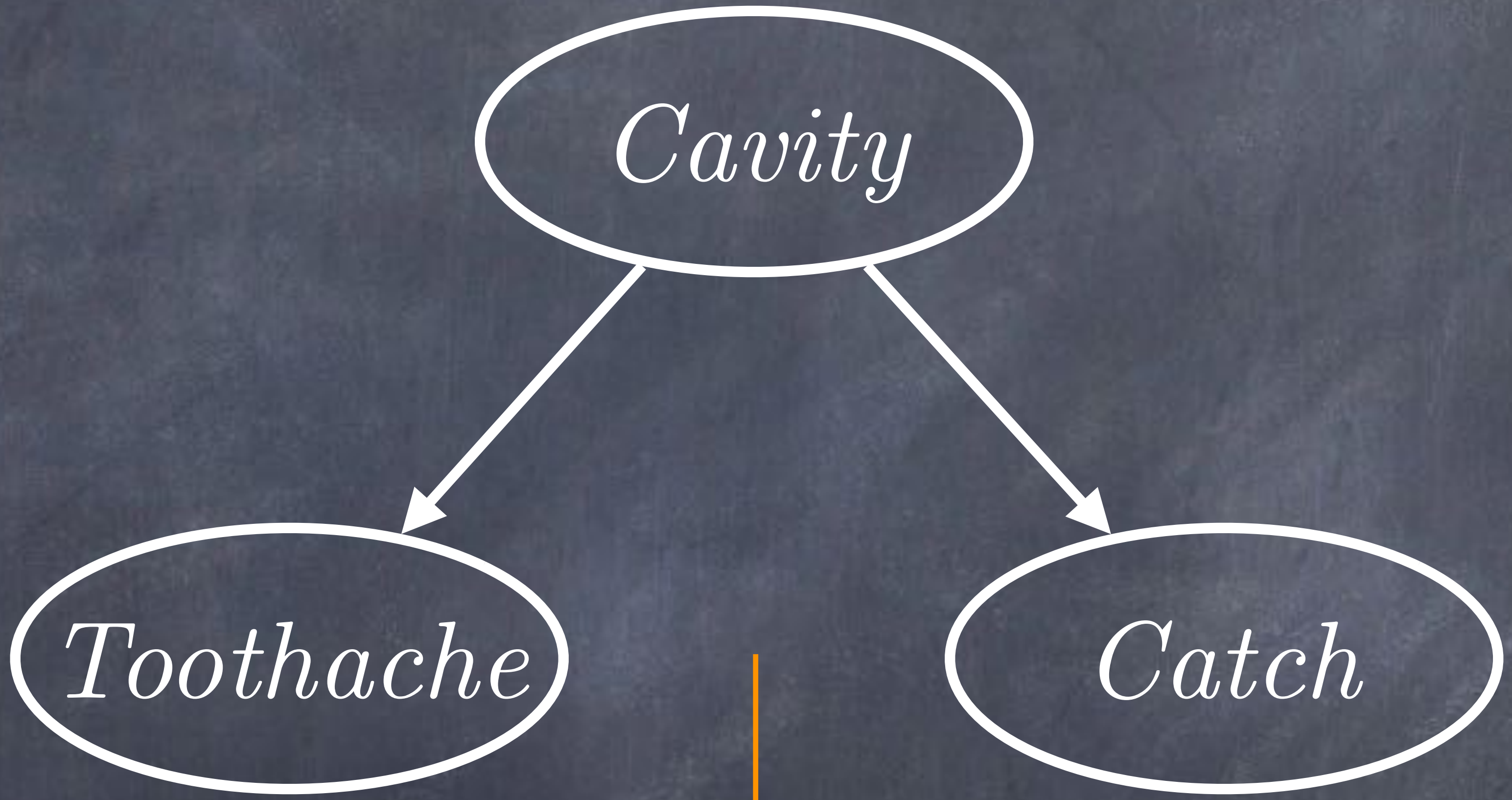


Bayesian Networks

"directly influences"

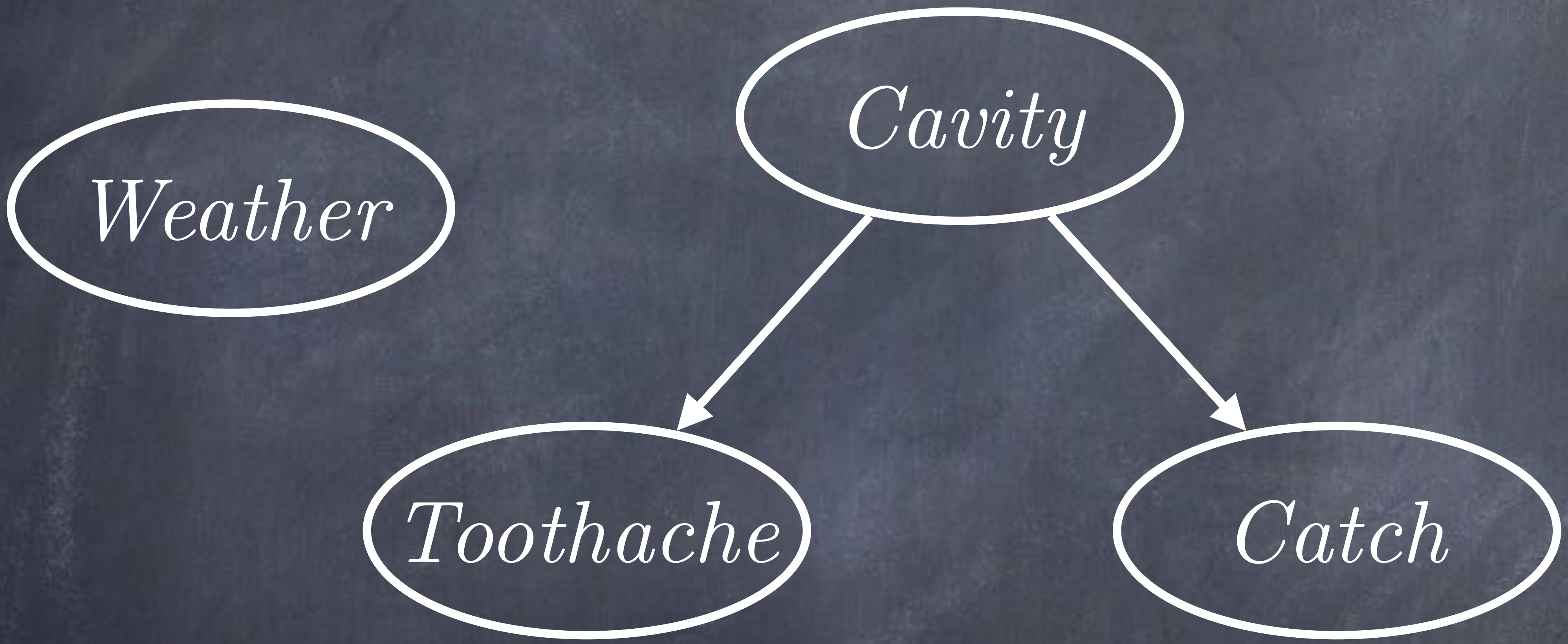


Bayesian Networks

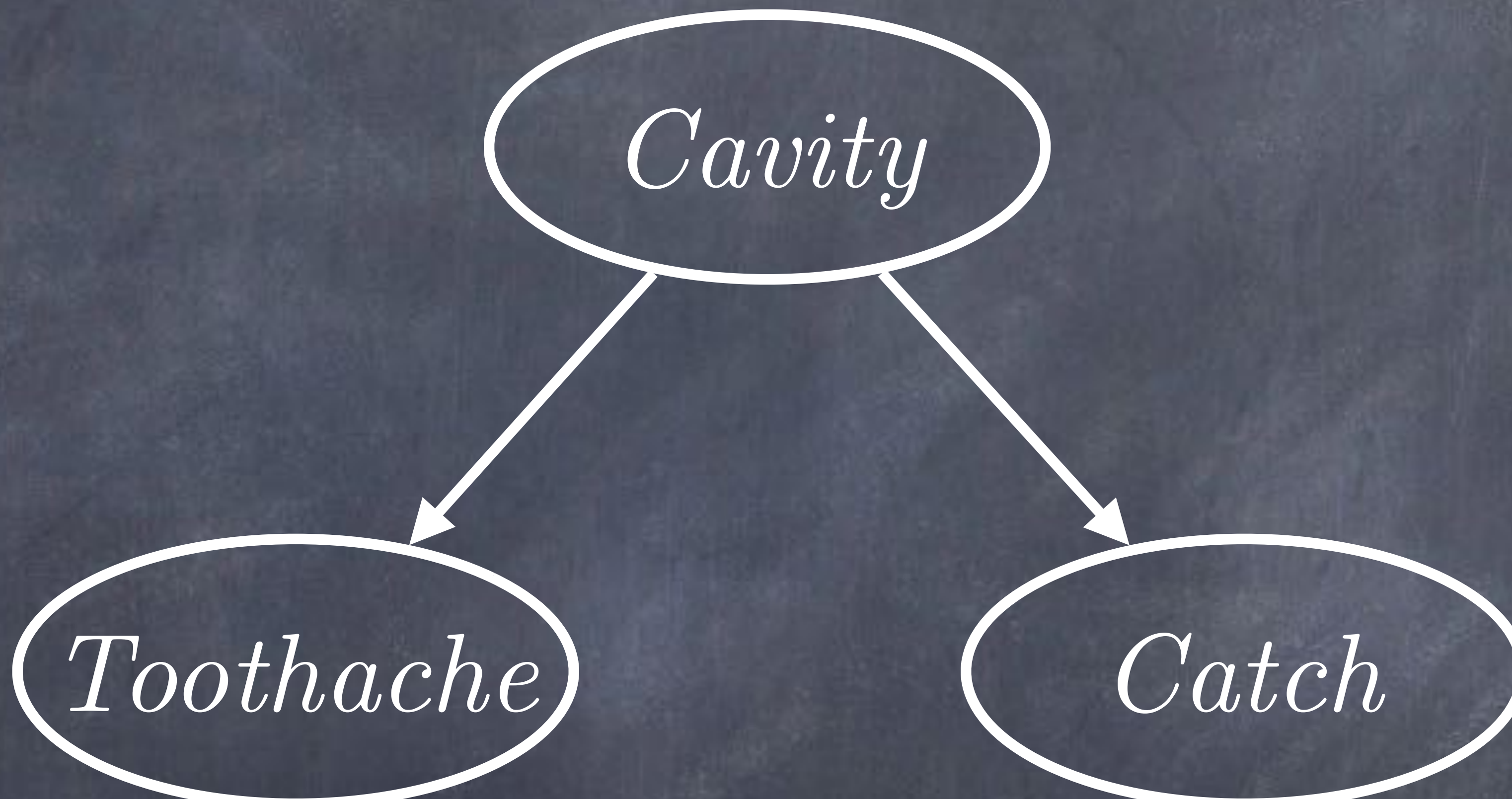


no "direct influence"

Bayesian Networks



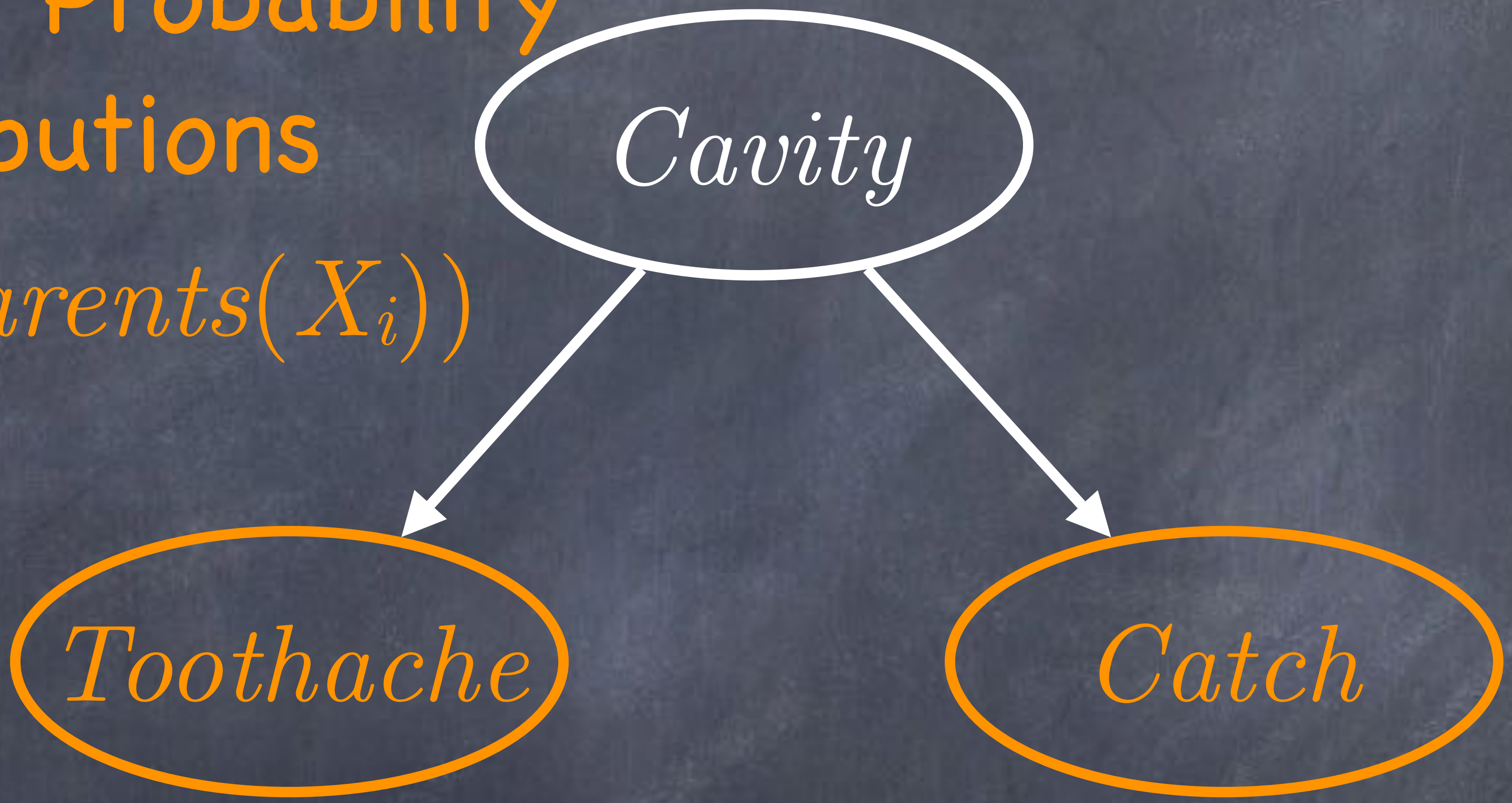
Bayesian Networks



Bayesian Networks

Conditional Probability Distributions

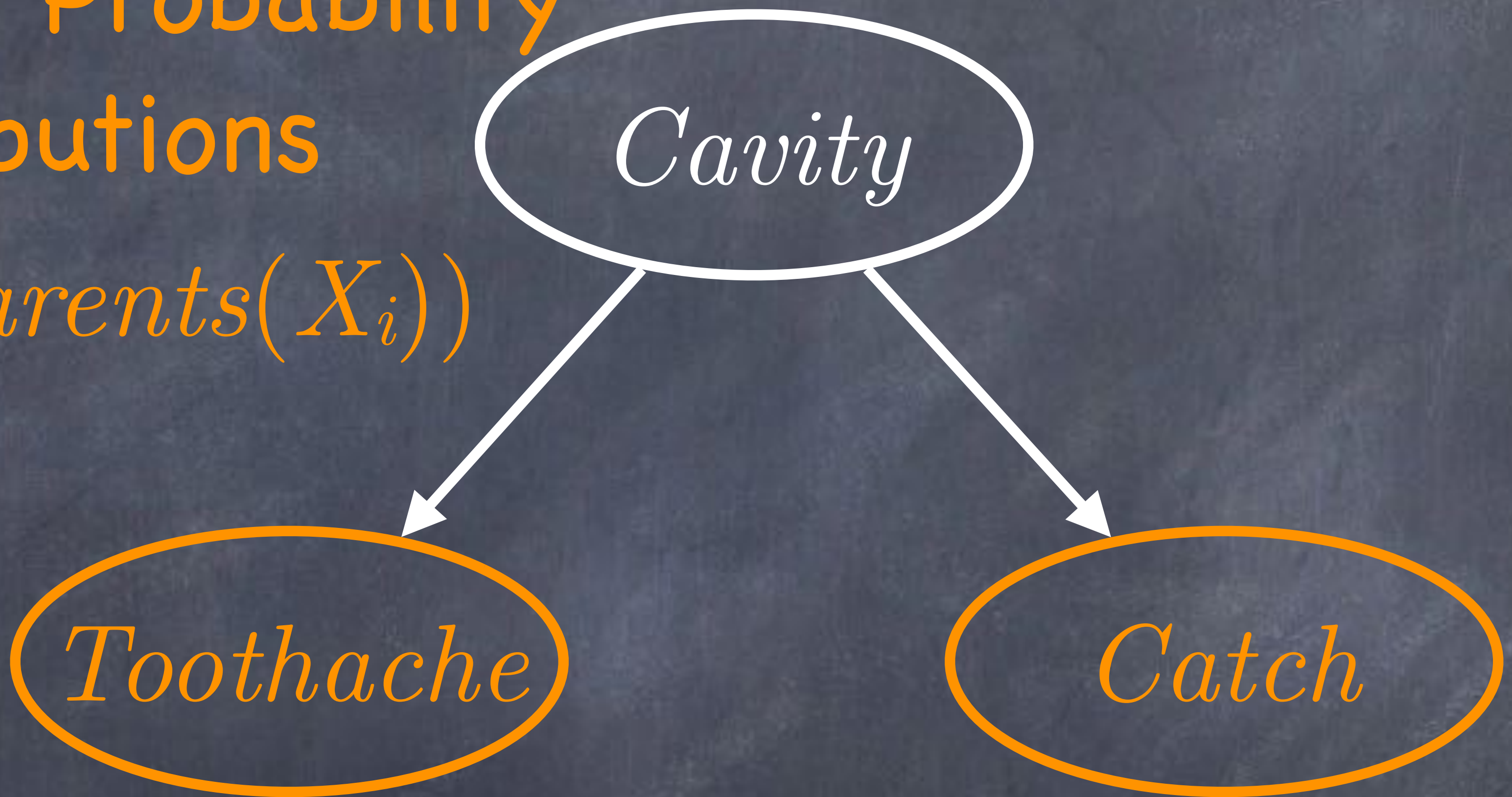
$$\mathbf{P}(X_i \mid Parents(X_i))$$



Bayesian Networks

Conditional Probability Distributions

$$\mathbf{P}(X_i \mid Parents(X_i))$$



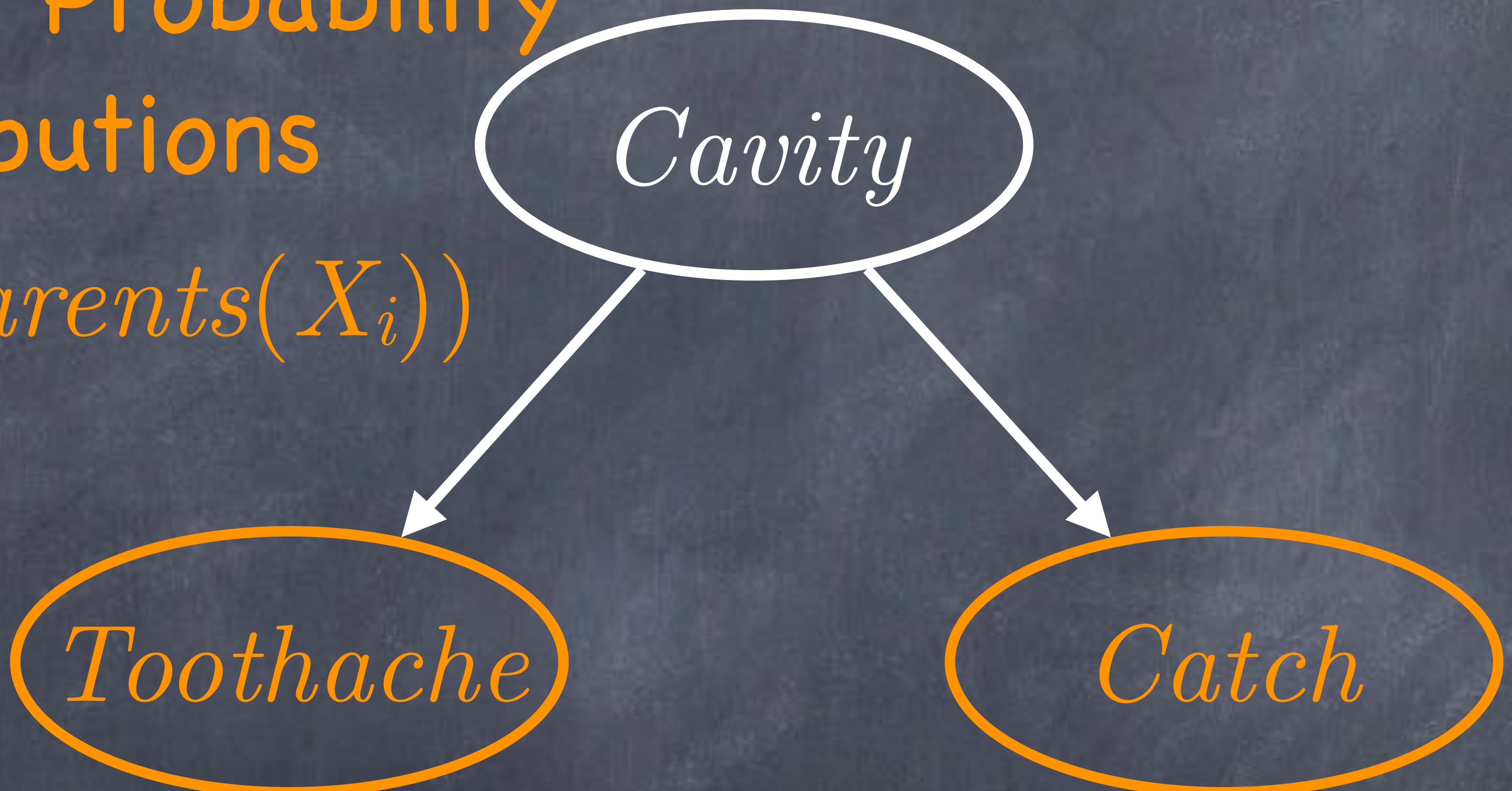
$$\mathbf{P}(Toothache|Cavity)$$

$Cavity$	$toothache$	$\neg toothache$
$cavity$		
$\neg cavity$		

Bayesian Networks

Conditional Probability Distributions

$$\mathbf{P}(X_i \mid Parents(X_i))$$



$$\mathbf{P}(Toothache|Cavity)$$

$Cavity$	$toothache$	$\neg toothache$
$cavity$		
$\neg cavity$		

$$\mathbf{P}(Catch|Cavity)$$

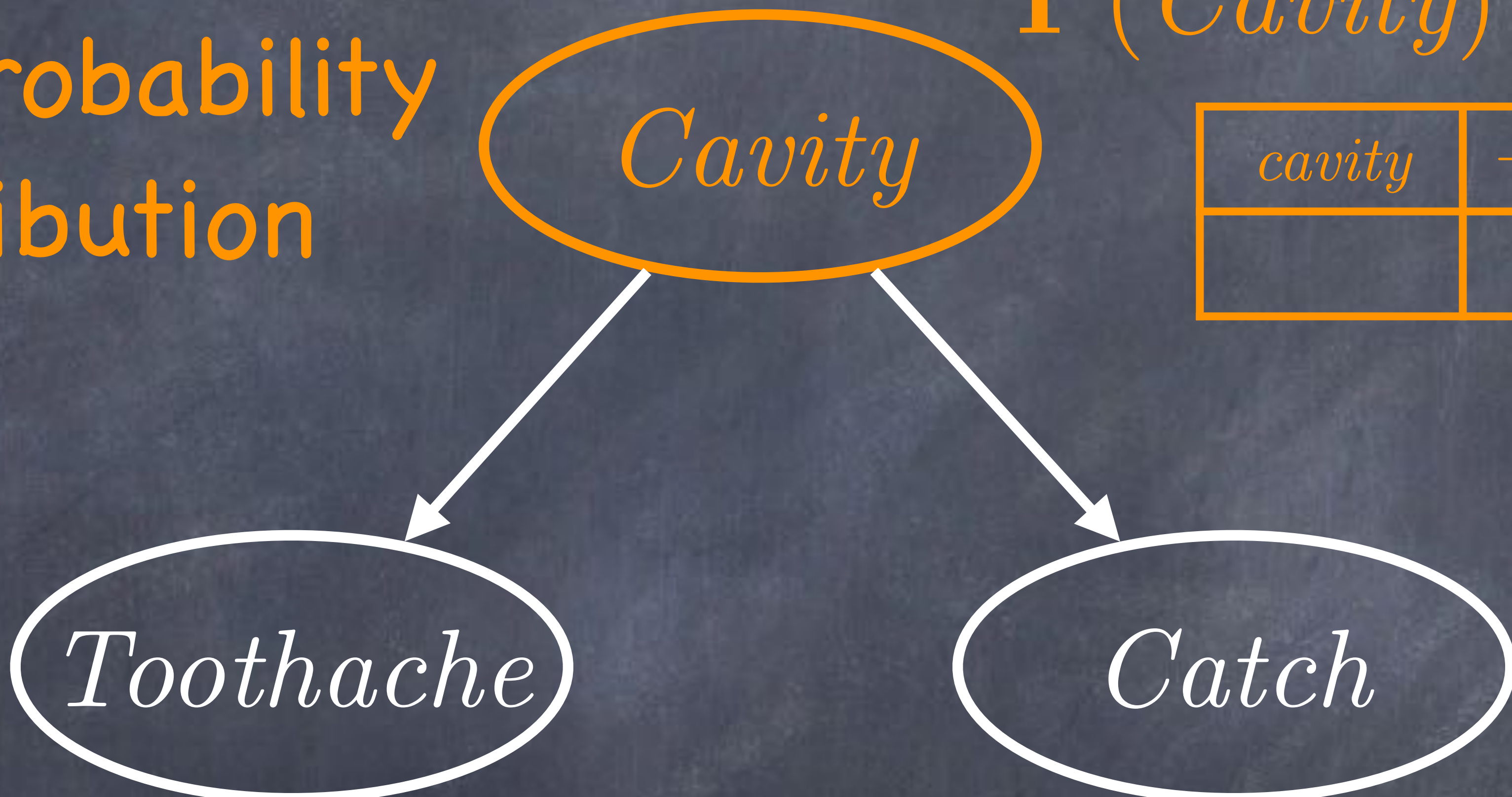
$Cavity$	$catch$	$\neg catch$
$cavity$		
$\neg cavity$		

Bayesian Networks

Prior Probability Distribution

$\mathbf{P}(Cavity)$

$cavity$	$\neg cavity$



$\mathbf{P}(Toothache|Cavity)$

$Cavity$	$toothache$	$\neg toothache$
$cavity$		
$\neg cavity$		

$\mathbf{P}(Catch|Cavity)$

$Cavity$	$catch$	$\neg catch$
$cavity$		
$\neg cavity$		

Bayesian Networks

- Nodes correspond to a random variables
- Link from X to Y iff X “directly influences” Y
  - No link: no “direct influence”
- Non-root nodes store the conditional distribution: $\mathbf{P}(X_i \mid Parents(X_i))$
- Root nodes store their prior $\mathbf{P}(X_i)$

Bayesian Networks How-To

- Select random variables required to model the domain
- Add links from causes to effects ("directly influences")
  - No cycles (see book)
- Add conditional probability distributions for $\mathbf{P}(X_i \mid Parents(X_i))$ and priors $\mathbf{P}(X_i)$


THE OFFICIAL
METALLICA
ILLUSTRATED CHRONICLE
SO WHAT!
THE GOOD, THE MAD AND THE UGLY
EDITED BY STEFFAN CHIRAZI

Probabilistic Inference

- Computing posterior probabilities for statements given observed evidence and probabilistic background knowledge

Probabilistic Inference (Single Variable)

Full Joint Prob. Dist.

$$\mathbf{P}(X \mid \mathbf{e}) = \alpha\, \mathbf{P}(X, \mathbf{e}) = \alpha \sum_{\mathbf{y}} \mathbf{P}(X, \mathbf{e}, \mathbf{y})$$

Query variable $X : Domain(X) = \{x_1, \ldots, x_m\}$

Evidence variables $\mathbf{E} : \{E_1, \ldots, E_k\}$

Observations $\mathbf{e} : \{e_1, \ldots, e_k\}$ s.t. $E_i = e_i$

Unobserved variables $\mathbf{Y} : \{Y_1, \ldots, Y_l\}$

$Domain(Y_i) = \{y_{i,1}, \ldots, y_{i,n_i}\}$

Semantics of Bayesian Networks

- Full joint distribution can be computed as the product of the separate conditional probabilities stored in the network

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

$$\mathrm{P}(\mathit{Toothache}, \mathit{Catch}, \mathit{Cavity})$$

Product Rule

$$= \mathrm{P}(\mathit{Toothache}, \mathit{Catch} | \mathit{Cavity})\ \mathrm{P}(\mathit{Cavity})$$

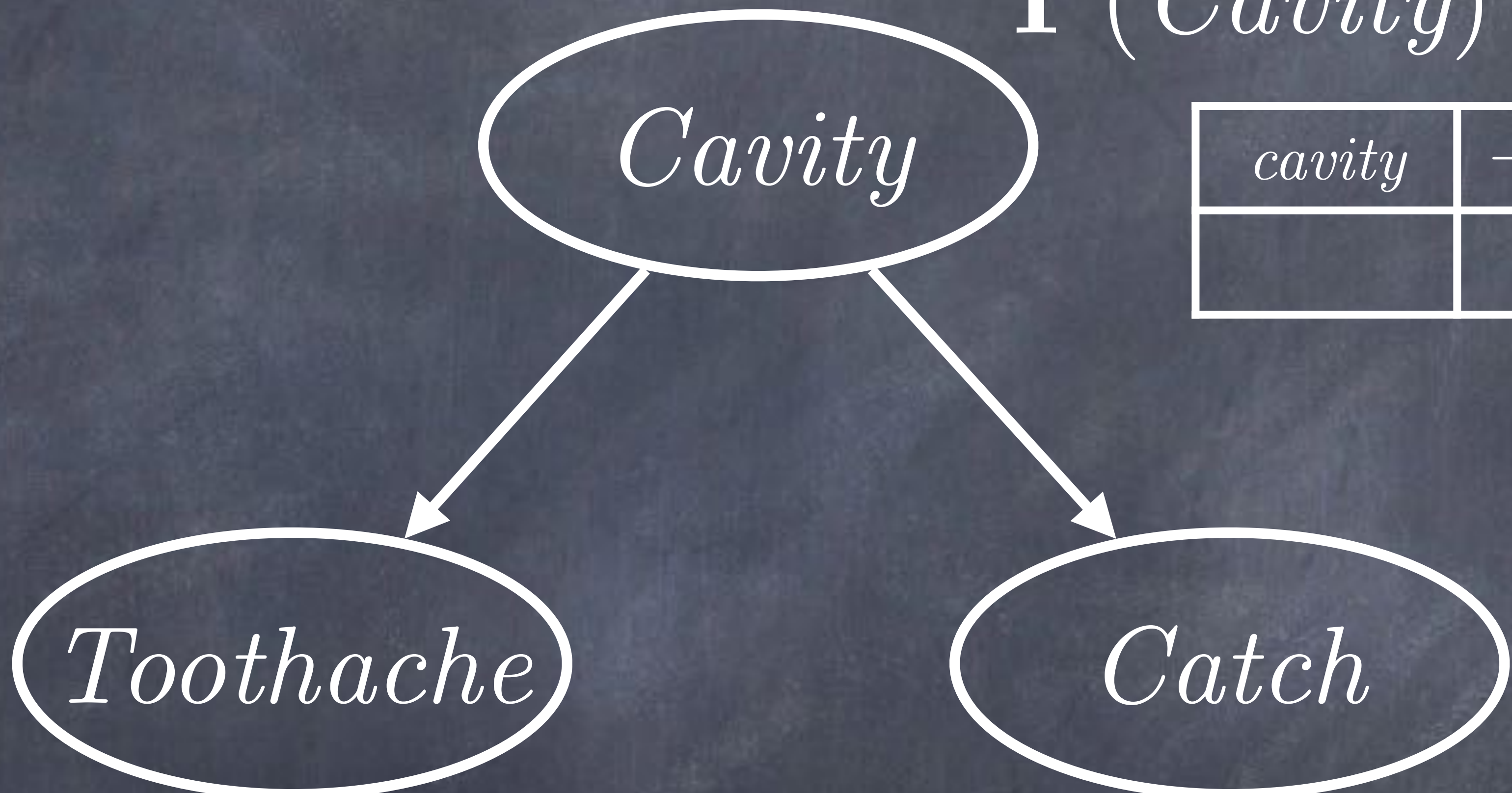
$$= \mathrm{P}(\mathit{Toothache} | \mathit{Cavity})\ \mathrm{P}(\mathit{Catch} | \mathit{Cavity})\ \mathrm{P}(\mathit{Cavity})$$

Conditional Independence

Bayesian Networks

$\mathbf{P}(Cavity)$

$cavity$	$\neg cavity$



$\mathbf{P}(Toothache|Cavity)$

$Cavity$	$toothache$	$\neg toothache$
$cavity$		
$\neg cavity$		

$\mathbf{P}(Catch|Cavity)$

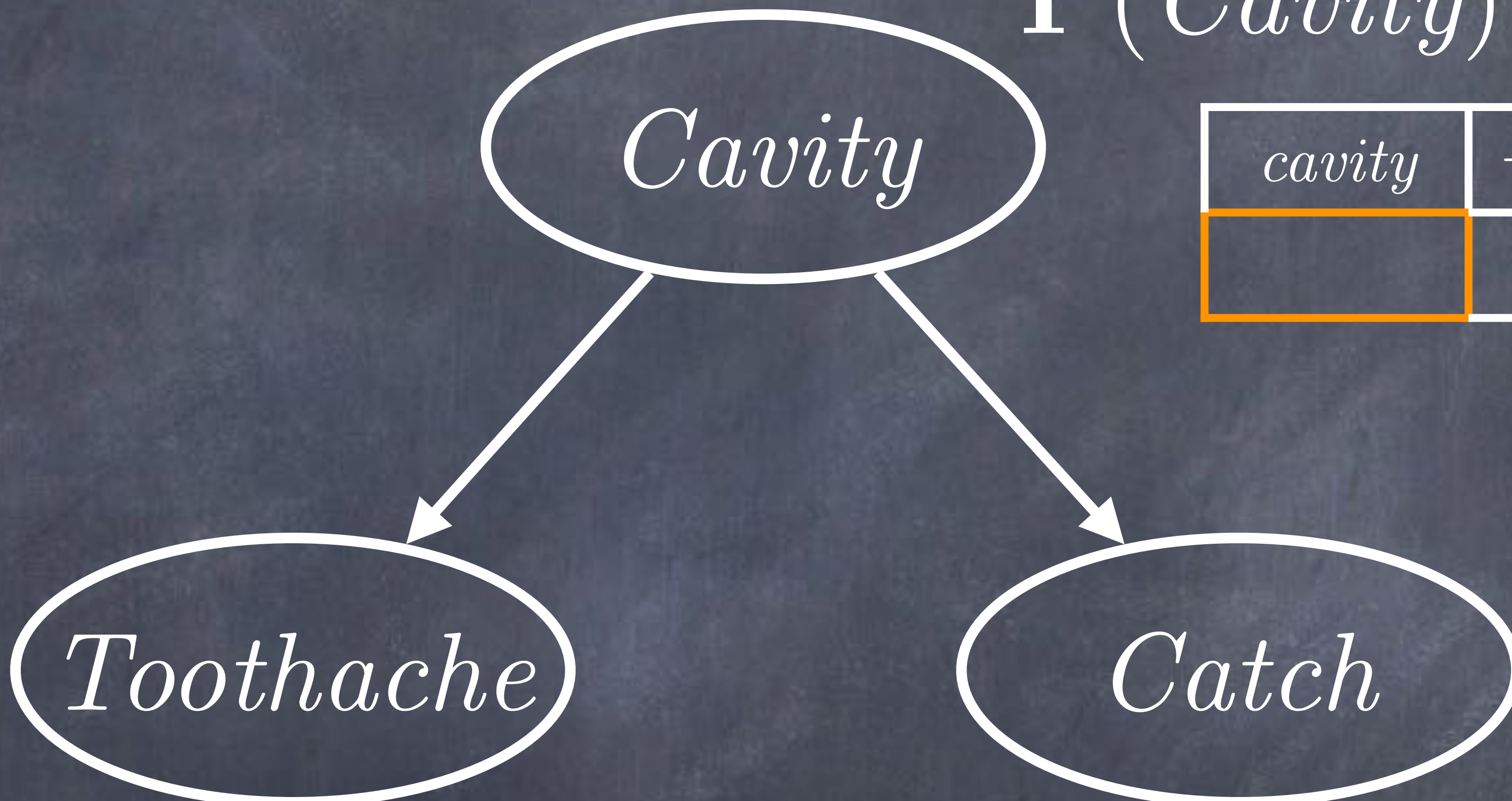
$Cavity$	$catch$	$\neg catch$
$cavity$		
$\neg cavity$		

Semantics of Bayesian Networks

- Full joint distribution can be computed as the product of the separate conditional probabilities stored in the network

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

$$P(toothache, cavity, catch) =$$
$$P(toothache|cavity)\,P(catch|cavity)\,P(cavity)$$



$\mathbf{P}(Cavity)$

$cavity$	$\neg cavity$

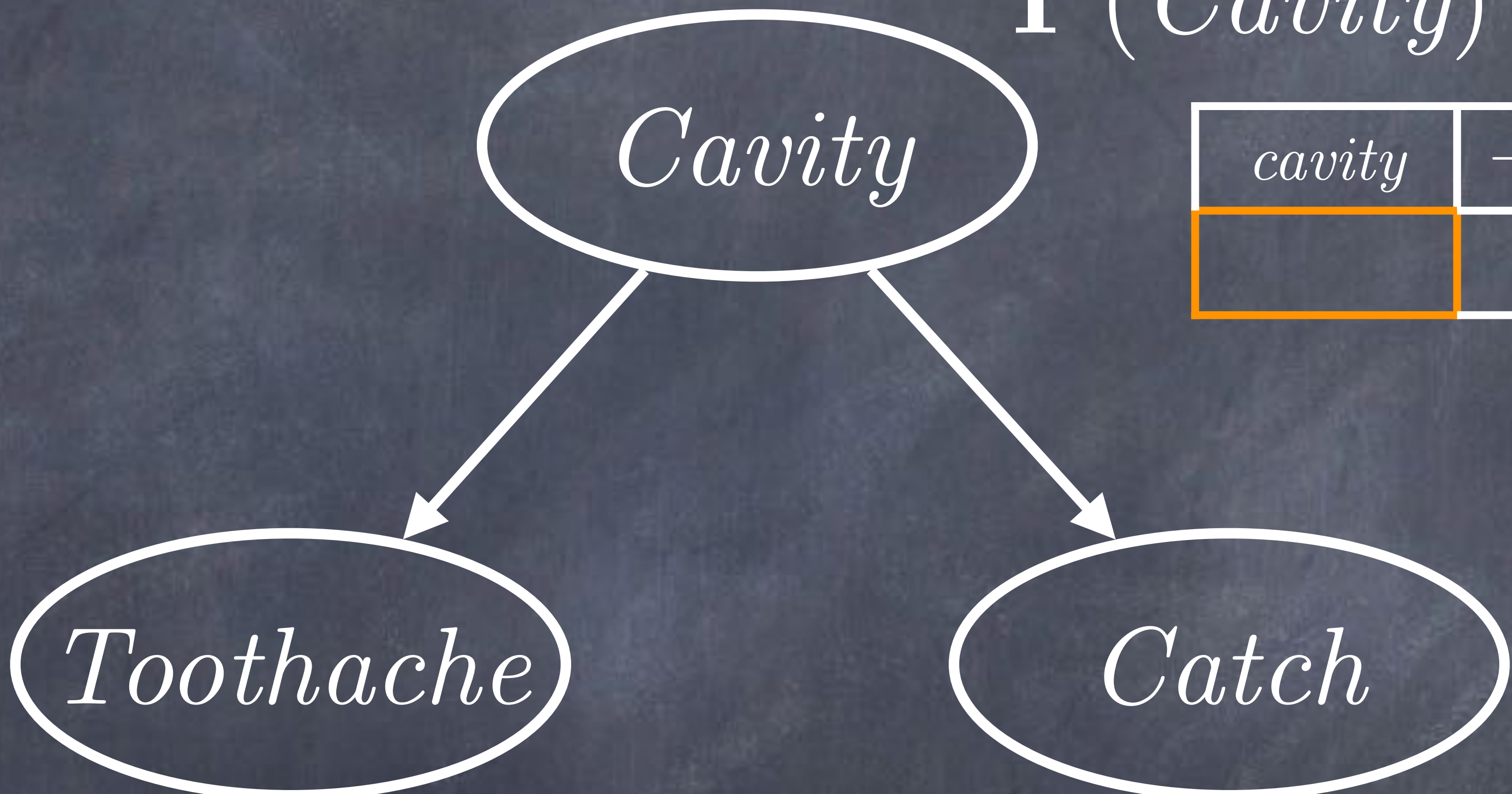
$\mathbf{P}(Toothache|Cavity)$

$Cavity$	$toothache$	$\neg toothache$
$cavity$		
$\neg cavity$		

$\mathbf{P}(Catch|Cavity)$

$Cavity$	$catch$	$\neg catch$
$cavity$		
$\neg cavity$		

$$P(\neg toothache, cavity, catch) =$$
$$P(\neg toothache | cavity) P(catch | cavity) P(cavity)$$



$\mathbf{P}(Cavity)$

$cavity$	$\neg cavity$

$\mathbf{P}(Toothache | Cavity)$

$Cavity$	$toothache$	$\neg toothache$
$cavity$		
$\neg cavity$		

$\mathbf{P}(Catch | Cavity)$

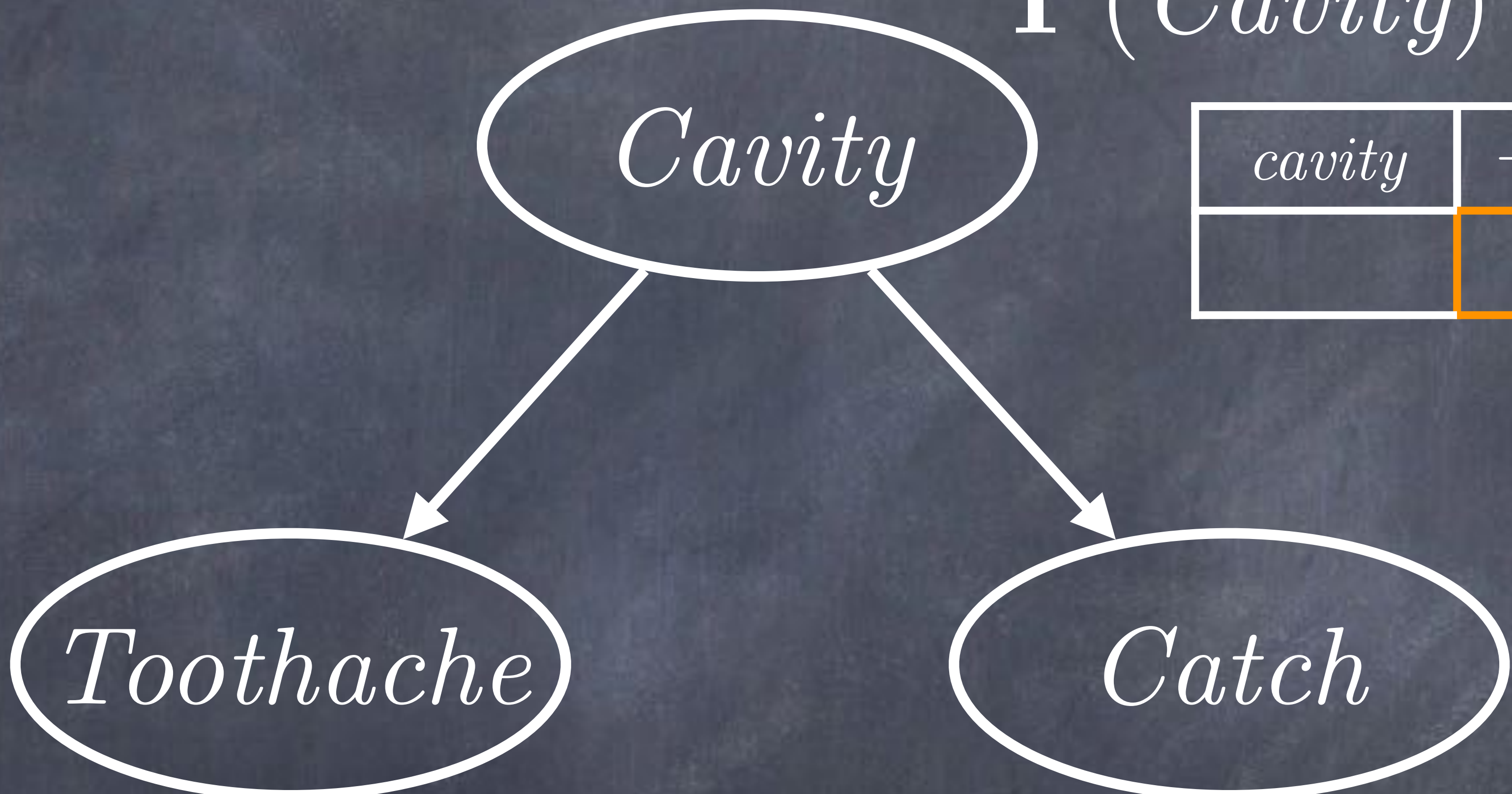
$Cavity$	$catch$	$\neg catch$
$cavity$		
$\neg cavity$		

$$P(\neg toothache, \neg cavity, \neg catch) =$$

$$P(\neg toothache | \neg cavity) P(\neg catch | \neg cavity) P(\neg cavity)$$

$\mathbf{P}(Cavity)$

$cavity$	$\neg cavity$



$\mathbf{P}(Toothache | Cavity)$

$Cavity$	$toothache$	$\neg toothache$
$cavity$		
$\neg cavity$		

$\mathbf{P}(Catch | Cavity)$

$Cavity$	$catch$	$\neg catch$
$cavity$		
$\neg cavity$		

	$toothache$		$\neg toothache$	
	$catch$	$\neg catch$	$catch$	$\neg catch$
$cavity$	0.108	0.012	0.072	0.008
$\neg cavity$	0.016	0.064	0.144	0.576

Semantics of Bayesian Networks

- Full joint distribution can be computed as the product of the separate conditional probabilities stored in the network

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

Inference in Bayesian Networks

$$\mathbf{P}(X \mid \mathbf{e}) = \alpha\, \mathbf{P}(X, \mathbf{e}) = \alpha \sum_{\mathbf{y}} \mathbf{P}(X, \mathbf{e}, \mathbf{y})$$

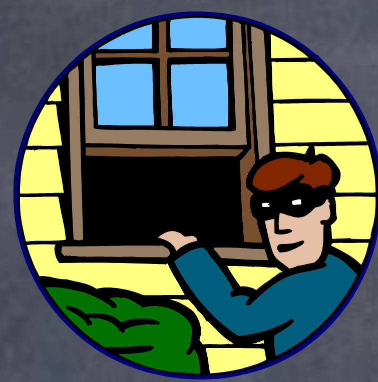
$$P(x_1, \ldots, x_n) = \prod_{i=1}^{n} P(x_i \mid parents(X_i))$$

Inference in Bayesian Networks

$$\mathbf{P}(X \mid \mathbf{e}) = \alpha\, \mathbf{P}(X, \mathbf{e}) = \alpha \sum_{\mathbf{y}} \mathbf{P}(X, \mathbf{e}, \mathbf{y})$$

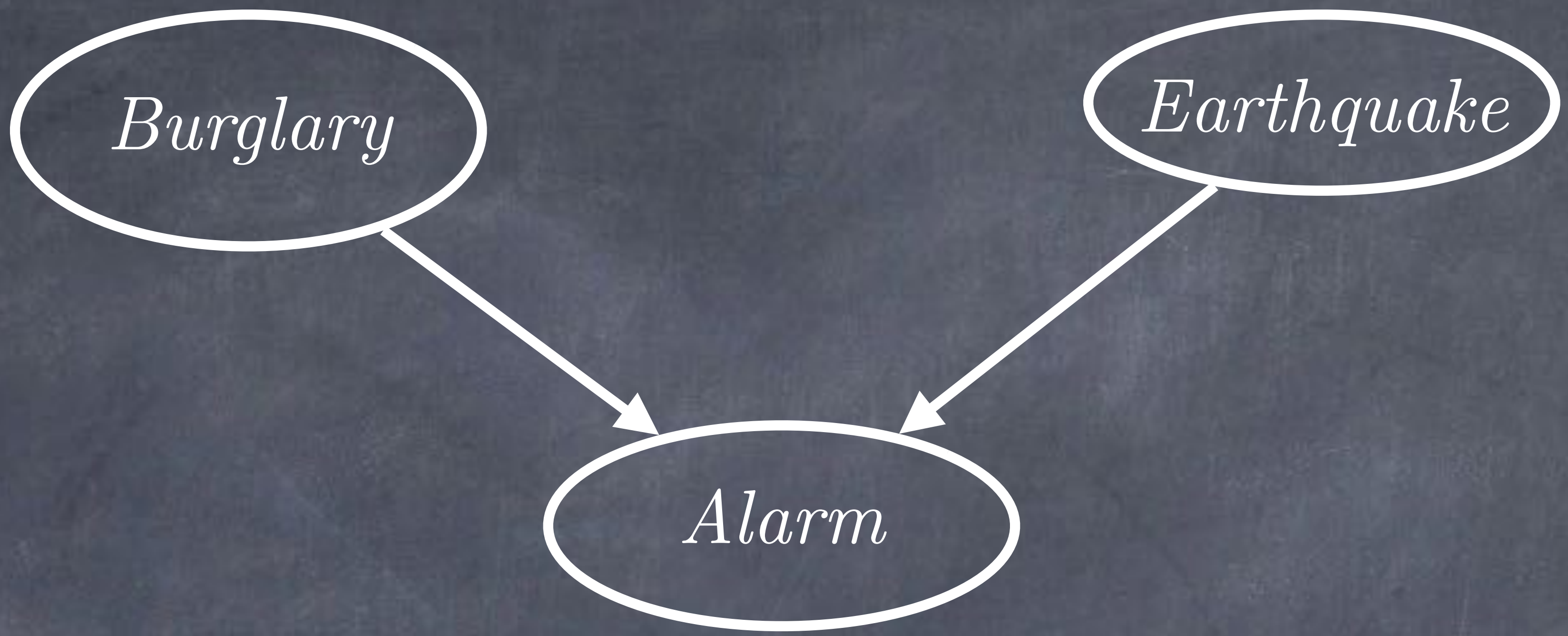
$$= \alpha \sum_{\mathbf{y}} \prod_{i=1}^{n} P(X_i \mid parents(X_i))$$

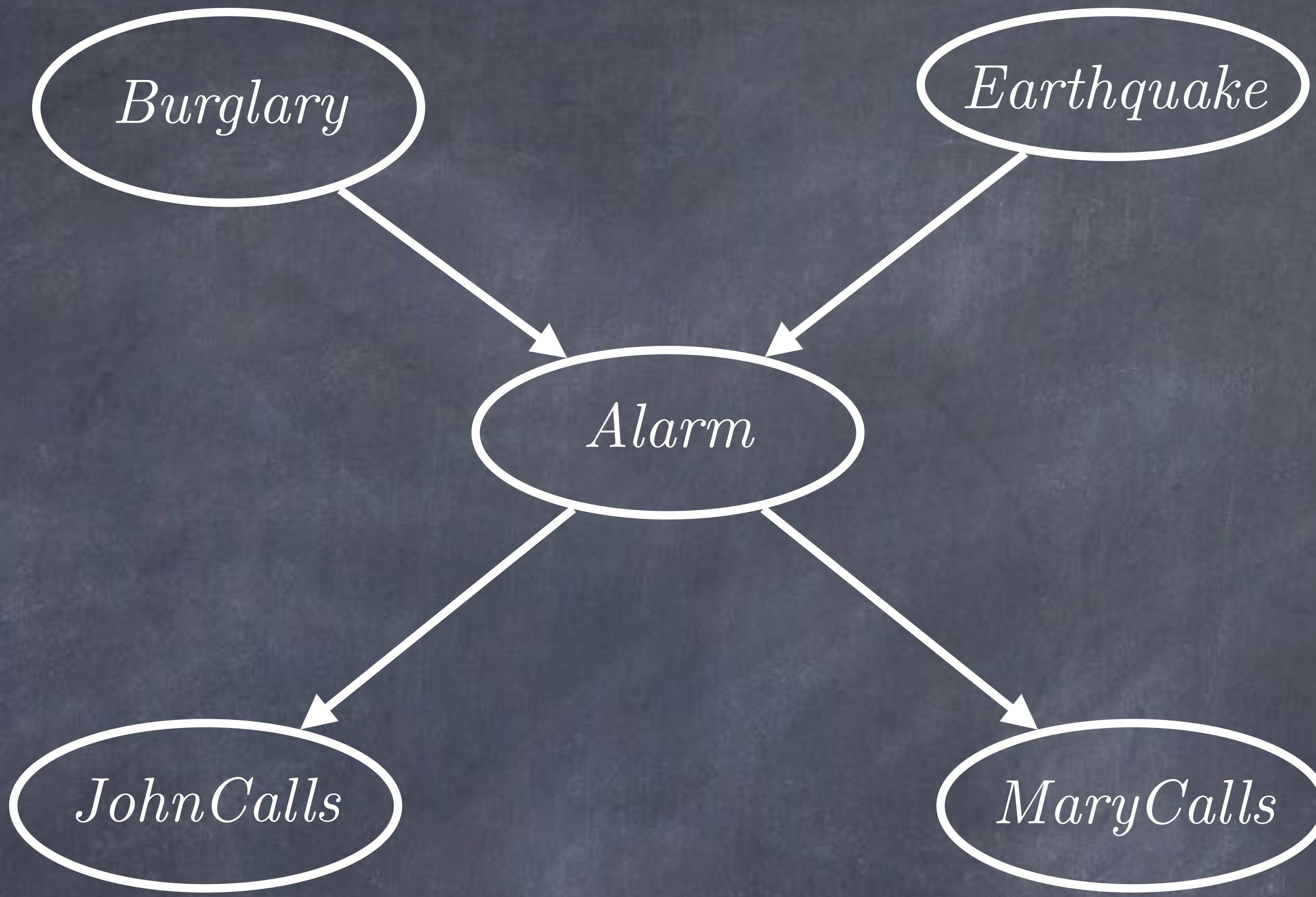
- "A query can be answered from a Bayesian Network by computing sums of products of conditional probabilities from the network."




HOME
ALARM

$Alarm$


Burglary
Earthquake
Alarm

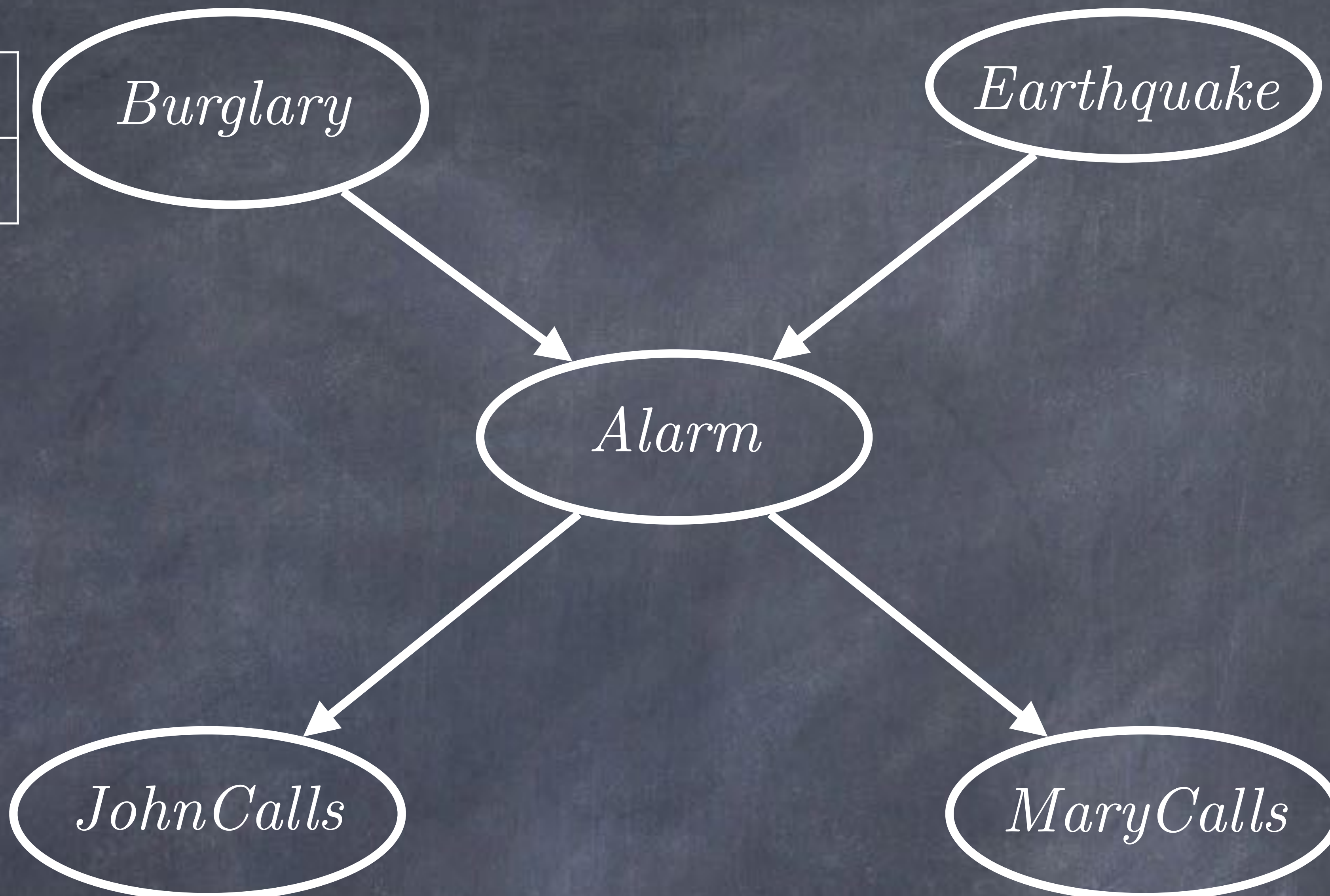

Burglary
Earthquake
Alarm
JohnCalls
MaryCalls

P(B)

$P(b)$
0.001

P(E)

$P(e)$
0.002

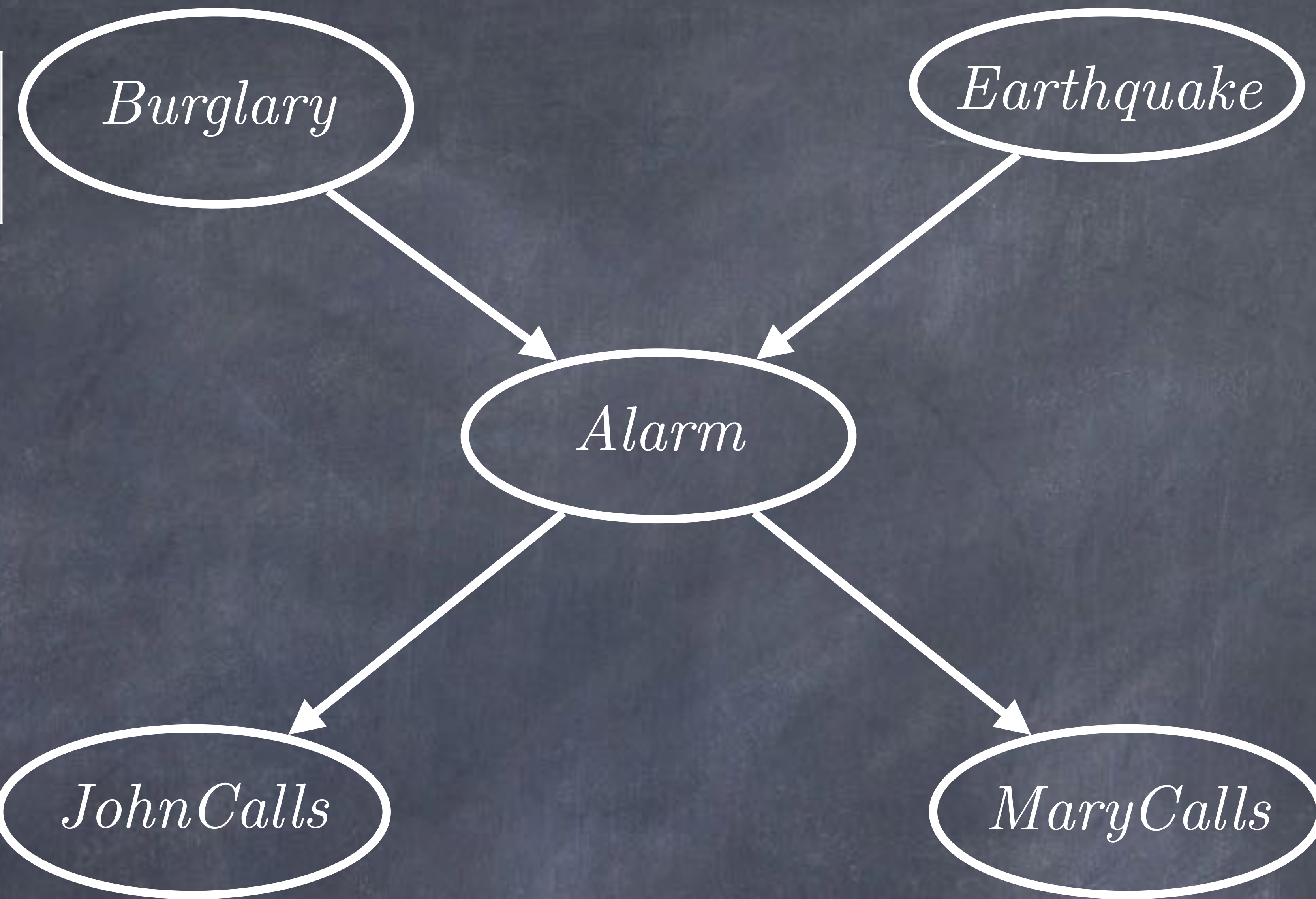


$\mathbf{P}(B)$

$P(b)$	$P(\neg b)$
0.001	1-0.001

$\mathbf{P}(E)$

$P(e)$	$P(\neg e)$
0.002	1-0.002

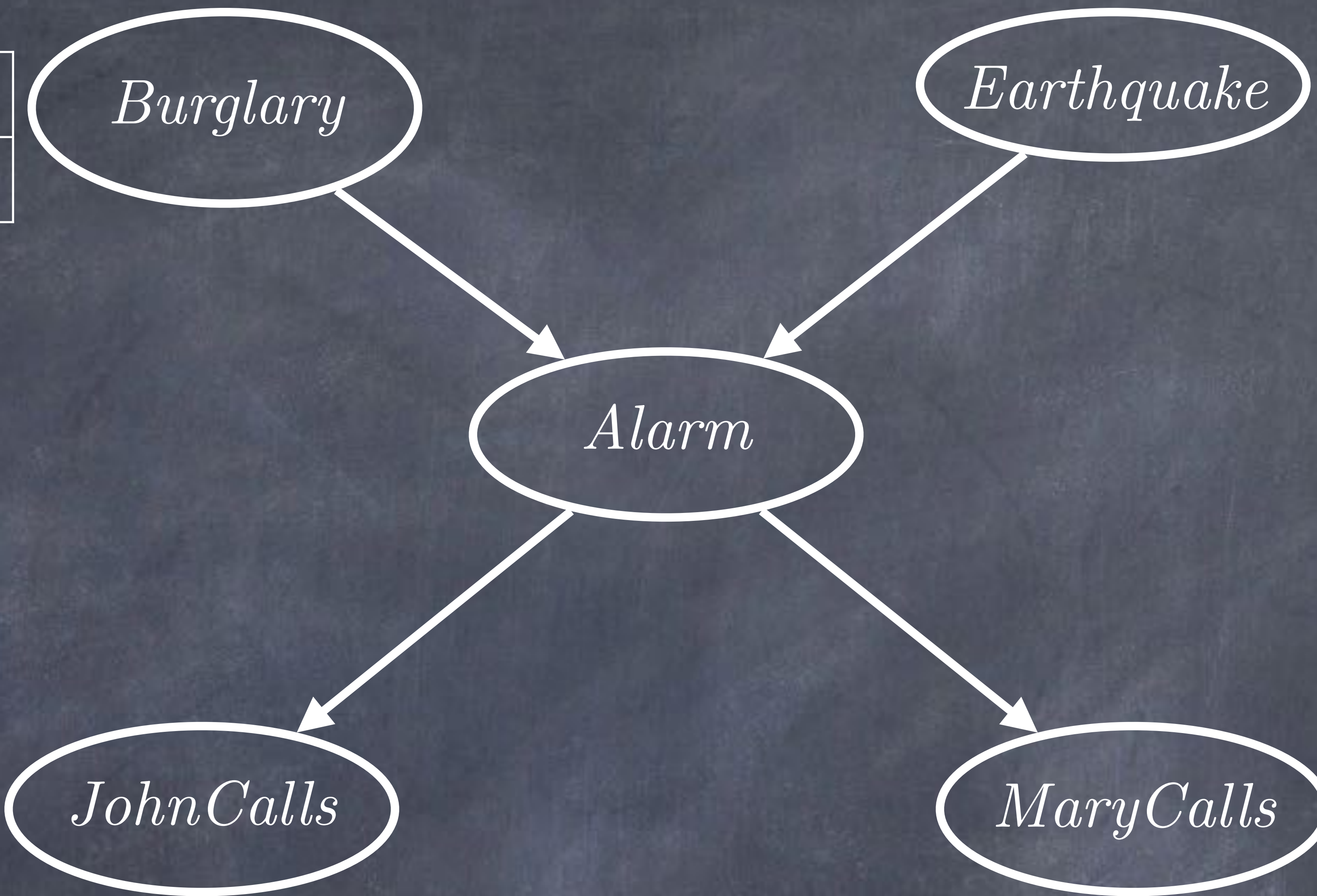


$\mathbf{P}(B)$

$P(b)$
0.001

$\mathbf{P}(E)$

$P(e)$
0.002

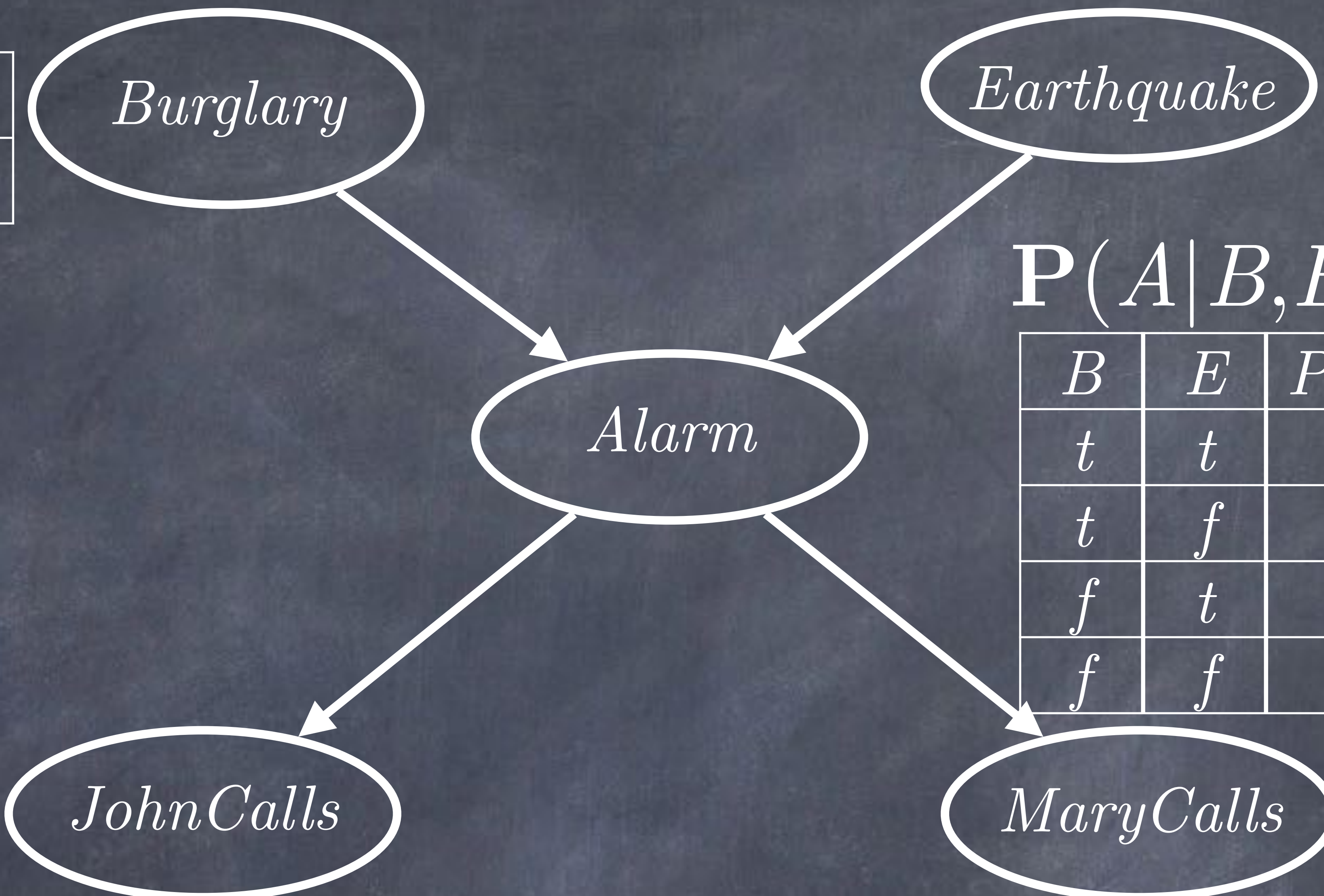


$\mathbf{P}(B)$

$P(b)$
0.001

$\mathbf{P}(E)$

$P(e)$
0.002



$\mathbf{P}(A|B,E)$

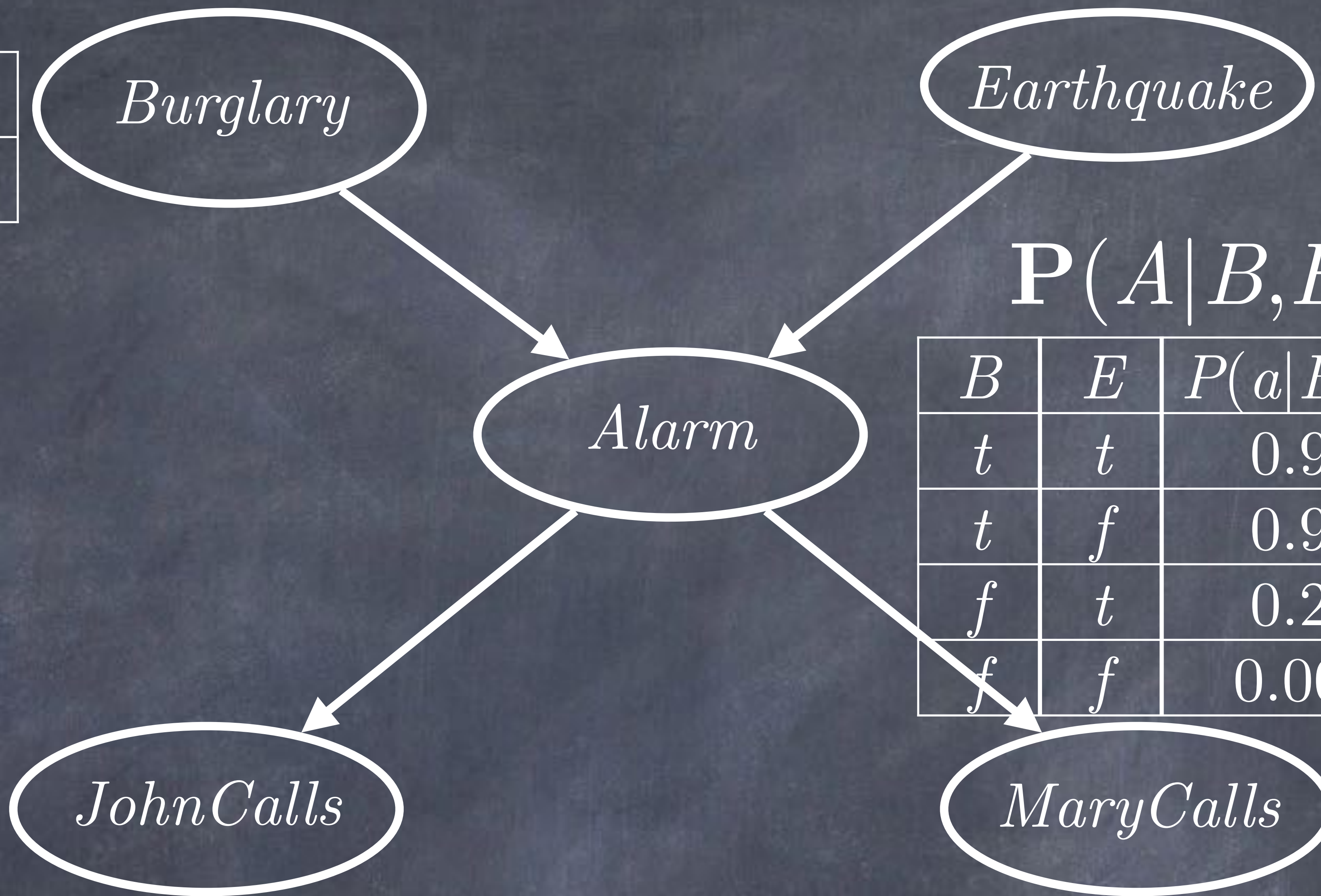
B	E	$P(a|B,E)$
t	t	0.95
t	f	0.94
f	t	0.29
f	f	0.001

$\mathbf{P}(B)$

$P(b)$
0.001

$\mathbf{P}(E)$

$P(e)$
0.002



$\mathbf{P}(A|B,E)$

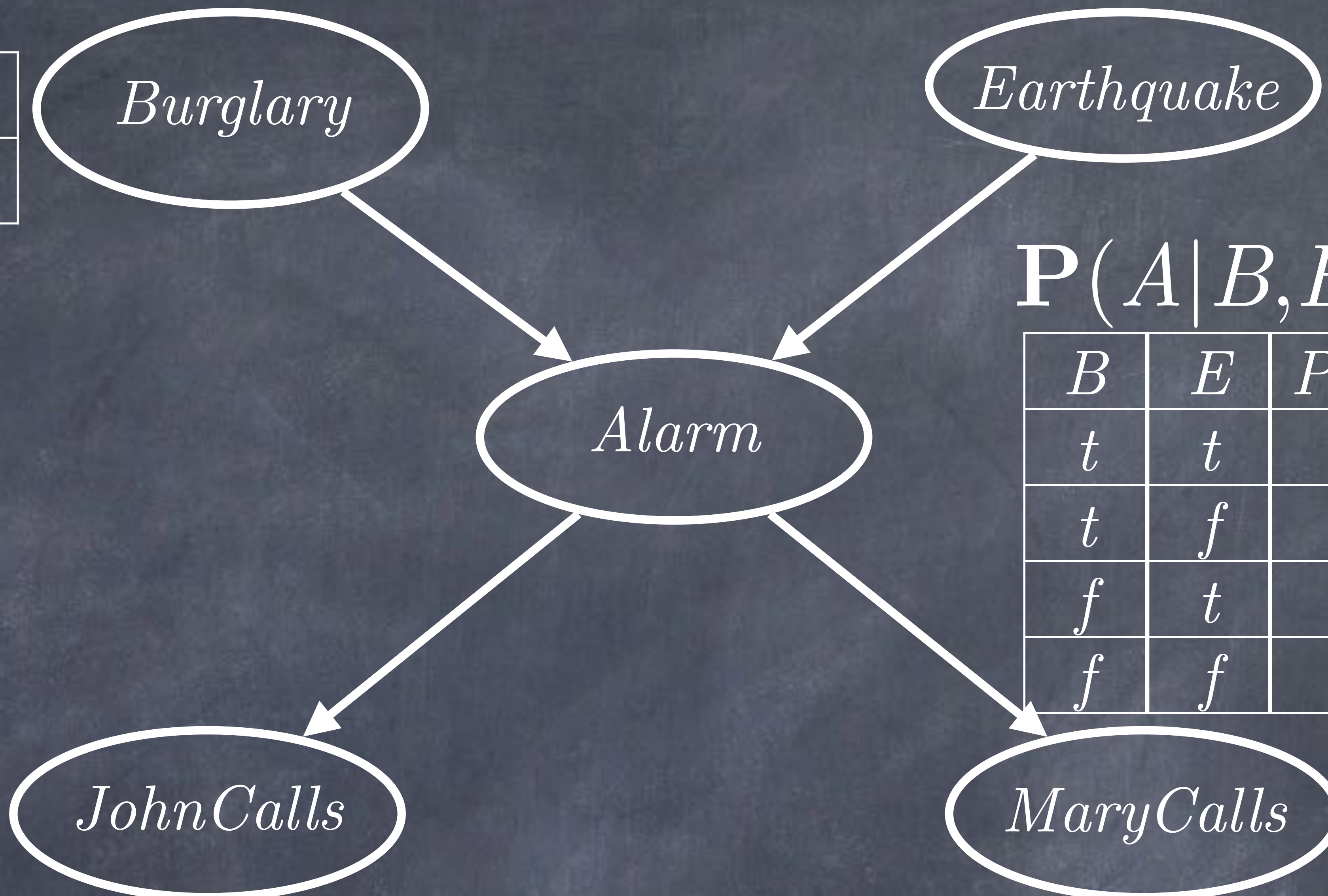
B	E	$P(a\|B,E)$	$P(\neg a\|B,E)$
t	t	0.95	1-0.95
t	f	0.94	1-0.94
f	t	0.29	1-0.29
f	f	0.001	1-0.001

$\mathbf{P}(B)$

$P(b)$
0.001

$\mathbf{P}(E)$

$P(e)$
0.002



$\mathbf{P}(A|B,E)$

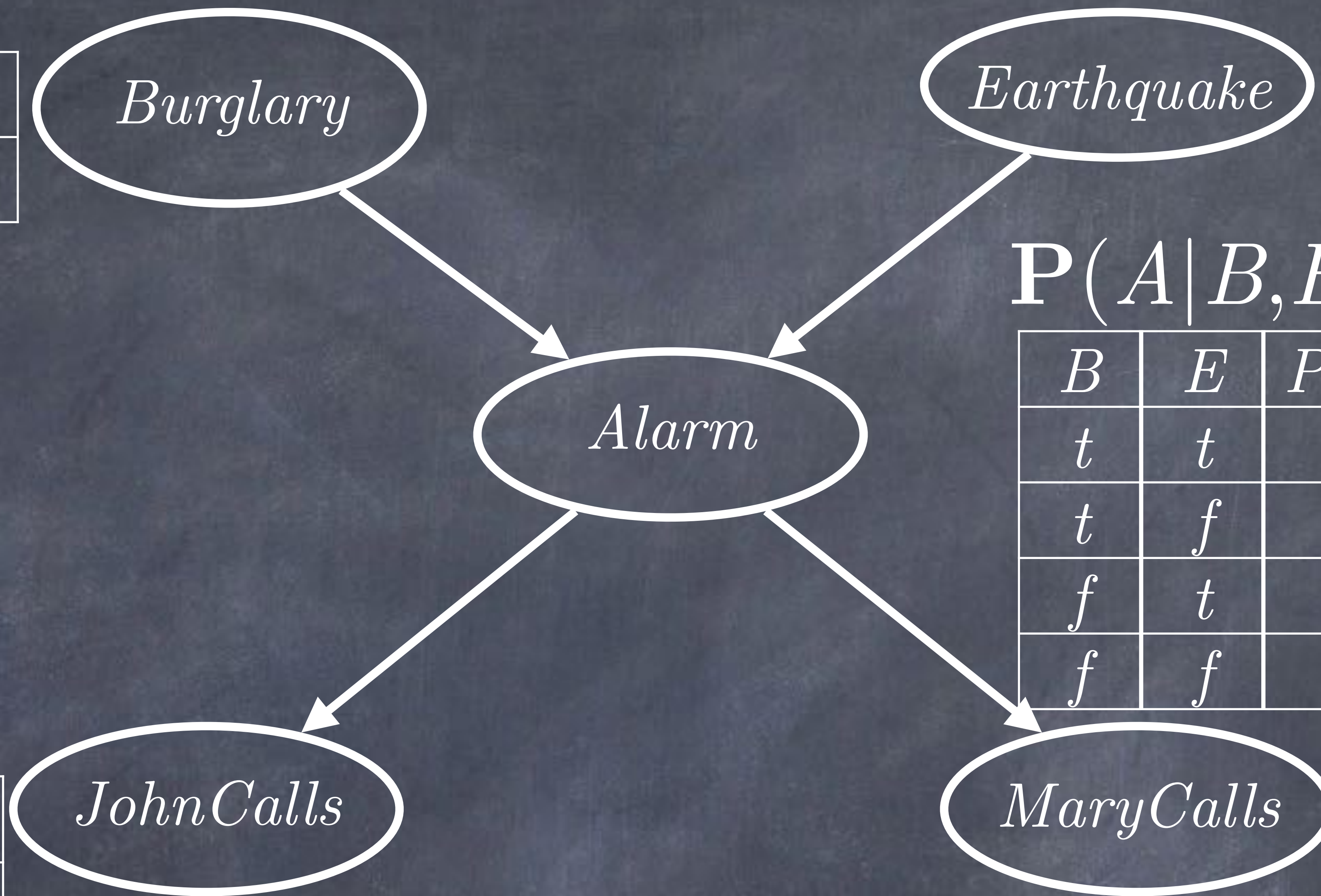
B	E	$P(a\|B,E)$
t	t	0.95
t	f	0.94
f	t	0.29
f	f	0.001

$\mathbf{P}(B)$

$P(b)$
0.001

$\mathbf{P}(E)$

$P(e)$
0.002



$\mathbf{P}(A|B,E)$

B	E	$P(a|B,E)$
t	t	0.95
t	f	0.94
f	t	0.29
f	f	0.001

$\mathbf{P}(J|A)$

A	$P(j|A)$
t	0.9
f	0.05

$\mathbf{P}(M|A)$

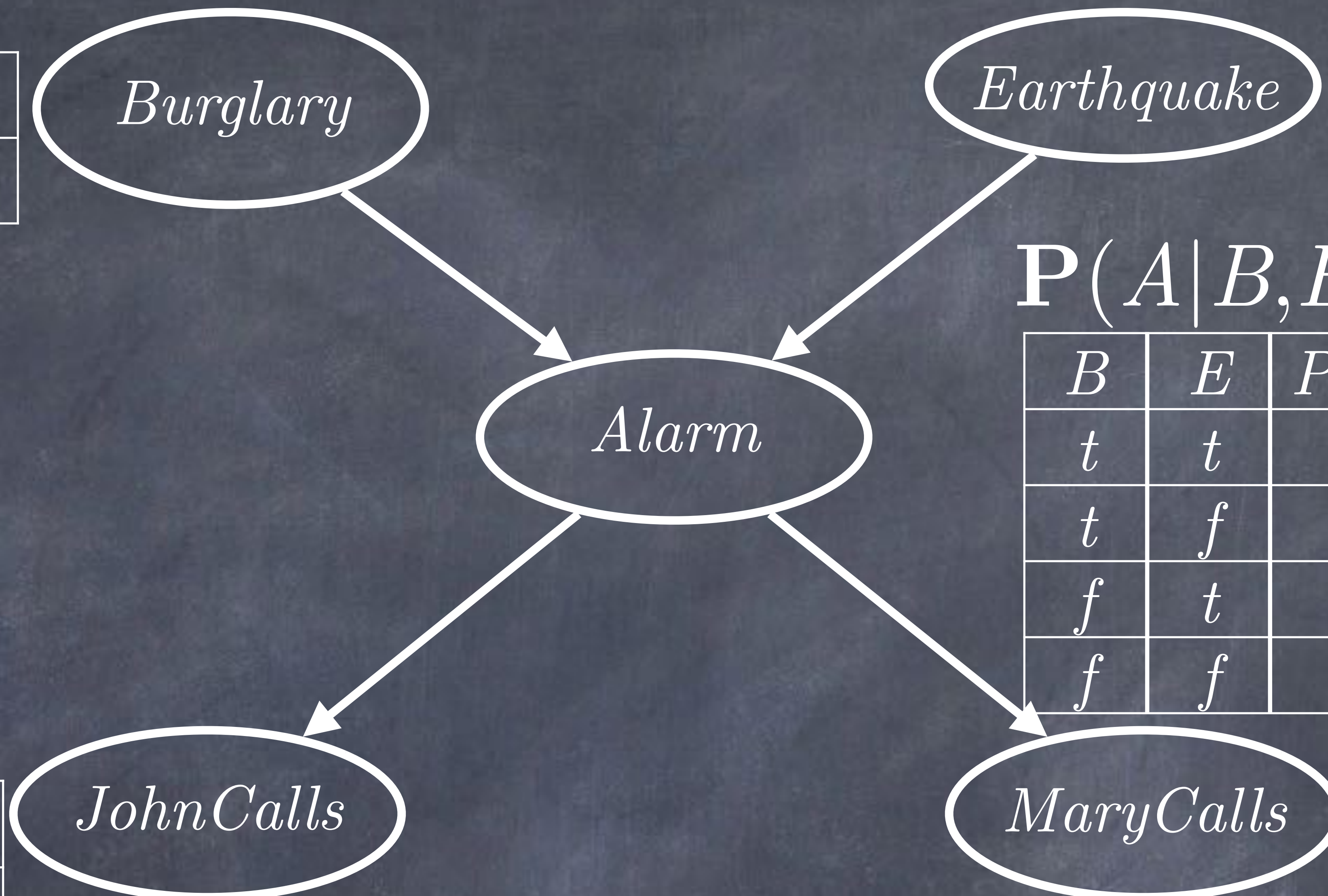
A	$P(m|A)$
t	0.7
f	0.01

$\mathbf{P}(B)$

$P(b)$
0.001

$\mathbf{P}(E)$

$P(e)$
0.002



$\mathbf{P}(A|B,E)$

B	E	$P(a\|B,E)$
t	t	0.95
t	f	0.94
f	t	0.29
f	f	0.001

$\mathbf{P}(J|A)$

A	$P(j\|A)$
t	0.9
f	0.05

$\mathbf{P}(M|A)$

A	$P(m\|A)$
t	0.7
f	0.01

Full joint distribution: 2^5=32 entries

Bayesian network: 10 entries

Assuming conditional independences encoded in the network

Exceptions:

Call from: John

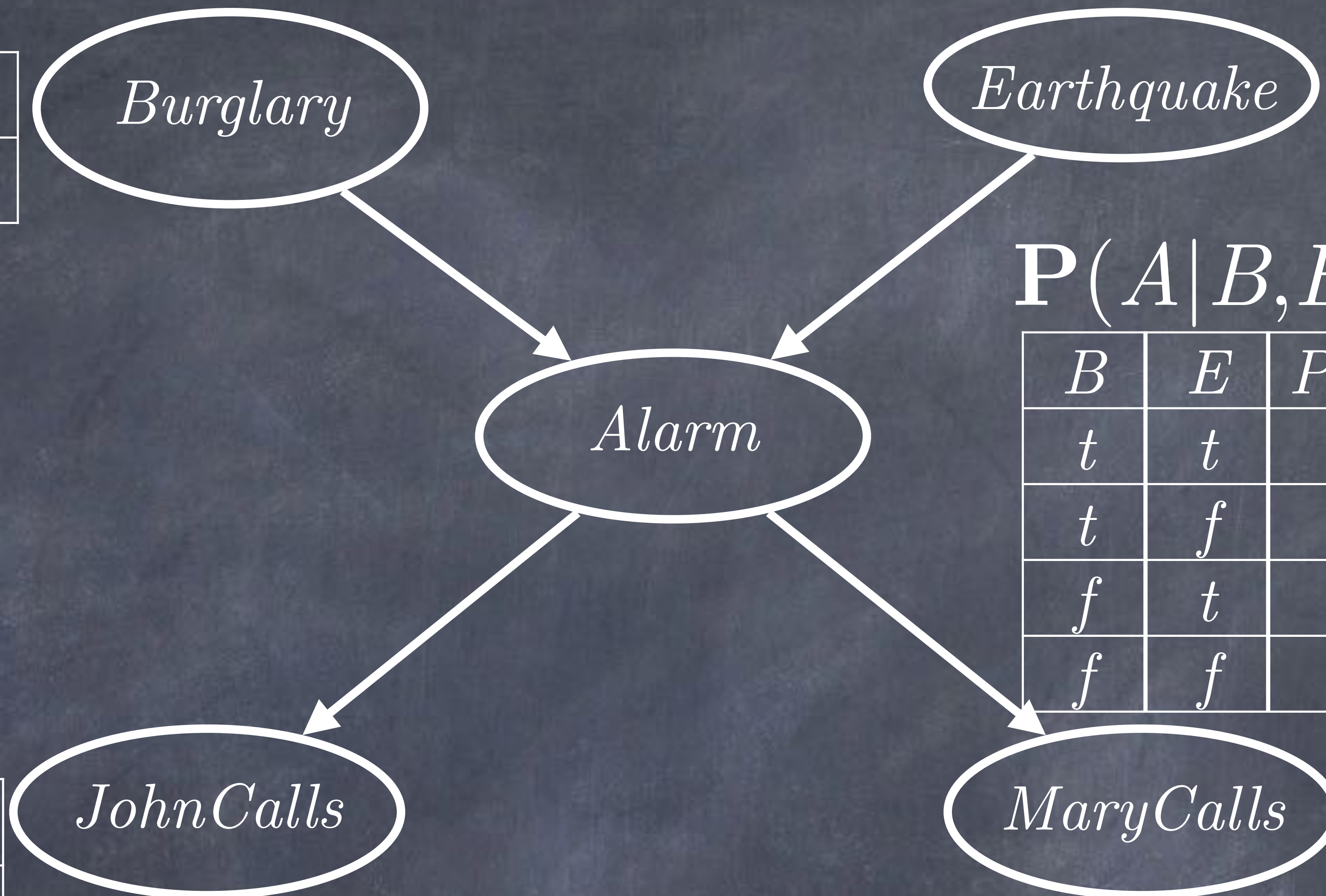
Call from: Mary

$\mathbf{P}(B)$

$P(b)$
0.001

$\mathbf{P}(E)$

$P(e)$
0.002



$\mathbf{P}(A|B,E)$

B	E	$P(a|B,E)$
t	t	0.95
t	f	0.94
f	t	0.29
f	f	0.001

$\mathbf{P}(J|A)$

A	$P(j|A)$
t	0.9
f	0.05

$\mathbf{P}(M|A)$

A	$P(m|A)$
t	0.7
f	0.01

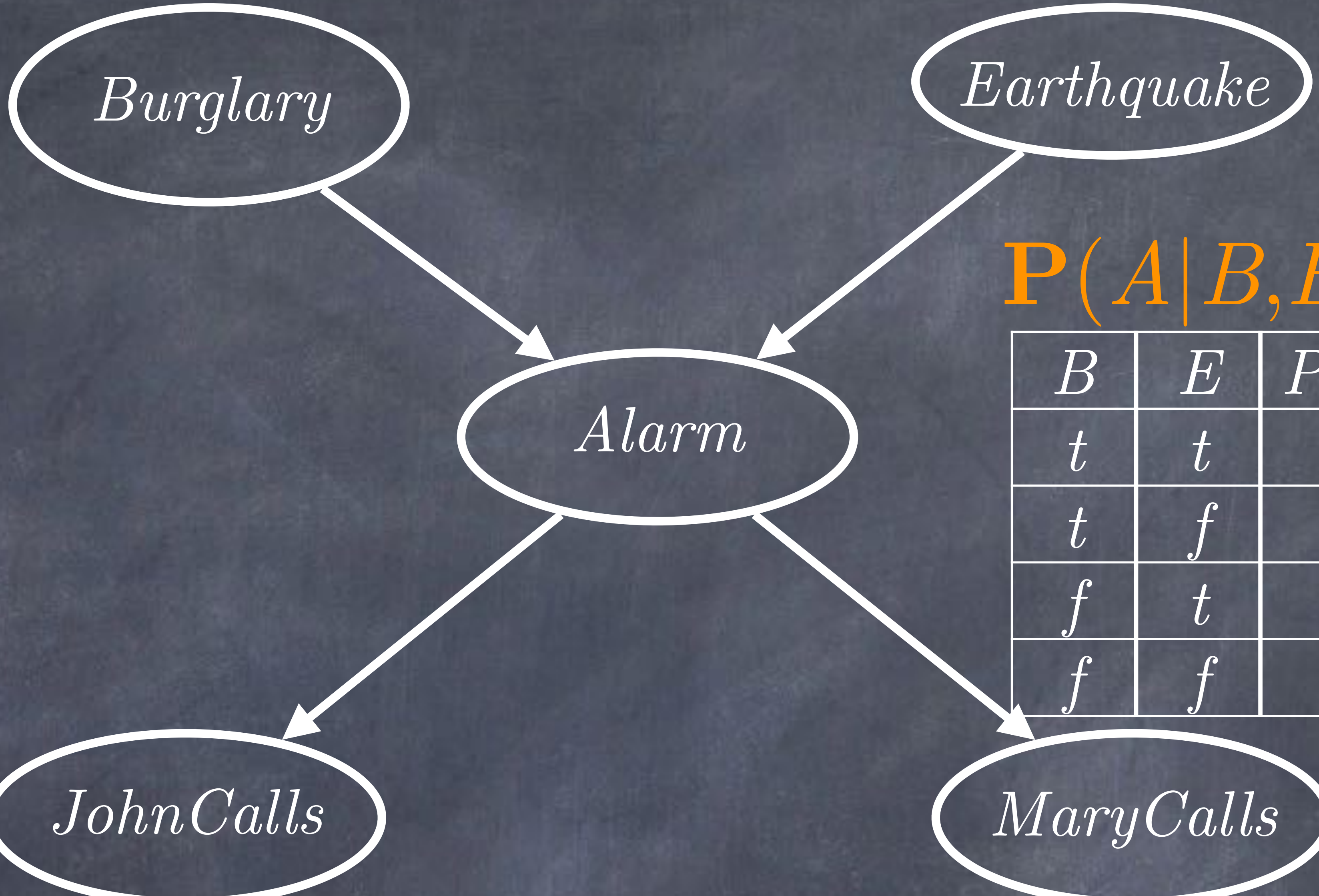
$$\mathbf{P}(Burglary \mid JohnCalls = True, MaryCalls = True)$$

$$\mathbf{P}(B \mid j, m)$$

$$\mathbf{P}(B \mid j, m) = \alpha\, \mathbf{P}(B, j, m) = \alpha \sum_{e} \sum_{a} \mathbf{P}(B, j, m, e, a)$$

$$\mathbf{P}(B \mid j, m) = \alpha\, \mathbf{P}(B, j, m) = \alpha \sum_{e} \sum_{a} \mathbf{P}(B, j, m, e, a)$$

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



$\mathbf{P}(B)$

$P(b)$
0.001

$\mathbf{P}(E)$

$P(e)$
0.002

$\mathbf{P}(A|B,E)$

B	E	$P(a|B,E)$
t	t	0.95
t	f	0.94
f	t	0.29
f	f	0.001

$\mathbf{P}(J|A)$

A	$P(j|A)$
t	0.9
f	0.05

$\mathbf{P}(M|A)$

A	$P(m|A)$
t	0.7
f	0.01

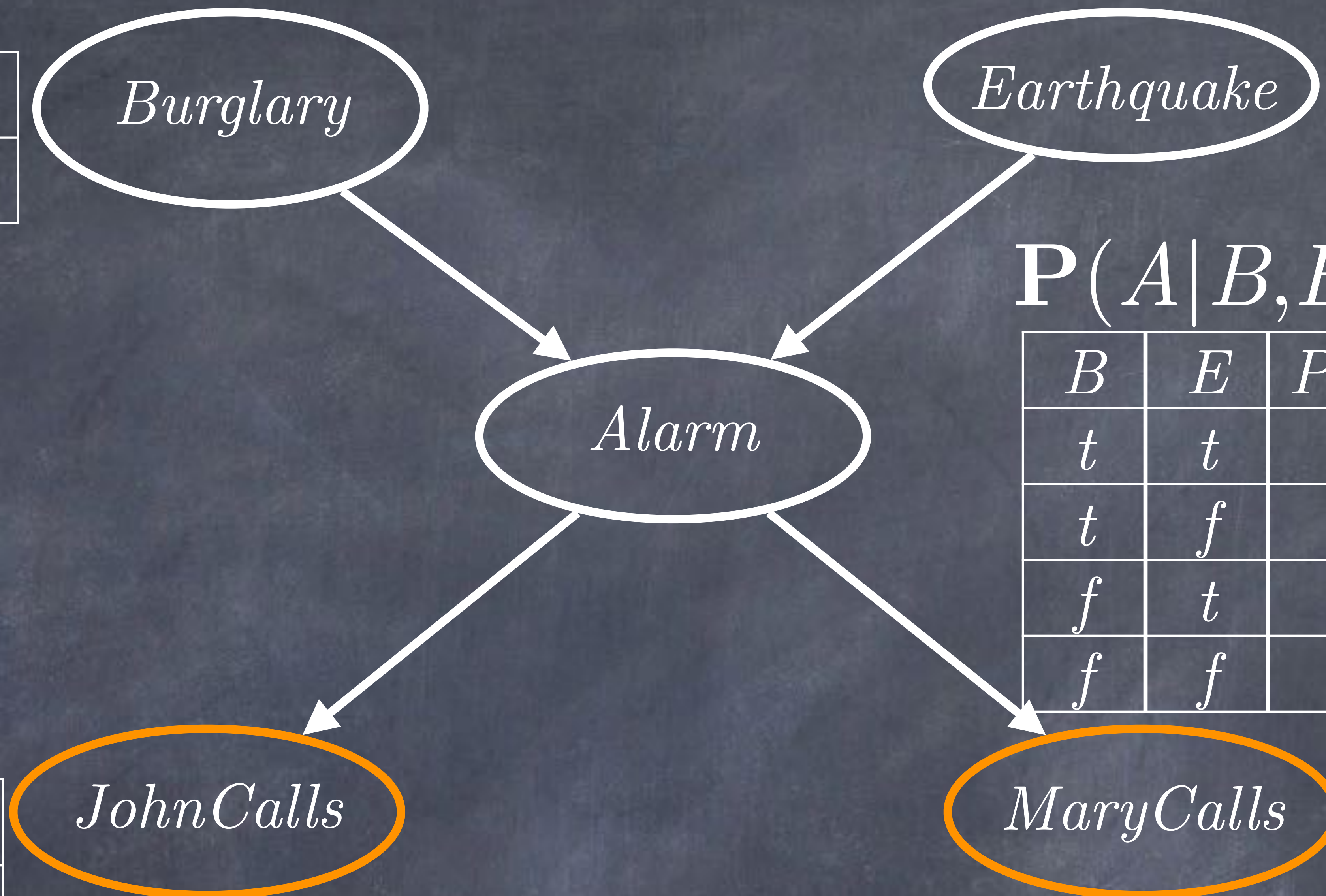
$$\mathbf{P}(B, E, A, J, M) = \alpha\, \mathbf{P}(B)\, \mathbf{P}(E)\, \mathbf{P}(A \mid B, E)\, \mathbf{P}(J \mid A)\, \mathbf{P}(M \mid A)$$

$\mathbf{P}(B)$

$P(b)$
0.001

$\mathbf{P}(E)$

$P(e)$
0.002



$\mathbf{P}(A|B,E)$

B	E	$P(a|B,E)$
t	t	0.95
t	f	0.94
f	t	0.29
f	f	0.001

$\mathbf{P}(J|A)$

A	$P(j|A)$
t	0.9
f	0.05

$\mathbf{P}(M|A)$

A	$P(m|A)$
t	0.7
f	0.01

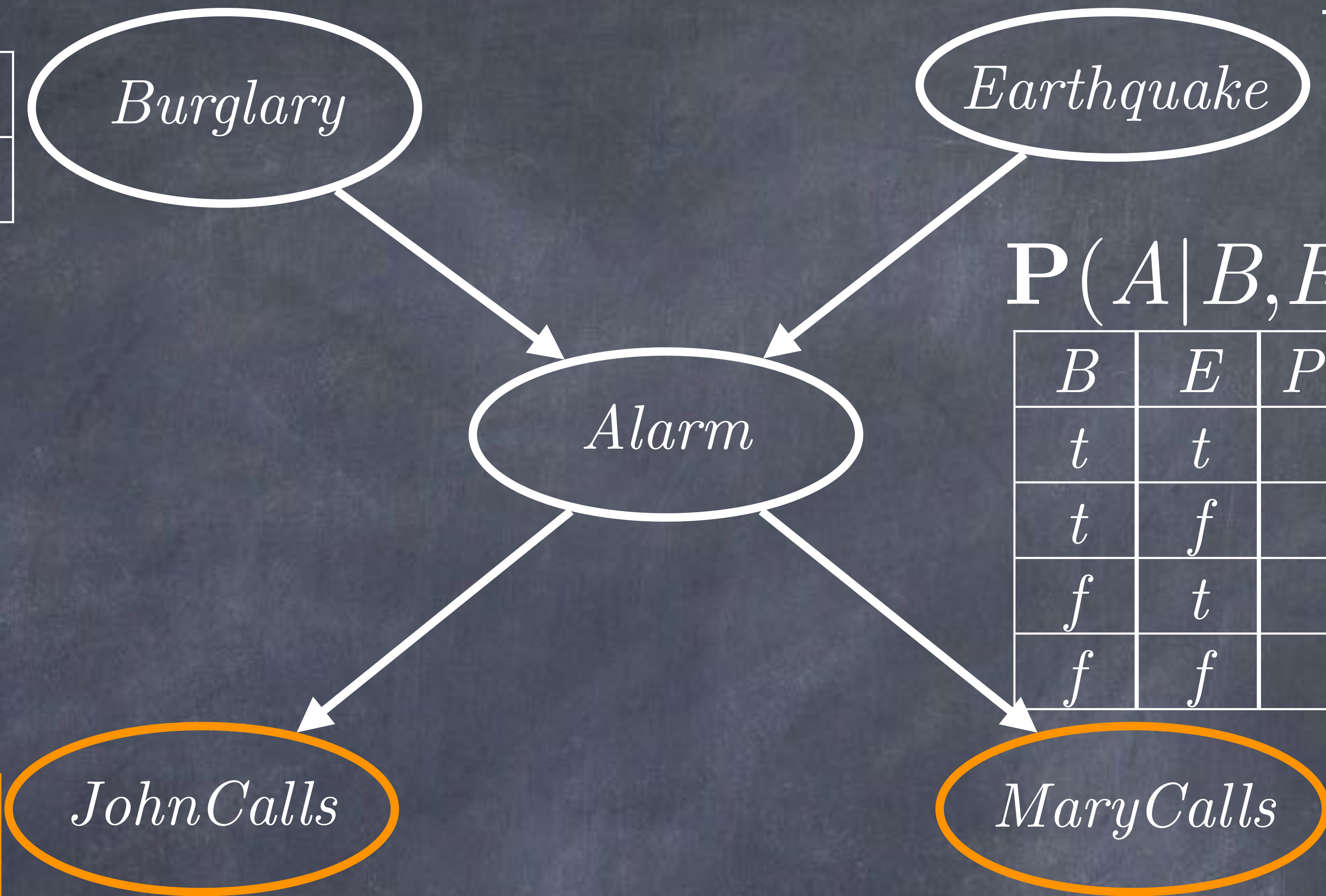
$$\mathbf{P}(B, E, A, \underline{J, M}) = \alpha\, \mathbf{P}(B)\, \mathbf{P}(E)\, \mathbf{P}(A \mid B, E)\, \mathbf{P}(\underline{J} \mid A)\, \mathbf{P}(\underline{M} \mid A)$$

$\mathbf{P}(B)$

$P(b)$
0.001

$\mathbf{P}(E)$

$P(e)$
0.002



$\mathbf{P}(A|B,E)$

B	E	$P(a\|B,E)$
t	t	0.95
t	f	0.94
f	t	0.29
f	f	0.001

$\mathbf{P}(j|A)$

A	$P(j\|A)$
t	0.9
f	0.05

$\mathbf{P}(m|A)$

A	$P(m\|A)$
t	0.7
f	0.01

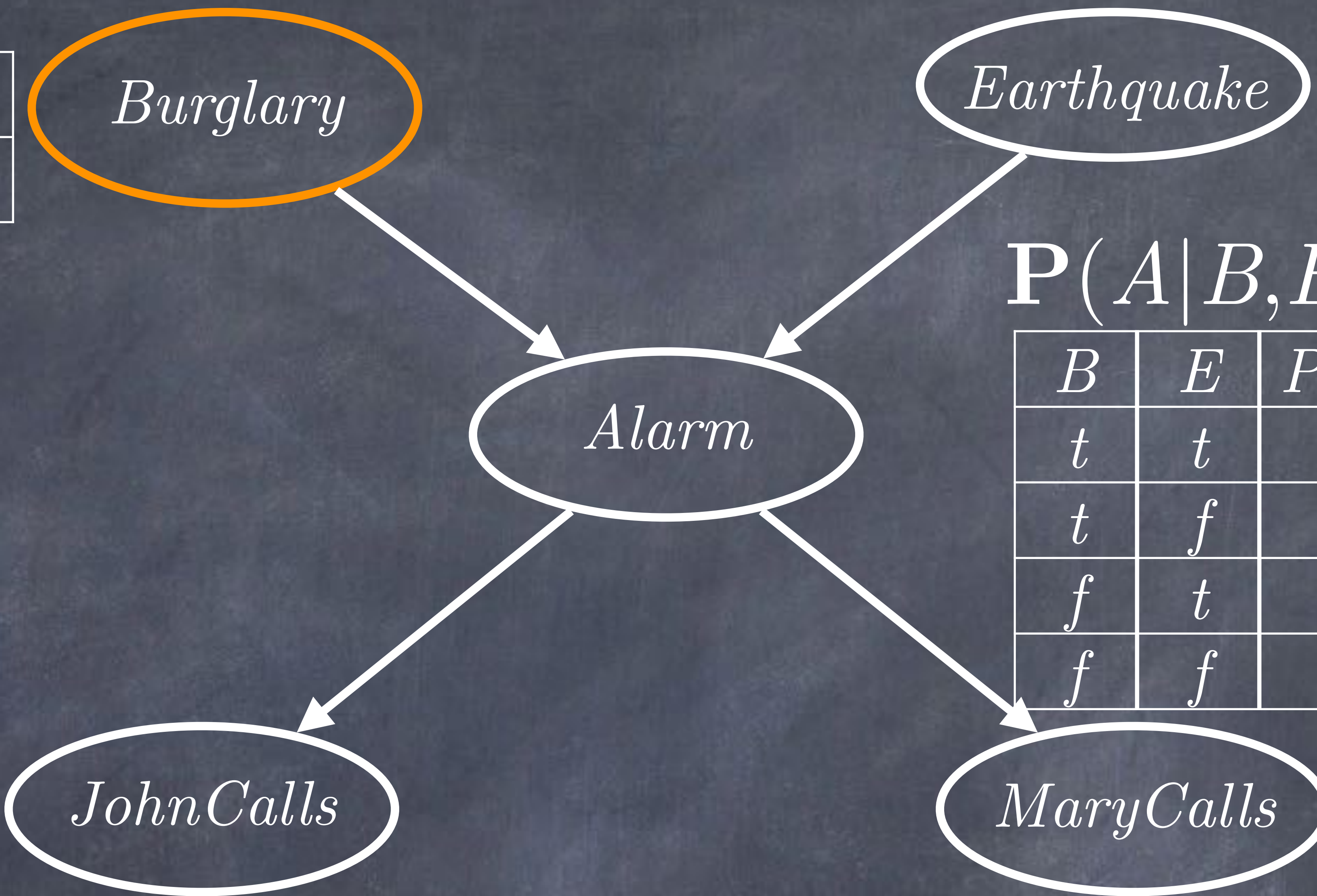
$$\mathbf{P}(B, E, A, \underline{j, m}) = \alpha\, \mathbf{P}(B)\, \mathbf{P}(E)\, \mathbf{P}(A \mid B, E)\, \mathbf{P}(\underline{j} \mid A)\, \mathbf{P}(\underline{m} \mid A)$$

$\mathbf{P}(B)$

$P(b)$
0.001

$\mathbf{P}(E)$

$P(e)$
0.002



$\mathbf{P}(A|B,E)$

B	E	$P(a|B,E)$
t	t	0.95
t	f	0.94
f	t	0.29
f	f	0.001

$\mathbf{P}(j|A)$

A	$P(j|A)$
t	0.9
f	0.05

$\mathbf{P}(m|A)$

A	$P(m|A)$
t	0.7
f	0.01

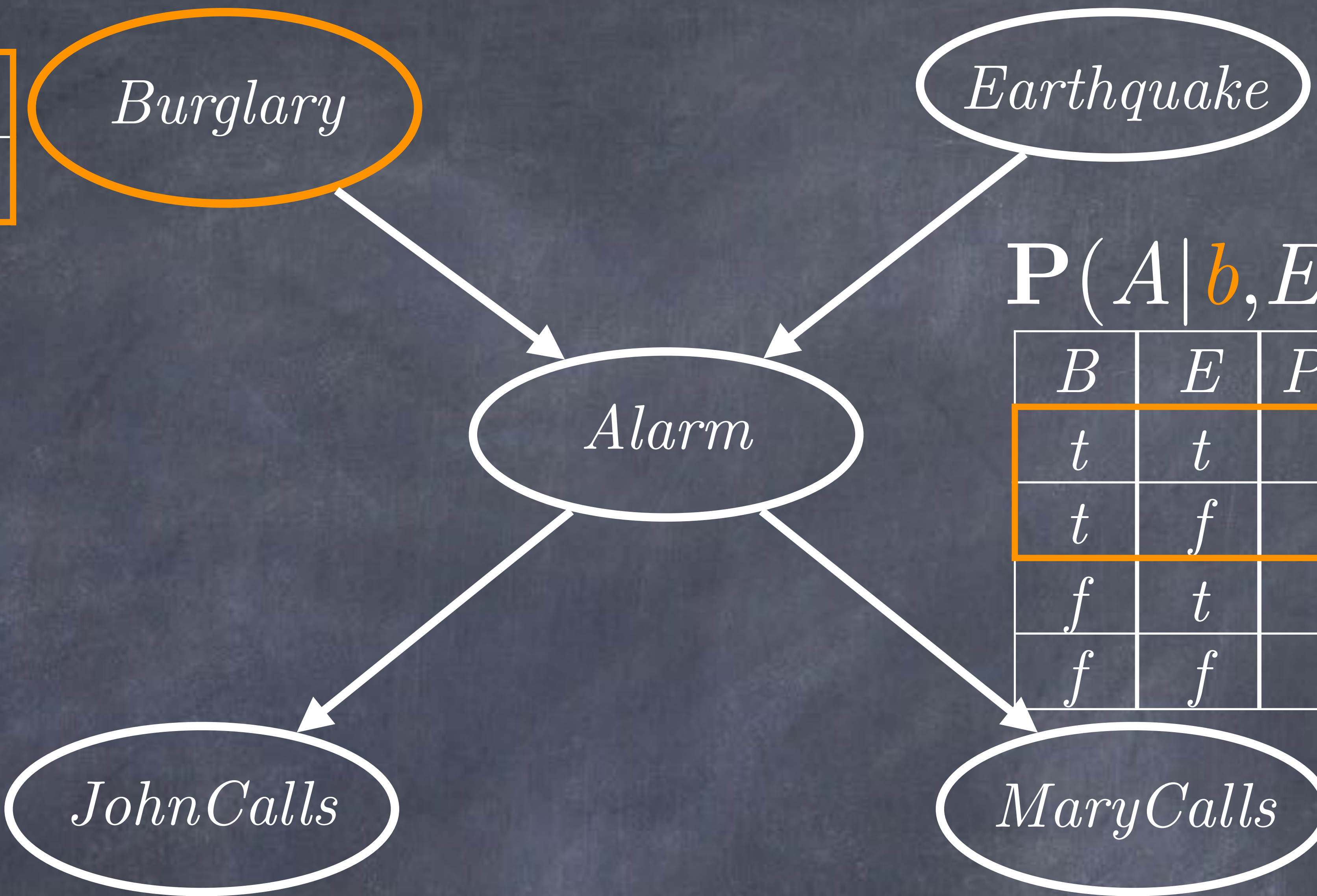
$$\mathbf{P}(\underline{B}, E, A, j, m) = \alpha\, \mathbf{P}(\underline{B})\, \mathbf{P}(E)\, \mathbf{P}(A \mid \underline{B}, E)\, \mathbf{P}(j \mid A)\, \mathbf{P}(m \mid A)$$

$\mathbf{P}(b)$

$P(b)$
0.001

$\mathbf{P}(E)$

$P(e)$
0.002



$\mathbf{P}(A|b,E)$

B	E	$P(a|B,E)$
t	t	0.95
t	f	0.94
f	t	0.29
f	f	0.001

$\mathbf{P}(j|A)$

A	$P(j|A)$
t	0.9
f	0.05

$\mathbf{P}(m|A)$

A	$P(m|A)$
t	0.7
f	0.01

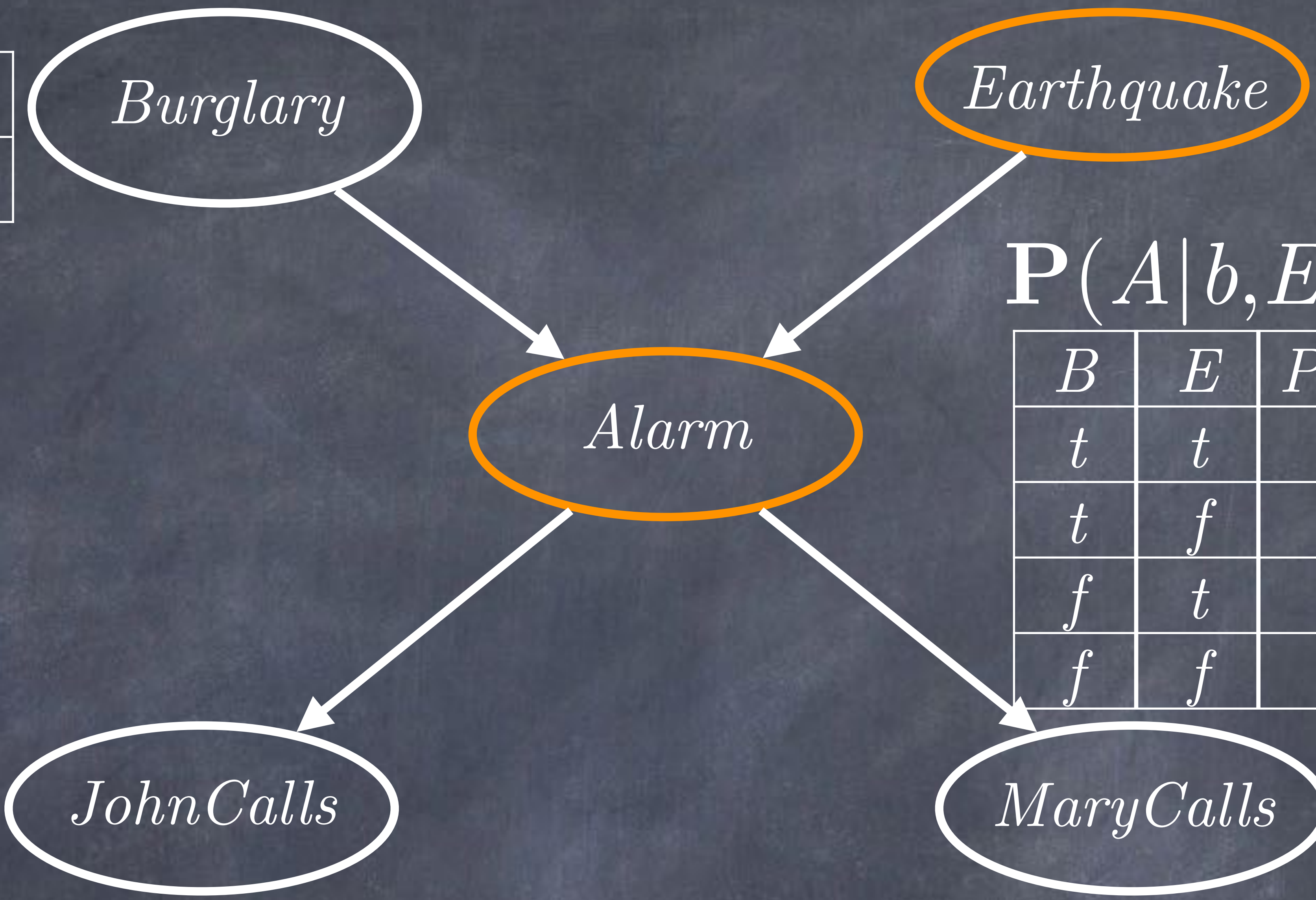
$$\mathbf{P}(\underline{b,}\, E, A, j, m) = \alpha\, \mathbf{P}(\underline{b})\, \mathbf{P}(E)\, \mathbf{P}(A \mid \underline{b,}\, E)\, \mathbf{P}(j \mid A)\, \mathbf{P}(m \mid A)$$

$\mathbf{P}(b)$

$P(b)$
0.001

$\mathbf{P}(E)$

$P(e)$
0.002



$\mathbf{P}(A|b,E)$

B	E	$P(a\|B,E)$
t	t	0.95
t	f	0.94
f	t	0.29
f	f	0.001

$\mathbf{P}(j|A)$

A	$P(j\|A)$
t	0.9
f	0.05

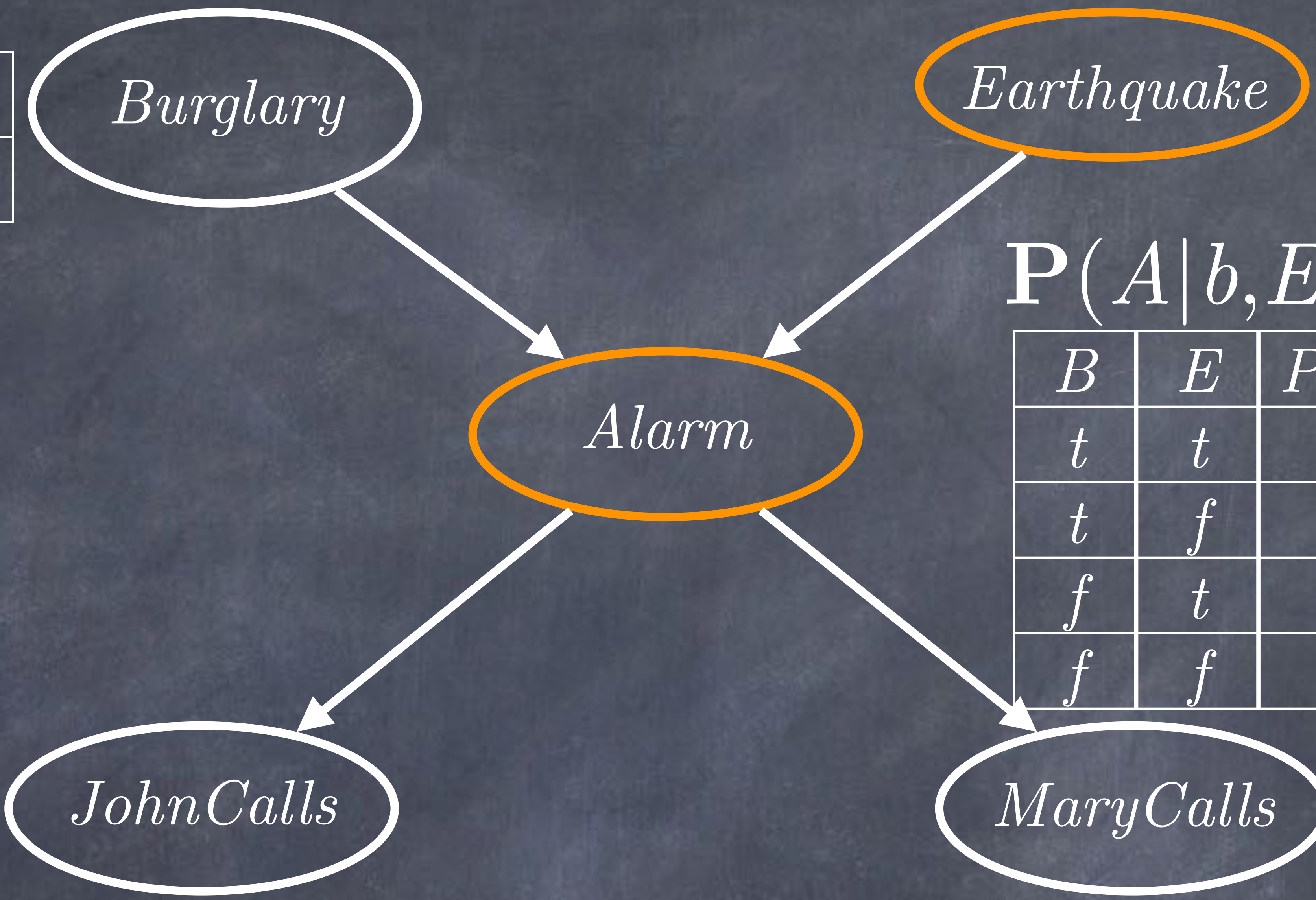
$\mathbf{P}(m|A)$

A	$P(m\|A)$
t	0.7
f	0.01

$$\mathbf{P}(b, \underline{E, A}, j, m) = \alpha\, \mathbf{P}(b)\, \mathbf{P}(\underline{E})\, \mathbf{P}(\underline{A} \mid b, \underline{E})\, \mathbf{P}(j \mid \underline{A})\, \mathbf{P}(m \mid \underline{A})$$

$\mathbf{P}(b)$

$P(b)$
0.001



$\mathbf{P}(E)$

$P(e)$
0.002

$\mathbf{P}(A|b,E)$

B	E	$P(a|B,E)$
t	t	0.95
t	f	0.94
f	t	0.29
f	f	0.001

$\mathbf{P}(j|A)$

A	$P(j|A)$
t	0.9
f	0.05

$\mathbf{P}(m|A)$

A	$P(m|A)$
t	0.7
f	0.01

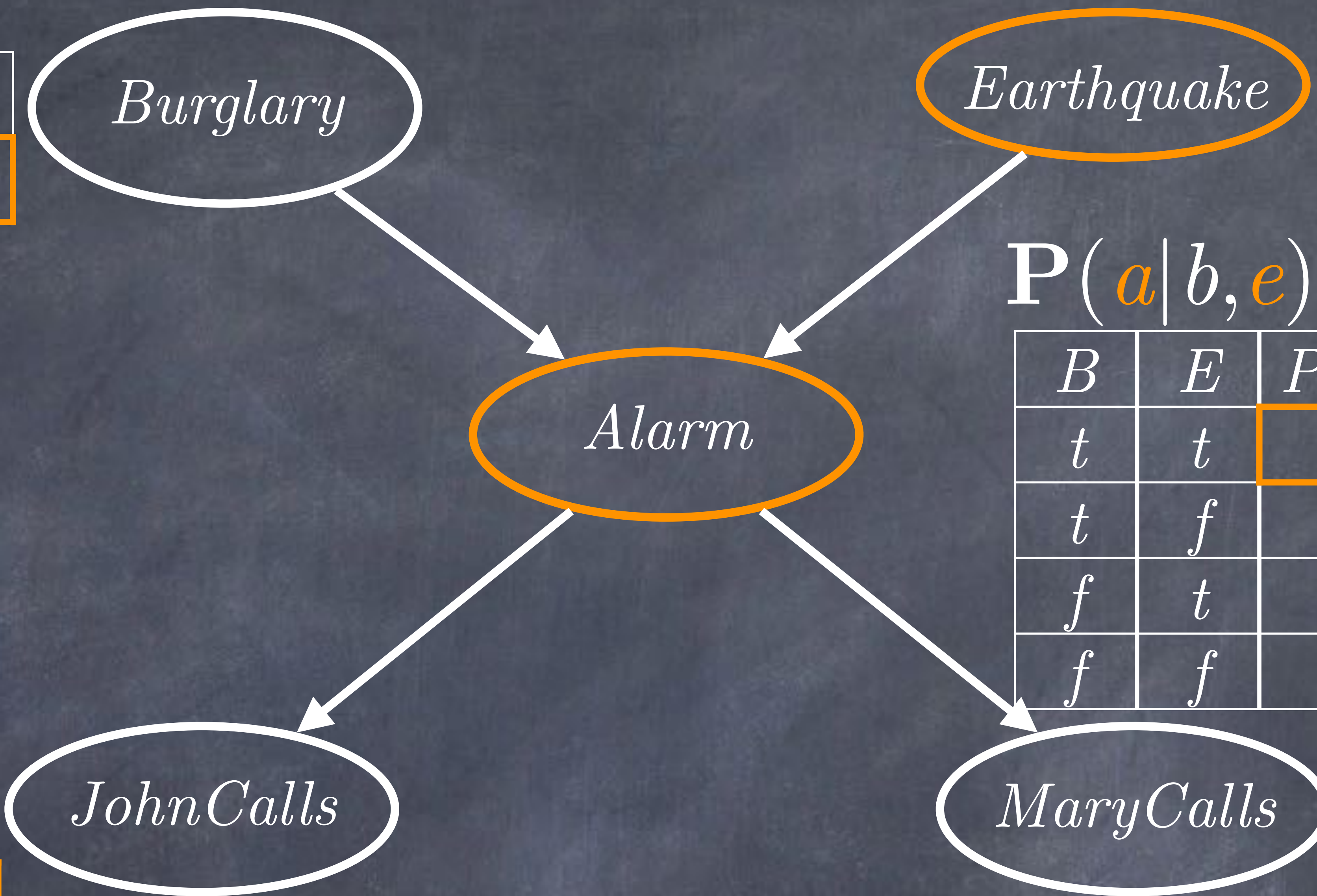
$$\mathbf{P}(b, j, m) = \alpha \sum_{e} \sum_{a} \mathbf{P}(b)\, \mathbf{P}(e)\, \mathbf{P}(a \mid b, e)\, \mathbf{P}(j \mid a)\, \mathbf{P}(m \mid a)$$

$\mathbf{P}(b)$

$P(b)$
$\boxed{0.001}$

$\mathbf{P}(e)$

$P(e)$
$\boxed{0.002}$



$\mathbf{P}(a|b,e)$

B	E	$P(a\|B,E)$
t	t	$\boxed{0.95}$
t	f	0.94
f	t	0.29
f	f	0.001

$\mathbf{P}(j|a)$

A	$P(j\|A)$
t	$\boxed{0.9}$
f	0.05

$\mathbf{P}(m|a)$

A	$P(m\|A)$
t	$\boxed{0.7}$
f	0.01

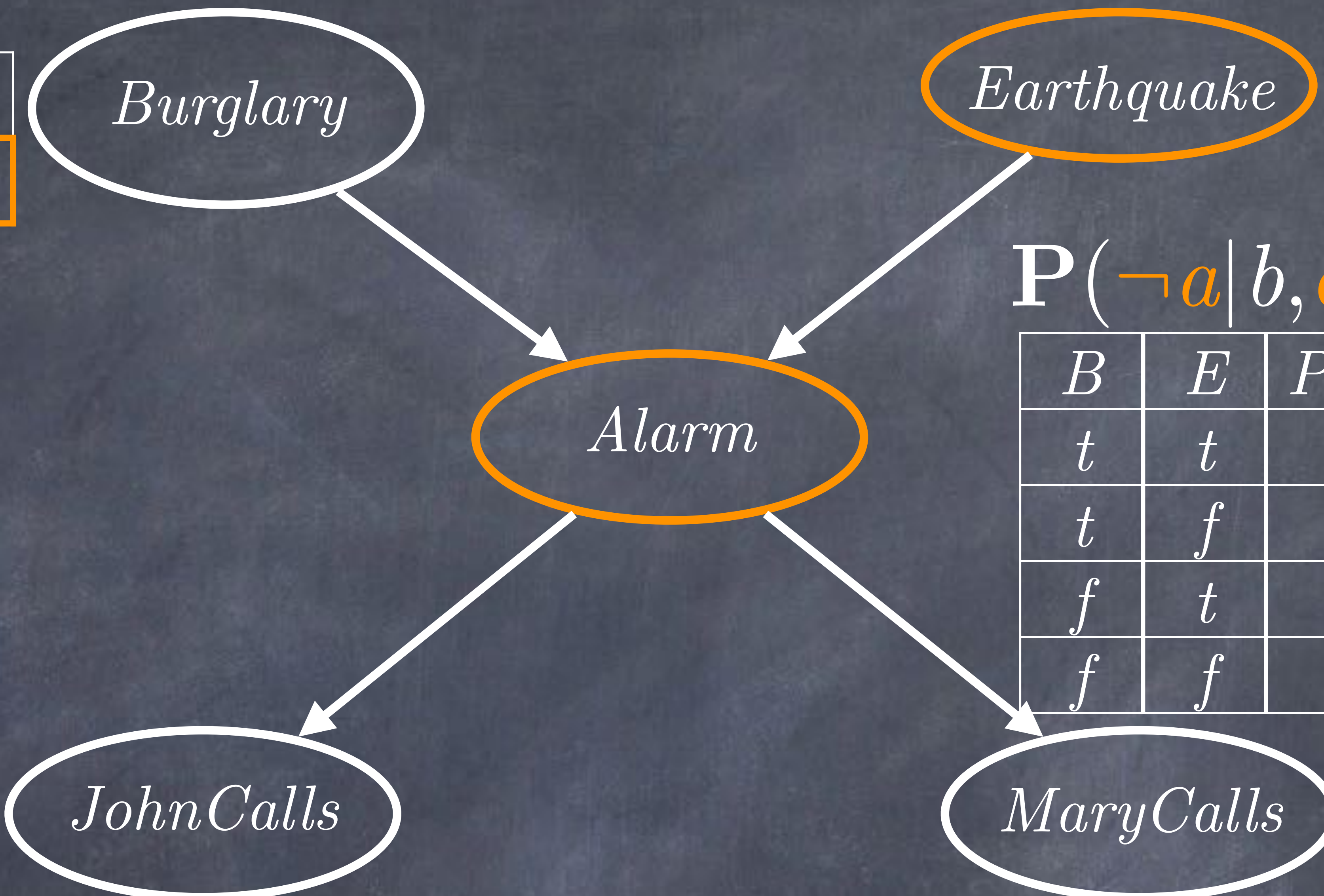
$$\mathbf{P}(b, j, m) = \alpha\, \mathbf{P}(b)\, \mathbf{P}(e)\, \mathbf{P}(a \mid b, e)\, \mathbf{P}(j \mid a)\, \mathbf{P}(m \mid a) +$$

$\mathbf{P}(b)$

$P(b)$
$\boxed{0.001}$

$\mathbf{P}(e)$

$P(e)$
$\boxed{0.002}$



$\mathbf{P}(\neg a | b, e)$

B	E	$P(a\|B,E)$	
t	t	0.95	1-0.95
t	f	0.94	
f	t	0.29	
f	f	0.001	

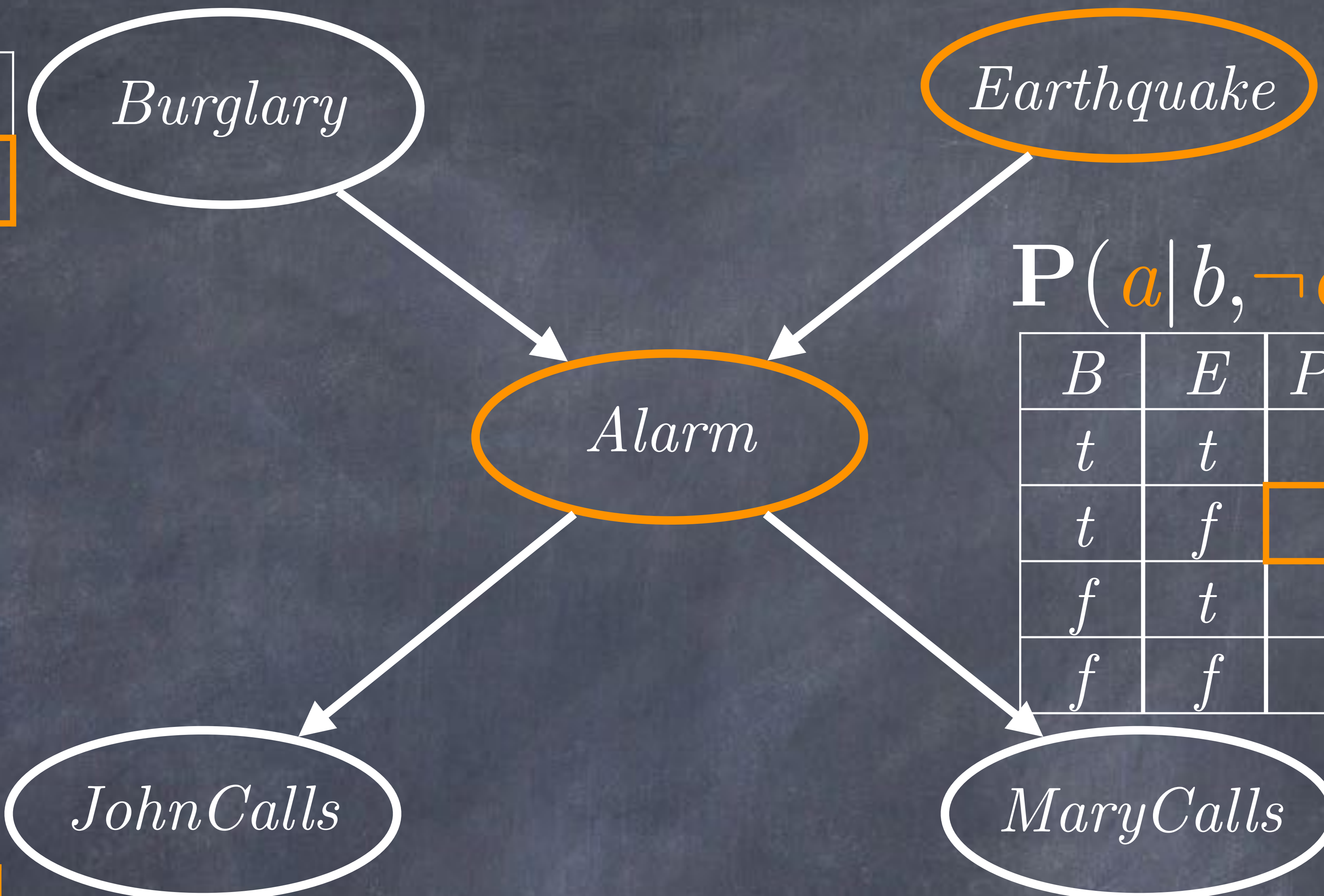
$\mathbf{P}(j | \neg a)$

A	$P(j\|A)$
t	0.9
f	$\boxed{0.05}$

$\mathbf{P}(m | \neg a)$

A	$P(m\|A)$
t	0.7
f	$\boxed{0.01}$

$$\mathbf{P}(b, j, m) = \alpha\, \mathbf{P}(b)\, \mathbf{P}(e)\, \mathbf{P}(a \mid b, e)\, \mathbf{P}(j \mid a)\, \mathbf{P}(m \mid a) +$$
$$\mathbf{P}(b)\, \mathbf{P}(e)\, \mathbf{P}(\neg a \mid b, e)\, \mathbf{P}(j \mid \neg a)\, \mathbf{P}(m \mid \neg a) +$$



$\mathbf{P}(b)$

$P(b)$
0.001

$\mathbf{P}(\neg e)$

$P(e)$	
0.002	1-0.002

$\mathbf{P}(a|b,\neg e)$

B	E	$P(a\|B,E)$
t	t	0.95
t	f	0.94
f	t	0.29
f	f	0.001

$\mathbf{P}(j|a)$

A	$P(j\|A)$
t	0.9
f	0.05

$\mathbf{P}(m|a)$

A	$P(m\|A)$
t	0.7
f	0.01

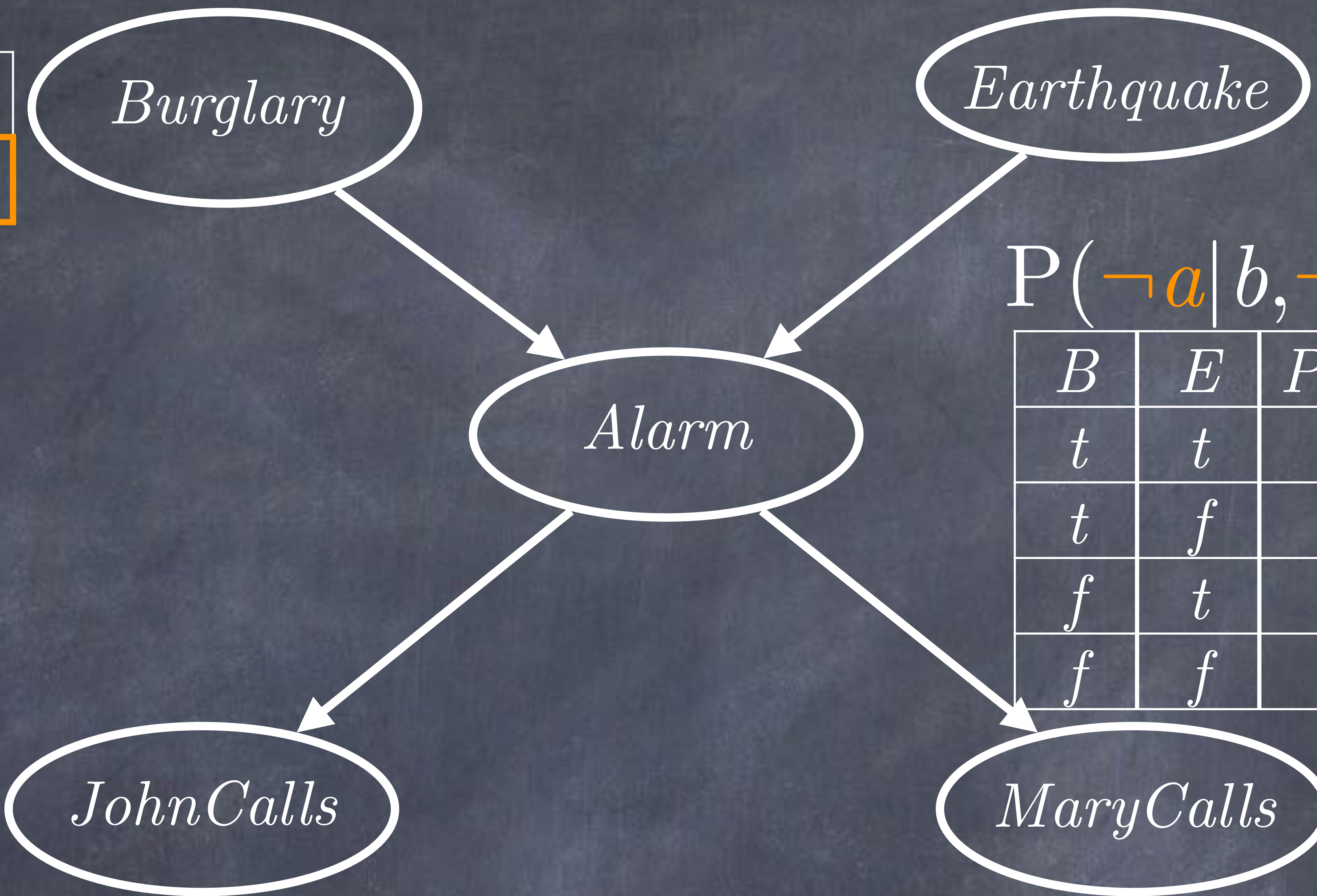
$$\mathbf{P}(b,j,m) = \alpha\, \mathbf{P}(b)\, \mathbf{P}(e)\, \mathbf{P}(a \mid b,e)\, \mathbf{P}(j \mid a)\, \mathbf{P}(m \mid a) +$$
$$\mathbf{P}(b)\, \mathbf{P}(e)\, \mathbf{P}(\neg a \mid b,e)\, \mathbf{P}(j \mid \neg a)\, \mathbf{P}(m \mid \neg a) +$$
$$\mathbf{P}(b)\, \mathbf{P}(\neg e)\, \mathbf{P}(a \mid b,\neg e)\, \mathbf{P}(j \mid a)\, \mathbf{P}(m \mid a) +$$

$P(b)$

$P(b)$
$\boxed{0.001}$

$P(\neg e)$

$P(e)$	
0.002	$\boxed{1-.002}$



$P(\neg a | b, \neg e)$

B	E	$P(a \mid B,E)$	
t	t	0.95	
t	f	0.94	$\boxed{1-.94}$
f	t	0.29	
f	f	0.001	

$P(j | \neg a)$

A	$P(j \mid A)$
t	0.9
f	$\boxed{0.05}$

$P(m | \neg a)$

A	$P(m \mid A)$
t	0.7
f	$\boxed{0.01}$

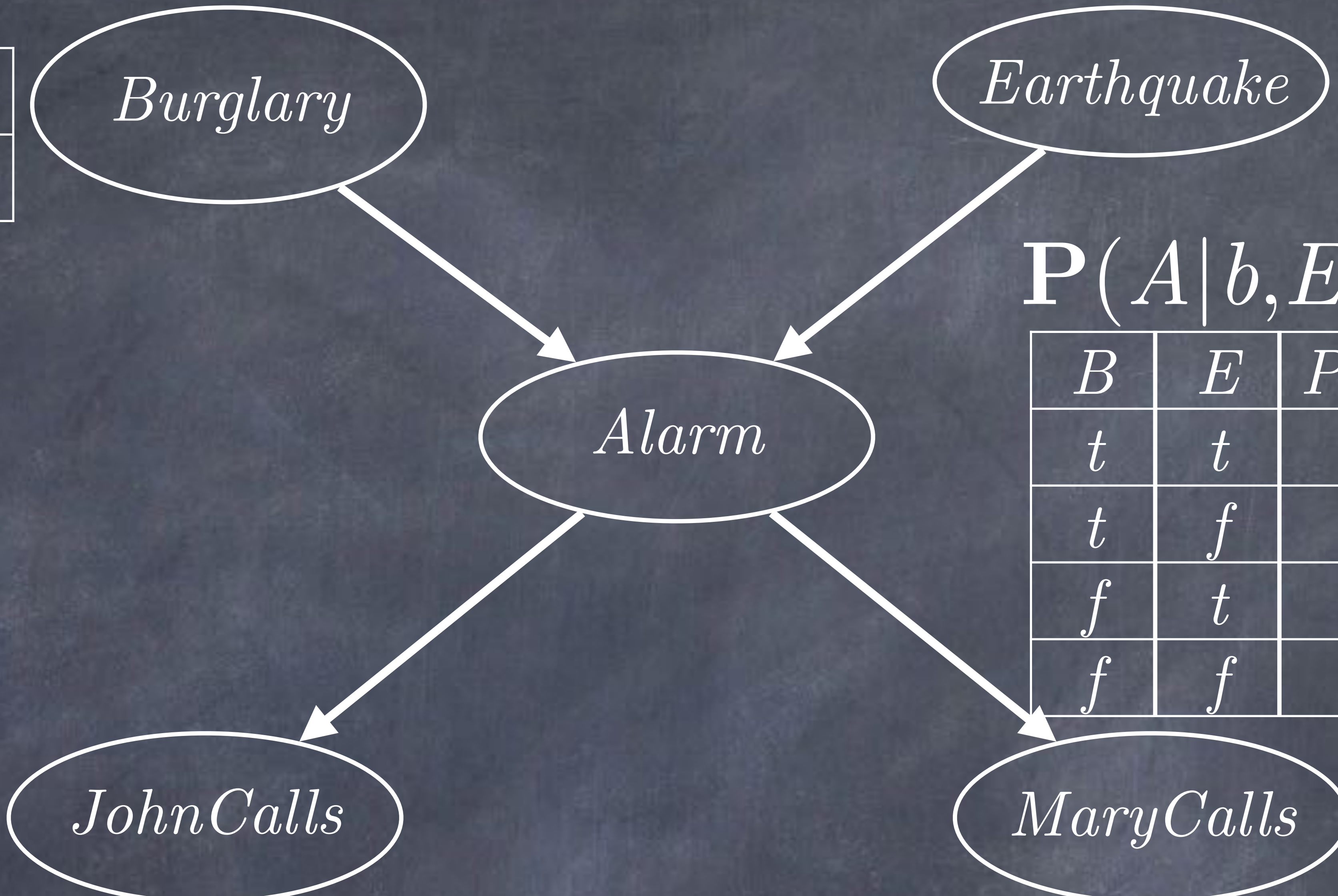
$$
\begin{aligned}
\mathbf{P}(b, j, m) = \alpha\, & \mathbf{P}(b)\,\mathbf{P}(e)\,\mathbf{P}(a \mid b, e)\,\mathbf{P}(j \mid a)\,\mathbf{P}(m \mid a) + \\
& \mathbf{P}(b)\,\mathbf{P}(e)\,\mathbf{P}(\neg a \mid b, e)\,\mathbf{P}(j \mid \neg a)\,\mathbf{P}(m \mid \neg a) + \\
& \mathbf{P}(b)\,\mathbf{P}(\neg e)\,\mathbf{P}(a \mid b, \neg e)\,\mathbf{P}(j \mid a)\,\mathbf{P}(m \mid a) + \\
& \mathbf{P}(b)\,\mathbf{P}(\neg e)\,\mathbf{P}(\neg a \mid b, \neg e)\,\mathbf{P}(j \mid \neg a)\,\mathbf{P}(m \mid \neg a)
\end{aligned}
$$

$\mathbf{P}(b)$

$P(b)$
0.001

$\mathbf{P}(E)$

$P(e)$
0.002



$\mathbf{P}(A|b,E)$

B	E	$P(a\|B,E)$
t	t	0.95
t	f	0.94
f	t	0.29
f	f	0.001

$\mathbf{P}(j|A)$

A	$P(j\|A)$
t	0.9
f	0.05

$\mathbf{P}(m|A)$

A	$P(m\|A)$
t	0.7
f	0.01

$$\mathbf{P}(b \mid j, m) = \alpha \sum_e \sum_a P(b)P(e)P(a|b,e)P(j|a)P(m|a)$$

Factored full joint dist.

Marginalization

$$\mathbf{P}(b \mid j, m) = \alpha \mathbf{P}(B, j, m) = \alpha \sum_e \sum_a \mathbf{P}(B, j, m, e, a)$$

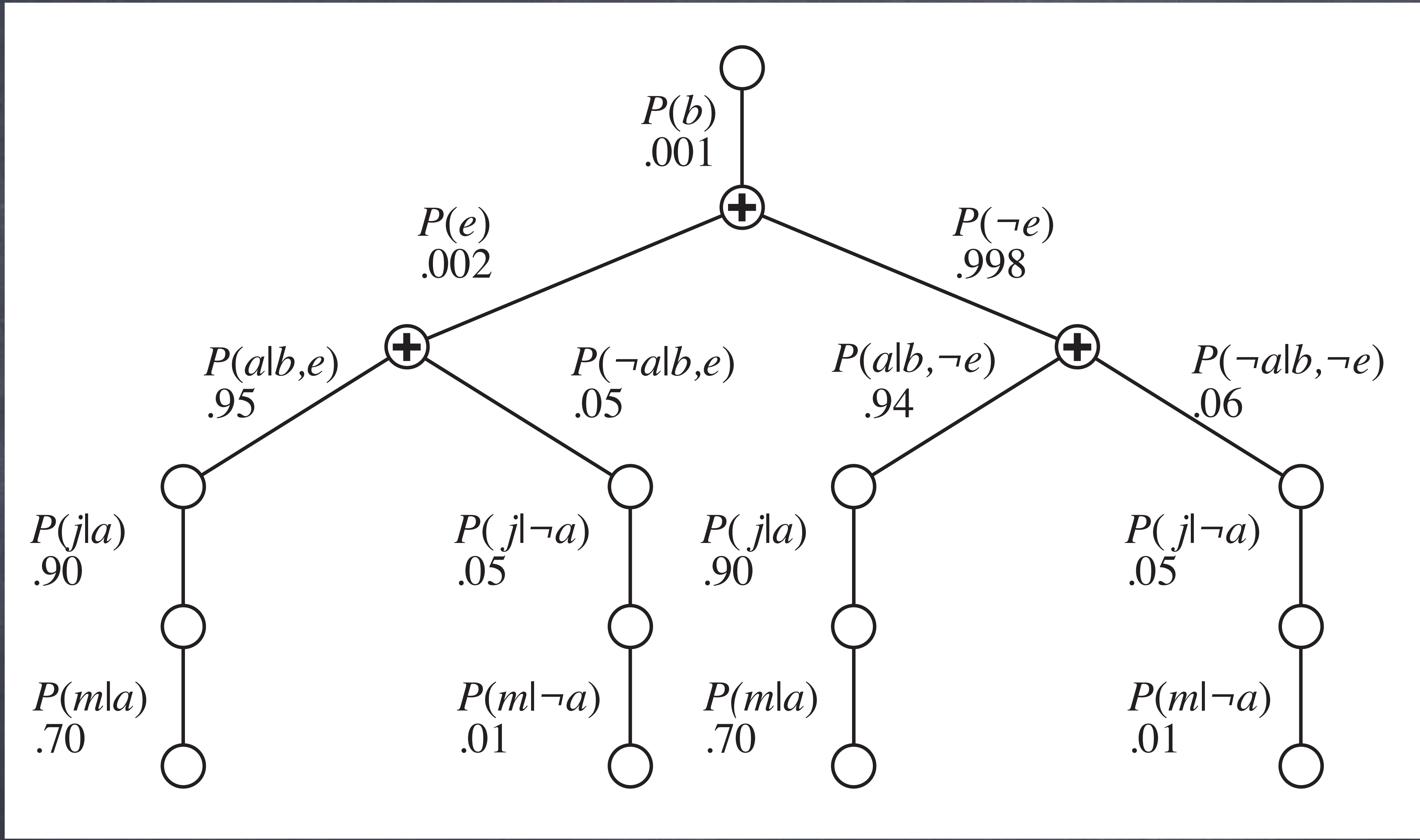
$$P(b \mid j, m) = \alpha \sum_e \sum_a P(b) P(e) P(a \mid b, e) P(j \mid a) P(m \mid a)$$

$$O(n2^n)$$

$$\mathbf{P}(b \mid j, m) = \alpha \mathbf{P}(B, j, m) = \alpha \sum_e \sum_a \mathbf{P}(B, j, m, e, a)$$

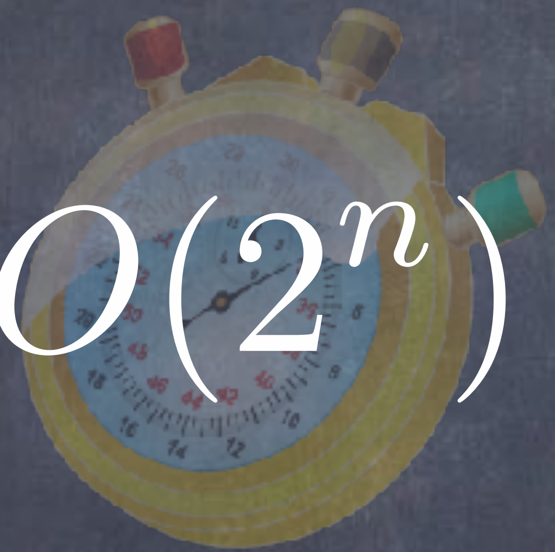
$$P(b \mid j, m) = \alpha \sum_e \sum_a P(b) P(e) P(a \mid b, e) P(j \mid a) P(m \mid a)$$

$$P(b \mid j, m) = \alpha P(b) \sum_e P(e) \sum_a P(a \mid b, e) P(j \mid a) P(m \mid a)$$



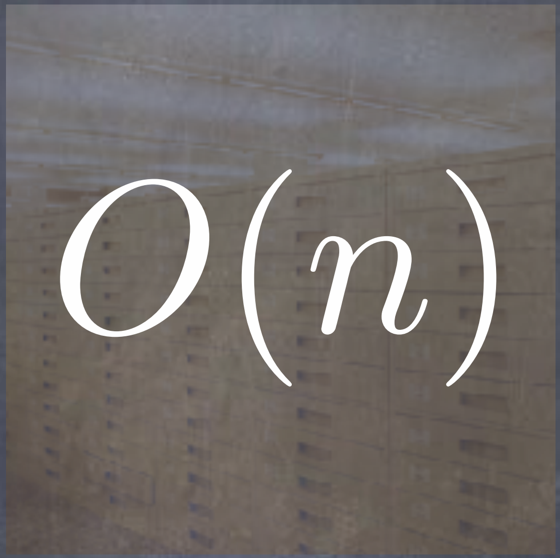
$$\mathbf{P}(B \mid j, m) = \alpha \langle 0.00059224, 0.0014919 \rangle \approx \langle 0.284, 0.716 \rangle$$

Exact Inference in BNs



$$O(2^n)$$

Time Complexity



$$O(n)$$

Space Complexity

Exact Inference in BNs

- Exact inference in BNs is NP-hard
- Can be shown to be as hard as computing the number of satisfying assignments of a propositional logic formula => #P-hard

Bayesian Networks Summary

- Independence assumptions make probabilistic inference easier
  - By factoring the joint distribution
- Bayesian Networks encode conditional independence assumptions among random variables
  - And store conditional probabilities

Inference in Bayesian Networks

$$
\begin{aligned}
\mathbf{P}(X \mid \mathbf{e}) &= \alpha\, \mathbf{P}(X, \mathbf{e}) = \alpha \sum_{\mathbf{y}} \mathbf{P}(X, \mathbf{e}, \mathbf{y}) \\
&= \alpha \sum_{\mathbf{y}} \prod_{i=1}^{n} P(X_i \mid parents(X_i))
\end{aligned}
$$

Inference in Bayesian Networks

- Exact inference with BNs is still hard
  - But we can do approximate inference efficiently (next time)
  - We can learn the conditional probabilities required to do inference from data (in a few weeks)

For Next Time:

AIMA 14.5; 14.7 fyi