Causal Data Science:
A general framework for data fusion and causal inference

Elias Bareinboim
Columbia University
Twitter: @eliasbareinboim

Machine Learning Research School
Bangkok, August, 2019

Logistics

Logistics

- Website
  - https://causalai.net/mlss19/ (soon)
  - slides – l: mlss19, p: mlss19

Logistics

- Website
  - https://causalai.net/mlss19/ (soon)
  - slides — l: mlss19, p: mlss19
- Instructor

Logistics

- Website
  - https://causalai.net/mlss19/ (soon)
  - slides – l: mlss19, p: mlss19
- Instructor
- Census

Logistics

- Website
  - https://causalai.net/mlss19/ (soon)
  - slides — l: mlss19, p: mlss19
- Instructor
- Census

1. prob. & statistics
2. graphical models
3. d-separation
4. experimental design
5. causal inference

Logistics

- Website
  - https://causalai.net/mlss19/ (soon)
  - slides — l: mlss19, p: mlss19
- Instructor
- Census
- Background

1. prob. & statistics
2. graphical models
3. d-separation
4. experimental design
5. causal inference

Logistics

- Website
  - https://causalai.net/mlss19/ (soon)
  - slides — l: mlss19, p: mlss19
- Instructor
- Census
- Background
- Software (email for access)

1. prob. & statistics
2. graphical models
3. d-separation
4. experimental design
5. causal inference

Logistics

- Website
  - https://causalai.net/mlss19/ (soon)
  - slides — l: mlss19, p: mlss19
- Instructor
- Census
- Background
- Software (email for access)
- Bibliography

1. prob. & statistics
2. graphical models
3. d-separation
4. experimental design
5. causal inference

References

References

- Causal Inference and the Data-fusion Problem
  - E. Bareinboim and J. Pearl
  - Proc. National Academy of Sciences, 2016

References

- Causal Inference and the Data-fusion Problem
  - E. Bareinboim and J. Pearl
  - Proc. National Academy of Sciences, 2016

References

- Causal Inference and the Data-fusion Problem
  - E. Bareinboim and J. Pearl
  - Proc. National Academy of Sciences, 2016

References

- Causal Inference and the Data-fusion Problem
  - E. Bareinboim and J. Pearl
  - Proc. National Academy of Sciences, 2016

References

- Causal Inference and the Data-fusion Problem
  - E. Bareinboim and J. Pearl
  - Proc. National Academy of Sciences, 2016

- The Book of Why: The New Science of Cause and Effect
  - J. Pearl and D. Mackenzie
  - Basic Books, 2018

References

- Causal Inference and the Data-fusion Problem
  - E. Bareinboim and J. Pearl
  - Proc. National Academy of Sciences, 2016
- The Book of Why: The New Science of Cause and Effect
  - J. Pearl and D. Mackenzie
  - Basic Books, 2018
- Causal Inference in Statistics: A primer
  - J. Pearl, M. Glymour, N. Jewel
  - Cambridge Press, 2016

References (cont'd)

References (cont'd)

- Causality: Models, Reasoning, and Inference
  - J. Pearl
  - Cambridge Press, 2000

References (cont'd)

- Causality: Models, Reasoning, and Inference
  - J. Pearl
  - Cambridge Press, 2000
- General Identifiability with Arbitrary Surrogate Experiments
  - S. Lee, J. Correa, and E. Bareinboim
  - Proc. of 35th UAI conference

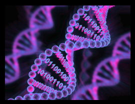
References (cont'd)

- Causality: Models, Reasoning, and Inference
  - J. Pearl
  - Cambridge Press, 2000
- General Identifiability with Arbitrary Surrogate Experiments
  - S. Lee, J. Correa, and E. Bareinboim
  - Proc. of 35th UAI conference
- Probabilistic Graphical Models
  - D. Koller and N. Friedman
  - MIT Press, 2009

Research in CausalAI - Big Picture

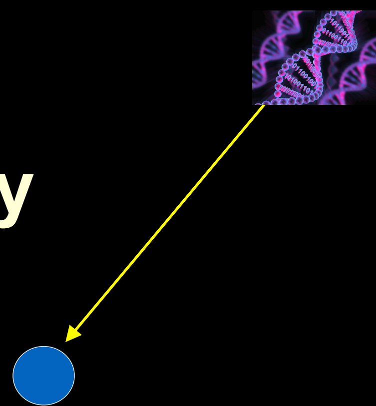
Research in CausalAI - Big Picture

Structural Causal Models



Research in CausalAI - Big Picture

Structural Causal Models



1. Explainability

(Effect identification and decomposition, Bias Analysis and Fairness, Robustness and Generalizability)

Research in CausalAI - Big Picture

Structural Causal Models

1. Explainability

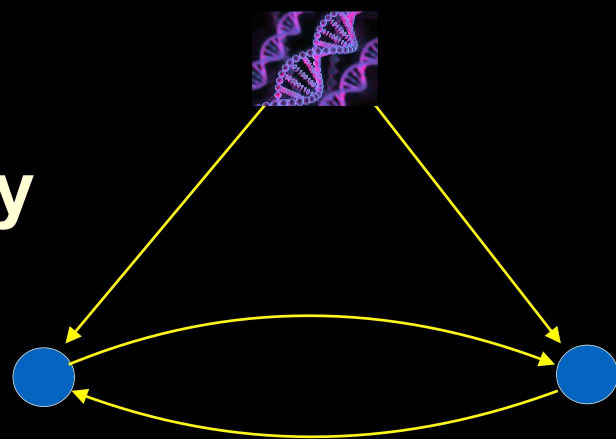
(Effect identification
and decomposition,
Bias Analysis and
Fairness, Robustness
and Generalizability)

2. Decision-Making

(Reinforcement Learning,
Randomized Clinical Trials,
Personalized Decision-Making)

Research in CausalAI - Big Picture

Structural Causal Models



1. Explainability

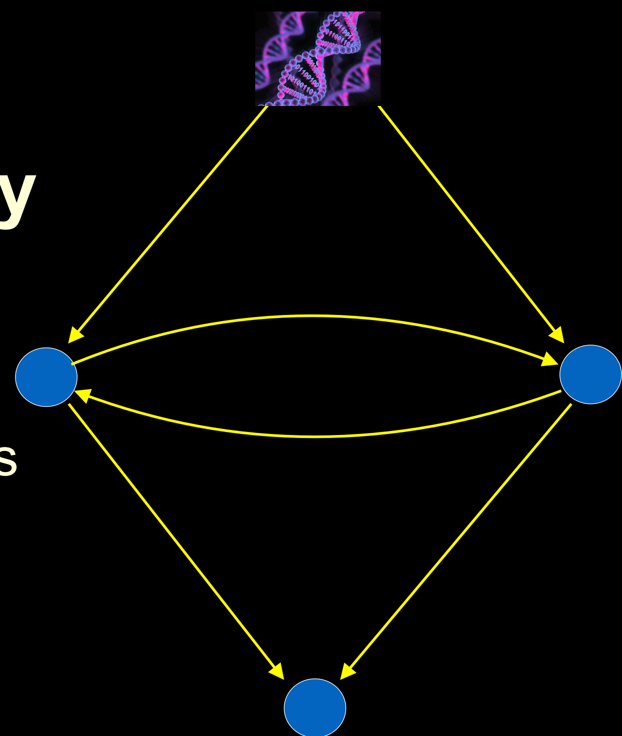
(Effect identification and decomposition, Bias Analysis and Fairness, Robustness and Generalizability)

2. Decision-Making

(Reinforcement Learning, Randomized Clinical Trials, Personalized Decision-Making)

Research in CausalAI - Big Picture

Structural Causal Models



1. Explainability

(Effect identification
and decomposition,
Bias Analysis and
Fairness, Robustness
and Generalizability)

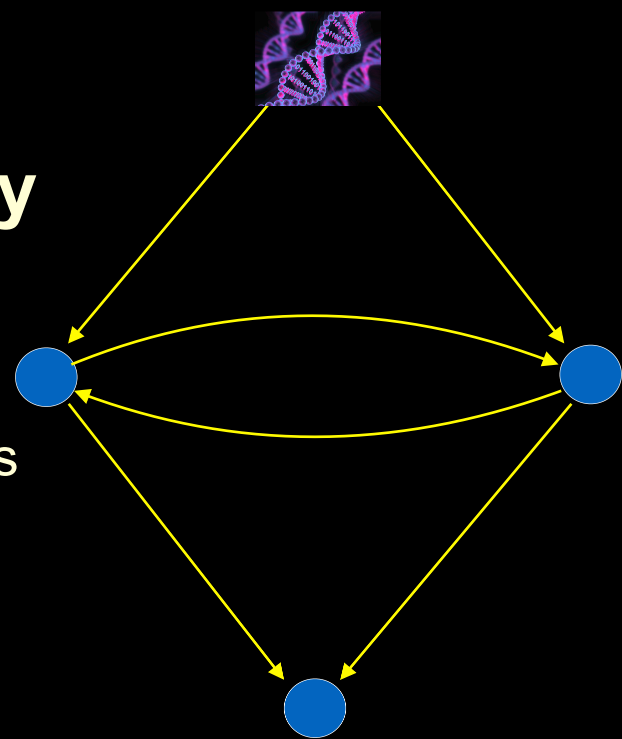
2. Decision-Making

(Reinforcement Learning,
Randomized Clinical Trials,
Personalized Decision-Making)

Applications

Research in CausalAI - Big Picture

Structural Causal Models



1. Explainability

(Effect identification and decomposition, Bias Analysis and Fairness, Robustness and Generalizability)

2. Decision-Making

(Reinforcement Learning, Randomized Clinical Trials, Personalized Decision-Making)

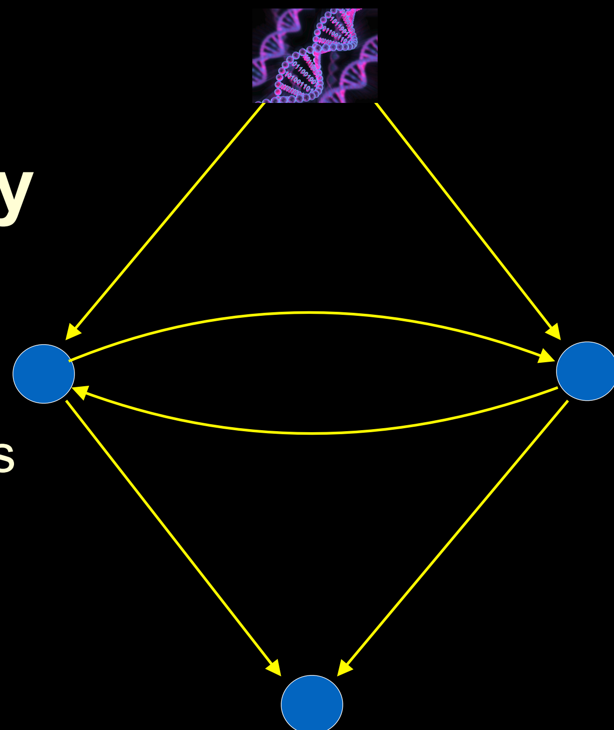
Applications

AI-ML:

Principles and tools for designing robust and adaptable learning systems.

Research in CausalAI - Big Picture

Structural Causal Models



1. Explainability

(Effect identification and decomposition, Bias Analysis and Fairness, Robustness and Generalizability)

2. Decision-Making

(Reinforcement Learning, Randomized Clinical Trials, Personalized Decision-Making)

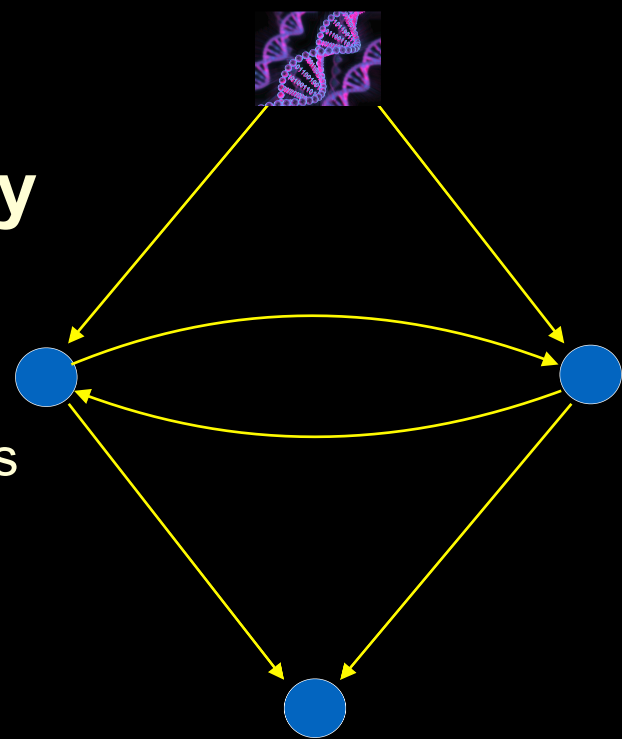
Applications

Data Science:
Principled ("scientific") inferences from large data collections.

AI-ML:
Principles and tools for designing robust and adaptable learning systems.

Research in CausalAI - Big Picture

Structural Causal Models



1. Explainability

(Effect identification and decomposition, Bias Analysis and Fairness, Robustness and Generalizability)

2. Decision-Making

(Reinforcement Learning, Randomized Clinical Trials, Personalized Decision-Making)

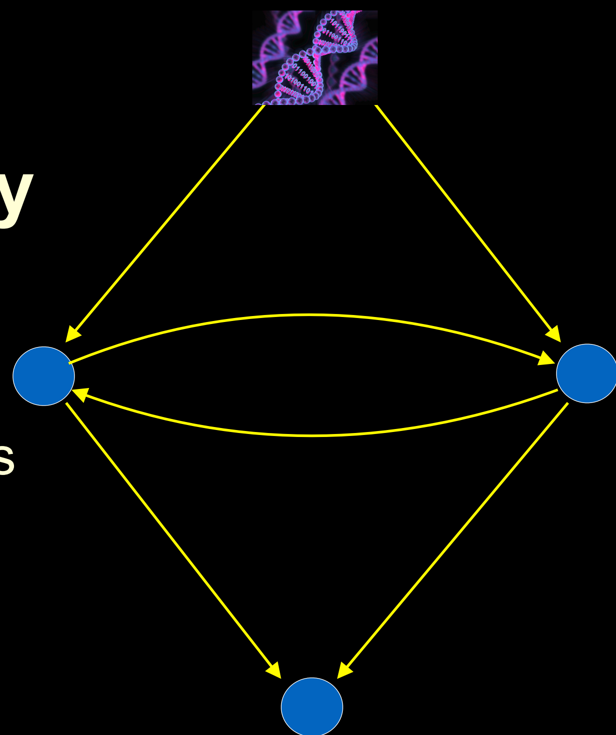
Applications

Data Science:
Principled ("scientific") inferences from large data collections.

AI-ML:
Principles and tools for designing robust and adaptable learning systems.

Research in CausalAI - Big Picture

Structural Causal Models



1. Explainability

(Effect identification and decomposition, Bias Analysis and Fairness, Robustness and Generalizability)

2. Decision-Making

(Reinforcement Learning, Randomized Clinical Trials, Personalized Decision-Making)

Applications

Data Science:
Principled (“scientific”) inferences from large data collections.

AI-ML:
Principles and tools for designing robust and adaptable learning systems.

MOTIVATION

The aspirations & challenges of modern Data Science (circa 2019)

The aspirations & challenges of modern Data Science (circa 2019)

- *"The ability to take data — to be able to understand it, to extract value from it, to visualize it, to communicate it - that's going to be hugely important in the next decades"*

  Hal Varian, chief economist at Google and UC Berkeley
  Professor of Information Sciences, Business, and Economics.

The aspirations & challenges of modern Data Science (circa 2019)

- *“The ability to take data — to be able to understand it, to extract value from it, to visualize it, to communicate it - that’s going to be hugely important in the next decades”*

  Hal Varian, chief economist at Google and UC Berkeley Professor of Information Sciences, Business, and Economics.

- *“Big data is not about the data!”*

  Gary King, Political Scientist, University Professor, Harvard University.

The aspirations & challenges of modern Data Science (circa 2019)

- *“The ability to take data — to be able to understand it, to extract value from it, to visualize it, to communicate it - that’s going to be hugely important in the next decades”*

  Hal Varian, chief economist at Google and UC Berkeley Professor of Information Sciences, Business, and Economics.

- *“Big data is not about the data!”*

  Gary King, Political Scientist, University Professor, Harvard University.

- *“Data Science is only as much of a science as it facilitates the interpretation of data - a two body problem, connecting data to reality”.*

  Judea Pearl, Professor of Computer Science & Statistics, UCLA.

CURRENT STATE OF AFFAIRS
(REPORT FROM THE TRENCHES)


EDITION: U.S. | INTERNATIONAL | MÉXICO | ARABIC
TV: CNN | CNNI | CNN en Español | HLN
CNN Health
SEARCH
POWERED BY Google
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
+1
More sharing
Privacy Policy Close
voiceFIVE
Are you aware of any of the following?
UnitedHealthcare
Humana
Aetna
Blue Cross
1 of 5
Only 5 questions? Easy.
Too easy
ADVERTISEMENT
COMSTOCK
Dr. Sanjay Gupta
August 15th, 2013
08:00 PM ET
Study: Heavy coffee drinking in people under 55 linked to early death


EDITION: U.S. | INTERNATIONAL | MÉXICO | ARABIC
TV: CNN | CNNi | CNN en Español | HLN
CNN Health
SEARCH
POWERED BY Google
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
+1
More sharing
Privacy Policy Close
voiceFive
Are you aware of any of the following?
UnitedHealthcare
Humana
Aetna
Blue Cross
1 of 5
Only 5 questions? Easy.
Too easy
ADVERTISEMENT
August 15th, 2013
08:00 PM ET
Study: Heavy coffee drinking in people under 55 linked to early death
Dr. Sanjay Gupta


EDITION: U.S. INTERNATIONAL MÉXICO ARABIC
TV: CNN CNNi CNN en Español HLN
Home TV & Video CNN Trends U.S. World
the chart
Dr. Sanjay Gupta
August 15th, 2013
08:00 PM ET
Study: Heavy
linked to early death
ACP American College of Physicians
Internal Medicine | Doctors for Adults
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
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: A A A
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.


CBSNEWS
Video
US
World
Politics
Entertainment
RELAX AND
ENJOY THE LUXURY.
A full-size sedan designed with you in mind.
By MICHELLE CASTILLO / CBS NEWS / February 15, 2013, 3: 36 PM
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
EDITION:
TV: CNN
Home
Dr. Sa
August 15t
08:00 PM E
UCLA Biomedical Lib/Serial
the Clinic
Journal Club
CME
PDF
Mortality
io, MD, PhD; and Frank
Size: A A A
ts (2)
icial and detrimental health effects,


Science News *... from universities, journals, and other research organizations*

One Drink Of Red Wine Or Alcohol Is Relaxing To Circulation, But Two Drinks Are Stressful

Feb. 13, 2008 — One drink of either red wine or alcohol slightly benefits the heart and blood vessels, but the positive effects on specific biological markers disappear with two drinks, say researchers at the Peter Munk Cardiac Centre of the Toronto General Hospital.

Related Topics

Health & Medicine
- Heart Disease
- Hypertension

Mind & Brain

Articles
- Mediterranean diet
- Drunkenness
- Coronary heart disease


CBSNEWS
Video
US
World
Politics
Entertainment
Science News
... from universities, journals, and other research organizations
One Drink Of Red Wine Or Alcohol Is Relaxing To Circulation, But Two Drinks Are Stressful
Feb. 13, 2008 — One drink of either red wine or alcohol slightly ... but the positive ... disappear with ... Peter Munk Ca... Hospital.
INSIDE THE EMOTIONAL LIVES OF BOYS
READ THE STORY NOW
TIME Magazine
Apps
Search TIME
Home
NewsFeed
U.S.
Politics
World
Business
Tech
Health
Science
Entertainment
Video
TIME 100
Photos
Current Issue
Archive
Covers
10 Questions
Subscribe
CLICK TO WATCH VIDEO
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
By John Cloud Monday, Aug. 30, 2010
Like 777
Tweet 44
+1 99
Share 1
Read Later
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.
Jodi Cobb / National Geographic Creative / Getty Images
Let's Build a Smarter Planet.
IBM
Is your business ready to launch a new business?
Smarter enterprises use cloud to reinvent themselves
The Magazine
Subscribe
Obama's Race for the Cure
What Boys Want
Foorohar: Yellen for America
Why China Needs More Children
In Texas it is illegal to take
41 Comments
Shares
Tweets
...ficial and detrimental health effects,
August 15th
08:00 PM E
Dr. Sa


CBSNEWS
Video
US
World
Politics
Entertainment
Science News
... from universities, journals, and other research organizations
One Drink Of Red W
Relaxing To Circulation, But Two
Drinks Are Stressf
Feb. 13, 2008 — One dr
alcohol slightly
but the positive
disappear with
Peter Munk Ca
Hospital.
INSIDE THE EM
OF BOYS
Home
Current Issue
Why I
One of the
the consi
By John Cl
In Texas it is illegal to take
41 Comments
Shares
August 15t
08:00 PM E
The New York Times Health | Science
Well
ASK WELL
Ask Well: Nighttime Urination
Bal Harbour Shops
DAVID YURMAN
BVLGARI
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.
Subscribe
What Boys Want
Why China Needs More Children


CBSNEWS
Video
US
World
Politics
Entertainment
Science News
from universities, journals, and other research organizations
One Drink Of Red W
Drinks Are Stressf
To Circulation, But Two
Feb. 13, 2008 — One dr
alcohol slightly
but the positive
disappear with
Peter Munk Ca
Hospital.
The New York
INSIDE THE EM
OF BOYS
Home
News
Current Issue
Why I
One of the
the consi
By John Cl
Like
THE CON
Gra
Re
By RON
PRINT
In Texas it is illegal to take
41 Comments
Shares
August 15th
08:00 PM E
Dr. Sa
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
Nuts and Death.

Other Applications

- Evaluating the impact of a bilingual education program

Other Applications

- Evaluating the impact of a bilingual education program
- Evaluating the impact of a new taxation program

Other Applications

- Evaluating the impact of a bilingual education program
- Evaluating the impact of a new taxation program
- Making a robot to perform a complex task in a new, unforeseen environment

Other Applications

- Evaluating the impact of a bilingual education program
- Evaluating the impact of a new taxation program
- Making a robot to perform a complex task in a new, unforeseen environment
- Diagnosing a patient and planning her/his optimal treatment

Other Applications

- Evaluating the impact of a bilingual education program
- Evaluating the impact of a new taxation program
- Making a robot to perform a complex task in a new, unforeseen environment
- Diagnosing a patient and planning her/his optimal treatment
- Evaluating the effectiveness of a new drug

Other Applications

- Evaluating the impact of a bilingual education program
- Evaluating the impact of a new taxation program
- Making a robot to perform a complex task in a new, unforeseen environment
- Diagnosing a patient and planning her/his optimal treatment
- Evaluating the effectiveness of a new drug
- Assessing the strength of the causal relationship between a set of genes and a certain type of cancer

Other Applications

- Evaluating the impact of a bilingual education program
- Evaluating the impact of a new taxation program
- Making a robot to perform a complex task in a new, unforeseen environment
- Diagnosing a patient and planning her/his optimal treatment
- Evaluating the effectiveness of a new drug
- Assessing the strength of the causal relationship between a set of genes and a certain type of cancer

Tangled web of cause-effect relationships

Why study causality?

Why study causality?

- Cognitive observation

Why study causality?

- Cognitive observation
  - human: sensors + agency + retrospection
  - ability to observe the world and infer the ‘next’ stage
  - ability to intervene in the world and change it
  - ability to reflect about what we did and infer new behavior

Why study causality?

- Cognitive observation
  layer 1 layer 2 layer 3
  - human: sensors + agency + retrospection
  - ability to observe the world and infer the ‘next’ stage
  - ability to intervene in the world and change it
  - ability to reflect about what we did and infer new behavior

Why study causality?

- Cognitive observation
  - layer 1 layer 2 layer 3
  - human: sensors + agency + retrospection
  - ability to observe the world and infer the 'next' stage
  - ability to intervene in the world and change it
  - ability to reflect about what we did and infer new behavior

- Scientific perspective
  - to be able to understand nature and the human being
  - relies on empirical data, logic/math, and scientific method

Why study causality?

- Cognitive observation

  layer 1 layer 2 layer 3

  - human: sensors + agency + retrospection
  - ability to observe the world and infer the 'next' stage
  - ability to intervene in the world and change it
  - ability to reflect about what we did and infer new behavior

- Scientific perspective
  - to be able to understand nature and the human being
  - relies on empirical data, logic/math, and scientific method

- Pragmatic perspective (e.g., AI, Econ, Medicine)
  - to be able to create "truly" intelligent systems
  - to be able to create and implement policies optimizing certain {social, economical, political} goals

OUTLINE (PART I)

- Introduction
  - Data, data, data…
  - Basic definitions
- Technical Results
  - The truncated product formula
  - The back-door adjustment formula
  - The front-door adjustment formula
  - The do-calculus
- Tasks
  - Policy evaluation
  - Generalizability & Robustness
  - Decision-making & RL

OUTLINE (PART I)

- Introduction
  * Data, data, data…
  * Basic definitions
- Technical Results
  * The truncated product formula
  * The back-door adjustment formula
  * The front-door adjustment formula
  * The do-calculus
- Tasks
  * Policy evaluation
  * Generalizability & Robustness
  * Decision-making & RL

Data, data, data…

Ex 1. Data analysis in a 2x2 table...

(Simpson, 1951; Ch. 6, Pearl, 2000)

	surv. (Y)	not-surv. (¬Y)		Recovery Rate
drug (X)	20	20	40	50%
no-drug (¬X)	16	24	40	40%
	36	44		

Stratifying by sex...

	surv. (Y)	not-surv. (¬Y)		Recovery Rate
drug (X)	20	20	40	50%
no-drug (¬X)	16	24	40	40%
	36	44		

male (¬F)	surv. (Y)	not-surv.		Recovery Rate
drug (X)	18	12	30	60%
no-drug (¬X)	7	3	10	70%
	25	15	40	

female (F)	surv. (Y)	not-surv. (¬Y)		Recovery Rate
drug (X)	2	8	10	20%
no-drug (¬X)	9	21	30	30%
	11	29	40	

Which table should the physician consult?

	surv. (Y)	not-surv. (¬Y)		Recovery Rate
drug (X)	20	20	40	50%
no-drug (¬X)	16	24	40	40%
	36	44		

$P(Y \mid F, X) < P(Y \mid F, \neg X)$

$P(Y \mid \neg F, X) < P(Y \mid \neg F, \neg X)$

male (¬F)	surv. (Y)	not-surv.		Recovery Rate
drug (X)	18	12	30	60%
no-drug (¬X)	7	3	10	70%
	25	15	40	

female (F)	surv. (Y)	not-surv. (¬Y)		Recovery Rate
drug (X)	2	8	10	20%
no-drug (¬X)	9	21	30	30%
	11	29	40	

Which table should the physician consult?

	surv. (Y)	not-surv. (¬Y)		Recovery Rate
drug (X)	20	20	40	50%
no-drug (¬X)	16	24	40	40%
	36	44		

$P(Y \mid F, X) < P(Y \mid F, \neg X)$

$P(Y \mid \neg F, X) < P(Y \mid \neg F, \neg X)$

but

$P(Y \mid X) > P(Y \mid \neg X)$!!

male (¬F)	surv. (Y)	not-surv.		Recovery Rate
drug (X)	18	12	30	60%
no-drug (¬X)	7	3	10	70%
	25	15	40	

female (F)	surv. (Y)	not-surv. (¬Y)		Recovery Rate
drug (X)	2	8	10	20%
no-drug (¬X)	9	21	30	30%
	11	29	40	

WOULD THE RESULT BE DIFFERENT WITH A DIFFERENT STORY?

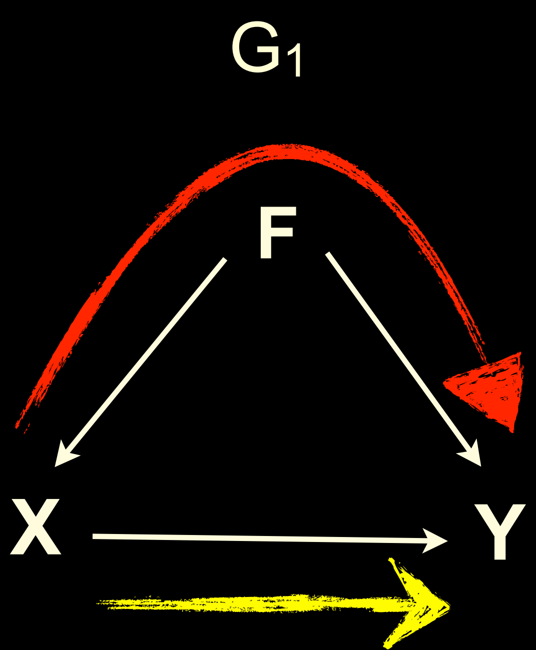
male (¬F)	surv. (Y)	not-surv.		Recovery Rate
drug (X)	18	12	30	60%
no-drug	7	3	10	70%
	25	15	40	

female (F)	surv. (Y)	not-surv.		Recovery Rate
drug (X)	2	8	10	20%
no-drug	9	21	30	30%
	11	29	40	

WOULD THE RESULT BE DIFFERENT WITH A DIFFERENT STORY?

male (¬F)	surv. (Y)	not-surv.		Recovery Rate
drug (X)	18	12	30	60%
no-drug	7	3	10	70%
	25	15	40	

female (F)	surv. (Y)	not-surv.		Recovery Rate
drug (X)	2	8	10	20%
no-drug	9	21	30	30%
	11	29	40	

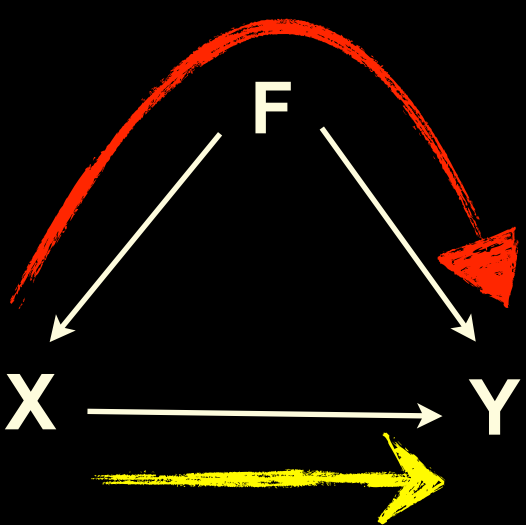


WOULD THE RESULT BE DIFFERENT WITH A DIFFERENT STORY?

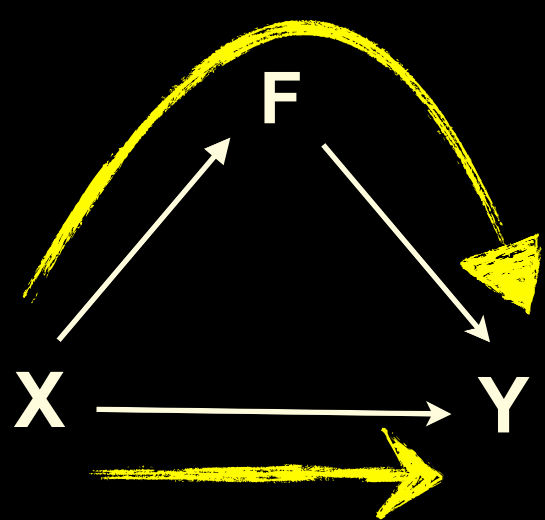
male (¬F)	surv. (Y)	not-surv.		Recovery Rate
drug (X)	18	12	30	60%
no-drug	7	3	10	70%
	25	15	40	

female (F)	surv. (Y)	not-surv.		Recovery Rate
drug (X)	2	8	10	20%
no-drug	9	21	30	30%
	11	29	40	

G_1



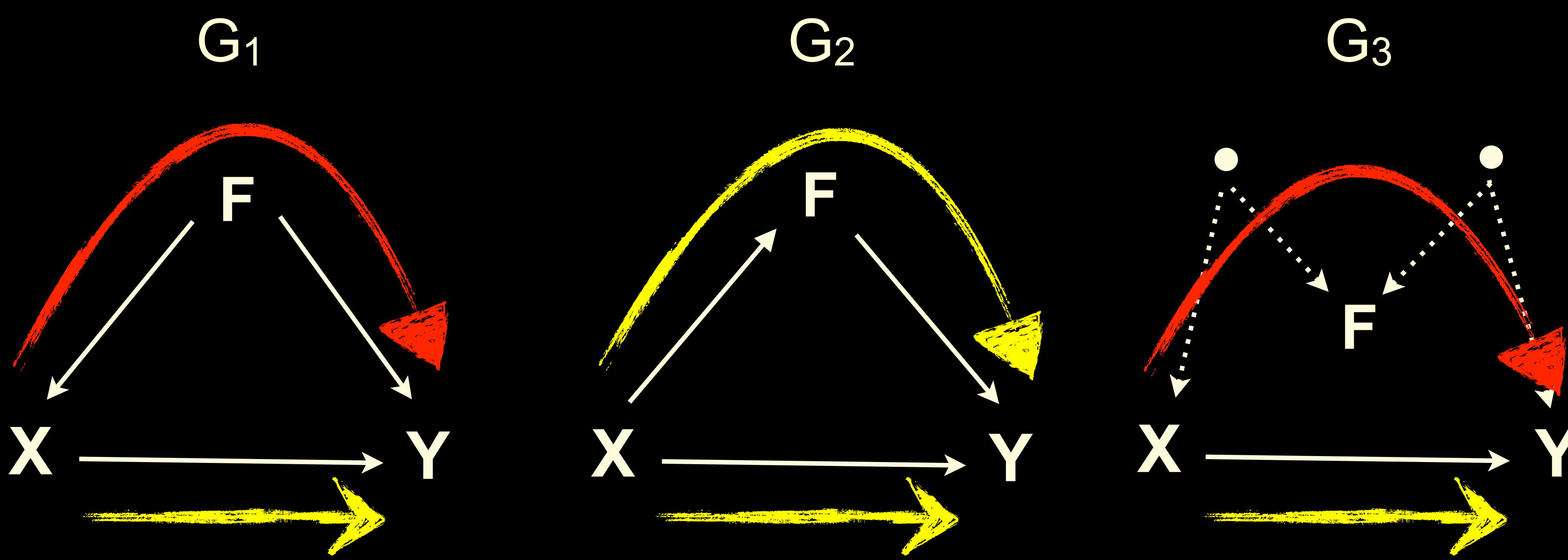
G_2



WOULD THE RESULT BE DIFFERENT WITH A DIFFERENT STORY?

male (¬F)	surv. (Y)	not-surv.		Recovery Rate
drug (X)	18	12	30	60%
no-drug	7	3	10	70%
	25	15	40	

female (F)	surv. (Y)	not-surv.		Recovery Rate
drug (X)	2	8	10	20%
no-drug	9	21	30	30%
	11	29	40	



Ex 2. Trivial data analysis (Thrun)?

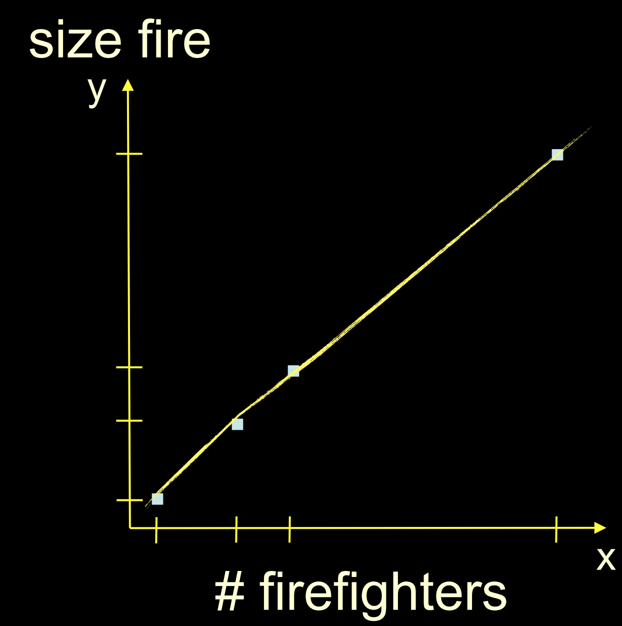
#Firefighters (X)	#Size Fire (Y)
10	100
40	400
60	600
200	2000

Is

$Y = \beta X + \epsilon$ a good fit?

Simple linear regression...

#Firefighters (X)	#Size Fire (Y)
10	100
40	400
60	600
200	2000

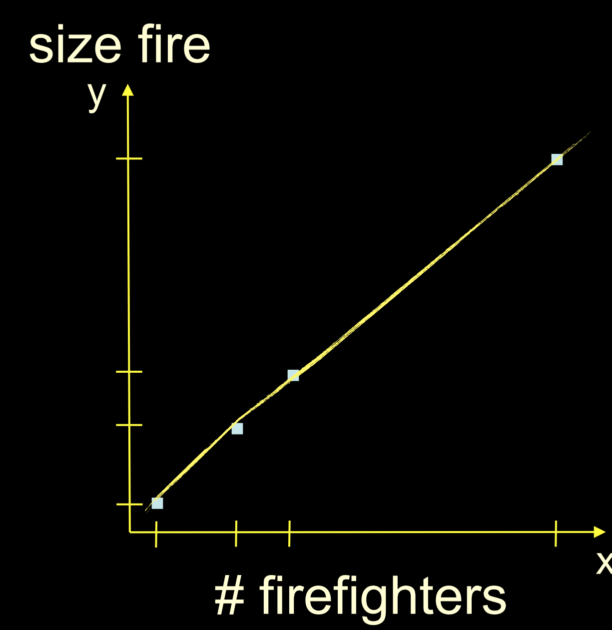


Is

$Y = \beta X + \epsilon$ a good fit?

Simple linear regression...

#Firefighters (X)	#Size Fire (Y)
10	100
40	400
60	600
200	2000



Is

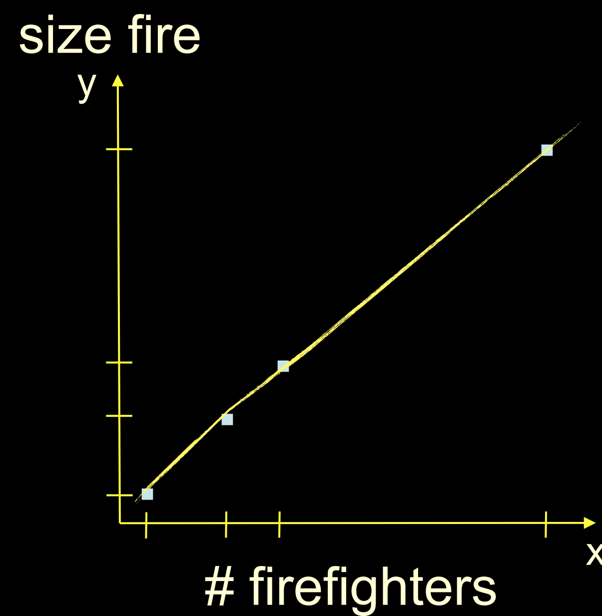
$Y = \beta X + \epsilon$ a good fit?

X and Y are perfectly correlated...

but does X cause Y?

Simple linear regression...

#Firefighters (X)	#Size Fire (Y)
10	100
40	400
60	600
200	2000



Is

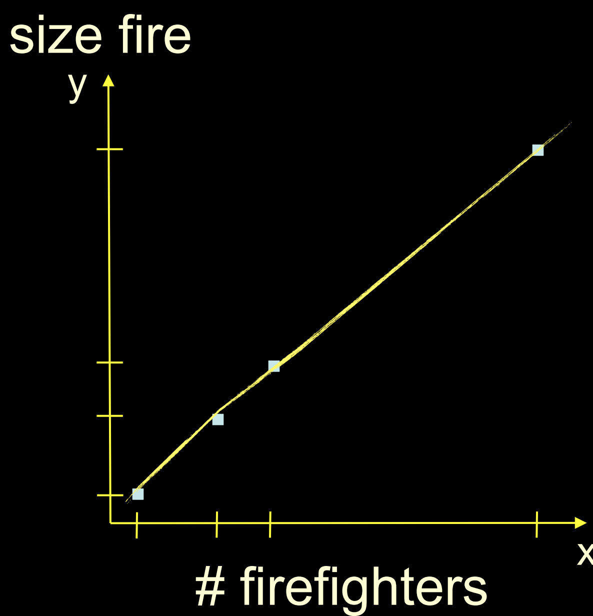
$Y = \beta X + \epsilon$ a good fit?

after all,

what's the meaning of β?

Simple linear regression...

#Firefighters (X)	#Size Fire (Y)
10	100
40	400
60	600
200	2000



Is

$Y = \beta X + \epsilon$ a good fit?

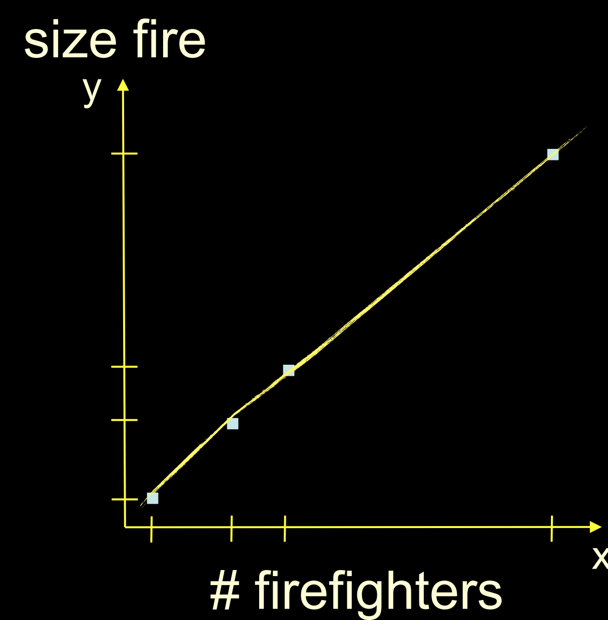
Lasso?

after all,

what's the meaning of β?

Simple linear regression...

#Firefighters (X)	#Size Fire (Y)
10	100
40	400
60	600
200	2000



Is

$Y = \beta X + \epsilon$ a good fit?

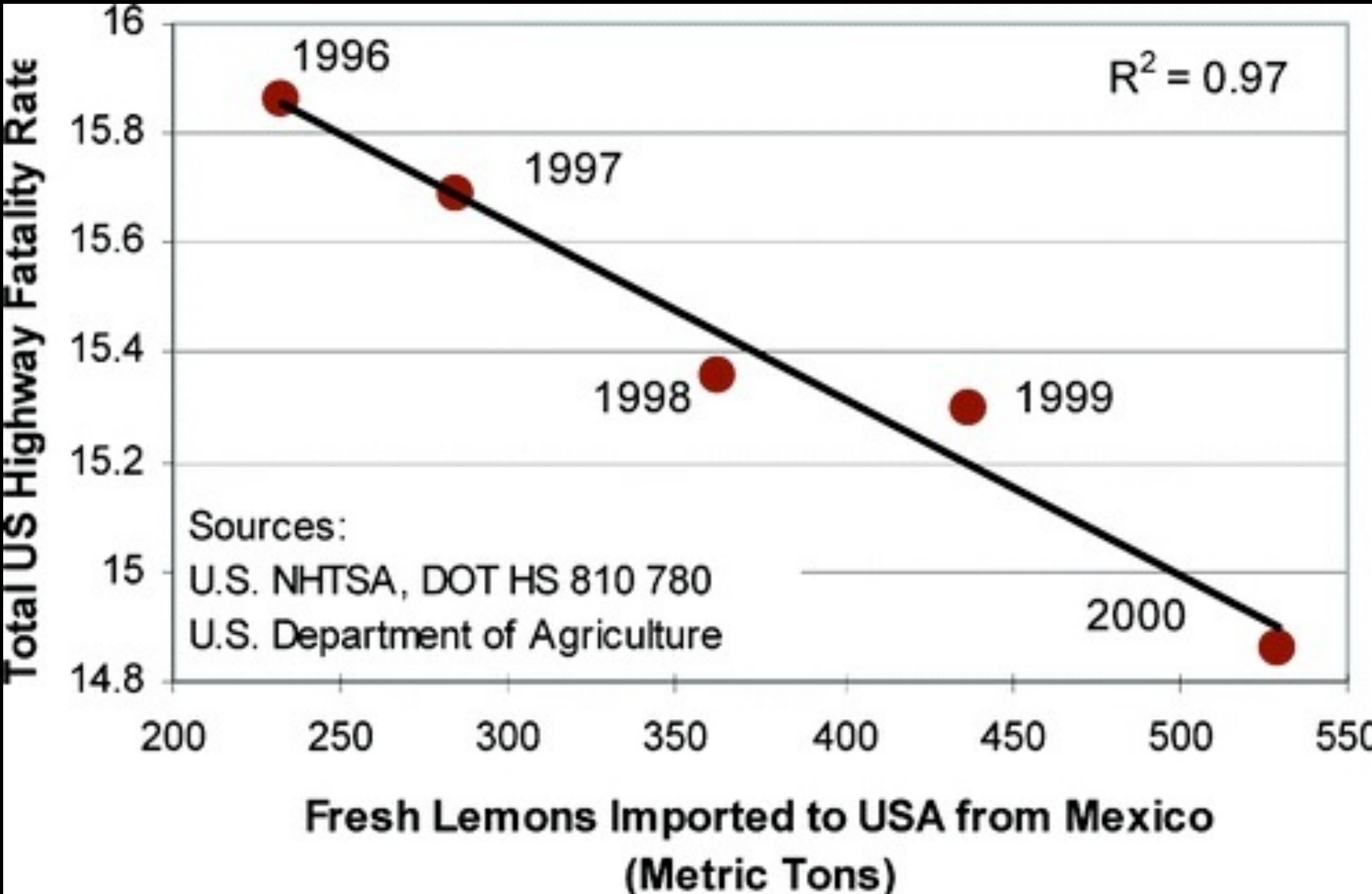
Lasso?

DNN?

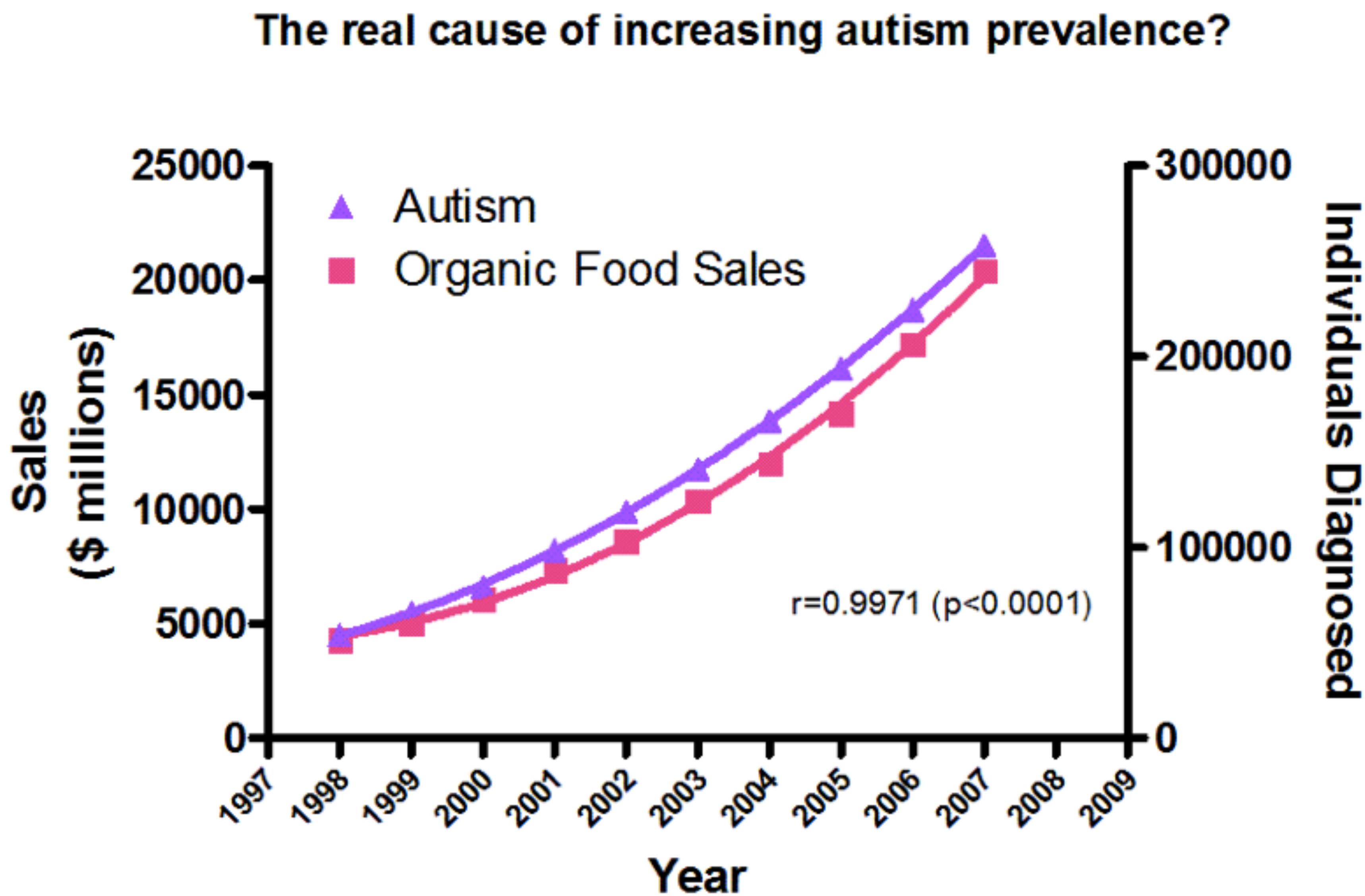
after all,

what's the meaning of β?

Ex 2b. Trivial data analysis (?)



Ex 2b. Trivial data analysis (?)



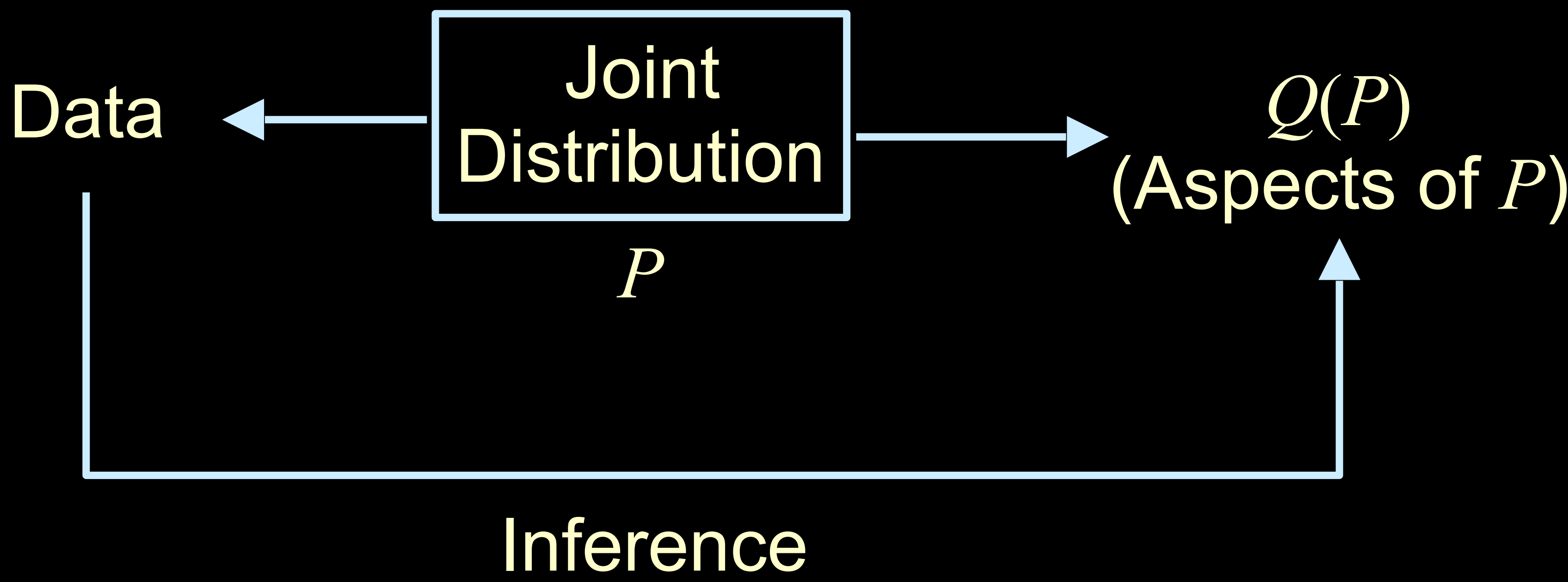
Sources: Organic Trade Association, 2011 Organic Industry Survey; U.S. Department of Education, Office of Special Education Programs, Data Analysis System (DANS), OMB# 1820-0043: "Children with Disabilities Receiving Special Education Under Part B of the Individuals with Disabilities Education Act

What’s going on here...?

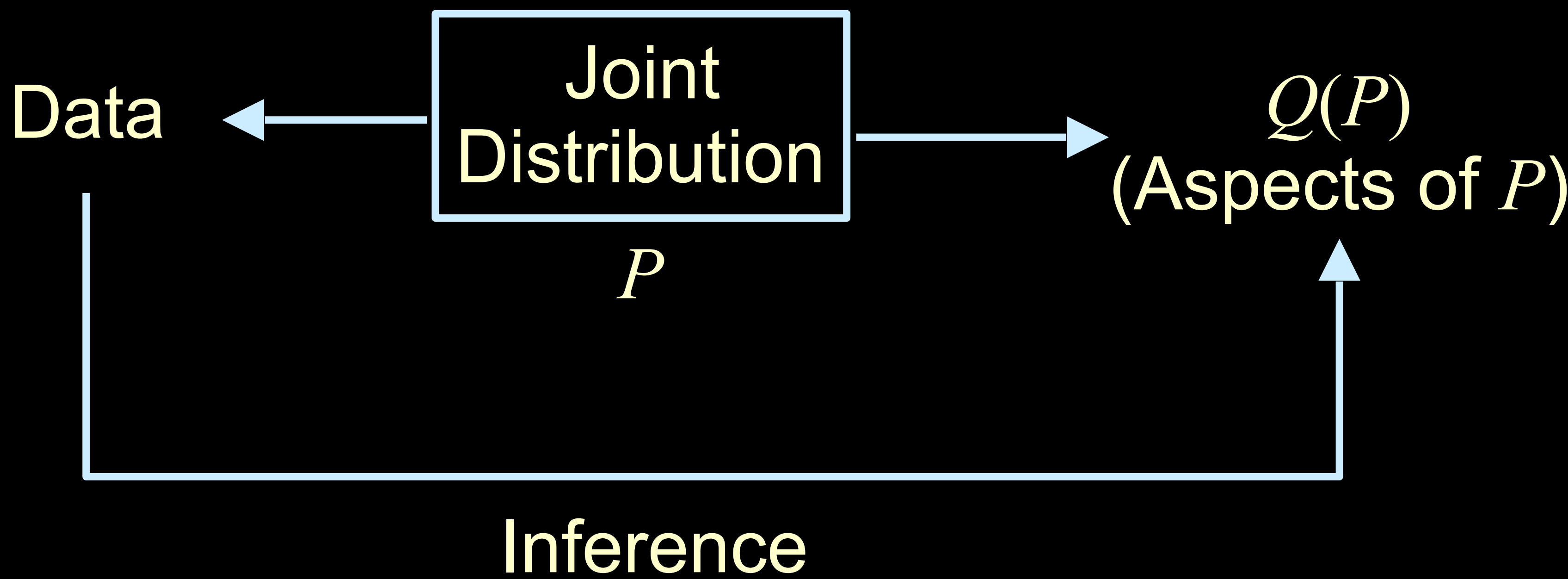
What’s going on here...?

🤷🏻‍♂️

TRADITIONAL STATISTICAL/ MACHINE LEARNING PARADIGM

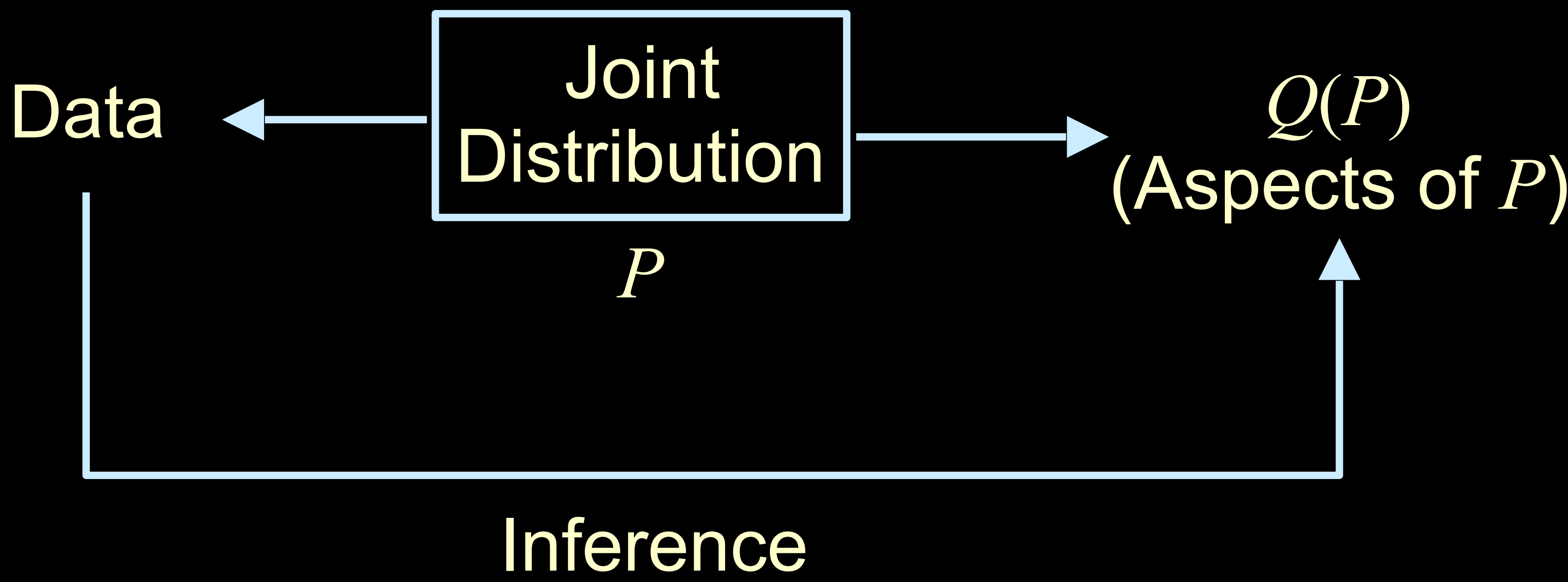


TRADITIONAL STATISTICAL/ MACHINE LEARNING PARADIGM



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

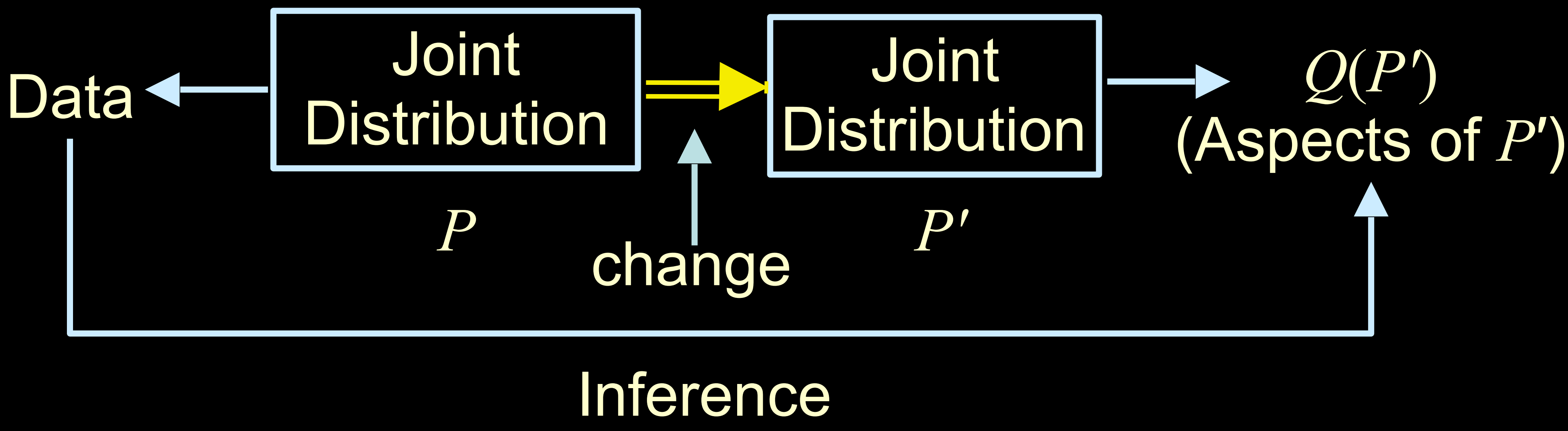
TRADITIONAL STATISTICAL/ MACHINE LEARNING PARADIGM



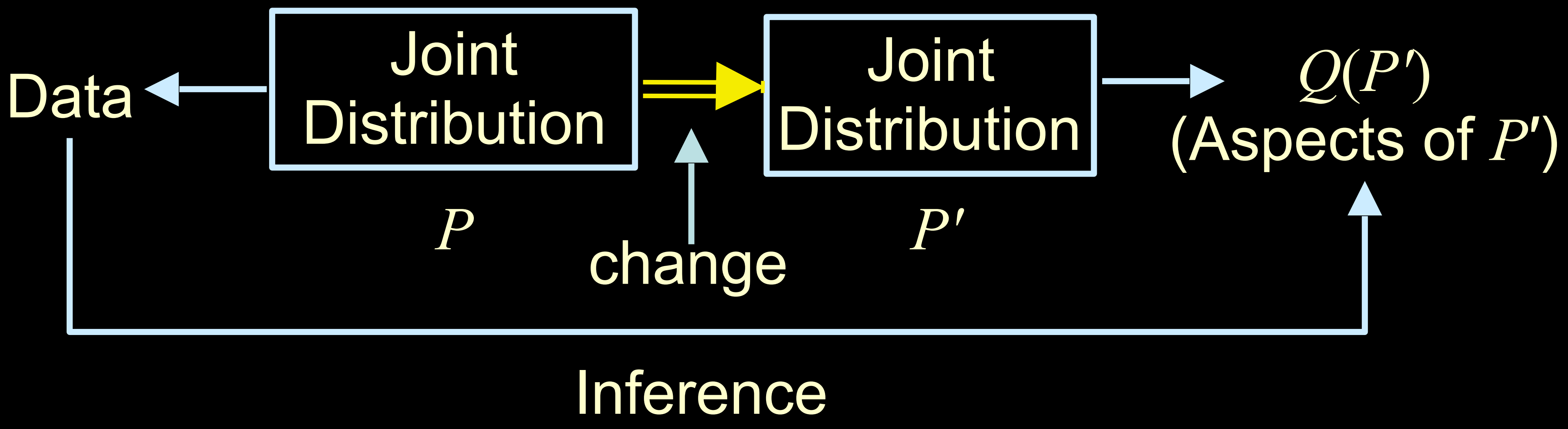
e.g., Infer whether customers who bought product A would also buy product B — or, compute $Q = P(B \mid A)$.

- capturing regularities of the original P (LLN, CLT)

FROM STATISTICAL TO CAUSAL ANALYSIS



FROM STATISTICAL TO CAUSAL ANALYSIS

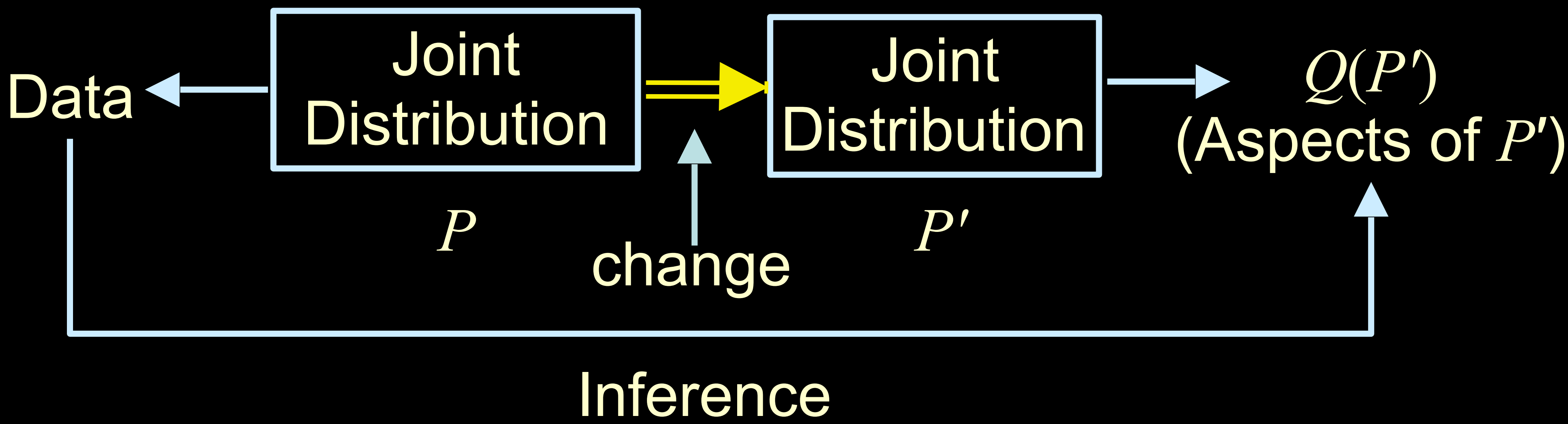


e.g., Estimate $P'(sales)$ if we double the price.

Estimate $P'(cancer)$ if we ban smoking.

How does P change to P'? New oracle

FROM STATISTICAL TO CAUSAL ANALYSIS



e.g., Estimate $P'(sales)$ if we double the price.

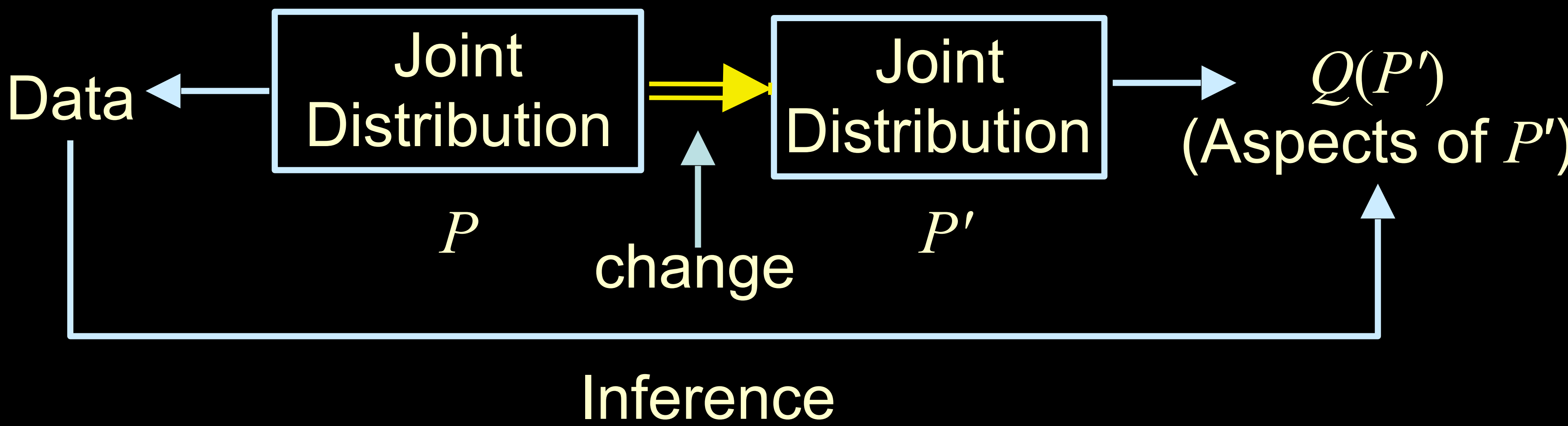
Estimate $P'(cancer)$ if we ban smoking.

What about the CLT?

How does P change to P'? New oracle

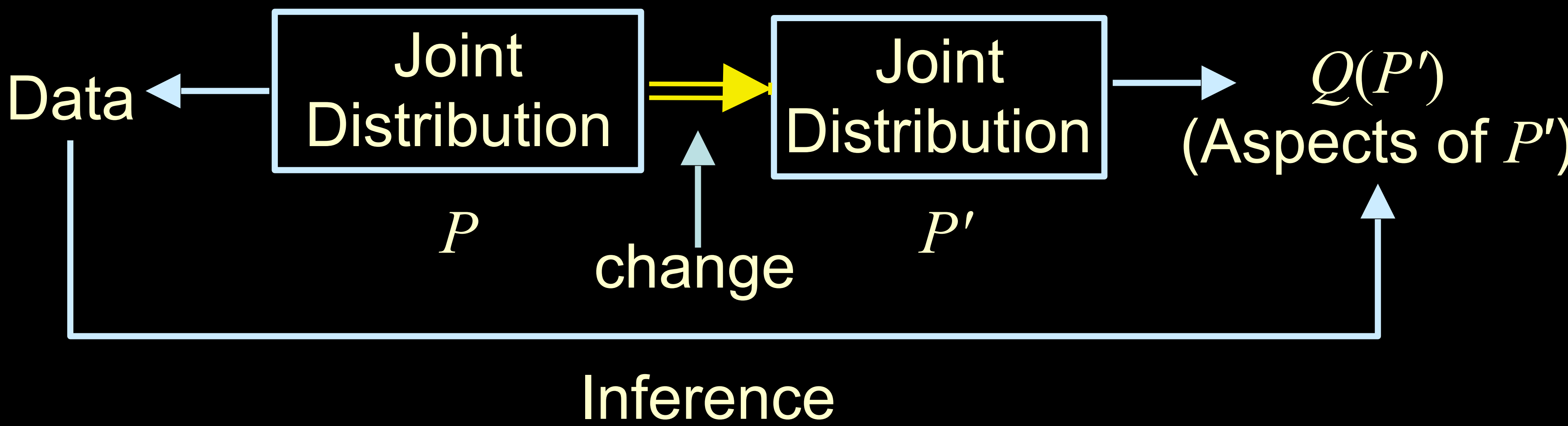
FROM STATISTICS TO COUNTERFACTUALS

Probability and statistics deal with static relations



FROM STATISTICS TO COUNTERFACTUALS

Probability and statistics deal with static relations

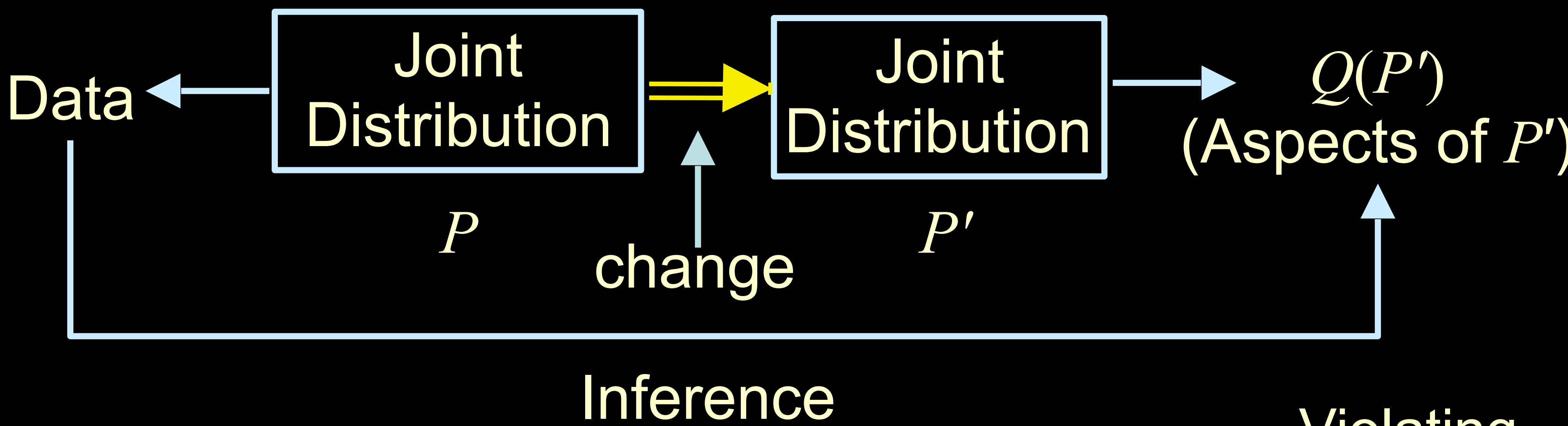


What happens when P changes?

e.g., Estimate the probability that a patient who took a drug and died would be alive had he not taken the drug.

FROM STATISTICS TO COUNTERFACTUALS

Probability and statistics deal with static relations



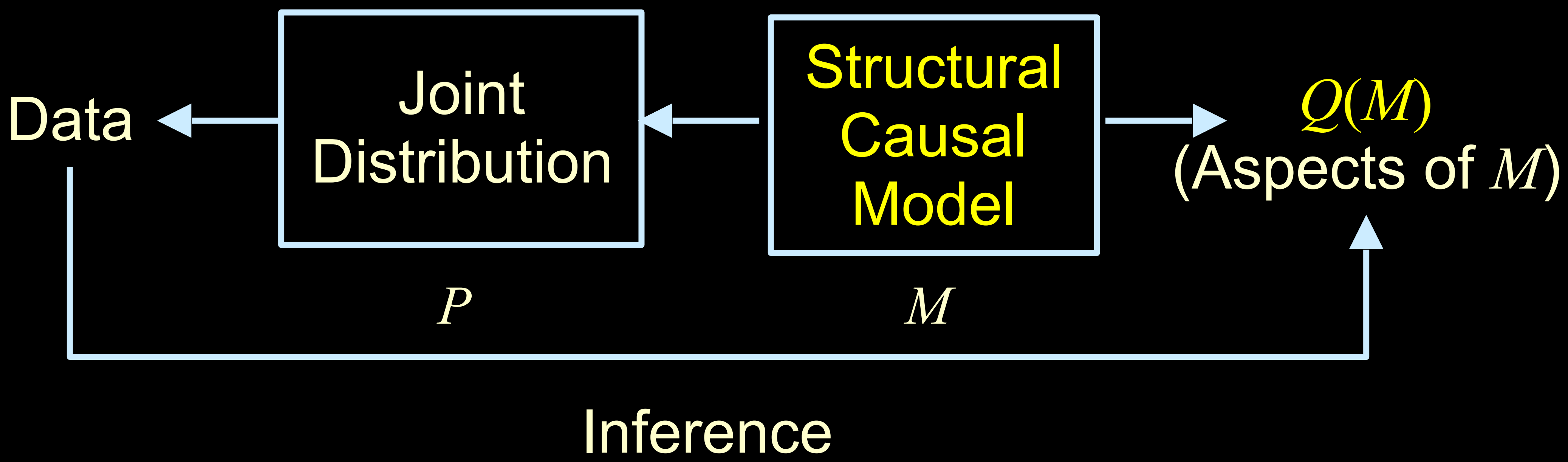
Violating
the laws
of physics?

What happens when P changes?

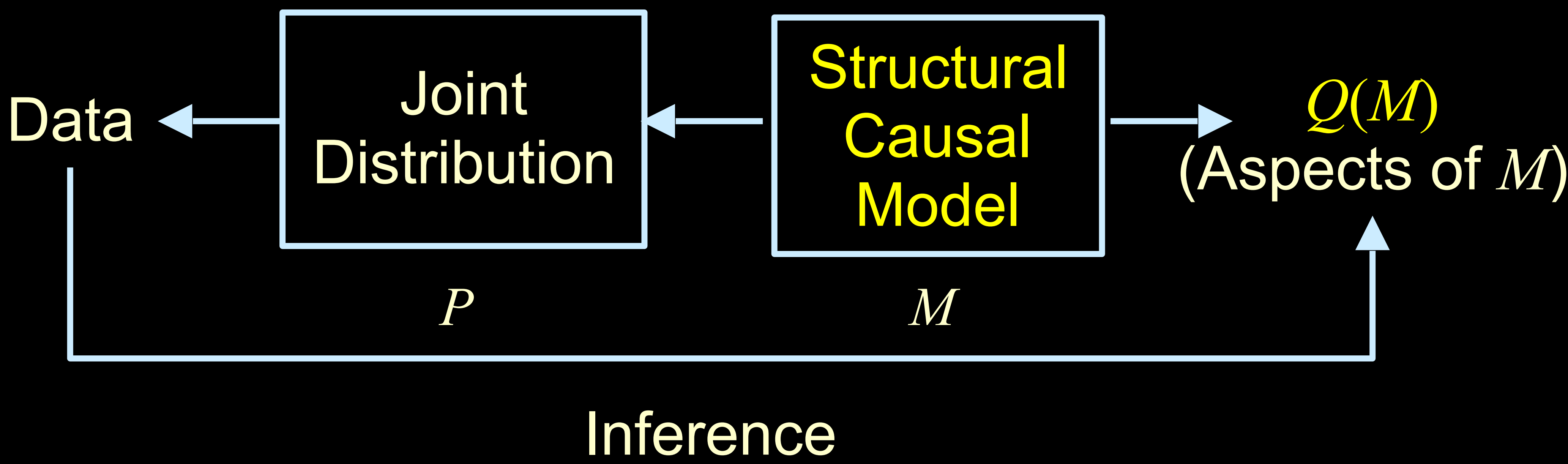
e.g., Estimate the probability that a patient who took a drug and died would be alive had he not taken the drug.

Instead of thinking about multiple individual distributions, let's dig deeper…

THE STRUCTURAL MODEL PARADIGM

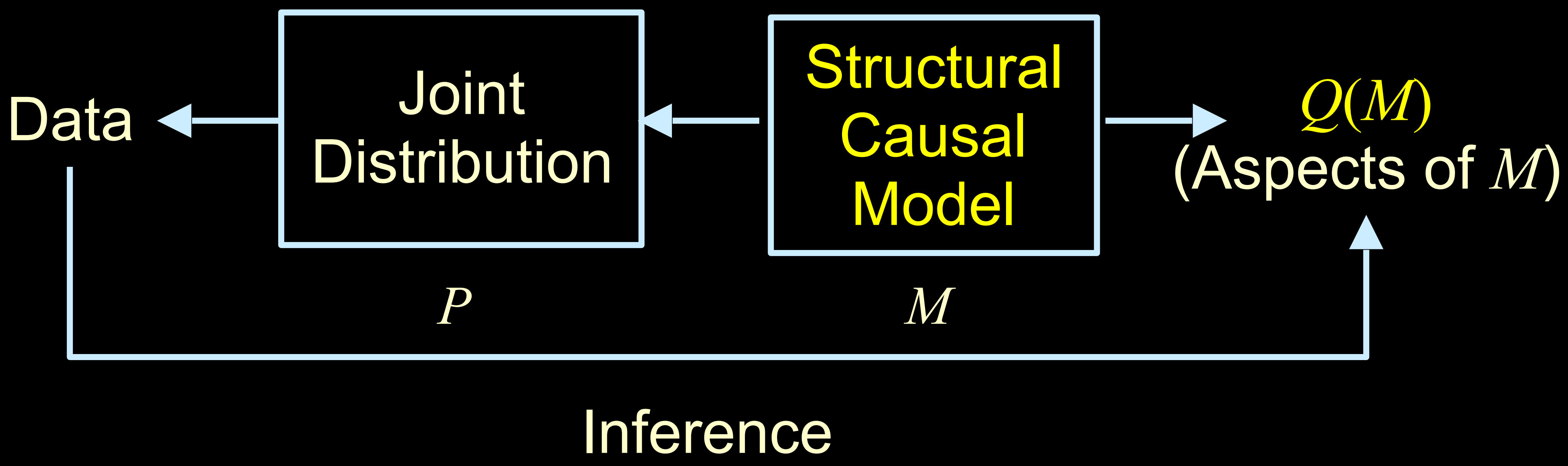


THE STRUCTURAL MODEL PARADIGM



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

THE STRUCTURAL MODEL PARADIGM



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

P — model of data
M — model of reality

REPRESENTING THE DATA GENERATING MODEL

- Processes

REPRESENTING THE DATA GENERATING MODEL

- Processes

  Edu $\leftarrow f_E\,(SES,\, U_e)$

REPRESENTING THE DATA GENERATING MODEL

- Processes

$$\text{Edu} \leftarrow f_E\,(SES, U_e)$$
$$\text{Income} \leftarrow f_I\,(Edu, SES, U_i)$$

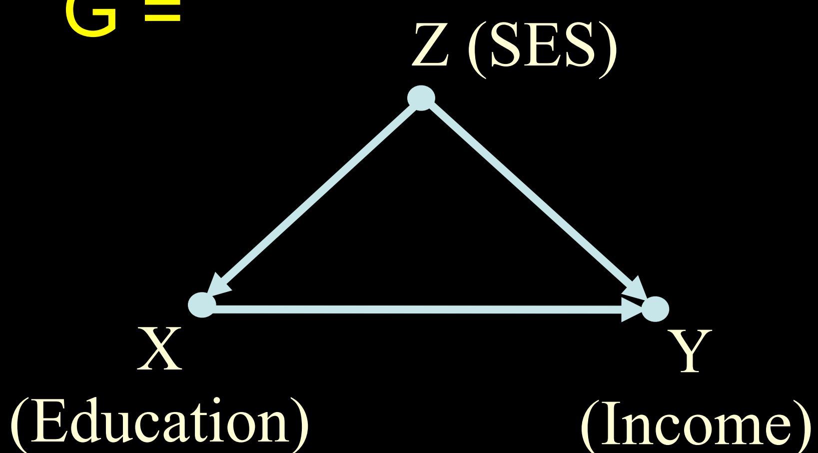
REPRESENTING THE DATA GENERATING MODEL

- Processes

$$\text{Edu} \leftarrow f_E\,(SES, U_e)$$

$$\text{Income} \leftarrow f_I\,(Edu, SES, U_i)$$

G =



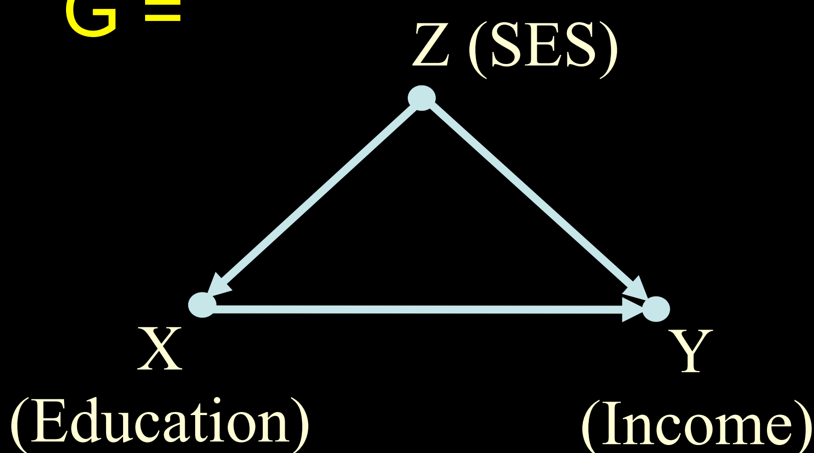
P(Z, X, Y)
(observational)

REPRESENTING THE DATA GENERATING MODEL

- Processes

$$\text{Edu} \leftarrow f_E(SES, U_e)$$
$$\text{Income} \leftarrow f_I(Edu, SES, U_i)$$

G =



P(Z, X, Y)
(observational)

- Intervention

REPRESENTING THE DATA GENERATING MODEL

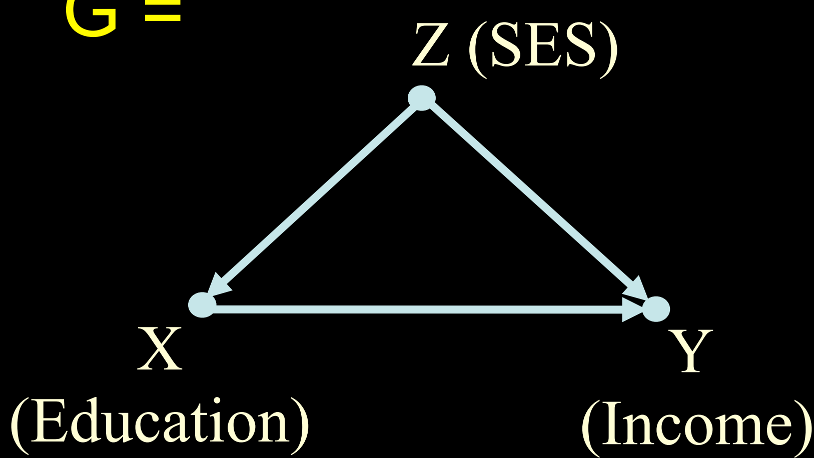
- Processes

$$\text{Edu} \leftarrow f_E\,(\text{SES}, U_e)$$
$$\text{Income} \leftarrow f_I\,(\text{Edu}, \text{SES}, U_i)$$

- Intervention

$$\text{Edu} \leftarrow \text{Yes}$$

G =



P(Z, X, Y)
(observational)

REPRESENTING THE DATA GENERATING MODEL

- Processes

  Edu $\leftarrow f_E$ (SES, U_e)

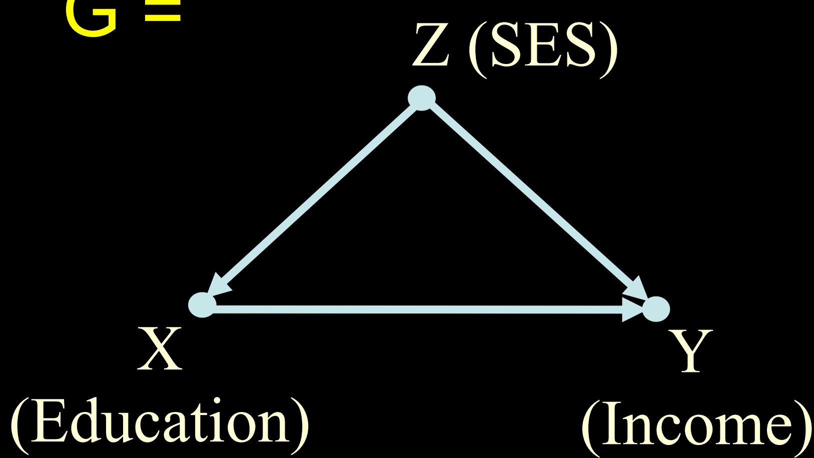
  Income $\leftarrow f_I$ (Edu, SES, U_i)

- Intervention

  Edu $\leftarrow$ Yes

  Income $\leftarrow f_I$ (Edu, SES, U_i)

G =



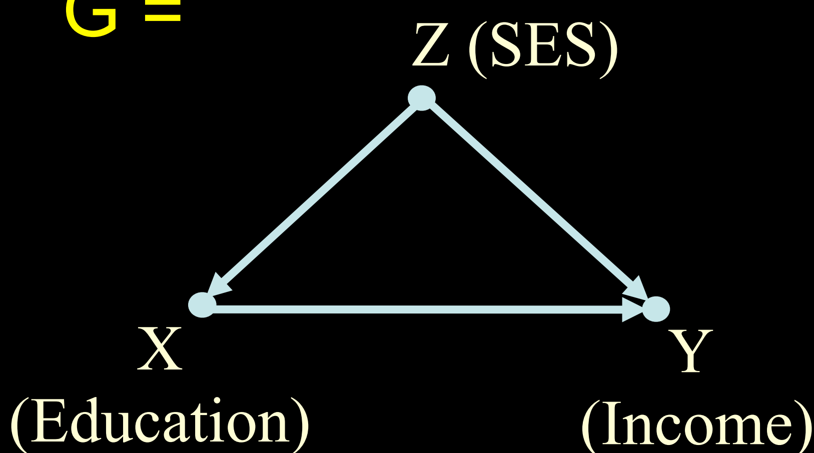
P(Z, X, Y)
(observational)

REPRESENTING THE DATA GENERATING MODEL

- Processes

$$\text{Edu} \leftarrow f_E(\text{SES}, U_e)$$
$$\text{Income} \leftarrow f_I(\text{Edu}, \text{SES}, U_i)$$

$G =$

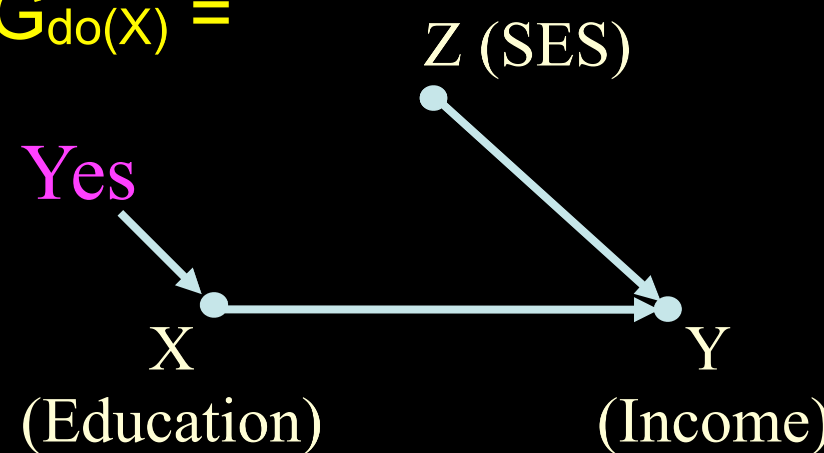


$P(Z, X, Y)$
(observational)

- Intervention

$$\text{Edu} \leftarrow \text{Yes}$$
$$\text{Income} \leftarrow f_I(\text{Edu}, \text{SES}, U_i)$$

$G_{do(X)} =$



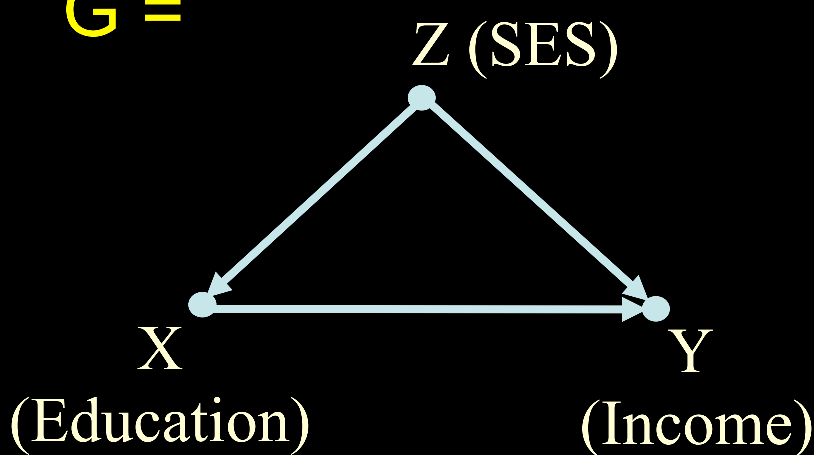
$P(Z, Y \mid do(X = \text{Yes}))$
(interventional)

REPRESENTING THE DATA GENERATING MODEL

- Processes

$$\text{Edu} \leftarrow f_E(SES, U_e)$$
$$\text{Income} \leftarrow f_I(Edu, SES, U_i)$$

$G =$

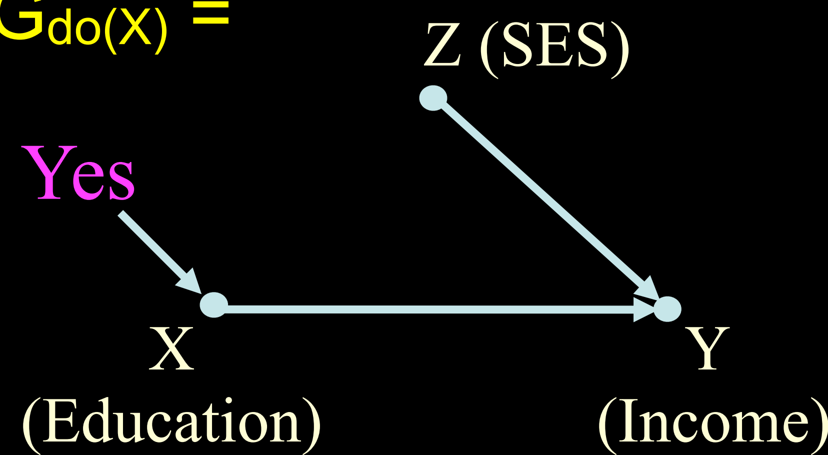


$P(Z, X, Y)$
(observational)

- Intervention

$$\text{Edu} \leftarrow \text{Yes}$$
$$\text{Income} \leftarrow f_I(Edu, SES, U_i)$$

$G_{do(X)} =$



$P(Z_{X=Yes}, Y_{X=Yes}) = P(Z, Y \mid do(X = Yes))$
(interventional)

REPRESENTING
THE DATA GENERATING MODEL

REPRESENTING THE DATA GENERATING MODEL

Definition: A structural causal model M (or, data generating model) is a tuple $(V, U, F, P(u))$, where

REPRESENTING THE DATA GENERATING MODEL

Definition: A **structural causal model M** (or, data generating model) is a tuple $(V, U, F, P(u))$, where

- $V = \{V_1, ..., V_n\}$ are endogenous variables,

REPRESENTING THE DATA GENERATING MODEL

Definition: A structural causal model M (or, data generating model) is a tuple $(V, U, F, P(u))$, where

- $V = \{V_1,...,V_n\}$ are endogenous variables,
- $U = \{U_1,...,U_m\}$ are exogenous variables,

REPRESENTING THE DATA GENERATING MODEL

Definition: A structural causal model M (or, data generating model) is a tuple $(V, U, F, P(u))$, where

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

  for each V_i, $V_i \leftarrow f_i(Pa_i, U_i)$, where $Pa_i \subset V$, $U_i \subset U$.

REPRESENTING THE DATA GENERATING MODEL

Definition: A structural causal model M (or, data generating model) is a tuple $(V, U, F, P(u))$, where

- $V = \{V_1,...,V_n\}$ are endogenous variables,
- $U = \{U_1,...,U_m\}$ are exogenous variables,
- $F = \{f_1,..., f_n\}$ are functions determining V,
  for each V_i, $\boxed{V_i \leftarrow f_i(Pa_i, U_i)}$, where $Pa_i \subset V$, $U_i \subset U$.

REPRESENTING THE DATA GENERATING MODEL

Definition: A structural causal model M (or, data generating model) is a tuple $(V, U, F, P(u))$, where

- $V = \{V_1,...,V_n\}$ are endogenous variables,
- $U = \{U_1,...,U_m\}$ are exogenous variables,
- $F = \{f_1,..., f_n\}$ are functions determining V,
  for each V_i, $V_i \leftarrow f_i(Pa_i, U_i)$, where $Pa_i \subset V$, $U_i \subset U$.
- $P(u)$ is a distribution over U.

REPRESENTING THE DATA GENERATING MODEL

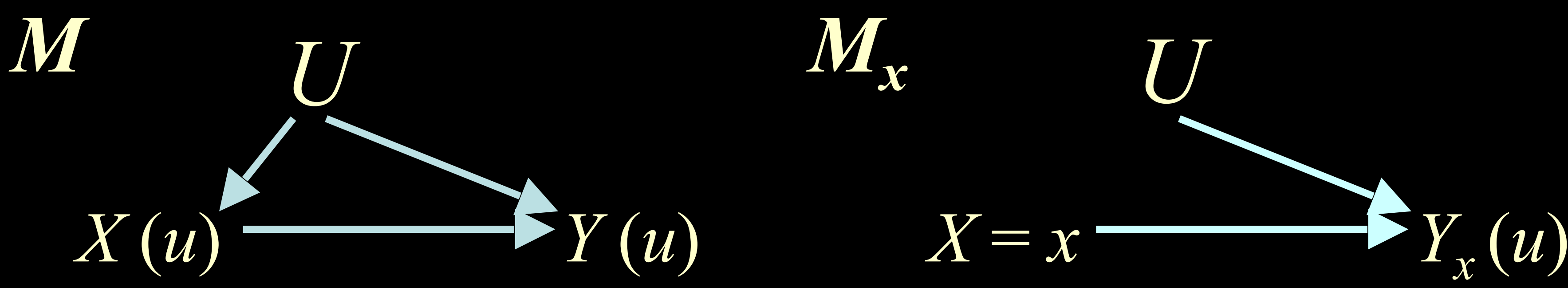
Definition: A structural causal model M (or, data generating model) is a tuple $(V, U, F, P(u))$, where

- $V = \{V_1,...,V_n\}$ are endogenous variables,
- $U = \{U_1,...,U_m\}$ are exogenous variables,
- $F = \{f_1,..., f_n\}$ are functions determining V,
  for each V_i, $V_i \leftarrow f_i(Pa_i, U_i)$, where $Pa_i \subset V$, $U_i \subset U$.
- $P(u)$ is a distribution over U.

(Axiomatic characterization [Halpern, Galles, Pearl, 1998].)

COUNTERFACTUALS IN A NUTSHELL

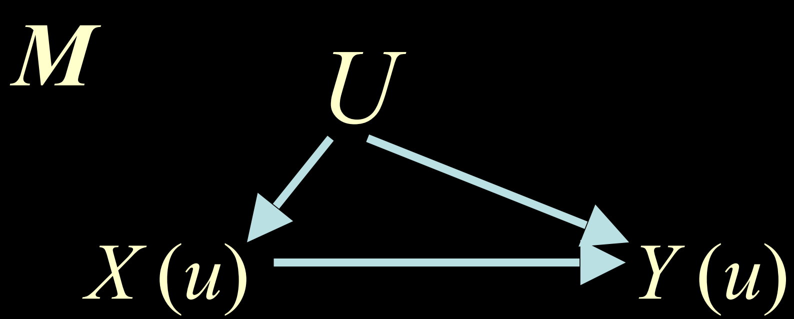
Definition:
Given a SCM model M, the potential outcome $\boldsymbol{Y_x(u)}$ for unit u is equal to the solution for Y in a mutilated model M_x, in which the equation for X is replaced by $X = x$.

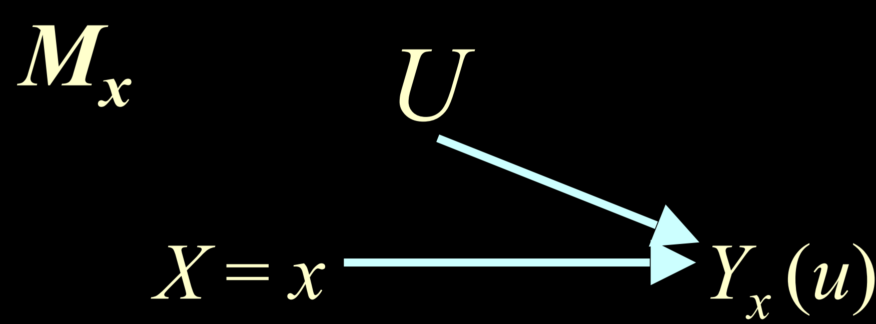


COUNTERFACTUALS IN A NUTSHELL

Definition:

Given a SCM model M, the potential outcome $\boldsymbol{Y_x(u)}$ 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 Law of Structural Counterfactuals:

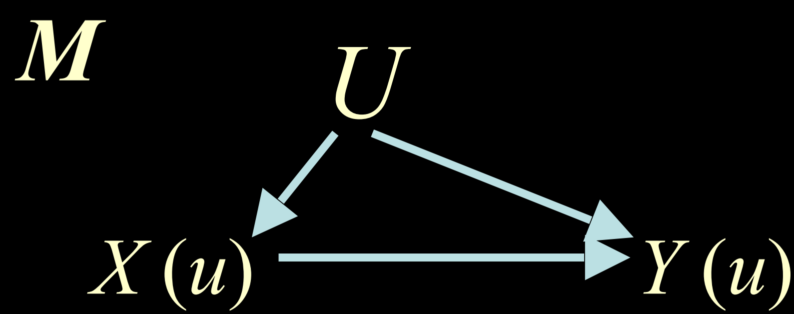
$$\boxed{Y_x(u) \triangleq Y_{M_x}(u)}$$

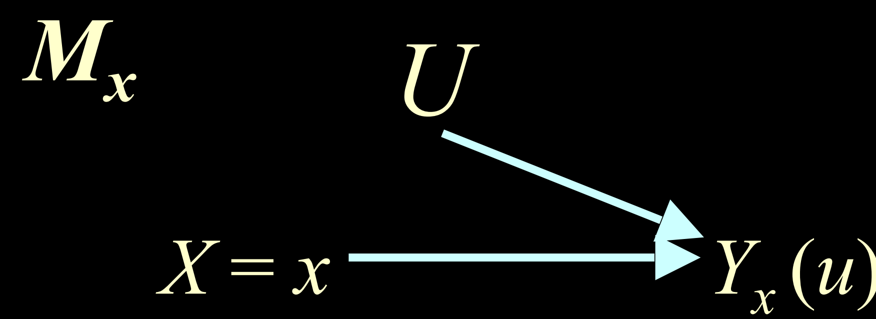
COUNTERFACTUALS IN A NU

If $Y = V_i$,
and $X = V \setminus Y$,
$Y_x(u) = f_i(pa_i, u)$

Definition:

Given a SCM model M, the potential outcome $\boldsymbol{Y_x(u)}$ 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 Law of Structural Counterfactuals:

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

SCM → PEARL (CAUSAL) HIERARCHY

SCM → CAUSAL HIERARCHY

	Level (Symbol)	Typical Activity	Typical Question	Examples
1	Associational $P(y \mid x)$	Seeing	What is? How would seeing X change my belief in Y?	What does a symptom tell us about the disease?

SCM → CAUSAL HIERARCHY

	Level (Symbol)	Typical Activity	Typical Question	Examples
1 👀	Associational $P(y \mid x)$	Seeing ML - (Un)Supervised DNN, Bayes Net Regression	What is? How would seeing X change my belief in Y?	What does a symptom tell us about the disease?

SCM → CAUSAL HIERARCHY

	Level (Symbol)	Typical Activity	Typical Question	Examples
1 👀	Associational $P(y \mid x)$	Seeing ML - (Un)Supervised DNN, Bayes Net Regression	What is? How would seeing X change my belief in Y?	What does a symptom tell us about the disease?
2 💪	Interventional $P(y \mid do(x), c)$	Doing	What if? What if I do X?	What if I take aspirin, will my headache be cured?

SCM → CAUSAL HIERARCHY

	Level (Symbol)	Typical Activity	Typical Question	Examples
1 👀	Associational $P(y \mid x)$	Seeing ML - (Un)Supervised DNN, Bayes Net Regression	What is? How would seeing X change my belief in Y?	What does a symptom tell us about the disease?
2 💪	Interventional $P(y \mid do(x), c)$	Doing ML - Reinforcement Causal Bayes Net	What if? What if I do X?	What if I take aspirin, will my headache be cured?

SCM → CAUSAL HIERARCHY

Level (Symbol)	Typical Activity	Typical Question	Examples
1 👀 Associational $P(y \mid x)$	Seeing ML - (Un)Supervised DNN, Bayes Net Regression	What is? How would seeing X change my belief in Y?	What does a symptom tell us about the disease?
2 💪 Interventional $P(y \mid do(x), c)$	Doing ML - Reinforcement Causal Bayes Net	What if? What if I do X?	What if I take aspirin, will my headache be cured?
3 🧠 Counterfactual $P(y_x \mid x', y')$	Imagining, Retrospection	Why? What if I had acted differently?	Was it the aspirin that stopped my headache?

SCM → CAUSAL HIERARCHY

Level (Symbol)	Typical Activity	Typical Question	Examples
1 Associational 👀 $P(y \mid x)$	Seeing ML - (Un)Supervised DNN, Bayes Net Regression	What is? How would seeing X change my belief in Y?	What does a symptom tell us about the disease?
2 Interventional 💪 $P(y \mid do(x), c)$	Doing ML - Reinforcement Causal Bayes Net	What if? What if I do X?	What if I take aspirin, will my headache be cured?
3 Counterfactual 🧠 $P(y_x \mid x', y')$	Imagining, Retrospection Structural Causal Model	Why? What if I had acted differently?	Was it the aspirin that stopped my headache?

SCM → CAUSAL HIERARCHY

	Level (Symbol)	Typical Activity	Typical Question	Examples
1 👀	Associational $P(y \mid x)$	Seeing ML - (Un)Supervised DNN, Bayes Net Regression	part 1	
2 💪	Interventional $P(y \mid do(x), c)$	Doing ML - Reinforcement Causal Bayes Net	What if? What if I do X?	What if I take aspirin, will my headache be cured?
3 🧠	Counterfactual $P(y_x \mid x', y')$	Imagining, Retrospection Structural Causal Model	Why? What if I had acted differently?	Was it the aspirin that stopped my headache?

SCM → CAUSAL HIERARCHY

	Level (Symbol)	Typical Activity	Typical Question	Examples
1 👀	Associational $P(y \mid x)$	Seeing ML - (Un)Supervised DNN, Bayes Net Regression	part 1	
2 💪	Interventional $P(y \mid do(x), c)$	Doing ML - Reinforcement Causal Bayes Net	part 2	
3 🧠	Counterfactual $P(y_x \mid x', y')$	Imagining, Retrospection Structural Causal Model	Why? What if I had acted differently?	Was it the aspirin that stopped my headache?

SCM → CAUSAL HIERARCHY

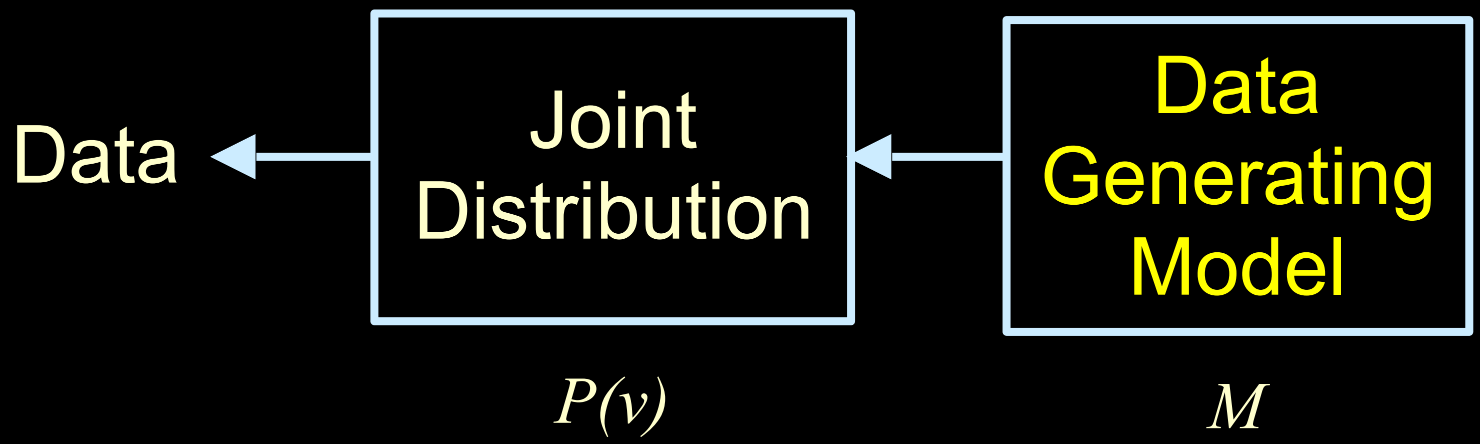
Level (Symbol)	Typical Activity	Typical Question	Examples
1 Associational 👀 $P(y \mid x)$	Seeing ML - (Un)Supervised DNN, Bayes Net Regression	part 1	
2 Interventional 💪 $P(y \mid do(x), c)$	Doing ML - Reinforcement Causal Bayes Net	part 2	
3 Counterfactual 🧠 $P(y_x \mid x', y')$	Imagining, Retrospection Structural Causal Mod	part 3	

FIRST LAYER OF THE CAUSAL HIERARCHY

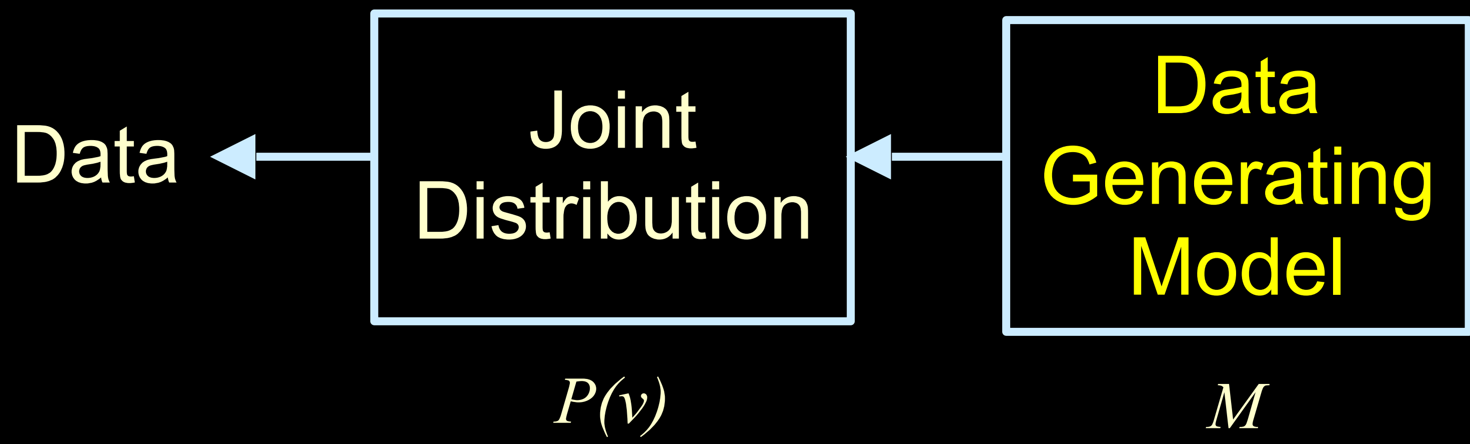
ASSOCIATIONAL

(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 an observational 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*.

ENCODING PROBABILISTIC INVARIANCE (GRAPHOIDS)

ENCODING PROBABILISTIC INVARIANCE (GRAPHOIDS)

Statements of the form:

ENCODING PROBABILISTIC INVARIANCE (GRAPHOIDS)

Statements of the form:

"X is irrelevant to Y given that we know Z",

ENCODING PROBABILISTIC INVARIANCE (GRAPHOIDS)

Statements of the form:

"X is irrelevant to Y given that we know Z",

where X, Y, and Z are sets of variables.

ENCODING PROBABILISTIC INVARIANCE (GRAPHOIDS)

Statements of the form:

"X is irrelevant to Y given that we know Z",

where X, Y, and Z are sets of variables.

ENCODING PROBABILISTIC INVARIANCE (GRAPHOIDS)

Statements of the form:

“X is irrelevant to Y given that we know Z”,

where X, Y, and Z are sets of variables.

The notion of irrelevance can be considered under different interpretations, including probabilistic, relational, and correlational.

ENCODING PROBABILISTIC INVARIANCE (GRAPHOIDS)

Statements of the form:

"X is irrelevant to Y given that we know Z",

where X, Y, and Z are sets of variables.

The notion of irrelevance can be considered under different interpretations, including probabilistic, relational, and correlational.

ENCODING PROBABILISTIC INVARIANCE (GRAPHOIDS)

Statements of the form:

“X is irrelevant to Y given that we know Z”,

where X, Y, and Z are sets of variables.

The notion of irrelevance can be considered under different interpretations, including probabilistic, relational, and correlational.

(See also J. Pearl (1988) and P. David (1979).)

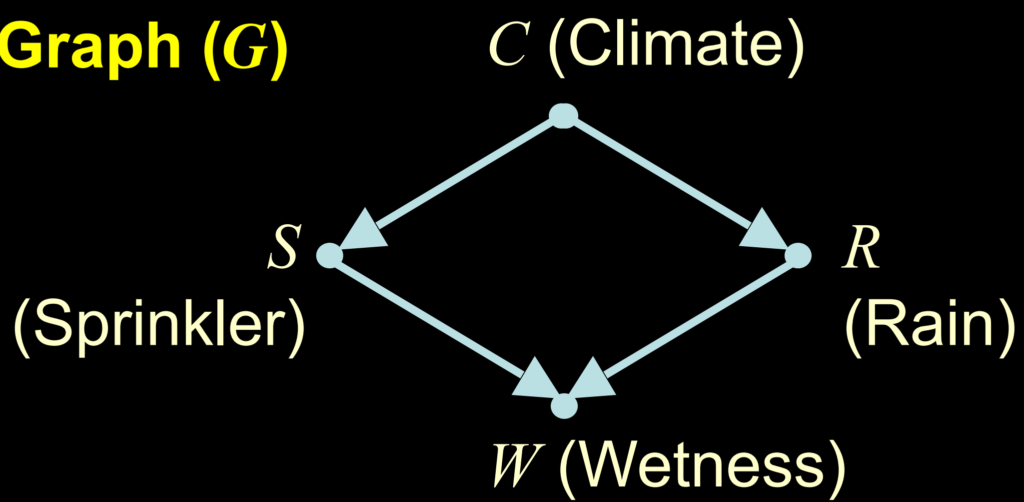
THE LAW OF CONDITIONAL INDEPENDENCE

SCM (M)

$C \leftarrow f_c(U_c)$
$S \leftarrow f_S(C, U_S)$
$R \leftarrow f_R(C, U_R)$
$W \leftarrow f_W(S, R, U_W)$

THE LAW OF CONDITIONAL INDEPENDENCE

Graph (G)

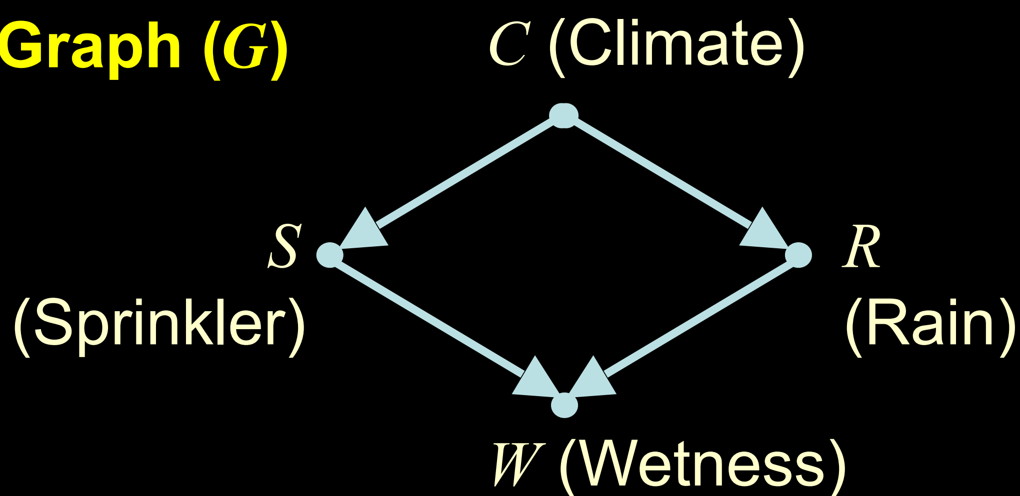


SCM (M)

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

THE LAW OF CONDITIONAL INDEPENDENCE

Graph (G)



SCM (M)

$$C \leftarrow f_c(U_c)$$
$$S \leftarrow f_S(C, U_S)$$
$$R \leftarrow f_R(C, U_R)$$
$$W \leftarrow f_W(S, R, U_W)$$

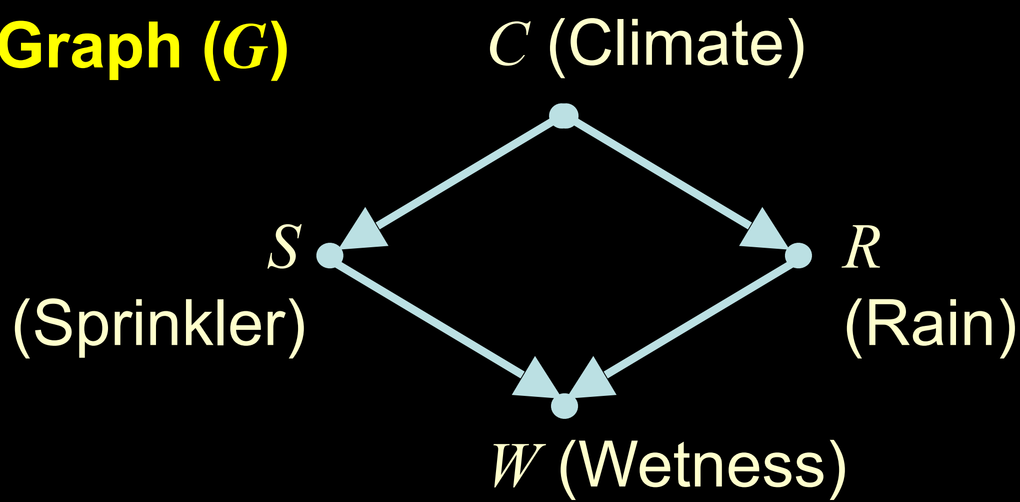
*** Surprising result:**

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: READING CONSTRAINTS IMPLIED IN THE DATA

Graph (*G*)



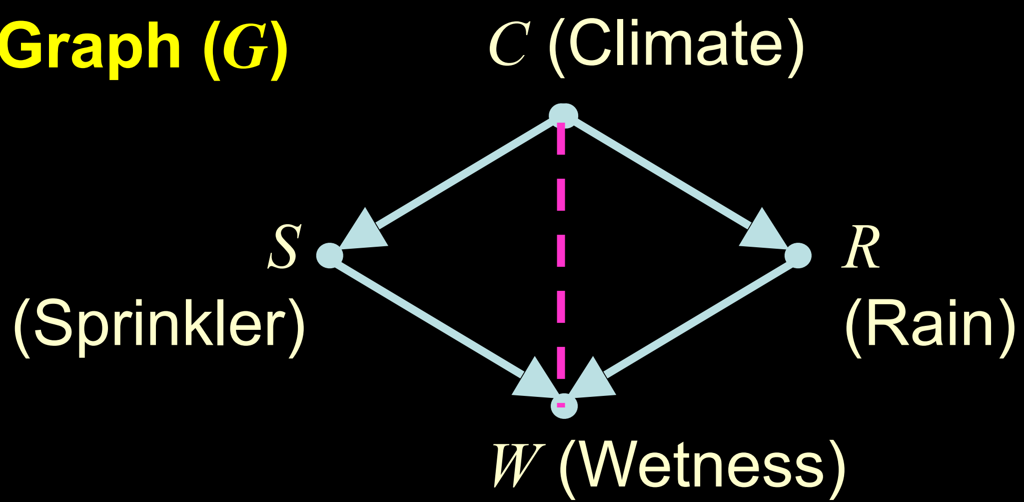
SCM (*M*)

$C \leftarrow f_c(U_c)$
$S \leftarrow f_S(C, U_S)$
$R \leftarrow f_R(C, U_R)$
$W \leftarrow f_W(S, R, U_W)$

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

D-SEPARATION: READING CONSTRAINTS IMPLIED IN THE DATA

Graph (G)



SCM (M)

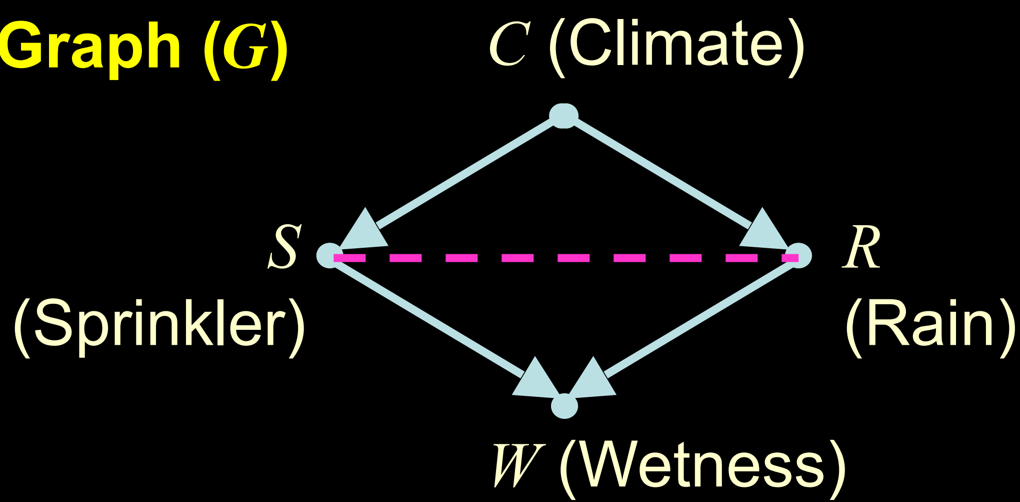
$C \leftarrow f_C(U_C)$
$S \leftarrow f_S(C, U_S)$
$R \leftarrow f_R(C, U_R)$
$W \leftarrow f_W(S, R, U_W)$

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

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

D-SEPARATION: READING CONSTRAINTS IMPLIED IN THE DATA

Graph (G)



SCM (M)

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

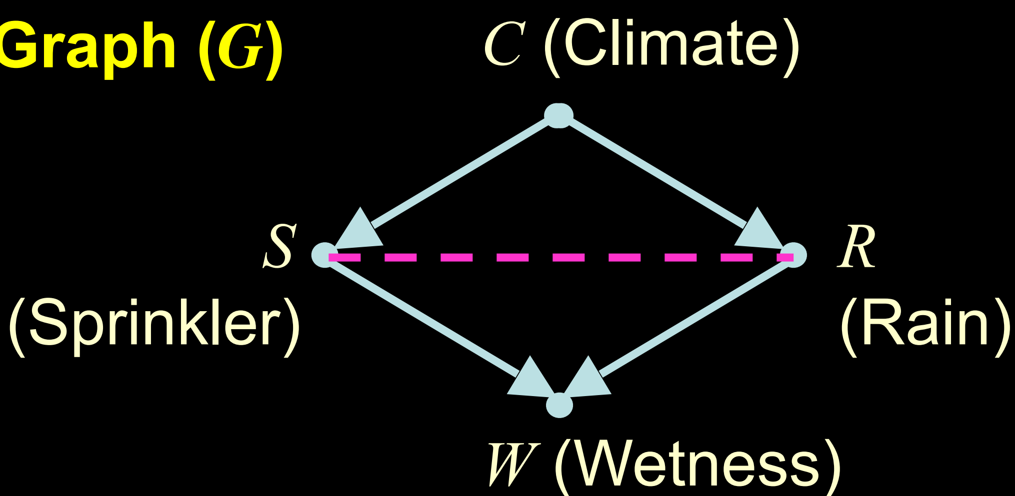
Every missing arrow advertises an independency, conditional on a separating set.

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

$S \perp\!\!\!\perp R \mid C$

D-SEPARATION: READING CONSTRAINTS IMPLIED IN THE DATA

Graph (G)



SCM (M)

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

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

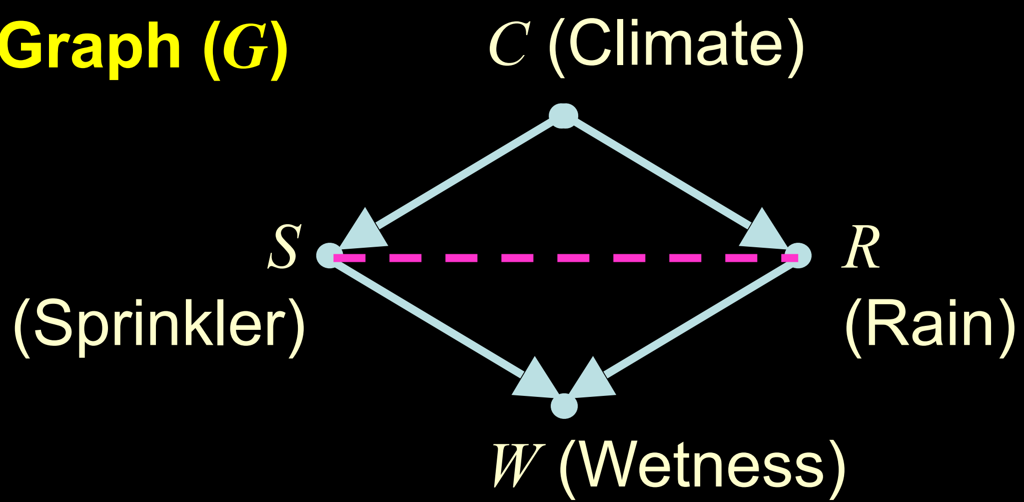
$C \perp\!\!\!\perp W \mid (S, R)$

$S \perp\!\!\!\perp R \mid C$

Applications:

D-SEPARATION: READING CONSTRAINTS IMPLIED IN THE DATA

Graph (G)



SCM (M)

$C \leftarrow f_c(U_c)$
$S \leftarrow f_S(C, U_S)$
$R \leftarrow f_R(C, U_R)$
$W \leftarrow f_W(S, R, U_W)$

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

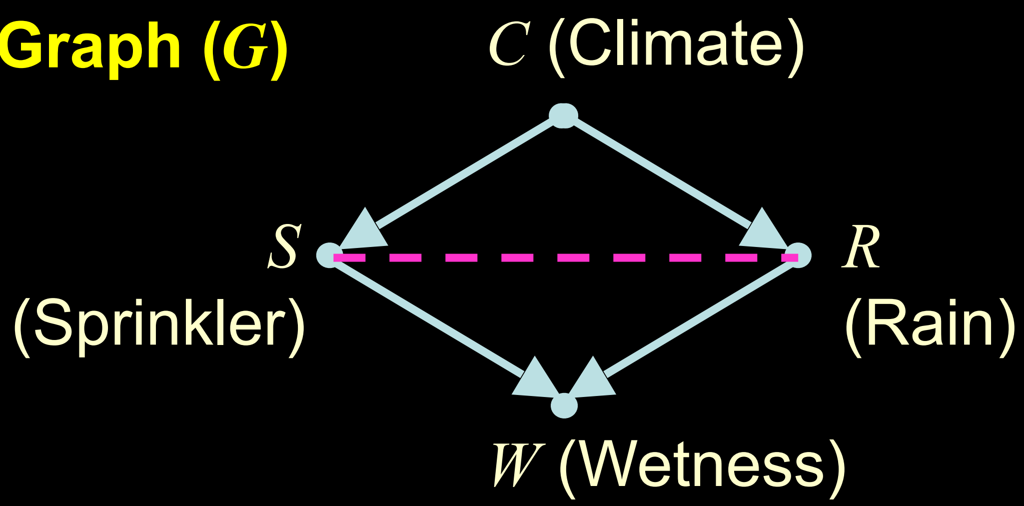
$C \perp\!\!\!\perp W \mid (S, R)$

$S \perp\!\!\!\perp R \mid C$

Applications:

D-SEPARATION: READING CONSTRAINTS IMPLIED IN THE DATA

Graph (G)



SCM (M)

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

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

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

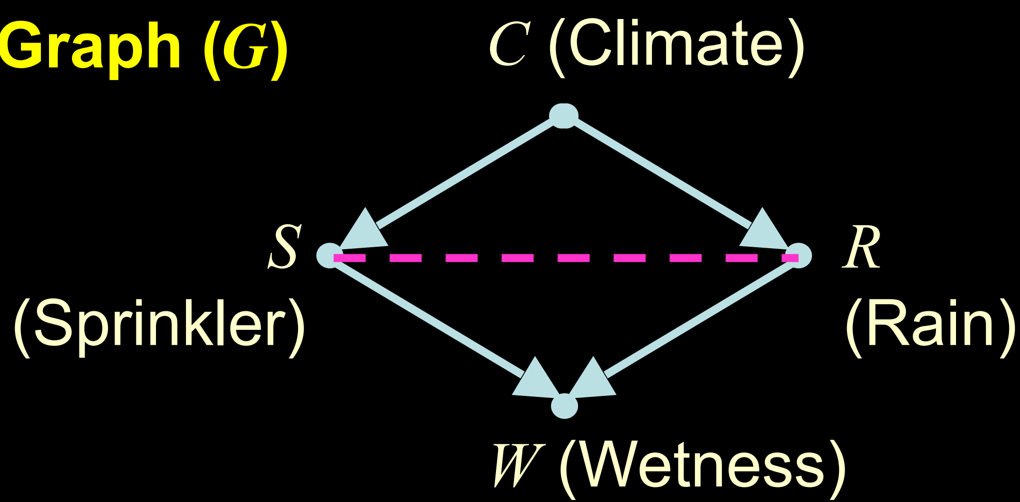
$S \perp\!\!\!\perp R \mid C$

Applications:

1. Structural learning

D-SEPARATION: READING CONSTRAINTS IMPLIED IN THE DATA

Graph (G)



SCM (M)

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

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

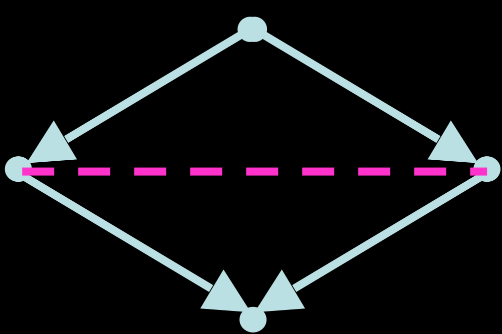
$$C \perp\!\!\!\perp W \mid (S, R)$$
$$S \perp\!\!\!\perp R \mid C$$

Applications:

1. Structural learning
2. Model testing

D-SEPARATION: READING CONSTRAINTS IMPLIED IN THE DATA

Graph (G)



SCM (M)

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

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

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

$S \perp\!\!\!\perp R \mid C$

Applications:

1. Structural learning
2. Model testing
3. (Efficient) Probabilistic inference

GRAPH SEPARATION (D-SEPARATION)

GRAPH SEPARATION (D-SEPARATION)

$(X \perp\!\!\!\perp Y \mid Z)$

normal valve

$x \longleftarrow z \longrightarrow y$

$x \longrightarrow z \longrightarrow y$

$x \longleftarrow z \longleftarrow y$

abnormal valve

$x \longrightarrow z \longleftarrow y$

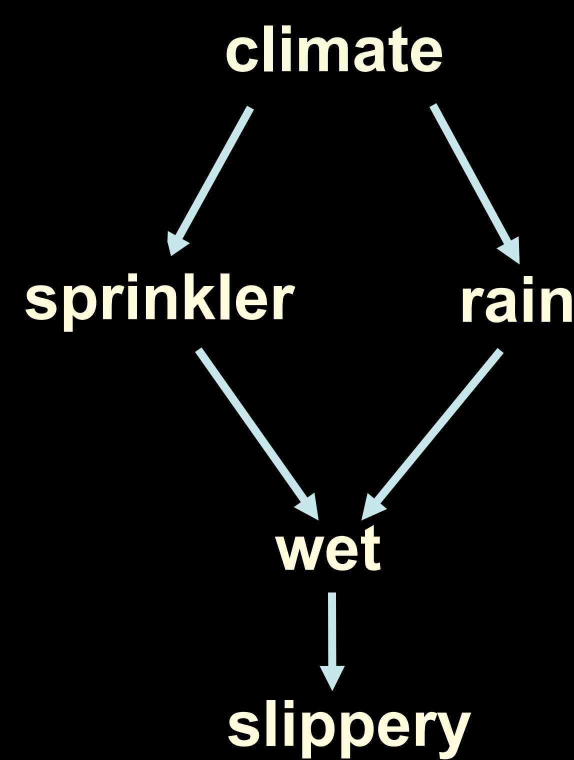
$$\begin{array}{ccccc} x & \longrightarrow & z & \longleftarrow & y \\ & & \downarrow & & \\ & & w & & \end{array}$$

$(X \perp\!\!\!\perp Y)$

$(X \not\perp\!\!\!\perp Y \mid Z)$

GRAPH SEPARATION (D-SEPARATION)

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



GRAPH SEPARATION (D-SEPARATION)

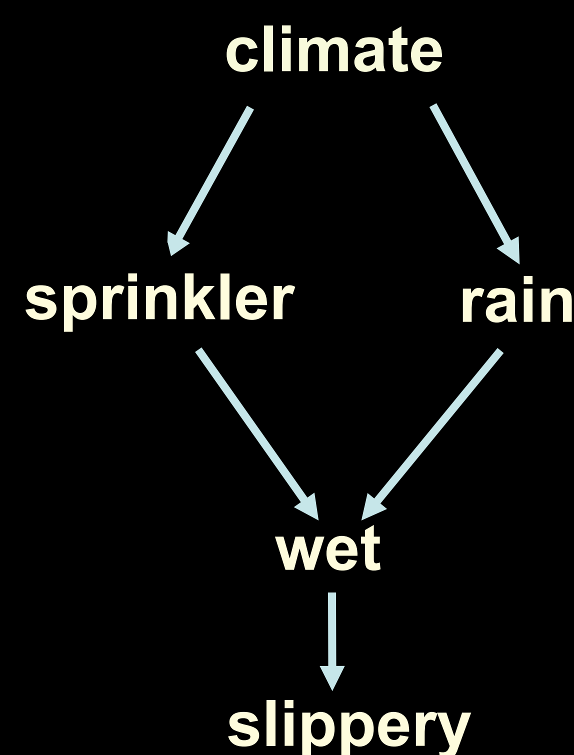
normal valve	abnormal valve
$x \longleftarrow z \longrightarrow y$	$x \longrightarrow z \longleftarrow y$
$x \longrightarrow z \longrightarrow y$	$x \longrightarrow z \longleftarrow y$, $z \longrightarrow w$
$x \longleftarrow z \longleftarrow y$	

$(X \perp\!\!\!\perp Y \mid Z)$

$(X \perp\!\!\!\perp Y)$

$(X \not\!\perp\!\!\!\perp Y \mid Z)$

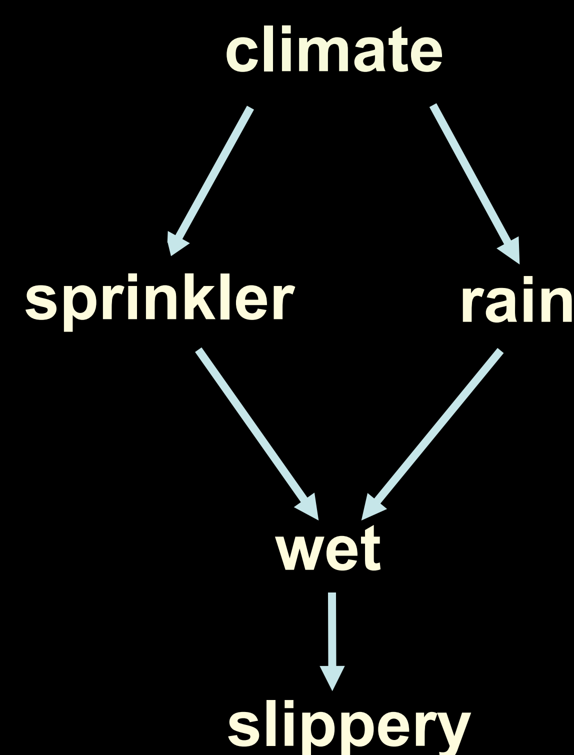
$CI_1 : (\text{Wet} \perp\!\!\!\perp \text{Sprinkler})$



GRAPH SEPARATION (D-SEPARATION)

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

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

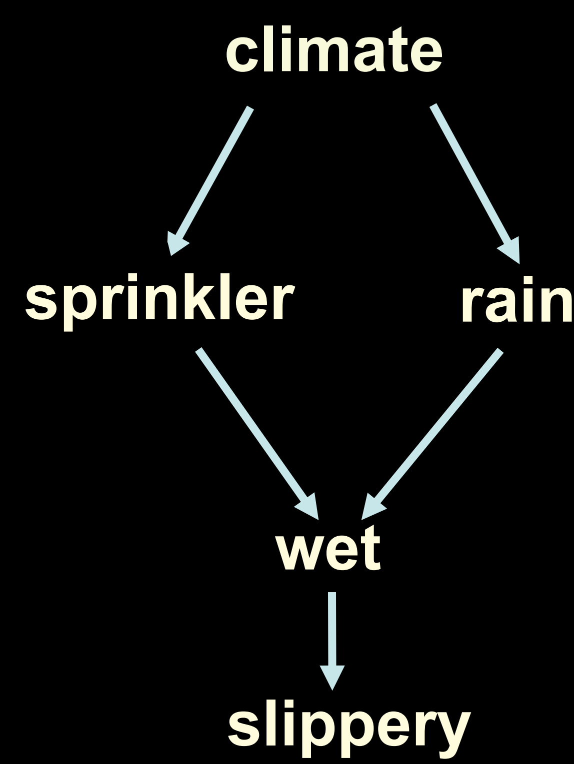


GRAPH SEPARATION (D-SEPARATION)

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

$CI_1 : (\text{Wet} \perp\!\!\!\perp \text{Sprinkler})$

$CI_2 : (\text{Wet} \perp\!\!\!\perp \text{Climate} \mid \text{Sprinkler})$

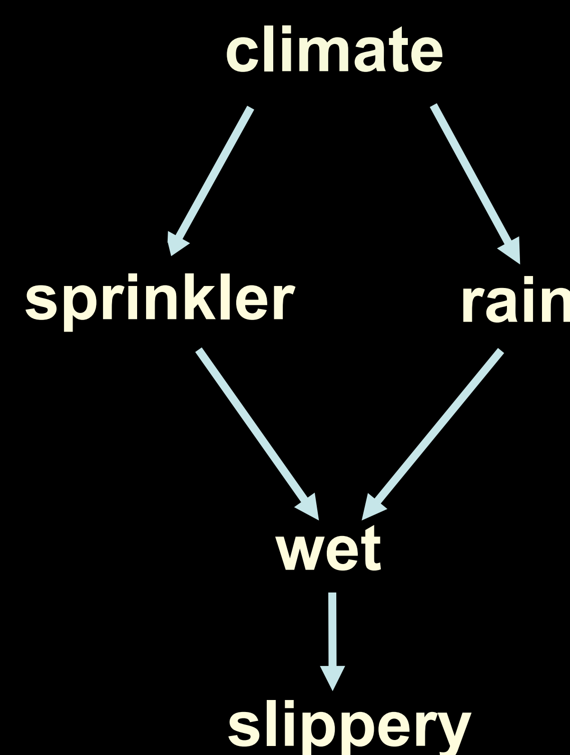


GRAPH SEPARATION (D-SEPARATION)

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

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

✗ CI_2 : (Wet $\perp\!\!\!\perp$ Climate | Sprinkler)



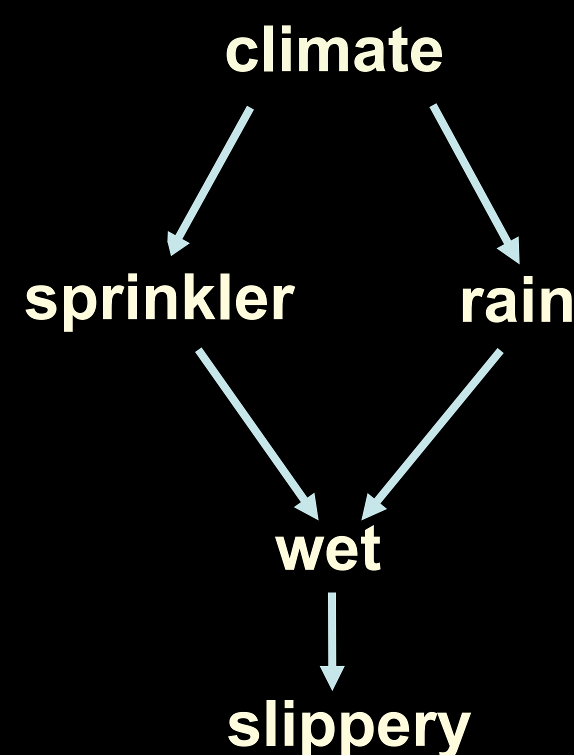
GRAPH SEPARATION (D-SEPARATION)

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

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

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

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



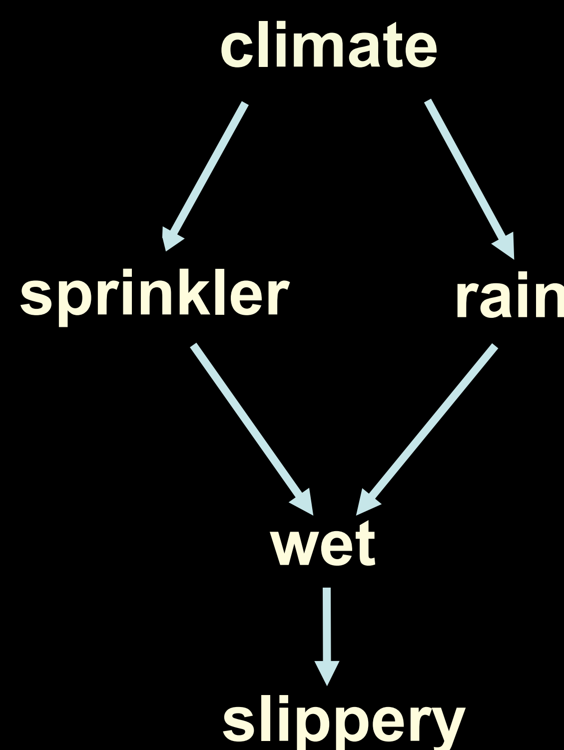
GRAPH SEPARATION
(D-SEPARATION)

	normal valve	abnormal valve	
$(X \perp\!\!\!\perp Y \mid Z)$	$x \longleftarrow z \longrightarrow y$	$x \longrightarrow z \longleftarrow y$	$(X \perp\!\!\!\perp Y)$
	$x \longrightarrow z \longrightarrow y$	$\begin{tikzcd} x \arrow[r] & z \arrow[d] & y \arrow[l] \\ & w & \end{tikzcd}$	$(X \not\perp\!\!\!\perp Y \mid Z)$
	$x \longleftarrow z \longleftarrow y$		

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

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

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



GRAPH SEPARATION (D-SEPARATION)

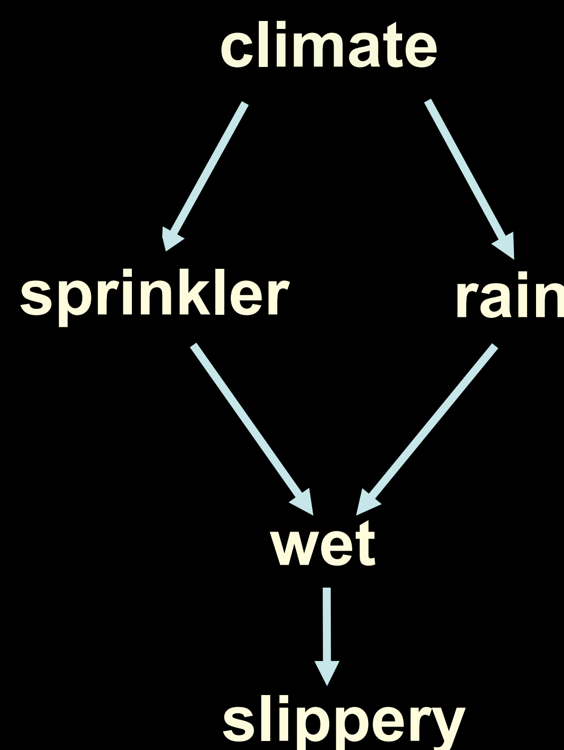
	normal valve	abnormal valve	
$(X \perp\!\!\!\perp Y \mid Z)$	$x \longleftarrow z \longrightarrow y$	$x \longrightarrow z \longleftarrow y$	$(X \perp\!\!\!\perp Y)$
	$x \longrightarrow z \longrightarrow y$	$x \longrightarrow z \longleftarrow y$, $z \longrightarrow w$	$(X \not\perp\!\!\!\perp Y \mid Z)$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

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

✗ CI_2 : (Wet $\perp\!\!\!\perp$ Climate | Sprinkler)

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

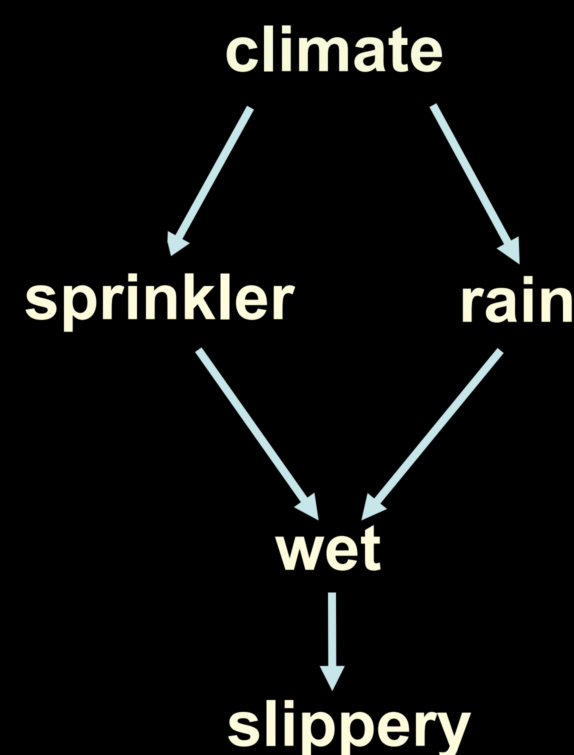
CI_4 : (Season $\perp\!\!\!\perp$ Wet | Sprinkler, Rain)



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



GRAPH SEPARATION (D-SEPARATION)

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

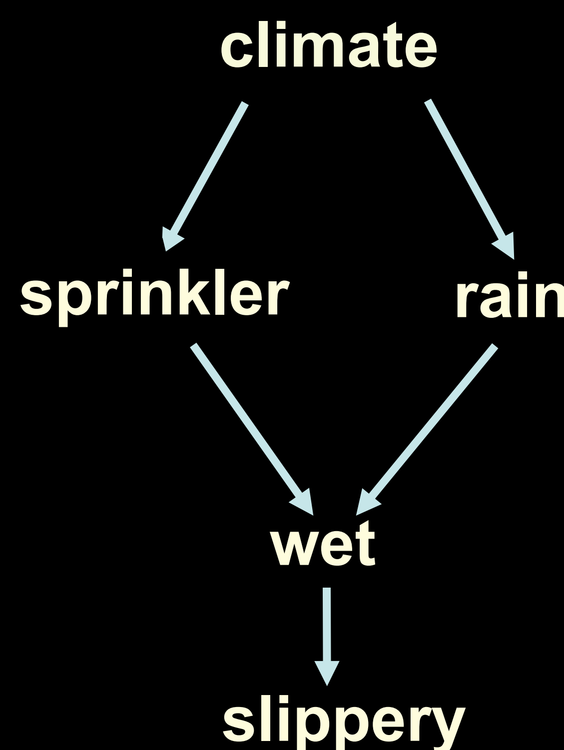
✘ CI_1 : (Wet $\perp\!\!\!\perp$ Sprinkler)

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

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

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

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



GRAPH SEPARATION (D-SEPARATION)

	normal valve	abnormal valve	
$(X \perp\!\!\!\perp Y \mid Z)$	$x \leftarrow z \rightarrow y$	$x \rightarrow z \leftarrow y$	$(X \perp\!\!\!\perp Y)$
	$x \rightarrow z \rightarrow y$	$\begin{tikzcd} x \arrow[r] & z \arrow[d] & y \arrow[l] \\ & w & \end{tikzcd}$	$(X \not\perp\!\!\!\perp Y \mid Z)$
	$x \leftarrow z \leftarrow y$		

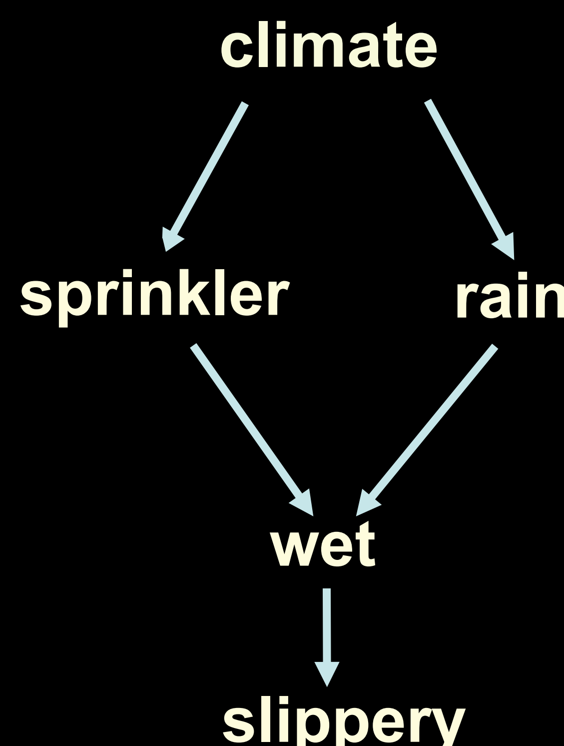
✗ CI_1 : (Wet $\perp\!\!\!\perp$ Sprinkler)

✗ CI_2 : (Wet $\perp\!\!\!\perp$ Climate | Sprinkler)

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

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

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



THE SECOND LAYER ON CAUSAL HIERARCHY:

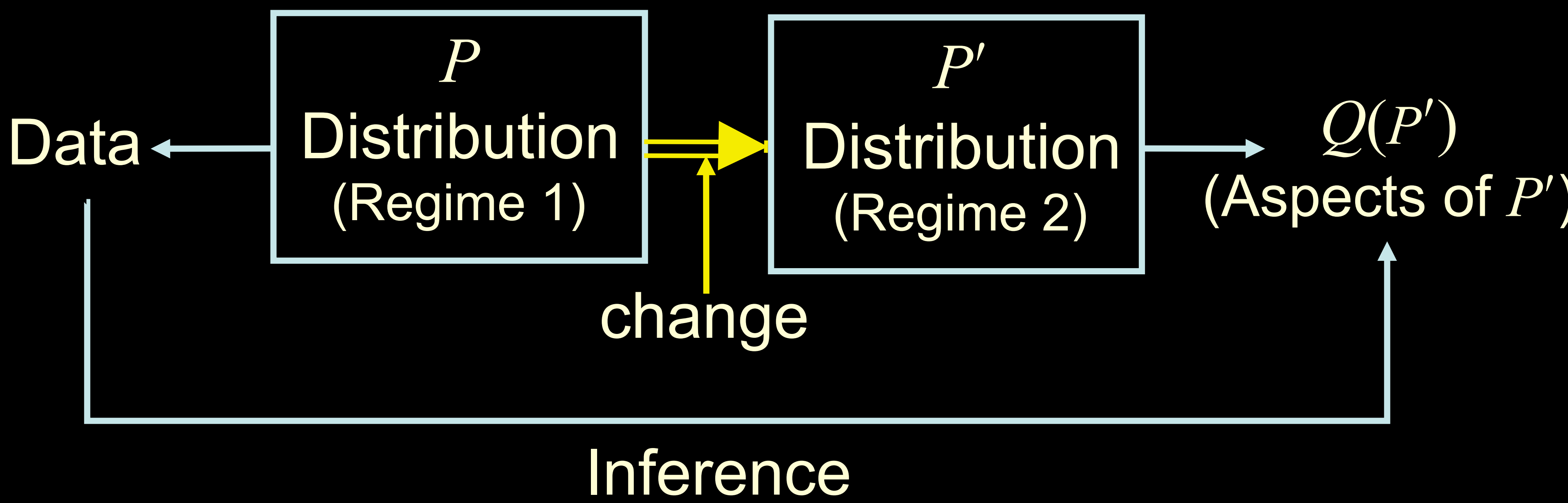
INTERVENTIONAL

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

OUTLINE

- Introduction
  * Data, data, data…
  * Basic definitions
- Technical Results
  * The truncated product formula
  * The back-door adjustment formula
  * The front-door adjustment formula
  * The do-calculus
- Tasks
  * Policy evaluation
  * Generalizability & Robustness
  * Decision-making & RL

CAUSAL INFERENCE: MOVING ACROSS 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).

Observation 1:

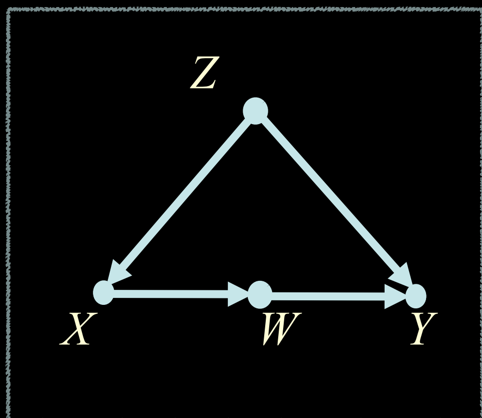
The distribution alone tells us nothing about change; it just describes static conditions of a population (under a specific regime).

Observation 2:

We need to be able to represent “change”, or how the population reacts when it undergoes change in regimes.

THE BIG PICTURE: THE CHALLENGE OF CAUSAL INFERENCE

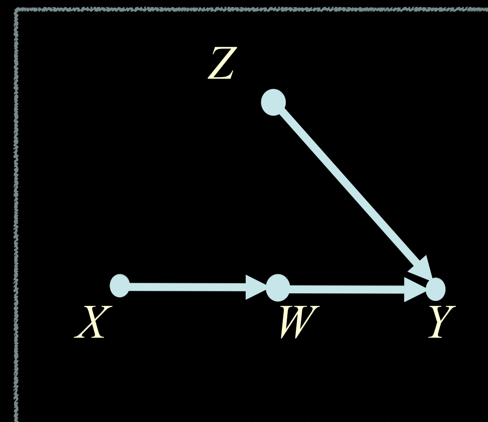
Real world



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

change

Alternative world

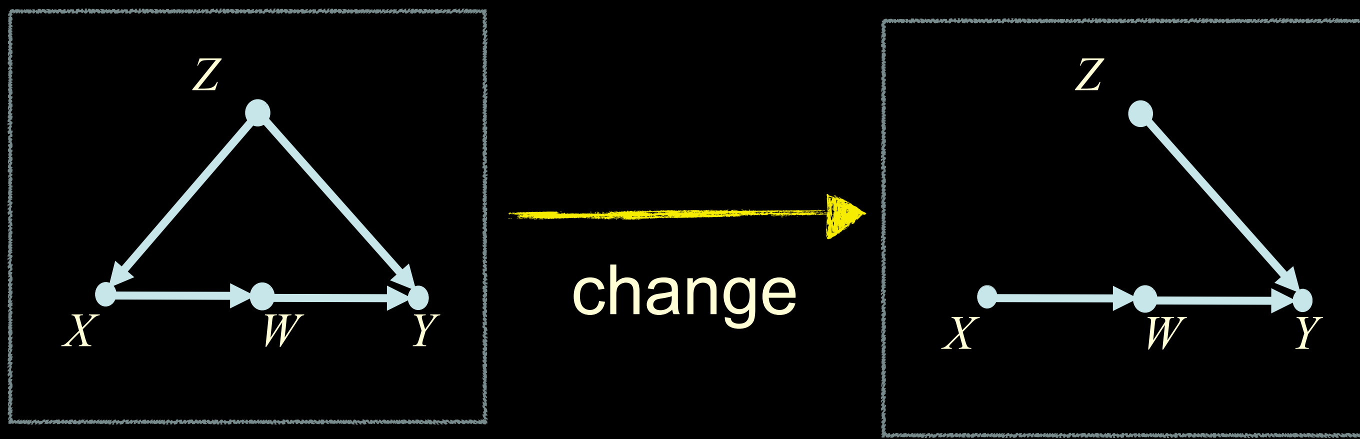


$P(y \mid do(x))$

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

THE BIG PICTURE: THE CHALLENGE OF CAUSAL INFERENCE

Real world | Alternative world

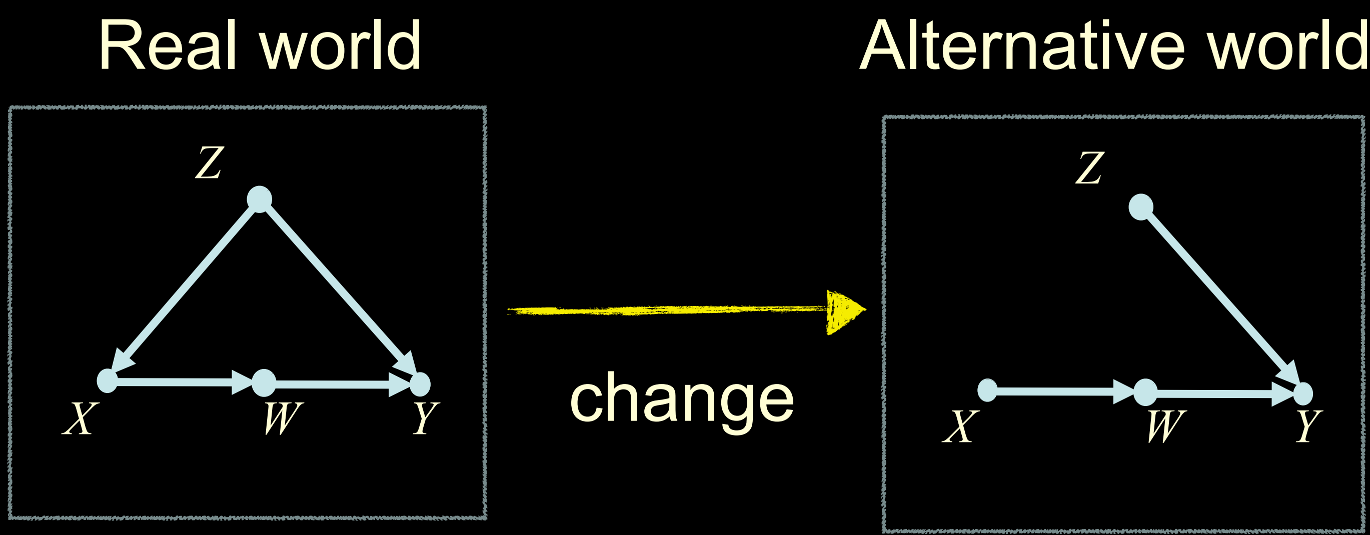


$P(z, x, w, y)$ | $P(y \mid do(x))$

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

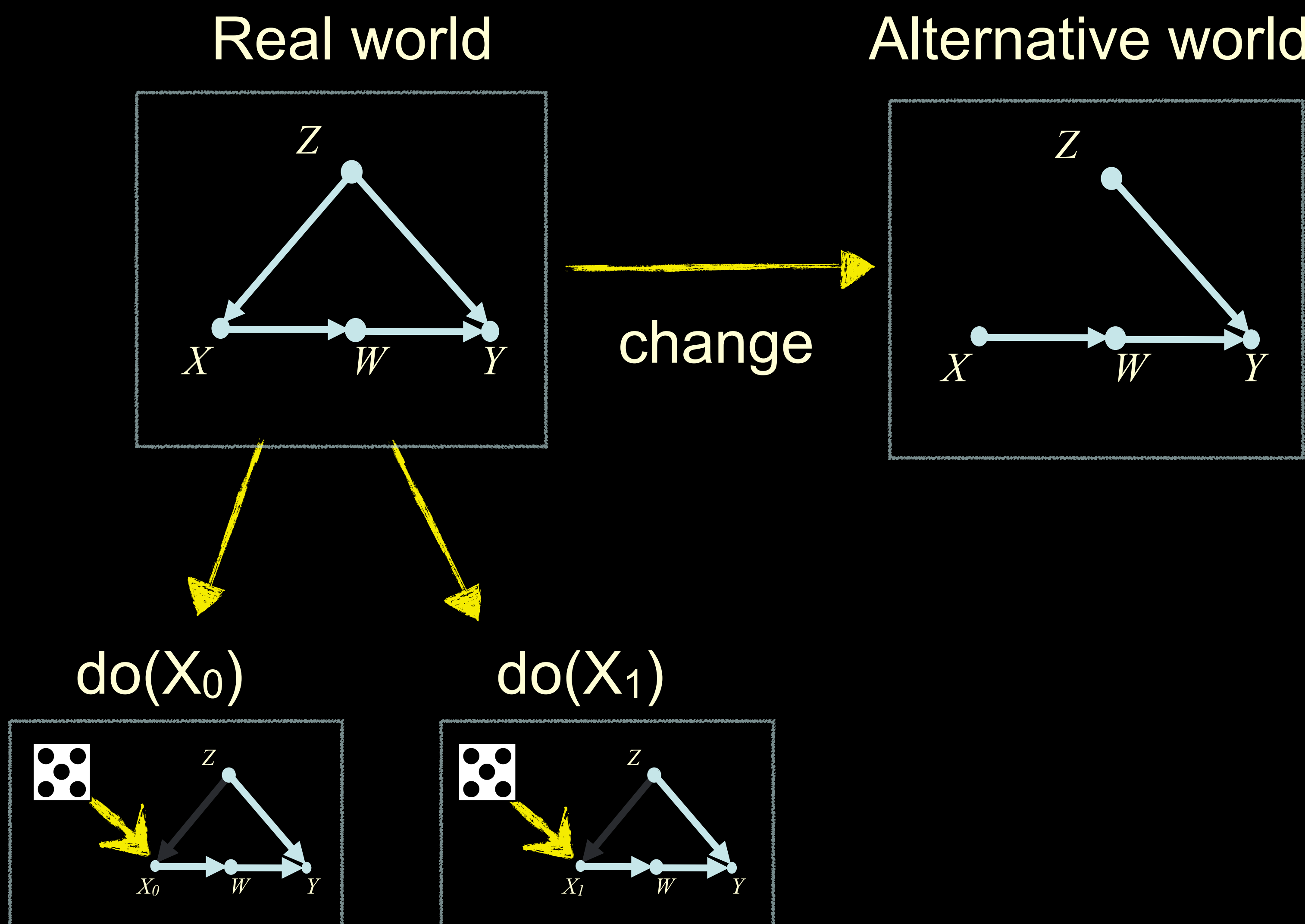
- Goal: how much Y changes with X if we vary X between two different constants free from the influence of Z.
- This is the definition of causal effect.

METHOD FOR COMPUTING CAUSAL EFFECTS: RANDOMIZED EXPERIMENTS



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

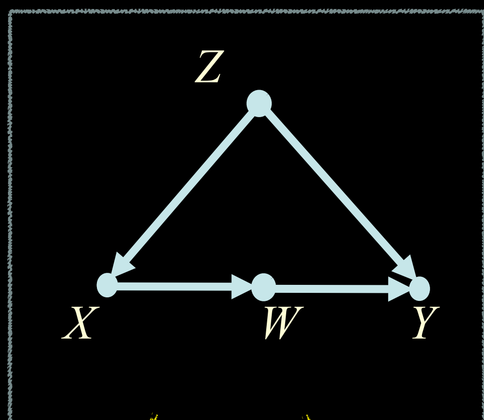
METHOD FOR COMPUTING CAUSAL EFFECTS: RANDOMIZED EXPERIMENTS



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

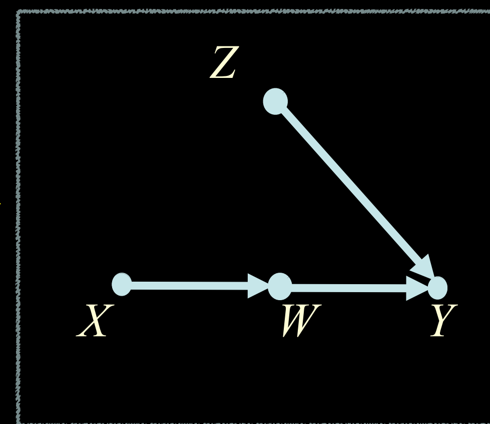
METHOD FOR COMPUTING CAUSAL EFFECTS: RANDOMIZED EXPERIMENTS

Real world



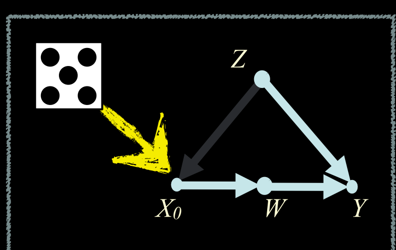
change

Alternative world

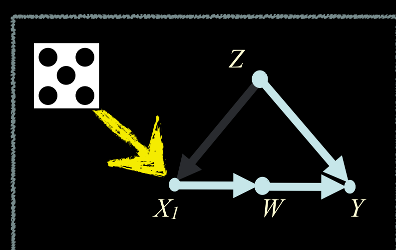


Z: age, sex
X: action
W: mediator
Y: outcome

do(X_0)



do(X_1)

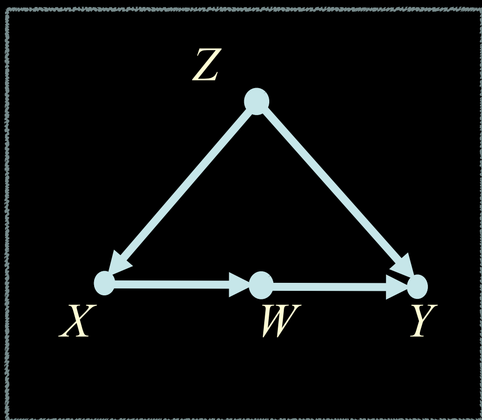


Randomization:

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

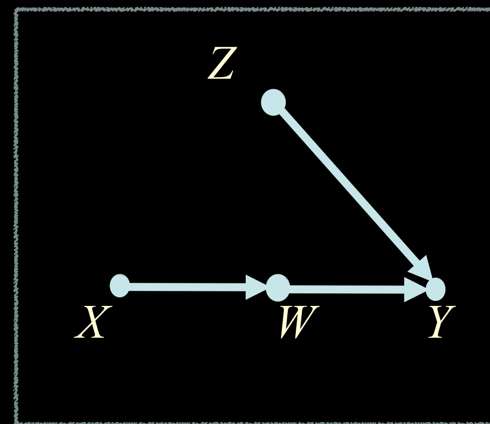
COMPUTING CAUSAL EFFECTS FROM OBSERVATIONAL DATA

Real world



change

Alternative world

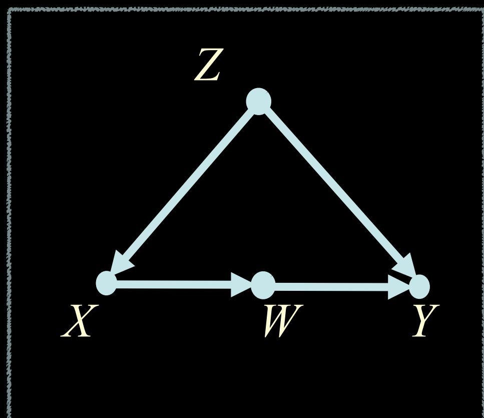


Z: age, sex
X: action
W: mediator
Y: outcome

P(z, x, w, y) ? P(y | do(x))

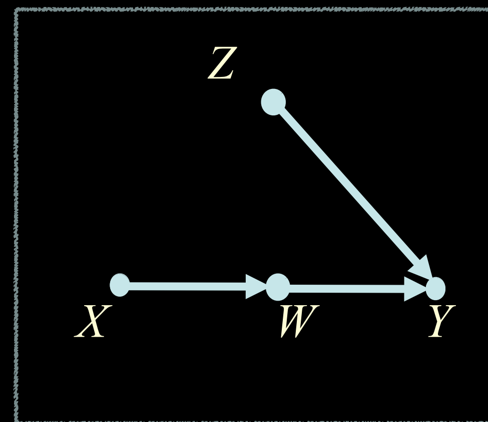
COMPUTING CAUSAL EFFECTS FROM OBSERVATIONAL DATA

Real world



change

Alternative world



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

P(z, x, w, y) ? P(y | do(x))

Questions:

* What is the relationship between P(z, x, w, y) and P(y | do(x))?
* Is P(y | do(x)) = P(y | x)?

EFFECT OF INTERVENTIONS

EFFECT OF INTERVENTIONS

- Def. 3.2.1 (Causal Effect)

Given two disjoint sets of variables, X and Y, the causal effect of X on Y, denoted as $P(y \mid do(x))$, is a function from X to the space of probability distributions of Y.

EFFECT OF INTERVENTIONS

- Def. 3.2.1 (Causal Effect)

Given two disjoint sets of variables, X and Y, the causal effect of X on Y, denoted as P(y | do(x)), is a function from X to the space of probability distributions of Y.

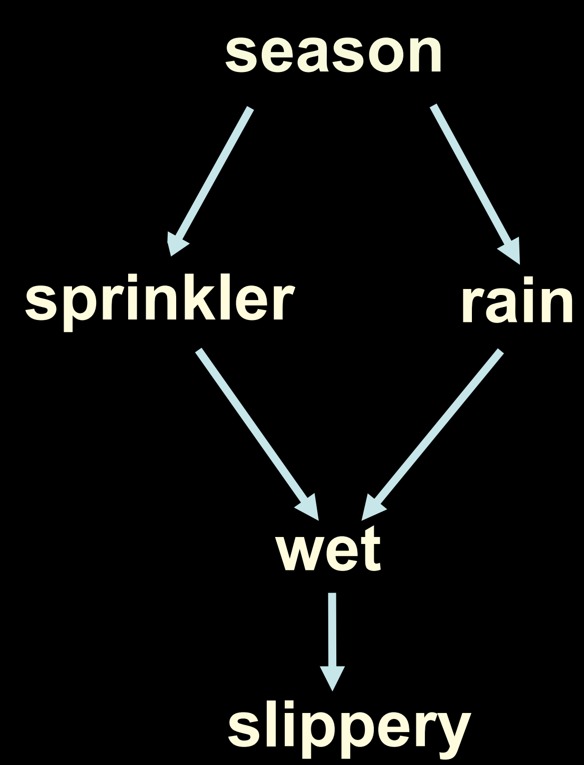
For each realization x of X, P(y | do(x)) gives the probability Y = y induced by deleting from the model all equations corresponding to variables in X and substituting X = x in the remaining equations.

COMPUTING CAUSAL EFFECTS FROM OBSERVATIONAL DATA

Queries:

Q_1 = Pr(wet | Sprinkler = on)

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

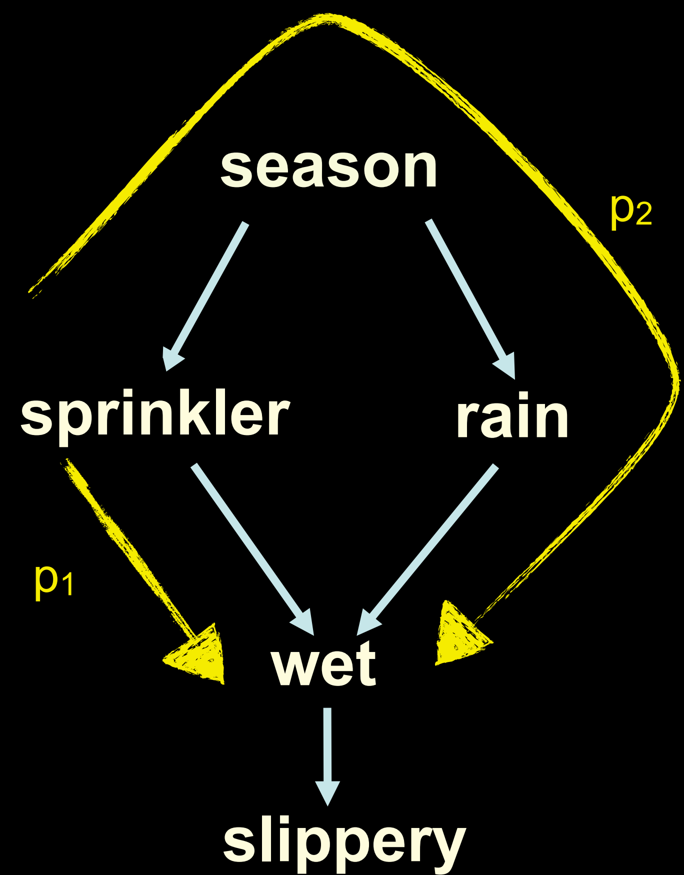


COMPUTING CAUSAL EFFECTS FROM OBSERVATIONAL DATA

Queries:

Q_1 = Pr(wet | Sprinkler = on)
= P(p_1) + P(p_2)

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

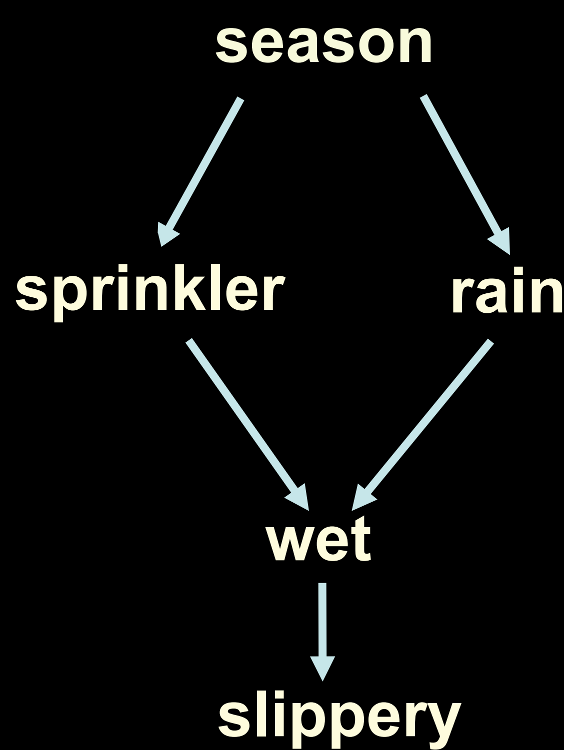


COMPUTING CAUSAL EFFECTS FROM OBSERVATIONAL DATA

Queries:

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

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

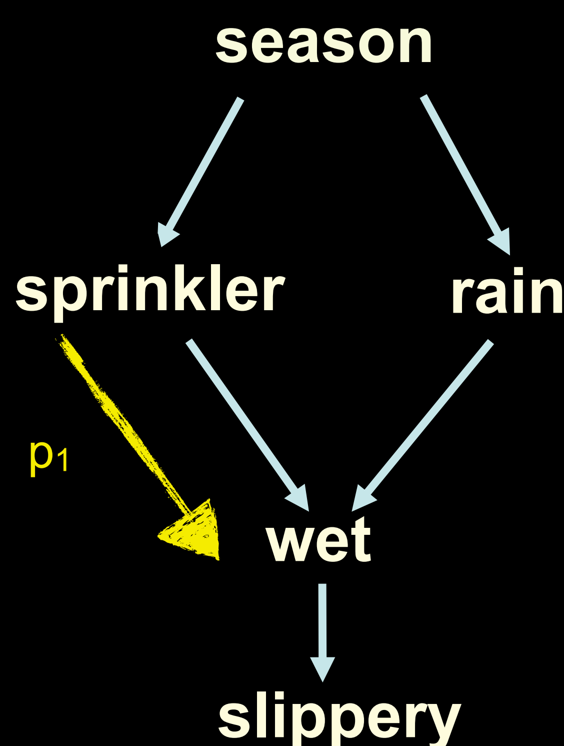


COMPUTING CAUSAL EFFECTS FROM OBSERVATIONAL DATA

Queries:

Q_1 = Pr(wet | Sprinkler = on)
= P(p_1) + P(p_2)

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

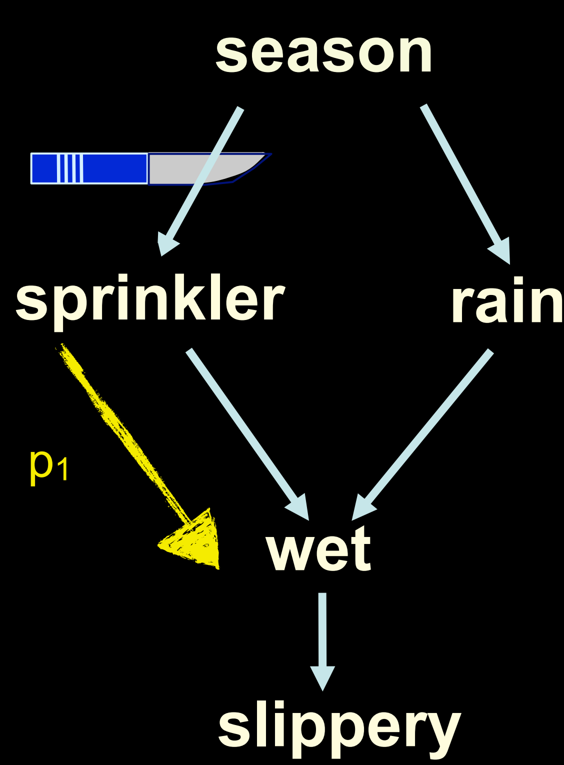


COMPUTING CAUSAL EFFECTS FROM OBSERVATIONAL DATA

Queries:

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

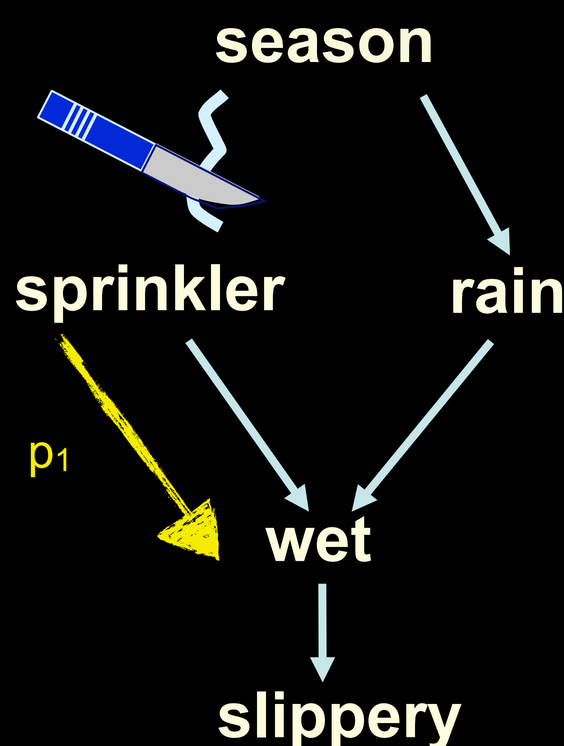
Q_2 = Pr(wet | do(Sprinkler = on))



COMPUTING CAUSAL EFFECTS FROM OBSERVATIONAL DATA

Queries:

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

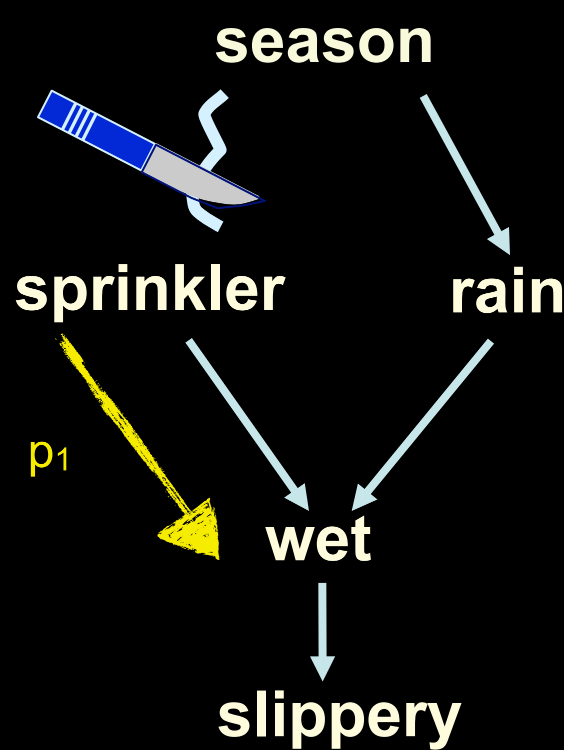


Q_2 = Pr(wet | do(Sprinkler = on))
= $P(p_1)$

COMPUTING CAUSAL EFFECTS FROM OBSERVATIONAL DATA

Queries:

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



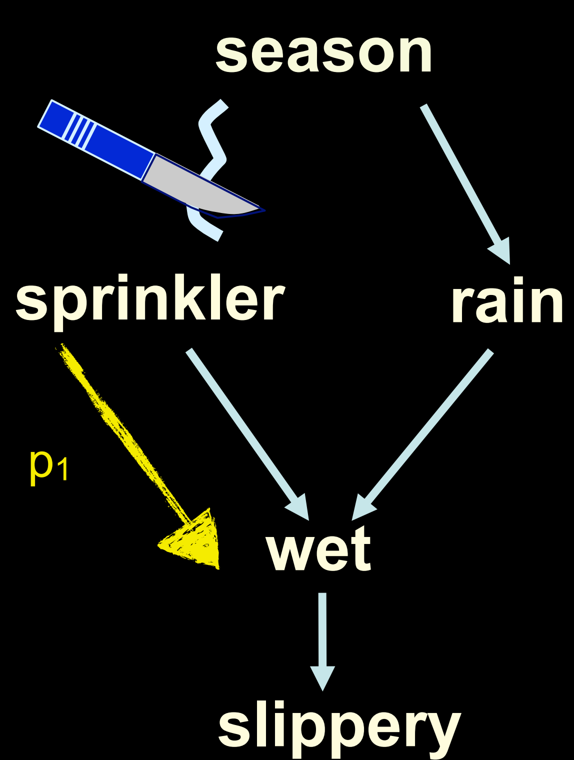
Q_2 = Pr(wet | do(Sprinkler = on))
= $P(p_1)$

COMPUTING CAUSAL EFFECTS FROM OBSERVATIONAL DATA

Queries:

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

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



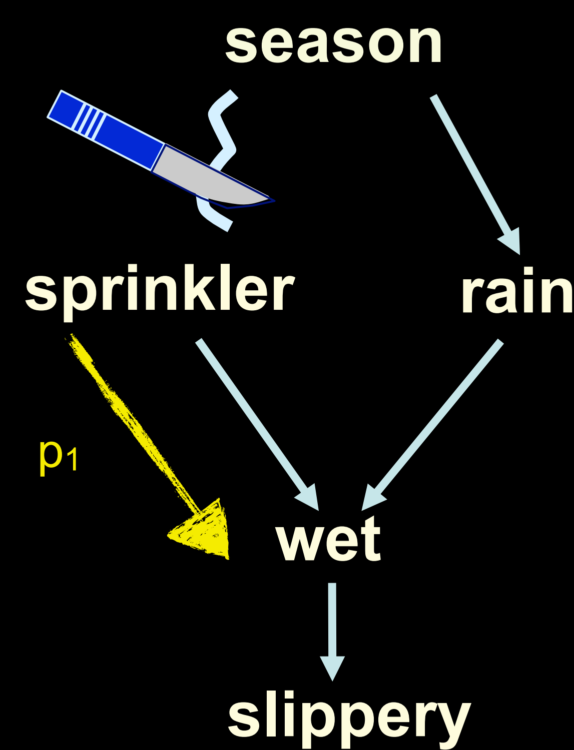
$$\sum_{Se,Ra,Sl} P(Se)\, P(Sp \mid Se)\, P(Ra \mid Se)\, P(We \mid Sp, Ra)\, P(Sl \mid We)$$

COMPUTING CAUSAL EFFECTS FROM OBSERVATIONAL DATA

Queries:

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

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



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

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:

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

THE IDENTIFIABILITY PROBLEM

THE IDENTIFIABILITY PROBLEM

- Def. 3.2.2 (Causal Effect Identifiability)

The causal effect of X on Y is said to be identifiable from a graph G if the quantity P(y | do(x)) can be computed uniquely from a positive probability of the observed variables.

THE IDENTIFIABILITY PROBLEM

- Def. 3.2.2 (Causal Effect Identifiability)

The causal effect of X on Y is said to be identifiable from a graph G if the quantity P(y | do(x)) can be computed uniquely from a positive probability of the observed variables.

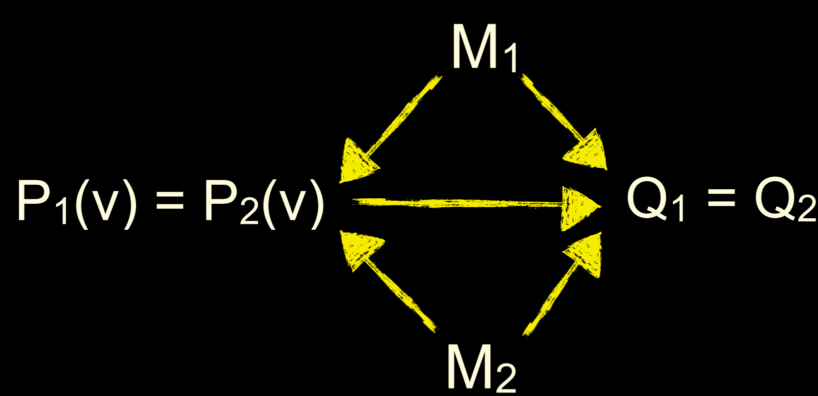
That is, if $P_{M1}(y \mid do(x)) = P_{M2}(y \mid do(x))$ for every pair of models M_1 and M_2 with $P_{M1}(v) = P_{M2}(v) > 0$ and $G(M_1) = G(M_2) = G$.

THE IDENTIFIABILITY PROBLEM

- Def. 3.2.2 (Causal Effect Identifiability)

The causal effect of X on Y is said to be identifiable from a graph G if the quantity $P(y \mid do(x))$ can be computed uniquely from a positive probability of the observed variables.

That is, if $P_{M1}(y \mid do(x)) = P_{M2}(y \mid do(x))$ for every pair of models M_1 and M_2 with $P_{M1}(v) = P_{M2}(v) > 0$ and $G(M_1) = G(M_2) = G$.

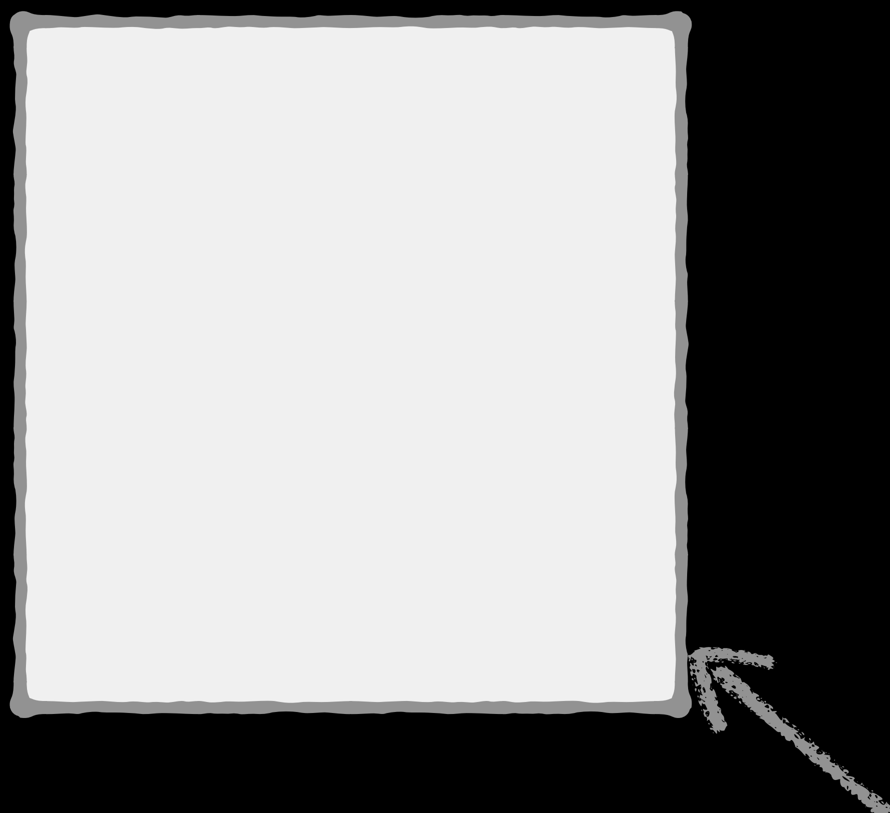


MEANING OF IDENTIFIABILITY

$P(y|do(x))$ is identifiable in G

MEANING OF IDENTIFIABILITY

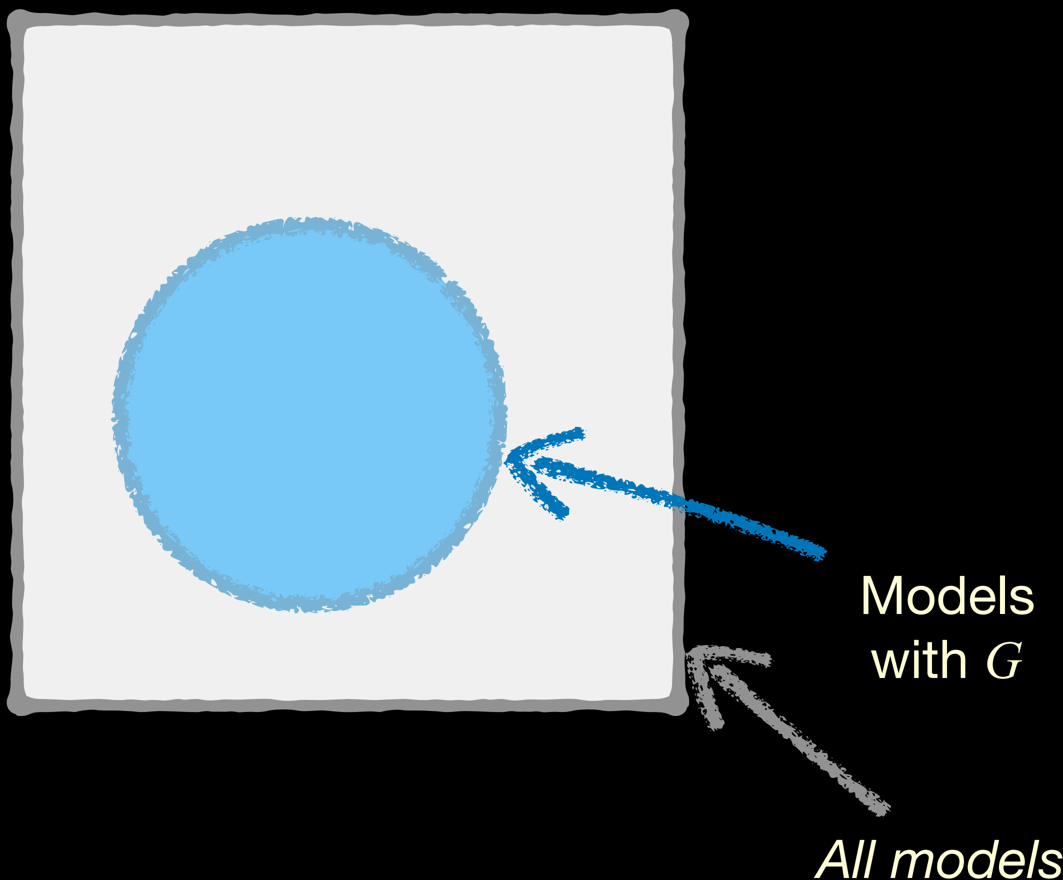
$P(y|do(x))$ is identifiable in *G*



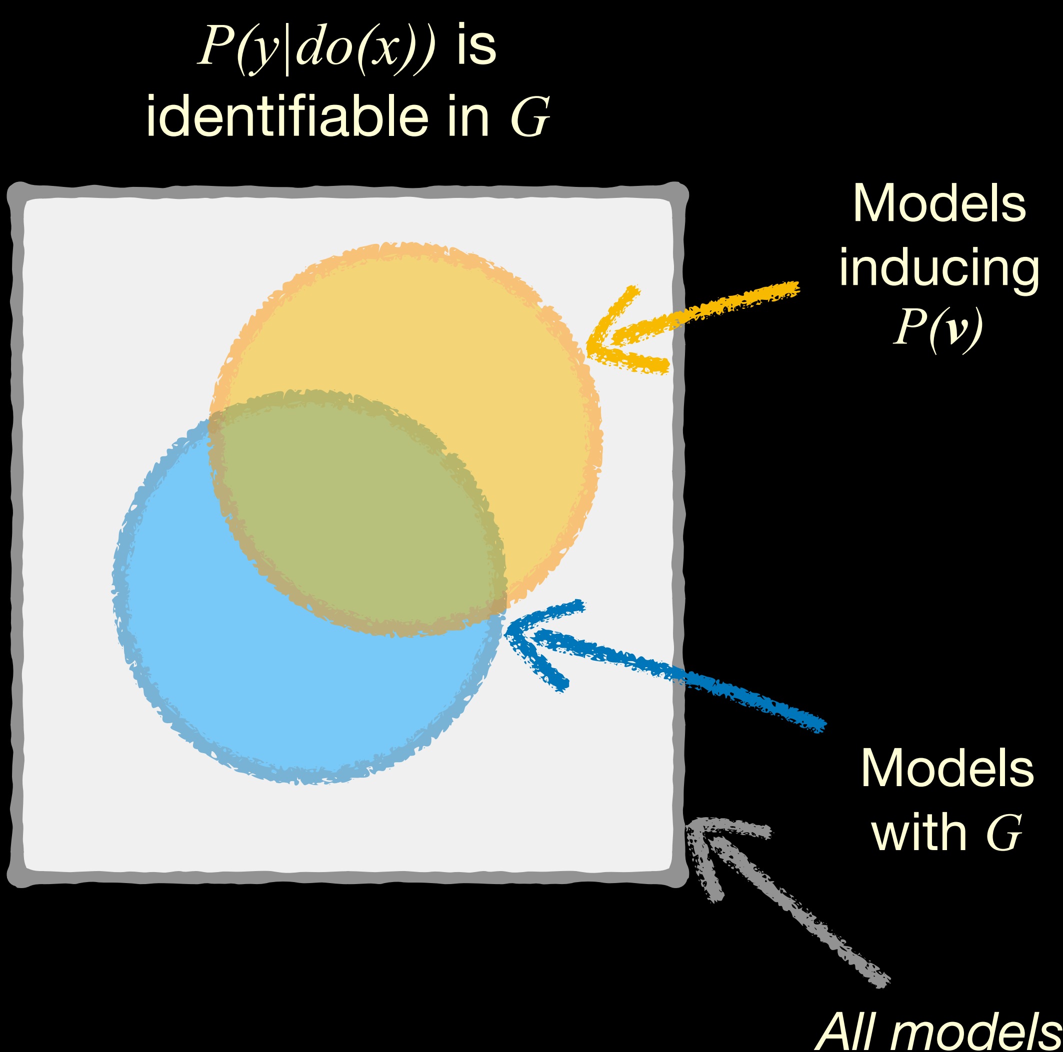
All models

MEANING OF IDENTIFIABILITY

$P(y|do(x))$ is identifiable in G

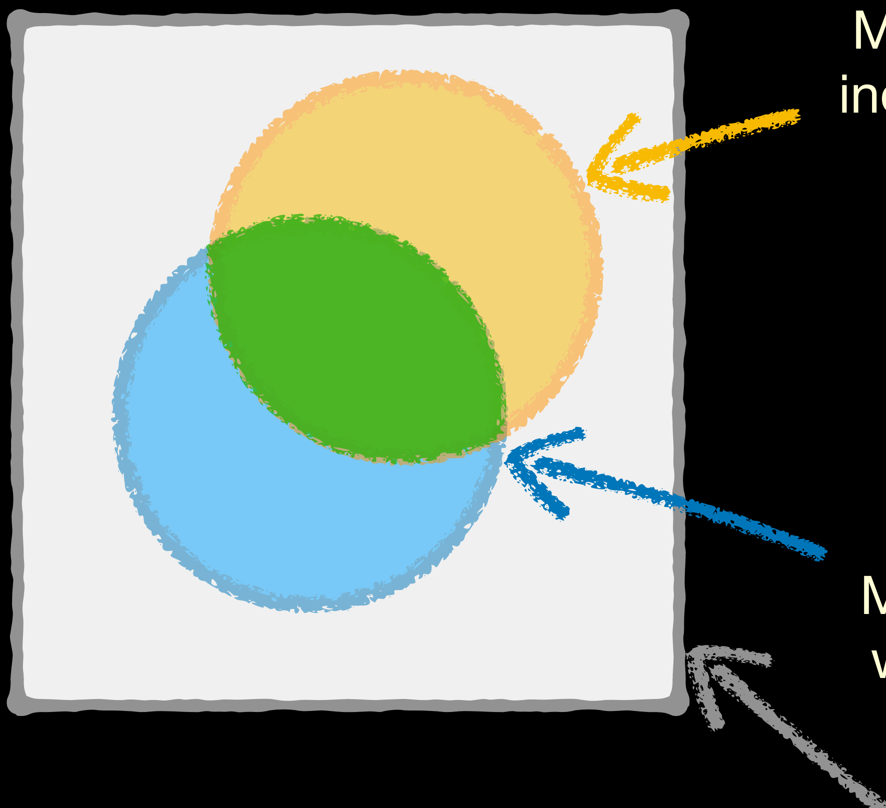


MEANING OF IDENTIFIABILITY



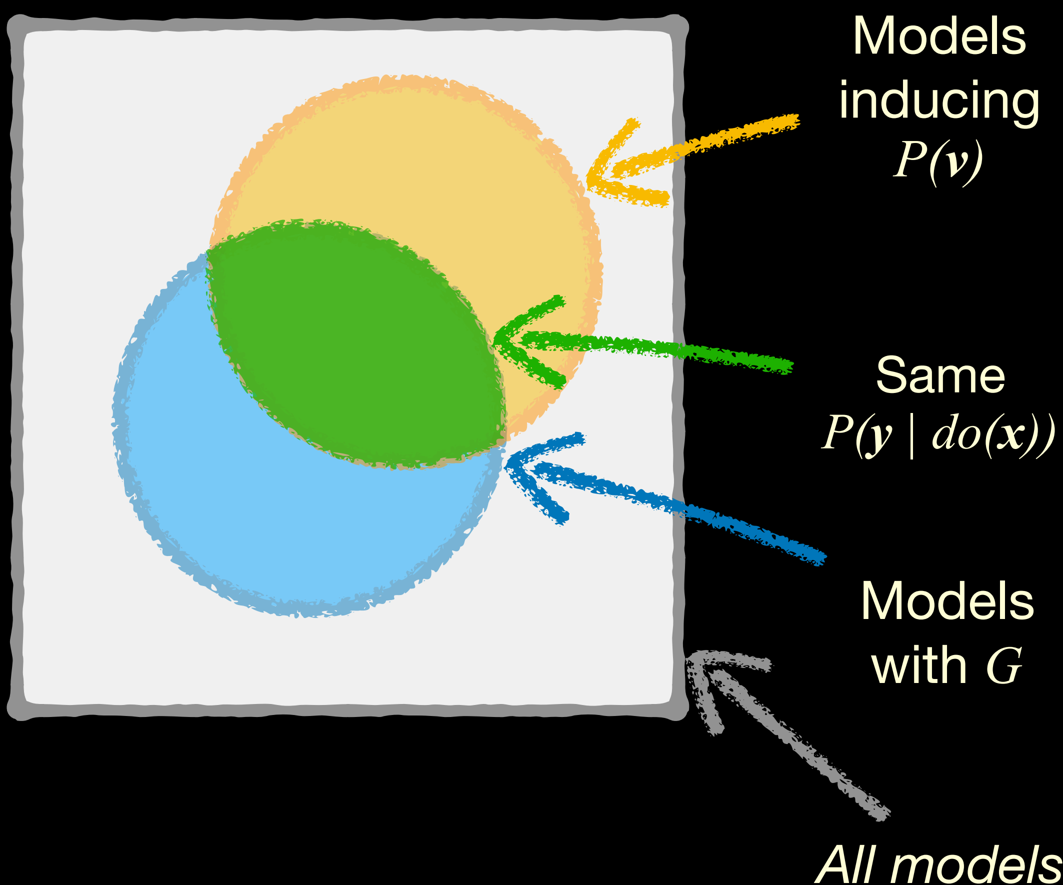
MEANING OF IDENTIFIABILITY

$P(y|do(x))$ is identifiable in G



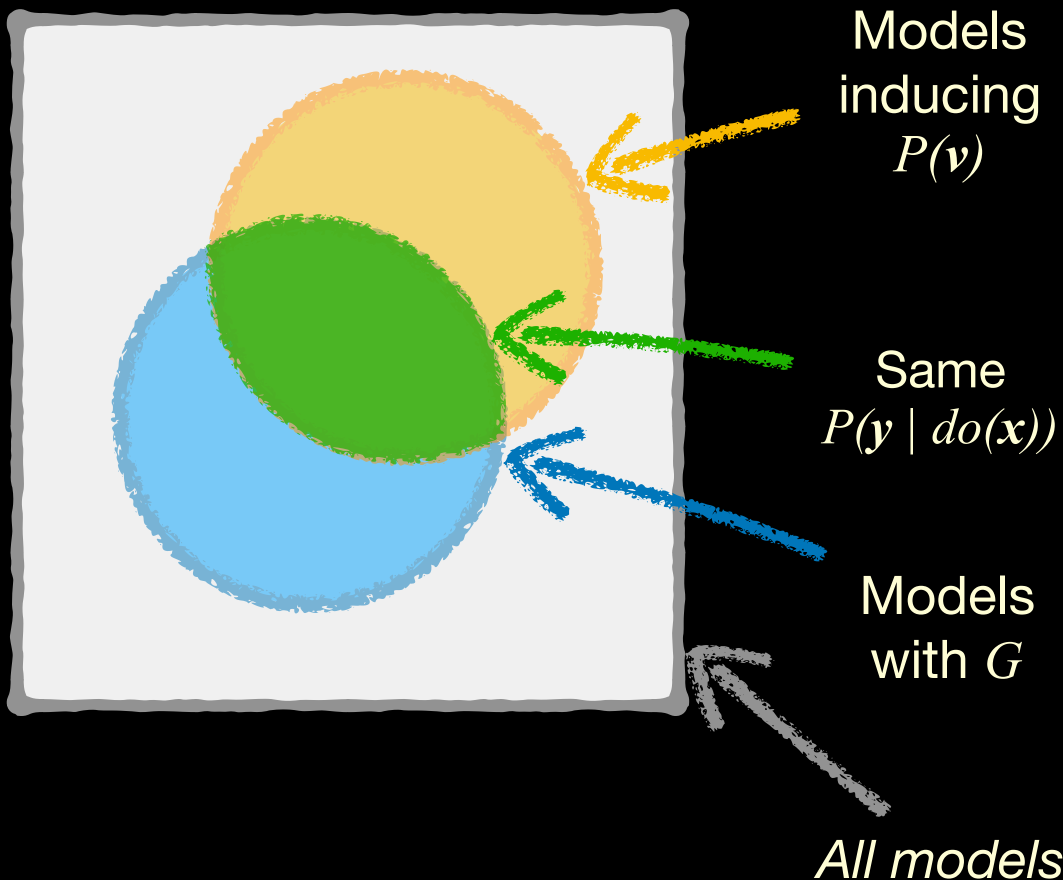
MEANING OF IDENTIFIABILITY

$P(y|do(x))$ is identifiable in G



MEANING OF IDENTIFIABILITY

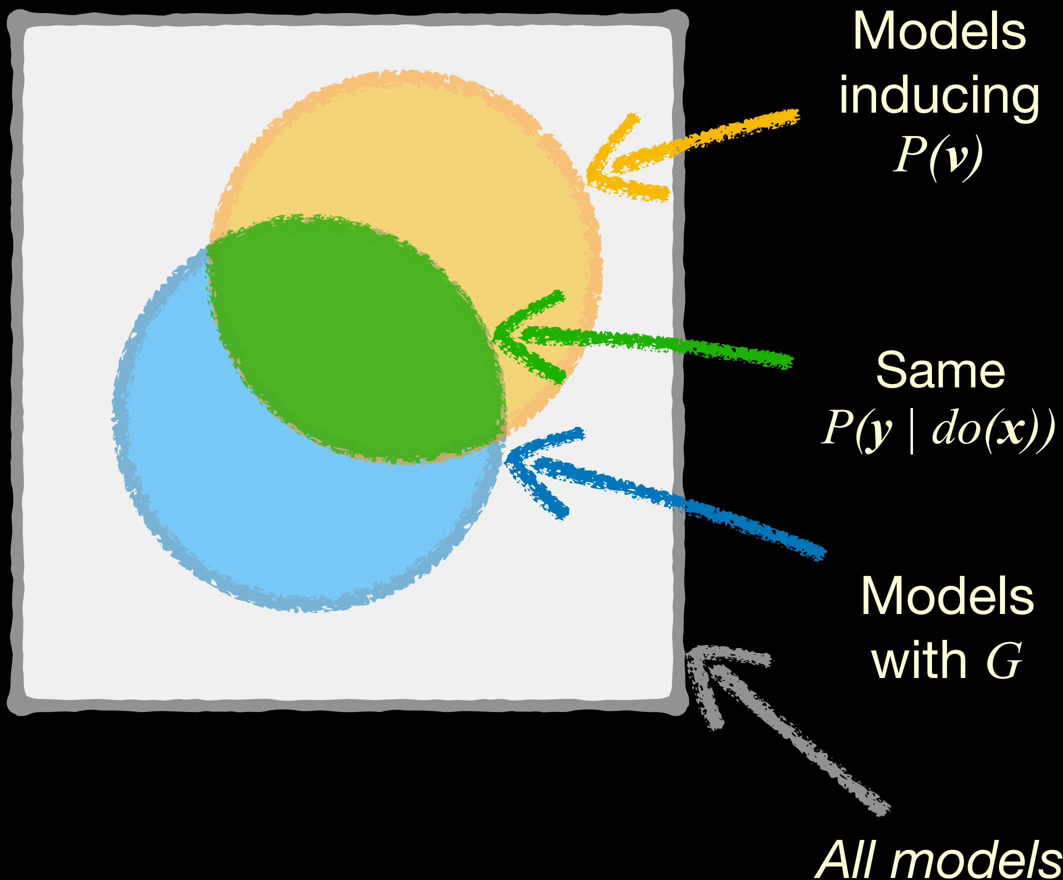
$P(y|do(x))$ is identifiable in G



Formally,
for any two causal models (encoding the unobserved nature), N_1, N_2, s.t.

MEANING OF IDENTIFIABILITY

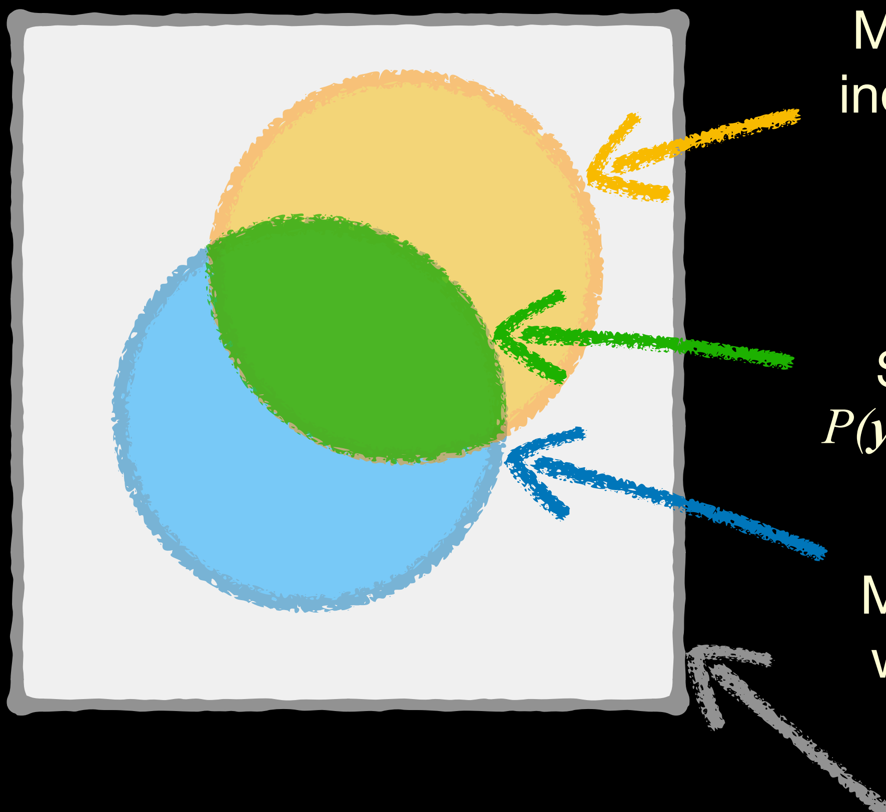
P(y|do(x)) is identifiable in *G*



Formally,
for any two causal models (encoding the unobserved nature), N_1, N_2, s.t.
$G(N_1) = G(N_2) = G$,

MEANING OF IDENTIFIABILITY

$P(y|do(x))$ is identifiable in G



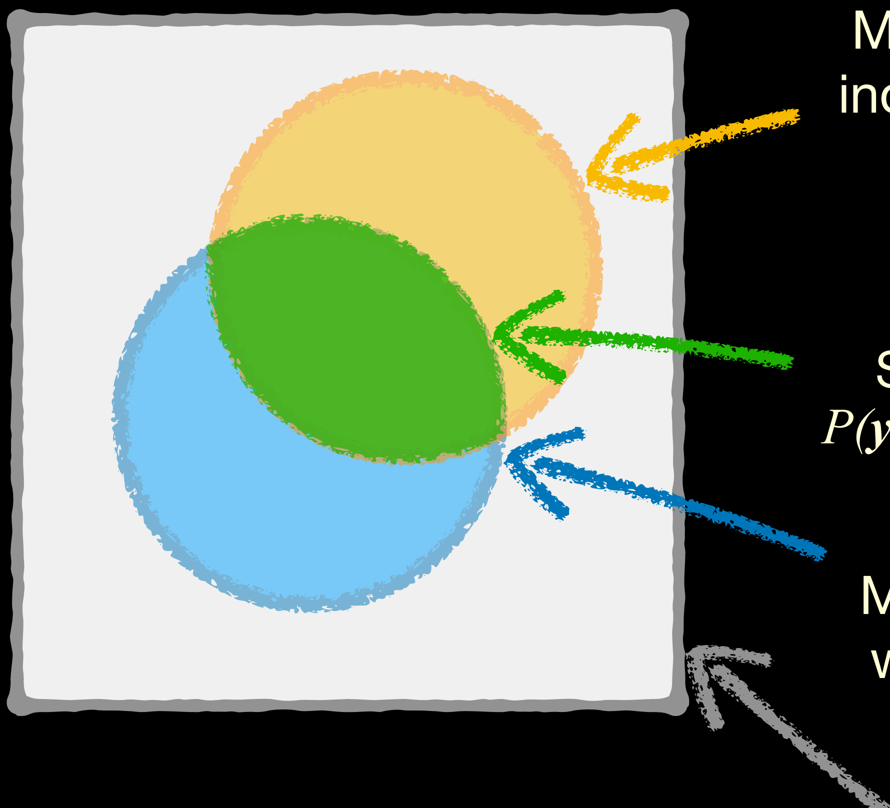
Formally, for any two causal models (encoding the unobserved nature), N_1, N_2, s.t.

$G(N_1) = G(N_2) = G$,

$P_1(v) = P_2(v)$

MEANING OF IDENTIFIABILITY

$P(y|do(x))$ is identifiable in G

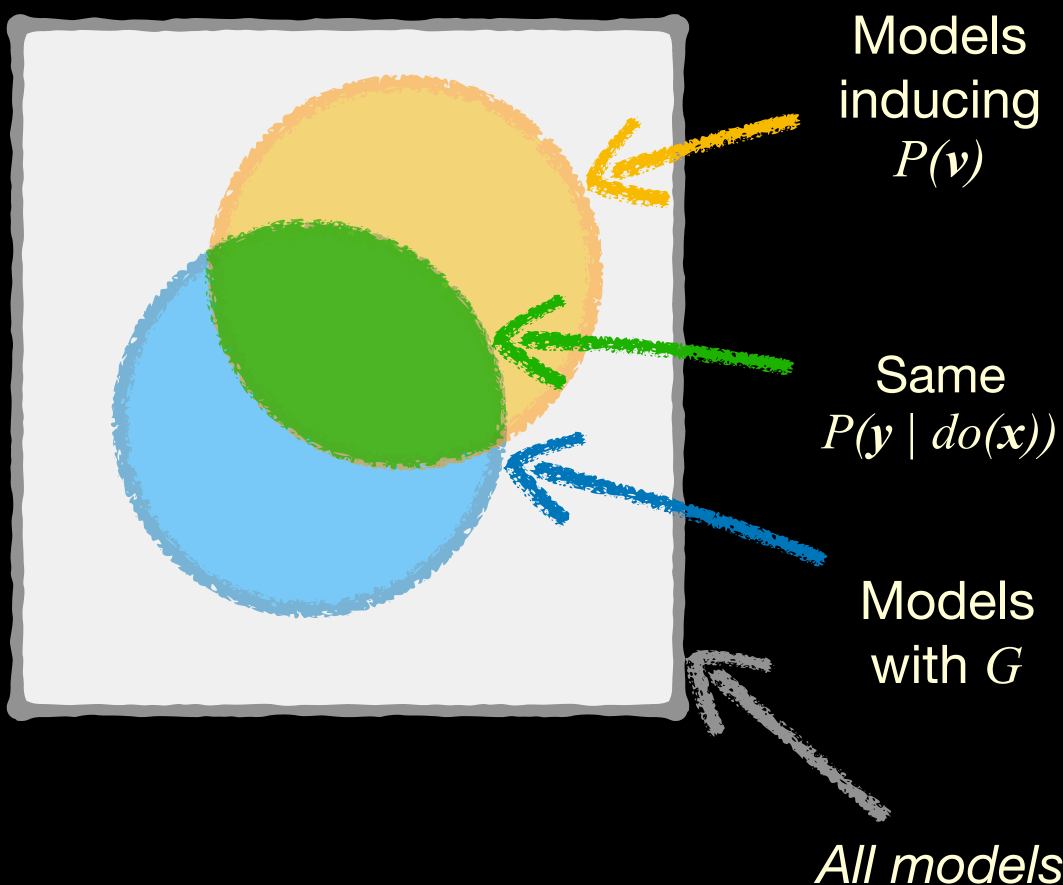


Formally,
for any two causal models (encoding the unobserved nature), N_1, N_2, s.t.
$G(N_1) = G(N_2) = G$,

$P_1(v) = P_2(v)$, then
$P_1(y \mid do(x)) = P_2(y \mid do(x))$.

MEANING OF IDENTIFIABILITY

$P(y|do(x))$ is identifiable in G



Formally, for any two causal models (encoding the unobserved nature), N_1, N_2, s.t.

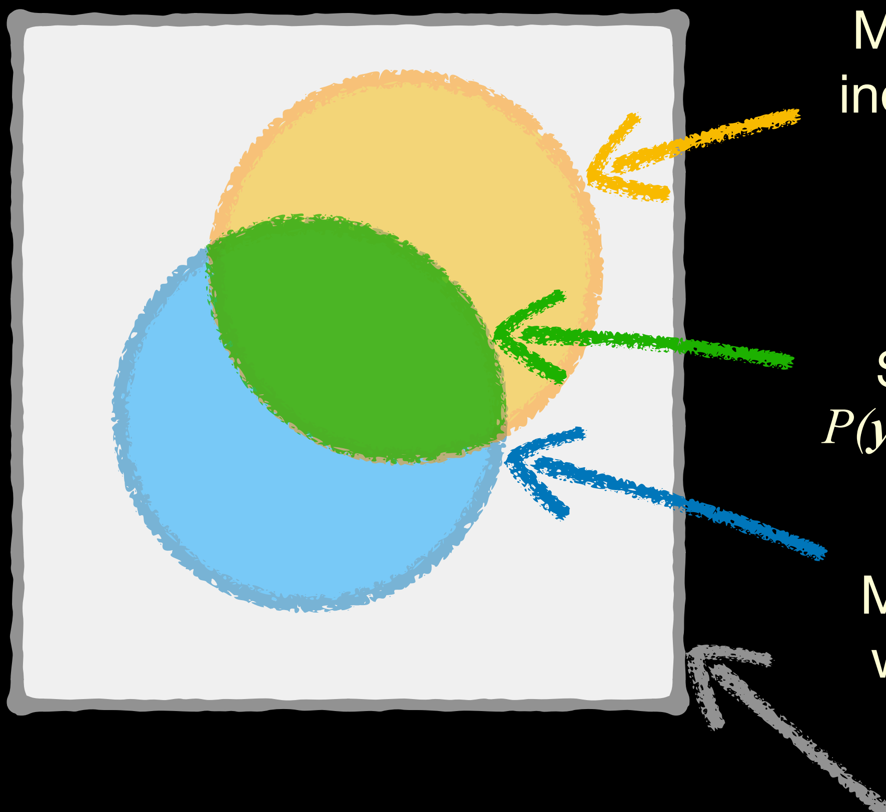
$G(N_1) = G(N_2) = G$,

$P_1(v) = P_2(v)$, then

$P_1(y \mid do(x)) = P_2(y \mid do(x))$.

MEANING OF IDENTIFIABILITY (II)

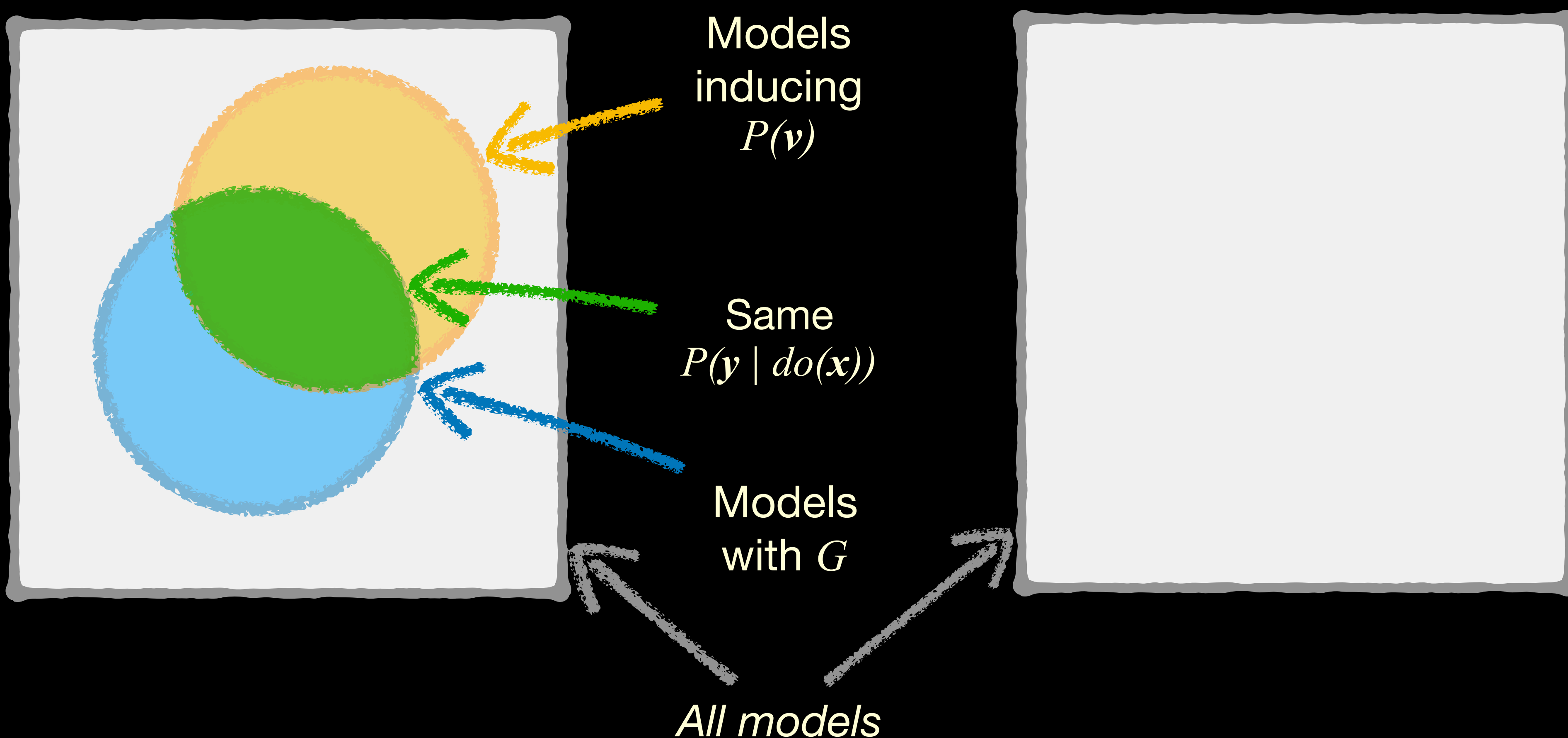
$P(y|do(x))$ is identifiability in G



MEANING OF IDENTIFIABILITY (II)

$P(y|do(x))$ is identifiability in G

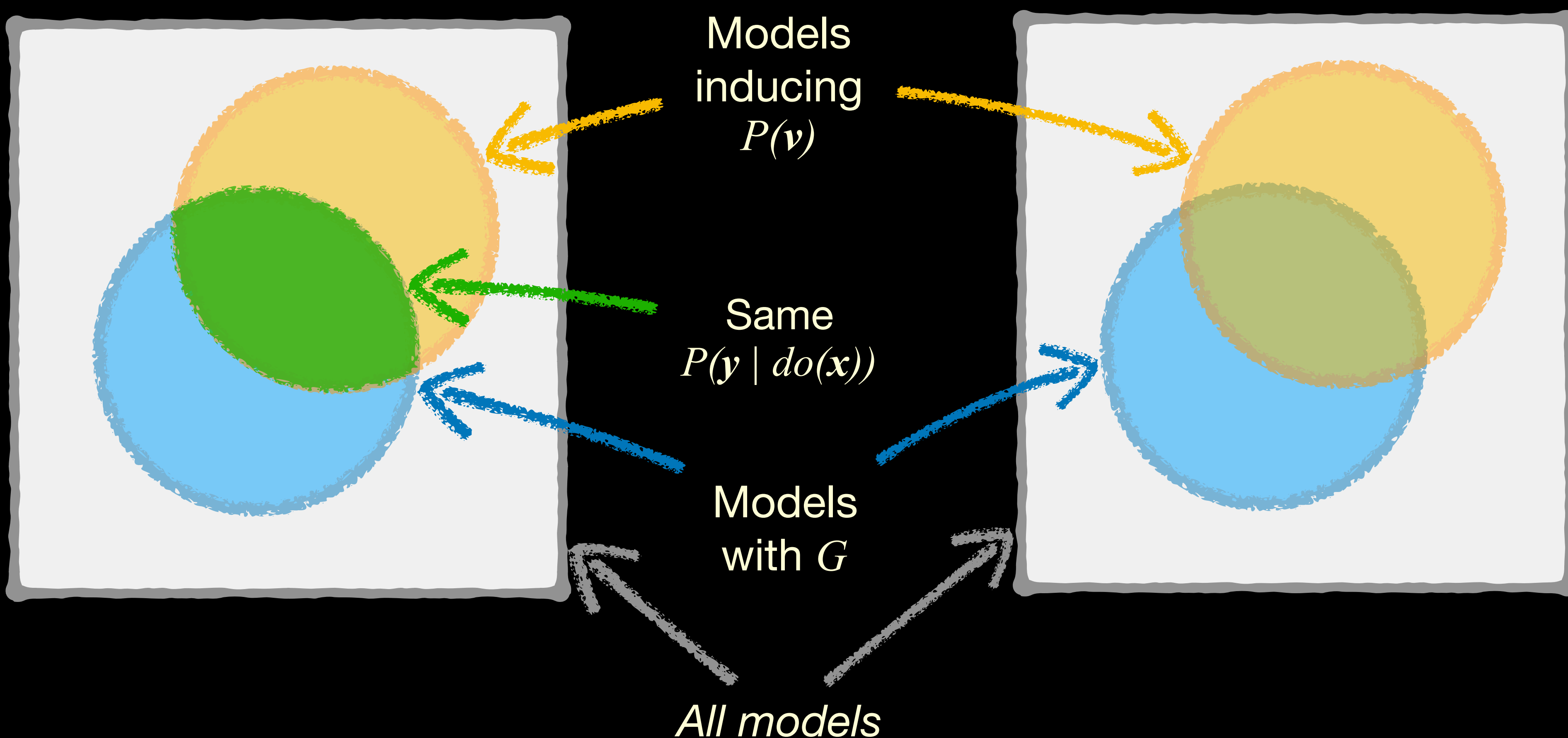
$P(y|do(x))$ is not identifiable in G



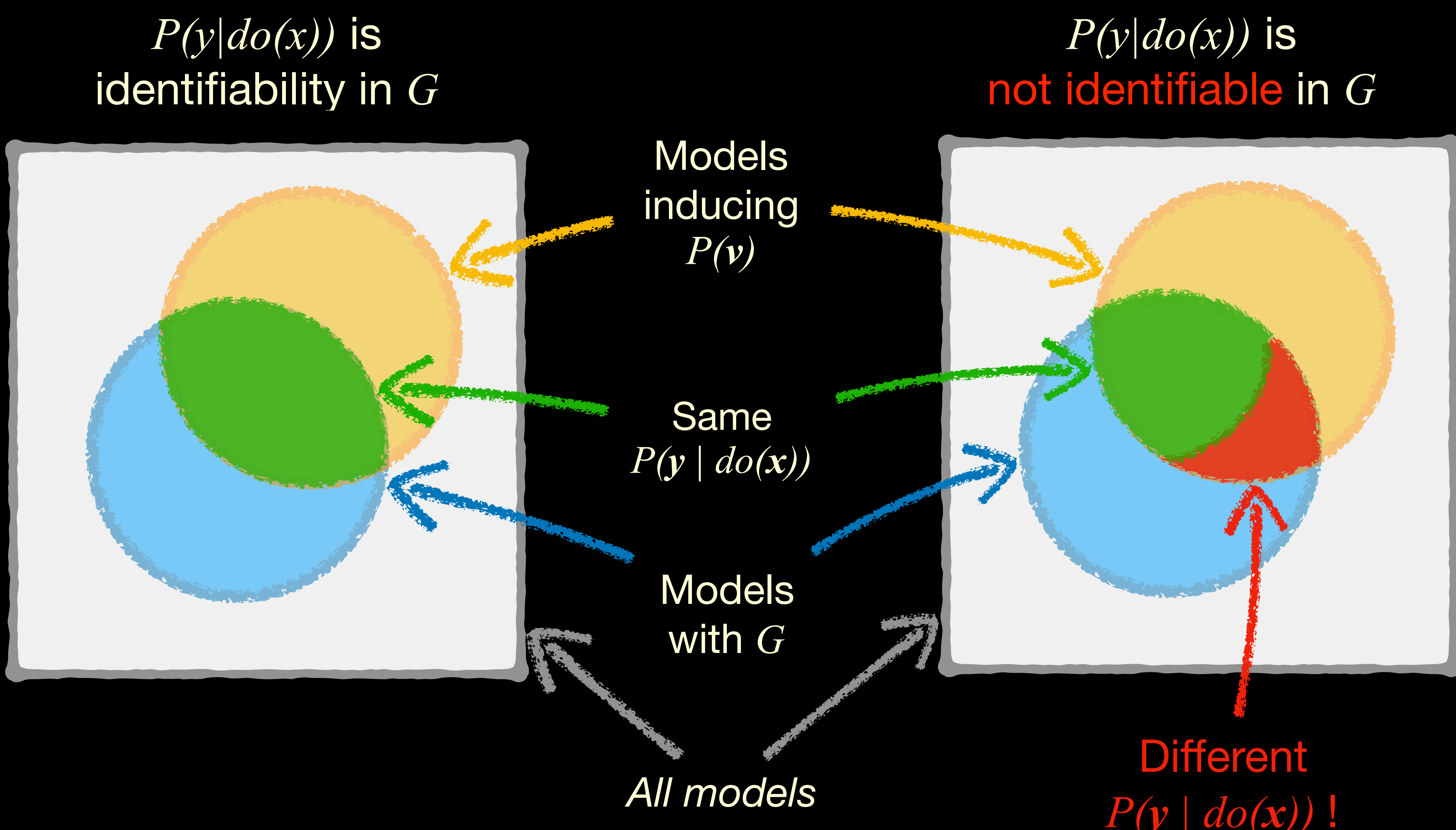
MEANING OF IDENTIFIABILITY (II)

$P(y|do(x))$ is identifiability in G

$P(y|do(x))$ is not identifiable in G



MEANING OF IDENTIFIABILITY (II)



ADJUSTMENT BY DIRECT PARENTS

- Theorem 3.2.5

Given a causal diagram G of any Markovian model in which a subset V of variables are measured, the causal effect $Q = P(y \mid do(x))$ is identifiable whenever $\{X, Y, Pa_X\} \subseteq V$, that is, whenever X, Y, and all parents of variables X are measured. The expression of Q is then obtained by adjustment for PA_X, or

ADJUSTMENT BY DIRECT PARENTS

- Theorem 3.2.5

Given a causal diagram G of any Markovian model in which a subset V of variables are measured, the causal effect $Q = P(y \mid do(x))$ is identifiable whenever $\{X, Y, Pa_x\} \subseteq V$, that is, whenever X, Y, and all parents of variables X are measured. The expression of Q is then obtained by adjustment for PA_x, or

$$Q = \sum_{PAx} P(y \mid x, PA_x = pa_x)\, P(PA_x = pa_x)$$

IDENTIFICATION IN MARKOVIAN MODELS

- Corollary 3.2.6

Given the causal diagram G of any Markovian model in which all variables are measured, the causal effect $Q = P(y \mid do(x))$ is identifiable for every subsets of variables X and Y and is obtained from the truncated factorization, i.e.,

IDENTIFICATION IN MARKOVIAN MODELS

- Corollary 3.2.6

Given the causal diagram G of any Markovian model in which all variables are measured, the causal effect Q = P(y | do(x)) is identifiable for every subsets of variables X and Y and is obtained from the truncated factorization, i.e.,

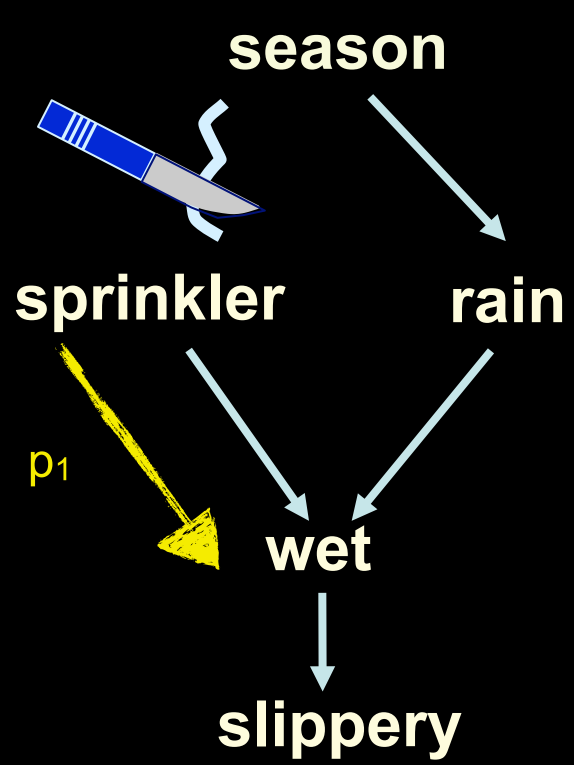
$$P(y \mid do(x)) = \sum_{V \setminus Y \setminus X} \prod_{\{i \mid X_i \notin X\}} P(x_i \mid pa_i)$$

IF SEASON IS LATENT, IS THE EFFECT STILL COMPUTABLE?

Queries:

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

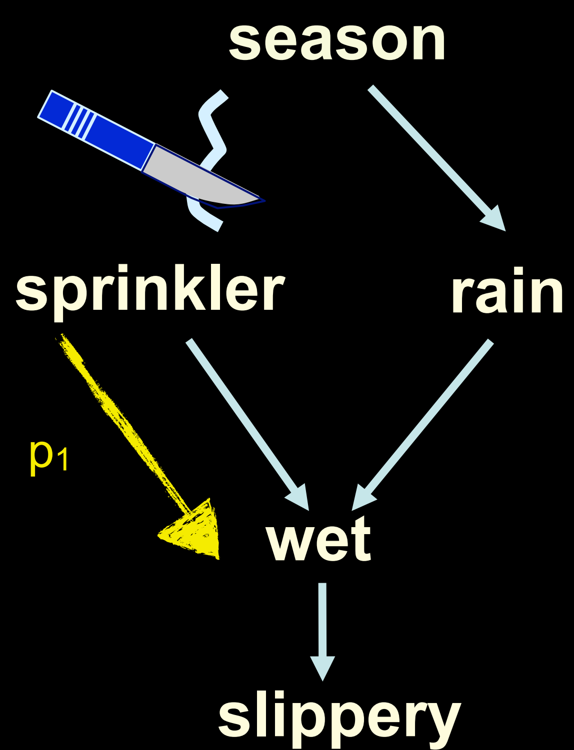
Q_2 = Pr(wet | do(Sprinkler = on))
= $P(p_1)$



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

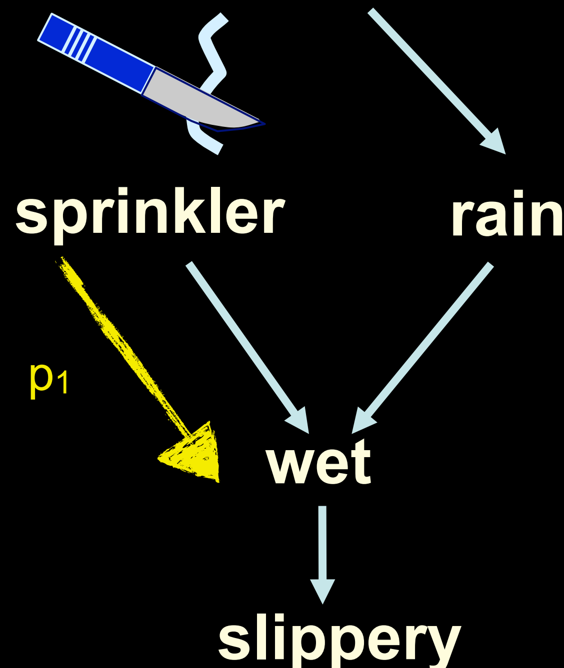
$$\sum_{Se,Ra,Sl} P(Se)\, \cancel{P(Sp \mid Se)}\, P(Ra \mid Se)\, P(We \mid Sp, Ra)\, P(Sl \mid We)$$

IF SEASON IS LATENT, IS THE EFFECT STILL COMPUTABLE?

Queries:

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

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

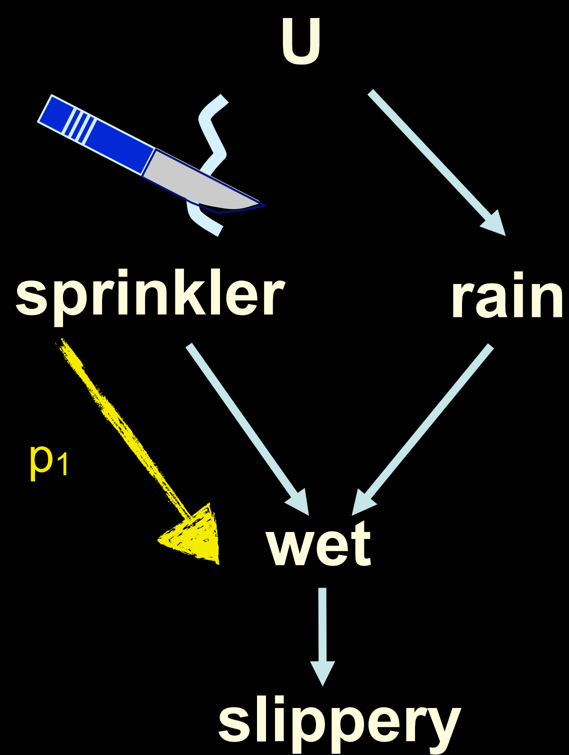


$$\sum_{Se,Ra,Sl} P(Se)\ \cancel{P(Sp \mid Se)}\ P(Ra \mid Se)\ P(We \mid Sp, Ra)\ P(Sl \mid We)$$

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

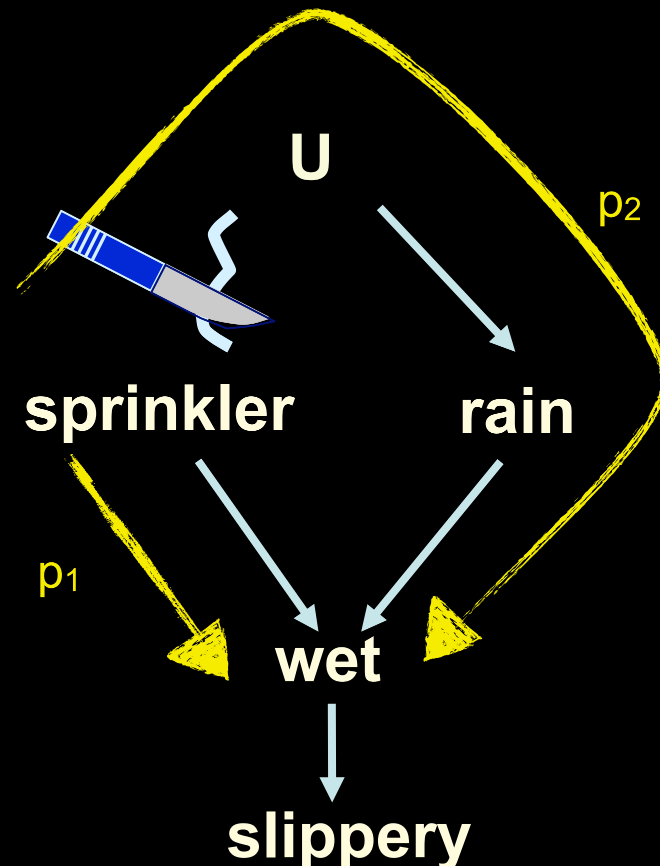
$$\sum_{Se,Ra,Sl} P(Se)\, \cancel{P(Sp \mid Se)}\, P(Ra \mid Se)\, P(We \mid Sp, Ra)\, P(Sl \mid We)$$

IF SEASON IS LATENT, IS THE EFFECT STILL COMPUTABLE?

Queries:

Q_1 = Pr(wet | Sprinkler = on)
 = P(p_1) + P(p_2)

Q_2 = Pr(wet | do(Sprinkler = on))
 = P(p_1)



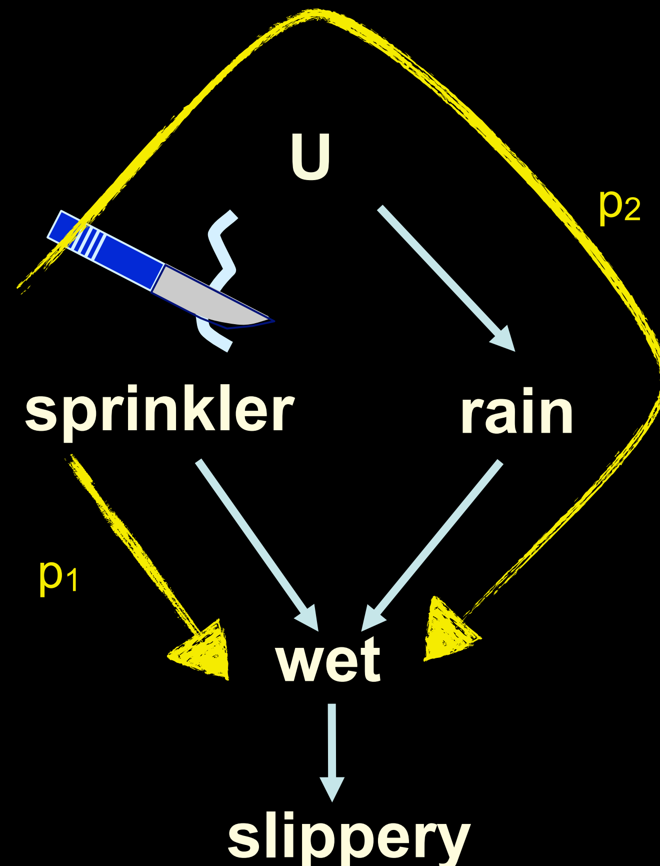
$$\sum_{Se,Ra,Sl} P(Se)\, P(Sp \mid Se)\, P(Ra \mid Se)\, P(We \mid Sp, Ra)\, P(Sl \mid We)$$

IF SEASON IS LATENT, IS THE EFFECT STILL COMPUTABLE?

Queries:

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

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



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

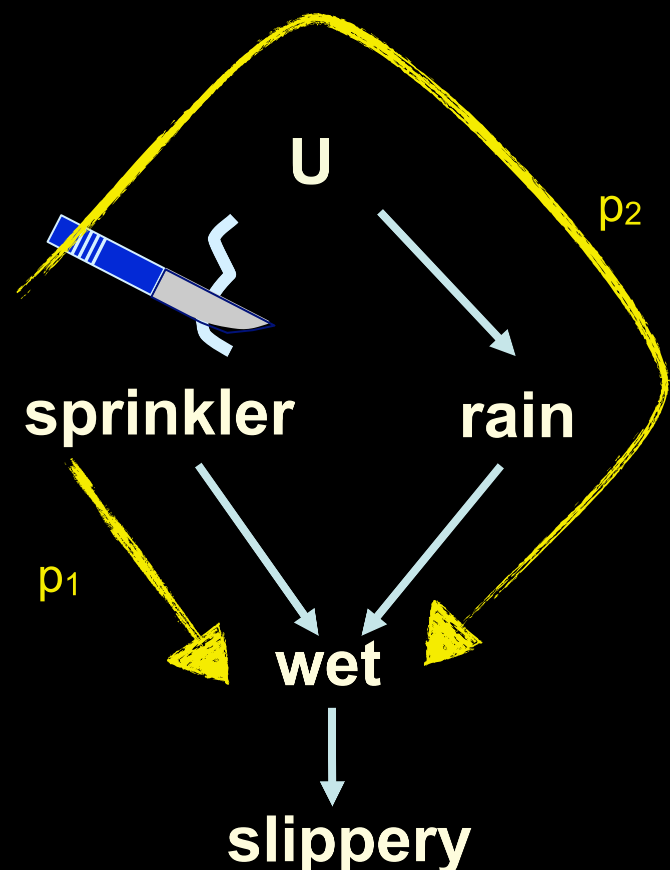
$$= \sum_{Se} P(We \mid Sp, Se)\, P(Se)$$

IF SEASON IS LATENT, IS THE EFFECT STILL COMPUTABLE?

Queries:

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

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



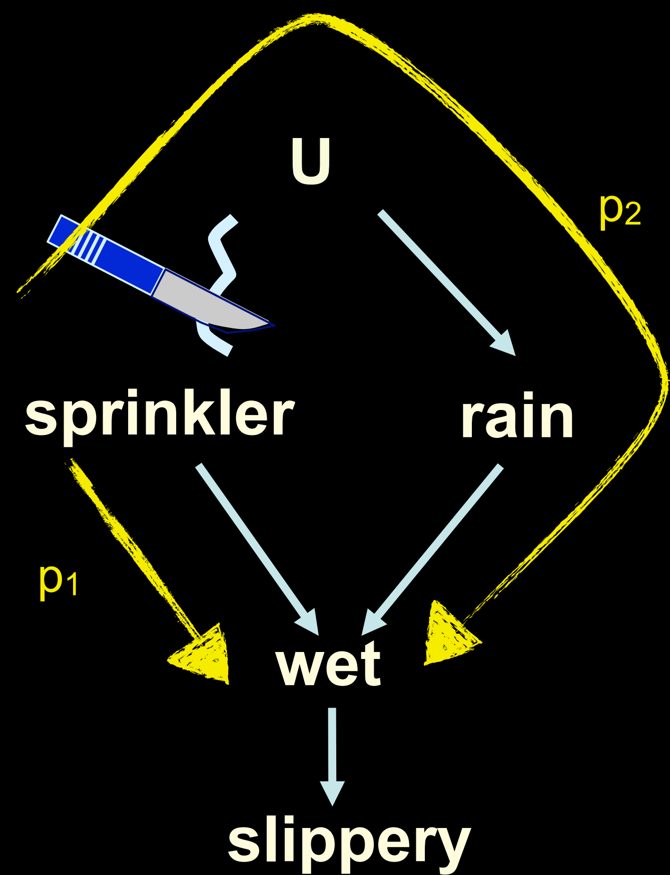
$$\sum_{Se,Ra,Sl} P(Se)\, \cancel{P(Sp \mid Se)}\, P(Ra \mid Se)\, P(We \mid Sp, Ra)\, P(Sl \mid We)$$

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

IF SEASON IS LATENT, IS THE EFFECT STILL COMPUTABLE?

Queries:

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



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

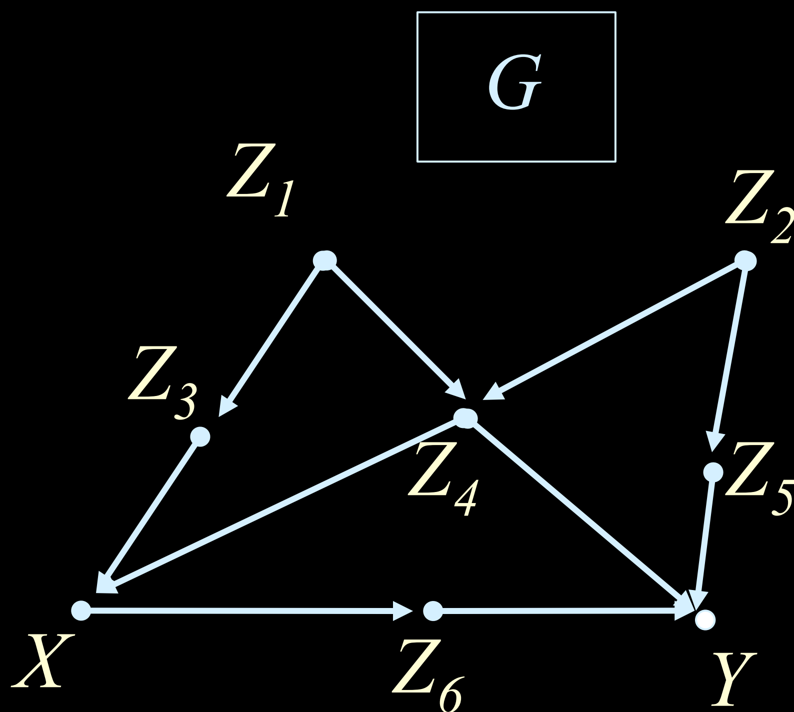
$$\sum_{Se,Ra,Sl} P(Se)\ \cancel{P(Sp \mid Se)}\ P(Ra \mid Se)\ P(We \mid Sp, Ra)\ P(Sl \mid We)$$

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

Adjustment formula

THE PROBLEM OF CONFOUNDING

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



Can Q be estimated if only a subset Z can be measured?

THE BACK-DOOR CRITERION

THE BACK-DOOR CRITERION

- Definition 3.3.1 (Back-door)

A set Z satisfies the back-door criterion relative to an ordered pair of variables (X, Y) in a DAG G if:

(i) no node in Z is a descendent of X; and
(ii) Z blocks every path between X and Y that contains an arrow into X.

THE BACK-DOOR CRITERION

- Definition 3.3.1 (Back-door)

A set Z satisfies the back-door criterion relative to an ordered pair of variables (X, Y) in a DAG G if:

(i) no node in Z is a descendent of X; and
(ii) Z blocks every path between X and Y that contains an arrow into X.

- Theorem 3.3.2 (Back-door Adjustment)

If a set Z satisfies the bd-c relative to (X, Y) then the causal effect of X on Y is identifiable and given by:

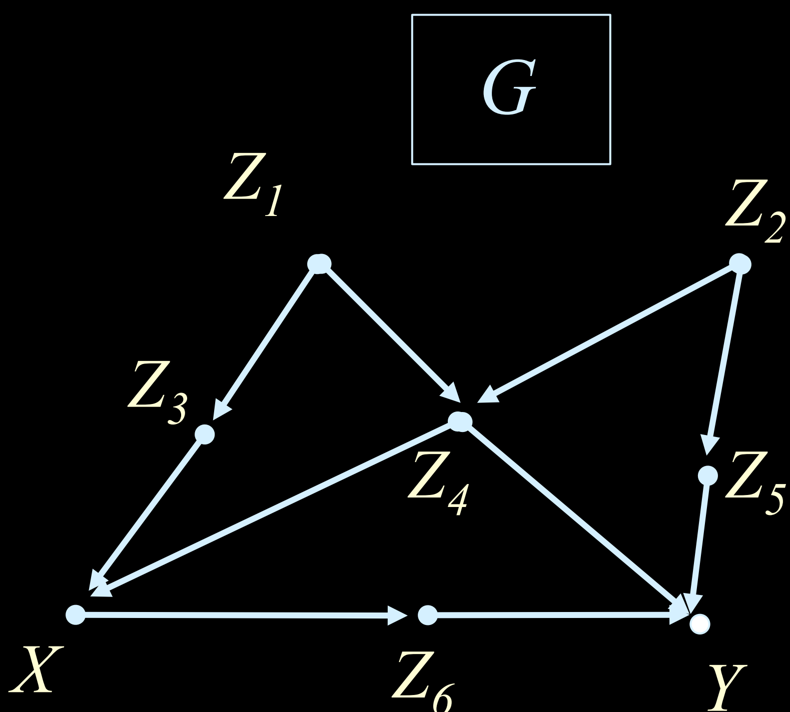
$$P(y \mid do(x)) = \sum_z P(y \mid x, z)\, P(z)$$

THE BACK-DOOR CRITERION
(graphical condition)

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

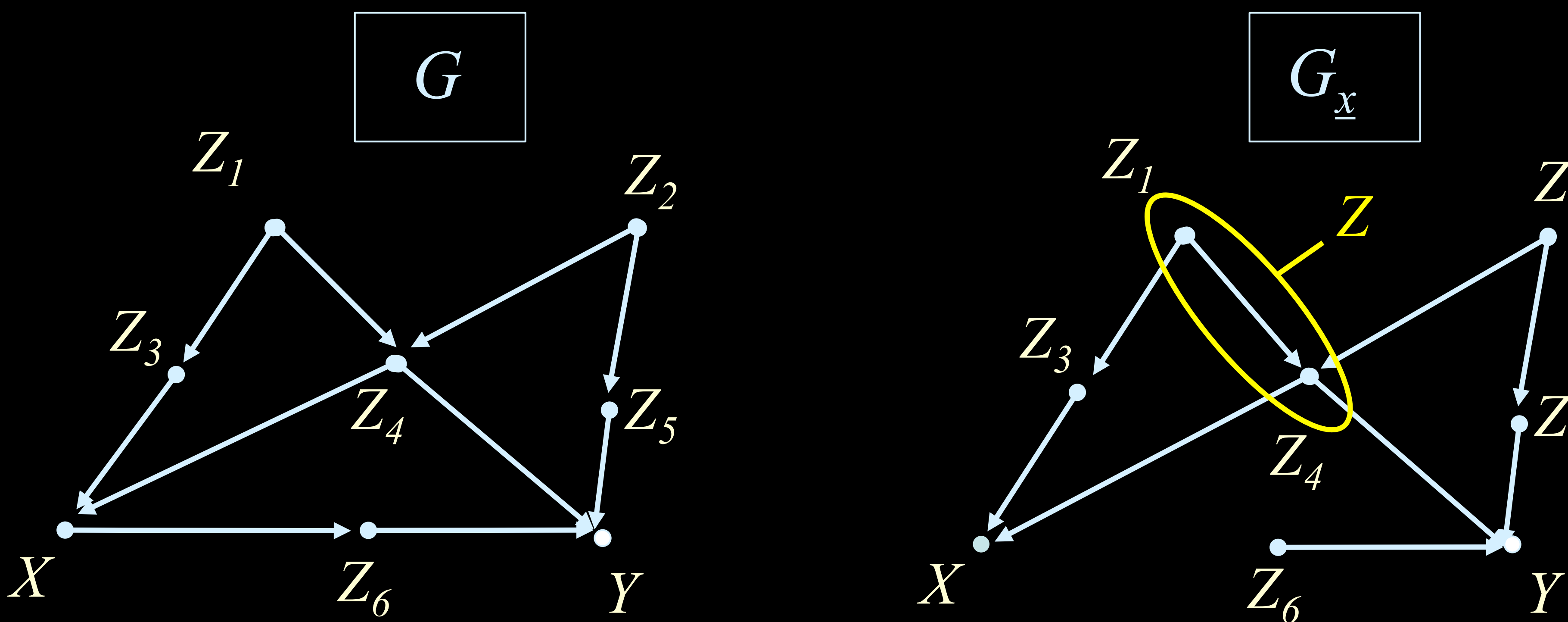
THE BACK-DOOR CRITERION
(graphical condition)

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

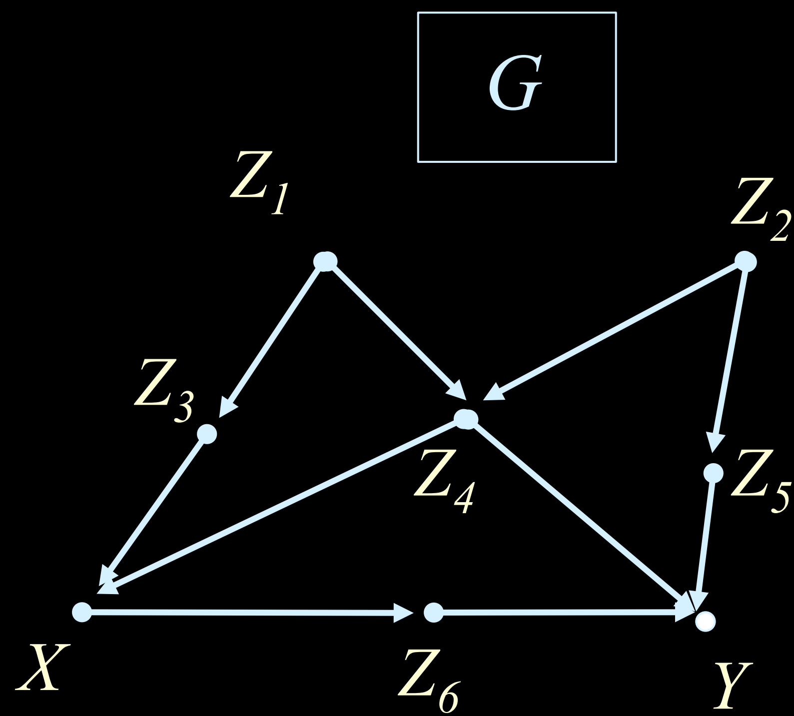


THE BACK-DOOR CRITERION
(graphical condition)

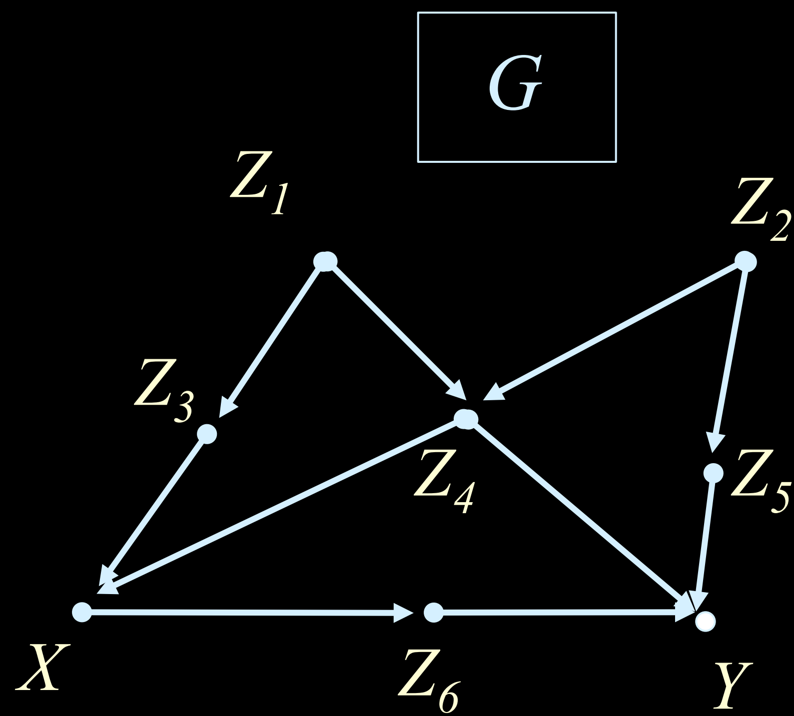
$P(y \mid do(x))$ is estimable if there is
a set Z of variables that d-separates X from Y in $G_{\underline{X}}$



GOING BEYOND ADJUSTMENT: THE FRONT-DOOR CRITERION



GOING BEYOND ADJUSTMENT: THE FRONT-DOOR CRITERION



Suppose Z_1, Z_2, Z_3, Z_4, and Z_5 are not measured.

Can we compute $Q = P(y \mid do(x))$?

THE FRONT-DOOR CRITERION

THE FRONT-DOOR CRITERION

- Definition 3.3.3 (Front-door)

A set Z satisfies is said to satisfy the front-door criterion relative to an ordered pair of variables (X, Y) if:

(i) Z intercepts all directed paths from X to Y;

(ii) There is no back-door path from X to Z; and

(iii) All back-door paths from Z to Y are blocked by X.

THE FRONT-DOOR CRITERION

- Definition 3.3.3 (Front-door)

A set Z satisfies is said to satisfy the front-door criterion relative to an ordered pair of variables (X, Y) if:

(i) Z intercepts all directed paths from X to Y;

(ii) There is no back-door path from X to Z; and

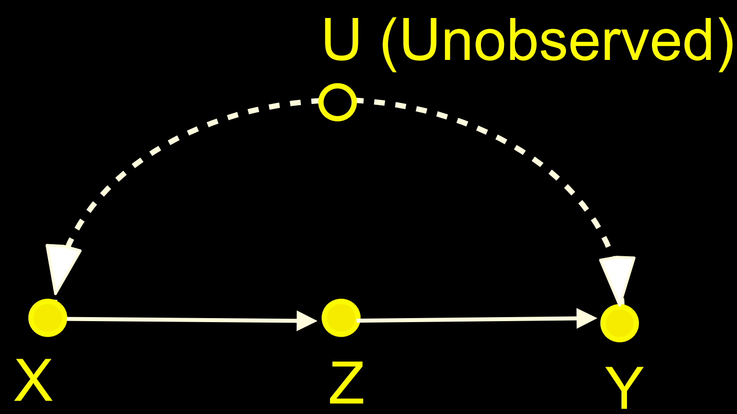
(iii) All back-door paths from Z to Y are blocked by X.

- Theorem 3.3.2 (Front-door Adjustment)

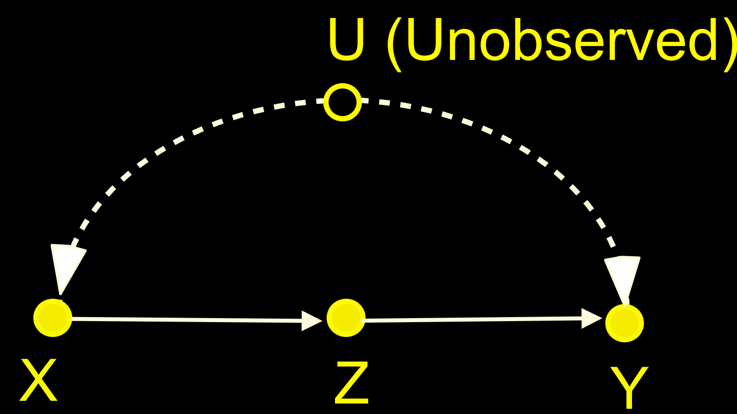
If a set Z satisfies the fd-c relative to (X, Y), then the causal effect of X on Y is identifiable and given by:

$$P(y \mid do(x)) = \sum_z P(z \mid x) \sum_{x'} P(y \mid x', z) P(x')$$

THE FRONT-DOOR CRITERION

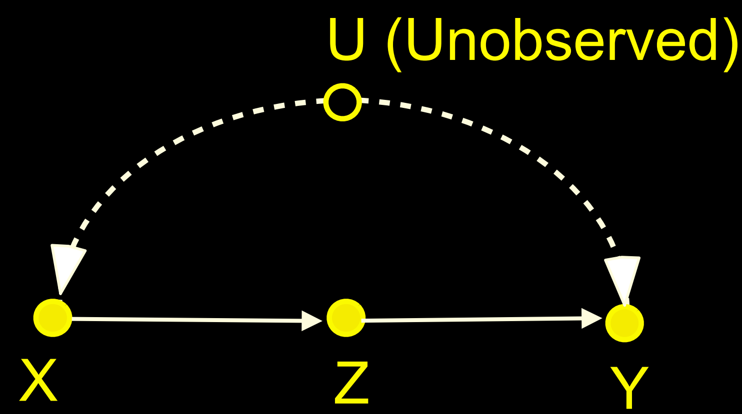


THE FRONT-DOOR CRITERION



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

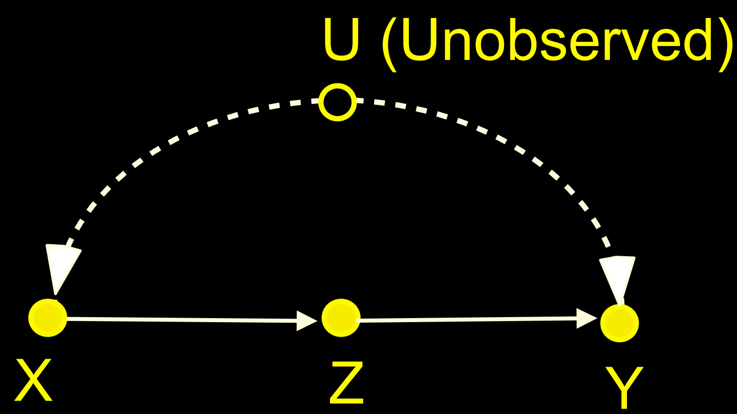
THE FRONT-DOOR CRITERION



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

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

THE FRONT-DOOR CRITERION



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

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

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

CAUSAL CALCULUS
(IDENTIFIABILITY REDUCED TO CALCULUS)

(or, do-calculus)

Theorem 3.4.1. The following transformations are valid for every interventional distribution generated by a structural causal model *M*:

CAUSAL CALCULUS (IDENTIFIABILITY REDUCED TO CALCULUS)

(or, do-calculus)

Theorem 3.4.1. 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 $(Y \perp\!\!\!\perp Z \mid X, W)_{G_{\overline{X}}}$

CAUSAL CALCULUS
(IDENTIFIABILITY REDUCED TO CALCULUS)

(or, do-calculus)

Theorem 3.4.1. 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

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

CAUSAL CALCULUS (IDENTIFIABILITY REDUCED TO CALCULUS)

(or, do-calculus)

Theorem 3.4.1. The following transformations are valid for every interventional distribution generated by a structural causal model *M*:

Rule 1: Ignoring observations

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

Rule 2: Action/observation exchange

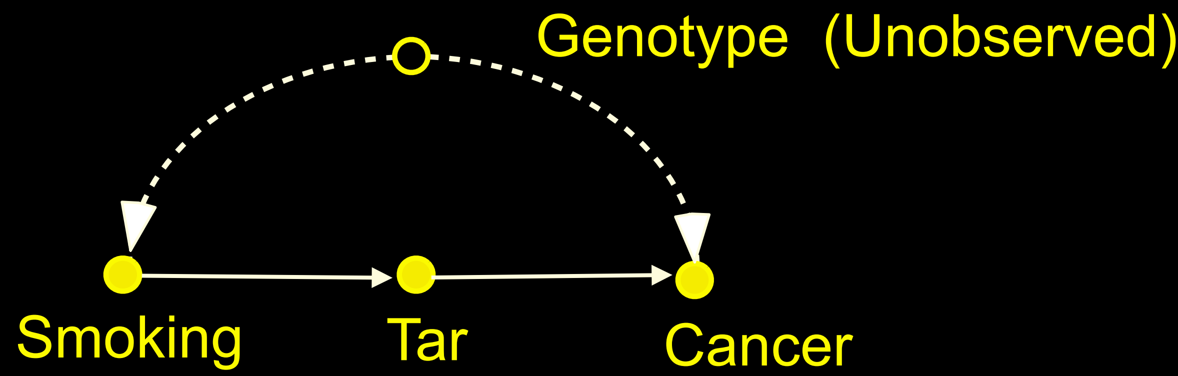
P(*y* | do(x), *do*(*z*), *w*) = *P*(*y* | do(x), *z*, *w*), if $(Y \perp\!\!\!\perp Z | X, W)_{G_{\overline{X}\underline{Z}}}$

Rule 3: Ignoring actions

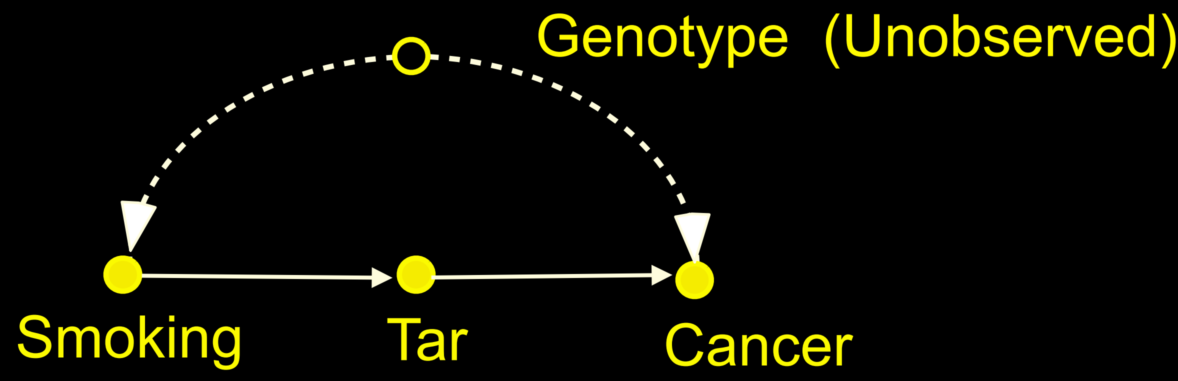
P(*y* | do(x), *do*(*z*), *w*) = *P*(*y* | *do*(*x*), *w*), if $(Y \perp\!\!\!\perp Z | X, W)_{G_{\overline{X}\,\overline{Z(W)}}}$

where Z(W) is the set of Z-nodes that are not ancestors of any W-node in G_X.

DERIVATION IN CAUSAL CALCULUS

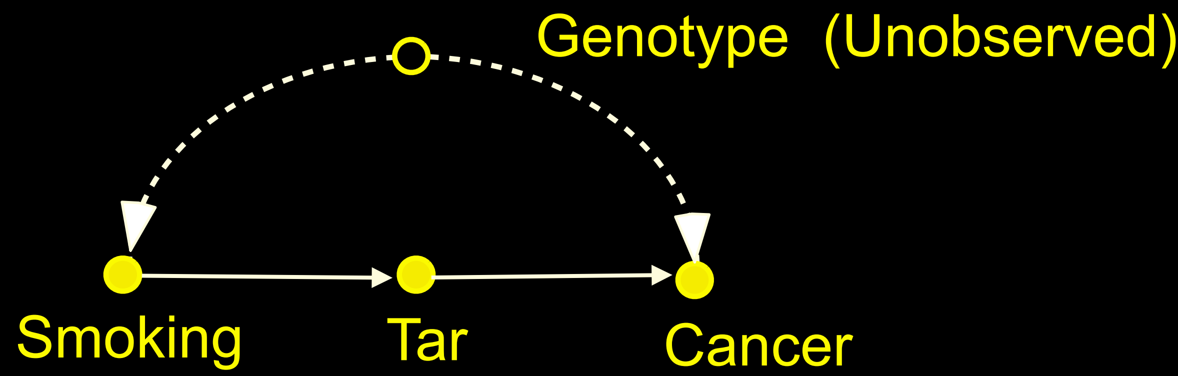


DERIVATION IN CAUSAL CALCULUS

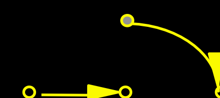


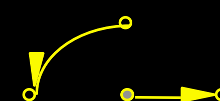
$P(c \mid do(s))$

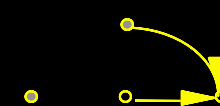
DERIVATION IN CAUSAL CALCULUS



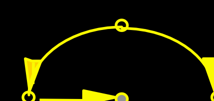
$P(c \mid do(s)) = \Sigma_t\, P(c \mid do(s), t)\, P(t \mid do(s))$ Probability Axioms

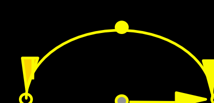
$= \Sigma_t\, P(c \mid do(s), do(t))\, P(t \mid do(s))$ Rule 2 

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

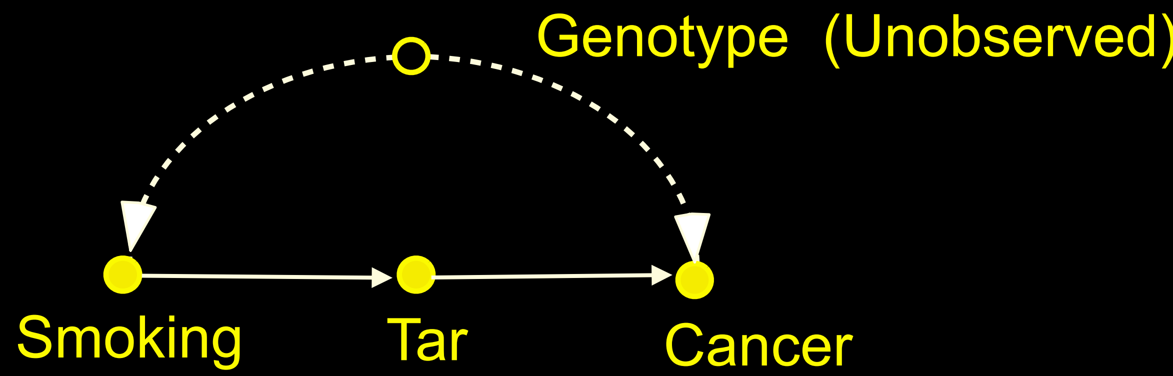
$= \Sigma_t\, P(c \mid do(t))\, P(t \mid s)$ Rule 3 

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

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

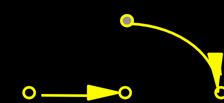
$= \Sigma_{s'} \Sigma_t\, P(c \mid t, s')\; P(s')\, P(t \mid s)$ Rule 3 

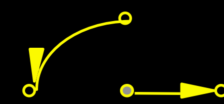
DERIVATION IN CAUSAL CALCULUS

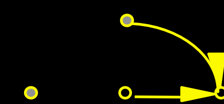


query

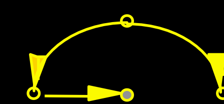
$P(c \mid do(s)) = \Sigma_t P(c \mid do(s), t)\, P(t \mid do(s))$ Probability Axioms

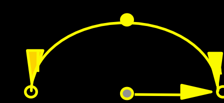
$= \Sigma_t P(c \mid do(s), do(t))\, P(t \mid do(s))$ Rule 2 

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

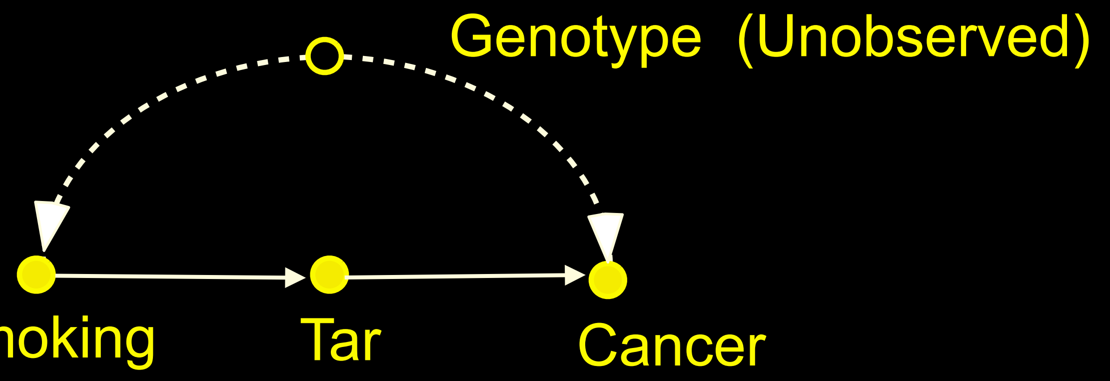
$= \Sigma_t P(c \mid do(t))\, P(t \mid s)$ Rule 3 

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

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

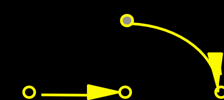
$= \Sigma_{s'} \Sigma_t P(c \mid t, s')\, P(s')\, P(t \mid s)$ Rule 3 

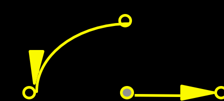
DERIVATION IN CAUSAL CALCULUS

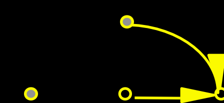


query

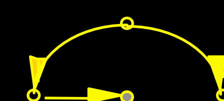
$$P(c \mid do(s)) = \Sigma_t P(c \mid do(s), t)\, P(t \mid do(s))$$ Probability Axioms

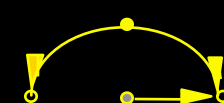
$$= \Sigma_t P(c \mid do(s), do(t))\, P(t \mid do(s))$$ Rule 2 

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

estimand (f(P(z, x, y))

CBNs WITH LATENT VARIABLES

Definition (Semi-Markovian Model):
The distribution generated by an intervention *do*(*X*=*x*) in a Semi-Markovian model *M* is given by the (extended) truncated factorization product, namely,

$$P(v_1, v_2, \ldots, v_n \mid do(x)) = \sum_{u} \prod_{i} P(v_i \mid pa_i, u_i) \prod_{i} P(u_i)$$

Example:

* Assume that in our original graph Z was unobserved.

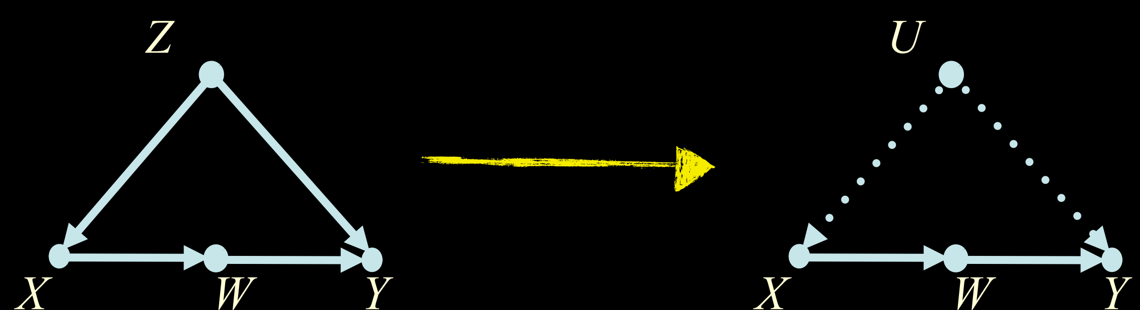
CBNs WITH LATENT VARIABLES

Definition (Semi-Markovian Model):
The distribution generated by an intervention *do*(*X*=*x*) in a Semi-Markovian model *M* is given by the (extended) truncated factorization product, namely,

$$P(v_1, v_2, \ldots, v_n \mid do(x)) = \sum_u \prod_i P(v_i \mid pa_i, u_i) \prod_i P(u_i)$$

Example:

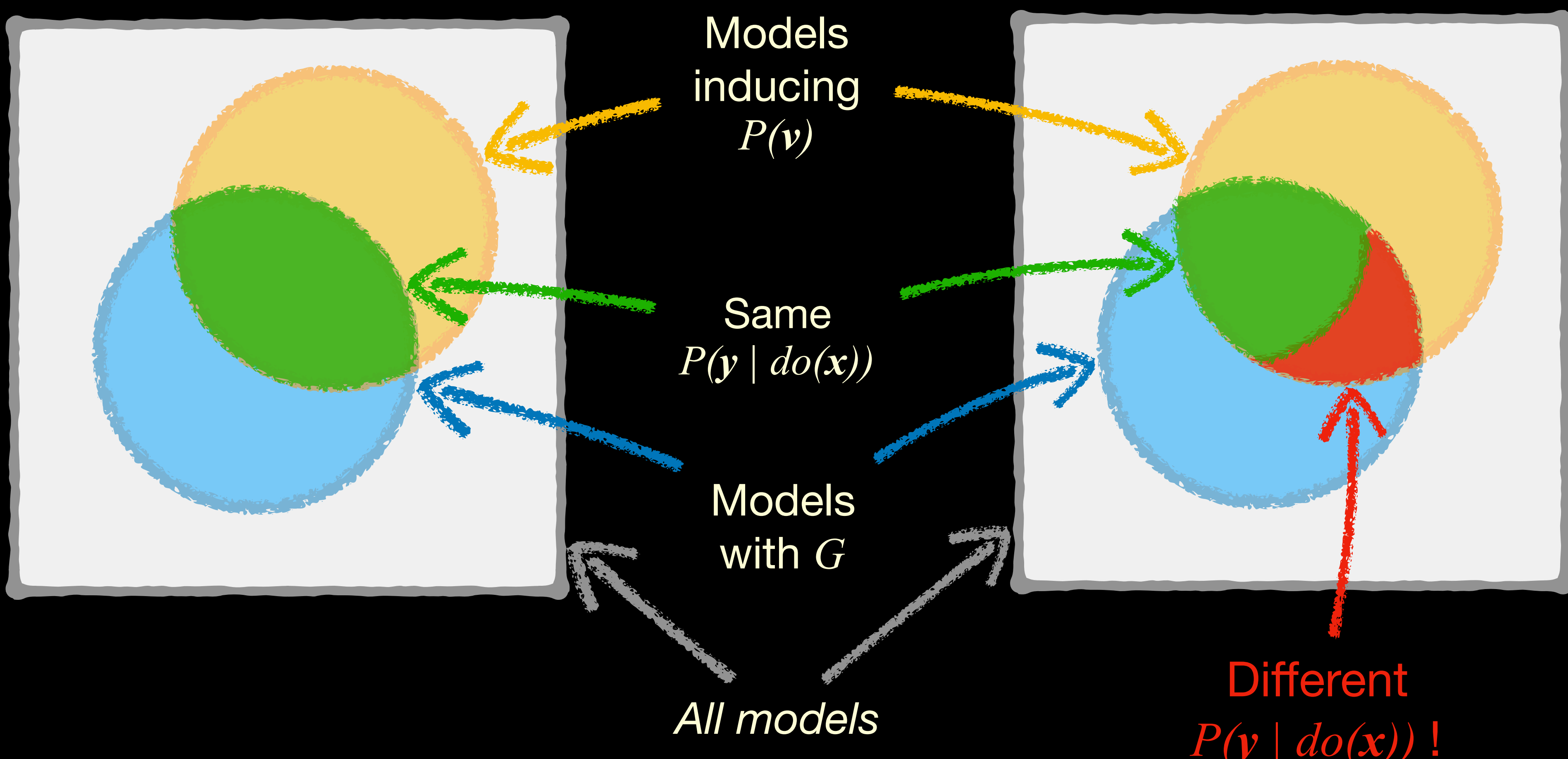
* Assume that in our original graph Z was unobserved.



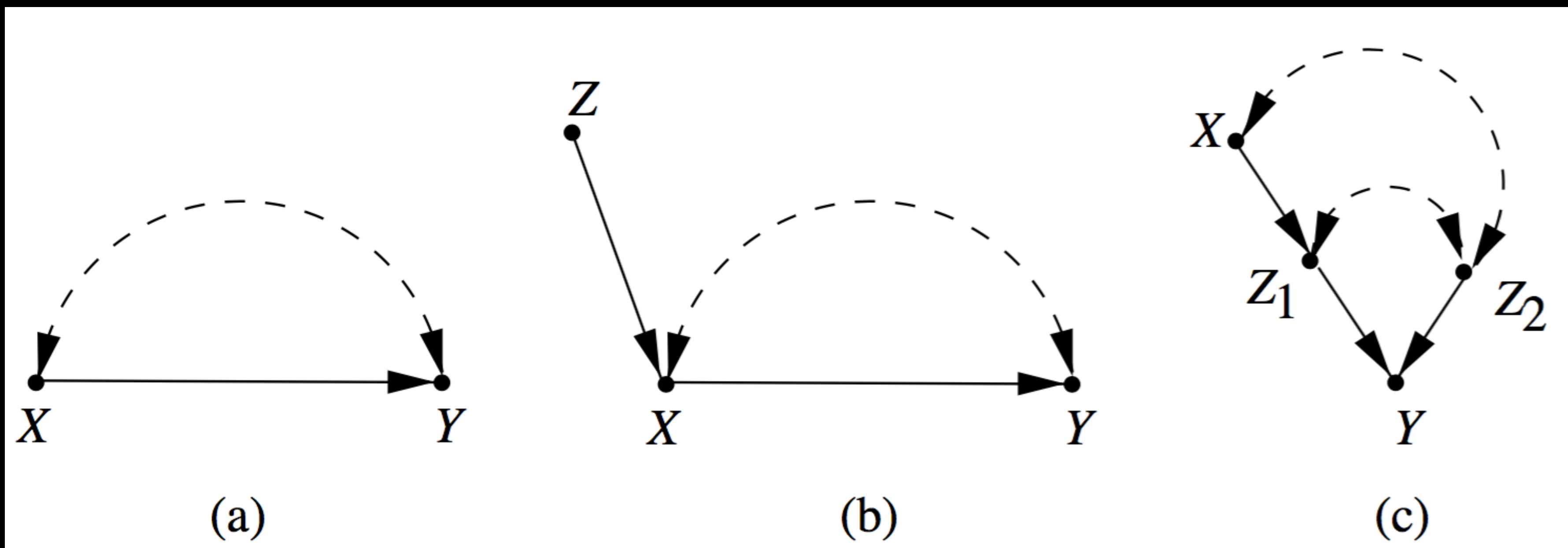
NON-IDENTIFIABLE MODELS

$P(y|do(x))$ is identifiability in G

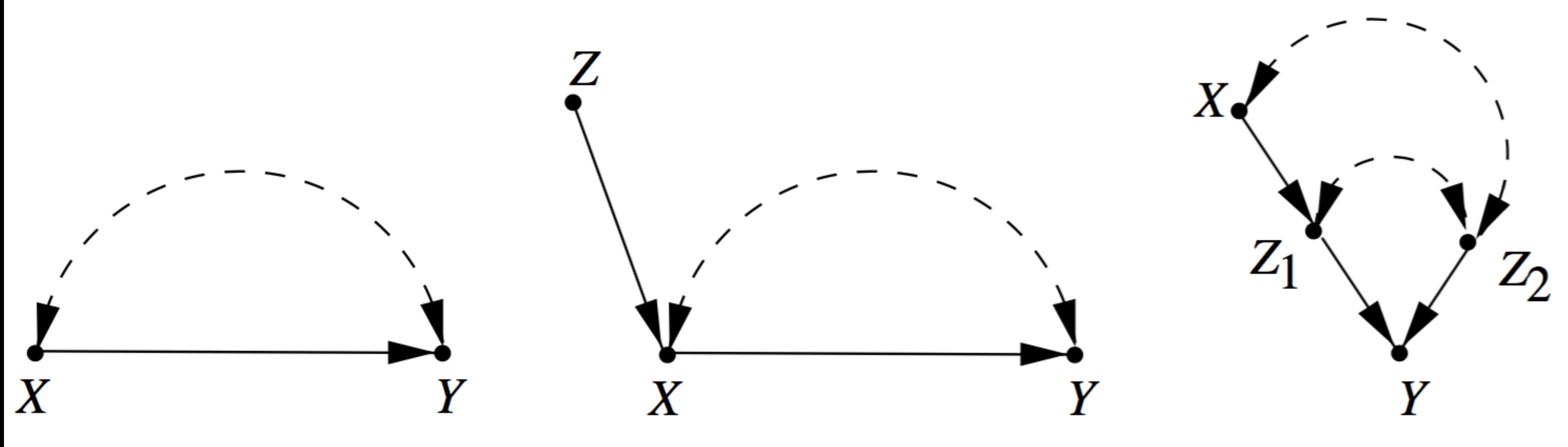
$P(y|do(x))$ is not identifiable in G



NON-IDENTIFIABLE MODELS (EXAMPLES)



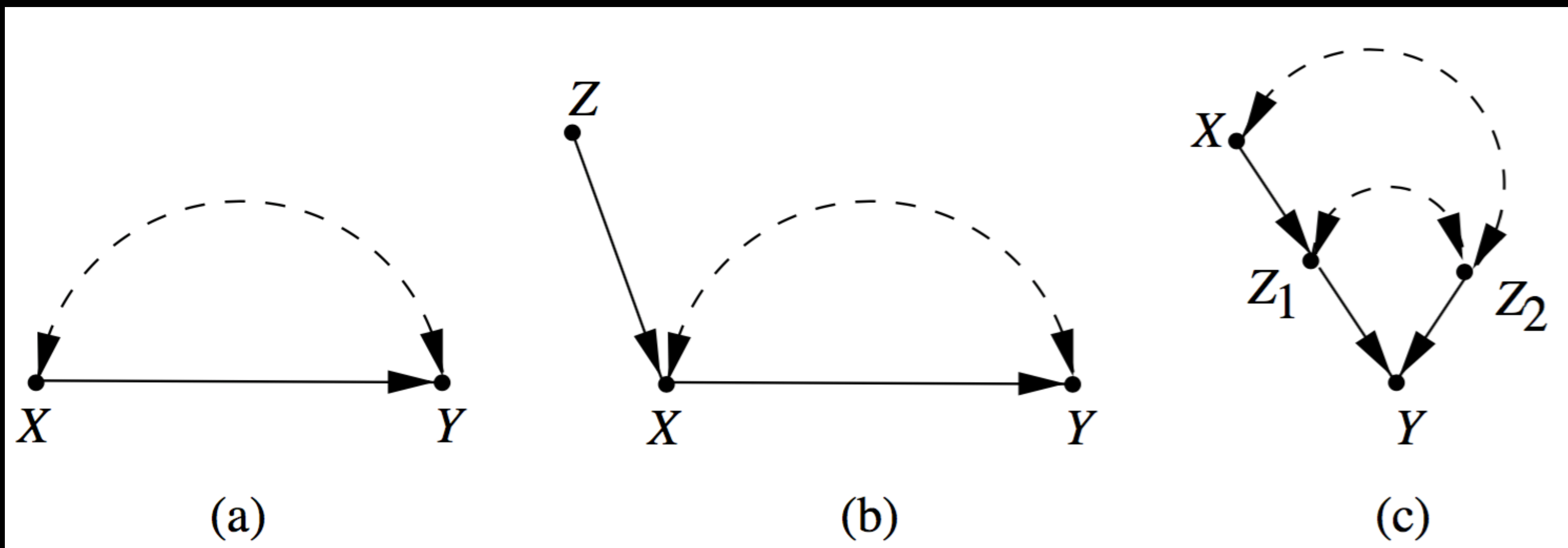
NON-IDENTIFIABLE MODELS (EXAMPLES)



(a) (b) (c)

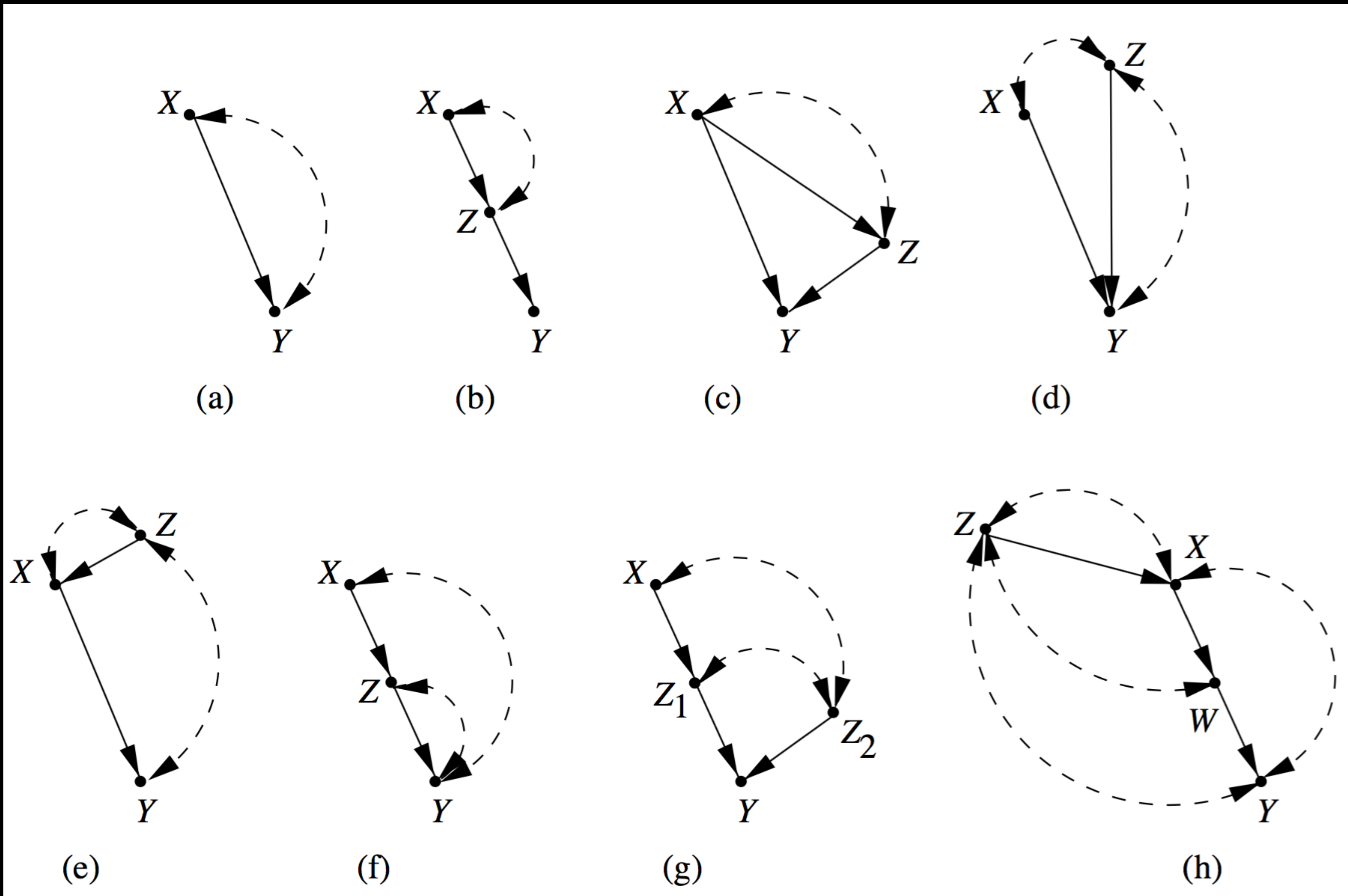
- Bow-arc observation

NON-IDENTIFIABLE MODELS (EXAMPLES)

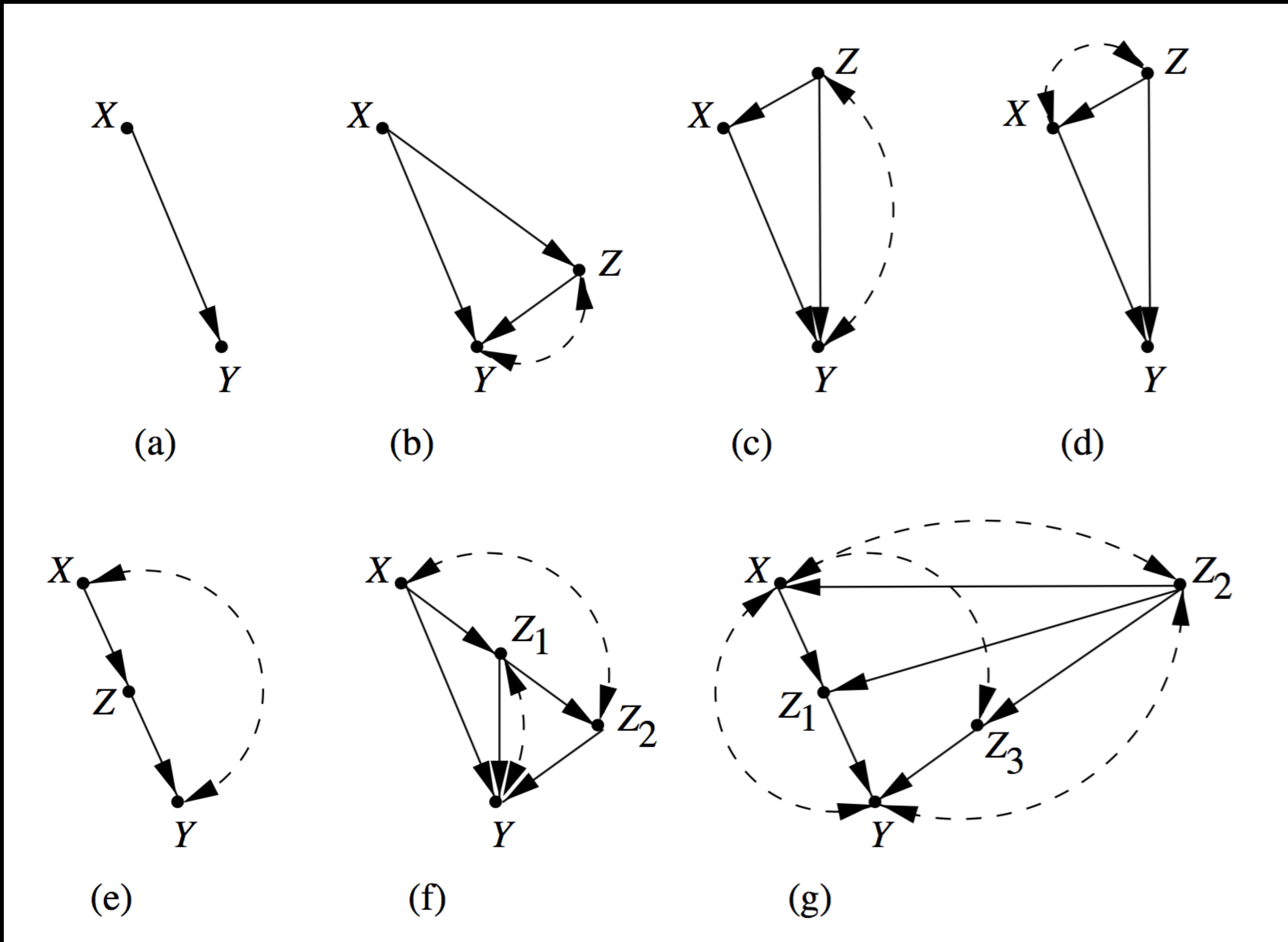


- Bow-arc observation
- $P(y \mid do(x, z_2))$ versus $P(y \mid do(x), z_2)$

NON-IDENTIFABLE MODELS (EXAMPLES)



IDENTIFIABLE MODELS (EXAMPLES)



IDENTIFYING THE EFFECT OF CONDITIONAL ACTIONS

IDENTIFYING THE EFFECT OF CONDITIONAL ACTIONS

$P(y \mid do(X = g(z)))$ = the distribution of Y under the policy $do(X = g(z))$.

IDENTIFYING THE EFFECT OF CONDITIONAL ACTIONS

$P(y \mid do(X = g(z))$ = the distribution of Y under the policy $do(X = g(z))$.

To compute $P(y \mid do(X = g(z)))$, we condition on Z and write,

IDENTIFYING THE EFFECT OF CONDITIONAL ACTIONS

$P(y \mid do(X = g(z))$ = the distribution of Y under the policy $do(X = g(z))$.

To compute $P(y \mid do(X = g(z)))$, we condition on Z and write,

$$P(y \mid do(X = g(z)))$$
$$= \sum_z P(y \mid do(X = g(z)), z) \; P(z \mid do(X = g(z)))$$
$$= \sum_z P(y \mid do(X = g(z)), z) \; P(z)$$
$$= E_z[P(y \mid do(x), z) \mid_{x = g(z)}]$$

IDENTIFYING THE EFFECT OF CONDITIONAL ACTIONS

$P(y \mid do(X = g(z))$ = the distribution of Y under the policy $do(X = g(z))$.

To compute $P(y \mid do(X = g(z)))$, we condition on Z and write,

$$P(y \mid do(X = g(z)))$$

$$= \sum_z P(y \mid do(X = g(z)), z) \; P(z \mid do(X = g(z)))$$

$$= \sum_z P(y \mid do(X = g(z)), z) \; P(z)$$

$$= E_z[P(y \mid do(x), z) \mid_{x = g(z)}]$$

(Conditioning on Z might create dependencies that will prevent the successful reduction of $P(y \mid do(x), z)$ to a do-free expression.)

IDENTIFYING THE EFFECT OF STOCHASTIC POLICIES

IDENTIFYING THE EFFECT OF STOCHASTIC POLICIES

Stochastic policy: enforce the intervention do(X = x) with probability P*(x | z)

IDENTIFYING THE EFFECT OF STOCHASTIC POLICIES

Stochastic policy: enforce the intervention $do(X = x)$ with probability $P^*(x \mid z)$

Given $Z = z$, the intervention $do(X = x)$ will occur with probability $P^*(x \mid z)$ and will produce a causal effect given by $P(y \mid do(x), z)$. Averaging over x and z gives

IDENTIFYING THE EFFECT OF STOCHASTIC POLICIES

Stochastic policy: enforce the intervention do(X = x) with probability P*(x | z)

Given Z = z, the intervention do(X = x) will occur with probability P*(x | z) and will produce a causal effect given by P(y | do(x), z). Averaging over x and z gives

$$P(y)_{\mid P^*(x \mid z)} = \sum_x \sum_z P(y \mid do(x), z)\, P^*(x \mid z)\, P(z)$$

IDENTIFYING THE EFFECT OF STOCHASTIC POLICIES

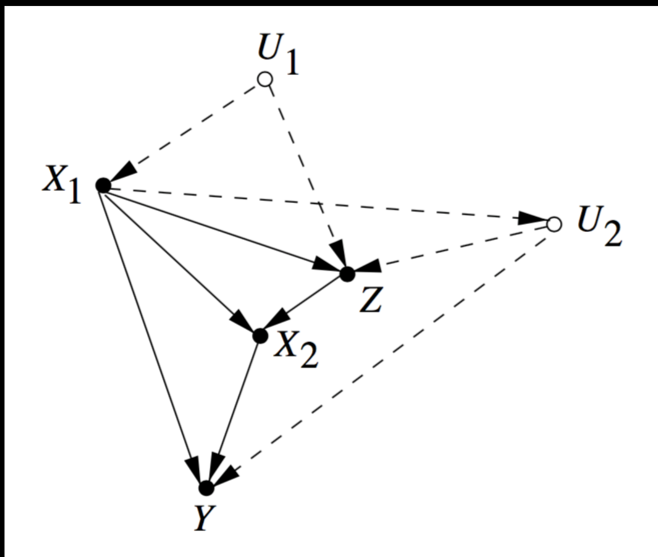
Stochastic policy: enforce the intervention do(X = x) with probability P*(x | z)

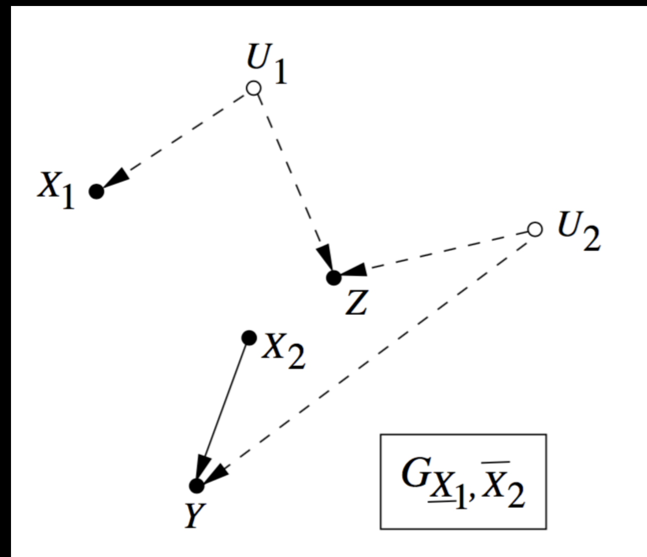
Given Z = z, the intervention do(X = x) will occur with probability P*(x | z) and will produce a causal effect given by P(y | do(x), z). Averaging over x and z gives

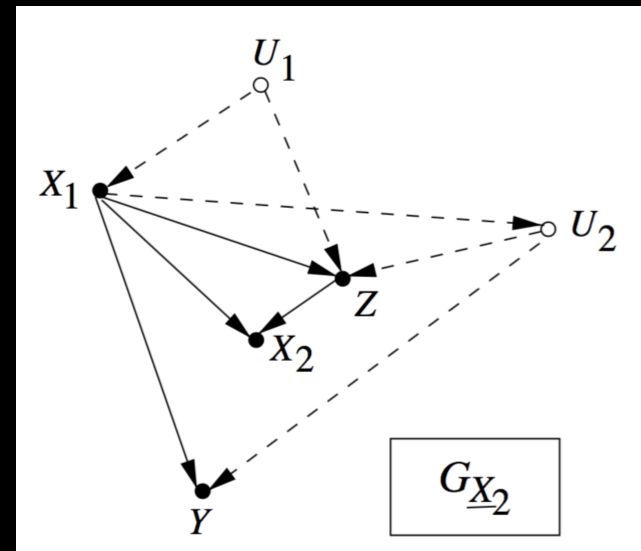
$$P(y)_{\mid P^*(x \mid z)} = \sum_x \sum_z P(y \mid do(x), z)\, P^*(x \mid z)\, P(z)$$

P*(x | z) is specified externally. Therefore, the identifiability of P(y | do(x), z) is a necessary and sufficient condition for the identifiability of any stochastic policy that shapes the distribution of X by the outcome Z.

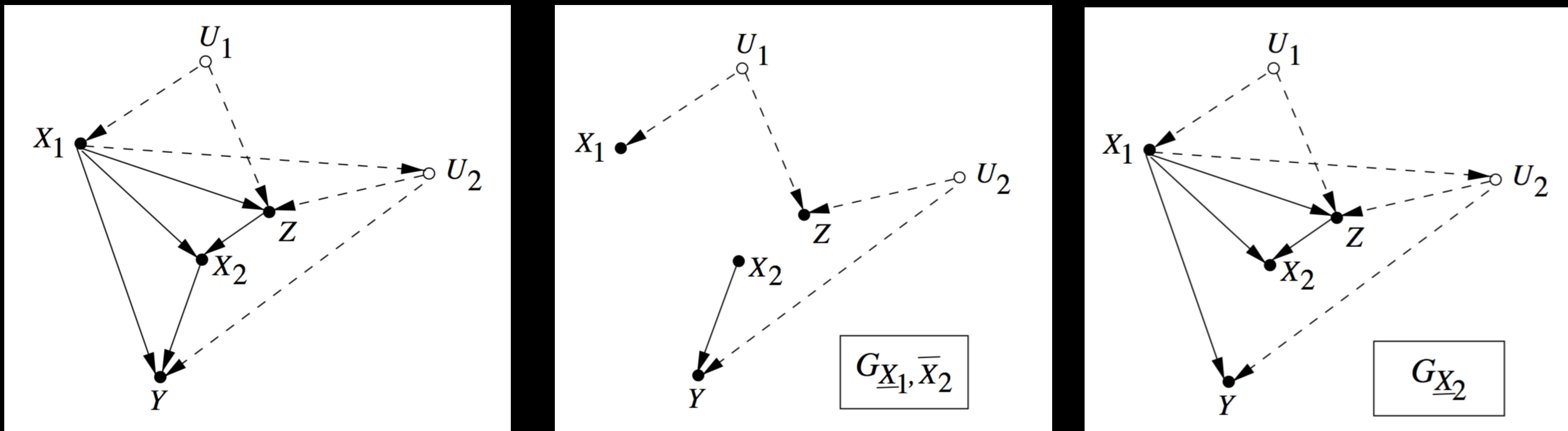
IDENTIFYING THE EFFECT OF PLANS





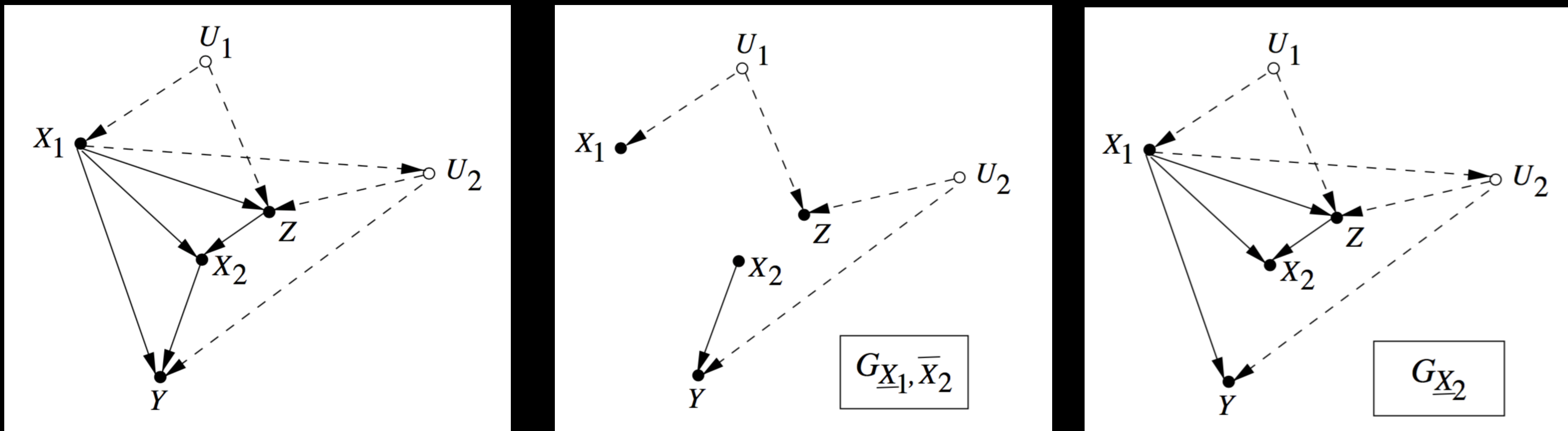


IDENTIFYING THE EFFECT OF PLANS



$$P(y \mid do(x_1, x_2)) = P(y \mid x_1, do(x_2))$$

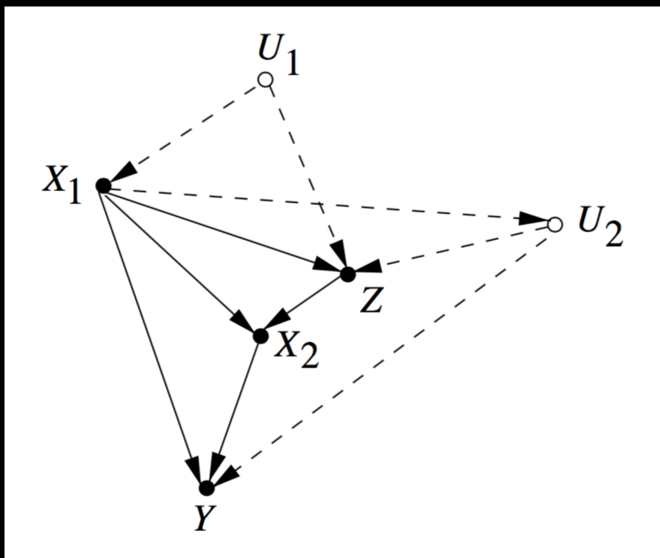
IDENTIFYING THE EFFECT OF PLANS

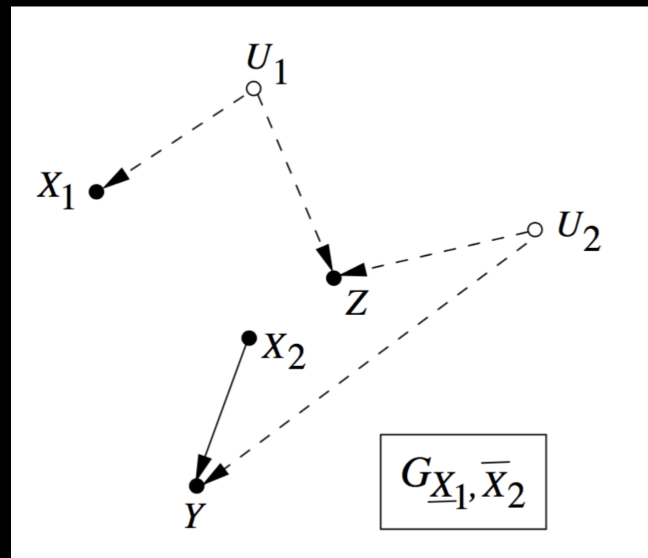


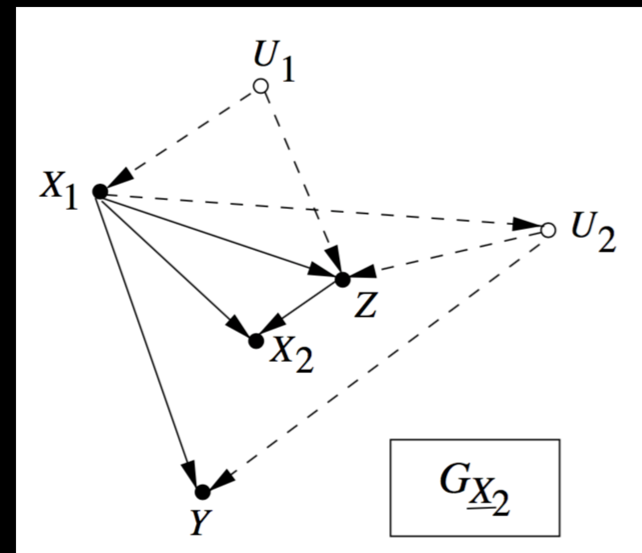
$$P(y \mid do(x_1, x_2)) = P(y \mid x_1, do(x_2))$$

$$= \sum_z P(y \mid x_1, do(x_2), z)\, P(z \mid x_1, do(x_2))$$

IDENTIFYING THE EFFECT OF PLANS





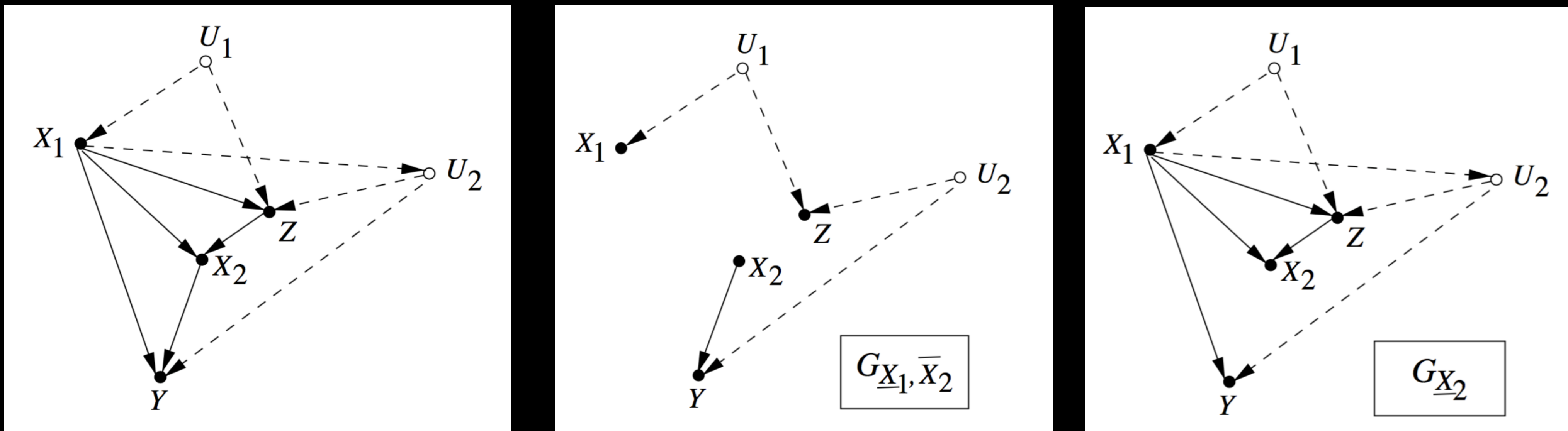


$$P(y \mid do(x_1, x_2)) = P(y \mid x_1, do(x_2))$$

$$= \sum_z P(y \mid x_1, do(x_2), z)\, P(z \mid x_1, do(x_2))$$

$$= \sum_z P(y \mid x_1, do(x_2), z)\, P(z \mid x_1)$$

IDENTIFYING THE EFFECT OF PLANS



$$P(y \mid do(x_1, x_2)) = P(y \mid x_1, do(x_2))$$

$$= \sum_z P(y \mid x_1, do(x_2), z)\, P(z \mid x_1, do(x_2))$$

$$= \sum_z P(y \mid x_1, do(x_2), z)\, P(z \mid x_1)$$

$$= \sum_z P(y \mid x_1, x_2, z)\, P(z \mid x_1)$$

Simpson’s Paradox
(Pearson, 1899, Yule, 1903, Simpson 1951)

Simpson's Paradox
(Pearson, 1899, Yule, 1903, Simpson 1951)

- Pearson's shock: "spurious correlation"

We are thus forced to the conclusion that a mixture of heterogenous groups, each of which exhibits in itself no organic correlation, will exhibit a greater or less amount of correlation. This correlation may properly be called spurious, yet as it is almost impossible to guarantee the absolute homogeneity of any community, our results for correlation are always liable to an error, the amount of which cannot be foretold. To those who persist on looking upon all correlation as cause and effect, the fact that correlation can be produced between two quite uncorrelated characters A and B by taking an artificial mixture of two closely allied races, must come as rather a shock.

[Pearson, Lee, Brandy-Moore (1899)]

- Causation = perfect correlation
- "Not all correlations are causal" (Aldrich, 1994)

Simpson's Paradox
(Pearson, 1899, Yule, 1903, Simpson 1951)

- e.g., UC Berkeley's alleged sex bias in graduate admission (Science, 1975)

Overall, data showed a higher rate of admission among male applicants, but, broken by departments, data showed a slight bias in favor of admitting female applicants.

- e.g., "reverse regression" (1970-1980)

Should one, in salary discrimination cases, compare salaries of equally qualified men and women, or, instead, compare qualifications of equally paid men and woman. (Opposite conclusions.)

Simpson's Paradox

(Robot: seeing and making sense)

	surv. (Y)	not-surv. (¬Y)		Recovery Rate
drug (X)	20	20	40	50%
no-drug (¬X)	16	24	40	40%
	36	44		

$P(Y \mid F, X) < P(Y \mid F, \neg X)$

$P(Y \mid \neg F, X) < P(Y \mid \neg F, \neg X)$

but

$P(Y \mid X) > P(Y \mid \neg X)$!!

male (¬F)	surv. (Y)	not-surv.		Recovery Rate
drug (X)	18	12	30	60%
no-drug (¬X)	7	3	10	70%
	25	15	40	

female (F)	surv. (Y)	not-surv. (¬Y)		Recovery Rate
drug (X)	2	8	10	20%
no-drug (¬X)	9	21	30	30%
	11	29	40	

Simpson's paradox

- Our tables implied what was deemed a paradox:

Simpson's paradox

- Our tables implied what was deemed a paradox:

$$P(Y \mid F, X) < P(Y \mid F, \neg X)$$

Simpson's paradox

- Our tables implied what was deemed a paradox:

$$P(Y \mid F, X) < P(Y \mid F, \neg X)$$

$$P(Y \mid \neg F, X) < P(Y \mid \neg F, \neg X)$$

Simpson's paradox

- Our tables implied what was deemed a paradox:

$$P(Y \mid F, X) < P(Y \mid F, \neg X)$$

$$P(Y \mid \neg F, X) < P(Y \mid \neg F, \neg X)$$

and

Simpson's paradox

- Our tables implied what was deemed a paradox:

$$P(Y \mid F, X) < P(Y \mid F, \neg X)$$

$$P(Y \mid \neg F, X) < P(Y \mid \neg F, \neg X)$$

and

$$P(Y \mid X) > P(Y \mid X)$$

Simpson's paradox

- Our tables implied what was deemed a paradox:

$$P(Y \mid F, X) < P(Y \mid F, \neg X)$$

$$P(Y \mid \neg F, X) < P(Y \mid \neg F, \neg X)$$

and

$$P(Y \mid X) > P(Y \mid X)$$

Dissolving the "paradox" with do-operator

- Our tables implied a deemed paradox:

$$P(Y \mid F, X) < P(Y \mid F, \neg X)$$

$$P(Y \mid \neg F, X) < P(Y \mid \neg F, \neg X)$$

and

$$P(Y \mid X) > P(Y \mid X)$$

- But if we re-write this using the do-notation:

$$P(Y \mid F, do(X)) < P(Y \mid F, do(\neg X))$$

$$P(Y \mid \neg F, do(X)) < P(Y \mid \neg F, do(\neg X))$$

Dissolving the “paradox” with do-operator

- Our tables implied a deemed paradox:

$$P(Y \mid F, X) < P(Y \mid F, \neg X)$$

$$P(Y \mid \neg F, X) < P(Y \mid \neg F, \neg X)$$

and

$$P(Y \mid X) > P(Y \mid X)$$

- But if we re-write this using the do-notation:

$$P(Y \mid F, do(X)) < P(Y \mid F, do(\neg X))$$

$$P(Y \mid \neg F, do(X)) < P(Y \mid \neg F, do(\neg X))$$

- It will not be possible to obtain:

$$P(Y \mid do(X)) > P(Y \mid do(\neg X))$$

Proof

Proof

- $P(Y \mid do(X)) = P(Y \mid do(X), F)\, P(F \mid do(X))$
  $+ P(Y \mid do(X), \neg F)\, P(\neg F \mid do(X))$

Proof

- $P(Y \mid do(X)) = P(Y \mid do(X), F)\, P(F \mid do(X))$

$$+ P(Y \mid do(X), \neg F)\, P(\neg F \mid do(X))$$

$$= P(Y \mid do(X), F)\, P(F) + P(Y \mid do(X), \neg F)\, P(\neg F)$$

Proof

- $$P(Y \mid do(X)) = P(Y \mid do(X), F)\, P(F \mid do(X)) + P(Y \mid do(X), \neg F)\, P(\neg F \mid do(X))$$
  $$= P(Y \mid do(X), F)\, P(F) + P(Y \mid do(X), \neg F)\, P(\neg F)$$

- $P(Y \mid do(\neg X)) = P(Y \mid do(\neg X), F)\, P(F) + P(Y \mid do(\neg X), \neg F)\, P(\neg F)$

Proof

- $P(Y \mid do(X)) = P(Y \mid do(X), F)\, P(F \mid do(X))$
  $\quad + P(Y \mid do(X), \neg F)\, P(\neg F \mid do(X))$
  $\quad = P(Y \mid do(X), F)\, P(F) + P(Y \mid do(X), \neg F)\, P(\neg F)$

- $P(Y \mid do(\neg X)) = P(Y \mid do(\neg X), F)\, P(F) + P(Y \mid do(\neg X), \neg F)\, P(\neg F)$

- And if assumptions hold:

Proof

- $P(Y \mid do(X)) = P(Y \mid do(X), F)\, P(F \mid do(X)) + P(Y \mid do(X), \neg F)\, P(\neg F \mid do(X))$
  $= P(Y \mid do(X), F)\, P(F) + P(Y \mid do(X), \neg F)\, P(\neg F)$

- $P(Y \mid do(\neg X)) = P(Y \mid do(\neg X), F)\, P(F) + P(Y \mid do(\neg X), \neg F)\, P(\neg F)$

- And if assumptions hold:

$$P(Y \mid do(X), F) < P(Y \mid do(\neg X), F)$$
$$P(Y \mid do(X), \neg F) < P(Y \mid do(\neg X), \neg F)$$

Proof

- $$P(Y \mid do(X)) = P(Y \mid do(X), F)\, P(F \mid do(X)) + P(Y \mid do(X), \neg F)\, P(\neg F \mid do(X))$$
  $$= P(Y \mid do(X), F)\, P(F) + P(Y \mid do(X), \neg F)\, P(\neg F)$$

- $P(Y \mid do(\neg X)) = P(Y \mid do(\neg X), F)\, P(F) + P(Y \mid do(\neg X), \neg F)\, P(\neg F)$

- And if assumptions hold:

$$P(Y \mid do(X), F) < P(Y \mid do(\neg X), F)$$
$$P(Y \mid do(X), \neg F) < P(Y \mid do(\neg X), \neg F)$$

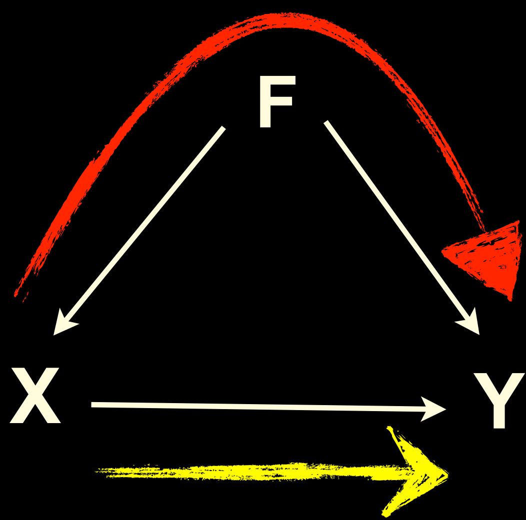
- Then,

$$P(Y \mid do(X)) < P(Y \mid do(\neg X))$$

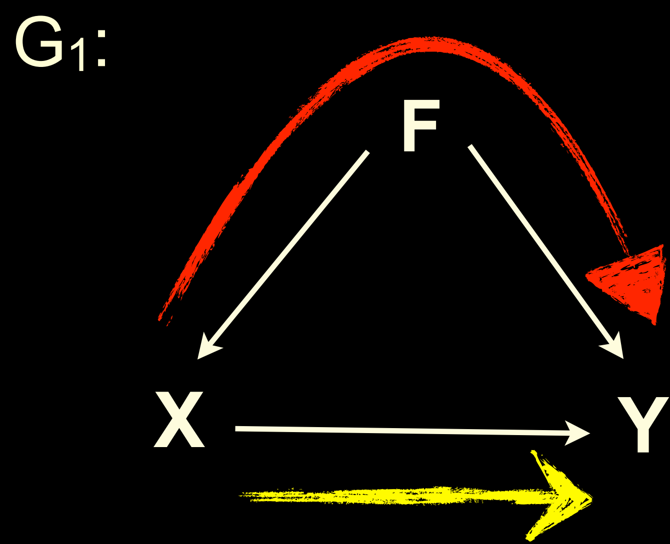
WOULD DECISION-MAKING BE DIFFERENT WITH A DIFFERENT STORY?

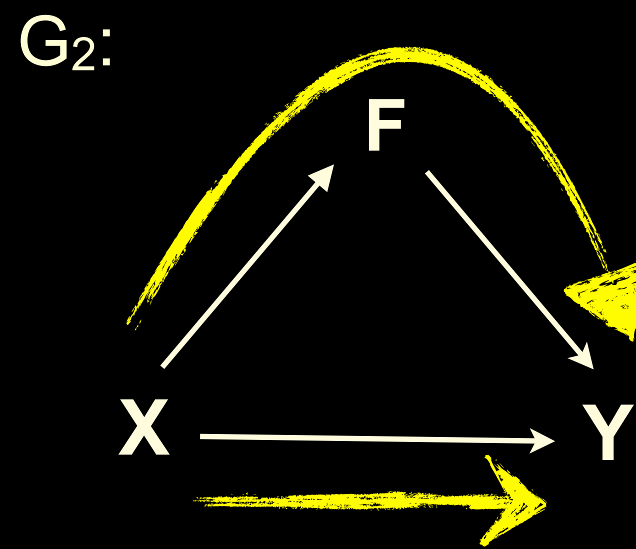
WOULD DECISION-MAKING BE DIFFERENT WITH A DIFFERENT STORY?

G_1:



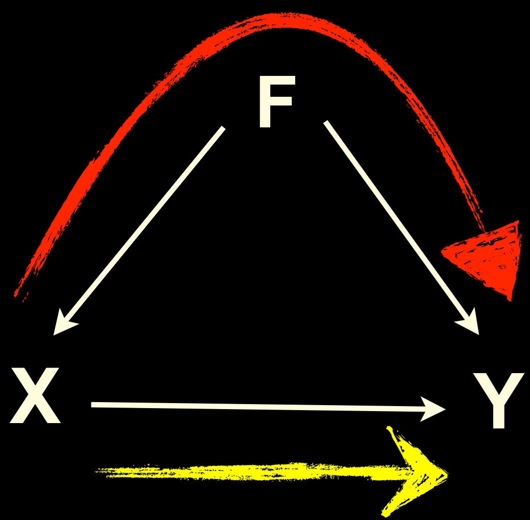
WOULD DECISION-MAKING BE DIFFERENT WITH A DIFFERENT STORY?



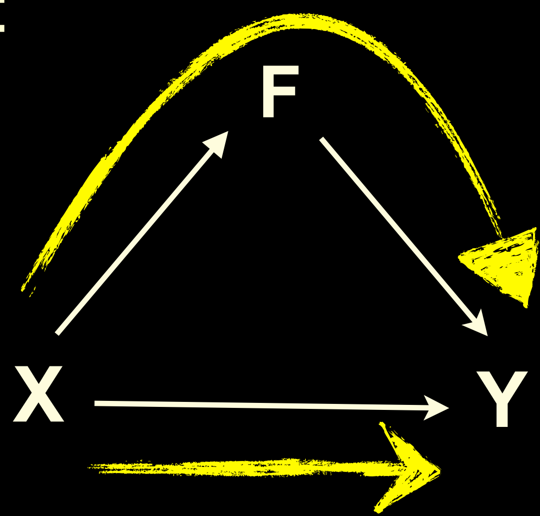


WOULD DECISION-MAKING BE DIFFERENT WITH A DIFFERENT STORY?

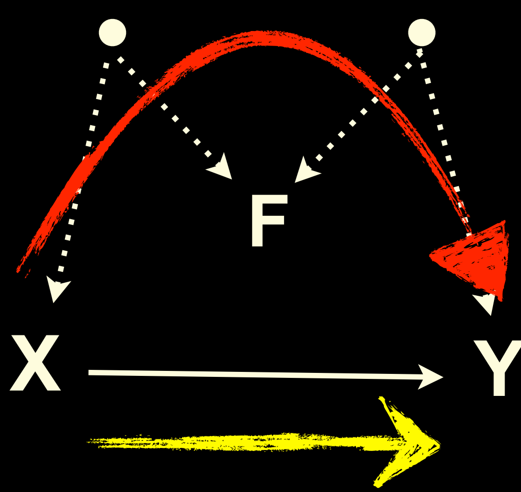
G_1:



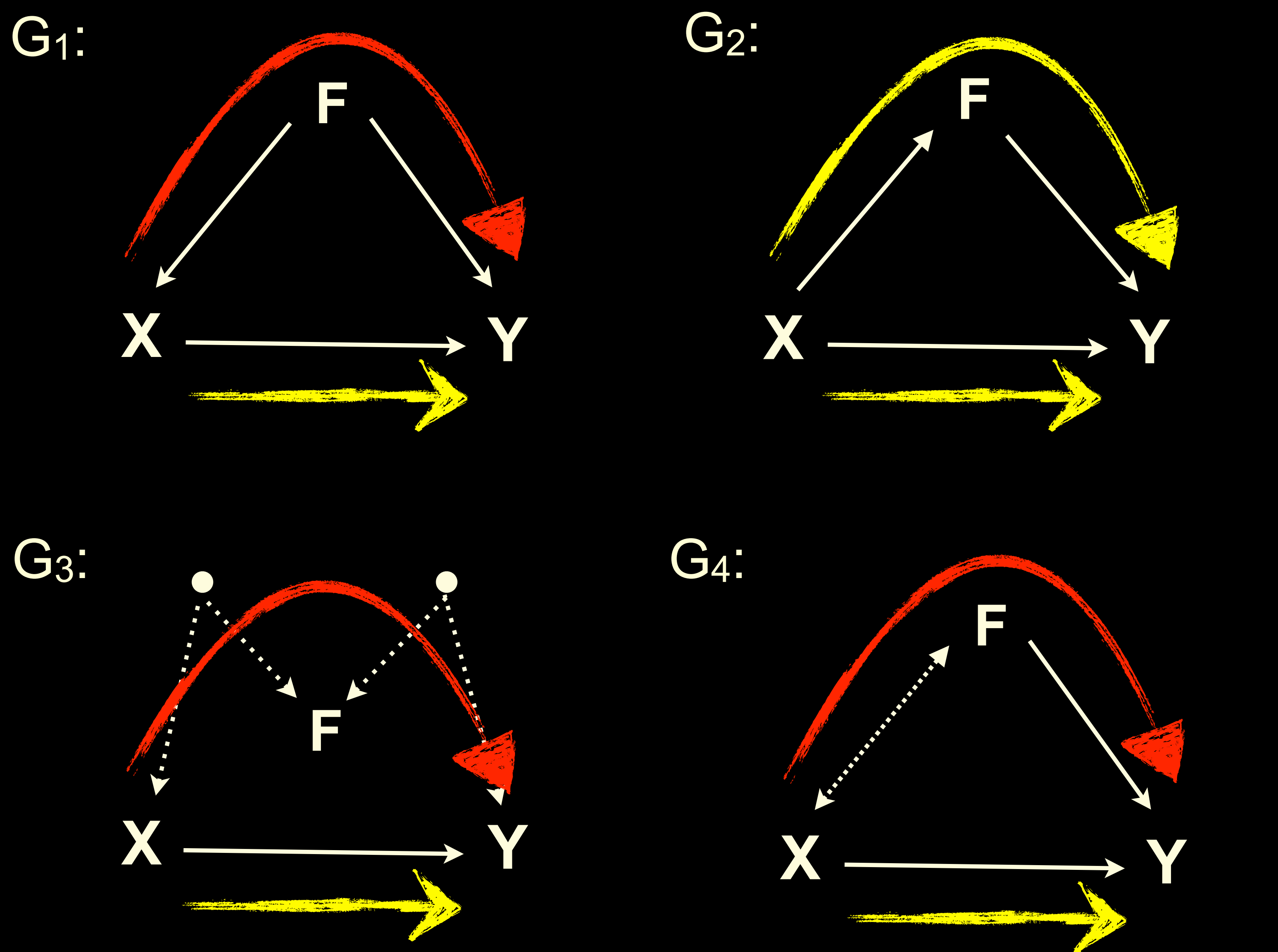
G_2:



G_3:



WOULD DECISION-MAKING BE DIFFERENT WITH A DIFFERENT STORY?



THE SURE THING PRINCIPLE
(formalizing Savage, 1972)

THE SURE THING PRINCIPLE
(formalizing Savage, 1972)

Theorem 6.1.1. An action X that increases the probability of an event Y in each subpopulation must also increase the probability of Y in the population as a whole, provided that the action does not change the distribution of the sub-populations.

THE SURE THING PRINCIPLE
(formalizing Savage, 1972)

Theorem 6.1.1. An action X that increases the probability of an event Y in each subpopulation must also increase the probability of Y in the population as a whole, provided that the action does not change the distribution of the sub-populations.

Summary.

- **Any** statistical relationship can be reversed by including additional factors in the analysis
- “Reversal” is possible in the calculus of proportions, but impossible in the calculus of causes
- People thinks causes, not proportions

SUMMARY

SUMMARY

- The do-calculus provides a syntactical characterization to the problem of policy evaluation for static and dynamic interventions.
- The problem of confounding and identification is essentially solved, non-parametrically.
- Simpson’s Paradox is mathematized and dissolved.
- Applications are pervasive in the social and health sciences as well as in statistics, machine learning, and artificial intelligence.

OUTLINE

- Introduction
  * Data, data, data…
  * Basic definitions
- Technical Results
  * The truncated product formula
  * The back-door adjustment formula
  * The front-door adjustment formula
  * The do-calculus
- Tasks
  * Policy evaluation
  * Generalizability and Robustness
  * Decision-making & RL

OUTLINE

SUMMARY - BIG PICTURE

1. query

2. model

3. data

SUMMARY - BIG PICTURE

1 query

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

2 model

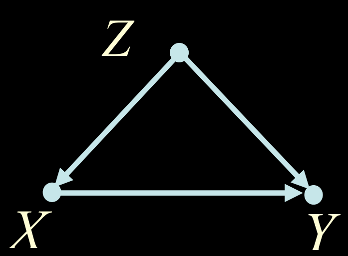
3 data

SUMMARY - BIG PICTURE

1 query

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

2 model



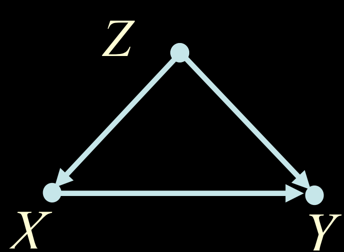
3 data

SUMMARY - BIG PICTURE

1 query

Q = P*(y | do(x))

2 model



3 data

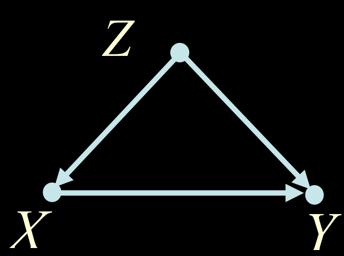
P(x, y, z)

SUMMARY - BIG PICTURE

1 query

Q = P*(y | do(x))

2 model

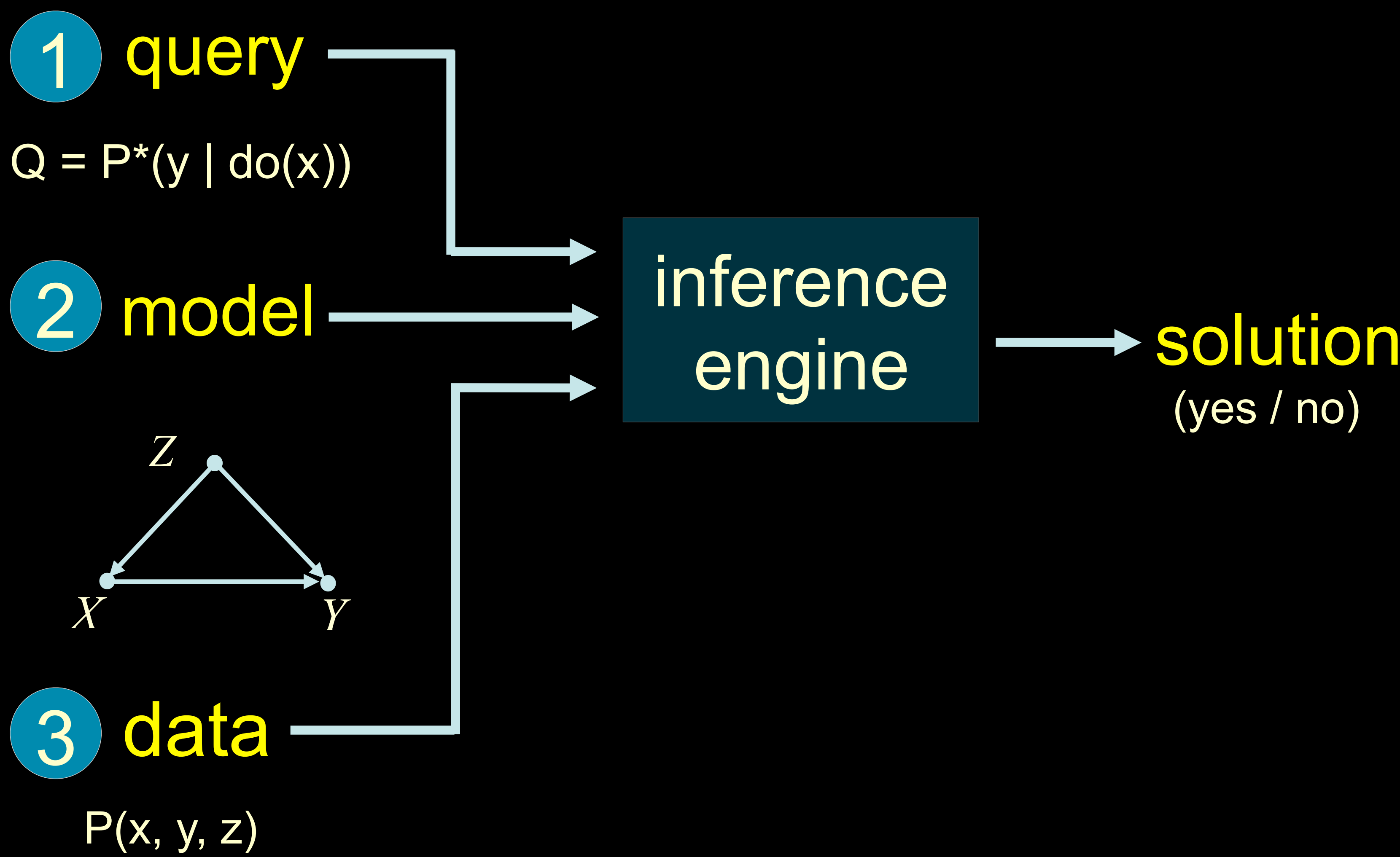


3 data

P(x, y, z)

inference engine

SUMMARY - BIG PICTURE



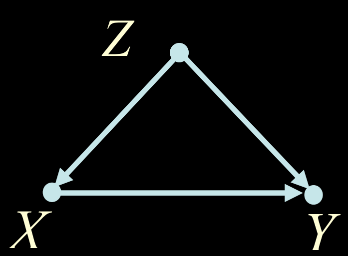
SUMMARY - BIG PICTURE

With the current scientific knowledge about the problem (2) and the available data (3), is it possible to answer the research question (1)?

1 query

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

2 model



3 data

$P(x, y, z)$

inference engine

solution

(yes / no)

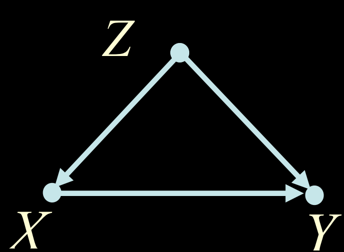
SUMMARY - BIG PICTURE

With the current scientific knowledge about the problem (2) and the available data (3), is it possible to answer the research question (1)?

1 query

Q = P*(y | do(x))

2 model



3 data

P(x, y, z)

inference engine

solution

(yes / no)

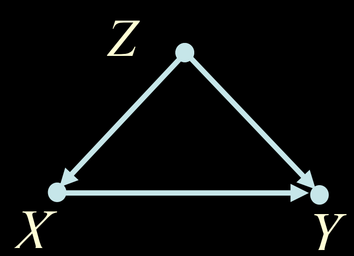
SUMMARY - BIG PICTURE

With the current scientific knowledge about the problem (2) and the available data (3), is it possible to answer the research question (1)?

1 query

Q = P*(y | do(x))

2 model



3 data

P(x, y, z)

inference engine

solution
(yes / no)

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

MOTIVATION FOR DATA-FUSION

MOTIVATION FOR DATA-FUSION

Target population Π^*

Query of interest $Q = P^*(y \mid do(x))$

(a) **US**

Census data available

MOTIVATION FOR DATA-FUSION

Target population Π^*

Query of interest $Q = P^*(y \mid do(x))$

(a) **US**	(b) **New York**
Census data available	Survey data resembling target

MOTIVATION FOR DATA-FUSION

Target population Π^*

Query of interest $Q = P^*(y \mid do(x))$

(a) **US**	(b) **New York**	(c) **Los Angeles**
Census data available	Survey data resembling target	Survey data younger population

MOTIVATION FOR DATA-FUSION

Target population Π^*

Query of interest $Q = P^*(y \mid do(x))$

(a) **US** Census data available	(b) **New York** Survey data resembling target	(c) **Los Angeles** Survey data younger population
(d) **Boston** Age not recorded Mostly successful lawyers	(e) **San Francisco** High post-treatment blood pressure	(f) **Texas** Mostly Spanish subjects High attrition
(g) **Arkansas** Randomized trial College students	(h) **Utah** RCT, paid volunteers, mainly unemployed	(i) **Wyoming** Natural experiment young athletes

CAUSAL DATA SCIENCE
(“All data is not created equal”)

CAUSAL DATA SCIENCE (“All data is not created equal”)

- Heterogenous datasets are pervasive in the empirical sciences since the data is collected:

CAUSAL DATA SCIENCE
(“All data is not created equal”)

- Heterogenous datasets are pervasive in the empirical sciences since the data is collected:

  - under different experimental conditions,

CAUSAL DATA SCIENCE (“All data is not created equal”)

- Heterogenous datasets are pervasive in the empirical sciences since the data is collected:

  - under different experimental conditions,
  - the underlying populations are different,

CAUSAL DATA SCIENCE (“All data is not created equal”)

- Heterogenous datasets are pervasive in the empirical sciences since the data is collected:

  - under different experimental conditions,
  - the underlying populations are different,
  - the sampling procedure is not random,

CAUSAL DATA SCIENCE
(“All data is not created equal”)

- Heterogenous datasets are pervasive in the empirical sciences since the data is collected:

  - under different experimental conditions,
  - the underlying populations are different,
  - the sampling procedure is not random,
  - the treatment assignment is not random,

CAUSAL DATA SCIENCE
("All data is not created equal")

- Heterogenous datasets are pervasive in the empirical sciences since the data is collected:

  - under different experimental conditions,
  - the underlying populations are different,
  - the sampling procedure is not random,
  - the treatment assignment is not random,
  - many variables are not measured.

CAUSAL DATA SCIENCE (“All data is not created equal”)

- Heterogenous datasets are pervasive in the empirical sciences since the data is collected:

  - under different experimental conditions,
  - the underlying populations are different,
  - the sampling procedure is not random,
  - the treatment assignment is not random,
  - many variables are not measured.

- All these dimensions are now formalized.

CAUSAL DATA SCIENCE
("All data is not created equal")

- Heterogenous datasets are pervasive in the empirical sciences since the data is collected:

  - under different experimental conditions,
  - the underlying populations are different,
  - the sampling procedure is not random,
  - the treatment assignment is not random,
  - many variables are not measured.

- All these dimensions are now formalized.

- And there are conditions and algorithms to decide what is "entailed" from a certain data collection.

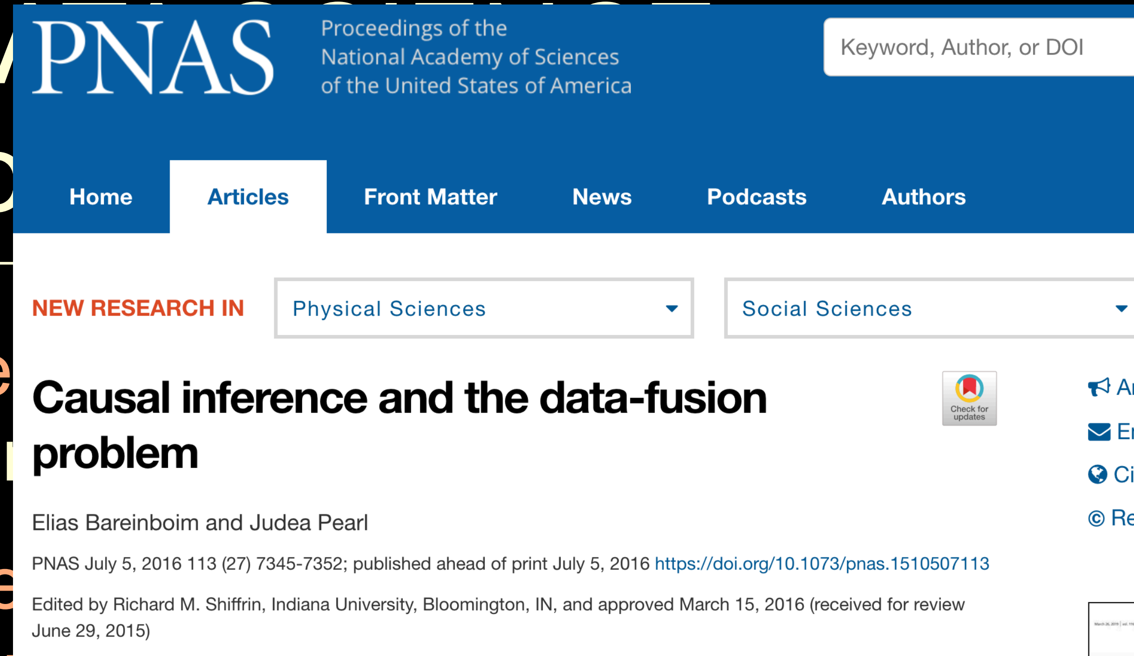
CAUSAL DA
(“All data is no

- Heterogenous datase
  empirical sciences si

  - under different expe
  - the underlying populations are different,
  - the sampling procedure is not random,
  - the treatment assignment is not random,
  - many variables are not measured.

- All these dimensions are now formalized.

- And there are conditions and algorithms to decide what is “entailed” from a certain data collection.



PNAS Proceedings of the National Academy of Sciences of the United States of America

Keyword, Author, or DOI

Home | Articles | Front Matter | News | Podcasts | Authors

NEW RESEARCH IN Physical Sciences | Social Sciences

Causal inference and the data-fusion problem

Elias Bareinboim and Judea Pearl

PNAS July 5, 2016 113 (27) 7345-7352; published ahead of print July 5, 2016 https://doi.org/10.1073/pnas.1510507113

Edited by Richard M. Shiffrin, Indiana University, Bloomington, IN, and approved March 15, 2016 (received for review June 29, 2015)

CAUSAL DATA SCIENCE (“All data is not created equal”)

- Heterogenous datasets are pervasive in the empirical sciences since the data is collected:

  - under different experimental conditions,
  - the underlying populations are different,
  - the sampling procedure is not random,
  - the treatment assignment is not random,
  - many variables are not measured.

- All these dimensions are now formalized.

- And there are conditions and algorithms to decide what is “entailed” from a certain data collection.

HETEROGENOUS DATASETS

Target $Q = P^*(y \mid do(x))$	Dataset 1	Dataset 2	…	Dataset n

Population	Los Angeles	New York	Texas
Obs. / Exp.	Experimental	Observational	Experimental
Treat. Assign.	Randomized Z_1	-	Randomized Z_2
Sampling	Selection on Age	Selection on SES	-
Measured	X_1, Z_1, W, M, Y_1	X_1, X_2, Z_1, N, Y_2	X_2, Z_1, W, L, M, Y_1

HETEROGENOUS DATASETS

Target	Dataset 1	Dataset 2	…	Dataset n
Target $Q = P^*(y \mid do(x))$				

		Los Angeles	New York	Texas
d_1	Population	Los Angeles	New York	Texas
	Obs. / Exp.	Experimental	Observational	Experimental
	Treat. Assign.	Randomized Z_1	-	Randomized Z_2
	Sampling	Selection on Age	Selection on SES	-
	Measured	X_1, Z_1, W, M, Y_1	X_1, X_2, Z_1, N, Y_2	X_2, Z_1, W, L, M, Y_1

HETEROGENOUS DATASETS

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

Dataset 1 | Dataset 2 | … | Dataset n

d_1				
	Population	Los Angeles	New York	Texas
	Obs. / Exp.	Experimental	Observational	Experimental
	Treat. Assign.	Randomized Z_1	-	Randomized Z_2
	Sampling	Selection on Age	Selection on SES	-
	Measured	X_1, Z_1, W, M, Y_1	X_1, X_2, Z_1, N, Y_2	X_2, Z_1, W, L, M, Y_1

HETEROGENOUS DATASETS

	Dataset 1	Dataset 2 …	Dataset n
Target $Q = P^*(y \mid do(x))$			

d_1	Population	Los Angeles	New York	Texas
d_2	Obs. / Exp.	Experimental	Observational	Experimental
	Treat. Assign.	Randomized Z_1	-	Randomized Z_2
	Sampling	Selection on Age	Selection on SES	-
	Measured	X_1, Z_1, W, M, Y_1	X_1, X_2, Z_1, N, Y_2	X_2, Z_1, W, L, M, Y_1

HETEROGENOUS DATASETS

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

Dataset 1 | Dataset 2 | … | Dataset n

		Dataset 1	Dataset 2	Dataset n
d_1	Population	Los Angeles	New York	Texas
d_2	Obs. / Exp.	Experimental	Observational	Experimental
	Treat. Assign.	Randomized Z_1	-	Randomized Z_2
	Sampling	Selection on Age	Selection on SES	-
	Measured	X_1, Z_1, W, M, Y_1	X_1, X_2, Z_1, N, Y_2	X_2, Z_1, W, L, M, Y_1

HETEROGENOUS DATASETS

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

Dataset 1 | Dataset 2 | … | Dataset n

		Dataset 1	Dataset 2	Dataset n
d_1	Population	Los Angeles	New York	Texas
d_2	Obs. / Exp.	Experimental	Observational	Experimental
	Treat. Assign.	Randomized Z_1	-	Randomized Z_2
d_3	Sampling	Selection on Age	Selection on SES	-
	Measured	X_1, Z_1, W, M, Y_1	X_1, X_2, Z_1, N, Y_2	X_2, Z_1, W, L, M, Y_1

HETEROGENOUS DATASETS

Target	Dataset 1	Dataset 2	…	Dataset n
Target $Q = P^*(y \mid do(x))$				

d_1	Population	Los Angeles	New York	Texas
d_2	Obs. / Exp.	Experimental	Observational	Experimental
	Treat. Assign.	Randomized Z_1	-	Randomized Z_2
d_3	Sampling	Selection on Age	Selection on SES	-
	Measured	X_1, Z_1, W, M, Y_1	X_1, X_2, Z_1, N, Y_2	X_2, Z_1, W, L, M, Y_1

HETEROGENOUS DATASETS

Target $Q = P^*(y \mid do(x))$	Dataset 1	Dataset 2	…	Dataset n

d_1	Population	Los Angeles	New York	Texas
d_2	Obs. / Exp.	Experimental	Observational	Experimental
	Treat. Assign.	Randomized Z_1	-	Randomized Z_2
d_3	Sampling	Selection on Age	Selection on SES	-
d_4	Measured	X_1, Z_1, W, M, Y_1	X_1, X_2, Z_1, N, Y_2	X_2, Z_1, W, L, M, Y_1

HETEROGENOUS DATASETS

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

Dataset 1 | Dataset 2 | … | Dataset n

d_1	Population	Los Angeles	New York	Texas
d_2	Obs. / Exp.	Experimental	Observational	Experimental
	Treat. Assign.	Randomized Z_1	-	Randomized Z_2
d_3	Sampling	Selection on Age	Selection on SES	-
d_4	Measured	X_1, Z_1, W, M, Y_1	X_1, X_2, Z_1, N, Y_2	X_2, Z_1, W, L, M, Y_1

DATA-FUSION CHALLENGES

Dimensions

DATA-FUSION CHALLENGES

Description of each dataset: tuple (d_1, d_2, d_3, d_4)
(population, obs./exp., sampling, measure.)

Dimensions

DATA-FUSION CHALLENGES

Description of each dataset: tuple (d_1, d_2, d_3, d_4)
(population, obs./exp., sampling, measure.)

Dimensions

1. *Causal Inference* (observational studies)
   (d_1, Observ., d_3, d_4) $\rightarrow$ (d_1, Experiment(X), d_3, d_4)

DATA-FUSION CHALLENGES

Description of each dataset: tuple (d_1, d_2, d_3, d_4)
(population, obs./exp., sampling, measure.)

Dimensions

1. *Causal Inference* (observational studies)
   $(d_1, \text{Observ.}, d_3, d_4) \rightarrow (d_1, \text{Experiment}(X), d_3, d_4)$

2. *Transportability* (External Validity)
   $(\text{Bonobos}, d_2, d_3, d_4) \rightarrow (\text{Humans}, d_2, d_3, d_4)$

DATA-FUSION CHALLENGES

Description of each dataset: tuple (d_1, d_2, d_3, d_4)
(population, obs./exp., sampling, measure.)

Dimensions

1. *Causal Inference* (observational studies)
   (d_1, Observ., d_3, d_4) $\longrightarrow$ (d_1, Experiment(X), d_3, d_4)

2. *Transportability* (External Validity)
   (Bonobos, d_2, d_3, d_4) $\longrightarrow$ (Humans, d_2, d_3, d_4)

3. *Sampling Selection Bias*
   (d_1, d_2, Select(Salary), d_4) $\longrightarrow$ (d_1, d_2, {}, d_4)

DATA-FUSION CHALLENGES

Description of each dataset: tuple (d_1, d_2, d_3, d_4)
(population, obs./exp., sampling, measure.)

Dimensions

1. *Causal Inference* (observational studies)
   $(d_1, \text{Observ.}, d_3, d_4) \rightarrow (d_1, \text{Experiment}(X), d_3, d_4)$

2. *Transportability* (External Validity)
   $(\text{Bonobos}, d_2, d_3, d_4) \rightarrow (\text{Humans}, d_2, d_3, d_4)$

3. *Sampling Selection Bias*
   $(d_1, d_2, \text{Select}(\text{Salary}), d_4) \rightarrow (d_1, d_2, \{\}, d_4)$

Data-fusion:

DATA-FUSION CHALLENGES

Description of each dataset: tuple (d_1, d_2, d_3, d_4)
(population, obs./exp., sampling, measure.)

Dimensions

1. *Causal Inference* (observational studies)
   $(d_1, \text{Observ.}, d_3, d_4) \rightarrow (d_1, \text{Experiment}(X), d_3, d_4)$

2. *Transportability* (External Validity)
   $(\text{Bonobos}, d_2, d_3, d_4) \rightarrow (\text{Humans}, d_2, d_3, d_4)$

3. *Sampling Selection Bias*
   $(d_1, d_2, \text{Select}(\text{Salary}), d_4) \rightarrow (d_1, d_2, \{\}, d_4)$

Data-fusion:
$\{ (d_1, d_2, d_3, d_4) \} \rightarrow (d'_1, d'_2, d'_3, d'_4)$

DATA-FUSION IN 3-D

1. Experimental conditions
   - Preliminary. Observational to Experimental regime
2. Environmental conditions
3. Sampling conditions

DATA-FUSION IN 3-D

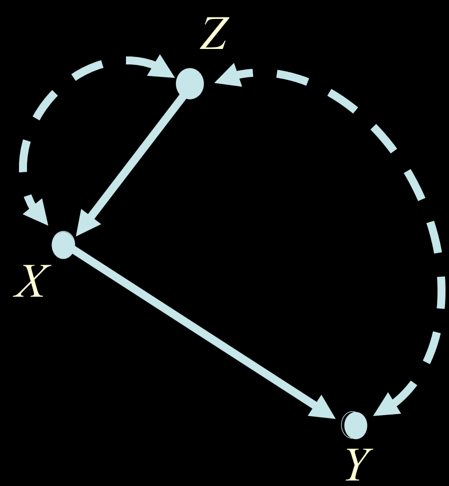
THE DATA-FUSION PROBLEM

EXPERIMENTAL CONDITIONS (D = 1) —
GENERAL IDENTIFIABILITY

THE CHALLENGE OF EXPERIMENTAL IDENTIFIABILITY

(also called z-ID)

G:



Z: Diet
X: Cholesterol level
Y: Heart Attack

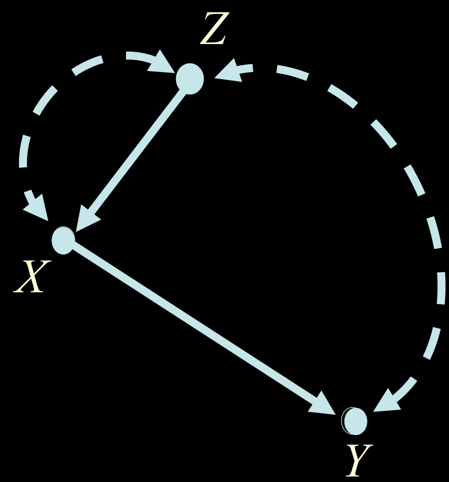
Measured:

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?

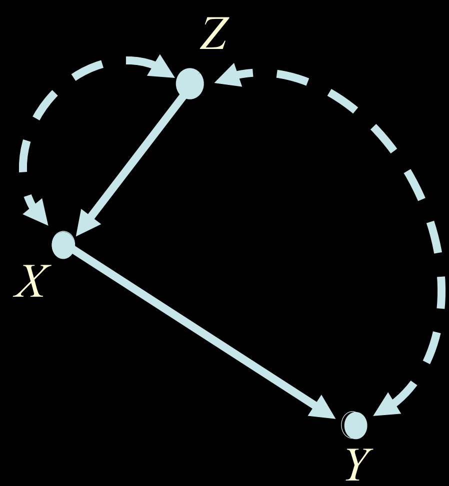
G:



Z: Diet
X: Cholesterol level
Y: Heart Attack

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

G:

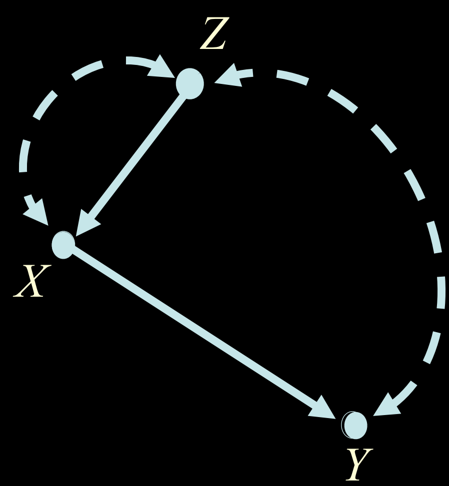


Z: Diet
X: Cholesterol level
Y: Heart Attack

Measured:

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

G:

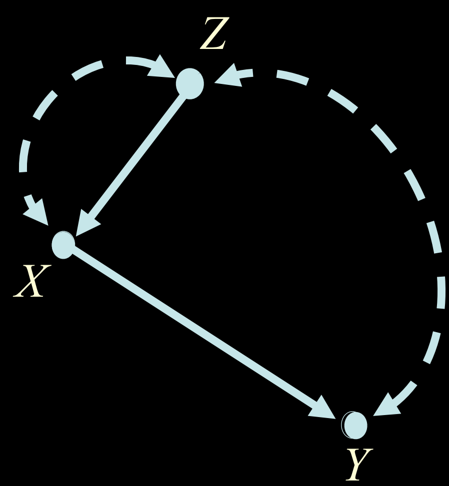


Z: Diet
X: Cholesterol level
Y: Heart Attack

Measured:

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

G:



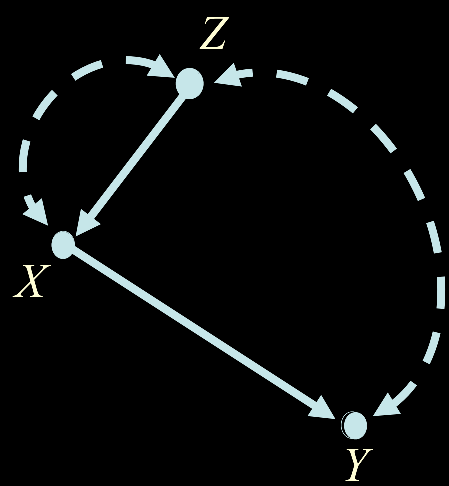
Z: Diet
X: Cholesterol level
Y: Heart Attack

Measured:

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

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

G:



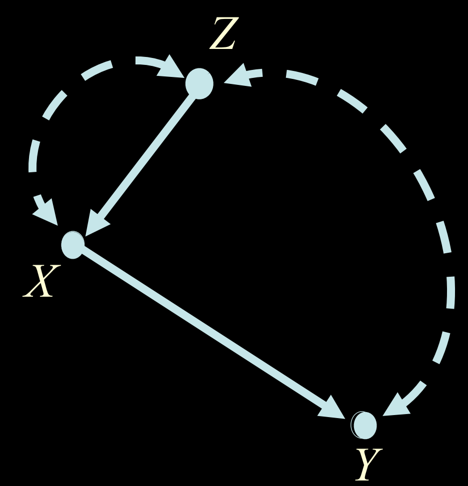
Z: Diet
X: Cholesterol level
Y: Heart Attack

Measured:

Observational study: *P(x, y, z)*
Experimental study: *P(x, y | do(z))*

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

G:



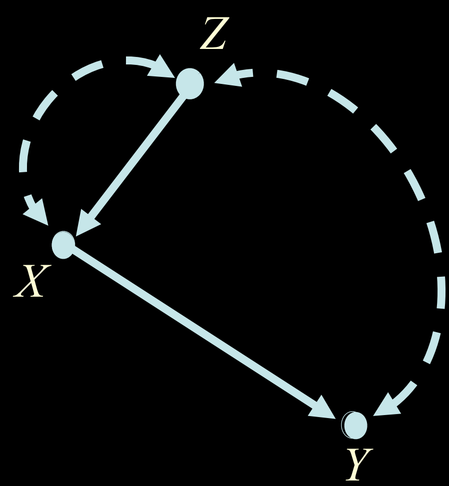
Z: Diet
X: Cholesterol level
Y: Heart Attack

Measured:

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

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

G:



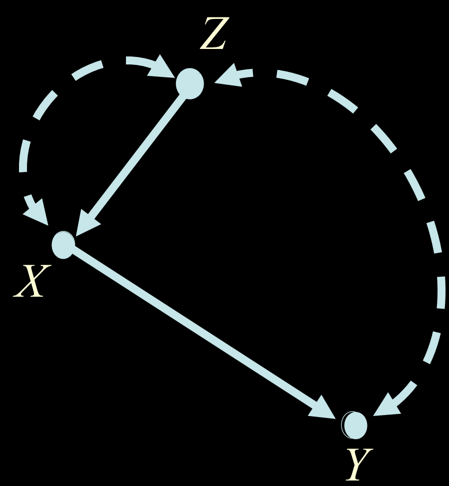
Z: Diet
X: Cholesterol level
Y: Heart Attack

Measured:

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

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

G:



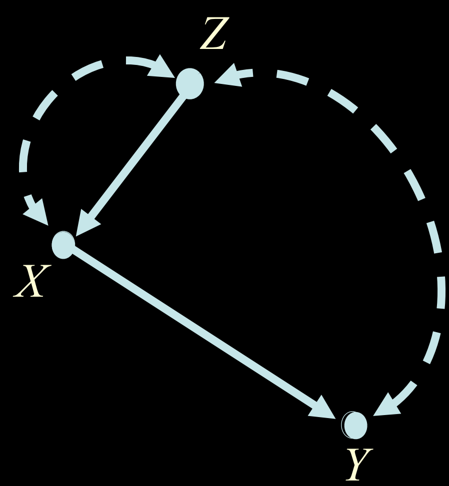
Z: Diet
X: Cholesterol level
Y: Heart Attack

Measured:

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

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

G:



Z: Diet
X: Cholesterol level
Y: Heart Attack

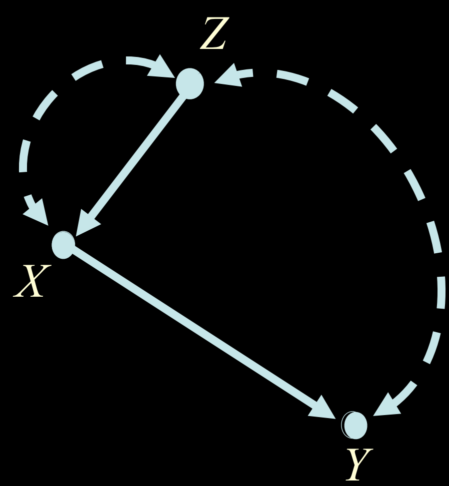
Measured:

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

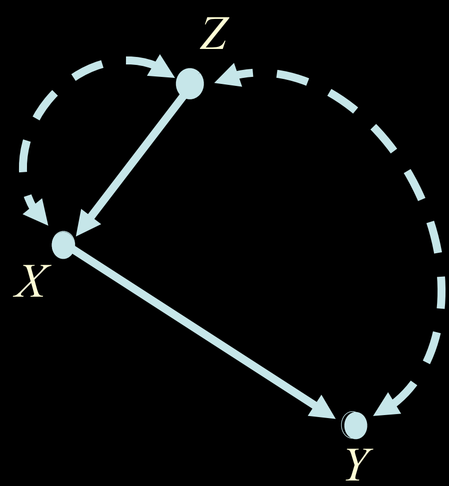
Measured:

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

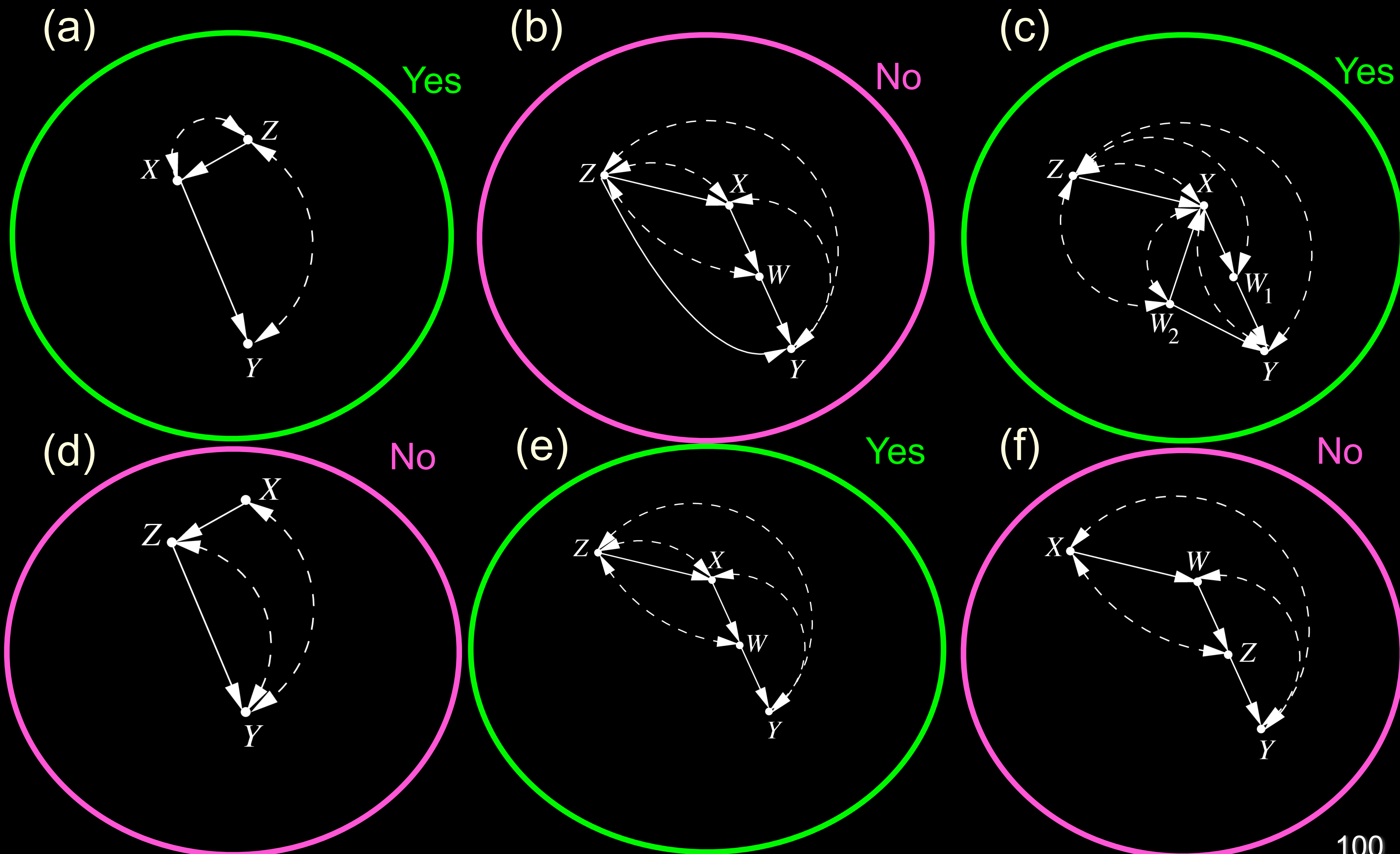
Measured:

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

WHICH MODEL LICENSES THE z-IDENTIFICATION OF THE CAUSAL EFFECT $X \rightarrow Y$?



SUMMARY D=1: POLICY EVALUATION RESULTS

SUMMARY D=1: 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.

SUMMARY D=1: 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 by a combination of observational and experimental studies.)
- The causal calculus is complete for this task.

SUMMARY D=1: 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 by a combination of observational and experimental studies.)
- The causal calculus is complete for this task.
- The algorithm is complete.

SUMMARY D=1: 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 by a combination of observational and experimental studies.)
- The causal calculus is complete for this task.
- The algorithm is complete.
- References

  Pearl (1993, 1995) – Back-door criterion & *do*-calculus
  Tian & Pearl (2002) – General ID algorithm
  Shpitser & Pearl (2006) – Completeness (observational)
  Bareinboim & Pearl (2012) – Completeness (obs. + experimental)
  Lee, Correa & Bareinboim (2019) – Completeness (arbitrary exp.)

THE DATA-FUSION PROBLEM

ENVIRONMENTAL CONDITIONS (D = 2) —
TRANSPORTABILITY, EXTERNAL VALIDITY, META-ANALYSIS

THE TRANSPORTABILITY PROBLEM

Question:

Is it possible to predict the effect of X on Y in a target population $\prod^*$ (where no experiments are feasible), using experimental findings from a different population $\prod$?

Answer: Sometimes yes.

THE TRANSPORTABILITY PROBLEM

Question:

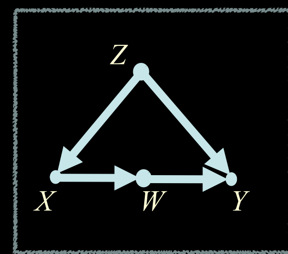
Is it possible to predict the effect of X on Y in a target population $\prod^*$ (where no experiments are feasible), using experimental findings from a different population $\prod$?

Answer: Sometimes yes.

Our goal is to formally characterize when and how.

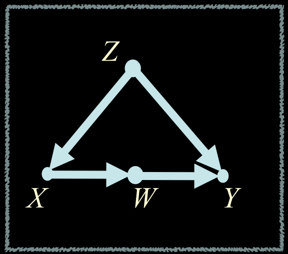
MOVING FROM THE “LAB” TO THE “REAL WORLD”...

Lab



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

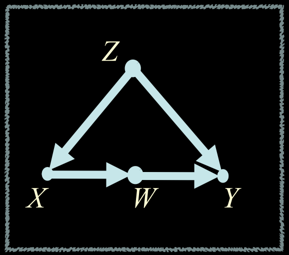
Lab



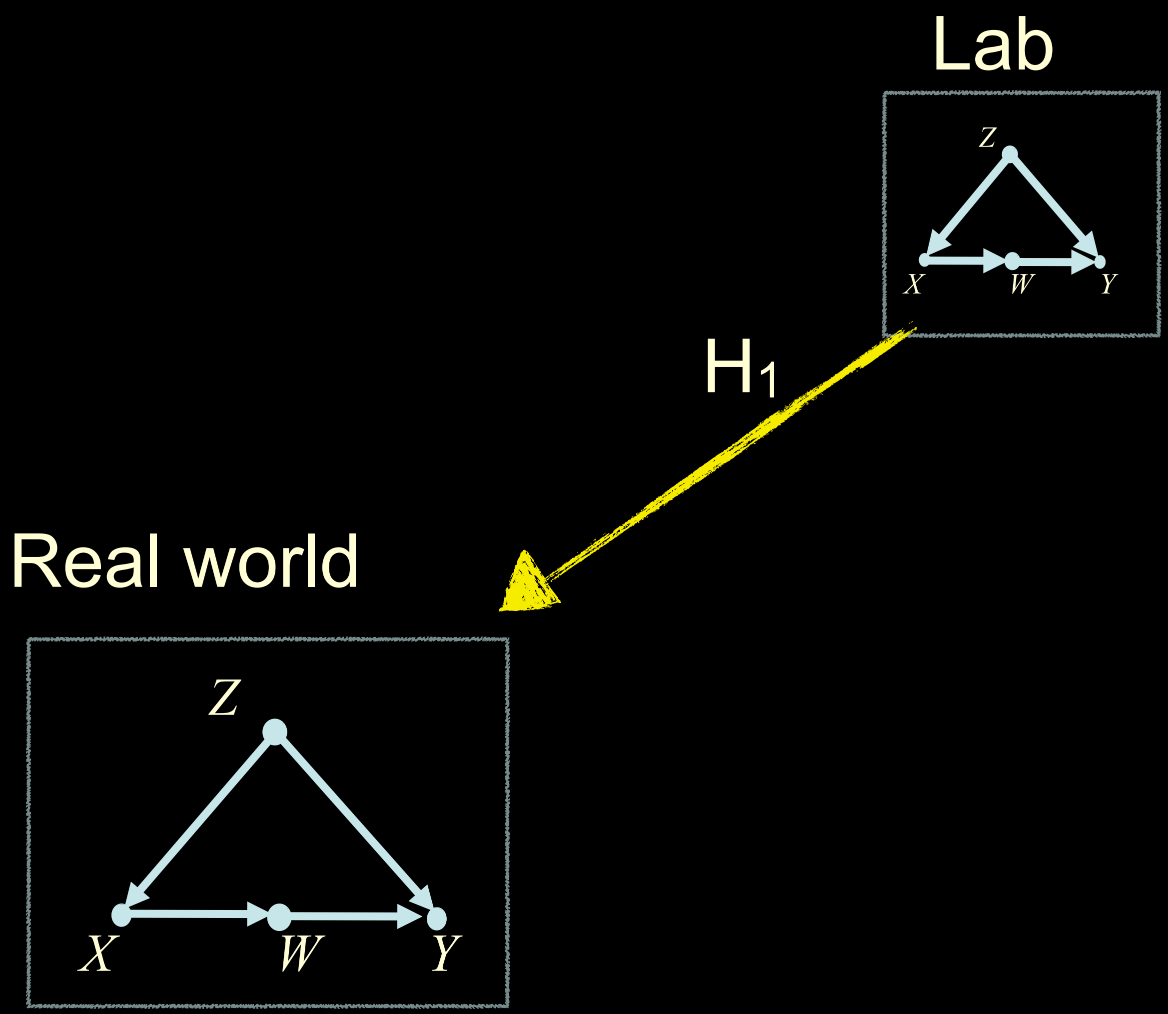
* The lab stands for any environment, population, domain, setting.

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

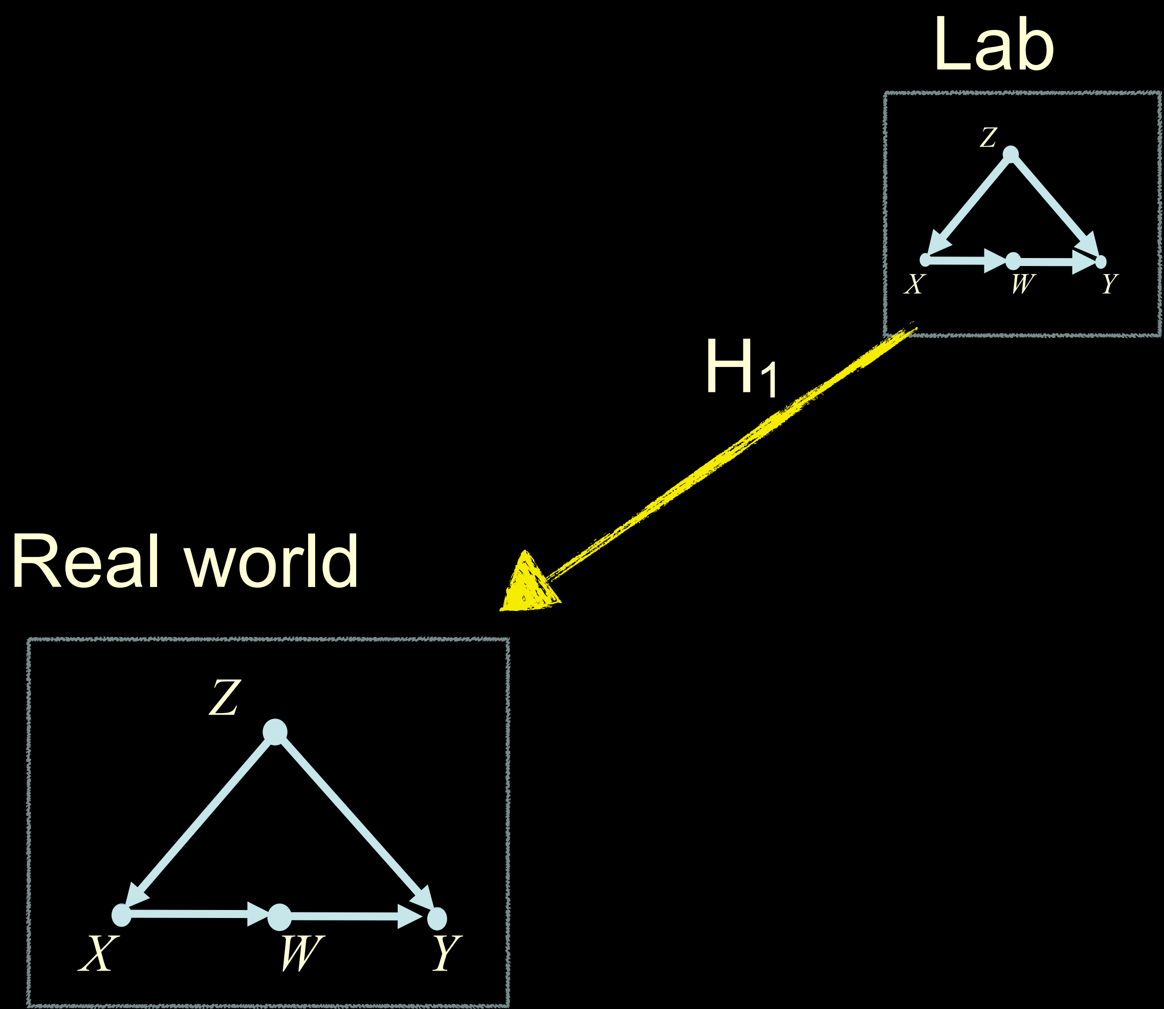
Lab



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

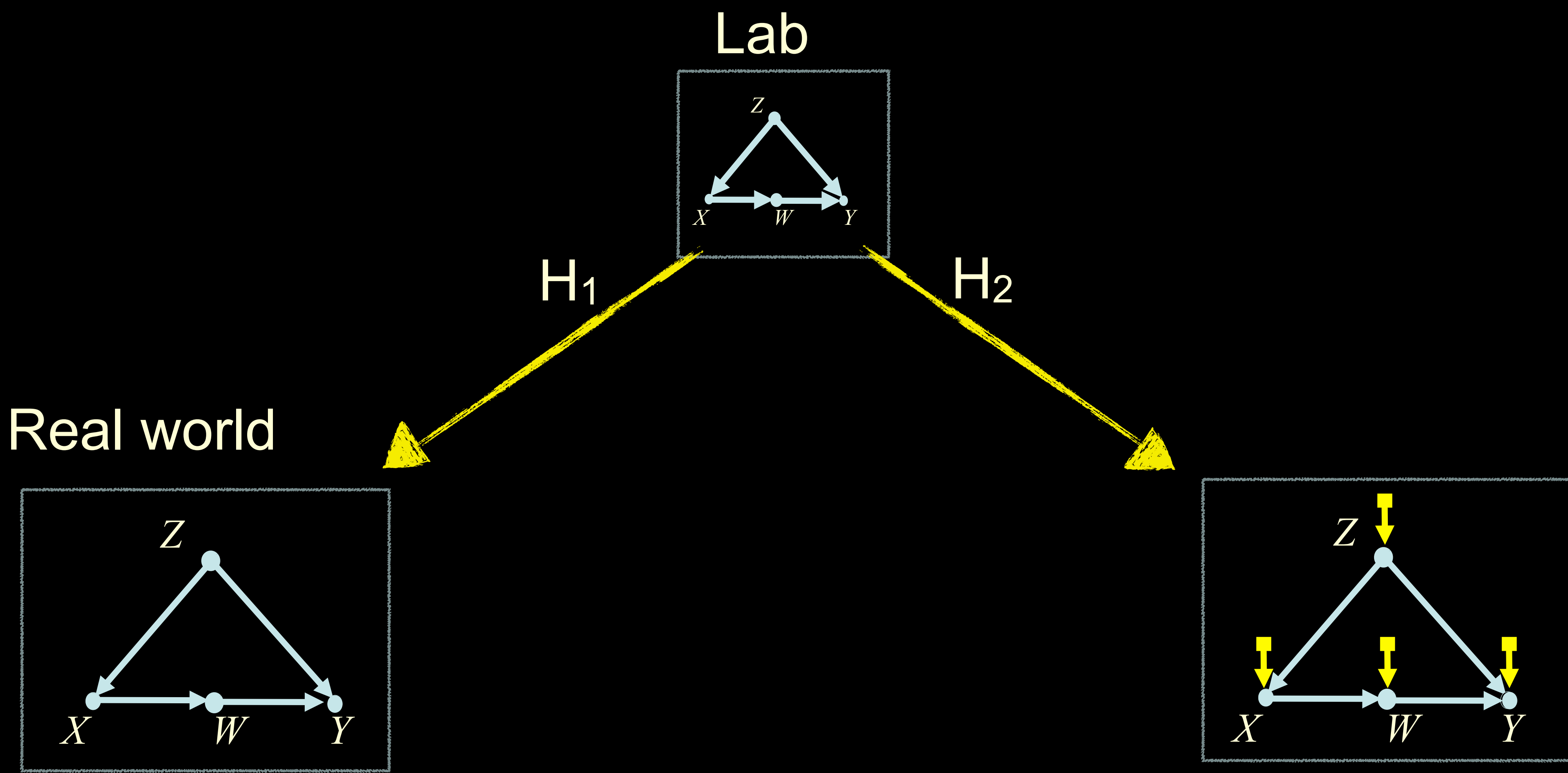


MOVING FROM THE "LAB" TO THE "REAL WORLD"...



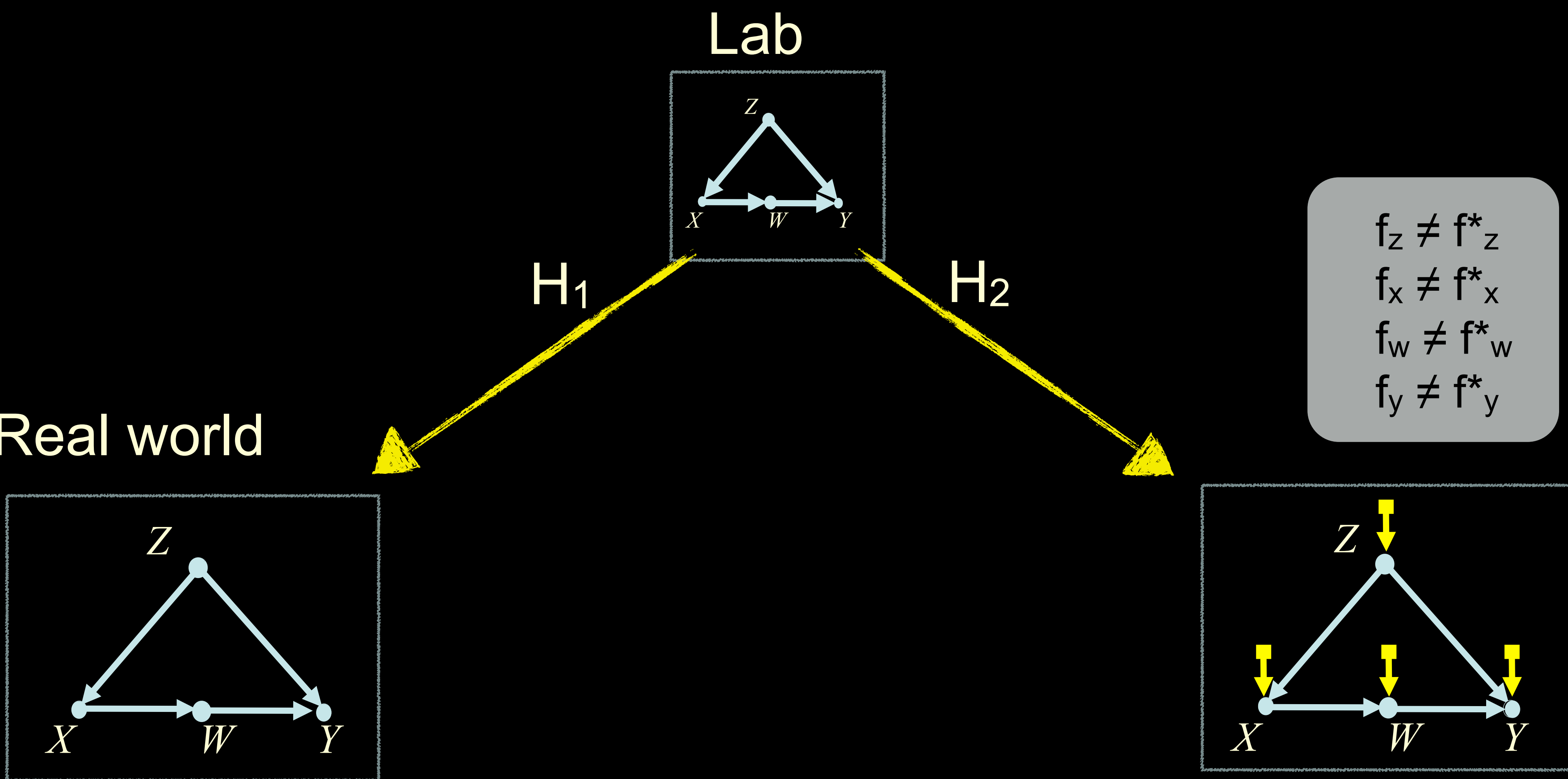
Everything is assumed to be
the same, trivially transportable!

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



Everything is assumed to be
the same, trivially transportable!

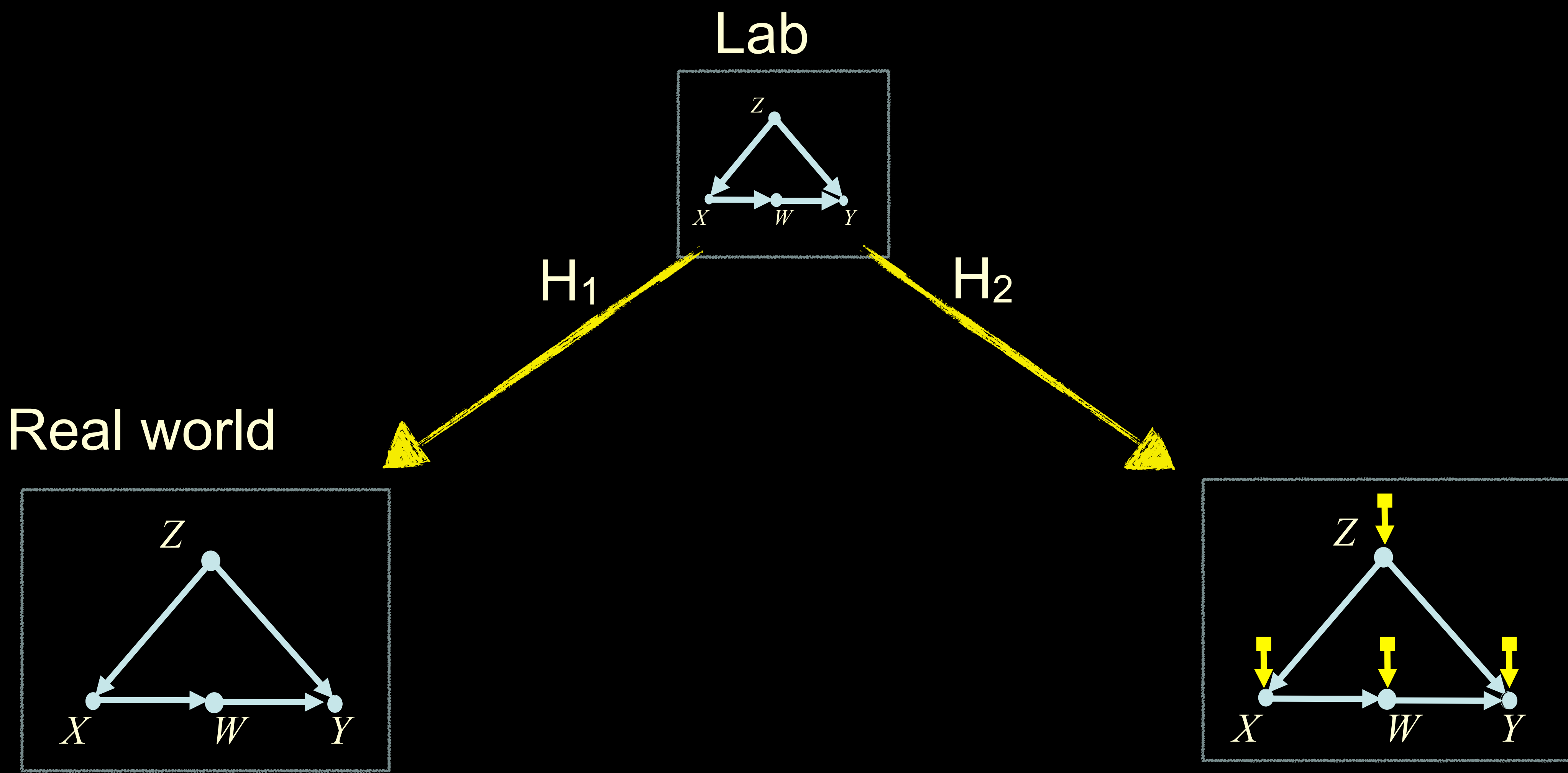
MOVING FROM THE “LAB” TO THE “REAL WORLD”...



$f_Z \neq f^*_Z$
$f_X \neq f^*_X$
$f_W \neq f^*_W$
$f_Y \neq f^*_Y$

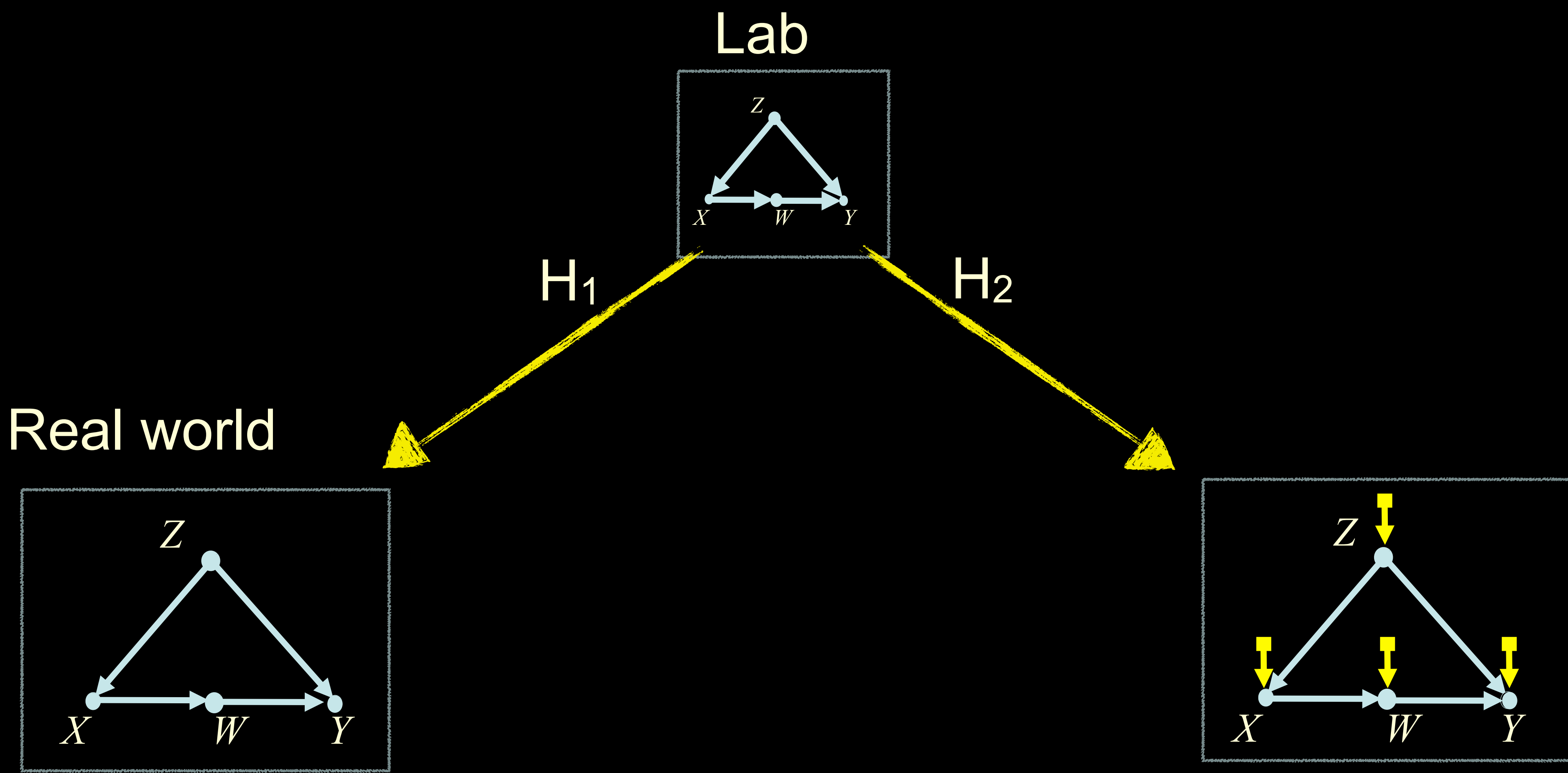
Everything is assumed to be
the same, trivially transportable!

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



Everything is assumed to be
the same, trivially transportable!

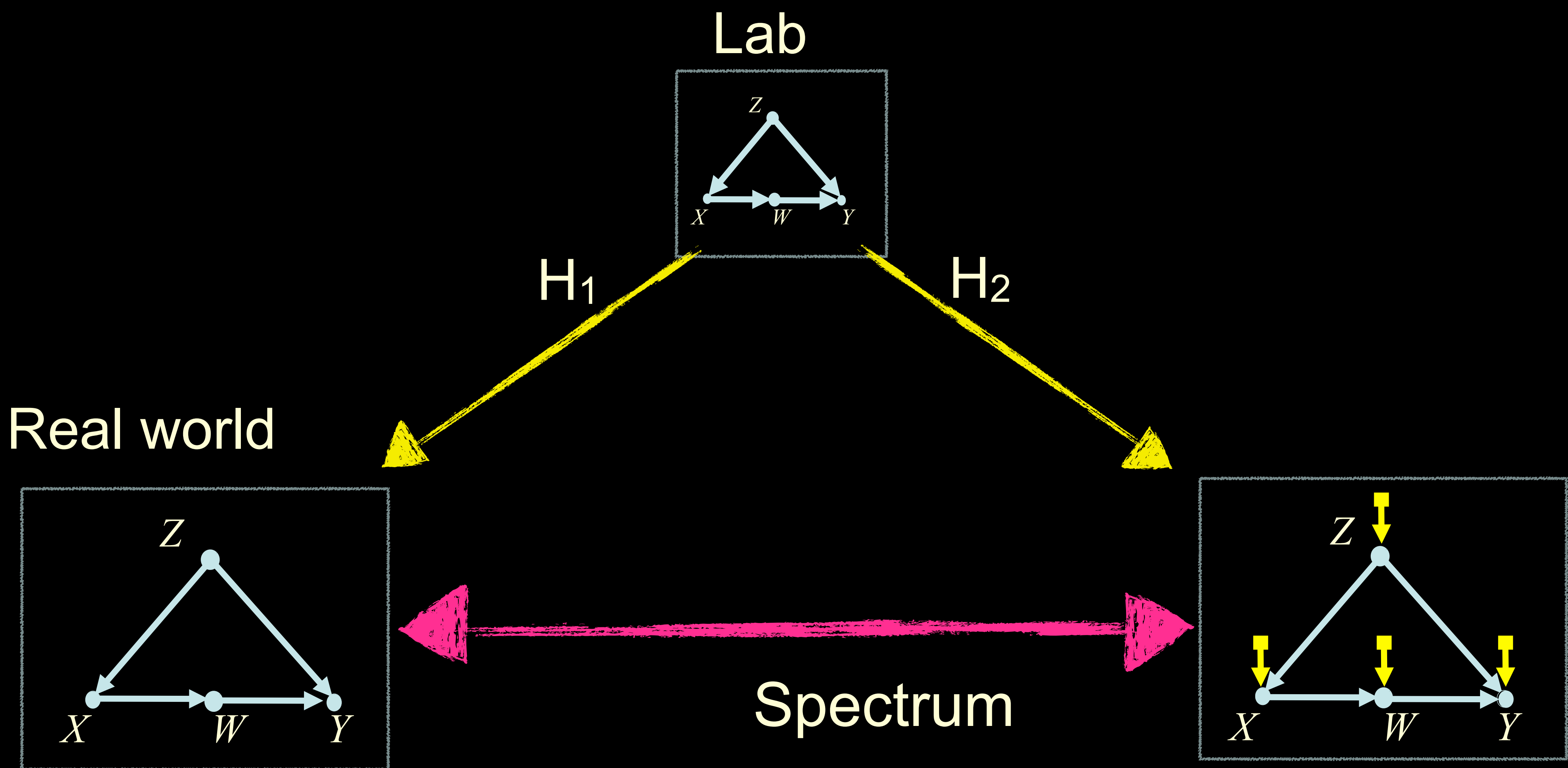
MOVING FROM THE "LAB" TO THE "REAL WORLD"...



Everything is assumed to be the same, trivially transportable!

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

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

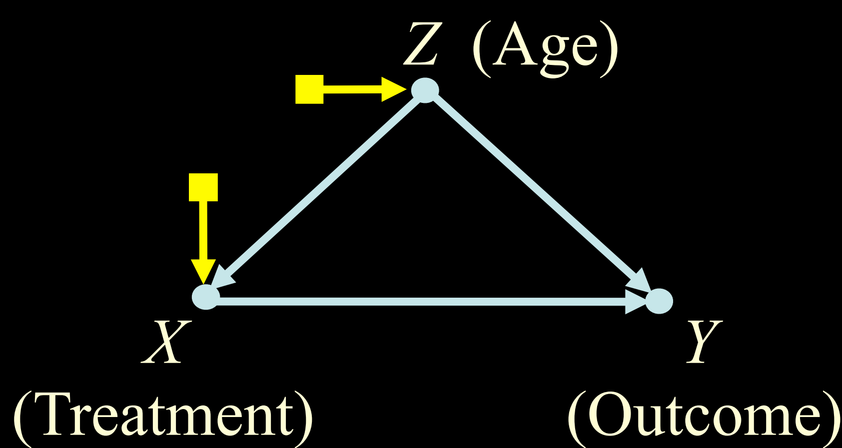


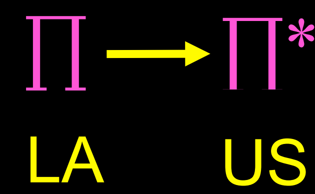
Everything is assumed to be the same, trivially transportable!

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

MOTIVATION

WHAT CAN EXPERIMENTS IN LA TELL US ABOUT US?

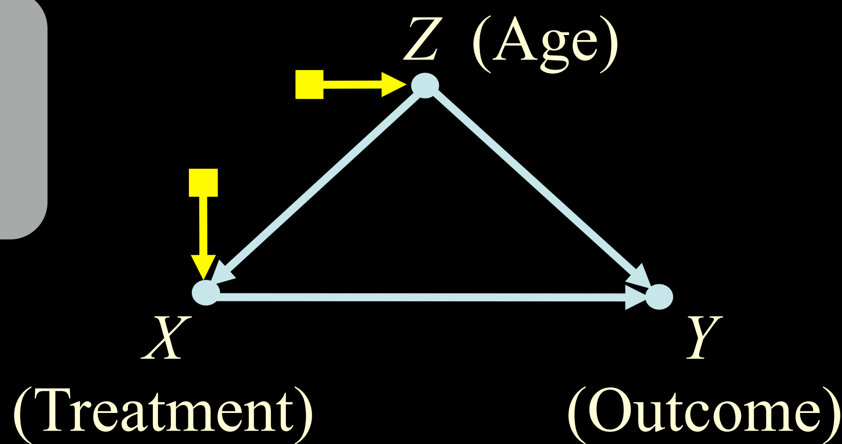




MOTIVATION

WHAT CAN EXPERIMENTS IN LA TELL US ABOUT US?

$f_z \neq f^*_z$
$f_x \neq f^*_x$
$f_y = f^*_y$

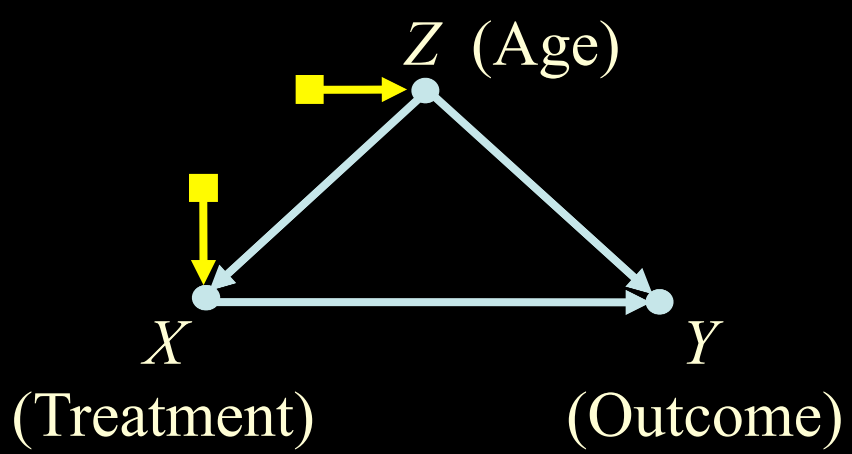


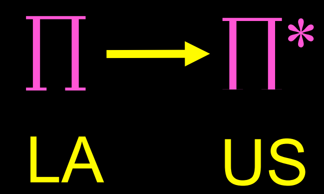
$\prod \longrightarrow \prod^*$

LA US

MOTIVATION

WHAT CAN EXPERIMENTS IN LA TELL US ABOUT US?

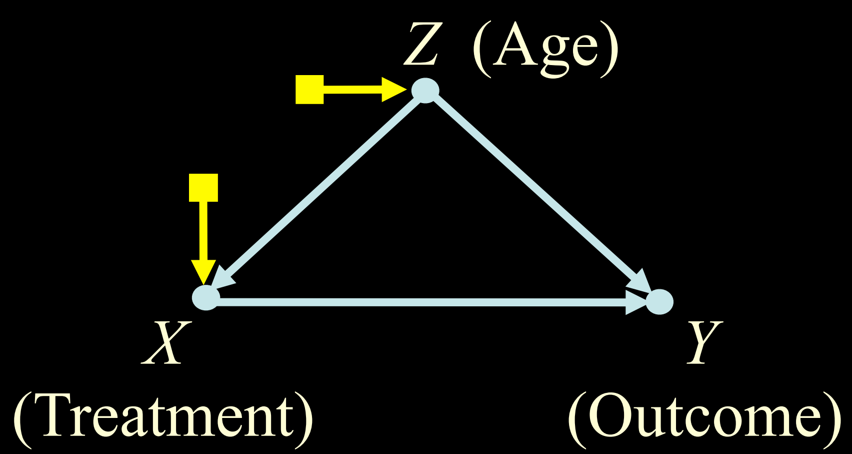




MOTIVATION

WHAT CAN EXPERIMENTS IN LA TELL US ABOUT US?

Input



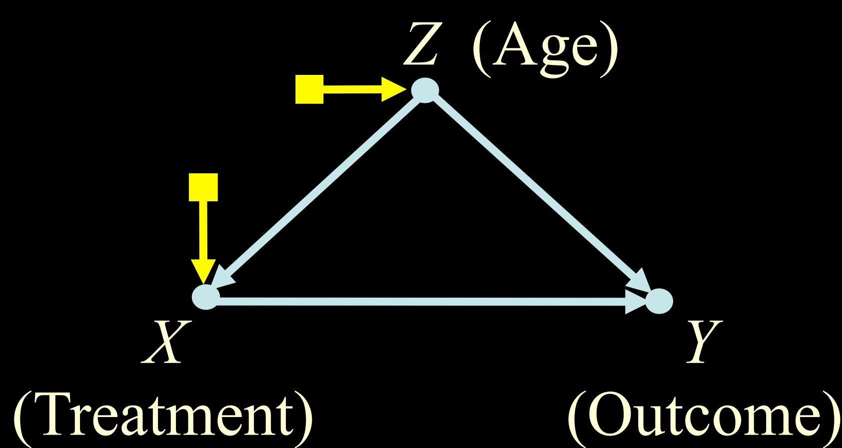
$\prod \longrightarrow \prod^*$

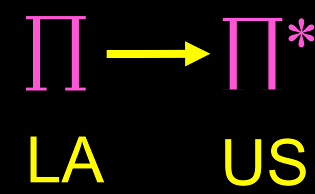
LA US

MOTIVATION

WHAT CAN EXPERIMENTS IN LA TELL US ABOUT US?

Input





Experimental study in LA

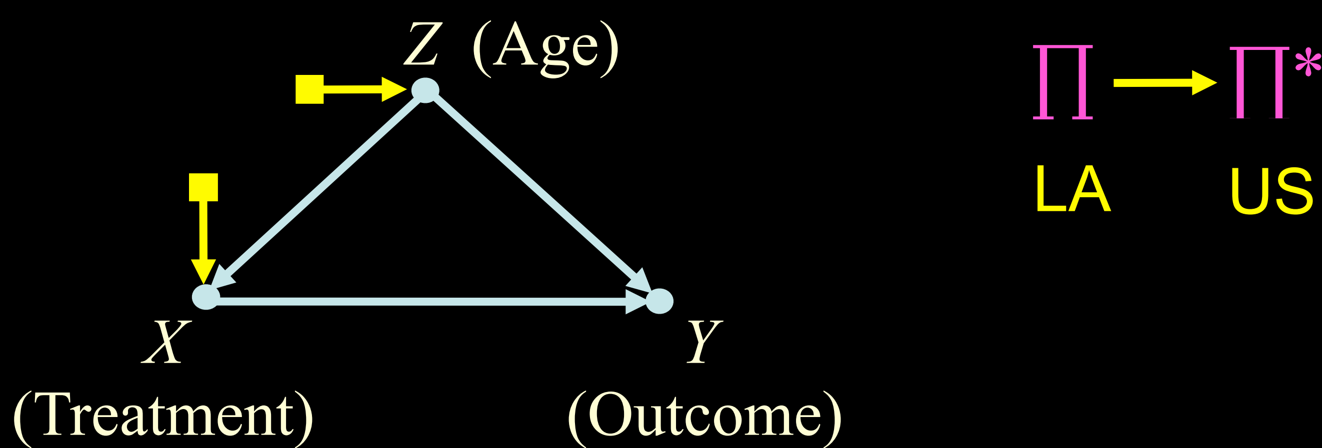
Measured: $P(x, y, z)$

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

MOTIVATION

WHAT CAN EXPERIMENTS IN LA TELL US ABOUT US?

Input



Experimental study in LA

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

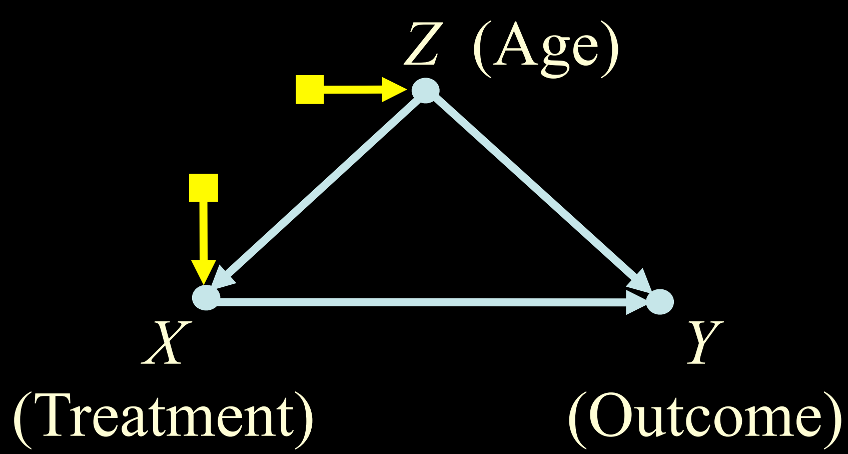
Observational study in US

Measured: $P^*(x, y, z)$
$[\ P^*(z) \neq P(z)\ ]$

MOTIVATION

WHAT CAN EXPERIMENTS IN LA TELL US ABOUT US?

Input



$\prod \longrightarrow \prod^*$

LA US

Experimental study in LA

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

Observational study in US

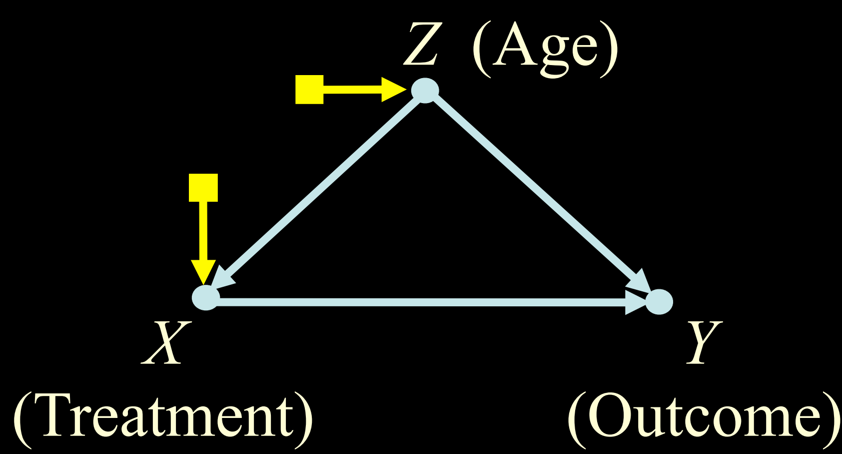
Measured: $P^*(x, y, z)$
$[\ P^*(z) \neq P(z)\ ]$

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

MOTIVATION

WHAT CAN EXPERIMENTS IN LA TELL US ABOUT US?

Input



$\prod \longrightarrow \prod^*$

LA US

Experimental study in LA

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

Observational study in US

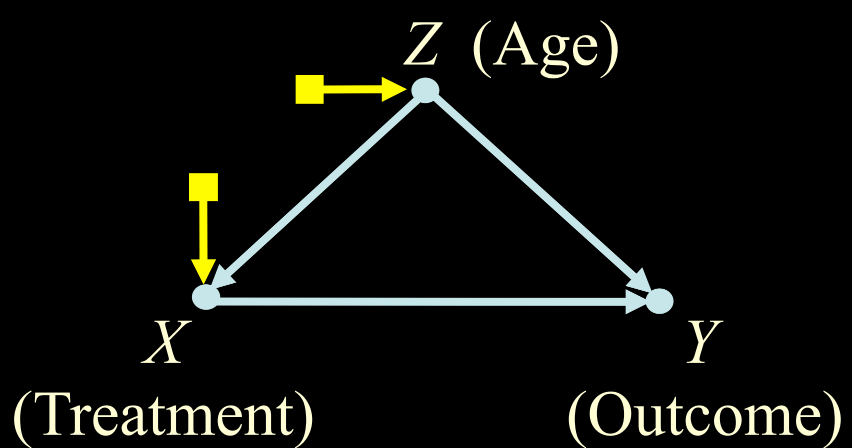
Measured: $P^*(x, y, z)$
$[\ P^*(z) \neq P(z)\ ]$

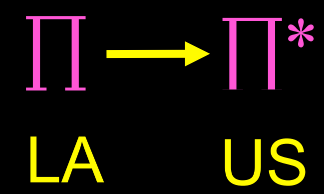
Output

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

MOTIVATION

WHAT CAN EXPERIMENTS IN LA TELL US ABOUT US?





Experimental study in LA

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

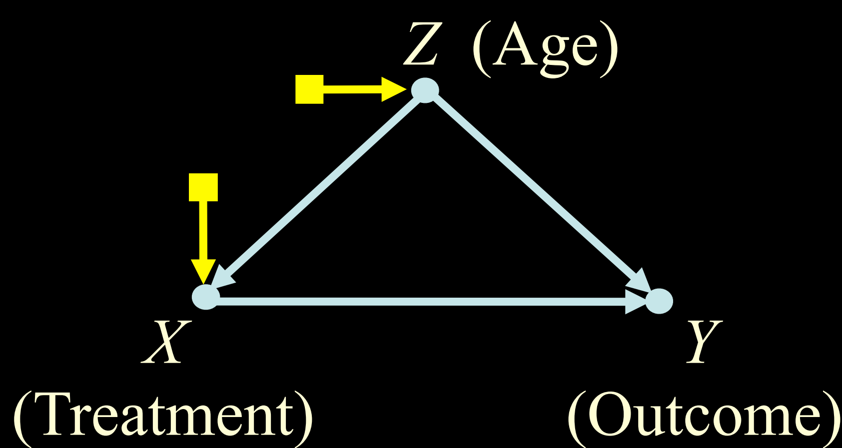
Observational study in US

Measured: $P^*(x, y, z)$
$[\ P^*(z) \neq P(z)\ ]$

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

MOTIVATION

WHAT CAN EXPERIMENTS IN LA TELL US ABOUT US?



$\prod \longrightarrow \prod^*$

LA US

Experimental study in LA

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

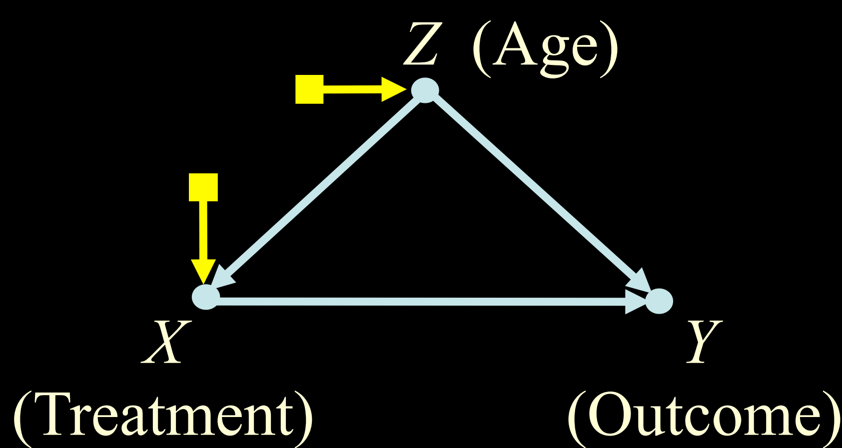
Observational study in US

Measured: $P^*(x, y, z)$
$[\, P^*(z) \neq P(z) \,]$

Needed: $Q = P^*(y \mid do(x)) = \sum_Z P(y \mid do(x), z)\, P^*(z)$

MOTIVATION

WHAT CAN EXPERIMENTS IN LA TELL US ABOUT US?



$\prod \longrightarrow \prod^*$

LA US

Experimental study in LA

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

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

Observational study in US

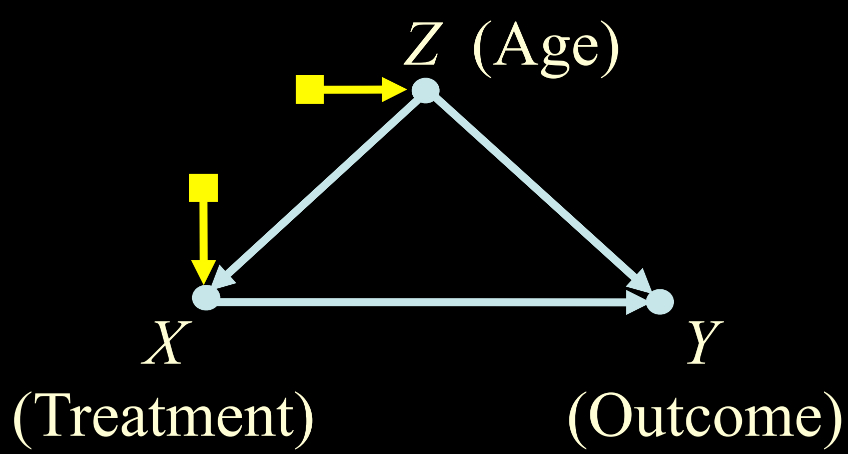
Measured: $P^*(x, y, z)$

$[\, P^*(z) \neq P(z) \,]$

Needed: $Q = P^*(y \mid do(x)) = \sum_z \underbrace{P(y \mid do(x), z)}_{\text{LA}} \underbrace{P^*(z)}_{\text{US}}$

MOTIVATION

WHAT CAN EXPERIMENTS IN LA TELL US ABOUT US?



$\Pi \longrightarrow \Pi^*$

LA US

Experimental study in LA

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

Observational study in US

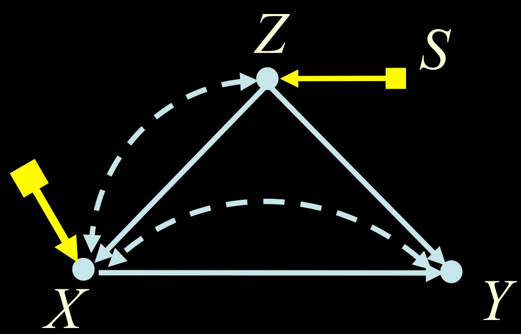
Measured: $P^*(x, y, z)$
$[\, P^*(z) \neq P(z) \,]$

Needed: $Q = P^*(y \mid do(x)) = \sum_z \underbrace{P(y \mid do(x), z)}_{\text{LA}} \; \underbrace{P^*(z)}_{\text{US}}$

Transport Formula (recalibration): $Q = F(P, P_{do}, P^*)$

TRANSPORT FORMULAS DEPEND ON THE CAUSAL STORY

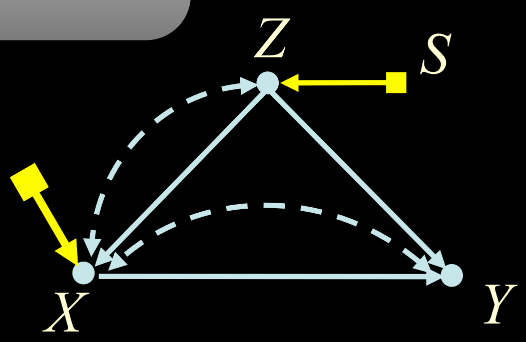
TRANSPORT FORMULAS DEPEND ON THE CAUSAL STORY



(a)

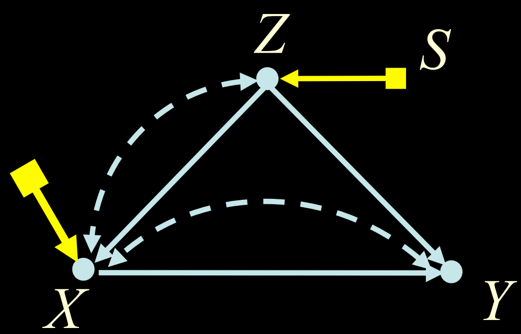
TRANSPORT FORMULAS DEPEND ON THE CAUSAL STORY

$z \leftarrow f_z(u_z, u_{xz})$
$x \leftarrow f_x(x, u_x, u_{xz}, u_{xy})$
$y \leftarrow f_y(x, z, u_y, u_{xy})$



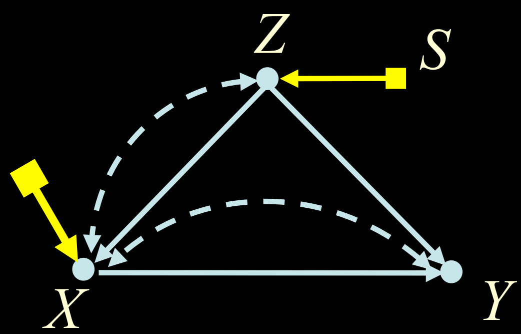
(a)

TRANSPORT FORMULAS DEPEND ON THE CAUSAL STORY



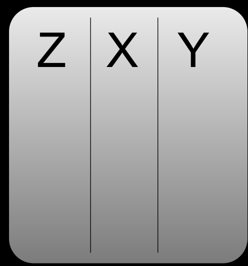
(a)

TRANSPORT FORMULAS DEPEND ON THE CAUSAL STORY

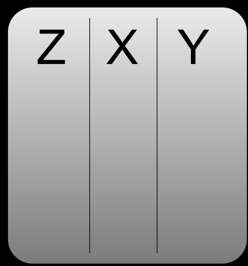


(a)

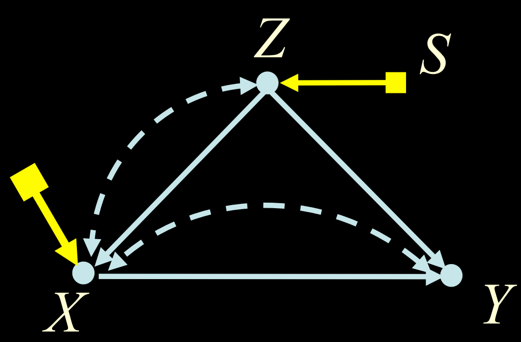
LA: P(y | do(x),z)



US: P*(x,y,z)

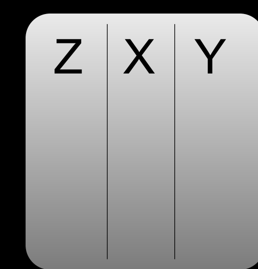


TRANSPORT FORMULAS DEPEND ON THE CAUSAL STORY

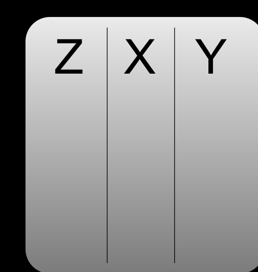


(a)

LA: P(y | do(x),z)



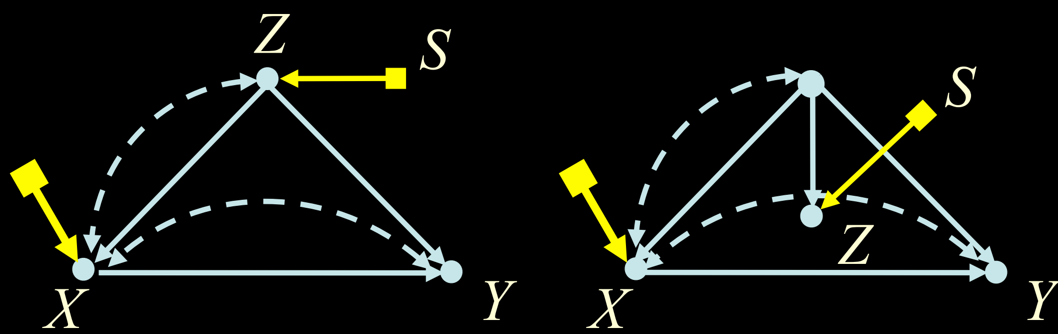
US: P*(x,y,z)



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



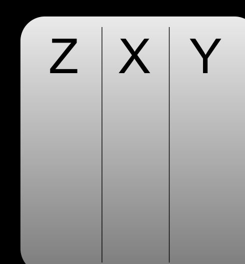
(a) (b)

a) ***Z*** represents age

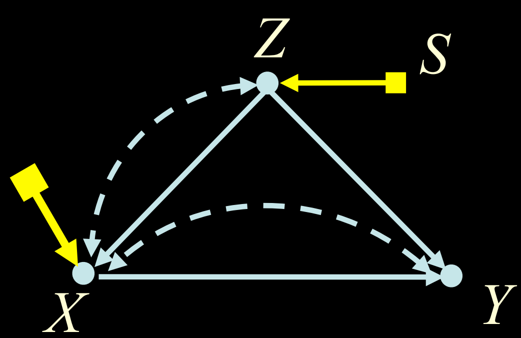
$$P^*(y \mid do(x)) = \sum_z P(y \mid do(x), z) P^*(z)$$

LA: P(y | do(x),z)

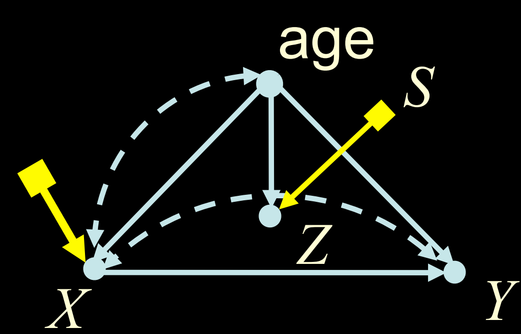
US: P*(x,y,z)



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

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

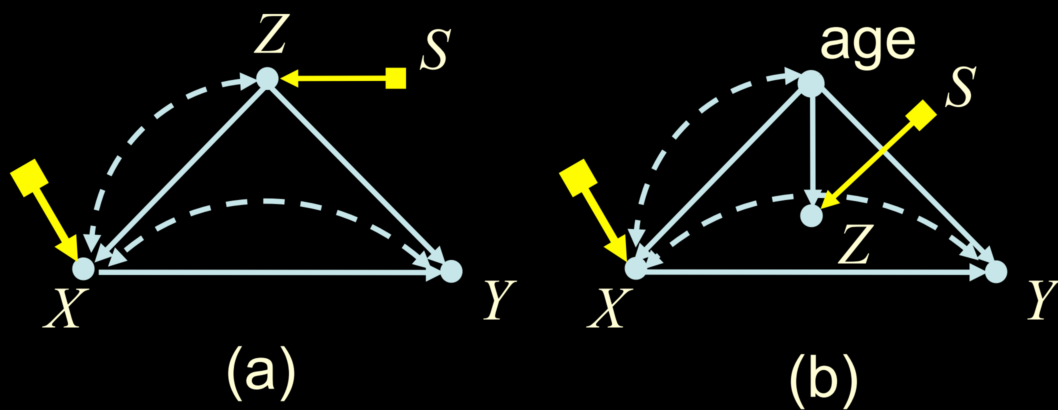
LA: P(y | do(x),z)

Z	X	Y

US: P*(x,y,z)

Z	X	Y

TRANSPORT FORMULAS DEPEND ON THE CAUSAL STORY



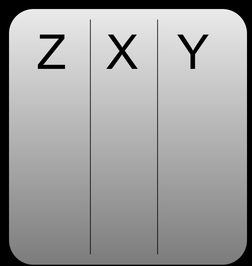
a) $\boldsymbol{Z}$ represents age

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

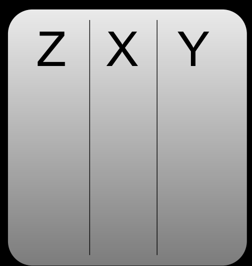
b) $\boldsymbol{Z}$ represents language skill

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

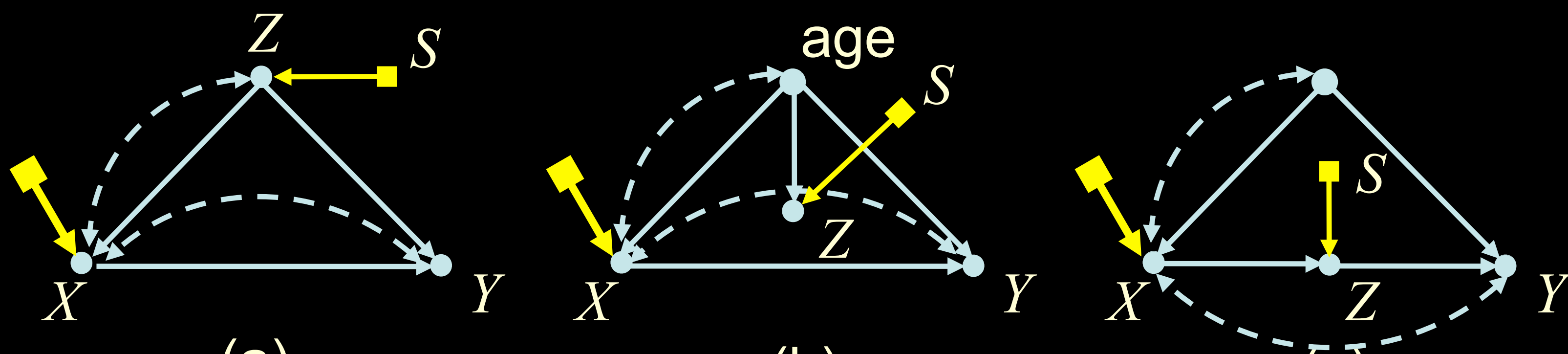
LA: P(y | do(x),z)



US: P*(x,y,z)



TRANSPORT FORMULAS DEPEND ON THE CAUSAL STORY



(a) (b) (c)

LA: P(y | do(x),z)

Z	X	Y

US: P*(x,y,z)

Z	X	Y

a) ***Z*** represents age

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

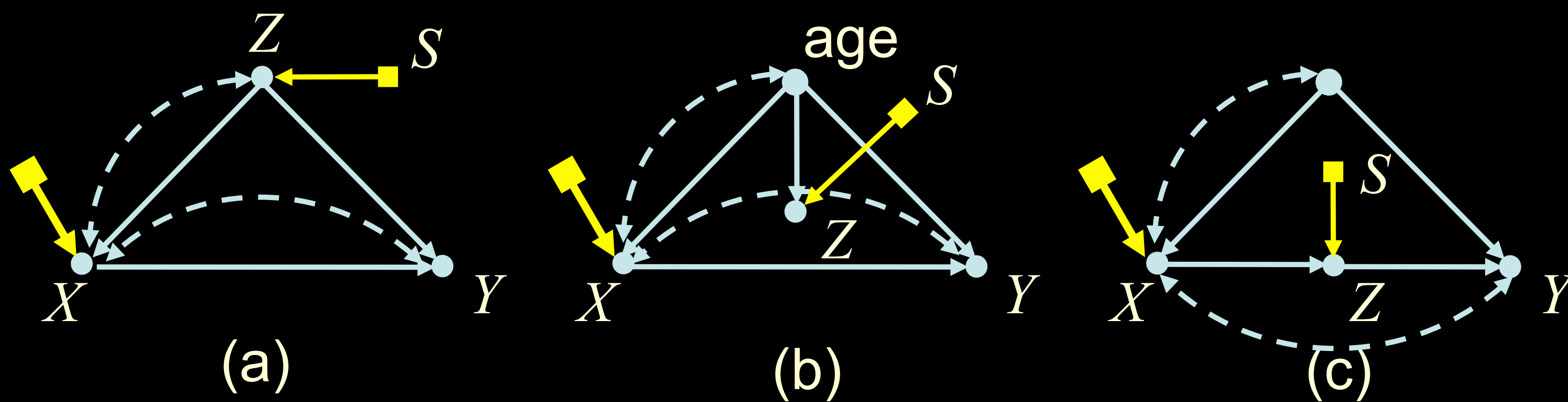
b) ***Z*** represents language skill

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

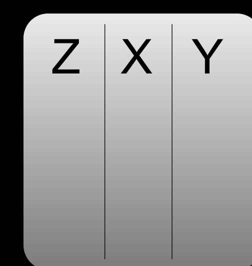
c) ***Z*** represents a bio-marker

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

TRANSPORT FORMULAS DEPEND ON THE CAUSAL STORY



LA: P(y | do(x),z)



US: P*(x,y,z)

a) Z represents age

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

b) Z represents language skill

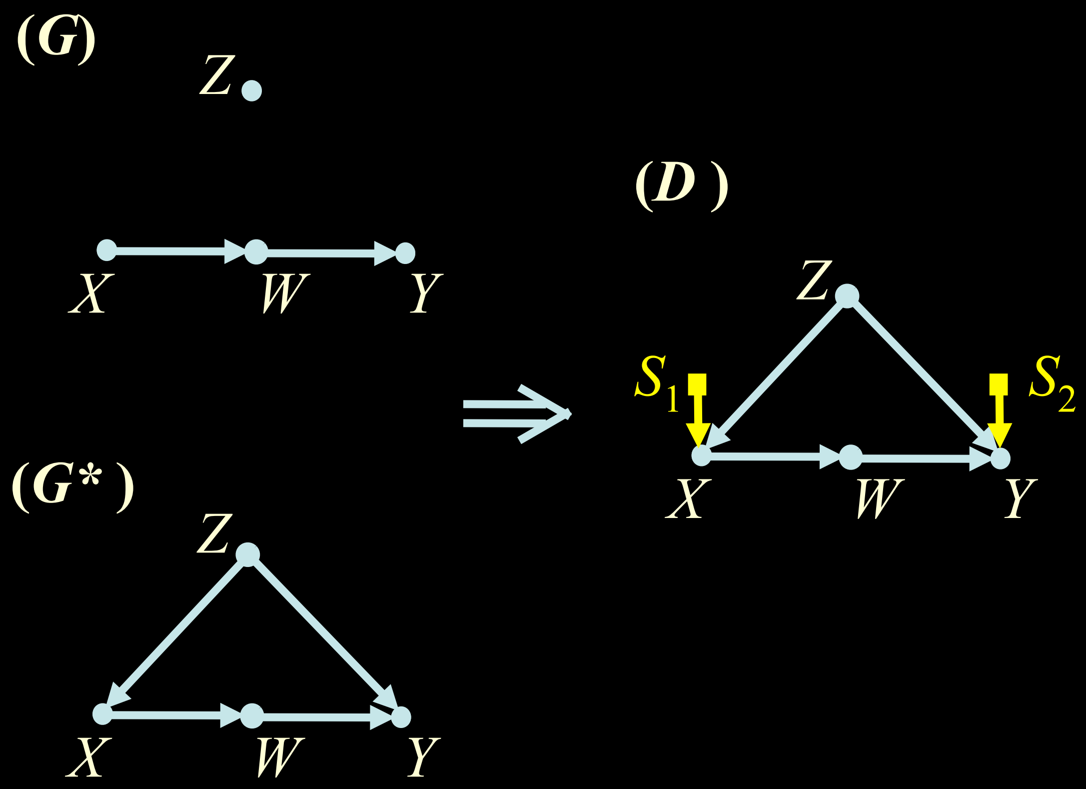
$$P^*(y \mid do(x)) = P(y \mid do(x))$$

c) Z represents a bio-marker

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

SELECTION DIAGRAMS

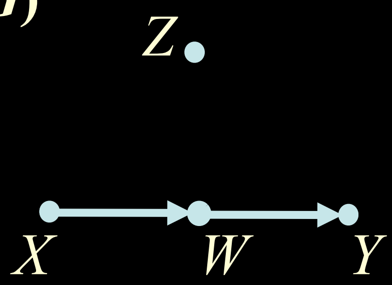
- How to encode disparities and commonalities about domains?



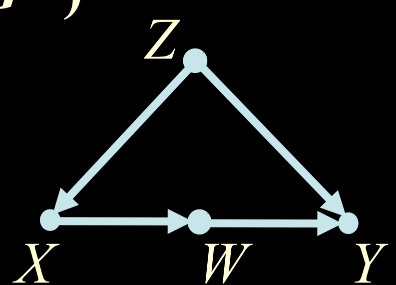
SELECTION DIAGRAMS

- How to encode disparities and commonalities about domains?

(*G*)

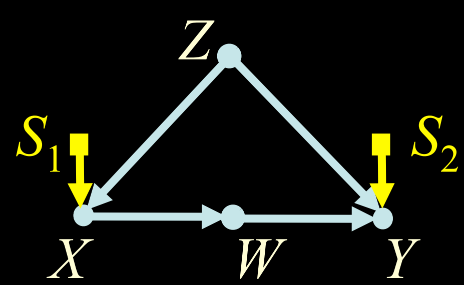


(*G)**



$\Longrightarrow$

(*D*)



$=$

$f_Z(u_Z) = f^*_Z(u_Z)$
$f_W(x, u_w) = f^*_W(x, u_w)$
$f_X(z, u_x) \neq f^*_W(x, u_x)$
$f_Y(w, z, u_y) \neq f^*_Y(w, z, u_y)$

RESULT 1: TRANSPORTABILITY REDUCED TO CALCULUS

Theorem:

A causal relation Q is transportable from Π to Π^* iff $Q(\Pi^*)$ is reducible, using the rules of causal calculus, to an expression in which S is separated from *do()*.

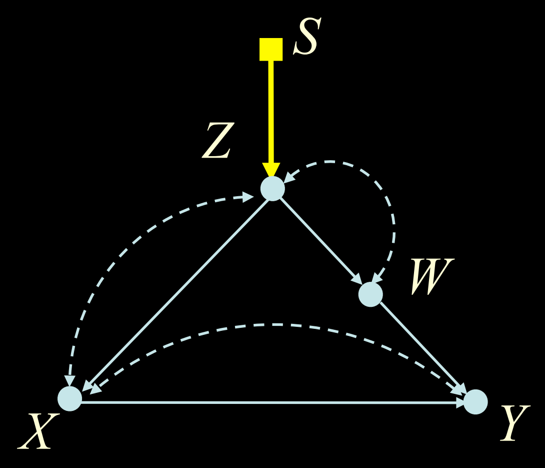
RESULT 1: TRANSPORTABILITY REDUCED TO CALCULUS

Theorem:

A causal relation $\boldsymbol{Q}$ is transportable from Π to Π^* iff $\boldsymbol{Q}(\Pi^*)$ is reducible, using the rules of causal calculus, to an expression in which $\boldsymbol{S}$ is separated from ***do()***.

Ex:

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

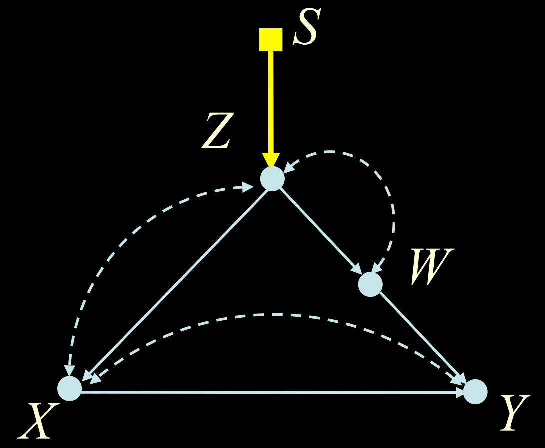


RESULT 1: TRANSPORTABILITY REDUCED TO CALCULUS

Theorem:
A causal relation Q is transportable from Π to Π^* iff $Q(\Pi^*)$ is reducible, using the rules of causal calculus, to an expression in which S is separated from *do()*.

Ex: query

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



RESULT 1: TRANSPORTABILITY REDUCED TO CALCULUS

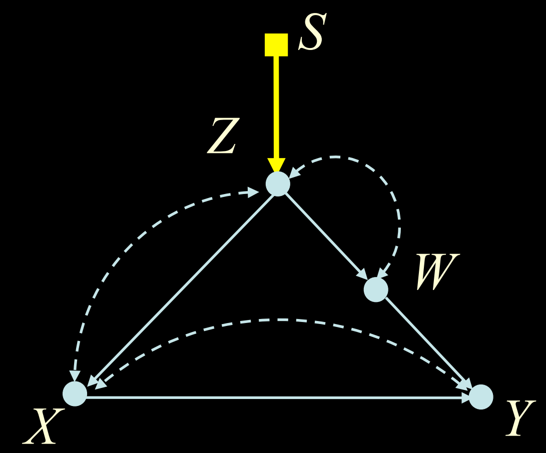
Theorem:

A causal relation Q is transportable from Π to Π^* iff $Q(\Pi^*)$ is reducible, using the rules of causal calculus, to an expression in which S is separated from *do()*.

Ex: query

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

data



RESULT 1: TRANSPORTABILITY REDUCED TO CALCULUS

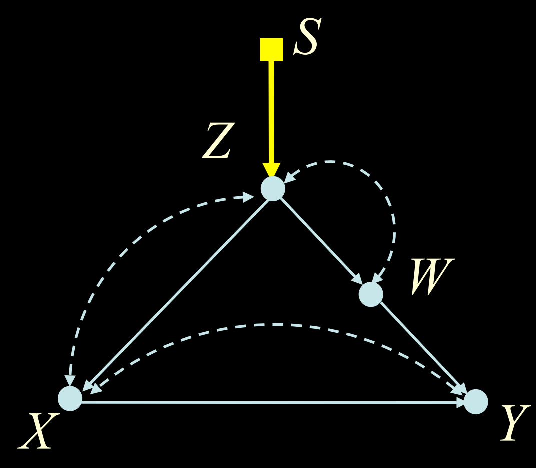
Theorem:

A causal relation Q is transportable from Π to Π^* iff $Q(\Pi^*)$ is reducible, using the rules of causal calculus, to an expression in which S is separated from $do()$.

Ex: query

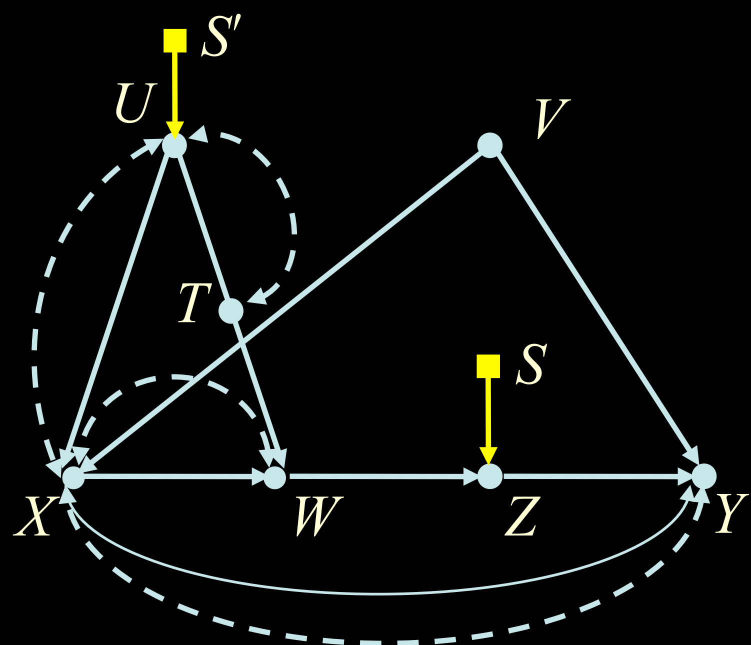
$$
\begin{aligned}
Q = P^*(y \mid do(x)) &= P(y \mid do(x), s) \\
&= \textstyle\sum_w P(y \mid do(x), s, w)\, P(w \mid do(x), s) \\
&= \textstyle\sum_w P(y \mid do(x), w)\, P(w \mid do(x), s) \\
&= \textstyle\sum_w P(y \mid do(x), w)\, P(w \mid s) \\
&= \textstyle\sum_w P(y \mid do(x), w)\, P^*(w)
\end{aligned}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

data



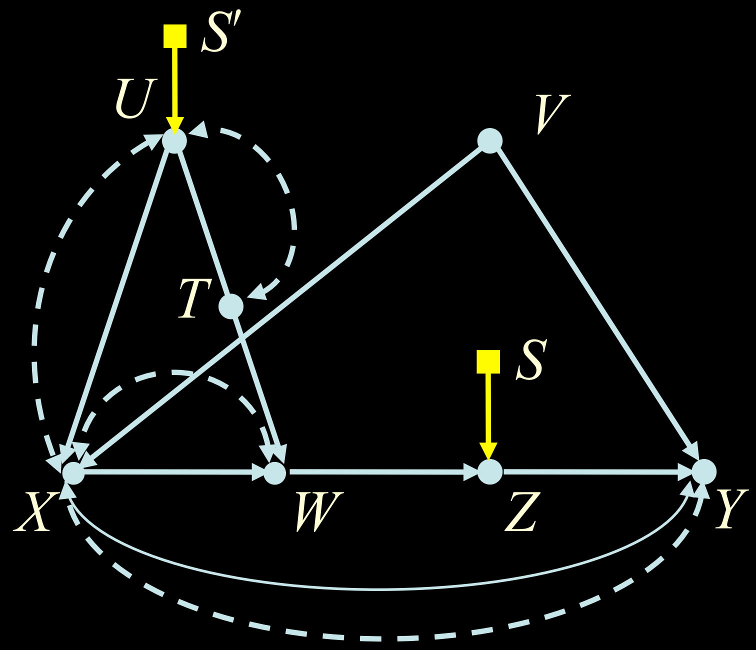
* Deterministic sep.: (Dawid, 1979, 2002; Constantinou & Dawid, 2016)

RESULT 2: ALGORITHM TO DETERMINE IF AN EFFECT IS TRANSPORTABLE

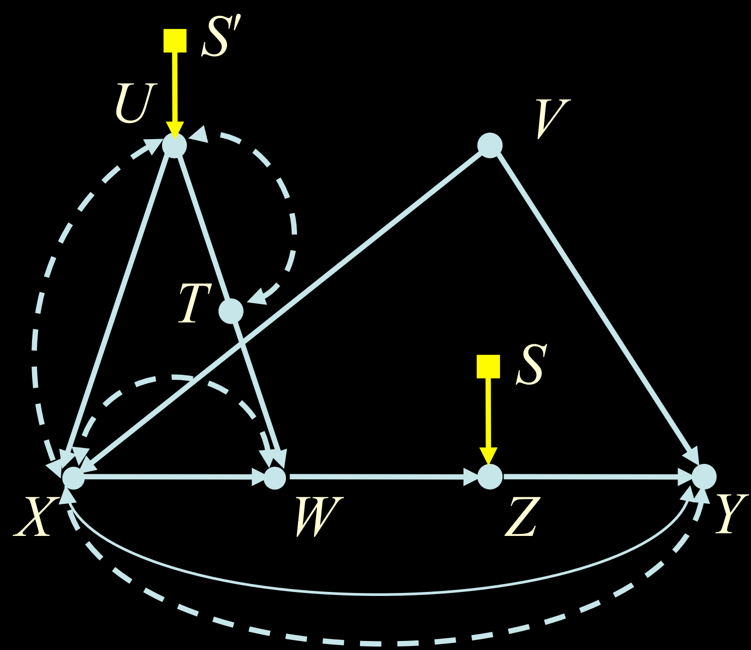


RESULT 2: ALGORITHM TO DETERMINE IF AN EFFECT IS TRANSPORTABLE

INPUT: Annotated Causal Graph



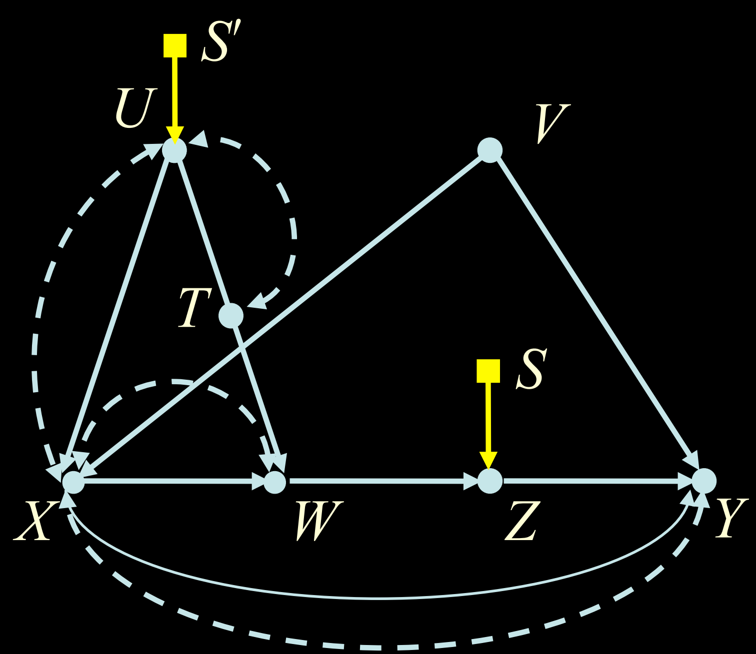
RESULT 2: ALGORITHM TO DETERMINE IF AN EFFECT IS TRANSPORTABLE



INPUT: Annotated Causal Graph

S ⟼ Factors creating differences

RESULT 2: ALGORITHM TO DETERMINE IF AN EFFECT IS TRANSPORTABLE

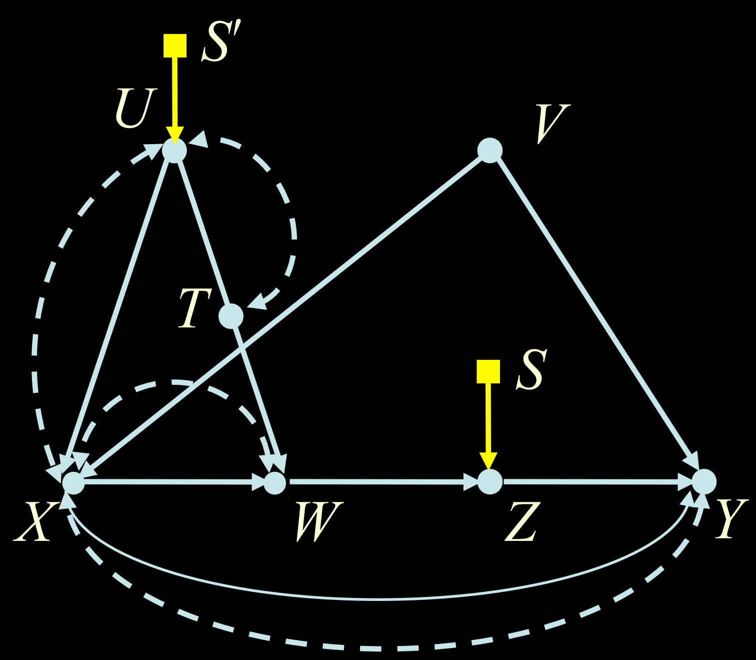


INPUT: Annotated Causal Graph

S ⟼ Factors creating differences

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

RESULT 2: ALGORITHM TO DETERMINE IF AN EFFECT IS TRANSPORTABLE



INPUT: Annotated Causal Graph

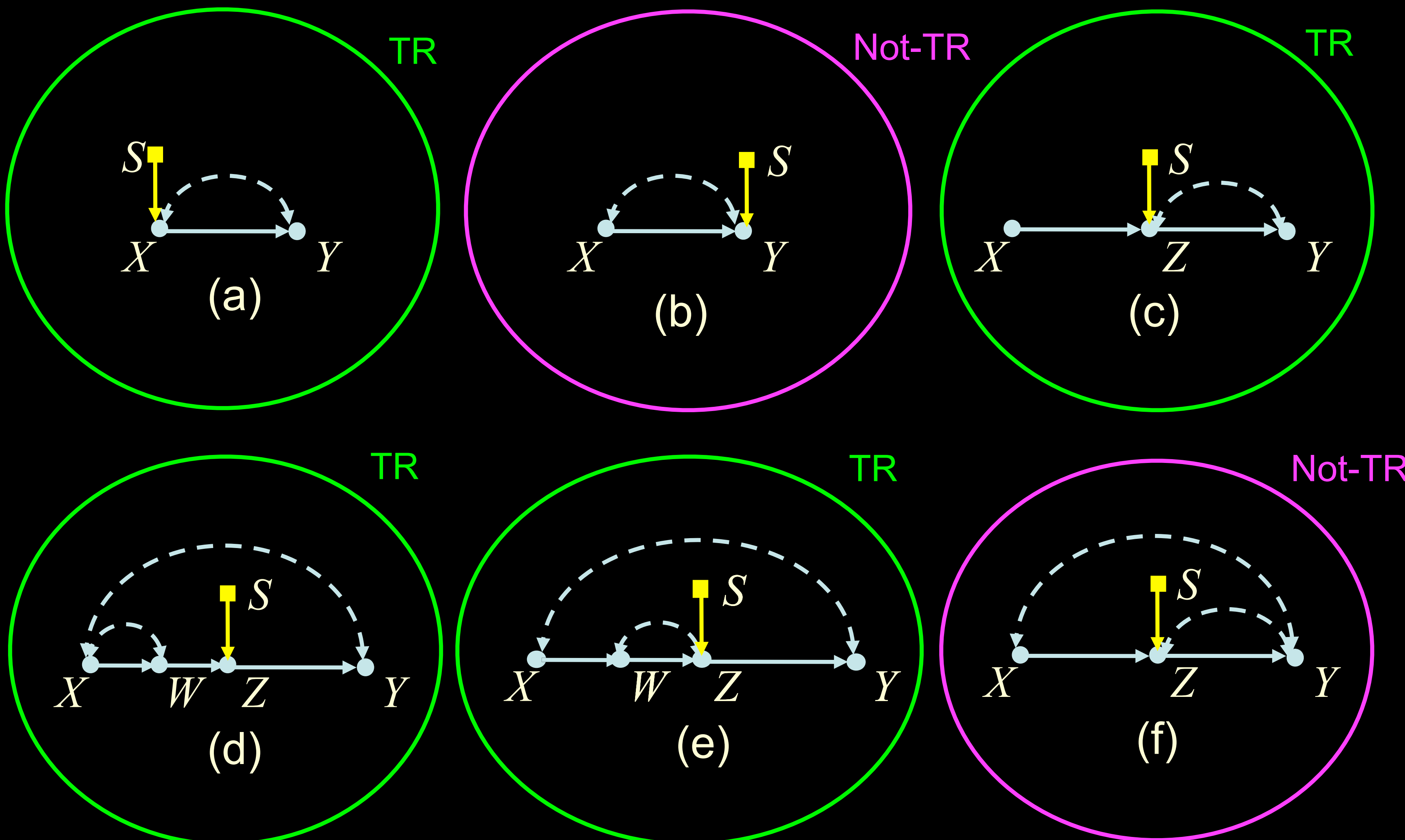
S ⟼ Factors creating differences

OUTPUT:

1. Transportable or not?
2. Yes = Transport formula
    1. Measurements to be taken in the experimental domain
    2. Measurements to be taken in the target domain
3. No = Not possible

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

RESULT 3: GRAPHICAL CONDITION FOR TRANSPORTABILITY

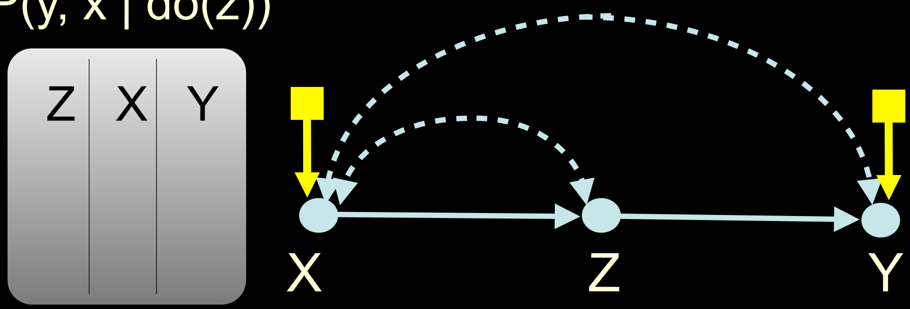


THE CHALLENGE OF GENERAL TRANSPORTABILITY

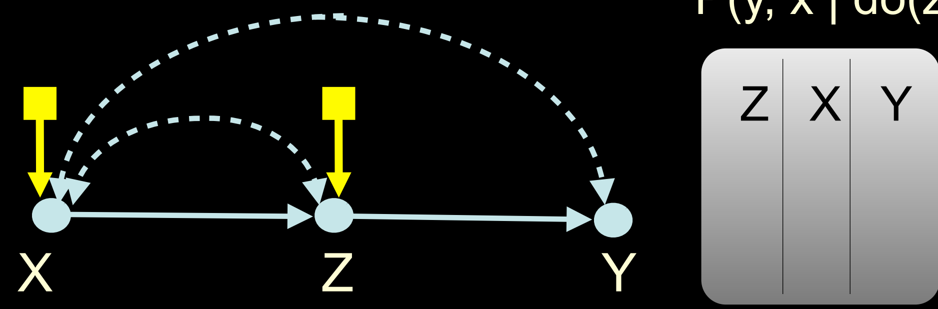
Assume we want to transport $Q = P^*(y \mid do(x))$ from $\{ \prod^a, \prod^b \}$ with experimental information $\{X, Z\}$ in $\prod^a$, $\{X, Z\}$ in $\prod^b$.

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

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



$\prod^a$ (=LA)



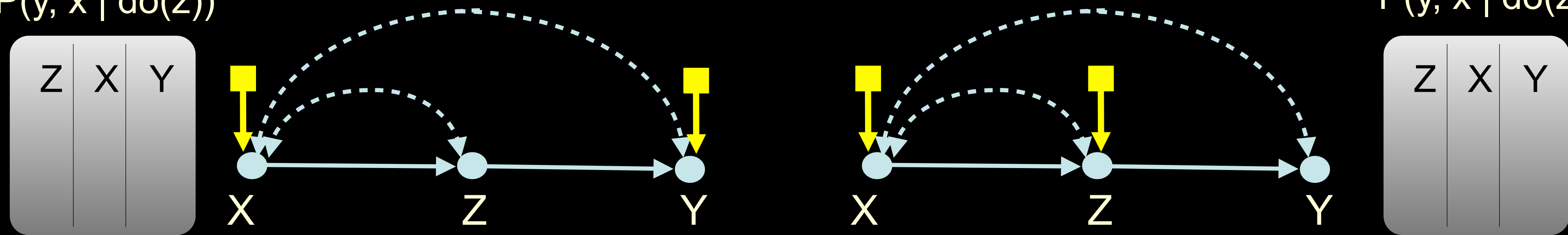
$\prod^b$ (=NYC)

THE CHALLENGE OF GENERAL TRANSPORTABILITY

Assume we want to transport $Q = P^*(y \mid do(x))$ from $\{ \prod^a, \prod^b \}$ with experimental information $\{X, Z\}$ in $\prod^a$, $\{X, Z\}$ in $\prod^b$.

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

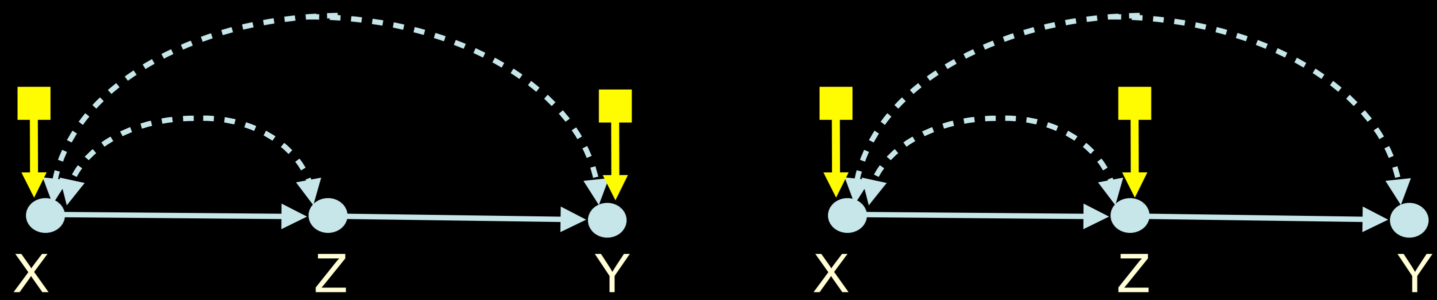
:NY
$P(y, z \mid do(x))$
$P(y, x \mid do(z))$



$\prod^a$ (=LA) $\prod^b$ (=NYC)

If we apply the result from limited transportability (1D), Q is (provably) not transportable from $\prod^a$ and $\prod^b$ separately.

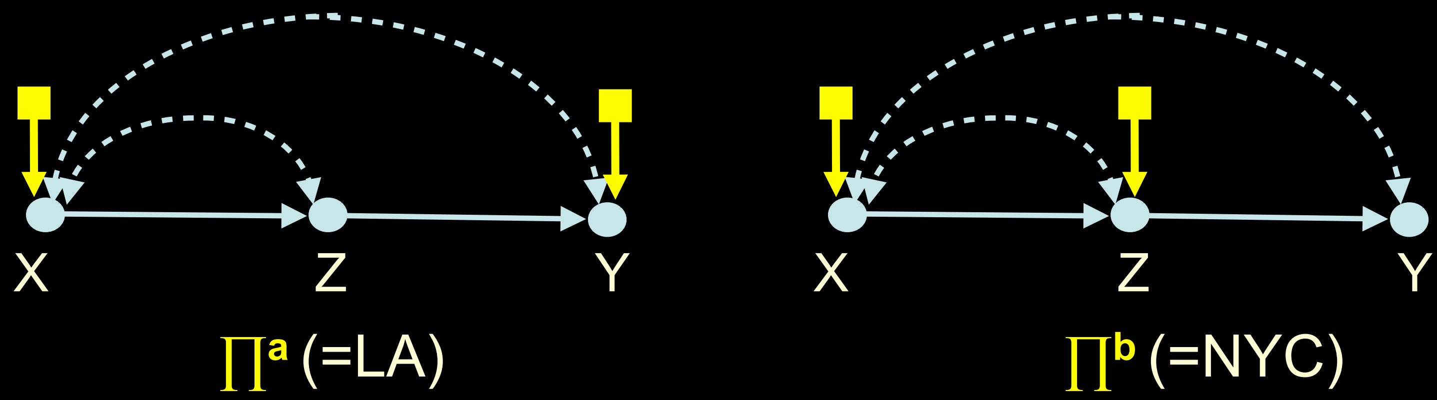
DERIVATION OF GENERAL TRANSPORTABILITY



$\prod^{a}$ (=LA) $\qquad$ $\prod^{b}$ (=NYC)

$Q = P^{*}(y \mid do(x)) = P(y \mid do(x), s^{a}, s^{b})$	Meta-TR Def.
$= \sum_{z} P(y \mid do(x), z, s^{a}, s^{b})\, P(z \mid do(x), s^{a}, s^{b})$	Probability Axioms
$= \sum_{z} P(y \mid do(x), do(z), s^{a}, s^{b})\, P(z \mid do(x), s^{a},$	Rule 2
$= \sum_{z} P(y \mid do(z), s^{a}, s^{b})\, P(z \mid do(x), s^{a}, s^{b})$	Rule 3
$= \sum_{z} P(y \mid do(z), s^{a})\, P(z \mid do(x), s^{a}, s^{b})$	Rule 1* in D_b
$= \sum_{z} P(y \mid do(z), s^{a})\, P(z \mid do(x), s^{b})$	Rule 1* in D_a
$= \sum_{z} P^{(b)}(y \mid do(z))\, P^{(a)}(z \mid do(x))$	Meta-TR Def.

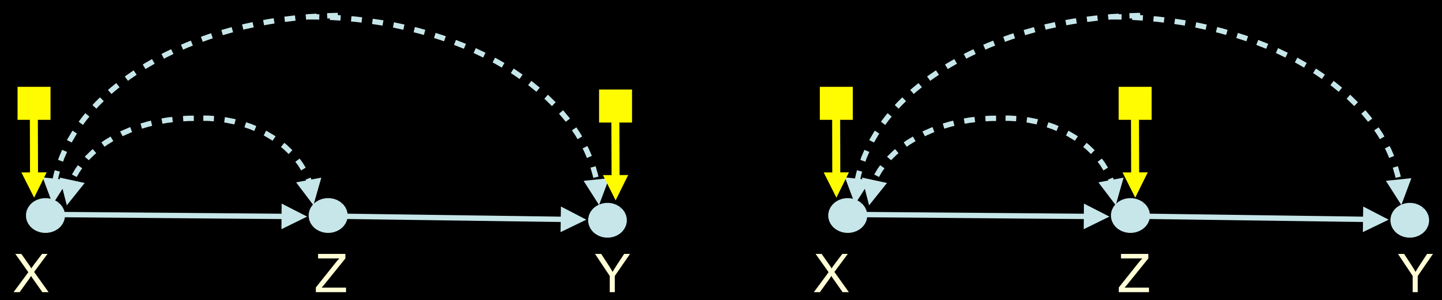
DERIVATION OF GENERAL TRANSPORTABILITY



query

$Q = P^*(y \mid do(x)) = P(y \mid do(x), s^a, s^b)$	Meta-TR Def.
$= \sum_z P(y \mid do(x), z, s^a, s^b)\, P(z \mid do(x), s^a, s^b)$	Probability Axioms
$= \sum_z P(y \mid do(x), do(z), s^a, s^b)\, P(z \mid do(x), s^a,$	Rule 2
$= \sum_z P(y \mid do(z), s^a, s^b)\, P(z \mid do(x), s^a, s^b)$	Rule 3
$= \sum_z P(y \mid do(z), s^a)\, P(z \mid do(x), s^a, s^b)$	Rule 1* in D_b
$= \sum_z P(y \mid do(z), s^a)\, P(z \mid do(x), s^b)$	Rule 1* in D_a
$= \sum_z P^{(b)}(y \mid do(z))\, P^{(a)}(z \mid do(x))$	Meta-TR Def.

DERIVATION OF GENERAL TRANSPORTABILITY



$\prod^{a}$ (=LA) $\qquad$ $\prod^{b}$ (=NYC)

query

$Q = P^*(y \mid do(x)) = P(y \mid do(x), s^a, s^b)$	Meta-TR Def.
$= \sum_z P(y \mid do(x), z, s^a, s^b)\, P(z \mid do(x), s^a, s^b)$	Probability Axioms
$= \sum_z P(y \mid do(x), do(z), s^a, s^b)\, P(z \mid do(x), s^a,$	Rule 2
$= \sum_z P(y \mid do(z), s^a, s^b)\, P(z \mid do(x), s^a, s^b)$	Rule 3
$= \sum_z P(y \mid do(z), s^a)\, P(z \mid do(x), s^a, s^b)$	Rule 1* in D_b
$= \sum_z P(y \mid do(z), s^a)\, P(z \mid do(x), s^b)$	Rule 1* in D_a
$= \sum_z P^{(b)}(y \mid do(z))\, P^{(a)}(z \mid do(x))$	Meta-TR Def.

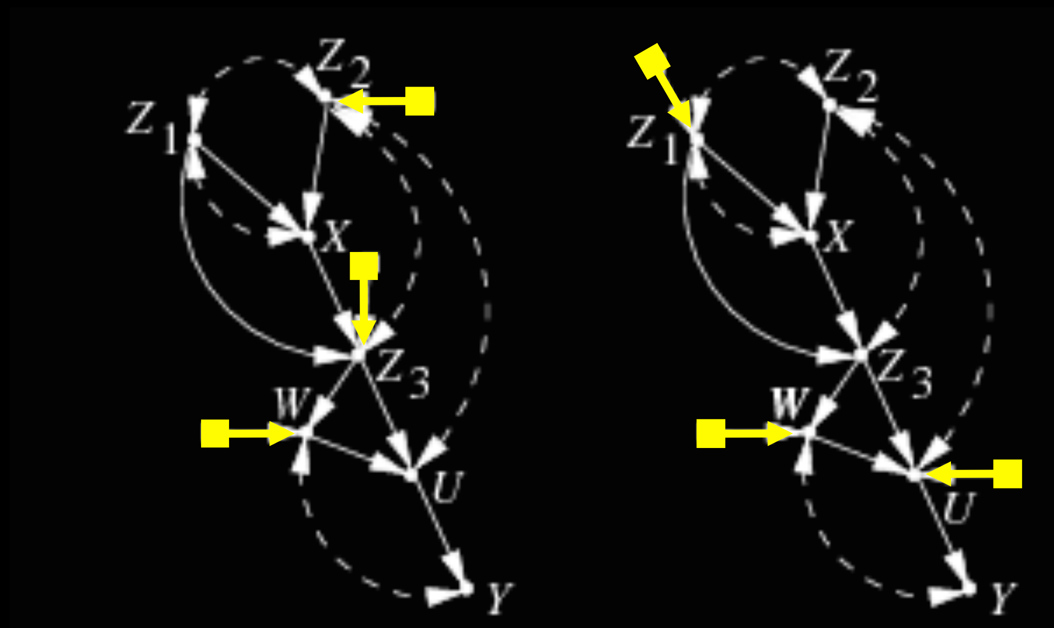
data $\qquad$ NYC $\qquad$ LA

A MORE CHALLENGING INSTANCE …

Assume we want to transport $Q = P^*(y \mid do(x))$ from $\{ \prod^a, \prod^b \}$ by intervening over $\{Z_1\}$ in $\prod^*$, $\{Z_2\}$ in $\prod^a$, $\{Z_1\}$ in $\prod^b$.

A MORE CHALLENGING INSTANCE …

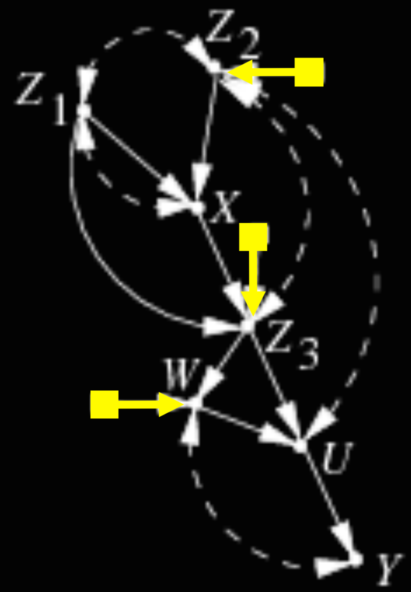
Assume we want to transport $Q = P^*(y \mid do(x))$ from $\{ \prod^a, \prod^b \}$ by intervening over $\{Z_1\}$ in $\prod^*$, $\{Z_2\}$ in $\prod^a$, $\{Z_1\}$ in $\prod^b$.



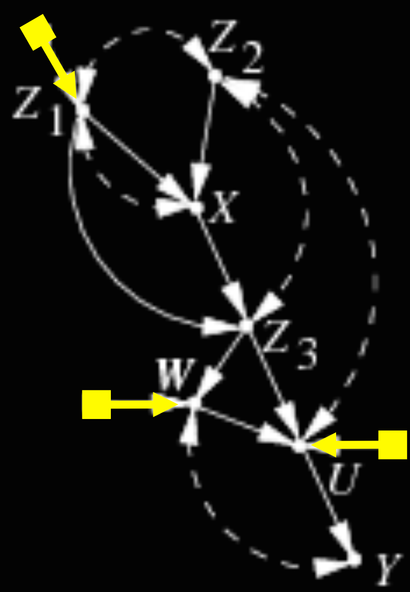
$\prod^a$ (=LA) $\qquad$ $\prod^b$ (=NYC)

A MORE CHALLENGING INSTANCE …

Assume we want to transport $Q = P^*(y \mid do(x))$ from $\{ \prod^a, \prod^b \}$ by intervening over $\{Z_1\}$ in $\prod^*$, $\{Z_2\}$ in $\prod^a$, $\{Z_1\}$ in $\prod^b$.



$\prod^a$ (=LA)



$\prod^b$ (=NYC)

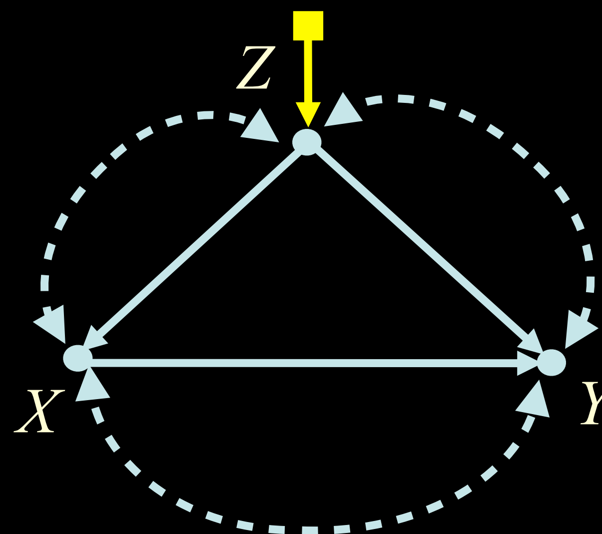
$$P^*_x(y) = \sum_{z_1,z_3,w,u} P^*(z_1) \Big(w_1^{(1)} \sum_{Z_{2'}} P^*_{z_1}(z_3|x, Z_2')P^*_{z_1}(Z_2') + w_2^{(1)} \sum_{Z_{2'}} P^{(b)}_{z_1}(z_3|x,Z_2')\, P^{(b)}_{z_1}(Z_2') \Big)$$

$$\Big(w_1^{(2)} \Big(\sum_{Z_{2'}} P^*_{z_1}(u|w, z_3, x, Z_2')\, P^*_{z_1}(z_3|x,Z_2')\, P^*_{z_1}(Z_2')\Big) \Big/ \Big(\sum_{Z_{2''}} P^*_{z_1}(z_3|x,Z_2'')\, P^*_{z_1}(Z_2'') \Big)$$

$$+ w_2^{(2)} P^{(a)}_{z_2}(u|w, z_3, x, Z_2') \Big)\, P^*(w|x,z_1,z_2,z_3)\, P^*(y|x,z_1,z_2,z_3,w, u)$$

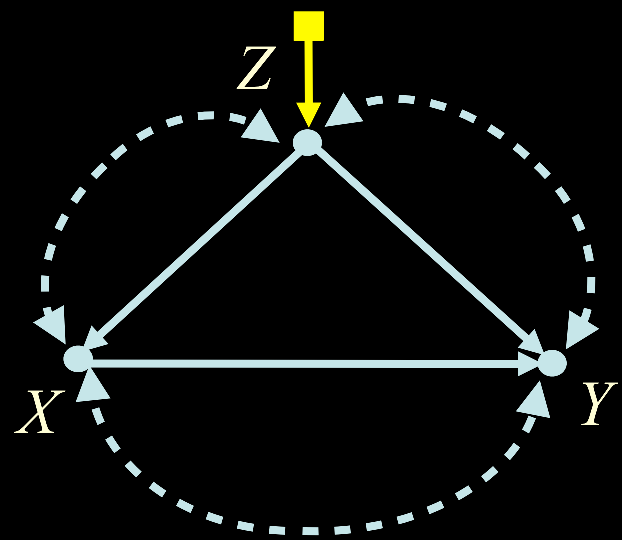
IS THE GOLD STANDARD SO GOLD?
(GENERALIZABILITY FROM EXPERIMENTS)

Before randomization

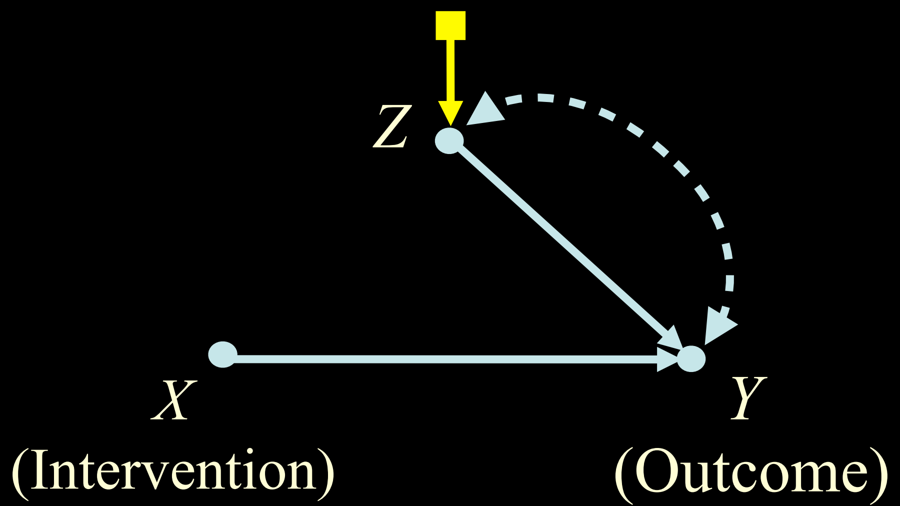


IS THE GOLD STANDARD SO GOLD?
(GENERALIZABILITY FROM EXPERIMENTS)

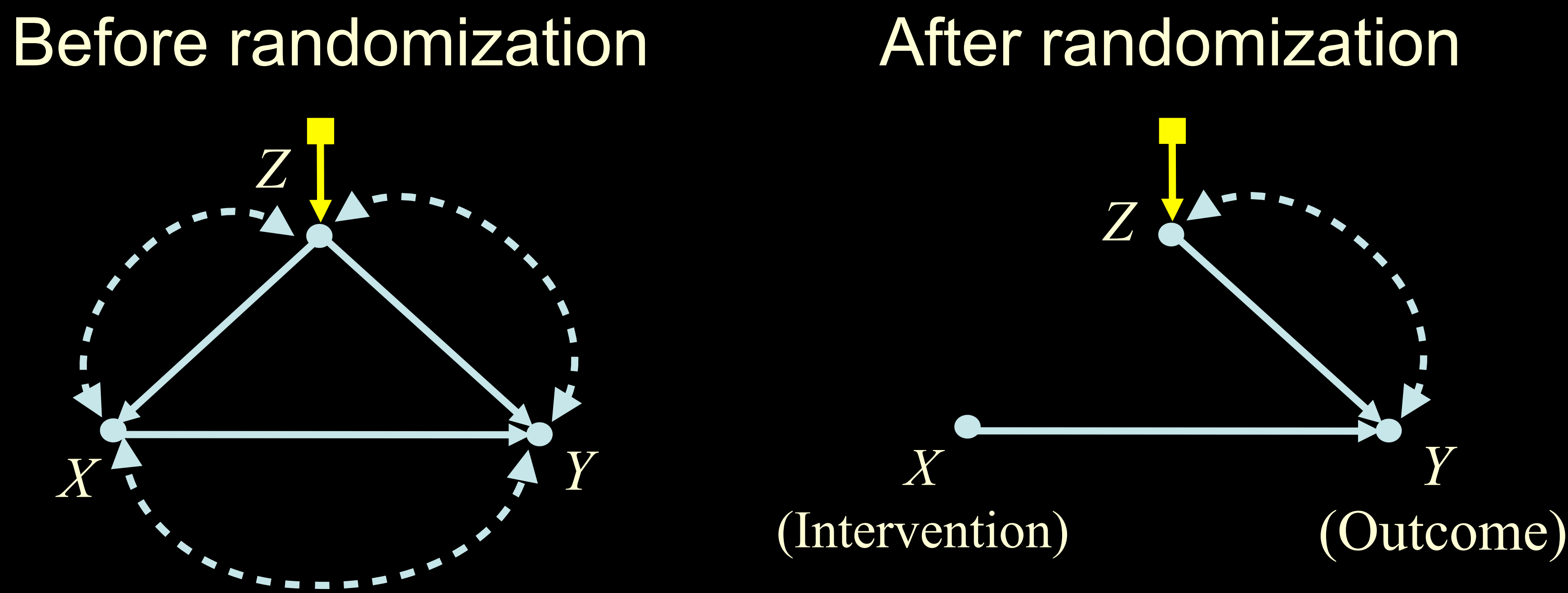
Before randomization



After randomization



IS THE GOLD STANDARD SO GOLD?
(GENERALIZABILITY FROM EXPERIMENTS)



The gold standard is not absolute, and strong assumptions are required to validate claims from trials.

SUMMARY D=2: POLICY GENERALIZATION

SUMMARY D=2: POLICY GENERALIZATION

- Nonparametric transportability can be determined provided that the problem instance is encoded in selection diagrams.

SUMMARY D=2: POLICY GENERALIZATION

- Nonparametric transportability can be determined provided that the problem instance is encoded in selection diagrams.
- When transportability is feasible, the transport formula can be derived in polynomial time.

SUMMARY D=2: POLICY GENERALIZATION

- Nonparametric transportability can be determined provided that the problem instance is encoded in selection diagrams.
- When transportability is feasible, the transport formula can be derived in polynomial time.
- The causal calculus and the algorithm are complete.

SUMMARY D=2: POLICY GENERALIZATION

- Nonparametric transportability can be determined provided that the problem instance is encoded in selection diagrams.
- When transportability is feasible, the transport formula can be derived in polynomial time.
- The causal calculus and the algorithm are complete.
- These results solve long-standing problems in the empirical sciences including external validity, meta-analysis, and heterogeneity.

SUMMARY D=2: POLICY GENERALIZATION

- Nonparametric transportability can be determined provided that the problem instance is encoded in selection diagrams.
- When transportability is feasible, the transport formula can be derived in polynomial time.
- The causal calculus and the algorithm are complete.
- These results solve long-standing problems in the empirical sciences including external validity, meta-analysis, and heterogeneity.
- References:

SUMMARY D=2: POLICY GENERALIZATION

- Nonparametric transportability can be determined provided that the problem instance is encoded in selection diagrams.
- When transportability is feasible, the transport formula can be derived in polynomial time.
- The causal calculus and the algorithm are complete.
- These results solve long-standing problems in the empirical sciences including external validity, meta-analysis, and heterogeneity.
- References:

Pearl & Bareinboim (2011) – semantics for transportability
Bareinboim & Pearl (2012) – completeness of limited TR
Bareinboim et al (2013) – multiple & limited + extensions
Bareinboim & Pearl (2014) – completeness of general TR

THE DATA-FUSION PROBLEM

SAMPLING CONDITIONS (D = 3)

THE SELECTION BIAS PROBLEM

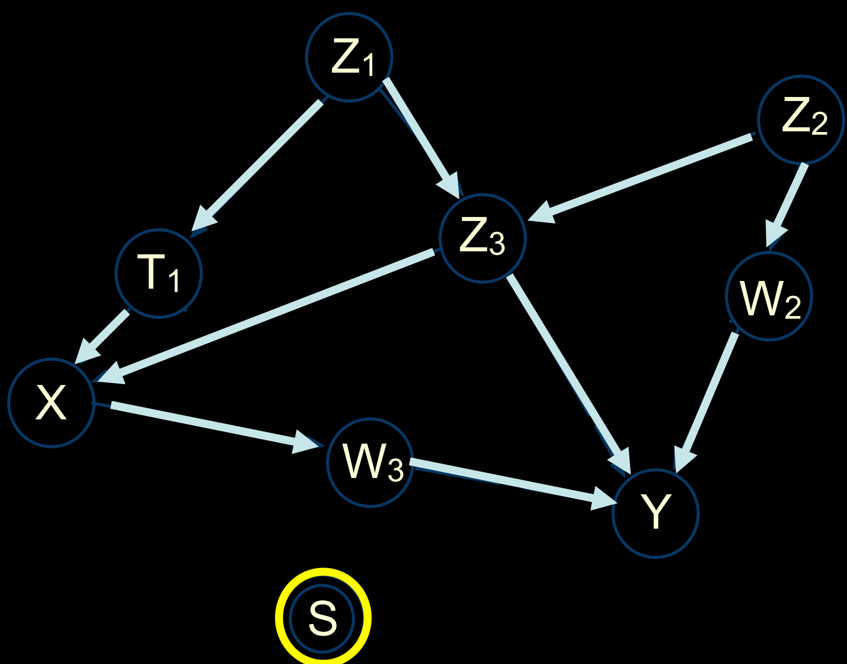
- Selection bias, caused by preferential exclusion of samples from the data, is a major obstacle to valid causal and statistical inferences;

THE SELECTION BIAS PROBLEM

- Selection bias, caused by preferential exclusion of samples from the data, is a major obstacle to valid causal and statistical inferences;

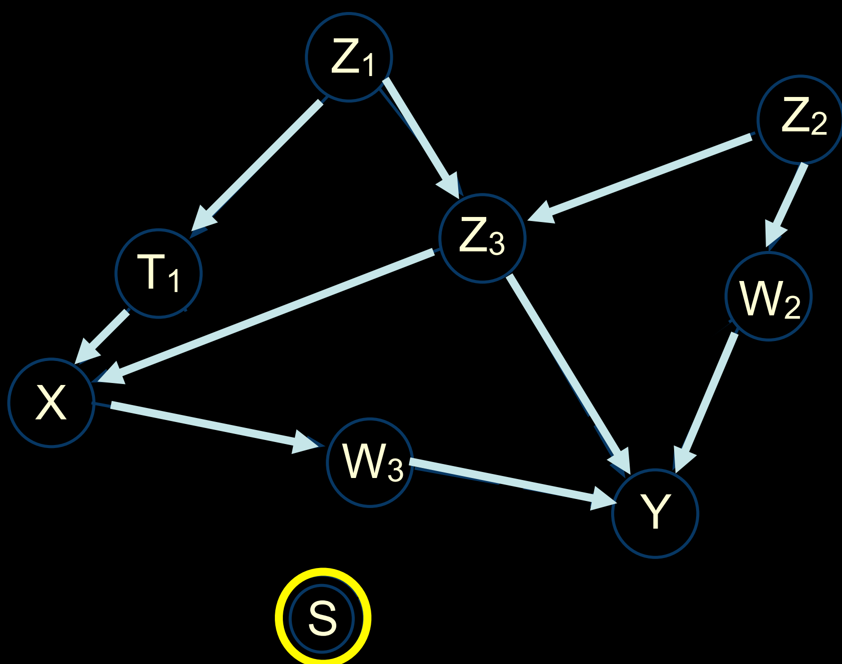
THE SELECTION BIAS PROBLEM

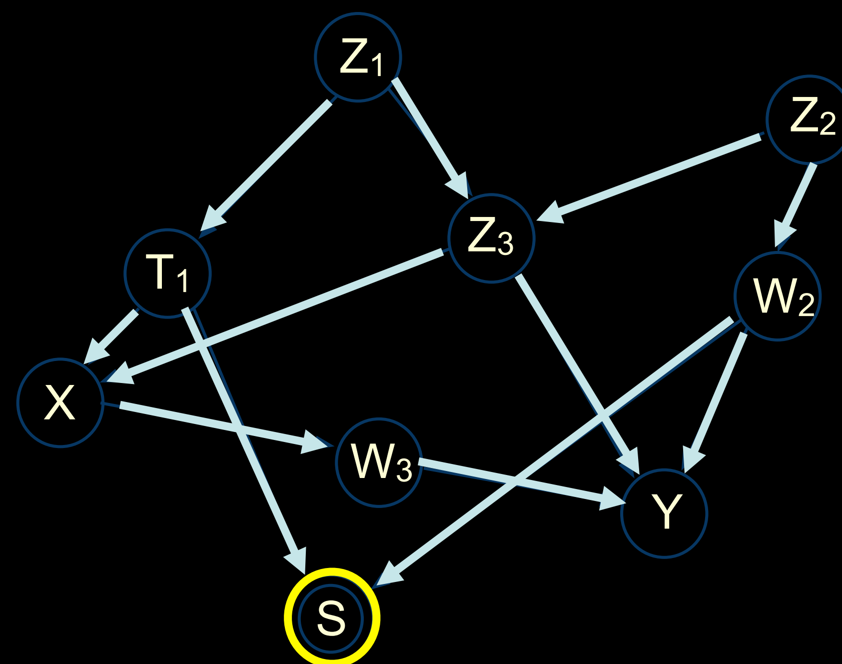
- Selection bias, caused by preferential exclusion of samples from the data, is a major obstacle to valid causal and statistical inferences;



THE SELECTION BIAS PROBLEM

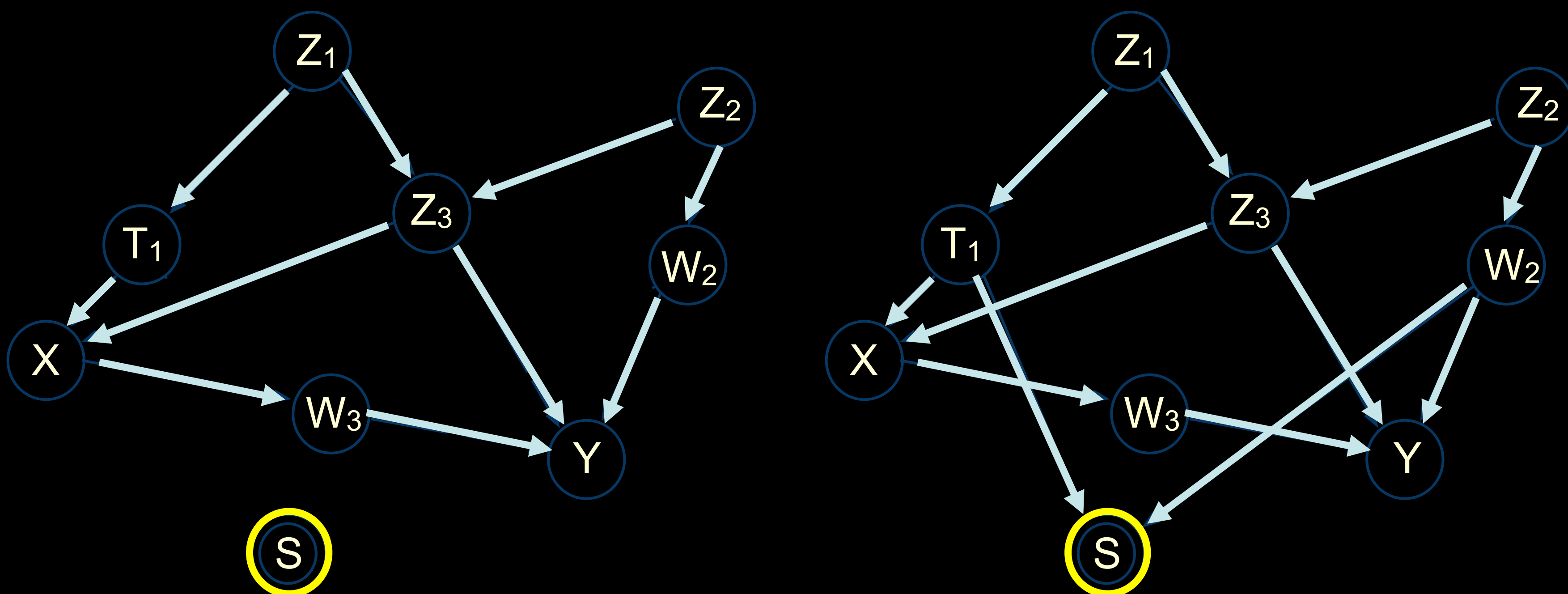
- Selection bias, caused by preferential exclusion of samples from the data, is a major obstacle to valid causal and statistical inferences;





THE SELECTION BIAS PROBLEM

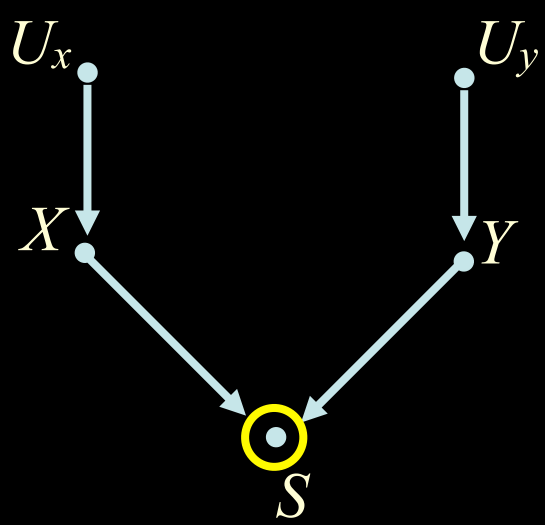
- Selection bias, caused by preferential exclusion of samples from the data, is a major obstacle to valid causal and statistical inferences;



- Under what conditions can we estimate the query (e.g., $P(y \mid x)$) from $P(v \mid S = 1)$.

THE ORIGIN OF SELECTION BIAS

G:



Under selection (S=1):

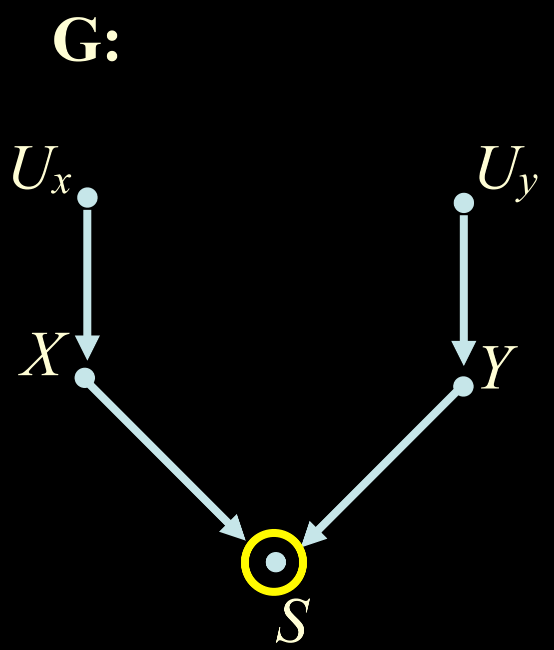
$P(X=1, Y=1 \mid S=1) = 1/2$
$P(X=0, Y=0 \mid S=1) = 1/2$
(correlation=1)

$P(Y=y|X=1,S=1) -$
$P(Y=y|X=0,S=1) = 1$

M:
$X = U_x$
$Y = U_y$
$S = \neg(X \oplus Y)$

$P(U_x,U_y) = 1/4$

THE ORIGIN OF SELECTION BIAS



Under selection (S=1):

$P(X=1, Y=1 \mid S=1) = 1/2$
$P(X=0, Y=0 \mid S=1) = 1/2$
(correlation=1)

$P(Y=y|X=1,S=1) -$
$P(Y=y|X=0,S=1) = 1$

but

$P(Y = y, X = x) = 1/4$

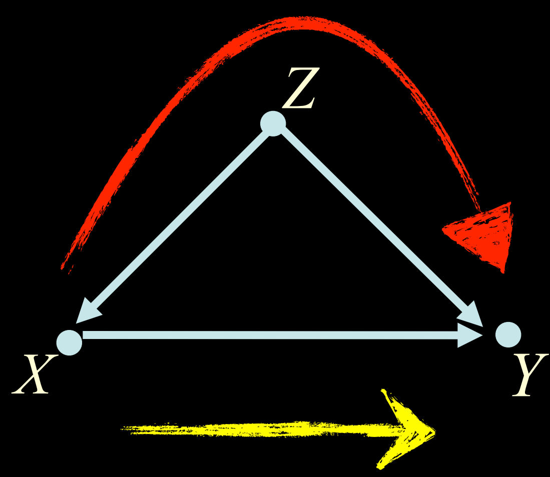
$P(Y=y \mid X=1) -$
$P(Y=y \mid X=0) = 0$

M:
$X = U_x$
$Y = U_y$
$S = \neg(X \oplus Y)$

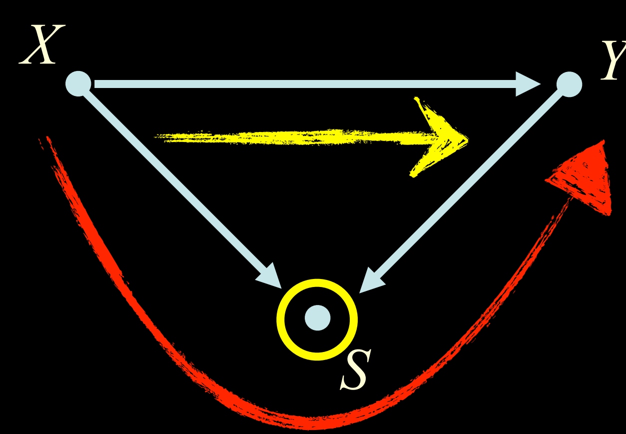
$P(U_x, U_y) = 1/4$

CONFOUNDING BIAS vs SELECTION BIAS

- Unblockable "flow" of information between treatment and outcome — spurious correlation.



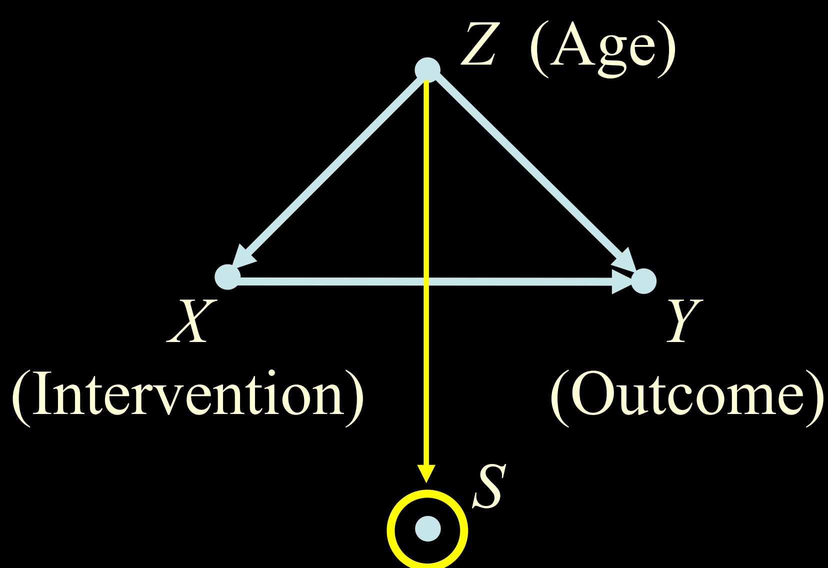
G_1



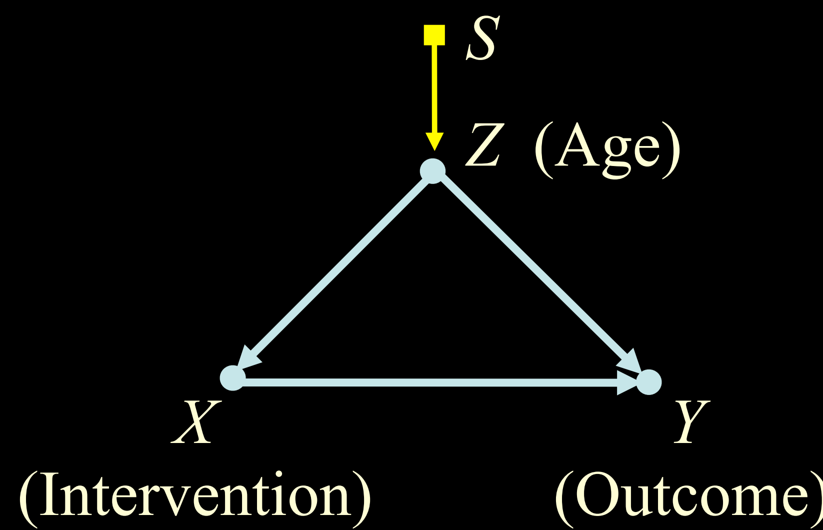
G_2

SELECTION VERSUS TRANSPORTABILITY (MAN-MADE VERSUS NATURE-MADE)

Selection Bias



Transportability



Z = sample-producing factors
S = selection indicator

Z = disparities across populations
S = difference-producing factors

LITERATURE ON SELECTION

LITERATURE ON SELECTION

1. Qualitative assumptions about the data-generating model (treatment-dependent, outcome-dependent).

LITERATURE ON SELECTION

1. Qualitative assumptions about the data-generating model (treatment-dependent, outcome-dependent).

LITERATURE ON SELECTION

1. Qualitative assumptions about the data-generating model (treatment-dependent, outcome-dependent).

   [Cornfield'51, Whittemore'78, Geng'92, Didelez'10]

LITERATURE ON SELECTION

1. Qualitative assumptions about the data-generating model (treatment-dependent, outcome-dependent).

   [Cornfield'51, Whittemore'78, Geng'92, Didelez'10]

LITERATURE ON SELECTION

1. Qualitative assumptions about the data-generating model (treatment-dependent, outcome-dependent).

   [Cornfield’51, Whittemore’78, Geng’92, Didelez’10]

2. Parametric assumptions about the data-generating model (e.g., linear, monotonic).

LITERATURE ON SELECTION

1. Qualitative assumptions about the data-generating model (treatment-dependent, outcome-dependent).

   [Cornfield'51, Whittemore'78, Geng'92, Didelez'10]

2. Parametric assumptions about the data-generating model (e.g., linear, monotonic).

LITERATURE ON SELECTION

1. Qualitative assumptions about the data-generating model (treatment-dependent, outcome-dependent).

   [Cornfield'51, Whittemore'78, Geng'92, Didelez'10]

2. Parametric assumptions about the data-generating model (e.g., linear, monotonic).

   [Heckman'79, Angrist'91, Pearl'13]

LITERATURE ON SELECTION

1. Qualitative assumptions about the data-generating model (treatment-dependent, outcome-dependent).

   [Cornfield'51, Whittemore'78, Geng'92, Didelez'10]

2. Parametric assumptions about the data-generating model (e.g., linear, monotonic).

   [Heckman'79, Angrist'91, Pearl'13]

LITERATURE ON SELECTION

1. Qualitative assumptions about the data-generating model (treatment-dependent, outcome-dependent).

   [Cornfield'51, Whittemore'78, Geng'92, Didelez'10]

2. Parametric assumptions about the data-generating model (e.g., linear, monotonic).

   [Heckman'79, Angrist'91, Pearl'13]

3. Quantitive assumptions about the data-generating model (i.e. $P(S = 1 \mid Z)$).

LITERATURE ON SELECTION

1. Qualitative assumptions about the data-generating model (treatment-dependent, outcome-dependent).

   [Cornfield’51, Whittemore’78, Geng’92, Didelez’10]

2. Parametric assumptions about the data-generating model (e.g., linear, monotonic).

   [Heckman’79, Angrist’91, Pearl’13]

3. Quantitive assumptions about the data-generating model (i.e. P(S = 1 | Z)).

LITERATURE ON SELECTION

1. Qualitative assumptions about the data-generating model (treatment-dependent, outcome-dependent).

   [Cornfield'51, Whittemore'78, Geng'92, Didelez'10]

2. Parametric assumptions about the data-generating model (e.g., linear, monotonic).

   [Heckman'79, Angrist'91, Pearl'13]

3. Quantitive assumptions about the data-generating model (i.e. $P(S = 1 \mid Z)$).

   [Elkan'01, Zadrozny'04, Cortes'08, Storkey'09]

SELECTION BIAS -- VARIANTS

- **Selection bias without external data (Odds Ratio):**
  Input: $P(v \mid S = 1)$;
  Output: $OR(y,x)$ recoverable? if so, how?

SELECTION BIAS -- VARIANTS

- **Selection bias without external data (Odds Ratio):**
  Input: $P(v \mid S = 1)$;
  Output: $OR(y,x)$ recoverable? if so, how?

- **Selection bias without external data:**
  Input: $P(v \mid S = 1)$;
  Output: $P(y \mid x)$ recoverable? if so, how?

SELECTION BIAS -- VARIANTS

- **Selection bias without external data (Odds Ratio):**
  Input: $P(v \mid S = 1)$;
  Output: $OR(y,x)$ recoverable? if so, how?

- **Selection bias without external data:**
  Input: $P(v \mid S = 1)$;
  Output: $P(y \mid x)$ recoverable? if so, how?

- **Selection bias with external data:**
  Input: $P(v \mid S = 1) + P(t)$, $T \subseteq V$;
  Output: $P(y \mid x)$ recoverable? if so, how?

SELECTION BIAS -- VARIANTS

- **Selection bias without external data (Odds Ratio):**
  Input: $P(v \mid S = 1)$;
  Output: $OR(y,x)$ recoverable? if so, how?

- **Selection bias without external data:**
  Input: $P(v \mid S = 1)$;
  Output: $P(y \mid x)$ recoverable? if so, how?

- **Selection bias with external data:**
  Input: $P(v \mid S = 1) + P(t)$, $T \subseteq V$;
  Output: $P(y \mid x)$ recoverable? if so, how?

- **Selection bias in causal inference:**
  Input: $P(v \mid S = 1) + P(t)$, $T \subseteq V$;
  Output: $P(y \mid do(x))$ recoverable? if so, how?

SELECTION BIAS -- VARIANTS

- **Selection bias without external data (Odds Ratio):**
  Input: $P(v \mid S = 1)$;
  Output: $OR(y,x)$ recoverable? if so, how?

- **Selection bias without external data:**
  Input: $P(v \mid S = 1)$;
  Output: $P(y \mid x)$ recoverable? if so, how?

- **Selection bias with external data:**
  Input: $P(v \mid S = 1) + P(t)$, $T \subseteq V$;
  Output: $P(y \mid x)$ recoverable? if so, how?

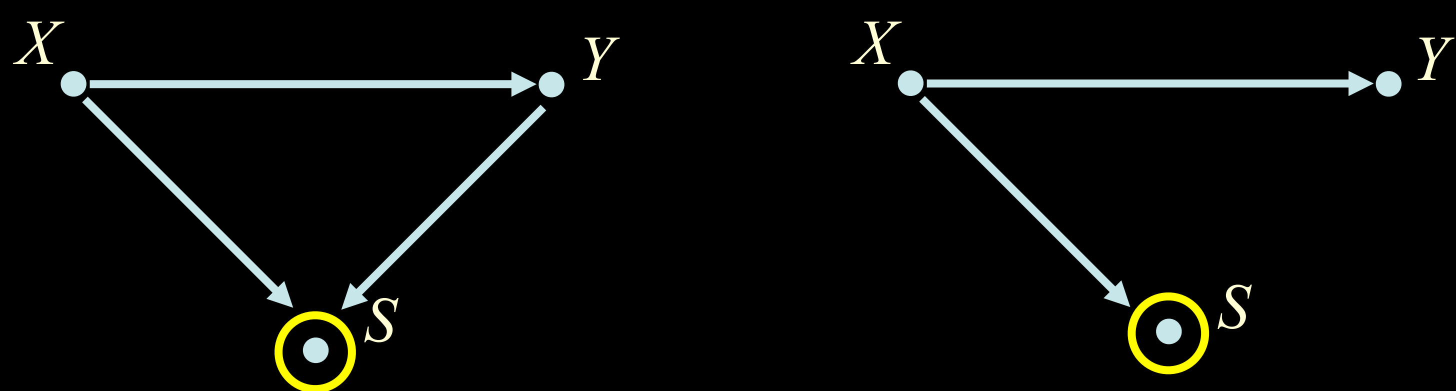
- **Selection bias in causal inference:**
  Input: $P(v \mid S = 1) + P(t)$, $T \subseteq V$;
  Output: $P(y \mid do(x))$ recoverable? if so, how?

- **Selection bias with instrumental variables:**
  Input: $P(v \mid S = 1) + P(t)$, $T \subseteq V$;
  Output: Bounds for $P(y \mid do(x))$ recoverable? if so, how?

SELECTION BIAS WITHOUT EXTERNAL INFORMATION

Question. Under what conditions can we estimate the distribution $P(y \mid x)$ from $P(v \mid S = 1)$.

Theorem. $Q = P(y \mid x)$ is recoverable from selection biased data if and only if $(S \perp\!\!\!\perp Y \mid X)$.

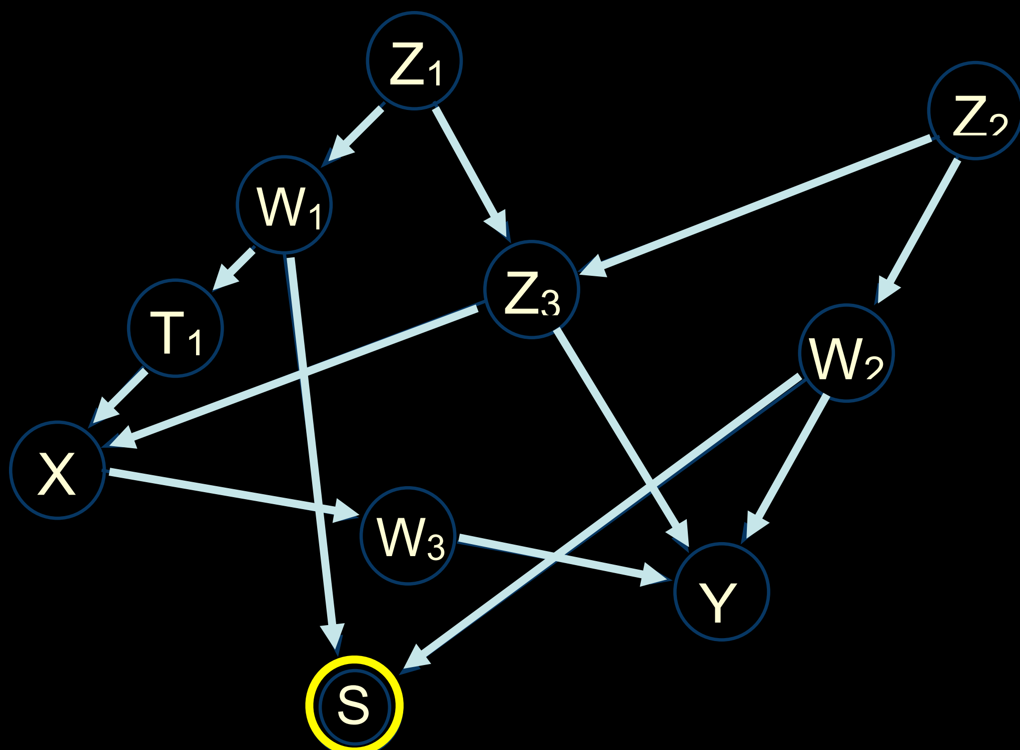


P(y | x) is not recoverable.

P(y | x) is recoverable.

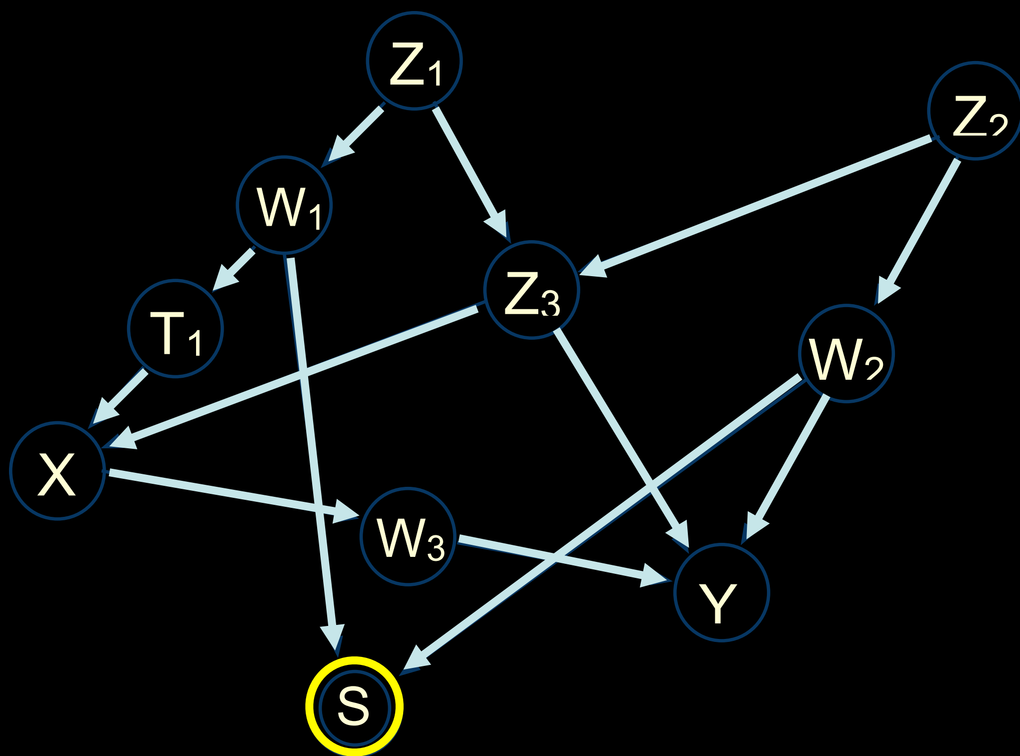
SELECTION WITH EXTERNAL INFORMATION

Assume we want to estimate $Q = P(y \mid x)$ from selection…



SELECTION WITH EXTERNAL INFORMATION

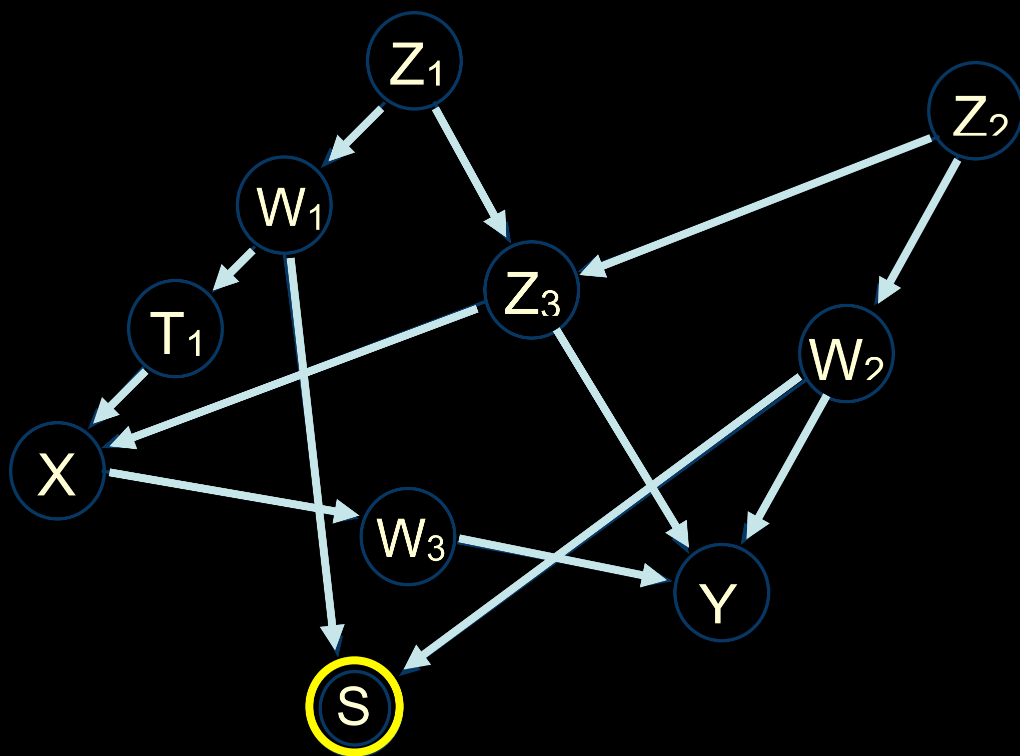
Assume we want to estimate $Q = P(y \mid x)$ from selection…



Q is not recoverable by the previous theorem…

SELECTION WITH EXTERNAL INFORMATION

Assume we want to estimate $Q = P(y \mid x)$ from selection…

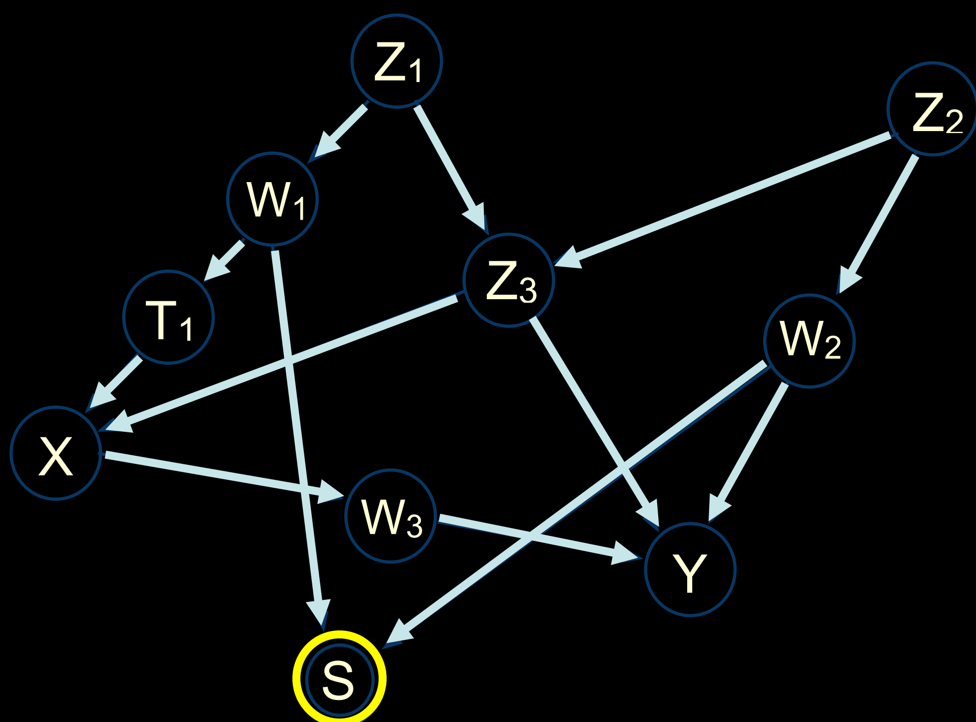


Q is not recoverable by the previous theorem…
but what if $P(W_1, W_2)$ is available?

RECOVERABILITY WITH EXTERNAL INFORMATION

Theorem. $P(y \mid x)$ is recoverable if there is a set C such that $(Y \perp\!\!\!\perp S \mid C, X)$ holds in G and $P(C, X)$ is estimable.

Moreover, $$P(y \mid x) = \sum_{c} P(y \mid x, c, S = 1)\, P(c \mid x)$$

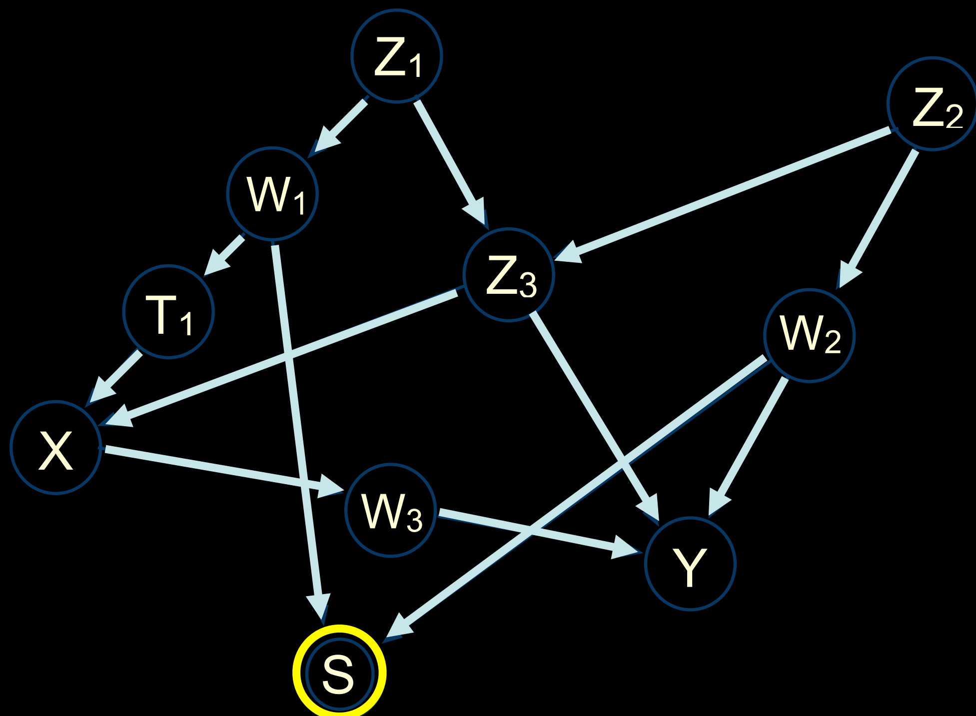


$$C = \{W_1, W_2\}?$$
$$= \{W_2, Z_1, Z_2\}?$$
$$= \{W_2, Z_3\}?$$

RECOVERABILITY WITH EXTERNAL INFORMATION

Theorem. $P(y \mid x)$ is recoverable if there is a set C such that $(Y \perp\!\!\!\perp S \mid C, X)$ holds in G and $P(C, X)$ is estimable.

Moreover, $P(y \mid x) = \sum_{c} P(y \mid x, c, S = 1)\, P(c \mid x)$



$C = \{W_1, W_2\}$? yes

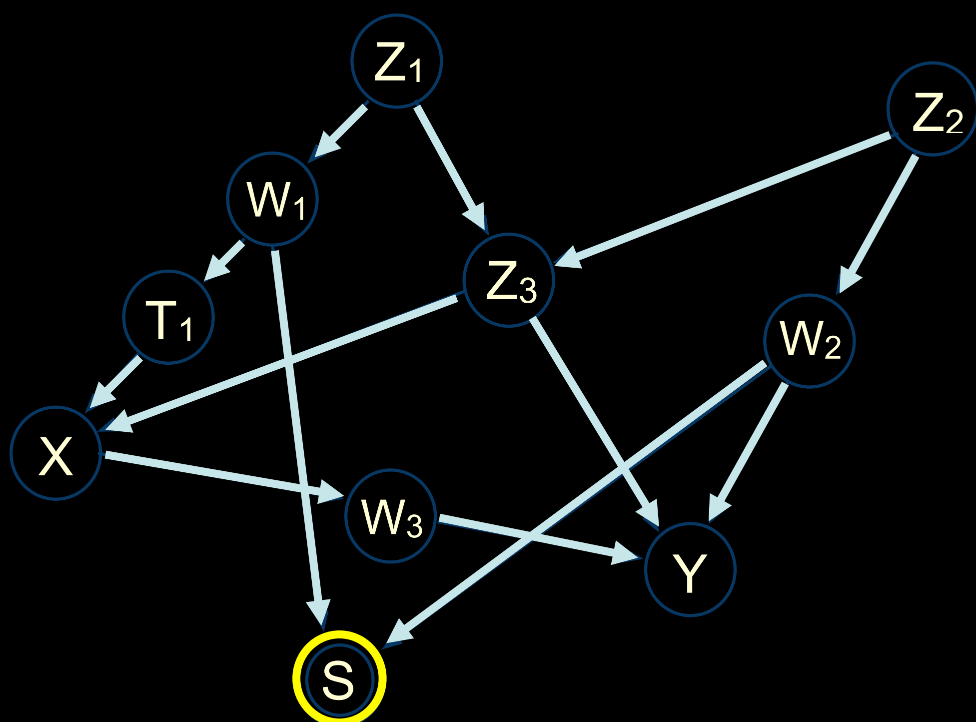
$= \{W_2, Z_1, Z_2\}$?

$= \{W_2, Z_3\}$?

RECOVERABILITY WITH EXTERNAL INFORMATION

Theorem. $P(y \mid x)$ is recoverable if there is a set C such that $(Y \perp\!\!\!\perp S \mid C, X)$ holds in G and $P(C, X)$ is estimable.

Moreover, $P(y \mid x) = \sum_{c} P(y \mid x, c, S = 1)\, P(c \mid x)$



$C = \{W_1, W_2\}$? yes

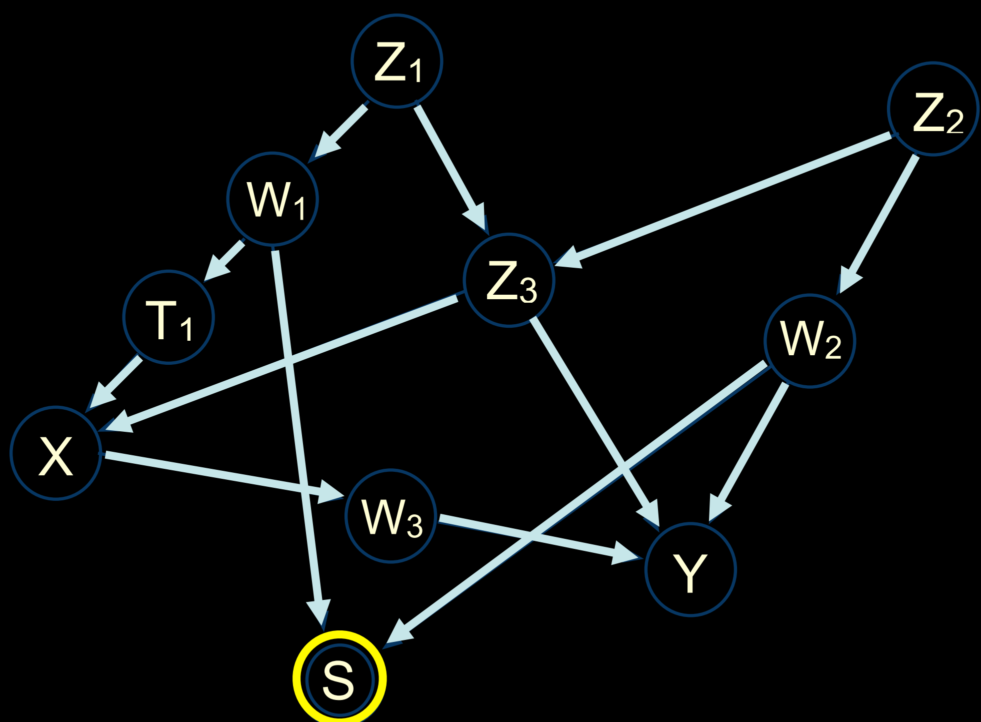
$= \{W_2, Z_1, Z_2\}$? no

$= \{W_2, Z_3\}$?

RECOVERABILITY WITH EXTERNAL INFORMATION

Theorem. $P(y \mid x)$ is recoverable if there is a set C such that $(Y \perp\!\!\!\perp S \mid C, X)$ holds in G and $P(C, X)$ is estimable.

Moreover, $P(y \mid x) = \sum_{c} P(y \mid x, c, S = 1)\, P(c \mid x)$



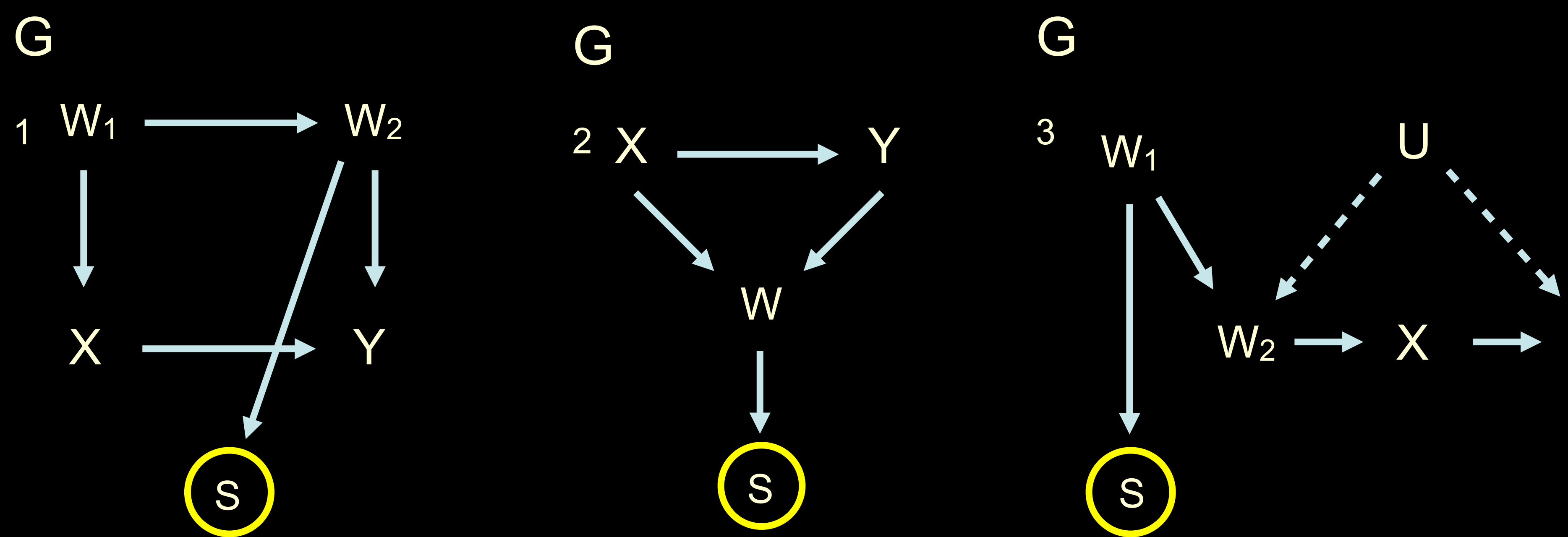
$C = \{W_1, W_2\}$? yes

$= \{W_2, Z_1, Z_2\}$? no

$= \{W_2, Z_3\}$? yes

SELECTION BIAS IN CAUSAL INFERENCE

We want to recover $Q = P(y \mid do(x))$.



$C = \{W_2\}$ recoverable
$= \{W_1\}$ not recoverable

$C = \{\ \}$ not recoverable
$= \{W\}$ not recoverable
$= \{W, X\}$ recoverable

recoverable
$(S \perp\!\!\!\perp Y \mid X, W_2)$
does not hold

SELECTION BIAS IN CAUSAL INFERENCE

Def. Let a set Z be partitioned into Z^+ and Z^- such that Z^+ contains all non-descendants of X and Z^- the descendants of X. Z is said to satisfy the s-backdoor criterion relative to an ordered pair of variables (X, Y) and an ordered pair (M, T) in a graph G_s if Z^+ and Z^- satisfy the following conditions:

(i) Z^+ blocks all backdoor paths from X to Y;
(ii) X and Z^+ block all paths between Z^- and Y;
(iii) X and Z block all paths between S and Y;
(iv) $Z \cup \{X, Y\} \subseteq M$, $Z \subseteq T$.

Thm. If a set Z satisfies the s-backdoor relative to the pairs (X, Y) and (M, T), then the effect of X on Y is identifiable and recoverable and is given by the formula

$$P(y \mid do(x)) = \sum_{z} P(y \mid x, z, S = 1)\, P(z)$$

SUMMARY (D=3): SELECTION BIAS

SUMMARY (D=3): SELECTION BIAS

- Nonparametric recoverability of selection from causal and statistical settings can be determined provided that an augmented causal graph is available.

SUMMARY (D=3): SELECTION BIAS

- Nonparametric recoverability of selection from causal and statistical settings can be determined provided that an augmented causal graph is available.
- When recoverability is feasible, the estimand can be derived in polynomial time.

SUMMARY (D=3): SELECTION BIAS

- Nonparametric recoverability of selection from causal and statistical settings can be determined provided that an augmented causal graph is available.
- When recoverability is feasible, the estimand can be derived in polynomial time.
- The result is complete for pure recoverability and sufficient for recoverability with external information.

SUMMARY (D=3): SELECTION BIAS

- Nonparametric recoverability of selection from causal and statistical settings can be determined provided that an augmented causal graph is available.
- When recoverability is feasible, the estimand can be derived in polynomial time.
- The result is complete for pure recoverability and sufficient for recoverability with external information.
- The back-door criterion can be generalized to handle selection bias.

SUMMARY (D=3): SELECTION BIAS

- Nonparametric recoverability of selection from causal and statistical settings can be determined provided that an augmented causal graph is available.
- When recoverability is feasible, the estimand can be derived in polynomial time.
- The result is complete for pure recoverability and sufficient for recoverability with external information.
- The back-door criterion can be generalized to handle selection bias.
- Stronger results can be obtained for the OR recoverability.

SUMMARY (D=3): SELECTION BIAS

- Nonparametric recoverability of selection from causal and statistical settings can be determined provided that an augmented causal graph is available.
- When recoverability is feasible, the estimand can be derived in polynomial time.
- The result is complete for pure recoverability and sufficient for recoverability with external information.
- The back-door criterion can be generalized to handle selection bias.
- Stronger results can be obtained for the OR recoverability.
- References:

Bareinboim, Pearl (2012) – OR completeness & IVs
Bareinboim, Tian, Pearl (2014) – Markovian completeness
Bareinboim, Tian (2015) – Causal queries
Correa, Bareinboim (2017) – Complete adjustment

CONCLUSIONS

CONCLUSIONS

- Data-fusion requires encoding of structural features of the data-generating model.

CONCLUSIONS

- Data-fusion requires encoding of structural features of the data-generating model.
- Both population-level and individual-level causal inference and fusion can be performed once the proper formal machinery is employed; algorithms and conditions are available.

CONCLUSIONS

- Data-fusion requires encoding of structural features of the data-generating model.
- Both population-level and individual-level causal inference and fusion can be performed once the proper formal machinery is employed; algorithms and conditions are available.

CONCLUSIONS

- Data-fusion requires encoding of structural features of the data-generating model.
- Both population-level and individual-level causal inference and fusion can be performed once the proper formal machinery is employed; algorithms and conditions are available.
- Principled framework for data-fusion — pooling and aggregating observational, experimental, and counterfactual information spread throughout heterogenous domains for population- and individual-level causal inference (understanding) and decision-making.

"Development of Western Science is based on two great achievements, the invention of the formal logical system (in Euclidean geometry) by the Greek philosophers, and the discovery of the possibility to find out causal relationships by systematic experiment (during the Renaissance)". Albert Einstein.

"Imagine how much harder physics would be if electrons had feelings!". Richard Feynman.

THANK YOU!