

# Causal Networks

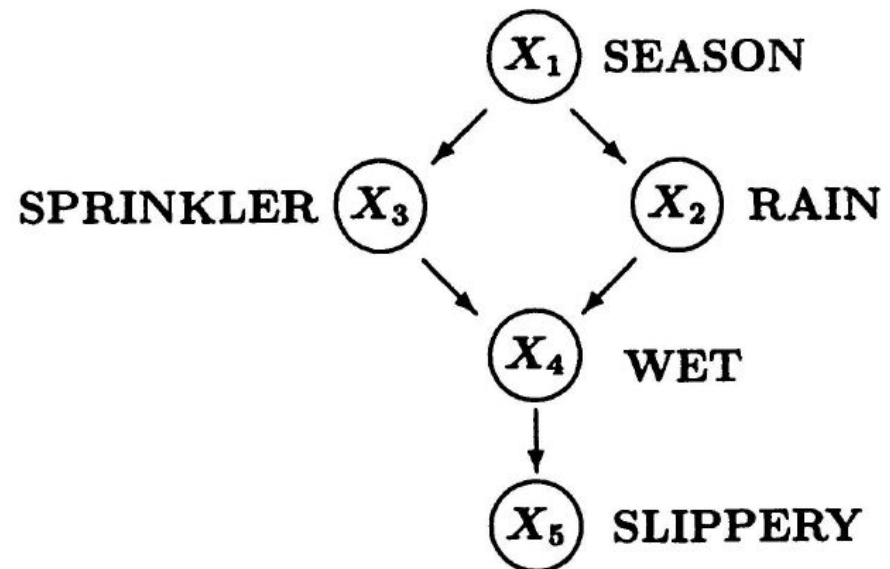
# Bayesian Networks are making „intelligent“ dialogs possible

Q1: If the season is dry,  
and the pavement is  
slippery, did it rain?

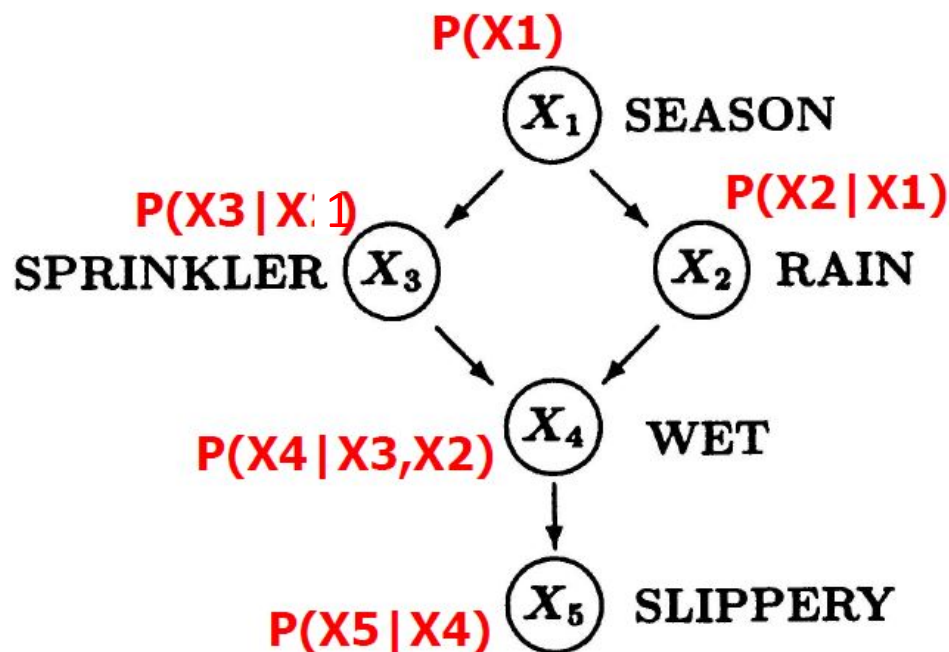
A1: Unlikely, it is more  
likely the sprinkler was  
ON.

Q2: But what if we SEE  
that the sprinkler is  
OFF?

A2: Then it is more likely  
that it rained



# Bayesian Networks are making „intelligent“ dialogs possible



X3	X2	wet=0	wet=1
0	0	0.9	0.1
0	1	0.1	0.9
1	0	0.2	0.8
1	1	0	1

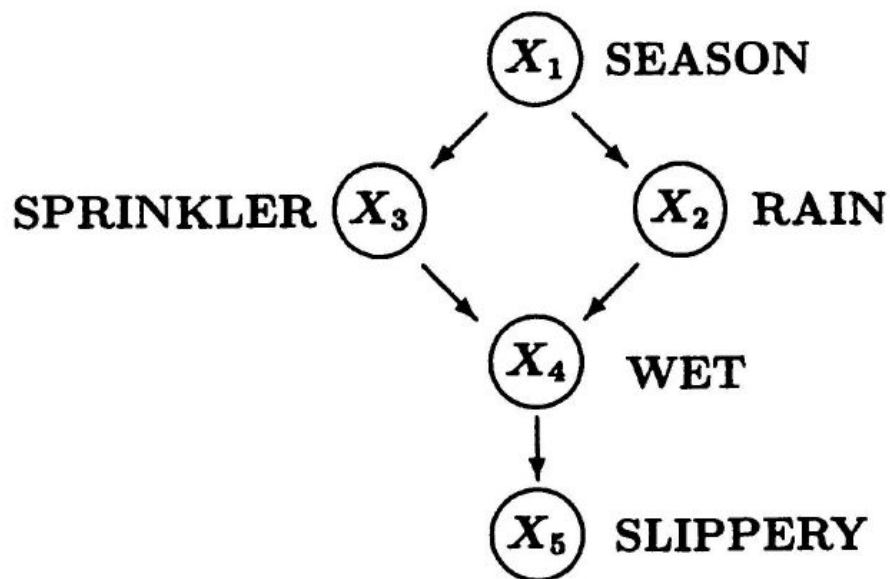
$$P(X_1, X_2, X_3, X_4, X_5) = P(X_1) P(X_2|X_1) P(X_3|X_1) P(X_4|X_3, X_2) P(X_5|X_4)$$

Conditional Independencies  $\longrightarrow$  Efficient Representation

# Bayesian Networks are making „intelligent“ dialogs possible

Q1: If the season is dry,  
and the pavement is  
slippery, did it rain?

Q2: But what if we SEE  
that the sprinkler is  
OFF?



$$P(X_1, X_2, X_3, X_4, X_5) = P(X_1) P(X_2|X_1) P(X_3|X_1) P(X_4|X_3, X_2) P(X_5|X_4)$$

Q1:  $\Pr(\text{rain}=\text{on} \mid \text{Slippery}=\text{yes}, \text{season}=\text{summer})?$

Q2:  $\Pr(\text{rain}=\text{on} \mid \text{Slippery}=\text{off}, \text{season}=\text{winter})?$

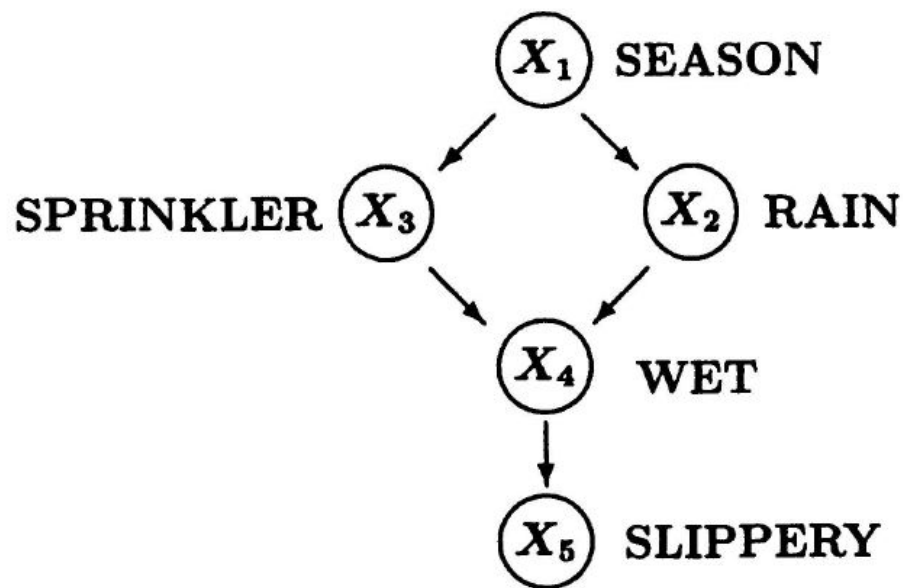
# Bayesian Networks are making „intelligent“ dialogs possible

Q2: But what if we SEE that the sprinkler is OFF?

A2: Then it is more likely that it rained

Q3: Do you mean that if we actually turn the sprinkler OFF, the rain will be more likely?

A3: No, the likelihood of rain would remain the same



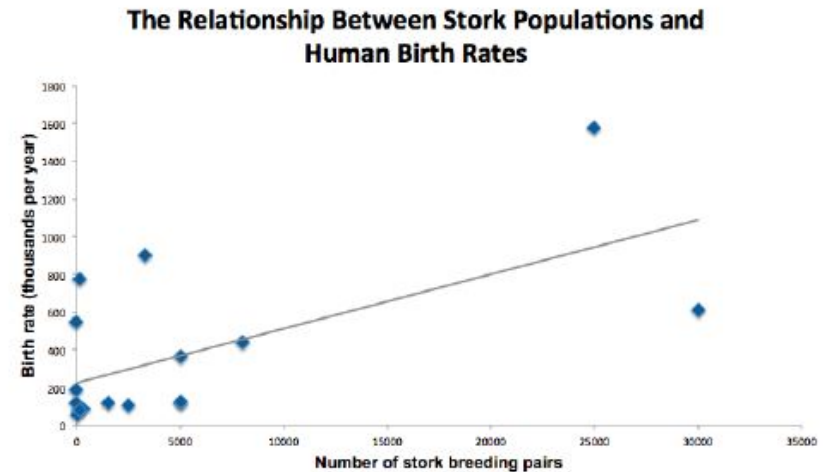
An **Observation** is different from an **Intervention**!

Bayesian Network model **associations/dependencies** and integrate **observations** (via **conditioning**).

In order to integrate **causalities** and **interventions**, we have to extend the concept.

# Do storks deliver babies?

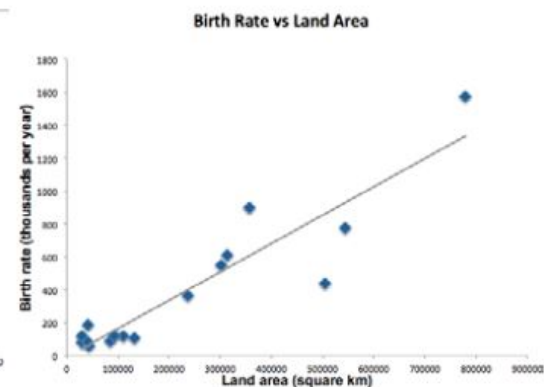
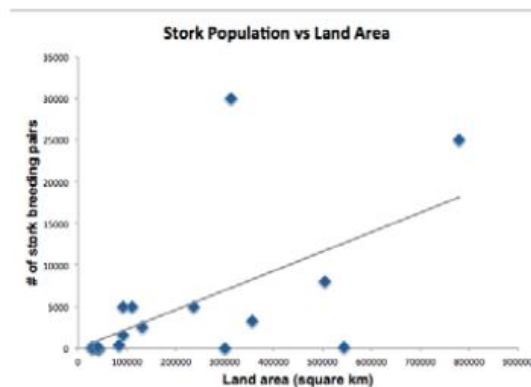
Country	Area (km <sup>2</sup> )	Storks (pairs)	Humans (10 <sup>6</sup> )	Birth rate (10 <sup>3</sup> /yr)
Albania	28,750	100	3.2	83
Austria	83,860	300	7.6	87
Belgium	30,520	1	9.9	118
Bulgaria	111,000	5000	9.0	117
Denmark	43,100	9	5.1	59
France	544,000	140	56	774
Germany	357,000	3300	78	901
Greece	132,000	2500	10	106
Holland	41,900	4	15	188
Hungary	93,000	5000	11	124
Italy	301,280	5	57	551
Poland	312,680	30,000	38	610
Portugal	92,390	1500	10	120
Romania	237,500	5000	23	367
Spain	504,750	8000	39	439
Switzerland	41,290	150	6.7	82
Turkey	779,450	25,000	56	1576



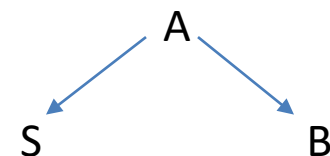
“Highly **statistically significant** degree of correlation between stork populations and birth rates” (or in technical terms,  $p = 0.008$ )

# Common Cause Principle

Country	Area (km <sup>2</sup> )	Storks (pairs)	Humans (10 <sup>6</sup> )	Birth rate (10 <sup>3</sup> /yr)
Albania	28,750	100	3.2	83
Austria	83,860	300	7.6	87
Belgium	30,520	1	9.9	118
Bulgaria	111,000	5000	9.0	117
Denmark	43,100	9	5.1	59
France	544,000	140	56	774
Germany	357,000	3300	78	901
Greece	132,000	2500	10	106
Holland	41,900	4	15	188
Hungary	93,000	5000	11	124
Italy	301,280	5	57	551
Poland	312,680	30,000	38	610
Portugal	92,390	1500	10	120
Romania	237,500	5000	23	367
Spain	504,750	8000	39	439
Switzerland	41,290	150	6.7	82
Turkey	779,450	25,000	56	1576



But a **simple variable that affects both** the birth rate and the stork population is the size of each country.



Common Cause Principle (Reichenbach, 1956): If there is a statistical dependence between variables X and Y (e.g S-B) then X causally influences Y, or Y causally influences X (e.g.  $S \rightarrow B$ ,  $B \rightarrow S$ ), or there exists Z causally influencing both (e.g.  $A \rightarrow B$  and  $A \rightarrow S$ ). We use this Reichenbach assumption in the following.

In practice, causalities may also arise for other reasons, e.g. when the variables follow a physical law and then only look as if they depend on each other.

# How to treat scurvy?

Experiment by James Lind (18th century)

- 12 scorput sailors treated with different aids (vinegar, cider, lemon)
- Only the sailors treated by lemon improved

Experiment: Data of 80 sailors with respect to Treatment, Recovery and Age

Combined	Recovery	No Recovery	Total	Recovery Rate
No lemons	20	20	40	50 %
Lemons	16	24	40	40 %
Total	36	44	80	

$$P(\text{recovery}|\text{lemons}) < P(\text{recovery}|\text{no lemons})$$

Old	Recovery	No Recovery	Total	Recovery Rate
No lemons	2	8	10	20 %
lemons	9	21	30	30 %
Total	11	29	40	

$$P(\text{recovery}|\text{lemons, old}) > P(\text{recovery}|\text{no lemons, old})$$

Young	Recovery	No Recovery	Total	Recovery Rate
No lemons	18	12	30	60 %
Lemons	7	3	10	70 %
Total	25	15	40	

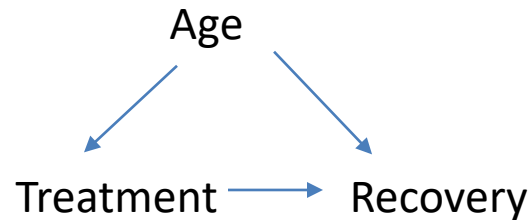
$$P(\text{recovery}|\text{lemons, young}) > P(\text{recovery}|\text{no lemons, young})$$

Simpsons Paradoxon: Reversal of association after considering a third variable

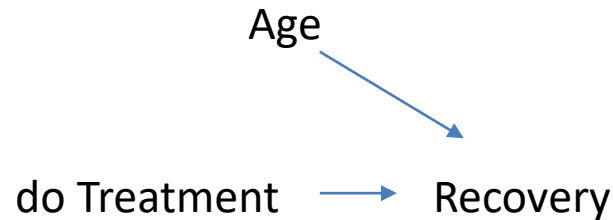


# Resolving the Paradoxon

Association Net (G,P)



Intervention Net ( $G_{do(T)}$ ,  $P_{do(T)}$ )



In  $G$  all possible causalities are modelled.

In  $G_{do(T)}$  all influences stemming from „natural causes“ of the intervention variables (Treatment) are removed.

- $P(\text{recovery} \mid \text{lemons})$  ist not the same as  $P_{do(\text{lemons})}(\text{recovery})$   
This is a common misinterpretation, the reason of the Simpson paradoxon
- We should treat scurvy with lemons if  
 $P_{do(\text{lemons})}(\text{recovery}) > P_{do(\text{no lemons})}(\text{recovery})$

# Using the do-operator

## Learn the Bayesian Network (G,P) from data (parametric estimation using the tables)

$$P(\text{old})=P(\text{young}) = 40/80=0.5$$

$$P(\text{lemons} \mid \text{old}) = (9+21)/40= 0.75, P(\text{lemons} \mid \text{young}) = (7+3)/40 =0.25$$

$$P(\text{recovery} \mid \text{old}, \text{lemons}) = 0.3 \quad P(\text{recovery} \mid \text{old}, \text{no lemons})= 0.2$$

$$P(\text{recovery} \mid \text{young}, \text{lemons}) = 0.7 \quad P(\text{recovery} \mid \text{young}, \text{no lemons}) = 0.6$$

Using only observations gives

$$0.4 = P(\text{recovery} \mid \text{lemons}) < P(\text{recovery} \mid \text{no lemons}) = 0.5$$

## Intervention Network: Estimate $P_{\text{do}(T)}$ (recovery) by using information about (G,P) and $G_{\text{do}(T)}$

$$\begin{aligned} P_{\text{do}(\text{lemon})}(\text{recovery}) &= P_{\text{do}(\text{lemon})}(\text{recovery}, \text{lemons}, \text{young}) + P_{\text{do}(\text{lemon})}(\text{recovery}, \text{lemons}, \text{old}) = \\ &= P_{\text{do}(\text{lem})}(\text{rec} \mid \text{lem}, \text{young}) P_{\text{do}(\text{lem})}(\text{lem}, \text{young}) + P_{\text{do}(\text{lem})}(\text{rec} \mid \text{lem}, \text{old}) P_{\text{do}(\text{lem})}(\text{lem}, \text{old}) \\ &= P_{\text{do}(\text{lem})}(\text{rec} \mid \text{lem}, \text{young}) P_{\text{do}(\text{lem})}(\text{young}) + P_{\text{do}(\text{lem})}(\text{rec} \mid \text{lem}, \text{old}) P_{\text{do}(\text{lem})}(\text{old}) \quad (\text{treatment lemon}) \\ &= P(\text{rec} \mid \text{lem}, \text{young}) P(\text{young}) + P(\text{rec} \mid \text{lem}, \text{old}) P(\text{old}) \quad (\text{age as influence for treatment is removed}) \\ &= 0.7 \times 0.5 + 0.3 \times 0.5 = 0.5 \end{aligned}$$

$$P_{\text{do}(\text{no lemon})}(\text{recovery}) = 0.4 \text{ (with the same method)}$$

We should treat scurvy with lemons.

Note that the same method was used for the famous data set for kidney stone recovery with 700 patients and two treatments in 1986.

# Causality versus Correlation

It's hard to separate out causality from correlation

DAGs can be viewed as a causal process: the parents "cause" the children to take different values

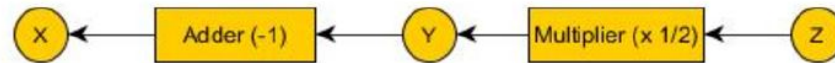
The below equations are equivalent and the graphs have same conditional independences, but the causalities are not the same. Graphs tells us something useful that equations miss.

Structural Equation Modeling (Structure Causal Models) is very popular in economics

$$Y=X+1, Z=Y*2$$



$$Y=Z/2, X=Y-1$$



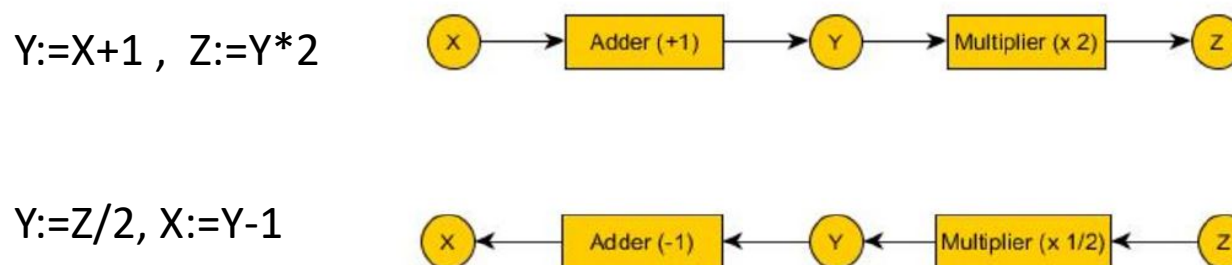
# Causality versus Correlation

It's hard to separate out causality from correlation

DAGs can be viewed as a causal process: the parents "cause" the children to take different values

The below equations are equivalent and the graphs have same conditional independences, but the causalities are not the same. Graphs tells us something useful that equations miss.

Structural Equation Modeling (Structure Causal Models) is very popular in economics



There is *observational data* ("seeing") and *interventional data* ("doing")

Usually the DAG is designed for observational data, but that ignores the possibility of unobserved variables, also without interventional data you can't distinguish the direction of causality.

Simplest external intervention: a single variable is forced to take some fixed value (in a graph remove arrows entering that variable)

# Pearl's Causal Calculus

## The Causal Calculus (do-calculus, Pearl's Causal Calculus, Calculus of Actions)

Shortly: Calculus to discuss causality in a formal language by Judea Pearl

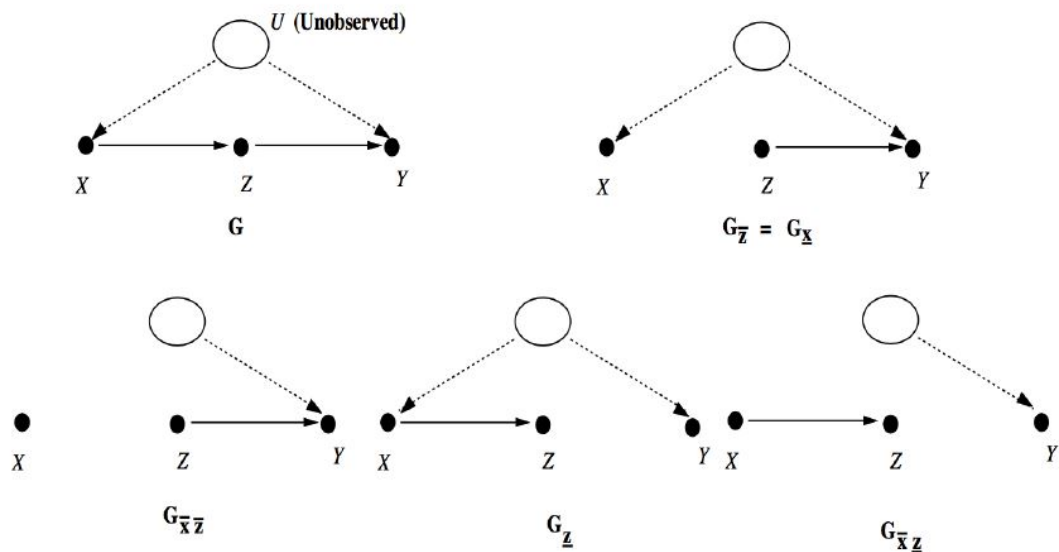
A new operator,  $do()$ , marks an action or an intervention in the model. In an algebraic model we replace certain functions with a constant  $X = x$ , and in a graph we remove edges going into the target of intervention, but preserve edges going out of the target.

The causal calculus uses Bayesian conditioning,  $p(y|x)$ , where  $x$  is observed variable, and causal conditioning,  $p(y|do(x))$ , where an action is taken to force a specific value  $x$ .

Goal is to generate probabilistic formulas for the effect of interventions in terms of the observed probabilities.

# Pearl's Causal Calculus

## Notations



Notation: a graph  $G$ ,  $W$ ,  $X$ ,  $Y$ ,  $Z$  are disjoint subsets of the variables.  $G_{\bar{X}}$  denotes the perturbed graph in which all edges pointing to  $X$  have been deleted, and  $G_{\bar{X}}$  denotes the perturbed graph in which all edges pointing from  $X$  have been deleted.  $Z(W)$  denote the set of nodes in  $Z$  which are not ancestors of  $W$

---

Image: Judea Pearl

# Pearl's Causal Calculus

## Pearl's 3 rules

Note that we use the short notation of Pearl in the following, i.e.

$$P(y \mid \underline{do}(x), \underline{z}, w) = P_{\underline{do}(x)}(y \mid z, w)$$

- ▶ Ignoring observations

$$p(y \mid \underline{do}(x), z, w) = p(y \mid \underline{do}(x), w) \text{ if } (Y \perp\!\!\!\perp Z \mid X, W)_{G_{\overline{X}}}$$

- ▶ Action/Observation exchange (the back-door criterion)

$$p(y \mid \underline{do}(x), \underline{do}(z), w) = p(y \mid \underline{do}(x), z, w) \text{ if } (Y \perp\!\!\!\perp Z \mid X, W)_{G_{\overline{X}, \underline{Z}}}$$

- ▶ Ignoring actions/interventions

$$p(y \mid \underline{do}(x), \underline{do}(z), w) = p(y \mid \underline{do}(x), w) \text{ if } (Y \perp\!\!\!\perp Z \mid X, W)_{G_{\overline{X}, \overline{Z(W)}}}$$

Notation: a graph  $G$ ,  $W$ ,  $X$ ,  $Y$ ,  $Z$  are disjoint subsets of the variables.  $G_{\overline{X}}$  denotes the perturbed graph in which all edges pointing to  $X$  have been deleted, and  $G_{\underline{X}}$  denotes the perturbed graph in which all edges pointing from  $X$  have been deleted.  $Z(W)$  denote the set of nodes in  $Z$  which are not ancestors of  $W$

# Pearl's Causal Calculus

## Intuition behind the Pearl's first rule

With condition  $(Y \perp\!\!\!\perp Z|X, W)_{G_{\overline{X}}}$  we have

$$p(y|do(x), z, w) = p(y|do(x), w)$$

- ▶ Let's start with a simple case where we assume that there are no  $W$  or  $X$ . We get a condition  $(Y \perp\!\!\!\perp Z)_G$ , so  $Y$  is independent of  $Z$ , that is,  $p(y|z) = p(y)$
- ▶ In the second case assume we have passively observed  $W$ , but no variable  $X$ :  $(Y \perp\!\!\!\perp Z|W)_G$ . Earlier we mentioned connection of d-separation and conditionally independent, that is,  $p(y|z, w) = p(y|w)$
- ▶ The third case assume we don't know  $W$ , but we have  $X$  that's value is set by intervention:  $(Y \perp\!\!\!\perp Z|X)_{G_{\overline{X}}}$ . By the same theorem, that is,  $p(y|z, do(x)) = p(y|do(x))$

Combining these gives the full rule.



# Pearl's Causal Calculus

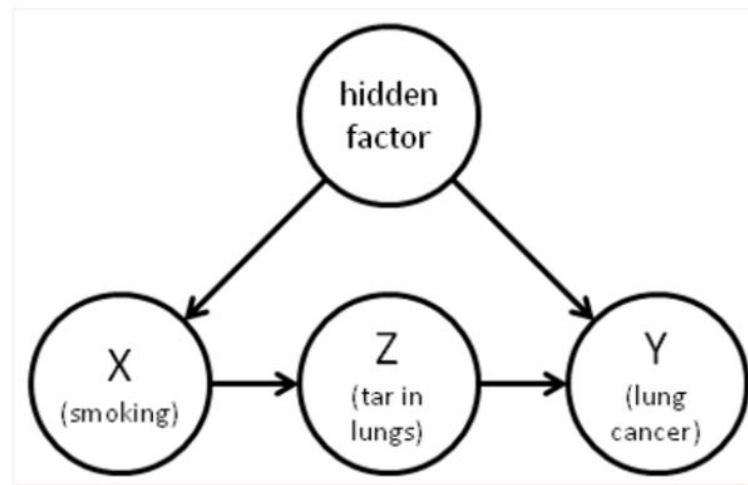
## Example: Smoking and lung cancer

### Randomized Controlled Trials (RCT)

- ▶ AKA Randomized Control Trial, Randomized clinical trial
- ▶ The participants in the trial are randomly allocated to either the group receiving the treatment under investigation or to the control group
- ▶ The control group removes the confounding factor of the placebo effect
- ▶ Double-blind studies remove further confounding factors
- ▶ Sometimes impractical or impossible

We can try to use causal calculus to analyze the probability that someone would get cancer given that they are smoking, without doing an actual RCT:

$$p(y|do(x))$$

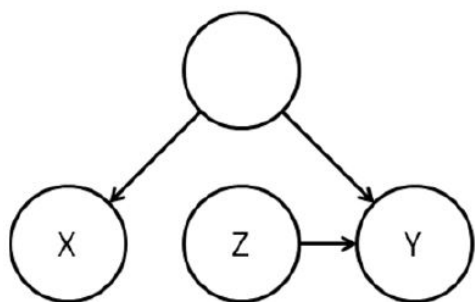


Note: We have no information about the hidden variable that could cause both smoking and cancer

# Pearl's Causal Calculus

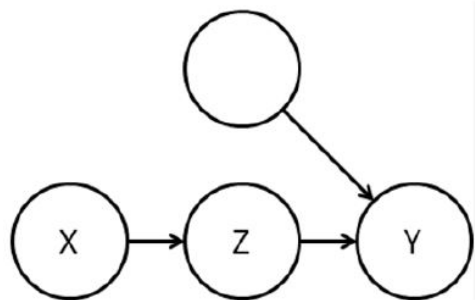
We can't try to apply rule 1 because there is no observations to ignore, we would just have  $p(y|do(x)) = p(y|do(x))$ .

Try apply rule 2: We would have  $p(y|do(x)) = p(y|x)$ , that is, the intervention doesn't matter. It's condition is  $(Y \perp\!\!\!\perp X)_{G_{\underline{X}}}$ :



$Y$  and  $X$  are not d-separated, because they have a common ancestor.  
 $\implies$  Rule 2 can't be applied

Try apply rule 3: We would have  $p(y|do(x)) = p(y)$ , that is, an intervention to force someone to smoke has no impact on whether they get cancer. It's condition is  $(Y \perp\!\!\!\perp X)_{G_{\overline{X}}}$ :

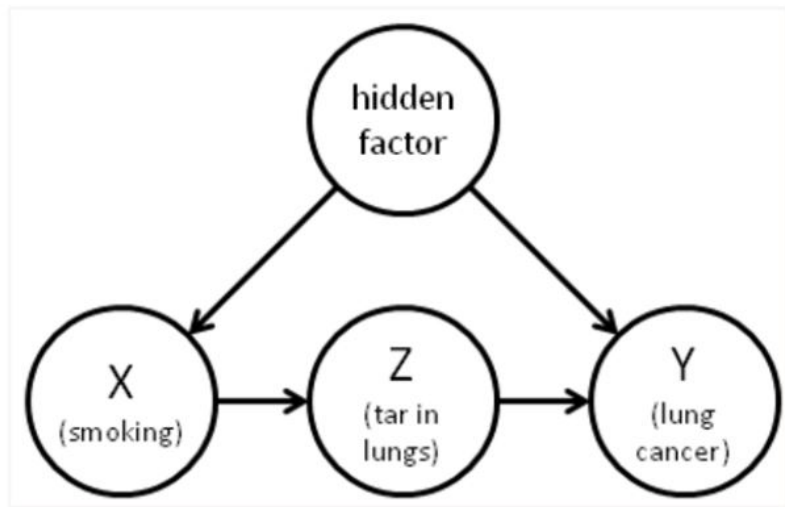


$Y$  and  $X$  are not d-separated, because we have unblocked path between them.  
 $\implies$  Rule 3 can't be applied

# Pearl's Causal Calculus

New attempt:

$$\begin{aligned} p(y|do(x)) &= \sum_z p(y|z, do(x))p(z|do(x)) \\ &= \sum_z p(y|z, do(x))p(z|x) && \text{(rule 2: } (Z \perp\!\!\!\perp X)_{G_{\underline{X}}} \text{)} \\ &= \sum_z p(y|do(z), do(x))p(z|x) && \text{(rule 2: } (Y \perp\!\!\!\perp Z|X)_{G_{\overline{X}, \underline{Z}}} \text{)} \\ &= \sum_z p(y|do(z))p(z|x) && \text{(rule 3: } (Y \perp\!\!\!\perp X|Z)_{G_{\overline{Z}, \overline{X}}} \text{)} \end{aligned}$$



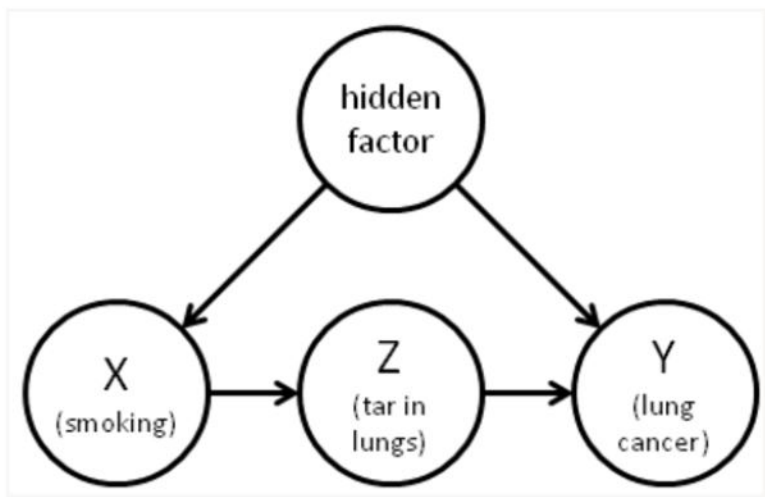
# Pearl's Causal Calculus

We can use the same approach to the first term on the right hand side:

$$\begin{aligned} p(y|do(z)) &= \sum_x p(y|x, do(z))p(x|do(z)) \\ &= \sum_x p(y|x, z)p(x) \end{aligned} \quad (\text{rule 2 + rule 3})$$

Finally we can combine these results:

$$p(y|do(x)) = \sum_{z, x'} p(y|x', z)p(z|x)p(x')$$



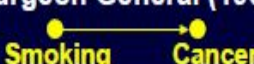
We can now compare  $p(y)$  and  $p(y|do(x))$ . The needed probabilities can be observed directly from experimental data: What part of smokers have lung cancer, how many of them have tar in their lungs etc.

# Pearl's Causal Calculus

**PREDICTING THE EFFECTS OF POLICIES**

---


1. Surgeon General (1964):



$P(c | do(s)) \approx P(c | s)$

---


2. Tobacco Industry:



$P(c | do(s)) = P(c)$

---


3. Combined:



$P(c | do(s)) = \text{noncomputable}$

---

4. Combined and refined:



$P(c | do(s)) = \text{computable}$

## Remarks about the do-calculus

- ▶ The analysis would have not worked if the graph had missed the tar variable,  $Z$ , because there is no general way to compute  $p(y|do(x))$  from any observed distributions whenever the causal model includes subgraph shown the figure below
- ▶ Causal Calculus can be used to analyze causality in more complicated (and more unethical) situations than RCT
- ▶ Causal Calculus can also be used to test whether unobserved variables are missed by removing all do terms from the relation
- ▶ Not all models are acyclic.

## General Properties of the do-calculus

The do-calculus is complete: Whenever a causal effect is estimatable from data, a sequence of steps using the three rules eliminates the do-operator.

The corresponding decision problem is tractable: Shpitser's algorithm decides if a solution exist in polynomial time.

Judea Pearl's "The Book of Why", 2018, coauthored with the journalist Dana MacKenzie describes the story "Is there scientific evidence on the health benefits of quitting smoking?". You'll find a good motivation about causation in future intelligent Systems in this award winning book.

# Pearl's Causal Calculus

The do-calculus is complete: Whenever a causal effect is estimatable from data, a sequence of steps using the three rules eliminates the do-operator.

The corresponding decision problem is tractable: Shpitser's algorithm decides if a solution exist in polynomial time.

Judea Pearl's "The Book of Why", 2018, coauthored with the journalist Dana MacKenzie describes the story "Is there scientific evidence on the health benefits of quitting smoking?"  
and also his vision about causation in intelligent Systems

# Mechanization of cognitive abilities with probabilistic networks

## **Associations (“Observation”)** Bayes, Belief, Conditional Probabilities

Are headaches an indication of a brain tumor? How likely is it that a customer who buys a Golf also wants a towbar?

## **Intervention (“Action”)** Do Operator , Model Revision, Causality

If I take aspirin, will my headaches be cured? What happens to our sales if we increase the price by 10%?

## **Imagination (“Understanding”)** Counterfactuals, Mining worlds that could have been

Was it aspirin that stopped my headache? What if I hadn't smoked in the last 2 years? What is the probability that a customer who bought a Golf would have bought the car if we had increased the price by 10% at the time?