# A (Brief) Introduction to Causal Inference

Evan A. Jones
University of Maryland – College Park

February 15, 2021

Prepared for GVPT's GSA Method Workshop, Spring 2021.

# The "Why?" and "What If?" Questions

- Understanding the world around us is an inherently human endeavor

May not know causal mechanics of physical world down to *functional* form

For instance, many adult may not know the Gas laws PV=nRT, but they have a correct, intuitive—or high-level—understanding of the these relationships just based on playing with balloons or filling your bike tires as a kid

# The "Why?" and "What If?" Questions

- Understanding the world around us is an inherently human endeavor
- Human children explore the world as scientists do (2, 4):
  - Asking questions
  - Forming hypotheses
  - Testing hypotheses via interventions (5)




May not know causal mechanics of physical world down to *functional* form

For instance, many adult may not know the Gas laws PV=nRT, but they have a correct, intuitive—or high-level—understanding of the these relationships just based on playing with balloons or filling your bike tires as a kid

# The "Why?" and "What If?" Questions

- Understanding the world around us is an inherently human endeavor
- Human children explore the world as scientists do (2, 4):
  - Asking questions
  - Forming hypotheses
  - Testing hypotheses via interventions (5)
- By adulthood, we have fairly solid causal intuition about the physical world




May not know causal mechanics of physical world down to *functional* form

For instance, many adult may not know the Gas laws PV=nRT, but they have a correct, intuitive—or high-level—understanding of the these relationships just based on playing with balloons or filling your bike tires as a kid

# The "Why?" and "What If?" Questions

- As researchers, we fit regressions all the time and interpret coefficients

As researchers, we go a step further. Seek to precisely uncover low-level relationships

Qualitative researchers also perform causal inference

Rather than uncovering functional forms or quantifying effects, seek out entire causal mechanisms or "configurations" of variables

So much of what we learn about interpreting regressions through telegraphic readings of the political science literature is simply wrong

Not a criticism of previous generations of scholars, causal inference simply not taught in most departments until recently

Moreover, causal inference still rarely taught in *statistics departments*

# The "Why?" and "What If?" Questions

- As researchers, we fit regressions all the time and interpret coefficients
- Take the following, for instance :

$$Y = \alpha + \beta x + \theta z + \epsilon$$



As researchers, we go a step further. Seek to precisely uncover low-level relationships

Qualitative researchers also perform causal inference

Rather than uncovering functional forms or quantifying effects, seek out entire causal mechanisms or "configurations" of variables

So much of what we learn about interpreting regressions through telegraphic readings of the political science literature is simply wrong

Not a criticism of previous generations of scholars, causal inference simply not taught in most departments until recently

Moreover, causal inference still rarely taught in *statistics departments*

# The "Why?" and "What If?" Questions

- As researchers, we fit regressions all the time and interpret coefficients
- Take the following, for instance :

$$Y = \alpha + \beta x + \theta z + \epsilon$$

- When can we interpret $\beta$ as a causal effect?

As researchers, we go a step further. Seek to precisely uncover low-level relationships

Qualitative researchers also perform causal inference

Rather than uncovering functional forms or quantifying effects, seek out entire causal mechanisms or "configurations" of variables

So much of what we learn about interpreting regressions through telegraphic readings of the political science literature is simply wrong

Not a criticism of previous generations of scholars, causal inference simply not taught in most departments until recently

Moreover, causal inference still rarely taught in *statistics departments*

# What is causality?



Ask audience to answer this question?

What distinguishes description and prediction from causal inference?

How can we move beyond observation, description, and prediction and towards answering causal and counterfactual questions?

This talk won't give you every tool you need to perform causal inference in your own project from start to finish

Field is much too large for that

Hope is to provide you with the philosophical and logical tool kit to engage this literature on your own, and prepare you for presentations to come

Relay how I first jumped into this literature, and why starting from "30,000" feet would have been better

# Outline

- Logic of Causal Inference
- Experiments vs the World
- Potential Outcomes vs Structural Causal Models

# Logic of Causal Inference

# Beyond Description

- Causal explanations are "more than mere descriptions . . . of the observed data" (1, p. 3)

Regression models, by themselves, are just descriptions of data, even if it has a high $R^2$, BIC, AIC, etc.,

# Beyond Description

- Causal explanations are "more than mere descriptions . . . of the observed data" (1, p. 3)

- Break down phenomena into constituent parts and define how parts interact to produce emergent behavior (*data-generating process*) (12)

Regression models, by themselves, are just descriptions of data, even if it has a high $R^2$, BIC, AIC, etc.,

# Beyond Description

- Causal explanations are "more than mere descriptions . . . of the observed data" (1, p. 3)
- Break down phenomena into constituent parts and define how parts interact to produce emergent behavior (*data-generating process*) (12)
- Once uncovered, causal mechanisms are powerful

Regression models, by themselves, are just descriptions of data, even if it has a high $R^2$, BIC, AIC, etc.,

# Beyond Description

We can answer counterfactual questions:

- (Incumbency effect) What *would have* been the election outcome if the candidate were an incumbent?

These are questions about a supposition or imagined state of reality

A different turn of historical events

A real-time intervention

A future event

# Beyond Description

We can answer counterfactual questions:

- (Incumbency effect) What *would have* been the election outcome if the candidate were an incumbent?

- (Resource curse) What *would have* been the GDP growth rate without oil?



These are questions about a supposition or imagined state of reality

A different turn of historical events

A real-time intervention

A future event

# Beyond Description

We can answer counterfactual questions:

- (Incumbency effect) What *would have* been the election outcome if the candidate were an incumbent?
- (Resource curse) What *would have* been the GDP growth rate without oil?
- (Democratic peace) *Would* the two countries have escalated conflict similarly if they were both autocratic?



These are questions about a supposition or imagined state of reality

A different turn of historical events

A real-time intervention

A future event

Causal mechanisms allow us to make unbiased predictions about *counterfactual situations* and the effect of *interventions* (10)



Must an intervention be something that is manipulable? Or can it represent a given state, such as a race, gender, height or age? Which camp do you fall into?

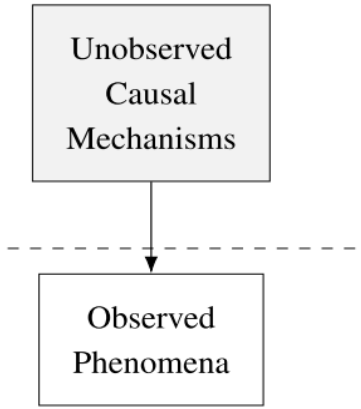
We are interested in study the causal effects of non-manipulable effects all the time, and we do indeed "manipulate" them in experiments, but only in the abstract. We cannot manipulate these states for individuals.

There is on-going debate about whether non-manipulable variables can take on causal interpretation

Worth reading

# The Ladder of Causation

Answering causal queries requires more than observing data. Why?




*Taken from (1, p. 6)*

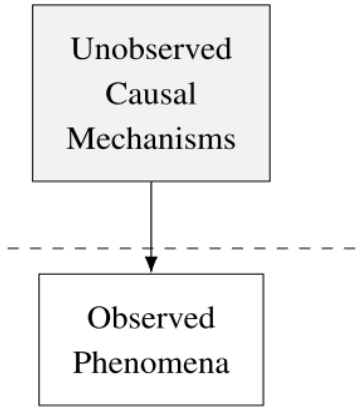
Reality is a garden of forking paths, we only see one potential outcome of many

We cannot observe all potential outcomes for every individual

# The Ladder of Causation

Answering causal queries requires more than observing data. Why?

1. Causal mechanisms are generally *unobservable* (8)




*Taken from (1, p. 6)*

Reality is a garden of forking paths, we only see one potential outcome of many

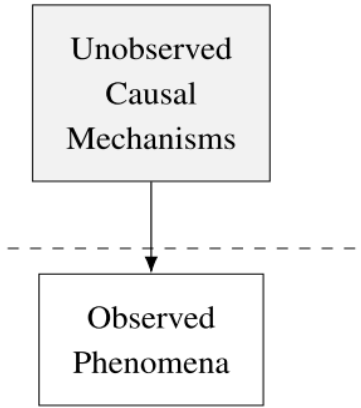
We cannot observe all potential outcomes for every individual

# The Ladder of Causation

Answering causal queries requires more than observing data. Why?

1. Causal mechanisms are generally *unobservable* (8)

2. Data represent single realized outcome of an intervention (traces of mechanism)




*Taken from (1, p. 6)*

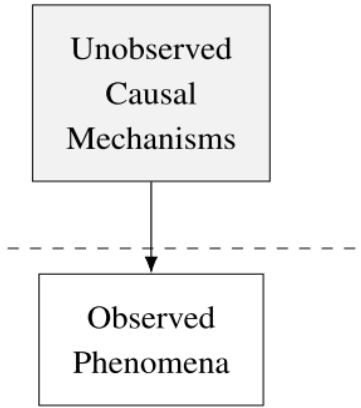
Reality is a garden of forking paths, we only see one potential outcome of many

We cannot observe all potential outcomes for every individual

# The Ladder of Causation

Answering causal queries requires more than observing data. Why?

1. Causal mechanisms are generally *unobservable* (8)
2. Data represent single realized outcome of an intervention (traces of mechanism)
3. Sample does not match the population/group we want to study




*Taken from (1, p. 6)*

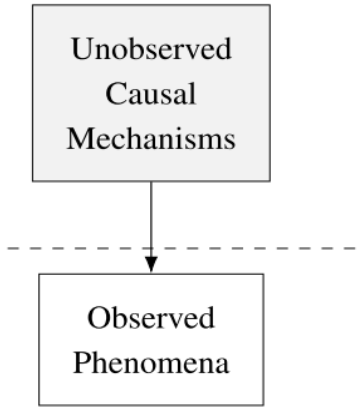
Reality is a garden of forking paths, we only see one potential outcome of many

We cannot observe all potential outcomes for every individual

# The Ladder of Causation

Answering causal queries requires more than observing data. Why?

1. Causal mechanisms are generally *unobservable* (8)
2. Data represent single realized outcome of an intervention (traces of mechanism)
3. Sample does not match the population/group we want to study
4. Data suggest paradoxical effects




*Taken from (1, p. 6)*

Reality is a garden of forking paths, we only see one potential outcome of many

We cannot observe all potential outcomes for every individual

# Simpson's Paradox

Believe the Election Was Stolen

| | | **Total** |
|---|---|---|
| **Misinfo** | Yes | 47% ($\frac{582}{1240}$) |
| | No | 60% ($\frac{456}{760}$) |
| | | $\mathbb{E}[Y\|T]$ |

- Social media data on user engagement
- Consumers of misinformation are *less* likely to believe the election was stolen
- Hmm... what is happening?

We are studying the effect of misinformation on beliefs about the election

Group by party, we see 800 out of 1240 people in sample who were exposed to misinformation were Democrats vs 440 Reps, so the 30% is up-weighted

Among non-exposed, far more Reps than Dems, so the 72$ is up-weighted

Which ATE is correct? Why and how do you know? Is the sample bad? How do you know? Where's the evidence? Everything is based on (plausible) assumptions (grounded prior knowledge).

Once we wipe the fog from our glasses, all assumptions—about sample size or otherwise—are rooted in a causal structure, as we'll see.

# Simpson's Paradox

Believe the Election Was Stolen

| Misinfo | D | R | Total |
|---|---|---|---|
| Yes | 30% ($\frac{240}{800}$) | 78% ($\frac{342}{440}$) | 47% ($\frac{582}{1240}$) |
| No | 11% ($\frac{16}{150}$) | 72% ($\frac{440}{610}$) | 60% ($\frac{456}{760}$) |
| | $\mathbb{E}[Y\|T,C{=}D]$ | $\mathbb{E}[Y\|T,C{=}R]$ | $\mathbb{E}[Y\|T]$ |

- When we group by party, the effect of misinformation flips

$$\underbrace{\frac{800}{1240}}_{\text{upweight}}(0.3) + \frac{440}{1240}(0.78) = 0.47$$

$$\frac{150}{760}(0.11) + \underbrace{\frac{610}{760}}_{\text{upweight}}(0.73) = 0.60$$

We are studying the effect of misinformation on beliefs about the election

Group by party, we see 800 out of 1240 people in sample who were exposed to misinformation were Democrats vs 440 Reps, so the 30% is up-weighted

Among non-exposed, far more Reps than Dems, so the 72$ is up-weighted

Which ATE is correct? Why and how do you know? Is the sample bad? How do you know? Where's the evidence? Everything is based on (plausible) assumptions (grounded prior knowledge).

Once we wipe the fog from our glasses, all assumptions—about sample size or otherwise—are rooted in a causal structure, as we'll see.

# Simpson's Paradox

Believe the Election Was Stolen

| Misinfo | D | R | Total |
|---|---|---|---|
| Yes | 30% ($\frac{240}{800}$) | 78% ($\frac{342}{440}$) | 47% ($\frac{582}{1240}$) |
| No | 11% ($\frac{16}{150}$) | 72% ($\frac{440}{610}$) | 60% ($\frac{456}{760}$) |
| | $\mathbb{E}[Y\|T,C{=}D]$ | $\mathbb{E}[Y\|T,C{=}R]$ | $\mathbb{E}[Y\|T]$ |

$$\mathbb{E}[Y|T{=}1]-\mathbb{E}[Y|T{=}0]{=}-0.13$$

$$\mathbb{E}[Y|T{=}1,C]-\mathbb{E}[Y|T{=}0,C]{=}0.12$$

**What is the effect of misinformation?**

- When we group by party, the effect of misinformation flips

$$\underbrace{\frac{800}{1240}}_{\text{upweight}}(0.3)+\frac{440}{1240}(0.78)=0.47$$

$$\frac{150}{760}(0.11)+\underbrace{\frac{610}{760}}_{\text{upweight}}(0.73)=0.60$$

We are studying the effect of misinformation on beliefs about the election

Group by party, we see 800 out of 1240 people in sample who were exposed to misinformation were Democrats vs 440 Reps, so the 30% is up-weighted

Among non-exposed, far more Reps than Dems, so the 72$ is up-weighted

Which ATE is correct? Why and how do you know? Is the sample bad? How do you know? Where's the evidence? Everything is based on (plausible) assumptions (grounded prior knowledge).

Once we wipe the fog from our glasses, all assumptions—about sample size or otherwise—are rooted in a causal structure, as we'll see.

# The Ladder of Causation

| Layer (Symbolic) | Typical Activity | Typical Question | Example | Statistics |
|---|---|---|---|---|

Based on table from (1, p. 8)

# The Ladder of Causation

| | Layer (Symbolic) | Typical Activity | Typical Question | Example | Statistics |
|---|---|---|---|---|---|
| $\mathcal{L}_1$ | Associational $P(y\|x)$ | Seeing | What is? How would seeing $X$ change my belief in $Y$? | What does a speech tell us about a politician's ideology? | Regression / Model fitting / MLE |

Based on table from (1, p. 8)

# The Ladder of Causation

| | Layer (Symbolic) | Typical Activity | Typical Question | Example | Statistics |
|---|---|---|---|---|---|
| $\mathcal{L}_1$ | Associational $P(y\|x)$ | Seeing | What is? How would seeing $X$ change my belief in $Y$? | What does a speech tell us about a politician's ideology? | Regression / Model fitting / MLE |
| $\mathcal{L}_2$ | Interventional $P(y\|do(x), c)$ | Doing | What if? What if I do $X$? | What if we tax carbon, will GHG emissions lower? | Experiment / Natural Experiment / Adjustment |

Based on table from (1, p. 8)

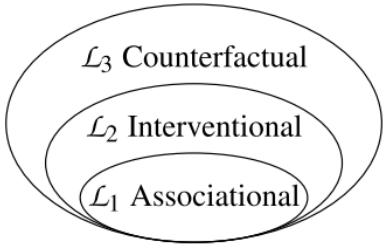
# The Ladder of Causation

| | Layer (Symbolic) | Typical Activity | Typical Question | Example | Statistics |
|---|---|---|---|---|---|
| $\mathcal{L}_1$ | Associational $P(y\|x)$ | Seeing | What is? How would seeing $X$ change my belief in $Y$? | What does a speech tell us about a politician's ideology? | Regression / Model fitting / MLE |
| $\mathcal{L}_2$ | Interventional $P(y\|do(x), c)$ | Doing | What if? What if I do $X$? | What if we tax carbon, will GHG emissions lower? | Experiment / Natural Experiment / Adjustment |
| $\mathcal{L}_3$ | Counterfactual $P(y_x\|x', y')$ | Imagining | Why? What if I had acted differently? | Was it the Russians that caused Trump to win? | (Adjustment) |

Based on table from (1, p. 8)

# The Problem of Causal Inference

- The layers of the causal hierarchy are nested



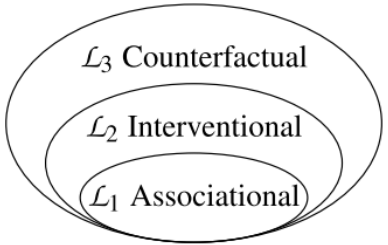

*Taken from (1, p. 6)*

What do we mean by "causal assumptions"?

Where do "causal assumptions" come from?

# The Problem of Causal Inference

- The layers of the causal hierarchy are nested
- *Generally impossible* to draw higher-layer inferences with only lower-layer information (1)



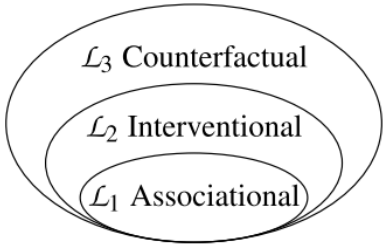

*Taken from (1, p. 6)*

What do we mean by "causal assumptions"?

Where do "causal assumptions" come from?

# The Problem of Causal Inference

- The layers of the causal hierarchy are nested
- *Generally impossible* to draw higher-layer inferences with only lower-layer information (1)
- Q: So what "information" allows us to move up the ladder?




*Taken from (1, p. 6)*

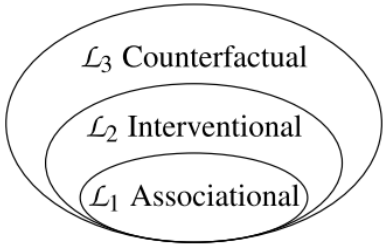


What do we mean by "causal assumptions"?

Where do "causal assumptions" come from?

# The Problem of Causal Inference

- The layers of the causal hierarchy are nested
- *Generally impossible* to draw higher-layer inferences with only lower-layer information (1)
- Q: So what "information" allows us to move up the ladder?
- A: *Causal assumptions*

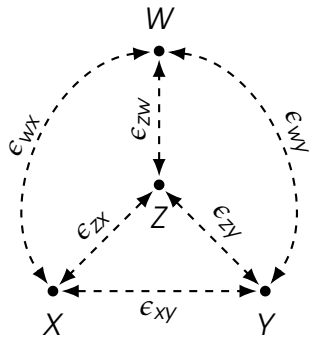



*Taken from (1, p. 6)*

What do we mean by "causal assumptions"?

Where do "causal assumptions" come from?

# The Problem of Causal Inference




Observed data ($\mathcal{L}_1$)

- At $\mathcal{L}_1$ we have a variable "salad"
- In terms of probability, all we know is $P(X, Y, Z, W)$
- Everything *could be* related to everything else
- Best we can do is estimate associations (correlations)



Explain DAG, nodes are variables, dashed edges represent potential relationships, arrows indicate direction of effect.

**Ask audience member what they study**

Imagine R dataframe with these variables, show it to your grandma or child. Does it mean anything to them? Let them play with some descriptive plots, can they tell you anything about direction of effects?

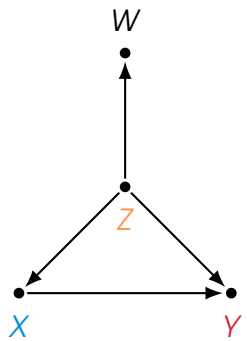
Causal assumptions are represented by the absence of relationships (edges)

Assuming NO relationship stricter than the contrary

Intervening frees of of causal assumptions because we know value of X is independent (a la randomization)

At $\mathcal{L}_3$ we observe some intervention, and want to know what would have happened had we altered the intervention or possibly some other variable in the system.

# The Problem of Causal Inference




observed data ($\mathcal{L}_1$) +
causal assumptions

- With knowledge + additional evidence, we *assume* away some paths
- Arrows imply conditional dependencies:
  $\implies P(Y|Z,X)P(X,W|Z)P(Z)$
- Still *no intervention*, observed effect of $X$ on $Y$ depends on $Z$



Explain DAG, nodes are variables, dashed edges represent potential relationships, arrows indicate direction of effect.

**Ask audience member what they study**

Imagine R dataframe with these variables, show it to your grandma or child. Does it mean anything to them? Let them play with some descriptive plots, can they tell you anything about direction of effects?

Causal assumptions are represented by the absence of relationships (edges)

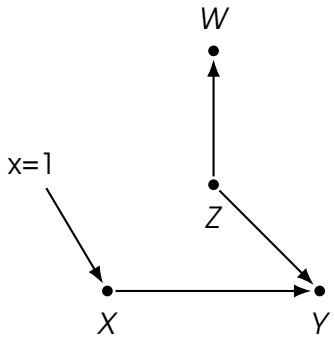
Assuming NO relationship stricter than the contrary

Intervening frees of of causal assumptions because we know value of X is independent (a la randomization)

At $\mathcal{L}_3$ we observe some intervention, and want to know what would have happened had we altered the intervention or possibily some other variable in the system.

# The Problem of Causal Inference

- We set the value of $X$:
  $do(x = 1)$




Intervention data ($\mathcal{L}_2$)



Explain DAG, nodes are variables, dashed edges represent potential relationships, arrows indicate direction of effect.

**Ask audience member what they study**

Imagine R dataframe with these variables, show it to your grandma or child. Does it mean anything to them? Let them play with some descriptive plots, can they tell you anything about direction of effects?

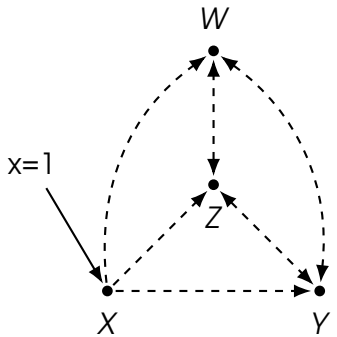
Causal assumptions are represented by the absence of relationships (edges)

Assuming NO relationship stricter than the contrary

Intervening frees of of causal assumptions because we know value of X is independent (a la randomization)

At $\mathcal{L}_3$ we observe some intervention, and want to know what would have happened had we altered the intervention or possibily some other variable in the system.

# The Problem of Causal Inference




Intervention data ($\mathcal{L}_2$)

- We set the value of $X$:
  $do(x = 1)$
- Causal assumptions rendered moot
- If $X$ influences $Y$, then a change in $X$ will appear as a change in $Y$

$$\mathbb{E}[Y|X = x_1] - \mathbb{E}[Y|X = x_0] \neq 0$$



Explain DAG, nodes are variables, dashed edges represent potential relationships, arrows indicate direction of effect.

**Ask audience member what they study**

Imagine R dataframe with these variables, show it to your grandma or child. Does it mean anything to them? Let them play with some descriptive plots, can they tell you anything about direction of effects?

Causal assumptions are represented by the absence of relationships (edges)

Assuming NO relationship stricter than the contrary

Intervening frees of of causal assumptions because we know value of X is independent (a la randomization)

At $\mathcal{L}_3$ we observe some intervention, and want to know what would have happened had we altered the intervention or possibily some other variable in the system.

# Simpson's Paradox Revisited

- What is the effect of misinformation on the belief that the 2020 election was fraudulent?

$$\mathbb{E}[Y|T=1] - \mathbb{E}[Y|T=0] = -0.13$$

or

$$\mathbb{E}[Y|T=1, C] - \mathbb{E}[Y|T=0, C] = 0.12$$



Does partisanship influence consumption of misinformation *and* belief in stop the steal? Very plausible, lot of evidence suggests this is the case

However, maybe you think exposure to/consumption of misinfo influences partianship. E.g. friends who used to be progressive, slowly become Republican after years of consuming conspiratorial content on the internet.

In scenario two, by comparing effects, we see that direct effect of misinformation is positive, but total effect is negative which suggests that indirect channel through partisanship is negative. Maybe, for average voter, exposure to misinfo turns then *away* from the Republicans because of content and leads them to become more left-leaning.

Both are plausible, require further validation of causal assumptions

# Simpson's Paradox Revisited

- What is the effect of misinformation on the belief that the 2020 election was fraudulent?

$$\mathbb{E}[Y|T=1] - \mathbb{E}[Y|T=0] = -0.13$$

or

$$\mathbb{E}[Y|T=1, C] - \mathbb{E}[Y|T=0, C] = 0.12$$

- It depends on your causal assumptions



Does partisanship influence consumption of misinformation *and* belief in stop the steal? Very plausible, lot of evidence suggests this is the case

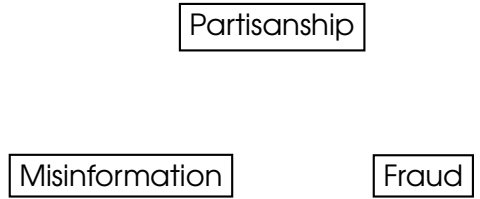
However, maybe you think exposure to/consumption of misinfo influences partianship. E.g. friends who used to be progressive, slowly become Republican after years of consuming conspiratorial content on the internet.

In scenario two, by comparing effects, we see that direct effect of misinformation is positive, but total effect is negative which suggests that indirect channel through partisanship is negative. Maybe, for average voter, exposure to misinfo turns then *away* from the Republicans because of content and leads them to become more left-leaning.

Both are plausible, require further validation of causal assumptions

# Simpson's Paradox Revisited

- What is the effect of misinformation on the belief that the 2020 election was fraudulent?




$$\mathbb{E}[Y|T=1] - \mathbb{E}[Y|T=0] = -0.13$$

or

$$\mathbb{E}[Y|T=1, C] - \mathbb{E}[Y|T=0, C] = 0.12$$

- It depends on your causal assumptions



Does partisanship influence consumption of misinformation *and* belief in stop the steal? Very plausible, lot of evidence suggests this is the case

However, maybe you think exposure to/consumption of misinfo influences partianship. E.g. friends who used to be progressive, slowly become Republican after years of consuming conspiratorial content on the internet.

In scenario two, by comparing effects, we see that direct effect of misinformation is positive, but total effect is negative which suggests that indirect channel through partisanship is negative. Maybe, for average voter, exposure to misinfo turns then *away* from the Republicans because of content and leads them to become more left-leaning.

Both are plausible, require further validation of causal assumptions

# Simpson's Paradox Revisited

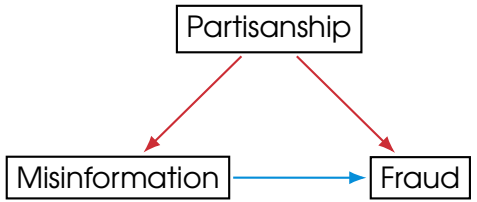
- What is the effect of misinformation on the belief that the 2020 election was fraudulent?

$$\mathbb{E}[Y|T=1] - \mathbb{E}[Y|T=0] = -0.13$$

or

$$\mathbb{E}[Y|T=1, C] - \mathbb{E}[Y|T=0, C] = 0.12$$

- It depends on your causal assumptions






Does partisanship influence consumption of misinformation *and* belief in stop the steal? Very plausible, lot of evidence suggests this is the case

However, maybe you think exposure to/consumption of misinfo influences partianship. E.g. friends who used to be progressive, slowly become Republican after years of consuming conspiratorial content on the internet.

In scenario two, by comparing effects, we see that direct effect of misinformation is positive, but total effect is negative which suggests that indirect channel through partisanship is negative. Maybe, for average voter, exposure to misinfo turns then *away* from the Republicans because of content and leads them to become more left-leaning.

Both are plausible, require further validation of causal assumptions

# Simpson's Paradox Revisited

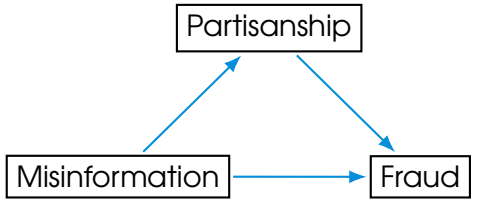
- What is the effect of misinformation on the belief that the 2020 election was fraudulent?

$$\mathbb{E}[Y|T=1] - \mathbb{E}[Y|T=0] = -0.13$$

or

$$\mathbb{E}[Y|T=1, C] - \mathbb{E}[Y|T=0, C] = 0.12$$

- It depends on your causal assumptions






Does partisanship influence consumption of misinformation *and* belief in stop the steal? Very plausible, lot of evidence suggests this is the case

However, maybe you think exposure to/consumption of misinfo influences partianship. E.g. friends who used to be progressive, slowly become Republican after years of consuming conspiratorial content on the internet.

In scenario two, by comparing effects, we see that direct effect of misinformation is positive, but total effect is negative which suggests that indirect channel through partisanship is negative. Maybe, for average voter, exposure to misinfo turns then *away* from the Republicans because of content and leads them to become more left-leaning.

Both are plausible, require further validation of causal assumptions

# The Problem of Causal Inference

"The central question in the analysis of causal effects is the question of *identification*: can the controlled (post-intervention) distribution, $P(Y = y|do(x))$, be estimated from data governed by the pre-intervention distribution $P(X, Y, Z, W)$?"

- Pearl (2009, p. 108)

# The Problem of Causal Inference

- Thus, the key to causal inference is achieving identification

# The Problem of Causal Inference

- Thus, the key to causal inference is achieving identification
- In experiments, identification is built-in since *we* control the treatment

# The Problem of Causal Inference

- Thus, the key to causal inference is achieving identification
- In experiments, identification is built-in since *we* control the treatment
- In observational data, identification is tougher and, sometimes, *unachievable*

# The Problem of Causal Inference

- Thus, the key to causal inference is achieving identification
- In experiments, identification is built-in since *we* control the treatment
- In observational data, identification is tougher and, sometimes, *unachievable*
- So why not only do experiments?

# Experiments vs the World

# Experiments: Pros and Cons

## Pros

- Identification guaranteed
- Control over who you want to assign treatments to and how
- e.g. different types of sampling
- Can design the experiment to directly assess research question of interest

## Cons

- Ethical example cannot start a war
- Physical example limited to phenomena you can examine in a lab or survey
- Temporal example many causal chains stretch over long time scales, longitudinal studies expensive
- Transportability highly unlikely that the sample you achieved is perfectly random and free of selection bias, effects unbiased within the sample, but not population of interest

# Experiments: Pros and Cons

## Pros

- Identification guaranteed

$\Longrightarrow$ Internal validity

## Cons

Control over who you want to assign treatments to and how

e.g. different types of sampling

Can design the experiment to directly assess research question of interest

Ethical example cannot start a war

Physical example limited to phenomena you can examine in a lab or survey

Temporal example many causal chains stretch over long time scales, longitudinal studies expensive

Transportability highly unlikely that the sample you achieved is perfectly random and free of selection bias, effects unbiased within the sample, but not population of interest

# Experiments: Pros and Cons

## Pros

- Identification guaranteed

  $\Longrightarrow$ Internal validity

- Greater control over intervention

## Cons

Control over who you want to assign treatments to and how

e.g. different types of sampling

Can design the experiment to directly assess research question of interest

Ethical example cannot start a war

Physical example limited to phenomena you can examine in a lab or survey

Temporal example many causal chains stretch over long time scales, longitudinal studies expensive

Transportability highly unlikely that the sample you achieved is perfectly random and free of selection bias, effects unbiased within the sample, but not population of interest

# Experiments: Pros and Cons

## Pros

- Identification guaranteed

  $\Longrightarrow$ Internal validity

- Greater control over intervention

## Cons

- limits, limits, limits...



Control over who you want to assign treatments to and how

e.g. different types of sampling

Can design the experiment to directly assess research question of interest

Ethical example cannot start a war

Physical example limited to phenomena you can examine in a lab or survey

Temporal example many causal chains stretch over long time scales, longitudinal studies expensive

Transportability highly unlikely that the sample you achieved is perfectly random and free of selection bias, effects unbiased within the sample, but not population of interest

# Experiments: Pros and Cons

## Pros

- Identification guaranteed

  $\Longrightarrow$ Internal validity

- Greater control over intervention

## Cons

- limits, limits, limits...
  - ethical



Control over who you want to assign treatments to and how

e.g. different types of sampling

Can design the experiment to directly assess research question of interest

Ethical example cannot start a war

Physical example limited to phenomena you can examine in a lab or survey

Temporal example many causal chains stretch over long time scales, longitudinal studies expensive

Transportability highly unlikely that the sample you achieved is perfectly random and free of selection bias, effects unbiased within the sample, but not population of interest

# Experiments: Pros and Cons

## Pros

- Identification guaranteed

  $\Longrightarrow$ Internal validity

- Greater control over intervention

## Cons

- limits, limits, limits...
  - ethical
  - physical



Control over who you want to assign treatments to and how

e.g. different types of sampling

Can design the experiment to directly assess research question of interest

Ethical example cannot start a war

Physical example limited to phenomena you can examine in a lab or survey

Temporal example many causal chains stretch over long time scales, longitudinal studies expensive

Transportability highly unlikely that the sample you achieved is perfectly random and free of selection bias, effects unbiased within the sample, but not population of interest

# Experiments: Pros and Cons

## Pros

- Identification guaranteed

  $\Longrightarrow$ Internal validity

- Greater control over intervention

## Cons

- limits, limits, limits...
  - ethical
  - physical
  - temporal



Control over who you want to assign treatments to and how

e.g. different types of sampling

Can design the experiment to directly assess research question of interest

Ethical example cannot start a war

Physical example limited to phenomena you can examine in a lab or survey

Temporal example many causal chains stretch over long time scales, longitudinal studies expensive

Transportability highly unlikely that the sample you achieved is perfectly random and free of selection bias, effects unbiased within the sample, but not population of interest

# Experiments: Pros and Cons

## Pros

- Identification guaranteed

  $\Longrightarrow$ Internal validity

- Greater control over intervention

## Cons

- limits, limits, limits...
  - ethical
  - physical
  - temporal
  - external validity / transportability



Control over who you want to assign treatments to and how

e.g. different types of sampling

Can design the experiment to directly assess research question of interest

Ethical example cannot start a war

Physical example limited to phenomena you can examine in a lab or survey

Temporal example many causal chains stretch over long time scales, longitudinal studies expensive

Transportability highly unlikely that the sample you achieved is perfectly random and free of selection bias, effects unbiased within the sample, but not population of interest

# Experiments: Pros and Cons

## Pros

- Identification guaranteed

  $\Longrightarrow$ Internal validity

- Greater control over intervention

## Cons

- limits, limits, limits...
  - ethical
  - physical
  - temporal
  - external validity / transportability

- Causal mechanism still an assumption



Control over who you want to assign treatments to and how

e.g. different types of sampling

Can design the experiment to directly assess research question of interest

Ethical example cannot start a war

Physical example limited to phenomena you can examine in a lab or survey

Temporal example many causal chains stretch over long time scales, longitudinal studies expensive

Transportability highly unlikely that the sample you achieved is perfectly random and free of selection bias, effects unbiased within the sample, but not population of interest

# Observational Studies: Pros and Cons

## Pros

- Can study phenomena of interest

## Cons

# Observational Studies: Pros and Cons

## Pros

- Can study phenomena of interest
- ...in the population of interest

## Cons

# Observational Studies: Pros and Cons

## Pros

- Can study phenomena of interest
- ...in the population of interest

$\Longrightarrow$ external validity

## Cons

# Observational Studies: Pros and Cons

## Pros

- Can study phenomena of interest
- ...in the population of interest

$\Longrightarrow$ external validity

## Cons

- Identification challenged by

# Observational Studies: Pros and Cons

## Pros

- Can study phenomena of interest
- ...in the population of interest

$\Longrightarrow$ external validity

## Cons

- Identification challenged by
  - selection bias

# Observational Studies: Pros and Cons

## Pros

- Can study phenomena of interest
- ...in the population of interest

$\Longrightarrow$ external validity

## Cons

- Identification challenged by
  - selection bias
  - non-random treatment

# Observational Studies: Pros and Cons

## Pros

- Can study phenomena of interest
- ...in the population of interest

$\Longrightarrow$ external validity

## Cons

- Identification challenged by
  - selection bias
  - non-random treatment
  - data limitations

# Observational Studies: Pros and Cons

## Pros

- Can study phenomena of interest
- ...in the population of interest

$\Longrightarrow$ external validity

## Cons

- Identification challenged by
  - selection bias
  - non-random treatment
  - data limitations

# Observational Studies: Pros and Cons

## Pros

- Can study phenomena of interest
- ...in the population of interest

  $\Longrightarrow$ external validity

## Cons

- Identification challenged by
  - selection bias
  - non-random treatment
  - data limitations
- Identification may be impossible without more data or experiment

# Potential Outcomes vs Structural Causal Models

# Potential Outcomes

- Associated with Neyman (7) and Rubin (11)
- Widely adopted in social sciences and medicine
- Randomized experiment serve as its ruling paradigm

# Potential Outcomes

- Object of analysis is a unit-based response variable
  - patients
  - survey respondents
  - cities

# Potential Outcomes

- Object of analysis is a unit-based response variable
  - patients
  - survey respondents
  - cities

- Comparison between factual and counterfactual for each unit *i*

# Potential Outcomes

- Object of analysis is a unit-based response variable
  - patients
  - survey respondents
  - cities

- Comparison between factual and counterfactual for each unit $i$

- Denoted $Y_i(T_i)$

# Potential Outcomes

- Object of analysis is a unit-based response variable
  - patients
  - survey respondents
  - cities
- Comparison between factual and counterfactual for each unit $i$
- Denoted $Y_i(T_i)$
- "The value outcome $Y$ would obtain in experimental unit $i$ had treatment $T_i$ been $t$"

# Potential Outcomes

- Units: $i = 1, \ldots, N$
- "Treatment":
  - $T_i = 1$ if treated
  - $T_i = 0$ otherwise
- *Observed* outcome: $Y_i$
- Pre-treatment covariates: $X_i$
- Potential outcomes: $Y_i(1)$ and $Y_i(0)$

| **Voters** | **Contact** | **Turnout** | | **Age** | **Party ID** |
|---|---|---|---|---|---|
| $i$ | $T_i$ | $Y_i(1)$ | $Y_i(0)$ | $X_i$ | $X_i$ |
| 1 | 1 | 1 | ? | 19 | D |
| 2 | 0 | ? | 0 | 45 | D |
| 3 | 0 | ? | 1 | 36 | R |
| ⋮ | ⋮ | ⋮ | ⋮ | ⋮ | ⋮ |
| N | 1 | 0 | ? | 71 | R |

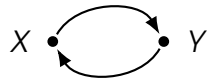
For each voter, we only observed one outcome

We cannot simultaneously observe two universes, one where individual i is given treatment and one where not, then compare

# Potential Outcomes Assumptions

Core Assumptions

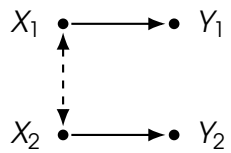
1. No simultaneity



# Potential Outcomes Assumptions

Core Assumptions

1. No simultaneity
2. No interference between units



# Potential Outcomes Assumptions

Core Assumptions

1. No simultaneity
2. No interference between units
3. Same version of treatment

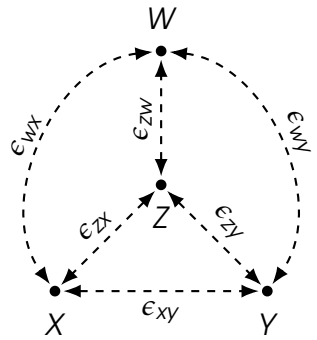
# Potential Outcomes Assumptions

Core Assumptions

1. No simultaneity
2. No interference between units
3. Same version of treatment

- Stable Unit Treatment Value Assumption (SUTVA)
- Potential violations:
  - feedback effects
  - spill-over effects
  - different treatment administration
- Observed outcome is random because treatment is random
- Multi-valued treatment: more potential outcomes for each unit

# Potential Outcomes Assumptions

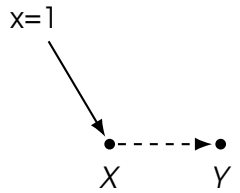



Observed data ($\mathcal{L}_1$)

Crux of PO is randomized treatment

- Causal mechanism too complex to rule out no omitted variable with certainty

# Potential Outcomes Assumptions




Crux of PO is randomized treatment

- Causal mechanism too complex to rule out no omitted variable with certainty
- Looks for "as-if" random treatments or proxy treatments
- Allows you to ignore possible confounders

Intervention data ($\mathcal{L}_2$)

# Potential Outcomes Research Designs

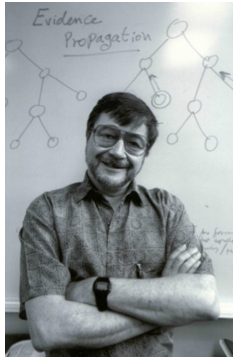
- Preferred research designs based on exogeneity assumption:
  - Instrumental Variables (IV)
  - Regression Discontinuity Design (RDD)
  - Difference-in-Difference (DiD)

- When we cannot find intervention data: matching

- Criticisms:
  - exogeneity assumption almost always untestable
  - finding guaranteed random treatments in the wild is extremely rare
  - OR the randomized "treatment" doesn't quite align with the theory we want to test

# Structural Causal Models

- Associated with Pearl (8) but many predecessors and successors
- Emerged from computer science field, but builds on:
  - structural equation models (SEM) (3)
  - potential outcomes
  - probabilistic graphical models (6, 13)
- The causal graph serves as ruling paradigm
- sometimes referred to as a "DAG" (directed acyclic graph)



# Structural Causal Models

- Based on a directed graph that displays casual relationships between variables

- Models sometimes defined as ordered triples $\langle U, V, E \rangle$:
  - Exogenous variables $U$
  - Endogenous variables $V$
  - Set of equations $E$ that defining relationships between $V$

# Structural Causal Models

- Based on a directed graph that displays casual relationships between variables

- Models sometimes defined as ordered triples $\langle U, V, E \rangle$:
  - Exogenous variables $U$
  - Endogenous variables $V$
  - Set of equations $E$ that defining relationships between $V$

- The models are probabilistic and represent a unique factorization of a joint probability distribution into conditional probabilities

# Structural Causal Models

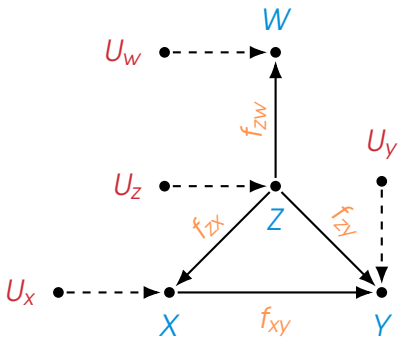
- Based on a directed graph that displays casual relationships between variables

- Models sometimes defined as ordered triples $\langle U, V, E \rangle$:
  - Exogenous variables $U$
  - Endogenous variables $V$
  - Set of equations $E$ that defining relationships between $V$

- The models are probabilistic and represent a unique factorization of a joint probability distribution into conditional probabilities

- Use *do*-calculus to achieve identification on observed data

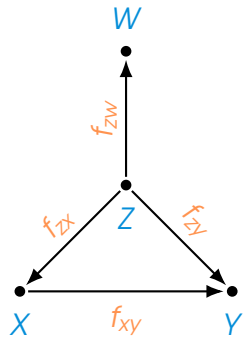
# SCM Assumptions




observed data ($\mathcal{L}_1$) + causal assumptions

- The notation seems scary, but we saw this before

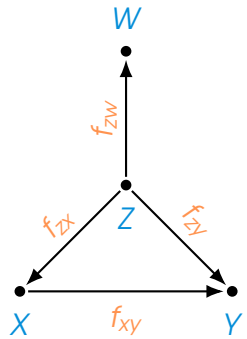
# SCM Assumptions




observed data ($\mathcal{L}_1$) + causal assumptions

- The notation seems scary, but we saw this before
  - $U$ are independent of what happens within the system
  - $V$ are dependent on system

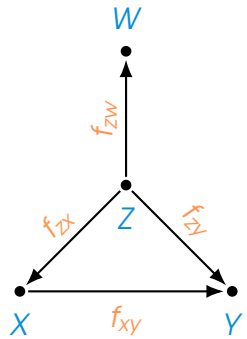
# SCM Assumptions




observed data ($\mathcal{L}_1$) + causal assumptions

- The notation seems scary, but we saw this before
  - $U$ are independent of what happens within the system
  - $V$ are dependent on system
  - $E$ represents functional relationships

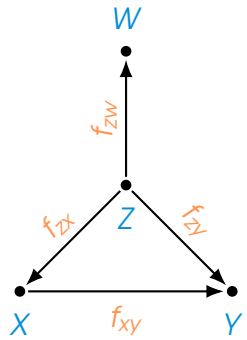
# SCM Assumptions




observed data ($\mathcal{L}_1$) + causal assumptions

- The notation seems scary, but we saw this before
  - $U$ are independent of what happens within the system
  - $V$ are dependent on system
  - $E$ represents functional relationships
- All assumptions are encoded into the graph itself

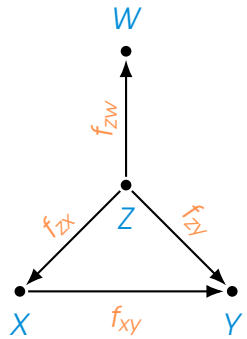
# SCM Assumptions




observed data ($\mathcal{L}_1$) + causal assumptions

- The notation seems scary, but we saw this before
  - $U$ are independent of what happens within the system
  - $V$ are dependent on system
  - $E$ represents functional relationships
- All assumptions are encoded into the graph itself
- Since the graph represents conditional probabilities, we can determine what variables to adjust for from it

# SCM Assumptions




observed data ($\mathcal{L}_1$) + causal assumptions

- The notation seems scary, but we saw this before
  - $U$ are independent of what happens within the system
  - $V$ are dependent on system
  - $E$ represents functional relationships
- All assumptions are encoded into the graph itself
- Since the graph represents conditional probabilities, we can determine what variables to adjust for from it
- Theory $\Longrightarrow$ assumptions

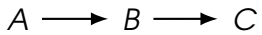
# Model Elements

All DAGs are built from three fundamental relationships

# Model Elements

Chain

- Straight line connections with arrows pointing from cause to effect

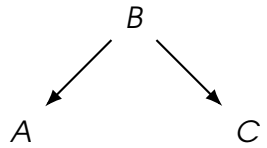



- *B* mediates effect of *A* on *C*

# Model Elements

Fork

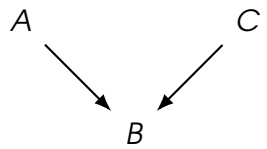
- One cause has multiple effects
- There exists spurious correlation between $A$ and $C$ due to $B$
- Eliminate by adjusting for $B$



# Model Elements

Collider

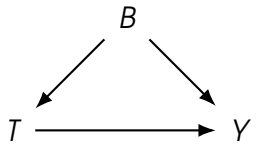
- Multiple causes affect one outcome
- Conditioning on $B$ often induces a non-causal negative relationship between $A$ and $C$
- Collider bias, wherein $B$ explains away correlation between $A$ and $C$



# Identification with DAGS

Identification is achieved via *do*-calculus

- Set of rules for determining a minimally-sufficient set of adjustment variables
- Examine all paths between treatment and outcome, control for confounders
- Not too complicated, but beyond scope of presentation




*B* confounds effect of *T* on *Y*

# SCM as a Language

- SCMs represents a *language* of causality
- All other approaches to causal inference can be encoded in a DAG (i.e. PO is subsumed by SCM)
- Can also be used to determine when and how to escape from selection bias
- Criticisms:
  - Encoding our theory into a DAG can be *hard*
  - Complex theory $\implies$ complex DAG

  $\hookrightarrow$ DAGs can become overwhelming, fast
  - do-calculus only guarantees identification if theory is correct
- Dagitty: tool that performs do-calculus for you, has `R` package too



- Critique of assumption true in all models

# Conclusion

# Conclusion

- Randomized experiments are considered a gold standard for causal inference
- But they are black boxes
- The key to causal inference on observational data is:
  - make stronger assumptions about the relationships between variables
  - Search for interventional $\mathcal{L}_2$ setups that match theory
- In SCM, we do the former and establish whether observational data is identified; if not, ask is it achievable and how?
- In PO, identifiability is guaranteed so long as we believe intervention is truly random
- Both require rigorous validation of assumptions
- Once identified, we can interpret $\beta$ as a causal effect



- All the work goes into developing good identification strategies and validating that strategy
- Once you achieve identification, there's still a question of the "correct" functional form.

# References I


- Barenboim, E. et al. (July 2020). *On Pearl's Hierarchy and the Foundations of Causal Inference*. Tech. rep. R-60. In: "Probabilistic and Causal Inference: The Works of Judea Pearl", ACM Books, in press. Causal Artificial Intelligence Lab, Columbia University.
- Buchsbaum, D. et al. (2012). "The power of possibility: causal learning, counterfactual reasoning, and pretend play". In: *Philosophical Transactions of the Royal Society B: Biological Sciences* 367.1599, pp. 2202–2212.
- Goldberger, A. (1973). "Structural equation models: An overview". In: *Structural Equation Models in the Social Sciences*. Ed. by A. Goldberger and O. Duncan. New York, NY: Seminar Press, pp. 1–18.
- Gopnik, A. (2012). "Scientific thinking in young children: Theoretical advances, empirical research, and policy implications". In: *Science* 337.6102, pp. 1623–1627.
- Gopnik, A. et al. (2004). "A theory of causal learning in children: Causal maps and bayes nets". In: *Psychological Review* 111.1, pp. 3–32.

# References II


- Lauritzen, S. (1996). *Graphical Models*. Oxford, UK: Clarendon Press.
- Neyman, Jerzy (1923). "On the application of probability theory to agricultural experiments. Essay on principles. Section 9". In: *Statistical Science* 5, pp. 465–480.
- Pearl, J. (2000). *Causality: models, reasoning, and inference*. New York, NY: Cambridge University Press.
- — (2009). "Causal inference in statistics: An overview". In: *Statistics Surveys* 3.September, pp. 96–146.
- — (2018). "Does obesity shorten life? Or is it the soda? On non-manipulable causes". In: *Journal of Causal Inference* 6.2.
- Rubin, Donald (1974). "Estimating Causal Effects of Treatments in Randomized and Nonrandomized Studies". In: *Journal of Educational Psychology* 66, pp. 688–701.

# References III


- Simon, A. (1953). "Causal ordering and identifiability". In: *Studies in Econometric Method*. Ed. by W. C. Hood and T. C. Koopmans. New York, NY: Wiley and Sons, Inc, pp. 49–74.
- Spirtes, P., C. Glymour, and R. Scheines (2000). *Causation, Prediction, and Search*. 2nd ed. Cambridge, MA: MIT Press.