

Causality

Francisco Villamil

Research Design for Social Sciences
MA Computational Social Science, UC3M
Fall 2025

Roadmap

Intro to explanation

Potential outcomes framework

Experiments

Causal models and diagrams

Back doors and front doors

Usual suspects

Paper discussion and next week

Explaining relationships

- Explanation vs. prediction

Explaining relationships

- Explanation vs. prediction
- If we want to explain, we need to know about how the real world works
 - **Causal model** (also think about **data generating process**)

Explaining relationships

- Explanation vs. prediction
- If we want to explain, we need to know about how the real world works
 - **Causal model** (also think about **data generating process**)
- We don't necessarily need that for prediction

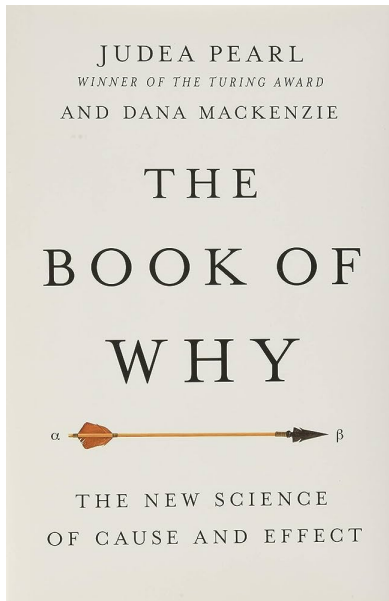
Explaining relationships

- No brute-force way into causal inference
 - Statistical/computational approaches to causality
1. Potential outcomes framework (Rubin model):
 - more present in economics, linked to experiments (but not only), formal notation and definition of causal estimands, ...

Explaining relationships

- No brute-force way into causal inference
 - Statistical/computational approaches to causality
1. Potential outcomes framework (Rubin model):
 - more present in economics, linked to experiments (but not only), formal notation and definition of causal estimands, ...
 2. Structural causal model:
 - use of DAGs, more popular in epidemiology, linked to observational data, increasingly popular in social sciences, ...

Potential outcomes vs causal models



Roadmap

Intro to explanation

Potential outcomes framework

Experiments

Causal models and diagrams

Back doors and front doors

Usual suspects

Paper discussion and next week

Explaining relationships

- Key thing: we want to know whether X actually *causes* Y
 - I.e., we want to do *causal inference*
- **Note** that this does not mean that X is the only cause of Y , but that **changing** X **alters** Y

Explaining relationships

- How could we observe causal relationships? Repeating history

Explaining relationships

- How could we observe causal relationships? Repeating history
- The 'fundamental problem of causal inference' is that we cannot
 - In other words, that for every unit of observation, we can only observe **either** $Y(X = 0)$ **or** $Y(X = 1)$

Explaining relationships

- How could we observe causal relationships? Repeating history
- The 'fundamental problem of causal inference' is that we cannot
 - In other words, that for every unit of observation, we can only observe **either** $Y(X = 0)$ **or** $Y(X = 1)$
- If we observe $Y(X = 1)$, causal inference essentially means trying to find as good an approximation to $Y(X = 0)$ as we can find
 - i.e., we want to find something that is valid as a *counterfactual*

Potential outcomes framework

- Also called Neyman–Rubin causal model
- An effect is the difference between the *actual world* and an *alternative reality* (counterfactual)
 - Causal effect of X , is $E(Y|X = 1) - E(Y|X = 0)$

Potential outcomes framework



What is the effect of smoking on life expectancy?

Potential outcomes framework

- Gary: smoker, doesn't exercise, but is vegetarian. We can wait and see how long he lives:

$$E(LExp|S = 1, G = Male, E = 0, V = 1)$$



Gary

Potential outcomes framework

- Gary: smoker, doesn't exercise, but is vegetarian. We can wait and see how long he lives:

$$E(LExp|S = 1, G = Male, E = 0, V = 1)$$

- The Q is, **what is the causal effect of smoking** on Gary?



Gary

Potential outcomes framework

- Gary: smoker, doesn't exercise, but is vegetarian. We can wait and see how long he lives:

$$E(LExp|S = 1, G = Male, E = 0, V = 1)$$

- The Q is, **what is the causal effect of smoking** on Gary?
- To know that, we need to estimate the life expectancy of an alternative Gary that is *exactly the same except* for the smoking:

$$E(LExp|S = 0, G = Male, E = 0, V = 1)$$



Gary

Potential outcomes framework

- Gary: smoker, doesn't exercise, but is vegetarian. We can wait and see how long he lives:

$$E(LExp|S = 1, G = Male, E = 0, V = 1)$$

- The Q is, **what is the causal effect of smoking** on Gary?
- To know that, we need to estimate the life expectancy of an alternative Gary that is *exactly the same except* for the smoking:

$$E(LExp|S = 0, G = Male, E = 0, V = 1)$$

- The problem is that the alternative Gary is unobservable: **missing data problem**



Gary

Potential outcomes framework

- That would be for Gary. What about the '**general**' effect of **smoking**?
- We'd need to find an alternative for every person (smoking and non-smoking), and just calculate the difference between the alternative and the reality:

$$E[LifeExp_i^1] - E[LifeExp_i^0]$$

which would be the **Average Treatment Effect** (or **ATE**)

Potential outcomes framework

- That would be for Gary. What about the **'general' effect of smoking?**
- We'd need to find an alternative for every person (smoking and non-smoking), and just calculate the difference between the alternative and the reality:

$$E[LifeExp_i^1] - E[LifeExp_i^0]$$

which would be the **Average Treatment Effect** (or **ATE**)

- Problem is we have missing data: we don't have $E[LifeExp_i^1]$ for non-smokers, and we don't have $E[LifeExp_i^0]$ for smokers

Potential outcomes framework

- Same goes with other quantities of interest we'll see:

Potential outcomes framework

- Same goes with other quantities of interest we'll see:
- **ATT**, or **average treatment effect on the treated**:

$$E[Y_i^1 | D_i = 1] - E[Y_i^0 | D_i = 1]$$

(effect of smoking among smokers)

Potential outcomes framework

- Same goes with other quantities of interest we'll see:
- **ATT**, or **average treatment effect on the treated**:

$$E[Y_i^1 | D_i = 1] - E[Y_i^0 | D_i = 1]$$

(effect of smoking among smokers)

- Or the **ATC** (or ATU), or **average treatment effect on the untreated**:

$$E[Y_i^1 | D_i = 0] - E[Y_i^0 | D_i = 0]$$

(effect of smoking among non-smokers)

Potential outcomes framework

- Same goes with other quantities of interest we'll see:
- **ATT**, or **average treatment effect on the treated**:

$$E[Y_i^1 | D_i = 1] - E[Y_i^0 | D_i = 1]$$

(effect of smoking among smokers)

- Or the **ATC** (or ATU), or **average treatment effect on the untreated**:

$$E[Y_i^1 | D_i = 0] - E[Y_i^0 | D_i = 0]$$

(effect of smoking among non-smokers)

- When is $ATT \neq ATC$? (Non-linearity, sampling bias)

PO example

	Confounder	Treatment	Unobservable		Realized	
	Age	Treated	Potential outcomes		ICE or δ_i^*	Outcome
ID	Z_i	X_i	Y_i^1	Y_i^0	$Y_i^1 - Y_i^0$	Y_i
1	Old	1	80	60	20	80
2	Old	1	75	70	5	75
3	Old	1	85	80	5	85
4	Old	0	70	60	10	60
5	Young	1	75	70	5	75
6	Young	0	80	80	0	80
7	Young	0	90	100	-10	100
8	Young	0	85	80	5	80

* ICE = individual causal effect

<https://www.andrewheiss.com/blog/2024/03/21/demystifying-ate-att-atu/>

PO example

	Confounder	Treatment	Unobservable		Realized
	Age	Treated	Potential outcomes		Outcome
ID	Z_i	X_i	Y_i^1	Y_i^0	Y_i
1	Old	1	80	60	80
2	Old	1	75	70	75
3	Old	1	85	80	85
4	Old	0	70	60	60
5	Young	1	75	70	75
6	Young	0	80	80	80
7	Young	0	90	100	100
8	Young	0	85	80	80

* ICE = individual causal effect

<https://www.andrewheiss.com/blog/2024/03/21/demystifying-ate-att-atu/>

PO example

	Confounder	Treatment	Unobservable		Realized	
	Age	Treated	Potential outcomes		ICE or δ_i^*	Outcome
ID	Z_i	X_i	Y_i^1	Y_i^0	$Y_i^1 - Y_i^0$	Y_i
1	Old	1	80	60	20	80
2	Old	1	75	70	5	75
3	Old	1	85	80	5	85
4	Old	0	70	60	10	60
5	Young	1	75	70	5	75
6	Young	0	80	80	0	80
7	Young	0	90	100	-10	100
8	Young	0	85	80	5	80

* ICE = individual causal effect

$$ATE = mean(20, 5, 5, 10, 5, 0, -10, 5) = 5$$

PO example

ID	Confounder	Treatment	Unobservable			Realized
	Age	Treated	Potential outcomes		ICE or δ_i^*	Outcome
	Z_i	X_i	Y_i^1	Y_i^0	$Y_i^1 - Y_i^0$	Y_i
1	Old	1	80	60	20	80
2	Old	1	75	70	5	75
3	Old	1	85	80	5	85
4	Old	0	70	60	10	60
5	Young	1	75	70	5	75
6	Young	0	80	80	0	80
7	Young	0	90	100	-10	100
8	Young	0	85	80	5	80

* ICE = individual causal effect

$$ATT = \text{mean}(20, 5, 5, 5) = 8.75$$

PO example

ID	Confounder	Treatment	Unobservable			Realized
	Age	Treated	Potential outcomes		ICE or δ_i^*	Outcome
	Z_i	X_i	Y_i^1	Y_i^0	$Y_i^1 - Y_i^0$	Y_i
1	Old	1	80	60	20	80
2	Old	1	75	70	5	75
3	Old	1	85	80	5	85
4	Old	0	70	60	10	60
5	Young	1	75	70	5	75
6	Young	0	80	80	0	80
7	Young	0	90	100	-10	100
8	Young	0	85	80	5	80

* ICE = individual causal effect

$$ATU = \text{mean}(10, 0, -10, 5) = 1.25$$

PO example

- *weighted mean of ATT and ATU = ATE*
- But we don't have both Y_0 and Y_1
- Two ways of looking into this observationally
 - (assuming **age** as the **only** confounder)

PO example

```
1 df = data.frame(  
2   age = rep(c("0", "Y"), each = 4),  
3   treatment = c(1, 1, 1, 0, 1, 0, 0, 0),  
4   outcome = c(80, 75, 85, 60, 75, 80, 100, 80))  
5  
6 m1 = lm(outcome ~ treatment, data = df)  
7 m2 = lm(outcome ~ treatment + factor(age), data = df)
```


PO example

	(1)	(2)
(Intercept)	80.000	71.875
	(6.016)	(9.305)
treatment	-1.250	4.167
	(8.509)	(9.610)
factor(age)Y		10.833
		(9.610)
Num.Obs.	8	8

Sometimes ATT is more useful in practice

Table 4: Different types of causal effects that can be found as ATEs and ATTs



ATE	ATT
Effect of mosquito bed nets for everyone in the country	Effect of mosquito bed nets for people who use bed nets
Effect of military service for typical applicants to the military*	Effect of military service for typical soldiers
Effect of a job training program on all residents in a state	Effect of a job training program on everyone who used it
Effect of a new cancer medicine on everyone in the world	Effect of a new cancer medicine on people with cancer
* Example via this Cross Validated response	

<https://www.andrewheiss.com/blog/2024/03/21/demystifying-ate-att-atu/>

Roadmap

Intro to explanation

Potential outcomes framework

Experiments

Causal models and diagrams

Back doors and front doors

Usual suspects

Paper discussion and next week

Estimating causal effects

- **So how do we solve this missing data problem?**
 - intervening in treatment assignment through randomization

Estimating causal effects

- **So how do we solve this missing data problem?**
 - intervening in treatment assignment through randomization
- 'No causation without manipulation' (Rubin)
- Potential outcomes framework initially developed for *experimental data*: randomized controlled trials are the gold-standard in approximating the alternative reality (counterfactual)
- But there are also problems or limitations:
 - issues in experimental design (next)
 - more importantly: not all experiments are feasible or ethical

Randomization issues

- Obviously, the basic of any experiment is that **treatment assignment is random**
- It's not frequent, but could happen that this randomization is not well done
- Also it might not let us detect the effect, and having statistical issues, especially when using *block* randomization, or *unit* vs. *cluster* randomization
- Also could be an issue when doing *block randomization*

SUTVA

- SUTVA stands for **Stable Unit Treatment Value Assumption**, and it is a key assumption in experimental designs
 - It is basically that the outcome in one unit is **not** affected by treatment assignment in other units
- Diffusion effects among subjects?
 - One solution could be to think about *unit of observation*
- (This problem is also discussed in causal inference with observational data)

Attrition

- **Attrition** is just the case when ‘participants leave the study’
- More generally, when some of the units in the experiment do not complete it
- The key question is, to what extent is this biasing the results?

External validity

- To what extent can we **generalize the results of an experiment**?
 - i.e., how much do we really learn with this experiment?
- This is a more general issue that we will also discuss with observational-data studies, but perhaps very relevant for experiments because of the setting it usually takes place
- Example: media exposure studies (or Guess *et al* 2023)
 - **Treatment validity?** Discuss concept of **bundled treatment**
 - **Outcome validity?** survey (hypothetical) questions vs behavioral outcomes, relationship with original Q

Treatment compliance

- Are all units assigned to treatment really exposed to it?
- In clinical trials, e.g. do they take the pill or spit it?
- How would this look like in an experiment when you pay (treated) individuals to watch TV or use Facebook?
- Concept of **intention-to-treat** (ITT) analyses and the **complier average causal effect** or **local average treatment effect** (LATE)

Roadmap

Intro to explanation

Potential outcomes framework

Experiments

Causal models and diagrams

Back doors and front doors

Usual suspects

Paper discussion and next week

So how to we approximate Y^0 ?

- Experiments are fine, but often not possible

So how to we approximate Y^0 ?

- Experiments are fine, but often not possible
- How do we do this with **observational data**?
 - We need to 'build' a counterfactual

So how to we approximate Y^0 ?

- Experiments are fine, but often not possible
- How do we do this with **observational data**?
 - We need to 'build' a counterfactual
- Basic idea: we come up with a strategy where the only variation we analyze is (according to us) *due* to the independent variable (cause) we are interested in
 - **Read it again:** it is actually the same idea as in the experimental method, where we use randomization to achieve that

So how to we approximate Y^0 ?

- Experiments are fine, but often not possible
- How do we do this with **observational data**?
 - We need to 'build' a counterfactual
- Basic idea: we come up with a strategy where the only variation we analyze is (according to us) *due* to the independent variable (cause) we are interested in
 - **Read it again:** it is actually the same idea as in the experimental method, where we use randomization to achieve that
- But in order to do that, we need to be clear about the **causal model** that is causing Y , so we know what we need to control for
 - And we're gonna use **causal diagrams** for that

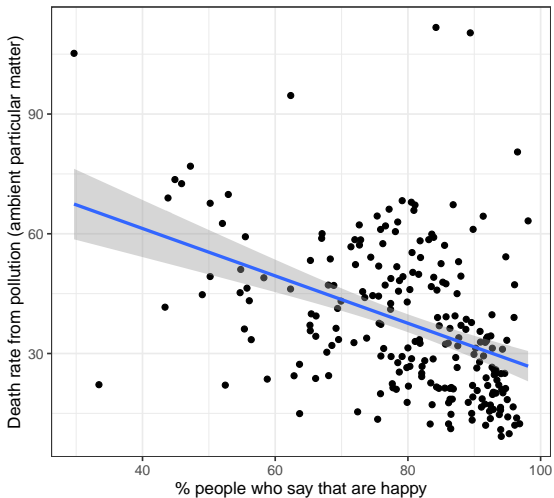
Example

- Let's say we want to know whether a cleaner environment makes people happier

Example

Environmental policies cause happiness!

(country-year data from ourworldindata.org)



Example

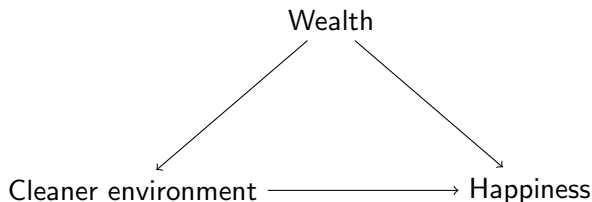
- Remember that our problem (the 'fundamental problem of causal inference' etc) is that we can observe these countries but not their **alternatives** without treatment (or with different levels of the treatment)
- So to infer causal models we need to **isolate the variation due to our treatment**, how?
 - Controlling the non-causal variation
- To get to this, we need to have a **causal model** to know what's 'contaminating' our outcome (so we know what we should be controlling for)

Our causal model

Cleaner environment \longrightarrow Happiness

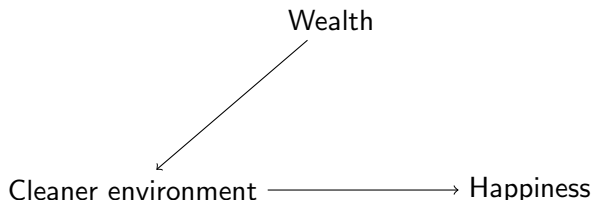
- This is our initial causal model: having a cleaner environment makes people happier (because they like looking into a blue sky without smog), and that's it. We do not have to control for anything nor do anything else.

Our causal model



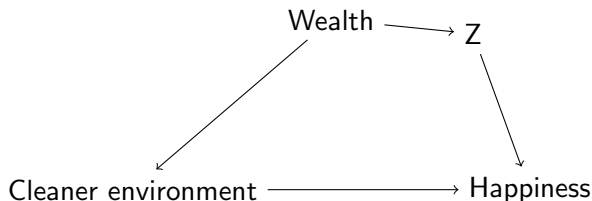
- Wait, but maybe it's about money, isn't it? Actually, wealthier countries tend to have cleaner environments and, at the same time, money causes happiness. **We need to control for wealth.**

Our causal model



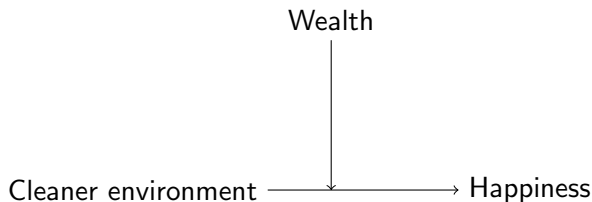
- Or perhaps it is not that money increases happiness *per se*, but that it does so through other **mediators**: wealth allows countries to focus on environment, which increases happiness. **Again, no need to control.** As long as this is the **only** mediator.

Our causal model



- We are happy with that model, but we're still missing something. Say we believe that money does not have any direct causal effect, but it does causes some other things (labour conditions, cultural offer, ... let's call them Z) and these, in turn, have an effect on happiness. **We need to control for wealth and all Z .**

Our causal model



- (Another thing would be if money **moderates** the relationship between environmental policies and happiness: spending resources to take care of our environment makes you happier only if you have enough money – this is an special case, we could talk about heterogenous effects)

Basics of causal inference

- So to come up with an strategy, we need to understand what's going on in terms of the data generating process

Basics of causal inference

- So to come up with an strategy, we need to understand what's going on in terms of the data generating process
 - This applies from the most basic strategy (add controls) to the more complicated ones (e.g. evaluating DiD or RDD)

Basics of causal inference

- So to come up with an strategy, we need to understand what's going on in terms of the data generating process
 - This applies from the most basic strategy (add controls) to the more complicated ones (e.g. evaluating DiD or RDD)
- Once we have that, we can **identify** an effect (in other words: isolating the causal variation from other sources of variation we are not interested in)

Causal models, mechanisms, and DAGs

- We will use Directed Acyclic Graphs (DAG) (causal diagrams), a graph where we link **variables** (nodes) with **causal effects** (arrows)

Causal models, mechanisms, and DAGs

- We will use Directed Acyclic Graphs (DAG) (causal diagrams), a graph where we link **variables** (nodes) with **causal effects** (arrows)

A few things:

Causal models, mechanisms, and DAGs

- We will use **Directed Acyclic Graphs (DAG)** (causal diagrams), a graph where we link **variables** (nodes) with **causal effects** (arrows)

A few things:

- Only one-directional causality (*acyclic*)
 - if you have feedback cycles, write multiple nodes for t_1 , t_2

Causal models, mechanisms, and DAGs

- We will use **Directed Acyclic Graphs (DAG)** (causal diagrams), a graph where we link **variables** (nodes) with **causal effects** (arrows)

A few things:

- Only one-directional causality (*acyclic*)
 - if you have feedback cycles, write multiple nodes for t_1 , t_2
- Sometimes: solid lines → observed, dashed → unobserved (U)

Causal models, mechanisms, and DAGs

- We will use **Directed Acyclic Graphs (DAG)** (causal diagrams), a graph where we link **variables** (nodes) with **causal effects** (arrows)

A few things:

- Only one-directional causality (*acyclic*)
 - if you have feedback cycles, write multiple nodes for t_1 , t_2
- Sometimes: solid lines → observed, dashed → unobserved (U)
- Treatment usually written as D (and Y the outcome)

Causal models, mechanisms, and DAGs

- We will use **Directed Acyclic Graphs (DAG)** (causal diagrams), a graph where we link **variables** (nodes) with **causal effects** (arrows)

A few things:

- Only one-directional causality (*acyclic*)
 - if you have feedback cycles, write multiple nodes for t_1 , t_2
- Sometimes: solid lines → observed, dashed → unobserved (U)
- Treatment usually written as D (and Y the outcome)
- Combine variables (usually B for background, or U for unknown)

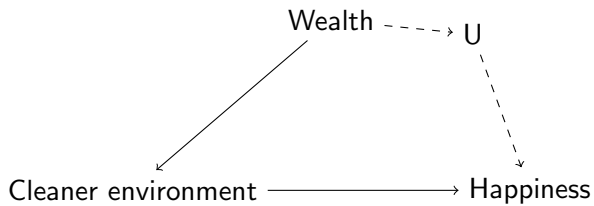
Causal models, mechanisms, and DAGs

- We will use **Directed Acyclic Graphs (DAG)** (causal diagrams), a graph where we link **variables** (nodes) with **causal effects** (arrows)

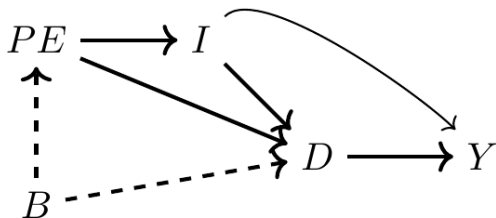
A few things:

- Only one-directional causality (*acyclic*)
 - if you have feedback cycles, write multiple nodes for t_1 , t_2
- Sometimes: solid lines → observed, dashed → unobserved (U)
- Treatment usually written as D (and Y the outcome)
- Combine variables (usually B for background, or U for unknown)
- No arrow means *no effect*, explicitly

This is a DAG



This is another DAG



- Y = earnings (outcome)
- D = college education (treatment)
- PE = parental education
- I = family income
- B = unobserved background factors (intelligence, abilities, home, etc)

from https://mixtape.scunning.com/03-directed_acyclical_graphs

Causal models, mechanisms, and DAGs

We use DAGs for mainly two things related to causal inference:

- Drawing up the **mechanism** that explains the outcome
- Come up with the strategy we need to **identify** the causal effect

Causal models, mechanisms, and DAGs

We use DAGs for mainly two things related to causal inference:

- Drawing up the **mechanism** that explains the outcome
- Come up with the strategy we need to **identify** the causal effect
 - The difference between the mechanism and the causal model is that not all intermediate steps are relevant for causal inference, even though they do work as an additional check

Mediation and moderation

- We usually find more than one variable present in a mechanism
- Two typical variables: mediator and moderator
- **Mediation**: a third variable explains the causal relationship between two variables (e.g. flu infection > immune reaction > fever)
- **Moderation**: a third variable changes the effect of one variable on another (e.g. how age changes the immune reaction)

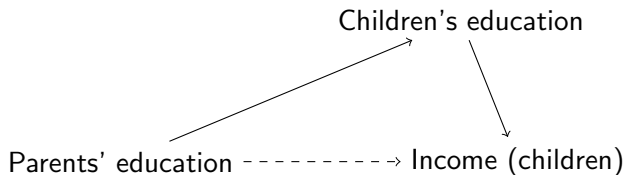
Example: income inequalities

Parents' education —————> Income (children)

- Say we want to explain income inequality, and we find that people whose parents went to university earn, on average, more. This would be the basic causal model.

(*Note:* in this case I use solid lines for direct effects and dashed lines for indirect effects, kind of)

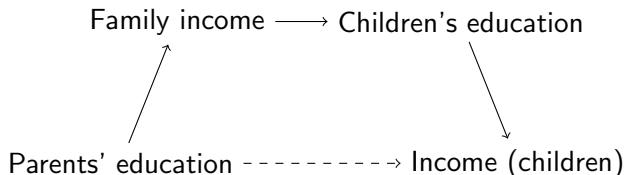
Example: income inequalities



- But *why* it is so? Someone comes and says: “It’s because parents with higher education are more likely to send their children to university and help them get through.”

(Note: in this case I use solid lines for direct effects and dashed lines for indirect effects, kind of)

Example: income inequalities



- And then someone comes and says: “It’s not only that, it’s money. Parents with higher education are richer and are able to send their kids to private schools and universities.”

(Note: in this case I use solid lines for direct effects and dashed lines for indirect effects, kind of)

DAGs and mechanisms

- We can draw the process we're trying to study

DAGs and mechanisms

- We can draw the process we're trying to study
- Main idea: focus on the *data generating process*

DAGs and mechanisms

- We can draw the process we're trying to study
- Main idea: focus on the *data generating process*
 - What had to happen for kids with university-degree parents to get richer?

DAGs and mechanisms

- We can draw the process we're trying to study
- Main idea: focus on the *data generating process*
 - What had to happen for kids with university-degree parents to get richer?
 - In other words, what mechanisms were in play behind what we see in the data?

DAGs and mechanisms

- We can draw the process we're trying to study
- Main idea: focus on the *data generating process*
 - What had to happen for kids with university-degree parents to get richer?
 - In other words, what mechanisms were in play behind what we see in the data?
- This matters when choosing our empirical strategy:

DAGs and mechanisms

- We can draw the process we're trying to study
- Main idea: focus on the *data generating process*
 - What had to happen for kids with university-degree parents to get richer?
 - In other words, what mechanisms were in play behind what we see in the data?
- This matters when choosing our empirical strategy:
 - Main identification strategy

DAGs and mechanisms

- We can draw the process we're trying to study
- Main idea: focus on the *data generating process*
 - What had to happen for kids with university-degree parents to get richer?
 - In other words, what mechanisms were in play behind what we see in the data?
- This matters when choosing our empirical strategy:
 - Main identification strategy
 - Additional checks or implications (testing the mechanism, heterogeneous effects, etc)

Roadmap

Intro to explanation

Potential outcomes framework

Experiments

Causal models and diagrams

Back doors and front doors

Usual suspects

Paper discussion and next week

Front doors and back doors

- We are interested in *identifying* the effect of pollution on happiness: that's out **front door**

Front doors and back doors

- We are interested in *identifying* the effect of pollution on happiness: that's out **front door**
- To do so, you have to make sure that the 'water flows' through this front door, and not through other 'pipe', or back door

Front doors and back doors

- We are interested in *identifying* the effect of pollution on happiness: that's out **front door**
- To do so, you have to make sure that the 'water flows' through this front door, and not through other 'pipe', or back door
- Pollution \rightarrow Happiness is our front door
- Pollution \leftarrow Wealth \rightarrow Happiness is a back door

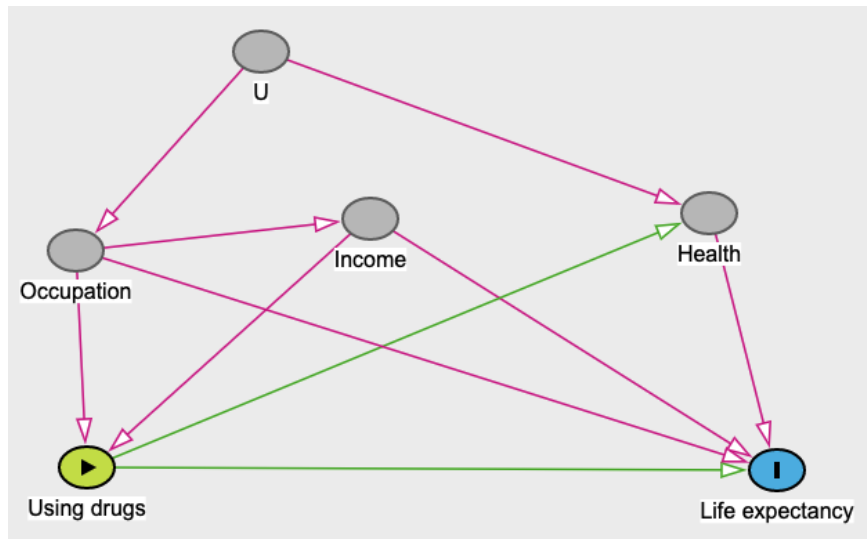
Front doors and back doors

- We are interested in *identifying* the effect of pollution on happiness: that's out **front door**
- To do so, you have to make sure that the 'water flows' through this front door, and not through other 'pipe', or back door
- Pollution \rightarrow Happiness is our front door
- Pollution \leftarrow Wealth \rightarrow Happiness is a back door
- How do you close it? Just controlling for Wealth

Front doors and back doors

- There are essentially two ways to do causal inference:
 1. Close all back doors and leave only the front door open
That's where DAGs help to identify these variables
 2. Using some other method where only the front door is opened
(Finding and analysing exogenous variation)

Front doors and back doors



Front doors and back doors

- $\text{Drugs} > \text{LifeExp}$
- $\text{Drugs} > \text{Health} > \text{LifeExp}$
- $\text{Drugs} < \text{Income} > \text{LifeExp}$
- $\text{Drugs} < \text{Occup} > \text{LifeExp}$
- $\text{Drugs} < \text{Occup} > \text{Income} > \text{LifeExp}$
- $\text{Drugs} < \text{Occup} < U > \text{Income} > \text{LifeExp}$
- $\text{Drugs} < \text{Occup} < U > \text{Health} > \text{LifeExp}$

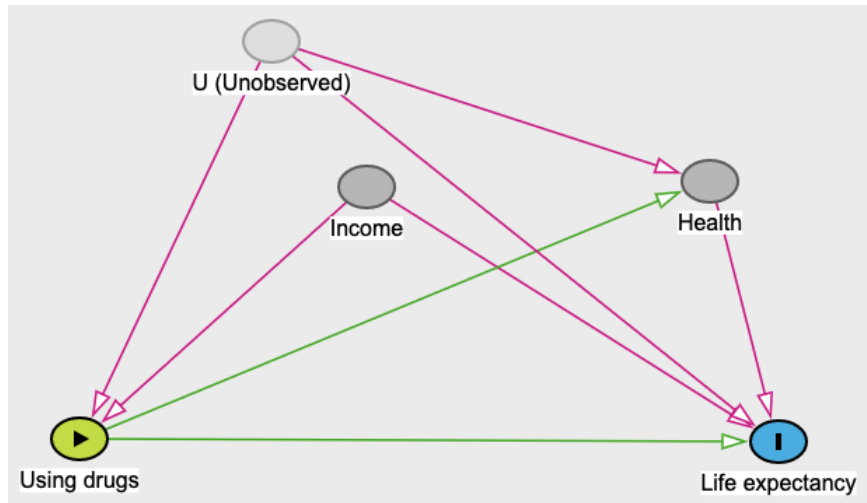
Front doors and back doors

- $\text{Drugs} > \text{LifeExp}$
- $\text{Drugs} > \text{Health} > \text{LifeExp}$
- $\text{Drugs} < \text{Income} > \text{LifeExp}$
- $\text{Drugs} < \text{Occup} > \text{LifeExp}$
- $\text{Drugs} < \text{Occup} > \text{Income} > \text{LifeExp}$
- $\text{Drugs} < \text{Occup} < U > \text{Income} > \text{LifeExp}$
- $\text{Drugs} < \text{Occup} < U > \text{Health} > \text{LifeExp}$

Front doors and back doors

- We just need to control for one of the variables in the path of a back door to close that path
- In this example, it would be enough to control for income and occupation
- This is the **back door criterion**

Front doors and back doors



Front doors and back doors

- $\text{Drugs} > \text{LifeExp}$
- $\text{Drugs} > \text{Health} > \text{LifeExp}$
- $\text{Drugs} < \text{Income} > \text{LifeExp}$
- $\text{Drugs} < U > \text{Income} > \text{LifeExp}$
- $\text{Drugs} < U > \text{Health} > \text{LifeExp}$
- $\text{Drugs} < U > \text{LifeExp}$

Front doors and back doors

- **What if we control for health?**
- We would be blocking part of the causal 'water flow' from drugs to life expectancy
- That's part of the mechanism: imagine that drugs has a direct effect, e.g. higher probability of dying on an accident, and an indirect effect through its effect on health
- (Unless you want to calculate the *direct effect*)

Front doors and back doors

- **What if we control for health?**
- We would be blocking part of the causal 'water flow' from drugs to life expectancy
- That's part of the mechanism: imagine that drugs has a direct effect, e.g. higher probability of dying on an accident, and an indirect effect through its effect on health
- (Unless you want to calculate the *direct effect*)
- We would have to control for **U**, which would close all other paths

Front doors and back doors

- $\text{Drugs} > \text{LifeExp}$
 - $\text{Drugs} > \text{Health} > \text{LifeExp}$
 - $\text{Drugs} < \text{Income} > \text{LifeExp}$
 - $\text{Drugs} < U > \text{Income} > \text{LifeExp}$
 - $\text{Drugs} < U > \text{Health} > \text{LifeExp}$
 - $\text{Drugs} < U > \text{LifeExp}$ (!)
-
- Problem? So?

Off topic: Controlling

- How does alcohol consumption affect health?
- Imagine we take data from a group of people:

```
1 df = data.frame(  
2   # In this group of people, one-third are rich  
3   rich = rbinom(500, 1, 0.3)) %>%  
4   # Rich people have 3x more money to buy whiskey  
5   mutate(whiskey = 3*rich + runif(500, 0, 4)) %>%  
6   # Health risk is worse if you drink more whiskey, but  
7     rich people have better health overall  
8   mutate(risk = -2*rich + .3*whiskey + rnorm(500, 2))
```


Off topic: Controlling

```
1 cor(df$whiskey, df$risk)
2 [1] -0.1150553
```

Off topic: Controlling

- Controlling for rich = look at the variation **not** explained by rich
- i.e., take the group prediction out (mean of whiskey/risk for rich or non-rich)

```
1 df = df %>%  
2   group_by(rich) %>%  
3   mutate(whiskey_resid = whiskey - mean(whiskey),  
4          risk_resid = risk - mean(risk)) %>%  
5   ungroup()
```

Off topic: Controlling

- The *true* model we created:

$$\text{risk} = -2 * \text{rich} + .3 * \text{whiskey} + \text{error}$$

```
1 cor(df$whiskey_resid, df$risk_resid)
2 [1] 0.3242735
```

Roadmap

Intro to explanation

Potential outcomes framework

Experiments

Causal models and diagrams

Back doors and front doors

Usual suspects

Paper discussion and next week

Usual suspects

- Confounding
- Reverse causality
- Bidirectional causation
- Selection bias
- Collider bias
- Post-treatment bias

Confounding

- Typical example: as the number of pirates in the oceans decreased, global mean temperature increased. Does it mean the disappearance of pirates is causing global warming?

Confounding

- Typical example: as the number of pirates in the oceans decreased, global mean temperature increased. Does it mean the disappearance of pirates is causing global warming?
- No, both are caused by the industrial revolution or technological development

Confounding

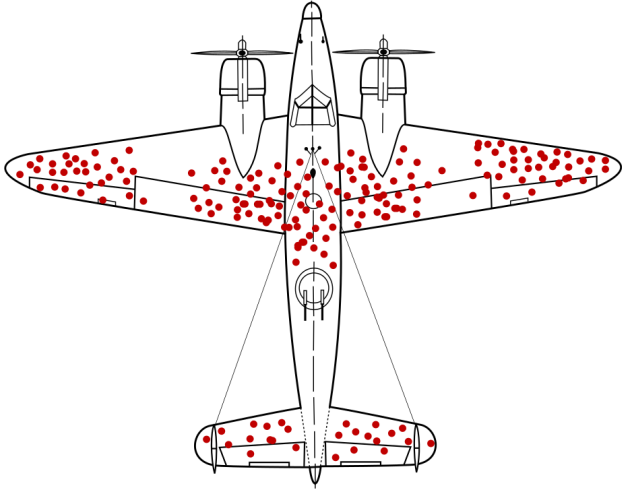
- Typical example: as the number of pirates in the oceans decreased, global mean temperature increased. Does it mean the disappearance of pirates is causing global warming?
- No, both are caused by the industrial revolution or technological development
- Months when people eat more ice-creams, also more people drown in the beach. Ice-creams causing drownings?

Reverse causality

- Many examples where correlations we think imply a particular causal effect might be explained by its reverse: Violent videogames making teenagers violence? Drug use causes psychological problems?
- “Hospitals make people sick.” If you collect data on illness development, you might find that people fare worse if they go to the hospital. Obviously, it's a case of reverse causality: being sick causes going to the hospital.

Bidirectional causation

- (there are endogenous cycles, *not the same as* reverse causality)
- Political values and voting: the way you think makes you vote in a particular way, but the way you vote can also affect the way you think (group influence, cognitive processes, etc)
- Can be closely related to selection bias: imagine we go to Madrid Rio and we measure if people doing exercises are more likely to be overweight than those lying around
- We probably don't find any result. Does it mean exercise does not decrease overweight? No, it's probably bidirectional causation: overweight makes people more likely to exercise, and exercise reduces overweight



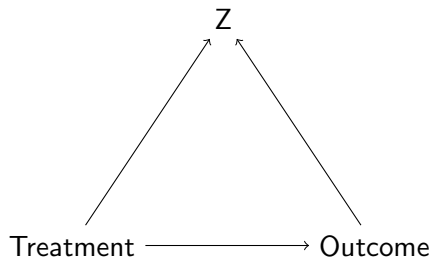
Selection bias

- Our observations are not representative
- Famous example from World War II airplanes
- Many examples: advice from successful CEOs, ex-heroin addicts more likely to do sports, etc
- Why?
 - Sampling
 - Attrition (\approx survivorship bias)
 - etc

Selection bias in causal inference

- Selection bias in statistics: sampling issue
- Quite different in causality: we're dealing with selection into treatment
- Remember example from HIV treatments studies

Collider bias



Collider bias

- Are smart people weirdos?
- We have 1,000 people, with **randomly** distributed intelligence and social skills

```
1  set.seed(7891234)
2  df = data.frame(
3    intelligence = rnorm(1000, mean = 5, sd = 1.5),
4    social_skills = rnorm(1000, mean = 5, sd = 1.5))
5
```

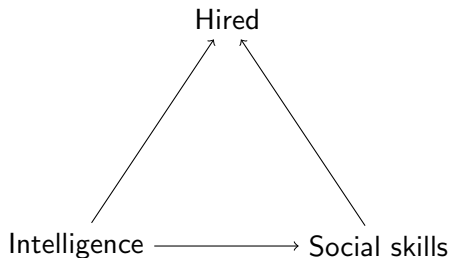
Collider bias

- No correlation

```
1 > cor(df$intelligence, df$social_skills)
2 [1] -0.03982793
3
```


Collider bias

- Now imagine that we have another variable, the probability of being hired in a company, which is caused by both intelligence and social skills:



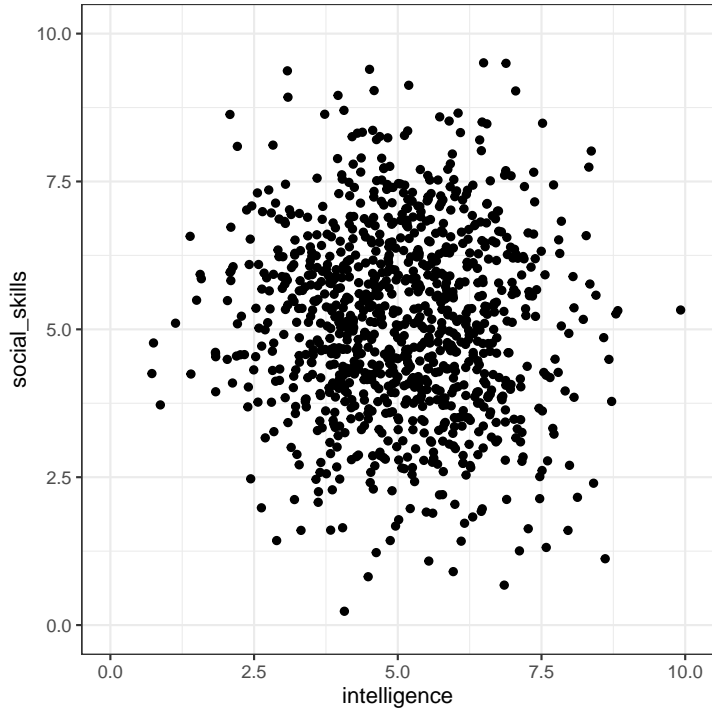
Collider bias

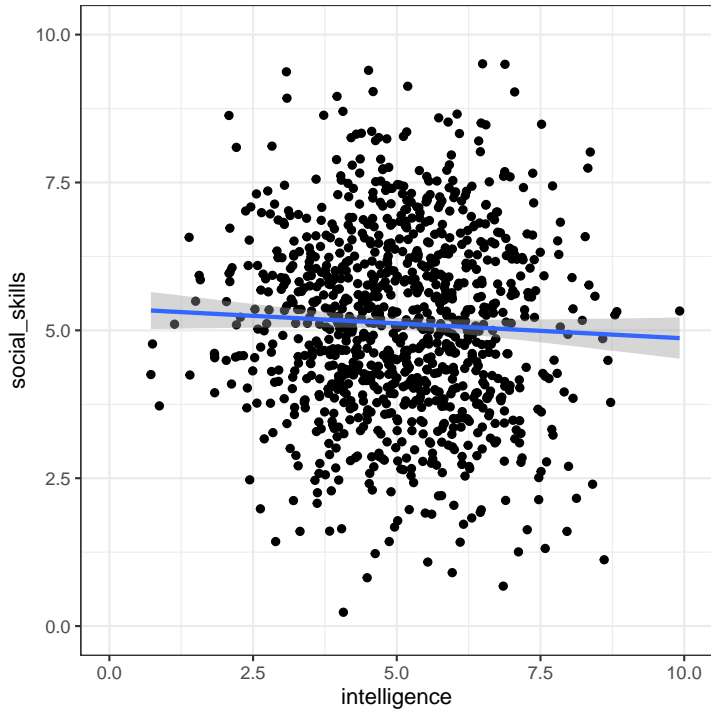
- This company values intelligence a bit more, but takes both into account. And hires the 25% best candidates:

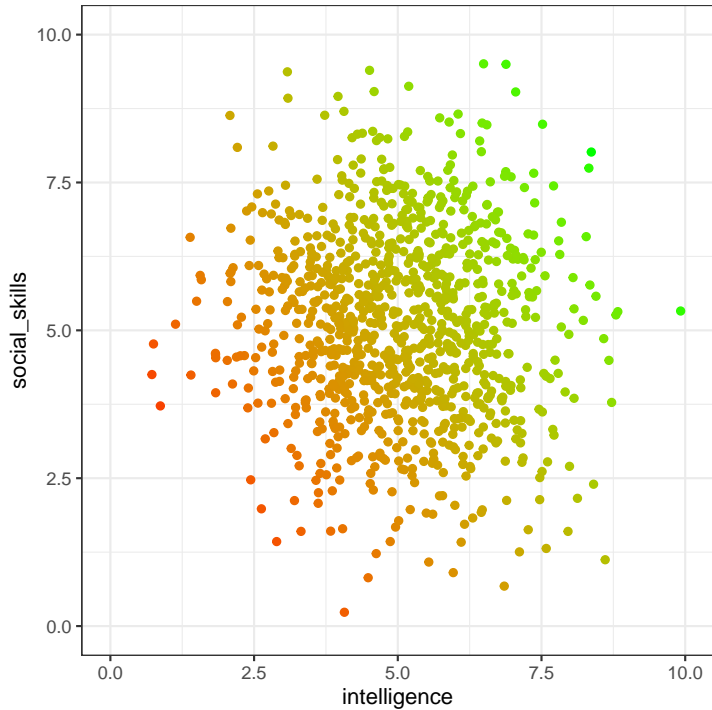
```
1 df = df %>%  
2   mutate(  
3     points = 3 * intelligence + 2 * social_skills) %>%  
4   mutate(hired_binary = ifelse(points > quantile(points  
    , 0.75), 1, 0))
```

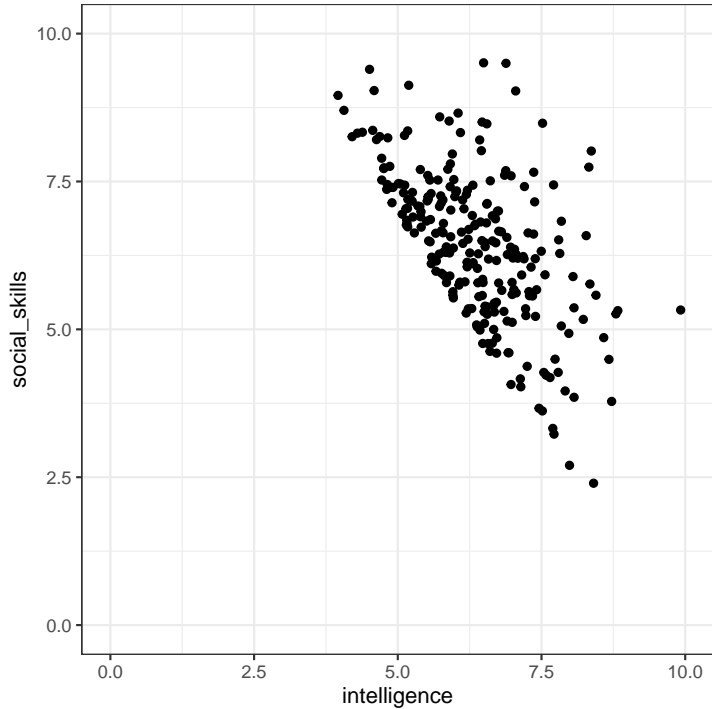
	Model 1	Model 2
(Intercept)	5.333***	
	(0.184)	
intelligence	-0.044	
	(0.035)	
hired_binary		
Num.Obs.	1000	
R2	0.002	

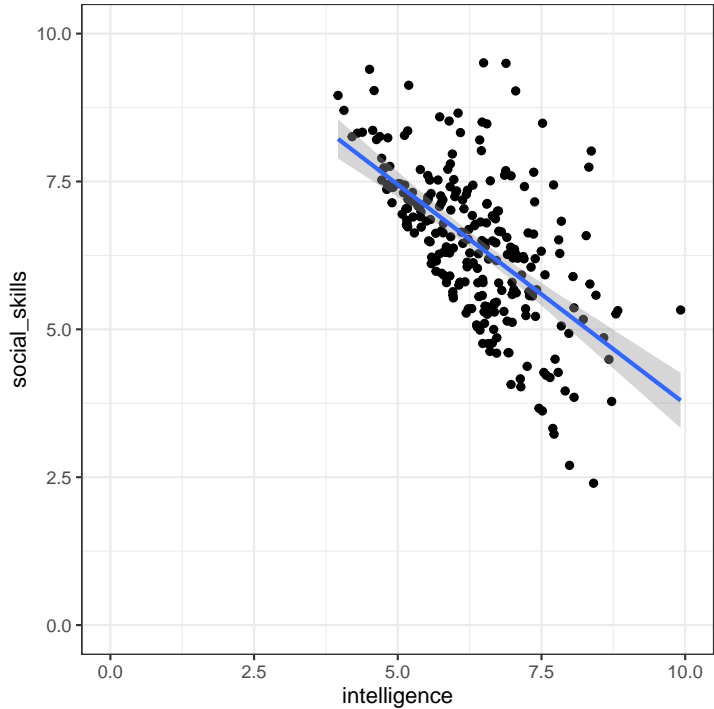
	Model 1	Model 2
(Intercept)	5.333***	6.820***
	(0.184)	(0.161)
intelligence	-0.044	-0.466***
	(0.035)	(0.033)
hired_binary		2.576***
		(0.110)
Num.Obs.	1000	1000
R2	0.002	0.356









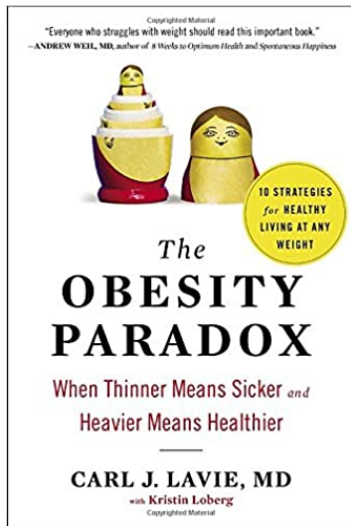


Collider bias

- A collider bias **opens** a path when you control for the variable

Collider bias

- Another example in life sciences where we can only use observational data, the 'obesity paradox'
 - Obesity reduces mortality among patients with some chronic heart diseases
- Collider bias?
 - <https://catalogofbias.org/biases/collider-bias/>



Collider bias

- Animated: <https://nickchk.com/causalgraphs.html>

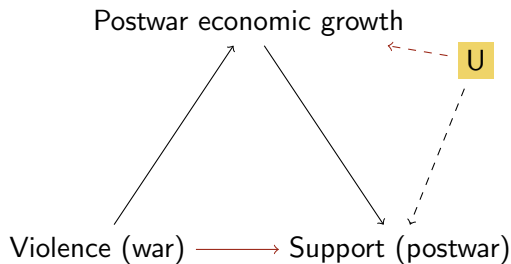
Post-treatment bias (collider again)

- We want to know whether suffering violence during a civil wars makes people more or less likely to support certain authorities decades after the war

Post-treatment bias (collider again)

- We want to know whether suffering violence during a civil wars makes people more or less likely to support certain authorities decades after the war
- And we say: well, the country develop economically after the war, so maybe it makes sense to control for local increase in GDPpc, because it will also affect support

Nice try, but...



Recap: what should **not** be controlled for

1. **Front-door paths**

- Blocking some of the effect through a mediator variable
- (There are almost always mediator variables, so you could potentially just eliminate all the effect you're trying to identify)

2. **Collider bias**

- Opens a new, uncontrolled-for path
- Sometimes you might be inadvertently controlling for a collider because of *selection* issues
- Extra care with post-treatment bias

Roadmap

Intro to explanation

Potential outcomes framework

Experiments

Causal models and diagrams

Back doors and front doors

Usual suspects

Paper discussion and next week


SOCIAL MEDIA

How do social media feed algorithms affect attitudes and behavior in an election campaign?

Andrew M. Guess^{1*}, Neil Malhotra², Jennifer Pan³, Pablo Barberá⁴, Hunt Allcott⁵, Taylor Brown⁴, Adriana Crespo-Tenorio⁴, Drew Dimmery^{4,6}, Deen Freelon⁷, Matthew Gentzkow⁸, Sandra González-Bailón⁹, Edward Kennedy¹⁰, Young Mie Kim¹¹, David Lazer¹², Devra Moehler⁴, Brendan Nyhan¹³, Carlos Velasco Rivera⁴, Jaime Settle¹⁴, Daniel Robert Thomas⁴, Emily Thorson¹⁵, Rebekah Tromble¹⁶, Arjun Wilkins⁴, Magdalena Wojcieszak^{17,18}, Beixian Xiong⁴, Chad Kiewiet de Jonge⁴, Annie Franco⁴, Winter Mason⁴, Natalie Jomini Stroud¹⁹, Joshua A. Tucker²⁰

We investigated the effects of Facebook's and Instagram's feed algorithms during the 2020 US election. We assigned a sample of consenting users to reverse-chronologically-ordered feeds instead of the default algorithms. Moving users out of algorithmic feeds substantially decreased the time they spent on the platforms and their activity. The chronological feed also affected exposure to content: The amount of political and untrustworthy content they saw increased on both platforms, the amount of content classified as uncivil or containing slur words they saw decreased on Facebook, and the amount of content from moderate friends and sources with ideologically mixed audiences they saw increased on Facebook. Despite these substantial changes in users' on-platform experience, the chronological feed did not significantly alter levels of issue polarization, affective polarization, political knowledge, or other key attitudes during the 3-month study period.

Do TJ policies cause backlash? Evidence from street name changes in Spain

Research and Politics
October-December 2021: 1–7
© The Author(s) 2021
Article reuse guidelines:
sagepub.com/journals-permissions
DOI: 10.1177/20531680211058550
journals.sagepub.com/home/rap


Francisco Villamil¹  and Laia Balcells²

Abstract

Memories of old conflicts often shape domestic politics long after these conflicts end. Contemporary debates about past civil wars and/or repressive regimes in different parts of the world suggest that these are sensitive topics that might increase political polarization, particularly when transitional justice policies are implemented and political parties mobilize discontentment with such policies. One such policy recently debated in Spain is removing public symbols linked to a past civil war and subsequent authoritarian regime (i.e., Francoism). However, the empirical evidence on its impact is still limited. This article attempts to fill this gap by examining the political consequences of street renaming. Using a difference-in-differences approach, we show that the removal of Francoist street names has contributed to an increase of electoral support for a new far-right party, Vox, mainly at the expense of a traditional right-wing conservative party, PP. Our results suggest that revisiting the past can cause a backlash among those ideologically aligned with the perpetrator, and that some political parties can capitalize on this.

Keywords

Transitional justice, voting, conflict memories, Spain

Next week (Oct 3)

- Think about the overall question (“Do TJ policies cause backlash?”) and about how much we can learn
 - Identification strategy?
 - Treatment validity?
 - Outcome?
 - Generalizing results? (across time and space)
 - Measurement, theory-empirics link, ...

(check Appendix!)