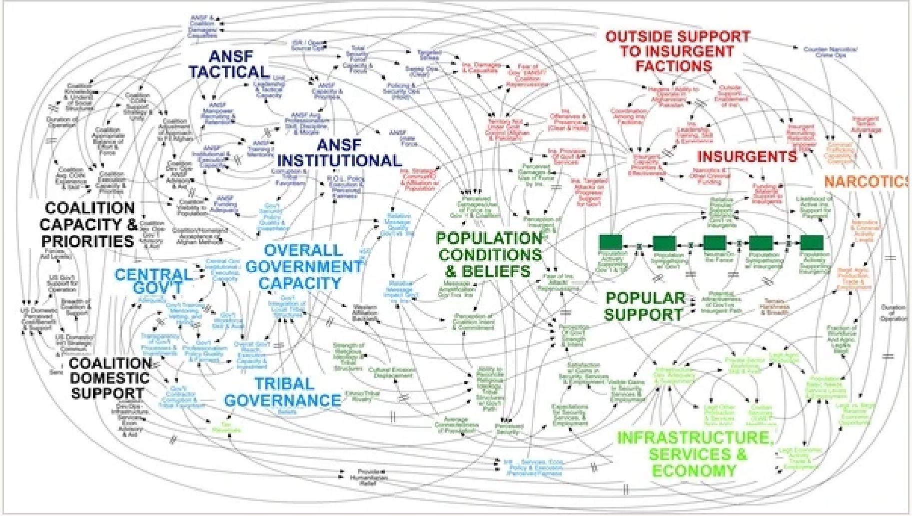
APPLIED CAUSAL GRAPHS

Foundations and new directions for causal graphs

William Lowe

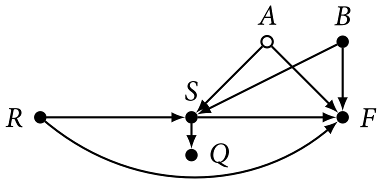
Hertie School Data Science Lab

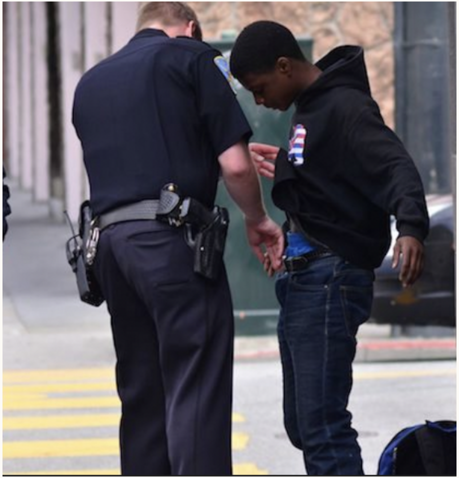
2024-04-23



“When we understand that slide, we’ll have won the war” (General Stanley McChrystal, 2009)

THE ADMINISTRATIVE DATA DAG



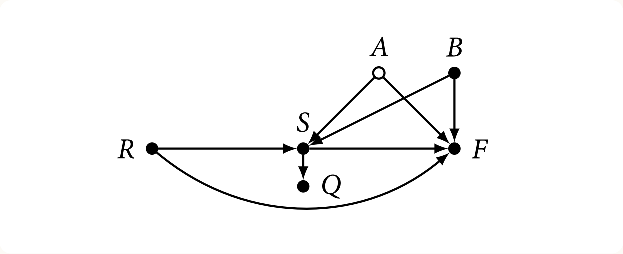


REALIZATION: RACE BIAS IN POLICING

- Race (R), Behaviour (B), Animosity (A)
- Stopped (S), Quantified in a report (Q)
- Force applied (F)

Knox, Lowe and Mummolo, 2020

THE ADMINISTRATIVE DATA DAG, AGAIN



Realization: Bicycle helmets and accidents

- Helmet (R), Behaviour (B), Traffic (A)
- Accident (S), Quantified in a hospital report (Q)
- Head injury (F)

e.g. Fernandez et al. 2024

"If I have an accident I'll wish I had been wearing an SUV"

“NOT ALL DISCIPLINES LOVE CAUSAL GRAPHS”

THERE ARE TWO TYPES OF CAUSAL PEOPLE

Row people

THERE ARE TWO TYPES OF CAUSAL PEOPLE

Row people

ESTIMAND FOCUSED

Focus on averages of contrasts over populations, e.g.

$$\begin{aligned} \text{ATE} &= \mathbb{E}[\Delta Y] \\ &= \mathbb{E}[Y^{(X=1)} - Y^{(X=0)}] \\ &= \mathbb{E}[Y^{(X=1)}] - \mathbb{E}[Y^{(X=0)}] \end{aligned}$$

THERE ARE TWO TYPES OF CAUSAL PEOPLE

Row people

ESTIMAND FOCUSED

Focus on averages of contrasts over populations, e.g.

$$\begin{aligned} \text{ATE} &= \mathbb{E}[\Delta Y] \\ &= \mathbb{E}[Y^{(X=1)} - Y^{(X=0)}] \\ &= \mathbb{E}[Y^{(X=1)}] - \mathbb{E}[Y^{(X=0)}] \end{aligned}$$

EXTENSIONAL FORMALISM

Lay out all the possibilities as potential outcomes and assert *independence relations* among them, e.g.

$$X \perp\!\!\!\perp (Y^{(X=1)}, Y^{(X=0)})$$

X	G	$Y^{(X=1)}$	$Y^{(X=0)}$	ΔY
1	M	Y_1		?
1	F	Y_2	??	?
$\vdots$		$\vdots$		?
1	F	Y_k		?
0	M		Y_{k+1}	?
0	F		Y_{k+2}	?
0	F	??	Y_{k+3}	?
0	M		Y_N	?

Identification: when you can remove the counterfactual quantities

THERE ARE TWO TYPES OF CAUSAL PEOPLE

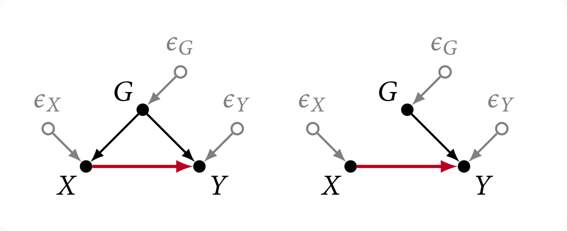
Column people

THERE ARE TWO TYPES OF CAUSAL PEOPLE

Column people

MECHANISM FOCUSED

Isolate the association from paths of interest



pre-intervention   post-intervention

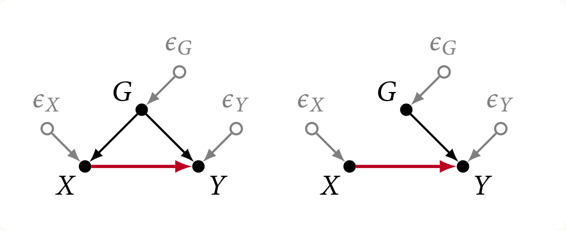
e.g. what would $X \longrightarrow Y$ be if $G \perp\!\!\!\perp X$?

THERE ARE TWO TYPES OF CAUSAL PEOPLE

Column people

MECHANISM FOCUSED

Isolate the association from paths of interest



pre-intervention post-intervention

e.g. what would $X \longrightarrow Y$ be if $G \perp\!\!\!\perp X$?

INTENSIONAL FORMALISM

Pre-intervention

$$X = f_X(G, \epsilon_X)$$
$$X = f_X(\epsilon_X)$$
$$Y = f_Y(X, G, \epsilon_Y)$$

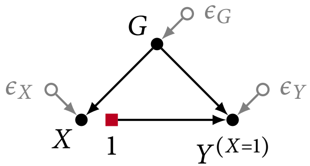
Post-intervention ($X \Rightarrow 1$)

$$X = 1$$
$$G = f_G(\epsilon_G)$$
$$Y = f_Y(1, G, \epsilon_Y)$$

Identification: when you can learn about *post* world from *pre* world

OK, MAYBE THREE

Single world intervention graph

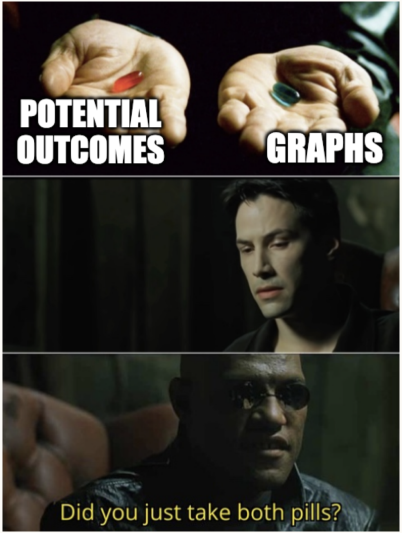


From the SWIG

$$X \not\perp\!\!\!\perp Y^{(X=1)} \qquad (G \text{ is a common cause})$$

$$X \perp\!\!\!\perp Y^{(X=1)} \mid G \qquad (\text{d-separation})$$

Richardson and Robins (2013)



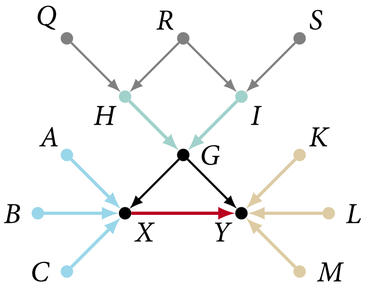
GRAPH FOUNDATIONS

THE WORLD MAY BE A COMPLICATED PLACE

Often we assume probability and wonder about causation. Let's do the reverse

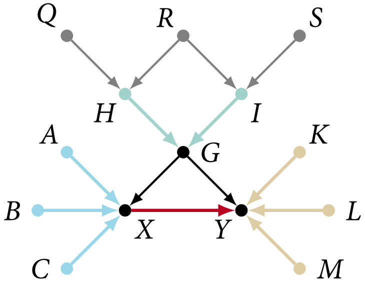
THE WORLD MAY BE A COMPLICATED PLACE

Often we assume probability and wonder about causation. Let's do the reverse



THE WORLD MAY BE A COMPLICATED PLACE

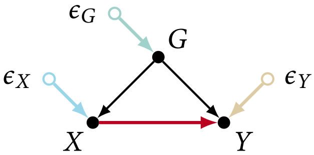
Often we assume probability and wonder about causation. Let's do the reverse



Focus

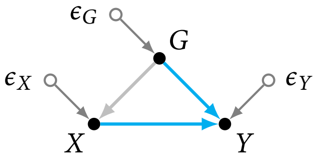
If we're interested in $X \longrightarrow Y$, not all of it is relevant

Bundle ancestors with no common causes into 'noise' terms $\epsilon_X, \epsilon_G, \epsilon_Y$



$[\epsilon_X, \epsilon_G, \epsilon_Y]$ picks out individuals

…BUT IT IS MADE UP OF MECHANISMS

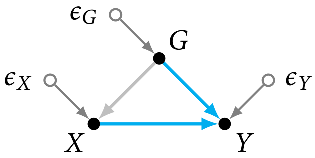


is a summary of *structural equations*

$$
\begin{aligned}
G &= f_G(\epsilon_G) \\
X &= f_X(G, \epsilon_X) \\
Y &= \mathbf{f}_Y(X, G, \epsilon_Y)
\end{aligned}
$$

Nature knows the details. The graph just shows her *joints*

…BUT IT IS MADE UP OF MECHANISMS



is a summary of *structural equations*

$$G = f_G(\epsilon_G)$$
$$X = f_X(G, \epsilon_X)$$
$$Y = \mathbf{f}_Y(X, G, \epsilon_Y)$$

Nature knows the details. The graph just shows her *joints*

Observable implications

Structural equations plus $[\epsilon_X, \epsilon_G, \epsilon_Y]$

> induce a joint probability distribution

> with a *causal* decomposition

$$P(Y, X, G) = P(G)P(X \mid G)P(Y \mid G, X)$$

> that reflects behaviour under interventions

> and has a (sometimes) distinctive *conditional independence structure*

> that connects it to *data*


CONDITIONAL INDEPENDENCE
DAG

CONNECTING PROBABILITIES AND GRAPHS

$$\text{independence} \iff \text{d-separation}$$

Definition 2.4.1 **(*d*-separation)** *A path p is blocked by a set of nodes Z if and only if*

1. *p contains a chain of nodes* $A \to B \to C$ *or a fork* $A \leftarrow B \to C$ *such that the middle node B is in Z (i.e., B is conditioned on), or*
2. *p contains a collider* $A \to B \leftarrow C$ *such that the collision node B is not in Z, and no descendant of B is in Z.*

CONNECTING PROBABILITIES AND GRAPHS

$$\text{independence} \iff \text{d-separation}$$

Definition 2.4.1 **(*d*-separation)** *A path p is blocked by a set of nodes Z if and only if*

1. *p contains a chain of nodes $A \rightarrow B \rightarrow C$ or a fork $A \leftarrow B \rightarrow C$ such that the middle node B is in Z (i.e., B is conditioned on), or*
2. *p contains a collider $A \rightarrow B \leftarrow C$ such that the collision node B is not in Z, and no descendant of B is in Z.*

Causal Markov Condition

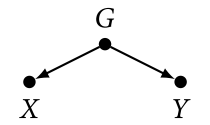
> All variables that are d-separated in the graph are independent of each other

Faithfulness

> All variables that are independent of each other are d-separated in the graph

PATHS OF FREE ASSOCIATION

PATHS OF FREE ASSOCIATION



common cause, fork

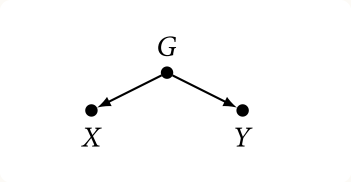
Implications

$$X \not\perp\!\!\!\perp Y$$

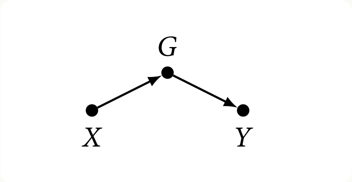
$$X \perp\!\!\!\perp Y \mid G$$

Conditioning on G removes association

PATHS OF FREE ASSOCIATION



common cause, fork



mediator

Implications

$$X \not\perp\!\!\!\perp Y$$

$$X \perp\!\!\!\perp Y \mid G$$

Conditioning on G removes association

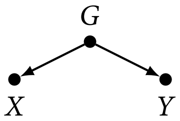
Implications

$$X \not\perp\!\!\!\perp Y$$

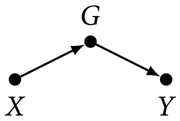
$$X \perp\!\!\!\perp Y \mid G$$

Same here. Which might be unfortunate

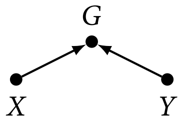
PATHS OF FREE ASSOCIATION



common cause, fork



mediator



collider, common effect

Implications

$$X \not\!\perp\!\!\!\perp Y$$
$$X \perp\!\!\!\perp Y \mid G$$

Conditioning on G removes association

Implications

$$X \not\!\perp\!\!\!\perp Y$$
$$X \perp\!\!\!\perp Y \mid G$$

Same here. Which might be unfortunate

Implications

$$X \perp\!\!\!\perp Y$$
$$X \not\!\perp\!\!\!\perp Y \mid G$$

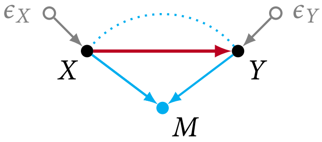
Terrible, non-intuitive, and *all the good stuff is here*

THE GOOD STUFF

COLLIDER BIAS

A REGRESSION MODEL, FROM LOW ORBIT

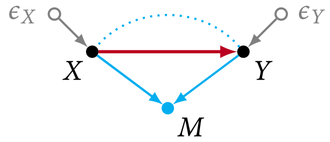
HOW NOT TO LEARN ABOUT $X \longrightarrow Y$

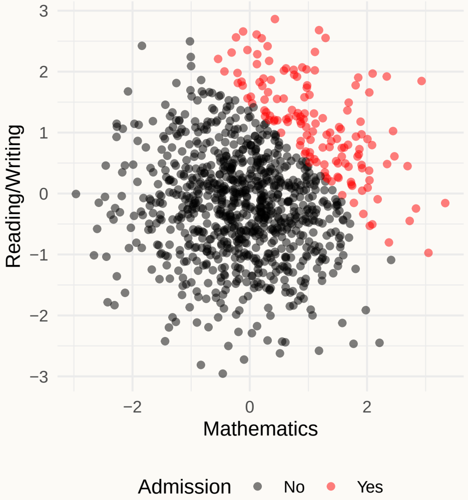


COLLIDER BIAS

A REGRESSION MODEL, FROM LOW ORBIT

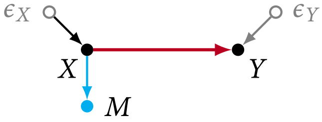
HOW NOT TO LEARN ABOUT $X \longrightarrow Y$



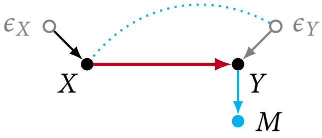


INVISIBLE COLLIDER BIAS

NO COLLIDER BIAS

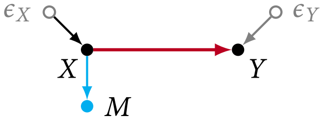


COLLIDER BIAS

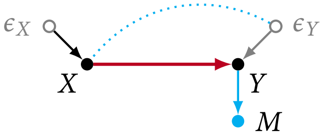


INVISIBLE COLLIDER BIAS

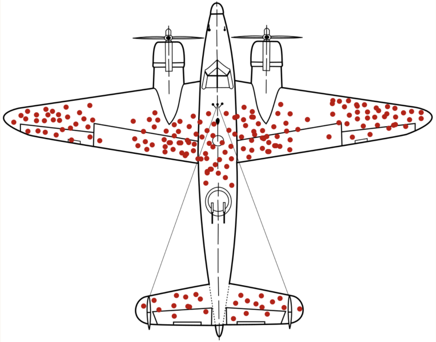
NO COLLIDER BIAS



COLLIDER BIAS



IN PLANE LANGUAGE: SURVIVAL BIAS



Not actually Wald's problem or plane, but hey… (Mangel & Samaniego, 1984)

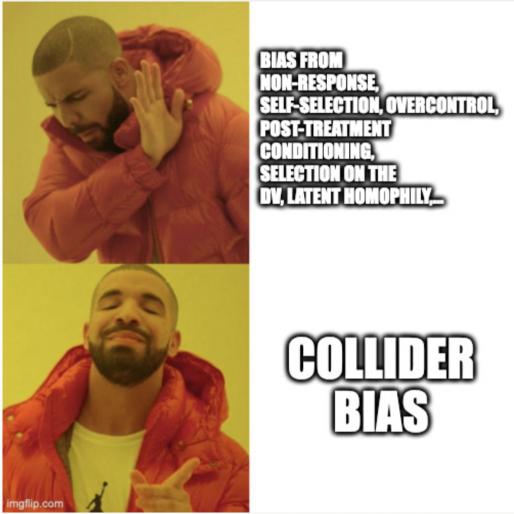
ALL COLLIDER BIAS ALL THE TIME

No really, all of it

> Elwert and Winship (2014)
> Hernán et al. (2004)

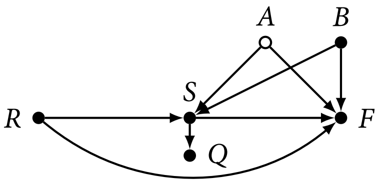
USUALLY IT'S THE PROBLEM…

- Non-response
- overcontrol
- attrition
- selection on the dependent variable
- survival bias
- latent homophily



COLLIDER BIAS AND ‘BIAS’

I FOUGHT THE LAW

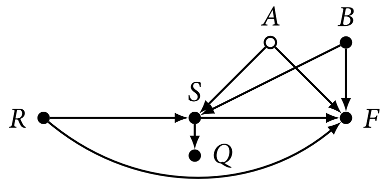


Abstract

This paper explores racial differences in police use of force. On non-lethal uses of force, blacks and Hispanics are more than fifty percent more likely to experience some form of force in interactions with police. Adding controls that account for important context and civilian behavior reduces, but cannot fully explain, these disparities. On the most extreme use of force – officer-involved shootings – we find no racial differences in either the raw data or when contextual factors are taken into account. We argue that the patterns in the data are consistent with a model in which police officers are utility maximizers, a fraction of which have a preference for discrimination, who incur relatively high expected costs of officer-involved shootings.

R. Fryer (2018) 'An empirical analysis of racial differences in police use of force'

I FOUGHT THE LAW



Abstract

This paper explores racial differences in police use of force. On non-lethal uses of force, blacks and Hispanics are more than fifty percent more likely to experience some form of force in interactions with police. Adding controls that account for important context and civilian behavior reduces, but cannot fully explain, these disparities. On the most extreme use of force – officer-involved shootings – we find no racial differences in either the raw data or when contextual factors are taken into account. We argue that the patterns in the data are consistent with a model in which police officers are utility maximizers, a fraction of which have a preference for discrimination, who incur relatively high expected costs of officer-involved shootings.

R. Fryer (2018) 'An empirical analysis of racial differences in police use of force'

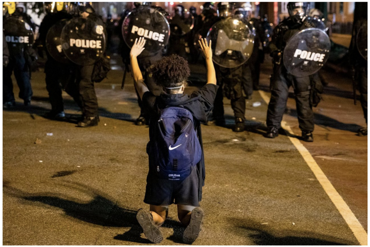
OPINION | COMMENTARY

The Myth of Systemic Police Racism

Hold officers accountable who use excessive force. But there's no evidence of widespread racial bias.

By Heather Mac Donald
June 2, 2020 1:44 pm ET

PRINT TEXT 2,534



MOST POPULAR NEWS

1. U.S. Withholds Sanctions on a Very Close Putin Associate: His Reputed Girlfriend

2. U.S. Drone Startups See an Opening in Ukraine

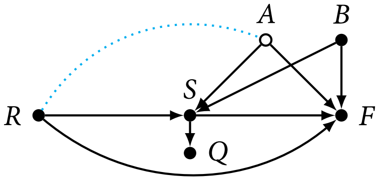
3. Twitter and Musk Are in Discussions to Strike a Deal

4. Saudi Royals Are Selling Homes, Yachts and Art as Crown Prince Cuts Income

5. Teaching Your Old Car New Tech Tricks


H. Mac Donald (2020) Wall Street Journal

ACAB (ALL COLLIDERS ARE BIASING)?



Abstract

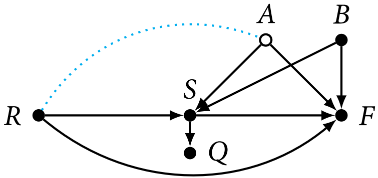
This paper explores racial differences in police use of force. On non-lethal uses of force, blacks and Hispanics are more than fifty percent more likely to experience some form of force in interactions with police. Adding controls that account for important context and civilian behavior reduces, but cannot fully explain, these disparities. On the most extreme use of force – officer-involved shootings – we find no racial differences in either the raw data or when contextual factors are taken into account. We argue that the patterns in the data are consistent with a model in which police officers are utility maximizers, a fraction of which have a preference for discrimination, who incur relatively high expected costs of officer-involved shootings.

R. Fryer (2018) 'An empirical analysis of racial differences in police use of force'

I GOT NINETY NINE PROBLEMS...

- This is a *mediation* problem
- But on a *subset* of the population because Q is a selection node
- Police records are *implicitly conditioned* on S
- Collider bias between *R*, *A*, *B*
- Estimands are rather unclear

ACAB (ALL COLLIDERS ARE BIASING)?



Abstract

This paper explores racial differences in police use of force. On non-lethal uses of force, blacks and Hispanics are more than fifty percent more likely to experience some form of force in interactions with police. Adding controls that account for important context and civilian behavior reduces, but cannot fully explain, these disparities. On the most extreme use of force – officer-involved shootings – we find no racial differences in either the raw data or when contextual factors are taken into account. We argue that the patterns in the data are consistent with a model in which police officers are utility maximizers, a fraction of which have a preference for discrimination, who incur relatively high expected costs of officer-involved shootings.

R. Fryer (2018) 'An empirical analysis of racial differences in police use of force'

I GOT NINETY NINE PROBLEMS...

- This is a *mediation* problem
- But on a *subset* of the population because Q is a selection node
- Police records are *implicitly conditioned* on S
- Collider bias between R, A, B
- Estimands are rather unclear

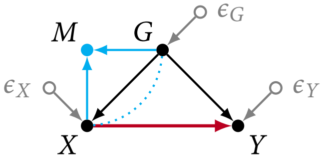
AND THOSE WERE JUST SIX OF THEM

Equal observed rates of force by race imply strongly *biased policing*

Knox, Lowe and Mummolo, 2020

COLLIDER BIAS FOR GOOD

COLLIDER BIAS FOR GOOD: MATCHING



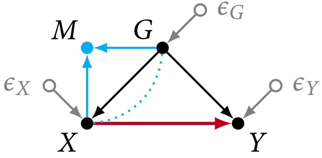
THE MATCHING MECHANISM

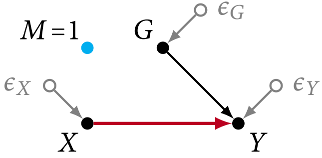
$$M = \mathbf{f}(X, G)$$

Exact one-to-one matching for the ATT:

- For every $X_i = 1$ look for an $X_j = 0$ with $G_j = G_i$
- If you find one, set M_i and M_j to 1, else 0

COLLIDER BIAS FOR GOOD: MATCHING





The matching mechanism

$$M = \mathbf{f}(X, G)$$

Exact one-to-one matching for the ATT:

- For every $X_i = 1$ look for an $X_j = 0$ with $G_j = G_i$
- If you find one, set M_i and M_j to 1, else 0

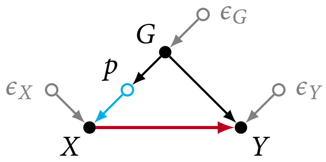
The matching process

Condition on M by removing all cases where $M = 0$

$$G \perp\!\!\!\perp X \mid M = 1$$

See Mansournia et al. (2013) for the case-control version

COLLIDER BIAS FOR GOOD: PROPENSITY SCORES



Known propensity scores

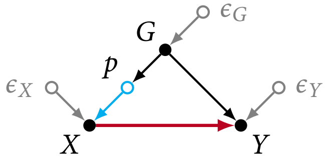
Condition on them to close

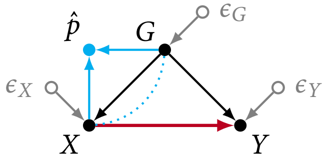
$$X \longleftarrow G \longrightarrow Y$$

"p is a balancing score" just means

$$G \perp\!\!\!\perp X \mid p$$

COLLIDER BIAS FOR GOOD: PROPENSITY SCORES





Known propensity scores

Condition on them to close

$$X \longleftarrow G \longrightarrow Y$$

"p is a balancing score" just means

$$G \perp\!\!\!\perp X \mid p$$

Estimated propensity scores

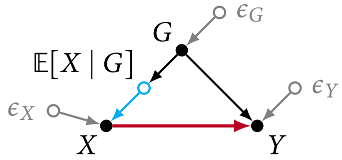
The propensity score estimator

$$\hat{p} = \mathbf{f}(X, G)$$

Condition on $\hat{p}$ for collider bias that cancels

$$X \longleftarrow G$$

COLLIDER BIAS FOR GOOD: MUNDLAK DEVICE

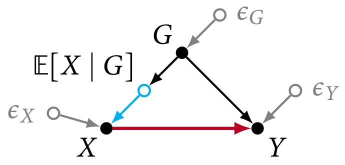


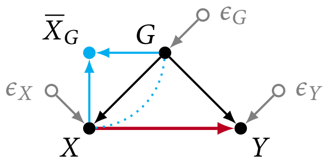
FRISCH-WAUGH-LOVELL

Intuition:

- Removing G's influence on X isolates ϵ_X
- General, but path blocking depends on functional details

COLLIDER BIAS FOR GOOD: MUNDLAK DEVICE





FRISCH-WAUGH-LOVELL

Intuition:

- Removing G's influence on X isolates ϵ_X
- General, but path blocking depends on functional details

MUNDLAK DEVICE

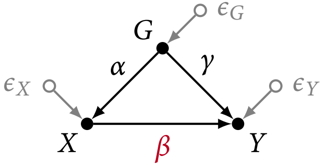
When G is a group, it moves $\mathbb{E}[X \mid G] \approx \overline{X}_G$

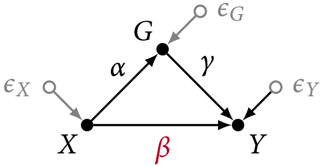
- Mundlak: Safely give G a *random effect*
- BW: Estimate *contextual effects* from

$$X = \underbrace{(X - \overline{X}_G)}_{\epsilon_X} + \overline{X}_G$$

BUT ENOUGH OF THE GOOD NEWS

CANCEL CULTURE





In a linear system, $\alpha\gamma - \beta = 0$ means $X \perp\!\!\!\perp Y$

MEASUREMENT FAILURE

Whenever a government seeks to rely on a previously observed statistical regularity for control purposes, that regularity will collapse

(Goodhart, 1981)

The more any quantitative social indicator is used for social decision-making, the more subject it will be to corruption pressures and the more apt it will be to distort and corrupt the social processes it is intended to monitor.

(Campbell, 1979)

MEASUREMENT FAILURE

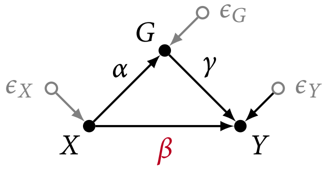
Whenever a government seeks to rely on a previously observed statistical regularity for control purposes, that regularity will collapse

(Goodhart, 1981)

The more any quantitative social indicator is used for social decision-making, the more subject it will be to corruption pressures and the more apt it will be to distort and corrupt the social processes it is intended to monitor.

(Campbell, 1979)

Causal measurement models

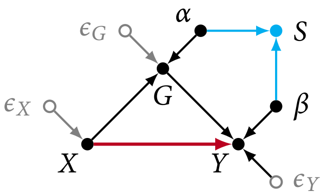


- X ability, Y exam results
- G test preparation services

$P(Y \mid X)$ is an *item response function*

with differential item functioning *at best*

CONTROL MANUFACTURES UNFAITHFULNESS



A parametric example from Shalizi (2021)

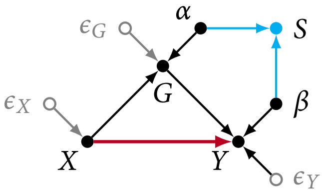
$$X = \epsilon_X$$
$$G = \alpha_0 + X\alpha_X + \epsilon_G$$
$$Y = (X - G) + \epsilon_Y$$

(So $\beta = 1$ and $\gamma = -1$)

Milton Friedman's thermostat

- X is outdoor temperature
- Y is indoor temperature
- G is the effect of the central heating system
- Y^* is the desired indoor temperature

CONTROL MANUFACTURES UNFAITHFULNESS



A parametric example from Shalizi (2021)

$$X = \epsilon_X$$
$$G = \alpha_0 + X\alpha_X + \epsilon_G$$
$$Y = (X - G) + \epsilon_Y$$

(So $\beta = 1$ and $\gamma = -1$)

Milton Friedman's thermostat

- X is outdoor temperature
- Y is indoor temperature
- G is the effect of the central heating system
- Y^* is the desired indoor temperature

In equilibrium ($Y = Y^*$) $\alpha_X = 1$ and $\alpha_0 = -Y^*$

- $X \perp\!\!\!\perp Y$
- $G \not\perp\!\!\!\perp X$
- $G \not\perp\!\!\!\perp Y$, but the better control is, the closer it gets to independent

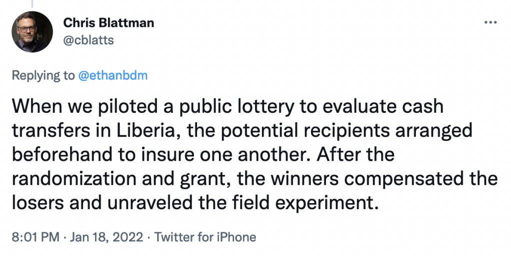
Only *out of equilibrium* can we 'see' the graph

EFFECTIVE CONTROL LOOKS LIKE…NOTHING

bjamngirl
@bjamngirlAA

I'm a medical coder for 10+ years. I have yet to see a patient getting treated for measles, mumps, tuberculosis, diphtheria, rubella, or pertussis. These vaccines have no need to be given.

RESEARCHERS RANDOMIZE – PEOPLE OPTIMIZE


Chris Blattman
@cblatts

Replying to @ethanbdm

When we piloted a public lottery to evaluate cash transfers in Liberia, the potential recipients arranged beforehand to insure one another. After the randomization and grant, the winners compensated the losers and unraveled the field experiment.

8:01 PM · Jan 18, 2022 · Twitter for iPhone

- Efficient market structures
- Regulation to offset negative outcomes
- Feedback control

The causal graph is *timescale specific* (Weinberger, 2020)

"Nana Otafrija Pallbearing & Waiting Services have evaluated your randomized controlled trial on behalf of West Africa"

NEW DIRECTIONS

A fundamental tension

Researchers randomize. People optimize, strategize, and generally create order.

NEW DIRECTIONS

A fundamental tension

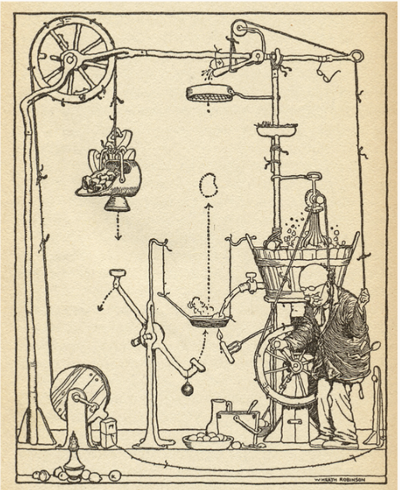
Researchers randomize. People optimize, strategize, and generally create order.

OK, but what about other directions?

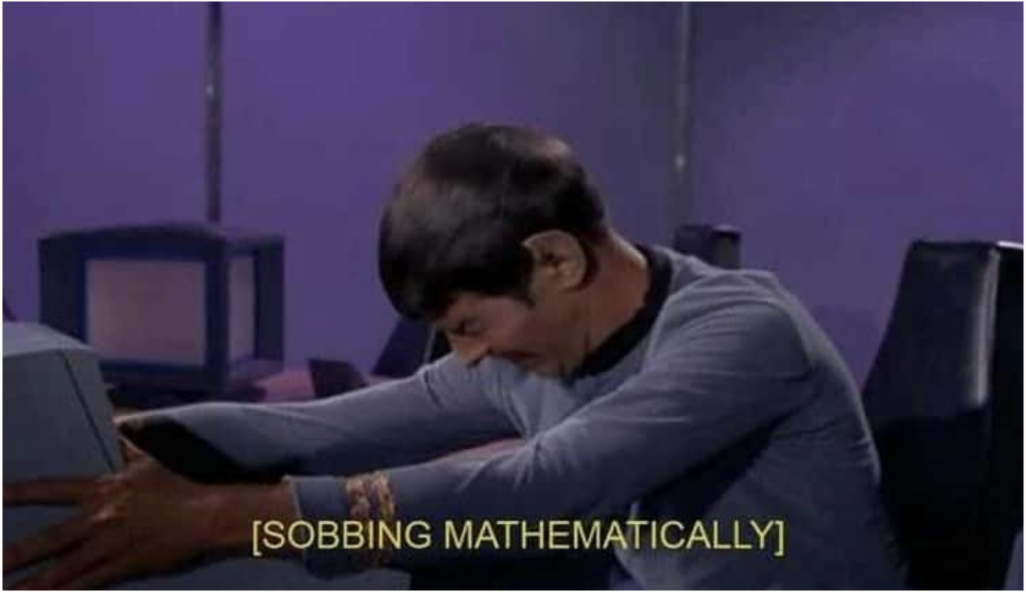
Causal accounts of

- measurement models
- hierarchical data structure
- mediation (no really)
- machine learning

And whatever else we come up with on a chilly Tuesday in Alex



The pancake-making machine


[SOBBING MATHEMATICALLY]

REFERENCES

Campbell, D. T. (1979).Assessing the impact of planned social change. *Evaluation and Program Planning*, *2*(1), 67–90.

Elwert, F., & Winship, C. (2014).Endogenous selection bias: The problem of conditioning on a collider variable. *Annual Review of Sociology*, *40*(1), 31–53.

Fernandez, A. M., Li, K. D., Patel, H. V., Allen, I. E., Ghaffar, U., Hakam, N., & Breyer, B. N. (2024).Electric bicycle injuries and hospitalizations. *JAMA Surgery*.

Fryer, R. G. (2018).An empirical analysis of racial differences in police use of force. *Journal of Political Economy*, *127*(3), 1210–1261.

Goodhart, C. (1981). Problems of monetary management: The U.K. experience. In A. S. Courakis (Ed.), *Inflation, depression, and economic policy in the West*. Barnes and Noble Books

Hernán, M. A., Hernández-Díaz, S., & Robins, J. M. (2004).A structural approach to selection bias. *Epidemiology*, *15*(5), 615–625.

REFERENCES

Knox, D., Lowe, W., & Mummolo, J. (2020).Administrative records mask racially biased policing. *American Political Science Review*, *114*(3), 619–637.

Mac Donald, H. (2020).The myth of systemic police racism [newspaper]. *Wall Street Journal: Opinion: Commentary*.

Mangel, M., & Samaniego, F. J. (1984).Abraham Wald's work on aircraft survivability. *Journal of the American Statistical Association*, *79*(386), 259–267.

Mansournia, M. A., Hernán, M. A., & Greenland, S. (2013).Matched designs and causal diagrams. *International Journal of Epidemiology*, *42*(3), 860–869.

Richardson, T. S., & Robins, J. M. (2013). *Single world intervention graphs (SWIGs): A unification of the counterfactual and graphical approaches to causality*.

Shalizi, C. (2021, March 26). *Regression, thermostats, causal inference: Some finger exercises*. Three-toed sloth.

Weinberger, N. (2020).Near-decomposability and the timescale relativity of causal representations. *Philosophy of Science*, *87*(5), 841–856.