Microsoft Research

Each year Microsoft Research hosts hundreds of influential speakers from around the world including leading scientists, renowned experts in technology, book authors, and leading academics, and makes videos of these lectures freely available.

CAUSES AND COUNTERFACTUALS: CONCEPTS, PRINCIPLES AND TOOLS

Judea Pearl
Elias Bareinboim
University of California, Los Angeles
{judea, eb}@cs.ucla.edu

NIPS 2013 Tutorial

CAUSES AND COUNTERFACTUALS: CONCEPTS, PRINCIPLES AND TOOLS

Judea Pearl
Elias Bareinboim
University of California, Los Angeles
{judea, eb}@cs.ucla.edu

NIPS 2013 Tutorial

OUTLINE

Concepts:

* Causal inference – a paradigm shift
* The two fundamental laws

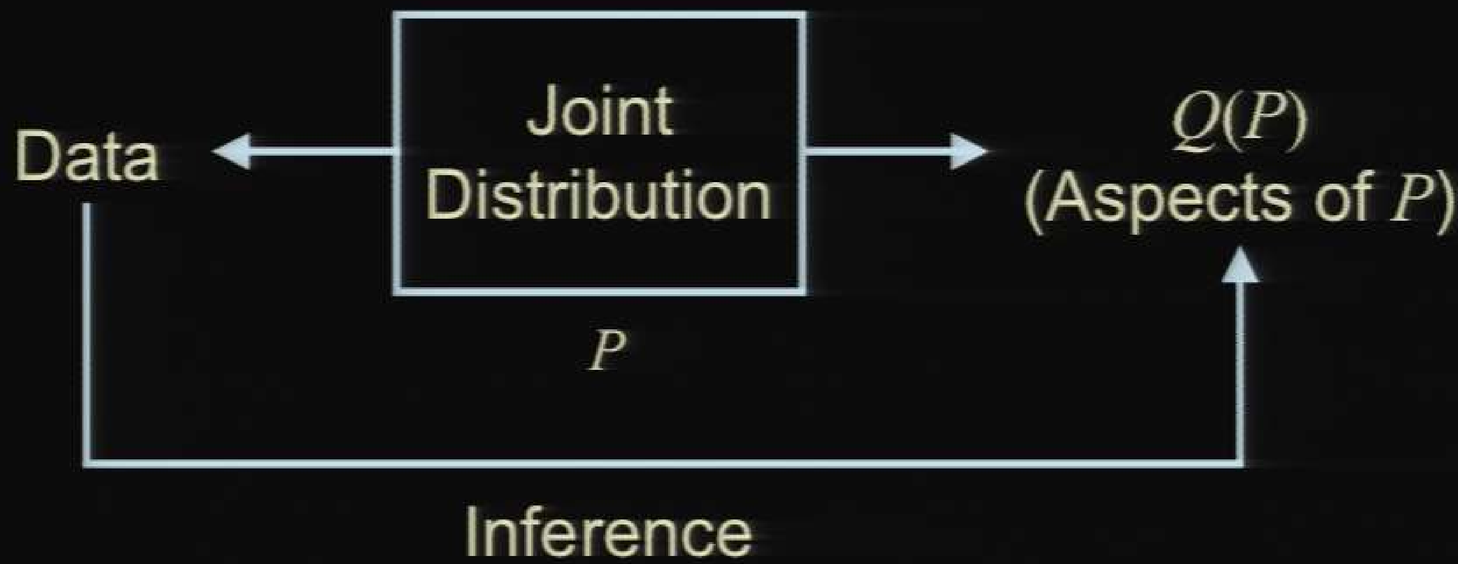
Basic tools:

* Graph separation
* The truncated product formula
* The back-door adjustment formula
* The do-calculus

Capabilities:

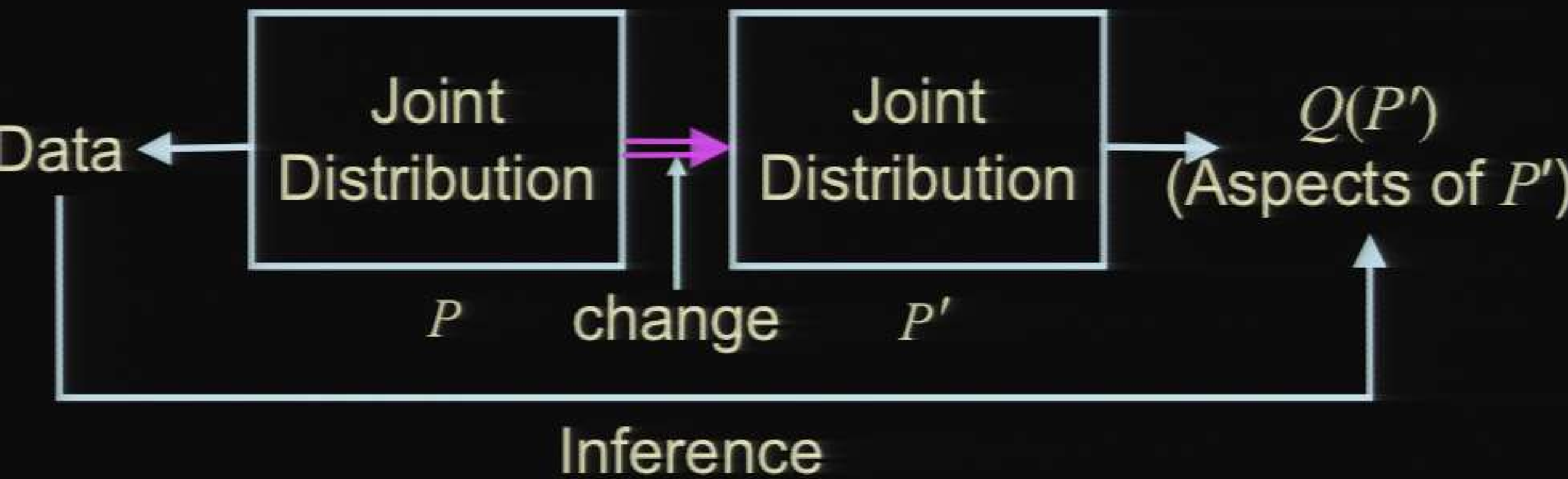
* Policy evaluation
* Transportability
* Mediation
* Missing Data

TRADITIONAL STATISTICAL INFERENCE PARADIGM



e.g.,
Infer whether customers who bought product A would also buy product B.
$Q = P(B \mid A)$

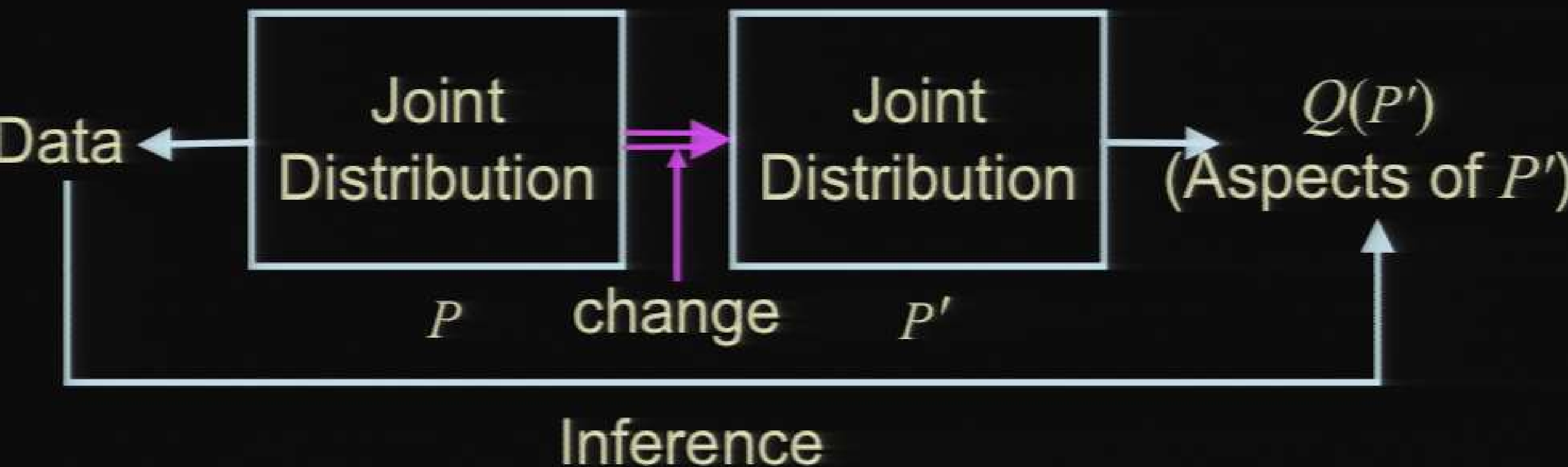
FROM STATISTICAL TO CAUSAL ANALYSIS: 1. THE DIFFERENCES



e.g., Estimate $P'(sales)$ if we double the price.
How does P change to P'? New oracle
e.g., Estimate $P'(cancer)$ if we ban smoking.

FROM STATISTICAL TO CAUSAL ANALYSIS: 1. THE DIFFERENCES

What remains invariant when P changes say, to satisfy $P'(price=2)=1$

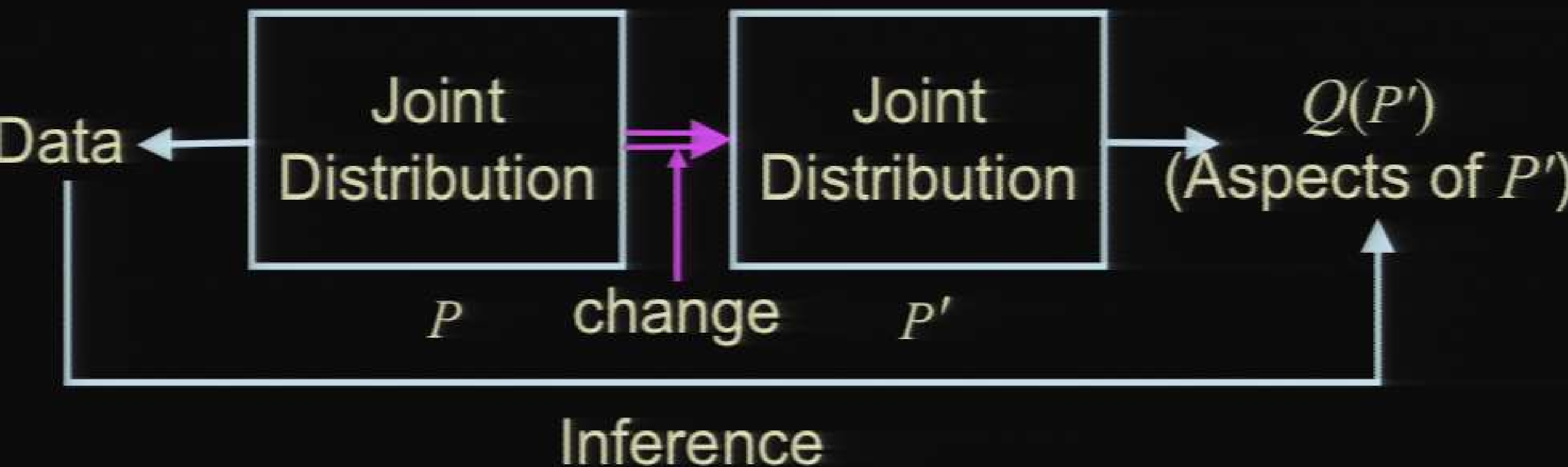


Note: $P'(sales) \neq P\,(sales \mid price = 2)$

e.g., Doubling price $\neq$ seeing the price doubled.

FROM STATISTICAL TO CAUSAL ANALYSIS: 1. THE DIFFERENCES

What remains invariant when P changes say, to satisfy $P'(price=2)=1$



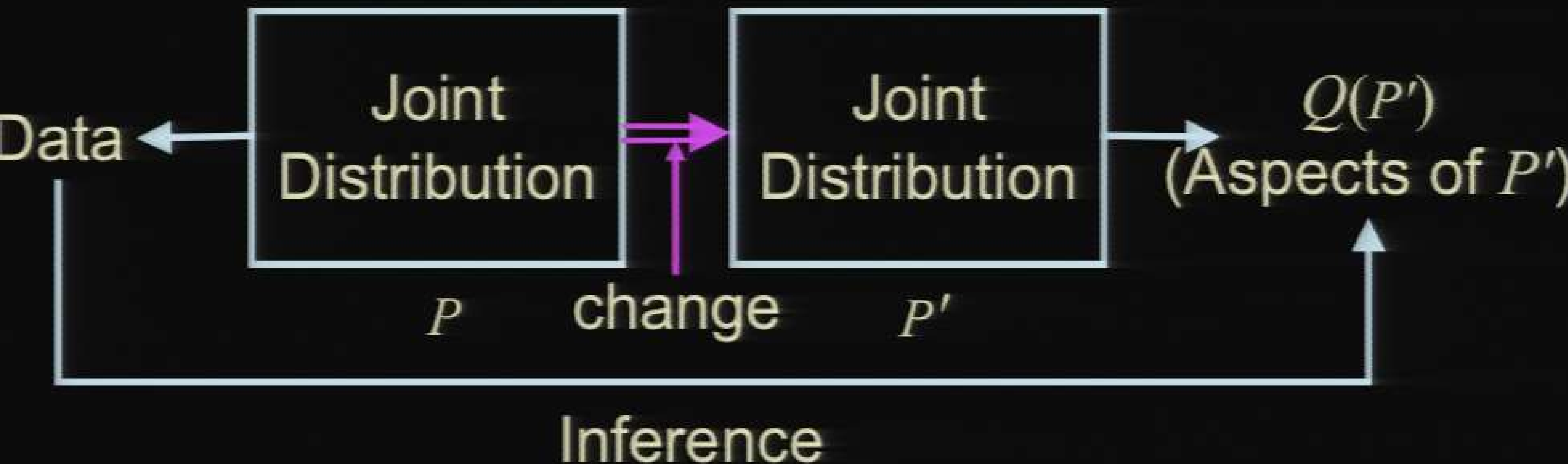
Note: $P'(sales) \neq P\,(sales \mid price = 2)$

e.g., Doubling price $\neq$ seeing the price doubled.

P does not tell us how it ought to change.

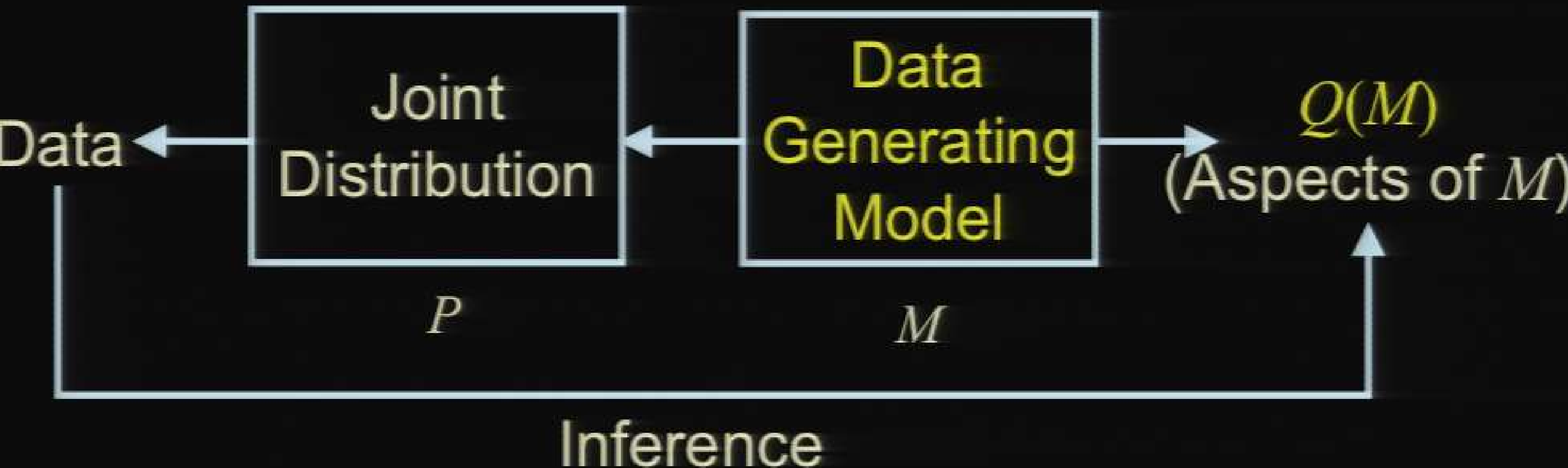
FROM STATISTICAL TO COUNTERFACTUALS: 1. THE DIFFERENCES

Probability and statistics deal with static relations



What happens when P changes?
e.g., Estimate the probability that a customer who bought A would buy A if we were to double the price.

THE STRUCTURAL MODEL PARADIGM



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

•

WHAT KIND OF QUESTIONS SHOULD THE NEW ORACLE ANSWER

THE CAUSAL HIERARCHY

- Observational Questions:
  "What if we see A"
- Action Questions:
  "What if we do A?"
- Counterfactuals Questions:
  "What if we did things differently?"
- Options:
  "With what probability?"

WHAT KIND OF QUESTIONS SHOULD THE NEW ORACLE ANSWER
THE CAUSAL HIERARCHY

- Observational Questions:
  "What if we see A" — Bayes Networks
- Action Questions:
  "What if we do A?" — Causal Bayes Networks
- Counterfactuals Questions:
  "What if we did things differently?" — Functional Causal Diagrams
- Options:
  "With what probability?"

GRAPHICAL REPRESENTATIONS

FROM STATISTICAL TO CAUSAL ANALYSIS: 2. THE SHARP BOUNDARY

1. Causal and associational concepts do not mix.

CAUSAL	ASSOCIATIONAL
Spurious correlation	Regression
Randomization / Intervention	Association / Independence
"Holding constant" / "Fixing"	"Controlling for" / Conditioning
Confounding / Effect	Odds and risk ratios
Instrumental variable	Collapsibility / Granger causality
Ignorability / Exogeneity	Propensity score

2.

3.

4.

FROM STATISTICAL TO CAUSAL ANALYSIS: 3. THE MENTAL BARRIERS

1. Causal and associational concepts do not mix.

CAUSAL	ASSOCIATIONAL
Spurious correlation	Regression
Randomization / Intervention	Association / Independence
"Holding constant" / "Fixing"	"Controlling for" / Conditioning
Confounding / Effect	Odds and risk ratios
Instrumental variable	Collapsibility / Granger causality
Ignorability / Exogeneity	Propensity score

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

$$\left.\begin{array}{r}\text{data}\\ \text{causal assumptions (or experiments)}\end{array}\right\} \Rightarrow \text{causal conclusions}$$

FROM STATISTICAL TO CAUSAL ANALYSIS: 3. THE MENTAL BARRIERS

1. Causal and associational concepts do not mix.

CAUSAL	ASSOCIATIONAL
Spurious correlation	Regression
Randomization / Intervention	Association / Independence
"Holding constant" / "Fixing"	"Controlling for" / Conditioning
Confounding / Effect	Odds and risk ratios
Instrumental variable	Collapsibility / Granger causality
Ignorability / Exogeneity	Propensity score

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

$$\left.\begin{array}{r}\text{data}\\ \text{causal assumptions (or experiments)}\end{array}\right\} \Rightarrow \text{causal conclusions}$$

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

FROM STATISTICAL TO CAUSAL ANALYSIS: 3. THE MENTAL BARRIERS

1. Causal and associational concepts do not mix.

CAUSAL	ASSOCIATIONAL
Spurious correlation	Regression
Randomization / Intervention	Association / Independence
"Holding constant" / "Fixing"	"Controlling for" / Conditioning
Confounding / Effect	Odds and risk ratios
Instrumental variable	Collapsibility / Granger causality
Ignorability / Exogeneity	Propensity score

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

$$\left.\begin{array}{r}\text{data}\\ \text{causal assumptions (or experiments)}\end{array}\right\} \Rightarrow \text{causal conclusions}$$

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 $(x \;\square\!\!\rightarrow Y)$)

THE NEW ORACLE: STRUCTURAL CAUSAL MODELS THE WORLD AS A COLLECTION OF SPRINGS

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

- $V = \{V_1, ..., V_n\}$ are endogenous variables
- $U = \{U_1, ..., U_m\}$ are background variables
- $F = \{f_1, ..., f_n\}$ are functions determining V,
  $v_i = f_i(v, u)$

THE NEW ORACLE: STRUCTURAL CAUSAL MODELS THE WORLD AS A COLLECTION OF SPRINGS

Definition: A structural causal model is a 4-tuple $<V,U, F, P(u)>$, where

- $V = \{V_1,...,V_n\}$ are endogenous variables
- $U = \{U_1,...,U_m\}$ are background variables
- $F = \{f_1,..., f_n\}$ are functions determining V,

  $v_i = f_i(v, u)$ e.g., $y = \alpha + \beta x + u_Y$ Not regression!!!!

THE NEW ORACLE: STRUCTURAL CAUSAL MODELS THE WORLD AS A COLLECTION OF SPRINGS

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

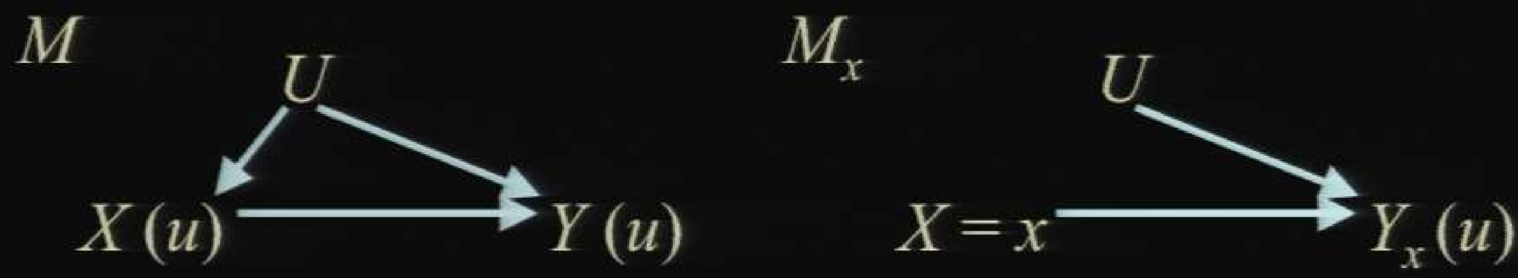
- $V = \{V_1, \ldots, V_n\}$ are endogenous variables
- $U = \{U_1, \ldots, U_m\}$ are background variables
- $F = \{f_1, \ldots, f_n\}$ are functions determining V,
  $v_i = f_i(v, u)$ e.g., $y = \alpha + \beta x + u_Y$ Not regression!!!!
- $P(u)$ is a distribution over U

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

COUNTERFACTUALS ARE EMBARRASSINGLY SIMPLE

Definition:

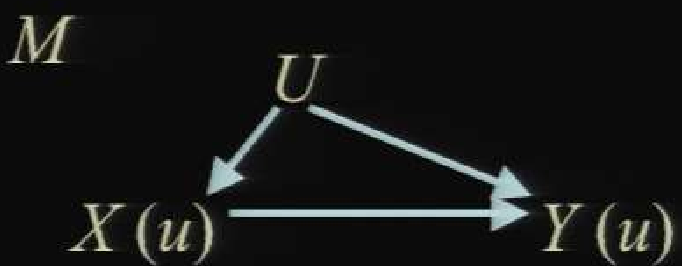
Given a SCM model M, the potential outcome $Y_x(u)$ for unit u is equal to the solution for Y in a mutilated model M_x, in which the equation for X is replaced by $X = x$.

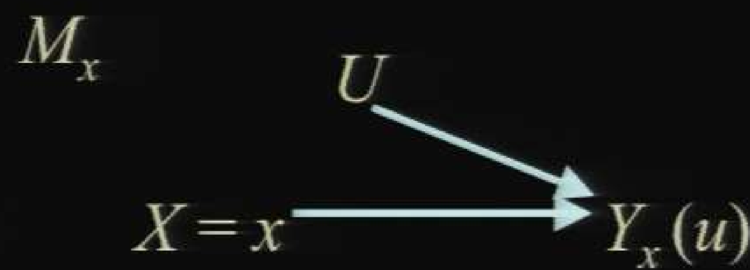


COUNTERFACTUALS ARE EMBARRASSINGLY SIMPLE

Definition:

Given a SCM model M, the potential outcome $Y_x(u)$ for unit u is equal to the solution for Y in a mutilated model M_x, in which the equation for X is replaced by $X = x$.



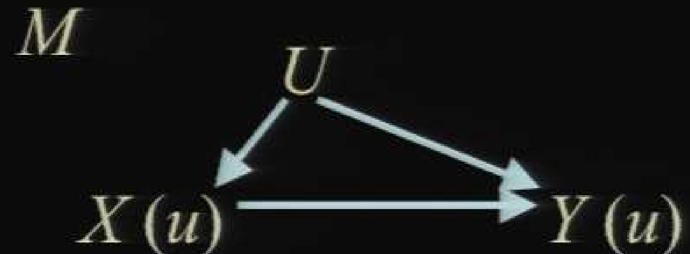


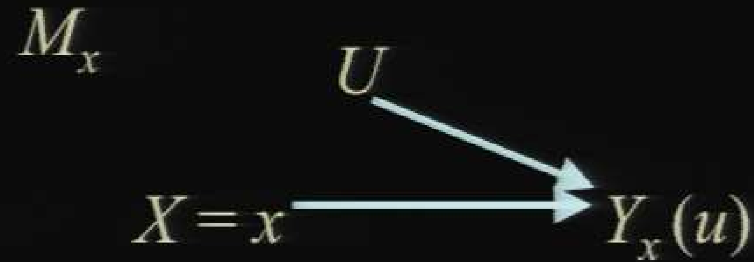
The Fundamental Equation of Counterfactuals:

$$\boxed{Y_x(u) \triangleq Y_{M_x}(u)}$$

EFFECTS OF INTERVENTIONS ARE EMBARRASSINGLY SIMPLE

Definition:
Given a SCM model M, the effect of setting X to x, $P(Y = y \mid do\,(X{=}x))$, is equal to the probability of $Y = y$ in a mutilated model M_x, in which the equation for X is replaced by $X = x$.

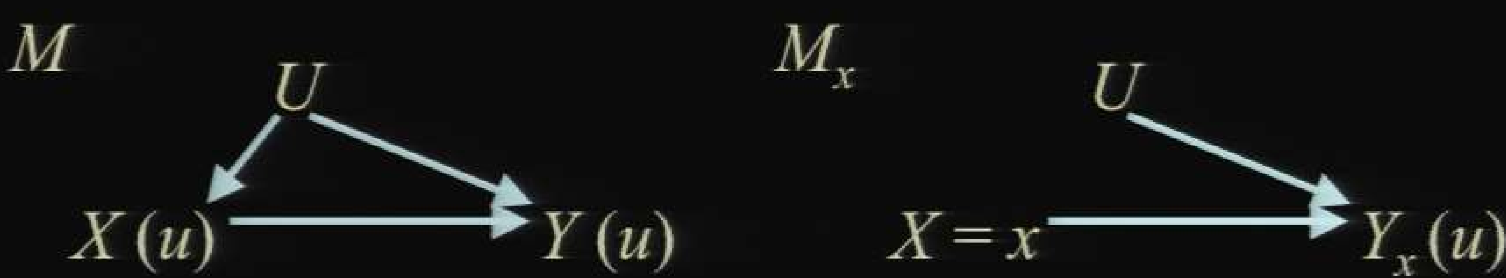




EFFECTS OF INTERVENTIONS ARE EMBARRASSINGLY SIMPLE

Definition:

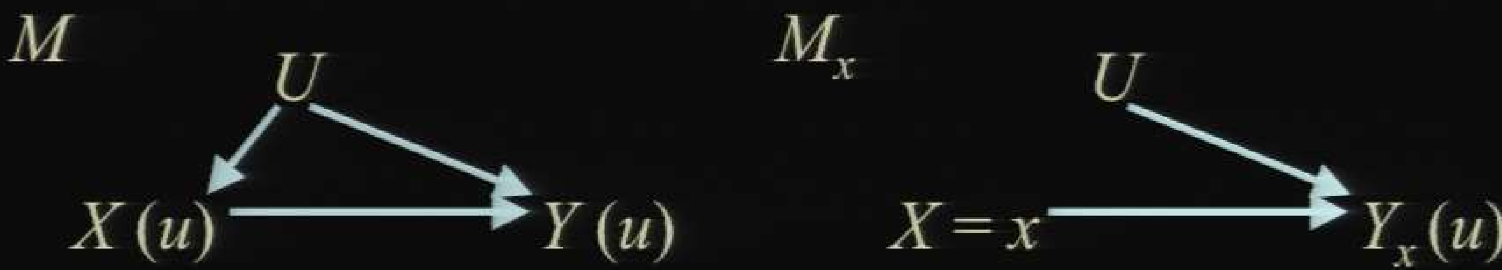
Given a SCM model M, the effect of setting X to x, $P(Y = y \mid do\,(X{=}x))$, is equal to the probability of $Y = y$ in a mutilated model M_x, in which the equation for X is replaced by $X = x$.



The Fundamental Equation of Interventions:

$$\boxed{P(Y = y \mid do(X = x)) \triangleq P_{M_x}(Y = y)}$$

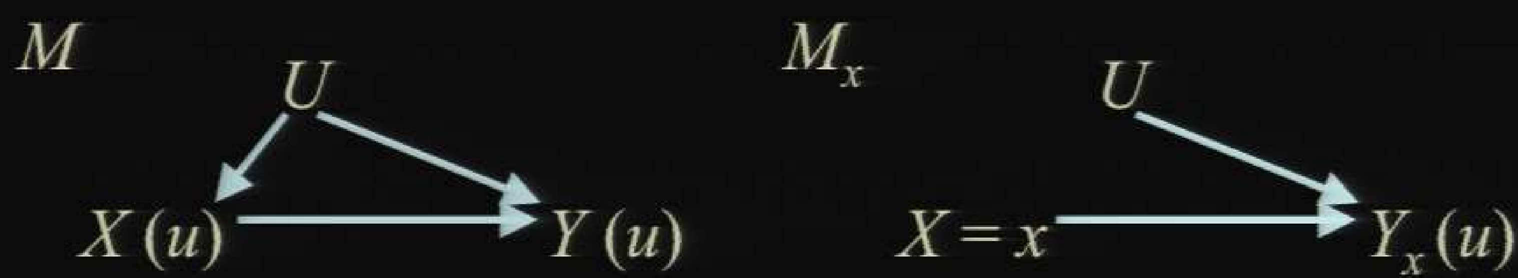
COMPUTING THE EFFECTS OF INTERVENTIONS



The Fundamental Equation of Interventions:

$$\boxed{P(Y = y \mid do(X = x)) \triangleq P_{M_x}(Y = y)}$$

COMPUTING THE EFFECTS OF INTERVENTIONS



The Fundamental Equation of Interventions:

$$\boxed{P(Y = y \mid do(X = x)) \stackrel{\Delta}{=} P_{M_x}(Y = y)}$$

$$P(x, y, u) = P(u)\cancel{P(x \mid u)}P(y \mid x, u)$$

$$P(y, u \mid do(x)) = P(u)P(y \mid x, u)$$

THE TWO FUNDAMENTAL LAWS OF CAUSAL INFERENCE

THE TWO FUNDAMENTAL LAWS OF CAUSAL INFERENCE

1. The Law of Counterfactuals (and Interventions)

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

(M generates and evaluates all counterfactuals.)

THE TWO FUNDAMENTAL LAWS OF CAUSAL INFERENCE

1. The Law of Counterfactuals (and Interventions)

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

(M generates and evaluates all counterfactuals.)

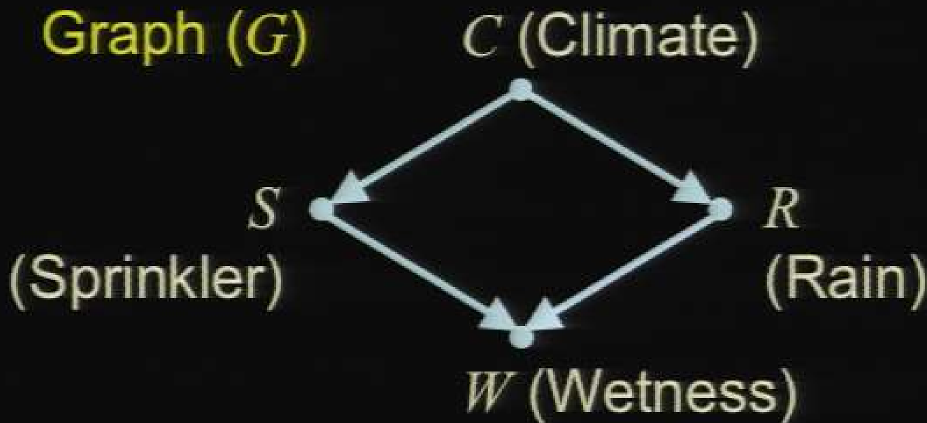
2. The Law of Conditional Independence (d-separation)

$$\boxed{(X \text{ sep } Y \mid Z)_{G(M)} \Rightarrow (X \perp\!\!\!\perp Y \mid Z)_{P(v)}}$$

(Separation in the model ⇒ independence in the distribution.)

THE LAW OF CONDITIONAL INDEPENDENCE

Graph (G)

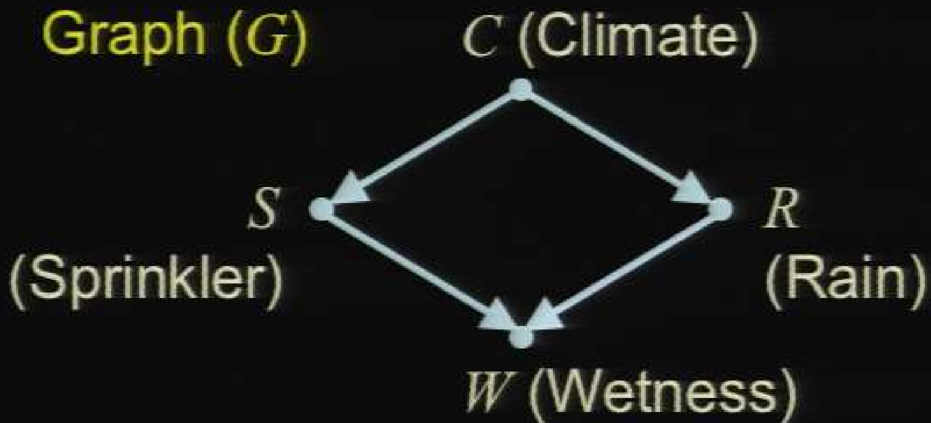


Model (M)

$$C = f_C(U_C)$$
$$S = f_S(C, U_S)$$
$$R = f_R(C, U_R)$$
$$W = f_W(S, R, U_W)$$

THE LAW OF CONDITIONAL INDEPENDENCE

Graph (G)



Model (M)

$$C = f_C(U_C)$$
$$S = f_S(C, U_S)$$
$$R = f_R(C, U_R)$$
$$W = f_W(S, R, U_W)$$

Gift of the Gods

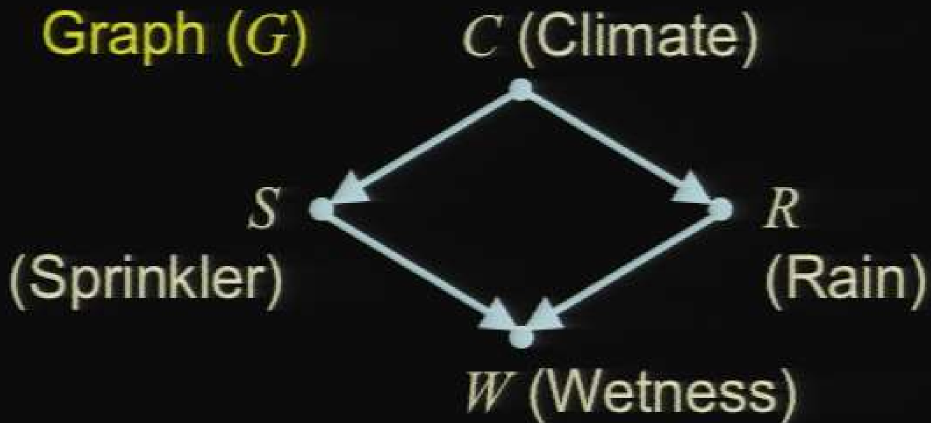
If the U's are independent, the observed distribution $P(C,R,S,W)$ satisfies constraints that are:

(1) independent of the f's and of $P(U)$,

(2) readable from the graph.

D-SEPARATION: NATURE'S LANGUAGE FOR COMMUNICATING ITS STRUCTURE

Graph (G)



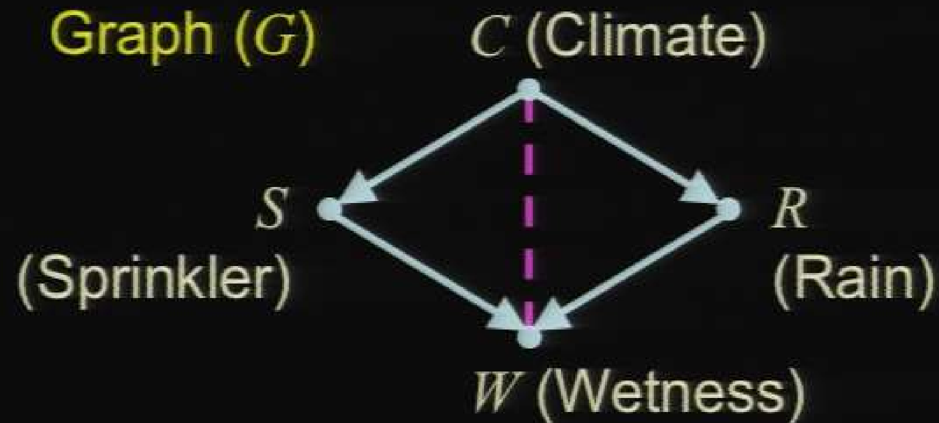
Model (M)

$$C = f_C(U_C)$$
$$S = f_S(C, U_S)$$
$$R = f_R(C, U_R)$$
$$W = f_W(S, R, U_W)$$

Every missing arrow advertises an independency, conditional on a separating set.

D-SEPARATION: NATURE'S LANGUAGE FOR COMMUNICATING ITS STRUCTURE

Graph (G)



Model (M)

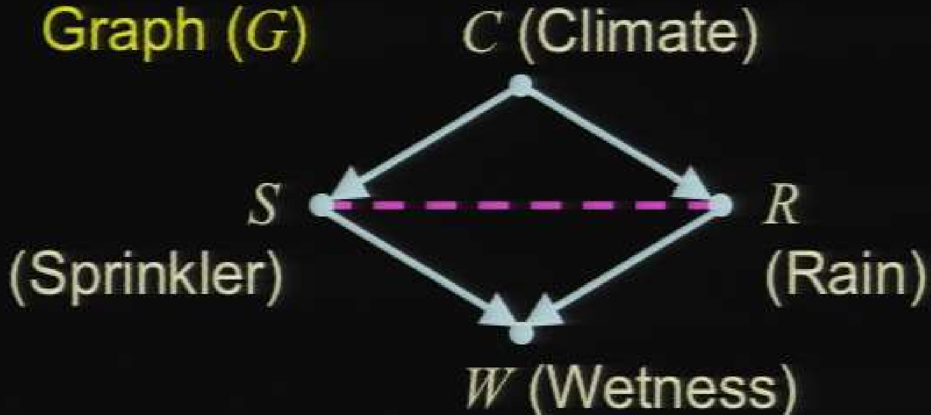
$$C = f_C(U_C)$$
$$S = f_S(C, U_S)$$
$$R = f_R(C, U_R)$$
$$W = f_W(S, R, U_W)$$

Every missing arrow advertises an independency, conditional on a separating set.

e.g., $C \perp\!\!\!\perp W \mid (S, R)$

D-SEPARATION: NATURE'S LANGUAGE FOR COMMUNICATING ITS STRUCTURE

Graph (G)



Model (M)

$$C = f_C(U_C)$$
$$S = f_S(C, U_S)$$
$$R = f_R(C, U_R)$$
$$W = f_W(S, R, U_W)$$

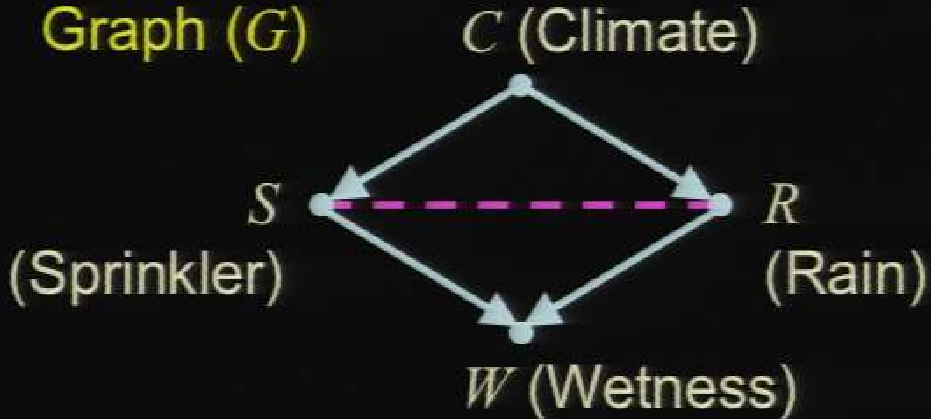
Every missing arrow advertises an independency, conditional on a separating set.

e.g., $C \perp\!\!\!\perp W \mid (S, R)$ $\qquad$ $S \perp\!\!\!\perp R \mid C$

Applications:

D-SEPARATION: NATURE'S LANGUAGE FOR COMMUNICATING ITS STRUCTURE

Graph (G)



Model (M)

$$C = f_C(U_C)$$
$$S = f_S(C, U_S)$$
$$R = f_R(C, U_R)$$
$$W = f_W(S, R, U_W)$$

Every missing arrow advertises an independency, conditional on a separating set.

e.g., $C \perp\!\!\!\perp W \mid (S, R)$ $\qquad$ $S \perp\!\!\!\perp R \mid C$

Applications:

1. Model testing

D-SEPARATION: NATURE'S LANGUAGE FOR COMMUNICATING ITS STRUCTURE

Graph (G)

Model (M)

$$C = f_C(U_C)$$
$$S = f_S(C, U_S)$$
$$R = f_R(C, U_R)$$
$$W = f_W(S, R, U_W)$$

Every missing arrow advertises an independency, conditional on a separating set.

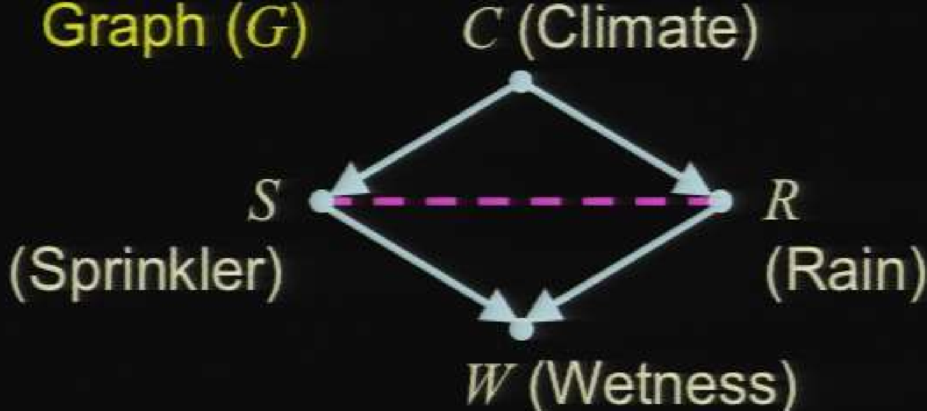
e.g., $C \perp\!\!\!\perp W \mid (S, R)$ $\qquad$ $S \perp\!\!\!\perp R \mid C$

Applications:

1. Model testing
2. Structure learning

D-SEPARATION: NATURE'S LANGUAGE FOR COMMUNICATING ITS STRUCTURE

Graph (G)



Model (M)

$$C = f_C(U_C)$$
$$S = f_S(C, U_S)$$
$$R = f_R(C, U_R)$$
$$W = f_W(S, R, U_W)$$

Every missing arrow advertises an independency, conditional on a separating set.

e.g., $C \perp\!\!\!\perp W \mid (S, R)$ $\qquad$ $S \perp\!\!\!\perp R \mid C$

Applications:

1. Model testing
2. Structure learning
3. Reducing "what if I do" questions to symbolic calculus

OUTLINE

Concepts:

* Causal inference – a paradigm shift
* The two fundamental laws

Basic tools:

* Graph separation
* The truncated product formula
* The back-door adjustment formula
* The do-calculus

Capabilities:

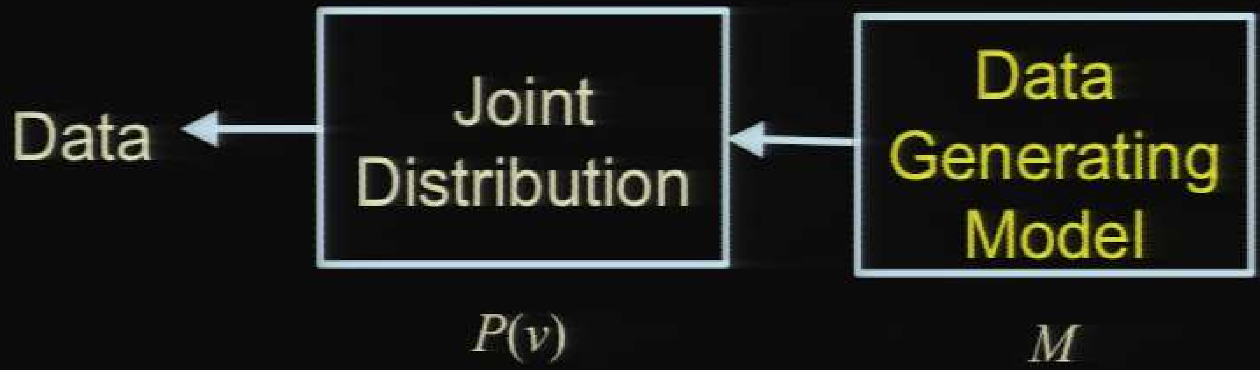
* Policy evaluation
* Transportability
* Mediation
* Missing Data

FIRST LAYER OF THE CAUSAL HIERARCHY

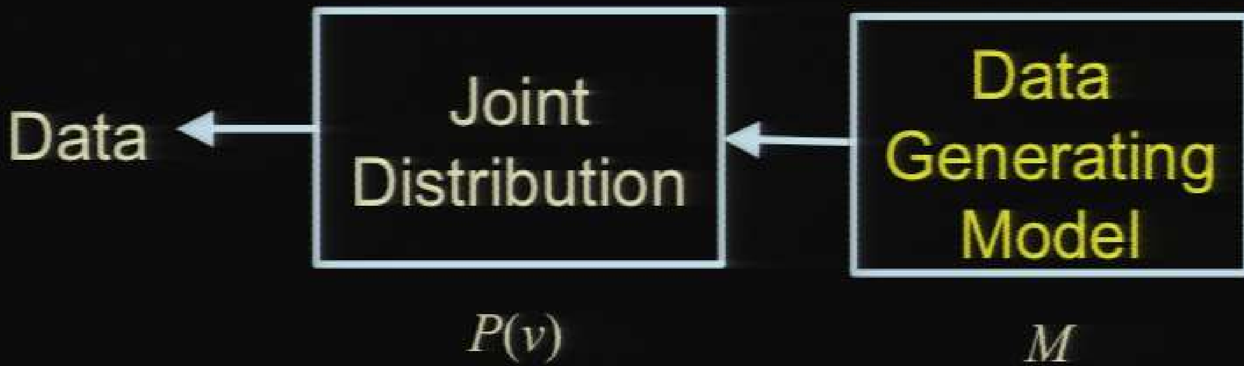
PROBABILITIES

(What if I see $X = x$?)

THE EMERGENCE OF THE FIRST LAYER



THE EMERGENCE OF THE FIRST LAYER



Theorem (PV, 1991). Every Markovian structural causal model M (recursive, with independent disturbances) induces a passive distribution $P(v_1, \ldots, v_n)$ that can be factorized as

$$P(v_1, v_2, \ldots, v_n) = \prod_i P(v_i \mid pa_i)$$

where pa_i are the (values of) the parents of V_i in the causal diagram associated with M.

OUTLINE

Concepts:

* Causal inference – a paradigm shift
* The two fundamental laws

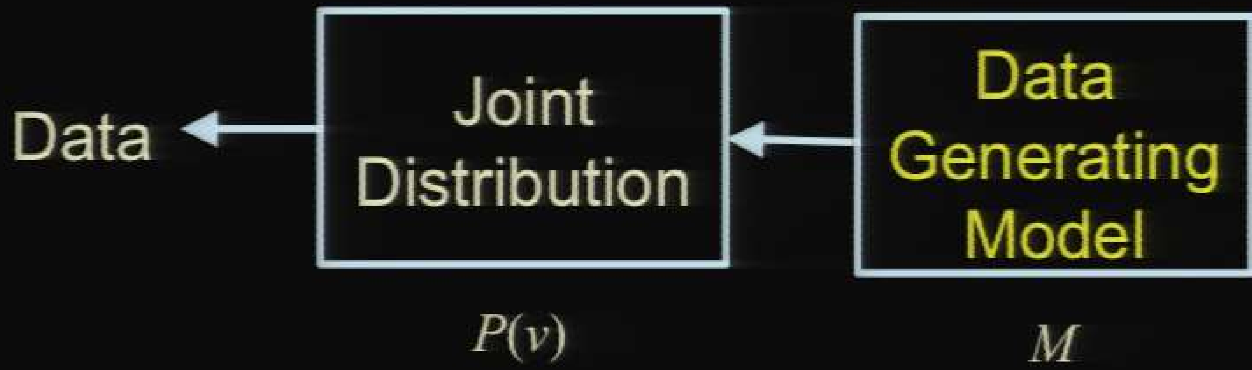
Basic tools:

* Graph separation
* The truncated product formula
* The back-door adjustment formula
* The do-calculus

Capabilities:

* Policy evaluation
* Transportability
* Mediation
* Missing Data

THE EMERGENCE OF THE FIRST LAYER



Theorem (PV, 1991). Every Markovian structural causal model M (recursive, with independent disturbances) induces a passive distribution $P(v_1,\ldots,v_n)$ that can be factorized as

$$P(v_1, v_2, \ldots, v_n) = \prod_i P(v_i \mid pa_i)$$

where pa_i are the (values of) the parents of V_i in the causal diagram associated with M.

OUTLINE

Concepts:

* Causal inference – a paradigm shift
* The two fundamental laws

Basic tools:

* Graph separation
* The truncated product formula
* The back-door adjustment formula
* The do-calculus

Capabilities:

* Policy evaluation
* Transportability
* Mediation
* Missing Data

TOOL 1. GRAPH SEPARATION (D-SEPARATION)

	normal valve	abnormal valve	
	$x \leftarrow z \rightarrow y$	$x \rightarrow z \leftarrow y$	$(X \perp\!\!\!\perp Y)$
$(X \perp\!\!\!\perp Y \mid Z)$	$x \rightarrow z \rightarrow y$	$\begin{array}{c} x \rightarrow z \leftarrow y \\ \downarrow \\ w \end{array}$	$(X \not\perp\!\!\!\perp Y \mid Z)$
	$x \leftarrow z \leftarrow y$		

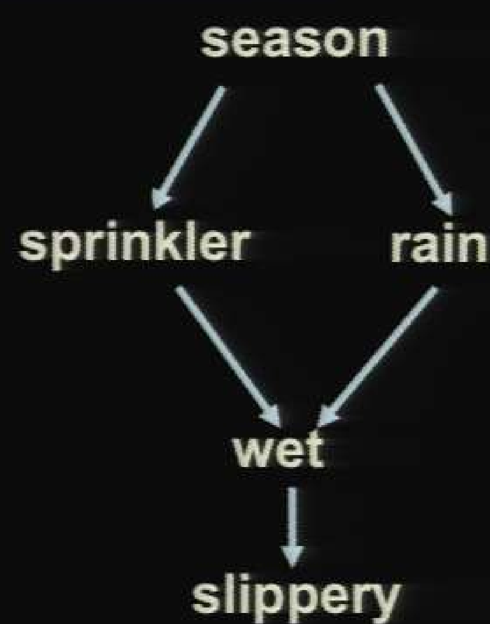
TOOL 1. GRAPH SEPARATION (D-SEPARATION)

	normal valve	abnormal valve	
	$x \leftarrow z \rightarrow y$	$x \rightarrow z \leftarrow y$	$(X \perp\!\!\!\perp Y)$
$(X \perp\!\!\!\perp Y \mid Z)$	$x \rightarrow z \rightarrow y$	$x \rightarrow z \leftarrow y$, $z \rightarrow w$	$(X \not\perp\!\!\!\perp Y \mid Z)$
	$x \leftarrow z \leftarrow y$		

TOOL 1. GRAPH SEPARATION (D-SEPARATION)

	normal valve	abnormal valve	
	$x \leftarrow z \rightarrow y$	$x \rightarrow z \leftarrow y$	$(X \perp\!\!\!\perp Y)$
$(X \perp\!\!\!\perp Y \mid Z)$	$x \rightarrow z \rightarrow y$	$x \rightarrow z \leftarrow y$, $z \rightarrow w$	$(X \not\!\perp\!\!\!\perp Y \mid Z)$
	$x \leftarrow z \leftarrow y$		

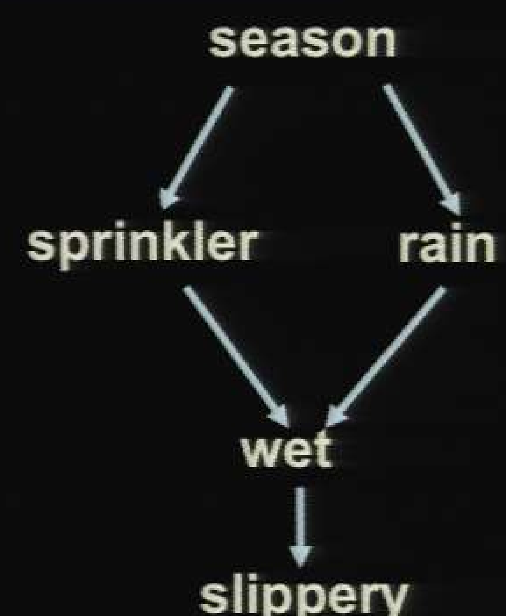
CI_1 : (Wet $\perp\!\!\!\perp$ Sprinkler)



TOOL 1. GRAPH SEPARATION (D-SEPARATION)

	normal valve	abnormal valve	
	$x \leftarrow z \rightarrow y$	$x \rightarrow z \leftarrow y$	$(X \perp\!\!\!\perp Y)$
$(X \perp\!\!\!\perp Y \mid Z)$	$x \rightarrow z \rightarrow y$	$x \rightarrow z \leftarrow y$, $z \rightarrow w$	$(X \not\perp\!\!\!\perp Y \mid Z)$
	$x \leftarrow z \leftarrow y$		

✘ CI_1 : (Wet $\perp\!\!\!\perp$ Sprinkler)



TOOL 1. GRAPH SEPARATION (D-SEPARATION)

	normal valve	abnormal valve	
$(X \perp\!\!\!\perp Y \mid Z)$	$x \leftarrow z \rightarrow y$	$x \rightarrow z \leftarrow y$	$(X \perp\!\!\!\perp Y)$
	$x \rightarrow z \rightarrow y$	$\begin{array}{c} x \rightarrow z \leftarrow y \\ \downarrow \\ w \end{array}$	$(X \not\perp\!\!\!\perp Y \mid Z)$
	$x \leftarrow z \leftarrow y$		

✘ CI_1 : (Wet $\perp\!\!\!\perp$ Sprinkler)

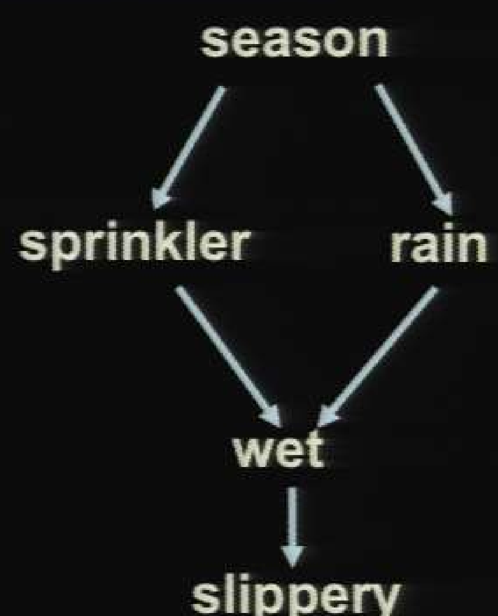
CI_2 : (Wet $\perp\!\!\!\perp$ Season | Sprinkler)

TOOL 1. GRAPH SEPARATION (D-SEPARATION)

	normal valve	abnormal valve	
	$x \leftarrow z \rightarrow y$	$x \rightarrow z \leftarrow y$	$(X \perp\!\!\!\perp Y)$
$(X \perp\!\!\!\perp Y \mid Z)$	$x \rightarrow z \rightarrow y$	$x \rightarrow z \leftarrow y$, $z \rightarrow w$	$(X \not\!\perp\!\!\!\perp Y \mid Z)$
	$x \leftarrow z \leftarrow y$		

✘ CI_1 : (Wet $\perp\!\!\!\perp$ Sprinkler)

✘ CI_2 : (Wet $\perp\!\!\!\perp$ Season | Sprinkler)



TOOL 1. GRAPH SEPARATION (D-SEPARATION)

	normal valve	abnormal valve	
	$x \leftarrow z \rightarrow y$	$x \rightarrow z \leftarrow y$	$(X \perp\!\!\!\perp Y)$
$(X \perp\!\!\!\perp Y \mid Z)$	$x \rightarrow z \rightarrow y$	$x \rightarrow z \leftarrow y$, $z \rightarrow w$	$(X \not\!\perp\!\!\!\perp Y \mid Z)$
	$x \leftarrow z \leftarrow y$		

✘ CI_1 : (Wet $\perp\!\!\!\perp$ Sprinkler)

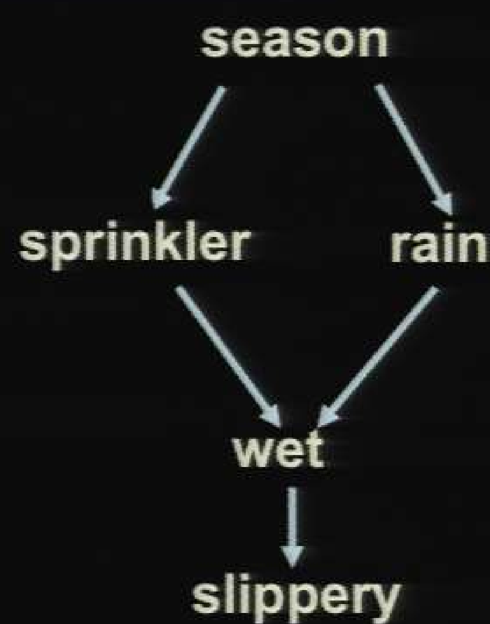
✘ CI_2 : (Wet $\perp\!\!\!\perp$ Season | Sprinkler)

CI_3 : (Rain $\perp\!\!\!\perp$ Slippery | Wet)

TOOL 1. GRAPH SEPARATION (D-SEPARATION)

	normal valve	abnormal valve	
	$x \leftarrow z \rightarrow y$	$x \rightarrow z \leftarrow y$	$(X \perp\!\!\!\perp Y)$
$(X \perp\!\!\!\perp Y \mid Z)$	$x \rightarrow z \rightarrow y$	$x \rightarrow z \leftarrow y$, $z \rightarrow w$	$(X \not\!\perp\!\!\!\perp Y \mid Z)$
	$x \leftarrow z \leftarrow y$		

- ✘ CI_1 : (Wet $\perp\!\!\!\perp$ Sprinkler)
- ✘ CI_2 : (Wet $\perp\!\!\!\perp$ Season | Sprinkler)
- ✔ CI_3 : (Rain $\perp\!\!\!\perp$ Slippery | Wet)
- ✔ CI_4 : (Season $\perp\!\!\!\perp$ Wet | Sprinkler, Rain)



TOOL 1. GRAPH SEPARATION (D-SEPARATION)

	normal valve	abnormal valve	
$(X \perp\!\!\!\perp Y \mid Z)$	$x \leftarrow z \rightarrow y$	$x \rightarrow z \leftarrow y$	$(X \perp\!\!\!\perp Y)$
	$x \rightarrow z \rightarrow y$	$x \rightarrow z \leftarrow y$, $z \rightarrow w$	$(X \not\perp\!\!\!\perp Y \mid Z)$
	$x \leftarrow z \leftarrow y$		

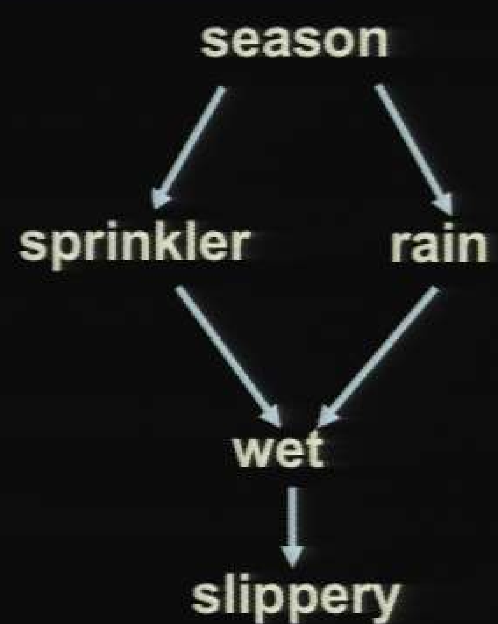
✘ CI_1 : (Wet $\perp\!\!\!\perp$ Sprinkler)

✘ CI_2 : (Wet $\perp\!\!\!\perp$ Season | Sprinkler)

✔ CI_3 : (Rain $\perp\!\!\!\perp$ Slippery | Wet)

✔ CI_4 : (Season $\perp\!\!\!\perp$ Wet | Sprinkler, Rain)

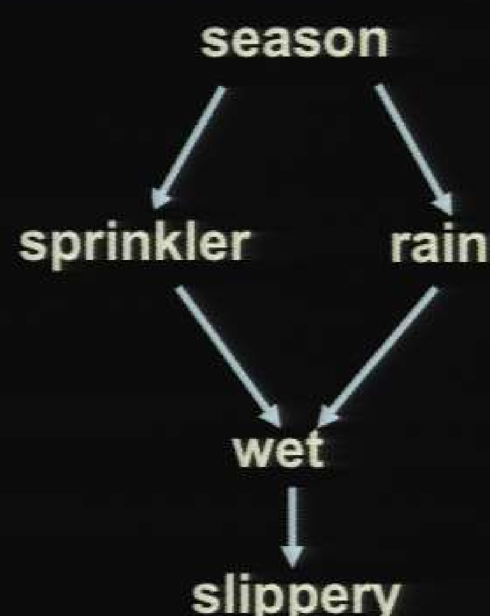
CI_5 : (Sprinkler $\perp\!\!\!\perp$ Rain | Season, Wet)



TOOL 1. GRAPH SEPARATION (D-SEPARATION)

	normal valve	abnormal valve	
$(X \perp\!\!\!\perp Y \mid Z)$	$x \leftarrow z \rightarrow y$	$x \rightarrow z \leftarrow y$	$(X \perp\!\!\!\perp Y)$
	$x \rightarrow z \rightarrow y$	$x \rightarrow z \leftarrow y$, $z \rightarrow w$	$(X \not\!\perp\!\!\!\perp Y \mid Z)$
	$x \leftarrow z \leftarrow y$		

- ✘ CI_1 : (Wet $\perp\!\!\!\perp$ Sprinkler)
- ✘ CI_2 : (Wet $\perp\!\!\!\perp$ Season | Sprinkler)
- ✔ CI_3 : (Rain $\perp\!\!\!\perp$ Slippery | Wet)
- ✔ CI_4 : (Season $\perp\!\!\!\perp$ Wet | Sprinkler, Rain)
- ✘ CI_5 : (Sprinkler $\perp\!\!\!\perp$ Rain | Season, Wet)

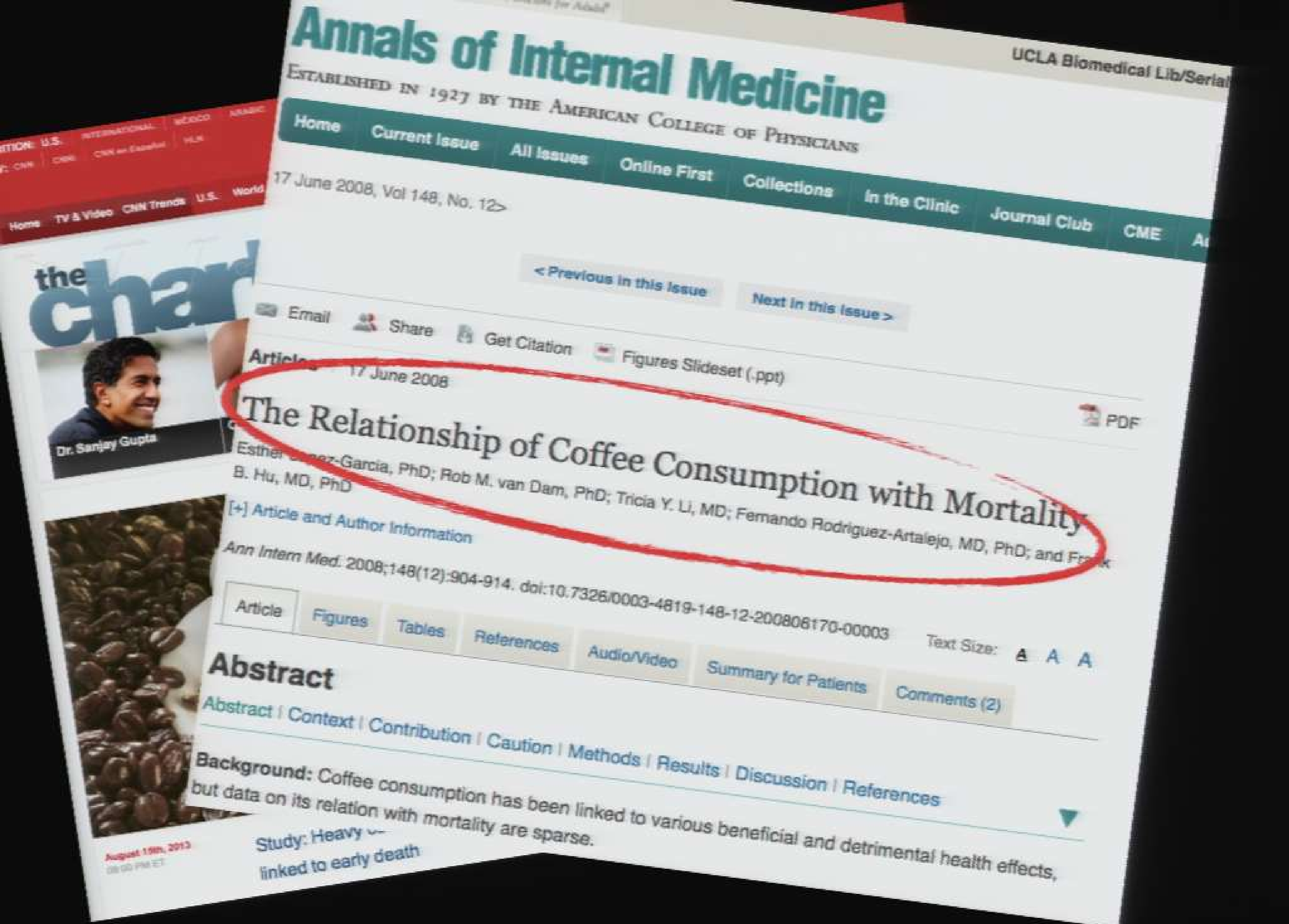


THE SECOND LAYER ON CAUSAL HIERARCHY:

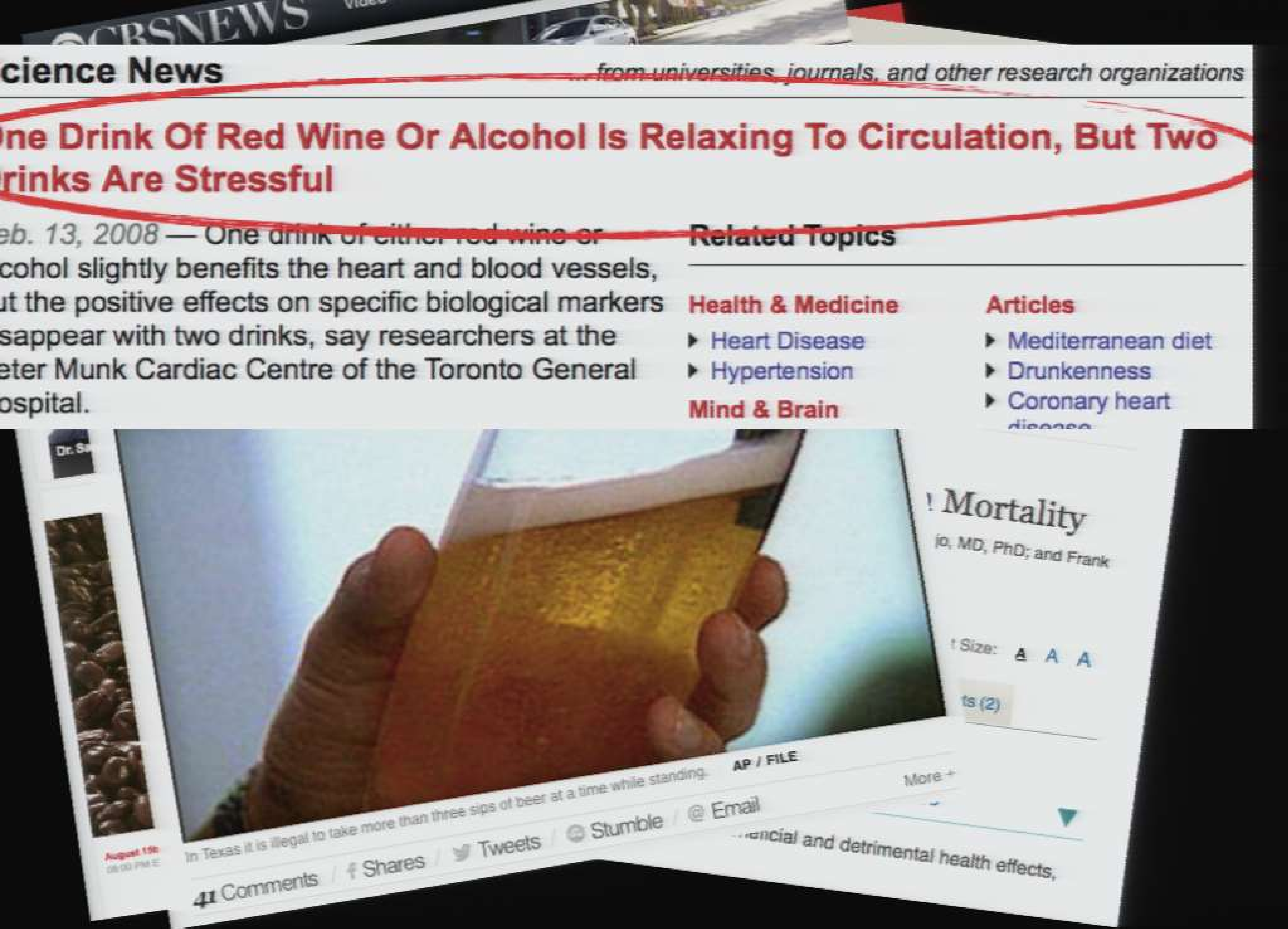
CAUSAL EFFECTS

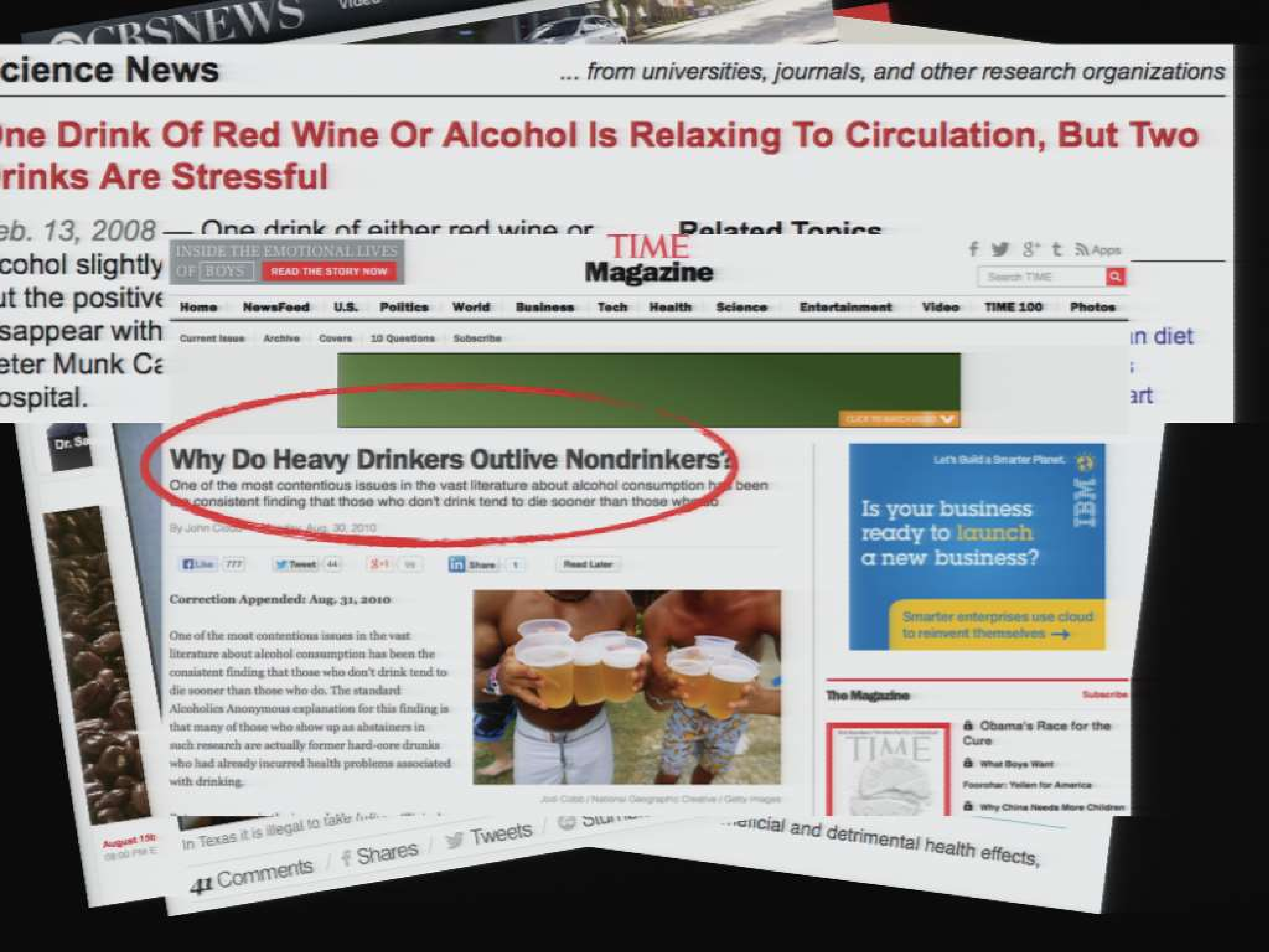
(What if I do $X=x$?)

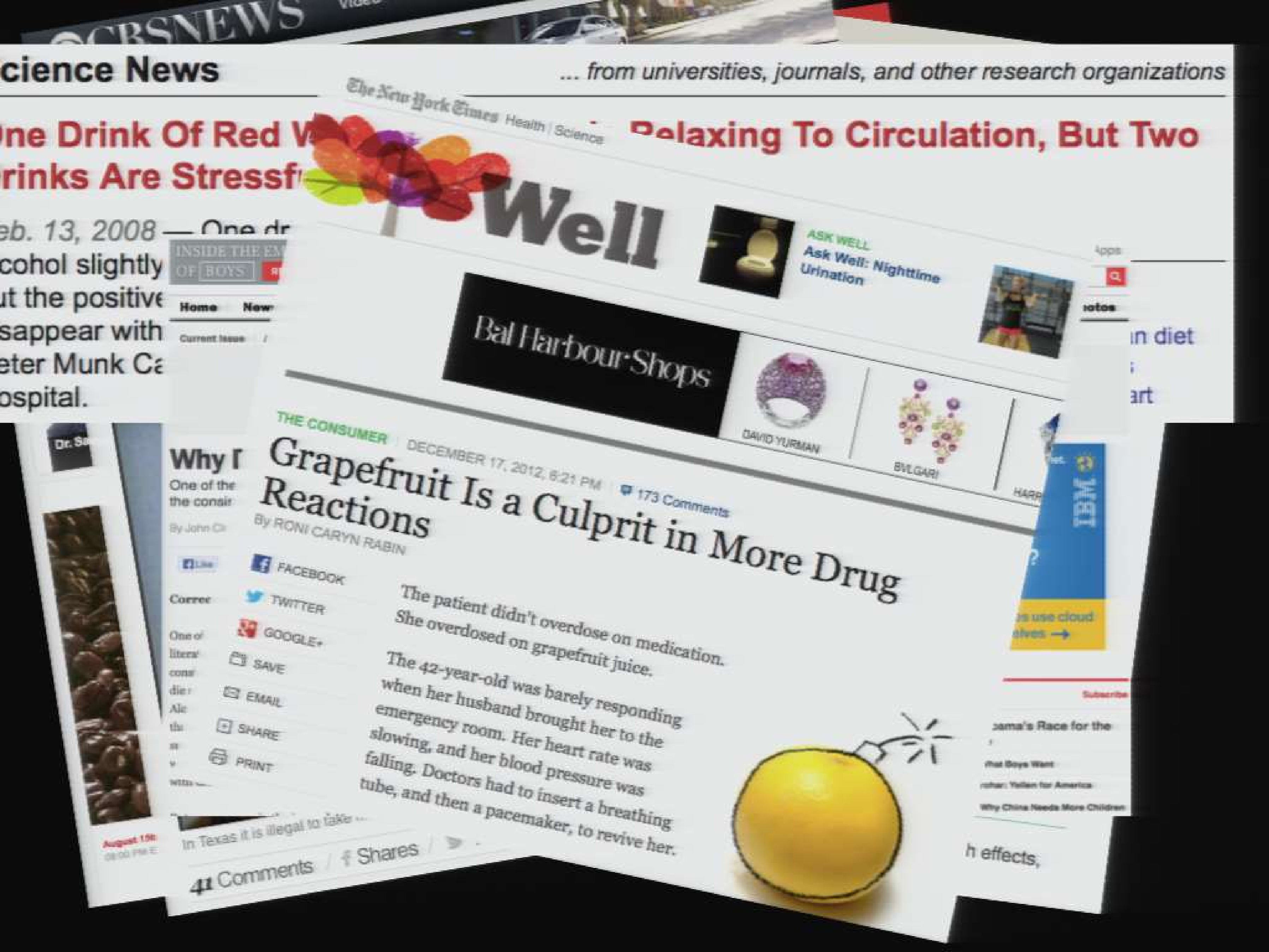
CNN Health
SEARCH
Home
TV & Video
CNN Trends
U.S.
World
Politics
Justice
Entertainment
Tech
Health
Living
Travel
Opinion
iReport
Money
Sports
the chart
Dr. Sanjay Gupta
Children's Health
Expert Doctor Q&A
Sleep
Sex and You
2.3k
Recommend
102
Tweet
31
Share
20
More sharing
voicefive
Are you aware of any of the following?
UnitedHealthcare
Humana
Aetna
Blue Cross
1 of 5
Dr. Sanjay Gupta
August 15th, 2013
Study: Heavy coffee drinking in people under 55 linked to early death


Annals of Internal Medicine
Established in 1927 by the American College of Physicians
UCLA Biomedical Lib/Serial
Home
Current Issue
All Issues
Online First
Collections
In the Clinic
Journal Club
CME
17 June 2008, Vol 148, No. 12>
< Previous in this Issue
Next in this Issue >
Email
Share
Get Citation
Figures Slideset (.ppt)
17 June 2008
PDF
The Relationship of Coffee Consumption with Mortality
-Garcia, PhD; Rob M. van Dam, PhD; Tricia Y. Li, MD; Fernando Rodriguez-Artalejo, MD, PhD; and Fr
B. Hu, MD, PhD
[+] Article and Author Information
Ann Intern Med. 2008;148(12):904-914. doi:10.7326/0003-4819-148-12-200806170-00003
Article
Figures
Tables
References
Audio/Video
Summary for Patients
Comments (2)
Text Size:
Abstract
Abstract | Context | Contribution | Caution | Methods | Results | Discussion | References
Background: Coffee consumption has been linked to various beneficial and detrimental health effects, but data on its relation with mortality are sparse.
Home
TV & Video
CNN Trends
U.S.
World
Dr. Sanjay Gupta
Study: Heavy
linked to early death

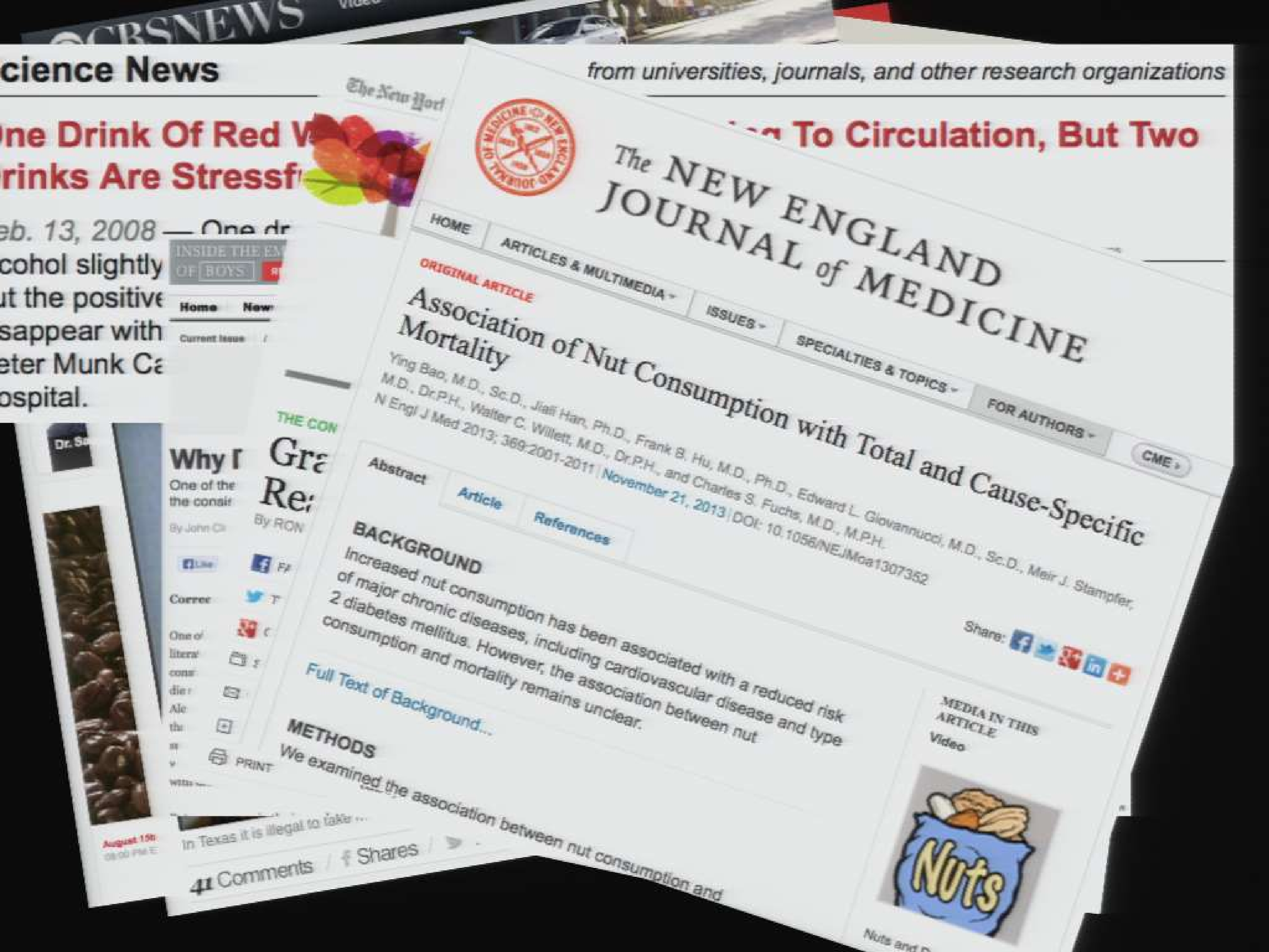

CBSNEWS
RELAX AND
ENJOY THE LUXURY.
February 15, 2013, 3: 36 PM
By MICHELLE CASTILLO / CBS NEWS
Alcohol causes 20,000 cancer deaths in the U.S. annually
In Texas it is illegal to take more than three sips of beer at a time while standing.
AP / FILE
More +
41 Comments / Shares / Tweets / Stumble / Email
Home
UCLA Biomedical Lib/Serials
the Clinic
Journal Club
CME
PDF
Mortality
MD, PhD; and Frank
and detrimental health effects,


CBSNEWS
cience News
...from universities, journals, and other research organizations
ne Drink Of Red Wine Or Alcohol Is Relaxing To Circulation, But Two
rinks Are Stressful
eb. 13, 2008 — One drink of either red wine or
cohol slightly benefits the heart and blood vessels,
ut the positive effects on specific biological markers
sappear with two drinks, say researchers at the
eter Munk Cardiac Centre of the Toronto General
ospital.
Related Topics
Health & Medicine
Heart Disease
Hypertension
Mind & Brain
Articles
Mediterranean diet
Drunkenness
Coronary heart
Mortality
io, MD, PhD; and Frank
Size: A A A
ts (2)
In Texas it is illegal to take more than three sips of beer at a time while standing.
AP / FILE
41 Comments
Shares
Tweets
Stumble
Email
More +
ancial and detrimental health effects,

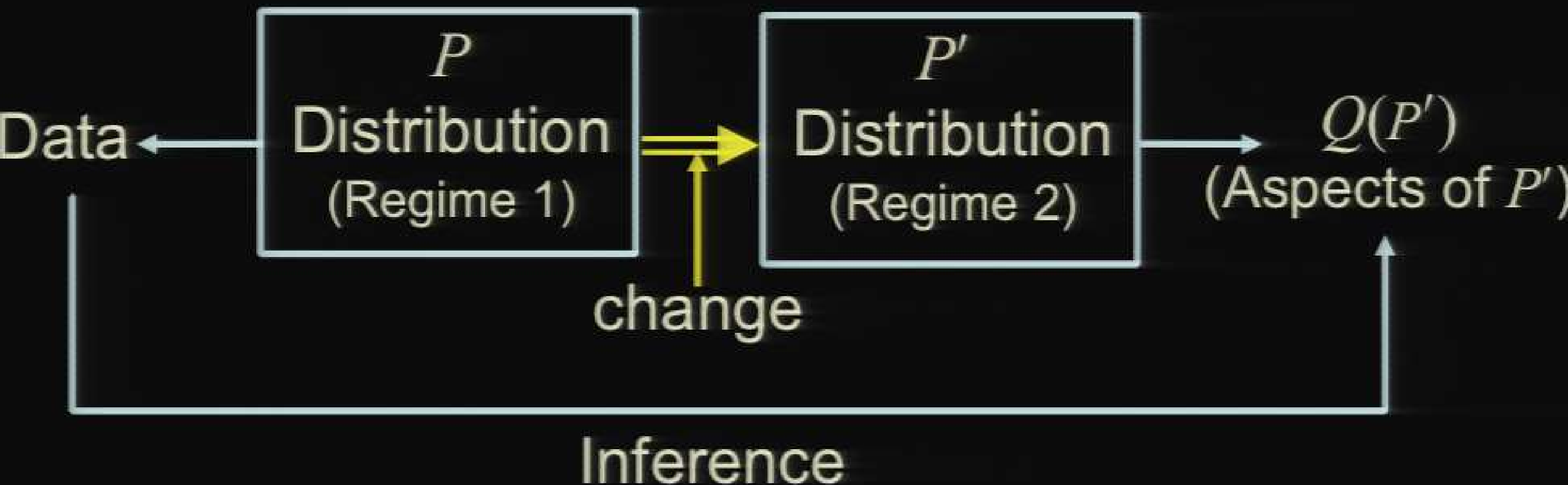

CBSNEWS
cience News
... from universities, journals, and other research organizations
ne Drink Of Red Wine Or Alcohol Is Relaxing To Circulation, But Two
rinks Are Stressful
eb. 13, 2008 — One drink of either red wine or
cohol slightly
it the positive
sappear with
eter Munk Ca
ospital.
Related Topics
n diet
art
INSIDE THE EMOTIONAL LIVES OF BOYS
READ THE STORY NOW
TIME Magazine
Home NewsFeed U.S. Politics World Business Tech Health Science Entertainment Video TIME 100 Photos
Current Issue Archive Covers 10 Questions Subscribe
Why Do Heavy Drinkers Outlive Nondrinkers?
One of the most contentious issues in the vast literature about alcohol consumption has been the consistent finding that those who don't drink tend to die sooner than those who do
Correction Appended: Aug. 31, 2010
One of the most contentious issues in the vast literature about alcohol consumption has been the consistent finding that those who don't drink tend to die sooner than those who do. The standard Alcoholics Anonymous explanation for this finding is that many of those who show up as abstainers in such research are actually former hard-core drunks who had already incurred health problems associated with drinking.
Is your business ready to launch a new business?
IBM
Smarter enterprises use cloud to reinvent themselves
The Magazine
Subscribe
Obama's Race for the Cure
TIME
In Texas it is illegal to take
41 Comments
Shares
Tweets
eficial and detrimental health effects,



CBSNEWS
cience News
... from universities, journals, and other research organizations
ne Drink Of Red W
elaxing To Circulation, But Two
rinks Are Stressf
eb. 13, 2008 — One dr
cohol slightly
ut the positive
sappear with
eter Munk Ca
ospital.
The New York Times Health | Science
Well
ASK WELL
Ask Well: Nighttime Urination
Bal Harbour Shops
DAVID YURMAN
BVLGARI
INSIDE THE EM
OF BOYS
Home
Current Issue
Why I
One of the
the consir
Correc
THE CONSUMER
DECEMBER 17, 2012, 6:21 PM
173 Comments
Grapefruit Is a Culprit in More Drug Reactions
By RONI CARYN RABIN
FACEBOOK
TWITTER
GOOGLE+
SAVE
EMAIL
SHARE
PRINT
The patient didn't overdose on medication. She overdosed on grapefruit juice.
The 42-year-old was barely responding when her husband brought her to the emergency room. Her heart rate was slowing, and her blood pressure was falling. Doctors had to insert a breathing tube, and then a pacemaker, to revive her.
In Texas it is illegal to take
41 Comments
Shares
n diet
art
IBM
h effects,
Subscribe


CBSNEWS
cience News
from universities, journals, and other research organizations
One Drink Of Red W
To Circulation, But Two
rinks Are Stressf
eb. 13, 2008 — One dr
cohol slightly
ut the positive
sappear with
eter Munk Ca
ospital.
The NEW ENGLAND JOURNAL of MEDICINE
HOME
ARTICLES & MULTIMEDIA
ISSUES
SPECIALTIES & TOPICS
FOR AUTHORS
CME
ORIGINAL ARTICLE
Association of Nut Consumption with Total and Cause-Specific Mortality
Ying Bao, M.D., Sc.D., Jiali Han, Ph.D., Frank B. Hu, M.D., Ph.D., Edward L. Giovannucci, M.D., Sc.D., Meir J. Stampfer, M.D., Dr.P.H., Walter C. Willett, M.D., Dr.P.H., and Charles S. Fuchs, M.D., M.P.H.
N Engl J Med 2013; 369:2001-2011 | November 21, 2013 | DOI: 10.1056/NEJMoa1307352
Abstract
Article
References
Share:
BACKGROUND
Increased nut consumption has been associated with a reduced risk of major chronic diseases, including cardiovascular disease and type 2 diabetes mellitus. However, the association between nut consumption and mortality remains unclear.
Full Text of Background...
METHODS
We examined the association between nut consumption and
MEDIA IN THIS ARTICLE
Video
Nuts
Why I
Gra
Re
By RON
THE CON
Home
Current Issue
PRINT
In Texas it is illegal to take ...
41 Comments
Shares

CAUSAL INFERENCE: MOVING BETWEEN REGIMES



- What happens when ***P*** changes?

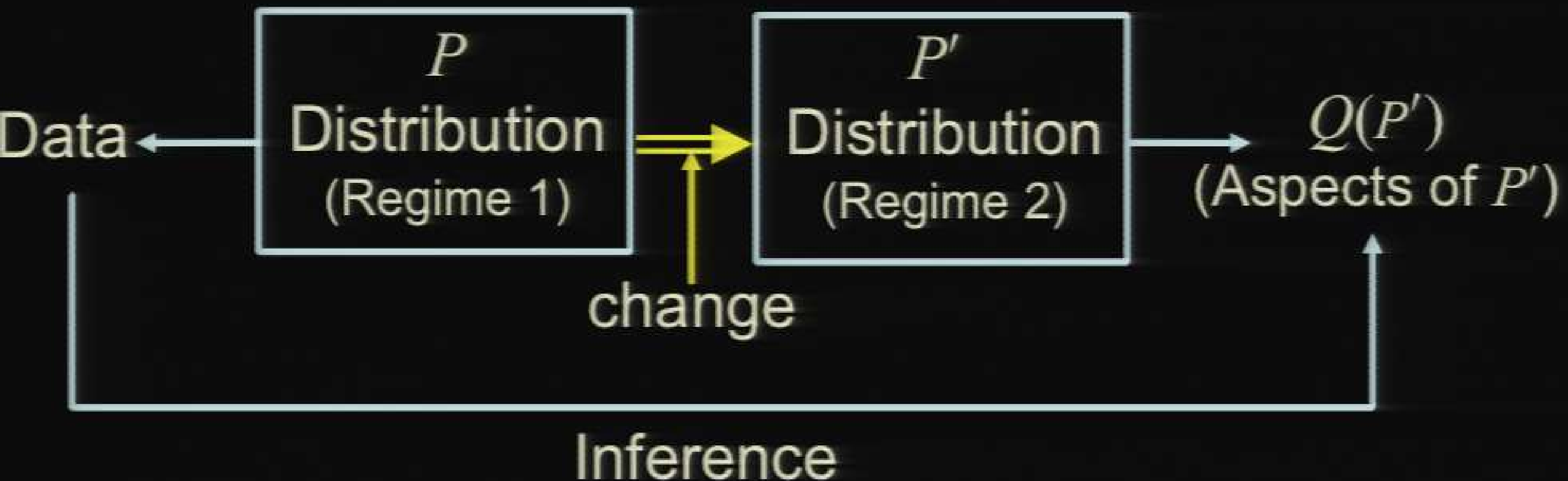
e.g., Infer whether less people would get cancer if we ban smoking.

$Q = P(Cancer = true \mid do(Smoking = no))$ Not an aspect of *P*.

Observation 1:

The distribution alone tells us nothing about change; it just describes static conditions of a population (under a specific regime).

CAUSAL INFERENCE: MOVING BETWEEN REGIMES



- What happens when ***P*** changes?

e.g., Infer whether less people would get cancer if we ban smoking.

$Q = P(Cancer = true \mid do(Smoking = no))$ Not an aspect of P.

Observation 1:

The distribution alone tells us nothing about change; it just describes static conditions of a population (under a specific regime).

THE BIG PICTURE: THE CHALLENGE OF CAUSAL INFERENCE

Real world — Alternative world

Z: age, sex
X: action
W: mediator
Y: outcome

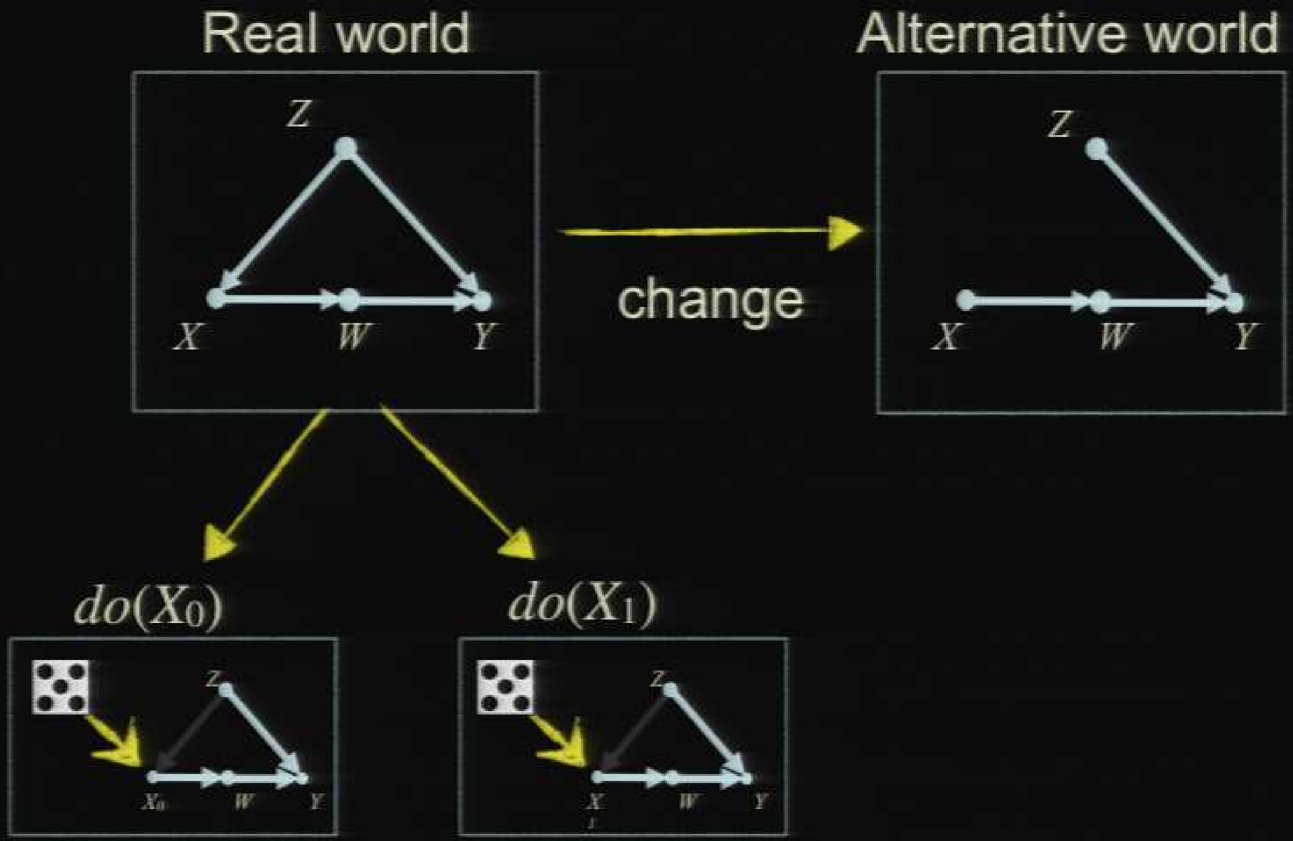
$P(z, x, w, y)$ — $P(y \mid do(x))$

- Goal: how much Y changes with X if we vary X between two different constants free from the influence of Z.
- This is the definition of causal effect.

METHOD FOR COMPUTING CAUSAL EFECTS: RANDOMIZED EXPERIMENTS

Z: age, sex
X: action
W: mediator
Y: outcome

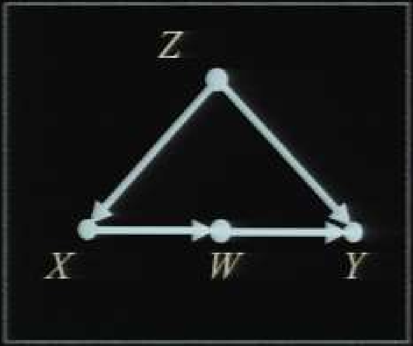
METHOD FOR COMPUTING CAUSAL EFECTS: RANDOMIZED EXPERIMENTS



Z: age, sex
X: action
W: mediator
Y: outcome

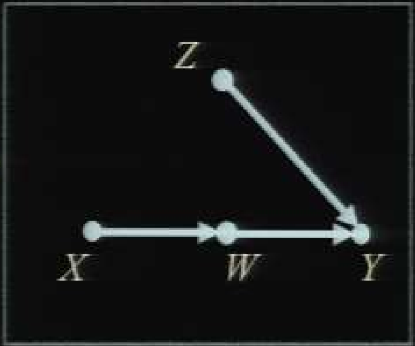
PROBLEM 1. COMPUTING EFFECTS FROM OBSERVATIONAL DATA

Real world



change

Alternative world



Z: age, sex
X: action
W: mediator
Y: outcome

$$P(z, x, w, y) \xrightarrow{?} P(y \mid do(x))$$

PROBLEM 1. COMPUTING EFFECTS FROM OBSERVATIONAL DATA

Real world | Alternative world

Z: age, sex
X: action
W: mediator
Y: outcome

$$P(z, x, w, y) \xrightarrow{?} P(y \mid do(x))$$

Questions:

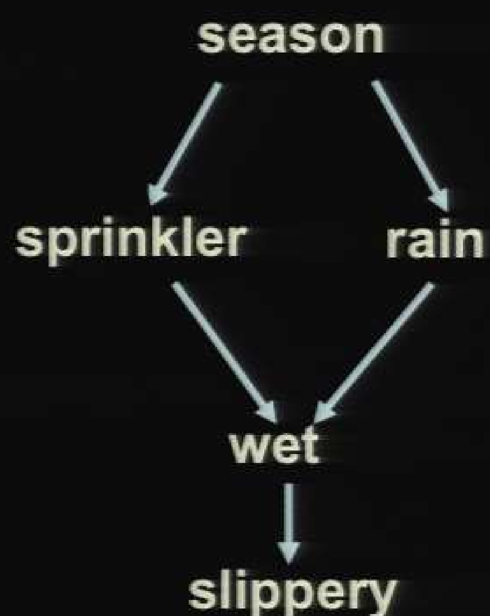
* What is the relationship between $P(z, x, w, y)$ and $P(y \mid do(x))$?
* Is $P(y \mid \mathrm{d}o(x)) = P(y \mid x)$?

COMPUTING CAUSAL EFFECTS FROM OBSERVATIONAL DATA

Queries:

Q_1 = Pr(wet | Sprinkler = on)

Q_2 = Pr(wet | do(Sprinkler = on))

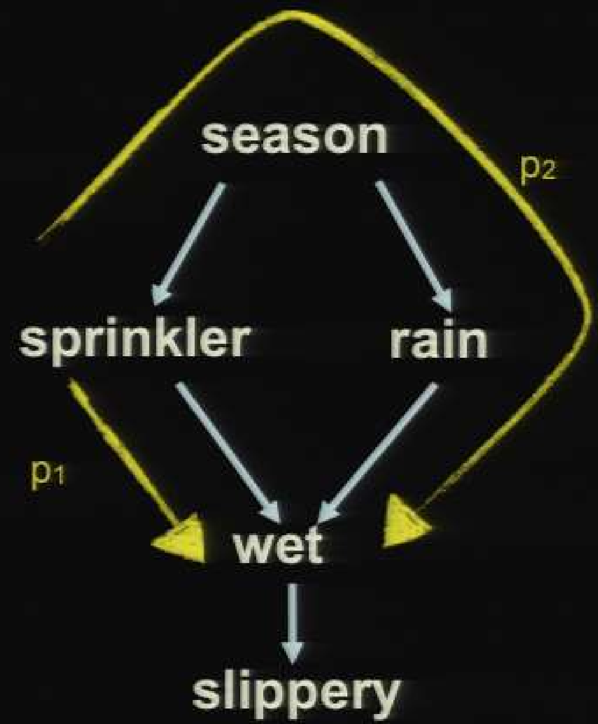


COMPUTING CAUSAL EFFECTS FROM OBSERVATIONAL DATA

Queries:

$Q_1 = \Pr(\text{wet} \mid \text{Sprinkler} = \text{on})$
$\quad = P(p_1) + P(p_2)$

$Q_2 = \Pr(\text{wet} \mid do(\text{Sprinkler} = \text{on}))$



COMPUTING CAUSAL EFFECTS FROM OBSERVATIONAL DATA

Queries:

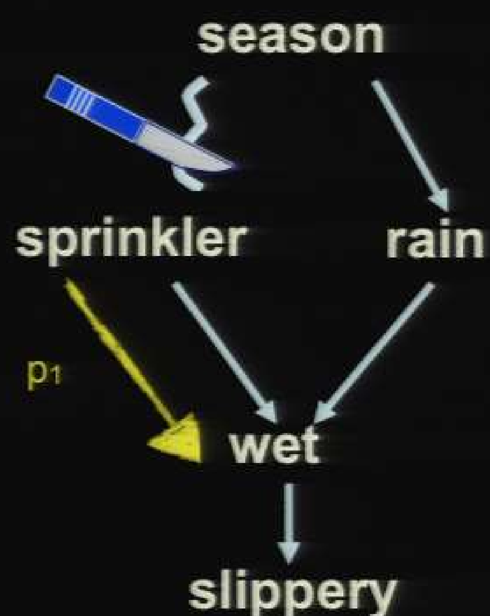
$Q_1 = \Pr(\text{wet} \mid \text{Sprinkler} = \text{on})$
$\quad = P(p_1) + P(p_2)$

$Q_2 = \Pr(\text{wet} \mid do(\text{Sprinkler} = \text{on}))$

COMPUTING CAUSAL EFFECTS FROM OBSERVATIONAL DATA

Queries:

$Q_1 = \Pr(wet \mid Sprinkler = on)$
$= P(p_1) + P(p_2)$



$Q_2 = \Pr(wet \mid do(Sprinkler = on))$
$= P(p_1)$

$\sum_{Se,Ra,Sl} P(Se)\, P(Sp \mid Se)\, P(Ra \mid Se)\, P(We \mid Sp, Ra)\, P(Sl \mid We)$

OUTLINE

Concepts:

* Causal inference – a paradigm shift
* The two fundamental laws

Basic tools:

* Graph separation
* The truncated product formula
* The back-door adjustment formula
* The do-calculus

Capabilities:

* Policy evaluation
* Transportability
* Mediation
* Missing Data

TOOL 2. TRUNCATED FACTORIZATION PRODUCT (OPERATIONALIZING INTERVENTIONS)

Corollary (Truncated Factorization, Manipulation Thm., G-comp.):
The distribution generated by an intervention $do(X{=}x)$ (in a Markovian model M) is given by the truncated factorization:

$$P(v_1, v_2, \ldots, v_n \mid do(x)) = \prod_{i \mid V_i \notin X} P(v_i \mid pa_i) \Bigg|_{X=x}$$

NO FREE LUNCH: ASSUMPTIONS ENCODED IN CBNs

Definition (Causal Bayesian Network):

$P(v)$: observational distribution
$P(v \mid do(x))$: experimental distribution
P^*: set of all observational and experimental distributions

A DAG G is called a Causal Bayesian Network compatible with P^* if and only if the following three conditions hold for every $P(v \mid do(x)) \in P^*$:

i. $P(v \mid do(x))$ is Markov relative to G;
ii. $P(v_i \mid do(x)) = 1$, for all $V_i \in X$;
iii. $P(v_i \mid \mathrm{pa}_i, do(x)) = P(v_i \mid pa_i)$, for all $V_i \notin X$.

TOOL 2. TRUNCATED FACTORIZATION PRODUCT (OPERATIONALIZING INTERVENTIONS)

Corollary (Truncated Factorization, Manipulation Thm., G-comp.):
The distribution generated by an intervention $do(X=x)$ (in a Markovian model M) is given by the truncated factorization:

$$P(v_1, v_2, \ldots, v_n \mid do(x)) = \prod_{i \mid V_i \notin X} P(v_i \mid pa_i) \Bigg|_{X=x}$$

NO FREE LUNCH: ASSUMPTIONS ENCODED IN CBNs

Definition (Causal Bayesian Network):

$P(v)$: observational distribution
$P(v \mid do(x))$: experimental distribution
P^*: set of all observational and experimental distributions

A DAG G is called a Causal Bayesian Network compatible with P* if and only if the following three conditions hold for every $P(v \mid do(x)) \in P^*$:

i. $P(v \mid do(x))$ is Markov relative to G;
ii. $P(v_i \mid do(x)) = 1$, for all $V_i \in X$;
iii. $P(v_i \mid \mathrm{pa}_i, do(x)) = P(v_i \mid pa_i)$, for all $V_i \notin X$.

TOOL 2. TRUNCATED FACTORIZATION PRODUCT (OPERATIONALIZING INTERVENTIONS)

Corollary (Truncated Factorization, Manipulation Thm., G-comp.):
The distribution generated by an intervention $do(X=x)$ (in a Markovian model M) is given by the truncated factorization:

$$P(v_1, v_2, \ldots, v_n \mid do(x)) = \prod_{i \mid V_i \notin X} P(v_i \mid pa_i) \Big|_{X=x}$$

NO FREE LUNCH: ASSUMPTIONS ENCODED IN CBNs

Definition (Causal Bayesian Network):

$P(v)$: observational distribution
$P(v \mid do(x))$: experimental distribution
P^*: set of all observational and experimental distributions

A DAG G is called a Causal Bayesian Network compatible with P^* if and only if the following three conditions hold for every $P(v \mid do(x)) \in P^*$:

i. $P(v \mid do(x))$ is Markov relative to G;
ii. $P(v_i \mid do(x)) = 1$, for all $V_i \in X$;
iii. $P(v_i \mid \mathrm{pa}_i, do(x)) = P(v_i \mid pa_i)$, for all $V_i \notin X$.

TOOL 2. TRUNCATED FACTORIZATION PRODUCT (OPERATIONALIZING INTERVENTIONS)

Corollary (Truncated Factorization, Manipulation Thm., G-comp.):
The distribution generated by an intervention $do(X=x)$ (in a Markovian model M) is given by the truncated factorization:

$$P(v_1, v_2, \ldots, v_n \mid do(x)) = \prod_{i \mid V_i \notin X} P(v_i \mid pa_i) \Bigg|_{X=x}$$

NO FREE LUNCH: ASSUMPTIONS ENCODED IN CBNs

Definition (Causal Bayesian Network):

$P(v)$: observational distribution
$P(v \mid do(x))$: experimental distribution
P^*: set of all observational and experimental distributions

A DAG G is called a Causal Bayesian Network compatible with P^* if and only if the following three conditions hold for every $P(v \mid do(x)) \in P^*$:

i. $P(v \mid do(x))$ is Markov relative to G;
ii. $P(v_i \mid do(x)) = 1$, for all $V_i \in X$;
iii. $P(v_i \mid \mathrm{pa}_i, do(x)) = P(v_i \mid pa_i)$, for all $V_i \notin X$.

OUTLINE

Concepts:

* Causal inference – a paradigm shift
* The two fundamental laws

Basic tools:

* Graph separation
* The truncated product formula
* The back-door adjustment formula
* The do-calculus

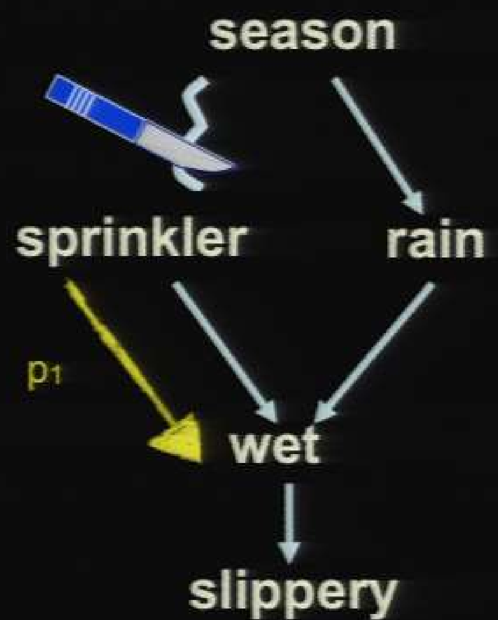
Capabilities:

* Policy evaluation
* Transportability
* Mediation
* Missing Data

IF SEASON IS LATENT, IS THE EFFECT STILL COMPUTABLE?

Queries:

$Q_1 = \Pr(\text{wet} \mid \text{Sprinkler} = \text{on})$
$\quad = P(p_1) + P(p_2)$



$Q_2 = \Pr(\text{wet} \mid do(\text{Sprinkler} = \text{on}))$
$\quad = P(p_1)$

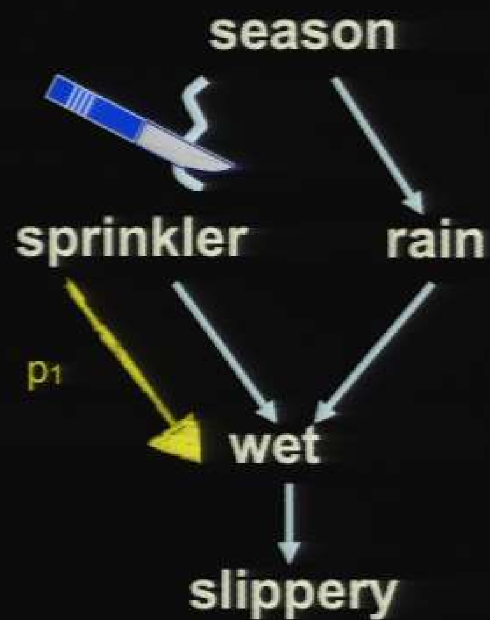
$$\sum_{Se,Ra,Sl} P(Se)\, P(\cancel{Sp \mid Se})\, P(Ra \mid Se)\, P(We \mid Sp, Ra)\, P(Sl \mid We)$$

$$= \sum_{Se} P(We \mid Sp, Se)\, P(Se)$$

IF SEASON IS LATENT, IS THE EFFECT STILL COMPUTABLE?

Queries:

$Q_1 = \Pr(\text{wet} \mid \text{Sprinkler} = \text{on})$
$= P(p_1) + P(p_2)$



$Q_2 = \Pr(\text{wet} \mid do(\text{Sprinkler} = \text{on}))$
$= P(p_1)$

$$\sum_{Se,Ra,Sl} P(Se)\, \cancel{P(Sp \mid Se)}\, P(Ra \mid Se)\, P(We \mid Sp, Ra)\, P(Sl \mid We)$$

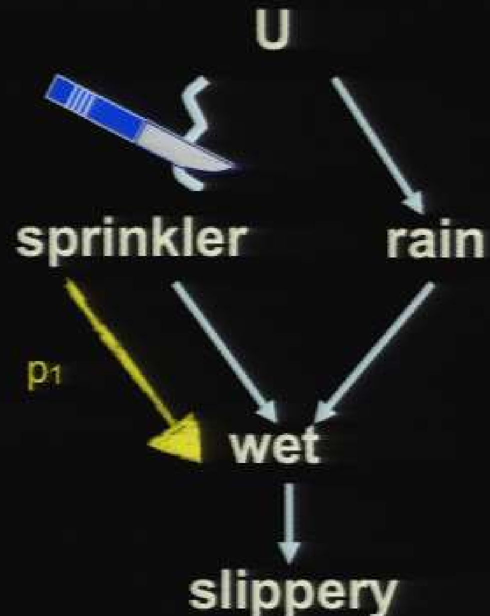
$$= \sum_{Se} P(We \mid Sp, Se)\, P(Se)$$

Adjustment for direct causes

IF SEASON IS LATENT, IS THE EFFECT STILL COMPUTABLE?

Queries:

$Q_1 = \Pr(\text{wet} \mid \text{Sprinkler} = \text{on})$
$\quad = P(p_1) + P(p_2)$



$Q_2 = \Pr(\text{wet} \mid do(\text{Sprinkler} = \text{on}))$
$\quad = P(p_1)$

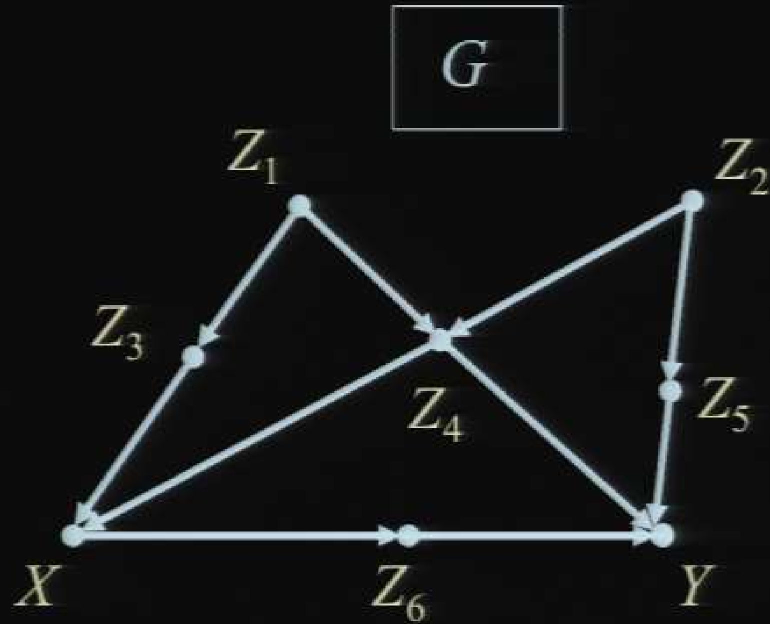
$\sum_{Se,Ra,Sl} P(Se)\, \cancel{P(Sp \mid Se)}\, P(Ra \mid Se)\, P(We \mid Sp, Ra)\, P(Sl \mid We)$

$= \sum_{Se} P(We \mid Sp, Se)\, P(Se)$

Adjustment for direct causes

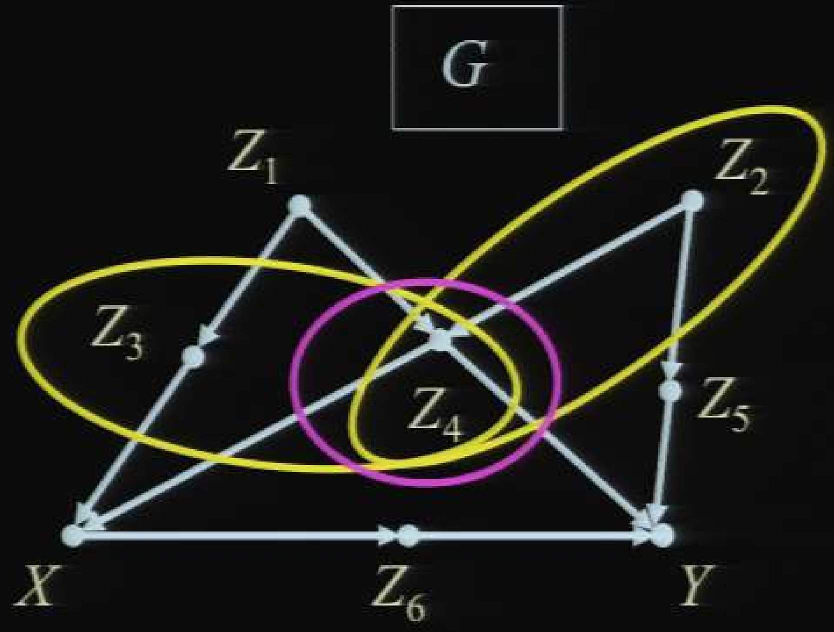
TOOL 3. BACK-DOOR CRITERION (THE PROBLEM OF CONFOUNDING)

Goal: Find the effect of X on Y, $P(y|do(x))$, given measurements on auxiliary variables $Z_1,\ldots, Z_k$



TOOL 3. BACK-DOOR CRITERION (THE PROBLEM OF CONFOUNDING)

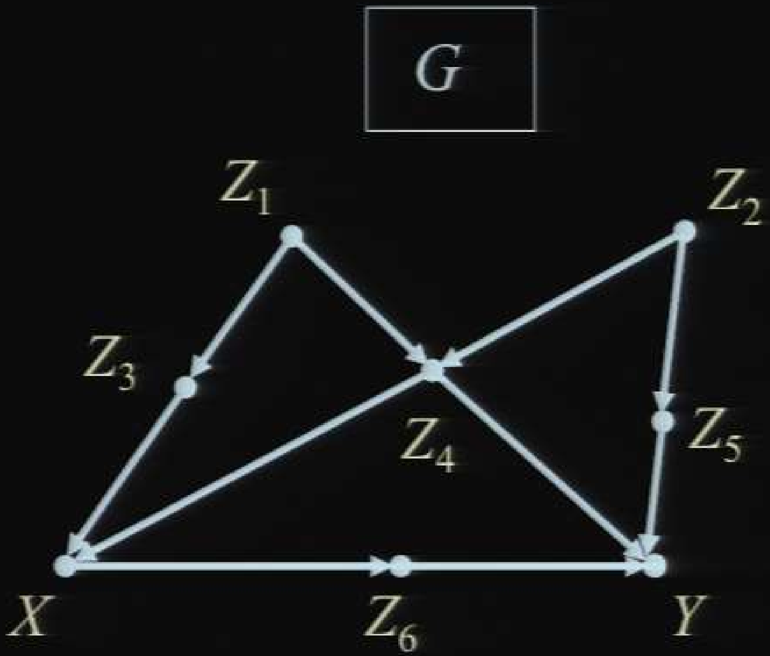
Goal: Find the effect of X on Y, $P(y|do(x))$, given measurements on auxiliary variables $Z_1,..., Z_k$

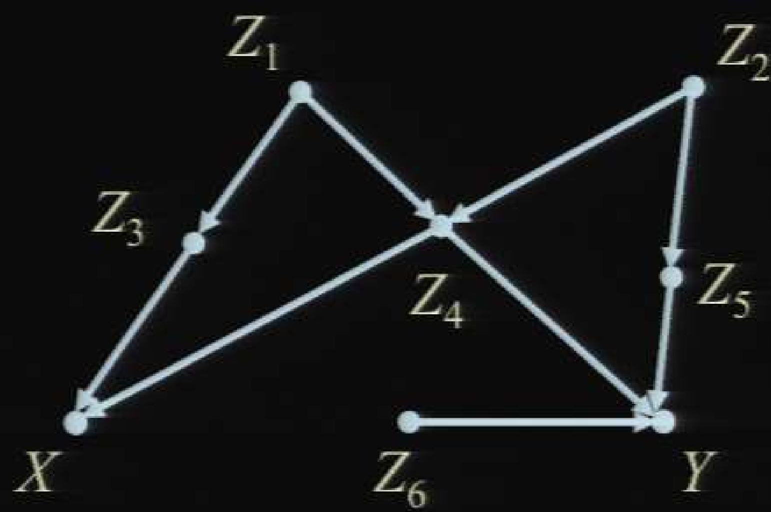


ELIMINATING CONFOUNDING BIAS THE BACK-DOOR CRITERION

$P(y \mid do(x))$ is estimable if

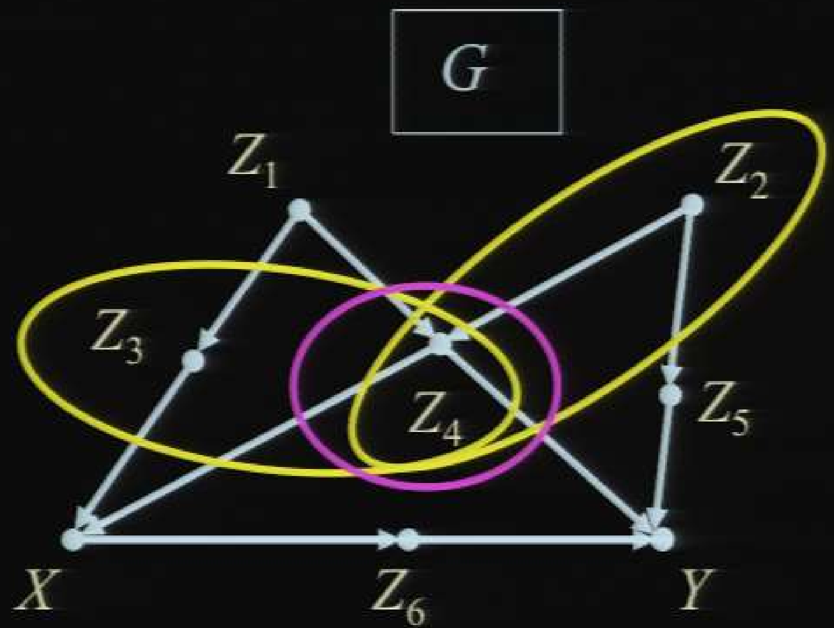
there is a set Z of variables that d-separates X from Y in $G_{\underline{X}}$





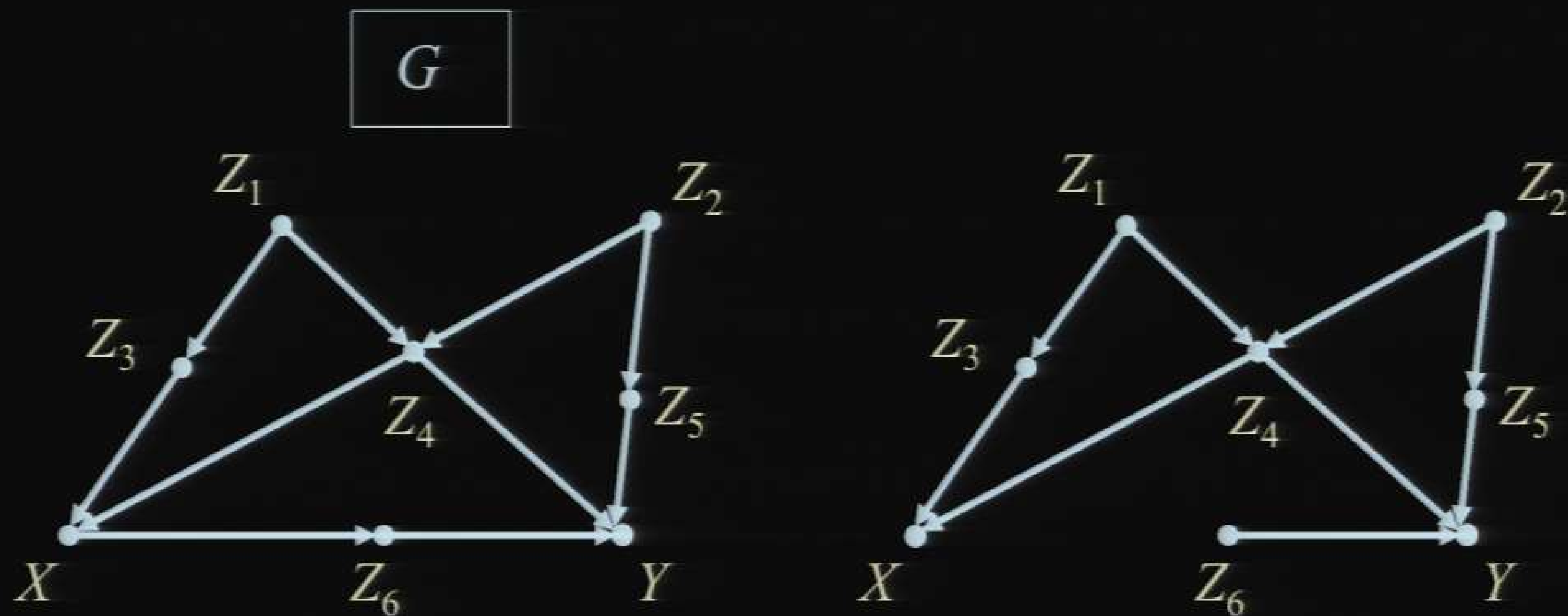
TOOL 3. BACK-DOOR CRITERION (THE PROBLEM OF CONFOUNDING)

Goal: Find the effect of X on Y, $P(y|do(x))$, given measurements on auxiliary variables $Z_1,\ldots, Z_k$



ELIMINATING CONFOUNDING BIAS
THE BACK-DOOR CRITERION

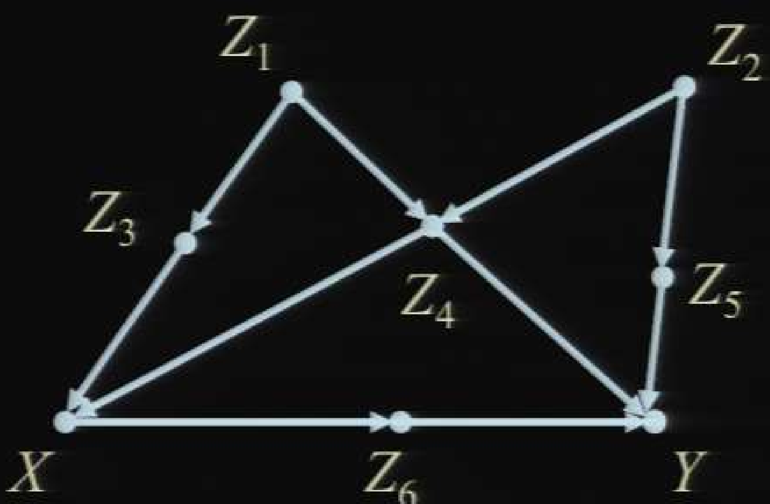
$P(y \mid do(x))$ is estimable if
there is a set Z of variables that d-separates X from Y in $G_{\underline{X}}$



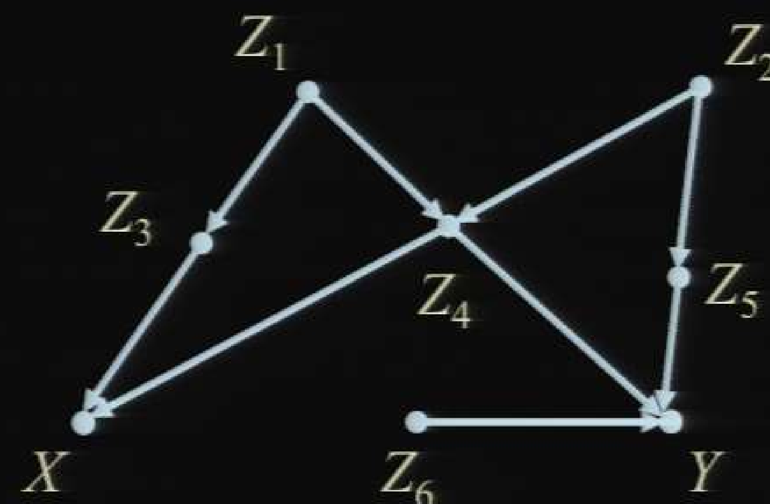
ELIMINATING CONFOUNDING BIAS THE BACK-DOOR CRITERION

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

G

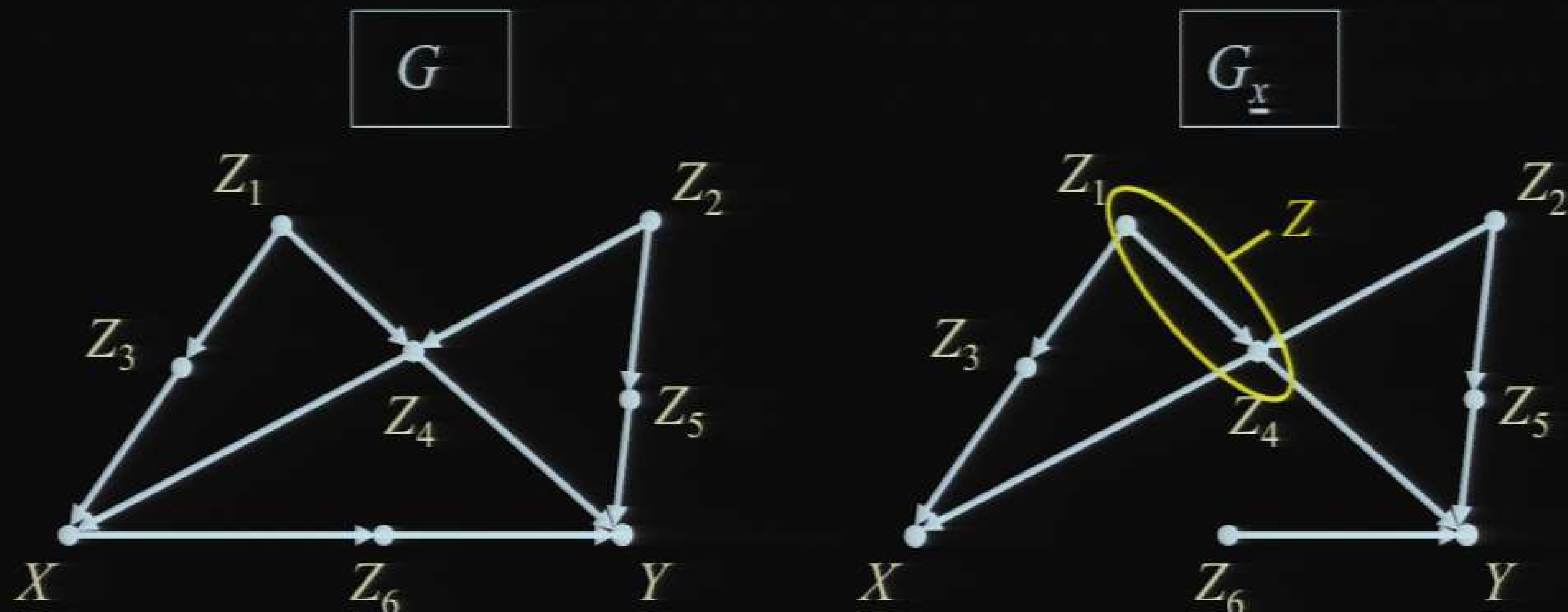


$G_{\underline{x}}$



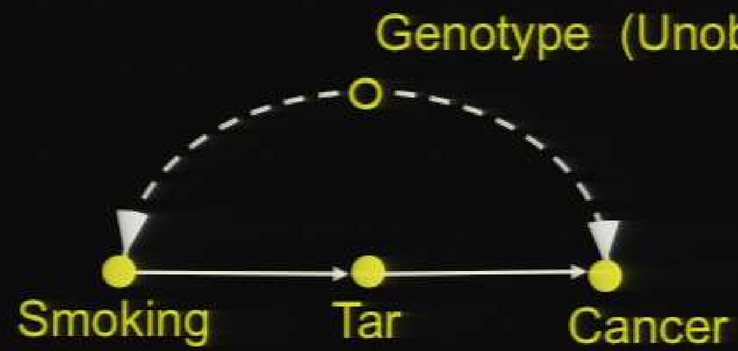
ELIMINATING CONFOUNDING BIAS THE BACK-DOOR CRITERION

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



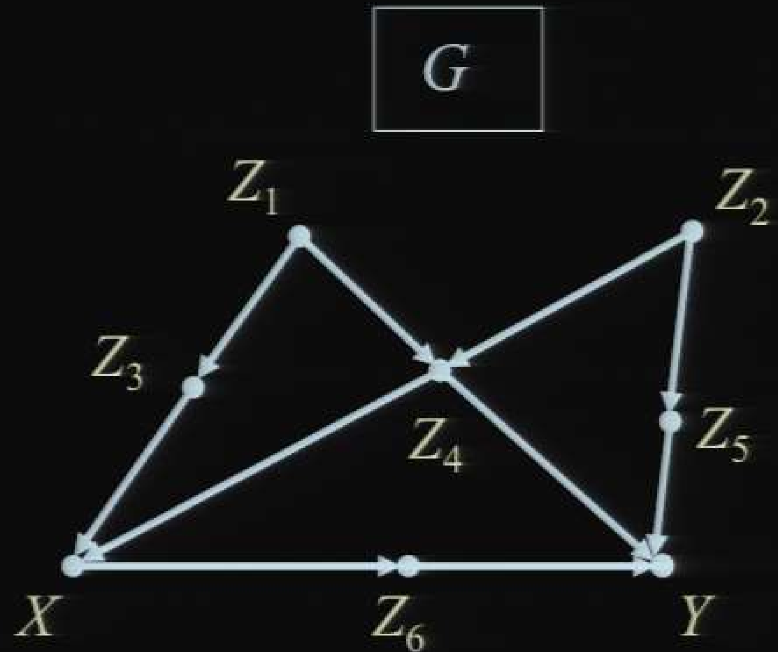
GOING BEYOND ADJUSTMENT

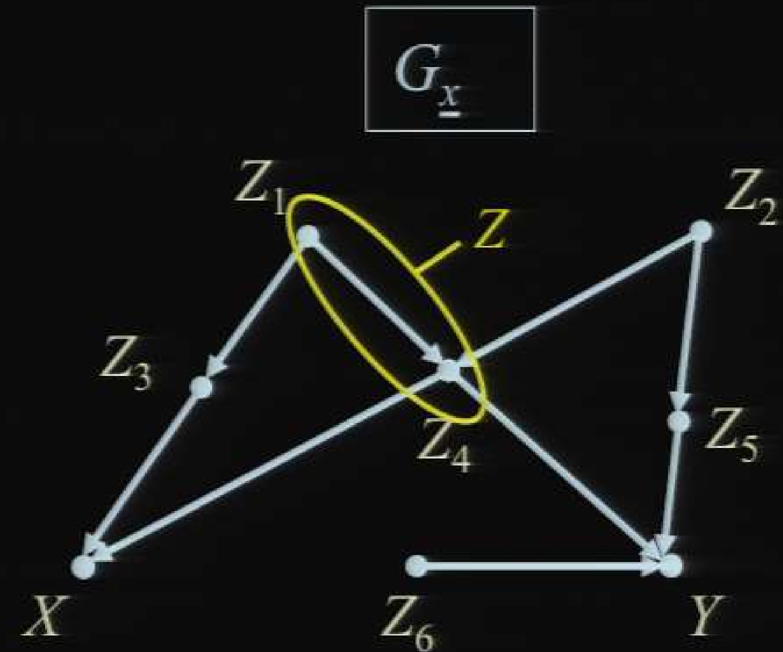
Goal: Find the effect of S on C, $P(c \mid do(s))$, given measurements on auxiliary variable T, and when latent variables confound the relationship S-C.



ELIMINATING CONFOUNDING BIAS THE BACK-DOOR CRITERION

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

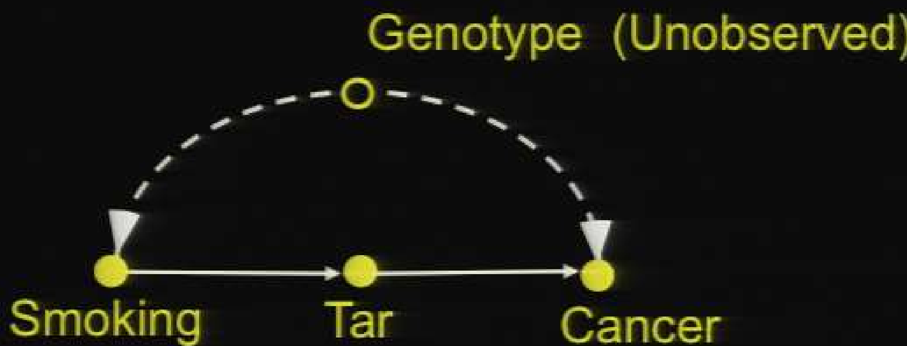




Moreover, $P(y \mid do(x)) = \sum_z P(y \mid x, z)\, P(z)$
("adjusting" for Z)

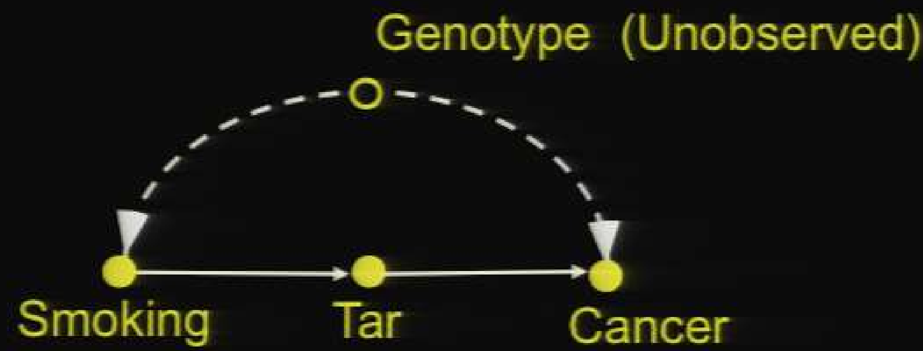
GOING BEYOND ADJUSTMENT

Goal: Find the effect of S on C, $P(c \mid do(s))$, given measurements on auxiliary variable T, and when latent variables confound the relationship S-C.



GOING BEYOND ADJUSTMENT

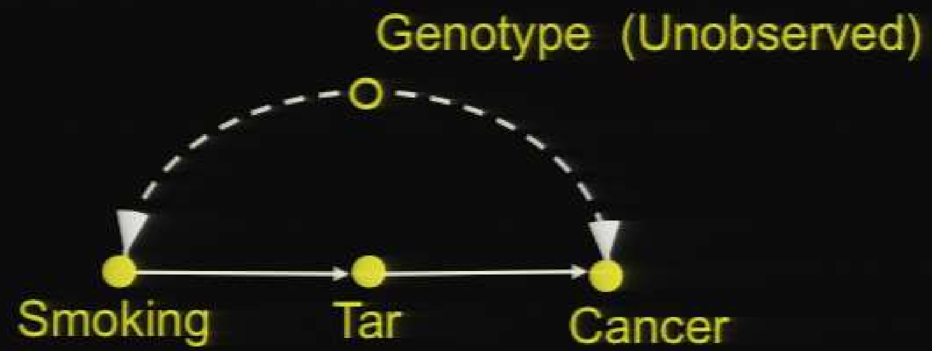
Goal: Find the effect of S on C, $P(c \mid do(s))$, given measurements on auxiliary variable T, and when latent variables confound the relationship S-C.



- What about the effect of S on T, $P(t \mid do(s))$?

GOING BEYOND ADJUSTMENT

Goal: Find the effect of S on C, $P(c \mid do(s))$, given measurements on auxiliary variable T, and when latent variables confound the relationship S-C.



- What about the effect of S on T, $P(t \mid do(s))$?
- What about the effect of T on C, $P(c \mid do(t))$?

OUTLINE

Concepts:

* Causal inference – a paradigm shift
* The two fundamental laws

Basic tools:

* Graph separation
* The truncated product formula
* The back-door adjustment formula
* The do-calculus

Capabilities:

* Policy evaluation
* Transportability
* Mediation
* Missing Data

TOOL 3. CAUSAL CALCULUS (IDENTIFIABILITY REDUCED TO CALCULUS)

The following transformations are valid for every interventional distribution generated by a structural causal model M:

Rule 1: Ignoring observations

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

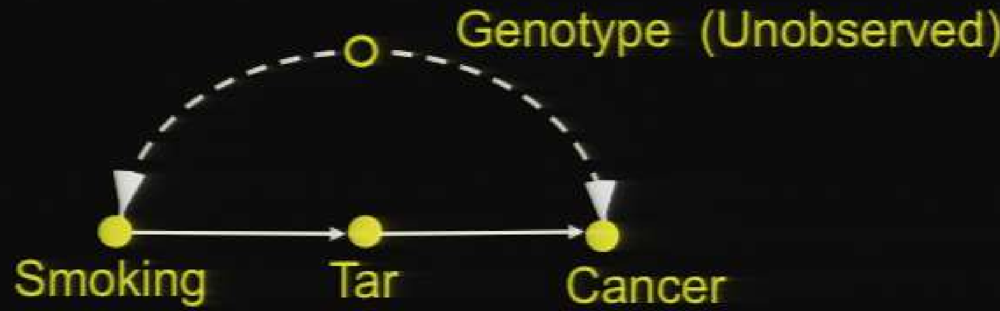
Rule 2: Action/observation exchange

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

Rule 3: Ignoring actions

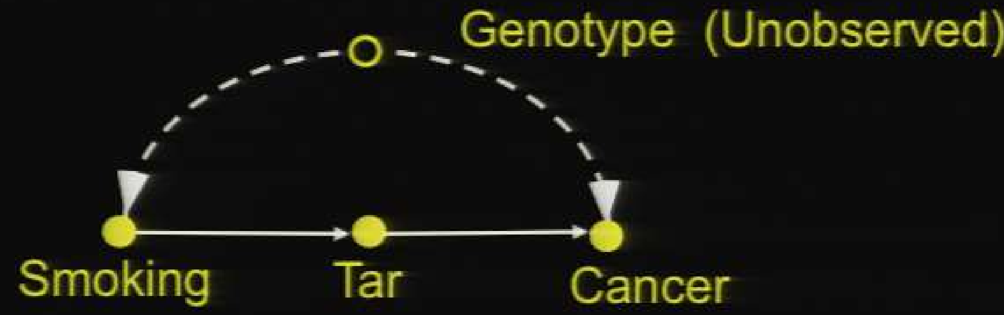
$P(y \mid do(x), do(z), w) = P(y \mid do(x), w)$, if $(Y \perp\!\!\!\perp Z \mid X, W)_{G_{\overline{X}\overline{Z(W)}}}$

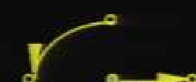
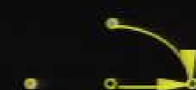
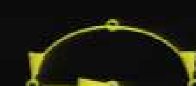
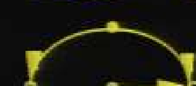
DERIVATION IN CAUSAL CALCULUS



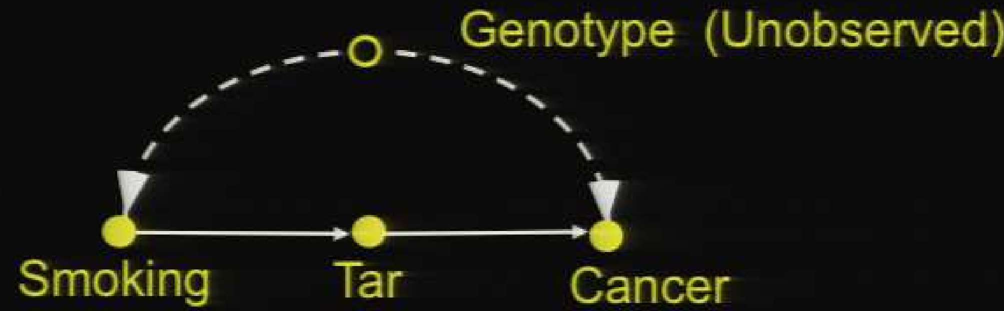
$P(c \mid do(s))$

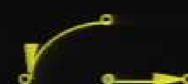
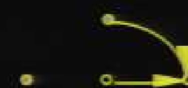
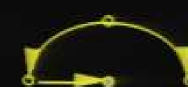
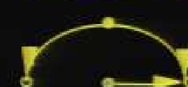
DERIVATION IN CAUSAL CALCULUS



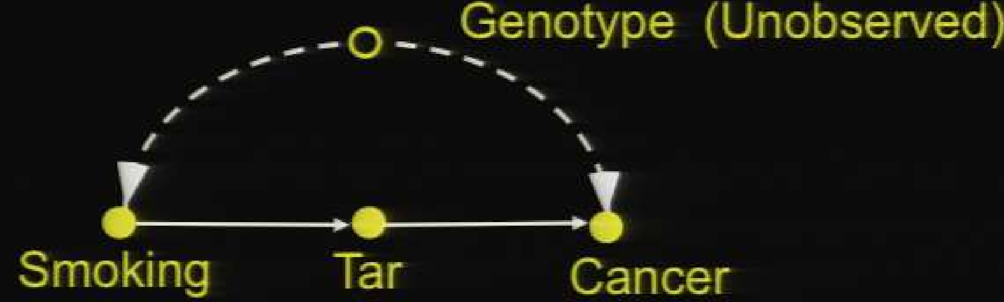
$P(c \mid do(s)) = \Sigma_t P(c \mid do(s), t)\, P(t \mid do(s))$	Probability Axioms	
$= \Sigma_t P(c \mid do(s), do(t))\, P(t \mid do(s))$	Rule 2	
$= \Sigma_t P(c \mid do(s), do(t))\, P(t \mid s)$	Rule 2	
$= \Sigma_t P(c \mid do(t))\, P(t \mid s)$	Rule 3	
$= \Sigma_{s'} \Sigma_t P(c \mid do(t), s')\, P(s' \mid do(t))\, P(t \mid s)$	Probability Axioms	
$= \Sigma_{s'} \Sigma_t P(c \mid t, s')\, P(s' \mid do(t))\, P(t \mid s)$	Rule 2	
$= \Sigma_{s'} \Sigma_t P(c \mid t, s')\, P(s')\, P(t \mid s)$	Rule 3	

DERIVATION IN CAUSAL CALCULUS



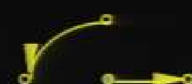
$P(c \mid do(s)) = \Sigma_t P(c \mid do(s), t)\, P(t \mid do(s))$	Probability Axioms	
$= \Sigma_t P(c \mid do(s), do(t))\, P(t \mid do(s))$	Rule 2	
$= \Sigma_t P(c \mid do(s), do(t))\, P(t \mid s)$	Rule 2	
$= \Sigma_t P(c \mid do(t))\, P(t \mid s)$	Rule 3	
$= \Sigma_{s'} \Sigma_t P(c \mid do(t), s')\, P(s' \mid do(t))\, P(t \mid s)$	Probability Axioms	
$= \Sigma_{s'} \Sigma_t P(c \mid t, s')\, P(s' \mid do(t))\, P(t \mid s)$	Rule 2	
$= \Sigma_{s'} \Sigma_t P(c \mid t, s')\, P(s')\, P(t \mid s)$	Rule 3	

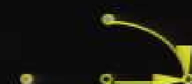
DERIVATION IN CAUSAL CALCULUS



$P(c \mid do(s)) = \Sigma_t P(c \mid do(s), t)\, P(t \mid do(s))$ Probability Axioms

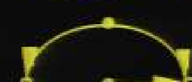
$= \Sigma_t P(c \mid do(s), do(t))\, P(t \mid do(s))$ Rule 2 

$= \Sigma_t P(c \mid do(s), do(t))\, P(t \mid s)$ Rule 2 

$= \Sigma_t P(c \mid do(t))\, P(t \mid s)$ Rule 3 

$= \Sigma_{s'} \Sigma_t P(c \mid do(t), s')\, P(s' \mid do(t))\, P(t \mid s)$ Probability Axioms

$= \Sigma_{s'} \Sigma_t P(c \mid t, s')\, P(s' \mid do(t))\, P(t \mid s)$ Rule 2 

$= \Sigma_{s'} \Sigma_t P(c \mid t, s')\, P(s')\, P(t \mid s)$ Rule 3 

TECHNICAL NOTE.
THE IDENTIFIABILITY PROBLEM

ID PROBLEM (decision): Given two models M_1 and M_2 compatible with G that agree on the observable distribution over V, $P_1(v) = P_2(v)$, decide whether they also agree in the target quantity $Q = P(y \mid do(x))$, i.e., whether the effect $P(y \mid do(x))$ is identifiable from G and $P(v)$.

$$
\begin{tikzcd}
 & M_1 \arrow[dl] \arrow[dr] & \\
P_1(v) = P_2(v) \arrow[rr] & & Q_1 = Q_2 \\
 & M_2 \arrow[ul] \arrow[ur] &
\end{tikzcd}
$$

TECHNICAL NOTE. THE IDENTIFIABILITY PROBLEM

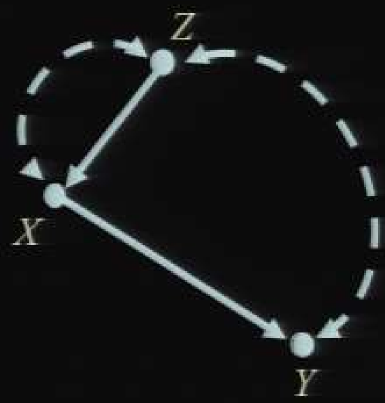
ID PROBLEM (decision): Given two models M_1 and M_2 compatible with $\boldsymbol{G}$ that agree on the observable distribution over V, $P_1(v) = P_2(v)$, decide whether they also agree in the target quantity $\boldsymbol{Q} = P(y \mid do(x))$, i.e., whether the effect $P(y \mid do(x))$ is identifiable from $\boldsymbol{G}$ and $P(v)$.

$$
\begin{tikzcd}
 & M_1 \arrow[dl] \arrow[dr] & \\
P_1(v) = P_2(v) \arrow[rr] & & Q_1 = Q_2 \\
 & M_2 \arrow[ul] \arrow[ur] &
\end{tikzcd}
$$

(i.e., $\exists f, f : P(v) \to P(y \mid do(x))$)

WHAT CAN EXPERIMENTS ON DIET REVEAL ABOUT THE EFFECT OF CHOLESTEROL ON HEART ATTACK?

G:



Z: Diet
X: Cholesterol level
Y: Heart Attack

Measured:

Observational study: $P(x, y, z)$

TECHNICAL NOTE. THE IDENTIFIABILITY PROBLEM

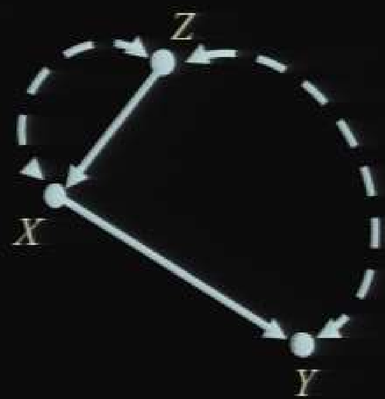
ID PROBLEM (decision): Given two models M_1 and M_2 compatible with G that agree on the observable distribution over V, $P_1(v) = P_2(v)$, decide whether they also agree in the target quantity $Q = P(y \mid do(x))$, i.e., whether the effect $P(y \mid do(x))$ is identifiable from G and $P(v)$.

$$
\begin{tikzcd}
 & M_1 \arrow[dl] \arrow[dr] & \\
P_1(v) = P_2(v) \arrow[rr] & & Q_1 = Q_2 \\
 & M_2 \arrow[ul] \arrow[ur] &
\end{tikzcd}
$$

(i.e., $\exists f, f: P(v) \rightarrow P(y \mid do(x))$)

WHAT CAN EXPERIMENTS ON DIET REVEAL ABOUT THE EFFECT OF CHOLESTEROL ON HEART ATTACK?

G:



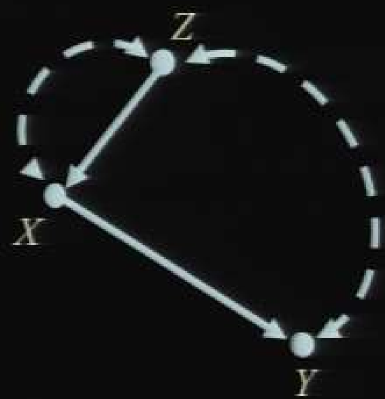
Z: Diet
X: Cholesterol level
Y: Heart Attack

Measured:

Observational study: $P(x, y, z)$

WHAT CAN EXPERIMENTS ON DIET REVEAL ABOUT THE EFFECT OF CHOLESTEROL ON HEART ATTACK?

G:



Z: Diet
X: Cholesterol level
Y: Heart Attack

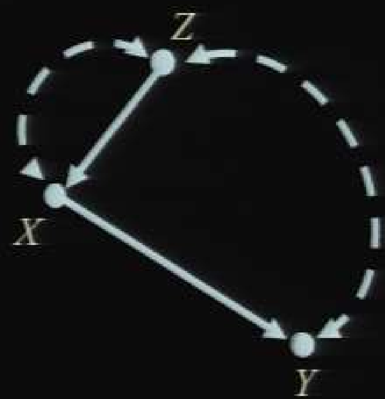
Measured:

Observational study: $P(x, y, z)$

Needed: $Q = P(y \mid do(x)) = ?$

WHAT CAN EXPERIMENTS ON DIET REVEAL ABOUT THE EFFECT OF CHOLESTEROL ON HEART ATTACK?

G:



Z: Diet
X: Cholesterol level
Y: Heart Attack

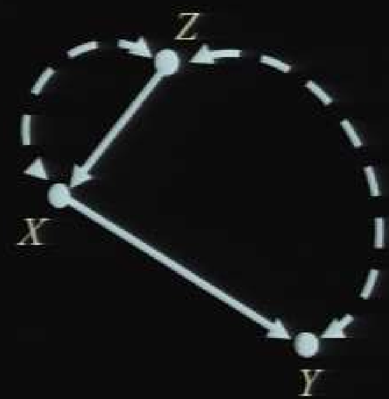
Measured:

Observational study: $P(x, y, z)$
Experimental study: $P(x, y \mid do(z))$

Needed: $Q = P(y \mid do(x)) = ?$

WHAT CAN EXPERIMENTS ON DIET REVEAL ABOUT THE EFFECT OF CHOLESTEROL ON HEART ATTACK?

G:



Z: Diet
X: Cholesterol level
Y: Heart Attack

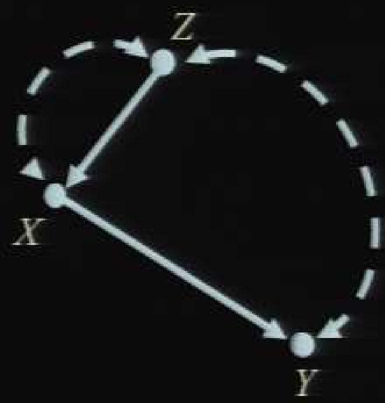
Measured:

Observational study: $P(x, y, z)$
Experimental study: $P(x, y \mid do(z))$

Needed: $Q = P(y \mid do(x)) = \,?\, = \dfrac{P(x, y \mid do(z))}{P(x \mid do(z))}$

WHAT CAN EXPERIMENTS ON DIET REVEAL ABOUT THE EFFECT OF CHOLESTEROL ON HEART ATTACK?

G:



Z: Diet
X: Cholesterol level
Y: Heart Attack

Measured:

Observational study: $P(x, y, z)$
Experimental study: $P(x, y \mid do(z))$

Needed: $Q = P(y \mid do(x)) = ? = \dfrac{P(x, y \mid do(z))}{P(x \mid do(z))}$

(i.e., $\exists\, f, f : P(v), P(v \mid do(z)) \rightarrow P(y \mid do(x))$)

OUTLINE

Concepts:

* Causal inference – a paradigm shift
* The two fundamental laws

Basic tools:

* Graph separation
* The truncated product formula
* The back-door adjustment formula
* The do-calculus

Capabilities:

* Policy evaluation
* Transportability
* Mediation
* Missing Data

SUMMARY OF POLICY EVALUATION RESULTS

- The estimability of any expression of the form

  $$Q = P(y_1, y_2, \ldots, y_n \mid do(x_1, x_2, \ldots, x_m), z_1, z_2, \ldots, z_k)$$

  can be determined given any causal graph $\boldsymbol{G}$ containing measured and unmeasured variables.

SUMMARY OF POLICY EVALUATION RESULTS

- The estimability of any expression of the form

  $$Q = P(y_1, y_2, \ldots, y_n \mid do(x_1, x_2, \ldots, x_m), z_1, z_2, \ldots, z_k)$$

  can be determined given any causal graph G containing measured and unmeasured variables.

- If Q is estimable, then its estimand can be derived in polynomial time (by estimable we mean either from observational or from experimental studies.)

SUMMARY OF POLICY EVALUATION RESULTS

- The estimability of any expression of the form

$$Q = P(y_1, y_2, \ldots, y_n \mid do(x_1, x_2, \ldots, x_m), z_1, z_2, \ldots, z_k)$$

  can be determined given any causal graph G containing measured and unmeasured variables.

- If Q is estimable, then its estimand can be derived in polynomial time (by estimable we mean either from observational or from experimental studies.)

- The algorithm is complete.

OUTLINE

Concepts:

* Causal inference – a paradigm shift
* The two fundamental laws

Basic tools:

* Graph separation
* The truncated product formula
* The back-door adjustment formula
* The do-calculus

Capabilities:

* Policy evaluation
* Transportability
* Mediation
* Missing Data

PROBLEM 2. GENERALIZABILITY AMONG POPULATIONS BREAK (TRANSPORTABILITY)

Question:

Is it possible to predict the effect of X on Y in a certain population Π^*, where no experiments can be conducted, using experimental data learned from a different population Π?

Answer: Sometimes yes.

HOW THIS PROBLEM IS SEEN IN OTHER SCIENCES?
(e.g., external validity, meta-analysis, ...)

- “Extrapolation across studies requires `some understanding of the reasons for the differences.’” (Cox, 1958)

HOW THIS PROBLEM IS SEEN IN OTHER SCIENCES?
(e.g., external validity, meta-analysis, ...)

- "Extrapolation across studies requires `some understanding of the reasons for the differences.'" (Cox, 1958)
- "`External validity' asks the question of generalizability: To what populations, settings, treatment variables, and measurement variables can this effect be generalized?" (Shadish, Cook and Campbell, 2002)

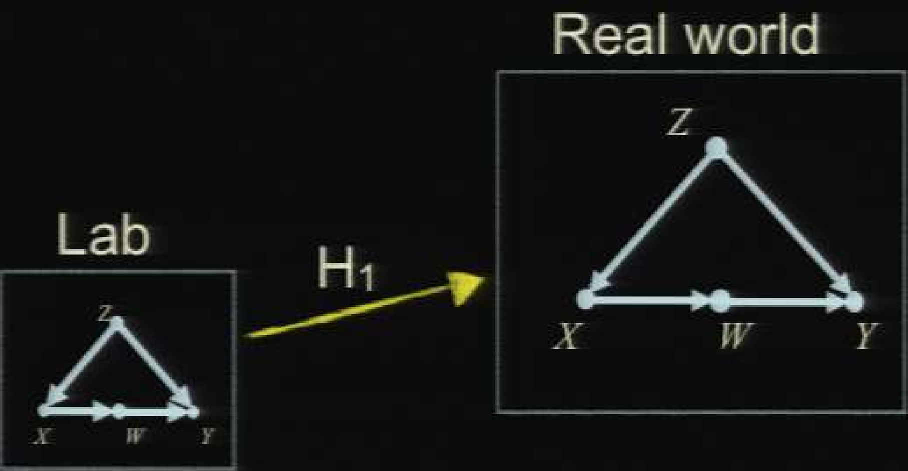
HOW THIS PROBLEM IS SEEN IN OTHER SCIENCES?
(e.g., external validity, meta-analysis, ...)

- "Extrapolation across studies requires `some understanding of the reasons for the differences.'" (Cox, 1958)
- "`External validity' asks the question of generalizability: To what populations, settings, treatment variables, and measurement variables can this effect be generalized?" (Shadish, Cook and Campbell, 2002)
- "An experiment is said to have "external validity" if the distribution of outcomes realized by a treatment group is the same as the distribution of outcome that would be realized in an actual program." (Manski, 2007)

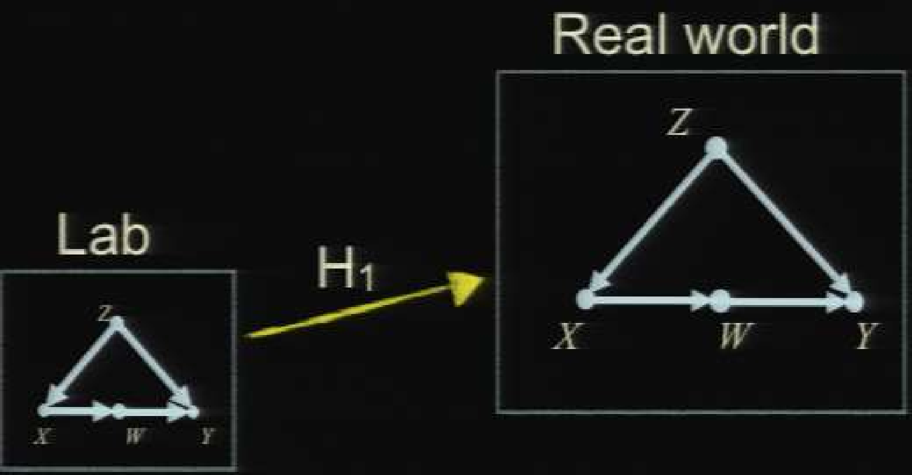
MOVING FROM THE “LAB” TO THE “REAL WORLD” ...

Lab

MOVING FROM THE "LAB" TO THE "REAL WORLD" ...

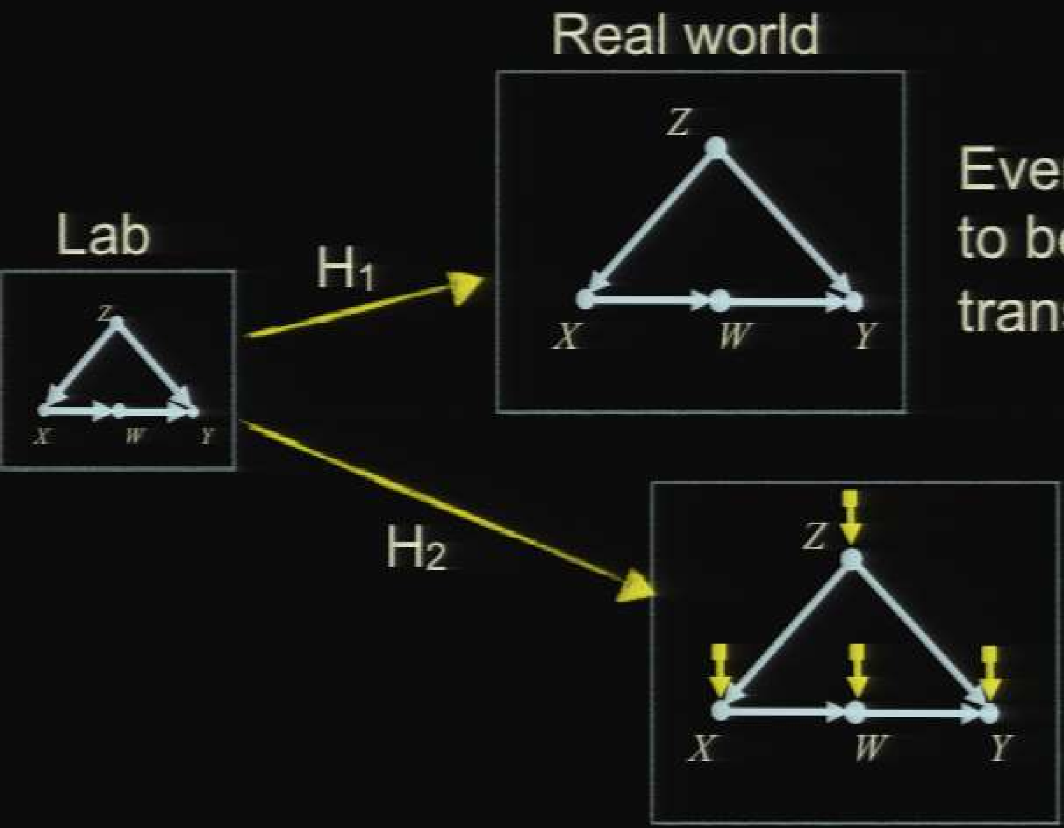


MOVING FROM THE "LAB" TO THE "REAL WORLD" ...



Everything is assumed to be the same, trivially transportable!

MOVING FROM THE "LAB" TO THE "REAL WORLD" ...



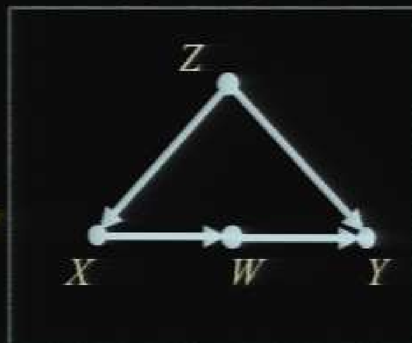
Everything is assumed to be the same, trivially transportable!

MOVING FROM THE "LAB" TO THE "REAL WORLD" ...

Lab

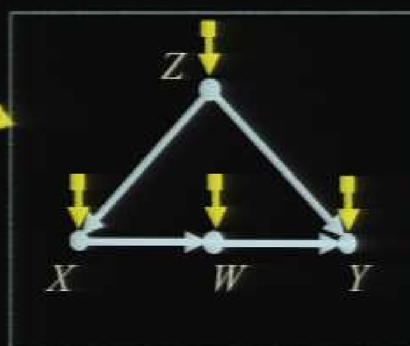
H_1

Real world



Everything is assumed to be the same, trivially transportable!

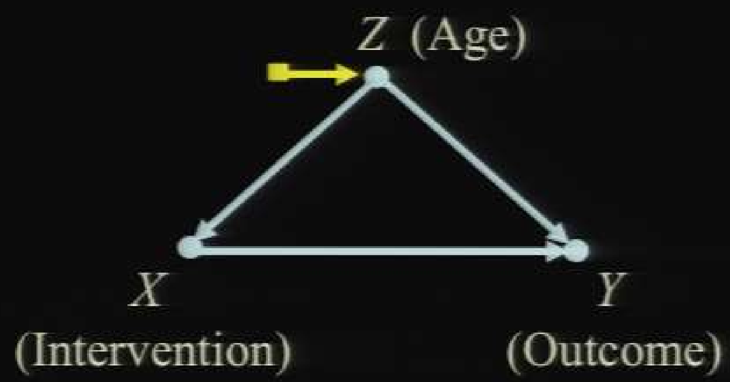
H_2



Everything is assumed to be different, not transportable...

MOTIVATION

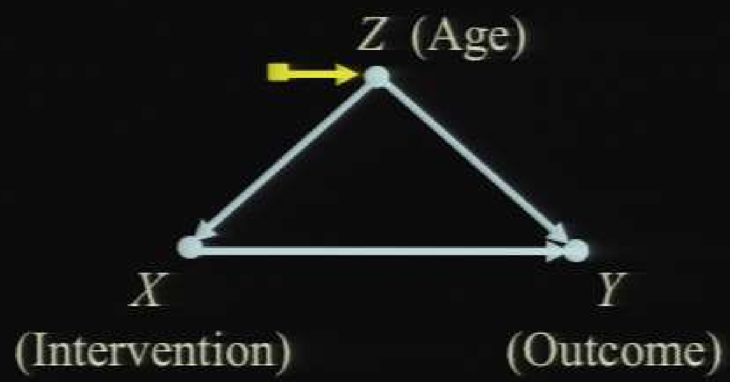
WHAT CAN EXPERIMENTS IN LA TELL US ABOUT NYC?



R: Π (LA) $\longrightarrow$ Π^* (NY)

MOTIVATION

WHAT CAN EXPERIMENTS IN LA TELL US ABOUT NYC?



R: Π (LA) ⟶ Π* (NY)

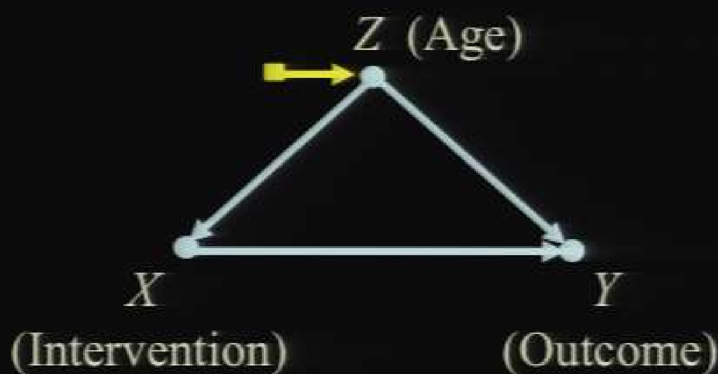
Experimental study in LA

Measured: $P(x, y, z)$

$P(y \mid do(x), z)$

MOTIVATION

WHAT CAN EXPERIMENTS IN LA TELL US ABOUT NYC?



R: Π (LA) $\longrightarrow$ Π^* (NY)

Experimental study in LA

Measured: $P(x, y, z)$

$P(y \mid do(x), z)$

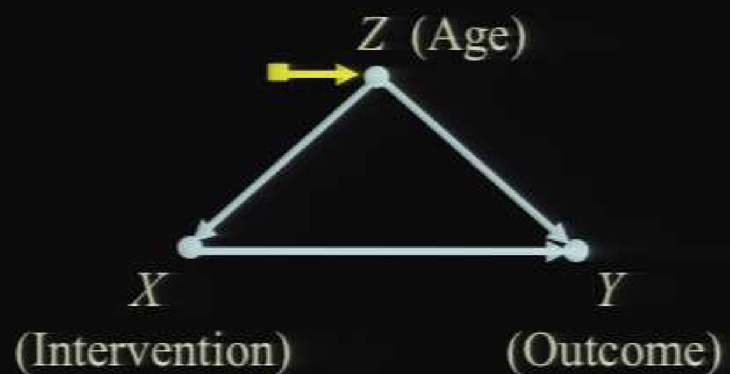
Observational study in NYC

Measured: $P^*(x, y, z)$

$P^*(z) \neq P(z)$

MOTIVATION

WHAT CAN EXPERIMENTS IN LA TELL US ABOUT NYC?



R: Π (LA) ⟶ Π* (NY)

Experimental study in LA

Measured: $P(x,y,z)$

$P(y \mid do(x), z)$

Observational study in NYC

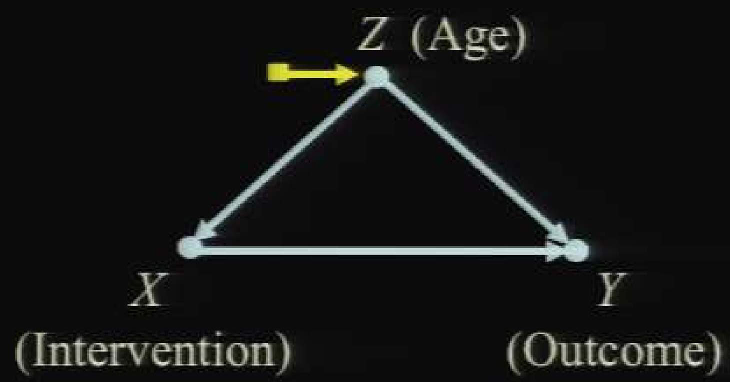
Measured: $P^*(x,y,z)$

$P^*(z) \neq P(z)$

Needed: R = $P^*(y \mid do(x)) = ?$

MOTIVATION

WHAT CAN EXPERIMENTS IN LA TELL US ABOUT NYC?



R: Π (LA) $\longrightarrow$ Π^* (NY)

Experimental study in LA

Measured: $P(x, y, z)$

$P(y \mid do(x), z)$

Observational study in NYC

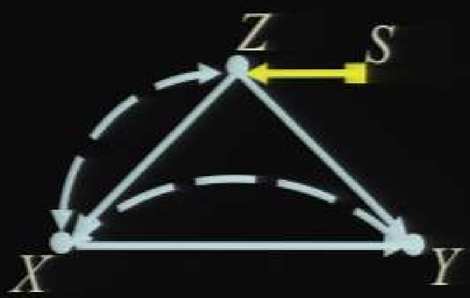
Measured: $P^*(x, y, z)$

$P^*(z) \neq P(z)$

Needed: R = $P^*(y \mid do(x))$ = ? $= \sum_z P(y \mid do(x), z) P^*(z)$

TRANSPORT FORMULAS DEPEND ON THE CAUSAL STORY

TRANSPORT FORMULAS DEPEND ON THE CAUSAL STORY

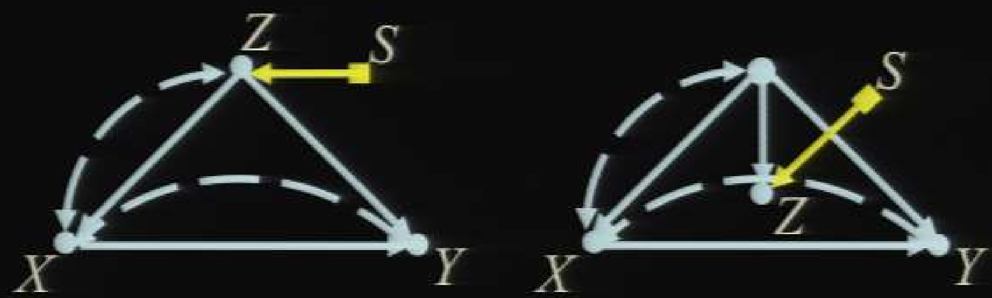


(a)

a) $\boldsymbol{Z}$ represents age

$$P^*(y \mid do(x)) = \sum_z P(y \mid do(x), z) P^*(z)$$

TRANSPORT FORMULAS DEPEND ON THE CAUSAL STORY

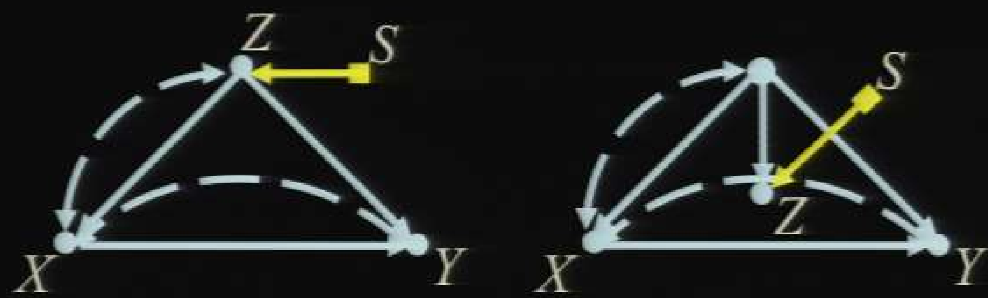


(a) (b)

a) ***Z*** represents age

$$P^*(y \mid do(x)) = \sum_z P(y \mid do(x), z) P^*(z)$$

TRANSPORT FORMULAS DEPEND ON THE CAUSAL STORY



(a) (b)

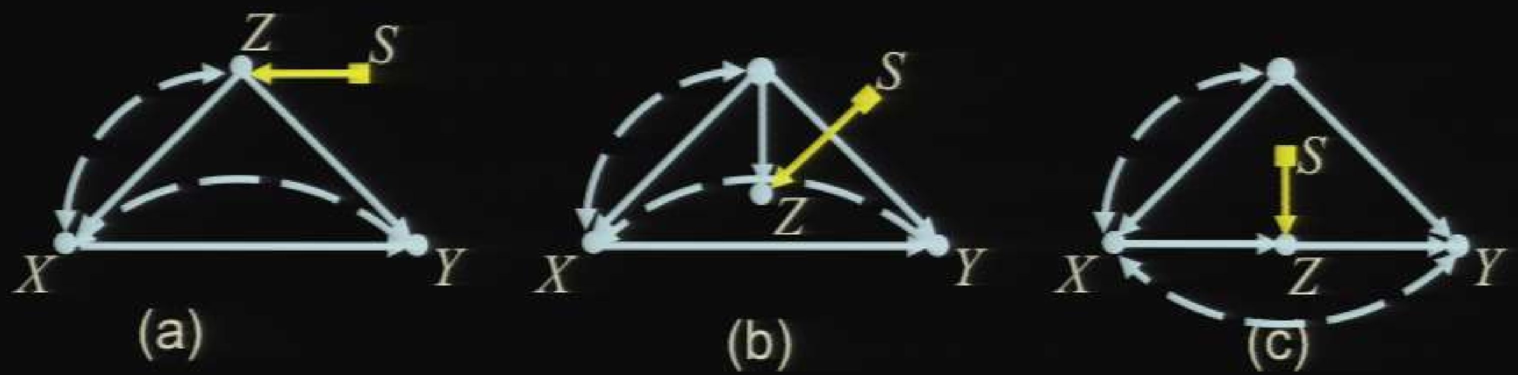
a) $\boldsymbol{Z}$ represents age

$$P^*(y \mid do(x)) = \sum_z P(y \mid do(x), z) P^*(z)$$

b) $\boldsymbol{Z}$ represents language skill

$$P^*(y \mid do(x)) = ?$$

TRANSPORT FORMULAS DEPEND ON THE CAUSAL STORY



a) $\boldsymbol{Z}$ represents age

$$P^*(y \mid do(x)) = \sum_z P(y \mid do(x), z) P^*(z)$$

b) $\boldsymbol{Z}$ represents language skill

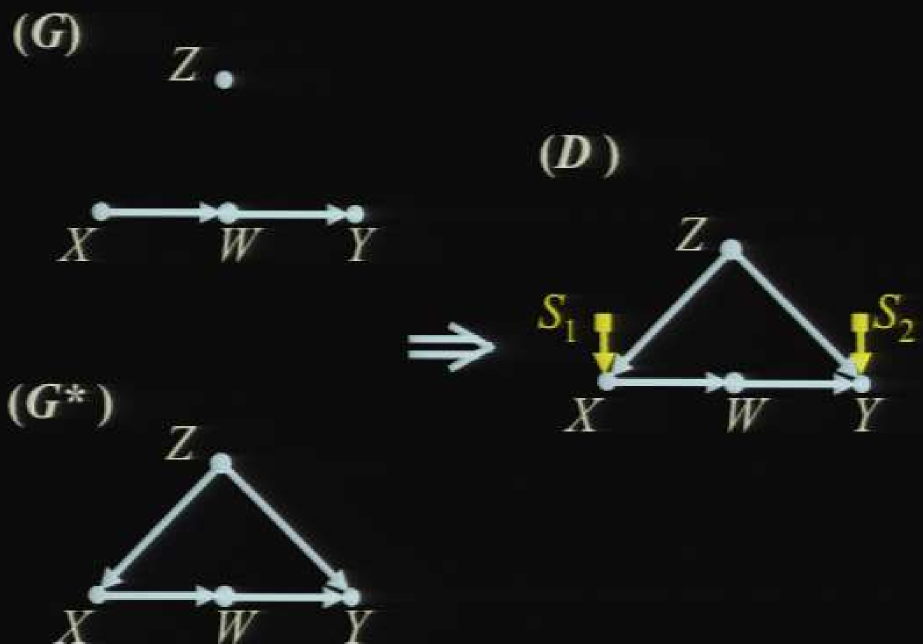
$$P^*(y \mid do(x)) = P(y \mid do(x))$$

c) $\boldsymbol{Z}$ represents a bio-marker

$$P^*(y \mid do(x)) = ?$$

SEMANTICS FOR TRANSPORTABILITY SELECTION DIAGRAMS

- How to encode disparities and commonalities about domains?

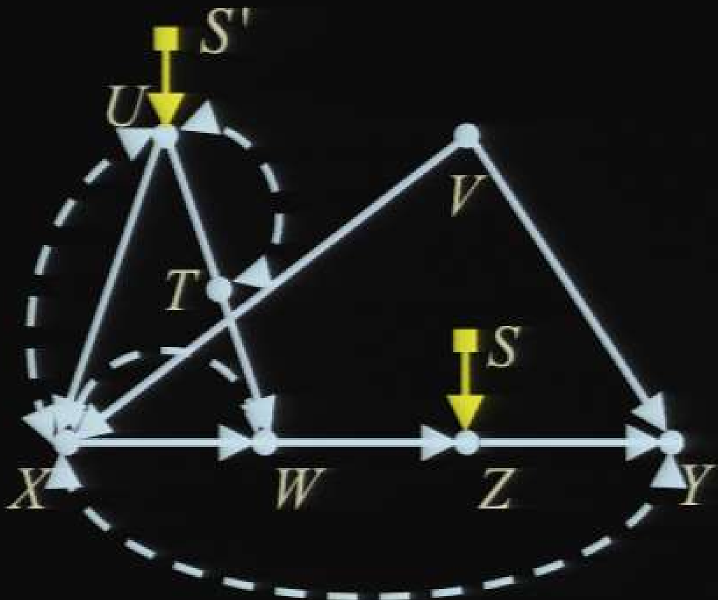


TRANSPORTABILITY REDUCED TO CALCULUS

Theorem

A causal relation $\boldsymbol{R}$ is transportable from $\prod$ to $\prod^*$ if and only if it is reducible, using the rules of *do*-calculus, to an expression in which $\boldsymbol{S}$ is separated from *do*( ).

RESULT: ALGORITHM TO DETERMINE IF AN EFFECT IS TRANSPORTABLE



INPUT: Annotated Causal Graph

S ■→ Factors creating differences

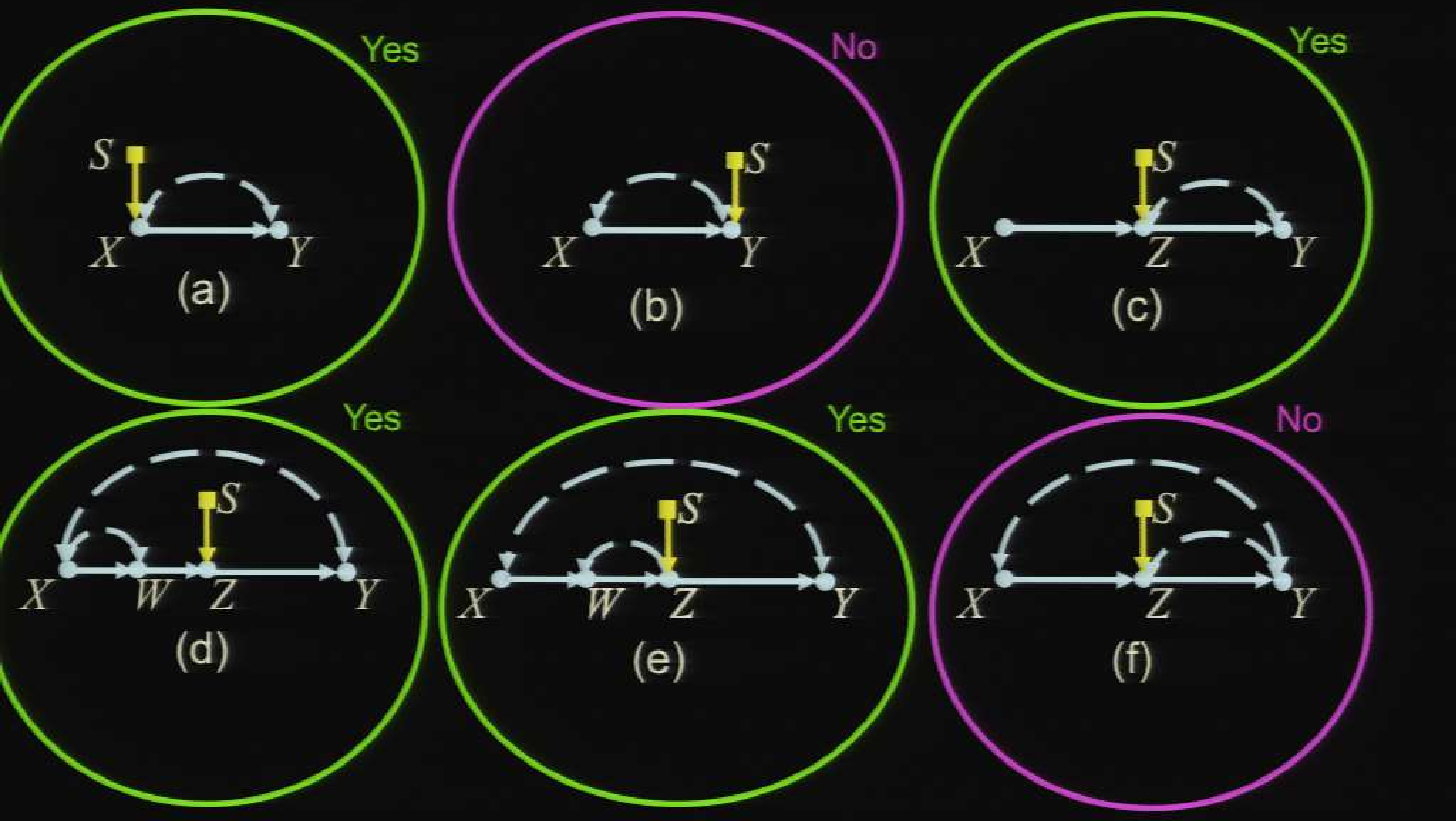
OUTPUT:

1. Transportable or not?
2. Measurements to be taken in the experimental study
3. Measurements to be taken in the target population
4. A transport formula

$$P^*(y \mid do(x)) =$$

$$\sum_z P(y \mid do(x), z) \sum_w P^*(z \mid w) \sum_t P(w \mid do(w), t) P^*(t)$$

WHICH MODEL LICENSES THE TRANSPORT OF THE CAUSAL EFFECT $X \rightarrow Y$



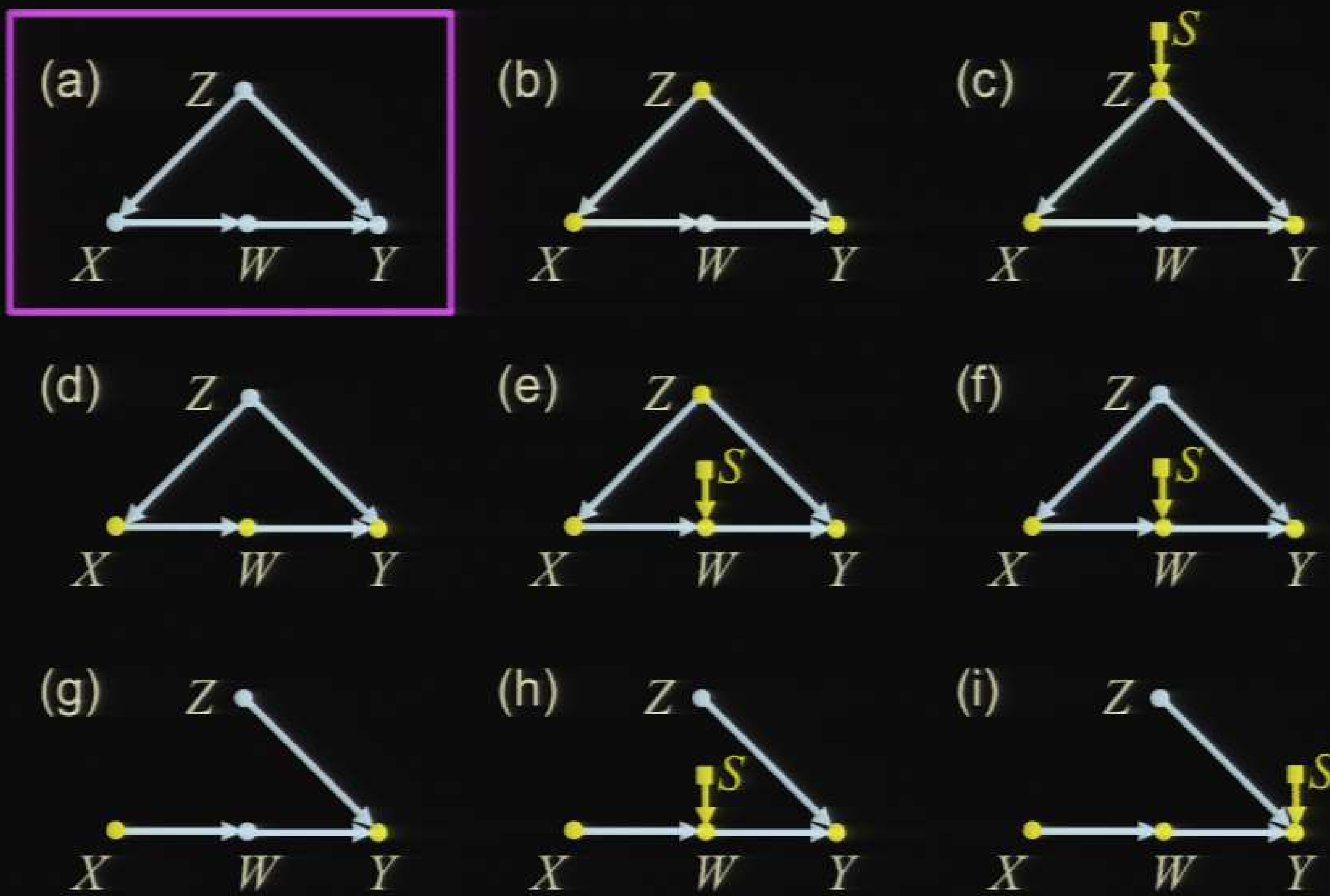
FROM META-ANALYSIS TO META-SYNTHESIS

The problem

How to combine results of several experimental and observational studies, each conducted on a different population and under a different set of experimental conditions, so as to construct an aggregate measure of effect size that is "better" than any one study in isolation.

META-SYNTHESIS AT WORK

Target population $\prod^*$ $\quad$ $R = P^*(y \mid do(x))$



SUMMARY OF TRANSPORTABILITY RESULTS

SUMMARY OF TRANSPORTABILITY RESULTS

- Nonparametric transportability of experimental results from multiple environments and limited experiments can be determined provided that commonalities and differences are encoded in selection diagrams.

SUMMARY OF TRANSPORTABILITY RESULTS

- Nonparametric transportability of experimental results from multiple environments and limited experiments can be determined provided that commonalities and differences are encoded in selection diagrams.
- When transportability is feasible, the transport formula can be derived in polynomial time.

SUMMARY OF TRANSPORTABILITY RESULTS

- Nonparametric transportability of experimental results from multiple environments and limited experiments can be determined provided that commonalities and differences are encoded in selection diagrams.
- When transportability is feasible, the transport formula can be derived in polynomial time.
- The algorithm is complete.
- The causal calculus is complete for this task.

OUTLINE

Concepts:

* Causal inference – a paradigm shift
* The two fundamental laws

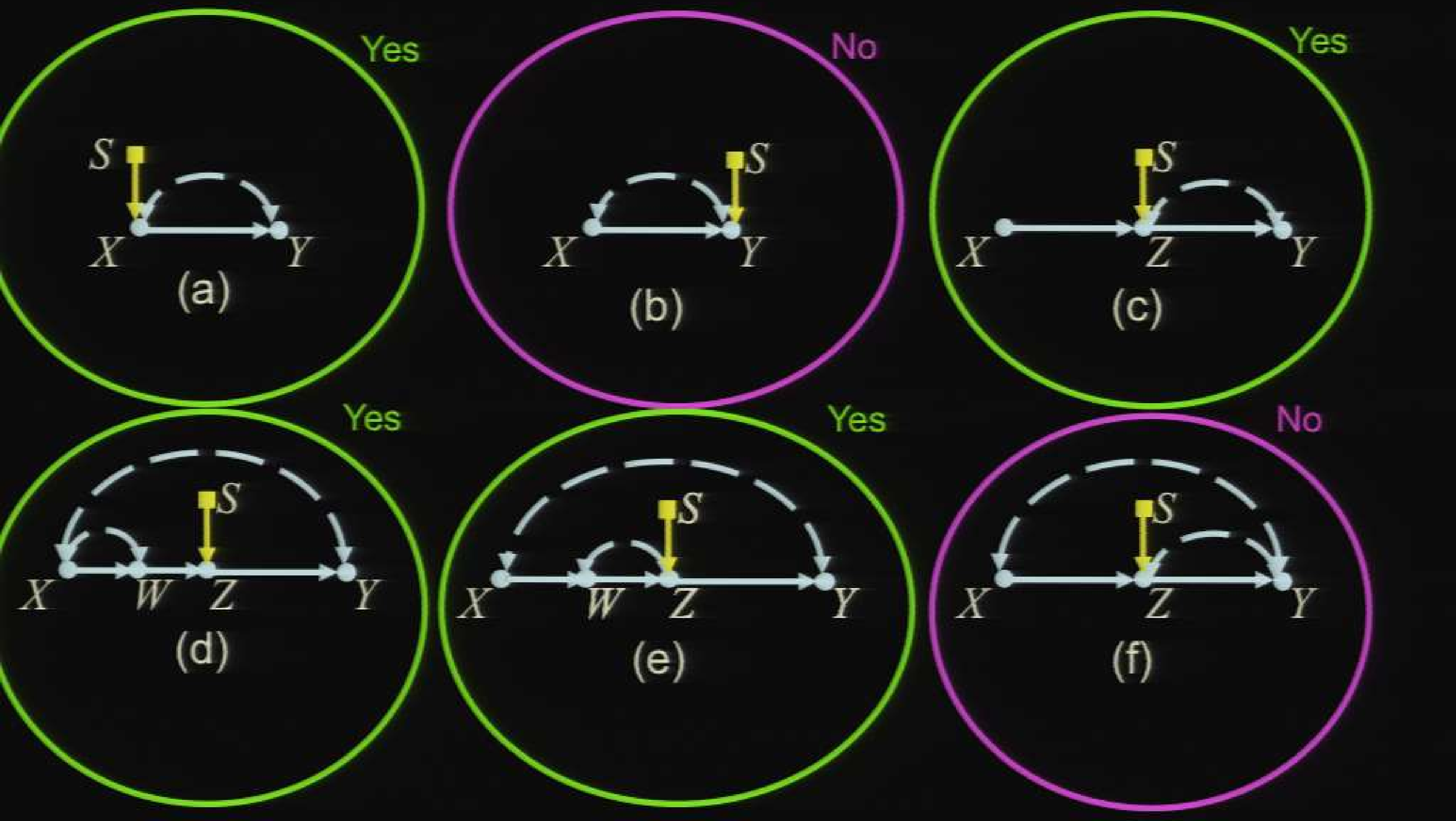
Basic tools:

* Graph separation
* The truncated product formula
* The back-door adjustment formula
* The do-calculus

Capabilities:

* Policy evaluation
* Transportability
* Mediation
* Missing Data

WHICH MODEL LICENSES THE TRANSPORT OF THE CAUSAL EFFECT $X \rightarrow Y$

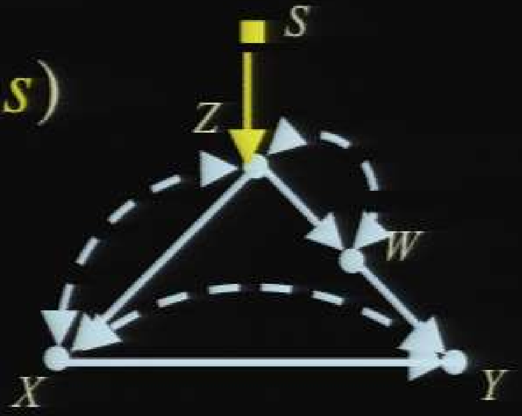


TRANSPORTABILITY REDUCED TO CALCULUS

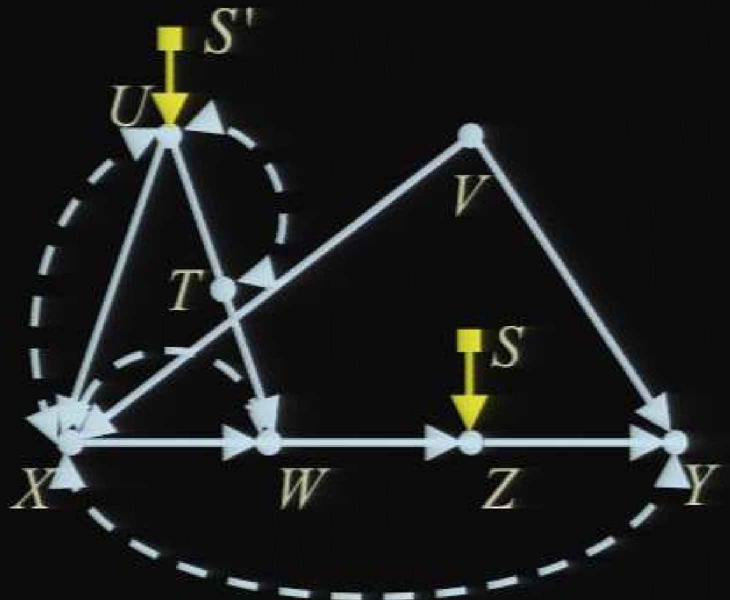
Theorem

A causal relation $\boldsymbol{R}$ is transportable from $\prod$ to $\prod^*$ if and only if it is reducible, using the rules of ***do***-calculus, to an expression in which $\boldsymbol{S}$ is separated from ***do***( ).

$$
\begin{aligned}
R &= P^*(y \mid do(x)) = P(y \mid do(x), s) \\
&= \sum_w P(y \mid do(x), s, w) P(w \mid do(x), s) \\
&= \sum_w P(y \mid do(x), w) P(w \mid s) \\
&= \sum_w P(y \mid do(x), w) P^*(w)
\end{aligned}
$$



RESULT: ALGORITHM TO DETERMINE IF AN EFFECT IS TRANSPORTABLE



INPUT: Annotated Causal Graph

S ⟼ Factors creating differences

OUTPUT:

1. Transportable or not?
2. Measurements to be taken in the experimental study
3. Measurements to be taken in the target population
4. A transport formula

$$P^*(y \mid do(x)) =$$

$$\sum_z P(y \mid do(x), z) \sum_w P^*(z \mid w) \sum_t P(w \mid do(w), t) P^*(t)$$

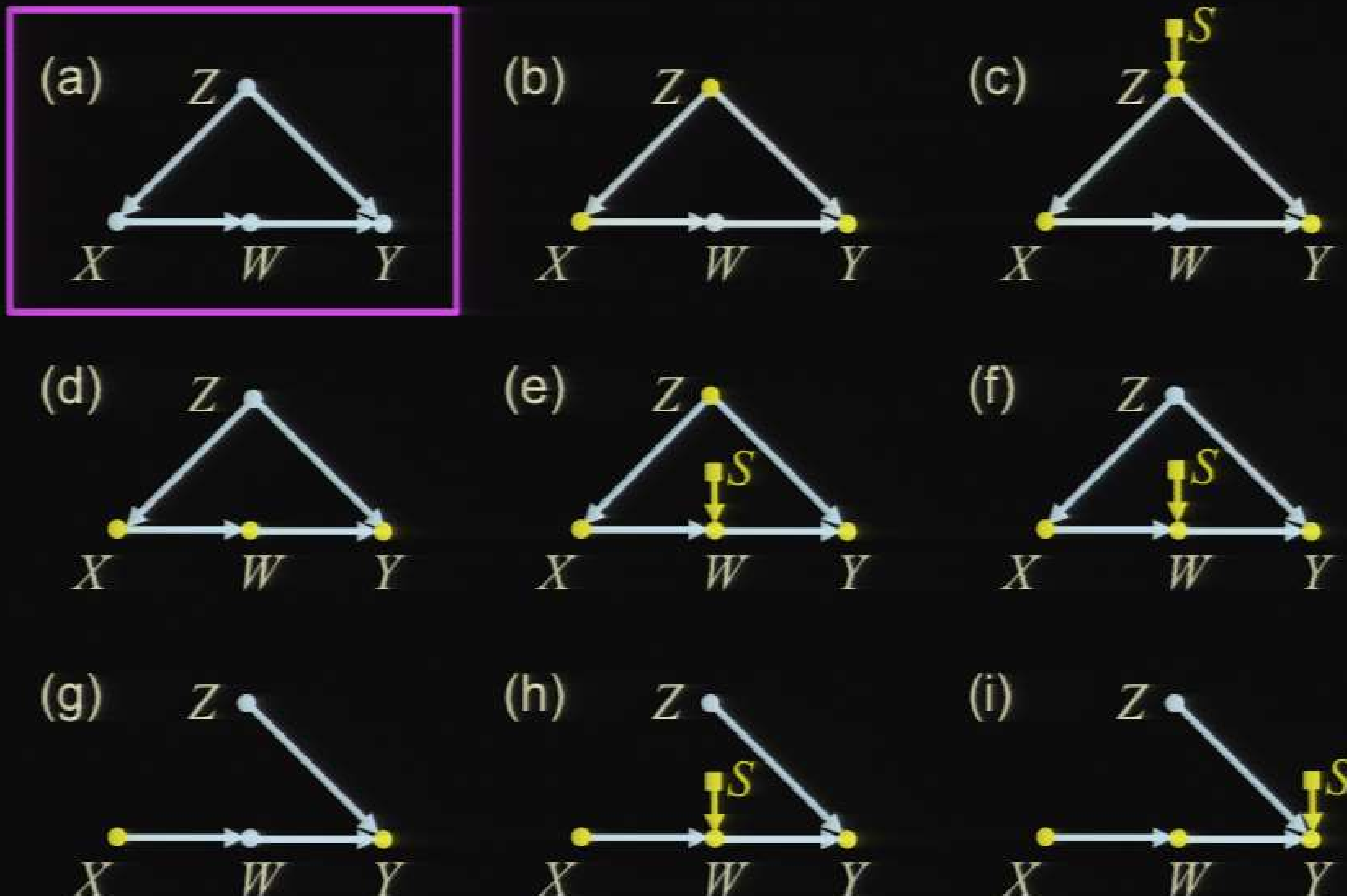
FROM META-ANALYSIS TO META-SYNTHESIS

The problem

How to combine results of several experimental and observational studies, each conducted on a different population and under a different set of experimental conditions, so as to construct an aggregate measure of effect size that is "better" than any one study in isolation.

META-SYNTHESIS AT WORK

Target population $\prod^*$ $R = P^*(y \mid do(x))$



SUMMARY OF TRANSPORTABILITY RESULTS

SUMMARY OF TRANSPORTABILITY RESULTS

- Nonparametric transportability of experimental results from multiple environments and limited experiments can be determined provided that commonalities and differences are encoded in selection diagrams.
- When transportability is feasible, the transport formula can be derived in polynomial time.
- The algorithm is complete.

MEDIATION: A GRAPHICAL-COUNTERFACTUAL SYMBIOSIS

1. Why decompose effects?
2. What is the definition 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?

OUTLINE

Concepts:

* Causal inference – a paradigm shift
* The two fundamental laws

Basic tools:

* Graph separation
* The truncated product formula
* The back-door adjustment formula
* The do-calculus

Capabilities:

* Policy evaluation
* Transportability
* Mediation
* Missing Data

MEDIATION: A GRAPHICAL-COUNTERFACTUAL SYMBIOSIS

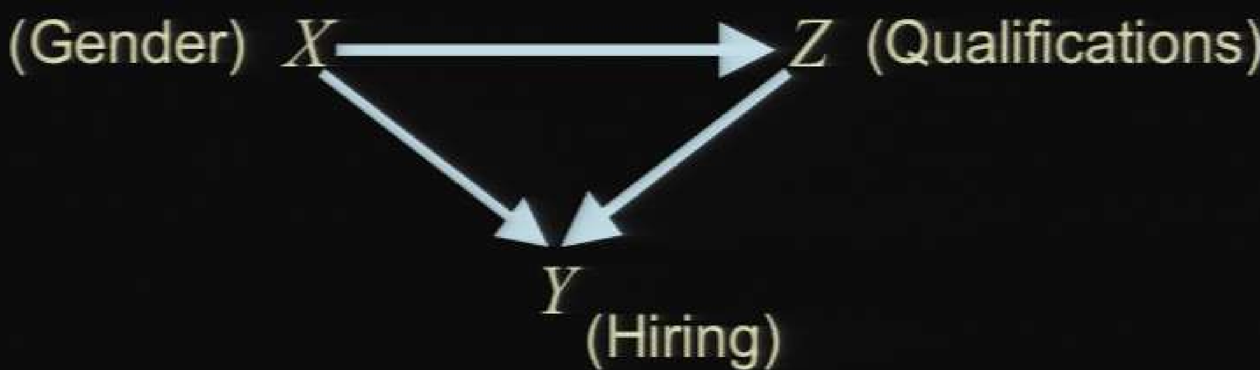
1. Why decompose effects?
2. What is the definition of direct and indirect effects?
3. What are the policy implications of direct and indirect effects?
4. When can direct and indirect effect be estimated consistently from experimental and nonexperimental data?

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: deactivate a mechanism, rather than fix a variable

LEGAL IMPLICATIONS OF DIRECT EFFECT

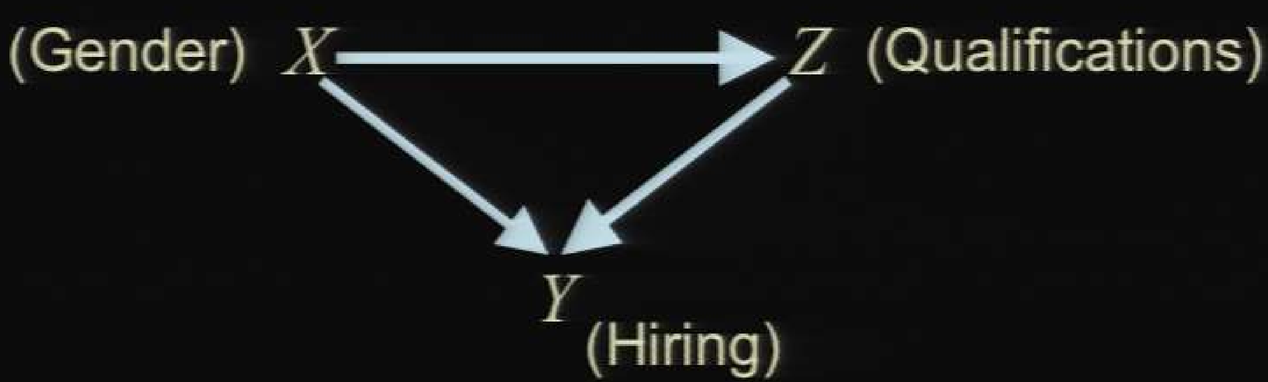
Can data prove an employer guilty of hiring discrimination?



What is the direct effect of X on Y?

LEGAL IMPLICATIONS OF DIRECT EFFECT

Can data prove an employer guilty of hiring discrimination?

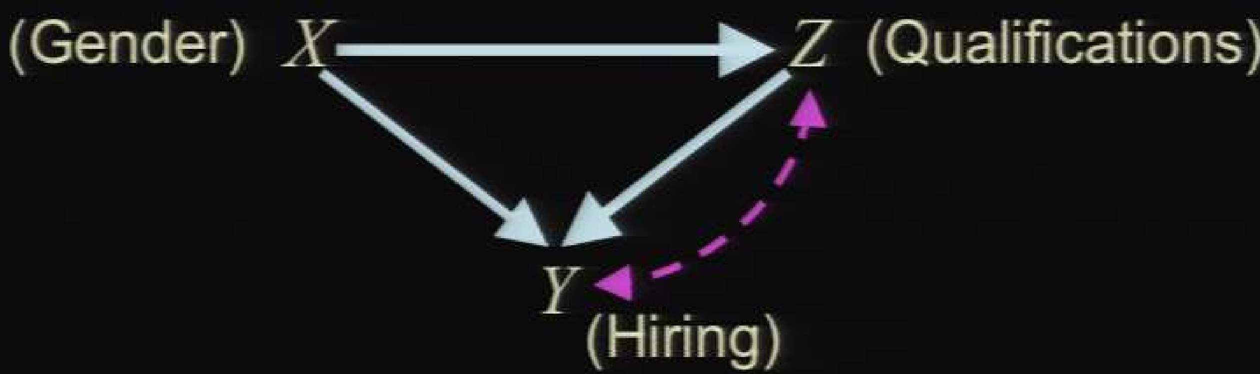


What is the direct effect of X on Y ?

Adjust for Z?

LEGAL IMPLICATIONS OF DIRECT EFFECT

Can data prove an employer guilty of hiring discrimination?



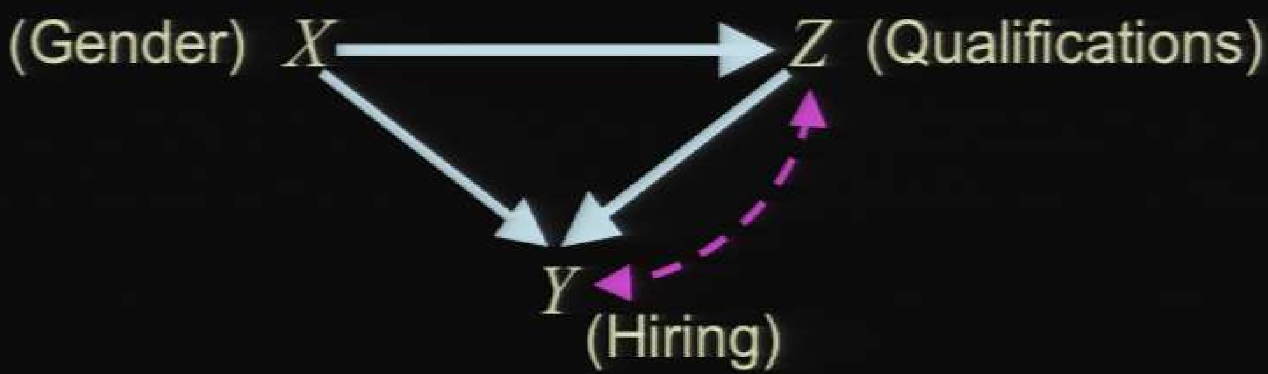
What is the direct effect of X on Y? (CDE)

$$E(Y|do(x_1),do(z)) - E(Y|do(x_0),do(z))$$

Adjust for Z? No! No!

LEGAL IMPLICATIONS OF DIRECT EFFECT

Can data prove an employer guilty of hiring discrimination?



What is the direct effect of X on Y ? (CDE)

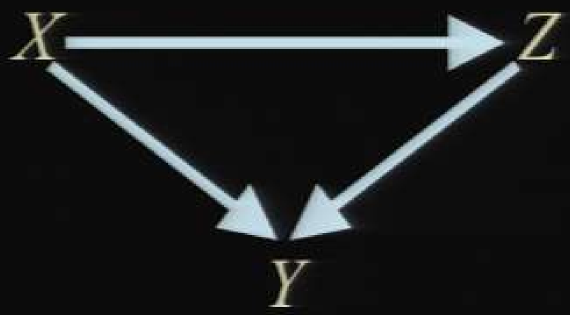
$$E(Y|do(x_1),do(z))-E(Y|do(x_0),do(z))$$

(z-dependent) Adjust for Z? No! No!

Identification is completely solved (Tian & Shpiser, 2006)

NATURAL INTERPRETATION OF AVERAGE DIRECT EFFECTS

Robins and Greenland (1992), Pearl (2001)



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

Natural 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 $= \beta(x_1 - x_0)$

DEFINITION OF INDIRECT EFFECTS

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

No controlled indirect effect

DEFINITION OF INDIRECT EFFECTS

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

No controlled indirect effect

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

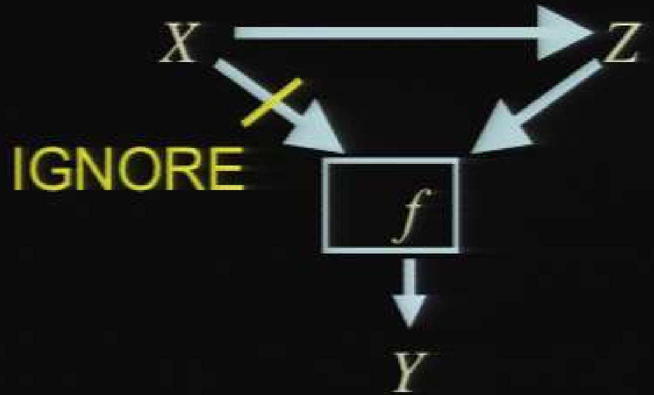
The expected change in Y when we keep X constant, say at x_0, and let Z change to whatever value it would have attained had X changed to x_1.

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

In linear models, $IE = TE - DE$

POLICY IMPLICATIONS OF INDIRECT EFFECTS

What is the indirect effect of X on Y?



Deactivating a link – a new type of intervention

THE MEDIATION FORMULAS IN UNCONFOUNDED MODELS

$$\begin{tikzcd} & Z \arrow[dr] & \\ X \arrow[ur] \arrow[rr] & & Y \end{tikzcd}$$

$z = f(x, u_1)$
$y = g(x, z, u_2)$
u_1 independent of u_2

$$DE = \sum_z [E(Y \mid x_1, z) - E(Y \mid x_0, z)] P(z \mid x_0)$$

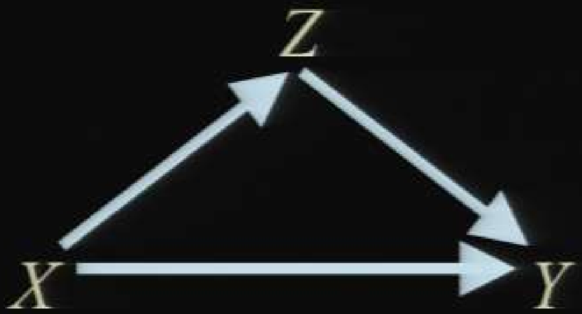
$$IE = \sum_z [E(Y \mid x_0, z)[P(z \mid x_1) - P(z \mid x_0)]$$

$$TE = E(Y \mid x_1) - E(Y \mid x_0) \qquad TE \neq DE + IE$$

$IE =$ Fraction of responses explained by mediation (sufficient)

$TE - DE =$ Fraction of responses owed to mediation (necessary)

THE MEDIATION FORMULAS IN UNCONFOUNDED MODELS



$z = f(x, u_1)$

$y = g(x, z, u_2)$

u_1 independent of u_2

$$DE = \sum_z [E(Y \mid x_1, z) - E(Y \mid x_0, z)] P(z \mid x_0)$$

$$IE = \sum_z [E(Y \mid x_0, z)[P(z \mid x_1) - P(z \mid x_0)]$$

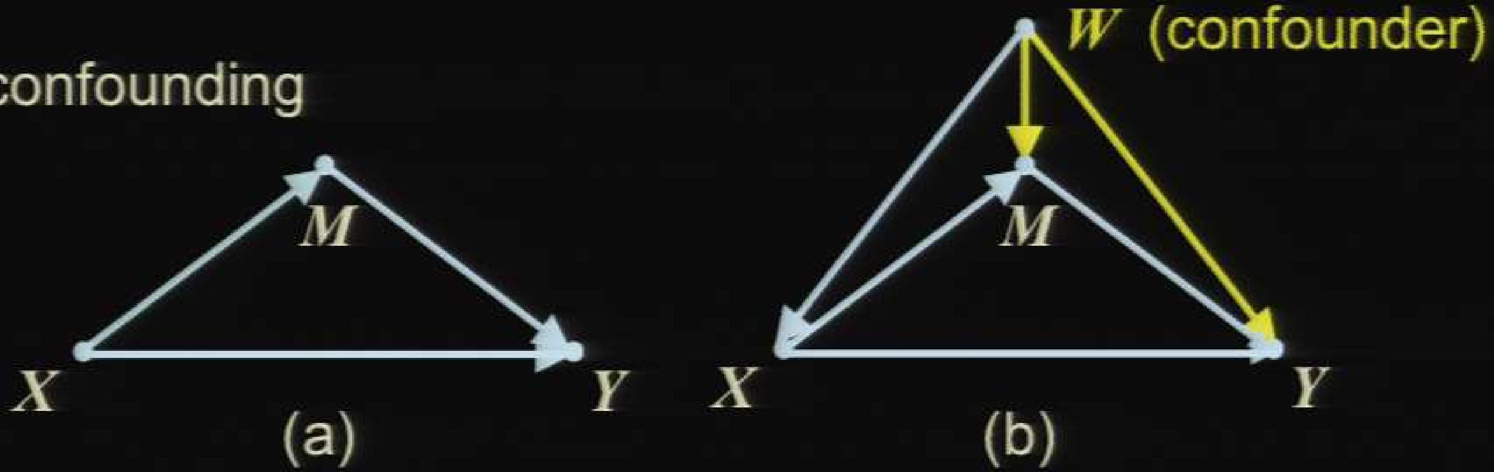
$$TE = E(Y \mid x_1) - E(Y \mid x_0)$$

$$TE \neq DE + IE$$

Complete identification conditions for confounded models with multiple mediators (Pearl 2001; Shpitser 2013).

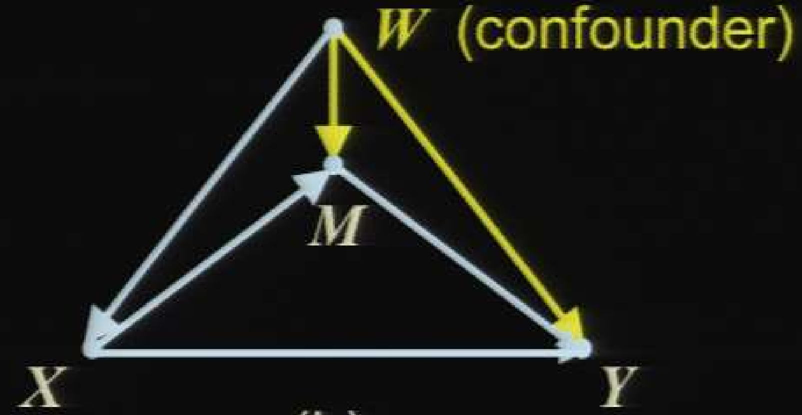
TRANSPARENT CONDITIONS OF NDE IDENTIFICATION

No confounding



TRANSPARENT CONDITIONS OF NDE IDENTIFICATION

No confounding



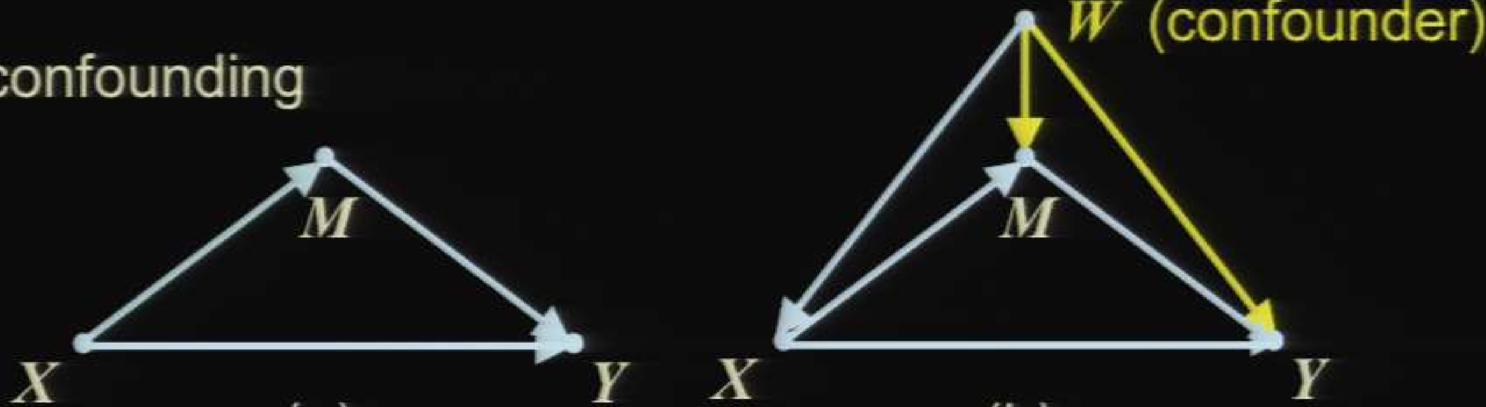
There exists a set W such that:

A-1 No member of W is a descendant of X.

A-2 W blocks all back-door paths from M to Y, disregarding the one through X.

TRANSPARENT CONDITIONS OF NDE IDENTIFICATION

No confounding



(a) (b)

There exists a set W such that:

A-1 No member of W is a descendant of X.

A-2 W blocks all back-door paths from M to Y, disregarding the one through X.

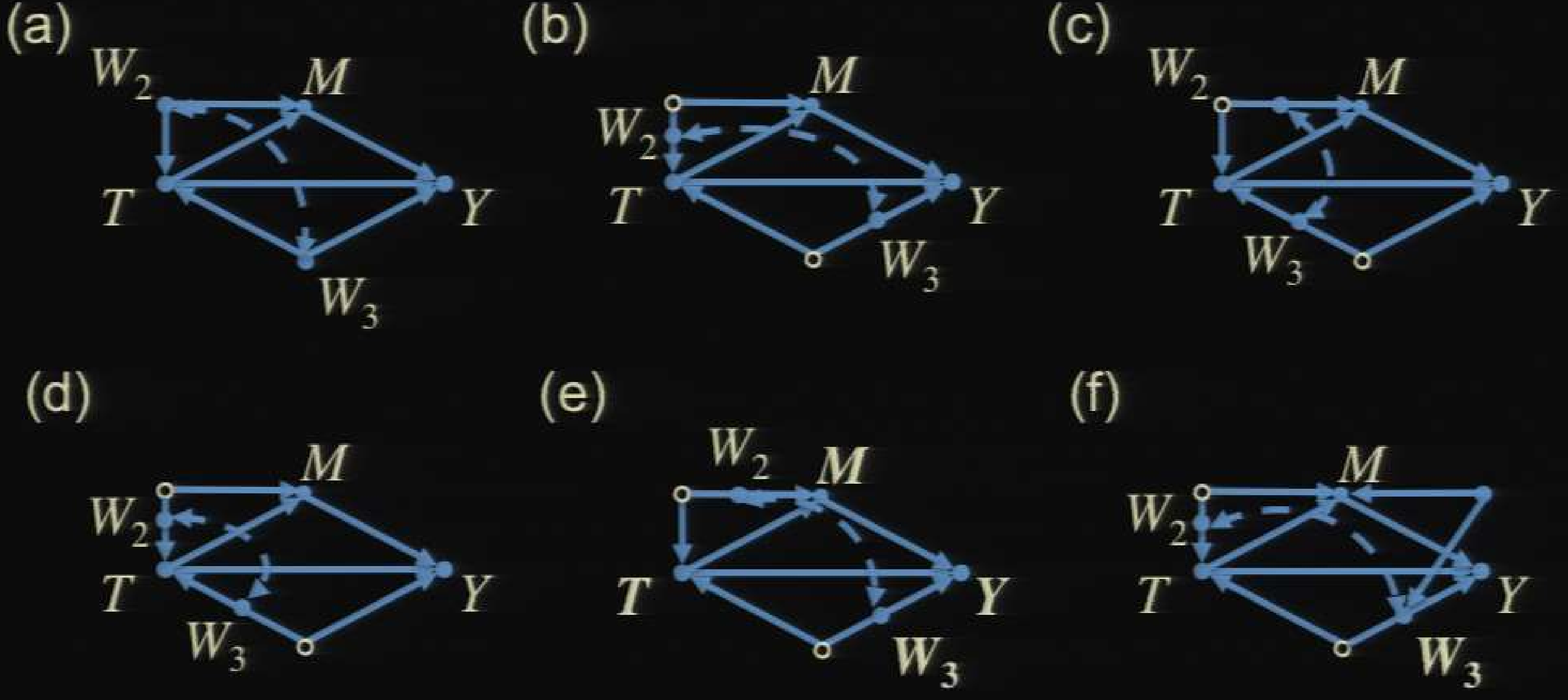
A-3 The W-specific effect of X on M is identifiable.

$$P(m \mid do(x), w)$$

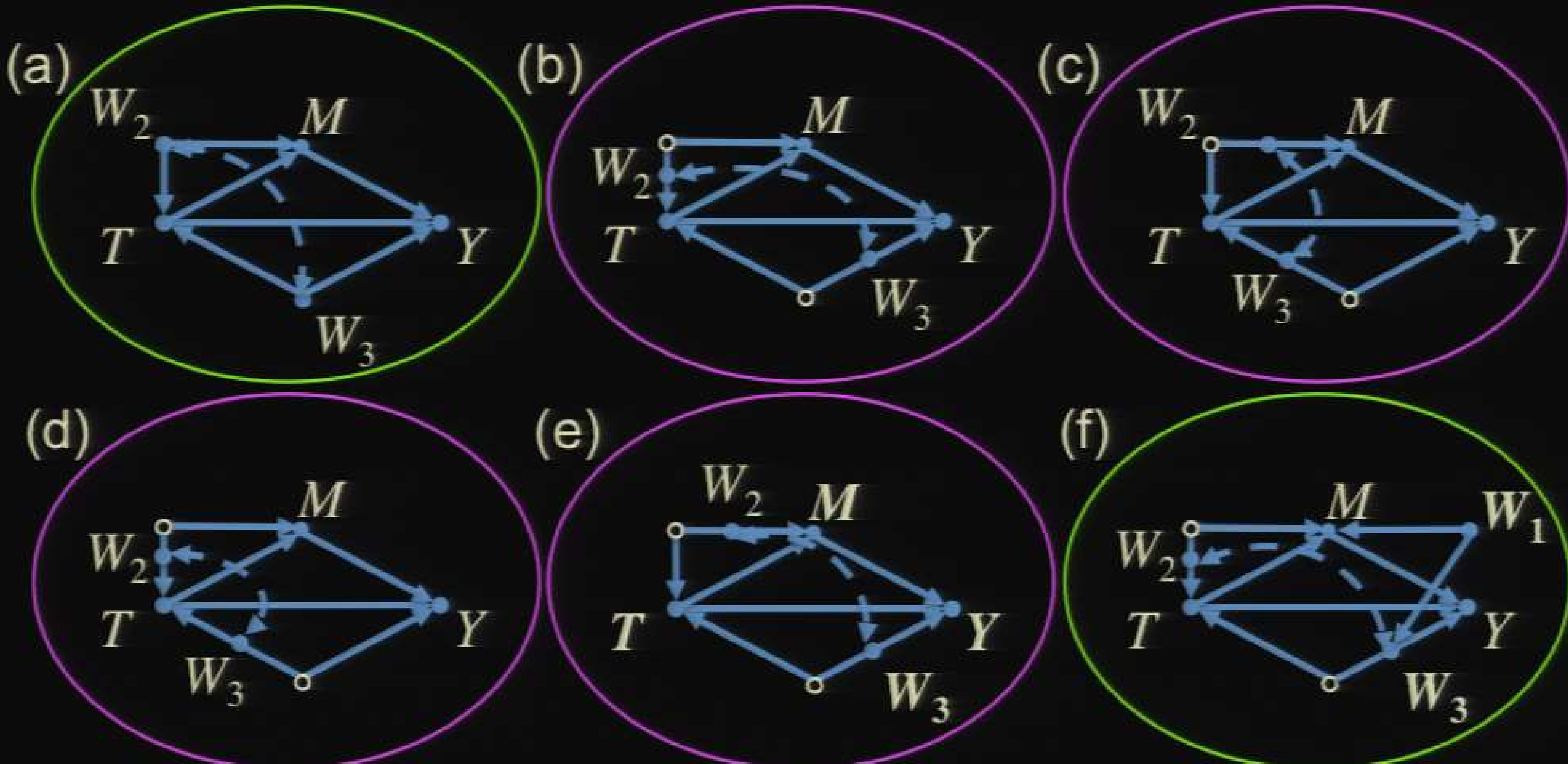
A-4 The W–specific effect of $\{X, M\}$ on Y is identifiable.

$$P(y \mid do(x, m), w)$$

WHEN CAN WE IDENTIFY MEDIATED EFFECTS?



WHEN CAN WE IDENTIFY MEDIATED EFFECTS?



SUMMARY OF RESULTS ON MEDIATION

SUMMARY OF RESULTS ON MEDIATION

- Ignorability is not required for identifying natural effects

SUMMARY OF RESULTS ON MEDIATION

- Ignorability is not required for identifying natural effects
- The nonparametric estimability of natural (and controlled) direct and indirect effects can be determined in polynomial time given any causal graph G with both measured and unmeasured variables.

SUMMARY OF RESULTS ON MEDIATION

- Ignorability is not required for identifying natural effects
- The nonparametric estimability of natural (and controlled) direct and indirect effects can be determined in polynomial time given any causal graph G with both measured and unmeasured variables.
- If NDE (or NIE) is estimable, then its estimand can be derived in polynomial time.

SUMMARY OF RESULTS ON MEDIATION

- Ignorability is not required for identifying natural effects
- The nonparametric estimability of natural (and controlled) direct and indirect effects can be determined in polynomial time given any causal graph G with both measured and unmeasured variables.
- If NDE (or NIE) is estimable, then its estimand can be derived in polynomial time.
- The algorithm is complete and was extended to any path-specific effect by Shpitser (2013).

OUTLINE

Concepts:

* Causal inference – a paradigm shift
* The two fundamental laws

Basic tools:

* Graph separation
* The truncated product formula
* The back-door adjustment formula
* The do-calculus

Capabilities:

* Policy evaluation
* Transportability
* Mediation
* Missing Data

MISSING DATA:
A CAUSAL INFERENCE PERSPECTIVE

(Mohan, Pearl & Tian 2013)

MISSING DATA: A CAUSAL INFERENCE PERSPECTIVE (Mohan, Pearl & Tian 2013)

- Pervasive in every experimental science.
- Huge literature, powerful software industry, deeply entrenched culture.

MISSING DATA: A CAUSAL INFERENCE PERSPECTIVE
(Mohan, Pearl & Tian 2013)

- Pervasive in every experimental science.
- Huge literature, powerful software industry, deeply entrenched culture.
- Current practices are based on statistical characterization (Rubin, 1976) of a problem that is inherently causal.

MISSING DATA: A CAUSAL INFERENCE PERSPECTIVE
(Mohan, Pearl & Tian 2013)

- Pervasive in every experimental science.
- Huge literature, powerful software industry, deeply entrenched culture.
- Current practices are based on statistical characterization (Rubin, 1976) of a problem that is inherently causal.
- Needed: (1) theoretical guidance, (2) performance guarantees, and (3) tests of assumptions.

WHAT CAN CAUSAL THEORY DO FOR MISSING DATA?

Q-1. What should the world be like, for a given statistical procedure to produce the expected result?

WHAT CAN CAUSAL THEORY DO FOR MISSING DATA?

Q-1. What should the world be like, for a given statistical procedure to produce the expected result?

Q-2. Can we tell from the postulated world whether any method can produce a bias-free result? How?

Q-3. Can we tell from data if the world does not work as postulated?

WHAT CAN CAUSAL THEORY DO FOR MISSING DATA?

Q-1. What should the world be like, for a given statistical procedure to produce the expected result?

Q-2. Can we tell from the postulated world whether any method can produce a bias-free result? How?

Q-3. Can we tell from data if the world does not work as postulated?

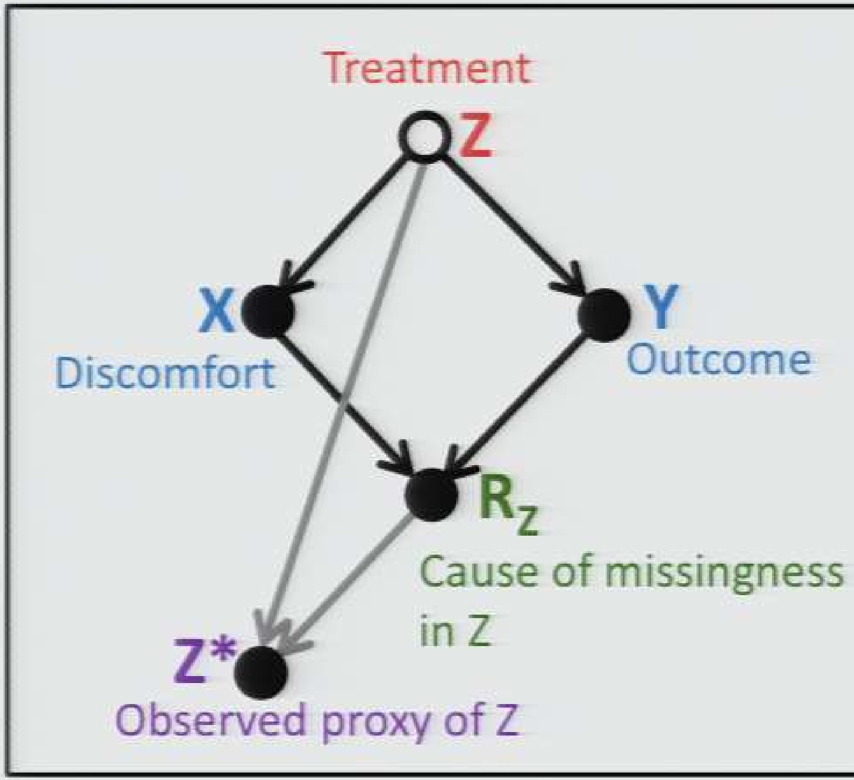
- To answer these questions, we need models of the world, i.e., process models.
- Statistical characterization of the problem is too crude, e.g., MCAR, MAR, MNAR.

Graphical Models for Inference With Missing Data

(From Mohan et al., NIPS-2013)

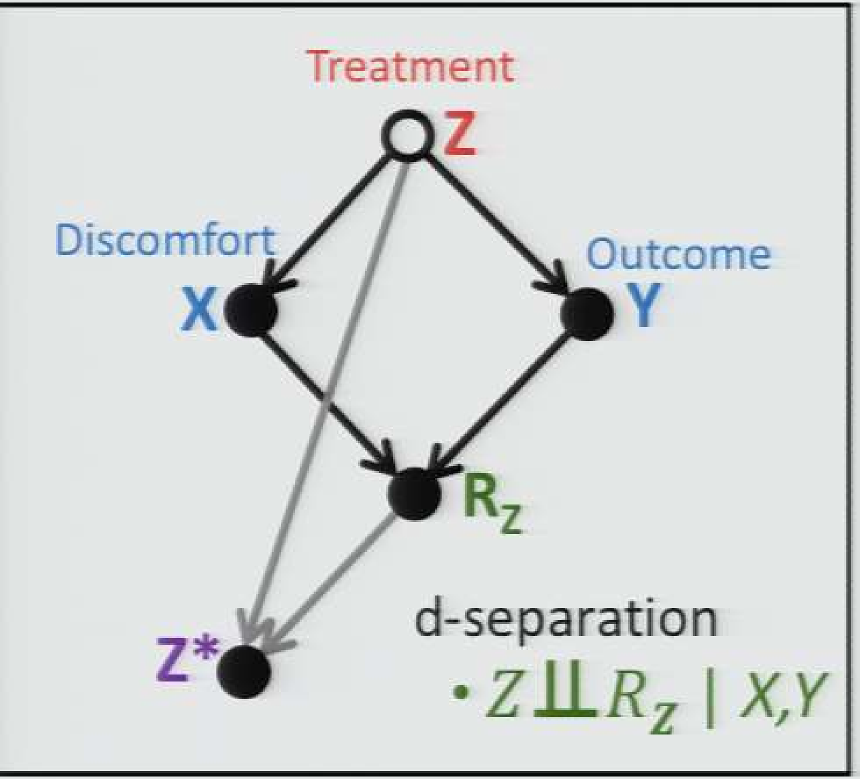
X	Y	Z*	R_Z	$P(Z^*,X,Y,R_Z)$
0	0	0	0	0.01
0	0	1	0	0.21
0	1	0	0	0.01
0	1	1	0	0.04
1	0	0	0	0.02
1	0	1	0	0.20
1	1	0	0	0.05
1	1	1	0	0.08
0	0	m	1	0.01
0	1	m	1	0.02
1	0	m	1	0.30
1	1	m	1	0.05

Distribution with missing values



Graph depicting the missingness process

Recoverability of Query (Q)



Is Q=P(X,Y,Z) recoverable?

$$Q = P(X,Y,Z)$$
$$= P(Z|X,Y)P(X,Y)$$
$$= P(Z|R_Z = 0, X, Y)P(X,Y)$$
$$= P(Z^*|R_Z = 0, X, Y)P(X,Y)$$

WHY GRAPHS?

$$x \longrightarrow y \longrightarrow z \longrightarrow w$$

1. Match the organization of human knowledge

WHY GRAPHS?

$$x \longrightarrow y \longrightarrow z \longrightarrow w$$

$$z \perp\!\!\!\perp x \mid y \quad w \perp\!\!\!\perp xy \mid z \;\Rightarrow\; x \perp\!\!\!\perp wz \mid y$$

1. Match the organization of human knowledge

1a. Guard veracity of assumptions

1b. Assure transparency of assumptions

1c. Assure transparency of their logical ramifications

WHY GRAPHS?

$$x \longrightarrow y \longrightarrow z \longrightarrow w$$

$$z \perp\!\!\!\perp x \mid y \quad w \perp\!\!\!\perp xy \mid z \;\Rightarrow\; x \perp\!\!\!\perp wz \mid y$$

1. Match the organization of human knowledge

1a. Guard veracity of assumptions

1b. Assure transparency of assumptions

1c. Assure transparency of their logical ramifications

2. Blueprints for simulation

RECOVERABILITY AND TESTABILITY

Recoverability

Given a missingness model G and data D, when is a quantity Q estimable from D without bias?

RECOVERABILITY AND TESTABILITY

Recoverability

Given a missingness model G and data D, when is a quantity Q estimable from D without bias?

Non-recoverability

Theoretical impediment to any estimation strategy

RECOVERABILITY AND TESTABILITY

Recoverability
Given a missingness model G and data D, when is a quantity Q estimable from D without bias?

Non-recoverability
Theoretical impediment to any estimation strategy

Testability
Given a model G, when does it have testable implications (refutable by some partially-observed data D')?

RECOVERABILITY AND TESTABILITY

Recoverability

Given a missingness model G and data D, when is a quantity Q estimable from D without bias?

Non-recoverability

Theoretical impediment to any estimation strategy

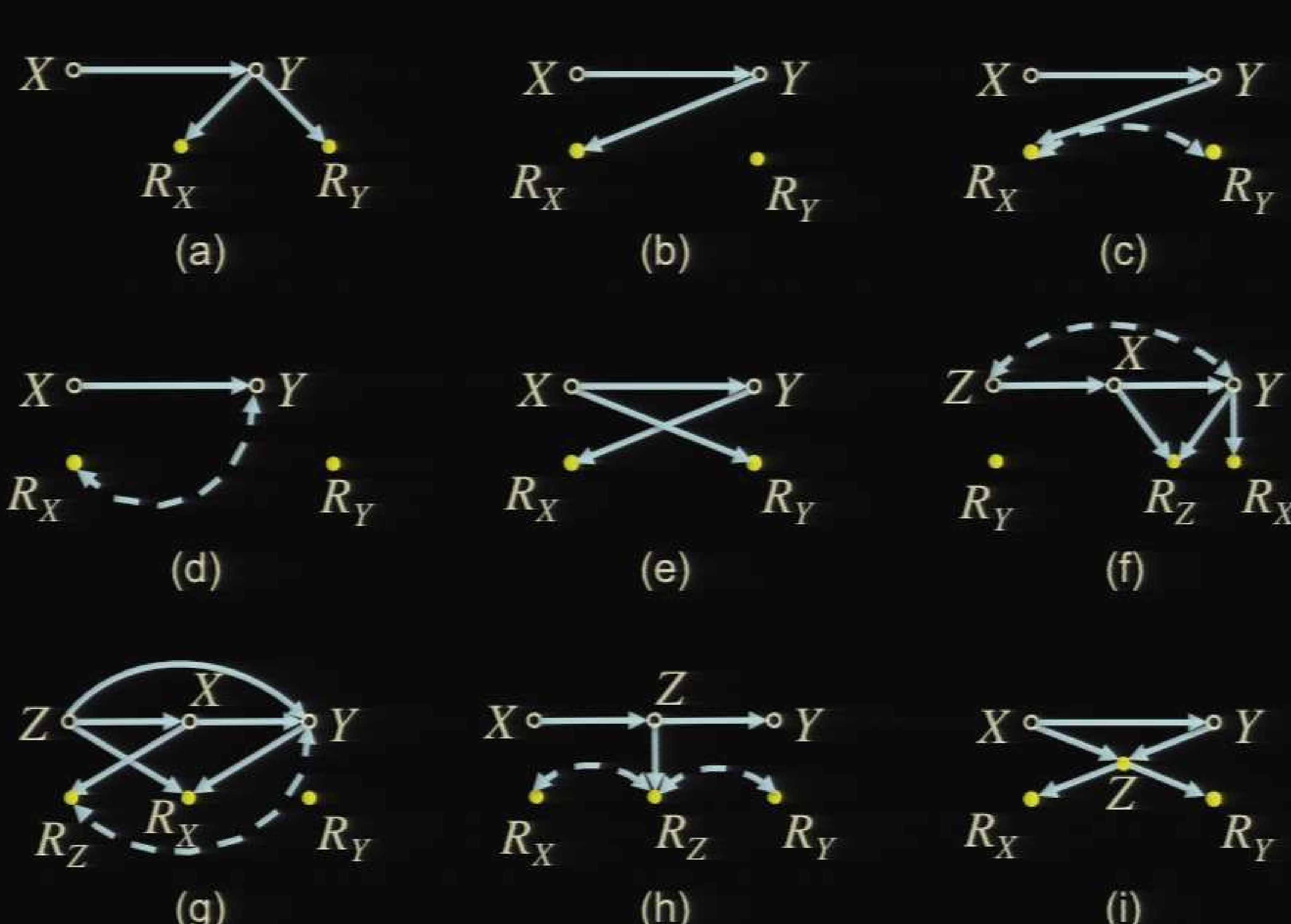
Testability

Given a model G, when does it have testable implications (refutable by some partially-observed data D')?

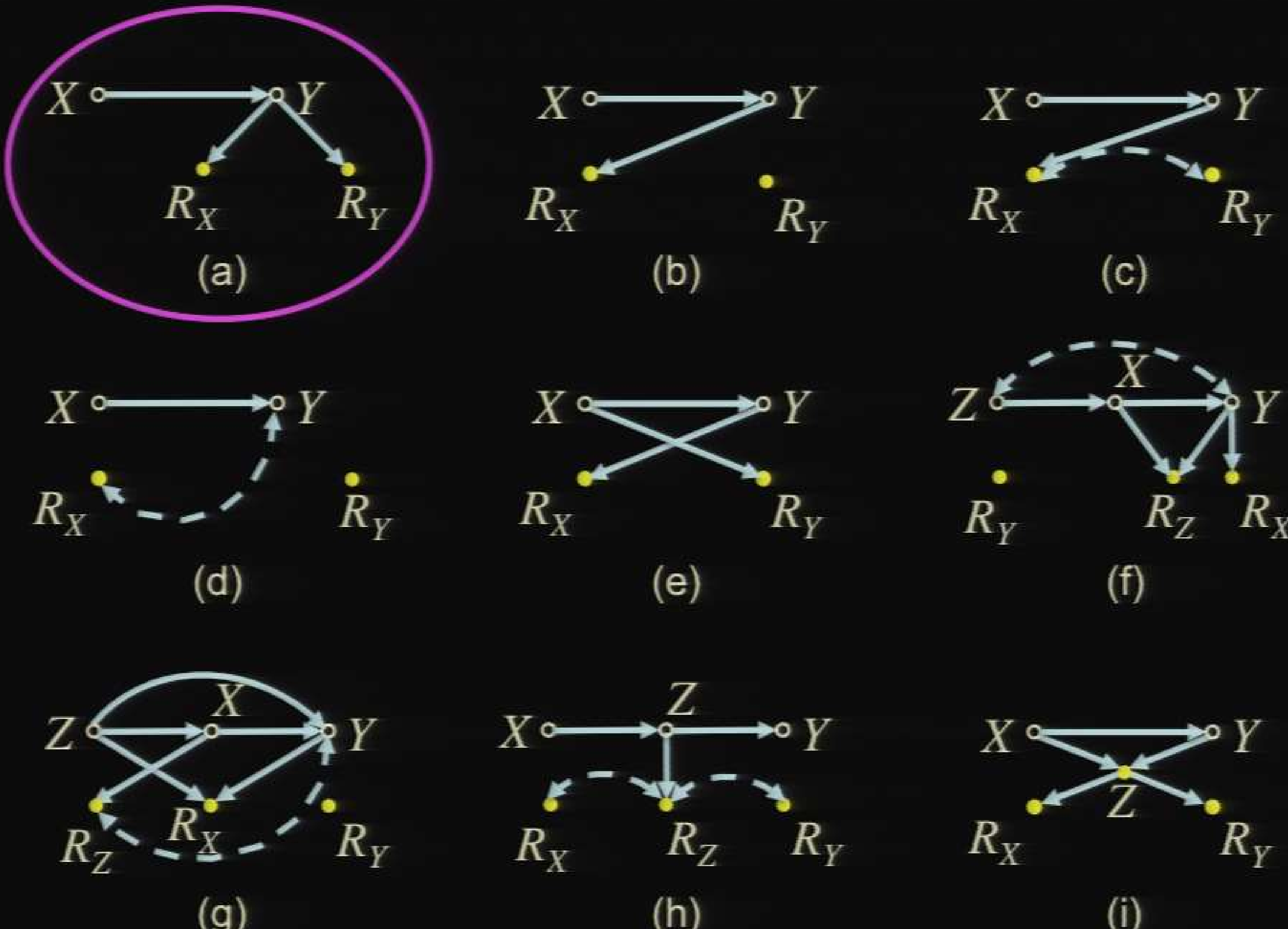
What is known about Recoverability and Testability?

MCAR	recoverable	almost testable
MAR	recoverable	uncharted
MNAR	uncharted	uncharted

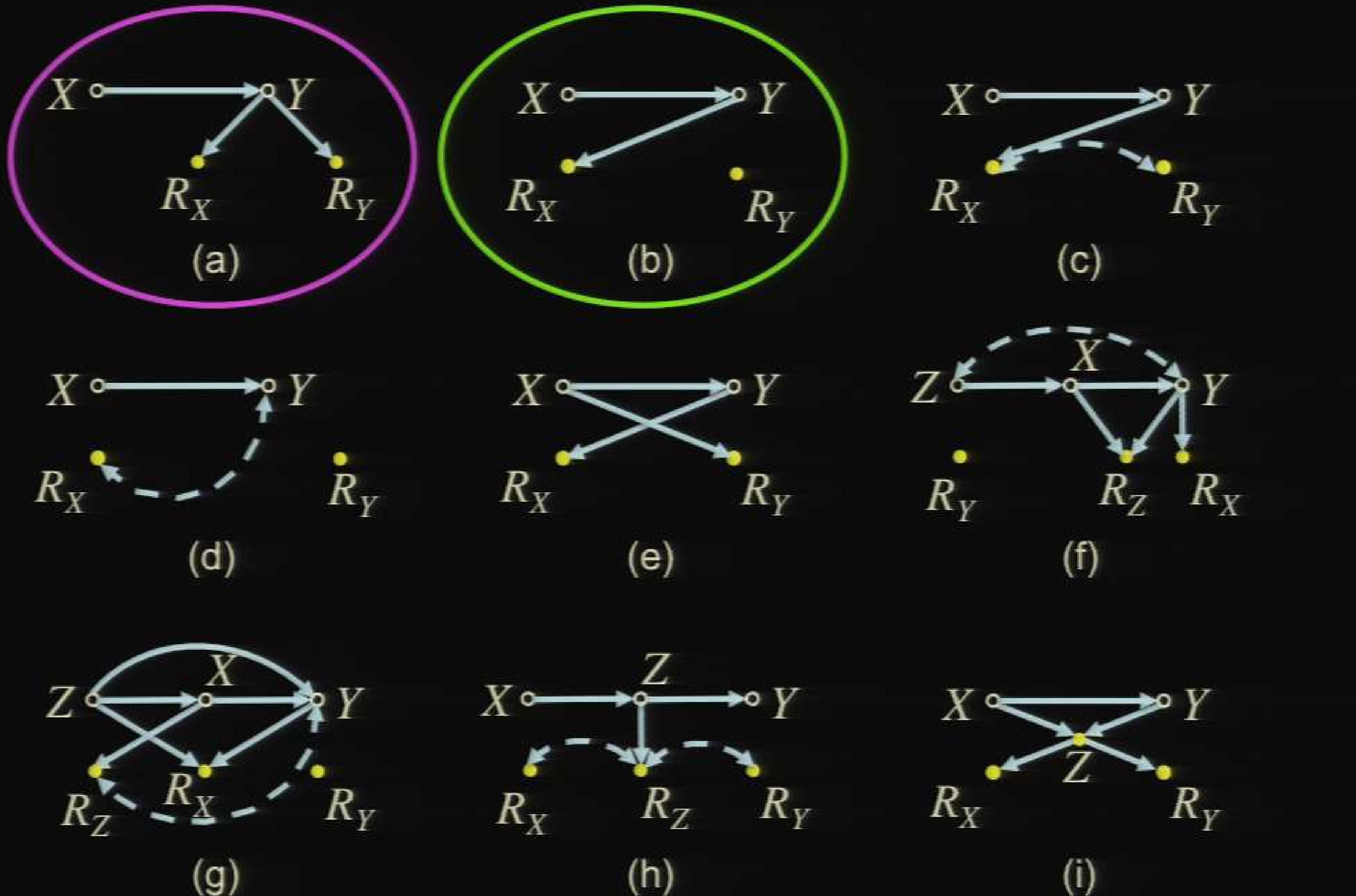
IS $P(X,Y)$ RECOVERABLE?



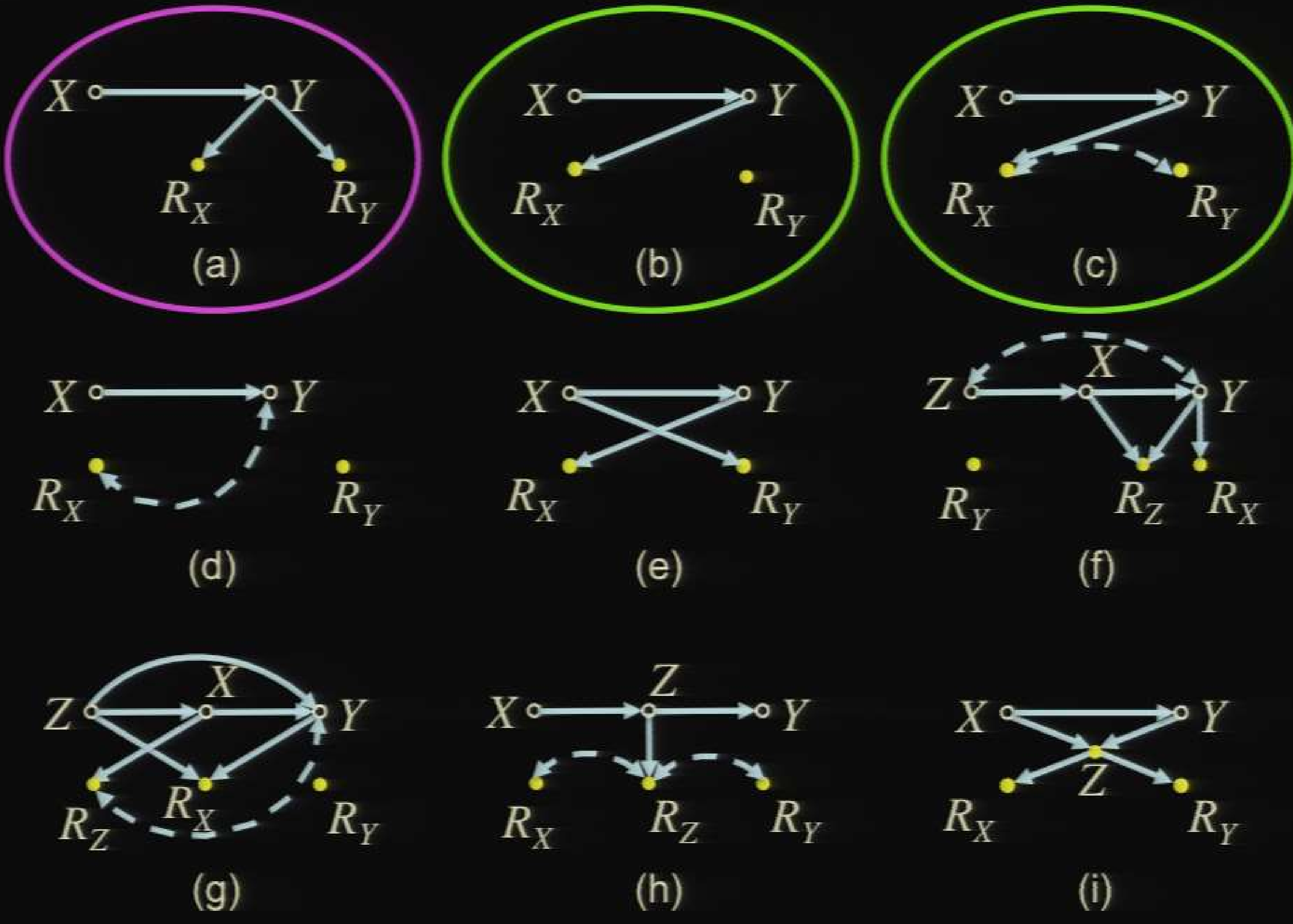
IS $P(X,Y)$ RECOVERABLE?



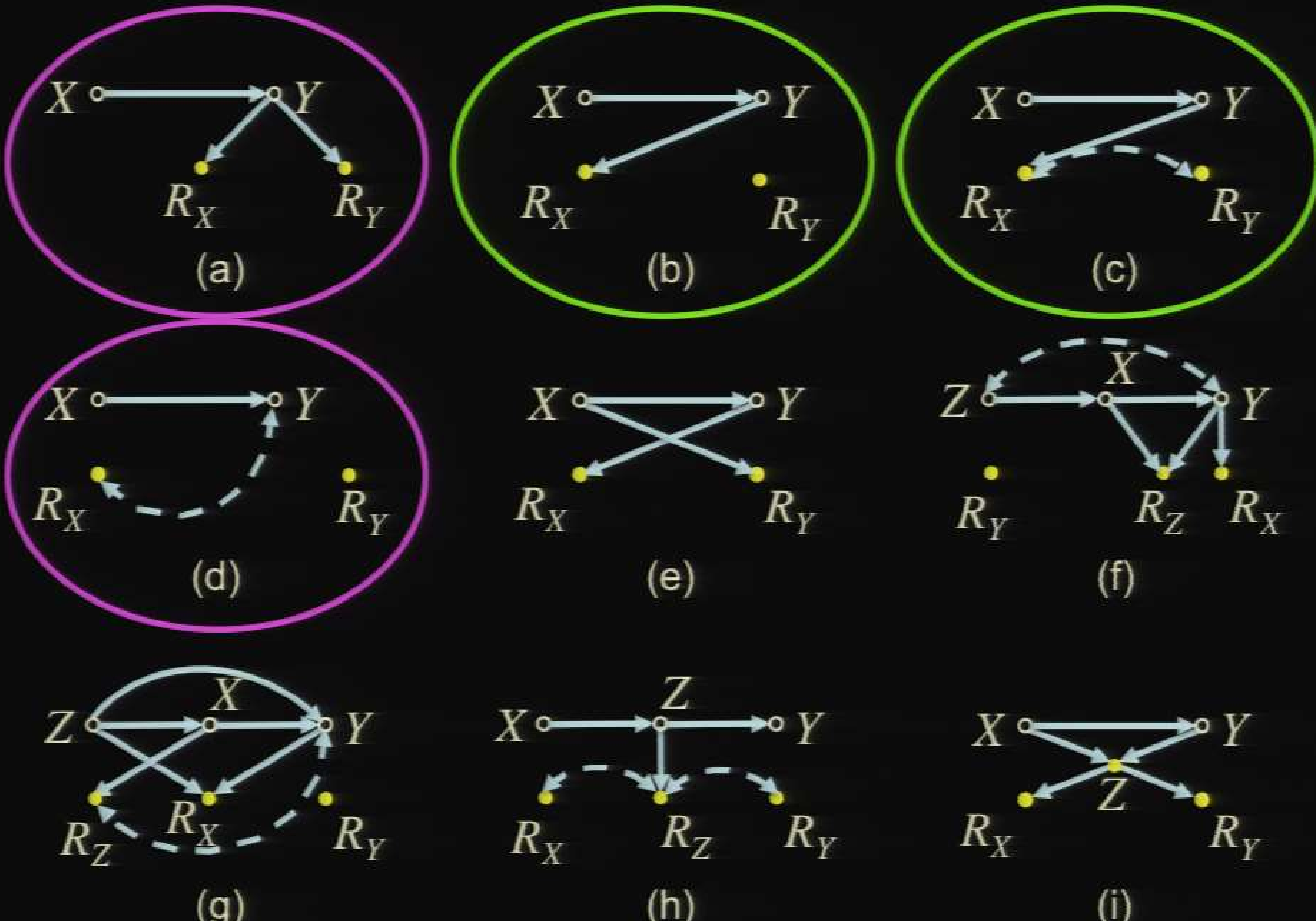
IS $P(X,Y)$ RECOVERABLE?



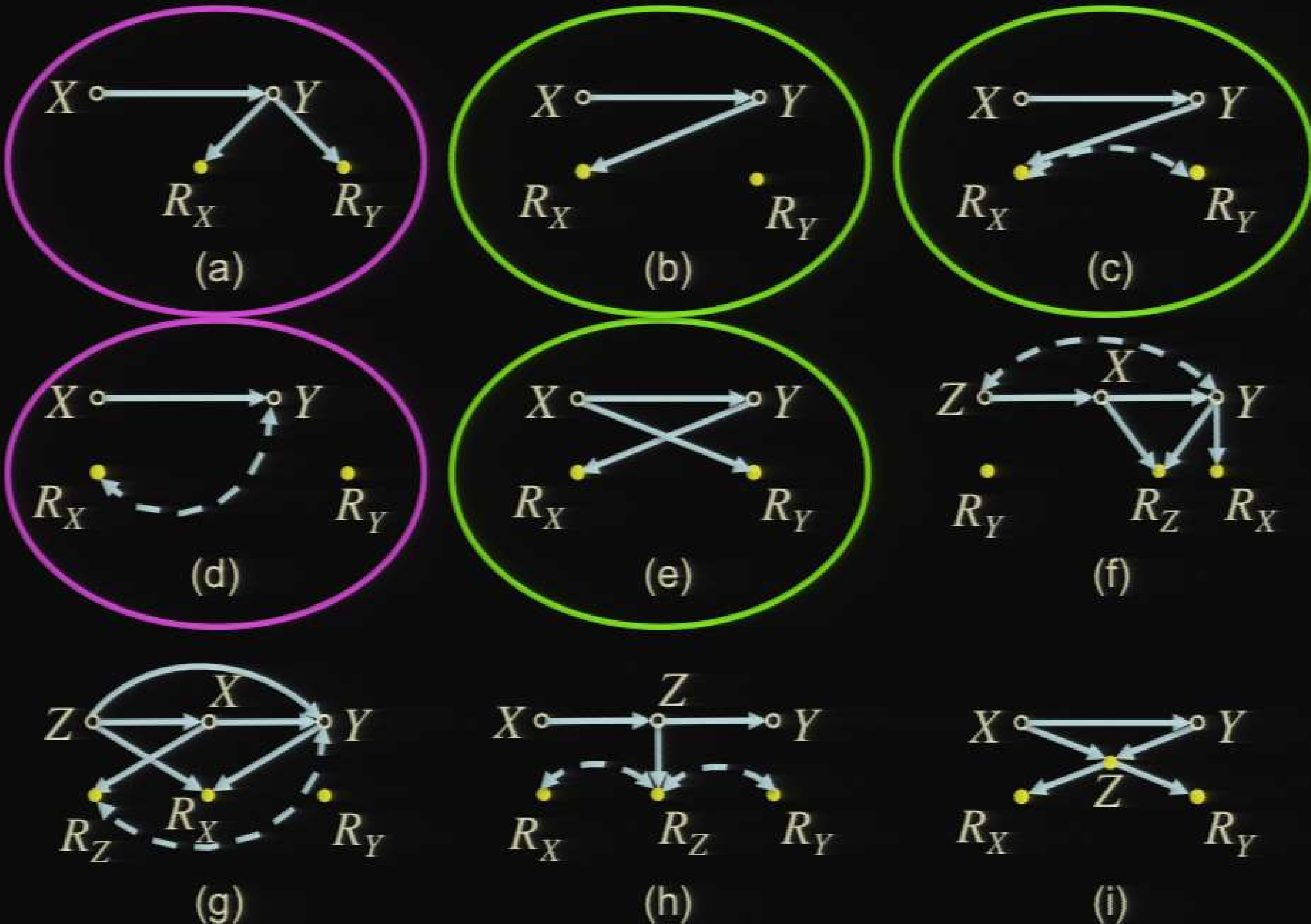
IS $P(X,Y)$ RECOVERABLE?



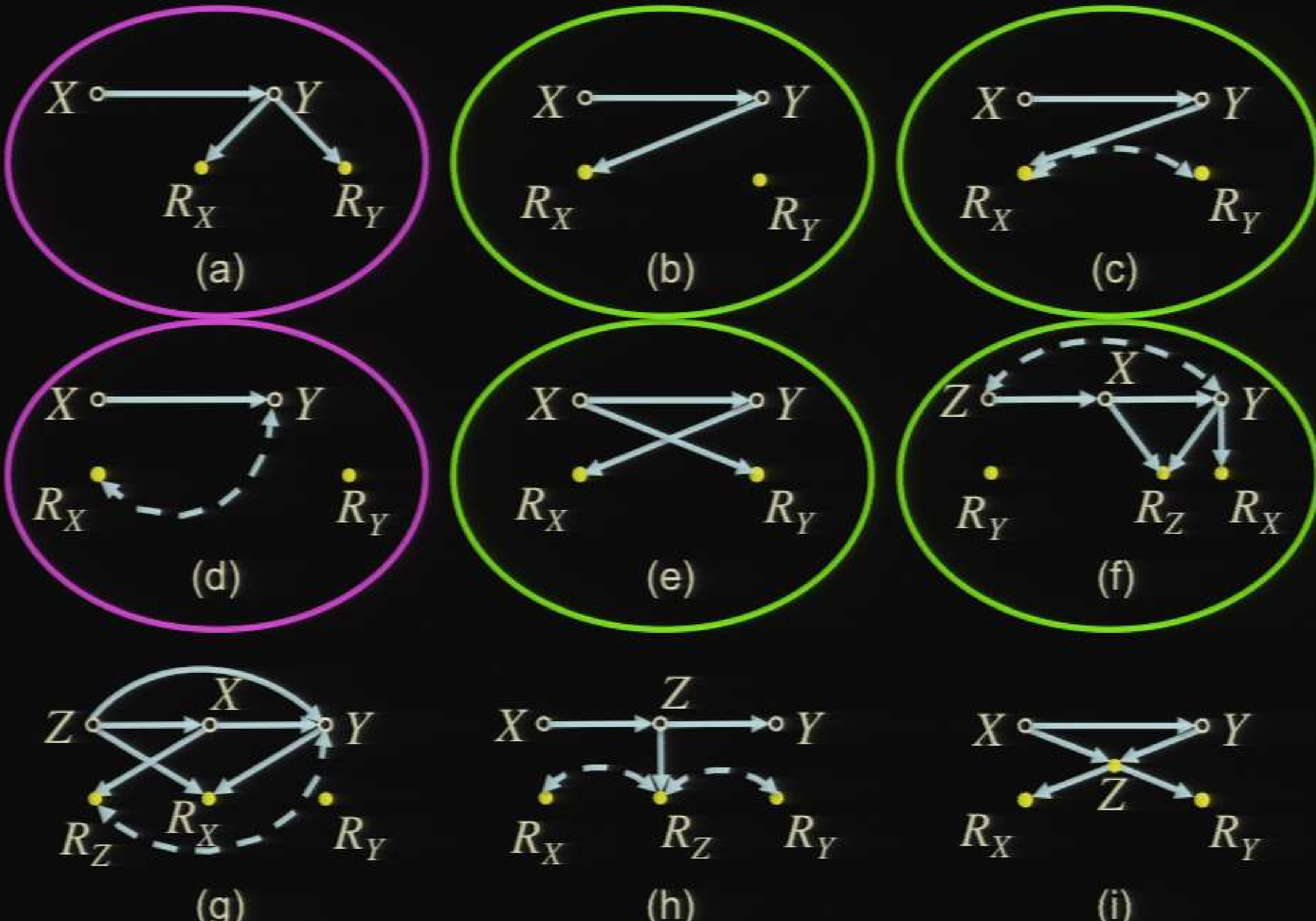
IS $P(X,Y)$ RECOVERABLE?



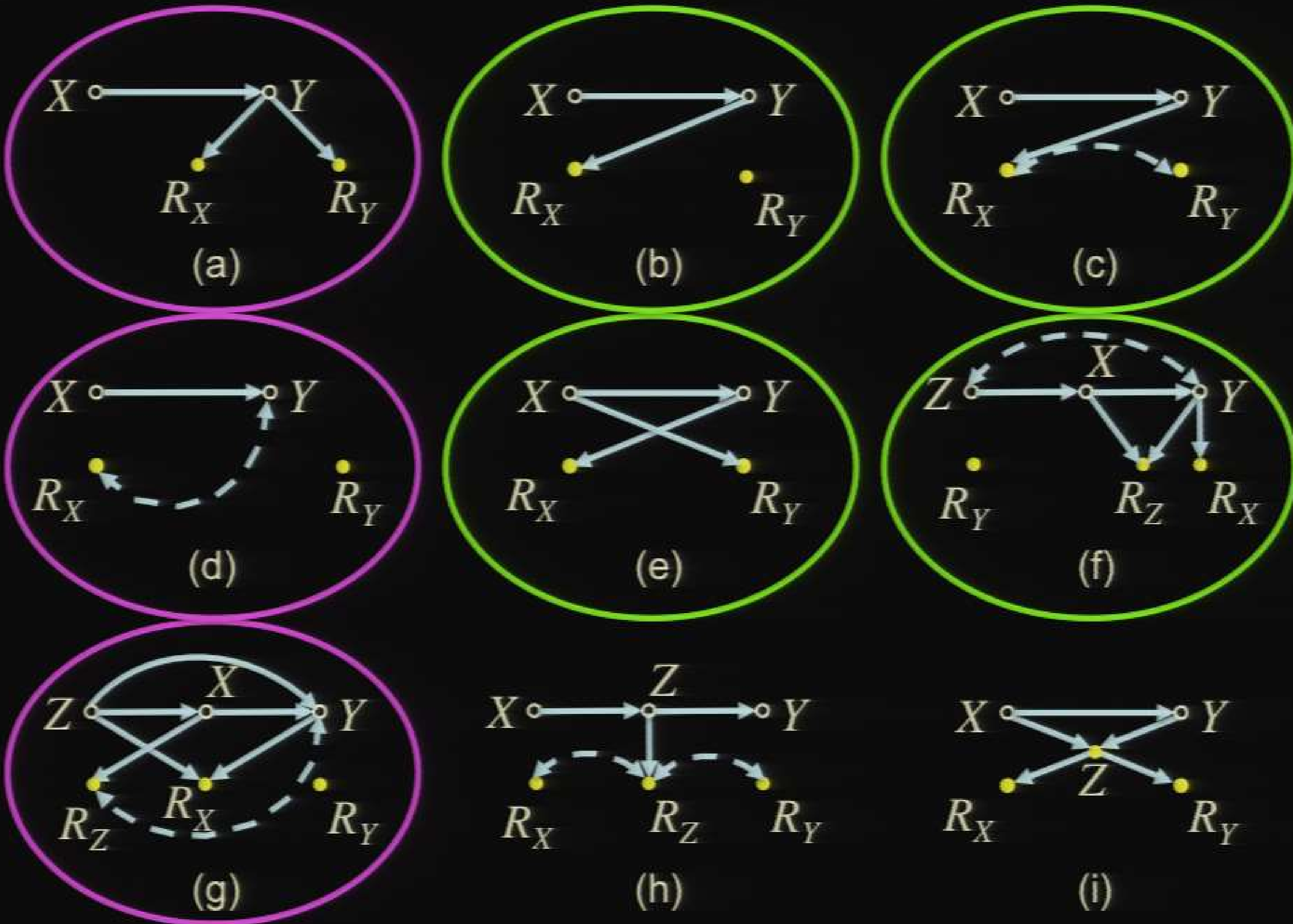
IS $P(X,Y)$ RECOVERABLE?



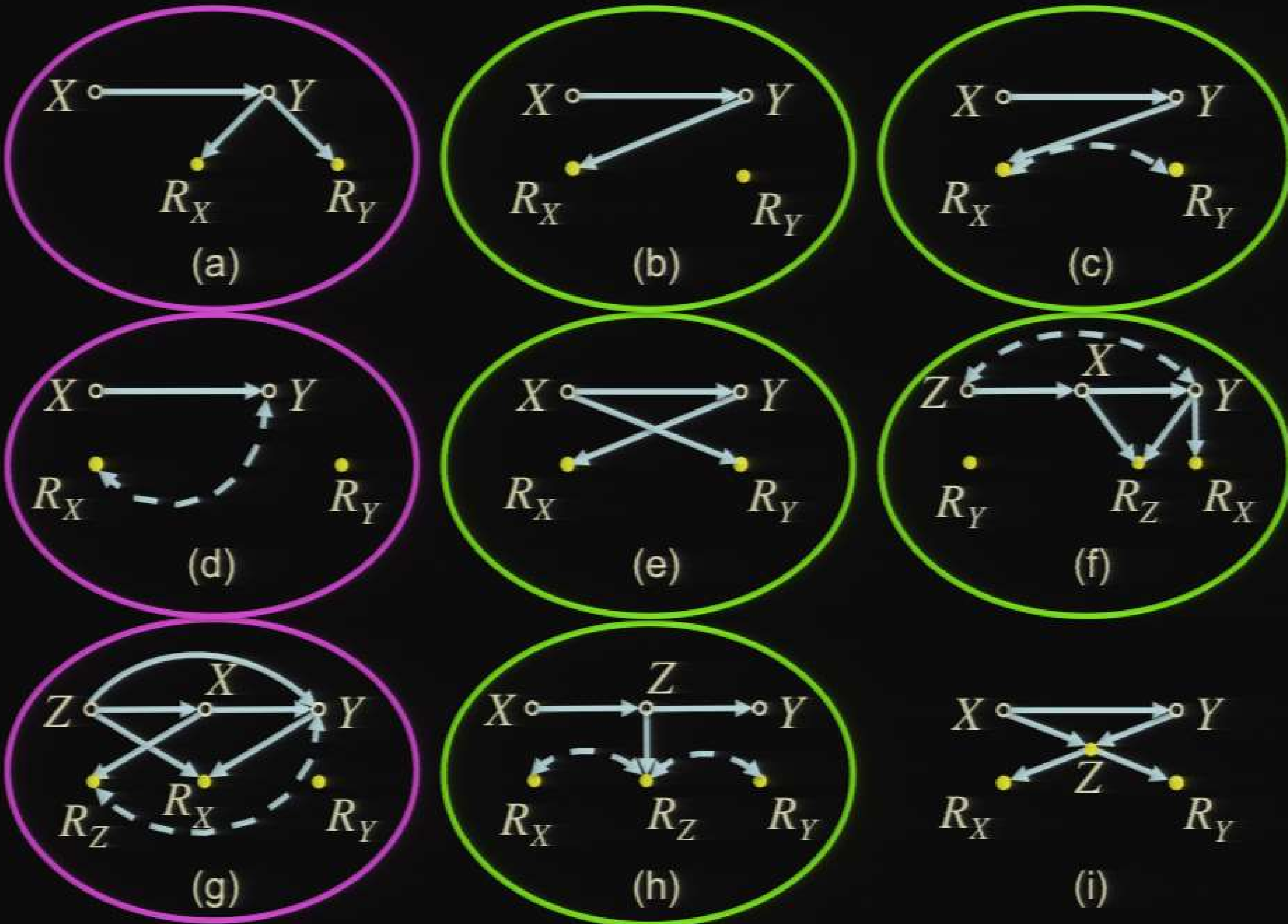
IS $P(X,Y)$ RECOVERABLE?



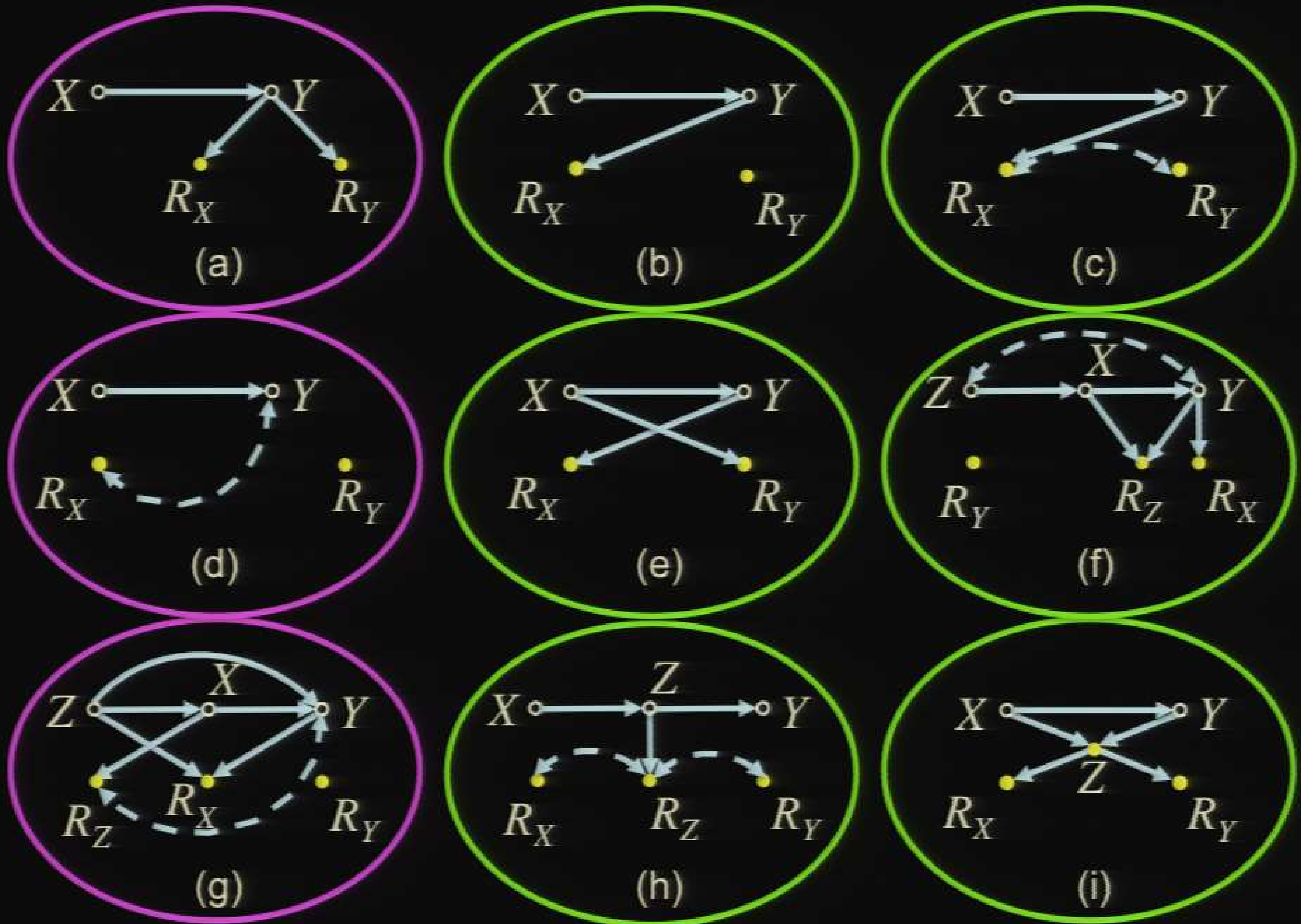
IS $P(X,Y)$ RECOVERABLE?



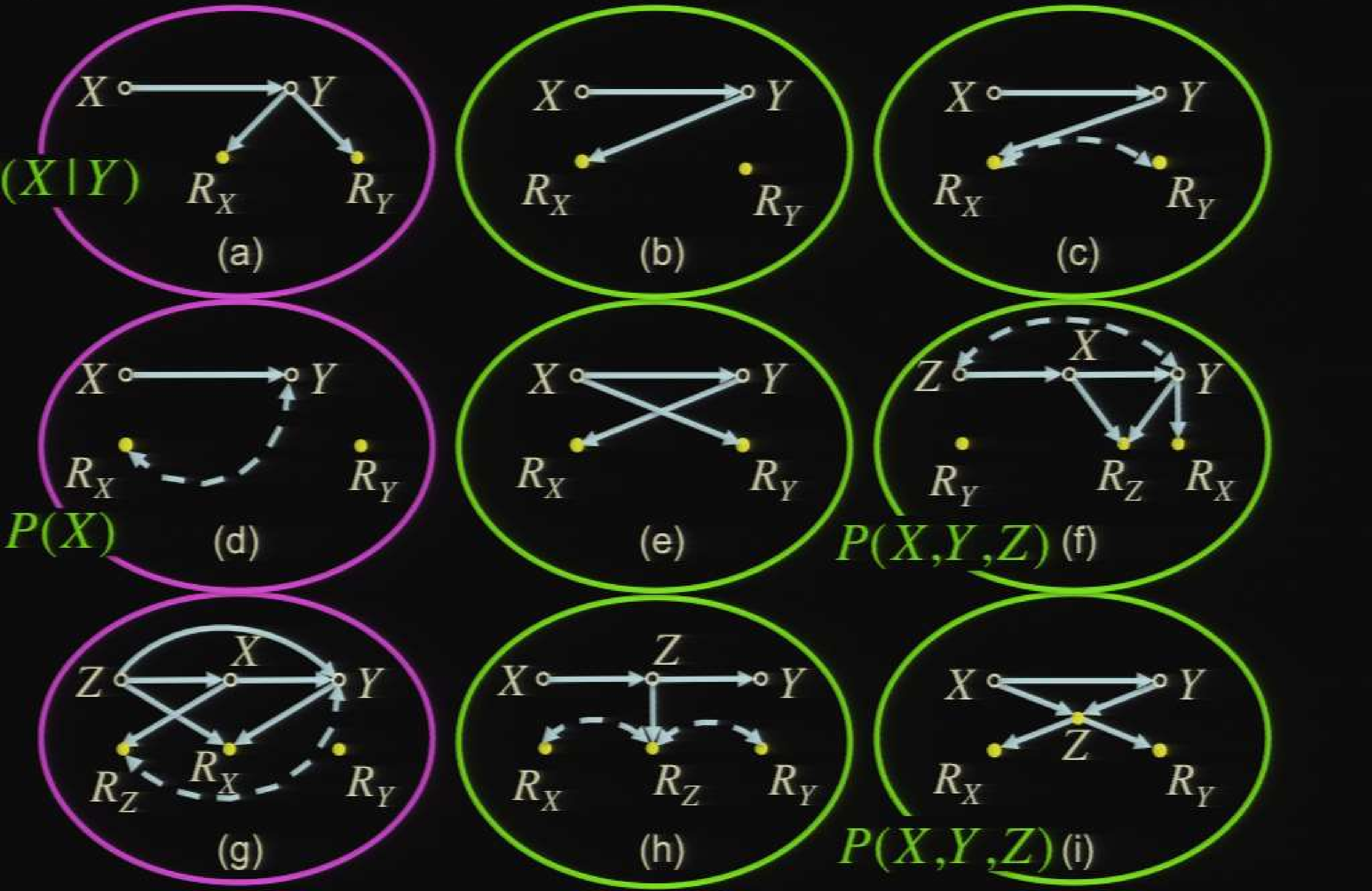
IS $P(X,Y)$ RECOVERABLE?



IS $P(X,Y)$ RECOVERABLE?



IS $P(X,Y)$ RECOVERABLE?



WHAT IF WE DON'T HAVE THE GRAPH?

WHAT IF WE DON'T HAVE THE GRAPH?

1. Constructing the graph requires less knowledge than deciding whether a problem lies in MCAR, MAR or MNAR.

WHAT IF WE DON'T HAVE THE GRAPH?

1. Constructing the graph requires less knowledge than deciding whether a problem lies in MCAR, MAR or MNAR.
2. Understanding what the world should be like for a given procedure to work is a precondition for deciding when model's details are not necessary.

(no universal estimator)

WHAT IF WE DON'T HAVE THE GRAPH?

1. Constructing the graph requires less knowledge than deciding whether a problem lies in MCAR, MAR or MNAR.
2. Understanding what the world should be like for a given procedure to work is a precondition for deciding when model's details are not necessary.
   (no universal estimator)
3. Knowing whether non-convergence is due to theoretical impediment or local optima, is extremely useful.

WHAT IF WE DON'T HAVE THE GRAPH?

1. Constructing the graph requires less knowledge than deciding whether a problem lies in MCAR, MAR or MNAR.
2. Understanding what the world should be like for a given procedure to work is a precondition for deciding when model's details are not necessary.
   (no universal estimator)
3. Knowing whether non-convergence is due to theoretical impediment or local optima, is extremely useful.
4. Graphs unveil when a model is testable.

CONCLUSIONS

CONCLUSIONS

1. Think nature, not data, not even experiment.

CONCLUSIONS

1. Think nature, not data, not even experiment.
2. Think hard, but only once – the rest is mechanizable.

CONCLUSIONS

1. Think nature, not data, not even experiment.
2. Think hard, but only once – the rest is mechanizable.
3. Speak a language in which the veracity of each assumption can be judged by users, and which tells you whether any of those assumptions can be refuted by data.

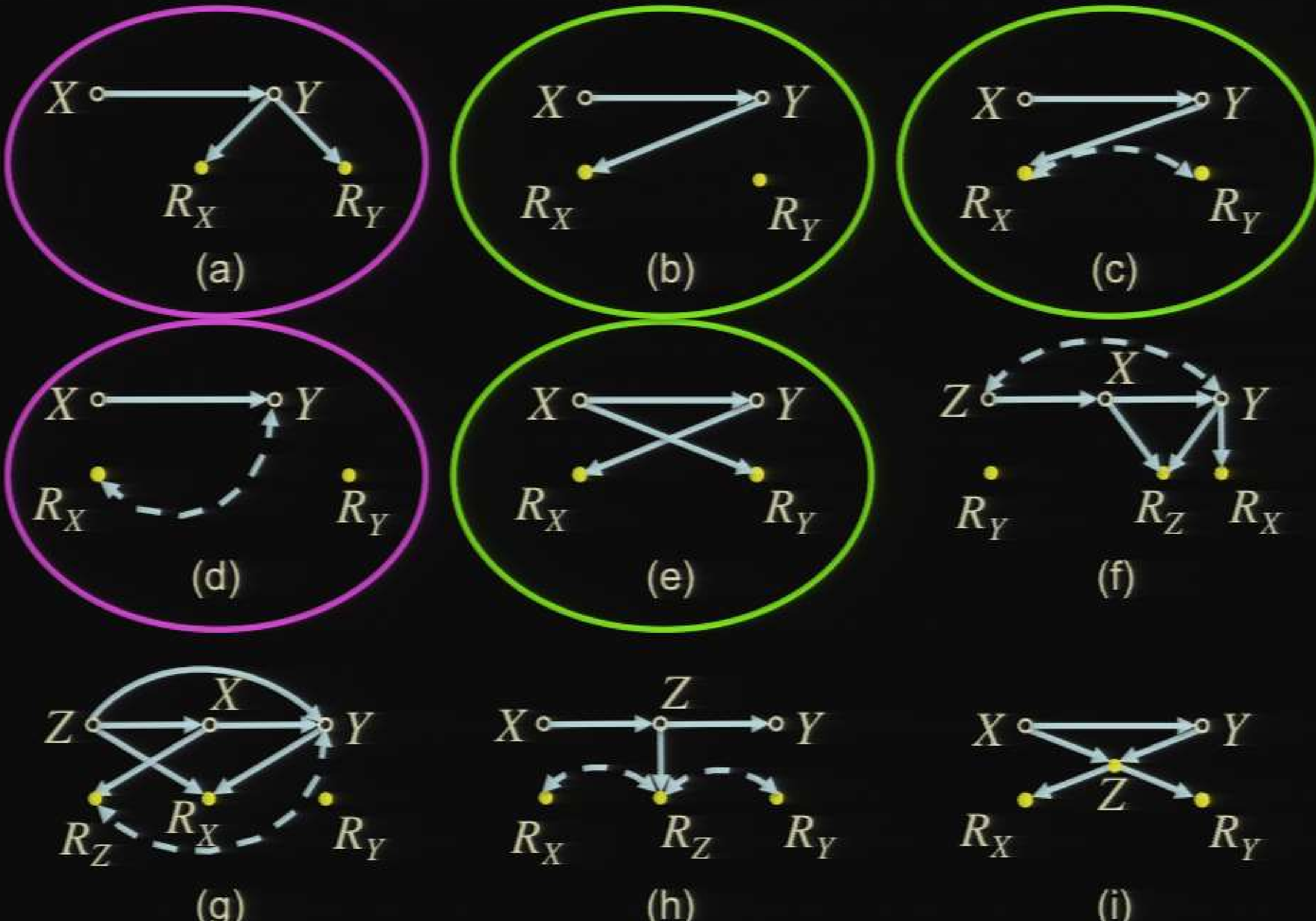
Thank you

CONCLUSIONS

WHAT IF WE DON'T HAVE THE GRAPH?

1. Constructing the graph requires less knowledge than deciding whether a problem lies in MCAR, MAR or MNAR.

IS $P(X,Y)$ RECOVERABLE?

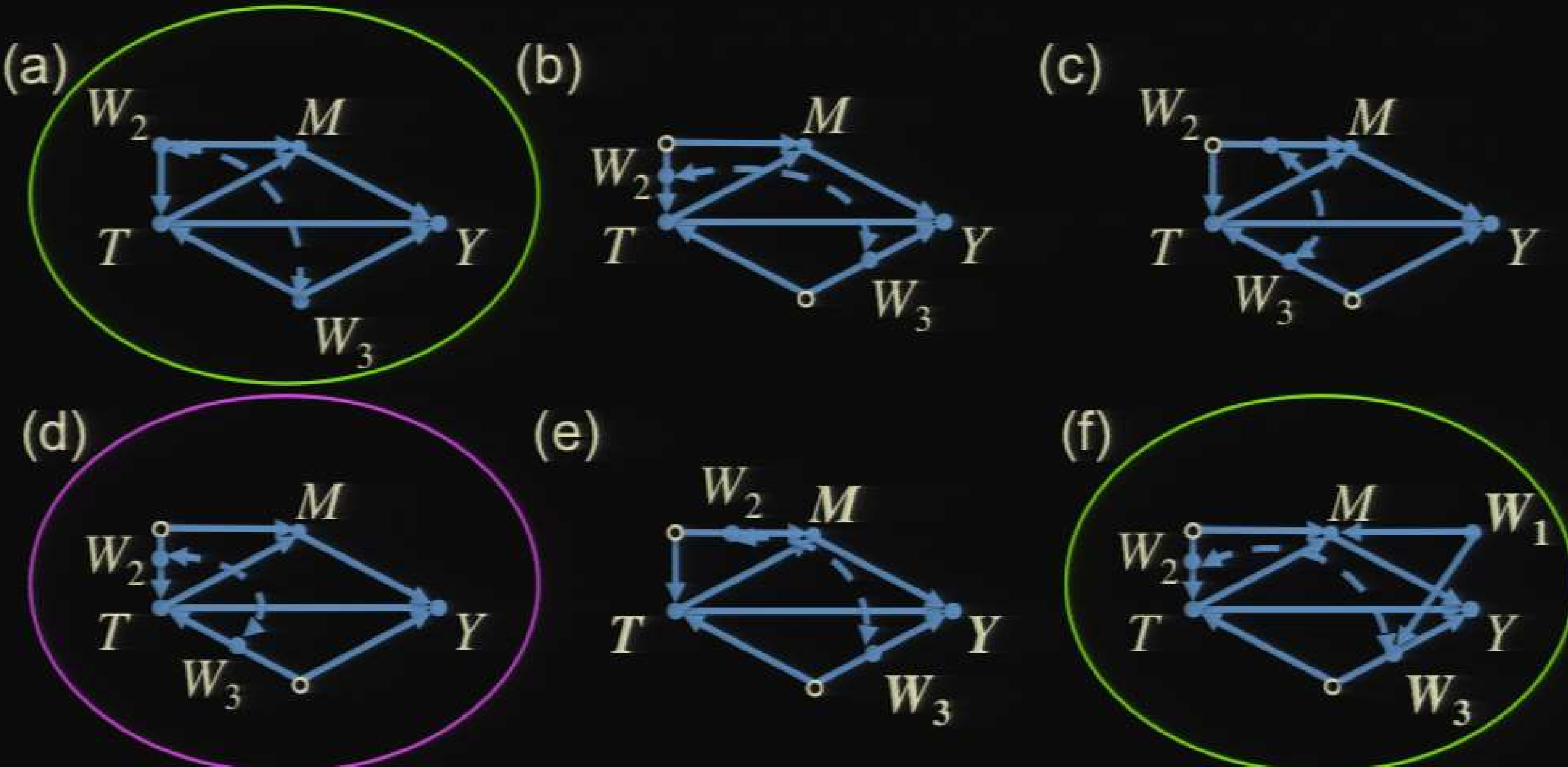


WHY GRAPHS?

$$x \longrightarrow y \longrightarrow z \longrightarrow w$$

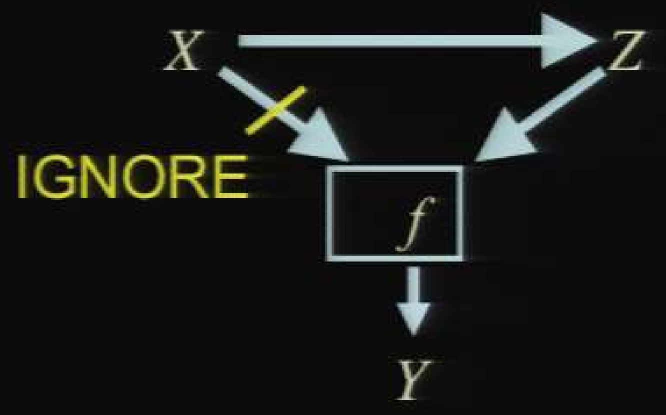
1. Match the organization of human knowledge

WHEN CAN WE IDENTIFY MEDIATED EFFECTS?



POLICY IMPLICATIONS OF INDIRECT EFFECTS

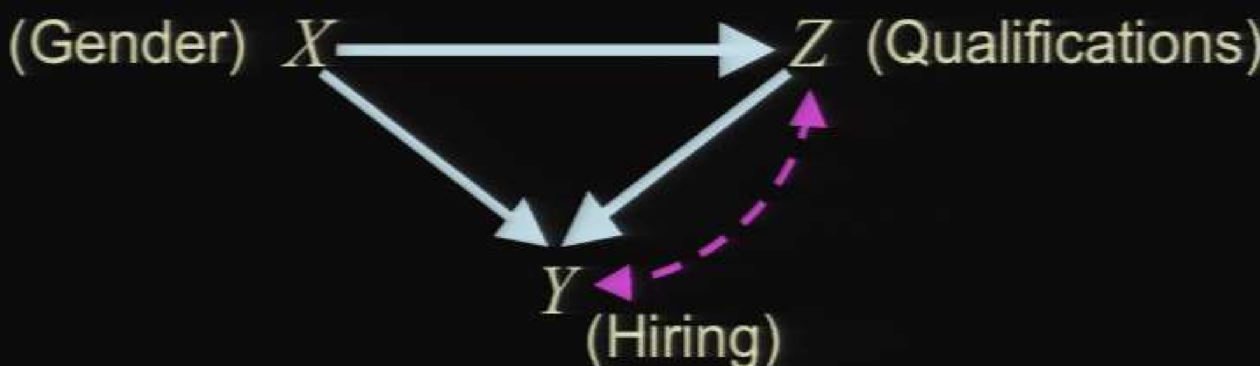
What is the indirect effect of X on Y?



Deactivating a link – a new type of intervention

LEGAL IMPLICATIONS OF DIRECT EFFECT

Can data prove an employer guilty of hiring discrimination?



What is the direct effect of X on Y? (CDE)

$$E(Y|do(x_1),do(z))-E(Y|do(x_0),do(z))$$

Adjust for Z? No! No!

OUTLINE

Concepts:

* Causal inference – a paradigm shift
* The two fundamental laws

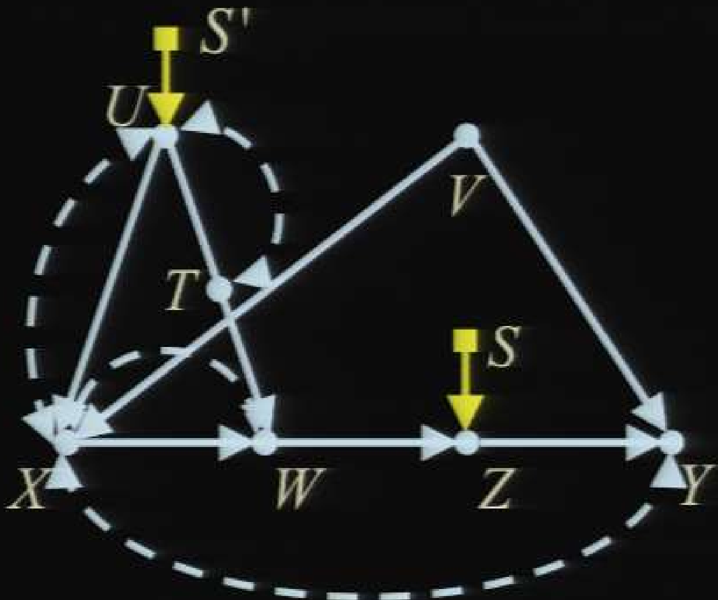
Basic tools:

* Graph separation
* The truncated product formula
* The back-door adjustment formula
* The do-calculus

Capabilities:

* Policy evaluation
* Transportability
* Mediation
* Missing Data

RESULT: ALGORITHM TO DETERMINE IF AN EFFECT IS TRANSPORTABLE



INPUT: Annotated Causal Graph

S → Factors creating differences

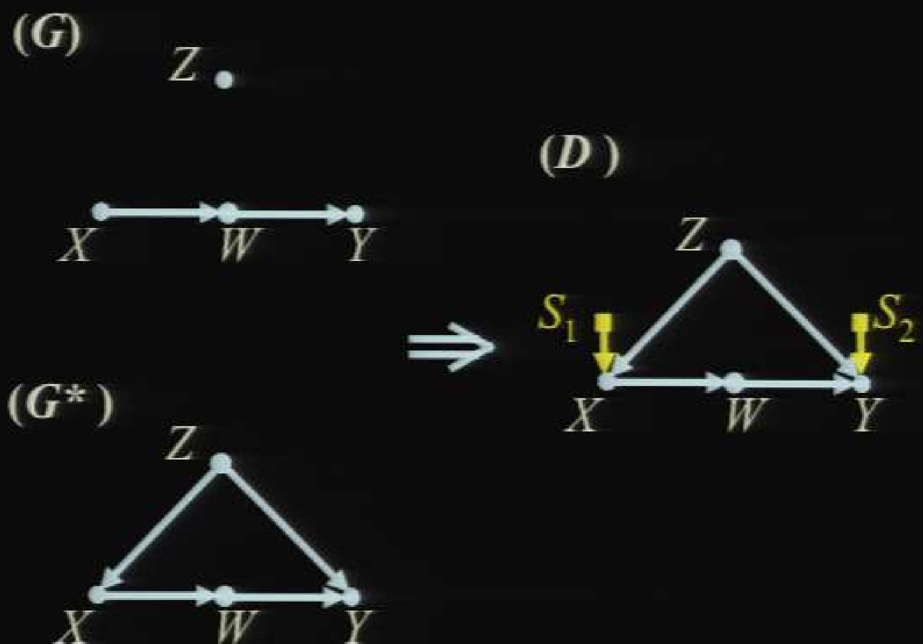
OUTPUT:

1. Transportable or not?
2. Measurements to be taken in the experimental study
3. Measurements to be taken in the target population
4. A transport formula

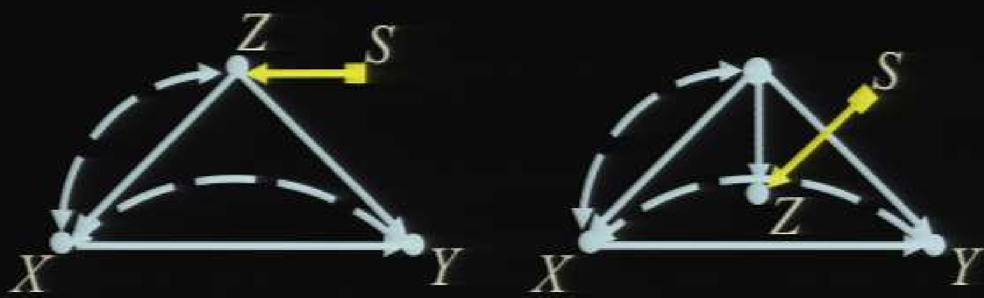
$$P^*(y \mid do(x)) = \sum_z P(y \mid do(x), z) \sum_w P^*(z \mid w) \sum_t P(w \mid do(w), t) P^*(t)$$

SEMANTICS FOR TRANSPORTABILITY SELECTION DIAGRAMS

- How to encode disparities and commonalities about domains?



TRANSPORT FORMULAS DEPEND ON THE CAUSAL STORY



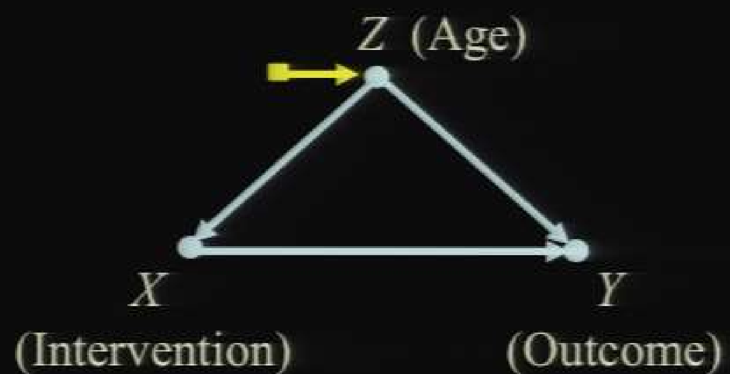
(a) (b)

a) $\boldsymbol{Z}$ represents age

$$P^*(y \mid do(x)) = \sum_z P(y \mid do(x), z) P^*(z)$$

MOTIVATION

WHAT CAN EXPERIMENTS IN LA TELL US ABOUT NYC?



R: Π (LA) $\longrightarrow$ Π^* (NY)

Experimental study in LA

Measured: $P(x, y, z)$

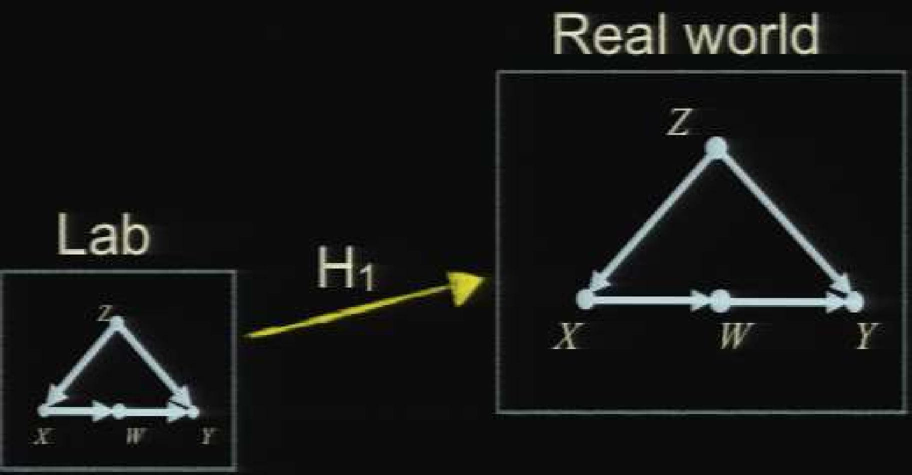
$P(y \mid do(x), z)$

Observational study in NYC

Measured: $P^*(x, y, z)$

$P^*(z) \neq P(z)$

MOVING FROM THE "LAB" TO THE "REAL WORLD" ...



SUMMARY OF POLICY EVALUATION RESULTS

- The estimability of any expression of the form

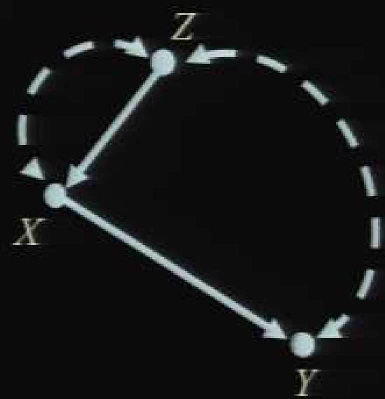
  $$Q = P(y_1, y_2, \ldots, y_n \mid do(x_1, x_2, \ldots, x_m), z_1, z_2, \ldots, z_k)$$

  can be determined given any causal graph G containing measured and unmeasured variables.

- If Q is estimable, then its estimand can be derived in polynomial time (by estimable we mean either from observational or from experimental studies.)

WHAT CAN EXPERIMENTS ON DIET REVEAL ABOUT THE EFFECT OF CHOLESTEROL ON HEART ATTACK?

G:



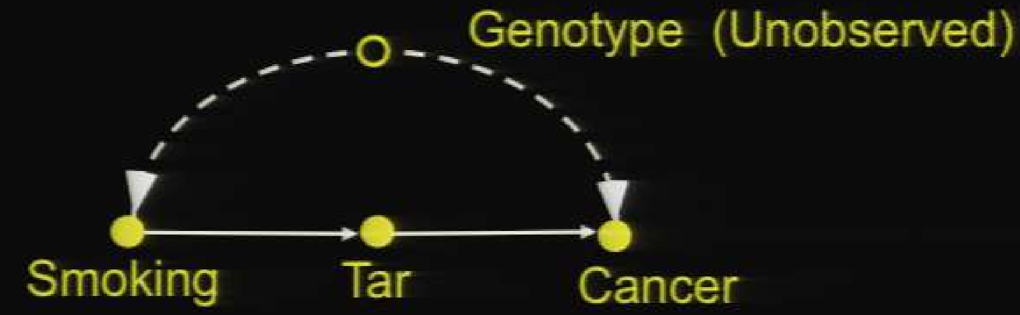
Z: Diet
X: Cholesterol level
Y: Heart Attack

Measured:

Observational study: $P(x, y, z)$
Experimental study: $P(x, y \mid do(z))$

Needed: $Q = P(y \mid do(x)) = ? = \dfrac{P(x, y \mid do(z))}{P(x \mid do(z))}$

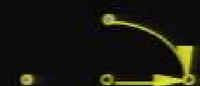
DERIVATION IN CAUSAL CALCULUS



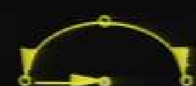
$P(c \mid do(s)) = \Sigma_t P(c \mid do(s), t)\, P(t \mid do(s))$ — Probability Axioms

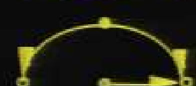
$= \Sigma_t P(c \mid do(s), do(t))\, P(t \mid do(s))$ — Rule 2 

$= \Sigma_t P(c \mid do(s), do(t))\, P(t \mid s)$ — Rule 2 

$= \Sigma_t P(c \mid do(t))\, P(t \mid s)$ — Rule 3 

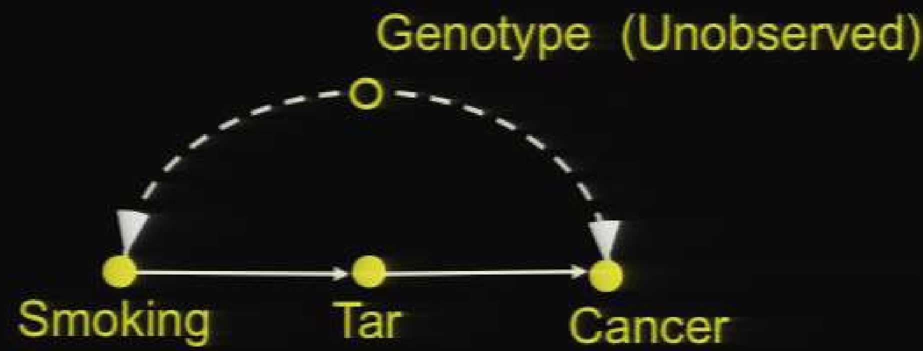
$= \Sigma_{s'} \Sigma_t P(c \mid do(t), s')\, P(s' \mid do(t))\, P(t \mid s)$ — Probability Axioms

$= \Sigma_{s'} \Sigma_t P(c \mid t, s')\, P(s' \mid do(t))\, P(t \mid s)$ — Rule 2 

$= \Sigma_{s'} \Sigma_t P(c \mid t, s')\, P(s')\, P(t \mid s)$ — Rule 3 

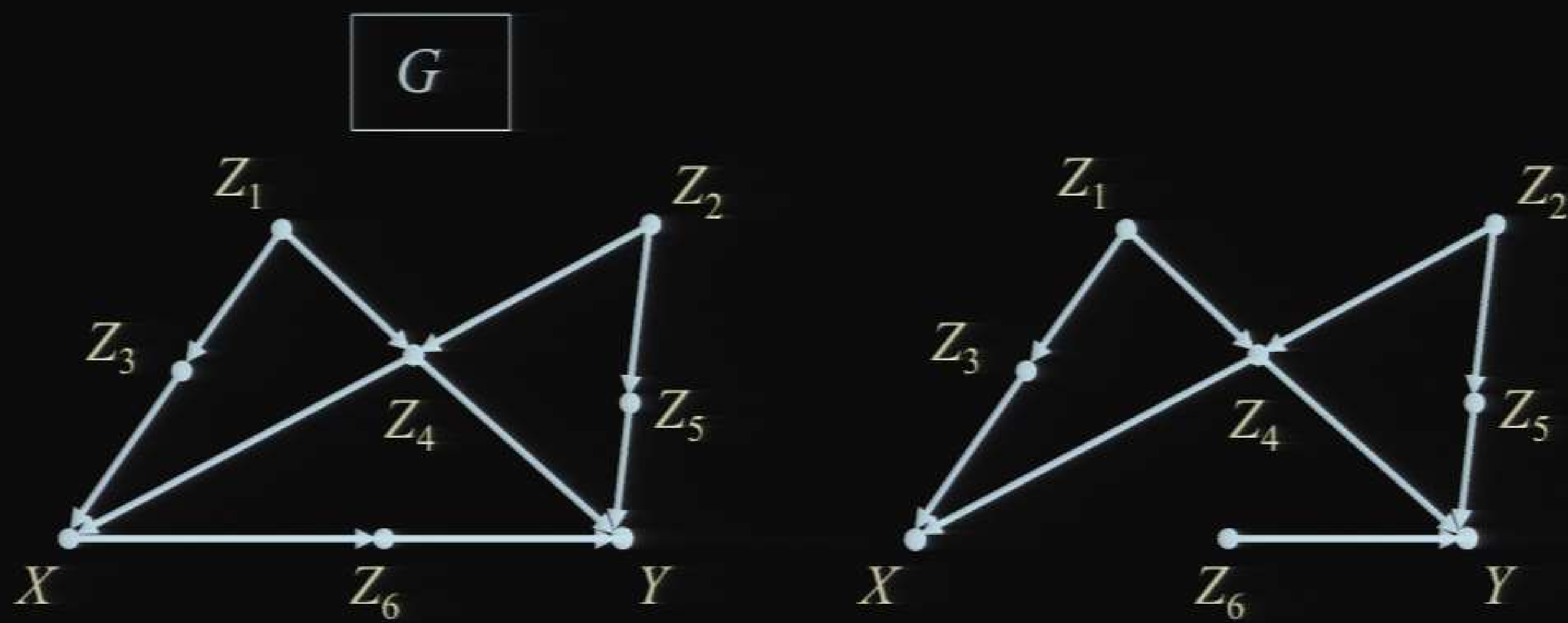
GOING BEYOND ADJUSTMENT

Goal: Find the effect of S on C, $P(c \mid do(s))$, given measurements on auxiliary variable T, and when latent variables confound the relationship S-C.



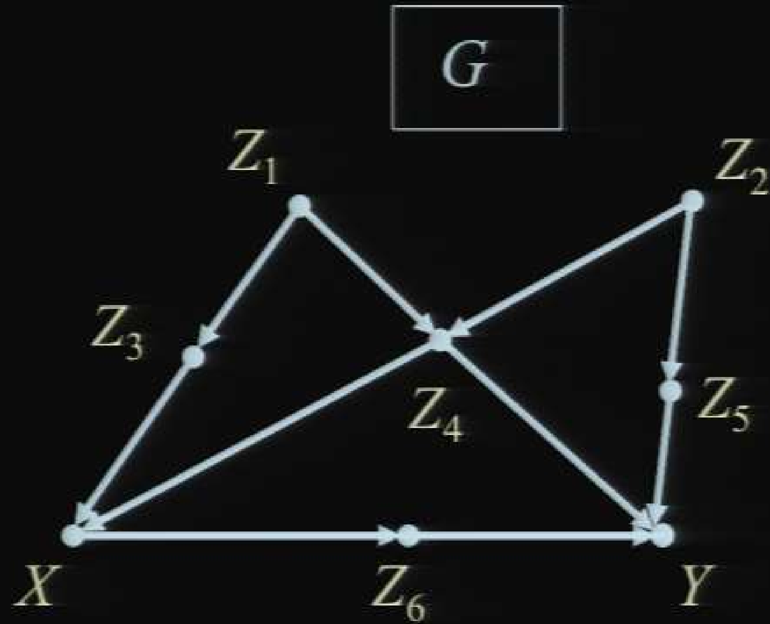
ELIMINATING CONFOUNDING BIAS THE BACK-DOOR CRITERION

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



TOOL 3. BACK-DOOR CRITERION (THE PROBLEM OF CONFOUNDING)

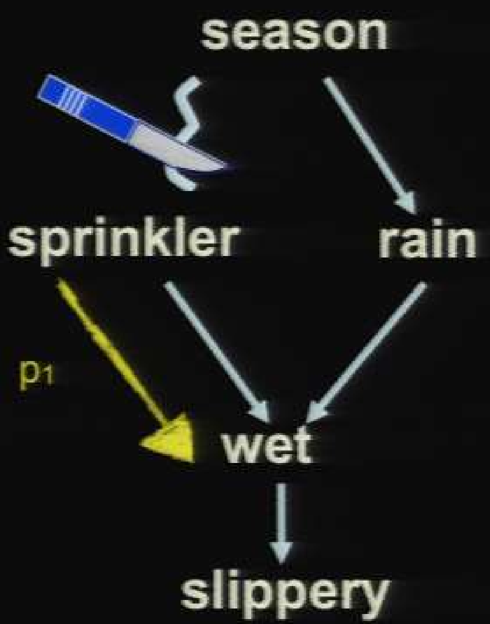
Goal: Find the effect of X on Y, $P(y|do(x))$, given measurements on auxiliary variables $Z_1, \ldots, Z_k$



IF SEASON IS LATENT, IS THE EFFECT STILL COMPUTABLE?

Queries:

$Q_1 = \Pr(\text{wet} \mid \text{Sprinkler} = \text{on})$
$\quad = P(p_1) + P(p_2)$



$Q_2 = \Pr(\text{wet} \mid do(\text{Sprinkler} = \text{on}))$
$\quad = P(p_1)$

$$\sum_{Se,Ra,Sl} P(Se)\, \cancel{P(Sp \mid Se)}\, P(Ra \mid Se)\, P(We \mid Sp, Ra)\, P(Sl \mid We)$$

$$= \sum_{Se} P(We \mid Sp, Se)\, P(Se)$$

Adjustment for direct causes

NO FREE LUNCH: ASSUMPTIONS ENCODED IN CBNs

Definition (Causal Bayesian Network):

$P(v)$: observational distribution
$P(v \mid do(x))$: experimental distribution
P^*: set of all observational and experimental distributions

A DAG G is called a Causal Bayesian Network compatible with P^* if and only if the following three conditions hold for every $P(v \mid do(x)) \in P^*$:

i. $P(v \mid do(x))$ is Markov relative to G;
ii. $P(v_i \mid do(x)) = 1$, for all $V_i \in X$;
iii. $P(v_i \mid \mathrm{pa}_i, do(x)) = P(v_i \mid pa_i)$, for all $V_i \notin X$.

COMPUTING CAUSAL EFFECTS FROM OBSERVATIONAL DATA

Queries:

$Q_1 = Pr(\text{wet} \mid \text{Sprinkler} = \text{on})$
$= P(p_1) + P(p_2)$

$Q_2 = Pr(\text{wet} \mid do(\text{Sprinkler} = \text{on}))$
$= P(p_1)$

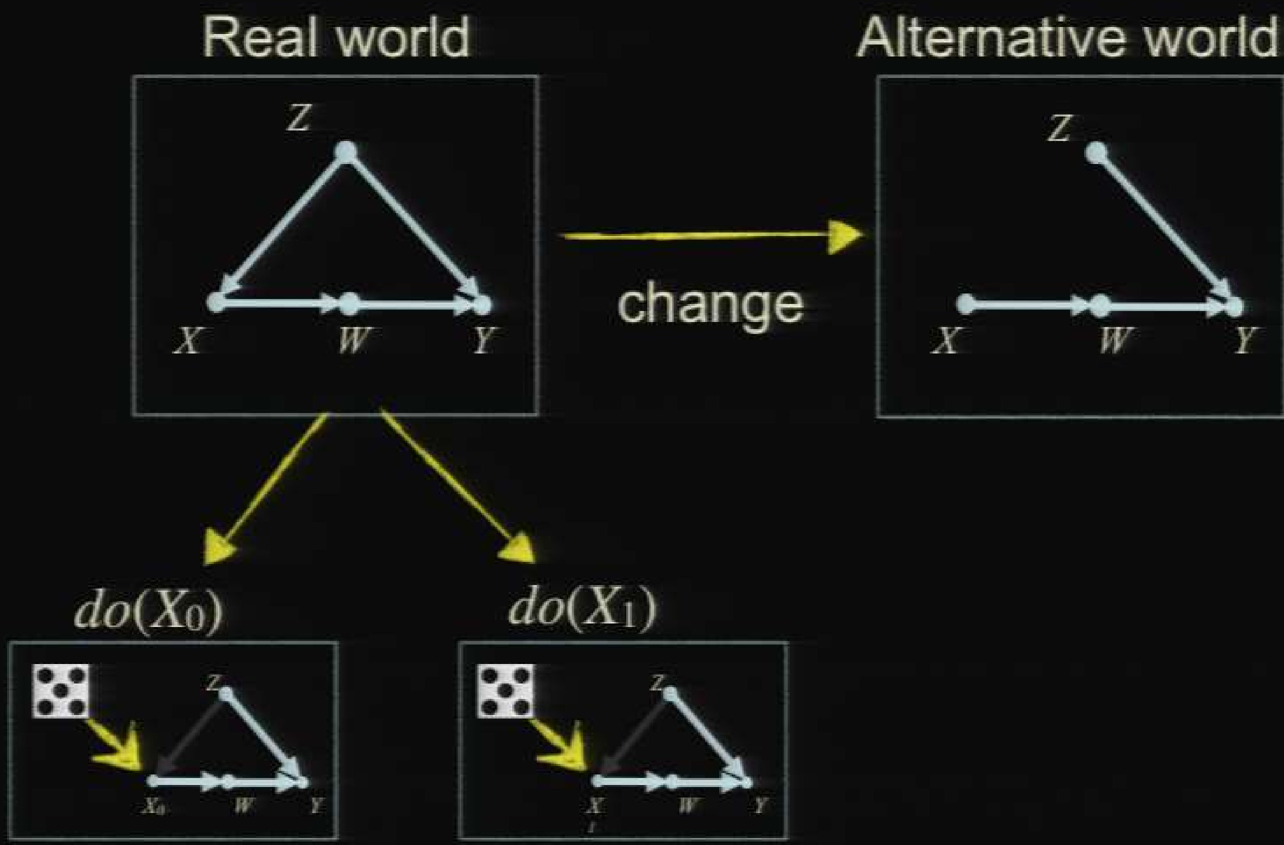
$$\sum_{Se,Ra,Sl} P(Se)\, P(Sp \mid Se)\, P(Ra \mid Se)\, P(We \mid Sp, Ra)\, P(Sl \mid We)$$

PROBLEM 1. COMPUTING EFFECTS FROM OBSERVATIONAL DATA

Z: age, sex
X: action
W: mediator
Y: outcome

$$P(z, x, w, y) \xrightarrow{?} P(y \mid do(x))$$

METHOD FOR COMPUTING CAUSAL EFECTS: RANDOMIZED EXPERIMENTS



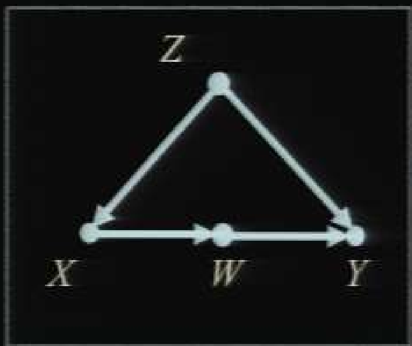
Z: age, sex
X: action
W: mediator
Y: outcome

Randomization:

$P(y \mid do(X_0))$ $\quad$ $P(y \mid do(X_1))$

THE BIG PICTURE: THE CHALLENGE OF CAUSAL INFERENCE

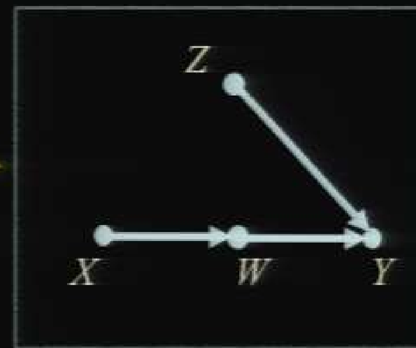
Real world



$P(z, x, w, y)$

change

Alternative world



$P(y \mid do(x))$

Z: age, sex
X: action
W: mediator
Y: outcome

- Goal: how much Y changes with X if we vary X between two different constants free from the influence of Z.
- This is the definition of causal effect.

METHOD FOR COMPUTING CAUSAL EFECTS: RANDOMIZED EXPERIMENTS

Z: age, sex
X: action
W: mediator
Y: outcome