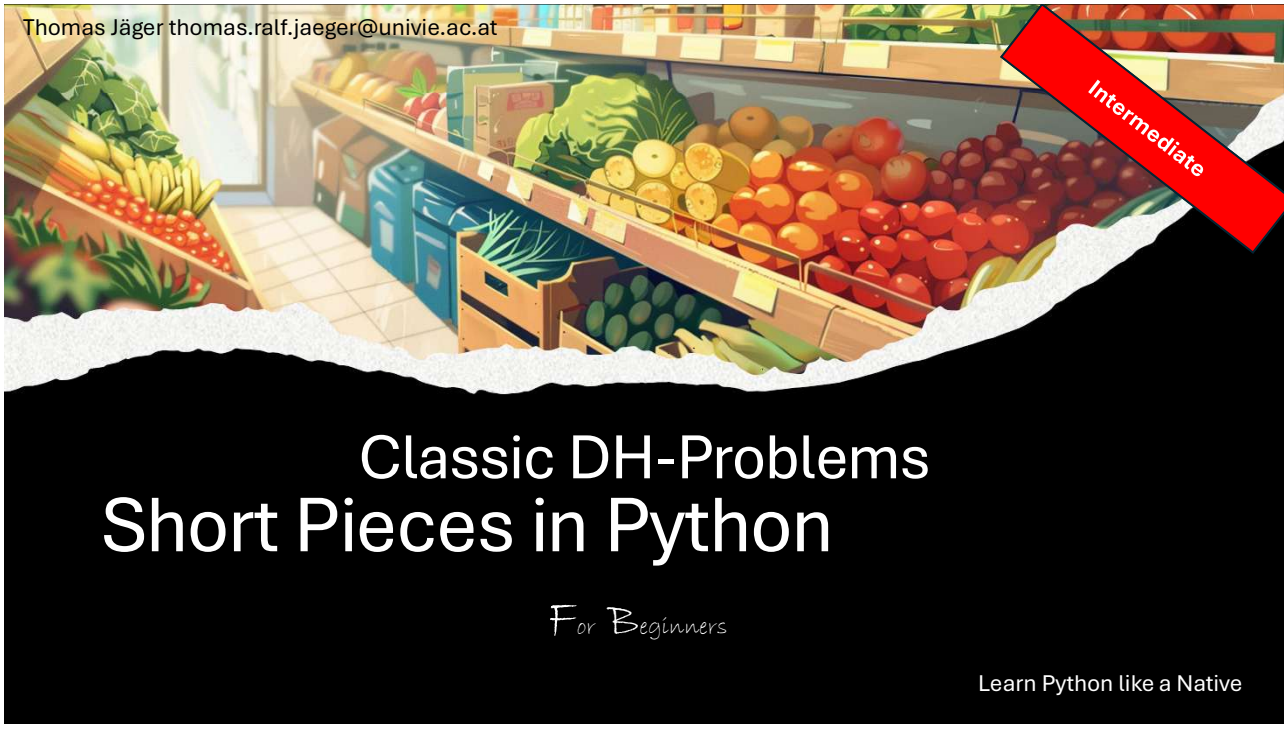
Thomas Jäger thomas.ralf.jaeger@univie.ac.at



Classic DH-Problems
Short Pieces in Python

For Beginners

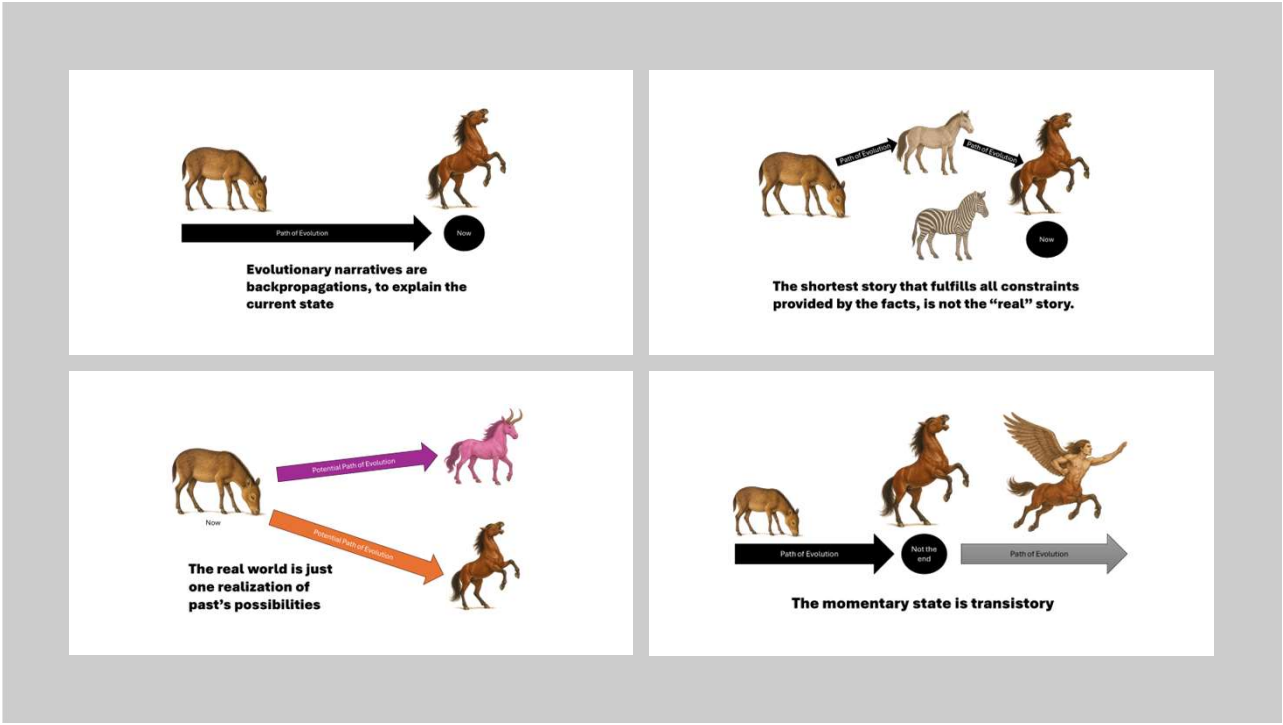
Learn Python like a Native

Programming:
- CP-Algorithm
- causalLearn
- DoWhy

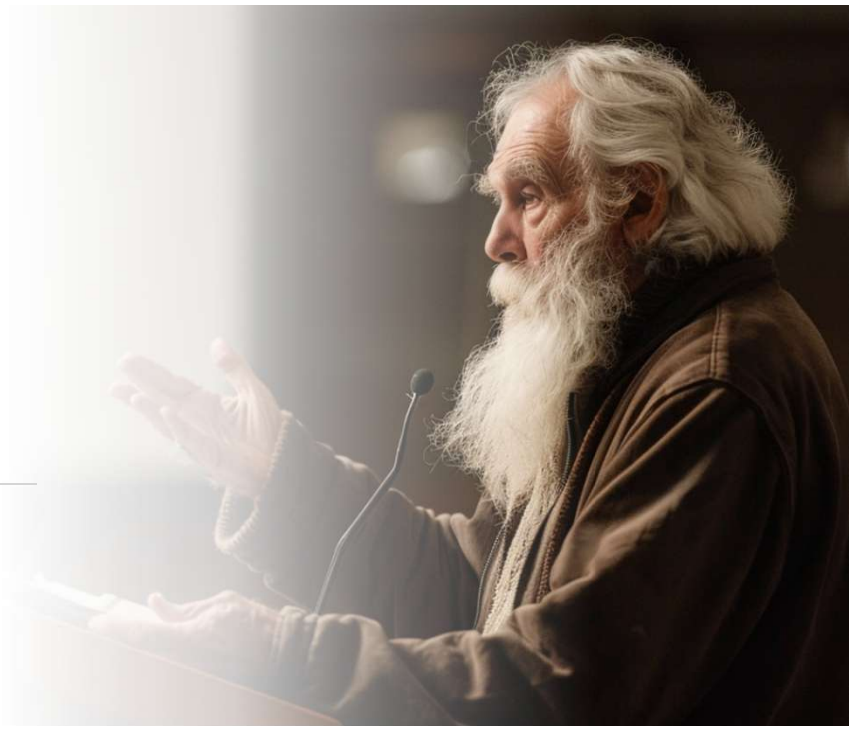
Modelling:
- Correlation
- Bayes
- Causal Graphs

Architecture:
- Pipelines
- Folders

Professional Knowledge:
- Pair Programming

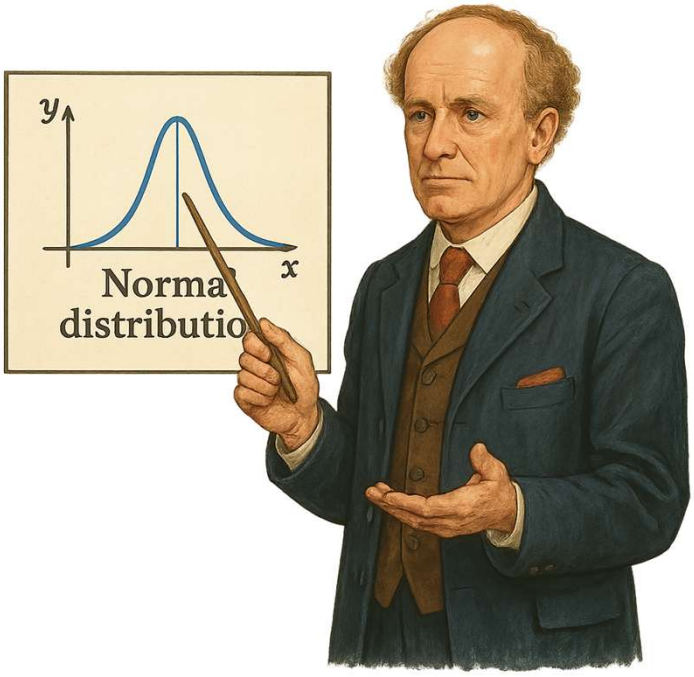


Evolutionary narratives are very often misapplied applications on humanity, to provide arguments (often for conservative topics)

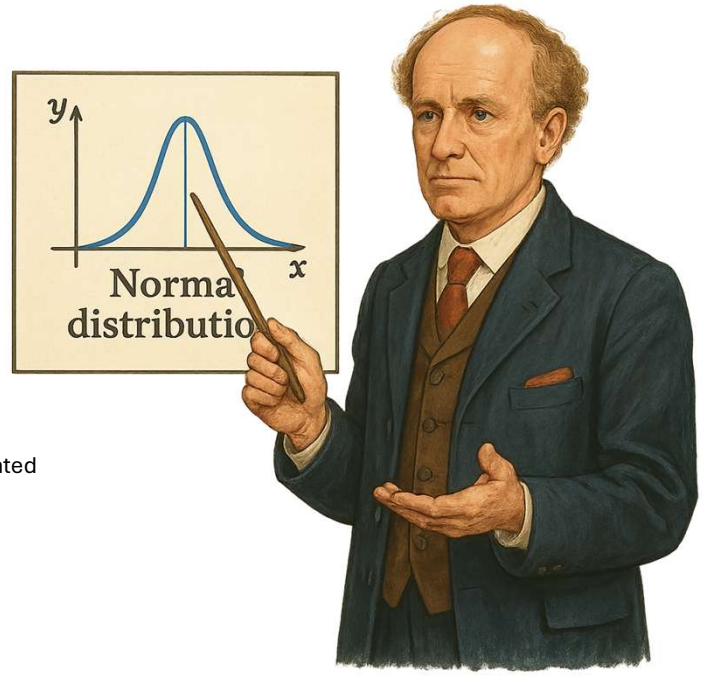


Karl Pearson:

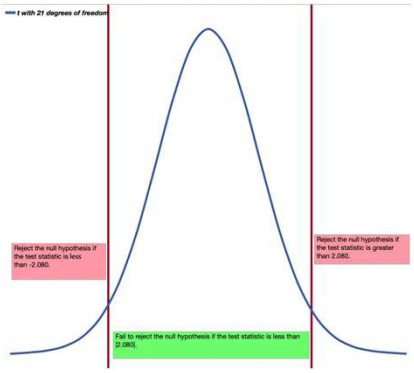
- *Mathematical Contributions to the Theory of Evolution* (18 papers).
- Founder of the biometrical school for inheritance.
- Was not interested in the causes of evolution (inheritance).



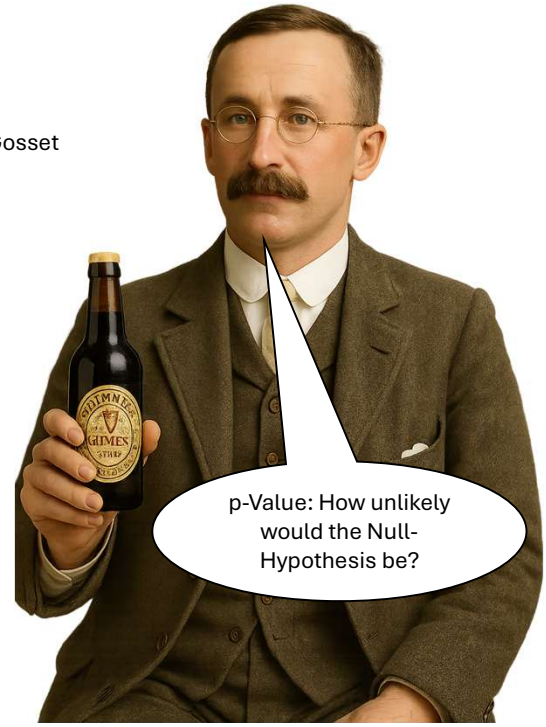
$$\rho_{x,y} = \frac{\mathbb{E}[(X-\mu_x)(Y-\mu_y)]}{\sigma_x \sigma_y}$$



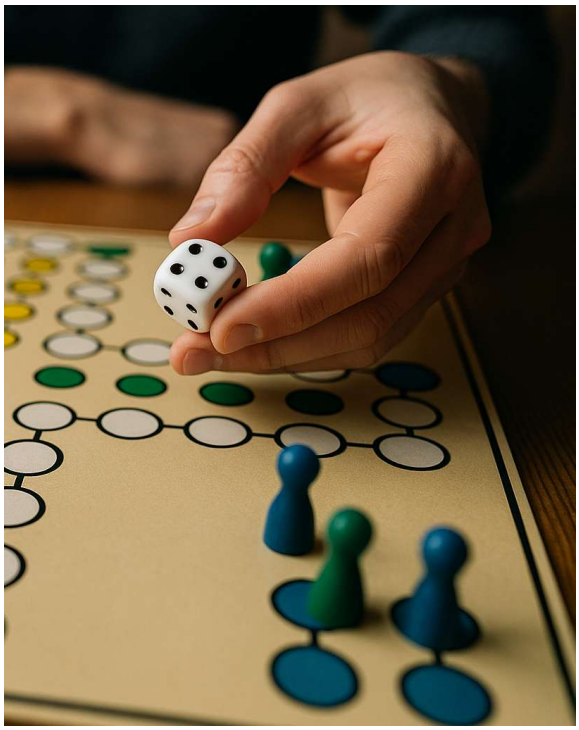
- $\rho_{x,y}$ is between -1 and 1
- -1 or 1 means: x and y are linear correlated
- 0 means x and y are not correlated



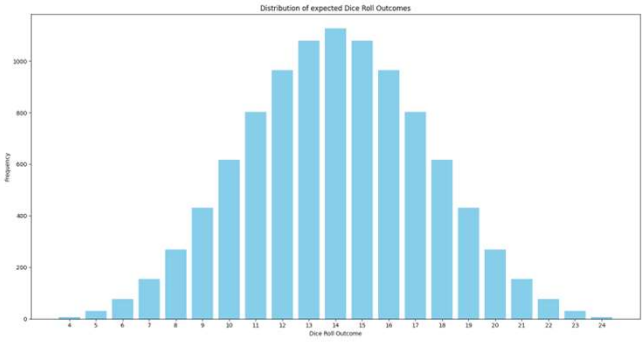
William Sealy Gosset
(Student 1908)

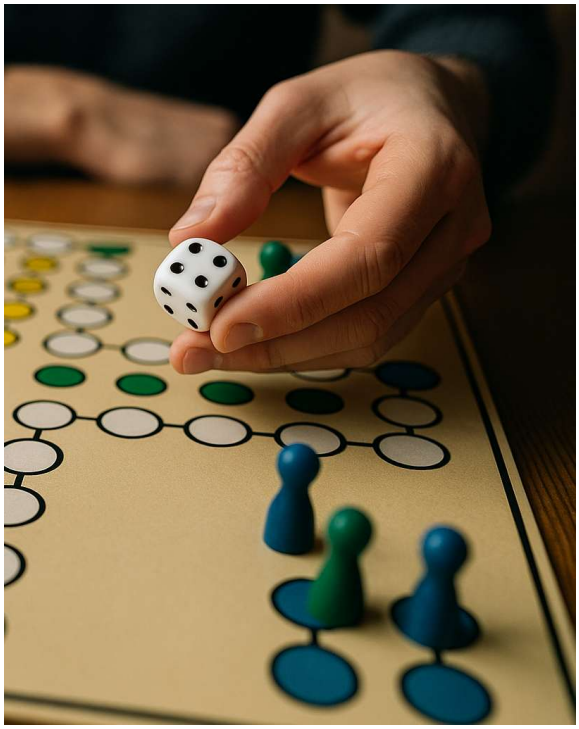


- The p-Value tries to falsify the hypothesis by proving the opposite of it is very likely to occur.
- A low p-Value says: the opposite of a hypotheses only rarely occurs.
- Rare does not mean impossible!!

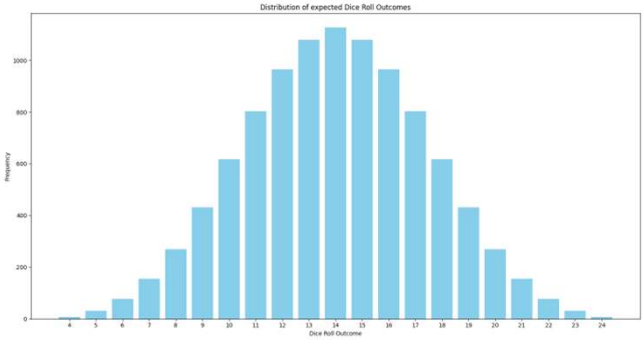


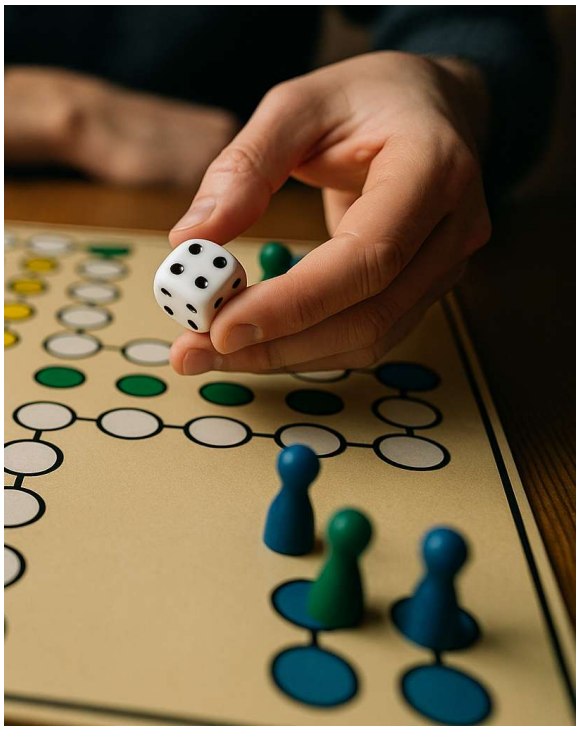
- The hypothesis that the dices are loaded, is the hypothesis to test.
- The hypothesis, that the dices are not loaded, is the 0-hypothesis.
- If the actual outcome of an experiment is unexpected under the 0-hypothesis, we have an argument for our hypothesis.



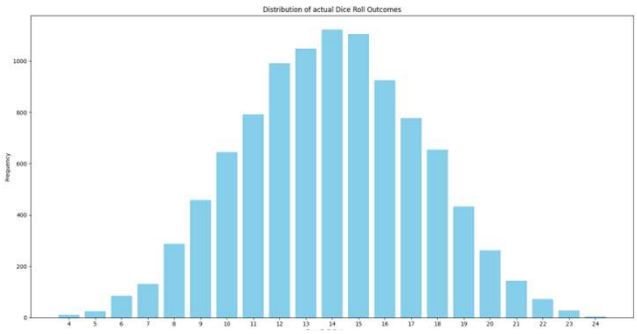


- If you roll 4 dices, you expect the sum of the dices to be 14, or close to 14.
- The lowest possible combination is 4, the highest possible outcome is 24.
- Numbers higher than 20 or lower than 8 are unexpected.
- A 24 (4 sixes) or a 4 (4 ones) should be extremely rare.
- The actual outcome might differ from expected results.
- The time when any combination of dices shows up first, is totally random.



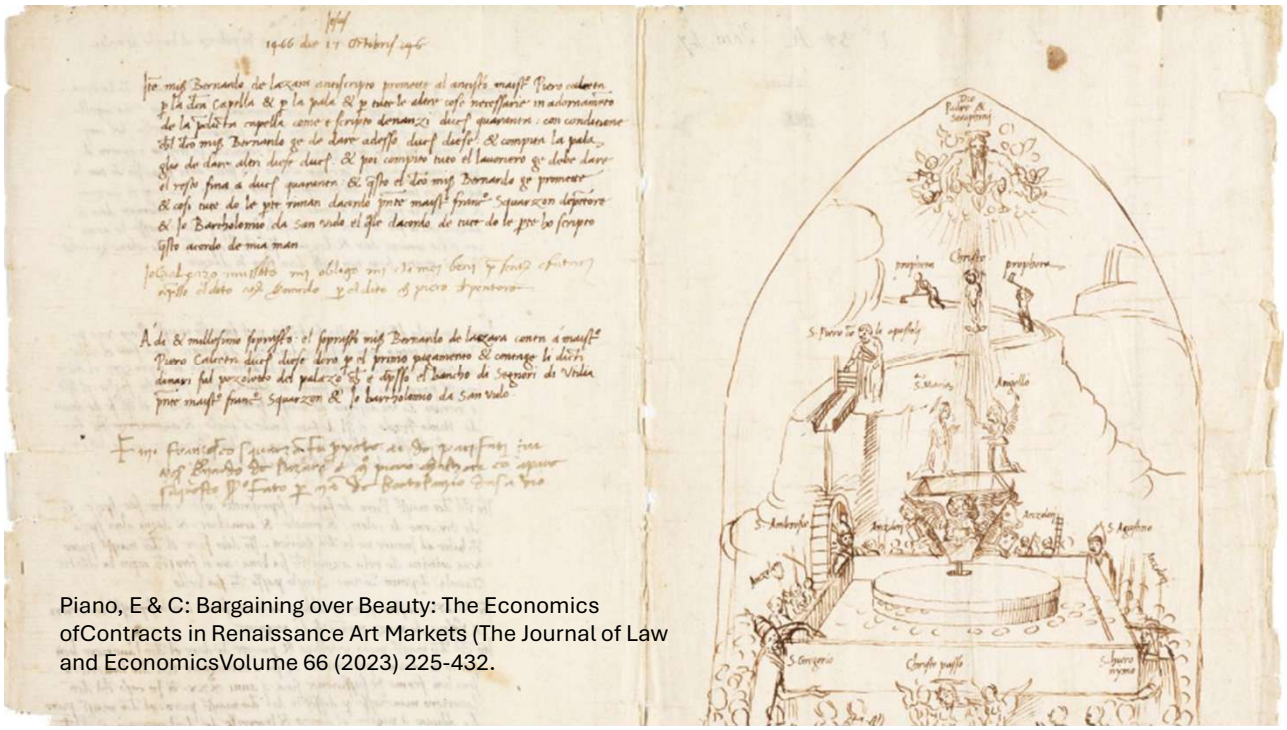


- The expected number of sum 24 should be 8 in 10.000 tries.
- The actual time sum 24 is reached in 10.000 tries is somewhere between 4 and 12.
- A throw of 4 dices might have four sixes any time in thousand throws. It might be the first, or the last of those throws and any time in between.
- In 1000 experiments the earliest time in any of the experiment, that a throw of 4 dices has sum 24, is expected to be close to one.



Thomas Bayes
An essay towards solving a problem in the doctrine of chances (posthum 1764)

$$P(A|B) = \frac{P(B|A) \cdot P(A)}{P(B)}$$

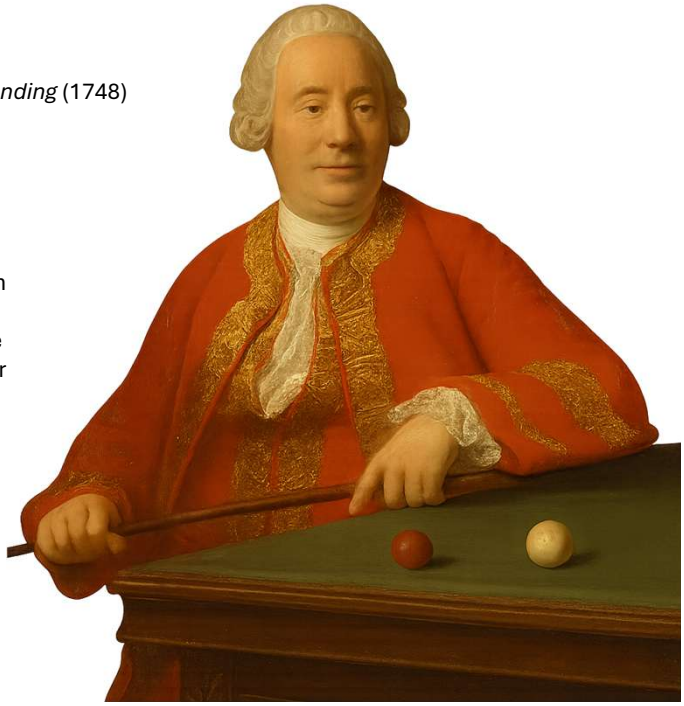


Piano, E & C: Bargaining over Beauty: The Economics ofContracts in Renaissance Art Markets (The Journal of Law and EconomicsVolume 66 (2023) 225-432.

- Data shows associations.
- Correlation does not show causation

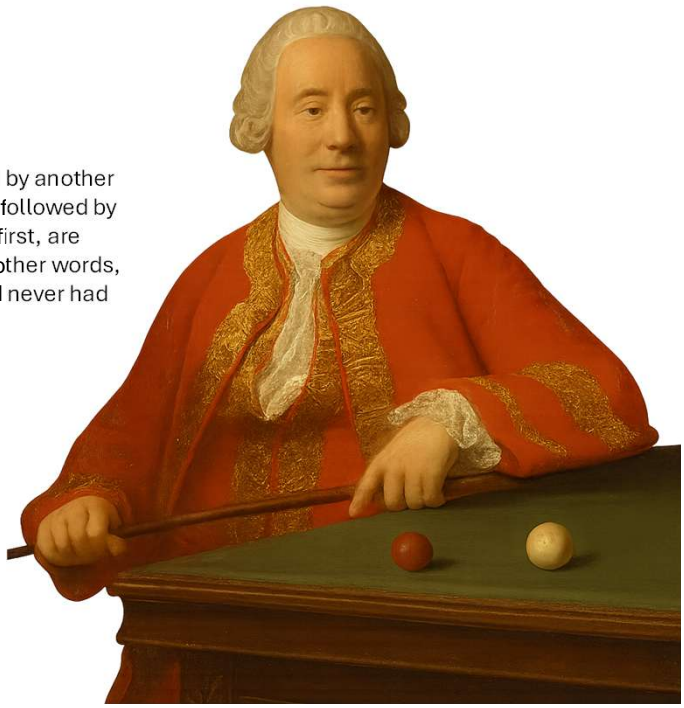
There is a secret tie or union among particular ideas, which causes the mind to conjoin them more frequently, and makes the one, upon its appearance, introduce the other. (David Hume)

David Hume
An Enquiry Concerning Human Understanding (1748)



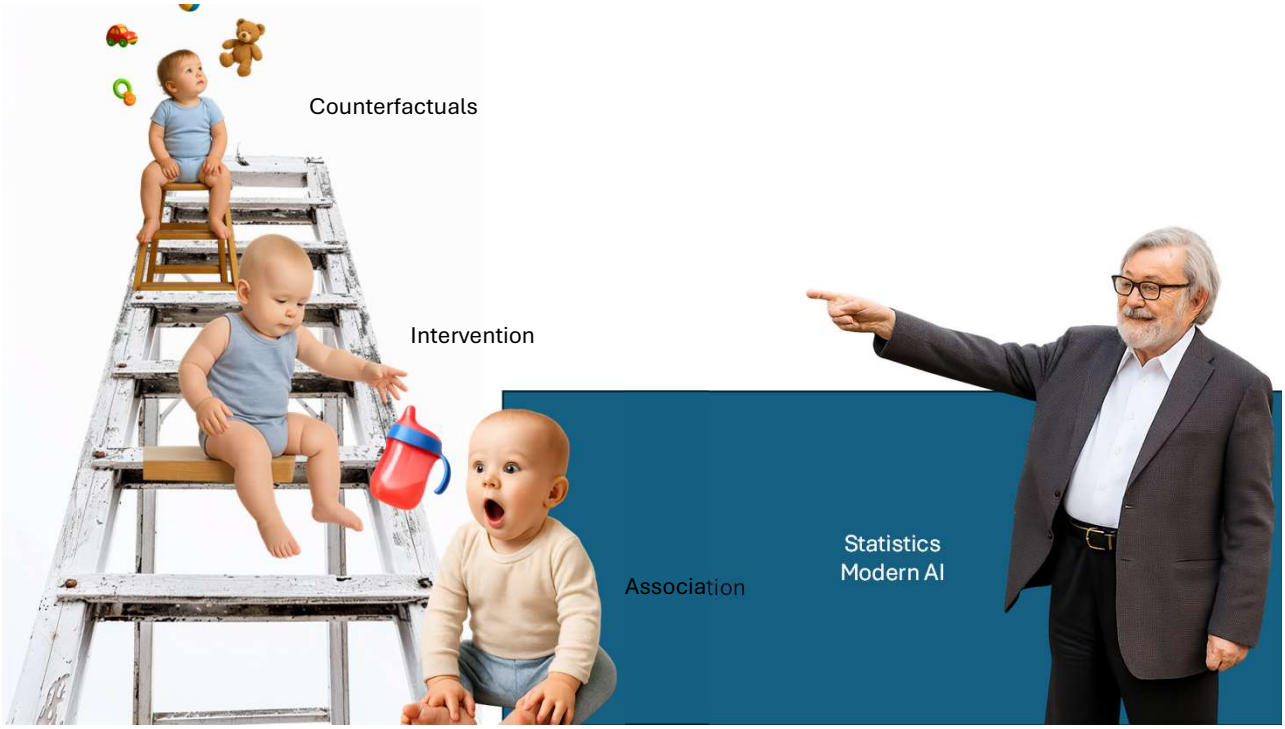
When I see, for instance, a billiard-ball moving in a straight line towards another; even suppose motion in the second ball should by accident be suggested to me, as the result of their contact or impulse; may I not conceive, that a hundred different events might as well follow from the cause? … All these suppositions are consistent and conceivable.

"We may define a cause to be an object followed by another and where all the objects, similar to the first, are followed by another and where all the objects, similar to the first, are followed by objects similar to the second. Or, in other words, where if the first object had not been, the second never had existed"



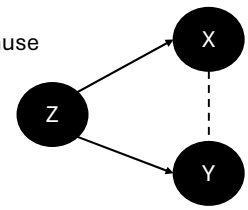
Judea Pearl
The Book of Why : The New Science of Cause and Effect. (2018)

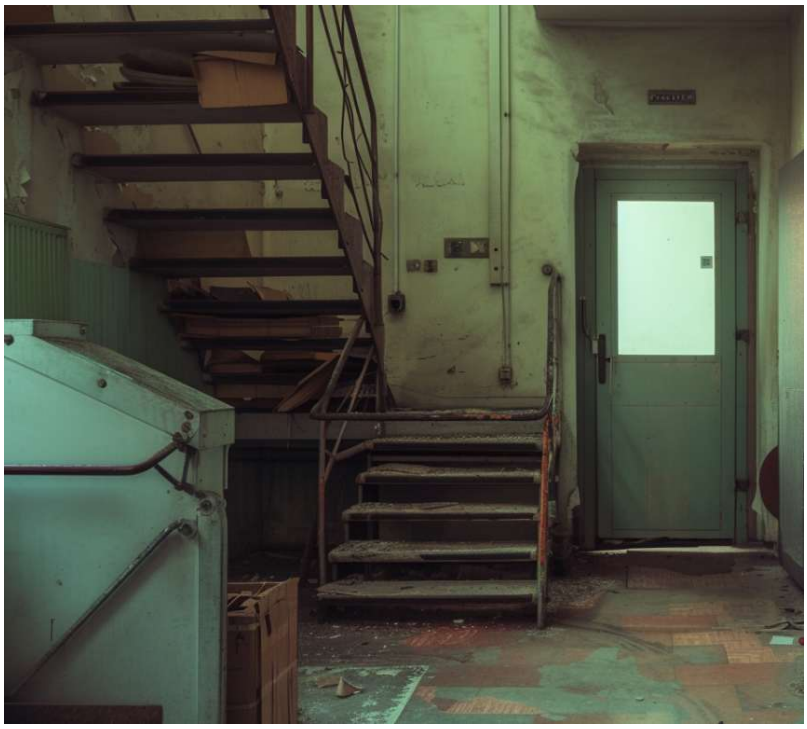
The Ladder of causation

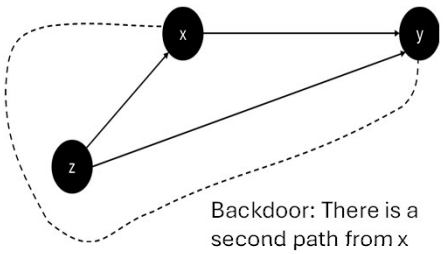


- A **fork** is a causal structure where **a single variable causes two others**:
- Z → X
- Z → Y
- **Z is a common cause** (a **confounder**) of both **X** and **Y**
- The path between **X and Y** is **opened** by **Z**, even though there's **no direct causal link** between X and Y

- X and Y are correlated
- You can calculate (Y|X).
- You can calculate (X|Y)
- It "looks" like X might cause Y or Y might cause X



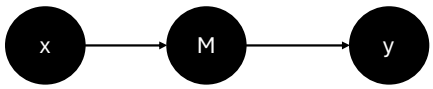




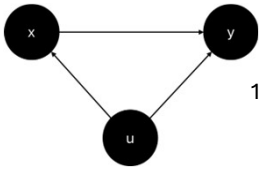
Backdoor: There is a second path from x to y (via z).

- A **backdoor path** is a **non-causal path** from a treatment (or exposure) variable **X** to an outcome variable **Y** that can **create spurious associations**.
- You need to control for variables that block these paths to isolate the true causal effect.

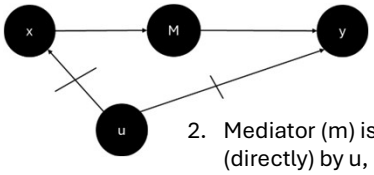
- A mediator is a node on the causal path between two nodes.
- If **X** causes **M**, and **M** causes **Y**, then **M is a mediator**.
- The treatment exerts some or all of its influence on the outcome through that mediator.
- Decomposing studies, how much of the effect is mediated.



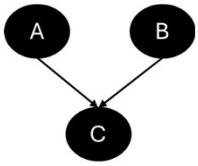
Front door opens new path via a **mediator** (a new front door) and avoids backdoor confounder.



1. unobserved confounder (u) affects x and y.



2. Mediator (m) is not affected (directly) by u, backdoor x via u to y is of no concern.



A **collider** is a variable **C** in a causal graph such that there are two (or more) variables, say **A** and **B**, that both have **directed edges into C.**

Colliders are critical in determining whether a path in a causal graph is blocked or open (which affects whether two variables are statistically independent).

Counterfactuals:

- Imaginary world, where something is decidedly different.
- If pigs could fly – what actually would be different?
- So – as pigs can't fly …?

ATE

Average Treatment Effect with the Do-Operator

- The **difference in expected outcomes between two levels of intervention**
- $ATE = E(Y|do(X = 1)) - E(Y|do(X = 0))$
- Example: Patients get pill / get placebo.

- The **Do-Operator (do(X = x))** represents an **intervention**: it forces the variable X to take the value x, **breaking its natural causes**.
- Instead of asking: *"What happens when we observe X = x?"*
  we ask: *"What happens when we **set** X = x?"*
- **Example:**
  - Observation: Higher education is associated with higher income.
  - **Causal question:** *What would happen to income if we actively increased education?* → do(education = high)
- This requires controlling for **confounding variables** that affect both education and income.

ATT

Average Treatment Effect on the treated

- The **difference in outcomes between two levels of intervention for those who got the treatment**
- $ATT = E\{(Y_i(1) - (Y_i(0)|X = 1\}$
- Example: What happens to Patients, that got the treatment?

ITE

Individual Treatment Effect

The **difference in outcomes for an individual**

- $ITE = E\{(Y_i(1) - (Y_i(0)\}$
- Example: What do we expect the effect to be on a patient?

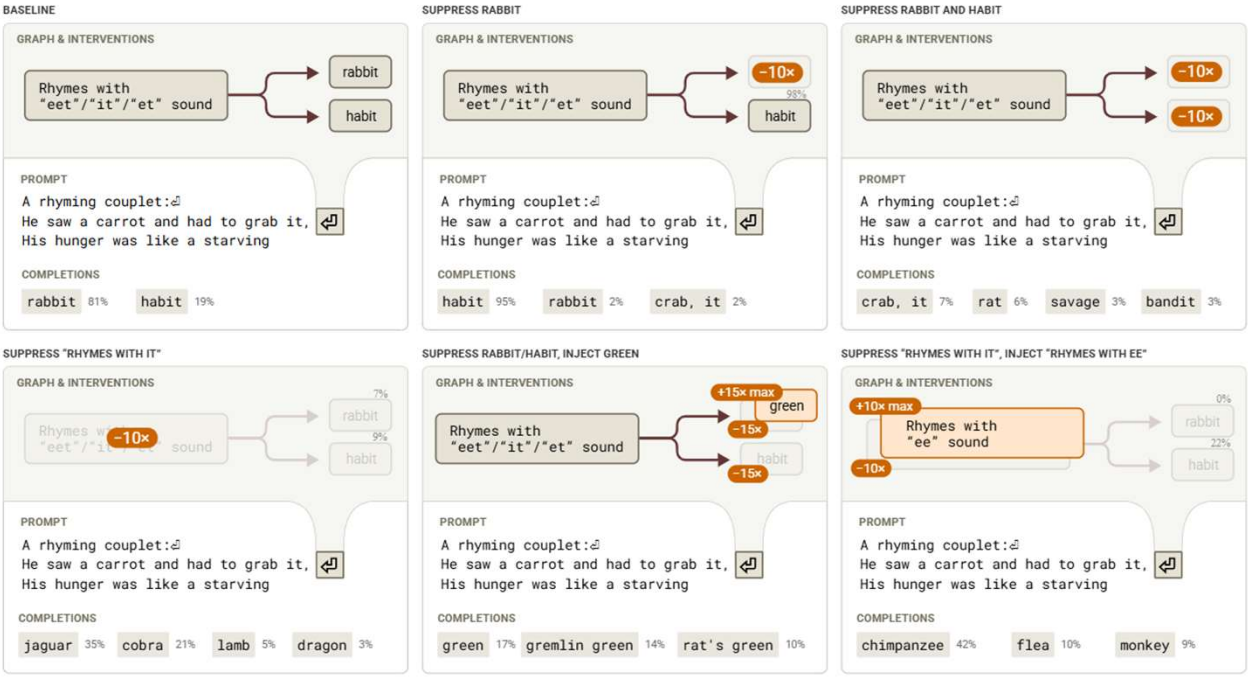
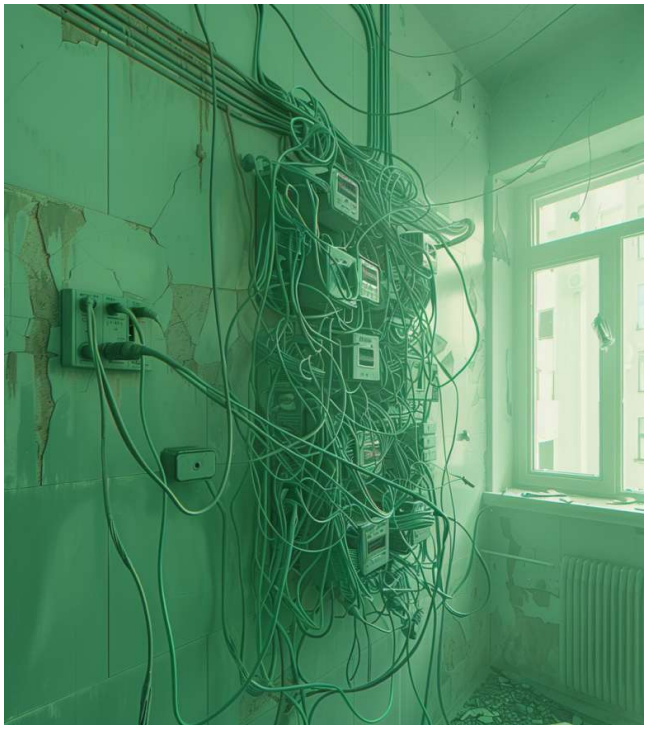


Figure 12: Interventions testing our understanding of the final token completion in the poetry example. Node activations are measured relative to the maximum baseline activation.



CP-Algorithm

1. **Start with a fully connected undirected graph** (between all observed variables)
2. **Remove edges for unconditional independence** (i.e., if two variables are marginally independent)
3. **Remove more edges using conditional independence**
    1. For each pair A–B, check all subsets C of adjacent variables
    2. If A ⫫ B | C, remove the edge between A and B
4. **Orient v-structures** (colliders):
    1. If A–B–C, and A and C are not connected, and A is **not** independent of C given B → then orient as A → B ← C
5. **Propagate orientation** using **Meek's rules**
    1. For example: avoid cycles, prevent new colliders, etc.