

# Explanation as an Aim of Science

# SCIENTIFIC KNOWLEDGE

## Explanation

*Understanding*

reasons for why the phenomenon to be explained is expected on a lawful basis (Hempel 1965)

"

## Prediction

*Forecasts*

reasons for expecting a phenomenon to occur in a particular way.

||

## Design

*Functioning artifacts*

reasons for expecting that a manipulation satisfies certain functions

Is explanation the same as prediction, just applied to phenomena already observed?

# The Deductive-Nomological (DN) Account

... a *DN* explanation answers the question “*Why* did the explanandum-phenomenon occur?” by showing that the phenomenon resulted from certain particular circumstances, specified in  $C_1, C_2, \dots, C_k$ , in accordance with the laws  $L_1, L_2, \dots, L_r$ . By pointing this out, the argument shows that, given the particular circumstances and the laws in question, the occurrence of the phenomenon *was to be expected*; and it is in this sense that the explanation enables us to *understand why* the phenomenon occurred.

Hempel 1965

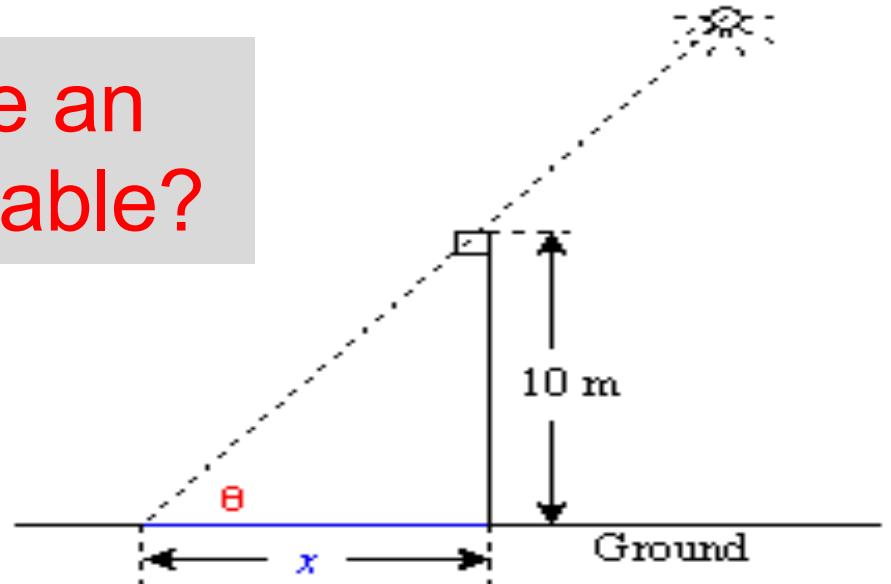
Question	<b>Why E?</b>	<b><i>Why was Mars at position X at time t?</i></b>
Law-like generalisations	$L_1 = ("If C_1, \dots, C_k \text{ then } E")$ , $L_2, \dots, L_r$	<i>Newton's laws of motion, the Newtonian inverse square law governing gravity</i>
Circumstances	$C_1, C_2, \dots, C_k$	<i>the mass of the sun, mass of Mars, their present position, their velocity</i>
Explanandum	E	<i>Mars at position X at time t</i>

# Summary

- Explanation as one aim of science
- Explanation provides understanding
- DN account: understanding a phenomenon achieved through deducing it from laws of nature

# Achieving Understanding

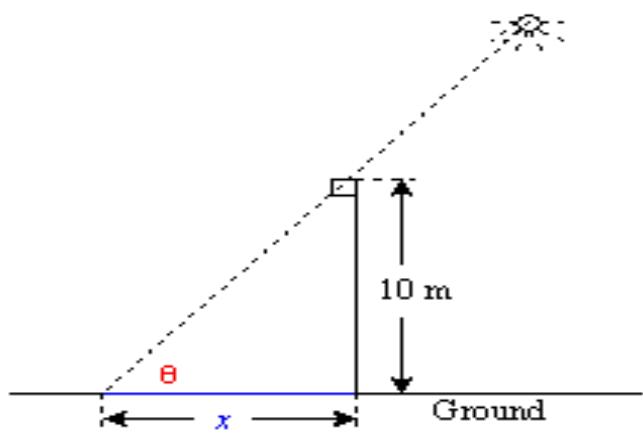
Which variable can provide an explanation of the other variable?



Please pause the video and answer the question before continuing!

- A. The shadow's length explains the height of the flagpole
- B. The flagpole's height explains the length of its shadow
- C. Either variable explains the other

## Explain the flagpole's length by its shadow?



- If we know the values of any two of the variables  $x$ ,  $h$ ,  $\theta$ , we can *calculate* the third
- However, the *productive relationships* between these variables is not symmetric:
  - Intervening on  $h$  or  $\theta$  will change  $x$
  - Intervening on  $x$  will *not* change  $h$  or  $\theta$
- To understand means to be able to say *what would have happen if things had been different*



James Woodward  
(\*1945)

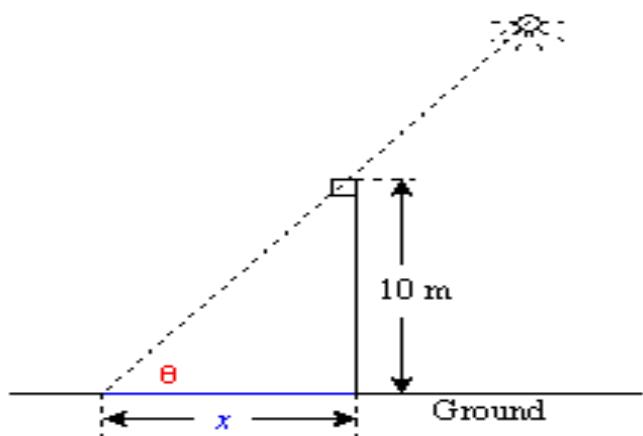


Judea Pearl  
(\*1936)



Michael Strevens  
(\*1965)

## Explain the flagpole's length by its shadow?



- If we know the values of any two of the variables  $x$ ,  $h$ ,  $\theta$ , we can *calculate* the third
- However, the *productive relationships* between these variables is not symmetric:
  - Intervening on  $h$  or  $\theta$  will change  $x$
  - Intervening on  $x$  will *not* change  $h$  or  $\theta$
- To understand means to be able to say *what would have happen if things had been different*
- Such *what-if* questions can be answered by tracing productive relationships
- Explanations help us understand by identifying the productive relationships – i.e. the relevant causes

Explain the  
pendulum's length  
by its period?

$$T = 2\pi * \sqrt{l/g}$$



$$l = T^2 * g / 4\pi^2$$

Explain the  
pendulum's length  
by its period

$$T = 2\pi * \sqrt{(\dots)}$$



$$l = T^2 * g / 4\pi^2$$

These cases satisfy the DN  
conditions, but intuitively do not  
constitute genuine explanations

→ DN account is not *sufficient* for  
explanation

(L) All biological males who take birth control pills regularly fail to get pregnant

(K) Mr. Jones is a biological male who has been taking birth control pills regularly

---

(E) Why does Mr. Jones fail to get pregnant?

(L) All biological males who take birth control pills regularly fail to get pregnant

(K) Mr. Smith has been taking birth control pills regularly for many years.

These cases satisfy the DN conditions, but intuitively do not constitute genuine explanations

→DN account is not *sufficient* for explanation

(E) Why?

# Singular Causal Explanations



- Vase stood on table
- 1 m above marble floor
- Hit table with my knee
- Vase fell...

# Singular Causal Explanations

When explaining a specific event, at least in everyday context we seem to do without laws.

# Singular Causal Explanations



These cases do *not* satisfy the DN conditions, but intuitively constitute genuine explanations  
→DN account is not *necessary* for explanation

To explain phenomenon E is to identify the *contributing cause* of E that makes a difference in the situation to be explained

Identifying the difference-making contributing causes is sufficient to answer *what if things had been different* questions – i.e.sufficient to provide understanding

# SCIENTIFIC KNOWLEDGE

## Explanation

*Understanding*

*Identify difference-making contributing cause for phenomenon to be explained*

## Prediction

*Forecasts*

*reasons for expecting a phenomenon to occur in a particular way.*

## Design

*Functioning artifacts*

*reasons for expecting that a manipulation satisfies certain functions*



Explanation NOT the same as prediction

# The Format of Explanation

Why did the vase break?

**Explanandum**

Features of phenomenon to be explained.



Because it was dropped.

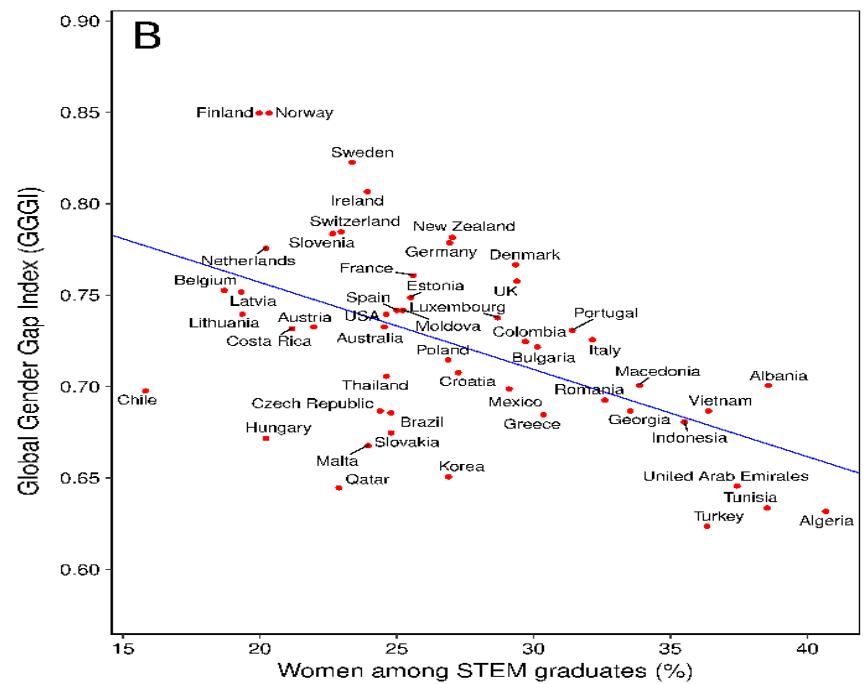
**Explanans**

Statements that increase understanding of explanandum

# Singular Explanandum



# General Explanandum



Contrastive Explanandum:

Why did the vase break *into fragments rather than just show fissures?*

Contrastive Explanans:

Because it was dropped *from height X onto a floor with stiffness Y rather than from height <X onto floor with stiffness <Y*

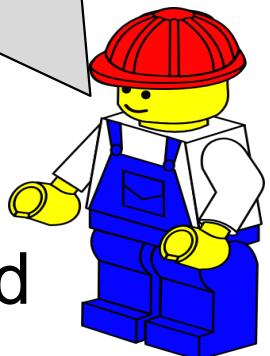


**Why did the crash occur  
with this driver while other  
drivers did traverse this  
curve safely?**

**Why did the crash occur  
in this situation while it doesn't  
occur in situations in which similar  
cars moving at similar speeds  
with similarly competent drivers  
traverse other curves?**



- (i) Because the driver was intoxicated
- (ii) Because the curve was too tightly banked



What makes  
Explanations powerful

1. Accuracy – whether the *explanans* describes the actual state/properties of the world
  - Explanation needs to identify only the difference-making contributing cause(s)

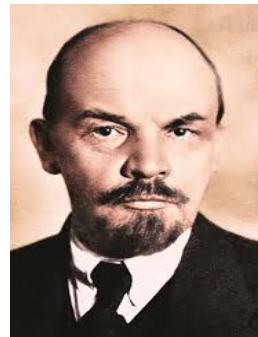
2. Precision (of the Explanandum) – the more precise the contrast is stated in the explanandum, the better the explanation.

3. Difference-Making (of the Explanans) – The *explanans* must identify all the contributing causes that produced the difference asked for in the *explanandum*.

4. Non-sensitivity (of the Explanans) – Some explanans causes are more *sensitive* to background causes than others

# Illustration: A Very Sensitive Explanans

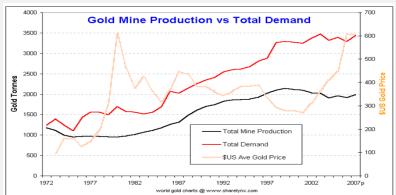
"I don't want to listen to [Beethoven's *Appassionata*] because it makes me want to stroke people's heads, and I have to smash those heads to bring the revolution to them."



4. Non-sensitivity (of the Explanans) – The less sensitive an accurate difference-making explanans, the more powerful the explanation

5. Cognitive Salience – The more easily a given explanation can be grasped, the more powerful it is.

What is the necessary amount of detail required in the *explanans* to explain the *explanandum*?



Aggregate demand & supply



Actual interactions on trading floors



Neural basis for decisions

Accuracy goes up



Salience goes up



# Summary

- Causal explanations, i.e. those that identify difference-making contributing causes of an explanandum, can be better or worse
- 5 dimensions of making them better

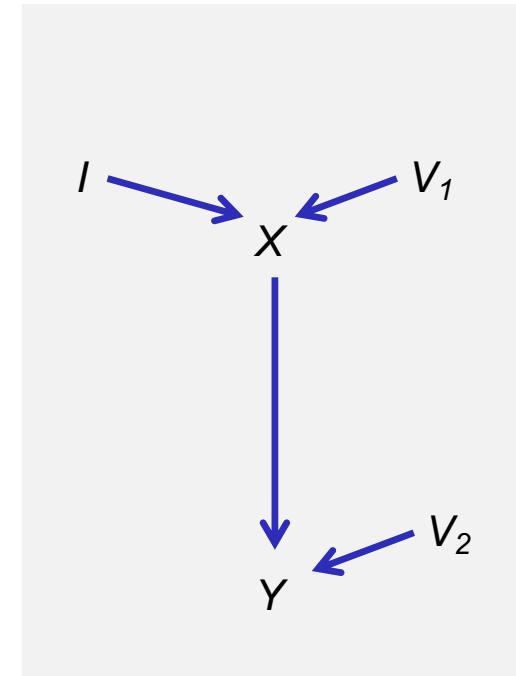
# What is Causation?

$X$  is a *direct cause* of  $Y$  with respect to a background variable set  $V$



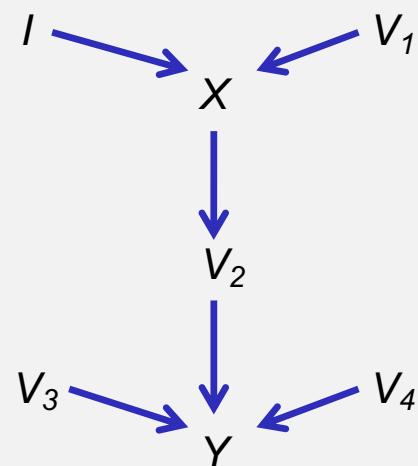
there is a *possible* intervention on  $X$  that will change  $Y$  when all other variables in  $V$  are held fixed.

(Woodward 2003)

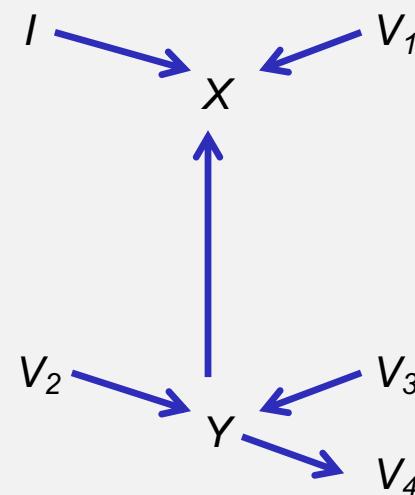


In which of the following models is X a direct cause of Y?

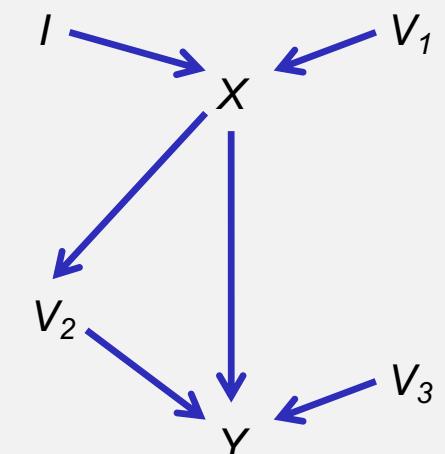
A



B



C



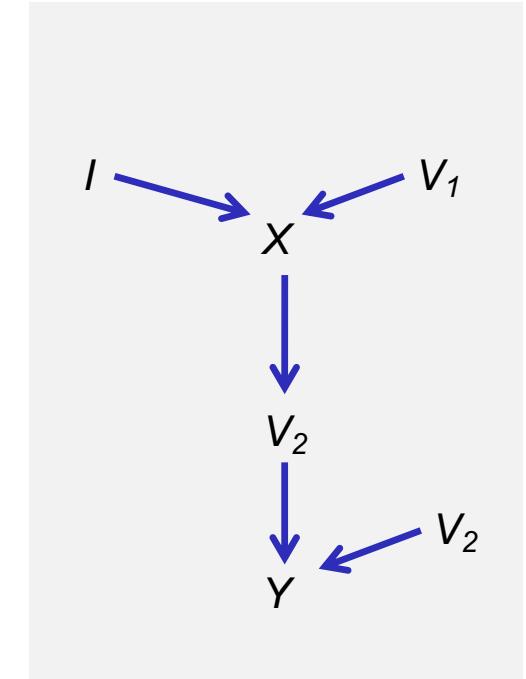
Pause the video and answer this question before continuing.

$X$  is a *contributing cause* of  $Y$  with respect to a background variable set  $V$



there is a causal chain, each link of which consists in a direct cause, extending from  $X$  to  $Y$

(Woodward 2003)



$X$  not a direct cause of  $Y$ ,  
but a contributing cause,  
wrt to  $\{V_i\}$

# Summary

- Manipulability account of *direct cause*
- Derived from that: account of *contributing cause*

How to learn about  
Causes?

# **Correlation**

**≠**

# **Causation**

## **Correlation**

- measures the association between two variables

## **Causation**

- measures the productive influence of one variable on another



David Hume  
(1711-1776)

We only observe correlation but never causation.



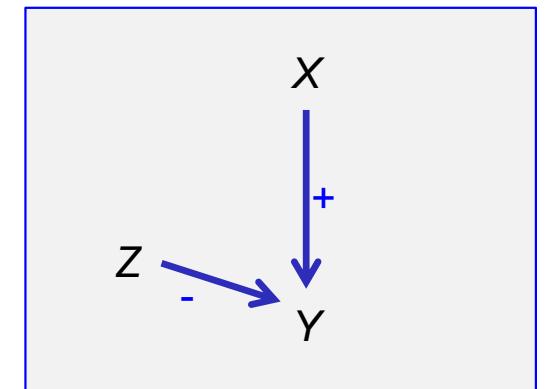
Judea Pearl  
(\*1936)

Observing correlation is an important kind of evidence for causation.

# Correlation Not Necessary for Causation

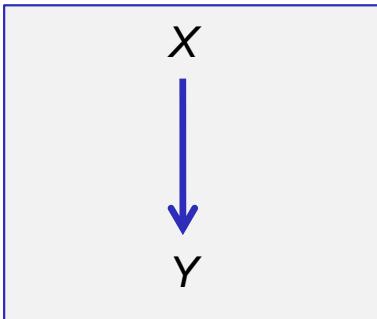
# Correlation Not Necessary for Causation

$$\text{Cov}(X, Y) = 0$$

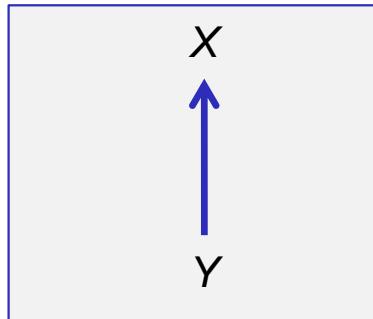


# Correlation Not Sufficient for Causation

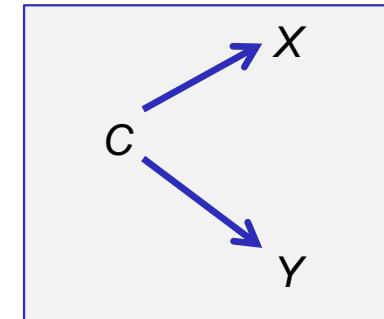
$$\text{Cov}(X, Y) \gg 0$$



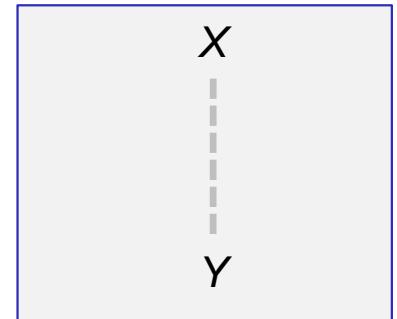
X causes Y



Y causes X



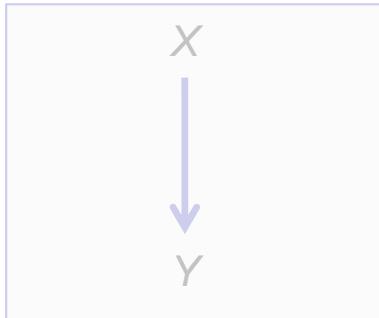
A common cause C  
causes both X and Y



X, Y  
independent but  
correlated

# Correlation Not Sufficient for Causation

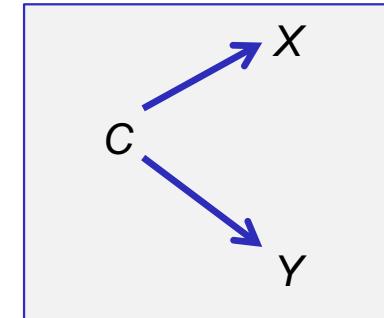
$$\text{Cov}(X, Y) \gg 0$$



X causes Y



Y causes X



A common cause C  
causes both X and Y



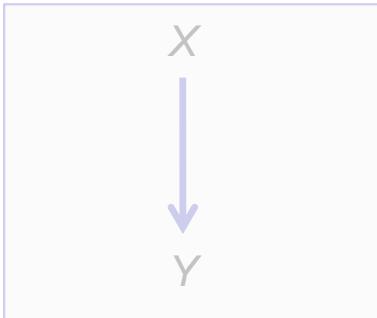
X, Y  
independent but  
correlated

## Examples:

- Per capita candy consumption correlated with divorce rate – common cause: age
- hormone replacement therapy correlated with coronary heart disease – common cause: socio-economic status

# Correlation Not Sufficient for Causation

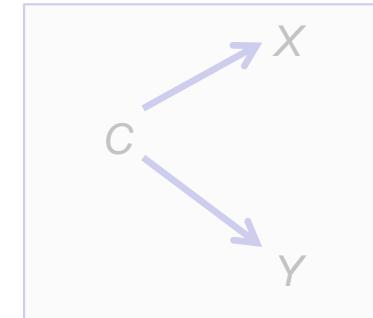
$$\text{Cov}(X, Y) \gg 0$$



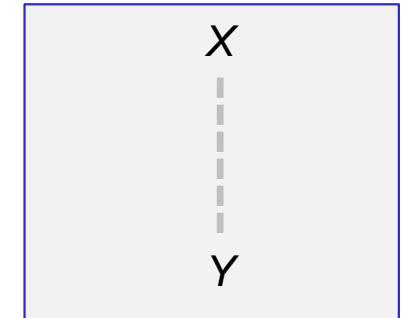
X causes Y



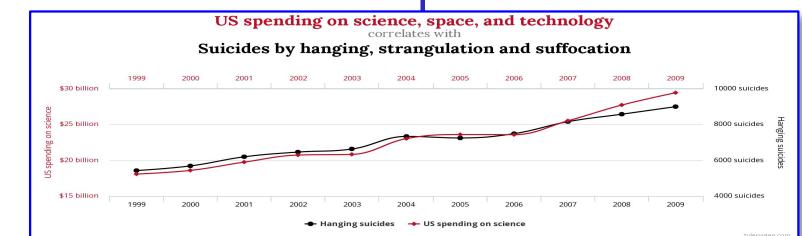
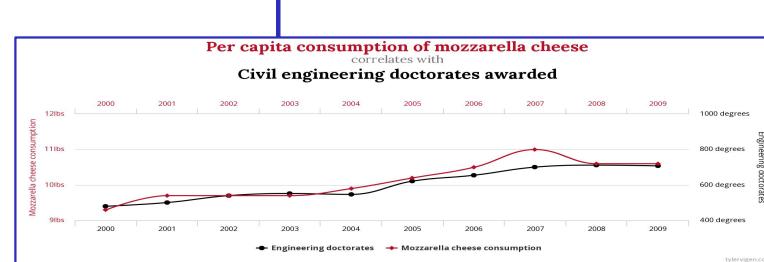
Y causes X



A common cause C  
causes both X and Y

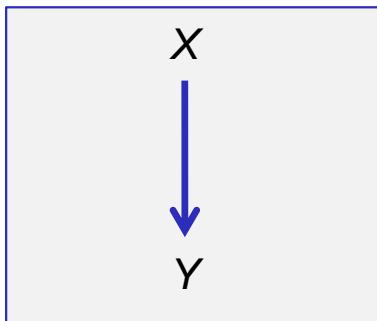


X, Y  
independent but  
correlated

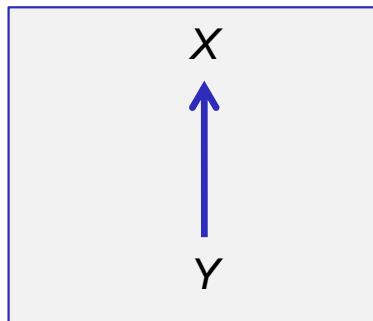


# Correlation Not Sufficient for Causation

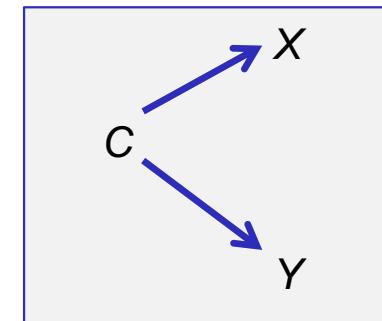
$$\text{Cov}(X, Y) \gg 0$$



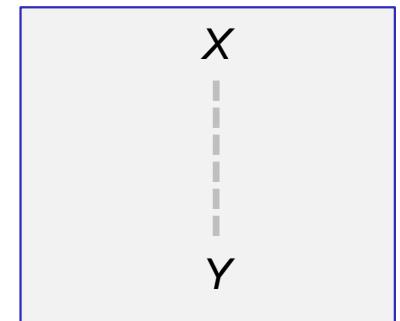
X causes Y



Y causes X



A common cause C  
causes both X and Y



X, Y  
independent but  
correlated

Many different causal models are compatible with correlation data! How to determine the correct one?

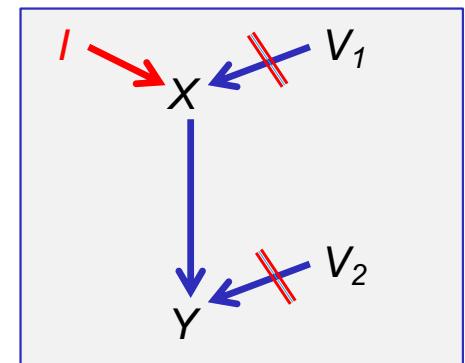
# Strategy 1: Controlled Experiments



## Mill's Method of Difference

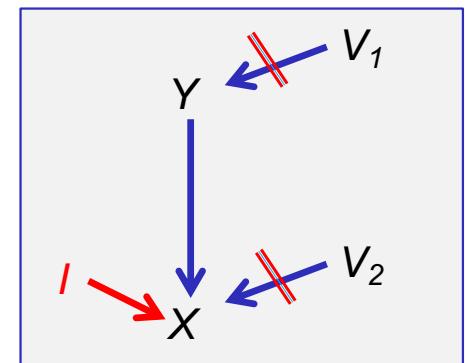
1. Control all background variables influencing X and Y
2. Intervene on hypothesized cause to see whether it makes a difference on hypothesised effect

# Strategy 1: Controlled Experiments



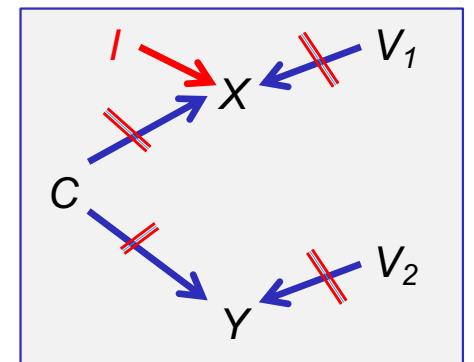
X causes Y

# Strategy 1: Controlled Experiments



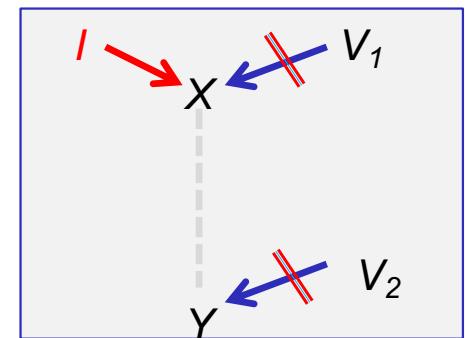
Y causes X

# Strategy 1: Controlled Experiments



A *common cause* C  
causes both X and Y

# Strategy 1: Controlled Experiments



$X, Y$   
independent but  
correlated

# Strategy 2: Instrumental Variable Analysis

For detecting causes from observational data

1. Observe correlation between X and Y
2. Find a variable Z that you know affects X, *but not* Y (the instrument)
3. Use the instrument Z instead of X when estimating the effect of X on Y

Example

$$\text{Cov}(\text{Smoking}, \text{Health}) >> 0$$



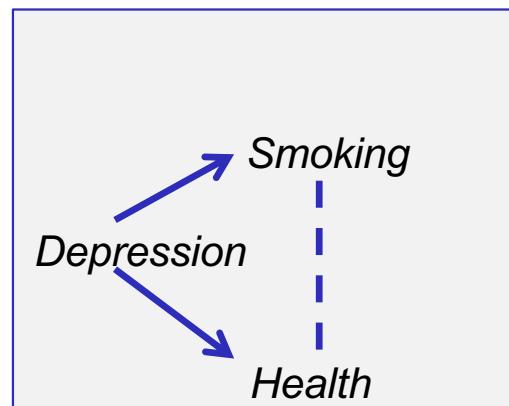
# Strategy 2: Instrumental Variable Analysis

For detecting causes from observational data

1. Observe correlation between X and Y
2. Find a variable Z that you know affects X, *but not* Y (the instrument)
3. Use the instrument Z instead of X when estimating the effect of X on Y

Example

$$\text{Cov}(\text{Smoking}, \text{Health}) >> 0$$



# Strategy 2: Instrumental Variable Analysis

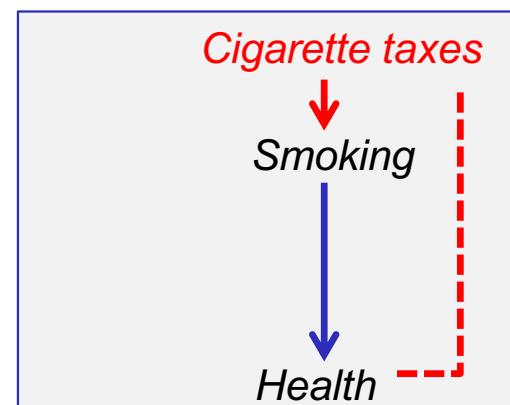
For detecting causes from observational data

1. Observe correlation between X and Y
2. Find a variable Z that you know affects X, *but not* Y (the instrument)
3. Use the instrument Z instead of X when estimating the effect of X on Y

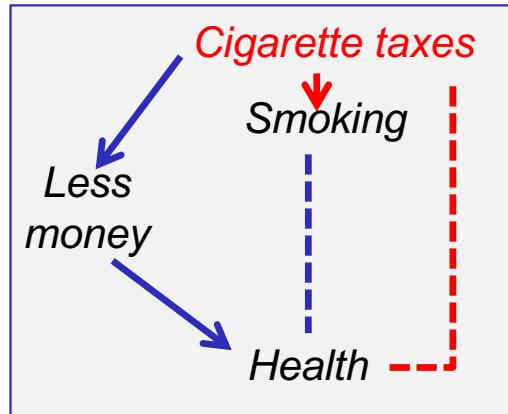
Example

$$\text{Cov}(\text{Smoking}, \text{Health}) >> 0$$

$$\text{Cov}(\text{Taxes}, \text{Health}) = ?$$



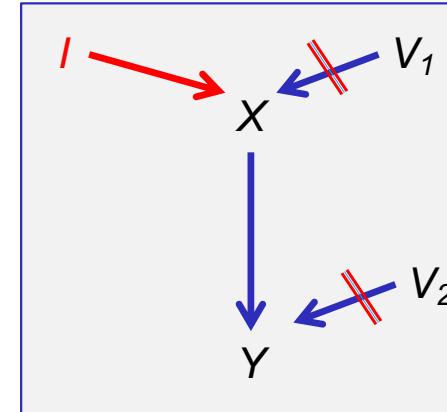
# These Strategies Require Causal Knowledge!



Need to know that tax increases do not cause health decreases



Nancy Cartwright  
(\*1944)



Need to know all relevant background conditions for control

**"No causes in, no causes out"**

# Summary

- Causes ≠ Correlations
- Correlations as evidence for causes
- Experimental & observational strategies for generating that evidence