Behavioral and Experimental Economics

Session 1: Why experiments? Theory, relevance, principles

Paolo Crosetto

How to tackle an issue

You observe a phenomenon you do not understand. For instance, you see *electric prices fluctuate a lot*. You want to understand it. What can you do?

- produce a theory: a causally linked series of statements about some factors driving prices. Eg: prices are driven by the amount of electricity produced globally.
- observe the market for a long time, record the prices and other events surrounding them, and use statistical techniques to fit the data. It fits: higher amounts produced are observed when prices are lower.

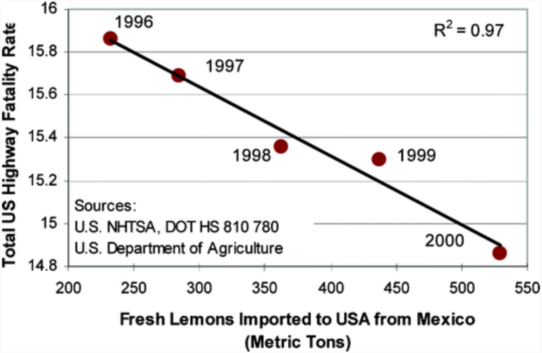
Have you proved your theory "*amount produced causes prices*"?

Correlation $\neq$ causation

Correlation does not imply causation.

$A \sim B$ could imply:

- causality! $A \Rightarrow B$: higher quantity $\Rightarrow$ lower prices
- **reverse** causality: $B \Rightarrow A$: lower prices $\Rightarrow$ higher quantity [why?]
- no causality: a spurious correlation



- a common **unobserved** cause: $C \Rightarrow A$ and $C \Rightarrow B$ but $A \not\Leftrightarrow B$

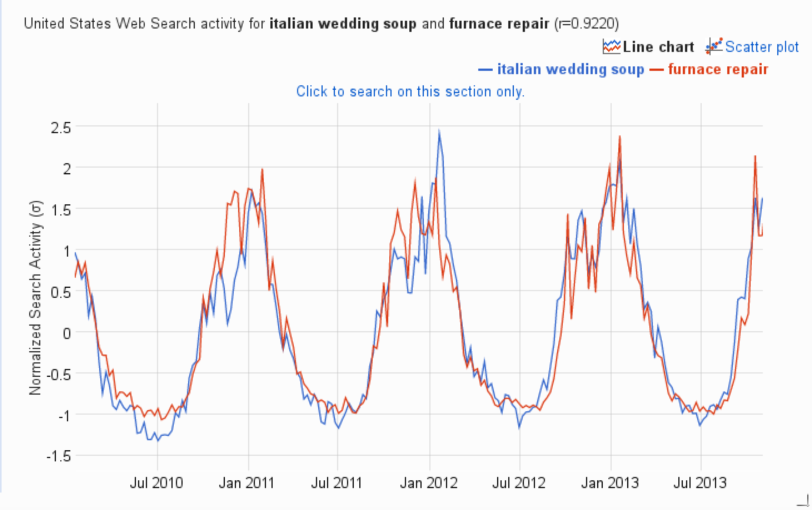
Reverse causality and COVID-19

In Italy and the United Kingdom, for example, where lockdowns have been repeatedly imposed, death totals per million remain among the worst in the world. Meanwhile, in the United States, states with the most harsh lockdown rules—such as New York, New Jersey, and Massachusetts are among the states with the worst total deaths.

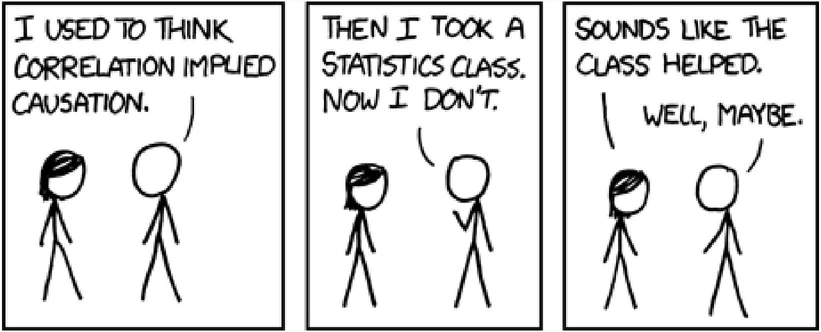
Is it **deaths** $\Rightarrow$ *lockdowns* or **lockdown** $\Rightarrow$ *deaths*?

Common unobserved cause

- a common unobserved cause ($C \Rightarrow A$ and $C \Rightarrow B$ but $A \nLeftrightarrow B$)



Correlation $\neq$ causation!



Test: if this *doesn't make you laugh*, you haven't understood. Study more.

Experiments

Experiments are the workhorse for unearthing causal relations in science

- Observe and measure the penomenon of interest [how do you do it?]
- Formulate *testable hypotheses* [what's that?]
- Control all possible confounding factors [what's that?]
- *Ceteris paribus*: keep *all* things equal, but *one* [it means ALL]
- Systematically vary the *one* variable of interest [how?]
- Compare results with hypothesis: rejected or not? [how?]
- Compare results to the real world: do they matter? [how?]

If the world does not accord with you, **you** are wrong, not the world.
[and that's fine. we learn.]

Experiments: control vs. relevance

In experiments we need *control* + *relevance*

Control: we need to be sure that *all possible confounds* are controlled for:

- Control over the sample;
- Control over the experimental design;
- Control over the previous knowledge of participants;
- Control over the beliefs of participants;
- ...

Relevance: we need to be sure that our experimental manipulation mimics reality; that is, that our experimental results can be assumed to work also *outside of the lab*

Control is also called *internal validity*: being sure that you are actually testing what you think you are.

Relevance is also called *external validity*: being sure that what you do matters for the problem at hand

Experiments

Observe and measure the penomenon of interest

- Observation must be *objective* [it's 9°C today *vs.* it's cold]
- Measure must follow some standards (loi Carrez)
- Both observation and measure must be **replicable**

Experiments

Formulate *testable hypotheses*

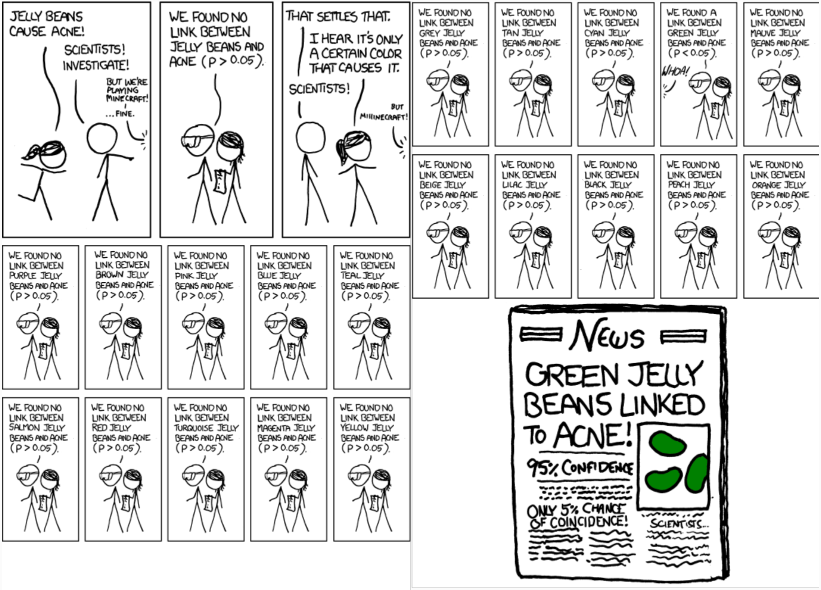
- Testable hypotheses must be apt to be **proven wrong**:
- they are **falsifiable**.
- You should lay them down *before* running the experiment...
- ...or else you risk *ex-post rationalization* (and go to hell!)
- because you will **always** find something...

The importance of ex-ante hypotheses and design

Figure: A pool game with infinite balls: you always score, the point is *which one*?

If you try hard enough, you *always* find *something*

- especially with the huge amount of data available today
- especially if you test *n* hypotheses on the same data
- especially if you look for subgroups, special cases...



Test: if this *doesn't make you laugh*, you haven't understood. Study more.

Scientific malpractices

The garden of forking paths the fact that in an analysis you took several dozen choices, and each one of them can bias your results or generate false positives; if you do not set a course beforehand, you can get lost.

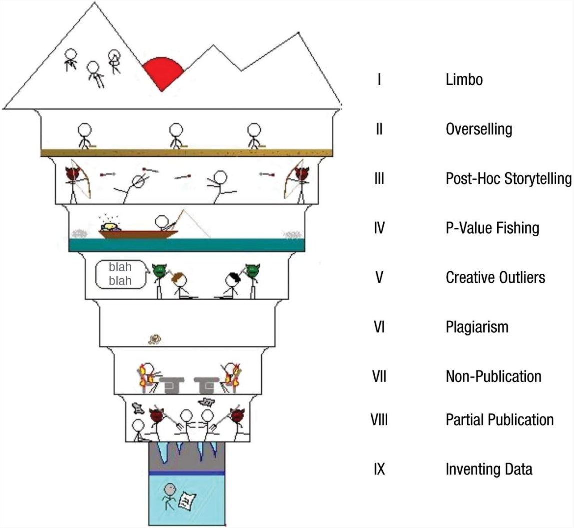
Multiple comparison problem the fact that if you make several comparisons, some of them are *bound* to be significant by pure chance (billiard, beans).

p-hacking the fact of running several tests, changing variables, changing econometric specification until you find some result that has the crucial $p < 0.05$ mark.

ex-post rationalization satisfied with your p-hacked result out of the garden of forking paths, you write up a story pretending that you *always* just looked for *that* result from the very start.

Further reading: A. Gelman, *The Garden of Forking Paths* (also on github)

shocking example: fMRI of a dead salmon here (and on github)



source: Nine circles of scientific hell (also on Github)

Experiments

Control all possible confounding factors

- Confounding factors are *possible factors influencing your variable of interest*...
- ...other than the one you are studying.
- The presence of these other factor invalidates your results: they might be spurious
- Think furnace repair and italian wedding soup

Controlling all confounding factors means reaching **internal validity**

- ...that is, being sure that results are really due to your manipulation only.

  Internal validity is the basic minimum without which any experiment is uninterpretable: Did in fact the experimental treatments make a difference in the experimental instance?

Confounds, game 1

A simple questionnaire

https://forms.gle/GnoTw3SvmX3VRZxM9

Possible confounds: experimenter demand effect

Definition (Experimenter demand effect)

Biases introduced in your experiment by the fact that the subjects act as if to do what they believe the experimenter wants of them, 'the right thing', 'the good thing'.

Example (Energy labeling)

You run an experiment comparing different energy labels on washing machines. You let the subject understand that out of the different labels, one was designed by the experimenter himself. Changing electricity consumption is not so difficult or costly (e.g.: unincentivized questionnaire). Subjects might deflate their declared consumption for *your* label because they want to *help* you.

- are we observing a real behavior or an artifact?
- if we prove causality, can we trust the proof?
- ways around:
  - neutral instructions
  - context independent choices: anonymized objects and labels
  - blind testing: the experimenter is a third person

Confounds, game 2

You have **10 euro** and you are matched with another player in this room.
One of the two players is randomly chosen to play *dictator*.
(S)he has to make a **transfer** $\in (0; 10)$ to the other player, *receiver*.
The amount chosen is **transferred** to the receiver.
End of the game.

https://forms.gle/kJk7zDtUGRpWn8B78

Possible confounds: incentive size effects (including size = 0)

Definition (Incentive size effects)

Biases introduced in your experiment by the fact that the small (or null) monetary values present in the experiment elicit qualitatively different behavior than the real-world high stakes

Example (Giving in trust games)

You just played a **dictator game**. Equilibrium (i.e. *rational* play for *selfish* agents) is for the the dictator to send zero. Most people in lab experiments *do* send something (and I bet you did).

- would this happen with complete strangers in the road and 100 rather than 10 euros?
- is the result an artifact of the very small incentives at stake?
- (note that things get worse for questionnaires: it is *free* to lie / misreport / make mistakes)
- Possible solutions: give appropriate incentives; prove behavior is unchanged as incentives change.

Confounds, game 3

A simple questionnaire

https://forms.gle/mN8jyXJqFix5zSWa7

Possible confounds: sample selection effects

Definition (Selection bias)

Selection bias is the selection of individuals, groups or data for analysis in such a way that proper randomization is not achieved, thereby ensuring that the sample obtained is not representative of the population intended to be analyzed [source: Wikipedia]

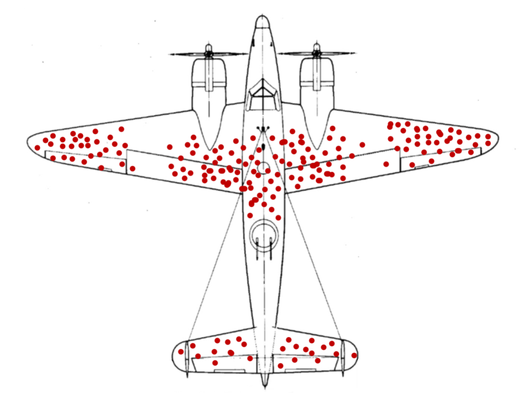
Example (An unexpected sample)

You expose 50 consumers to a questionnaire on their use of electricty. Turns out the most used means of transport is the bike, and 30% have photo-voltaic cells on their roofs. You meant the sample to be random, but due to word of mouth 78% of subjects come from a group of activists of the green party.

- selection effects can lead to biased and false claims
- selection is everywhere: **you** a random sample of the student population
- if you have selected subjects, you only observe *conditional* probabilities
- and these cannot be generalized
- solution: randomize, use large numbers, pay attention!

WW2 UK bomber planes

In 1943, British were having heavy casualties from planes shot down by Germans anti-aircraft fire. They decided to add armor to the planes in specific points (armoring all the plane means too much weight). From all the bombers that returned from missions in Germany, they collected data about all the hits.



Where would you add armor?

Test: if this *doesn't make you laugh*, you haven't understood. Study more.

Possible confounds: lack of credibility

Definition (Experimenter credibility)

The fact that experimental subjects believe the instructions to be the real rule of the game. Control about the *beliefs* the subjects have of what is going on in the laboratory.

Example (Electrocution)

A famous Social Psychology experiment involves questioning a subject telling him that he can choose whether to earn some money or not, knowing that if he does then in an adjacent room another subject will be (lightly) electrocuted. The experiment investigated the moral sense of people faced with immoral rules of the game. Nobody was really electrocuted, but a confederate (an actor acting on behalf of the experimenter) was making pain noises from an adjacent room.

- experimental results are strong only insofar as the scenario is believed
- that is, the script must be fully credible
- taking the money does not only mean not caring about the other suffering; it might also mean that you do not believe anyone is really suffering.
- credibility is key: beliefs guide actions and all we can observe are actions (and not beliefs!)

Experiments

Ceteris paribus: keep all things equal, but one

- This means *all* things
- Information; time of the day; room temperature; number of people; ...
- But also: awareness of being observed (Hawthorne effect, experimenter demand effect)
- (this is why medical test have a placebo condition rather than no intervention)

But can you really control everything? No, but you can randomise

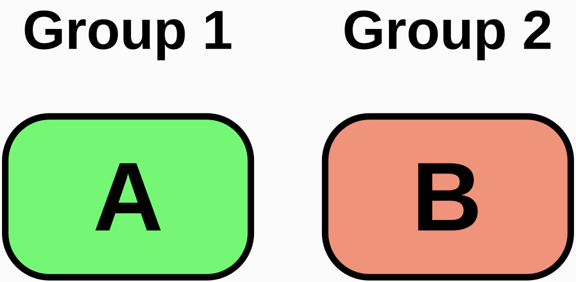
- Some factors can be controlled
- Most others are randomized: you let them vary *randomly* and rely on a high number of observations to smooth things out
- As $N \longrightarrow \infty$, random factors cancel each other out
- (example: die roll)

Experiments

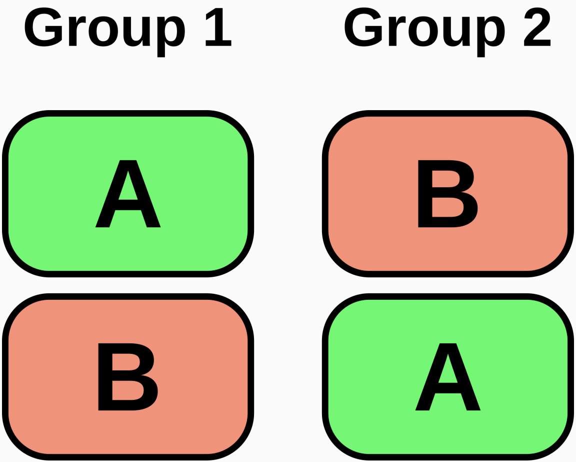
Systematically vary the *one* variable of interest

- These are what we call *Treatments*
- They can be varied within or between subjects
- Between: different subjects are subjected to different treatments (needs high N)
- Within: same subject is exposed to all treatments (needs less data, order effects)

Between-subjects



Within-subjects



Experiments

Compare results with hypothesis: rejected or not?

- Statistical testing can tell us if two samples are different (in mean, median, distribution, count...)
- Regression analysis or ANOVA can take care of multivariate analyses (control for common confounds as age, gender...)

Hypotheses are not proved true; they are not proven wrong

- Your tests tell us that you fail to reject the hypothesis.
- This does not mean it is true; it just means it is not random...
- ...and it means the hypothesis is **supported by the data**.
- other data might prove it wrong; or it might not always apply; or...

Good scientists try as hard as they can to prove themselves wrong

External validity

Definition (External Validity)

In scientific research, external validity is the degree to which it is warranted to generalize results to other contexts.[source: Wikipedia]

Example (Impact of advertisement)

You expose 45 subjects to a campaign to incentivise the use of buses.They report to you in a questionnaire that after seeing the ad they feel more inclined to take the bus by an average 20% more. Will the ad work also in the field?

External validity

Definition (External Validity)

In scientific research, external validity is the degree to which it is warranted to generalize results to other contexts.[source: Wikipedia]

Example (Impact of advertisement)

You expose 45 subjects to a campaign to incentivise the use of buses.They report to you in a questionnaire that after seeing the ad they feel more inclined to take the bus by an average 20% more. Will the ad work also in the field?

- representativeness of the sample [to represent different types of people]
- size of the sample [to exploit statistical inference]
- control group [to ascertain effect different from baseline or random]
- incentive compatibility [to make sure that subject do not lie to you or to themselves]
- double-blind testing [both the experimenter and the subject are blind as to the purpose of the test]
- more...
- if *one* of the ifs above is not fulfilled *then* you can arguably have *low* external validity

How to get external validity: get near to theory

How do you get external validity?

Get as far away from reality as possible and near to theory

- theory is generalizable by default
- so the closer your expeiment is to theory, the higher the chances of falsifying it
- if you fail, theory survives and you can apply it to further contextualized settings
- because imagine theory fails in *one* little setting
- then it can be because of some confounds, some external influence, while theory is still valid
- if the test is non-contextual then it is in principle stronger

How to get external validity: get near to the real world & replicate

How do you get external validity?

Get as much real context as possible, and then replicate in several different contexts

- running an experiment in the field, within the target population, fully contextualized
- makes sure the result is valid in *that specific field*
- *but* leaves open the possibility that it will *not* replicate in *other* contexts
- from which we deduce the need for replication

Different types of experiments

You can hence choose your pick of experiment

Natural: the exogenous manipulation and the control group happen randomly and naturally in the population. The expeirmenter can simply look at the real world data [very rare!]

Field: the experimenter creates a control group and an exogenous manipulation, randomizing some aspects and keeping others as they appear in nature. [also called Randomized Control Trial]

Lab: the experimenter takes full control and *recreates the setting in a lab*; synthetically creates a control group and a manipulation, and sees details [this is the gold standard in the hard sciences]

Natural experiments

If you are lucky enough, nature provides you with a ready-made experiment

Conditions:

- the manipulation must be *exogenous* (i.e. out of the control of the agent)
- there must be no selection (i.e. subjects being able to opt-in or -out)
- subjects must be unaware of the manipulation
- the manipulation must fall across lines that while not random, are so from the point of view of the agents

Example (Crime rate decline in the US and abortion laws)

According to Steve Levitt (book: Freakonomics) the legalization of abortion in the US reduced crime 20 years later. How can he prove it? via a natural experiment. Some US states legalized abortion before the others – and saw earlier declines. Moreover some states made it harder than others to perform an abortion – and in those states crime declined less. [more details on the video]

Field experiments: Randomized Control Trials (RCT)

If you are a medical doctor, what you do is RCT

- Take a population fo interest (e.g.: lung cancer patients)
- administer a drug to a treatment group, a placebo to a control group
- track effects of the drug
- done.

Control vs. Randomization

- The setting, drug, background health levels, conditions of exposure to the drug, etc... are controlled
- The sample, number of patients, hospitals, control/treatment, are all randomized

But can it be generalized to other patients? other settings? slightly different drug? not really..

Lab experiments

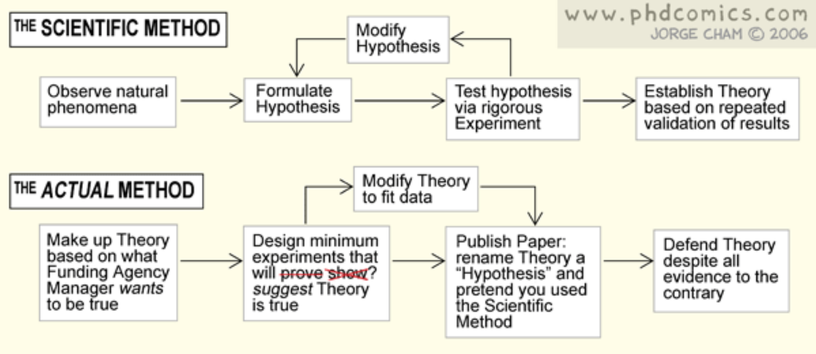
High internal validity, low external validity

- Recreate in a lab the *essential traits* of the setting to be studied
- That is: simplify, simplify, simplify!
- Keep only the minimal things that can prove your theory wrong
- Formally create one (or more) control groups
- done.

Control vs. Randomization

- All is controlled, synthetic, recreated in perfect (from the point of view of the theory) conditions
- Only allocation to treatment or control is randomized

Experiments



Test: if this *doesn't make you laugh*, you haven't understood. Study more.

Economics as a non-experimental science

The economic world is extremely complicated. There are millions of people and firms, thousands of prices and industries. One possible way of figuring out economic laws in such a setting is by controlled experiments.
A controlled experiment takes place when everything else but the item under investigation is held constant. Thus a scientist trying to determine whether saccharine causes cancer in rats will hold "other things equal" and only vary the amount of saccharine. Same air, same light, same type of rat. Economists have no such luxury when testing economic laws. They cannot perform the controlled experiments of chemists or biologists because they cannot easily control other important factors. Like astronomers or meteorologists, they generally must be content largely to observe.

Samuelson, Paul A., and William D. Nordhaus, 1965

Economics as a non-experimental science

What can we expect to learn from experiments in economics?

- Economies are indeed complicated
- Good data not always available
- Confounding factors abound
- Even if we could isolate a factor, how to run a proper experiment (e.g. exit the euro?)
- Even if we could run an experiment on a simple class game, how does this relate to the real world (external vailidity?)

Still, it is possible and there are things to be learned

Experimental Economics: peculiarities

ExEc is the use of the experimental method in economics

- Usually relies on simple abstraction of real problems (k.i.s.s – *keep it simple, stupid!*)
- Usually does *not* trust subjects to tell the truth unless properly incentivized to do so
- Usually relies on giving full knowledge to participants
- Never lies and always tells subjects the truth, *never deceiving them*

Experimental Economics: incentives

Experimental Economics: incentive-compatibility

Problem **hypothetical bias**

- Usually self-declared price > incentivized price
- The second decision is called **incentive-compatible**
- i.e., you have real consequences, and your best strategy is to reveal your true preferences

We go at great length to create incentive-compatible mechanisms

Experimental Economics: common knowledge

Rules of the game are common knowledge

- As in the previous example, all was clear *before* we set up the sale
- No surprises, no changes, no confederates (i.e. false participants)
- Usually instructions are read aloud to make sure that not only everyone gets the same
- But that everybody knows that everyone has heard the same (common knowledge)

Common knowledge levels the playing field

Experimental Economics: credibility and the ban on deception

How to be sure that everyone is playing the same game?

- Experimental Economists never lie to subjects
- (this is not the case in psychology experiments)
- Deception (lying or in other ways inducing false beliefs in the subjects) is banned
- This is made to make sure that the experiment is **credible**
- Subject can trust the experimenter do to exactly as he said
- Subjects' beliefs are aligned
- Internal validity is enhanced

Lying is usually the fastest route; no deception means hard work, but more trustworthy result


KEEP
CALM
it's
QUESTION
TIME