Causality

What is it really? Why do people talk about it? How do you do math with it?

Jaana Heino, Jukka Kohonen,
Miikka Miettinen, Mikko Rauhala

Dictionary definition

”Kausaalilaki (lat. causa = syy) lausuu sen periaatteen, että tapahtumat ovat määrätyissä syy- ja seuraussuhteissa toisiinsa.

Lisäksi kausaalilaki sisältää sen, että syysuhteet ovat pysyväisiä, s. o. että määrätyillä syillä on aina määrätyt seurauksensa.

Kausaalilain merkitys on siinä, että me ainoastaan pitämällä tapahtumia luonnossa syy- ja seuraussuhteessa olevina voimme edeltäpäin määrätä tapahtumain kulun ja siten hallita luontoa.

Siihen perustuu sekä jokapäiväinen että tieteellinen käsitystapa luonnosta.”

(”Tietosanakirja”, a Finnish encyclopedia from year 1912, on ”causality”)

- in plain English: ”it’s important, everybody uses it, but we (authors) don’t know how to say what it is, so we give you a circular definition”

Short History Of Causality

- ancient Greeks (xxx B.C.) talked about it, but knew little
- Hume (1711-1776):
  ”You can’t deduce it so don’t talk about it!”
- Kant (1724-1804):
  ”So take it *a priori*!”
- 20th century statisticians:
  ”You can’t distinguish it from statistical dependence
  – so don’t talk about it!”
- *Meanwhile Back In The Jungle: Everybody has been using causality in science and in everyday life all the time...*

Really, why use it?

- understand how things work and affect each other:
  ”A happened <u>because</u> of B, not by pure chance”
- do smart decisions, or ”interventions”:
  ”I’ll <u>increase</u> B in order to achieve A”
- Application fields: mechanics, electronics, all technology, economics, medicine, community planning, game theory, you name it… also: *autonomous causal agents* (robots etc.)

Causality *vs.* Statistical Dependence

- related but separate concepts, often confused
- causality is directional, dependence is not
- dependence says: ”if you look at the system and see A, you are *likely* to see also B”
- causality says: ”if you touch the system so that A happens, then you will (or are more likely to) also see B”
- so basically, causality is an implication (=>)
  – but between what propositions?

note: (linear) correlation is yet another thing – it’s a special case of statistical dependence

Work up your Intuition!

”If I pull this **lever**, then this **tank** will drain.”

Work up your Intuition!

”If I pull this **lever**, then this **tank** will drain.”

- *does it have to be ”I”? or at least ”a human”?*
  No! It can be ”Michael” or ”a randomly falling object” – as long as it is something independent

Work up your Intuition!

”If I pull this **lever**, then this **tank** will drain.”

- *does it have to be ”I”? or at least ”a human”?*
  No! It can be ”Michael” or ”a randomly falling object” – as long as it is something independent
- *is it symmetric?* **No!** Emptying the tank by ”other means” does not affect the lever.

Work up your Intuition!

”If I pull this **lever**, then this **tank** will drain.”

- *does it have to be ”I”? or at least ”a human”?*
  No! It can be ”Michael” or ”a randomly falling object” – as long as it is something independent
- *is it symmetric?* **No!** Emptying the tank by ”other means” does not affect the lever.
- *what are the relevant variables?* The **lever** position and the **tank** water content.

Work up your Intuition!

”If I pull this **lever**, then this **tank** will drain.”

- *does it have to be ”I”? or at least ”a human”?*
  No! It can be ”Michael” or ”a randomly falling object” – as long as it is something independent
- *is it symmetric?* **No!** Emptying the tank by ”other means” does not affect the lever.
- *what are the relevant variables?* The **lever** position and the **tank** water content.
- *”But how can you have asymmetric relation between two variables…”*

Indeed!

- typical statistical inference relies on symmetric concepts like

- *joint distribution*: P(AB)
- *independence*: $P(AB) = P(A)\,P(B)$
- *dependence*: $P(AB) \neq P(A)\,P(B)$
- *conditional probability* $P(A \mid B)$ – still essentially symmetric because of Bayes

- ”can’t talk about asymmetry with symmetric language only”: CORRECT
- ”can’t talk about asymmetry at all” INCORRECT

Why have a mathematical language for causality?

Same reason as for any other complicated problem: math gives you

- more precise, and
- more reliable

results than intuition alone.

How to have a mathematical language for causality?

separate tasks:

- formalism for saying causal statements
- rules for deducing things from causal statements
- rules for inducing causal statements from things

Pearl's formalism of Causality

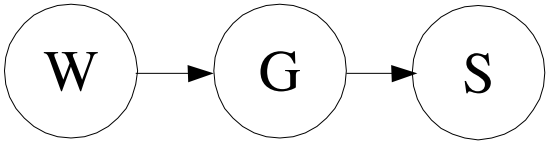
- Pearl's model of causality stems from the philosophical considerations outlined above.
- Pearl's causal networks and calculus can be generally described as extensions of the more traditional Bayesian variants.

Pearl's causal calculus

- Pearl's definition of causality comes from the effect that interventions external to the causal model have on other variables in the model.
- Given two disjoint sets of variables, X and Y, the causal effect of X on Y is denoted by $P(y \mid do(x))$. For each X=x, this function gives the probability of each Y=y. (Note the difference to the purely statistical relationship $P(y \mid x)$).
- If $P(y \mid do(x)) = P(y)$, X does not have any causal effect on Y.

Causal networks

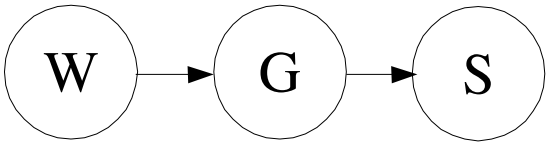
- Causal networks are directional acyclic graphs.
- Each node represents a variable in the model.
- Each arc represents a direct causal relationship.
- Longer paths represent indirect relationships.



W: Weather (Shine/Rain)
G: Grass (Dry/Wet)
S: Slipping (Yes/No)

Node functions

- For each node there is a function that gives its probability distribution, given the known distributions of its input nodes.

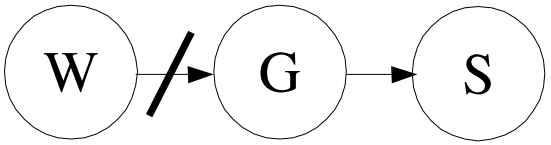


W: Weather (shine/rain)
G: Grass (dry/wet)
S: Slipping (yes/no)

$P(W = s) = 0.8$
$P(G = w \mid W = r) = 1$
$P(G = w \mid W = s) = 0.1$
$P(S = y \mid G = w) = 0.1$
$P(S = y \mid G = d) = 0.02$

Interventions

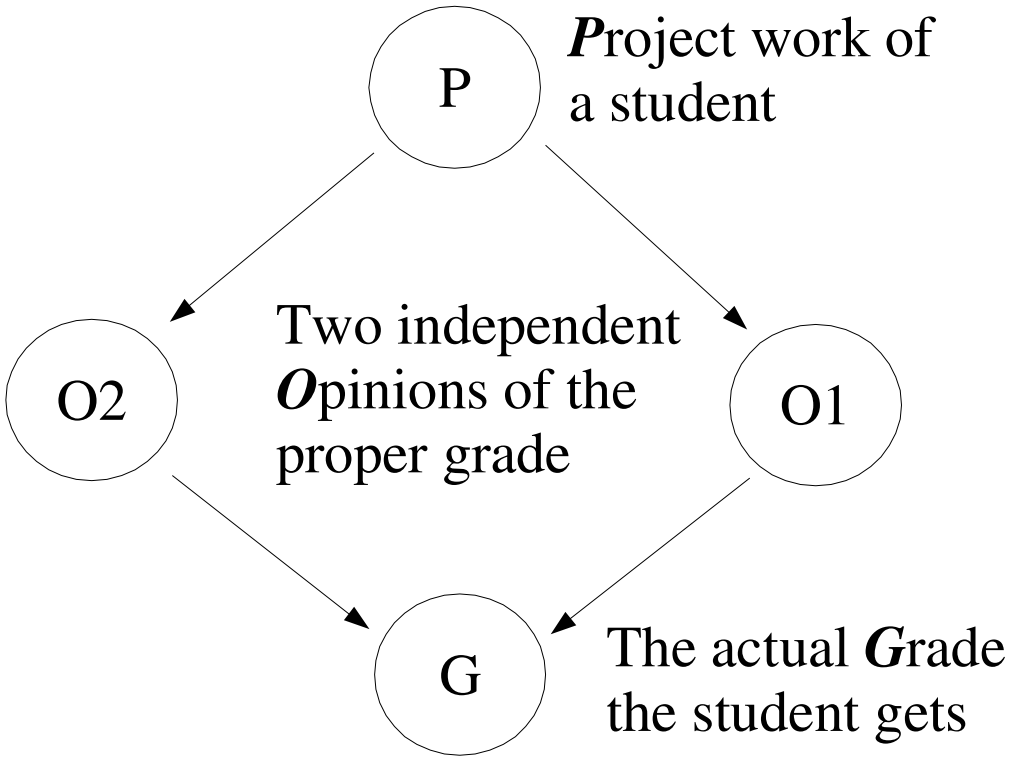
- Effects of interventions external to the model may be calculated by removing all arcs ending in the manipulated node, and setting its value manually.



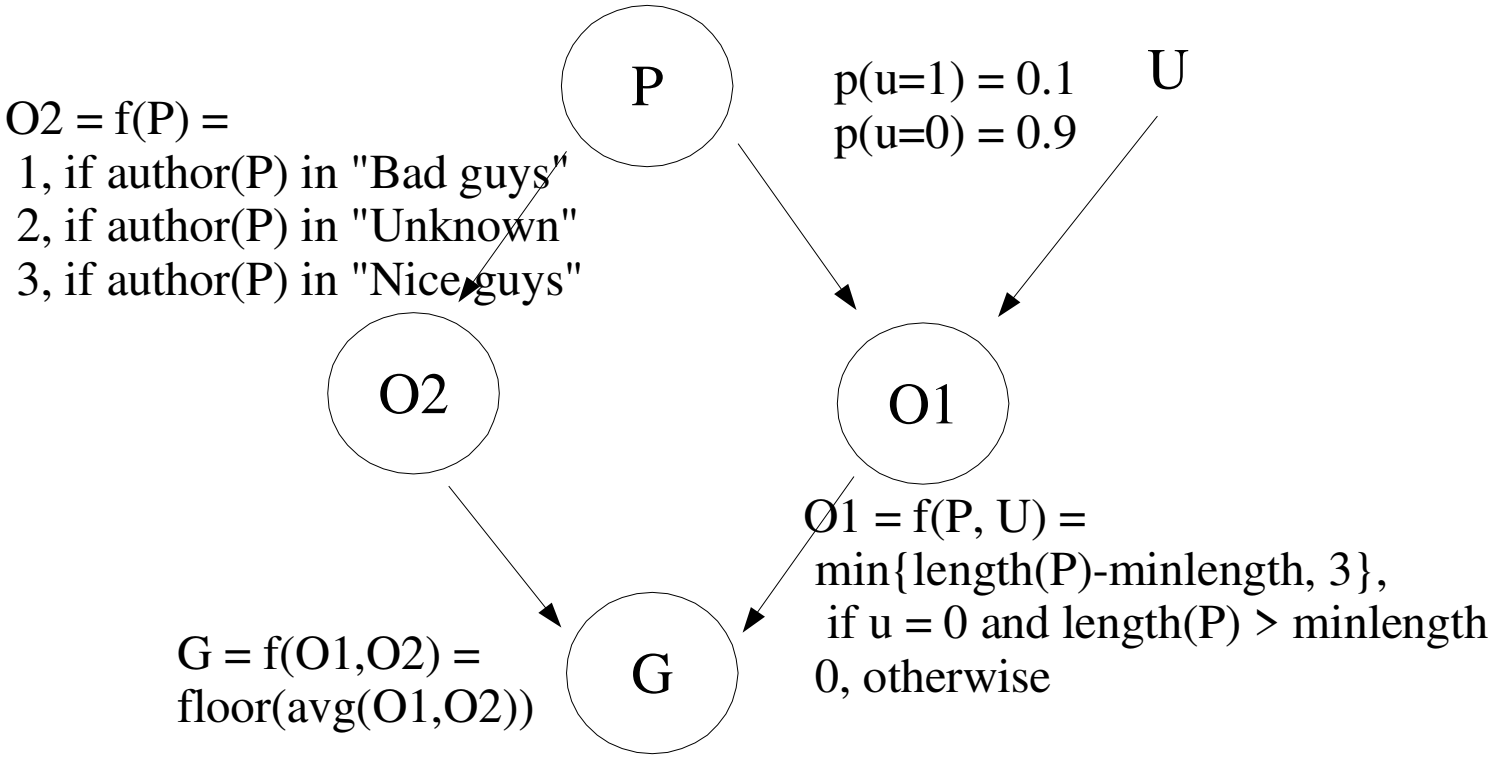
W: Weather (shine/rain)
G: Grass (dry/wet)
S: Slipping (yes/no)

Intervention: Grass is watered
$P(W = s) = 0.8$
$P(G = w) = 1$
$P(S = y \mid do(G = w)) = 0.1$
$P(S = y \mid do(G = d)) = 0.02$

Causal model, a simple example (structure of the DAG)

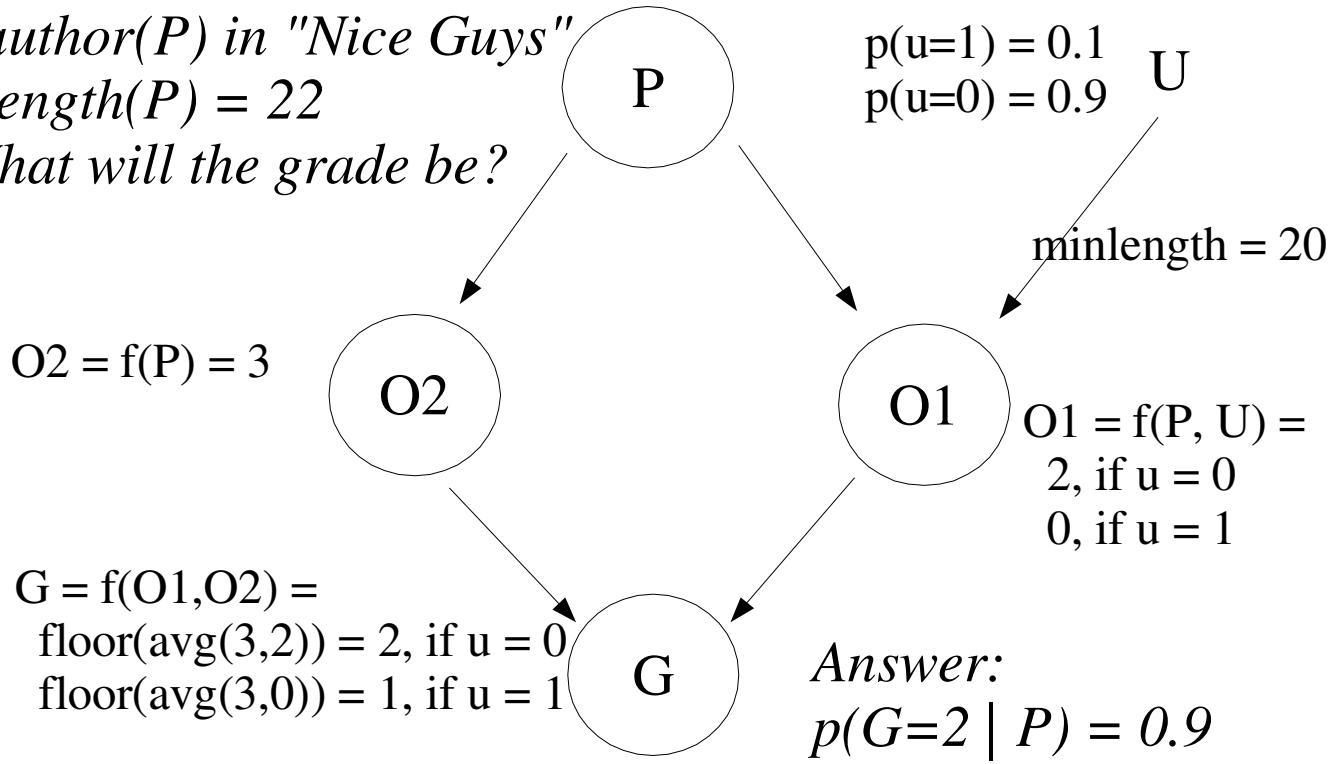


Causal model, a simple example (with the functions)



Causal model, a simple example (simple reasoning)

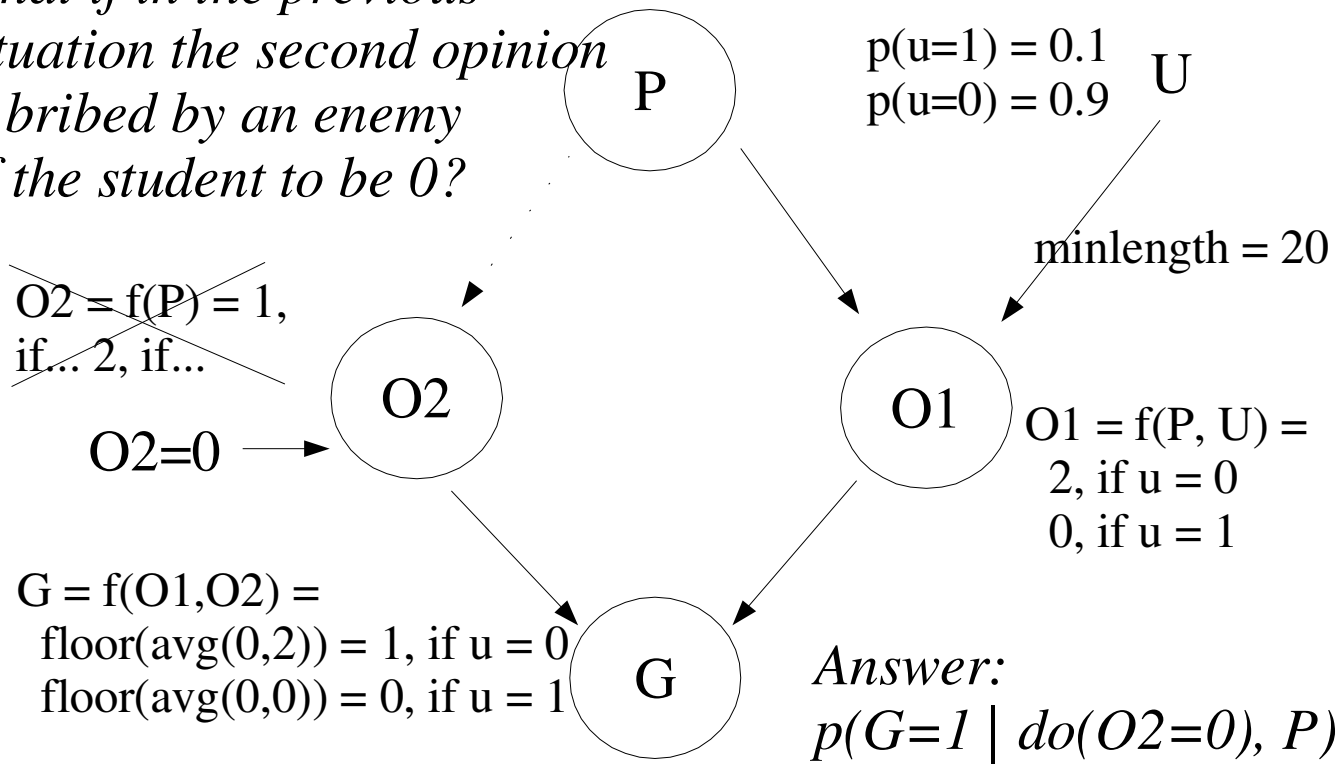
Consider P such that:
author(P) in "Nice Guys"
length(P) = 22
What will the grade be?



Answer:
$p(G=2 \mid P) = 0.9$
$p(G=1 \mid P) = 0.1$

Causal model, a simple example (intervention)

What if in the previous situation the second opinion is bribed by an enemy of the student to be 0?



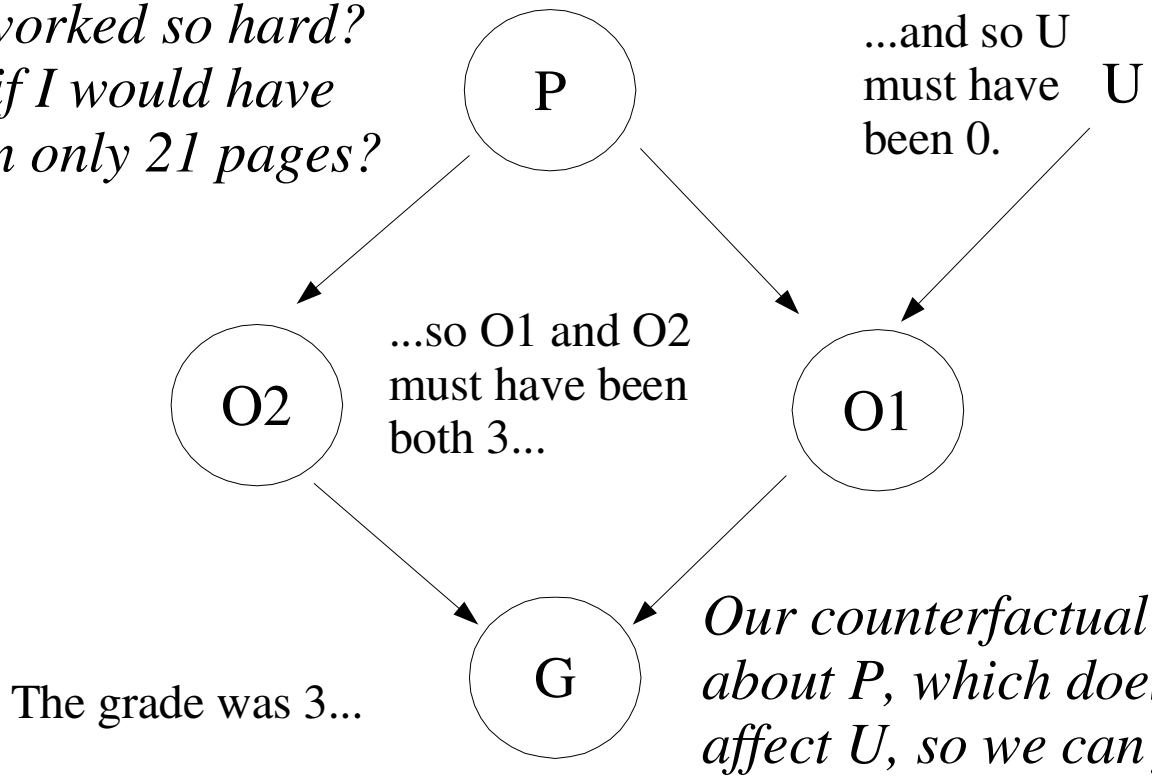
Answer:

$p(G=1 \mid do(O2=0), P) = 0.9$

$p(G=0 \mid do(O2=0), P) = 0.1$

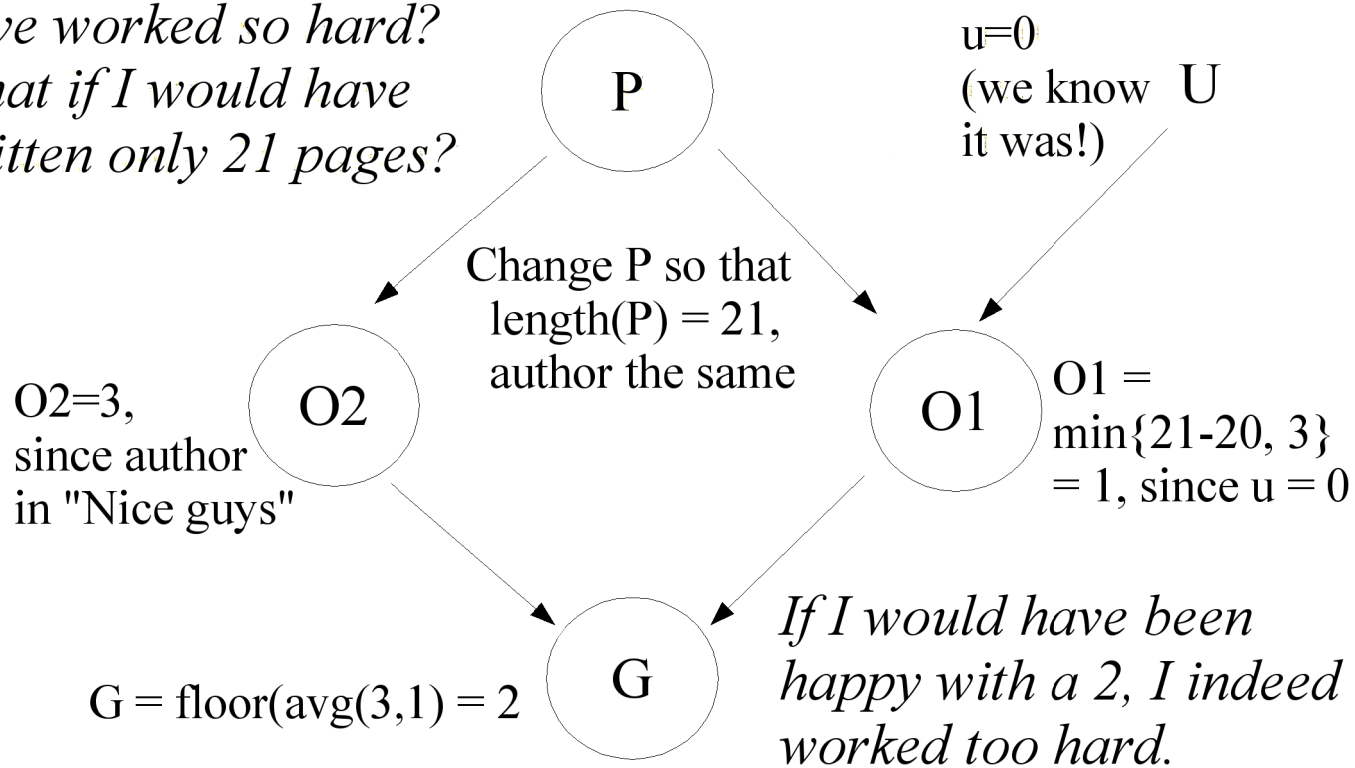
Causal model, a simple example (counterfactual 1.1)

I got a 3! Maybe I shouldn't have worked so hard? What if I would have written only 21 pages?



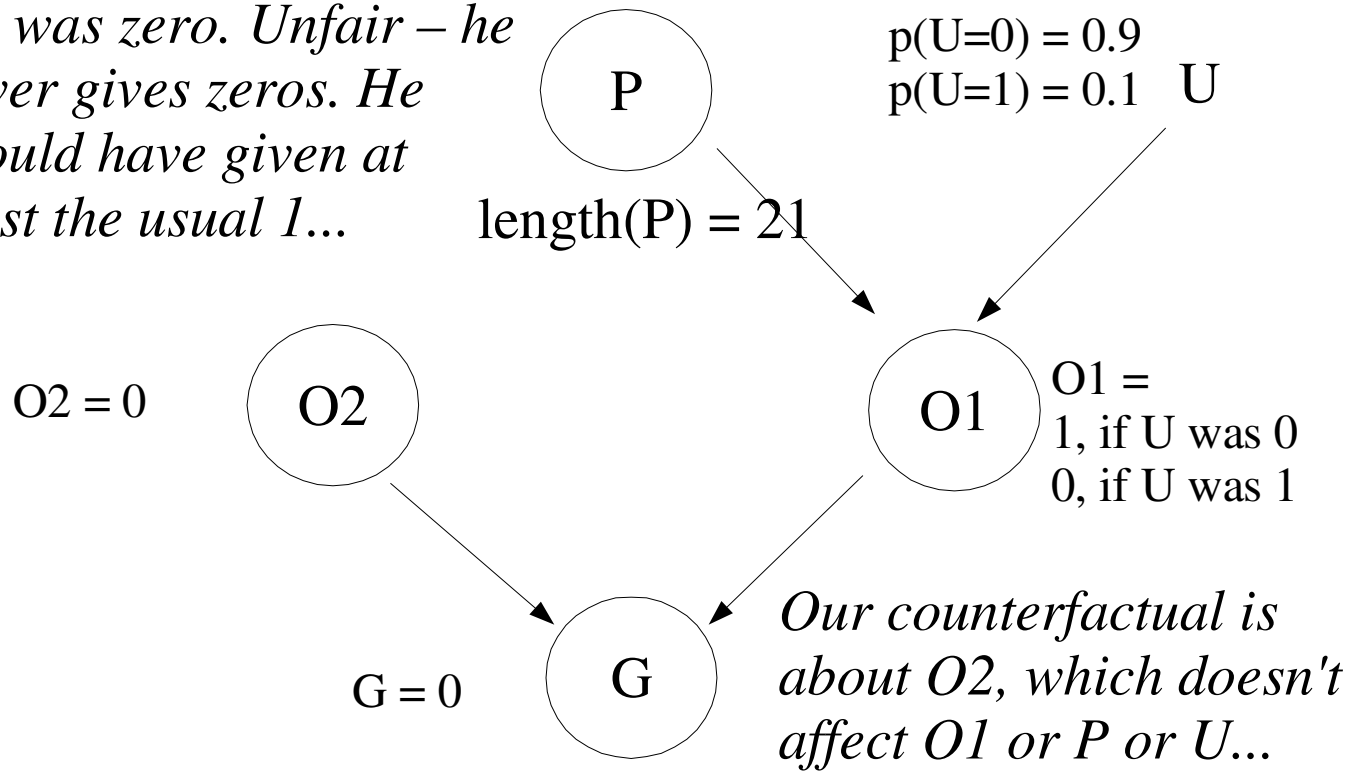
Causal model, a simple example (counterfactual 1.2)

I got a 3! Maybe I shouldn't have worked so hard? What if I would have written only 21 pages?



Causal model, a simple example (counterfactual 2.1)

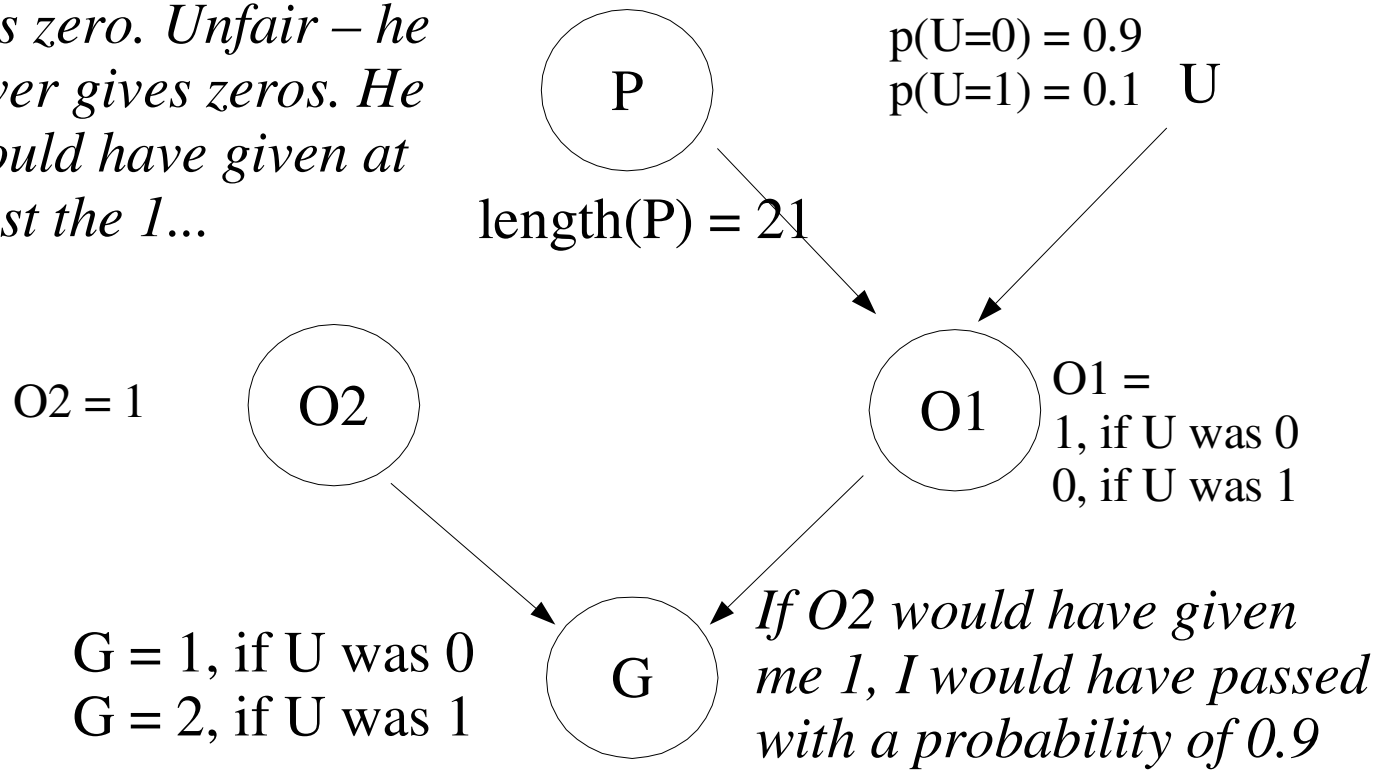
I got a 0, and moreover, O2 was zero. Unfair – he never gives zeros. He should have given at least the usual 1...



Our counterfactual is about O2, which doesn't affect O1 or P or U...

Causal model, a simple example (counterfactual 2.2)

I got a 0, and moreove, O2 was zero. Unfair – he never gives zeros. He should have given at least the 1...



How to come up with a causal model?

- Major sources of causal information are:
  - Teaching (i.e. pulling a network out of one's sleeve)
  - Interactive experimentation
  - Data (with certain assumptions)
- Note: teaching and experimentation major sources of causal information for humans, too