

# CAUSAL INFERENCE AS COMPUTATIONAL LEARNING

Judea Pearl  
University of California  
Los Angeles  
([www.cs.ucla.edu/~judea](http://www.cs.ucla.edu/~judea))

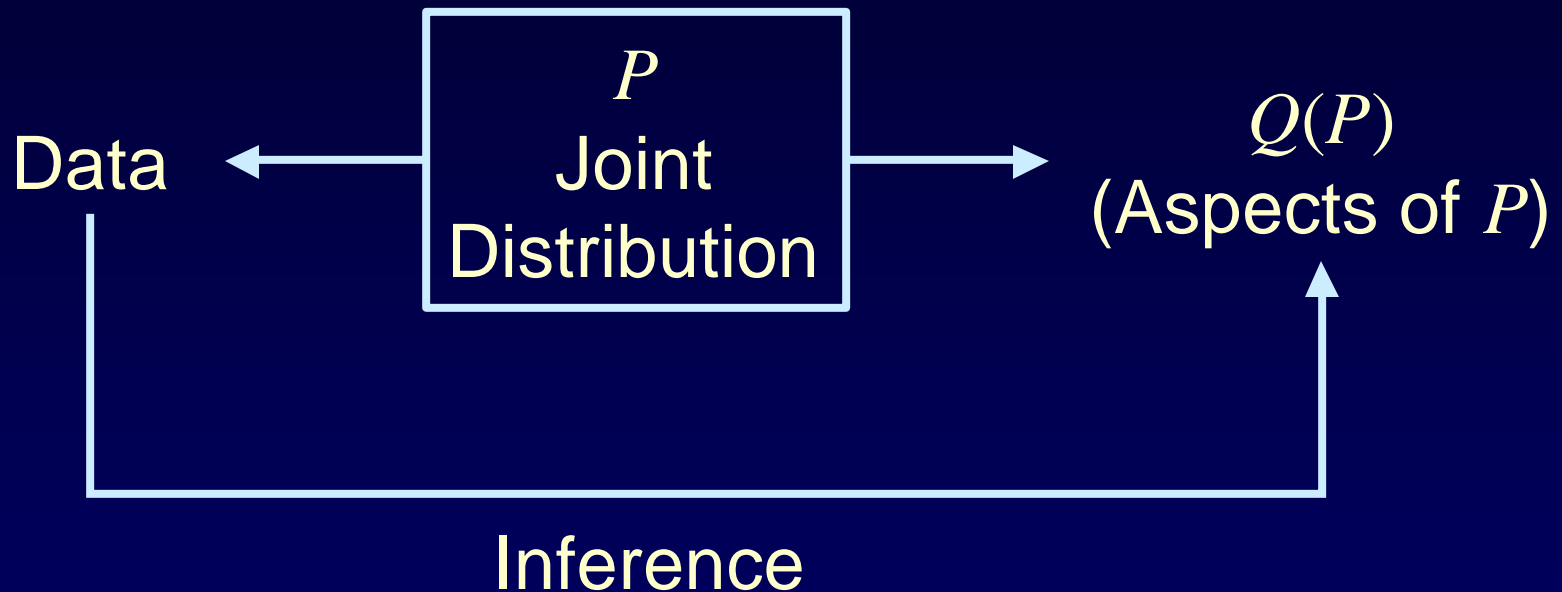
# OUTLINE

---

- Inference: Statistical vs. Causal distinctions and mental barriers
- Formal semantics for counterfactuals: definition, axioms, graphical representations
- Inference to three types of claims:
  1. Effect of potential interventions
  2. Attribution (Causes of Effects)
  3. Direct and indirect effects

# TRADITIONAL STATISTICAL INFERENCE PARADIGM

---



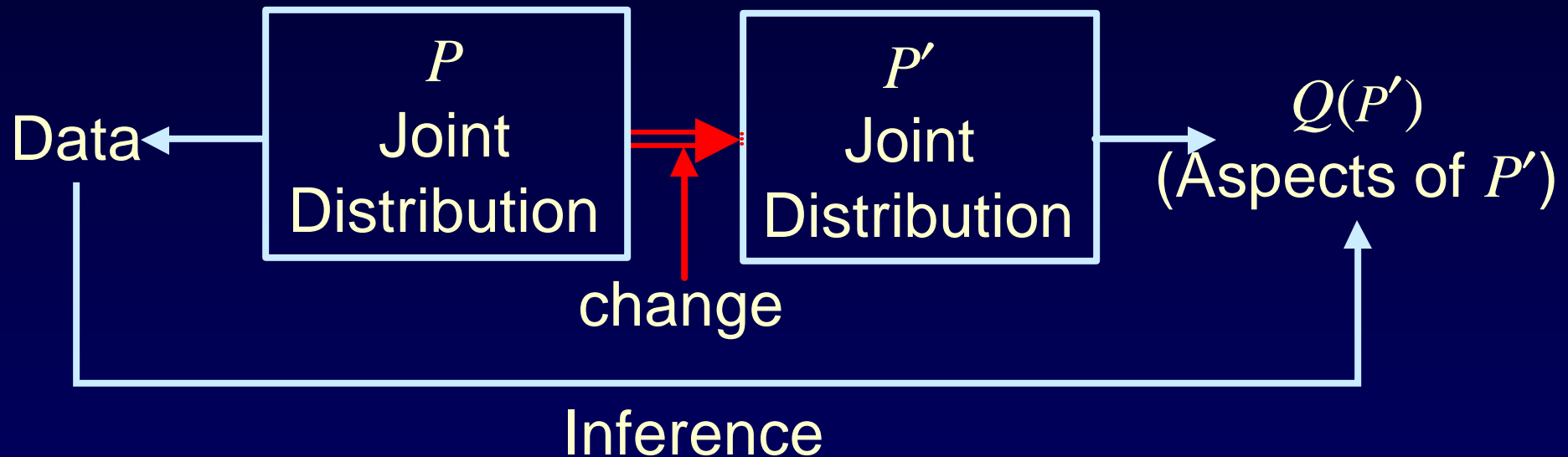
e.g.,

Infer whether customers who bought product  $A$  would also buy product  $B$ .

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

# FROM STATISTICAL TO CAUSAL ANALYSIS: 1. THE DIFFERENCES

Probability and statistics deal with static relations



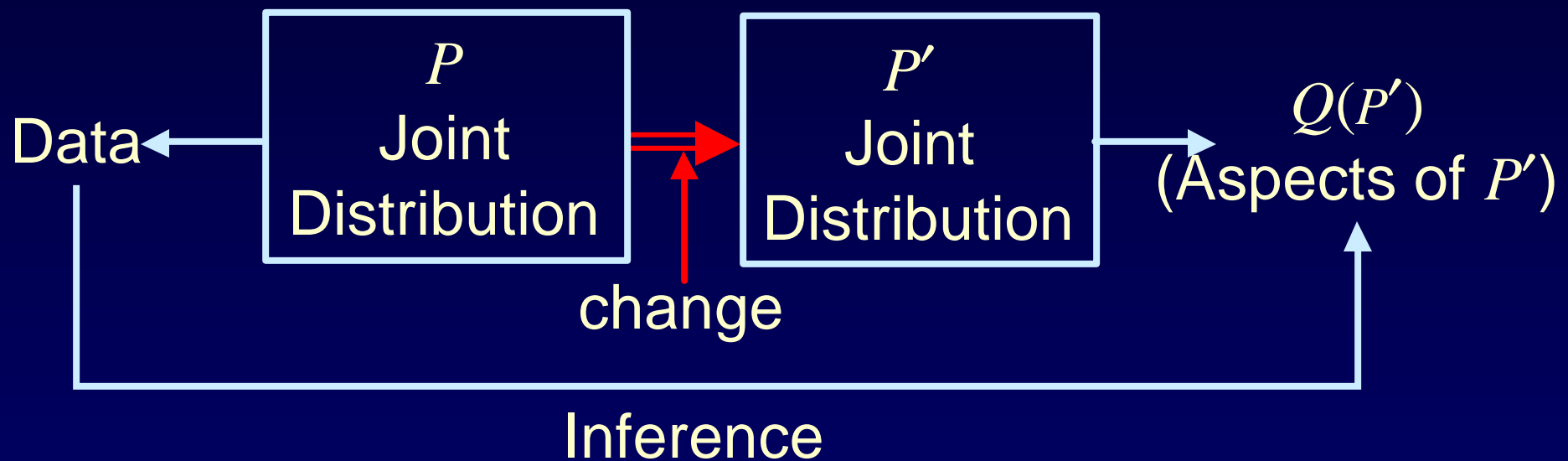
What happens when  $P$  changes?

e.g.,

Infer whether customers who bought product  $A$  would still buy  $A$  if we were to double the price.

# FROM STATISTICAL TO CAUSAL ANALYSIS: 1. THE DIFFERENCES

What remains invariant when  $P$  changes say, to satisfy  
 $P\mathcal{C}(price=2)=1$



Note:  $P\mathcal{C}(v) \neq P(v | price = 2)$

$P$  does not tell us how it ought to change  
e.g. Curing symptoms vs. curing diseases  
e.g. Analogy: mechanical deformation

# FROM STATISTICAL TO CAUSAL ANALYSIS: 1. THE DIFFERENCES (CONT)

---

1. Causal and statistical concepts do not mix.

## CAUSAL

Spurious correlation  
Randomization  
Confounding / Effect  
Instrument  
Holding constant  
Explanatory variables

## STATISTICAL

Regression  
Association / Independence  
“Controlling for” / Conditioning  
Odd and risk ratios  
Collapsibility  
Propensity score

2.

3.

4.

# FROM STATISTICAL TO CAUSAL ANALYSIS: 2. MENTAL BARRIERS

---

1. Causal and statistical concepts do not mix.

## CAUSAL

Spurious correlation  
Randomization  
Confounding / Effect  
Instrument  
Holding constant  
Explanatory variables

## STATISTICAL

Regression  
Association / Independence  
“Controlling for” / Conditioning  
Odd and risk ratios  
Collapsibility  
Propensity score

2. **No causes in – no causes out** (Cartwright, 1989)

statistical assumptions + data  
causal assumptions }  $\Rightarrow$  causal conclusions

3. Causal assumptions cannot be expressed in the mathematical language of standard statistics.

- 4.

# FROM STATISTICAL TO CAUSAL ANALYSIS: 2. MENTAL BARRIERS

---

1. Causal and statistical concepts do not mix.

## CAUSAL

Spurious correlation  
Randomization  
Confounding / Effect  
Instrument  
Holding constant  
Explanatory variables

## STATISTICAL

Regression  
Association / Independence  
“Controlling for” / Conditioning  
Odd and risk ratios  
Collapsibility  
Propensity score

2. **No causes in – no causes out** (Cartwright, 1989)

statistical assumptions + data  
causal assumptions }  $\Rightarrow$  causal conclusions

3. Causal assumptions cannot be expressed in the mathematical language of standard statistics.
4. **Non-standard mathematics:**
  - a) Structural equation models (Wright, 1920; Simon, 1960)
  - b) Counterfactuals (Neyman-Rubin ( $Y_x$ ), Lewis ( $x \boxrightarrow Y$ ))

# WHY CAUSALITY NEEDS SPECIAL MATHEMATICS

---

Scientific Equations (e.g., Hooke's Law) are non-algebraic

e.g., Pricing Policy: "Double the competitor's price"

Correct notation:

$$Y := 2X$$

$$X = 1$$

Process information

$$X = 1$$

$$Y = 2$$

The solution

Had  $X$  been 3,  $Y$  would be 6.

If we raise  $X$  to 3,  $Y$  would be 6.

Must "wipe out"  $X = 1$ .

# WHY CAUSALITY NEEDS SPECIAL MATHEMATICS

---

Scientific Equations (e.g., Hooke's Law) are non-algebraic

e.g., Pricing Policy: "Double the competitor's price"

Correct notation:

(or)

$$Y \leftarrow 2X$$

$$X = 1$$

Process information

$$X = 1$$

$$Y = 2$$

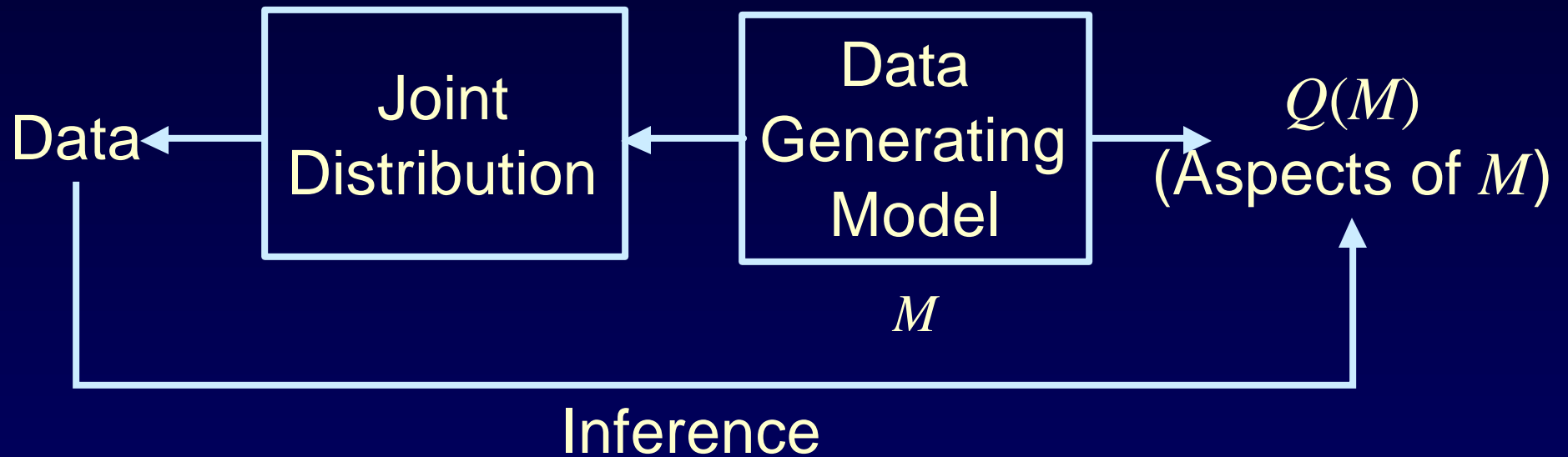
The solution

Had  $X$  been 3,  $Y$  would be 6.

If we raise  $X$  to 3,  $Y$  would be 6.

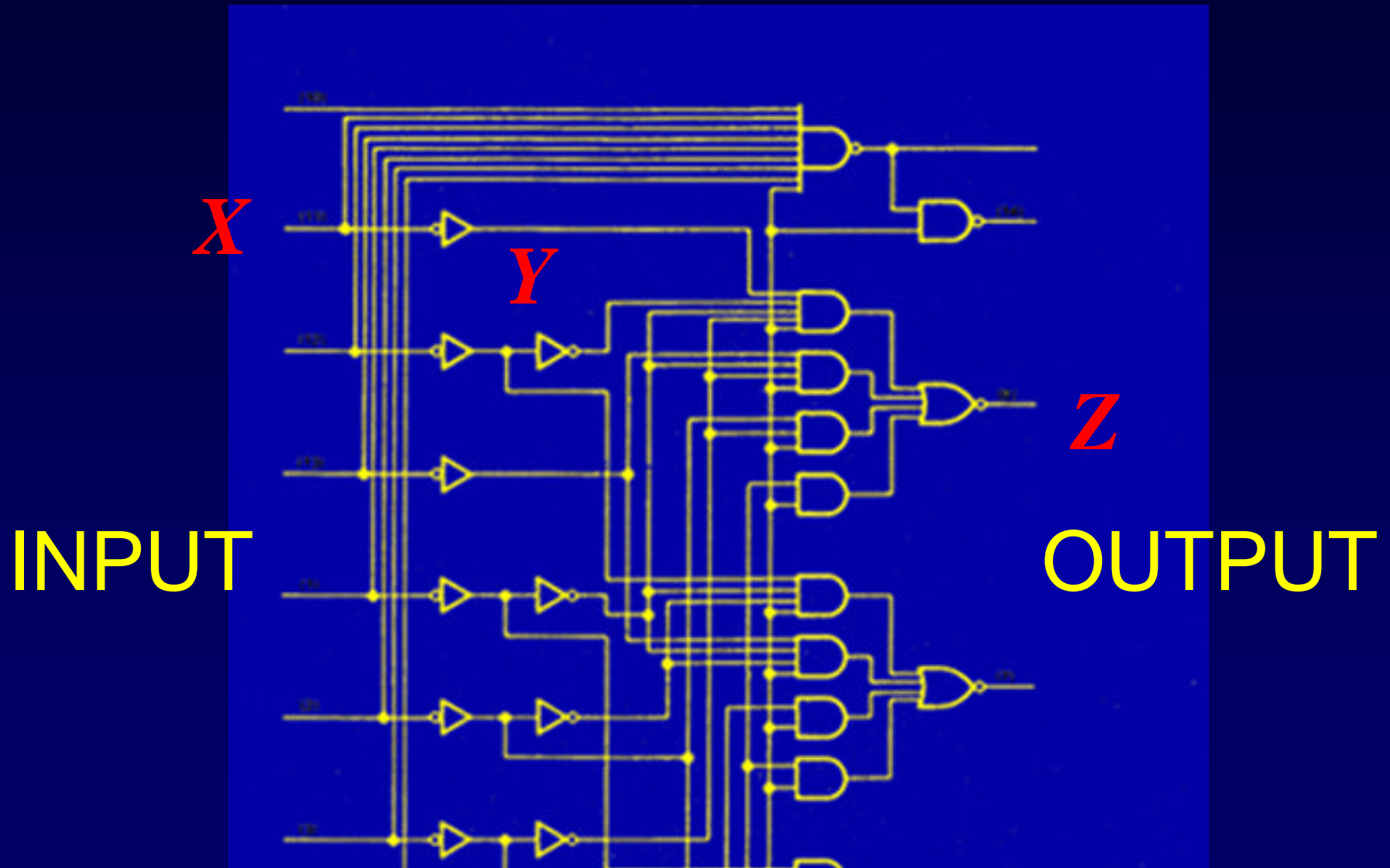
Must "wipe out"  $X = 1$ .

# THE STRUCTURAL MODEL PARADIGM



$M$  – Invariant strategy (mechanism, recipe, law, protocol) by which Nature assigns values to variables in the analysis.

# FAMILIAR CAUSAL MODEL ORACLE FOR MANIPULATION



# STRUCTURAL CAUSAL MODELS

---

Definition: A **structural causal model** is a 4-tuple  $\langle V, U, F, P(u) \rangle$ , where

- $V = \{V_1, \dots, V_n\}$  are observable variables
- $U = \{U_1, \dots, U_m\}$  are background variables
- $F = \{f_1, \dots, f_n\}$  are functions determining  $V$ ,  
 $v_i = f_i(v, u)$
- $P(u)$  is a distribution over  $U$

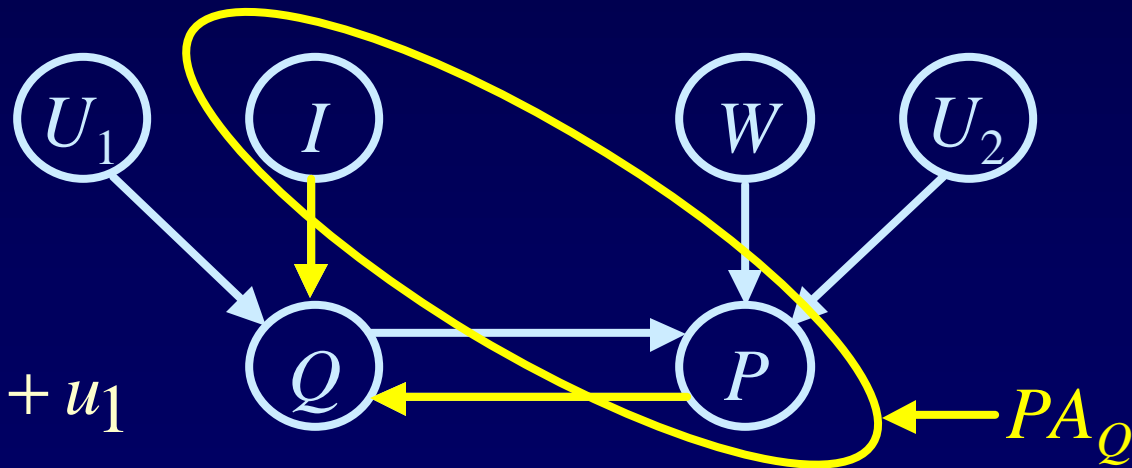
$P(u)$  and  $F$  induce a distribution  $P(v)$  over observable variables

# STRUCTURAL MODELS AND CAUSAL DIAGRAMS

The arguments of the functions  $v_i = f_i(v, u)$  define a graph

$$v_i = f_i(pa_i, u_i) \quad PA_i \subseteq V \setminus V_i \quad U_i \subseteq U$$

Example: Price – Quantity equations in economics



$$q = b_1 p + d_1 i + u_1$$

$$p = b_2 q + d_2 w + u_2$$

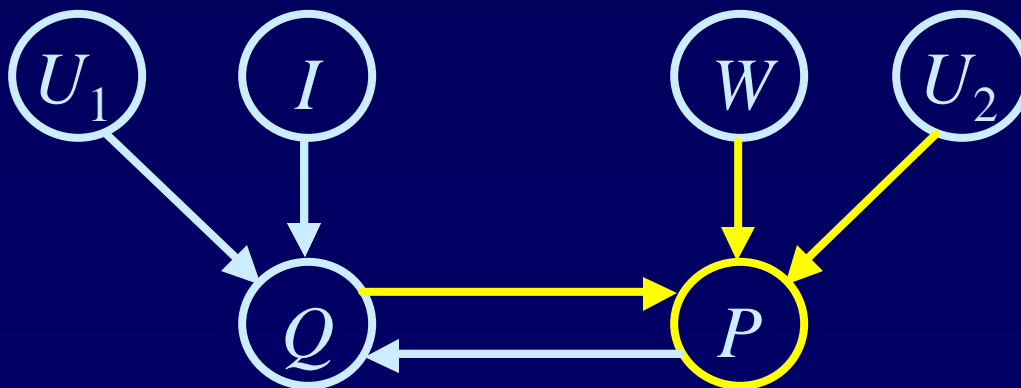
# STRUCTURAL MODELS AND INTERVENTION

Let  $X$  be a set of variables in  $V$ .

The action  $do(x)$  sets  $X$  to constants  $x$  regardless of the factors which previously determined  $X$ .

$do(x)$  replaces all functions  $f_i$  determining  $X$  with the constant functions  $X=x$ , to create a **mutilated model**  $M_x$

$$q = b_1 p + d_1 i + u_1$$
$$p = b_2 q + d_2 w + u_2$$

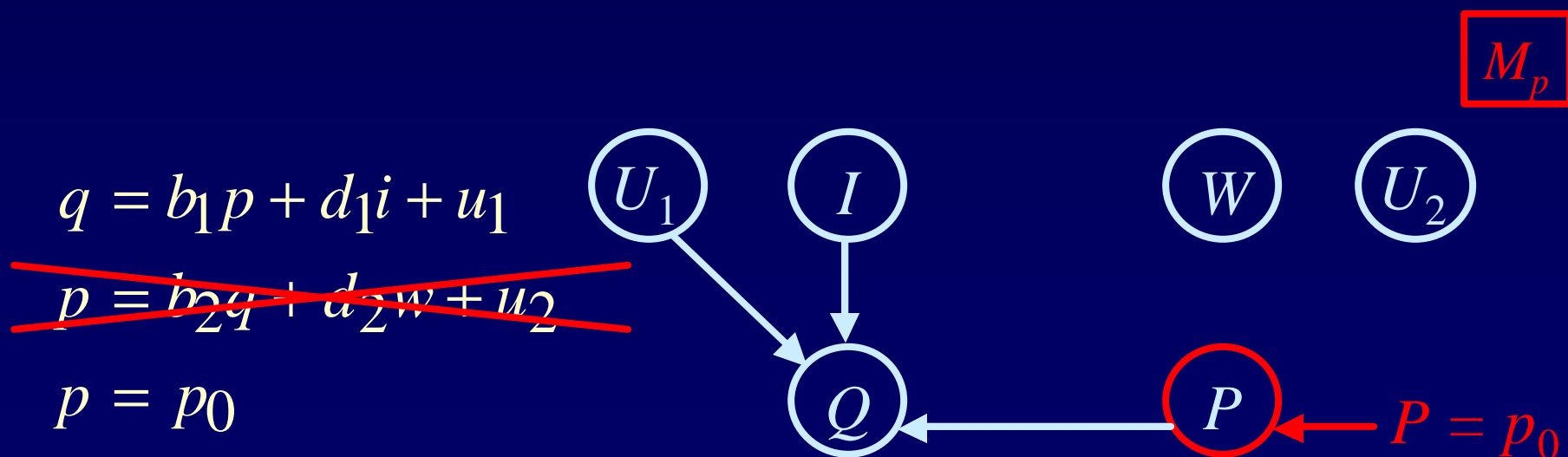


# STRUCTURAL MODELS AND INTERVENTION

Let  $X$  be a set of variables in  $V$ .

The action  $do(x)$  sets  $X$  to constants  $x$  regardless of the factors which previously determined  $X$ .

$do(x)$  replaces all functions  $f_i$  determining  $X$  with the constant functions  $X=x$ , to create a **mutilated model**  $M_x$



# CAUSAL MODELS AND COUNTERFACTUALS

## Definition:

The sentence: “ $Y$  would be  $y$  (in situation  $u$ ), had  $X$  been  $x$ ,” denoted  $Y_x(u) = y$ , means:

The solution for  $Y$  in a mutilated model  $M_x$ , (i.e., the equations for  $X$  replaced by  $X = x$ ) with input  $U = u$ , is equal to  $y$ .

The Fundamental Equation of Counterfactuals:

$$Y_x(u) = Y_{M_x}(u)$$

# CAUSAL MODELS AND COUNTERFACTUALS

## Definition:

The sentence: “ $Y$  would be  $y$  (in situation  $u$ ), had  $X$  been  $x$ ,” denoted  $Y_x(u) = y$ , means:

The solution for  $Y$  in a mutilated model  $M_x$ , (i.e., the equations for  $X$  replaced by  $X = x$ ) with input  $U = u$ , is equal to  $y$ .

- Joint probabilities of counterfactuals:

$$P(Y_x = y, Z_w = z) = \sum_{u: Y_x(u)=y, Z_w(u)=z} P(u)$$

The Fundamental Equation of Counterfactuals:

In particular:

$$P(y | do(x)) \triangleq \sum_{u: Y_x(u)=y} P(u)$$

$$PN(Y_{x'} = y' | x, y) = \sum_{u: Y_{x'}(u)=y'} P(u | x, y)$$

# AXIOMS OF CAUSAL COUNTERFACTUALS

---

$Y_x(u) = y$  :  $Y$  would be  $y$ , had  $X$  been  $x$  (in state  $U = u$ )

## 1. Definiteness

$$\exists x \in X \text{ s.t. } X_y(u) = x$$

## 2. Uniqueness

$$(X_y(u) = x) \ \& \ (X_y(u) = x') \Rightarrow x = x'$$

## 3. Effectiveness

$$X_{xw}(u) = x$$

## 4. Composition

$$W_x(u) = w \Rightarrow Y_{xw}(u) = Y_x(u)$$

## 5. Reversibility

$$(Y_{xw}(u) = y \ \& \ (W_{xy}(u) = w) \Rightarrow Y_x(u) = y$$

# INFERRING THE EFFECT OF INTERVENTIONS

---

## The problem:

To predict the impact of a proposed intervention using data obtained prior to the intervention.

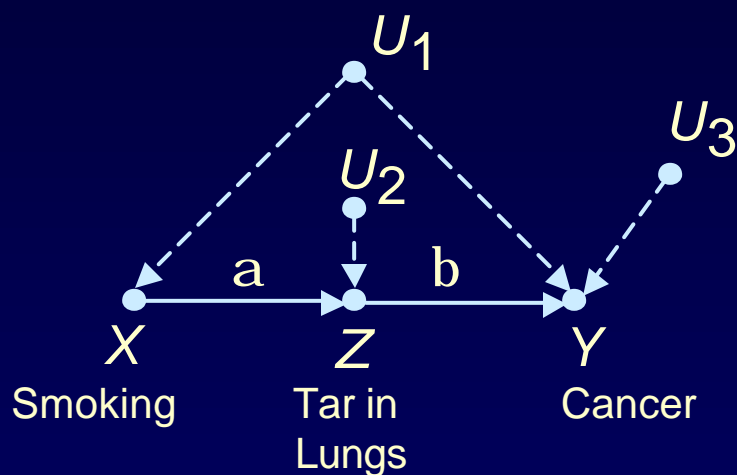
## The solution (conditional):

Causal Assumptions + Data  $\rightarrow$  Policy Claims

1. Mathematical tools for communicating causal assumptions formally and transparently.
2. Deciding (mathematically) whether the assumptions communicated are sufficient for obtaining consistent estimates of the prediction required.
3. Designing (or affirming) a set of formal expressions for the prediction that, if performed, would render a consistent estimate feasible.

# NON-PARAMETRIC STRUCTURAL MODELS

Given  $P(x,y,z)$ , should we ban smoking?



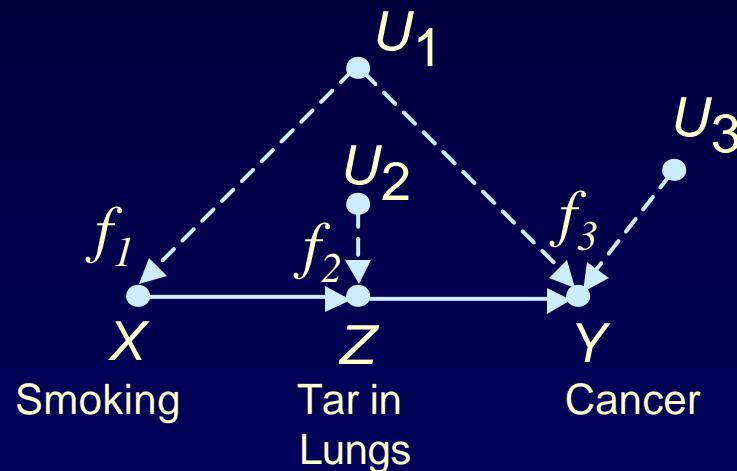
## Linear Analysis

$$X = u_1,$$

$$Z = \alpha X + u_2,$$

$$y = \beta Z + \gamma u_1 + u_3.$$

Find:  $\alpha \cdot \beta$



## Nonparametric Analysis

$$x = f_1(u_1),$$

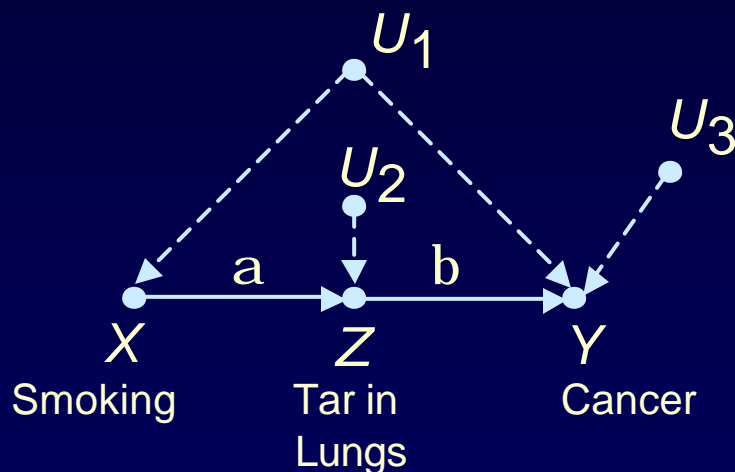
$$z = f_2(x, u_2),$$

$$y = f_3(z, u_1, u_3).$$

Find:  $P(y|do(x))$

# EFFECT OF INTERVENTION AN EXAMPLE

Given  $P(x,y,z)$ , should we ban smoking?



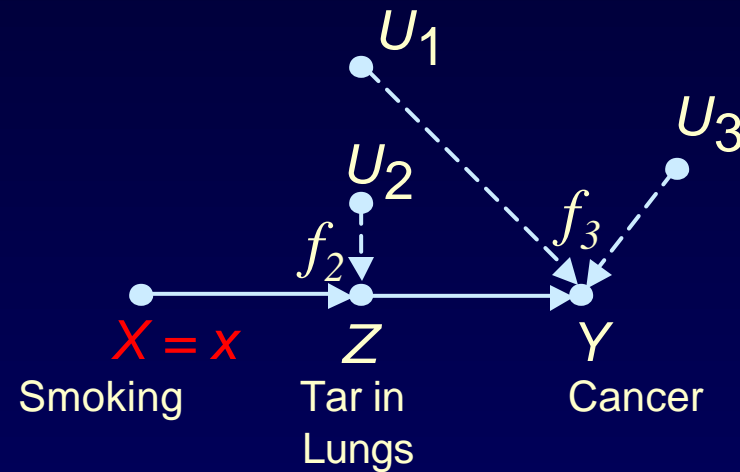
Linear Analysis

$$X = u_1,$$

$$Z = \alpha X + u_2,$$

$$y = \beta Z + \gamma u_1 + u_3.$$

Find:  $\alpha \cdot \beta$



Nonparametric Analysis

$$x = \text{const.}$$

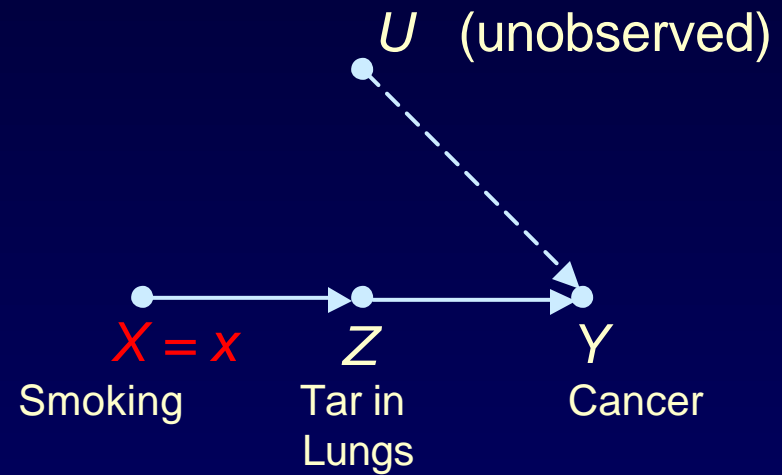
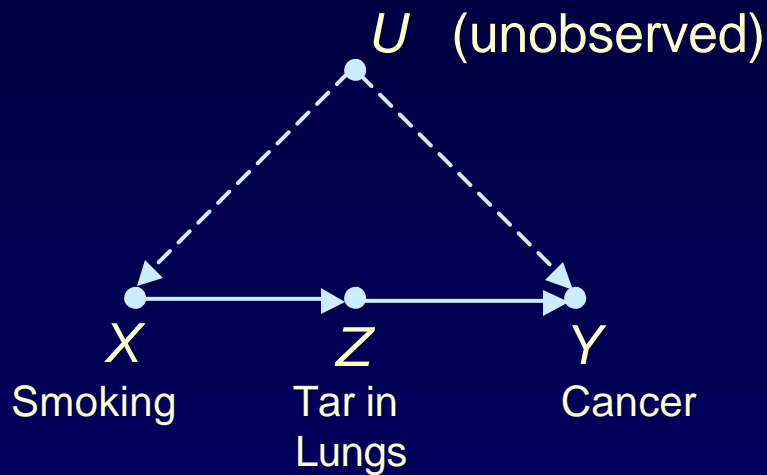
$$Z = f_2(x, u_2),$$

$$y = f_3(Z, u_1, u_3).$$

Find:  $P(y|do(x)) \stackrel{\Delta}{=} P(Y=y)$  in new model

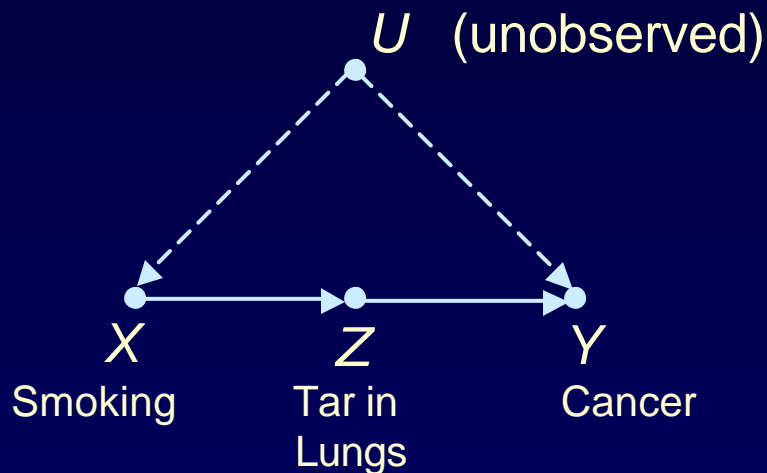
# EFFECT OF INTERVENTION AN EXAMPLE (cont)

Given  $P(x,y,z)$ , should we ban smoking?



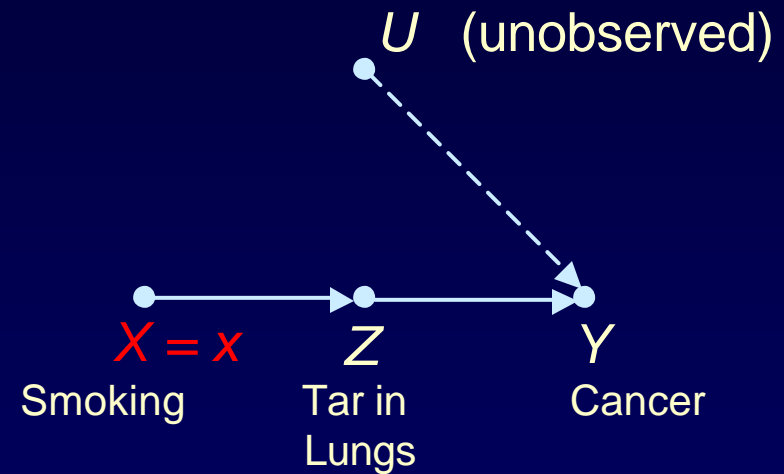
# EFFECT OF INTERVENTION AN EXAMPLE (cont)

Given  $P(x,y,z)$ , should we ban smoking?



**Pre-intervention**

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



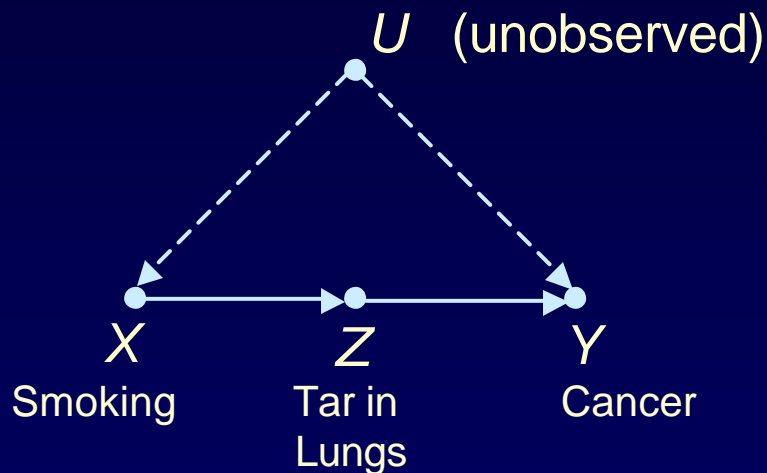
**Post-intervention**

$$P(y, z | do(x)) = \sum_u P(u)P(z | x)P(y | z, u)$$

•

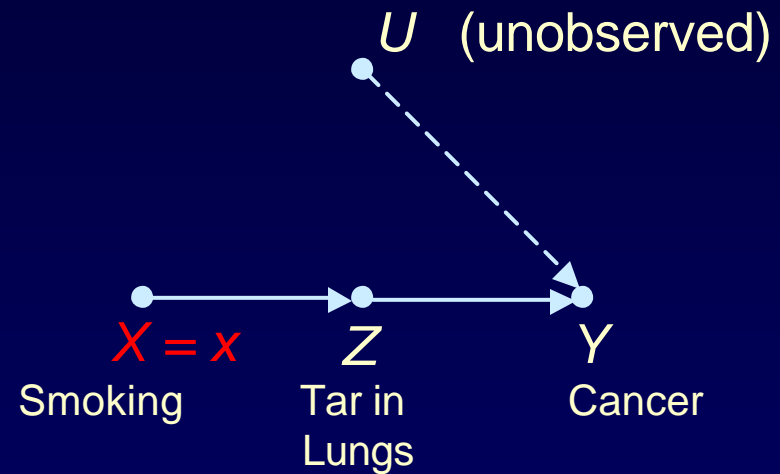
# EFFECT OF INTERVENTION AN EXAMPLE (cont)

Given  $P(x,y,z)$ , should we ban smoking?



**Pre-intervention**

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



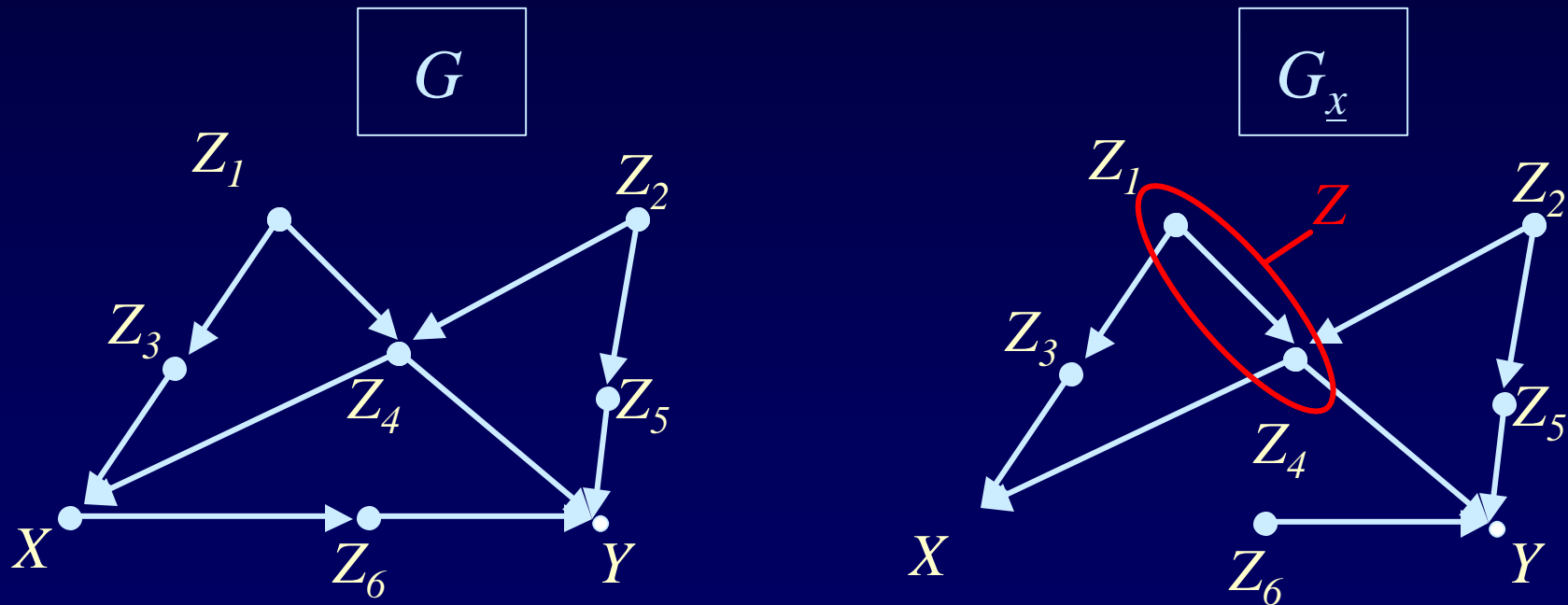
**Post-intervention**

$$P(y, z | do(x)) = \sum_u P(u)P(z | x)P(y | z, u)$$

To compute  $P(y, z | do(x))$ , we must eliminate  $u$ . (graphical problem).

# ELIMINATING CONFOUNDING BIAS A GRAPHICAL CRITERION

$P(y | do(x))$  is estimable if there is a set  $Z$  of variables such that  $Z$  *d-separates*  $X$  from  $Y$  in  $G_{\underline{x}}$ .



Moreover,  $P(y | do(x)) = \sum_z P(y | x, z) P(z)$   
("adjusting" for  $Z$ )

# ● RULES OF CAUSAL CALCULUS

---

**Rule 1:** Ignoring observations

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

if  $(Y \perp\!\!\!\perp Z \mid X, W)_{G_{\overline{X}}}$

**Rule 2:** Action/observation exchange

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

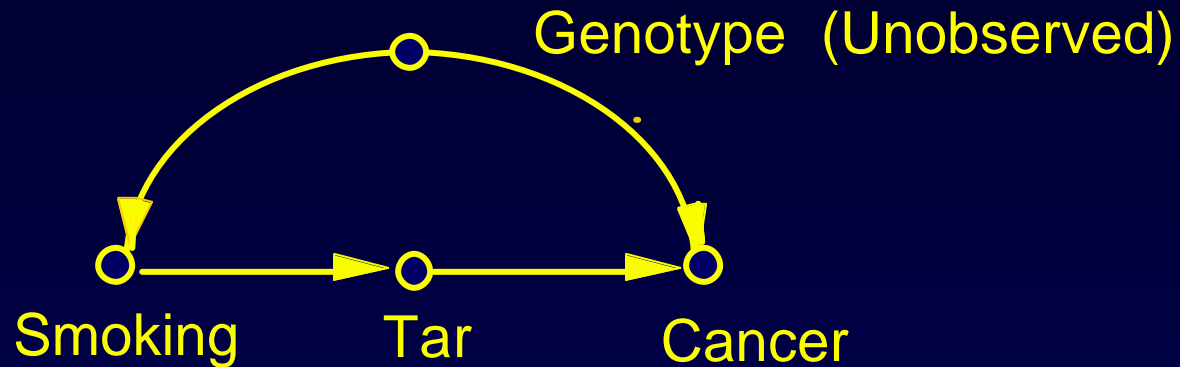
if  $(Y \perp\!\!\!\perp Z \mid X, W)_{G_{\overline{X}, \underline{Z}}}$

**Rule 3:** Ignoring actions

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

if  $(Y \perp\!\!\!\perp Z \mid X, W)_{G_{\overline{X}, \overline{Z(W)}}$

# DERIVATION IN CAUSAL CALCULUS

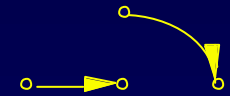


$$P(c | do\{s\}) = \sum_t P(c | do\{s\}, t) P(t | do\{s\})$$

Probability Axioms

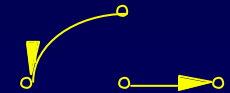
$$= S_t P(c | do\{s\}, do\{t\}) P(t | do\{s\})$$

Rule 2



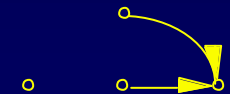
$$= S_t P(c | do\{s\}, do\{t\}) P(t | s)$$

Rule 2



$$= S_t P(c | do\{t\}) P(t | s)$$

Rule 3



$$= S_{s^c} S_t P(c | do\{t\}, s^c) P(s^c | do\{t\}) P(t | s)$$

Probability Axioms

$$= S_{s^c} S_t P(c | t, s^c) P(s^c | do\{t\}) P(t | s)$$

Rule 2



$$= S_{s^c} S_t P(c | t, s^c) P(s^c) P(t | s)$$

Rule 3



# INFERENCE ACROSS DESIGNS

---

## Problem:

Predict  $P(y \mid do(x))$  from a study in which only  $Z$  can be controlled.

## Solution:

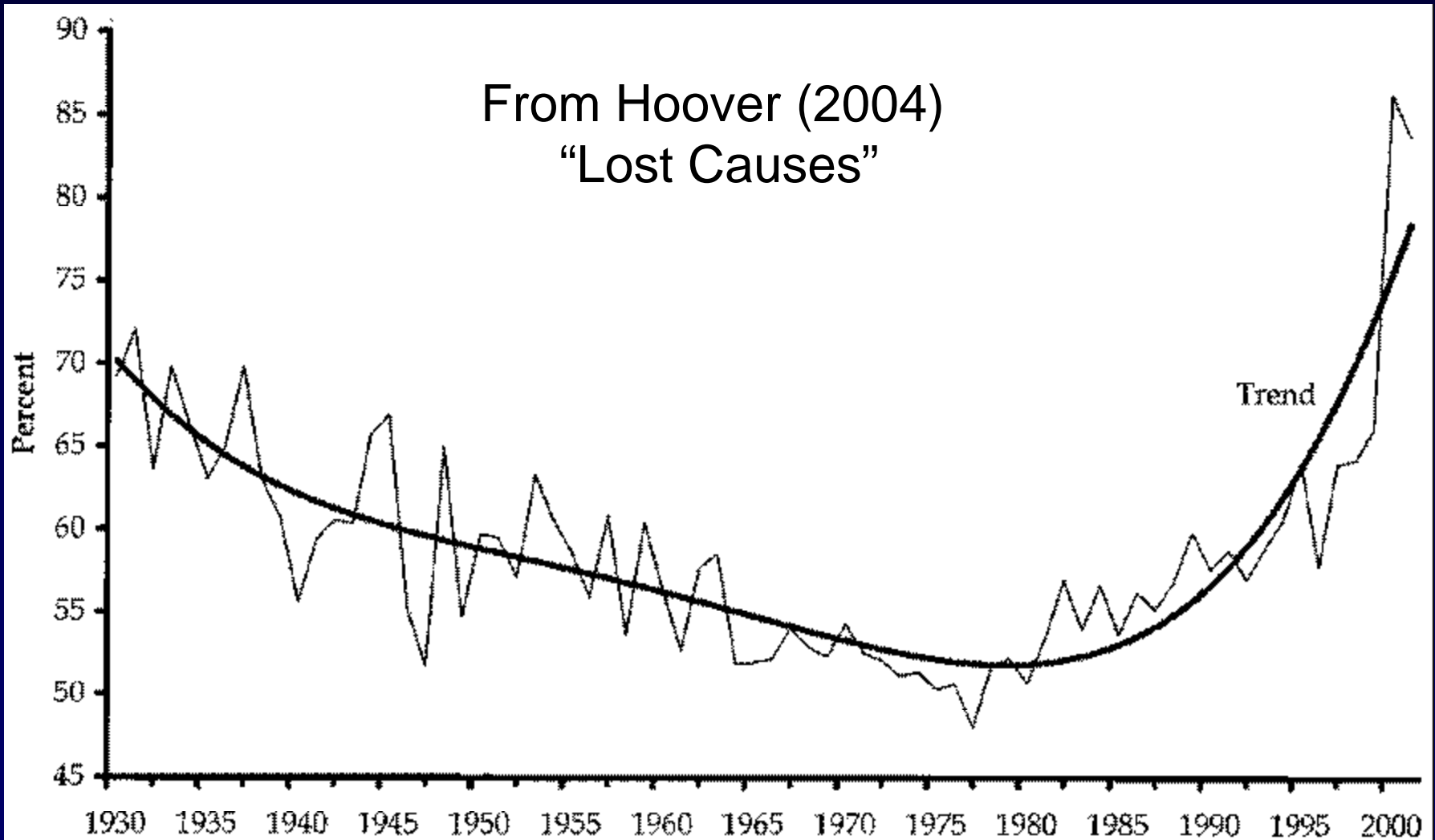
Determine if  $P(y \mid do(x))$  can be reduced to a mathematical expression involving only  $do(z)$ .

# COMPLETENESS RESULTS ON IDENTIFICATION

---

- *do*-calculus is complete
- Complete graphical criterion for identifying **causal effects** (Shpitser and Pearl, 2006).
- **Complete graphical** criterion for empirical testability of counterfactuals (Shpitser and Pearl, 2007).

# THE CAUSAL RENAISSANCE: VOCABULARY IN ECONOMICS



# THE CAUSAL RENAISSANCE: USEFUL RESULTS

---

1. Complete formal semantics of counterfactuals
2. Transparent language for expressing assumptions
3. Complete solution to causal-effect identification
4. Legal responsibility (bounds)
5. Imperfect experiments (universal bounds for IV)
6. Integration of data from diverse sources
7. Direct and Indirect effects,
8. Complete criterion for counterfactual testability

# EFFECT DECOMPOSITION (direct vs. indirect effects)

---

1. Why decompose effects?
2. What is the semantics of direct and indirect effects?
3. What are the policy implications of direct and indirect effects?
4. When can direct and indirect effect be estimated consistently from experimental and nonexperimental data?

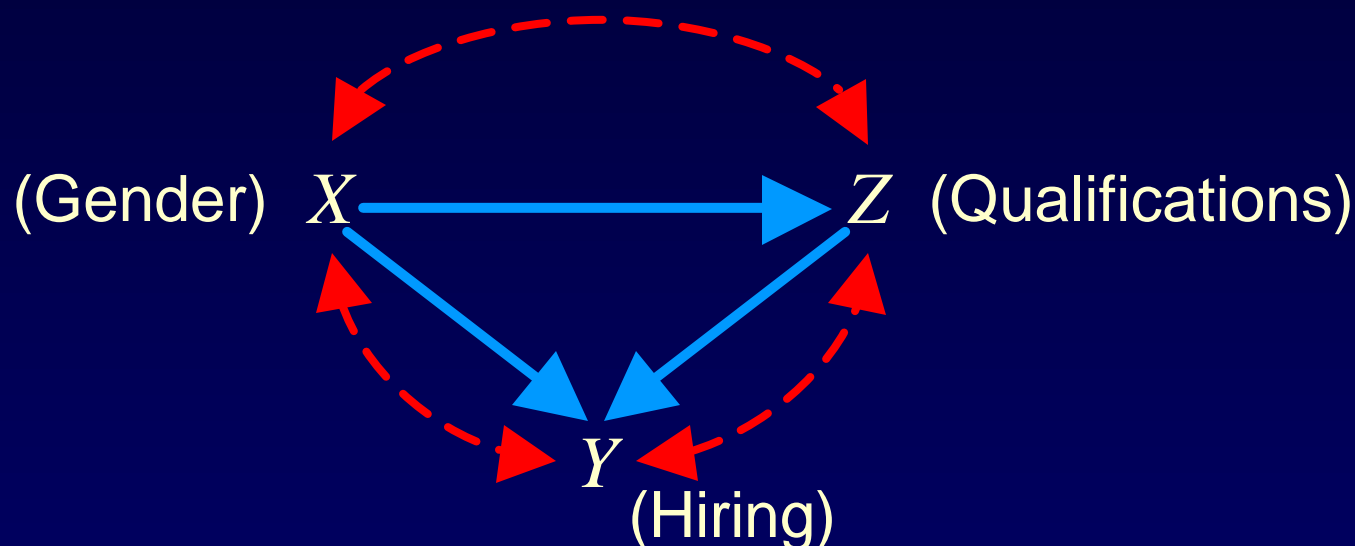
# WHY DECOMPOSE EFFECTS?

---

1. To understand how Nature works
2. To comply with legal requirements
3. To predict the effects of new type of interventions:  
**Signal routing**, rather than **variable fixing**

# LEGAL IMPLICATIONS OF DIRECT EFFECT

Can data prove an employer guilty of hiring discrimination?



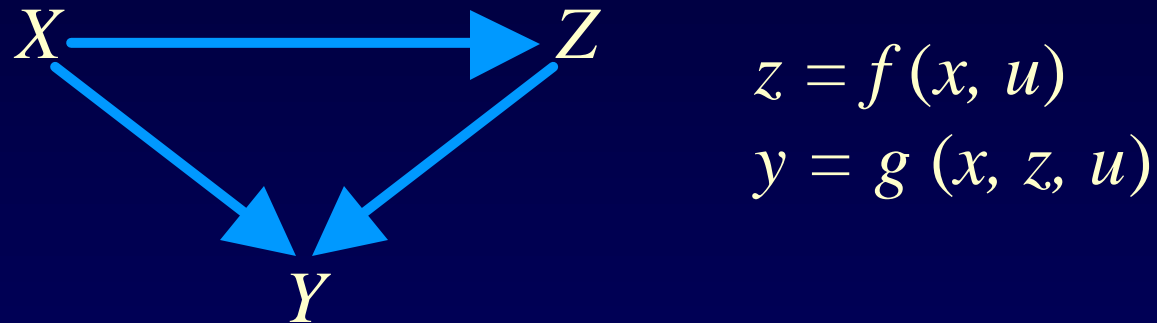
What is the direct effect of  $X$  on  $Y$ ?

$$E(Y \mid do(x_1), do(z)) - E(Y \mid do(x_0), do(z))$$

(averaged over  $z$ ) **Adjust for Z? No! No!**

# NATURAL SEMANTICS OF AVERAGE DIRECT EFFECTS

Robins and Greenland (1992) – “Pure”



**Average Direct Effect of  $X$  on  $Y$ :  $DE(x_0, x_1; Y)$**

The expected change in  $Y$ , when we change  $X$  from  $x_0$  to  $x_1$  and, for each  $u$ , we keep  $Z$  constant at whatever value it attained before the change.

$$E[Y_{x_1 Z_{x_0}} - Y_{x_0}]$$

In linear models,  $DE =$  Controlled Direct Effect

# SEMANTICS AND IDENTIFICATION OF NESTED COUNTERFACTUALS

Consider the quantity  $Q \triangleq E_u[Y_{xZ_{x^*}(u)}(u)]$

Given  $\langle M, P(u) \rangle$ ,  $Q$  is well defined

Given  $u$ ,  $Z_{x^*}(u)$  is the solution for  $Z$  in  $M_{x^*}$ , call it  $z$

$Y_{xZ_{x^*}(u)}(u)$  is the solution for  $Y$  in  $M_{xz}$

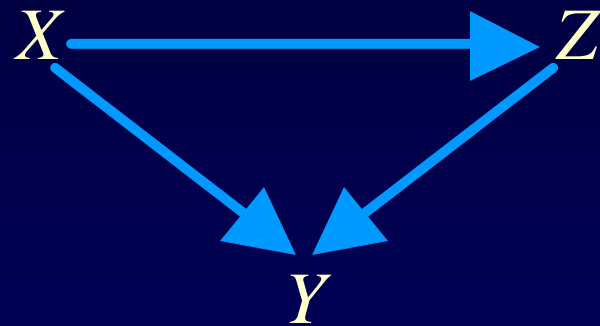
Can  $Q$  be estimated from  $\left\{ \begin{array}{l} \text{experimental} \\ \text{nonexperimental} \end{array} \right\}$  data?

**Experimental:** nest-free expression

**Nonexperimental:** subscript-free expression

# NATURAL SEMANTICS OF INDIRECT EFFECTS

---



$$z = f(x, u)$$

$$y = g(x, z, u)$$

**Indirect Effect of  $X$  on  $Y$ :  $IE(x_0, x_1; Y)$**

The expected change in  $Y$  when we keep  $X$  constant, say at  $x_0$ , and let  $Z$  change to whatever value it would have attained had  $X$  changed to  $x_1$ .

$$E[Y_{x_0 Z_{x_1}} - Y_{x_0}]$$

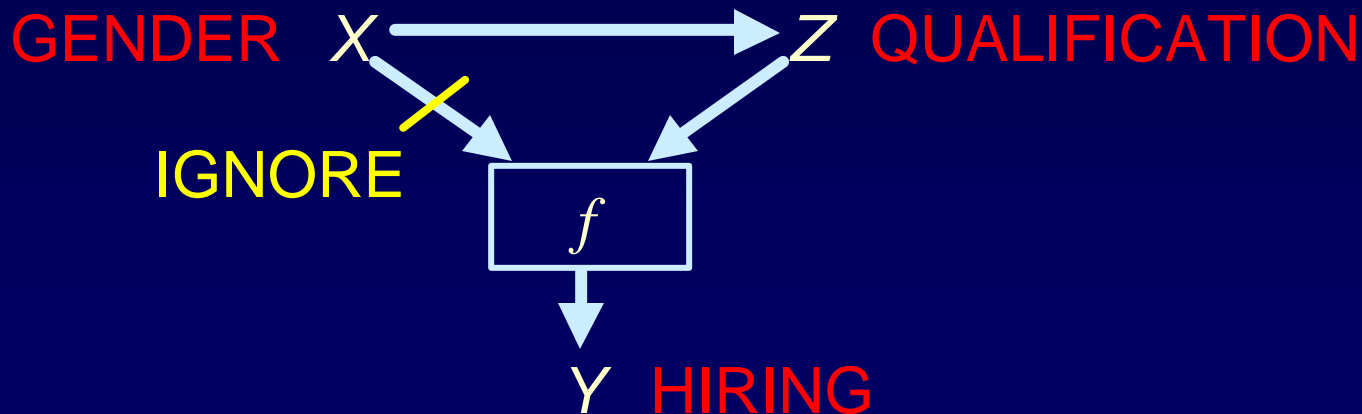
In linear models,  $IE = TE - DE$

# POLICY IMPLICATIONS OF INDIRECT EFFECTS

---

What is the **indirect** effect of  $X$  on  $Y$ ?

The effect of Gender on Hiring if sex discrimination is eliminated.



Blocking a link – a new type of intervention

# RELATIONS BETWEEN TOTAL, DIRECT, AND INDIRECT EFFECTS

---

**Theorem 5:** The total, direct and indirect effects obey  
The following equality

$$TE(x, x^*; Y) = DE(x, x^*; Y) - IE(x^*, x; Y)$$

In words, the total effect (on  $Y$ ) associated with the transition from  $x^*$  to  $x$  is equal to the **difference** between the direct effect associated with this transition and the indirect effect associated with the **reverse transition**, from  $x$  to  $x^*$ .

# EXPERIMENTAL IDENTIFICATION OF AVERAGE DIRECT EFFECTS

---

**Theorem:** If there exists a set  $W$  such that

$$Y_{xz} \perp\!\!\!\perp Z_{x^*} \mid W \text{ for all } z \text{ and } x$$

Then the average direct effect

$$DE(x, x^*; Y) = E(Y_{x, Z_{x^*}}) - E(Y_{x^*})$$

Is identifiable from experimental data and is given by

$$DE(x, x^*; Y) = \sum_{w, z} [E(Y_{xz} \mid w) - E(Y_{x^*z} \mid w)] P(Z_{x^*} = z \mid w) P(w)$$

# GRAPHICAL CONDITION FOR EXPERIMENTAL IDENTIFICATION OF DIRECT EFFECTS

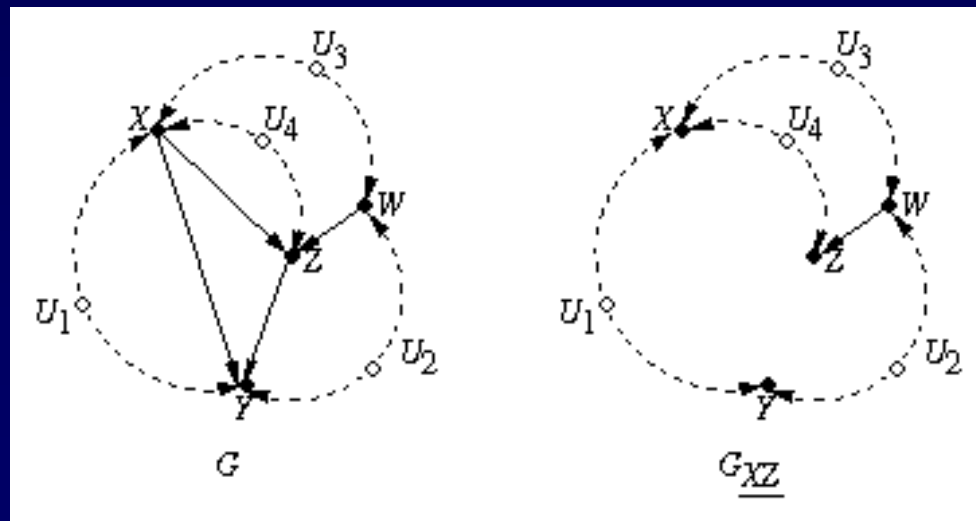
**Theorem:** If there exists a set  $W$  such that

$$(Y \perp\!\!\!\perp Z | W)_{G_{\underline{XZ}}} \text{ and } W \subseteq ND(X \cup Z)$$

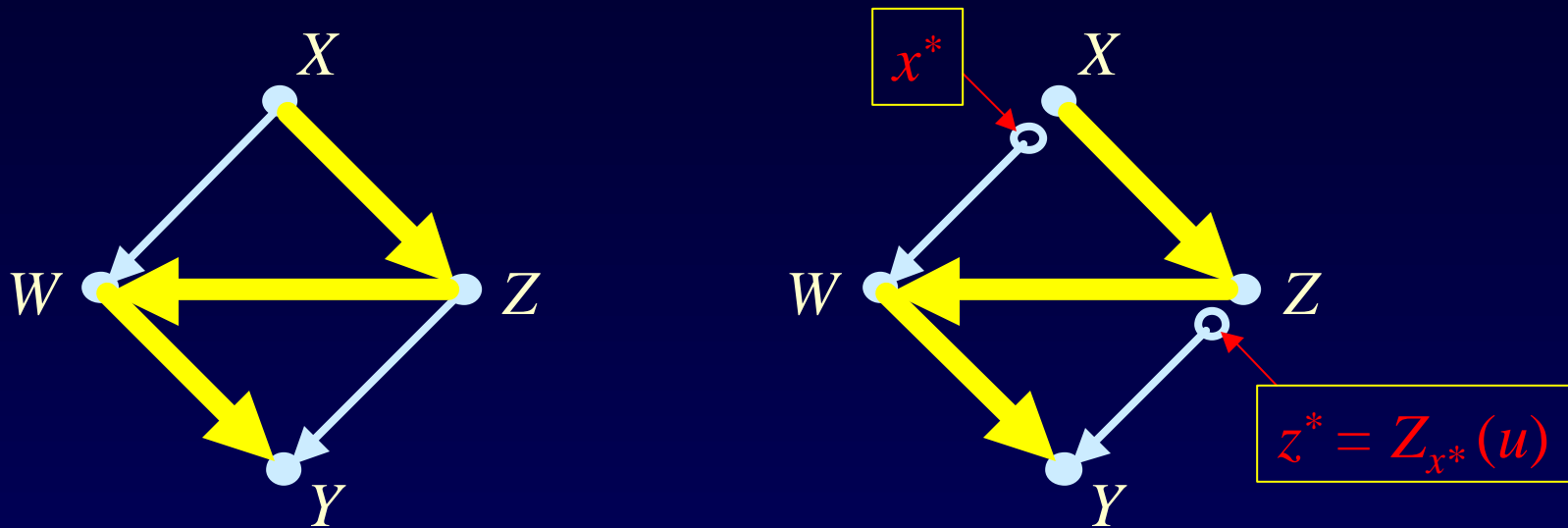
then,

$$DE(x, x^*; Y) = \sum_{w, z} [E(Y_{xz} | w) - E(Y_{x^*z} | w)] P(Z_{x^*} = z | w) P(w)$$

Example:



# GENERAL PATH-SPECIFIC EFFECTS (Def.)



Form a new model,  $M_g^*$ , specific to active subgraph  $g$

$$f_i^*(pa_i, u; g) = f_i(pa_i(g), pa_i^*(\bar{g}), u)$$

Definition:  $g$ -specific effect

$$E_g(x, x^*; Y)_M = TE(x, x^*; Y)_{M_g^*}$$

Nonidentifiable even in Markovian models

# SUMMARY OF RESULTS

---

1. Formal semantics of path-specific effects, based on **signal blocking**, instead of value fixing.
2. Path-analytic techniques extended to nonlinear and nonparametric models.
3. **Meaningful (graphical) conditions for estimating** direct and indirect effects from experimental and nonexperimental data.

# CONCLUSIONS

---

Structural-model semantics, enriched with logic and graphs, provides:

- Complete formal basis for causal and counterfactual reasoning
- Unifies the graphical, potential-outcome and structural equation approaches
- Provides friendly and formal solutions to century-old problems and confusions.



# WHY CAUSALITY NEEDS SPECIAL MATHEMATICS

---

SEM Equations are Non-algebraic

e.g., Pricing Policy: “Double the competitor’s price”

Correct notation:

$$Y := 2X$$

$$X = 1$$

Process information

$$X = 1$$

$$Y = 2$$

Static information

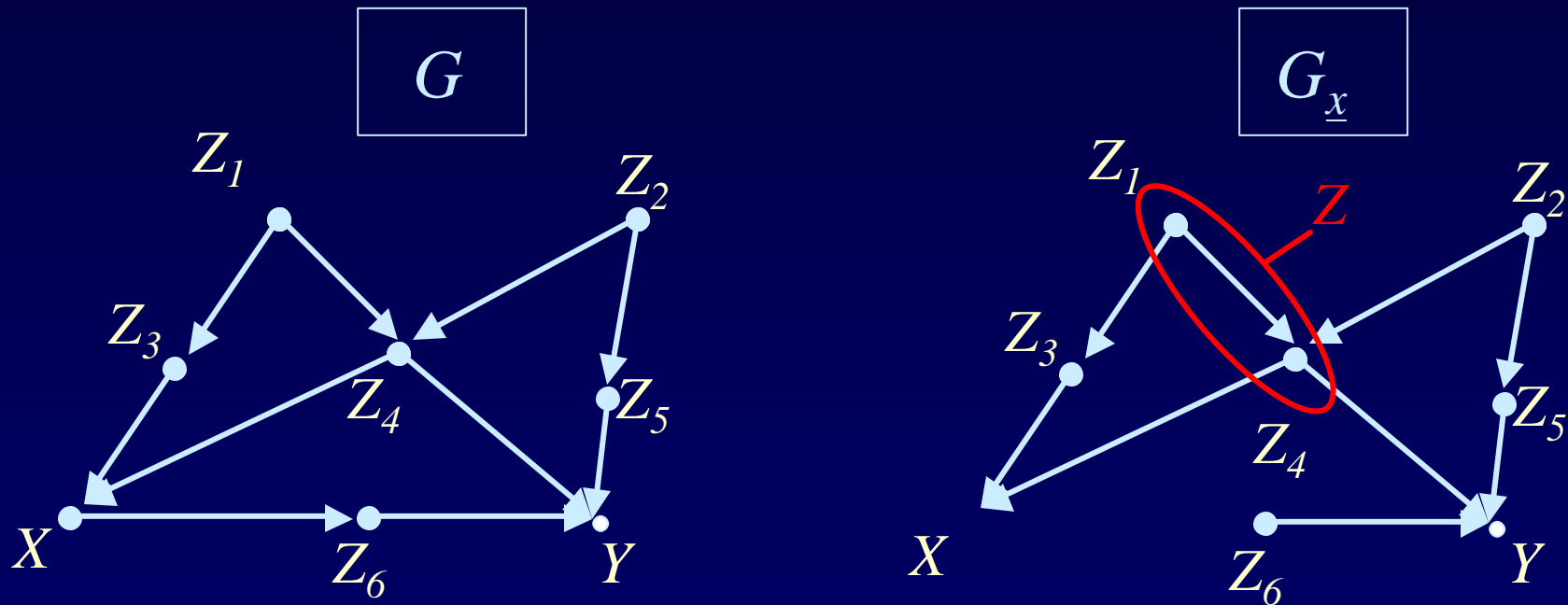
Had  $X$  been 3,  $Y$  would be 6.

If we raise  $X$  to 3,  $Y$  would be 6.

Must “wipe out”  $X = 1$ .

# ELIMINATING CONFOUNDING BIAS A GRAPHICAL CRITERION

$P(y | do(x))$  is estimable if there is a set  $Z$  of variables such that  $Z$  *d-separates*  $X$  from  $Y$  in  $G_{\underline{x}}$ .



Moreover,  $P(y | do(x)) = \sum_z P(y | x, z) P(z)$   
("adjusting" for  $Z$ )

# ● RULES OF CAUSAL CALCULUS

---

**Rule 1:** Ignoring observations

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

if  $(Y \perp\!\!\!\perp Z \mid X, W)_{G_{\overline{X}}}$

**Rule 2:** Action/observation exchange

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

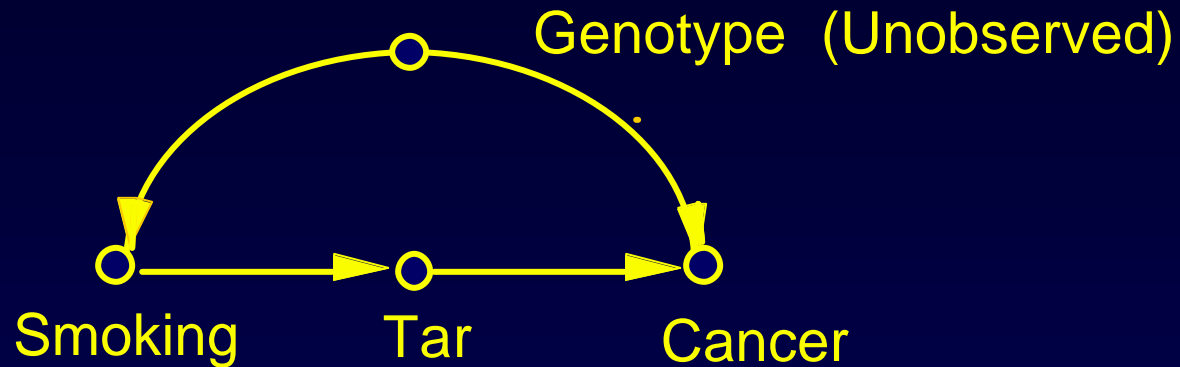
if  $(Y \perp\!\!\!\perp Z \mid X, W)_{G_{\overline{X}, \underline{Z}}}$

**Rule 3:** Ignoring actions

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

if  $(Y \perp\!\!\!\perp Z \mid X, W)_{G_{\overline{X}, \overline{Z(W)}}$

# DERIVATION IN CAUSAL CALCULUS

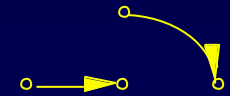


$$P(c | do\{s\}) = \sum_t P(c | do\{s\}, t) P(t | do\{s\})$$

Probability Axioms

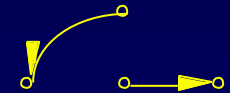
$$= S_t P(c | do\{s\}, do\{t\}) P(t | do\{s\})$$

Rule 2



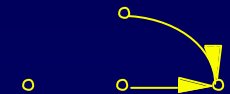
$$= S_t P(c | do\{s\}, do\{t\}) P(t | s)$$

Rule 2



$$= S_t P(c | do\{t\}) P(t | s)$$

Rule 3



$$= S_{s^c} S_t P(c | do\{t\}, s^c) P(s^c | do\{t\}) P(t | s)$$

Probability Axioms

$$= S_{s^c} S_t P(c | t, s^c) P(s^c | do\{t\}) P(t | s)$$

Rule 2



$$= S_{s^c} S_t P(c | t, s^c) P(s^c) P(t | s)$$

Rule 3



# INFERENCE ACROSS DESIGNS

---

## Problem:

Predict  $P(y \mid do(x))$  from a study in which only  $Z$  can be controlled.

## Solution:

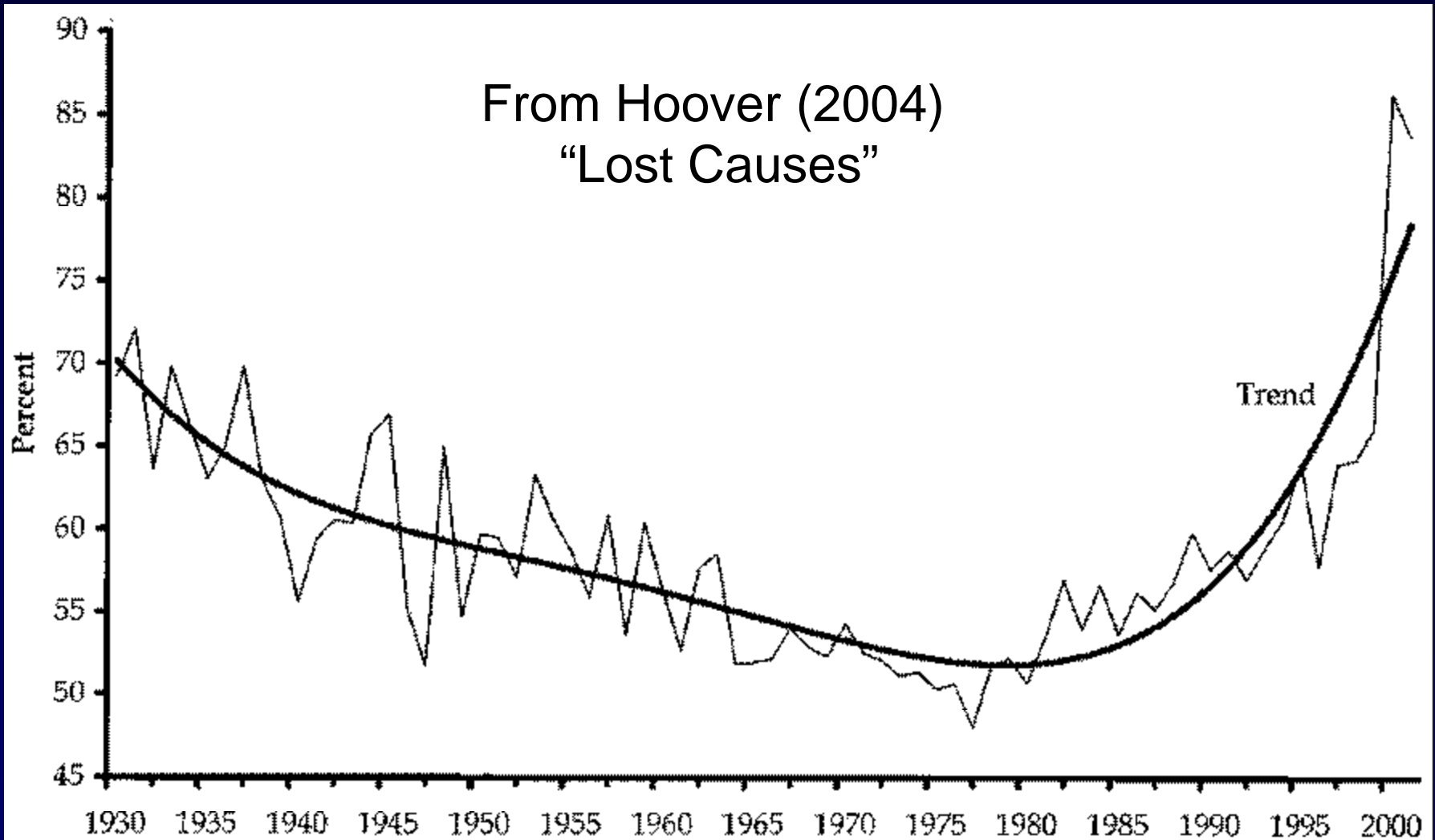
Determine if  $P(y \mid do(x))$  can be reduced to a mathematical expression involving only  $do(z)$ .

# COMPLETENESS RESULTS ON IDENTIFICATION

---

- *do*-calculus is complete
- Complete graphical criterion for identifying **causal effects** (Shpitser and Pearl, 2006).
- **Complete graphical** criterion for empirical testability of counterfactuals (Shpitser and Pearl, 2007).

# THE CAUSAL RENAISSANCE: VOCABULARY IN ECONOMICS



# THE CAUSAL RENAISSANCE: USEFUL RESULTS

---

1. Complete formal semantics of counterfactuals
2. Transparent language for expressing assumptions
3. Complete solution to causal-effect identification
4. Legal responsibility (bounds)
5. Imperfect experiments (universal bounds for IV)
6. Integration of data from diverse sources
7. Direct and Indirect effects,
8. Complete criterion for counterfactual testability

# DETERMINING THE CAUSES OF EFFECTS (The Attribution Problem)

---

- Your Honor! My client (Mr. A) died **BECAUSE** he used that drug.



# DETERMINING THE CAUSES OF EFFECTS (The Attribution Problem)

---

- Your Honor! My client (Mr. A) died **BECAUSE** he used that drug.



- Court to decide if it is **MORE PROBABLE THAN NOT** that A would be alive **BUT FOR** the drug!  
 $PN = P(? | A \text{ is dead, took the drug}) \geq 0.50$

# THE PROBLEM

---

## Semantical Problem:

1. What is the meaning of  $PN(x,y)$ :  
“Probability that event  $y$  would not have occurred if it were not for event  $x$ , given that  $x$  and  $y$  did in fact occur.”
  -

# THE PROBLEM

---

## Semantical Problem:

1. What is the meaning of  $PN(x,y)$ :  
“Probability that event  $y$  would not have occurred if it were not for event  $x$ , given that  $x$  and  $y$  did in fact occur.”

## Answer:

$$PN(x, y) = P(Y_{x'} = y' \mid x, y)$$

Computable from  $M$

# THE PROBLEM

---

## Semantical Problem:

1. What is the meaning of  $PN(x,y)$ :  
“Probability that event  $y$  would not have occurred if it were not for event  $x$ , given that  $x$  and  $y$  did in fact occur.”

## Analytical Problem:

2. Under what condition can  $PN(x,y)$  be learned from statistical data, i.e., observational, experimental and combined.

# TYPICAL THEOREMS

(Tian and Pearl, 2000)

- Bounds given combined nonexperimental and experimental data

$$\max \left\{ \frac{0}{P(x,y)} \right\} \leq PN \leq \min \left\{ \frac{1}{P(x,y)} \right\}$$

- Identifiability under monotonicity (Combined data)

$$PN = \frac{P(y | x) - P(y | x')}{P(y | x)} + \frac{P(y | x') - P(y_{x'})}{P(x,y)}$$

corrected Excess-Risk-Ratio

# CAN FREQUENCY DATA DECIDE LEGAL RESPONSIBILITY?

	<u>Experimental</u>		<u>Nonexperimental</u>	
	$do(x)$	$do(x')$	$x$	$x'$
Deaths ( $y$ )	16	14	2	28
Survivals ( $y'$ )	984	986	998	972
	1,000	1,000	1,000	1,000

- **Nonexperimental data:** drug usage predicts longer life
- **Experimental data:** drug has negligible effect on survival
- **Plaintiff:** Mr. A is special.
  1. He actually **died**
  2. He used the drug by **choice**
- Court to decide (given both data):  
Is it **more probable than not** that A would be alive **but for** the drug?

$$PN \triangleq P(Y_{x'} = y' | x, y) > 0.50$$

# TYPICAL THEOREMS

(Tian and Pearl, 2000)

- Bounds given combined nonexperimental and experimental data

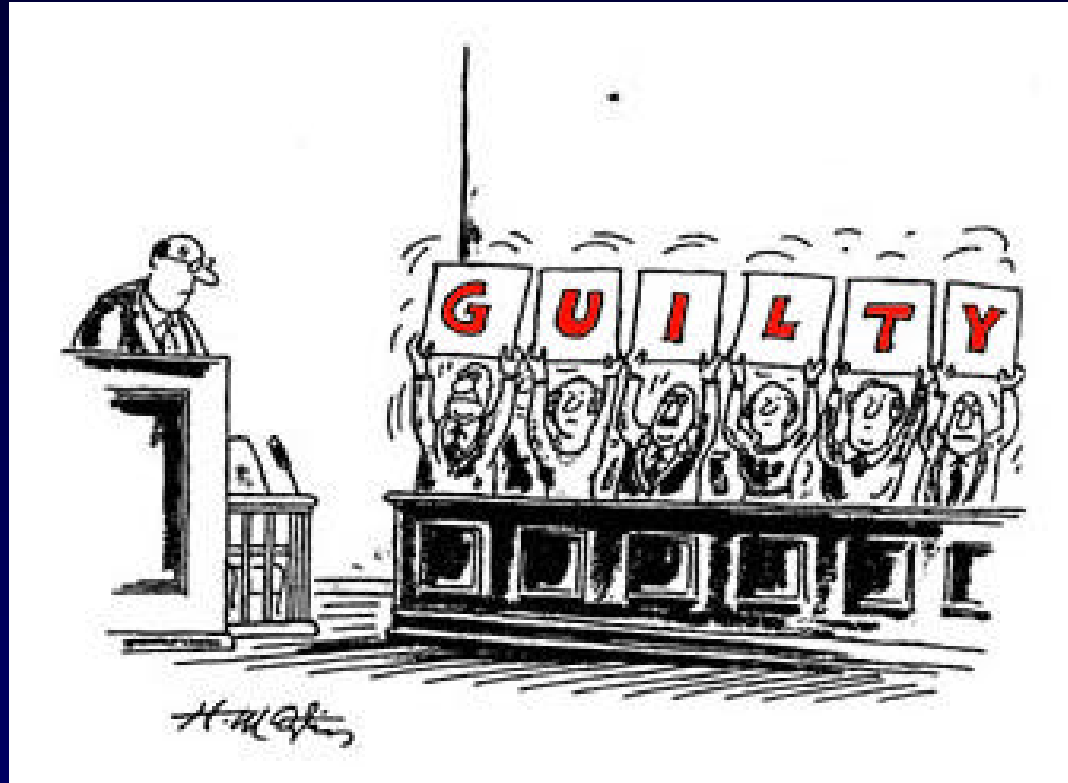
$$\max \left\{ \frac{0}{P(x,y)} \right\} \leq PN \leq \min \left\{ \frac{1}{P(x,y)} \right\}$$

- Identifiability under monotonicity (Combined data)

$$PN = \frac{P(y | x) - P(y | x')}{P(y | x)} + \frac{P(y | x') - P(y_{x'})}{P(x,y)}$$

corrected Excess-Risk-Ratio

# SOLUTION TO THE ATTRIBUTION PROBLEM



- WITH PROBABILITY ONE  $1 \leq P(y \in C_c | x, y) \leq 1$
- Combined data tell more than each study alone

# LEGAL DEFINITION OF DIRECT EFFECT (FORMALIZING DISCRIMINATION)

---

“The central question in any employment-discrimination case is whether the employer would have taken the **same action** had the employee been of **different race** (age, sex, religion, national origin etc.) and **everything else had been the same**”

[Carson versus Bethlehem Steel Corp. (70 FEP Cases 921, 7<sup>th</sup> Cir. (1996))]

$x$  = male,  $x^c$  = female  
 $y$  = hire,  $y^c$  = not hire  
 $z$  = applicant's qualifications

**NO DIRECT EFFECT**

$$Y_{x^c|Z_x} = Y_x$$