THE BOOK OF WHY (JUDEA PEARL)

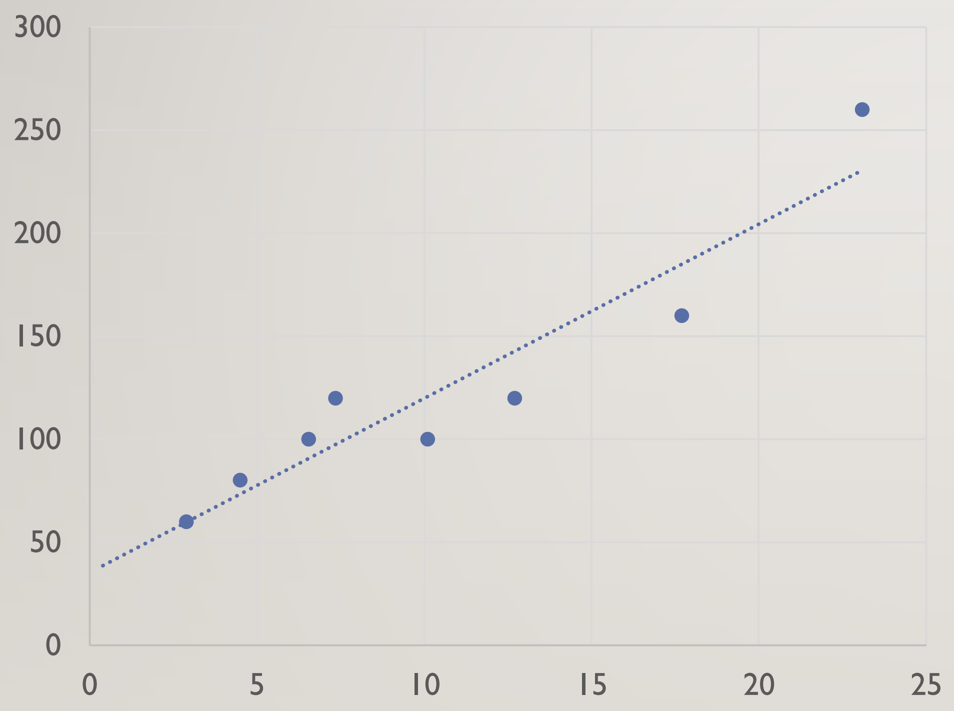
OR: WHAT THE DATA DOESN'T TELL YOU (BY PATRICK DAMMANN; 2020-01-16)

Image by enriquelopezgarre from Pixabay
https://pixabay.com/photos/animal-night-moon-sky-gallo-cloud-4166416/

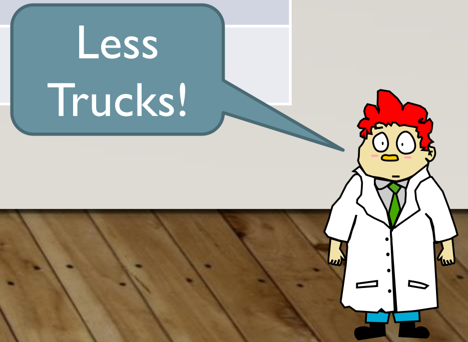
MR. ROBERT OTTER

- freshly elected, human major of some town
- by **absolutely** no means an android driven by state-of-the-art neural networks

TWO SCENARIOS (I: AIR POLLUTION IN DISTRICT)



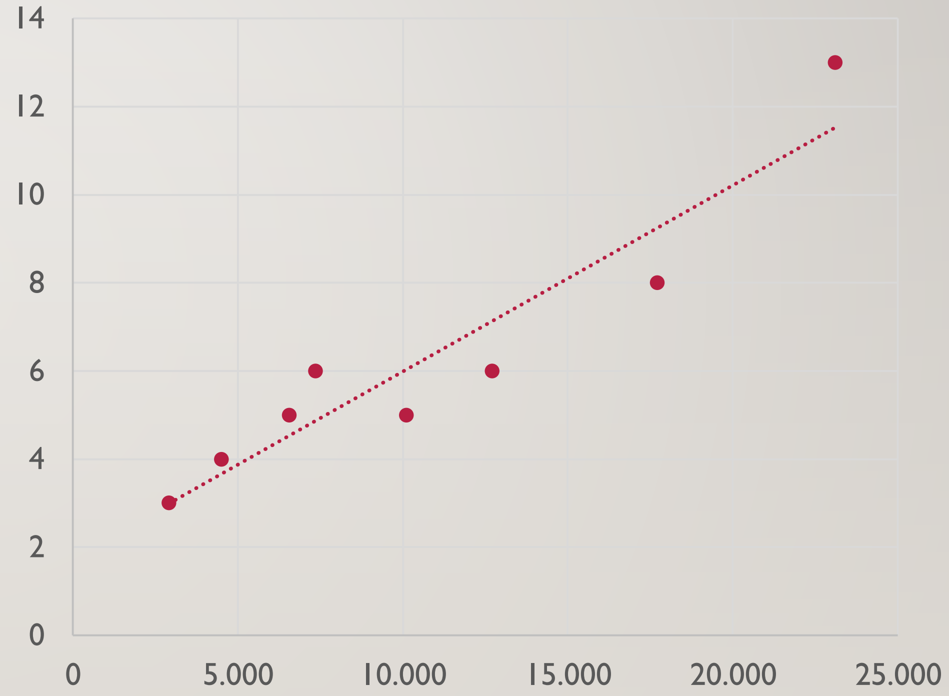
PM10 over mean in µg/m³	#Trucks passing / day
4,5	80
12,7	120
2,9	60
23,1	260
6,55	100
10,1	100
7,35	120
17,7	160



INHALES DEEPLY

TWO SCENARIOS (II: PRIVATE HOUSE FIRES)

Fire Damage in $	#Firefighters
4.500	4
12.700	6
2.900	3
23.100	13
6.550	5
10.100	5
7.350	6
17.700	8



WELL…

WHAT WENT WRONG?

MR. ROB OTTER

- data driven decision making
  - sees only correlation between two phenomena
  - tries to act based on what *is seen*, not what *might happen* or *might have happened*
- same action evokes dramatically different outcomes

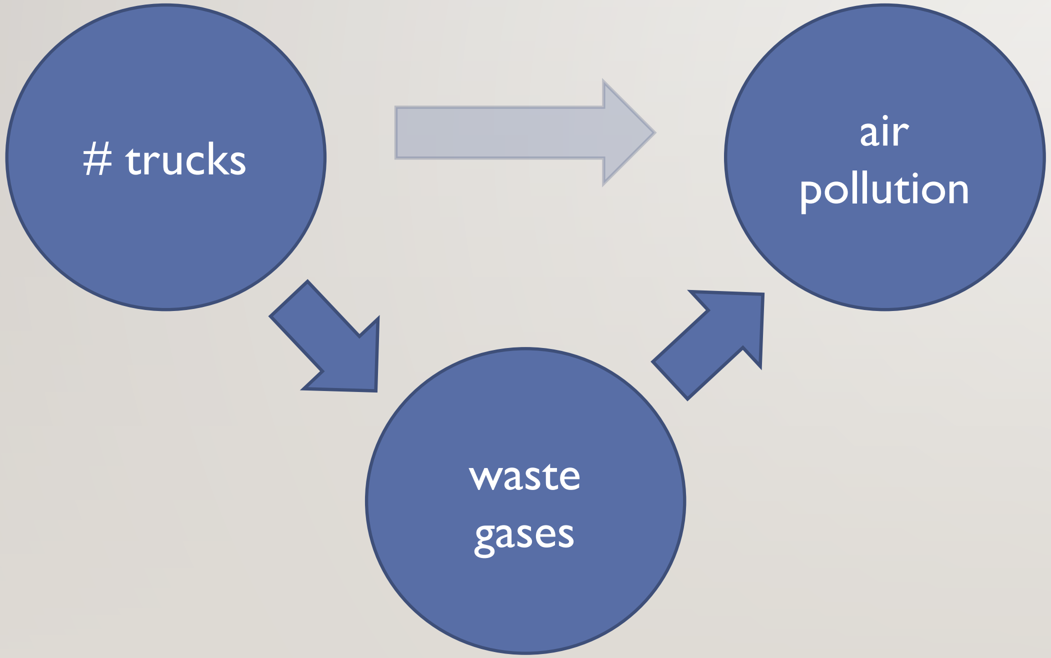
(OTHER) HUMANS

- see immediately what went wrong
- have a „*model of the world*“ in their head
- use more information than just data
  - but which?

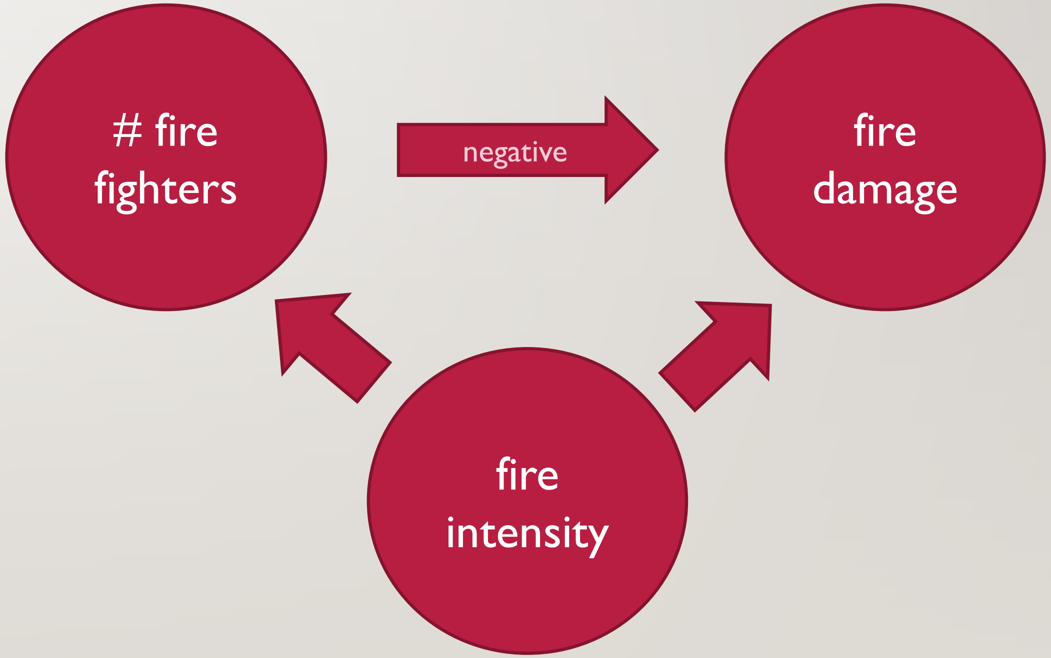
CAUSALITY

THE PROCESS BEHIND THE DATA

AIR POLLUTION PROBLEM



HOUSE FIRE PROBLEM



THE CAUSE-FREE DISTOPIAN WORLD (OF SCIENCE)

definition of "*science*"

(knowledge from) the careful study of the structure and behaviour of the physical world,
especially by watching, measuring, and doing experiments,
and the development of theories to describe **the results of these activities**

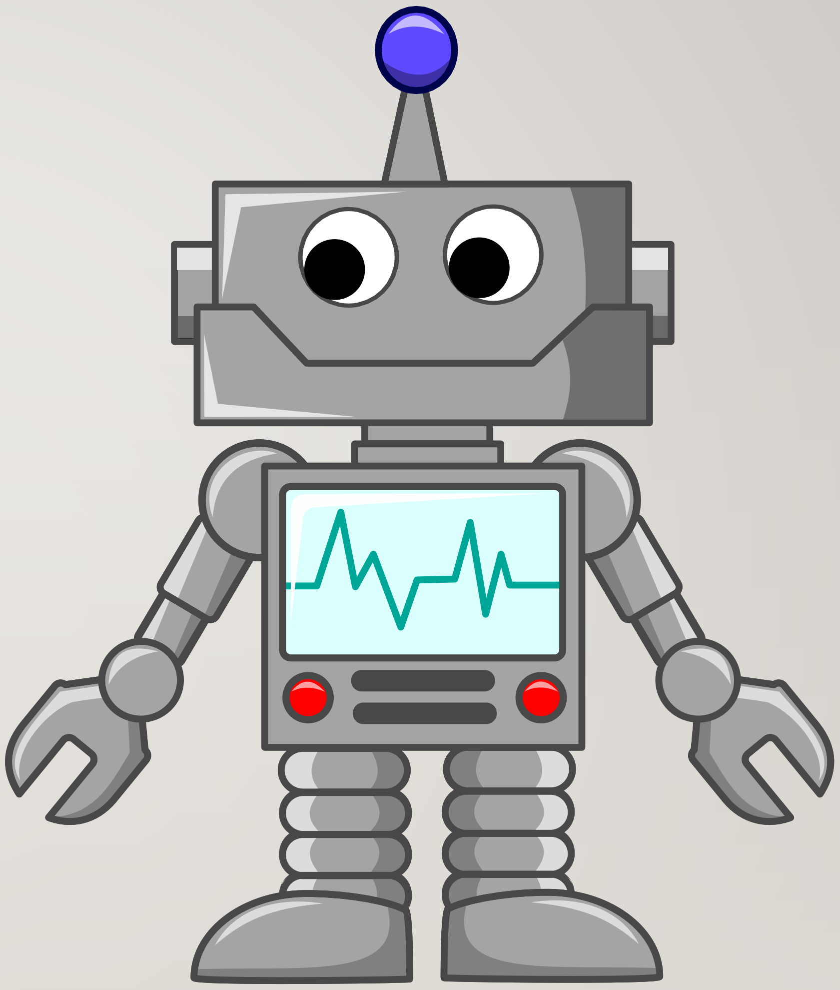
From: https://dictionary.cambridge.org/de/worterbuch/englisch/science

root of all evil: **statistics**

- needed by most sciences to **process data** and **deal with uncertainty**
- "correlation doesn't imply causation"
  - became "dogma"
  - talking about causality deemed unscientific
  - Pearson: causation just correlation with special (observable) properties
- "causal revolution" since the 1960s

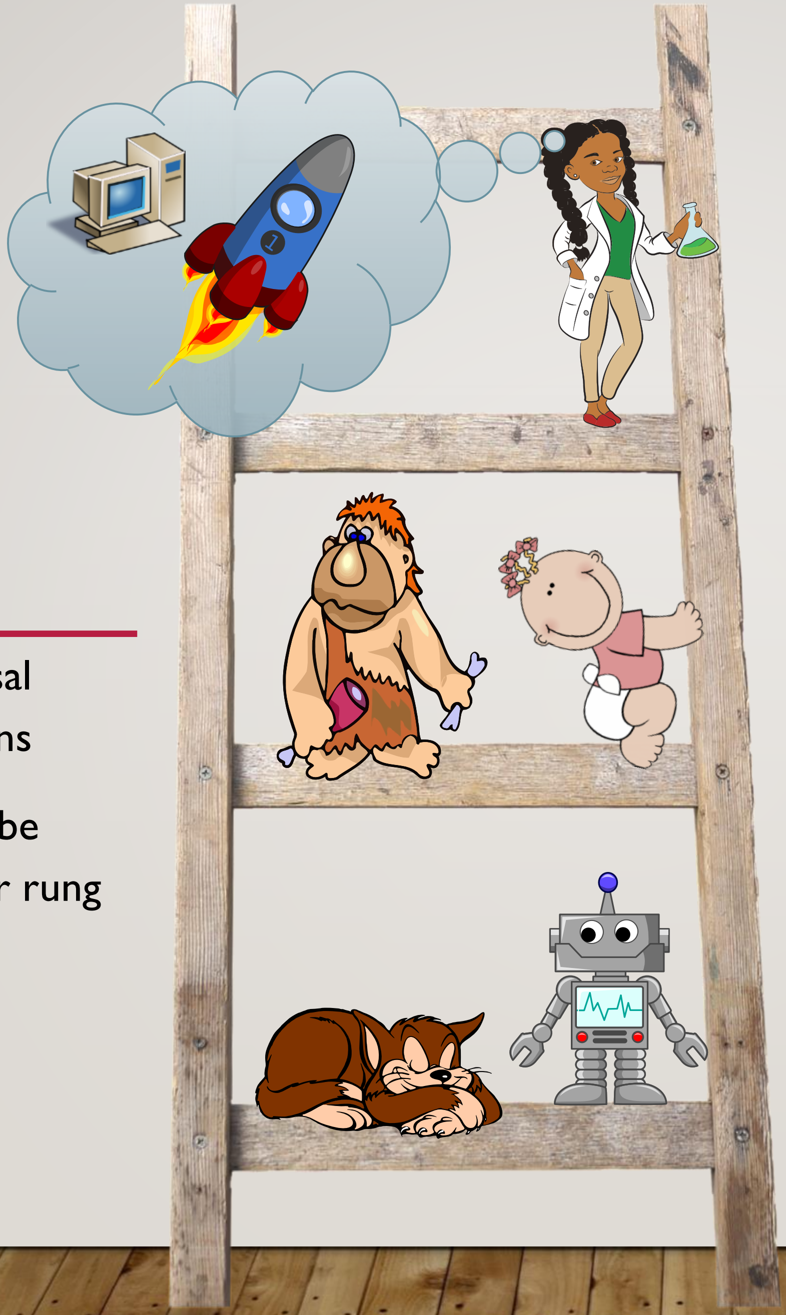
CAUSALITY IN ARTIFICIAL INTELIGENCE?

- Holy Grail: “Strong AI”
  - AI capable of everything human, but better…
  - implies causal thinking, the foundation of human thinking
- State of the Art: “Deep Learning”
  - mostly data-driven
  - decision-making highly unexplainable and based on seen experiences



THE LADDER OF CAUSATION

- metaphor for different causal problem classes or questions
- each rung's problems can't be solved by methods of lower rung
- proven mathematically



COUNTERFACTUALS
(Imagining/Understanding)
"What if I had done..?
Why?"

- Was it X that caused Y?
- What if X had not occurred?
- *Was it the pill that cured my headache?*
- *Would Kennedy be alive, if he hadn't been shot?*

INTERVENTION
(Doing)
"What if I do..? How?

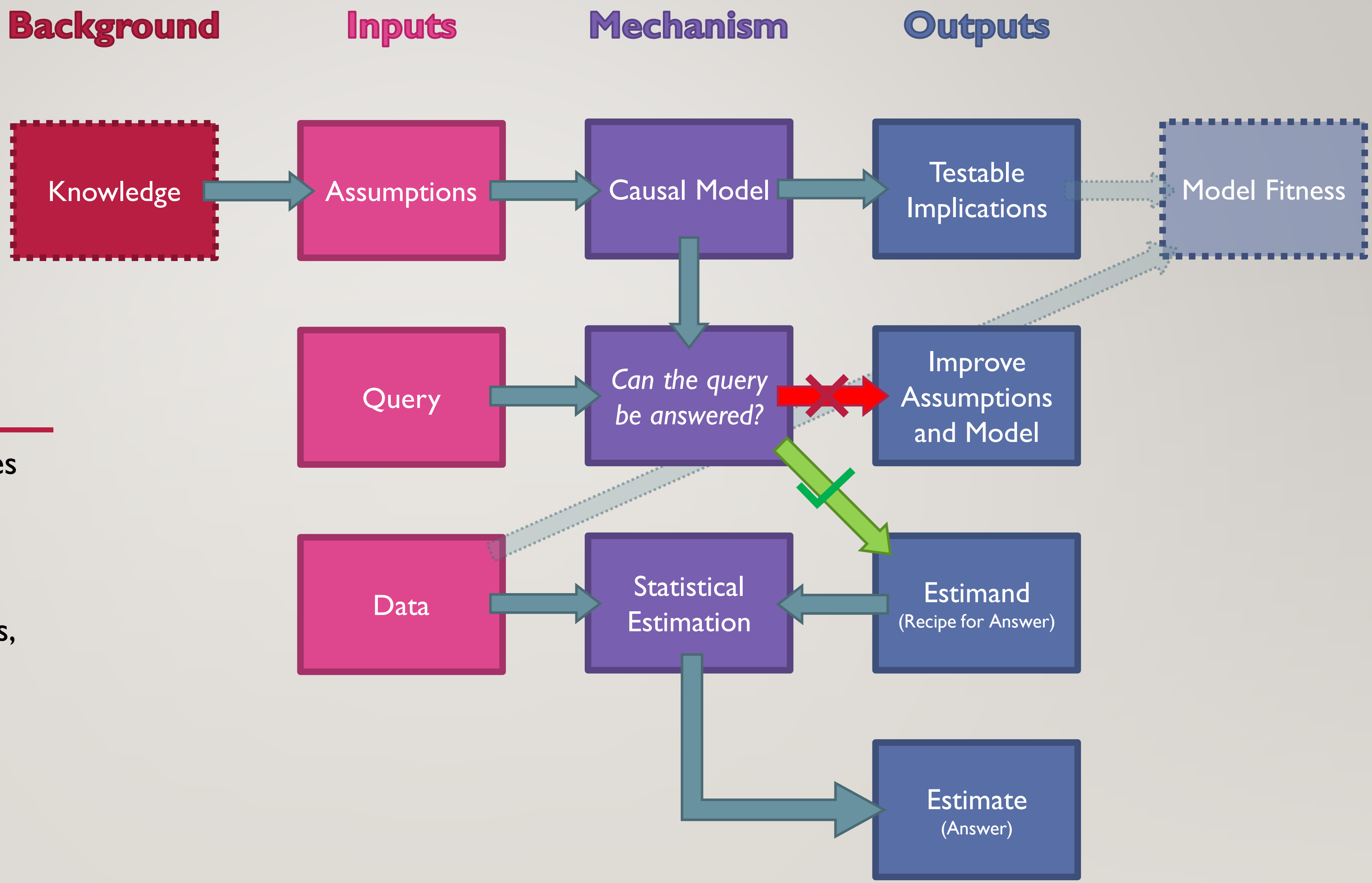
- What would Y be, if I do X?
- How can I make Y happen?
- *Will my headache be cured, if I take Ibuprofen?*
- *What if we ban cigarettes?*

ASSOCIATION
(Seeing)
"What if I see..?"

- How does seeing X change my belief in Y?
- *How likely is a customer who bought toothpaste to also buy dental floss?*
- *What does a symptom tell me about a disease?*

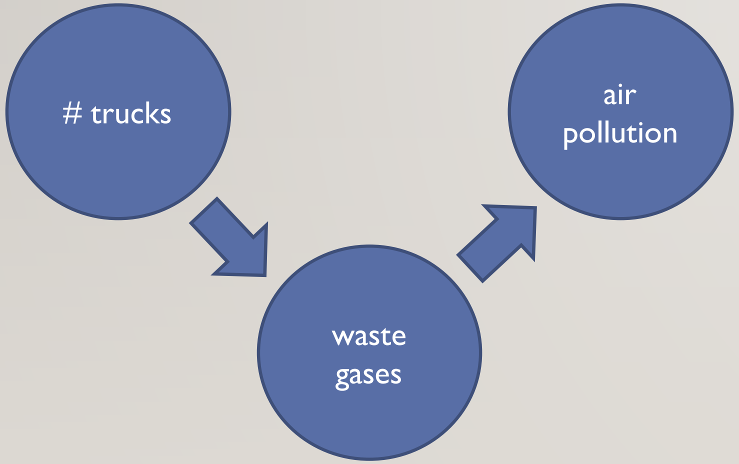
THE INFERENCE MACHINE

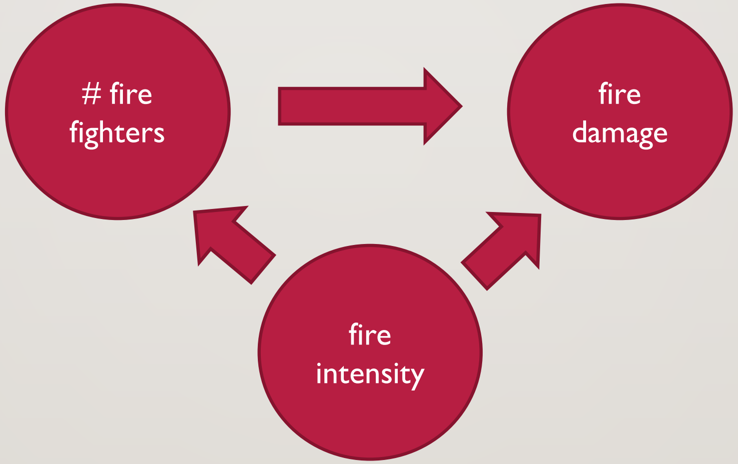
- flowchart on how causal queries should be answered, based on **knowledge** and **data**
- done subconsciously by humans, but clear modelling needed for e.g. implementation in AIs
- uses knowledge to transform queries to "Rung 1"-problems



CAUSAL DIAGRAMS

Causal Model

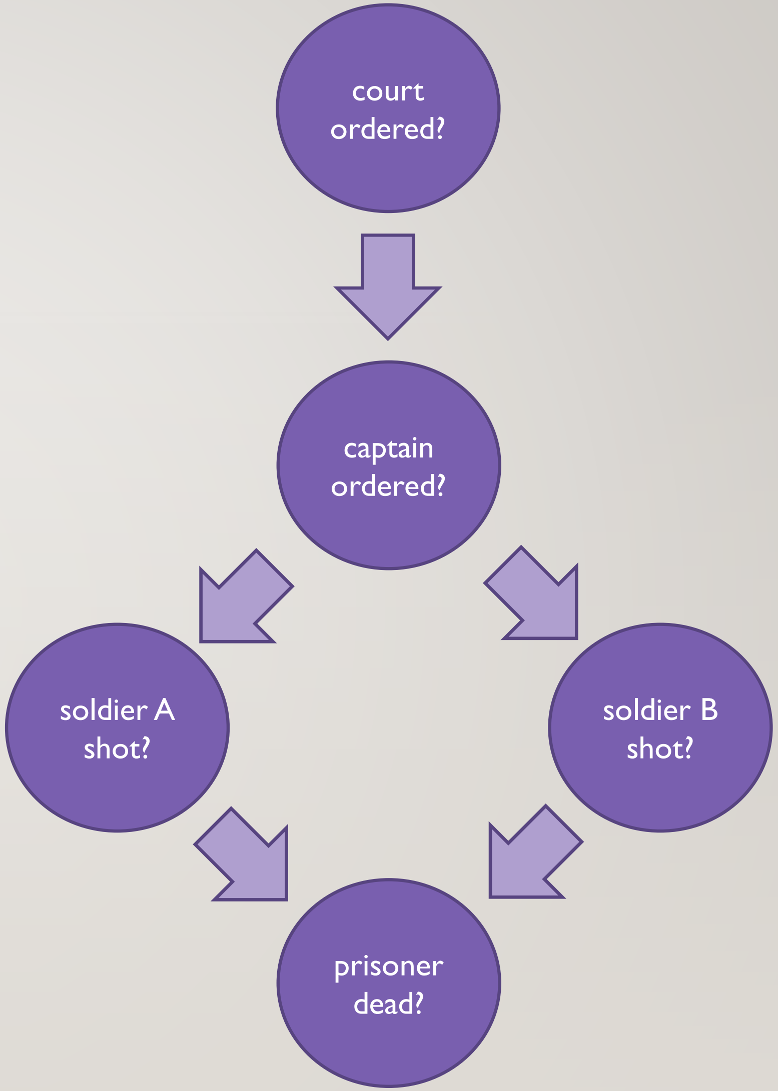




- nodes represent random variables, measuarable or not
- arrows represent causal relationships, from cause to effect
    - can be explicitly defined by (linear) coefficient or any function

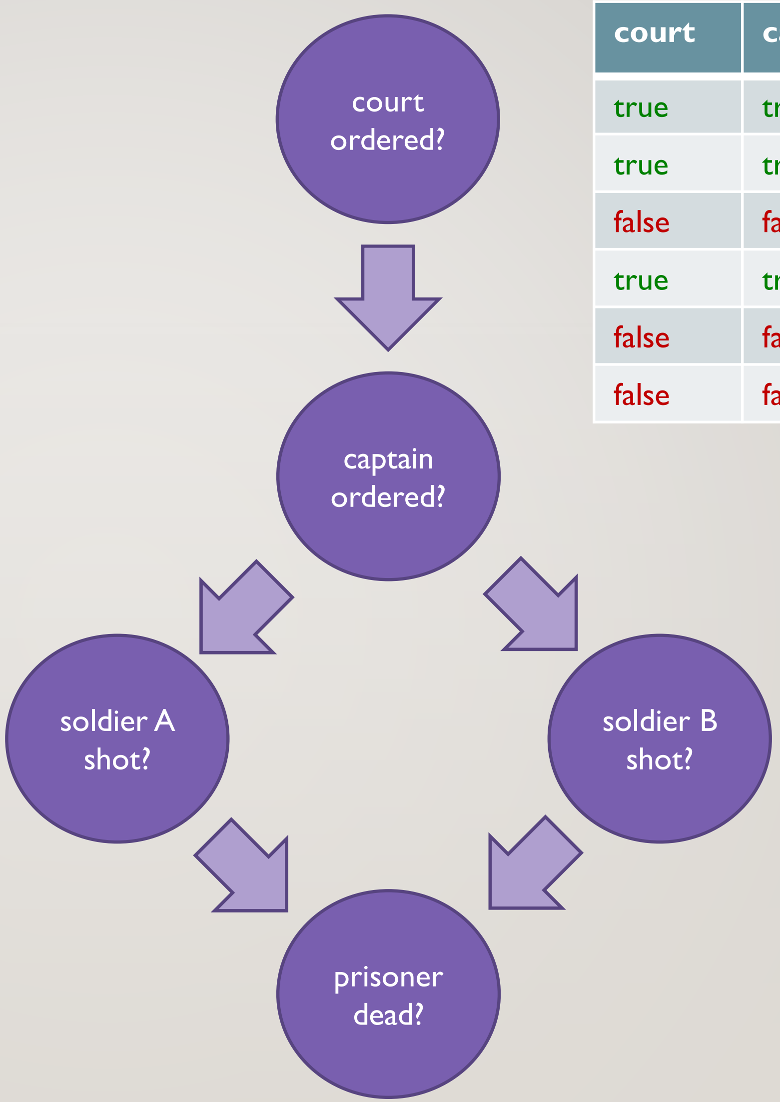
SHOOTING SQUAD (MODEL)

- prisoner shall be executed
- court gives shooting order
- captain orders his soldiers to shoot, iff court order occured
- soldier A and B shoot, iff they recieved order from captain
- prisoner dies, iff at least one bullet hits him
- all variables are boolean [true|false]
- all relationships are the identity



SHOOTING SQUAD (RUNG-1-QUERIES)

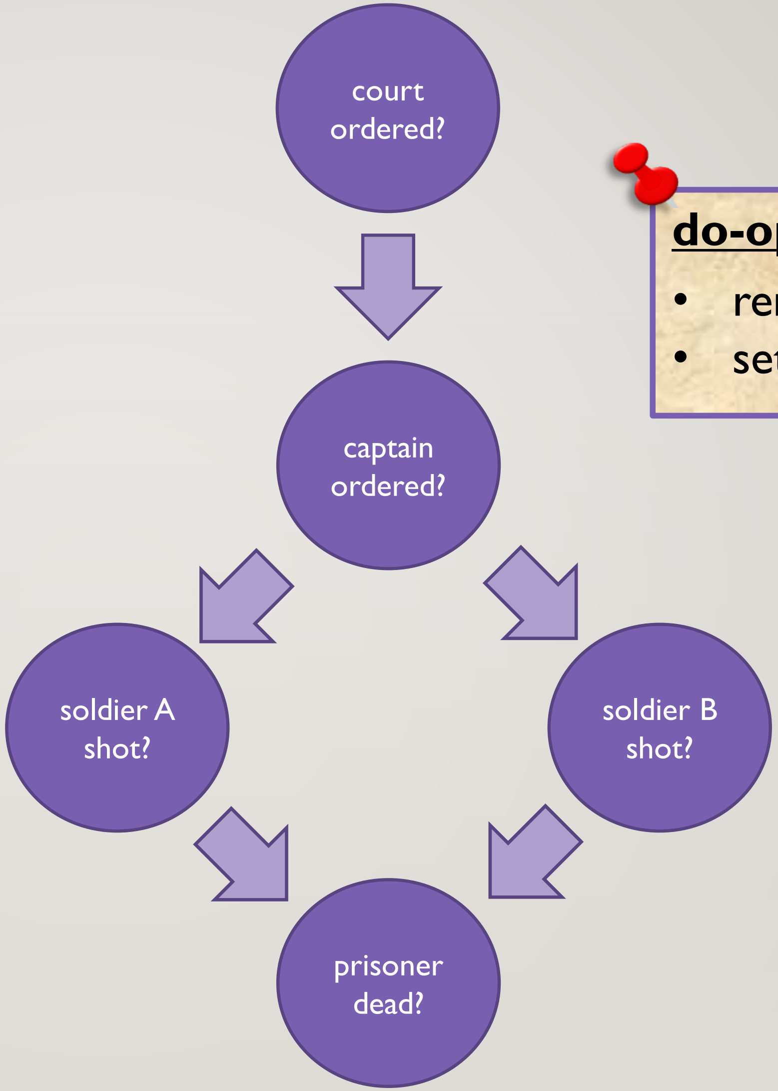
- Is the prisoner dead, when the court gave the order? $P(dead = 1 | court = 1) = 1$
- Did B shoot when A shot? $P(B = 1 | A = 1) = 1$
- 'classic' statistical queries
  - "If I observe X, how likely is it, that Y occurred?"
  - Can be answered by data alone



court	captain	sol. A	sol. B	dead
true	true	true	true	true
true	true	true	true	true
false	false	false	false	false
true	true	true	true	true
false	false	false	false	false
false	false	false	false	false

SHOOTING SQUAD (RUNG-2-QUERIES)

- If A decides to shoot on his own, is the prisoner dead? $P(dead = 1|do(A = 1)) = 1$
  - question breaks the rules, but still seems valid to humans
  - can't be formulated in statistical notation $P(dead = 1|???)$
  - must be formulated using the do-operator $P(dead = 1|do(A = 1))$
- If A decides to shoot on his own, did B shoot, too? $P(B = 1|do(A = 1)) = P(B = 1)$
  - highly unlikely, since most of the time, prisoners aren't shot

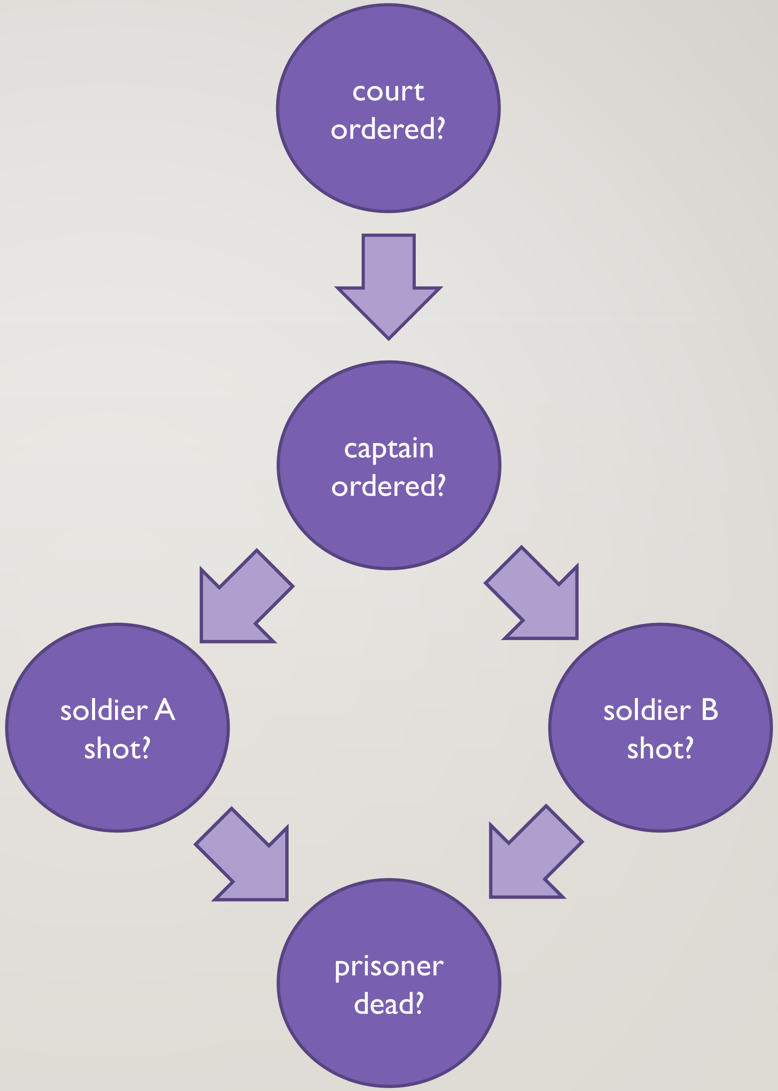


do-operator:

- remove all incoming arrows
- set variable to desired value

SHOOTING SQUAD (RUNG-3-QUERIES)

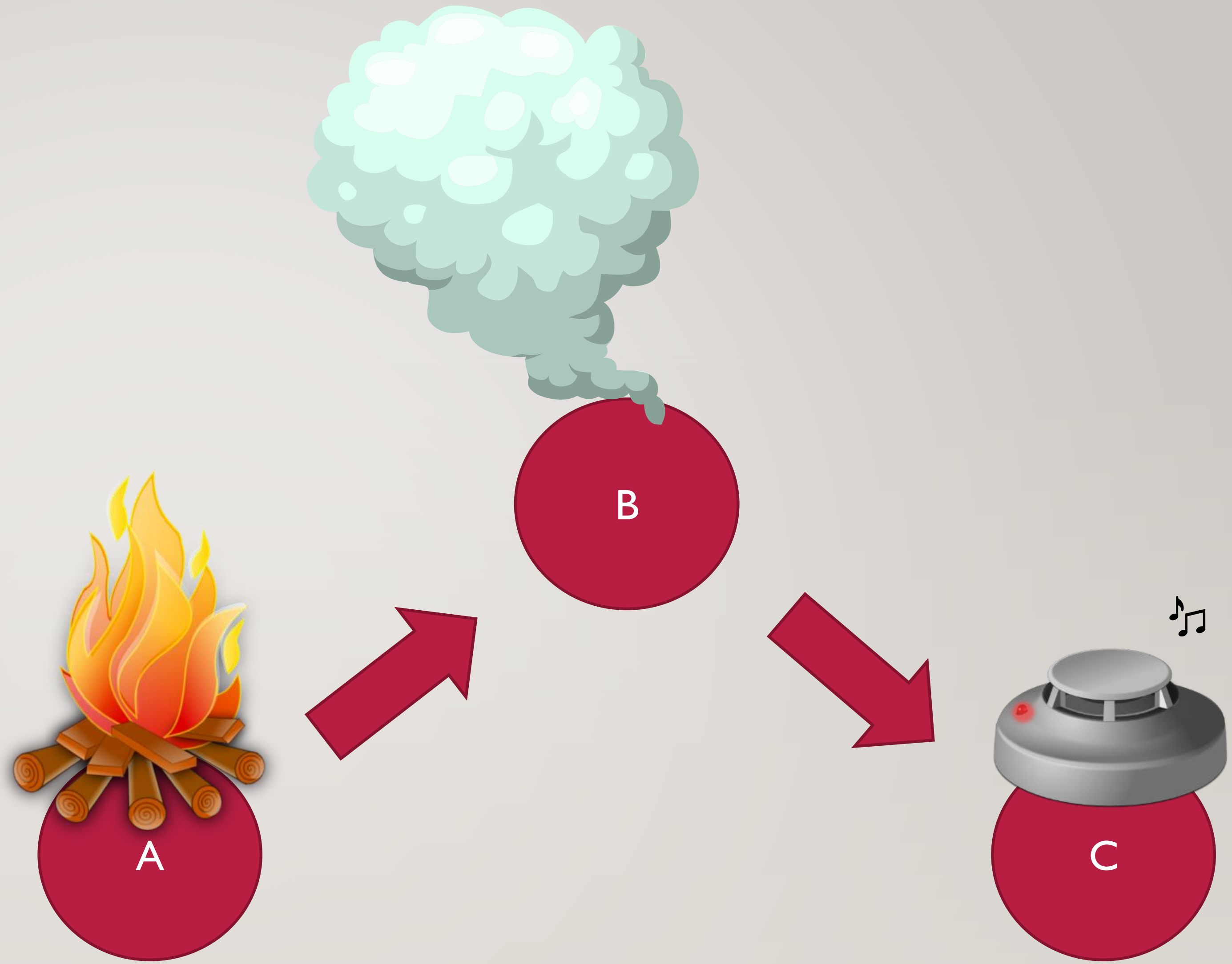
- couterfactual queries (also: potential outcome queries) are about certain individuals (or „worlds“), rather than a whole population
- The prisoner is dead. Would he also be, if soldier A hadn't shot?
  - new notation: $dead_{A=0}(dead = 1)$
  - algorithm (simplified):
    - simulate "normal" situation with our knowledge
    - use do-operator to intervene
    - check for changes



CORRELATION AND CONDITIONING: **CHAINS**

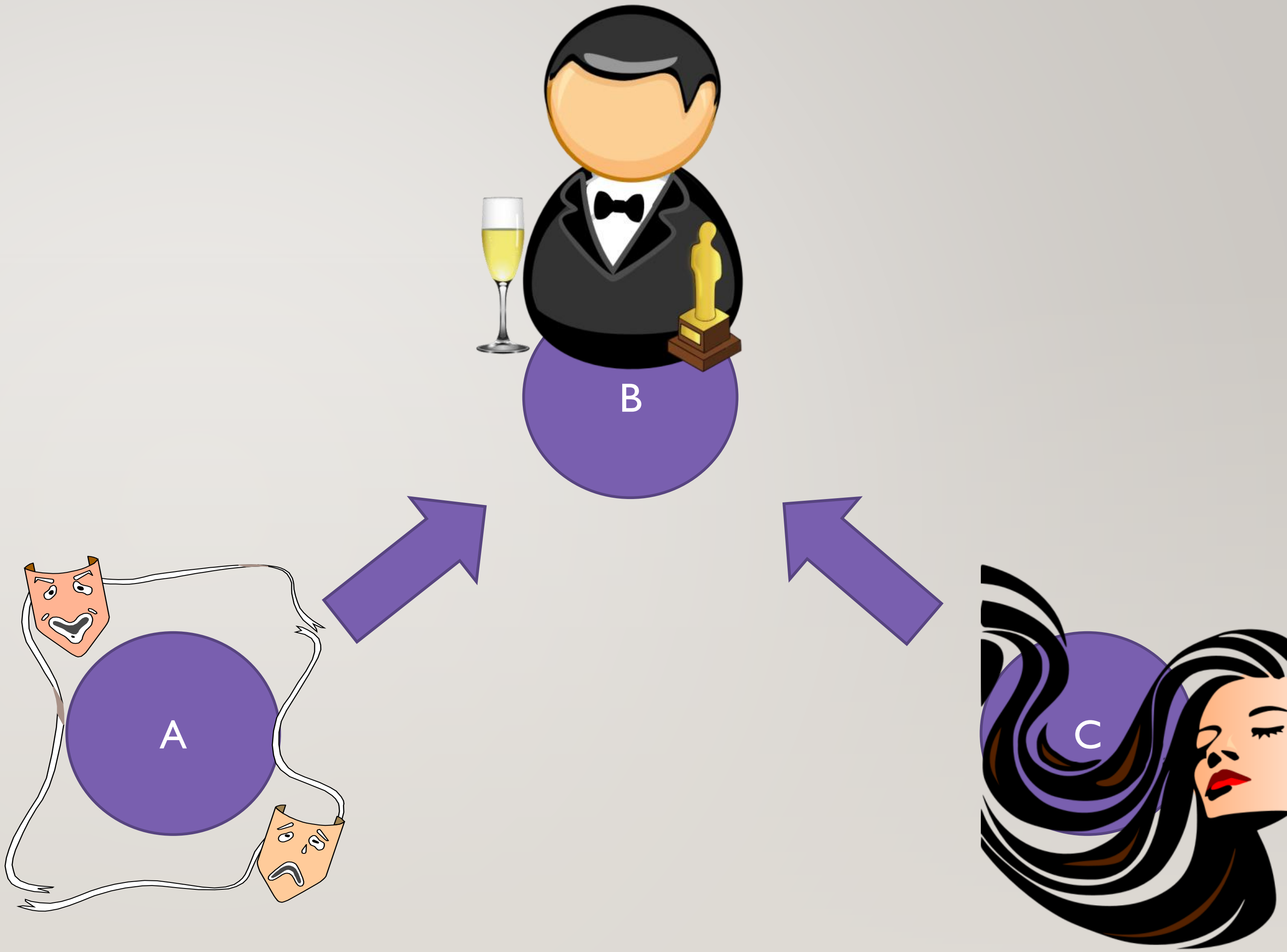
- „real“ correlation between A and C
- B is called **mediator**, which “transports” information from A to C
- fixing B decorrelates A and C
  - fixing called “**conditioning**”
- example fire alarm
  - fire produces smoke
  - smoke triggers the alarm
  - no direct, causal connection **A→C**
  - imagine “fail-chance” of 0.05
  - looking only at scenarios where smoke was present:

$$P(alarm|fire) = P(alarm|\neg fire)$$



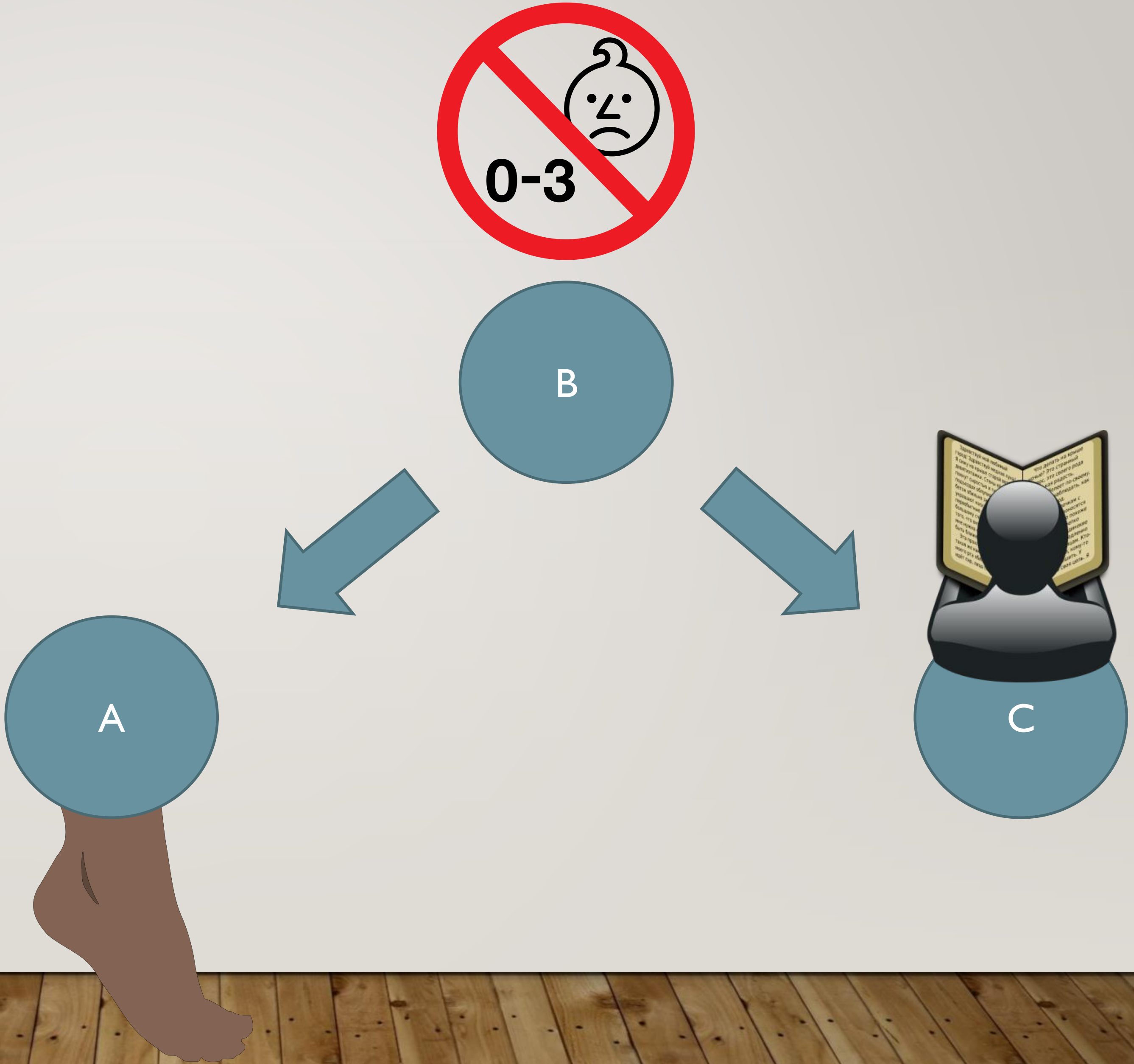
CORRELATION AND CONDITIONING: **COLLIDER**

- no correlation between A and C
- B is an effect of both A and B *(no naming)*
- fixing B correlates A and C
- example with Hollywood actors
  - B is rank on a list of most famous actors
  - getting famous is caused by talent and good looks
  - looking only at a certain segment of the rank list shows:
    - pretty people tend to be untalented
    - good actors tend to be unattractive
  - (of course this is highly simplified!)
- called "explain-away" effect



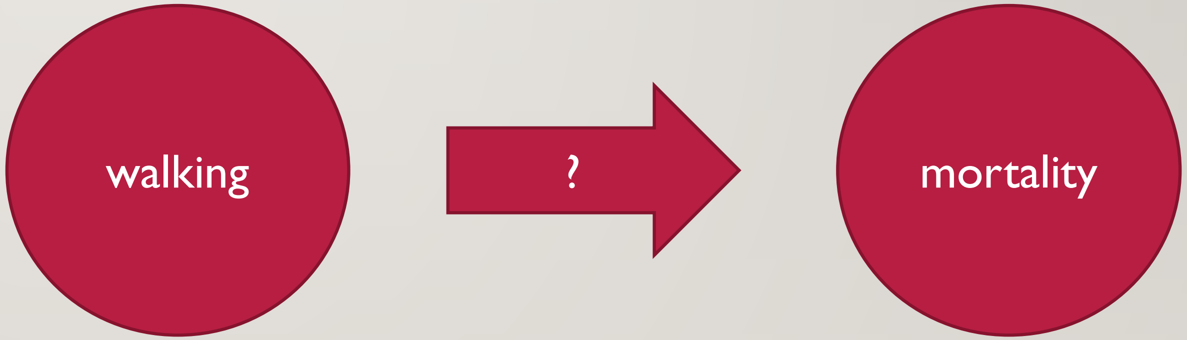
CORRELATION AND CONDITIONING: **FORK**

- „spurious" correlation between A and C
- B is called **confounder**, which is a common cause of A and C
- fixing B decorrelates A and C
- example with children
  - children with bigger feet tend to read better, which is obviously nonsense
  - but both are highly affected by age
  - by looking at a certain age group (stratum), the correlation vanishes



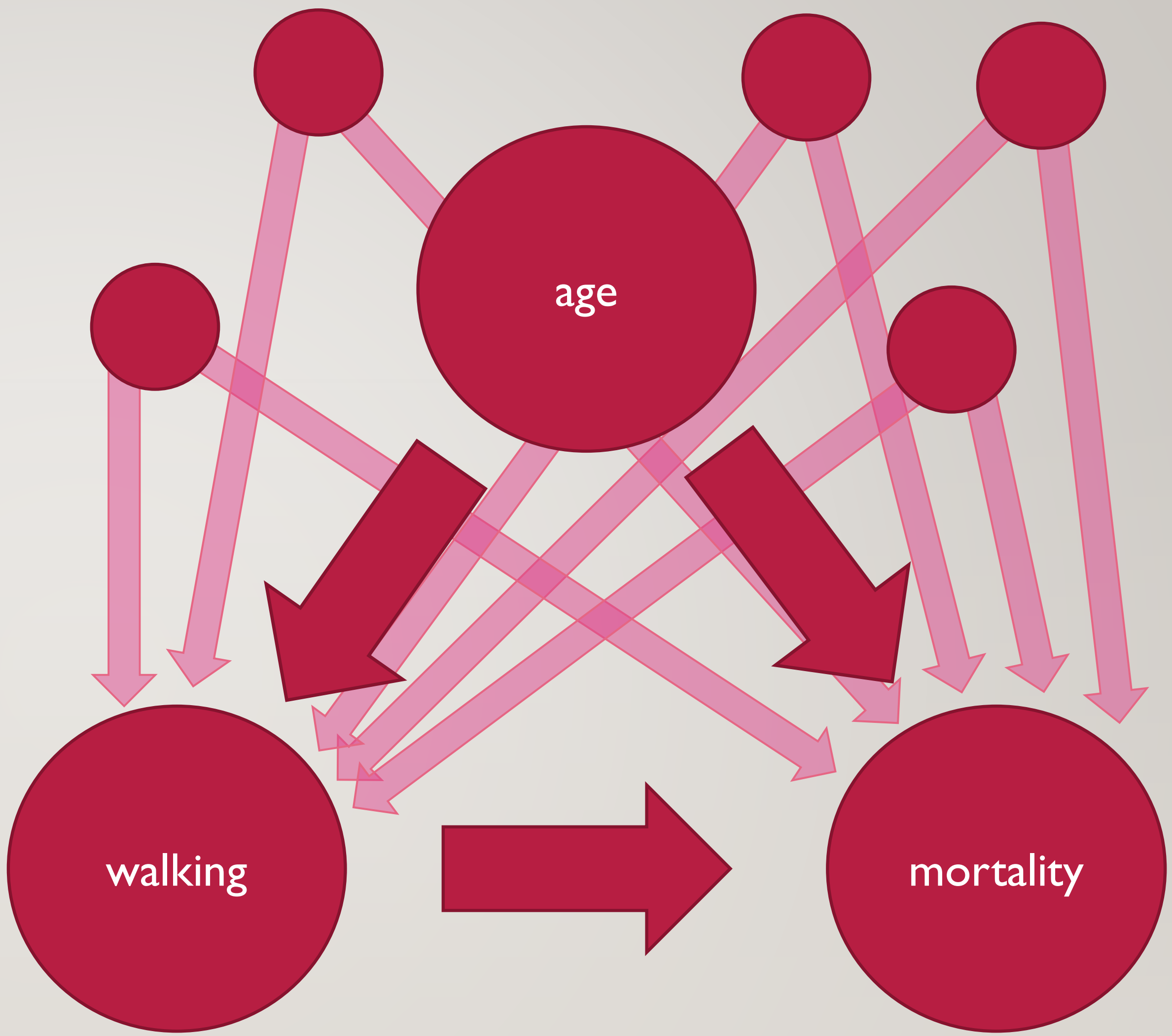
CONFOUNDERS: THE LURKING VARIABLE

- Study in *New England Journal of Medicine* led by *Robert Abbott* about effect on walking on average lifespan
  - *Intense Walkers* walked >2 miles/day
  - *Casual Walkers* walked <1 mile/day
  - all subject are from the same region (JP)
  - after 12 years, **43%** of the *Casual Walkers* died, while **21.5%** of the *Intense Walkers* died
  - since walking preferences were not prescribed, walking and mortality might have common cause



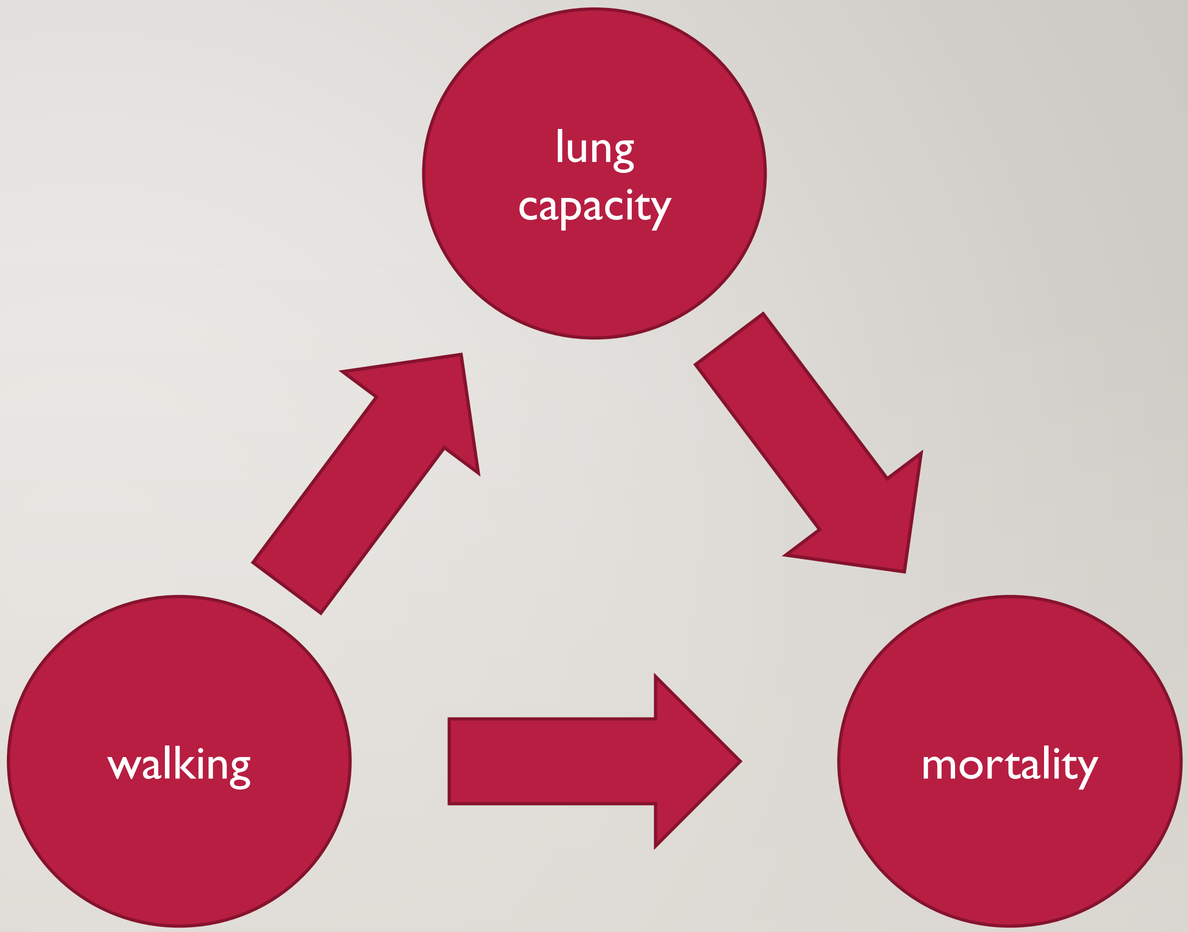
CONFOUNDERS:
THE LURKING VARIABLE

- after 12 years, **43%** of the *Casual Walkers* died, while **21.5%** of the *Intense Walkers* died
- since walking preferences were not prescribed, walking and mortality might have common cause
  - e.g. higher age, which prevents walking due to physical reasons and leads to earlier death
  - can be deconfounded by conditioning
  - age-adjusted values: 41% for *Casual Walkers*, 24% for *Intense Walkers*
  - researches also tried adjusting for physical condition, alcohol consumption, diet, etc.



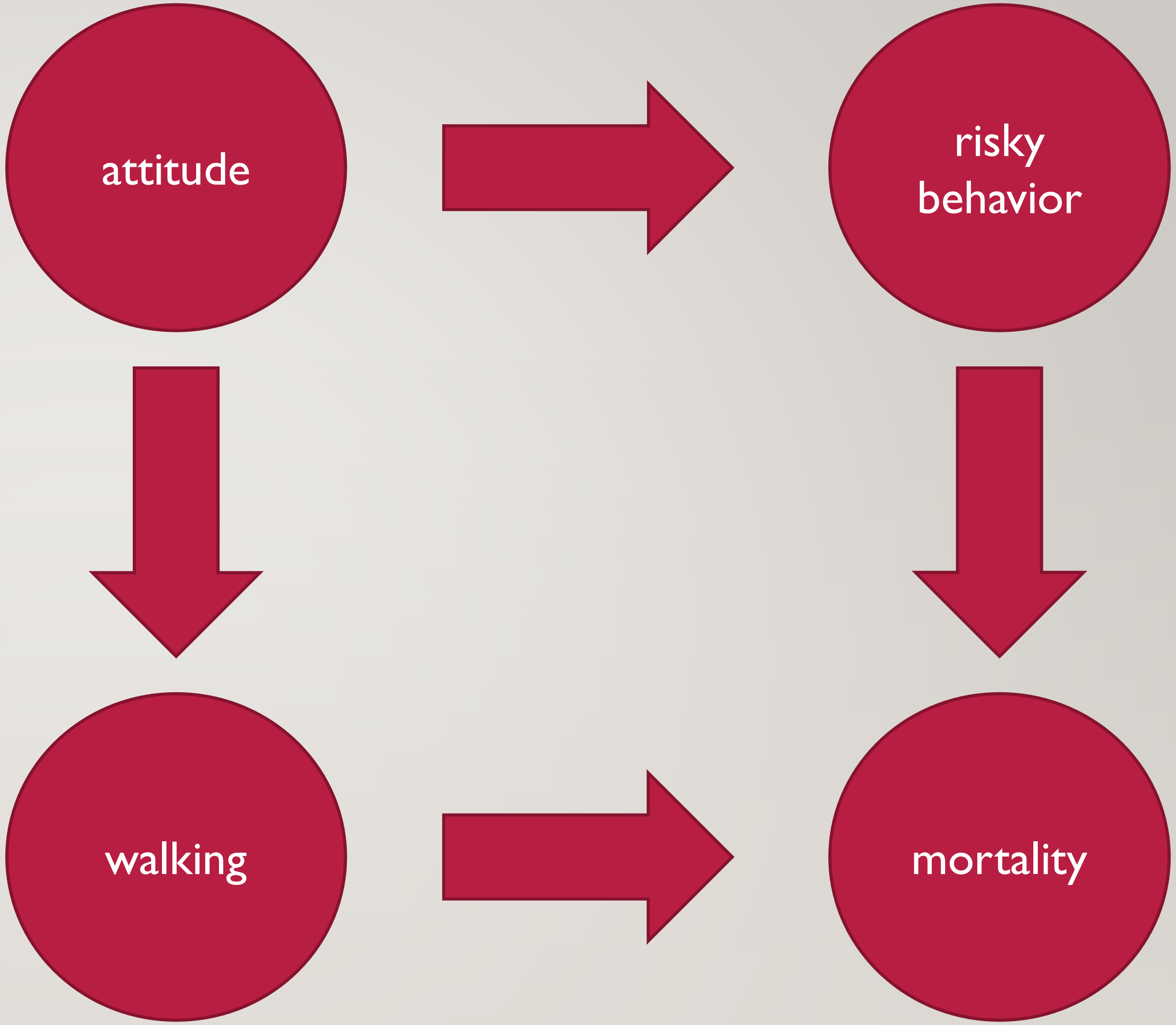
IS OVERDECON-FOUNDING POSSIBLE?

- short answer: **yes**
- imagine e.g. conditioning for „**lung capacity**“ *(fictional example)*
- might be a common cause, since it makes walking less exhausting and improves oxygen supply
- might also be a mediator, since it improves through cardio
- conditioning on mediator will shield off some of the effects *(remember Fire → Smoke → Alarm)*



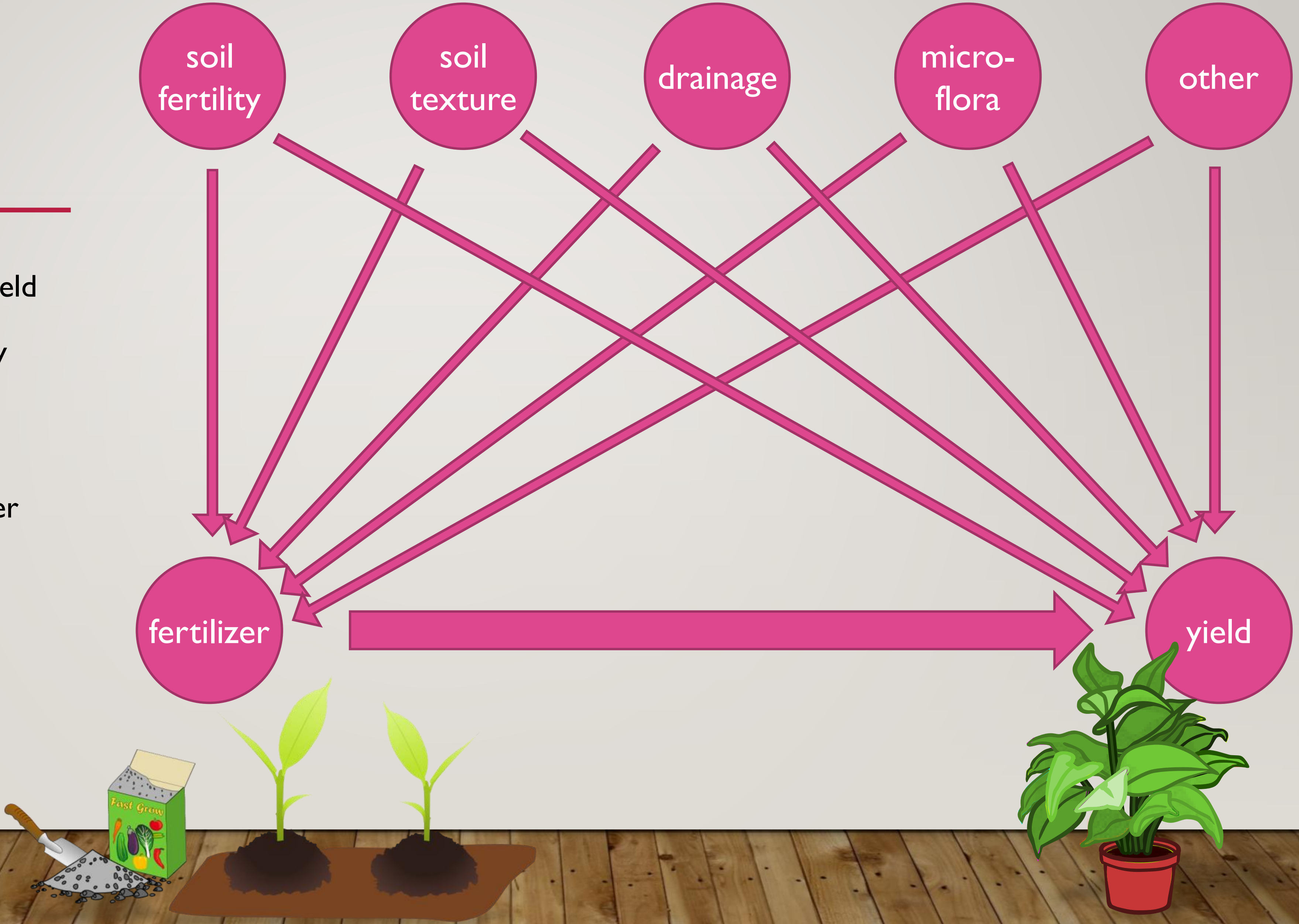
THE WORST CONFOUNDER?

- many aspects in human life are (partially) an effect of personal preferences
- cannot be measured and thus not be conditioned for in the data
- confounders with these properties need to be eliminated in another way



THE BEST FERTILIZER?

- a farmer wants to now which fertilizer gets him the most yield
- plant growth is determined by many factors
- (un)concious biases help the farmer to decide on a fertilizer based on his environment



THE BEST FERTILIZER?

What do we want to know?

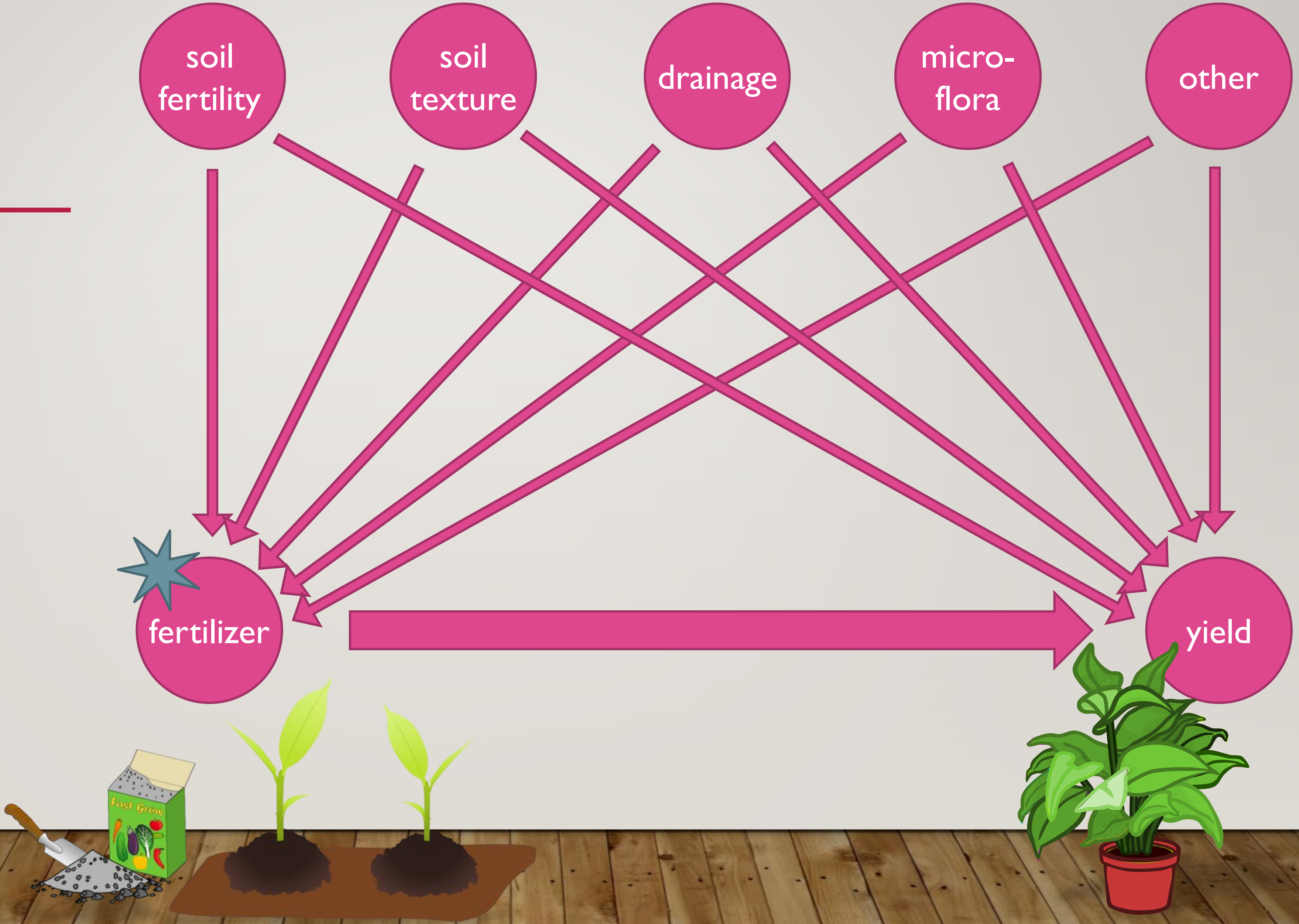
We need to compare

$P(yield|do(fertilizer = 1))$

to

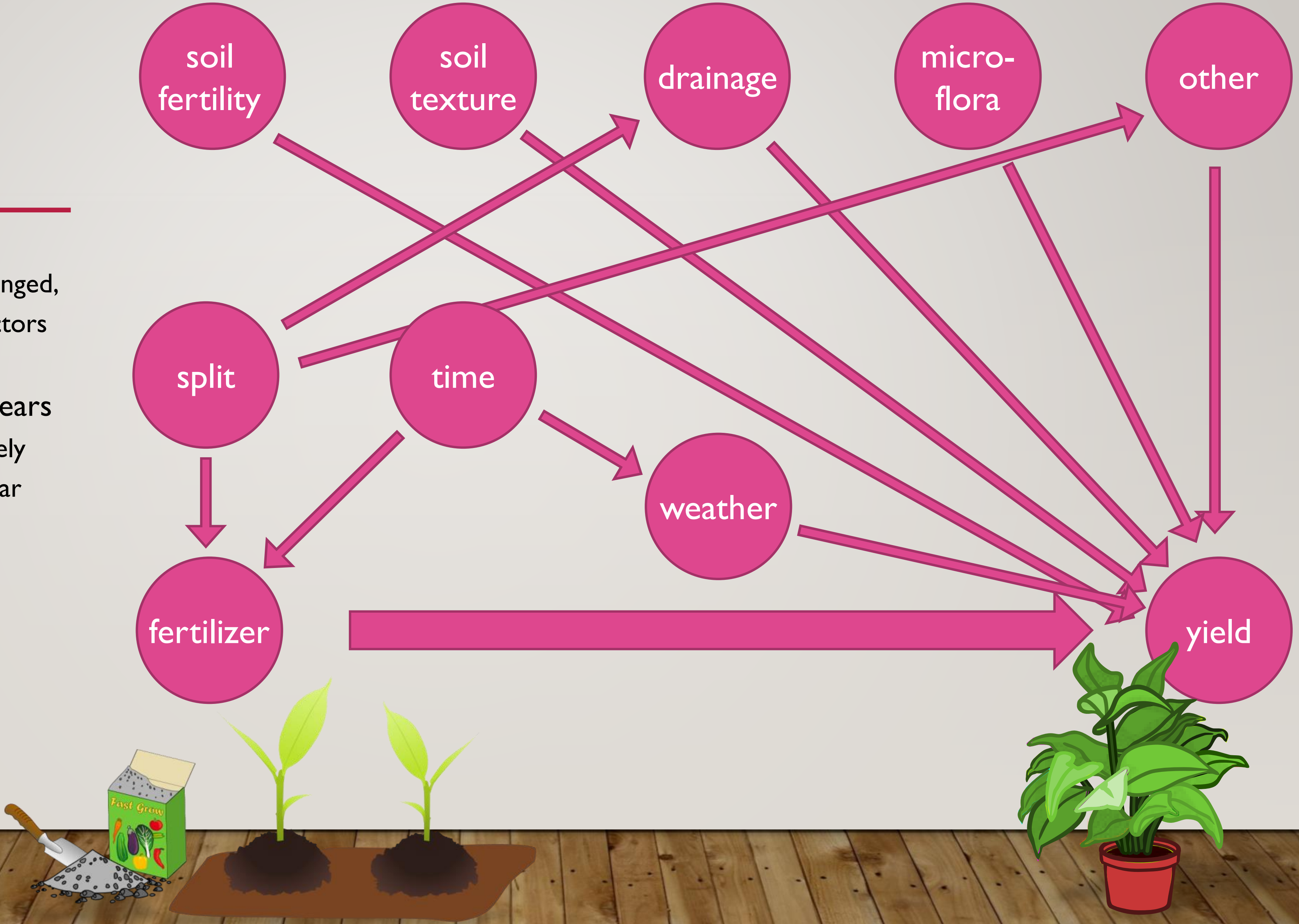
$P(yield|do(fertilizer = 2))$

But how to do that in real life?



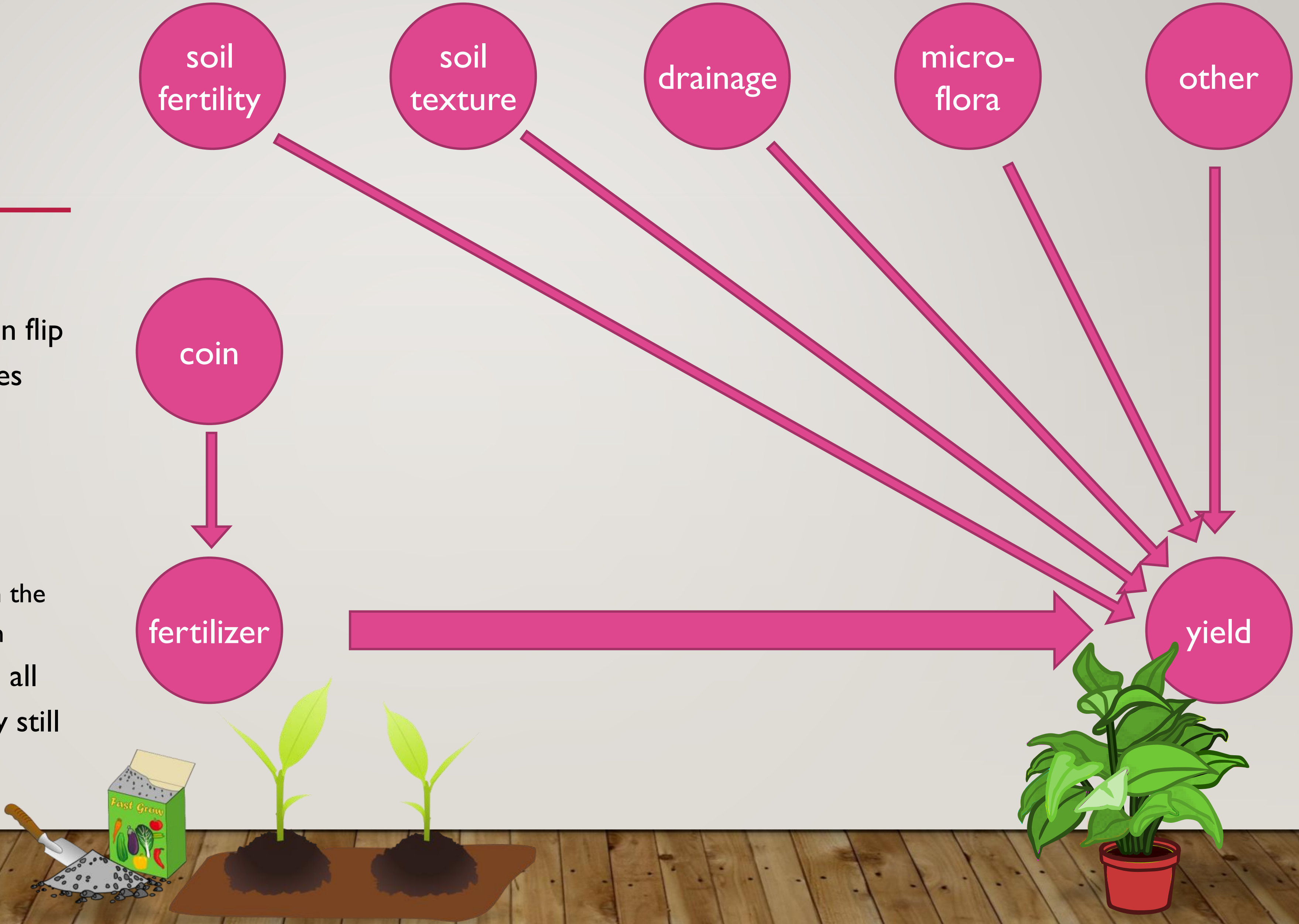
THE BEST FERTILIZER?

- 1st attempt: split field in two
  - even if the whole soil is changed, some location dependet factors will remain
- 2nd attempt: try in different years
  - at least weather will be surely different in the following year



THE BEST FERTILIZER?

- split field into **many** parts
- for each part, flip a coin
- choose fertilizer based on coin flip
- randomness removes all causes without adding new ones
- this is the idea of randomized controlled trials (RCTs)
  - good way of simulating an unconfounded model, when the confounders are not known
- randomization not possible in all settings, so causal analysis may still be required



FORMAL CRITERION FOR CONFOUNDING

- we can only meassure the value of Y for given values of X

$$P(Y|X)$$

- we want to know the causal effect of a fixed X on Y

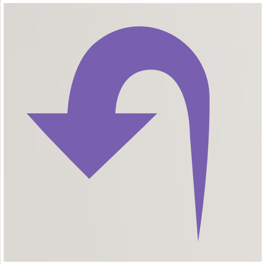
$$P(Y|do(X))$$

- if a variable causes both X and Y, those two differ, so confounding means

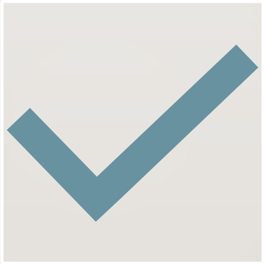
$$P(Y|X) \neq P(Y|do(X))$$

this was skipped in the presentation

BACKDOOR-CRITERION



backdoor-path: undirected path from X to Y, that starts with an arrow into X

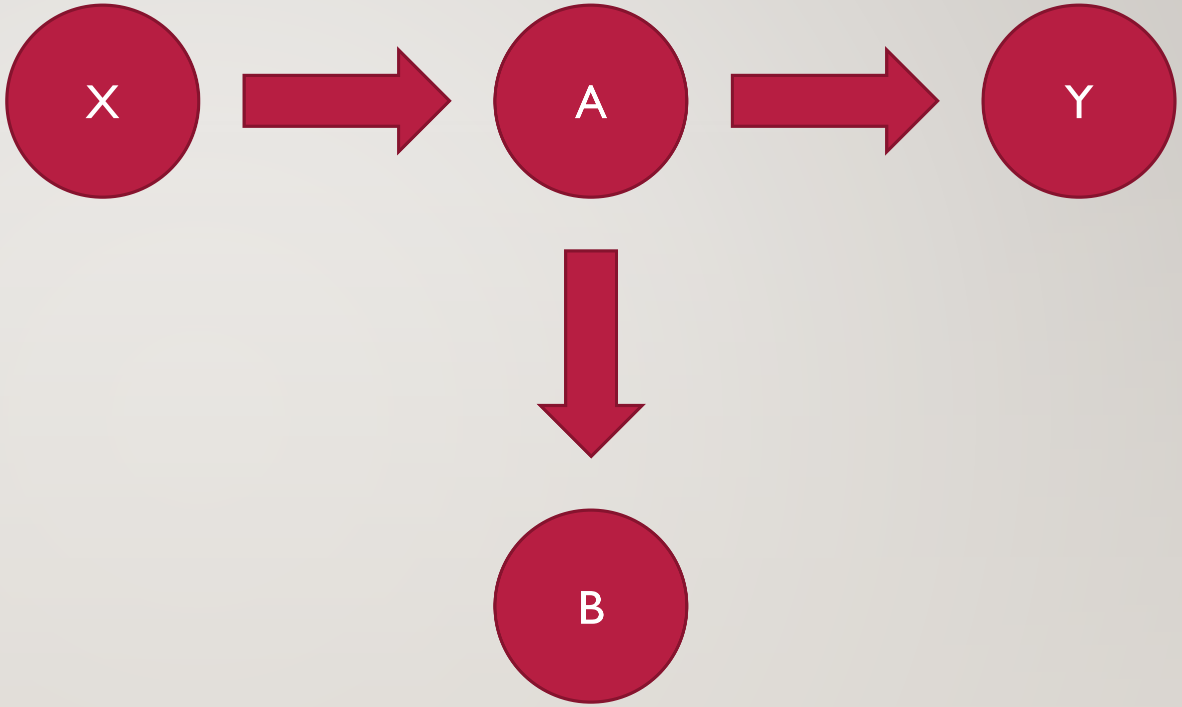


relation is unconfounded, if no unblocked backdoor-paths exist

blocking backdoor-paths must not be done bei conditioning on decendants of X

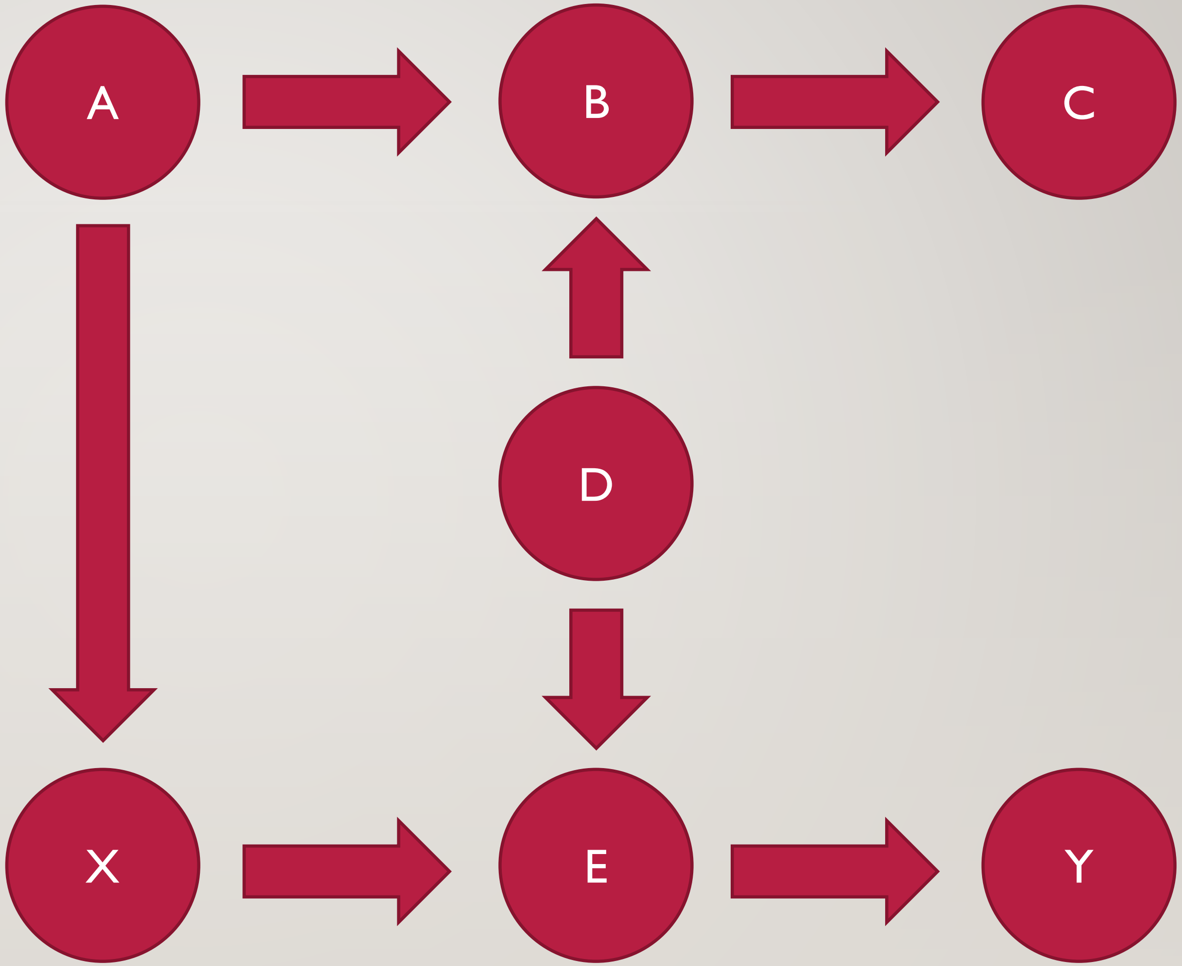
EXAMPLE 1

- meassure Y given X
- no paths into X
  - no confounding
- nothing to do here



EXAMPLE 2

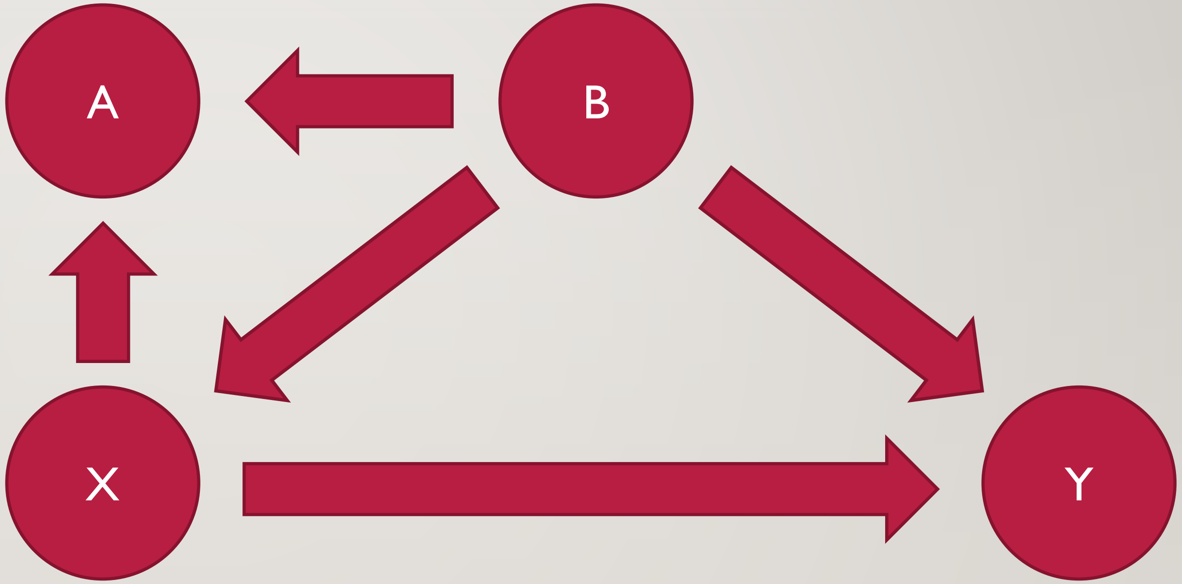
- meassure Y given X
- backdoor-path exists:
  X ←A→*B* ←D →*E* →*Y*
- already blocked by collider
  A→***B*** ←**D**
- nothing to do here



this was skipped in the presentation

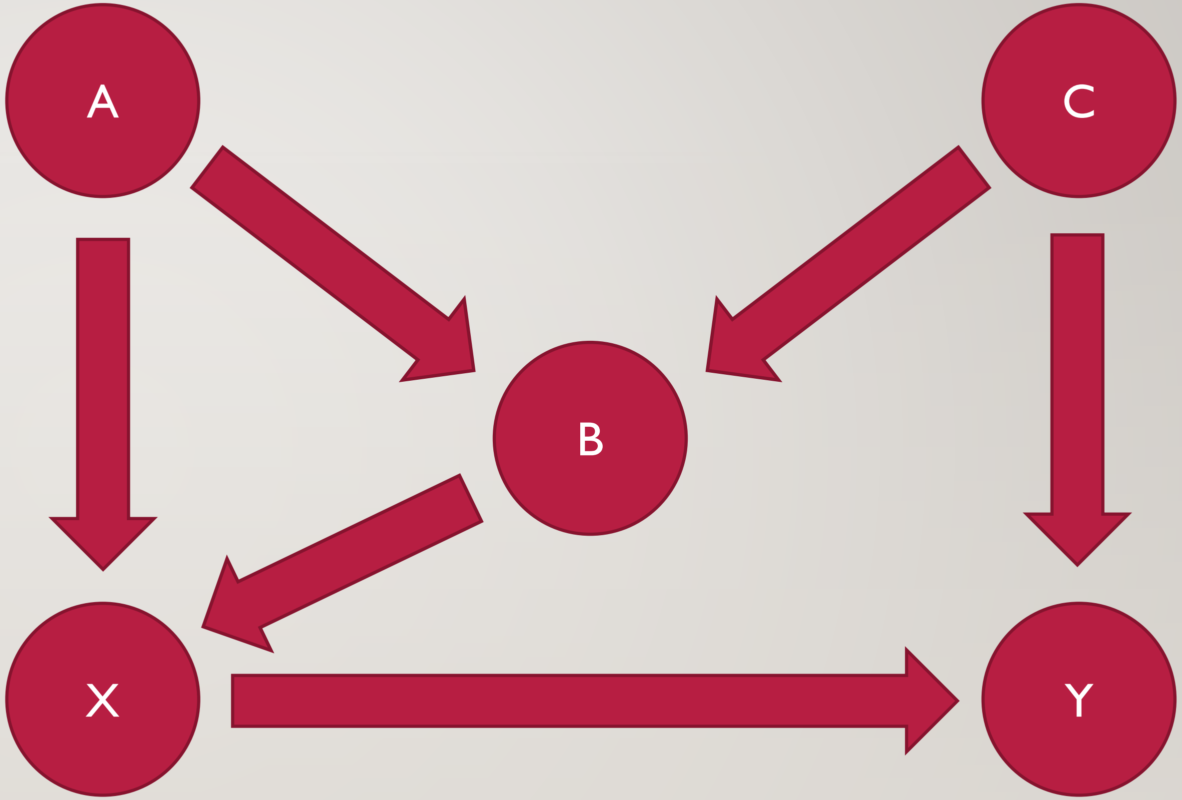
EXAMPLE 3

- meassure Y given X
- backdoor-path exists:
  $X \leftarrow B \rightarrow Y$
- blockable by conditioning for B, since it is a **fork**
- if B is unobservable, the true effect can‘t be meassured



EXAMPLE 4

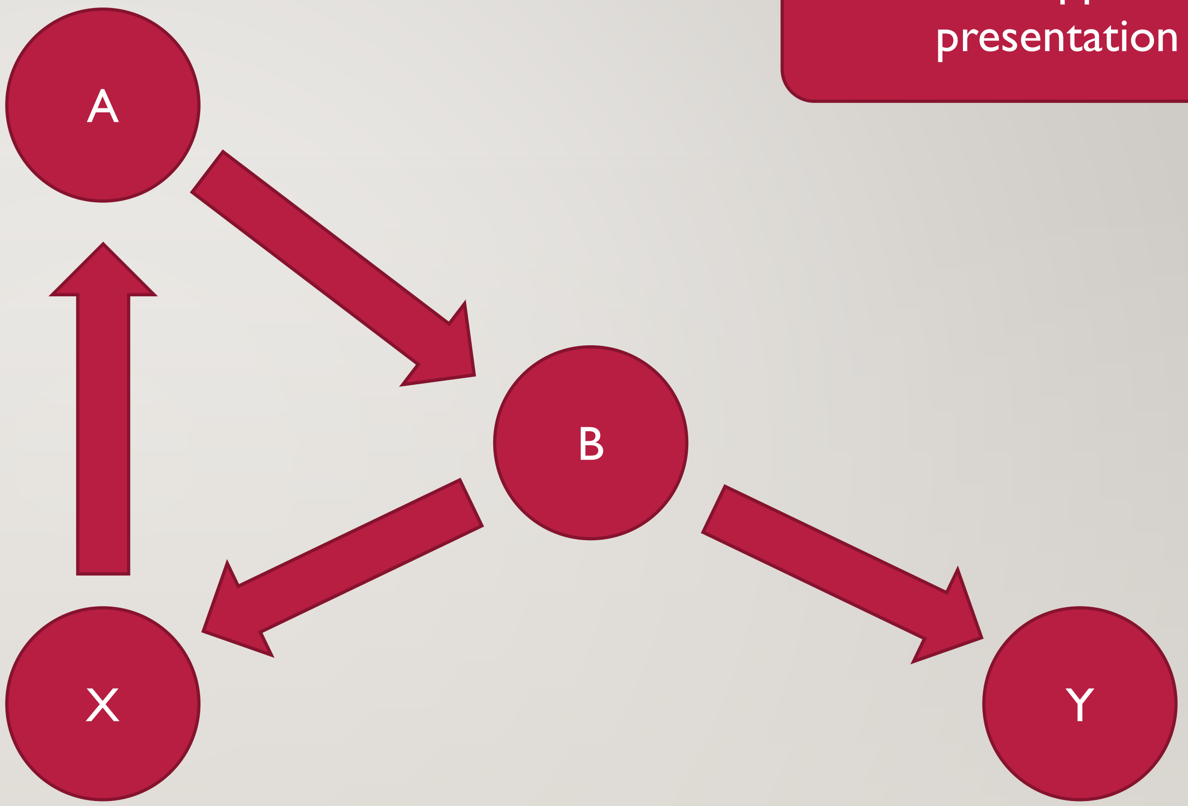
- measure Y given X
- 2 backdoor-paths exists:
  (1) X←A →B ←C → Y
  (2) X←B →C → Y
- (2) blockable via B [collider], but that opens (1), which conditioning on A or C closes that path
- conditioning on C alone would suffice



this was skipped in the presentation

EXAMPLE 5

- measure Y given X
- backdoor-path exists:
  $X \leftarrow B \rightarrow Y$
- can‘t control for B, since it is not only confounder, but also causal decendant of X
- backdoor-criterion not usable

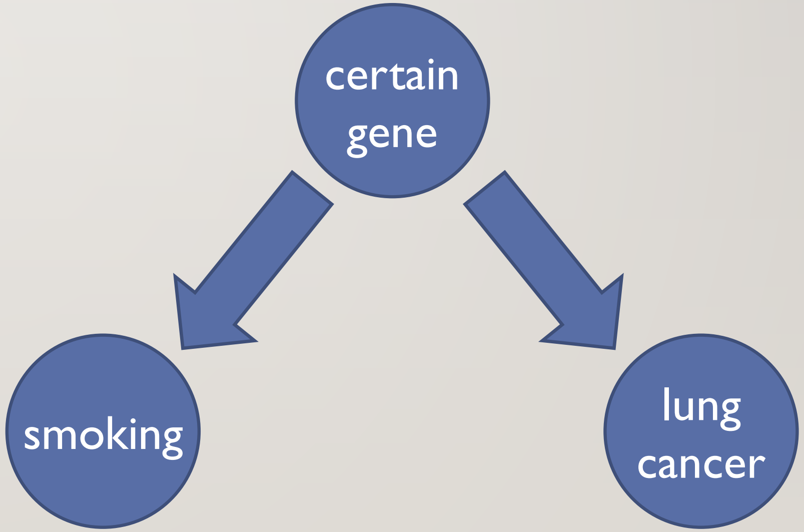


DOES SMOKING CAUSE LUNG CANCER?

ONE EXTREME:

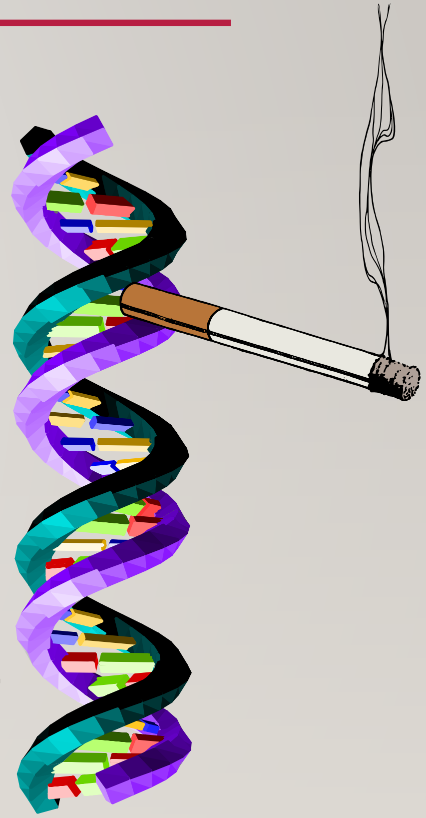
- RCTs not realizable in an ethical manner
- so, statistics couldn't find causation

THE OTHER EXTREME:



CORNFIELD‘S INEQUALITY

- $P(lung\ cancer|smoker) \approx 9 * P(lung\ cancer|\neg smoker)$
- assume, there is a gene that fully accounts for that
  - ⇒ that gene occurs 9 times more often in smokers than in none-smokers
  - ⇒ if 11% of non-smokers have the gene, 99% of all smokers have it
  - ⇒ it is mathematically impossible, that more than 11% have the gene
- highly implausible, that the gene is so tightly liked to ones decision to smoke
- knowledge today: gene exists, effect is much smaller than the direct causal effect



CONSEQUENCES (HILL'S CRITERIA)

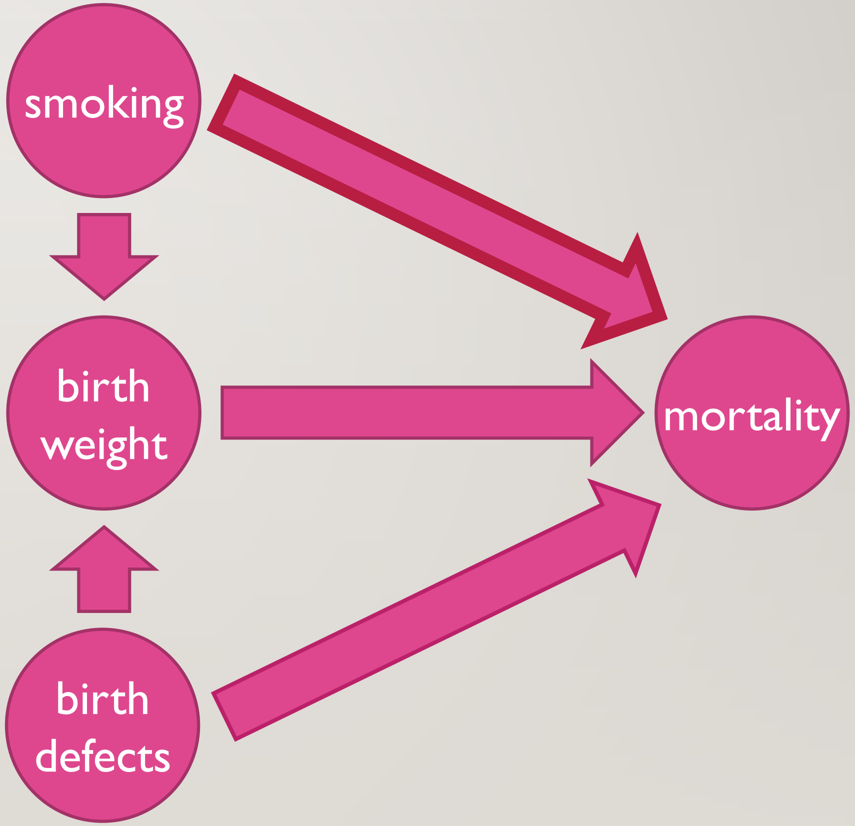
Committee: „Statistical methods **cannot** establish proof of a causal relationship in an association. The causal significance of an association is a matter of judgment which **goes beyond any statement of statistical probability**."

- five criterions, not necessary, not sufficient
    1. consistency
    2. strength of association
    3. specificity of association
    4. temporal relationship
    5. coherence
- four more in a later summary

BIRTH-WEIGHT PARADOX (UNSOLVED UNTIL 2006)

„smoking reduces child mortality, if the baby is born underweight“

- lower birth weight increases mortality
- smoking reduces birthweight
- other serious defects lower birth weight and increase mortality
- controlling for birth weight introduces collider bias and creates spurious correlation



WHERE DOES THE DATA FIT IN?

this was skipped in the presentation

- causal diagrams: structure of the data (so far)
- calculating results needs data
- one approach: **Bayesian Networks**
  - same structure as causal diagrams
  - each node stores probability ditribution for its values given the values of its parents
  - arrows don‘t model causal relationships, but direction of **forward probability**

FORWARD VS. BACKWARD PROBABILITY

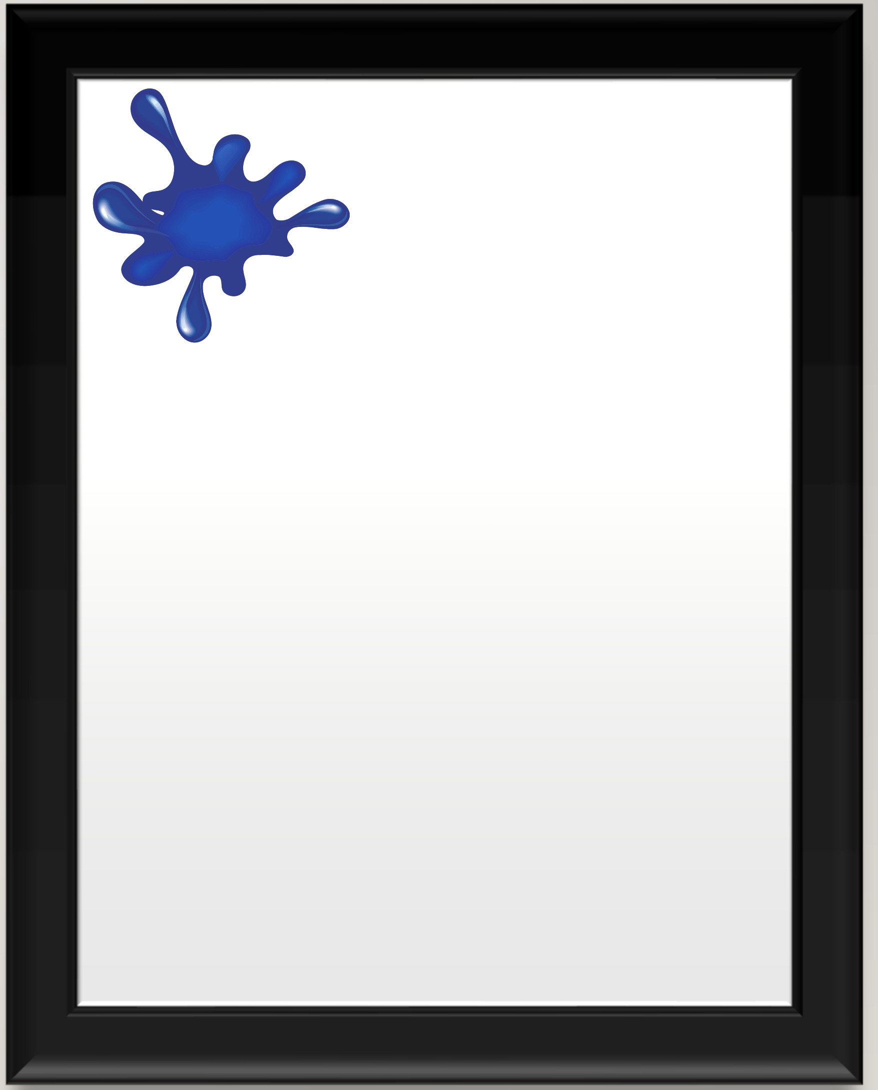
- imagine a canvas is shot with a paintball marker
  - assume hit chance is 100% and position is uniformly distributed
- what‘s the probability of the shot hitting in the upper 10cm of the canvas?

$$\boldsymbol{P(x \leq 10cm | H = h\ cm) = 10/h}$$

- what‘s the probability of the canvas being h cm high, if the shot landed on the upper 10cm?

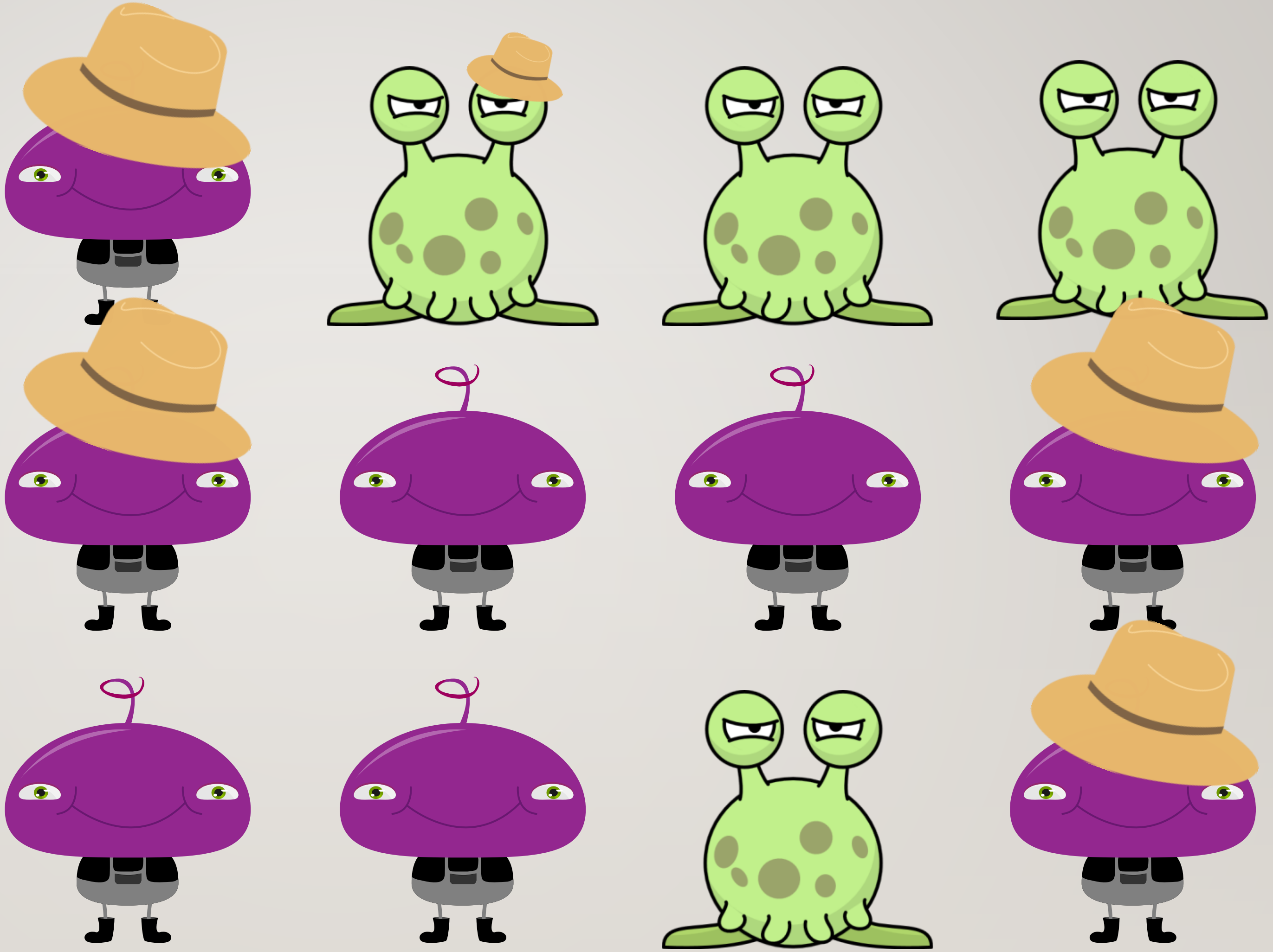
$$\boldsymbol{P(H = h\ cm | x \leq 10cm)} = ???$$

  - much harder problem, since it also requires “experience” about the world.
  - how large are canvases “usually”? 20-30cm? 70-80m?



RULE OF BAYES

- what are the odds that an alien is purple and wears a hat?
- first approach:
  - only look at purple aliens $(^2/_3)$
  - count the ones with a hat $(^1/_2)$
- second approach:
  - look at aliens with hats$(^5/_{12})$
  - count the purple ones $(^4/_5)$



RULE OF BAYES

- what are the odds that an alien is purple and wears a hat?
- first approach:
  - only look at purple aliens $({}^{2}/_{3})$
  - count the ones with a hat $({}^{1}/_{2})$
- second approach:
  - look at aliens with hats$({}^{5}/_{12})$
  - count the purple ones $({}^{4}/_{5})$

$$\frac{2}{3} * \frac{1}{2} = \frac{2}{6} = \frac{1}{3}$$

$$\frac{5}{12} * \frac{4}{5} = \frac{4}{12} = \frac{1}{3}$$

$$P(purple \wedge hat) = P(purple) * P(hat|purple)$$

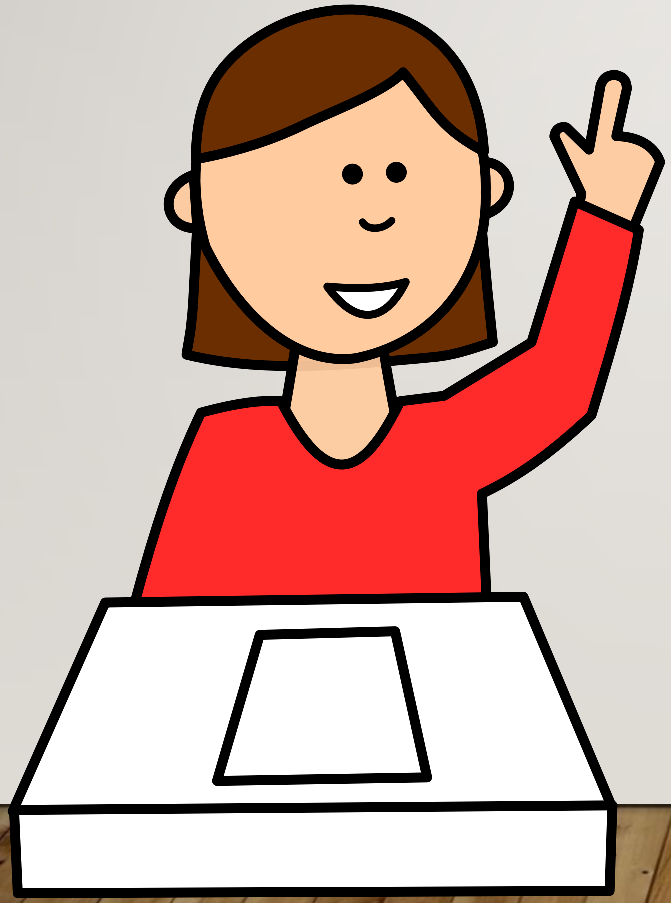
$$P(purple \wedge hat) = P(hat) * P(purple|hat)$$

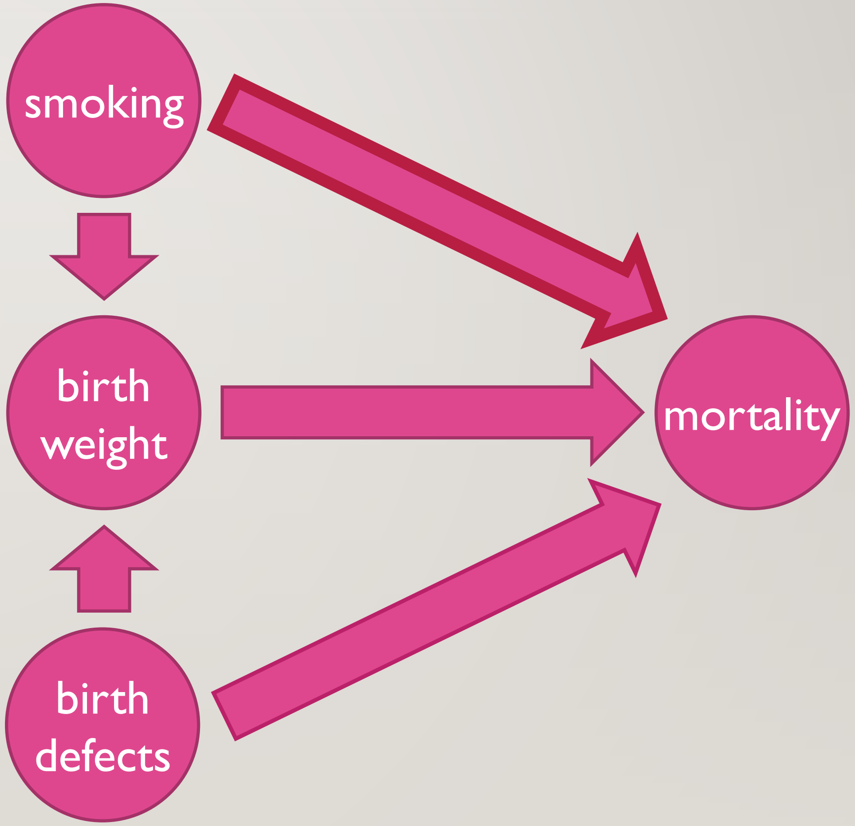
$$P(purple) * P(hat|purple) = P(hat) * P(purple|hat)$$

$$P(hat|purple) = \frac{P(hat) * P(purple|hat)}{P(purple)}$$

$$P(H = h\ cm|x \leq 10cm) = \frac{P(H = h\ cm) * P(x \leq 10cm|H = h\ cm)}{P(x \leq 10cm)}$$

THE END (OF MY PART)





SOURCES

Judea Pearl & Dana Mackenzie - The Book of Why

IMAGE SOURCES (LICENCED CC0):

https://pixabay.com/photos/animal-night-moon-sky-gallo-cloud-4166416/

https://unsplash.com/photos/YKW0JjP7rlU

http://www.publicdomainfiles.com/show_file.php?id=13932561625880

https://freesvg.org/black-suit-jacket-vector-illustration

https://pixabay.com/photos/apocalypse-war-disaster-destruction-2459465/

https://pixabay.com/photos/landscape-city-bank-heidelberg-3688428/

https://freesvg.org/cartoon-scientist-guy

https://freesvg.org/vector-image-of-container-carrying-truck

https://freesvg.org/comic-robot

https://freesvg.org/sleeping-cat-image

https://freesvg.org/stone-age-man

https://freesvg.org/baby-girl-crawling-vector-graphics

https://freesvg.org/young-female-scientist

https://freesvg.org/blue-and-red-rocket-with-engines-ignited-vector-graphics

https://freesvg.org/old-style-computer-vector-image

https://www.needpix.com/photo/download/1296094/head-wooden-ladder-ladder-angle-of-attack-head-rise-climb-up-old-shaky-vintage

https://freesvg.org/vector-image-of-wooden-fire

https://freesvg.org/grayscale-smoke-detector-vector-drawing

https://freesvg.org/misc-mental-figment

https://de.wikipedia.org/wiki/Datei:Age_warning_symbol.svg

https://freesvg.org/human-foot-vector-image

https://freesvg.org/male-book-reader

https://freesvg.org/movie-star-actor

https://freesvg.org/vector-image-of-woman-with-lustrous-long-hair

https://freesvg.org/thespian-frame

https://freesvg.org/red-pin

https://freesvg.org/potted-plant-image

https://freesvg.org/1536869662

https://freesvg.org/blue-stain-colored

https://freesvg.org/unimpressed-alien

https://freesvg.org/alien-baby-with-curly-hair-vector-image

https://freesvg.org/dsdna-linear

https://freesvg.org/colour-cigarette

https://freesvg.org/female-student-asking-a-question-vector-illustration