Explanation Trees for Causal Bayesian Networks

Ulf H. Nielsen,[1] Jean-Philippe Pellet,[1,2] and André Elisseeff[1]

July 11th, 2008

[1] Data Analytics Group
IBM Zurich Research Lab

[2] Pattern Analysis & Machine Learning Group
ETH Zurich

Outline

- **What?**

  Evidence Explanations in Bayesian Networks

- **How?**

  Building Causal Explanation Trees

- **So What?**

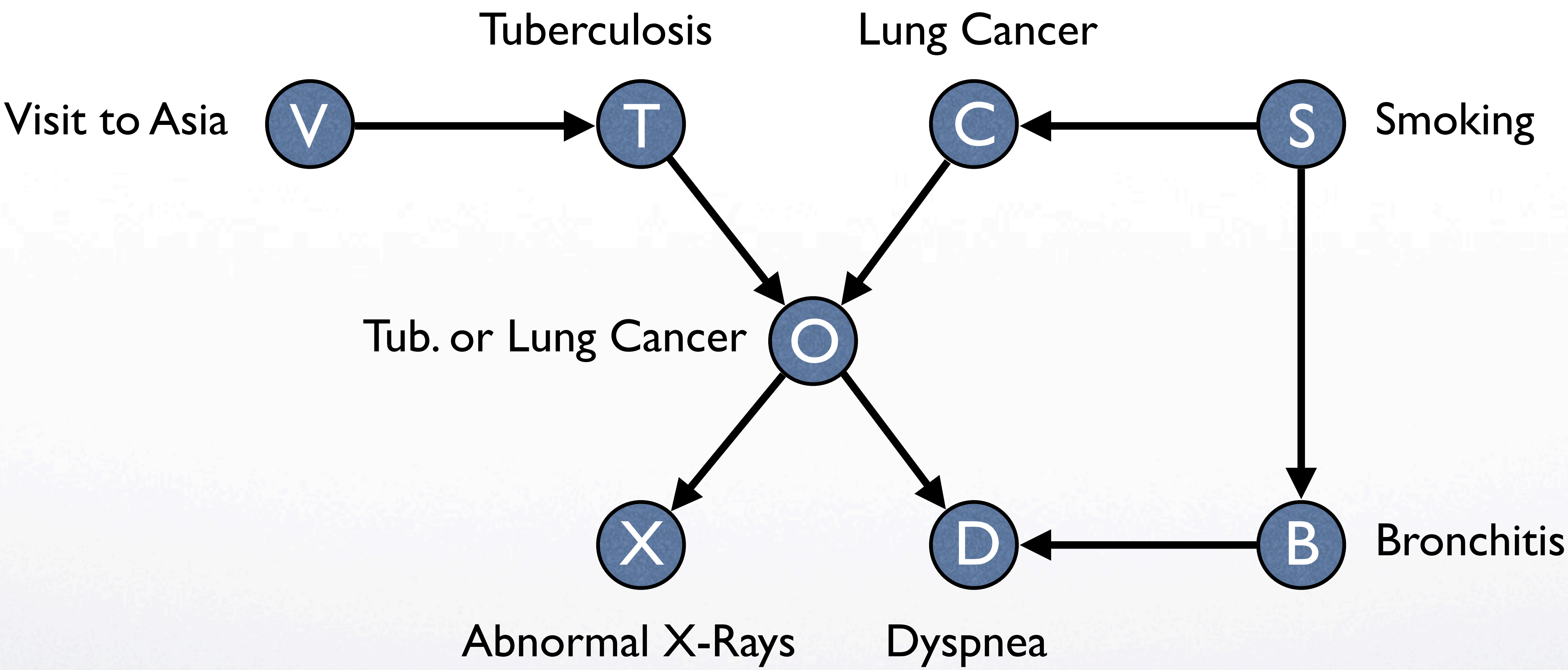
  Comparisons

I

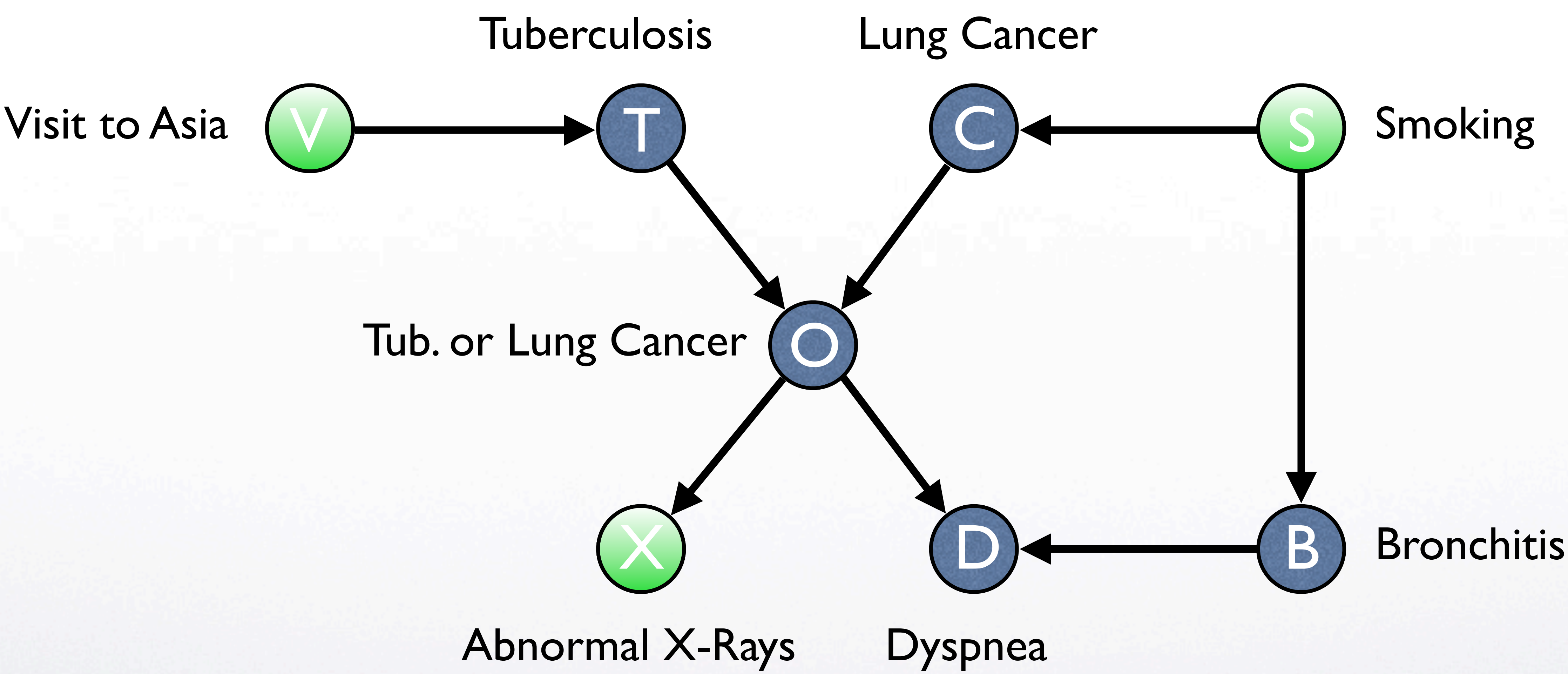
What?

Evidence Explanation in Bayesian Networks

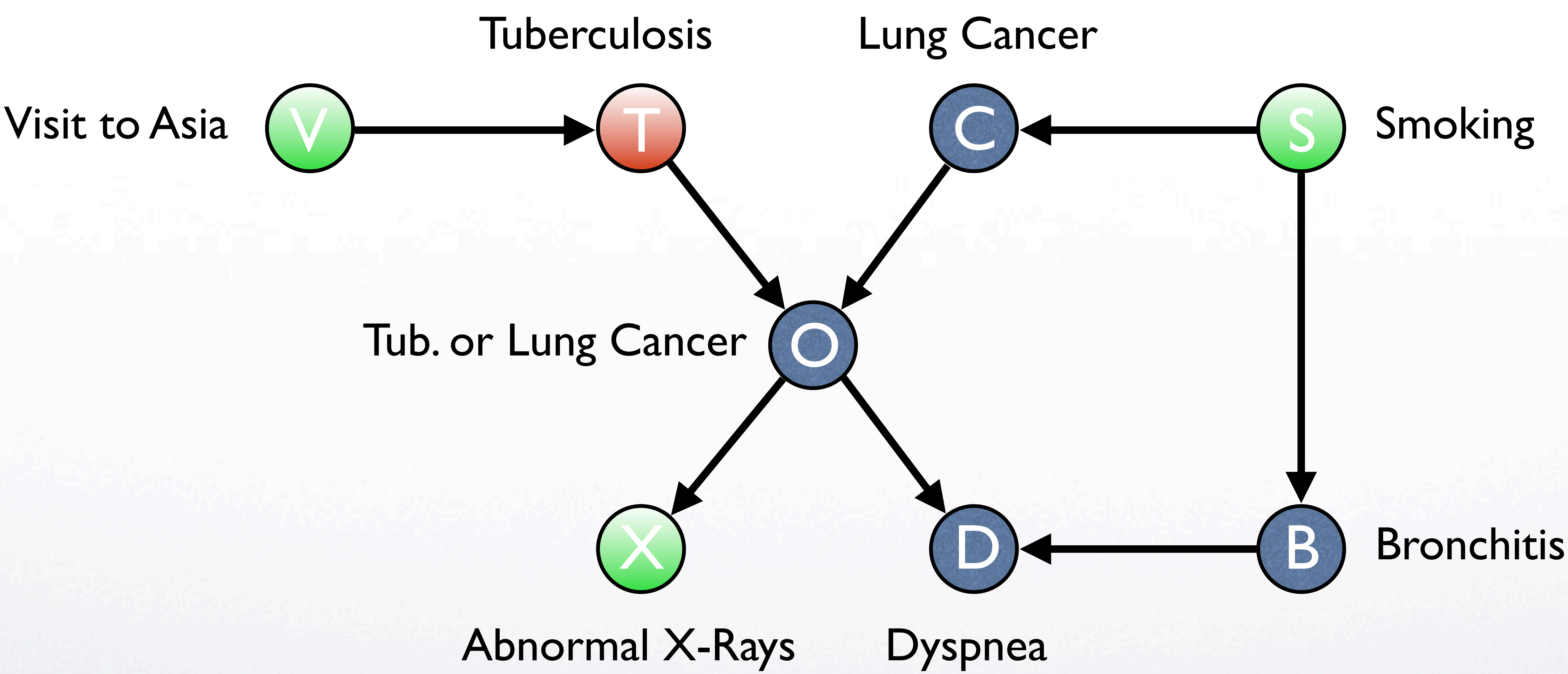
A Bayesian Network



A Bayesian Network



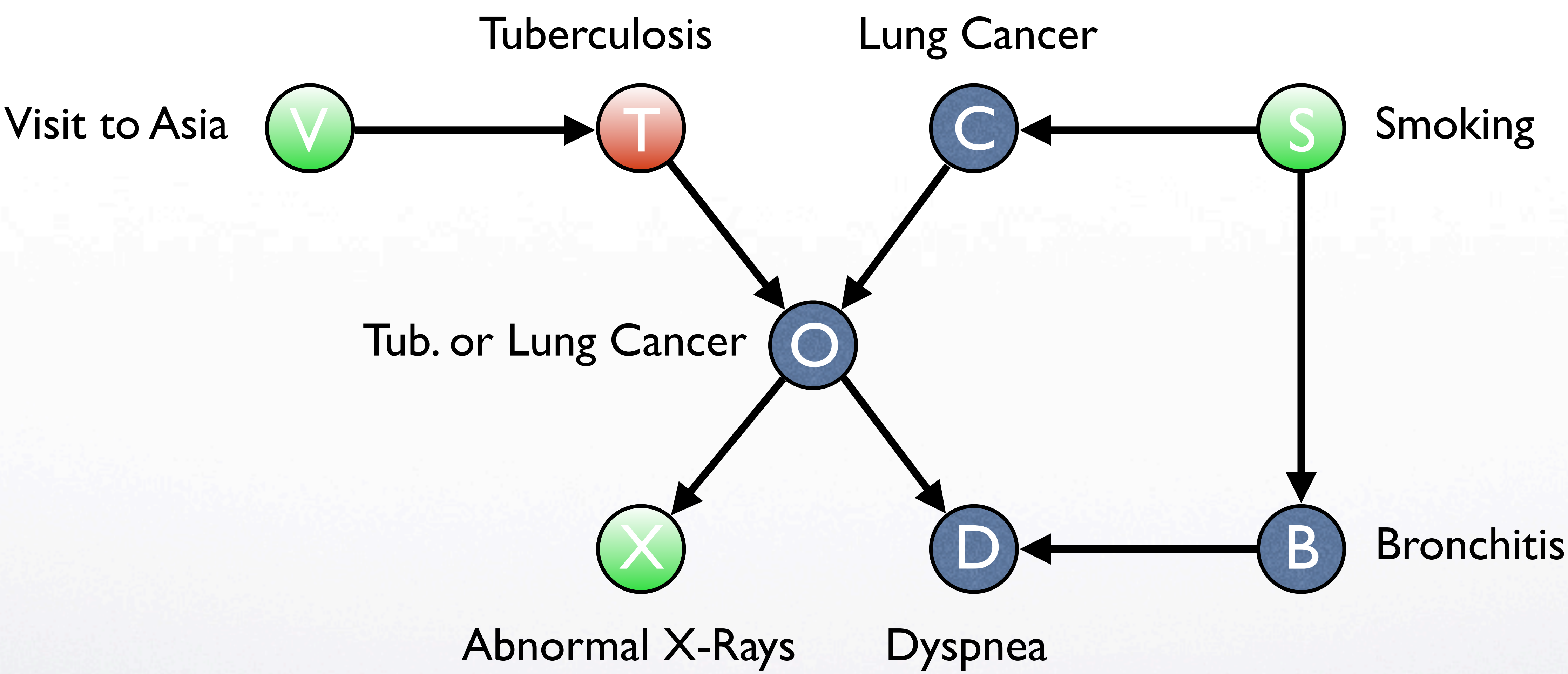
A Bayesian Network



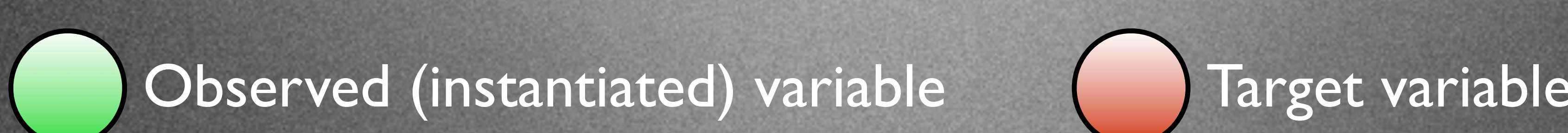
Observed (instantiated) variable

Target variable

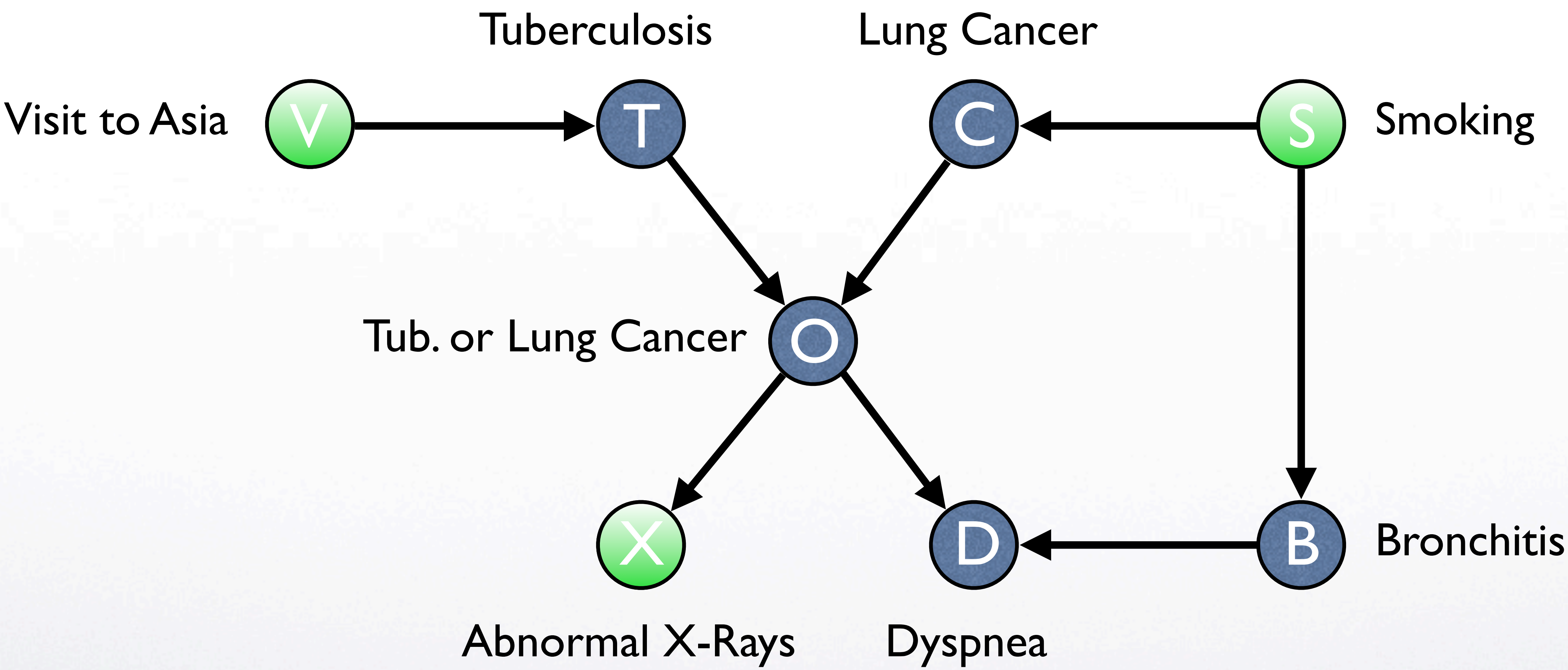
A Bayesian Network



Inference task: find P(*Target* | *Observed*) = P(*T* | *V, S, X*)

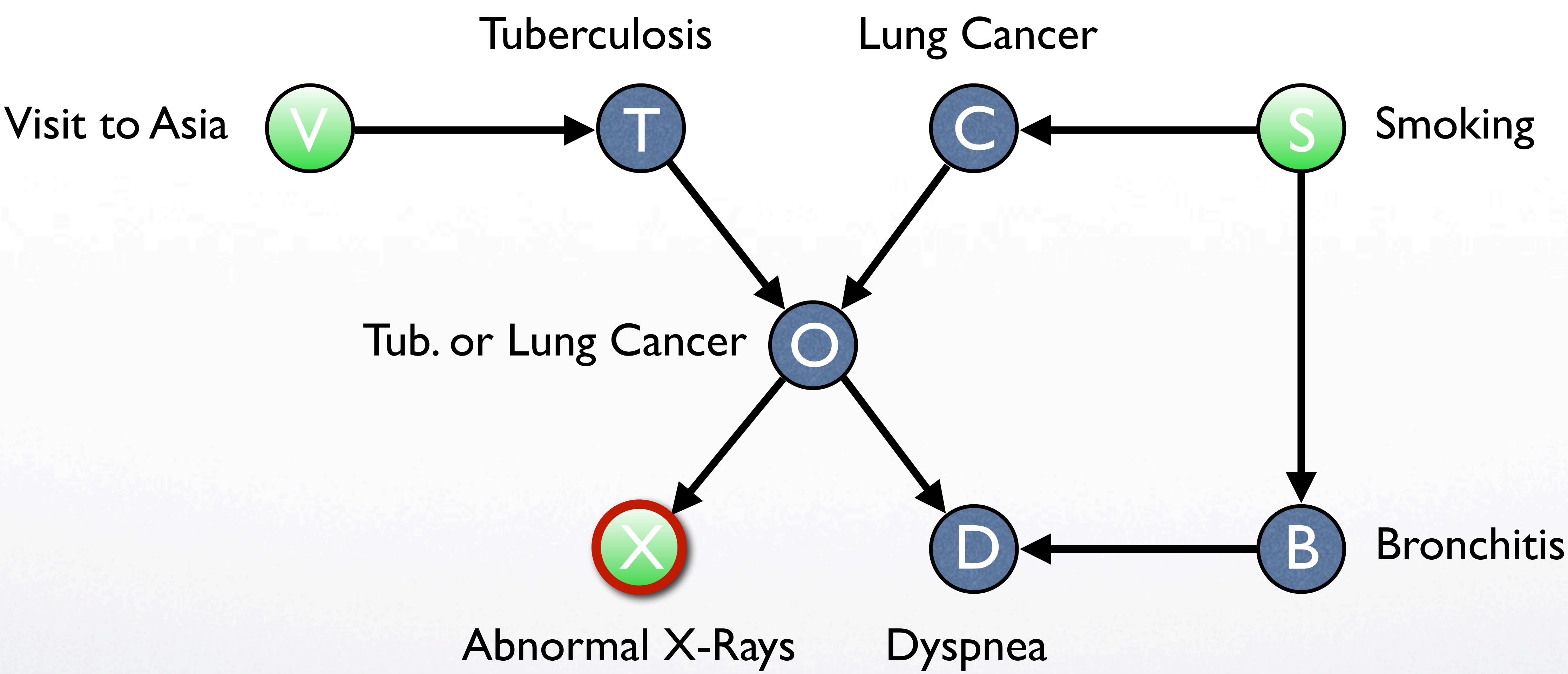


A Bayesian Network

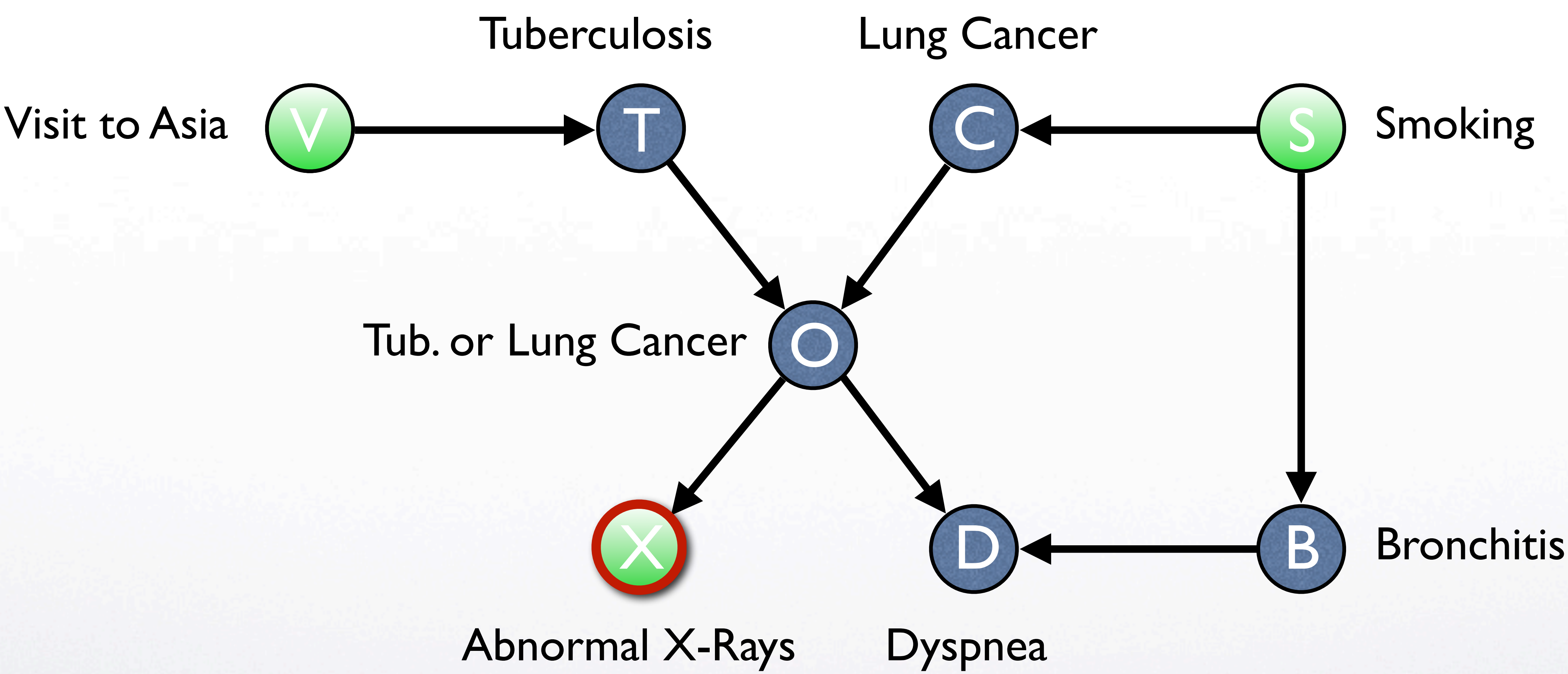


Observed (instantiated) variable

A Bayesian Network



A Bayesian Network



Evidence explanation task: why do we observe $X = x$?

Observed (instantiated) variable

Explanandum (observed)

Evidence Explanation

- Explanation = Set of variable assignments; e.g., "$Z = z$, $Y = y$ explain $X = x$." Call the explanation $\mathbf{H} = \mathbf{h}$
  - Which explanatory variables to choose?
  - Which assignments to choose?

- Interventionist desideratum: Intervening on the network according to the explanation favors the explanandum

$$P(X = x) < P\big(X = x \mid do(\mathbf{H} = \mathbf{h})\big)$$

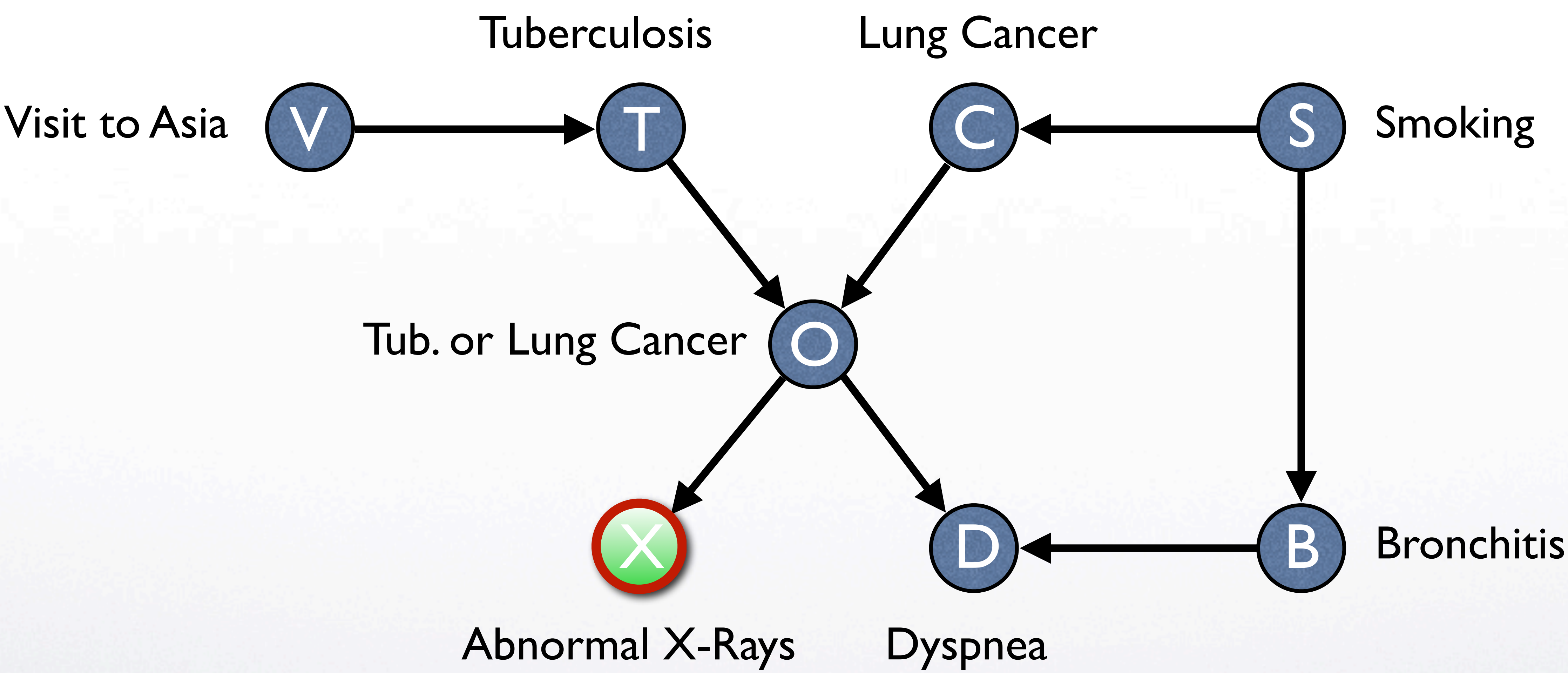
- To evaluate *do*-expression, we require a causal Bayesian network

Why Causality Here?

- Natural to explain effects with causes

Why Causality Here?

- Natural to explain effects with causes

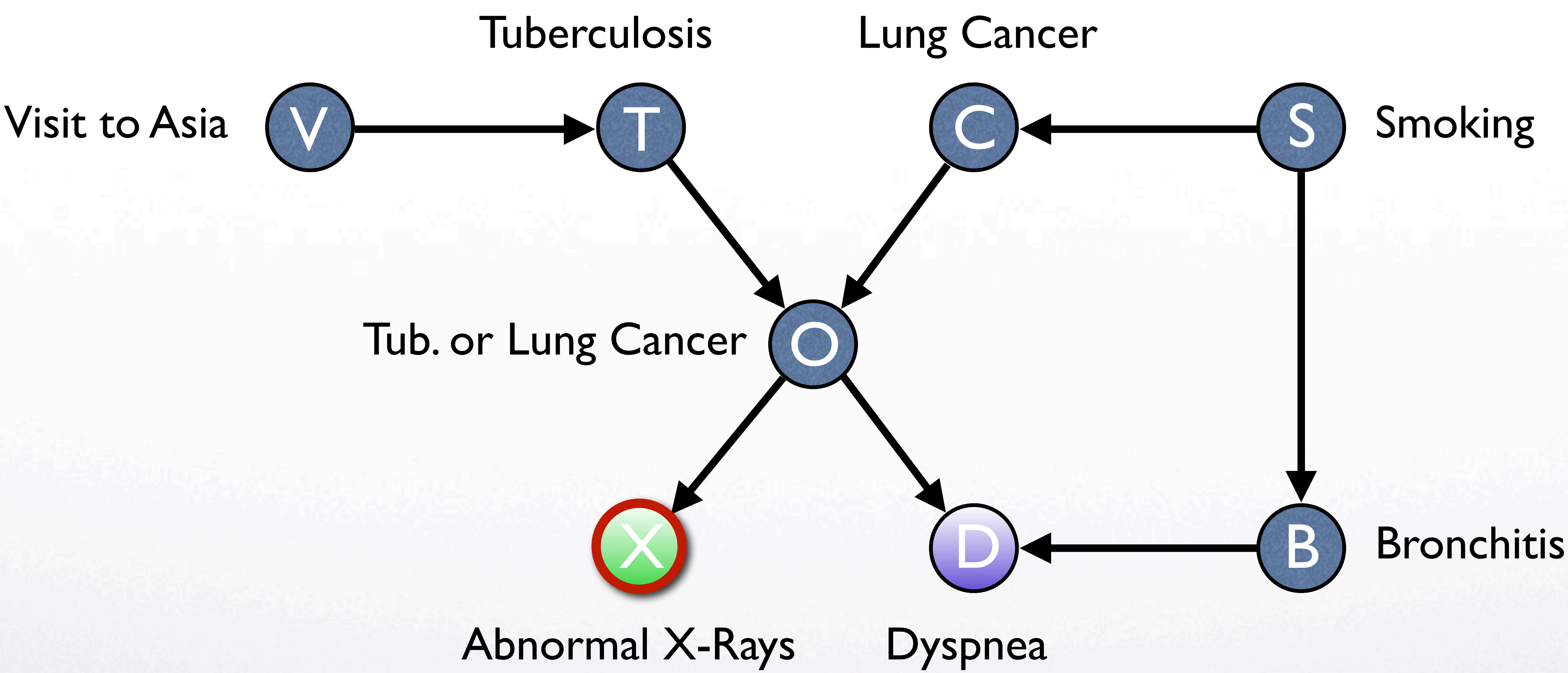


- Leads to intervention rules: “A ‘good’ way to obtain $X = x$ is to intervene on the system, setting **H** = **h**.”

Explanandum

Why Causality Here?

- Natural to explain effects with causes



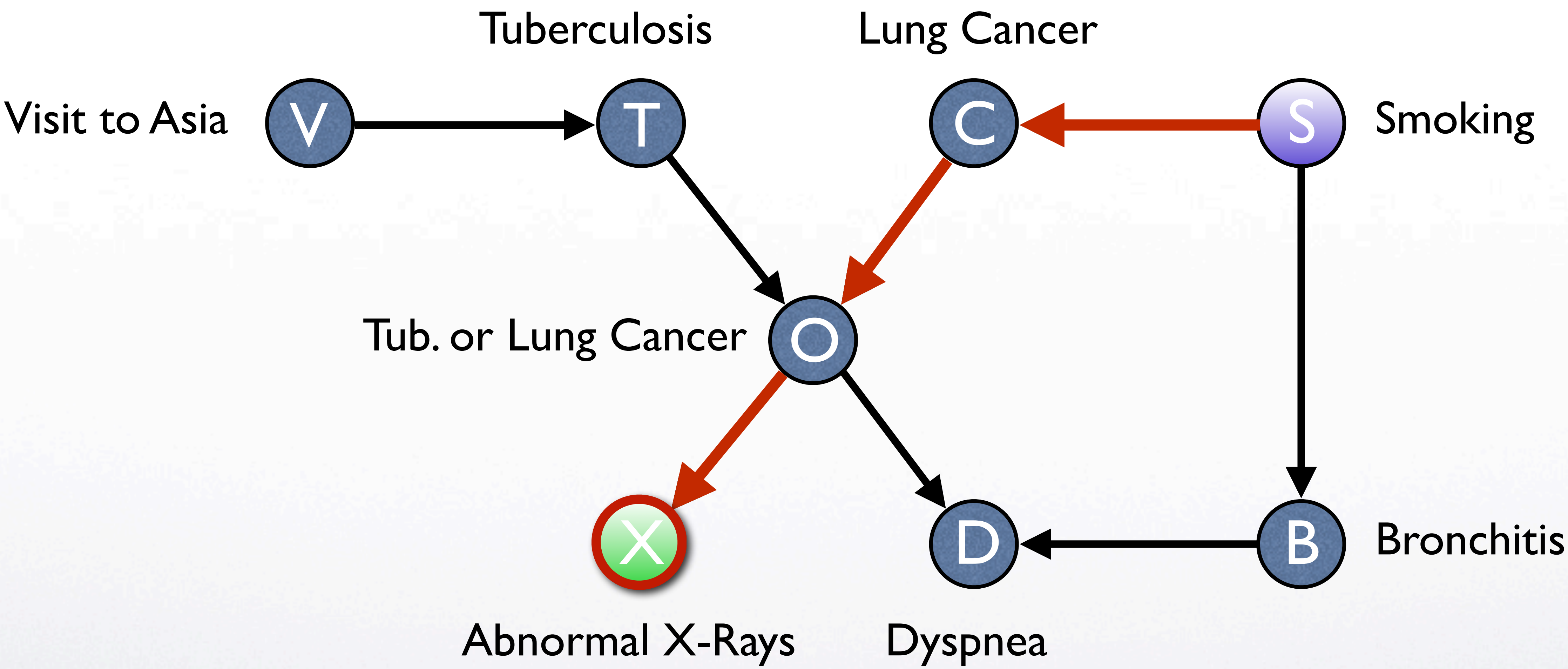
- Leads to intervention rules: “A ‘good’ way to obtain $X = x$ is to intervene on the system, setting $\mathbf{H} = \mathbf{h}$.”

Explanandum

Potential explanatory variable

Why Causality Here?

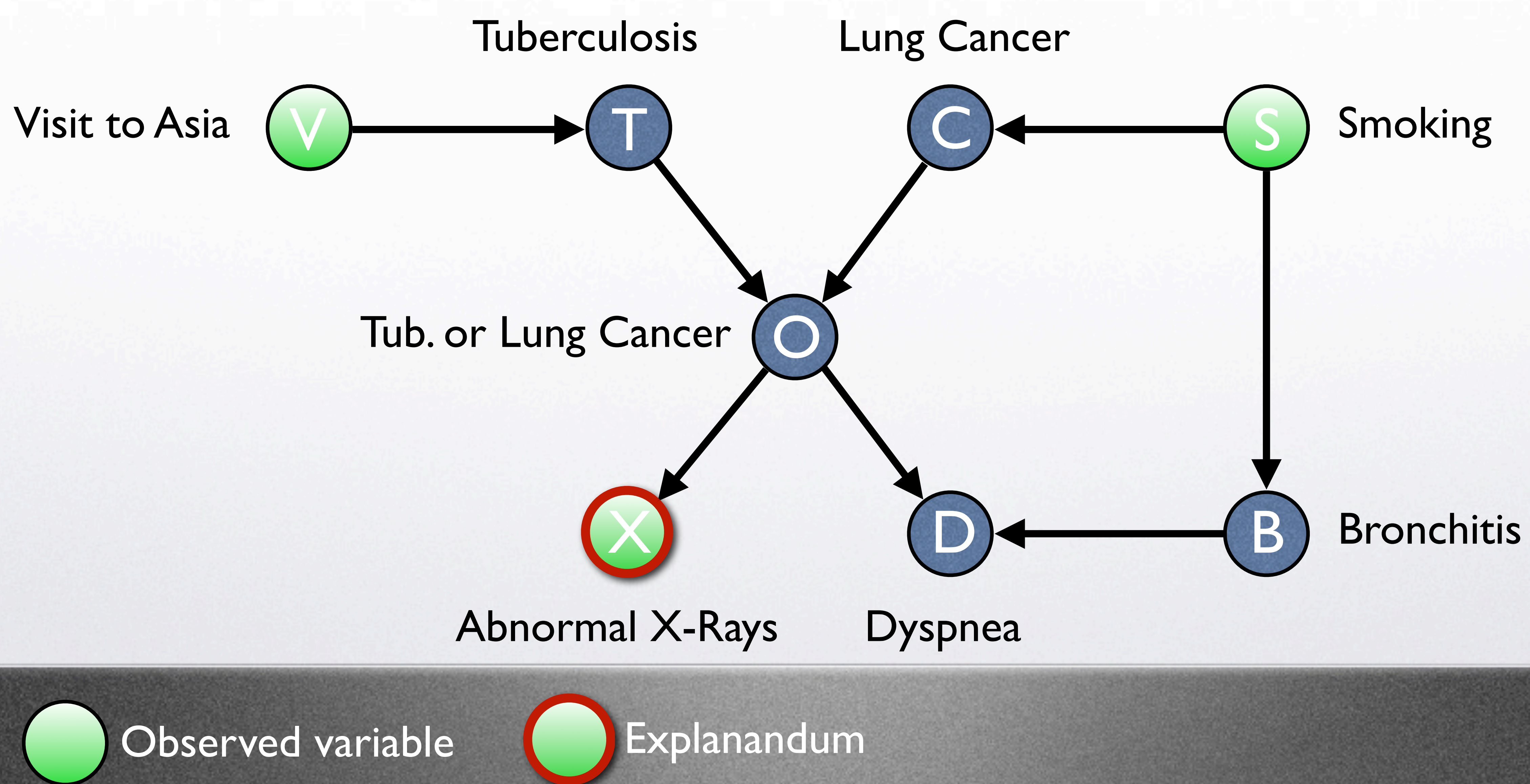
- Natural to explain effects with causes



- Leads to intervention rules: "A 'good' way to obtain $X = x$ is to intervene on the system, setting **H** = **h**."

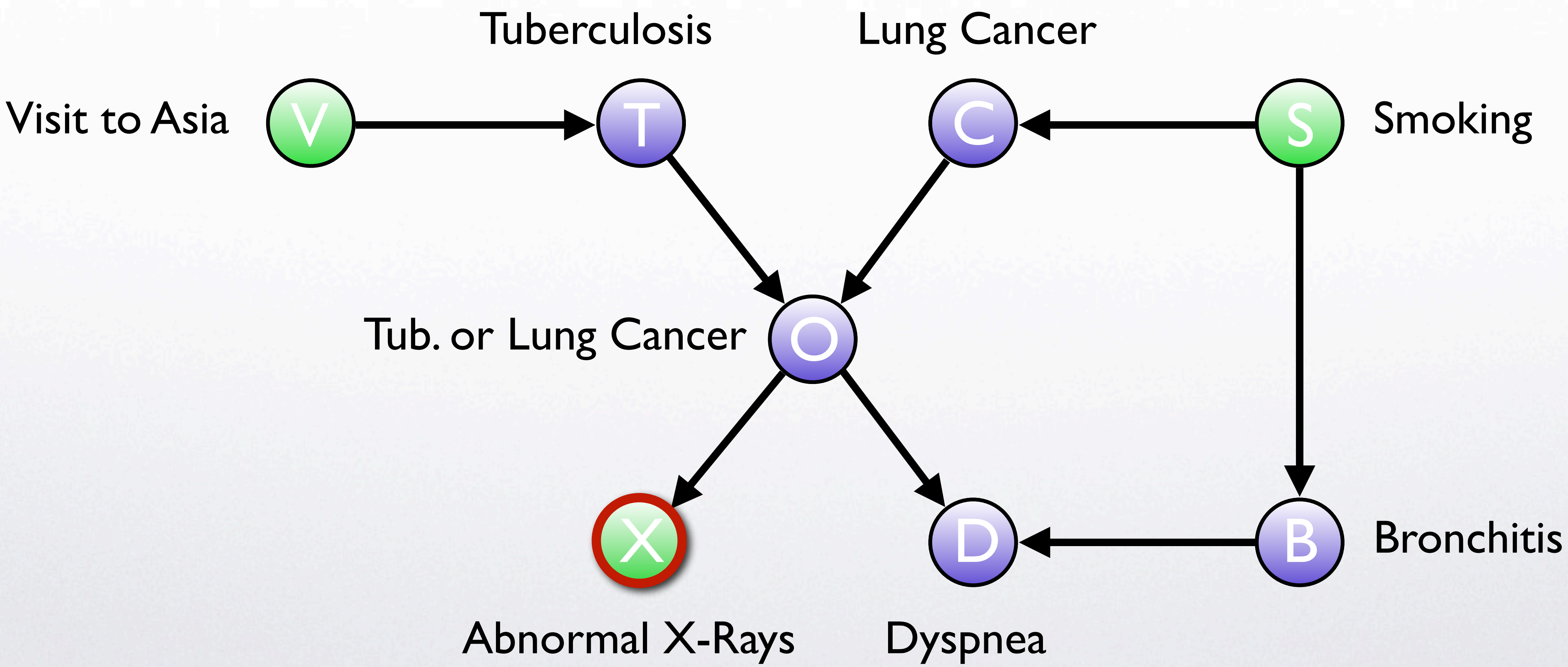
Other Approaches to Explanations

- Pearl's (1988) **Most Probable Explanation** (MPE):
  Find configuration that maximizes P(all unobserved | all observed)
  - Easy to compute
  - No distinction explanandum/observation; long, sensitive explanations



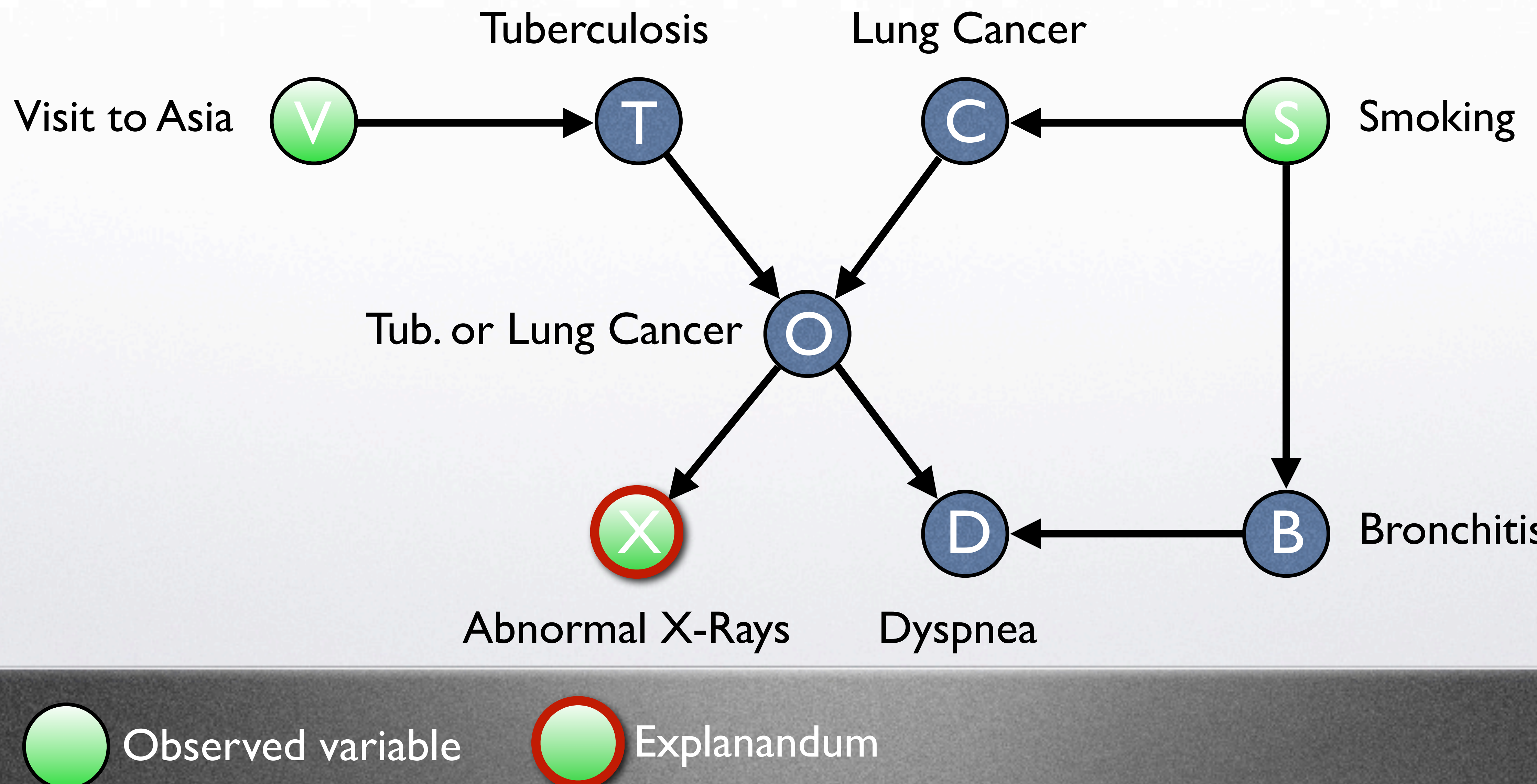
Other Approaches to Explanations

- Pearl's (1988) **Most Probable Explanation** (MPE):
  Find configuration that maximizes P(all unobserved | all observed)
    - Easy to compute
    - No distinction explanandum/observation; long, sensitive explanations



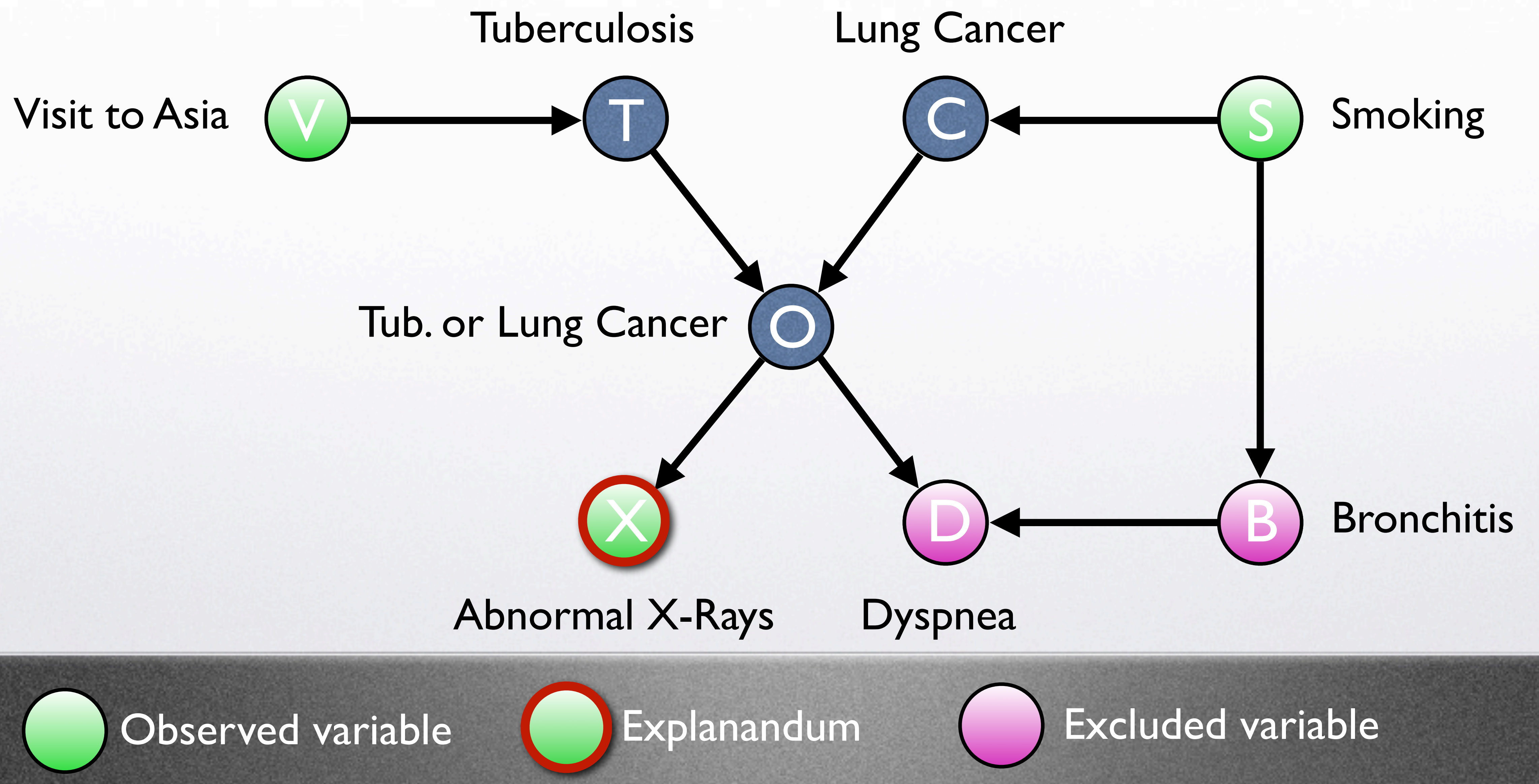
Other Approaches to Explanations

- **Partial Abduction**:
  Maximize $\sum_{excluded}$ P(unobserved, excluded | all observed)
  - More targeted explanations. Not as easy to compute as MPE
  - No distinction explanandum/observation; how to choose the excluded variables?



Other Approaches to Explanations

- **Partial Abduction**:
  Maximize $\sum_{excluded}$ P(unobserved, excluded | all observed)
  - More targeted explanations. Not as easy to compute as MPE
  - No distinction explanandum/observation; how to choose the excluded variables?



Other Approaches to Explanations

- **Partial Abduction** (continued)

  Maximize $\sum_{\text{excluded}}$ P(unobserved, excluded | all observed)

- Yuan and Lu (2007): subset search for the excluded variables; explanations ranked by Bayes' Factor

$$\mathrm{BF}(\mathbf{h}) = \frac{p(\mathbf{H} = \mathbf{h} \mid \text{observed})}{1 - p(\mathbf{H} = \mathbf{h} \mid \text{observed})}$$

  - Concise explanations
  - Scalability concerns, no distinction explanandum/observation

- **MPE/Partial abduction:** Use $p(\mathbf{H} = \mathbf{h} \mid \text{observed})$ criterion

  - We think $p(\text{explanandum} \mid \mathbf{H} = \mathbf{h})$ is more intuitive; $p(\text{explanandum} \mid do(\mathbf{H} = \mathbf{h}))$ even more so

II

How?

Building Causal Explanation Trees

Causal Relevance

- **Causal Information Flow** (Ay & Polani, 2006): causal counterpart of mutual information
- Conditional mutual information:

$$I(X,Y \mid \mathbf{Z} = \mathbf{z}) = \sum_x P(x|\mathbf{z}) \sum_y P(y|x,\mathbf{z}) \log \frac{P(y|x,\mathbf{z})}{P(y|\mathbf{z})}$$

- Causal information flow:

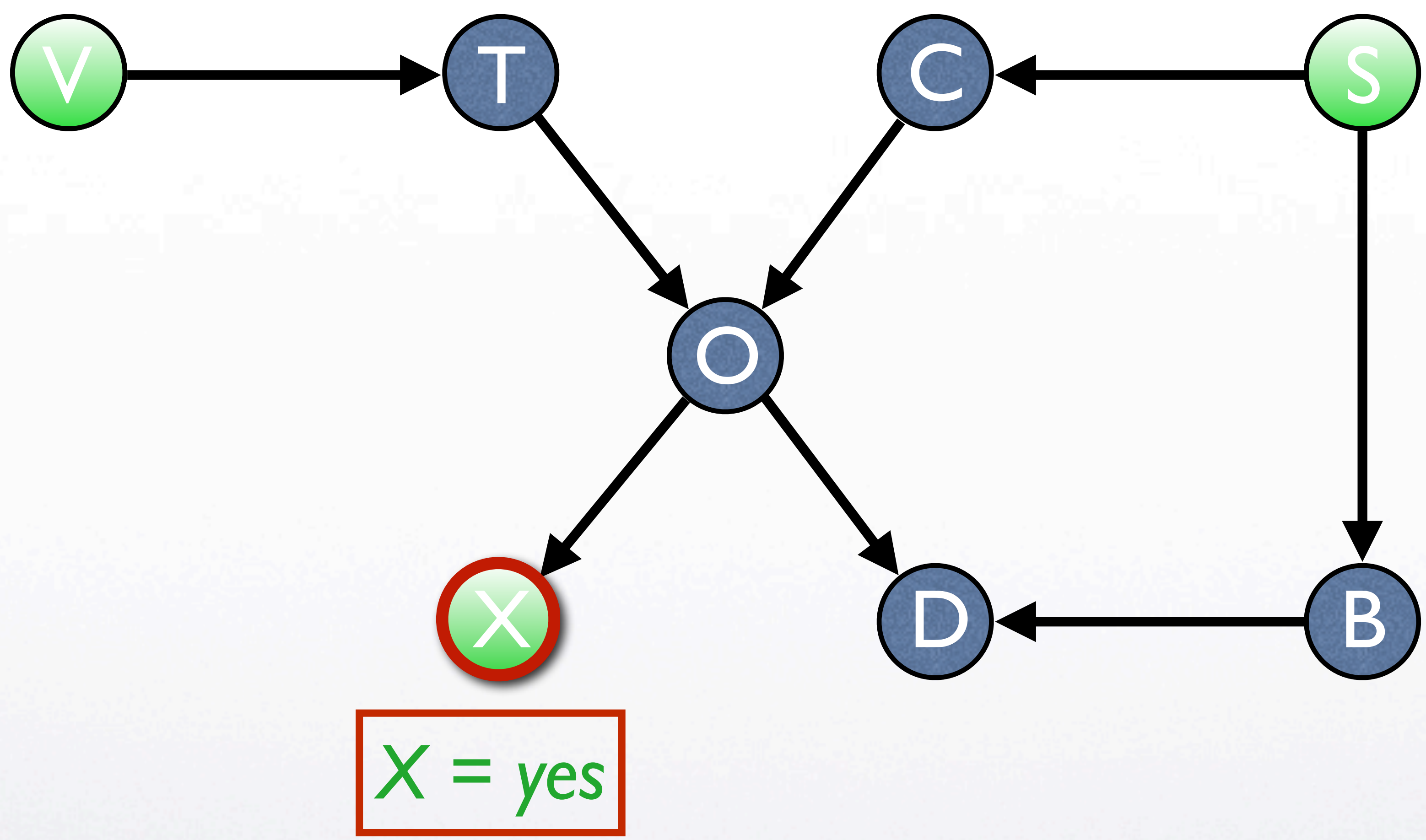
$$I(X \rightarrow Y \mid do(\mathbf{Z} = \mathbf{z})) = \sum_x P(x|do(\mathbf{z})) \sum_y P(y|do(x,\mathbf{z})) \log \frac{P(y|do(x,\mathbf{z}))}{P^*(y|do(\mathbf{z}))}$$

where $P^*(y|do(\mathbf{z})) = \sum_{x'} P(x'|do(\mathbf{z}))\ P(y|do(x',\mathbf{z}))$

- $I(X \rightarrow Y \mid do(\mathbf{Z} = \mathbf{z})) \geq 0$ iff X is an ancestor of Y and there is a directed path from X to Y not going through any node in $\mathbf{Z}$.

Causal Evidence Explanation

Context: we observe $V = yes$, $S = no$, $X = yes$.
We want to explain $X = yes$.



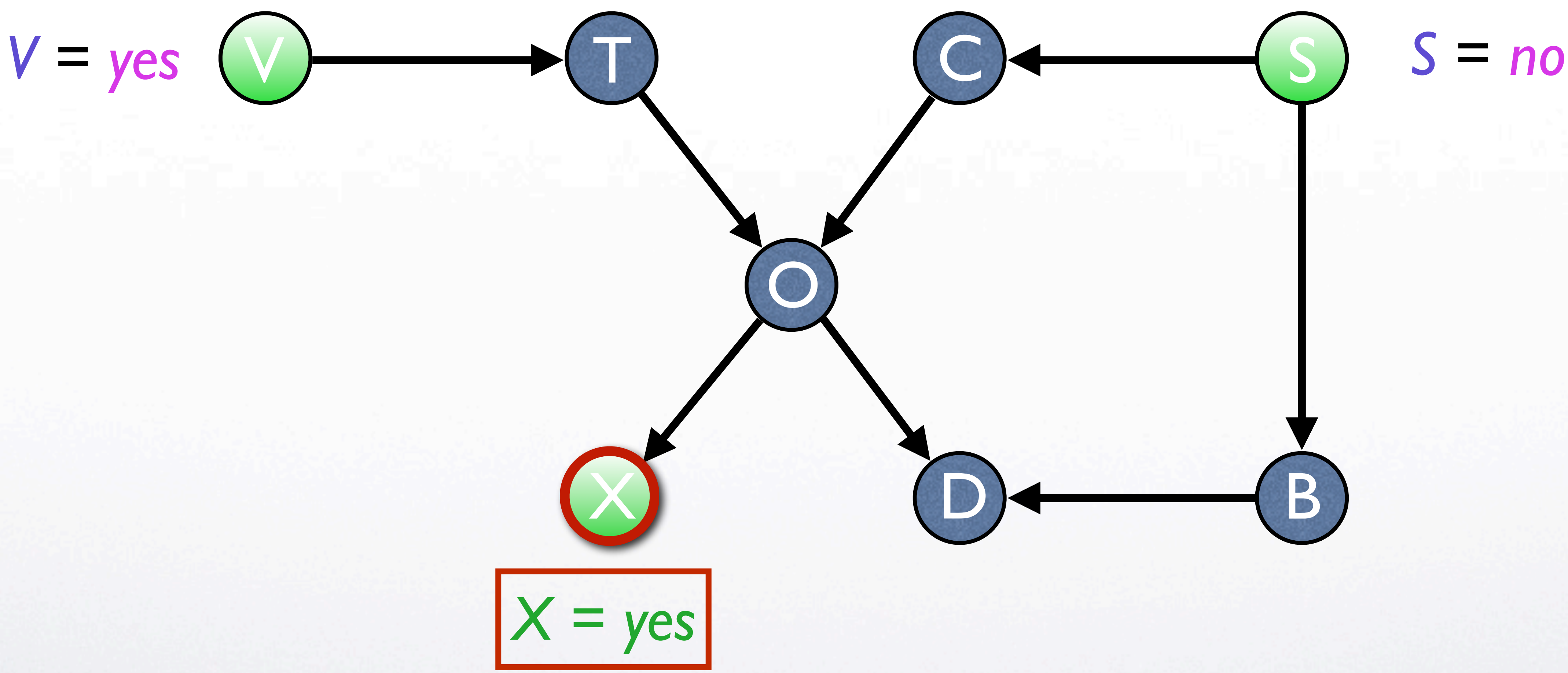
Find: explanation = Set of variable assignments.

Observed variable

Explanandum

Causal Evidence Explanation

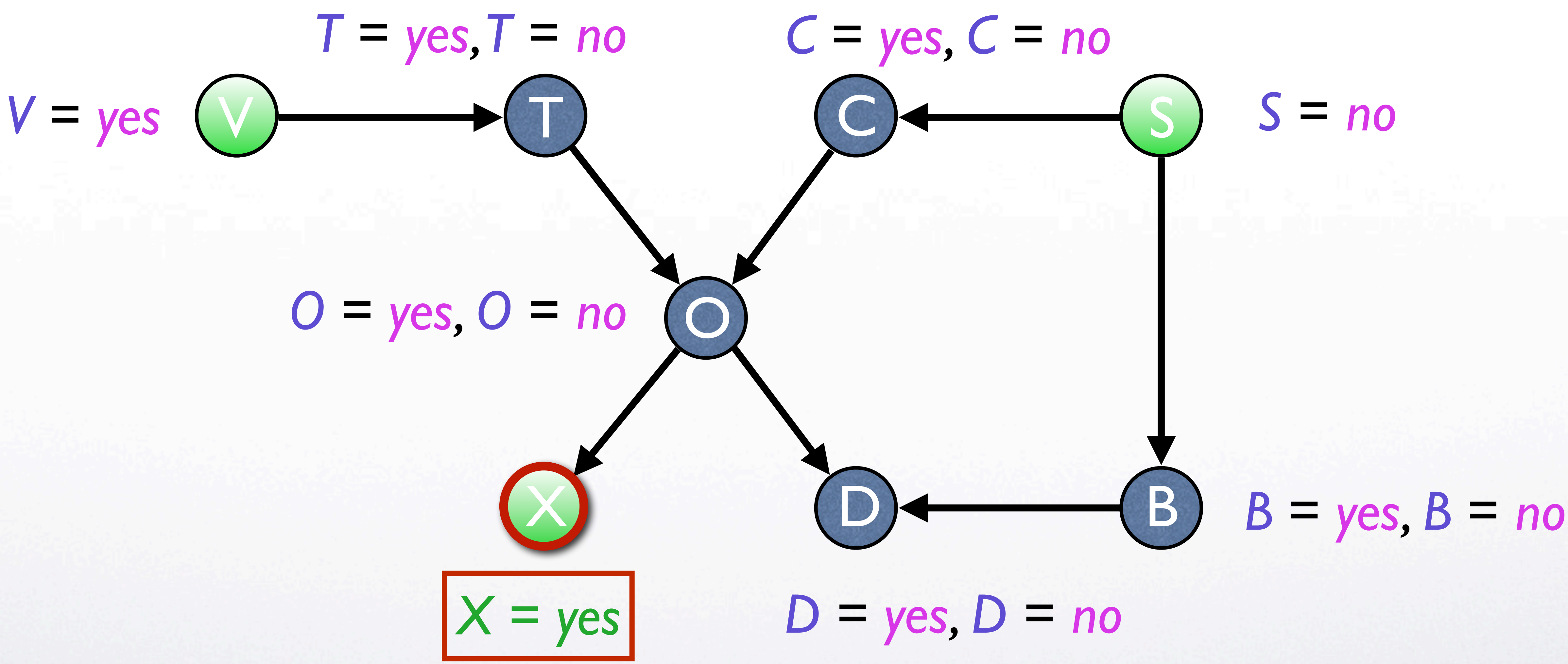
Context: we observe *V = yes*, *S = no*, *X = yes*.
We want to explain *X = yes*.



Find: explanation = Set of variable assignments.

Causal Evidence Explanation

Context: we observe V = *yes*, S = *no*, X = *yes*.
We want to explain X = *yes*.



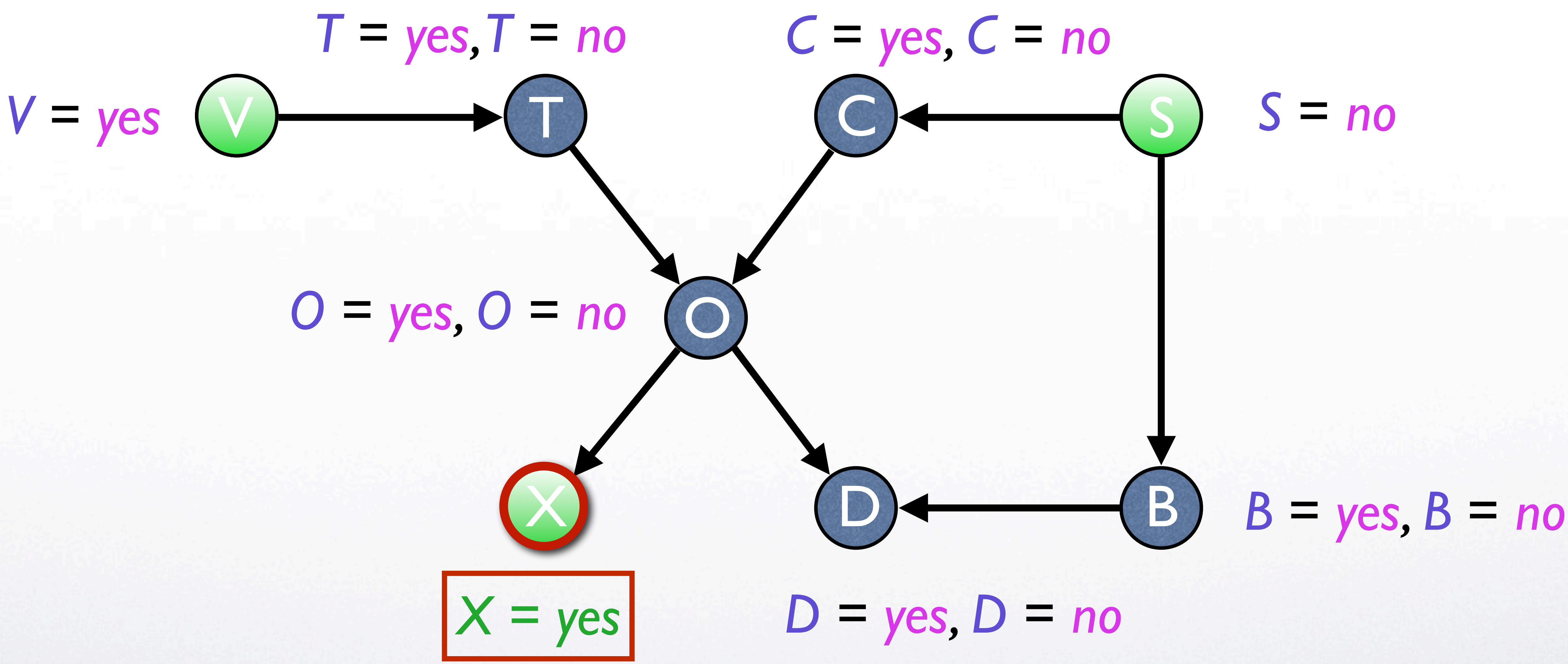
Find: explanation = Set of variable assignments.

Observed variable

Explanandum

Causal Evidence Explanation

Context: we observe $V = yes$, $S = no$, $X = yes$.
We want to explain $X = yes$.



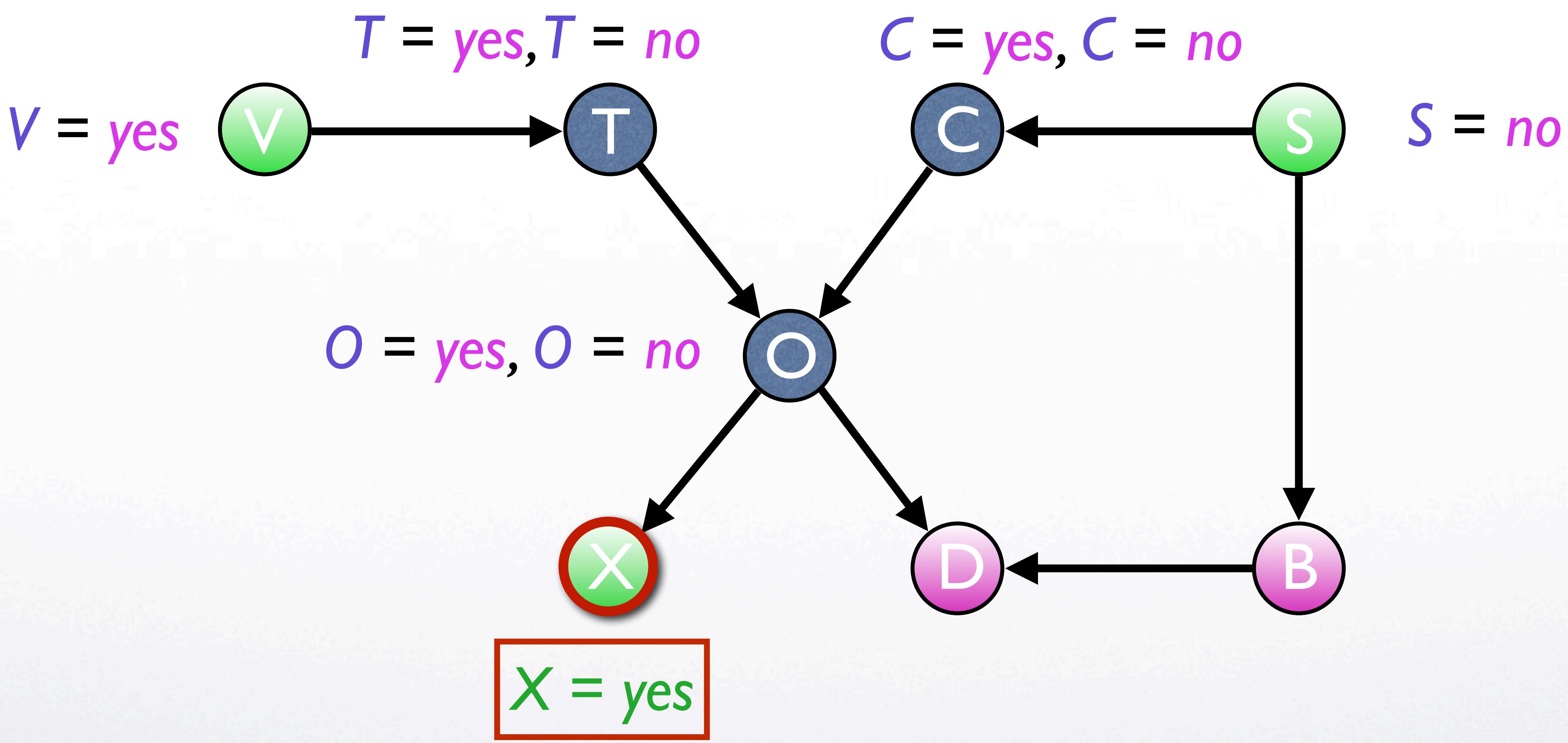
Find: explanation = Set of variable assignments.

Exhaustive Search: $2^2 \cdot 3^5 = 972$ inferences of type $P(X = x \mid do(Explanation))$

Observed variable
Explanandum

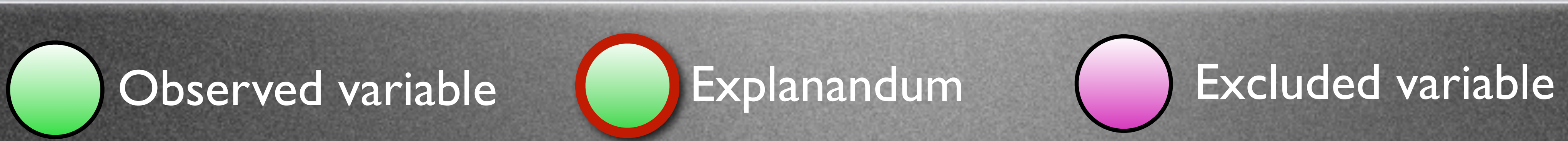
Causal Evidence Explanation

Context: we observe $V = yes$, $S = no$, $X = yes$.
We want to explain $X = yes$.



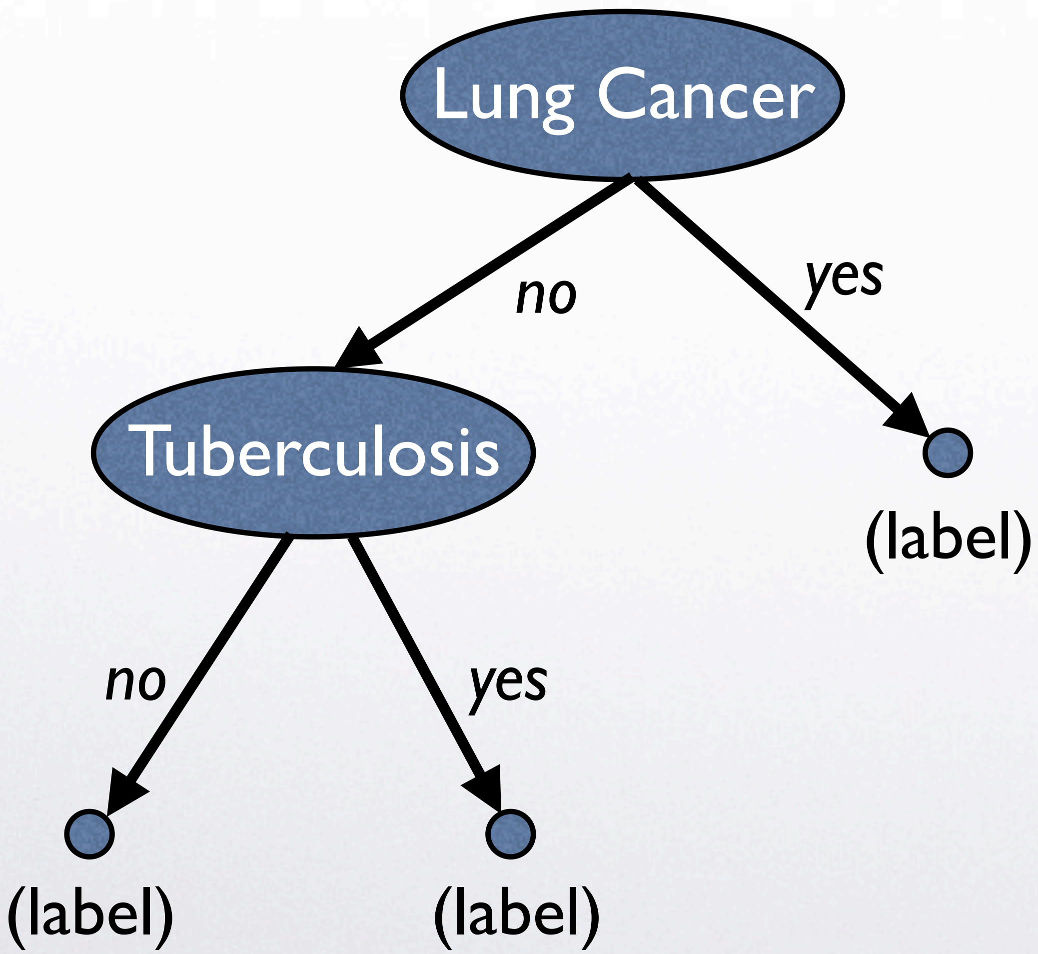
Find: explanation = Set of variable assignments.

Exhaustive Search: $2^2 \cdot 3^5 = 972$ inferences of type $P(X = x \mid do(Explanation))$



Explanation Trees

- Introduced by Flores (2005) in the context of partial abduction
- Allows compact representation of several explanations: explanation is a path from the root to a leaf



Represents:

Lung Cancer = yes

Lung Cancer = no, Tuberculosis = yes

Lung Cancer = no, Tuberculosis = no

Building the Explanation Tree

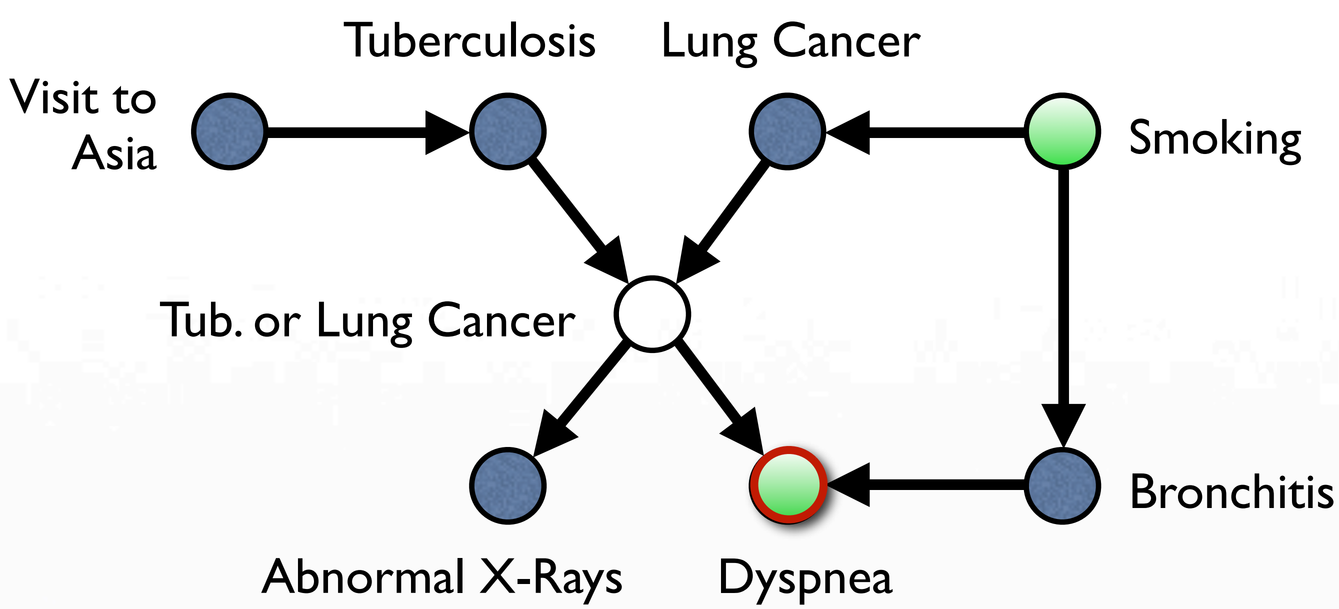
- Start with empty tree. Greedy selection: recursively select next node as variable X that maximizes

$$I(X \rightarrow \mathit{explanandum} \mid \mathit{observations}, do(\mathit{current\ path}))$$

- Add as outgoing edges the values for X
- Stopping criterion: minimal additional causal information flow

Building the Tree: Example

Explain: *Dyspnea = yes* | *Smoking = yes*



Observed variable

Explanandum

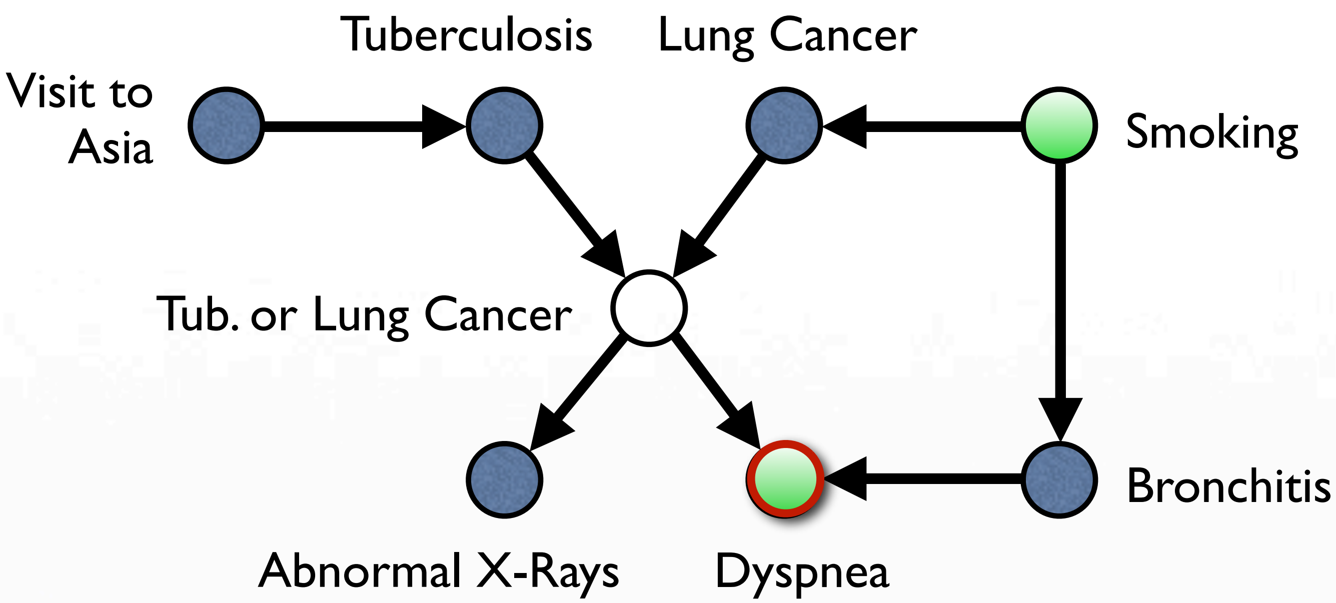
Candidate explanatory variable

Selected explanatory variable

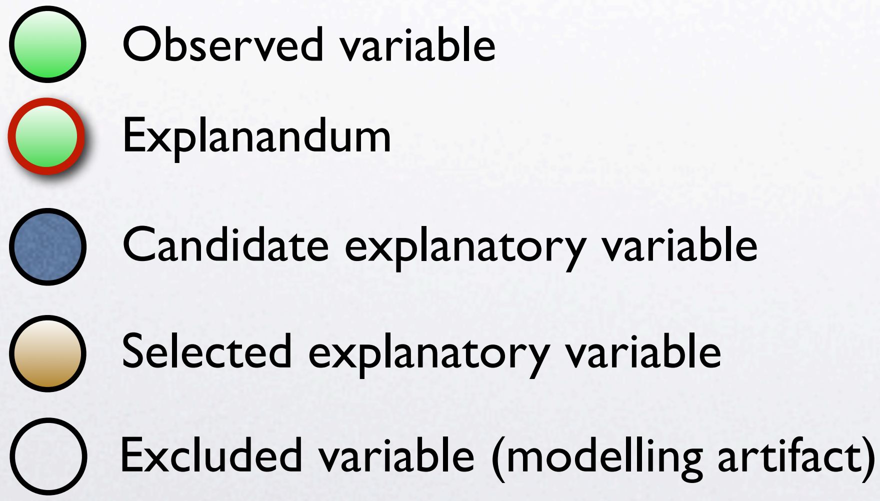
Excluded variable (modelling artifact)

Building the Tree: Example

Explain: *Dyspnea = yes* | *Smoking = yes*

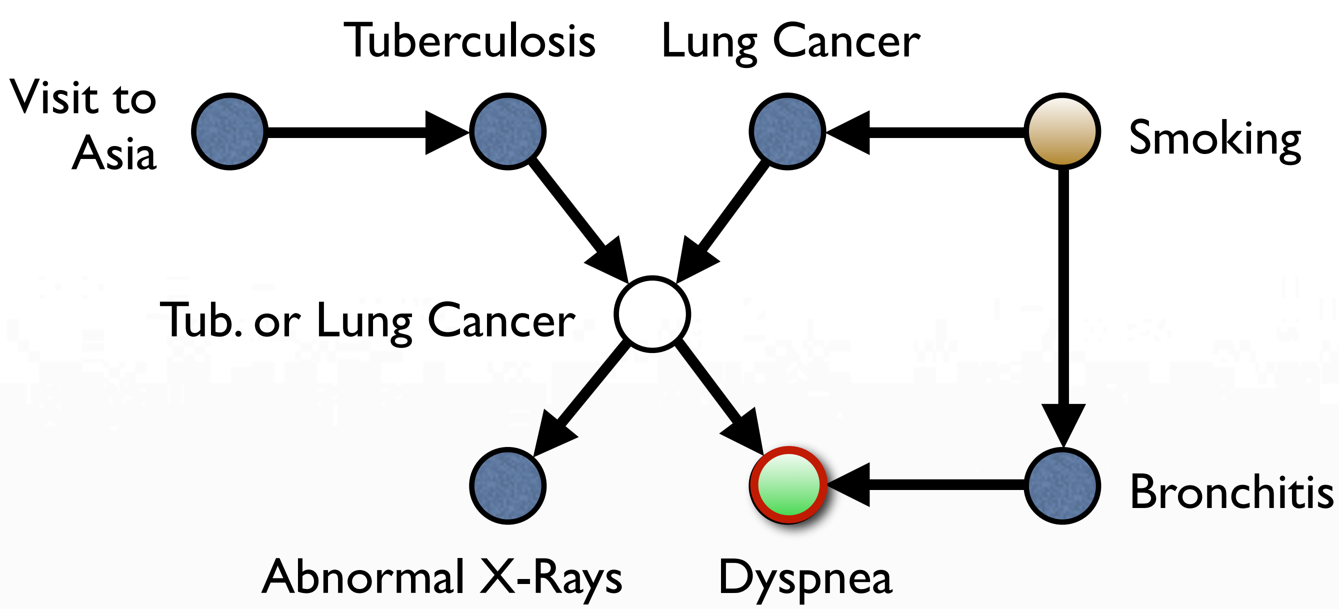


$\arg\max_X I(X \rightarrow D = yes \mid S = yes)$



Building the Tree: Example

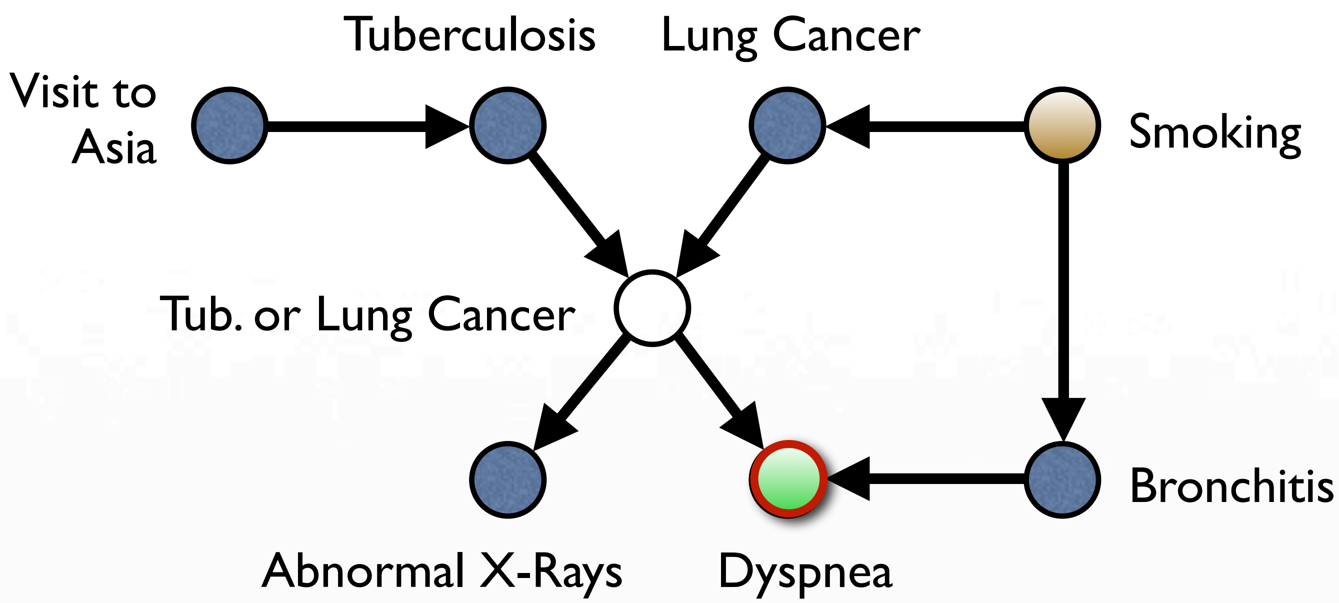
Explain: *Dyspnea = yes* | *Smoking = yes*

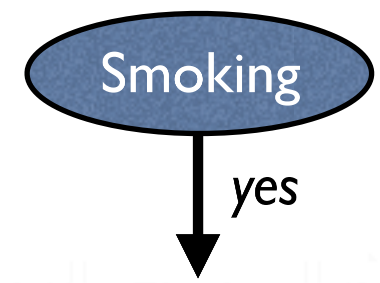


- Observed variable
- Explanandum
- Candidate explanatory variable
- Selected explanatory variable
- Excluded variable (modelling artifact)

Building the Tree: Example

Explain: *Dyspnea = yes* | *Smoking = yes*

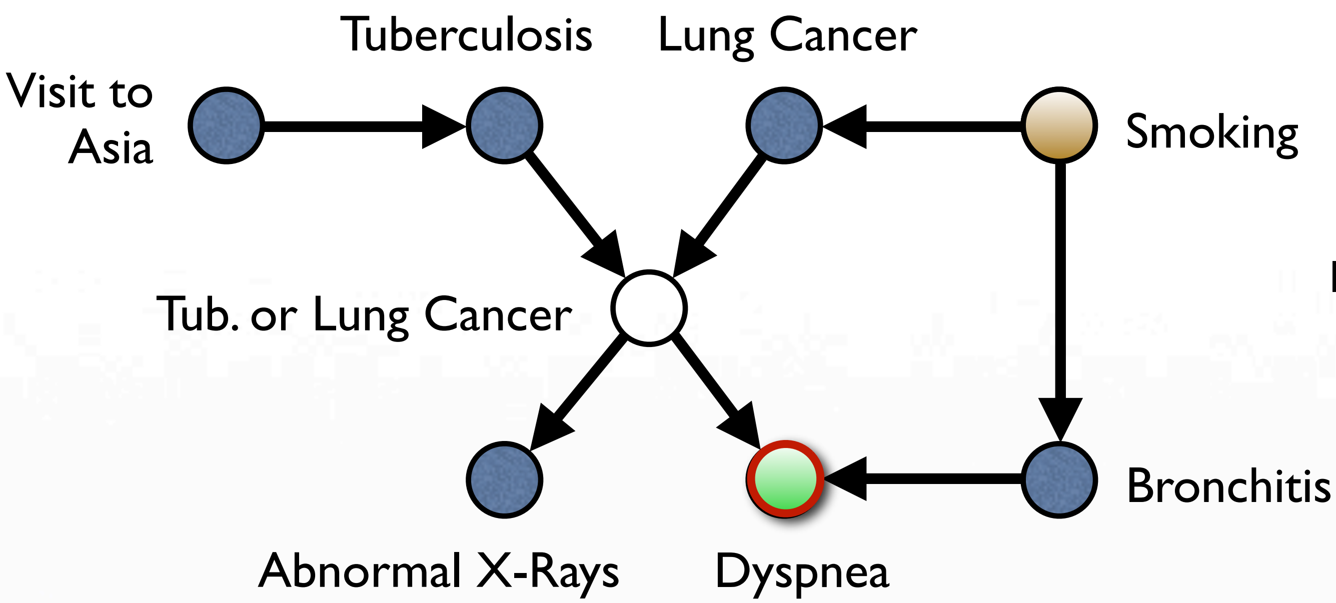




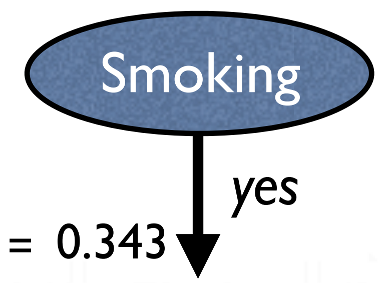
- Observed variable
- Explanandum
- Candidate explanatory variable
- Selected explanatory variable
- Excluded variable (modelling artifact)

Building the Tree: Example

Explain: *Dyspnea = yes* | *Smoking = yes*



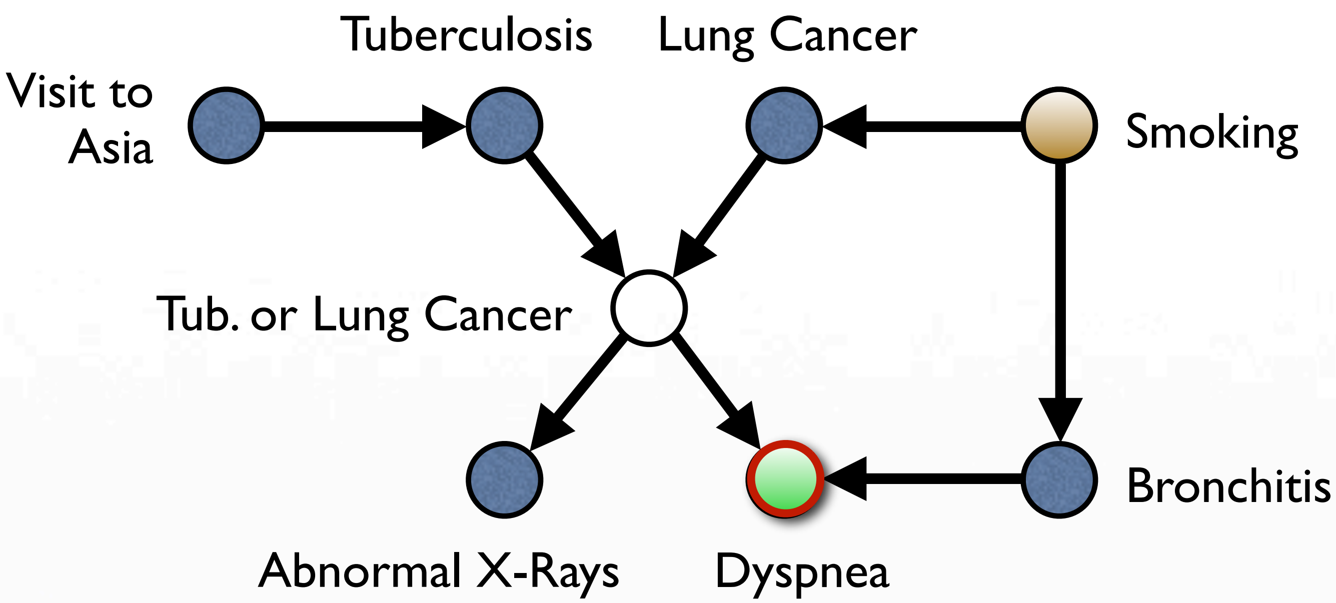
$$\log \frac{P(D = yes \mid do(S = yes))}{P(D = yes)} = 0.343$$

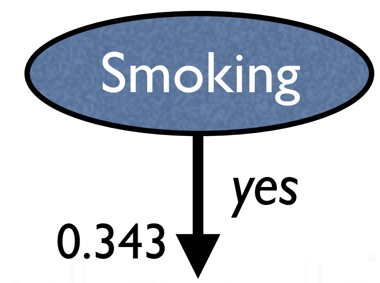


- Observed variable
- Explanandum
- Candidate explanatory variable
- Selected explanatory variable
- Excluded variable (modelling artifact)

Building the Tree: Example

Explain: *Dyspnea = yes* | *Smoking = yes*





$$\arg\max_X I(X \rightarrow D = yes \mid do(S = yes))$$

Observed variable

Explanandum

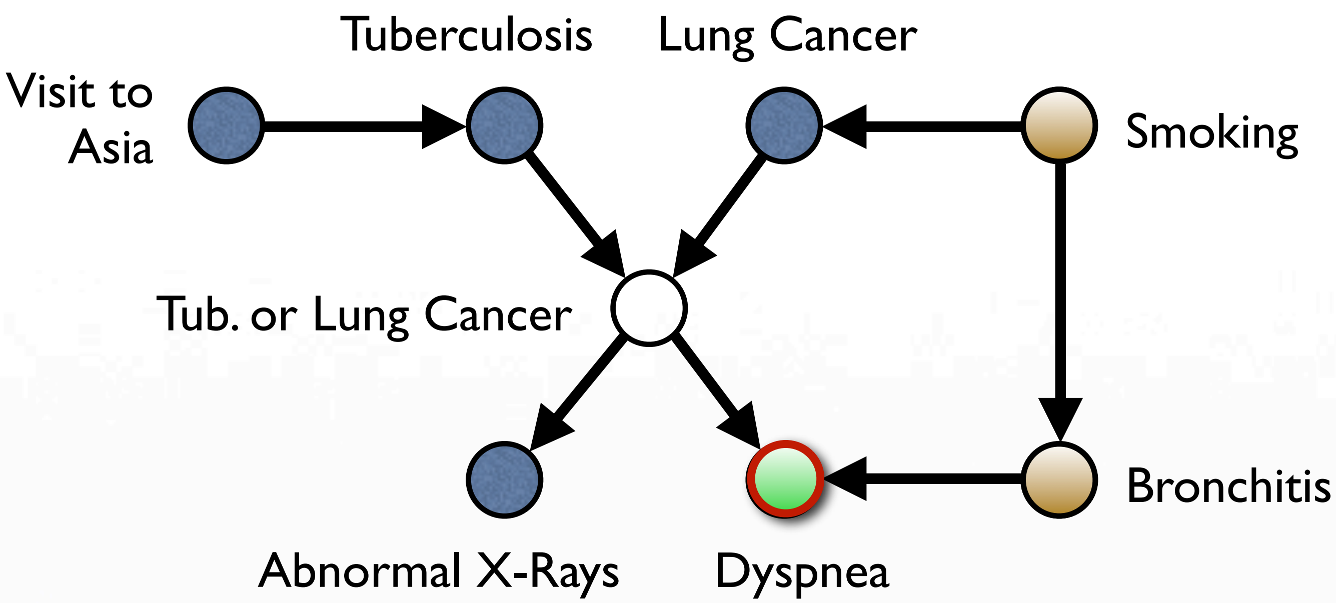
Candidate explanatory variable

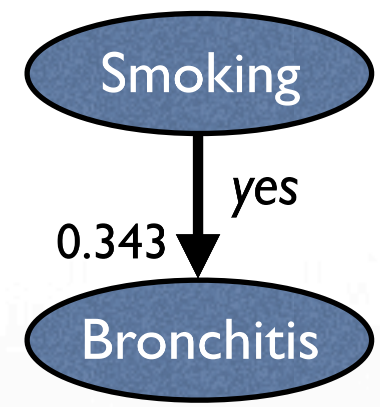
Selected explanatory variable

Excluded variable (modelling artifact)

Building the Tree: Example

Explain: *Dyspnea = yes* | *Smoking = yes*

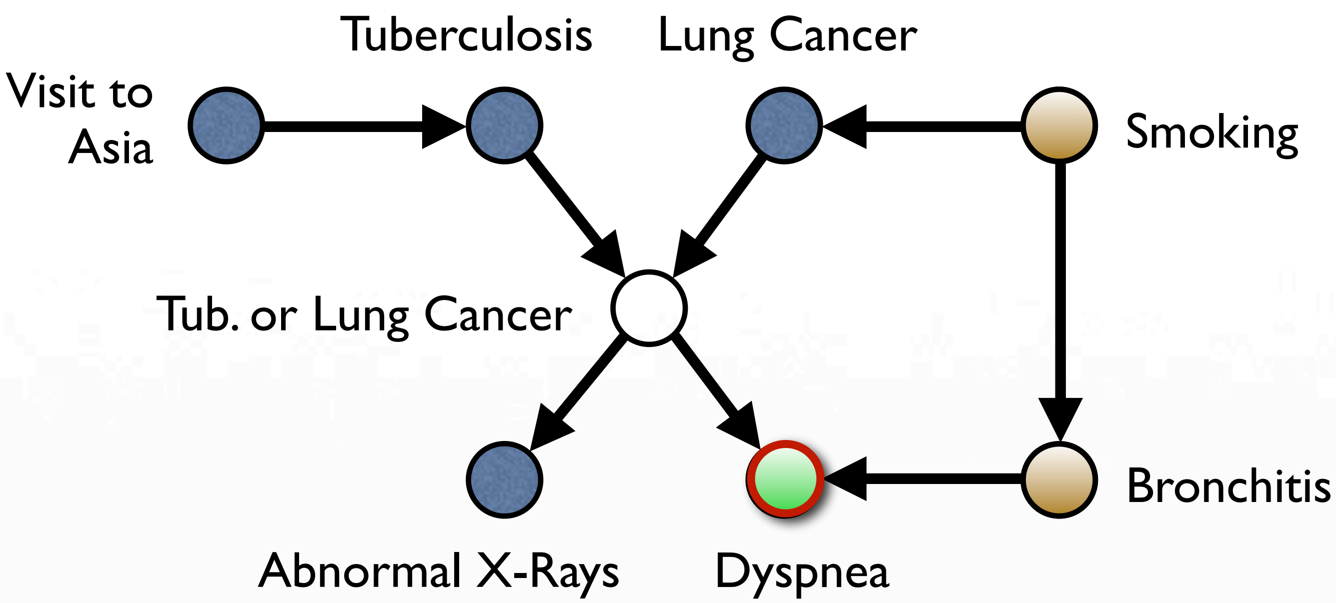


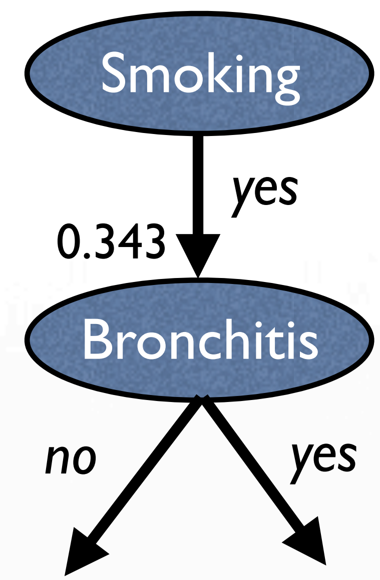


- Observed variable
- Explanandum
- Candidate explanatory variable
- Selected explanatory variable
- Excluded variable (modelling artifact)

Building the Tree: Example

Explain: *Dyspnea = yes* | *Smoking = yes*

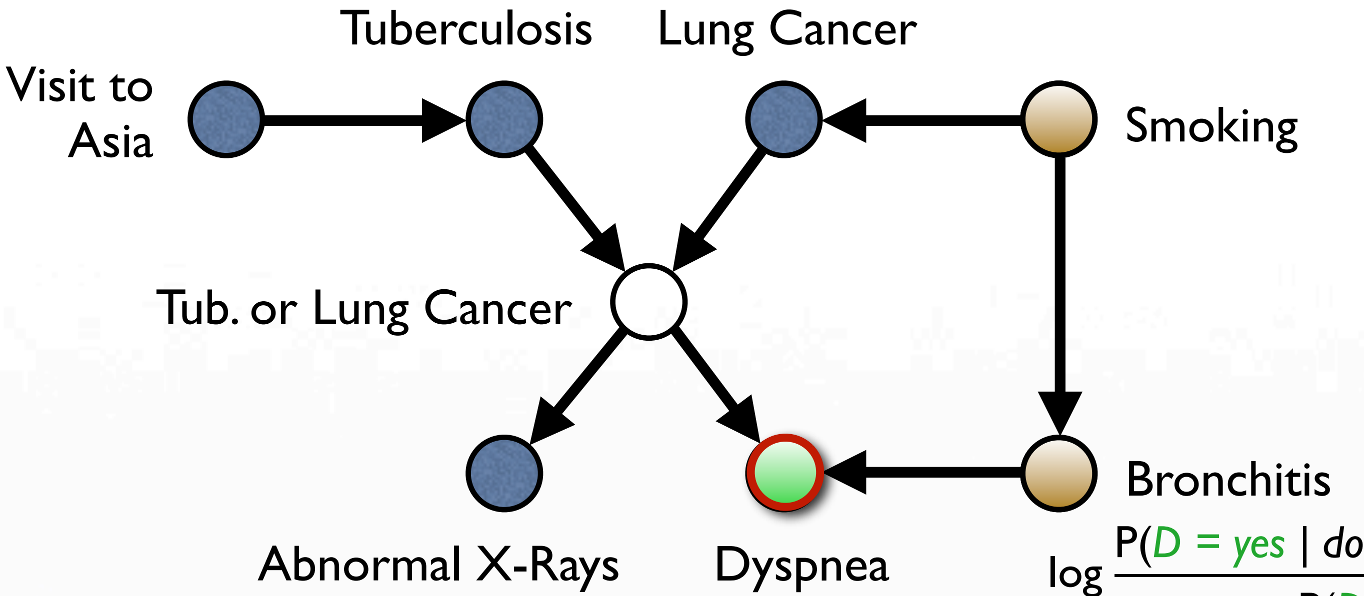




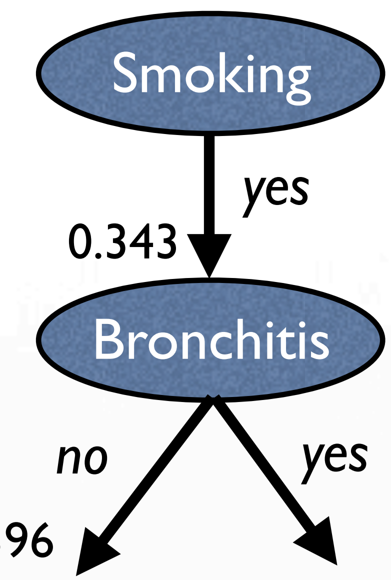
- Observed variable
- Explanandum
- Candidate explanatory variable
- Selected explanatory variable
- Excluded variable (modelling artifact)

Building the Tree: Example

Explain: *Dyspnea = yes* | *Smoking = yes*



$$\log \frac{P(D = yes \mid do(S = yes, B = no))}{P(D = yes)} = -1.396$$



Observed variable

Explanandum

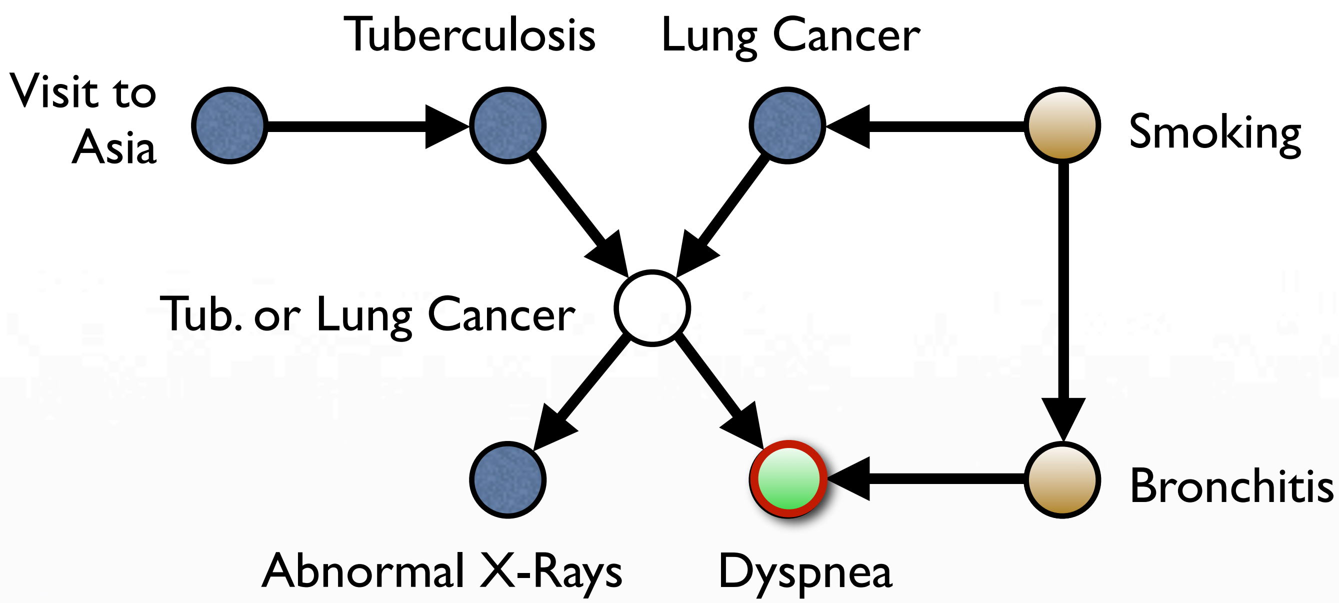
Candidate explanatory variable

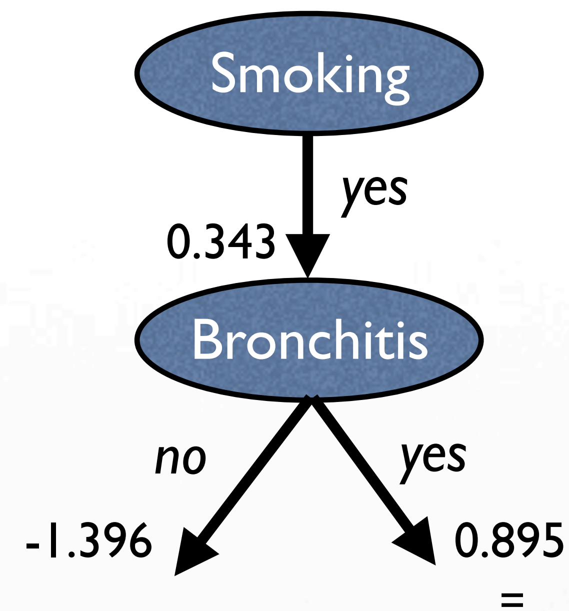
Selected explanatory variable

Excluded variable (modelling artifact)

Building the Tree: Example

Explain: *Dyspnea = yes* | *Smoking = yes*





$$\log \frac{P(D = yes \mid do(S = yes, B = yes))}{P(D = yes)}$$

Observed variable

Explanandum

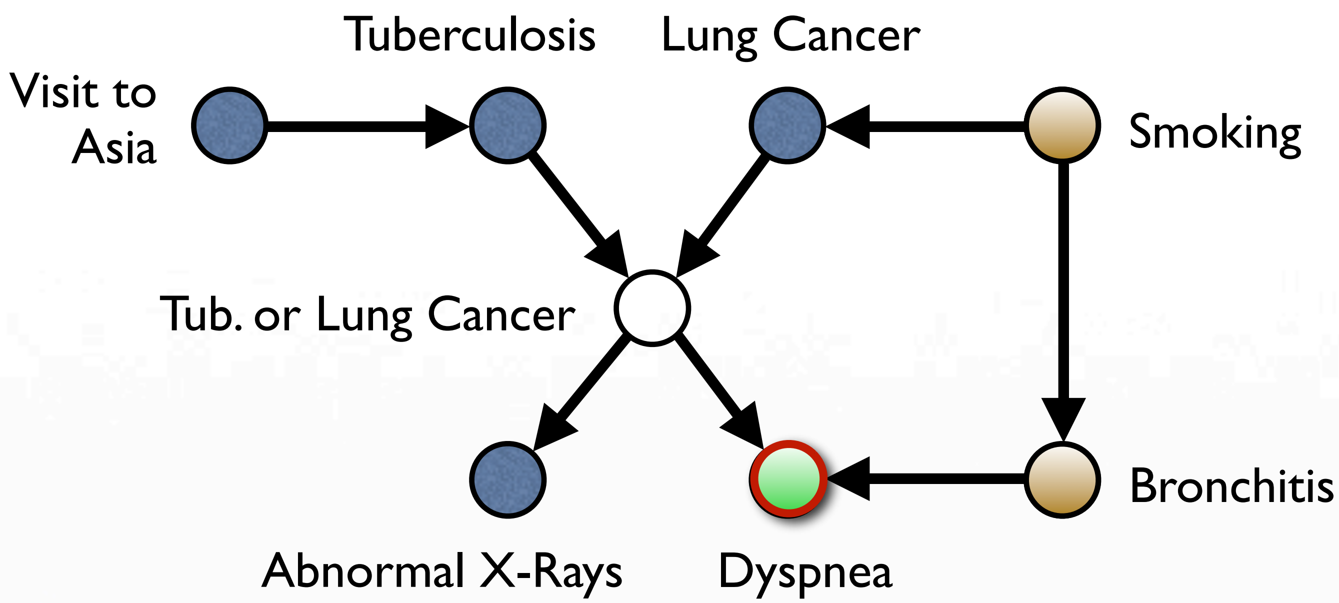
Candidate explanatory variable

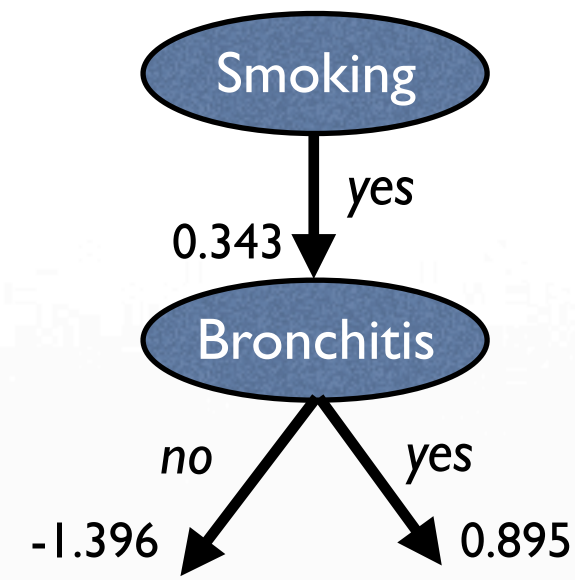
Selected explanatory variable

Excluded variable (modelling artifact)

Building the Tree: Example

Explain: *Dyspnea = yes* | *Smoking = yes*



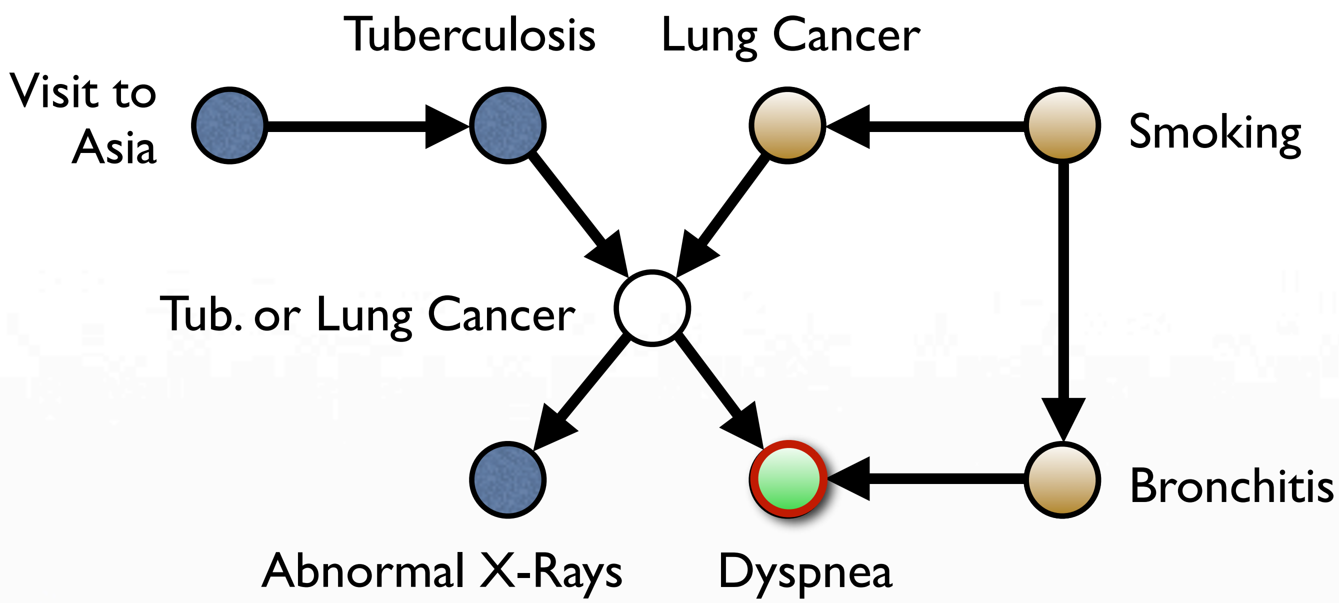


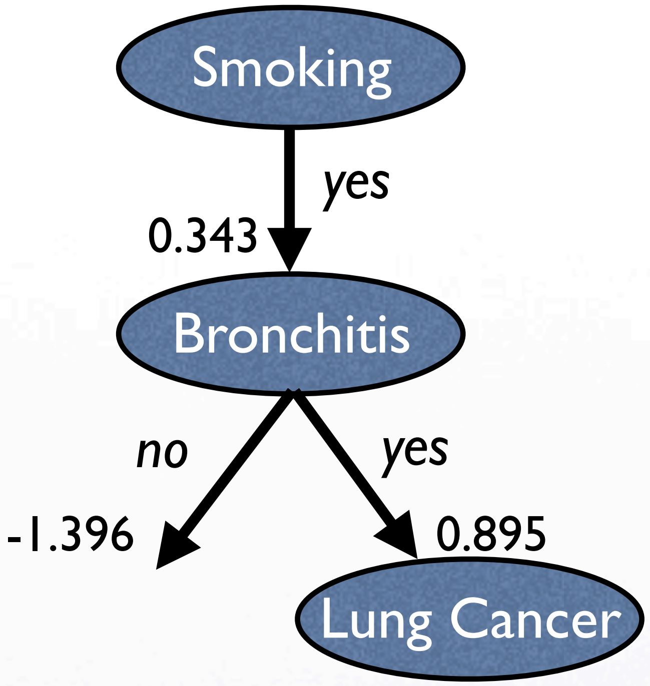
$\arg\max_X I(X \rightarrow D = yes \mid do(S = yes, B = yes))$

- Observed variable
- Explanandum
- Candidate explanatory variable
- Selected explanatory variable
- Excluded variable (modelling artifact)

Building the Tree: Example

Explain: *Dyspnea = yes* | *Smoking = yes*





Observed variable

Explanandum

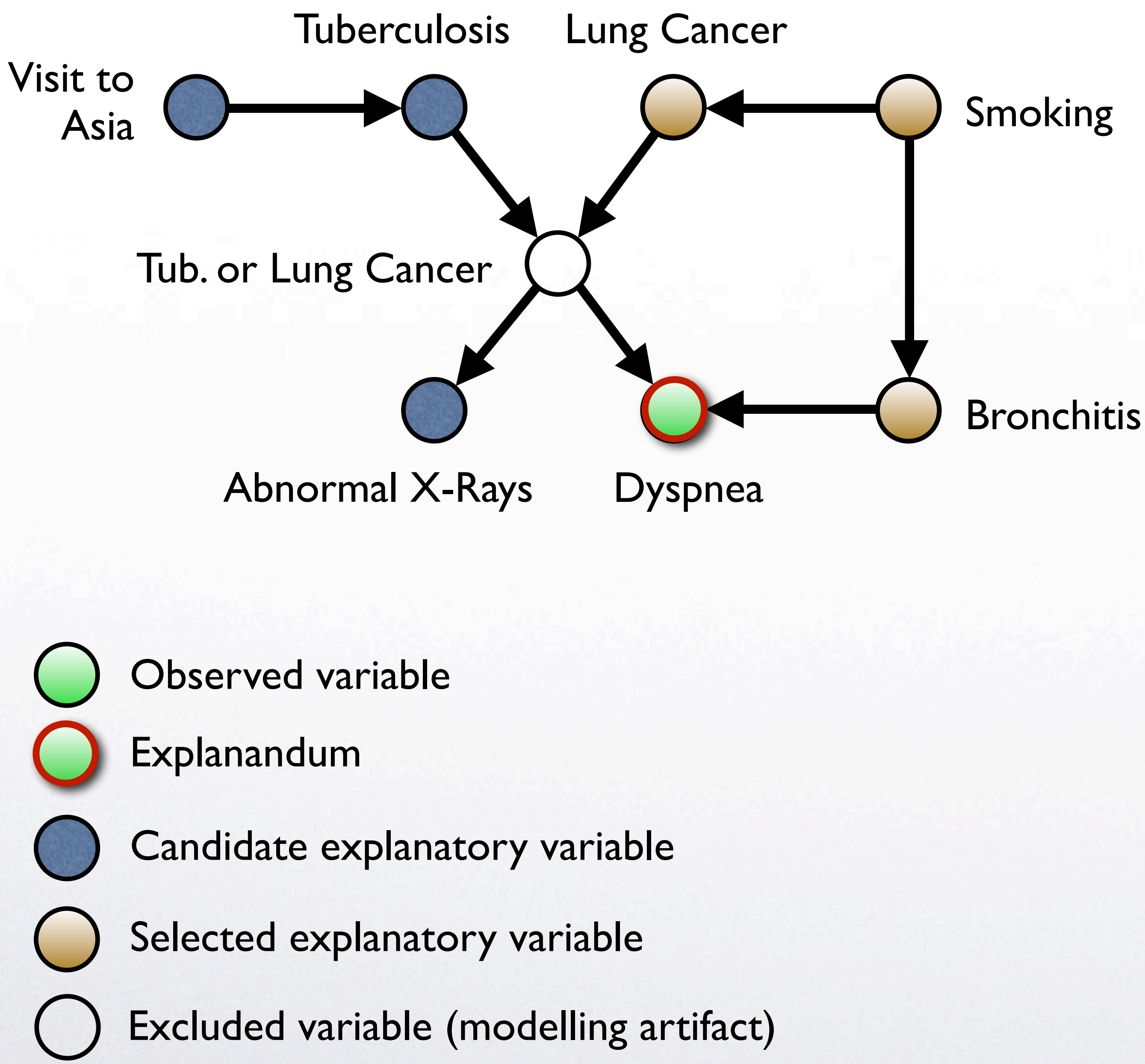
Candidate explanatory variable

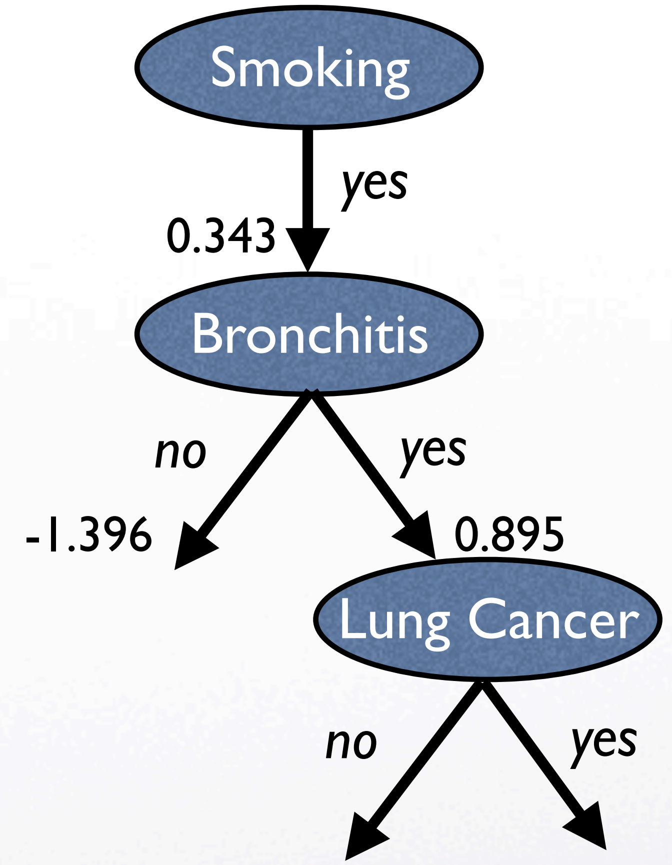
Selected explanatory variable

Excluded variable (modelling artifact)

Building the Tree: Example

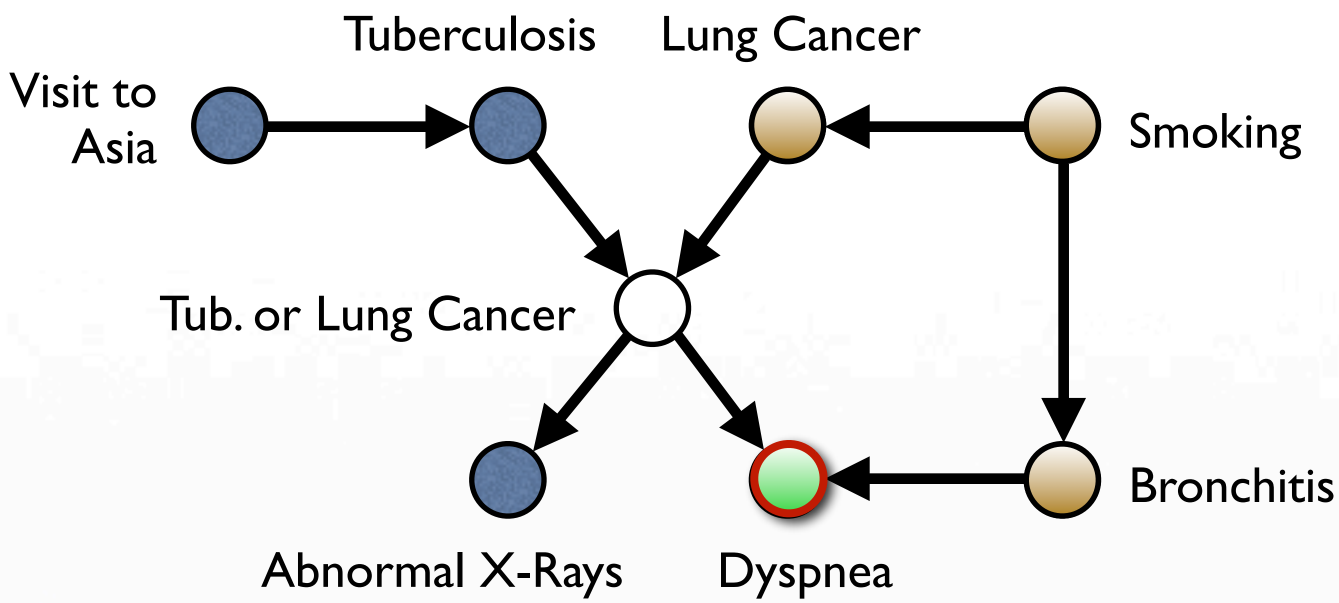
Explain: *Dyspnea = yes* | *Smoking = yes*

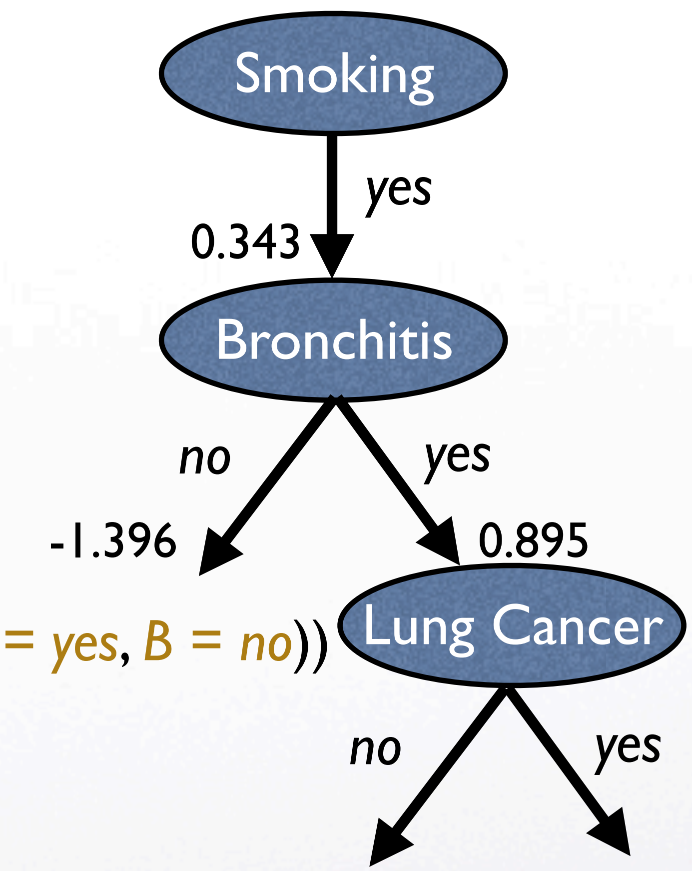




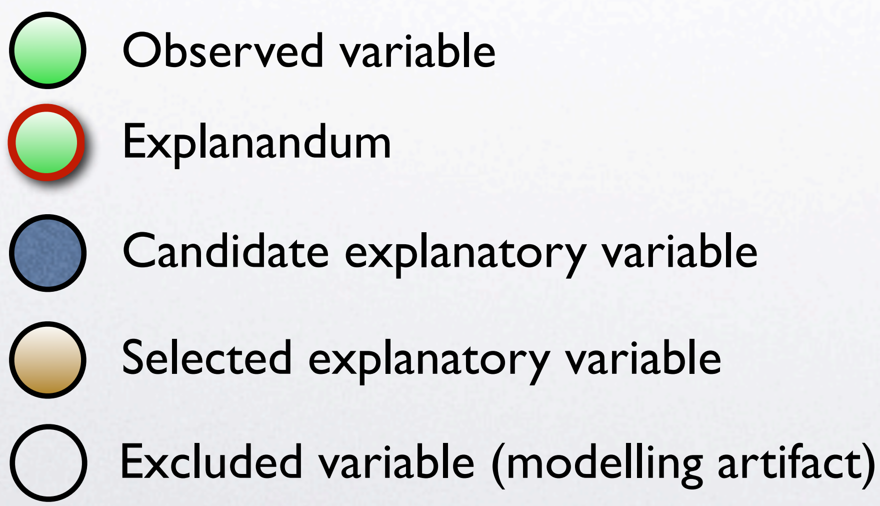
Building the Tree: Example

Explain: *Dyspnea = yes* | *Smoking = yes*



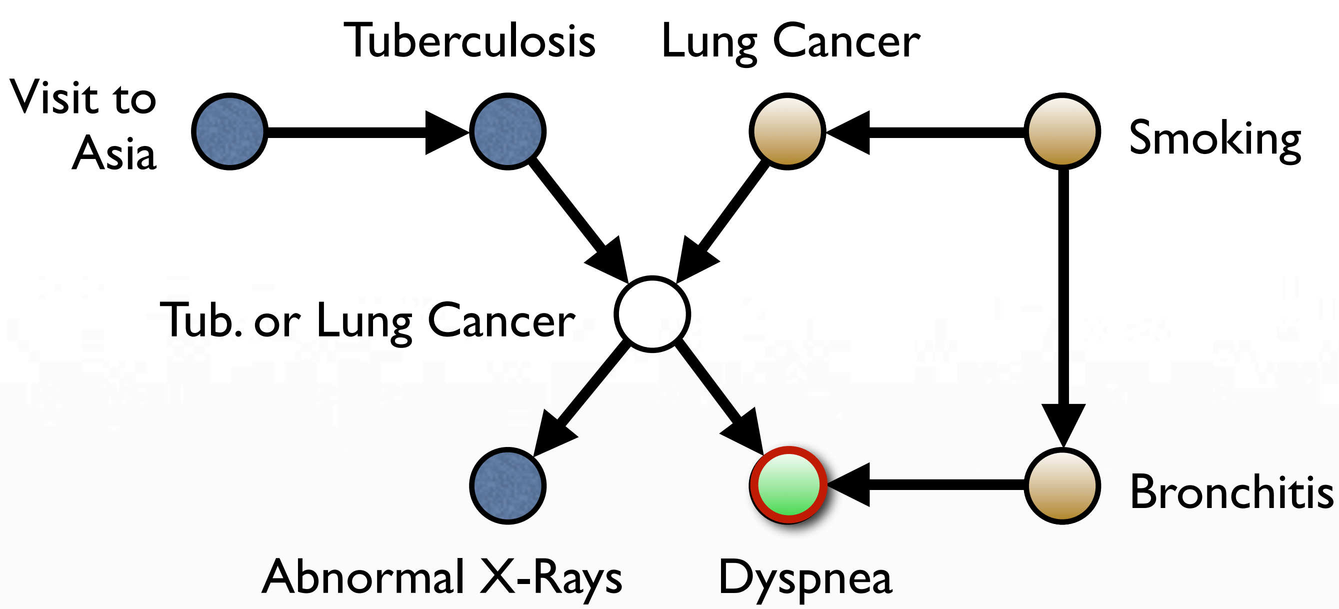


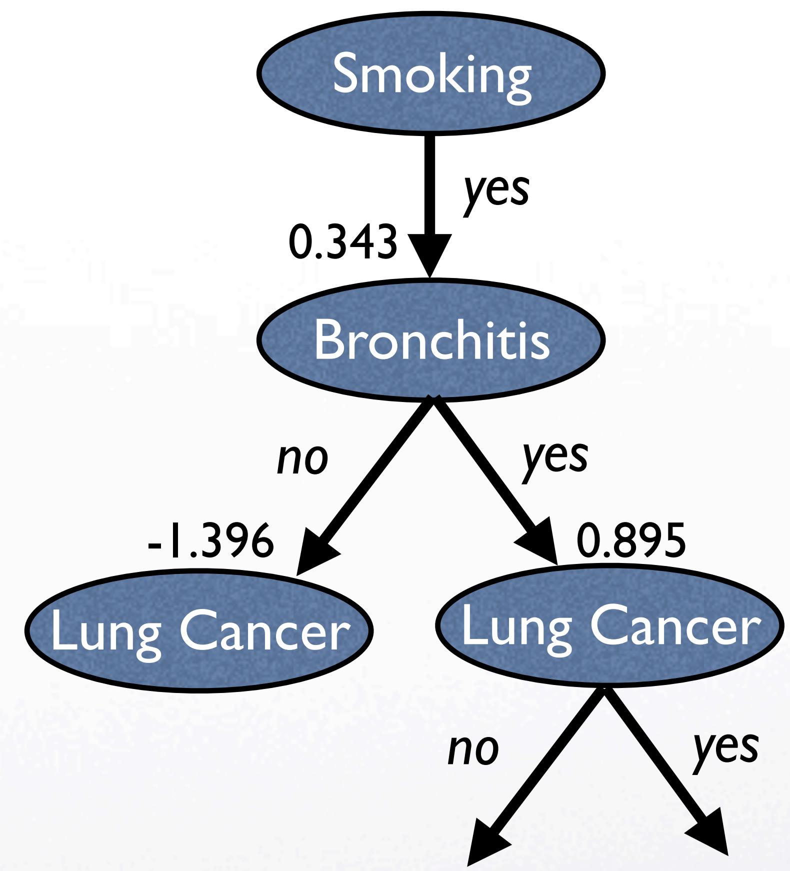
$\arg\max_X I(X \rightarrow D = yes \mid do(S = yes, B = no))$



Building the Tree: Example

Explain: *Dyspnea = yes* | *Smoking = yes*





Observed variable

Explanandum

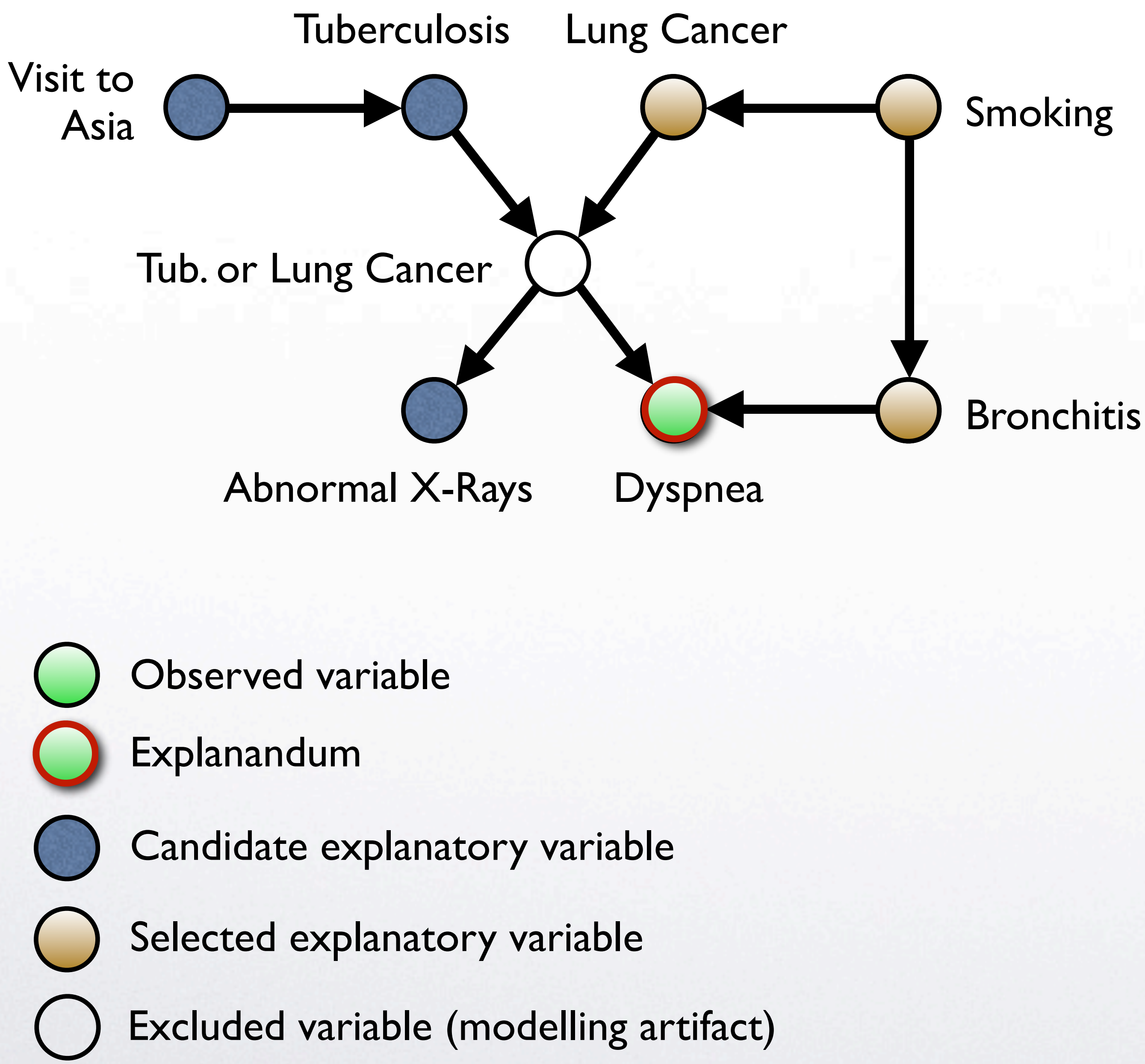
Candidate explanatory variable

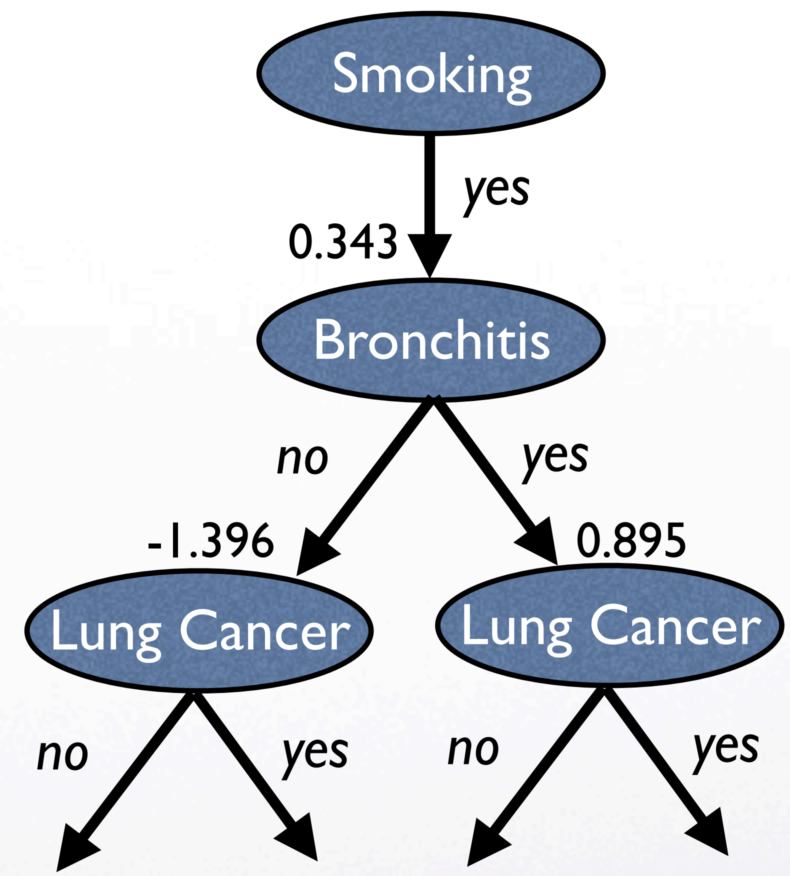
Selected explanatory variable

Excluded variable (modelling artifact)

Building the Tree: Example

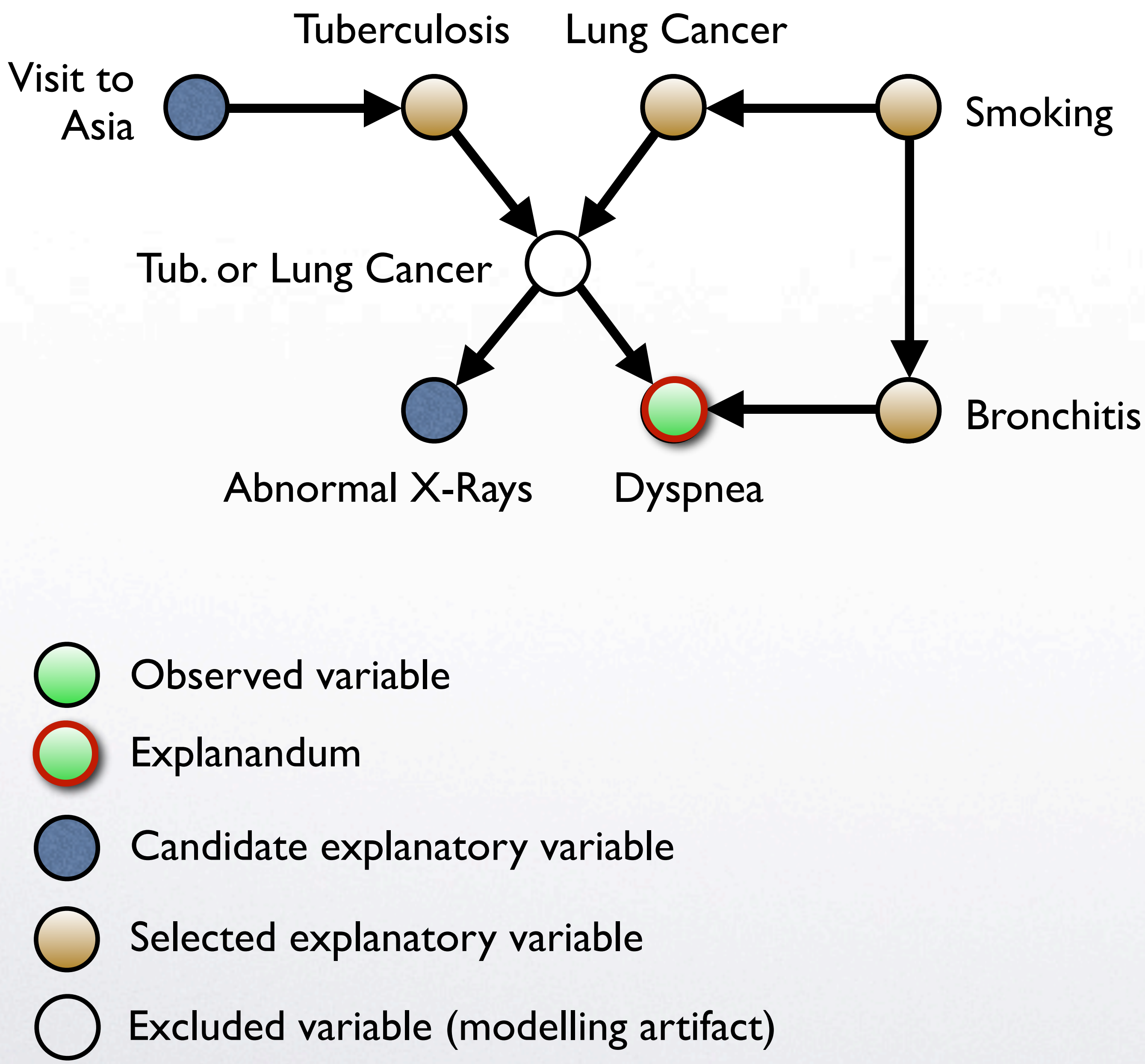
Explain: *Dyspnea = yes* | *Smoking = yes*

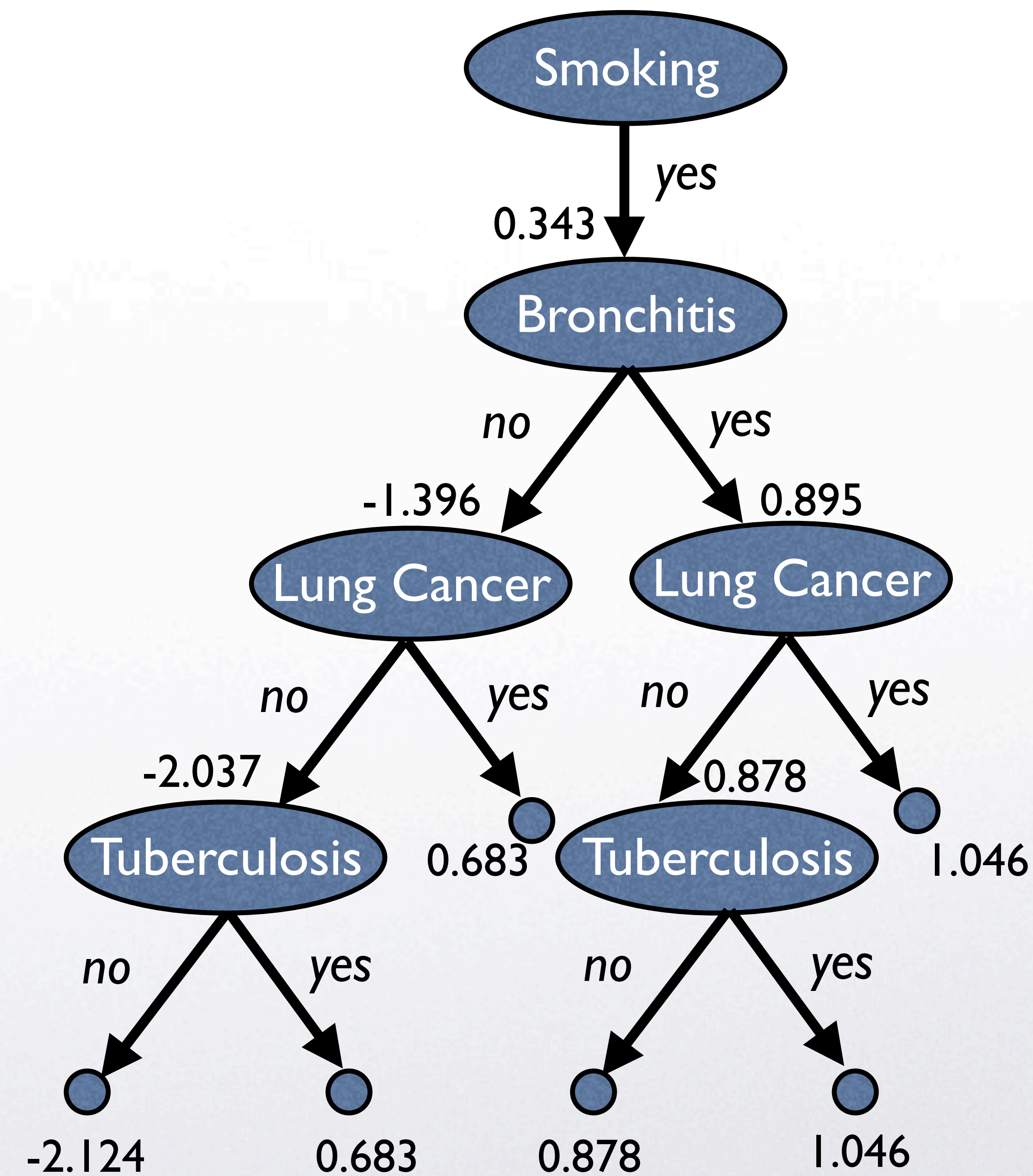




Building the Tree: Example

Explain: *Dyspnea = yes* | *Smoking = yes*





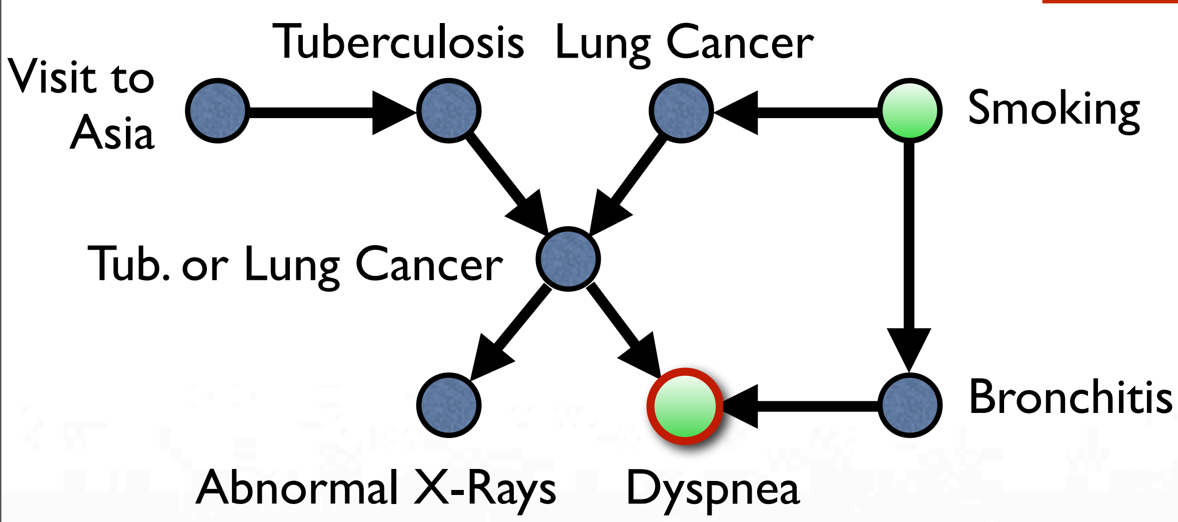
III

So What?

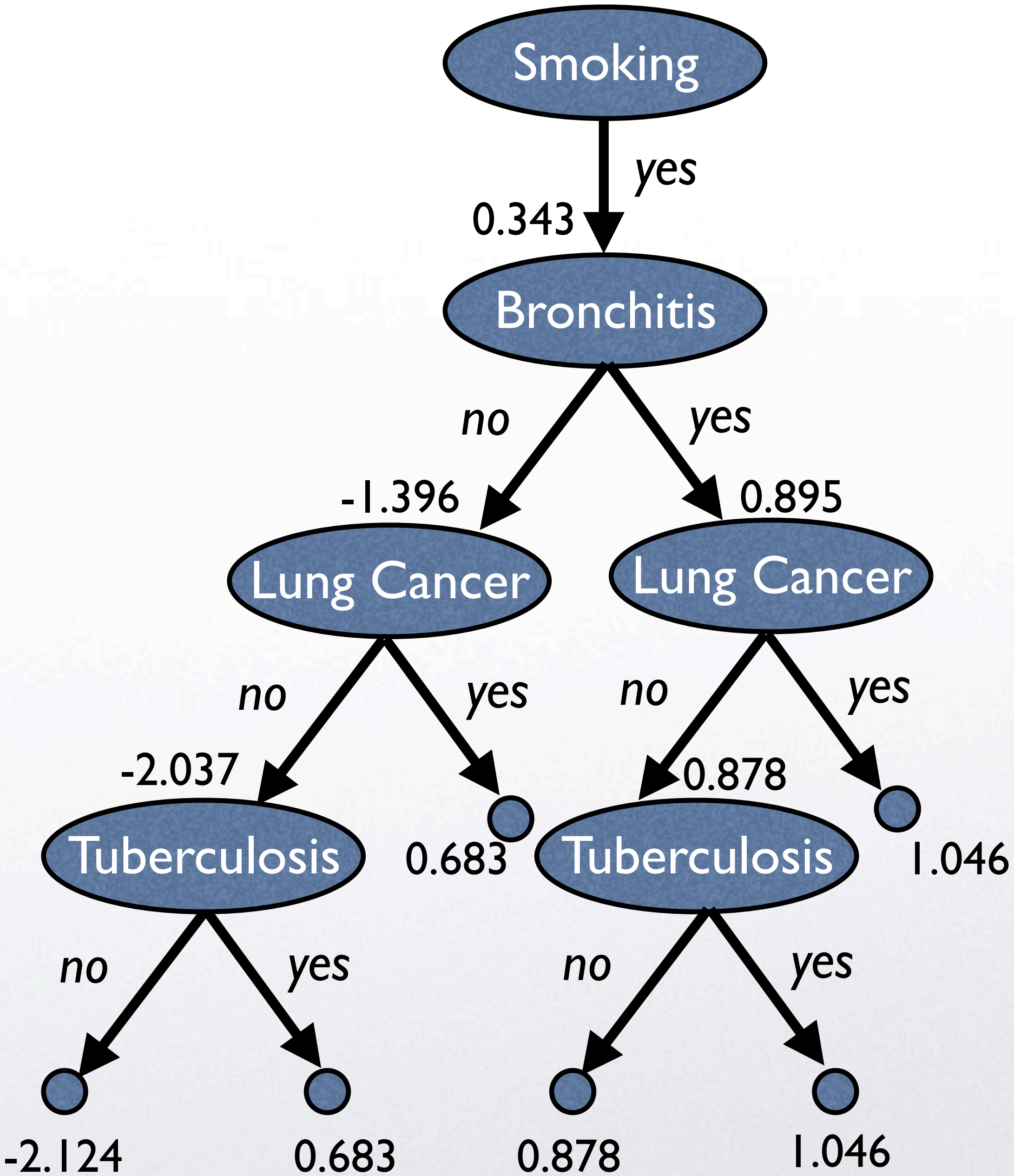
Comparisons

Asia: Example I

Explain: *Dyspnea = yes* | *Smoking = yes*

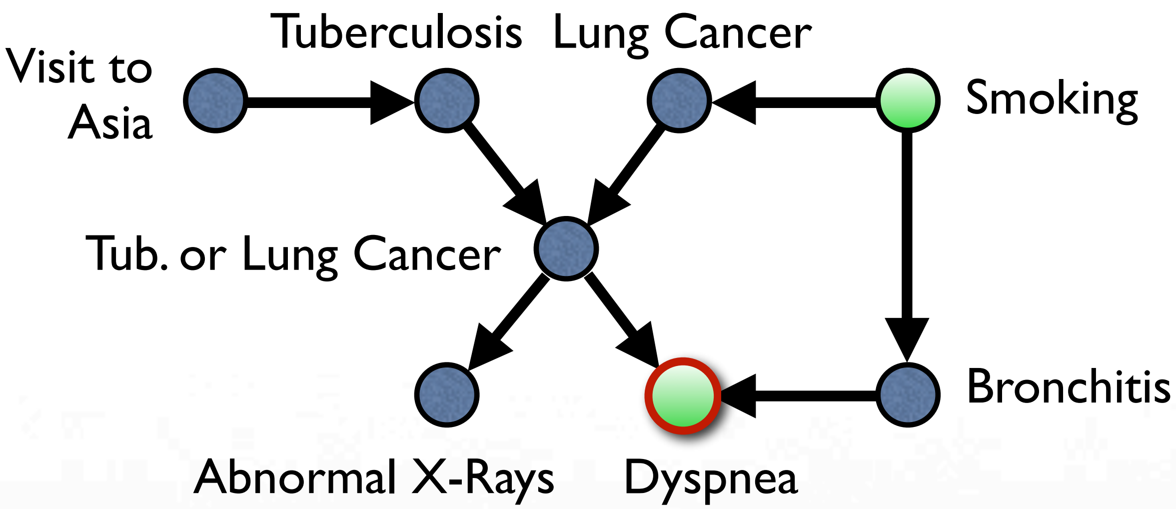


Causal explanation tree

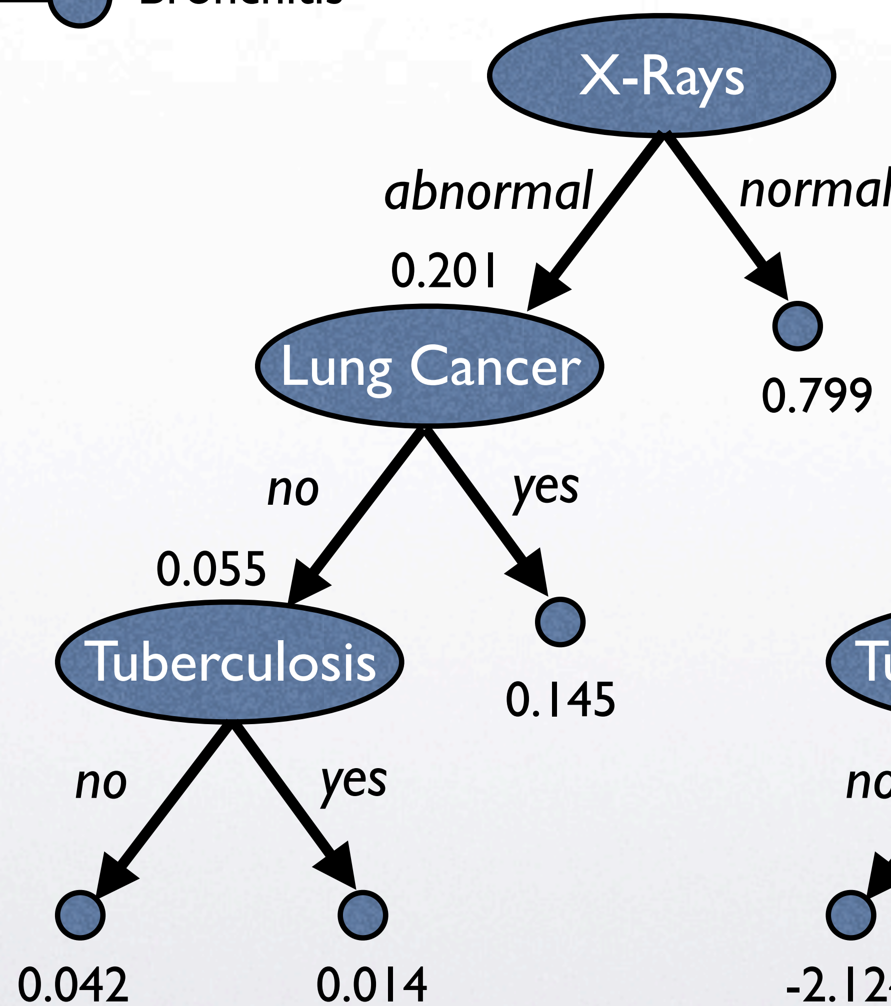


Asia: Example I

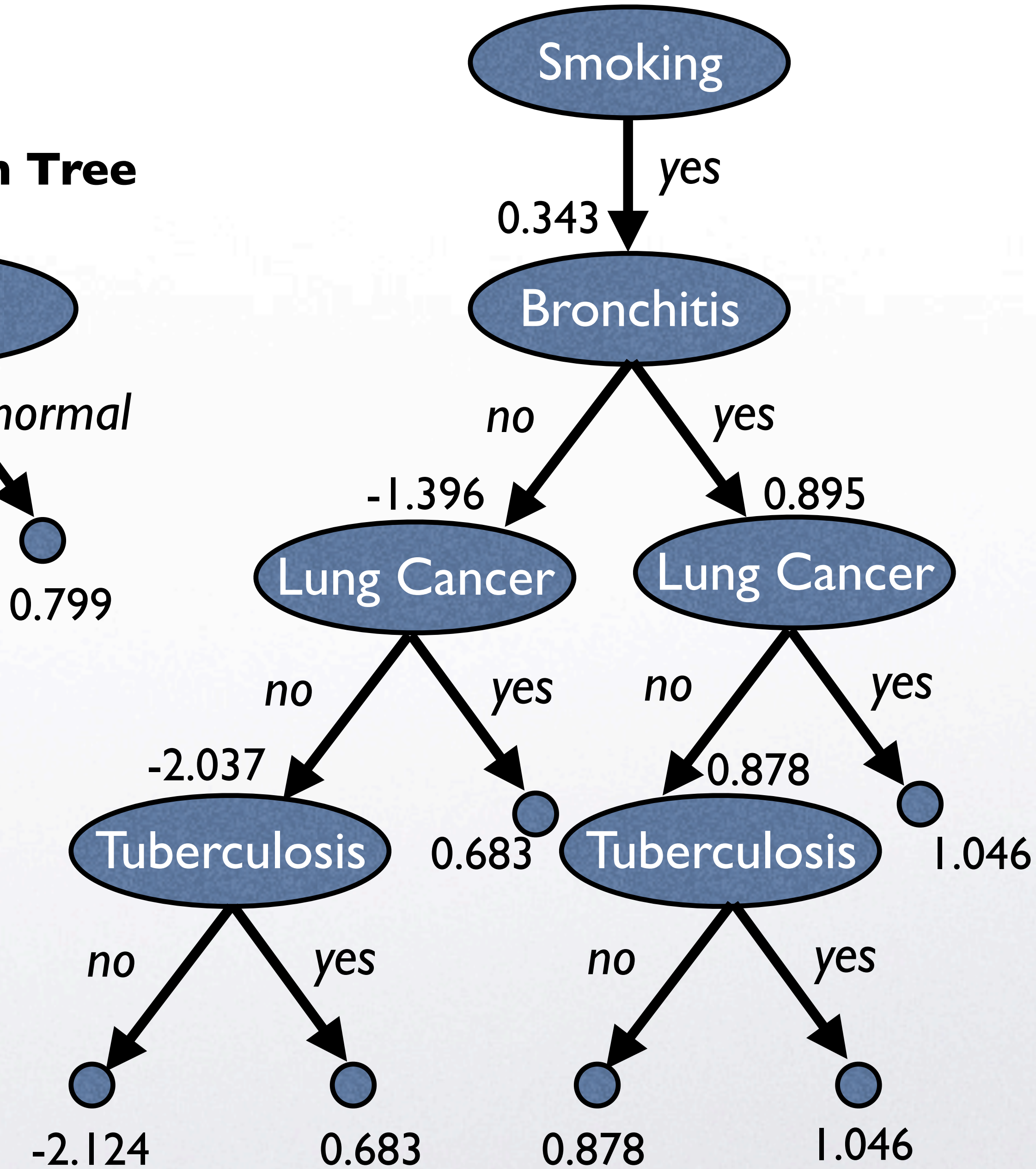
Explain: *Dyspnea = yes* | *Smoking = yes*



Explanation Tree

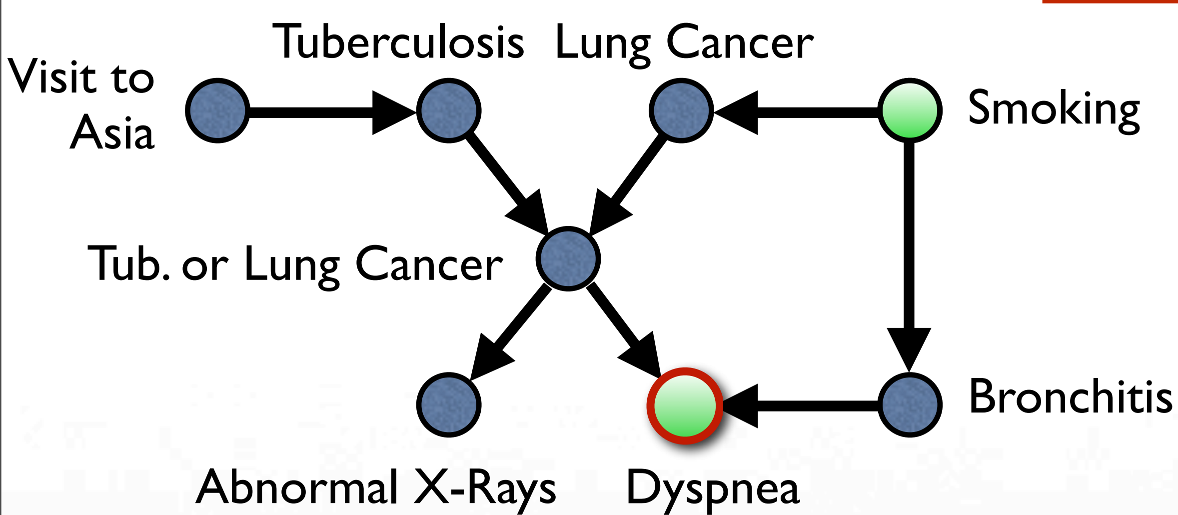


Causal explanation tree

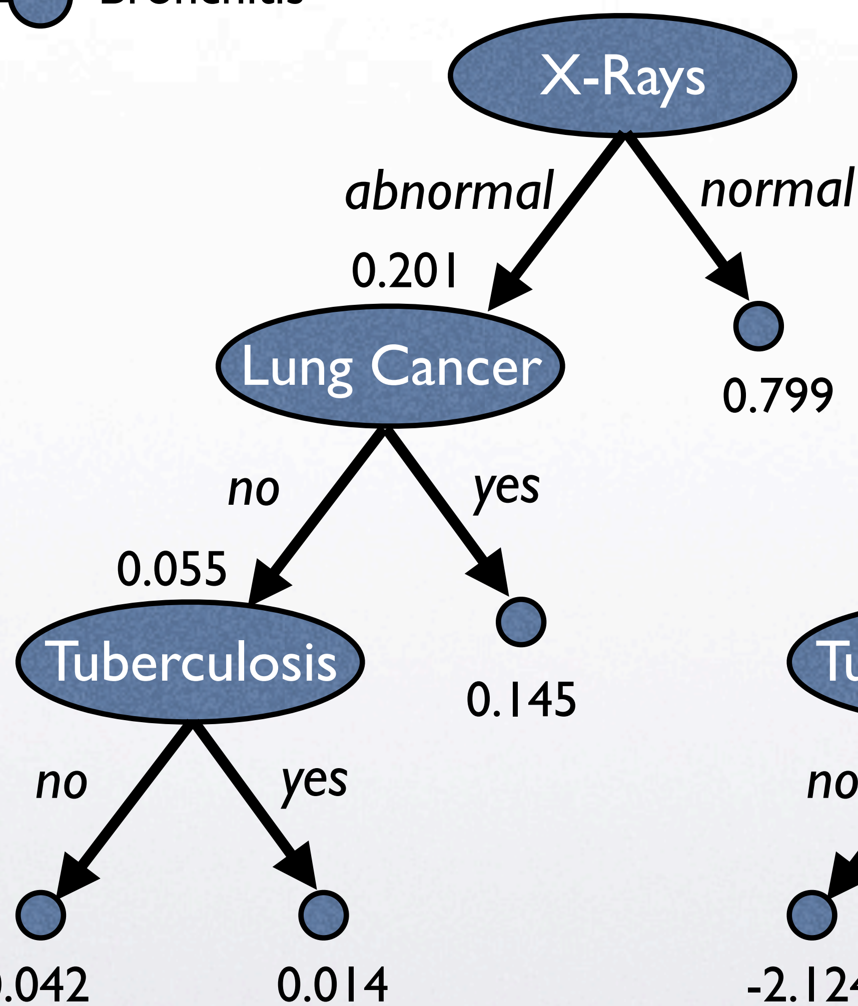


Asia: Example I

Explain: *Dyspnea = yes* | *Smoking = yes*

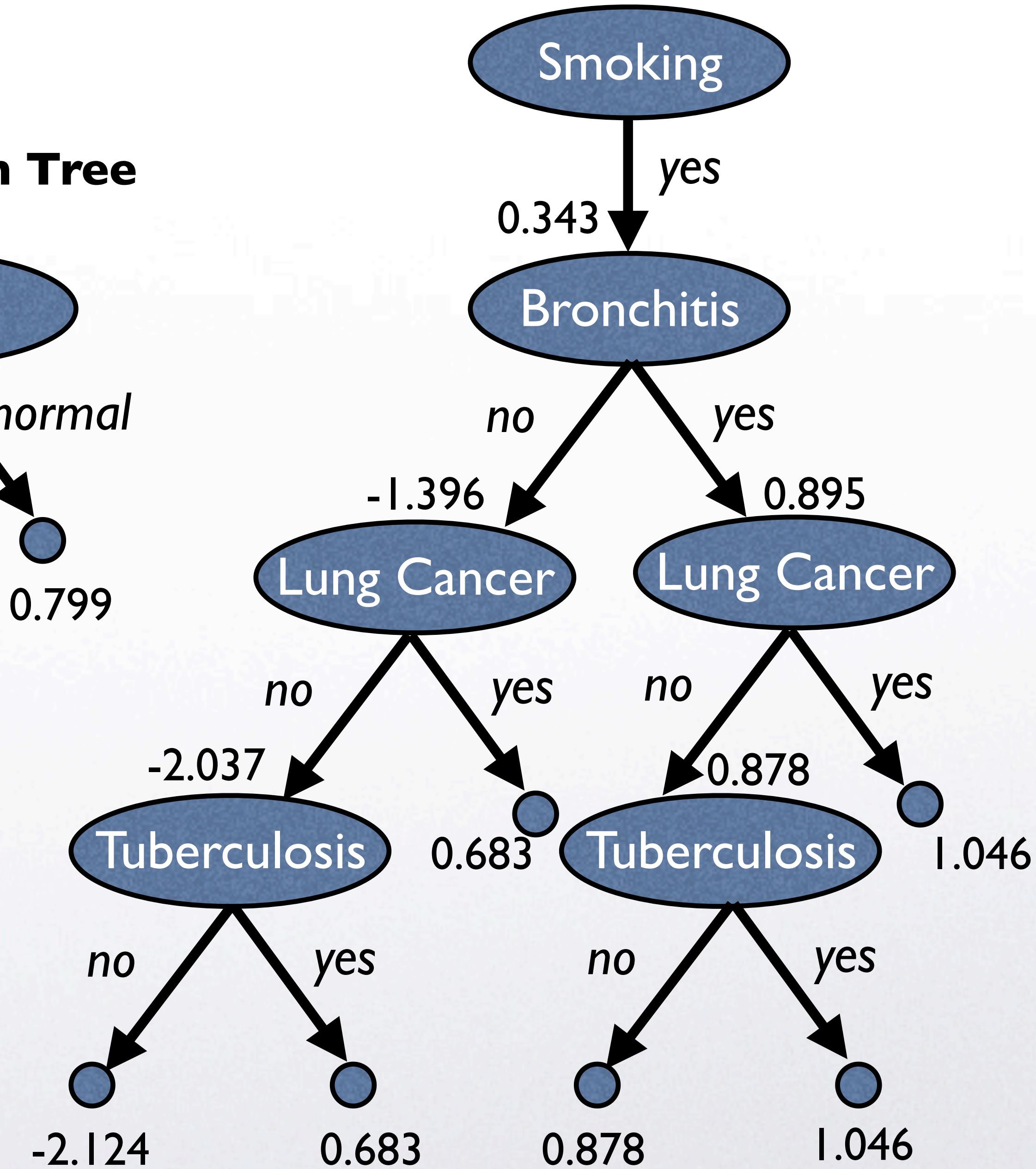


Explanation Tree



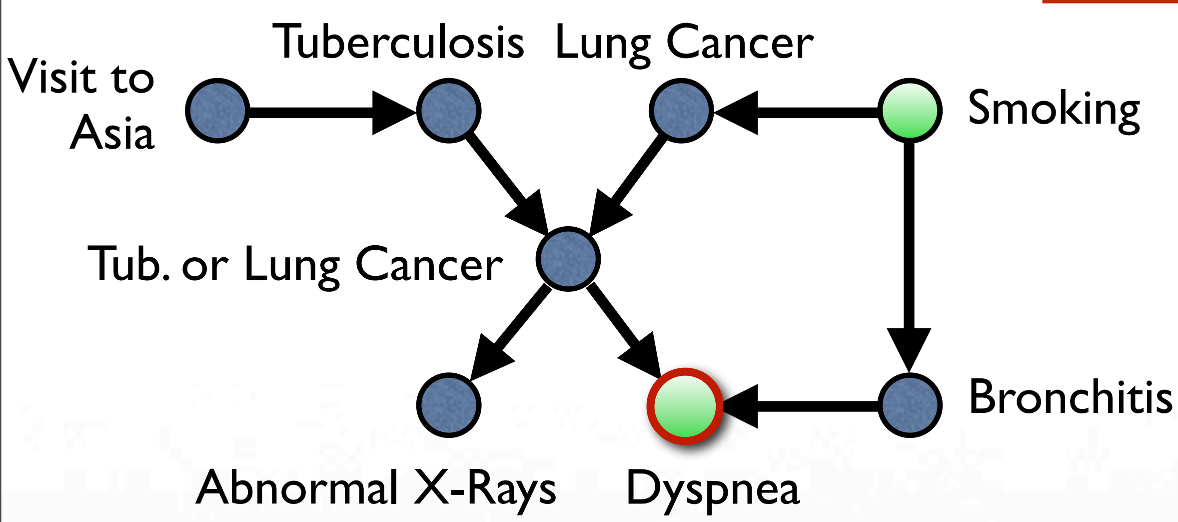
P(*X-Ray* = *normal*, *Lung cancer* = *no*, *Tuberculosis* = *no* | *Dyspnea = yes*) = 0.042

Causal explanation tree

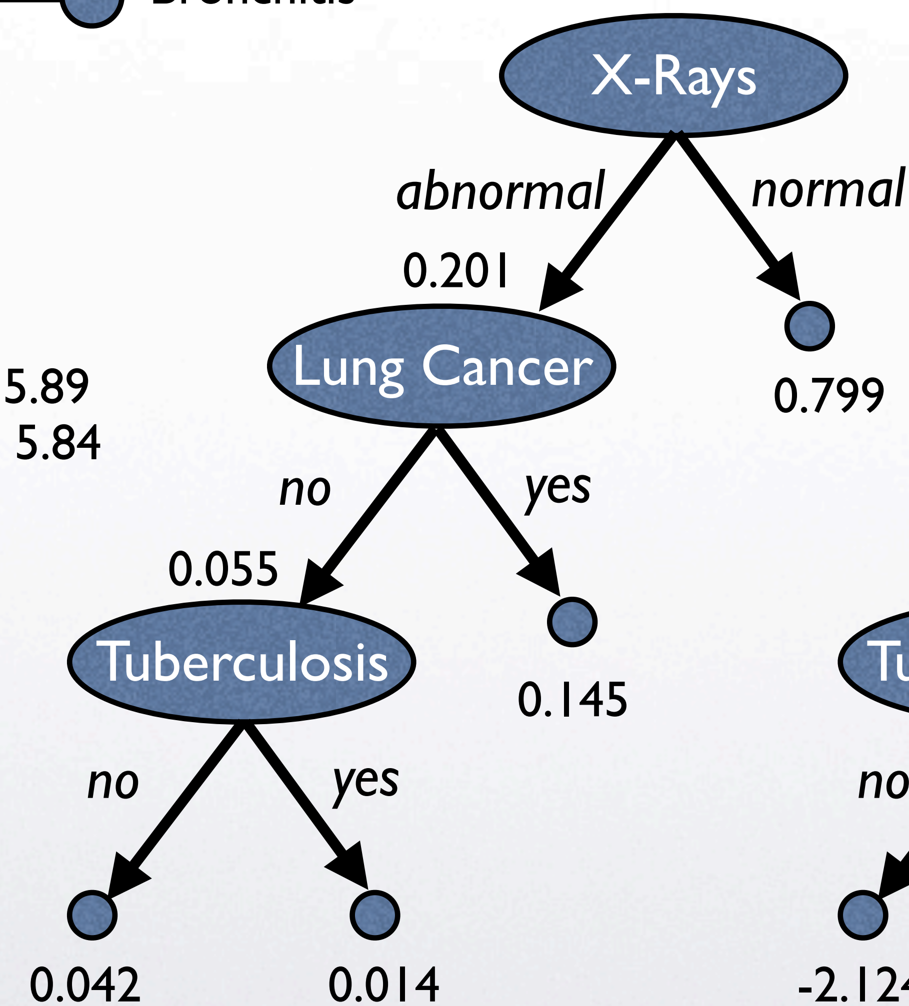


Asia: Example I

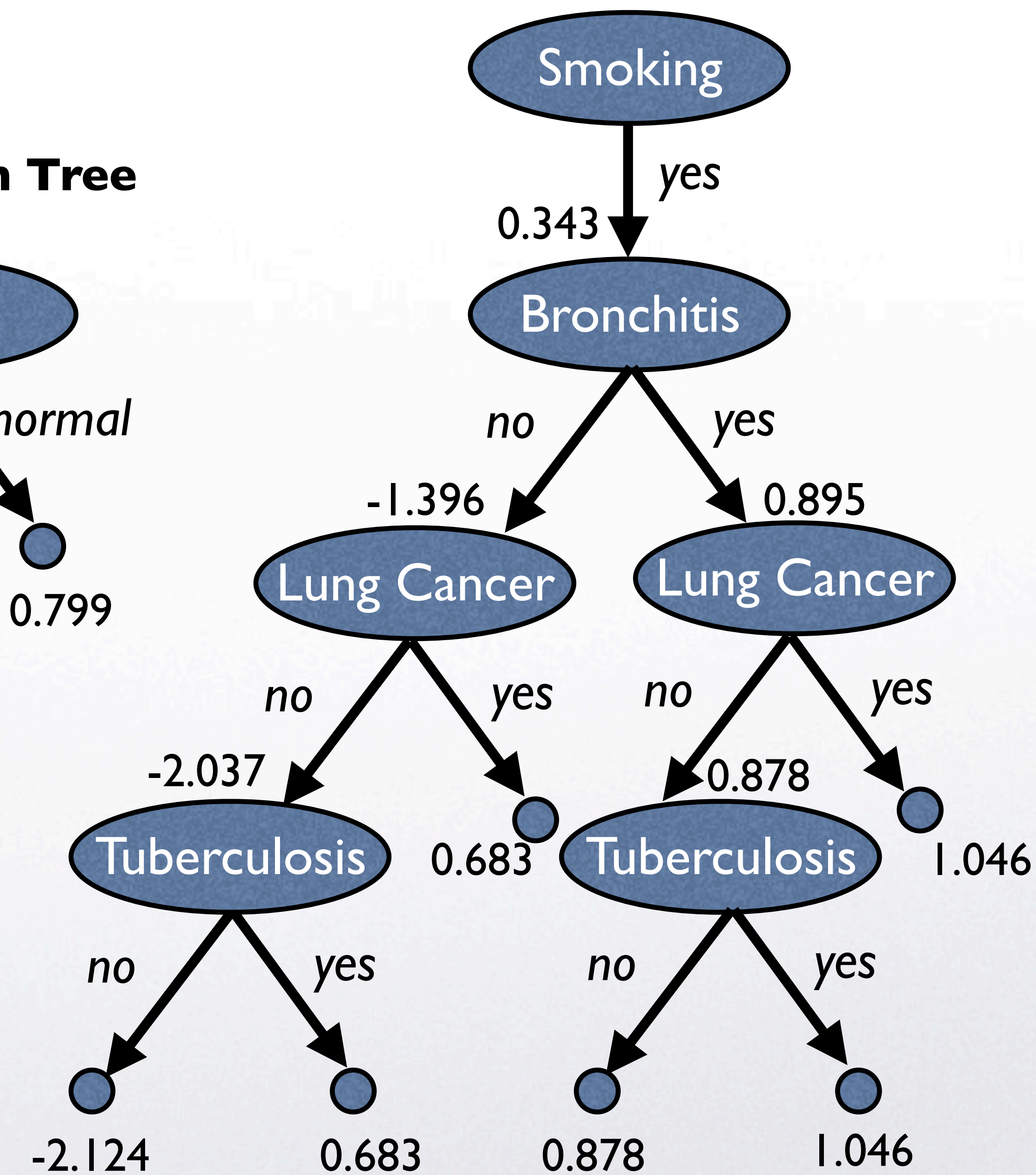
Explain: *Dyspnea = yes* | *Smoking = yes*



Explanation Tree



Causal explanation tree



Bayes' Factor

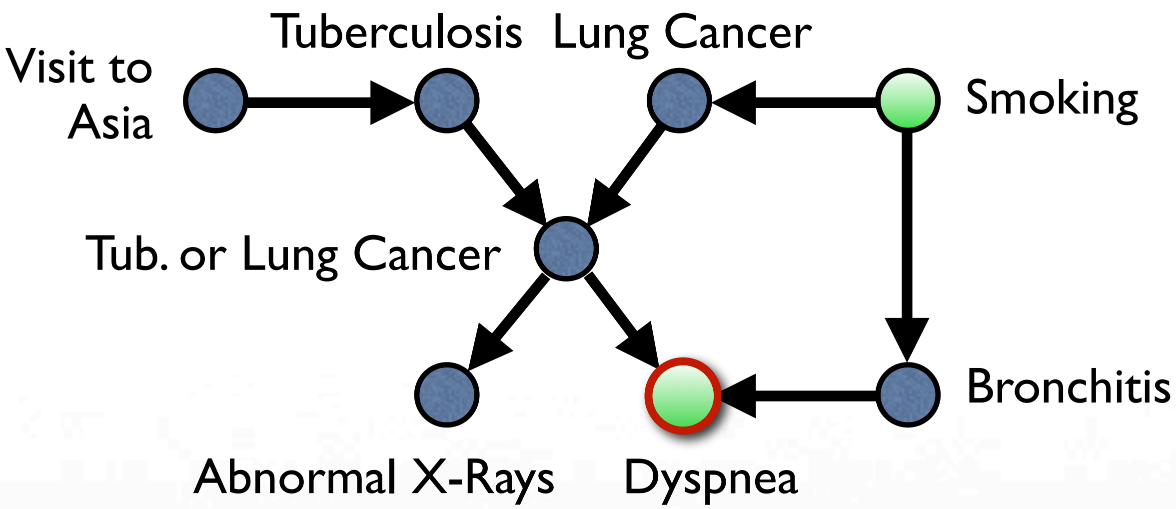
BF(*Bronchitis* = *yes*) = 6.14

BF(*Bronchitis* = *yes*, *Visit to Asia* = *no*) = 5.89

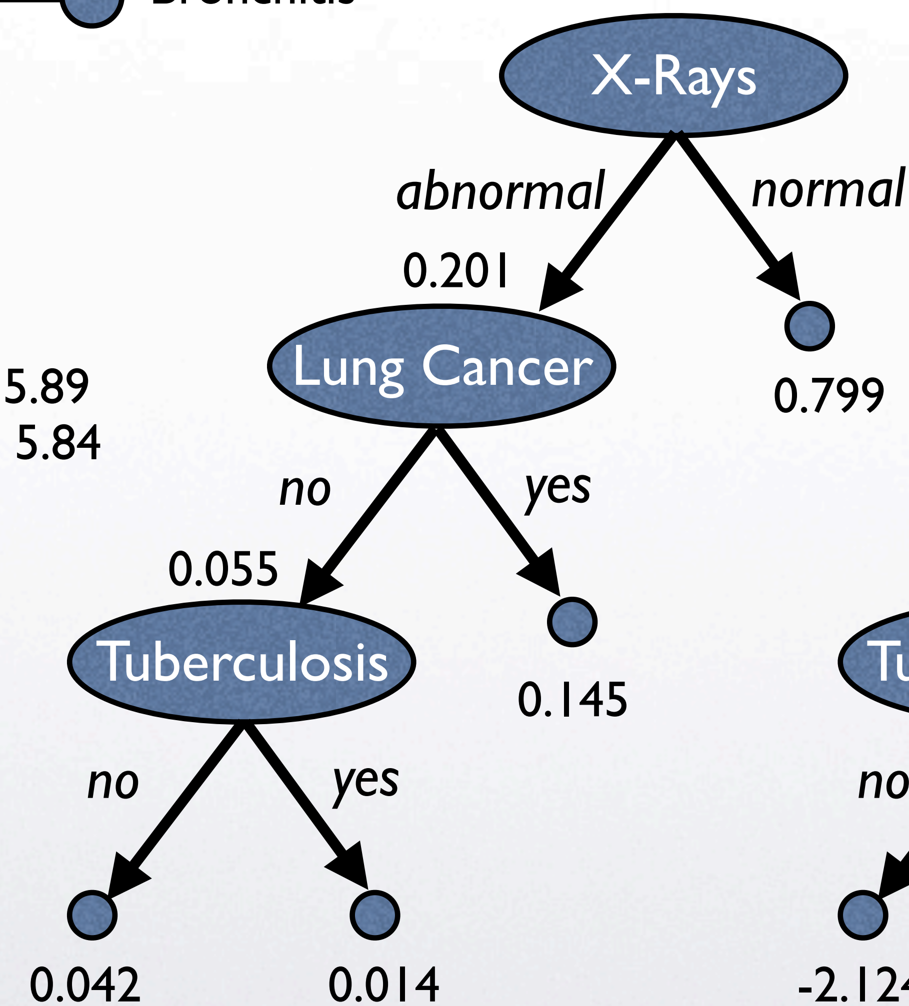
BF(*Bronchitis* = *yes*, *Tuberculosis* = *no*) = 5.84

Asia: Example I

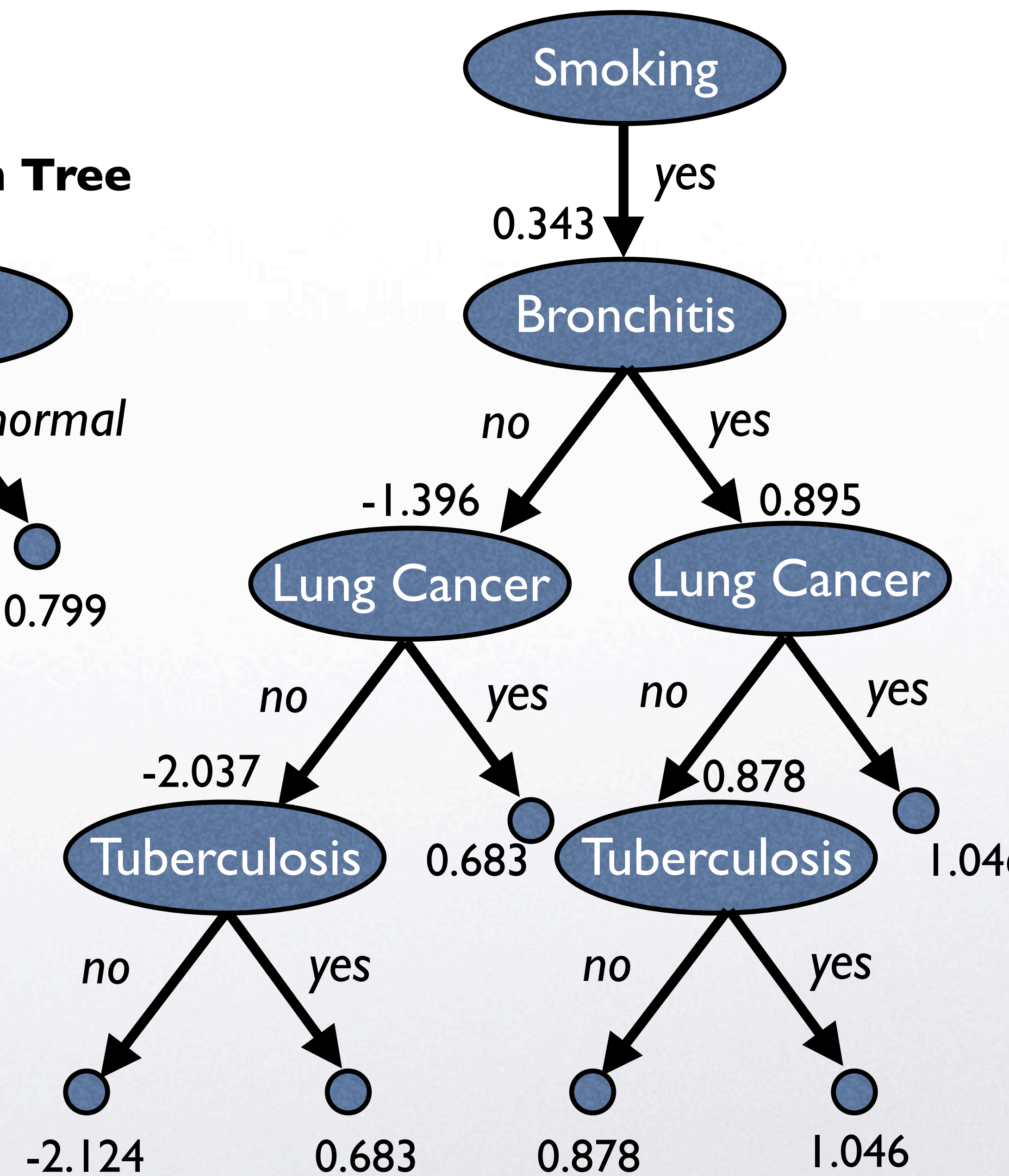
Explain: *Dyspnea = yes* | *Smoking = yes*



Explanation Tree



Causal explanation tree



Bayes' Factor

BF(*Bronchitis* = *yes*) = 6.14

BF(*Bronchitis* = *yes*, *Visit to Asia* = *no*) = 5.89

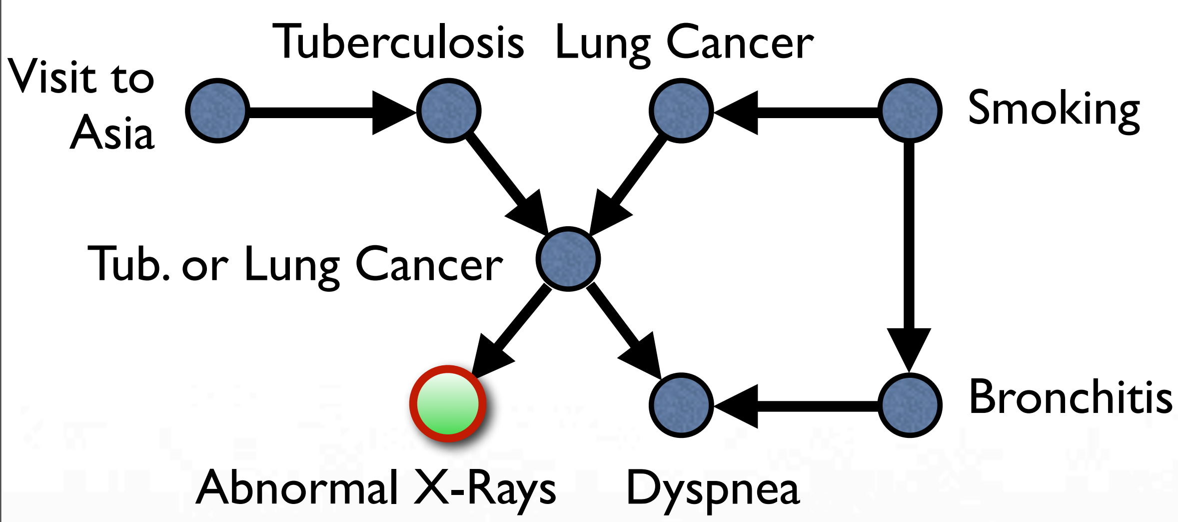
BF(*Bronchitis* = *yes*, *Tuberculosis* = *no*) = 5.84

MPE

P(*Bronchitis* = *yes*, *X-Ray* = *normal*,
Lung cancer = *no*, *Tuberculosis* = *no*,
Smoker = *yes*, *Visit to Asia* = *no*,
TbOrCa = *no* | *Dyspnea* = *yes*) = 0.46

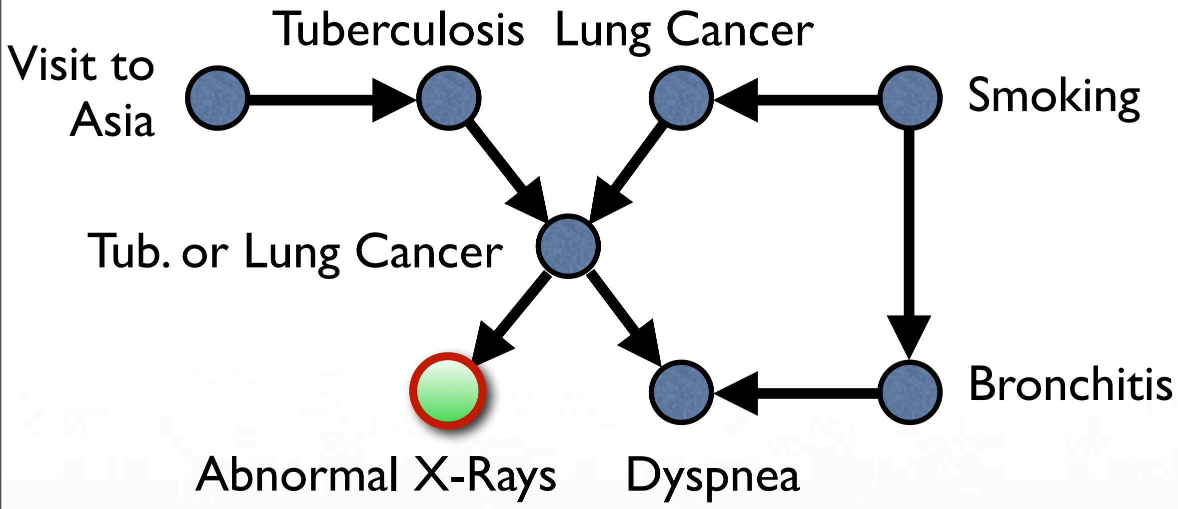
Asia: Example II

Explain: *X-Ray = abnormal*

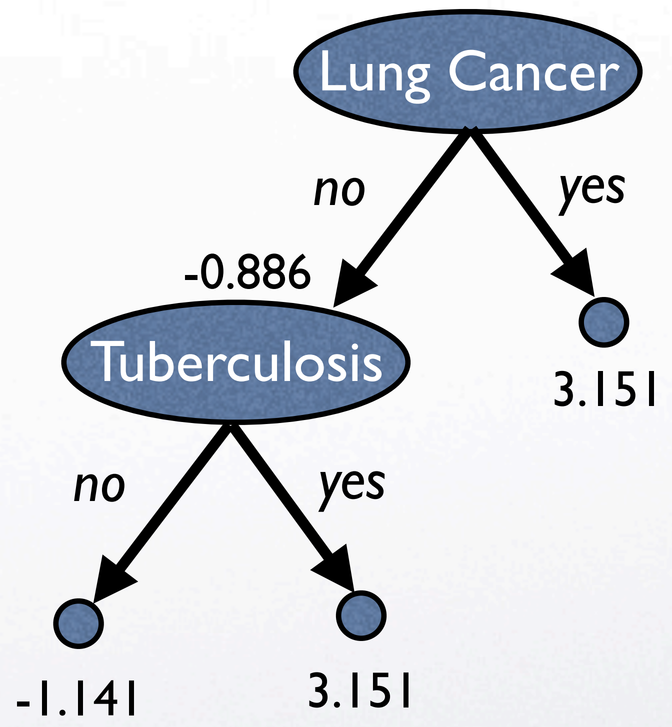


Asia: Example II

Explain: *X-Ray = abnormal*

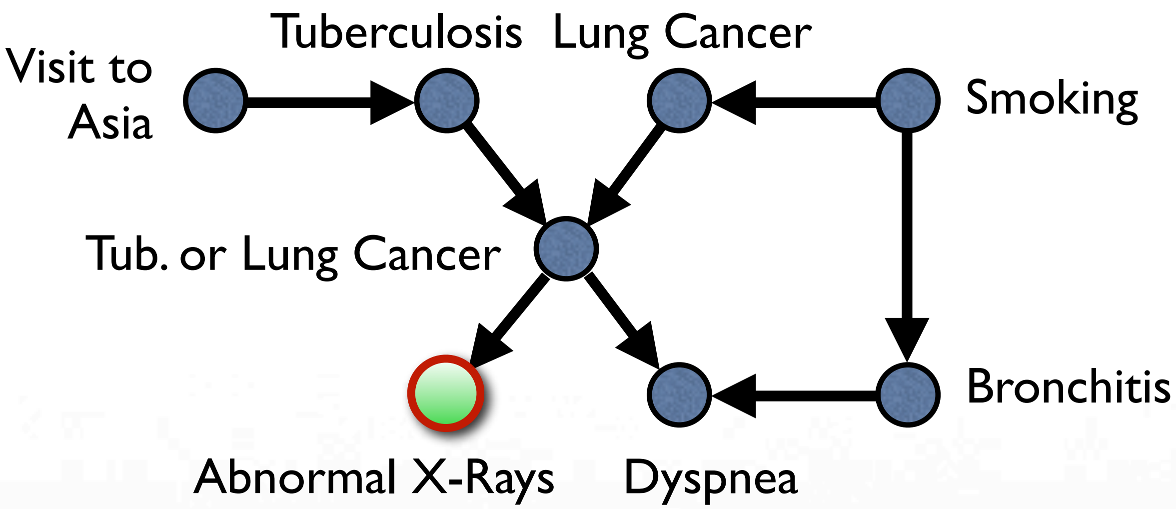


Causal explanation tree

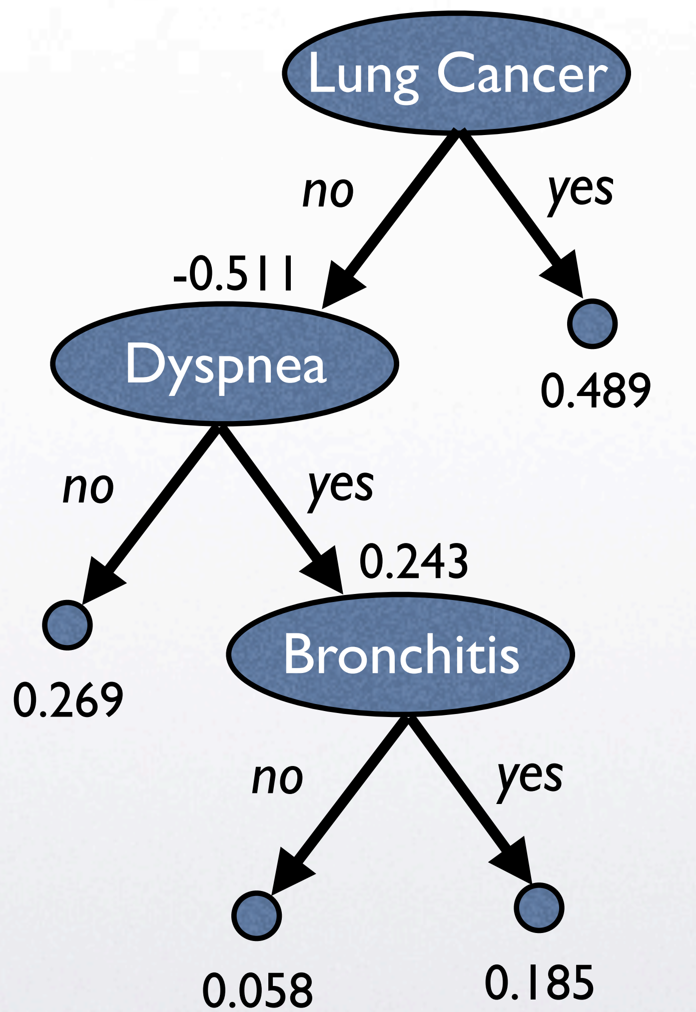


Asia: Example II

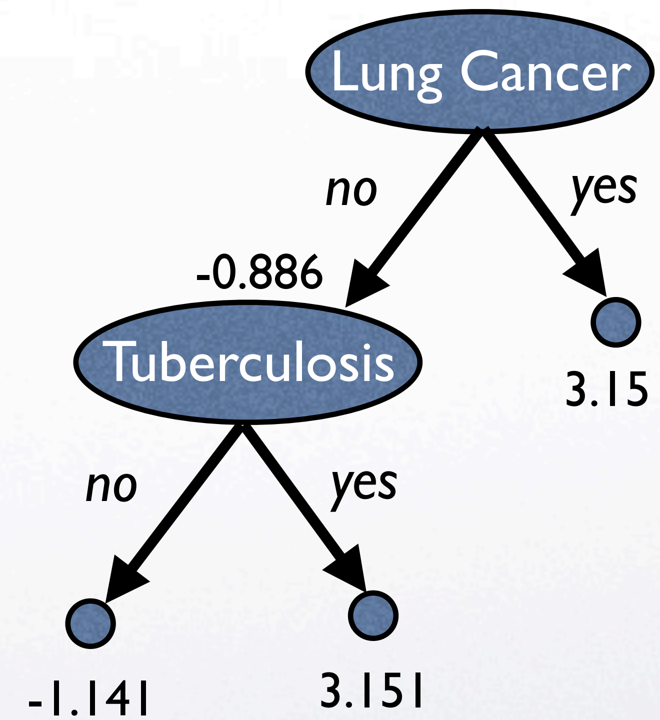
Explain: *X-Ray = abnormal*



Explanation Tree

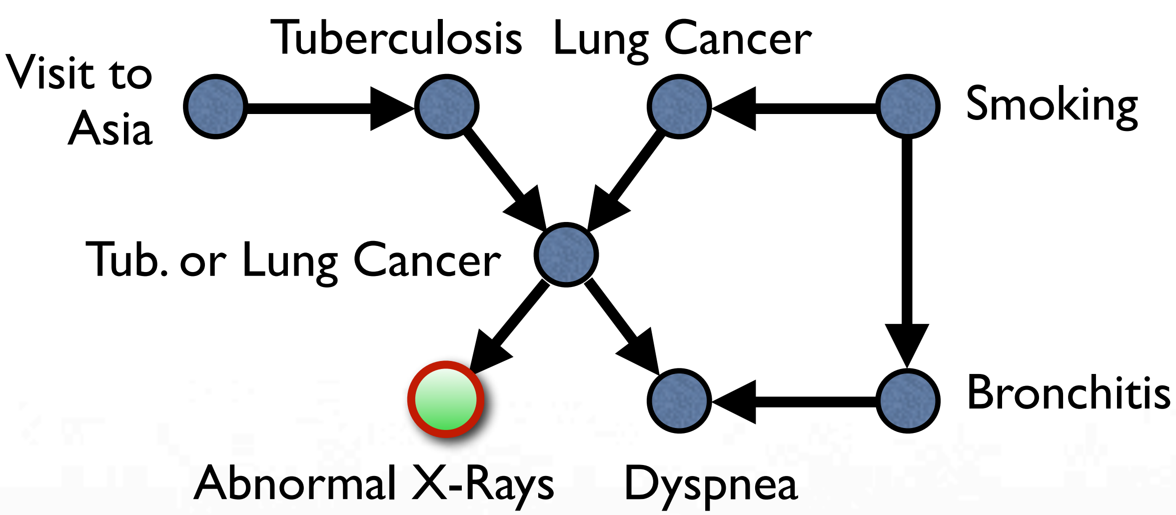


Causal explanation tree



Asia: Example II

Explain: *X-Ray = abnormal*



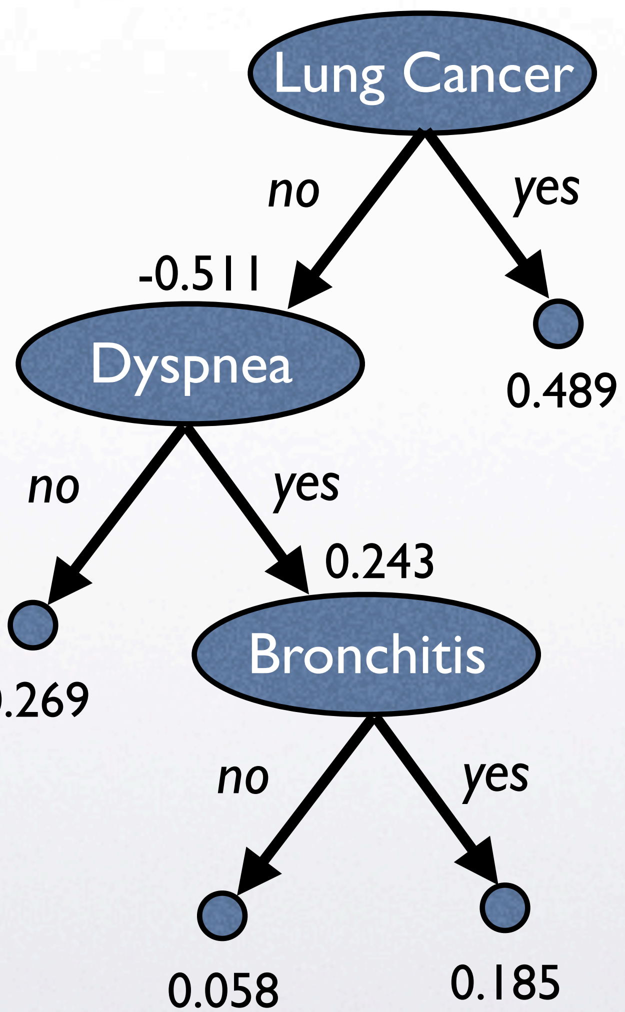
Bayes' Factor

BF(*TbOrCa* = *yes*) = 19.60

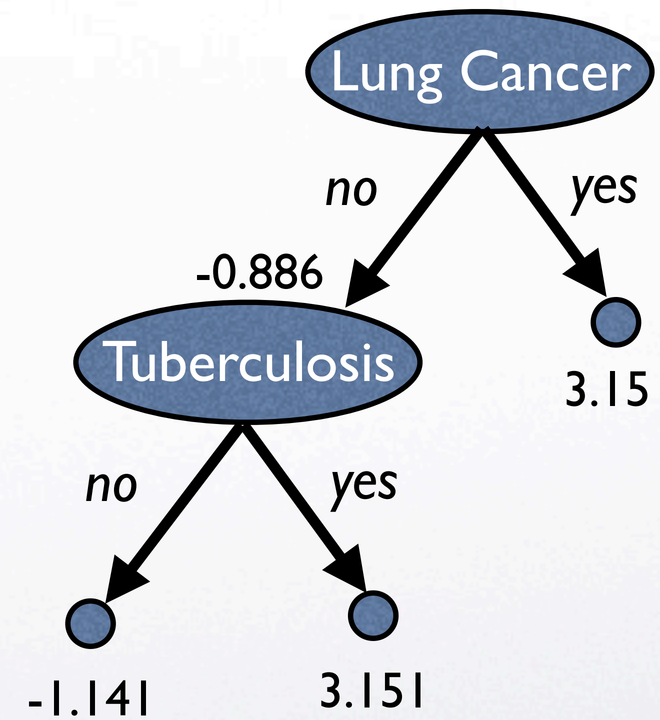
BF(*TbOrCa* = *yes*, *Visit to Asia* = *no*) = 19.21

BF(*TbOrCa* = *yes*, *Lung cancer* = *yes*) = 16.42

Explanation Tree

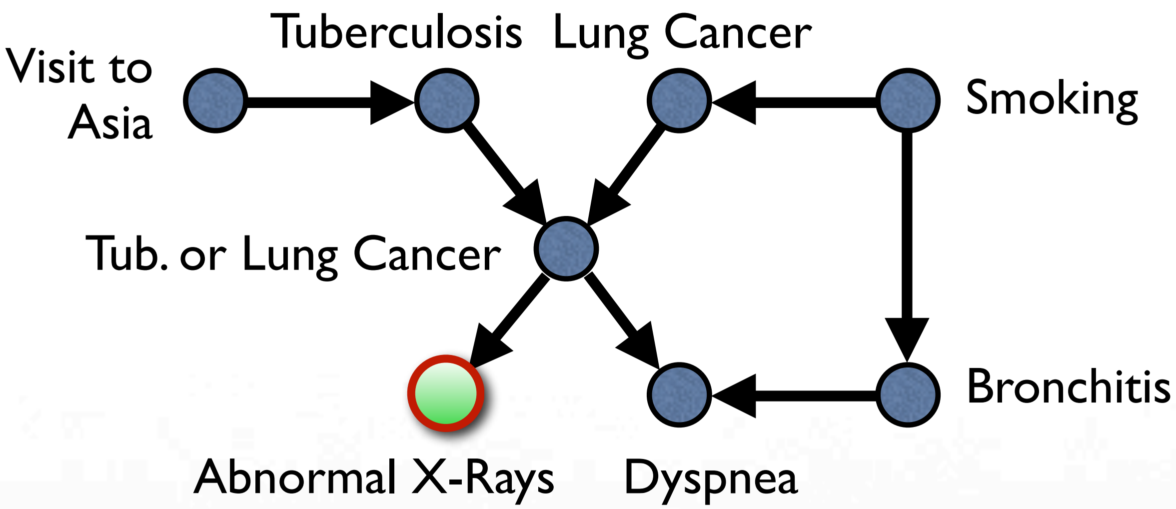


Causal explanation tree



Asia: Example II

Explain: *X-Ray = abnormal*



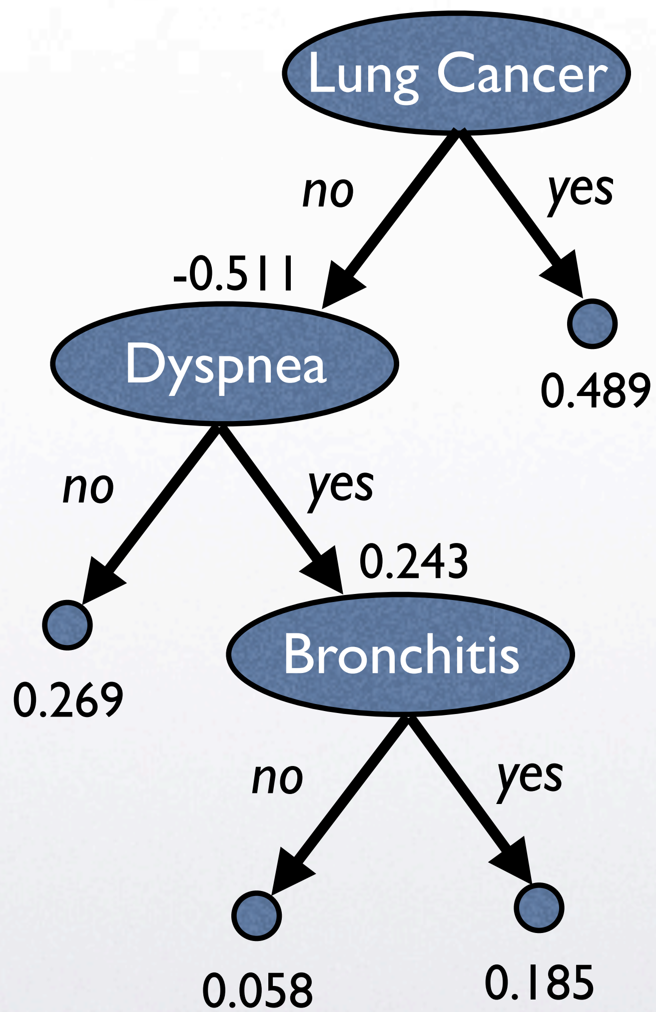
Bayes' Factor

BF(*TbOrCa* = *yes*) = 19.60
BF(*TbOrCa* = *yes*, *Visit to Asia* = *no*) = 19.21
BF(*TbOrCa* = *yes*, *Lung cancer* = *yes*) = 16.42

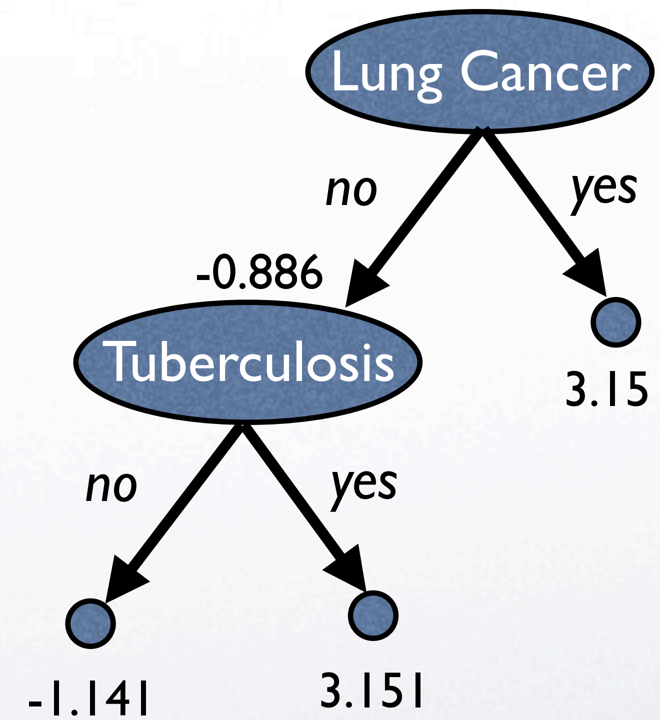
MPE

P(*Bronchitis* = *yes*, *Dyspnea* = *yes*,
Lung cancer = *yes*, *TbOrCa* = *yes*,
Smoker = *yes*, *Tuberculosis* = *yes*,
Visit to Asia = *no* | *X-Ray = abnormal*) = 0.24

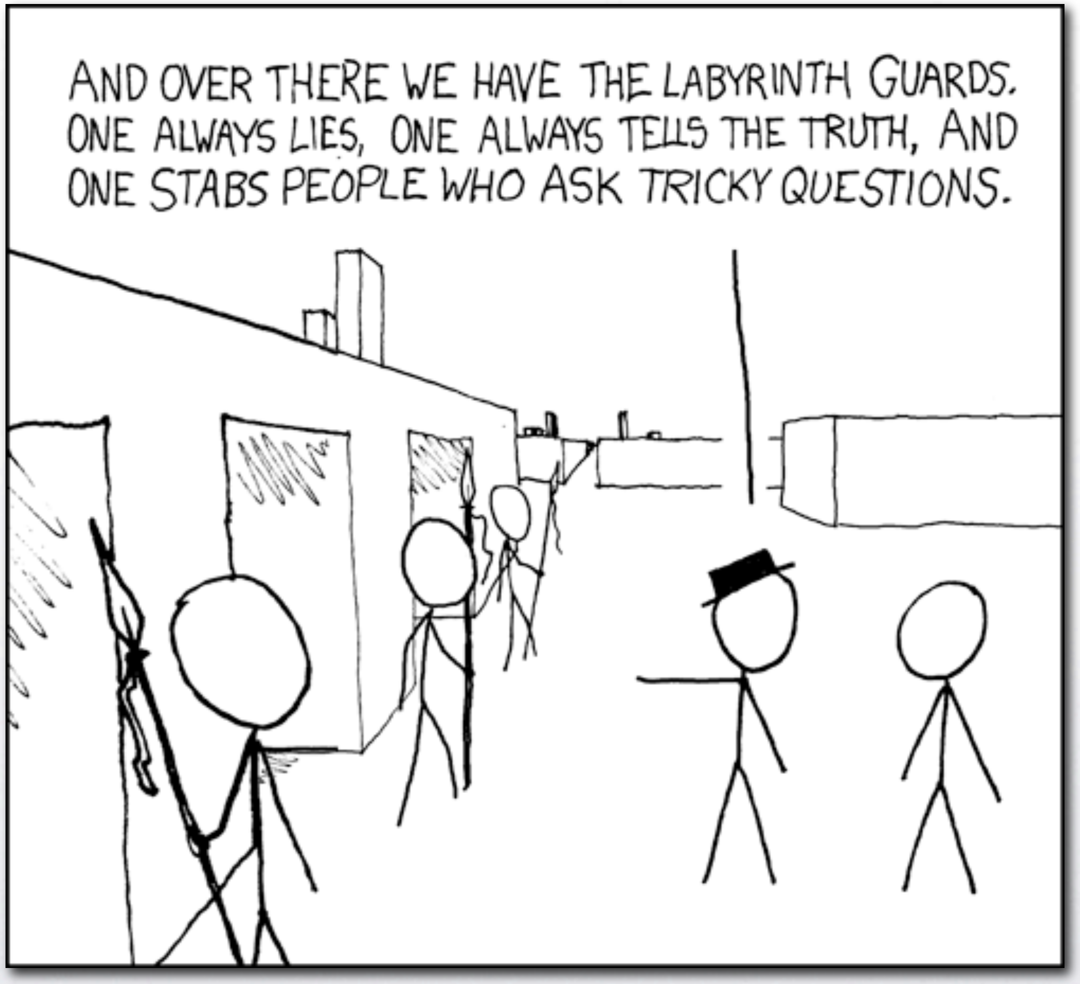
Explanation Tree



Causal explanation tree



Thank you!



xkcd.com