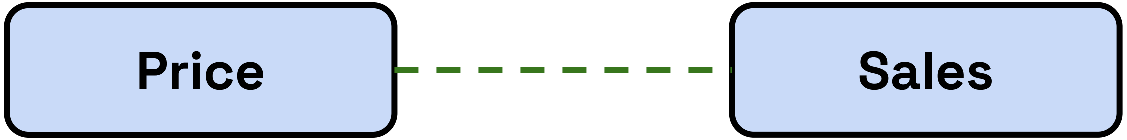
Ergodic

ODSC Europe 2024

A world model is all you need

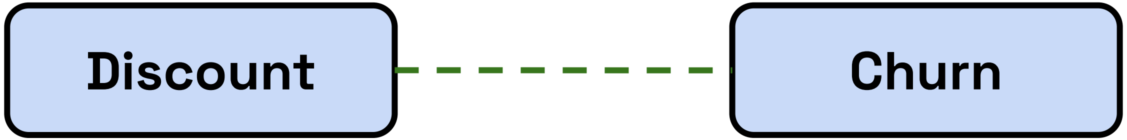
ergodic.ai

September 2024



When trying to estimate elasticity, a company noticed the following problem: there was a positive correlation between their price and the demand -
i.e.: in periods when they lowered the prices, sales actually lowered!

Should they increase their prices?



When modelling customer retention, a telco company was faced with the following problem: the higher the discount they offered, the more likely custumers were likely to churn.

Should they stop offering discounts?

While processing data from hospital admissions, healthcare practitioners noticed something quite alarming: there was a negative correlation between the incidence of diabetes and respiratory diseases.

Should we get flu as a prevention for diabetes?

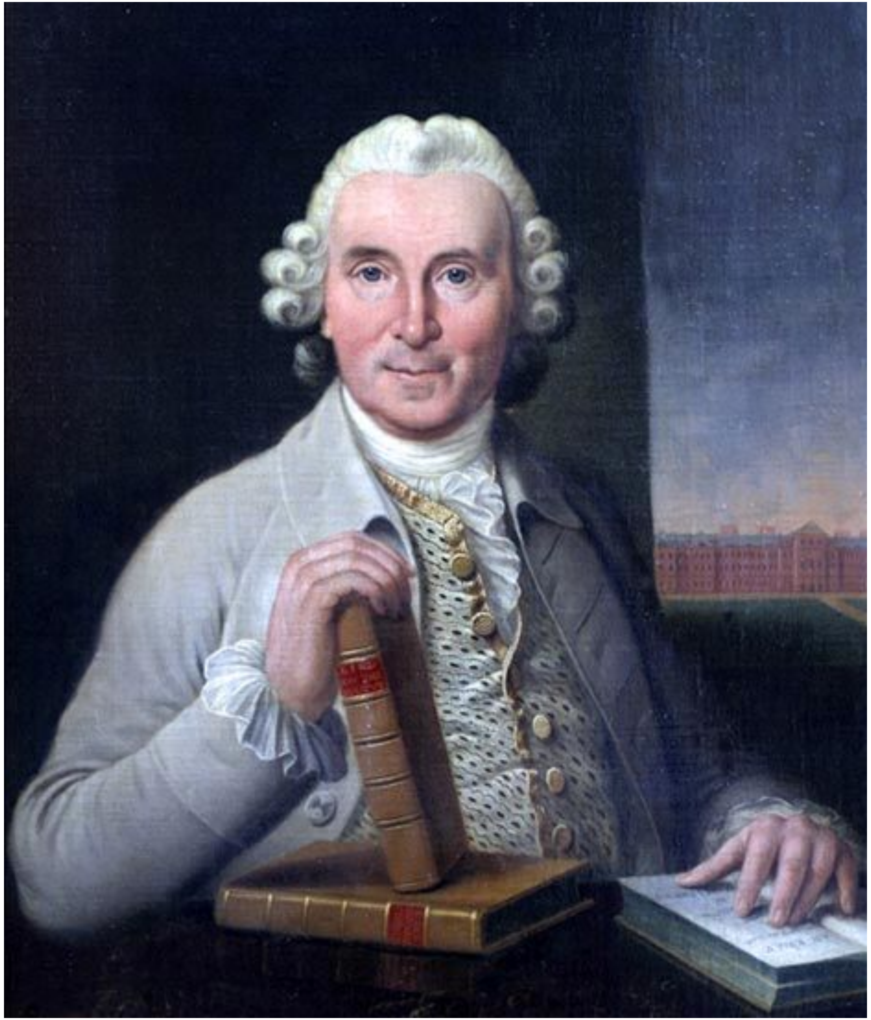
#1 Terra Nova

How we lost the cure for a disease to a spurious correlation

Scott, Bowers, Wilson, and Edgar Evans at [Amundsen's base] at the South Pole

[Robert Falcon] Scott left a [South Pole] base abundantly stocked with fresh meat, fruits, apples, and lime juice, and headed out on the ice for five months with no protection against scurvy, all the while confident he was not at risk. What happened? … In the second half of the nineteenth century, **the cure for scurvy was lost**. The story of how this happened is a striking demonstration of the **problem of induction**, and how progress in one field of study can lead to unintended steps backward in another.

Maciej Cegłowski [link]

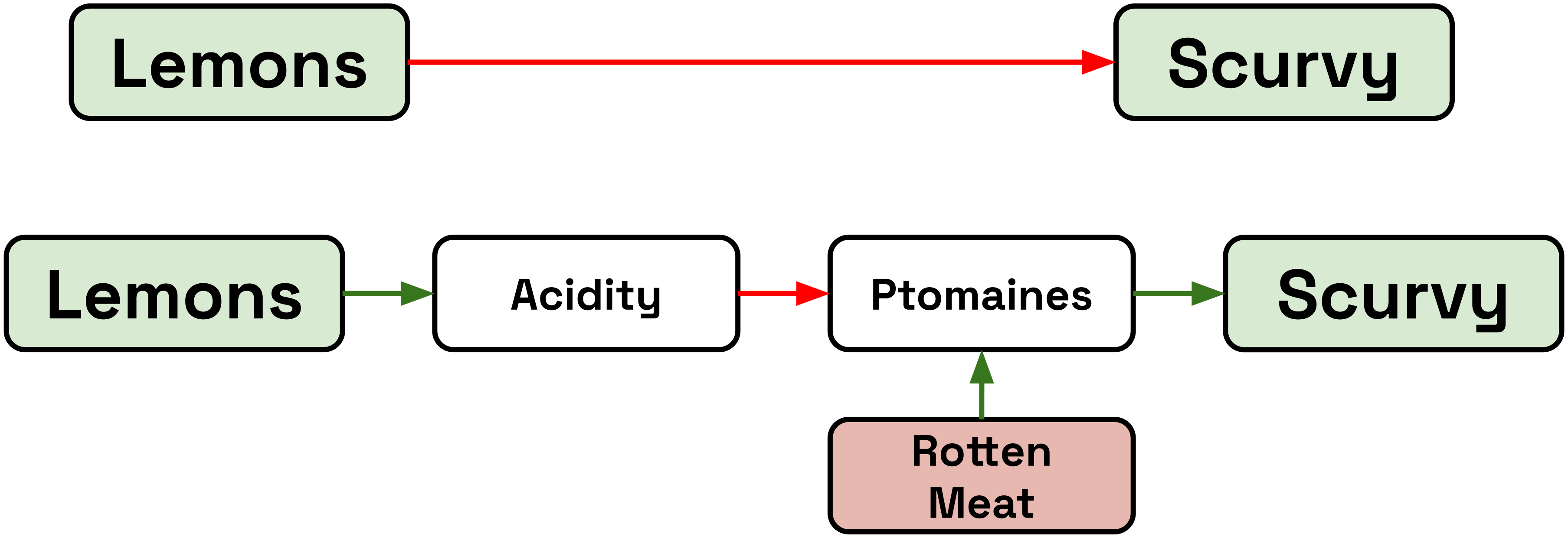


ergodic.ai

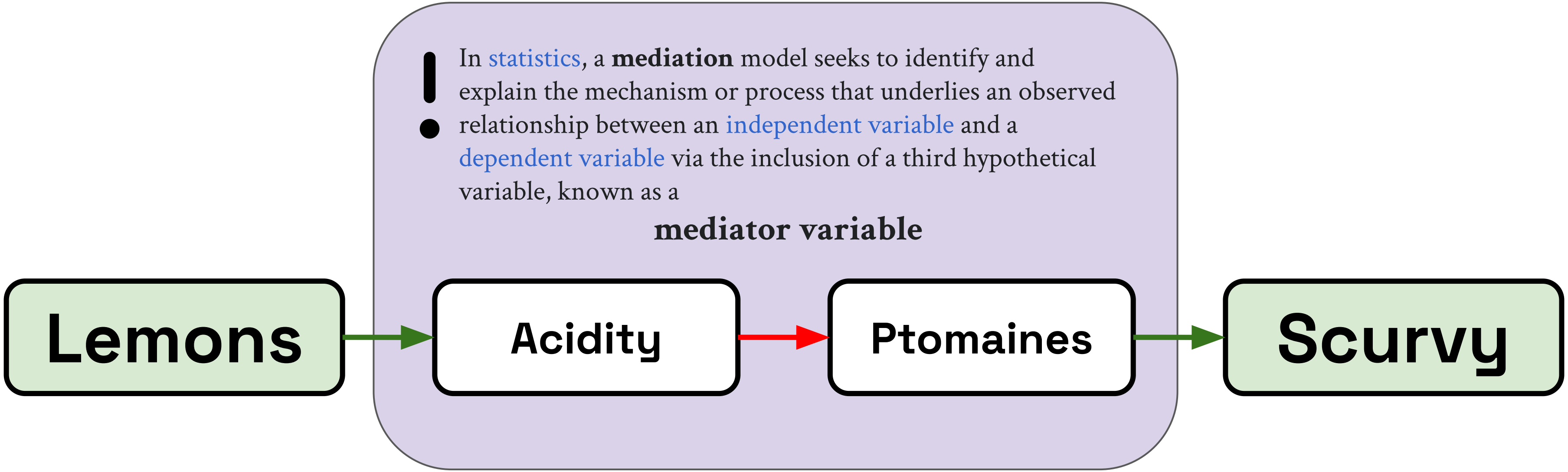
Lind **thought that scurvy was due to putrefaction of the body** that **could be helped by acids,** so he included an acidic dietary supplement in the experiment. This began after two months at sea when the ship was afflicted with scurvy. He divided twelve scorbutic sailors into **six groups of two**. They all received the same diet, but in addition group one was given a quart of cider daily, group two twenty-five drops of elixir of vitriol (sulfuric acid), group three six spoonfuls of vinegar, group four half a pint of seawater, group five two oranges and one lemon, and the last group a spicy paste plus a drink of barley water. The treatment of group five stopped after six days when they ran out of fruit, but by that time **one sailor was fit for duty while the other had almost recovered**. Apart from that, only group one showed any effect from its treatment.

Ergodic

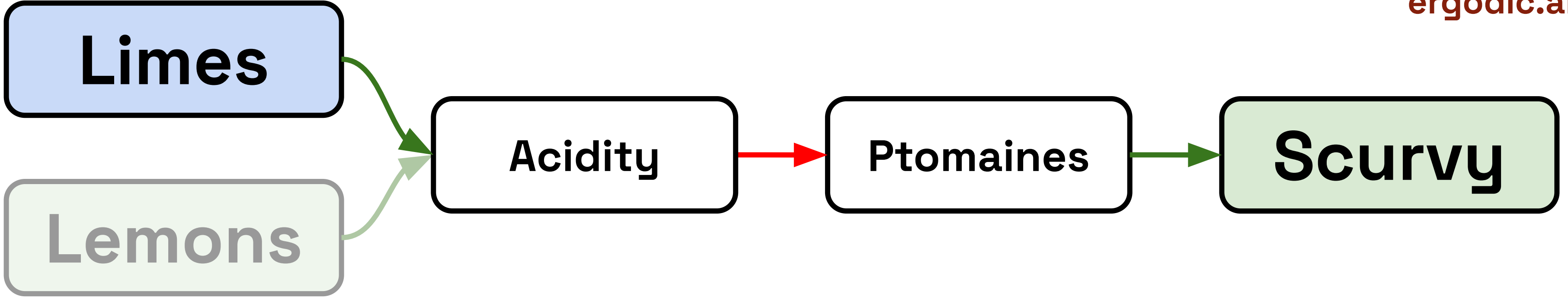
Randomized Control Trial



The leading theory on why lemons cured scurvy
was that it denatured the ***ptomaines*** in rotten meat

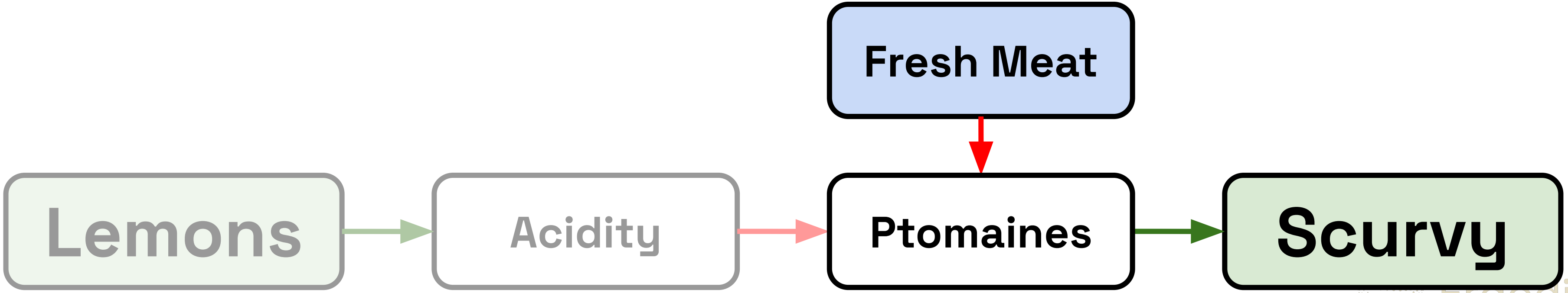


Acidity was believed to mediate the relationship between lemons and scurvy.

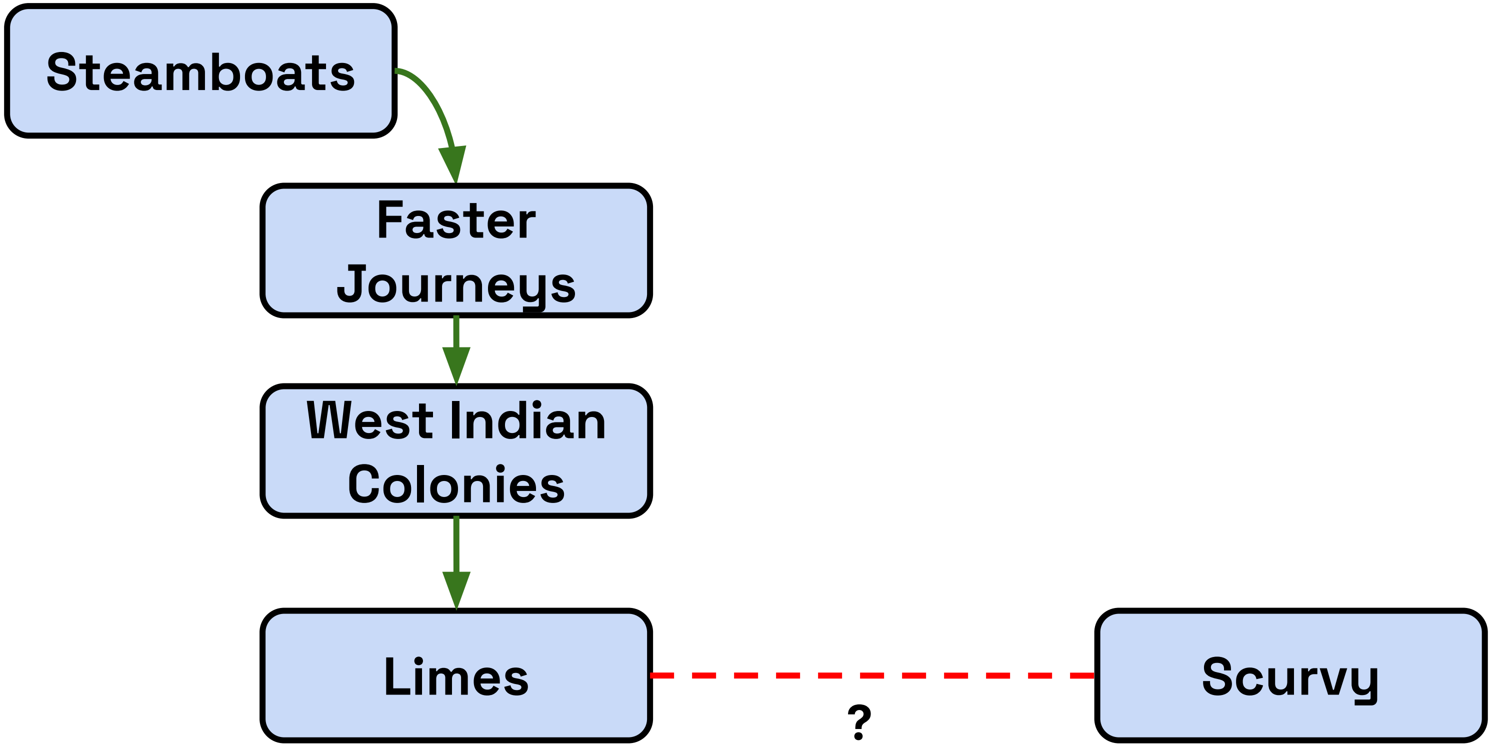


If this is the mediating effect, then we can also
substitute lemons for limes

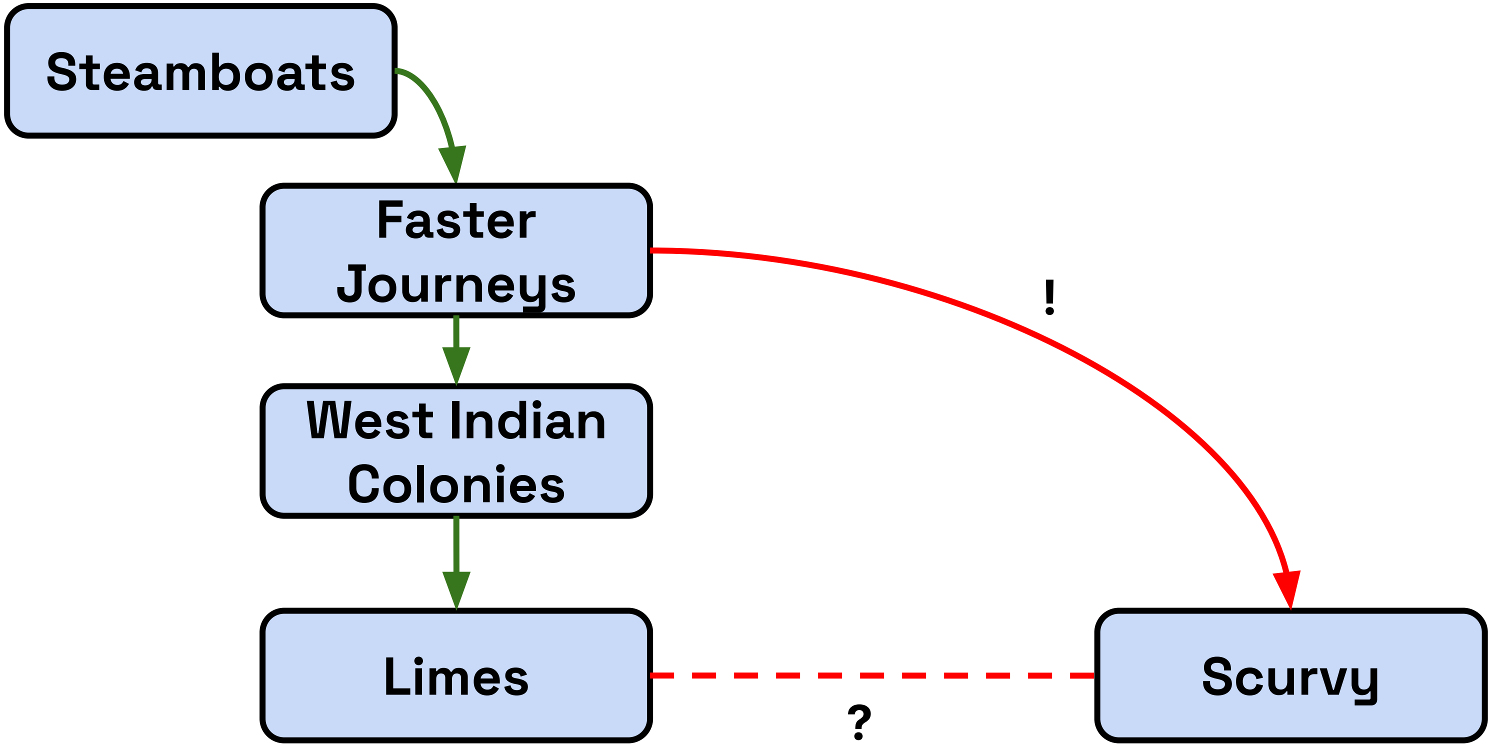
or add fresh sources of meat



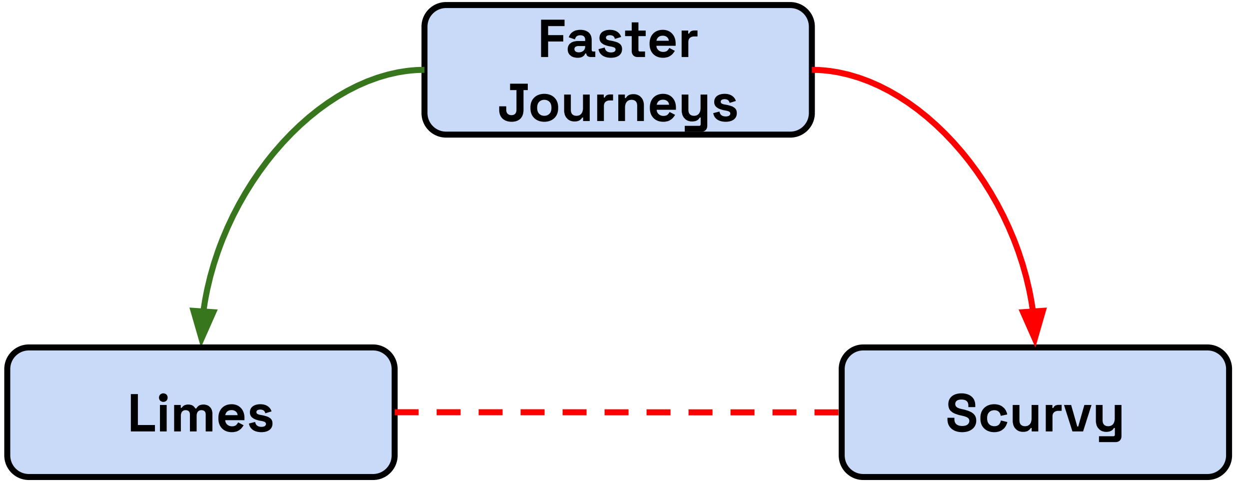
In 1860, naval authorities switched procurement from **Mediterranean lemons to West Indian limes**. The motives for this were mainly colonial - it was better to buy from British plantations than to continue importing lemons from Europe. Confusion in naming didn't help matters. Both "lemon" and "lime" were in use as a collective term for citrus, and though European lemons and sour limes are quite different fruits, their Latin names (citrus medica, var. limonica and citrus medica, var. acida) suggested that they were as closely related as green and red apples. Moreover, as there was a widespread belief that the antiscorbutic properties of lemons were due to their acidity, it made sense that the more acidic Caribbean limes would be even better at fighting the disease.



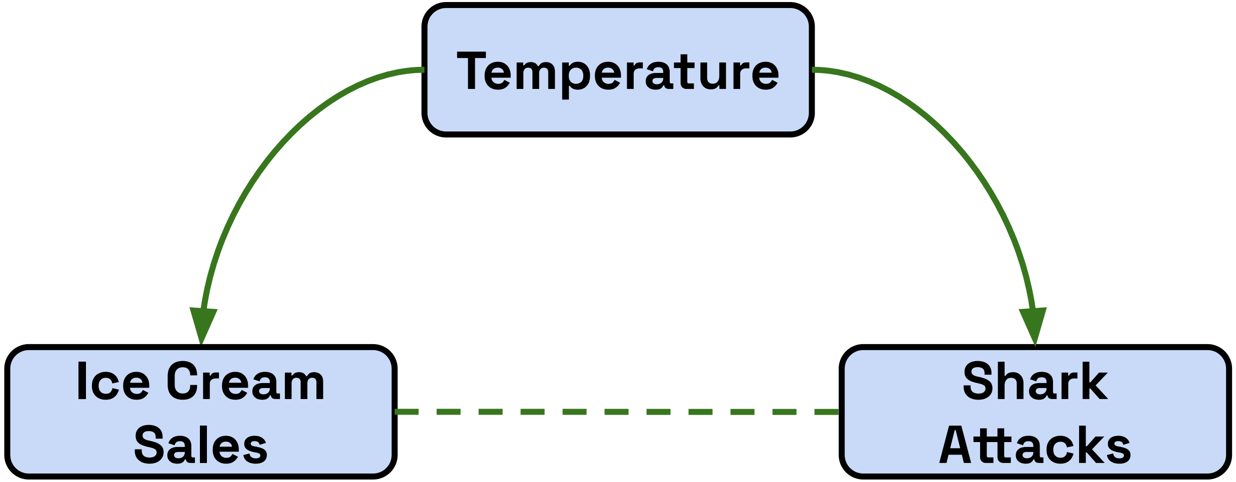
During the 1800s, British sailors - limeys - still believed to be protected from scurvy due their lime allowance.



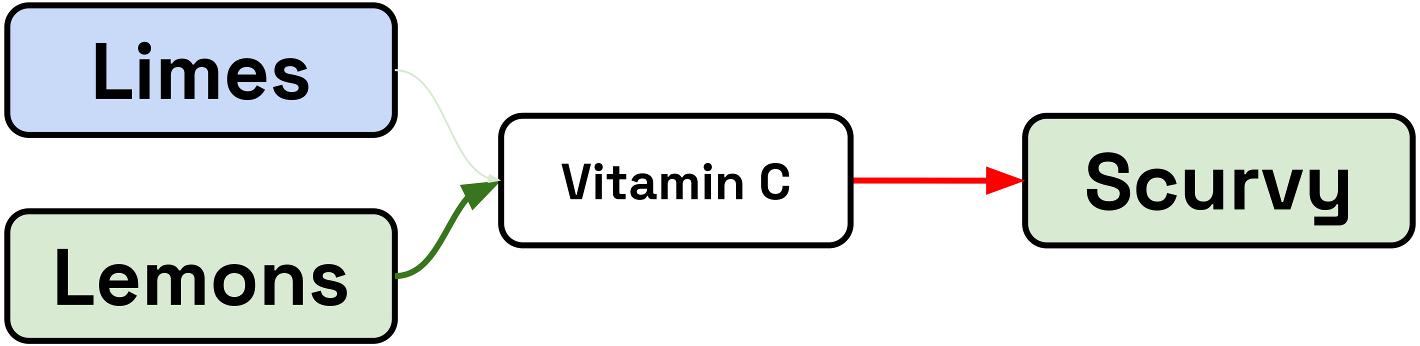
But in reality, most trips were short enough that they would have been protected anyway, as the disease wouldn't have time to set in.



A confounder is a variable that influences both the dependent variable and independent variable, causing a **spurious association**.



! **A confounder** is a variable that influences both the dependent variable and independent variable, causing a **spurious association**.

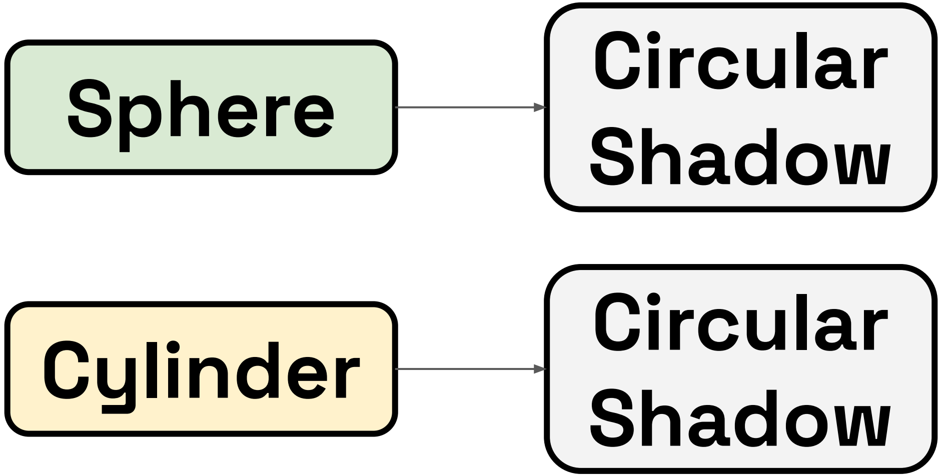


One of the simplest of diseases managed to utterly confound us for so long, at the cost of millions of lives, even after we had stumbled across an unequivocal cure. It makes you wonder how many incurable ailments of the modern world will turn out to have equally simple solutions, once we are able to see them in the correct light.

What will we be slapping our foreheads about sixty years from now, wondering how we missed something so obvious?

Maciej Cegłowski [link]

You see a circular shadow on the wall.



What is the probability that a Sphere will cast a circular shadow?

P(Observations | Mechanism) -> Easy!

What is the probability that the circular shadow was created by a Sphere?

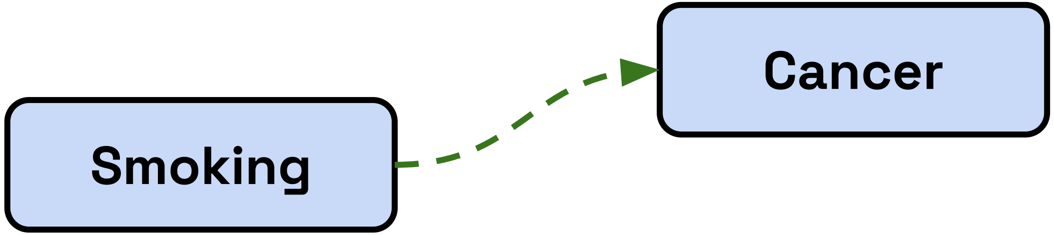
P(Mechanism | Observations) -> HARD!

P(Mechanism | Observations) =
P(Observations | Mechanism) * P(Mechanism) / P(Observation).

P(Observation) = "partition function Z" =
Sum(P(Observation | Mechanism) * P(Mechanism)) → Intractibility

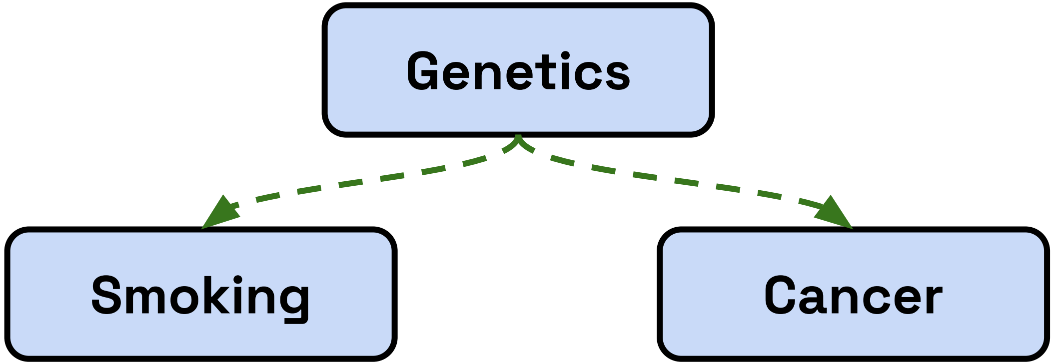
Intractability KILLS!

OBSERVATION -> People who smoke are more likely do die of lung cancer



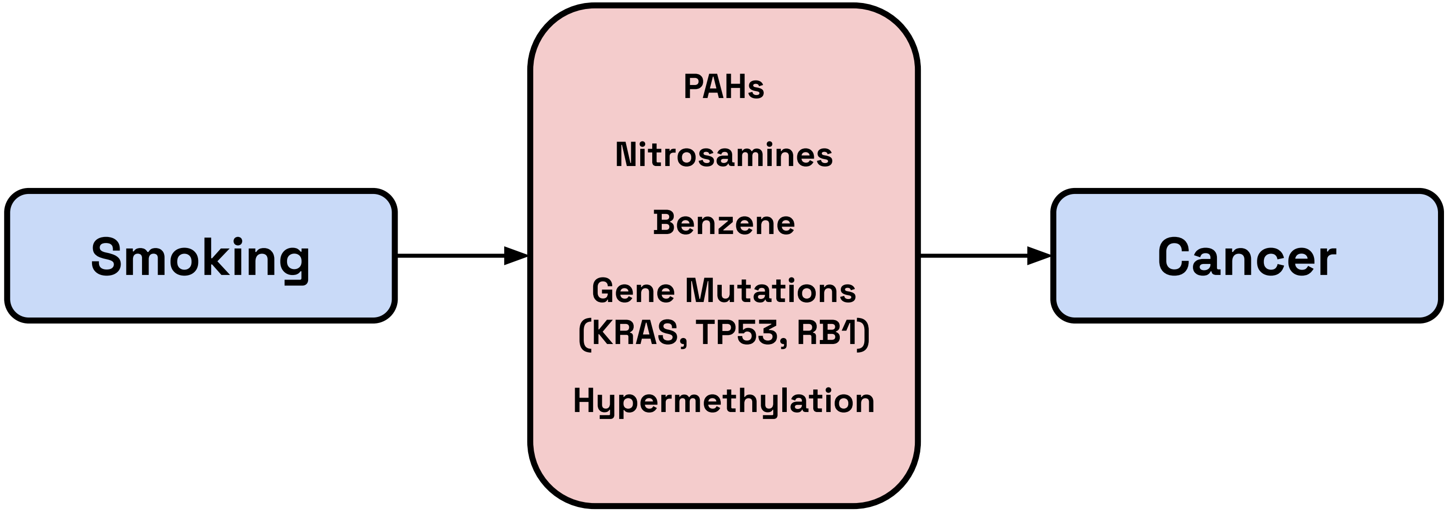
What is the probability that "Smoking causes cancer" (Mechanism) given the observation?

Intractability KILLS!



Knowing that there are multiple competing explanations that also explain the data...

World models save lives!

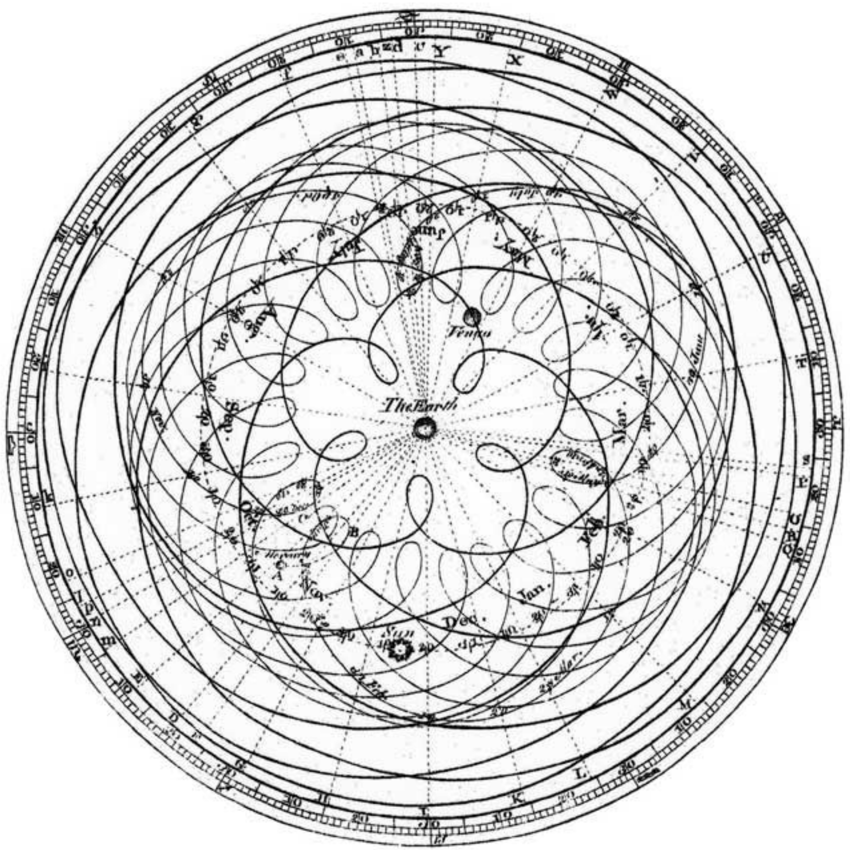


Conclusion #1

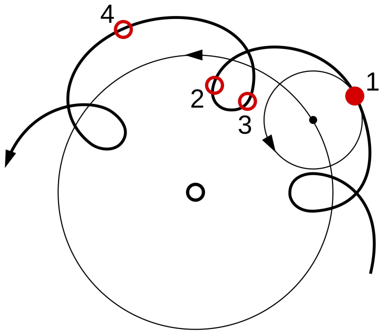
It’s hard to build a world model

#2 From Aristotle to Kepler

How a causal insight changed science forever



Epicycles are astronomical models that **predict** the movement of celestial bodies, assuming that the earth is in the center of the solar system. It's a perfect **predictive** model, although it completely fails at identifying the right underlying mechanisms.



Gravity generalizes. It both explains the movement of celestial bodies, and also how an apple falls from the tree.

Constructing the right world model is fundamental for generalization.

How did we go from epicycles to gravity?

1: A silver-nosed astronomer with a drunken elk pet

Tycho Brahe was wealthy and flamboyant, connected to royalty and given to throwing lavish and drunken parties. Famously, Tycho wore a silver prosthetic nose, having once lost part of his real nose in a foolish duel as a young man.

He was the proud owner of a pet elk, but the poor beast drank too much beer from the cups of guests during a festive dinner and died after tumbling down a set of stairs.

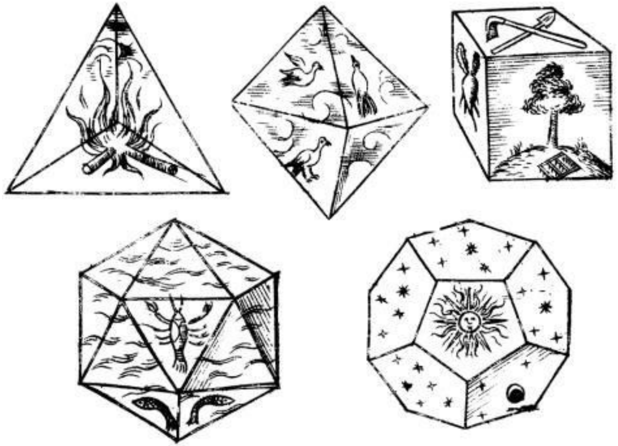
While holding court, Tycho often tossed scraps to a dwarf named Jepp who hid under the dinner table. Tycho leveraged his considerable reputation to solicit funding from Rudolf II, the Holy Roman Emperor, to build a new observatory near Prague.

BRIAN VENTRUDO [LINK]

2: A mystic mathematician who sought...

Kepler, by contrast, was an awkward, self-tortured neurotic. His mother was nearly burned at the stake as a witch, while his father, a drunk and a wife beater, supported the family for a time as a mercenary before disappearing when young Kepler was five years old. In school, Kepler was ridiculed by his classmates and he failed his attempt to become a Lutheran minister. So he turned to mathematics and made a living teaching and casting horoscopes to noblemen, an occupation for which he was rarely paid and routinely mocked when he bumbled into court with ill-fitting and food-stained clothes.

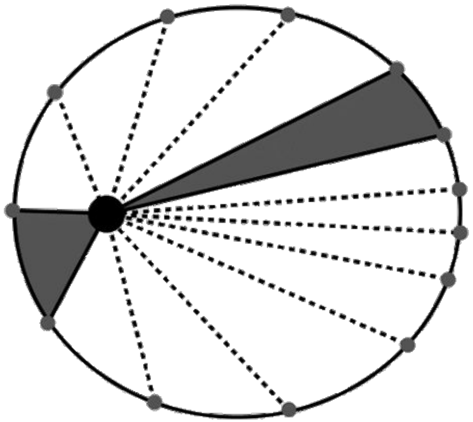
2: ...perfect symmetries,



Kepler claimed to have had an epiphany on 19 July 1595 (...) **He found that each of the five Platonic solids could be inscribed and circumscribed by spherical orbs**; nesting these solids, each encased in a sphere, within one another would **produce six layers, corresponding to the six known planets—Mercury, Venus, Earth, Mars, Jupiter, and Saturn**. By ordering the solids selectively—octahedron, icosahedron, dodecahedron, tetrahedron, cube—Kepler found that the spheres could be placed at intervals corresponding to the relative sizes of each planet's path, assuming the planets circle the Sun.

Kepler also

3: and a quest for beauty and simplicity



Kepler calculated and recalculated various approximations of Mars's orbit using an equant, but found that the models were still too complex and inaccurate.

In Kepler's religious view of the cosmos, the Sun (a symbol of God the Father) was the source of motive force in the Solar System. As a physical basis, Kepler supposed that the motive power radiated by the Sun weakens with distance, an assumption that perhaps would restore astronomical order. Based on Tycho's measurements, **he created a formula in which a planet's rate of motion is inversely proportional to its distance from the Sun**.

Verifying this relationship throughout the orbital cycle required very extensive calculation; **to simplify this task**, Kepler reformulated the proportion in terms of geometry:

planets sweep out equal areas in equal times

Kepler's decision to base his **causal explanation of planetary motion** on a distance-velocity law (...) marks a major shift from ancient to modern conceptions of science.

Kepler had begun with physical principles and had then derived a trajectory from it, rather than simply constructing new models.

Peter Barker and Bernard R. Goldstein, "Distance and Velocity in Kepler's Astronomy"

Conclusion #2

Physical Principles Before Models

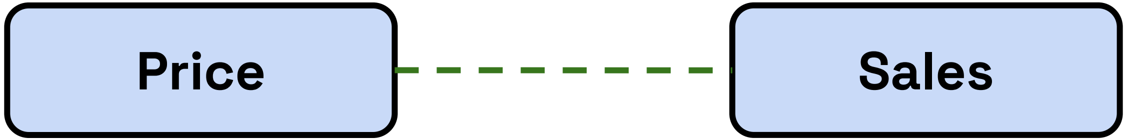
#3 Causality and the technocene

Changing the world, one causal insight at a time


WHITESNAKE
IS THIS LOVE


THE SHIRELLES
WILL YOU STILL LOVE ME TOMORROW
REMASTERED

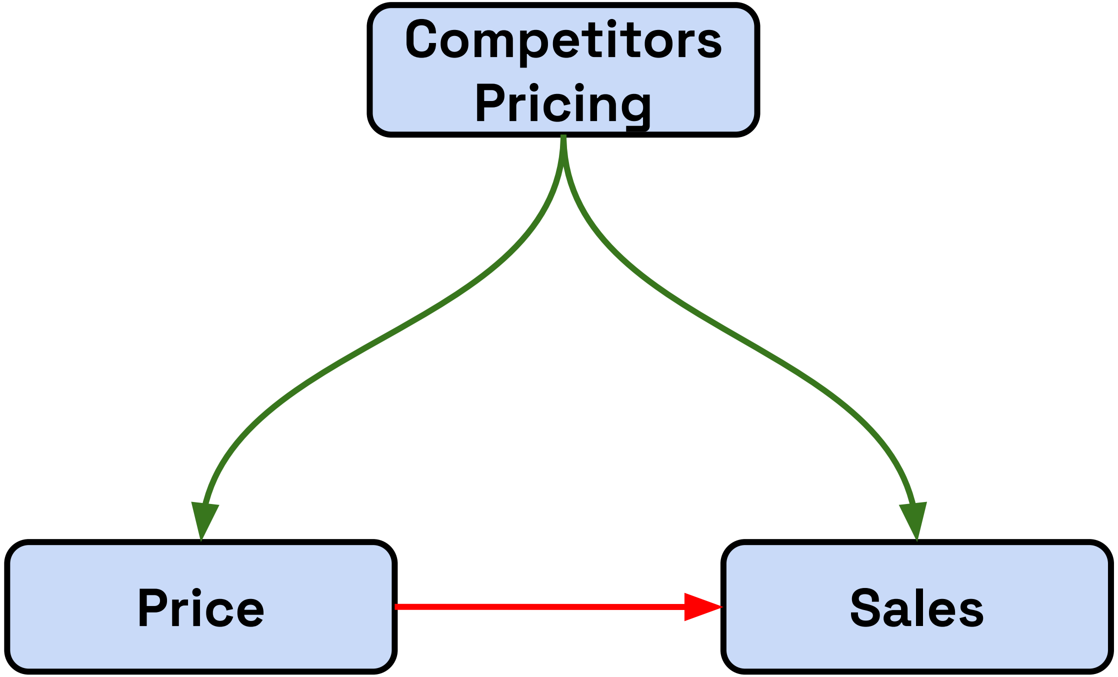

SHOULD I STAY
OR SHOULD I GO



When trying to estimate elasticity, a company noticed the following problem: there was a positive correlation between their price and the demand -
i.e.: in periods when they lowered the prices, sales actually lowered!

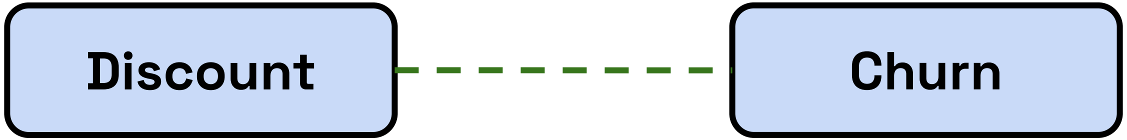
Should they increase their prices?

Simpson's paradox #1



As competitors decrease their prices, we react by decreasing our prices as well. But by decreasing their prices competitors also capture a bigger market share, thus reducing our sales.

This creates a **positive correlation** between our prices and our sales, although **the causal effect is negative**.

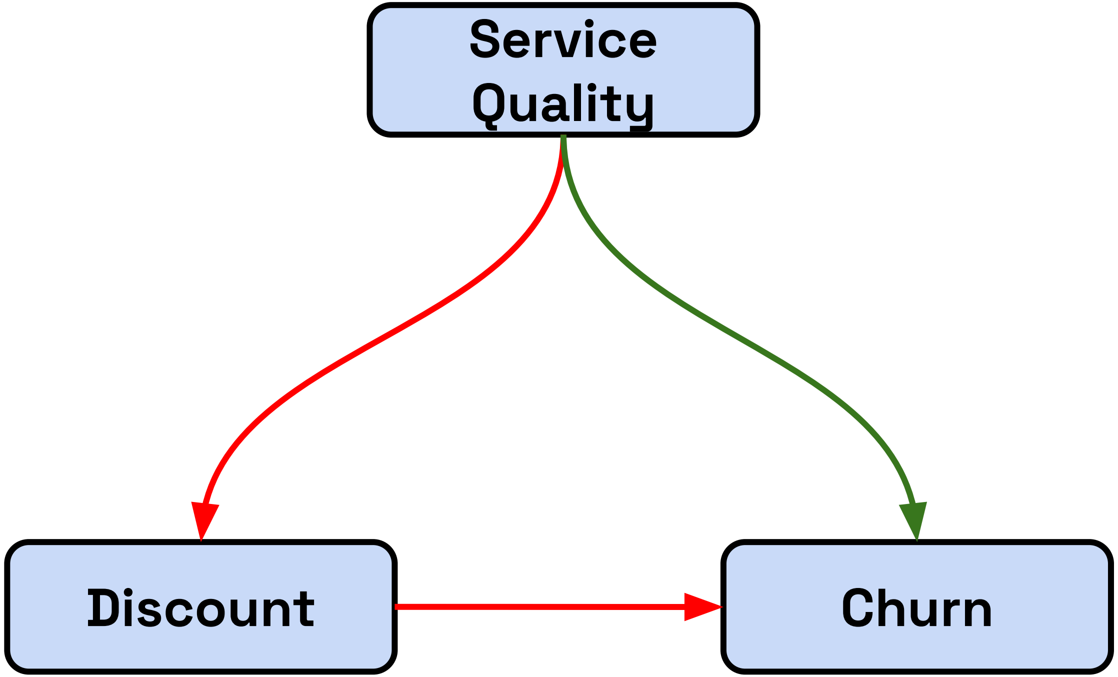


When modelling customer retention, a telco company was faced with the following problem: the higher the discount they offered, the more likely customers were likely to churn.

Should they stop offering discounts?

ergodic.ai

Simpson's paradox #2



Customers who faced problems with their service complain more, thus receiving a higher rate of discounts. These customers, however, are the first ones to cancel their membership when they have the chance.

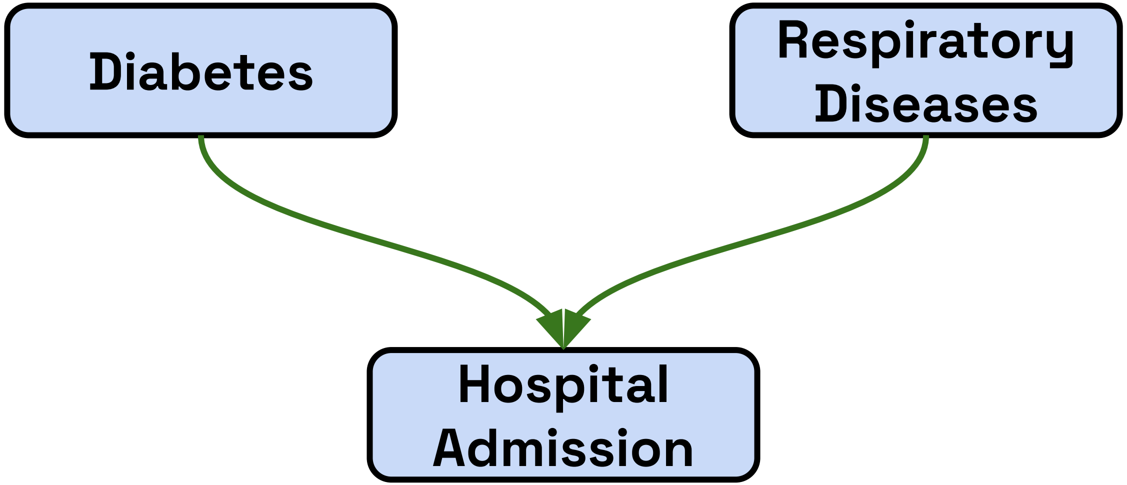
This creates a **positive correlation** between discounts and churn, although **the causal effect is negative**.

Ergodic

While processing data from hospital admissions, healthcare practitioners noticed something quite alarming: there was a negative correlation between the incidence of diabetes and respiratory diseases.

Should we get flu as a prevention for diabetes?

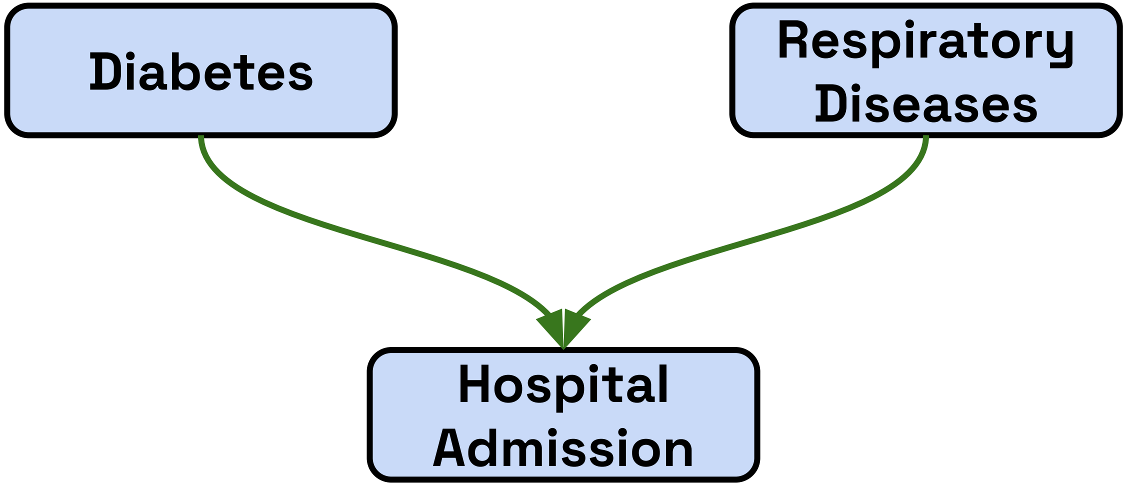
Collider bias



Patients are admitted to hospital whenever they require critical care of either one of these conditions.

Conditioning on hospital admissions creates a **negative correlation** between both risk factors, although there's **a priori no causal relationship between them**.

Collider bias

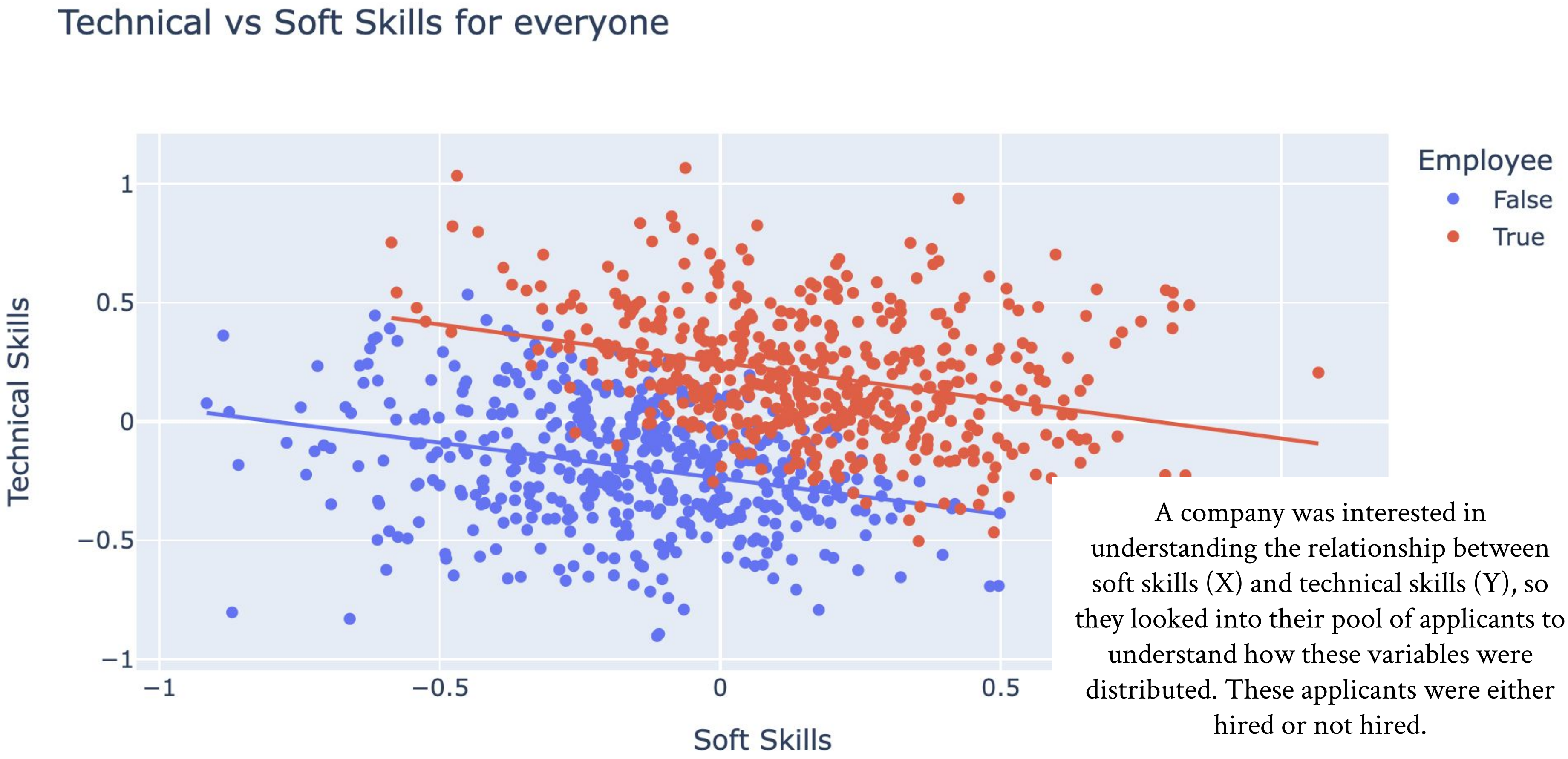


Patients are admitted to hospital whenever they require critical care of either one of these conditions.

Conditioning on hospital admissions creates a **negative correlation** between both risk factors, although there's **a priori no causal relationship between them**.

ergodic.ai

Collider bias

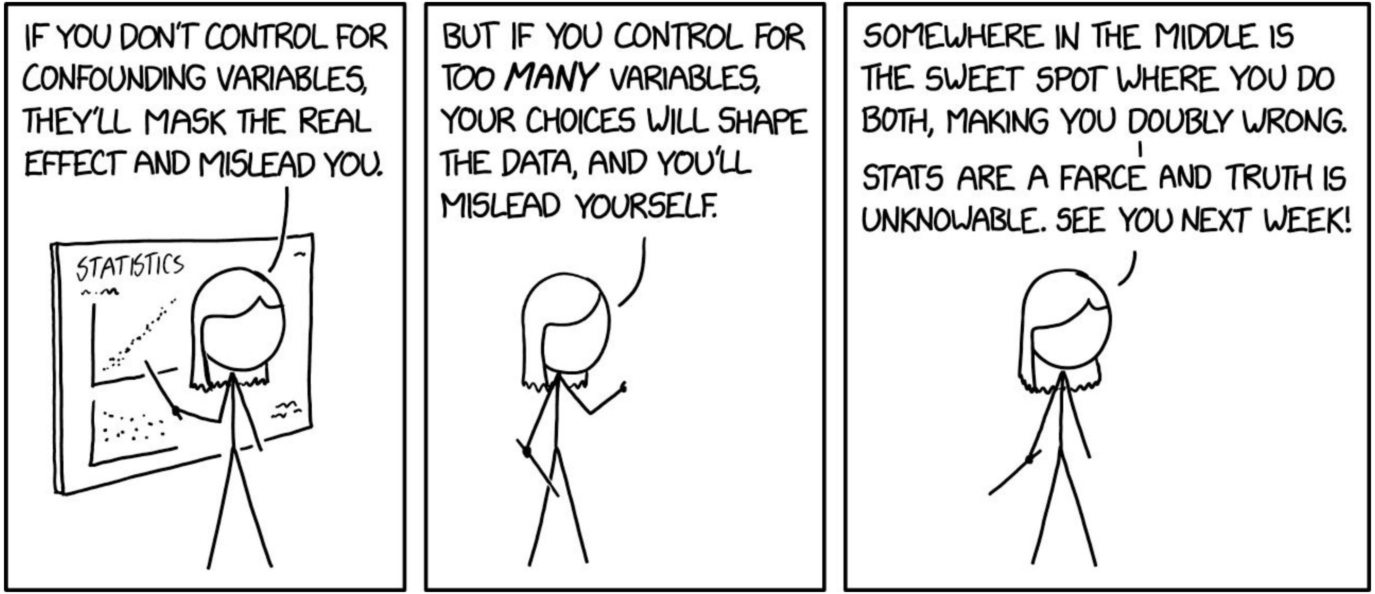


A company was interested in understanding the relationship between soft skills (X) and technical skills (Y), so they looked into their pool of applicants to understand how these variables were distributed. These applicants were either hired or not hired.

Conclusion #3

Control for Confounders

Don’t control for Colliders


IF YOU DON'T CONTROL FOR CONFOUNDING VARIABLES, THEY'LL MASK THE REAL EFFECT AND MISLEAD YOU.
STATISTICS
BUT IF YOU CONTROL FOR TOO *MANY* VARIABLES, YOUR CHOICES WILL SHAPE THE DATA, AND YOU'LL MISLEAD YOURSELF.
SOMEWHERE IN THE MIDDLE IS THE SWEET SPOT WHERE YOU DO BOTH, MAKING YOU DOUBLY WRONG.
STATS ARE A FARCE AND TRUTH IS UNKNOWABLE. SEE YOU NEXT WEEK!

What Is Causal AI ?

ŷ problems

Classical supervised ML

β̂ hat

causal inference?

What happens to Y if we change X ?

ŷ problems

Pattern recognition:

"Is this love?"

cost(model.predict(X), y)

There's a target **y**
(the label)
and our objective is to
reconstruct the label
using features **X**.

ŷ problems

Forecasting:

"Will you still love me tomorrow?"

cost(model.predict(X), y)

There's a target **y**
(the future realisations of a variable)
and our objective is to reconstruct the label using features **X**.

$\hat{\beta}$ hat problems

"Should I stay or should I go now?"

E(y | do(go)) = trouble
"If I go there will be trouble"

E(y | do(go)) = 2*trouble
"If I stay it will be double"

$$y = a + \beta * x + \ldots$$

Predicting **y** is less important than estimating the effect of **x** on **y**.

Causal Inference is required for optimal decision making!

But ML is awful at it!

why is vanilla ML awful at that?

$p(y \mid do(x)) \neq p(y \mid x)$

Estimating the impact of a change

Estimating joint distributions

X - Adam buys a rucksack
Y - Adam buys a laptop

$p(Y \mid X) \neq 0$

There's a big probability that Adam buys a laptop **knowing that** he bought a rucksack

$p(Y \mid do(X)) \sim 0$

If I offer Adam a Rucksack (**intervention**), it's very unlikely that he will buy a Laptop **because** of it.

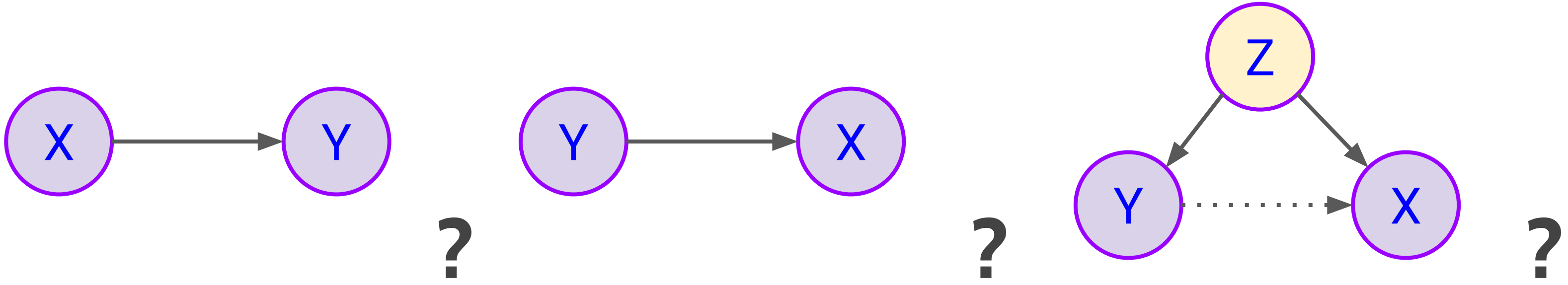
$p(X \mid do(Y)) \sim \text{high}$

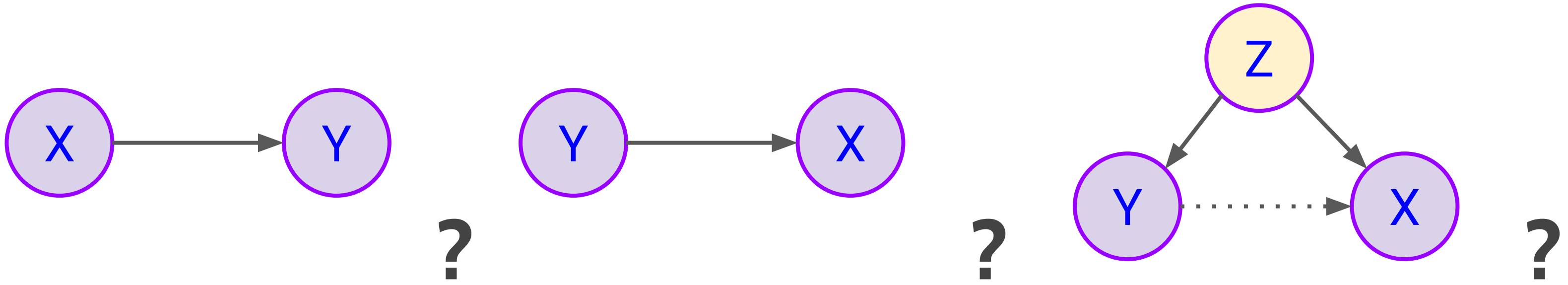
But if I offer him a laptop, he will probably want to buy a rucksack to keep it safe.

observational data tells us about joint distributions (“correlations”)

$p(y \cap x)$

but we need more



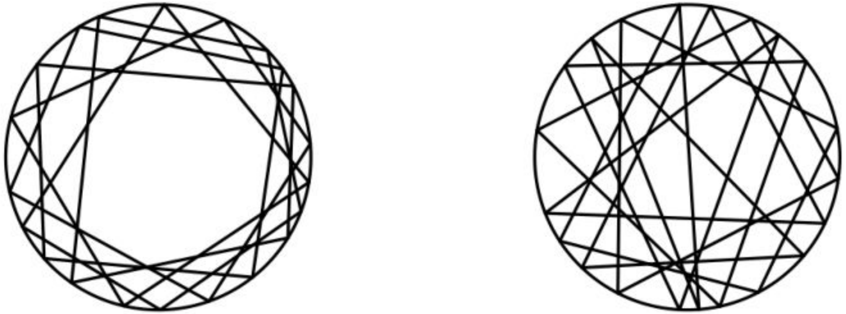


Causal structures are inductive biases not present in ML.

But absolutely necessary to make the right decisions!

Conclusion #4

Predicting Is Not Enough



A. Non-ergodic B. Ergodic

In mathematics, ergodicity expresses the idea that a point of a moving system, either a dynamical system or a stochastic process, will eventually **visit all parts of the space** that the system moves in, in a uniform and random sense.