L6. Experiments...
are still experimental in sociology

Inquiries in sociology

Olivier Godechot

Experiments

- Artificial setting in order to isolate and study the impact of some precise causal mechanisms.
  - MANIPULATION holding everything constant
  - Replicable
- Galileo's experiment for studying the fall of bodies
  - Bells at different intervals in order to produce regular rhythm with a marble
  - Distance between bells follows uneven numbers: 1 3 5 7
  - Correspond to square distance:
    - 1, 1+3=4, 1+3+5=9, 1+3+5+7=16
  - Distance proportional to square time

$$d(t)=(a/2).t^2$$

https://www.youtube.com/watch?v=eUbv78PHaro

Experiments in sociology and other social sciences

- Durkheim… sociology a non-experimental science … partially true
  - Not against experiment… but thinks it is difficult to artificially manipulate a society…
  - "When the phenomena can be artificially produced at will by the observer, the method is that of experimentation proper. When, on the other hand, the production of facts is something beyond our power to command, and we can only bring them together as they have been spontaneously produced, the method used is one of indirect experimentation, or the comparative method." (Durkheim, 1894)

- Experiments common in social psychology
  - Milgram experiments
    - https://www.youtube.com/watch?v=rdrKCilEhC0 (full) ; https://www.youtube.com/watch?v=Kzd6Ew3TraA (short)
  - Stanford experiment
  - Cook and Emerson (1978): social exchange theory, reciprocity, etc.

The credibility revolution

- Experimental economics and the development of RCT in social sciences
  - Behavioral
  - Public Policy (Duflo and al. / Gurgand / etc.)
- Angrist & Pischke (2010) : "The Credibility Revolution in Empirical Economics: How Better Research Design Is Taking the Con out of Econometrics"
  - Improve the research design
  - Rather than complexify the statistical model

- Compare to the counterfactual !
  - Opposition between treated group & control group.
  - In order to know the treatment effect…
    - ➔ treated group differs from control group only by the treatment.

- Two methods
  - Randomized control trials (RCT) experiments
 - Natural experiments

1. Randomized controlled trial experiments

Non-randomized experiments (versus randomized ones)

- Many experiments non-randomized
  - Milgram experiment
  - Prisoner dilemma
  - Ultimatum game
  - Within individual experiment
- Comparisons of results to theoretical benchmark
  - Milgram => Morality concerns
  - Prisoner dilemma / Ultimatum game => Rational actor in Game theory

- Effect of a combined change in manipulation and sample ambiguous
  - If results of experiments very robust: Change in manipulation => change in results
  - If results of experiments not very robust: Change in sample => change in results ??
  - Within individual experiment => no path dependency hypothesis
- Randomized control trial.
  - Checking if change in results are due to change in manipulation and not to change in samples
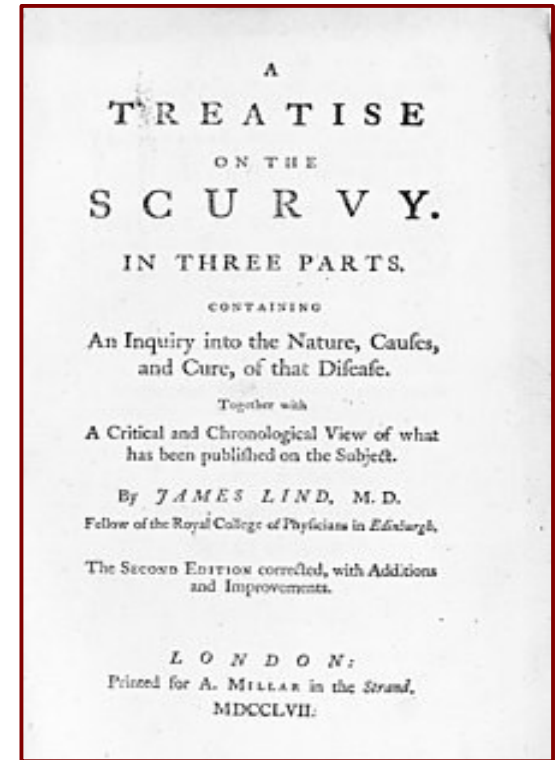
James Lind and the scurvy (1747)

"On the 20th of May 1747, I selected twelve patients in the scurvy, on board the Salisbury at sea. **Their cases were as similar as I could have them**. They all in general had putrid gums, the spots and lassitude, with weakness of the knees. They lay together in one place, being a proper apartment for the sick in the fore-hold; **and had one diet common to all**, viz. water gruel sweetened with sugar in the morning; fresh mutton-broth often times for dinner; at other times light puddings, boiled biscuit with sugar, etc., and for supper, barley and raisins, rice and currants, sago and wine or the like. **Two were ordered each a quart of cyder a day. Two others took twenty-five drops of elixir vitriol three times a day . . . Two others took two spoonfuls of vinegar three times a day . . . Two of the worst patients were put on a course of sea-water . . . Two others had each two oranges and one lemon given them every day . . . The two remaining patients, took . . . an electary recommended by a hospital surgeon . . .** The consequence was, that the **most sudden and visible good effects were perceived from the use of oranges and lemons**; one of those who had taken them, being at the end of six days fit for duty . . . The other was the best recovered of any in his condition; and . . . was appointed to attend the rest of the sick. Next to the oranges, I thought the cyder had the best effects . . ."
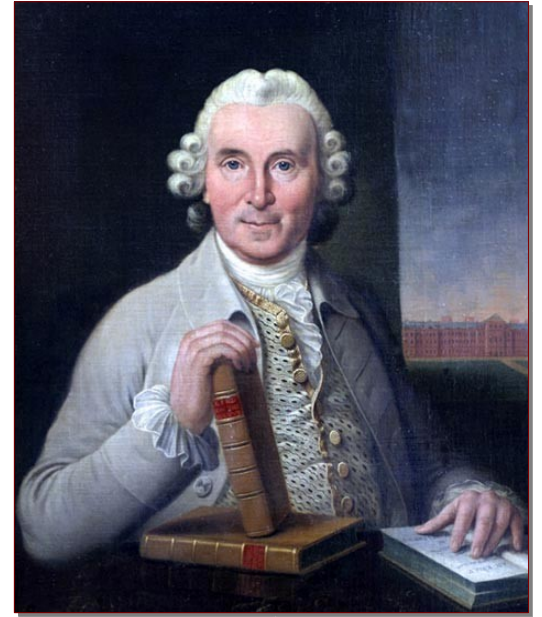
James Lind (1747)

- 12 sailors with scurvy. Divided in 6 groups
  - Same diet + supplementation
    - cider
    - Sulfuric acid
    - Vinegar
    - Herbs and spices decoction
    - Sea water
    - Oranges and lemon ➔ early cure
- First clinical trial (not totally randomized) … but not taken seriously
- Generalization of lemon in British Navy
- … and the role of C vitamin identified much later (1930)

A
TREATISE
ON THE
SCURVY.
IN THREE PARTS.
CONTAINING
An Inquiry into the Nature, Causes, and Cure, of that Disease.
Together with
A Critical and Chronological View of what has been published on the Subject.
By JAMES LIND, M.D.
Fellow of the Royal College of Physicians in Edinburgh.
The Second Edition corrected, with Additions and Improvements.
LONDON:
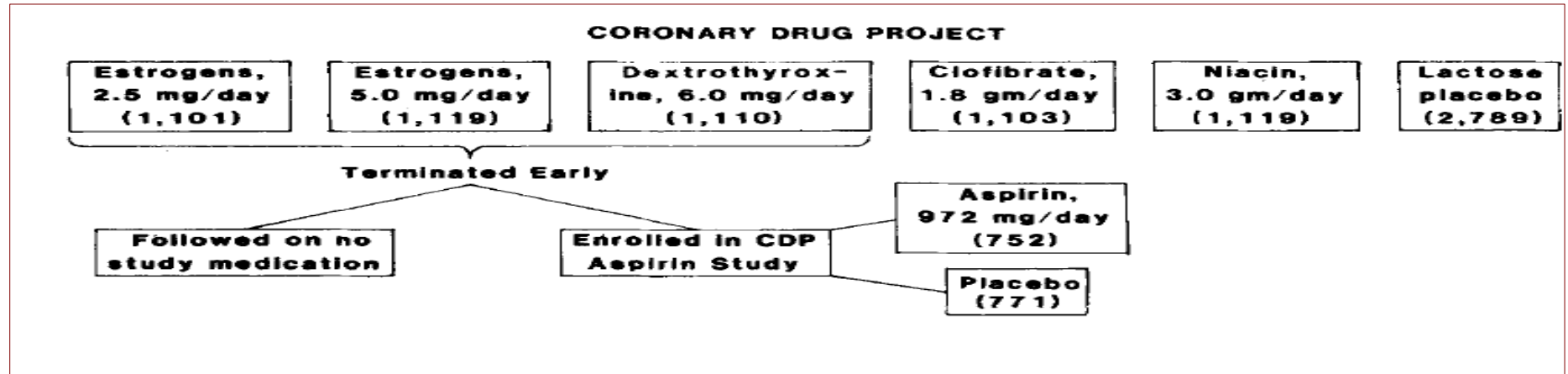Printed for A. Millar in the Strand.
MDCCLVII.

Randomized controlled trials (RCT) experiments

- End of 19th century
  - Peirce in psychology
    - Charles Sanders Peirce and Joseph Jastrow (1885). "On Small Differences in Sensation". Memoirs of the National Academy of Sciences. 3: 73–83
  - In education
- First RCT in medicine with a detailed research design:
  - 1948 "Streptomycin treatment of pulmonary tuberculosis"
- Now the standard of "evidence based" medicine

An example of evidence based medicine

Canner, Paul L., and al. "Fifteen-year mortality in Coronary Drug Project patients: long-term benefit with niacin." *Journal of the American College of Cardiology* 8.6 (1986): 1245-1255.

- Influence of different treatments for persons surviving Myocardial infarcts
- Randomized allocation of participants in different groups

CORONARY DRUG PROJECT

| Estrogens, 2.5 mg/day (1,101) | Estrogens, 5.0 mg/day (1,119) | Dextrothyrox-ine, 6.0 mg/day (1,110) | Clofibrate, 1.8 gm/day (1,103) | Niacin, 3.0 gm/day (1,119) | Lactose placebo (2,789) |

Terminated Early

| Followed on no study medication | Enrolled in CDP Aspirin Study |

| Aspirin, 972 mg/day (752) |
| Placebo (771) |

Results

Table 2. All-Cause Mortality (%) for a Mean Follow-up Period of 15 Years in the Estrogen, Clofibrate, Dextrothyroxine and Placebo Groups

Lipid-Lowering Drug	Drug		Placebo		z Value
	n	%	n	%	
Low dose estrogen	1,101	59.7	2,789	58.2	0.84
High dose estrogen	1,119	58.3	2,789	58.2	0.04
Clofibrate	1,103	57.8	2,789	58.2	−0.25
Dextrothyroxine	1,110	57.0	2,789	58.2	−0.67

Table 3. Mortality (%) by Cause for a Mean Follow-up Period of 15 Years in the Niacin and Placebo Groups

Cause of Death	Niacin	Placebo	z Value
All causes	52.0	58.2	−3.52
Coronary heart disease	36.5	41.3	−2.80
Cerebrovascular causes	1.4	1.6	−0.34
Other cardiovascular	4.5	4.8	−0.45
Cancer	4.0	4.4	−0.59
Other causes	2.9	3.0	−0.16
Unknown or not coded	2.7	3.0	−0.56
No. of patients	1,119	2,789	

Types of experiments

- Laboratory experiments
  - Milgram example / Ultimatum game / Social exchange theory
  - People enter a lab and know they participate to an experiment
  - Conditions are artificial (on purpose to be controllable and manipulable)
  - Not too costly but what relation with real life?
- Field experiments
  - Conducted in real life
  - Less artificial
  - Costly / Linked to public policy or firms
  - Ethical problems

- Other types in between lab or field
  - Online experiments
    - Could range from lab to field depending on the setting
    - Manipulation of people/behaviors online (easier)
    - With or without full awareness
  - Vignette experiments
    - Experiments within questionnaires
    - Very cheap
    - Respondents always sensible to the framing
  - Audit surveys

Field experiment:
Welcome in the real world…

Moen, P., Kelly, E. L., Fan, W., Lee, S. R., Almeida, D., Kossek, E. E., & Buxton, O. M. 2016. "Does a flexibility/support organizational initiative improve high-tech employees' well-being? Evidence from the work, family, and health network". *American Sociological Review*, *81*(1), 134-164.

© Slides adapted from Shi-Rong Lee

Research Questions

- Does providing workers real flexibility in the form of greater control over their working time and more supportive supervisors improve their subjective well-being?

  - Mechanisms

  - Sub-group analysis

The Work, Family, and Health Network (WFHN) and STAR Initiative

- STAR: Support. Transform. Achieve. Results
- Designed to:
  - Promote greater temporal flexibility (schedule control)
  - Provide greater supervisor support for family and personal issues
  - Reduce low value work and emphasize outcomes, not time at work

Real-World Research….

- Randomized field trial – promoting flexibility and supervisor support in an IT workforce (STAR)
- Natural Experiment – announcement of merger in middle of field trial
  - Early Survey Group was first interviewed prior to the merger announcement.
  - Late Survey Group was exposed to the merger announcement before the baseline survey and before STAR.
  - Hypothesis: STAR will have stronger effects on well-being outcomes for those in the Early Survey Group, who had completed or initiated STAR prior to the merger announcement.

Variables

- Independent variables
  - STAR
  - Early/Late Survey Group

- Dependent variables
  - Job satisfaction
  - Burnout
  - Perceived stress (non-supervisory employees only)
  - Psychological distress

Methods

- ## Sample

  - IT division of Fortune 500 firm called TOMO

  - 867 information technology (IT) workers

  - Managers and employees

- ## Data collection

  - Group-randomized field trial (56 study groups)

  - Three-wave data (Baseline - 6 months - 12 months)

- ## Analytical plan

  - Linear Mixed-Effects Models predicting outcomes at wave 3 (with lagged variables at baseline)

  - Multilevel Sober's test

Descriptive Statistics. 12 months later

	Panel A: Early Survey Group				
	STAR (N = 234)		Usual Practice (N = 219)		
	Mean/%	Std Dev	Mean/%	Std Dev	t-test
Dependent Variables					
Burnout by Wave 3	3.92	1.52	4.54	1.41	***
Job Satisfaction by Wave 3	4.12	0.71	3.89	0.77	**
Perceived Stress by Wave 3	8.18	2.56	8.80	2.85	*
Psychological Distress by Wave 3	10.12	3.00	10.83	3.58	*

	Panel B: Late Survey Group				
	STAR (N = 202)		Usual Practice (N = 210)		
	Mean/%	Std Dev	Mean/%	Std Dev	t-test
Dependent Variables					
Burnout by Wave 3	3.98	1.54	3.71	1.41	+
Job Satisfaction by Wave 3	4.10	0.75	4.01	0.80	
Perceived Stress by Wave 3	8.16	2.80	7.94	2.39	
Psychological Distress by Wave 3	10.03	3.19	9.87	2.68	

Online experiments

Online experiment

Salganik, Dodds & Watts. « Experimental study of inequality and unpredictability in an artificial cultural market. » *science* 311.5762 (2006): 854-856.

- Downloading website: 48 unknown songs
  - 14 341 participants: mostly teenagers
    - Randomized assignment of internet users towards one or the other platform
  - Aim of the test :
    - Role of social influence in inequality mechanisms: cumulative inequality dynamics and *Winner take all phenomenon.*

- Platform 1: no indication of other participants' downloading (independence)

- Platform 2: indication other participants' downloading.
  - Ventilation in 8 different world with different history of downloading
    - Experiment 1: in a random order (16*3)
    - Experiment 2: In one column following the reverse order of downloading indicator

Results

- More inequality under social influence (Figure 1)
- More unpredictability under social influence (Figure 2)
- Phenomenon stronger when order of downloading is more visible
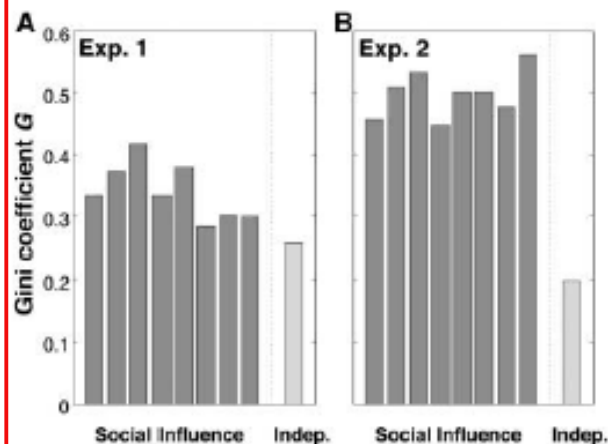
Fig. 1. Inequality of success for social influence (dark bars) and independent (light bars) worlds for (**A**) experiment 1 and (**B**) experiment 2. The success of a song is defined by m_i, its market share of downloads ($m_i = d_i / \sum_{i=1}^{S} d_k$, where d_i is song i's download count and S is the number of songs). Success inequality is defined by the Gini coefficient $G = \sum_{i=1}^{S} \sum_{j=1}^{S} |m_i - m_j| / 2S \sum_{k=1}^{S} m_k$, which represents the average difference in market share for two songs normalized to fall between 0 (complete equality) and 1 (maximum inequality). Differences between independent and social influence conditions are significant ($P < 0.01$) (18).

Fig. 2. Unpredictability of success for (**A**) experiment 1 and (**B**) experiment 2. In both experiments, success in the social influence condition was more unpredictable than in the independent condition. Moreover, the stronger social signal in experiment 2 leads to increased unpredictability. The measure of unpredictability u_i for a single song i is defined as the average difference in market share for that song between all pairs of realizations; i.e., $u_i = \sum_{j=1}^{W} \sum_{k=j+1}^{W} |m_{i,j} - m_{i,k}| / \binom{W}{2}$, where $m_{i,j}$ is song i's market share in world j and W is the number of worlds. The overall unpredictability measure $U = \sum_{i=1}^{S} u_i / S$ is then the average of this measure over all S songs. For the independent condition, we randomly split the single world into two subpopulations to obtain differences in market shares, and we then averaged the results over 1000 of these splits. All differences are significant ($P < 0.01$) (18).

Audit studies

Testing experiments.

Rivera, Lauren A., and András Tilcsik. 2016. "Class advantage, commitment penalty: The gendered effect of social class signals in an elite labor market." *American Sociological Review* 81(6): 1097-1131.

Figure 2. Combinations of Résumé Items that Together Signal Social Class Background

	Higher-class combination[a]	Lower-class combination[b]
Last name	Cabot	Clark[c]
Undergraduate athletic award	University athletic award[c]	University award for outstanding athletes on financial aid
Undergraduate extracurricular activity (2008-2011)	Peer mentor for first-year students[c]	Peer mentor for first-generation college students
Undergraduate extracurricular activity (2007-2011)	Sailing team	Track & field (relay team)[c]
Personal interests	Sailing, polo, classical music	Track & field,[c] pick-up-soccer, country music

Results

Table 2. Proportions of Applicants Receiving Interview Invitations by Gender and Social Class

	Interview Invitations / Applications	% Invited to Interview
Higher-class man	13/80	16.25
Higher-class woman	3/79	3.80
Lower-class man	1/78	1.28
Lower-class woman	5/79	6.33

Vignette experiments

Mize, Trenton D., and Bianca Manago. 2018. "Precarious sexuality: How men and women are differentially categorized for similar sexual behavior." *American Sociological Review* 83(2): 305-330.

Vignette

"*Michael* is currently single but has had multiple happy relationships *with women* in the past. *Michael* has only dated *women* and one of his relationships with a *woman named Emily* lasted for over two years. The other night, *Michael* met *Matt* and felt attracted to him. At the end of the night, *Michael and Matt* went home together and had a casual sexual encounter."

- 4 treatments:
  - (1) man with a heterosexual dating history but recent same-sex encounter,
  - (2) man with a gay dating history but recent different-sex encounter,
  - (3) woman with a heterosexual dating history but recent same-sex encounter
  - (4) woman with a gay dating history but recent different-sex encounter
- Question on attribution of sexual orientation
  - how likely they thought the target character was *heterosexual, bisexual*, or *gay/lesbian* (from 0 to 100)

Results

- Survey
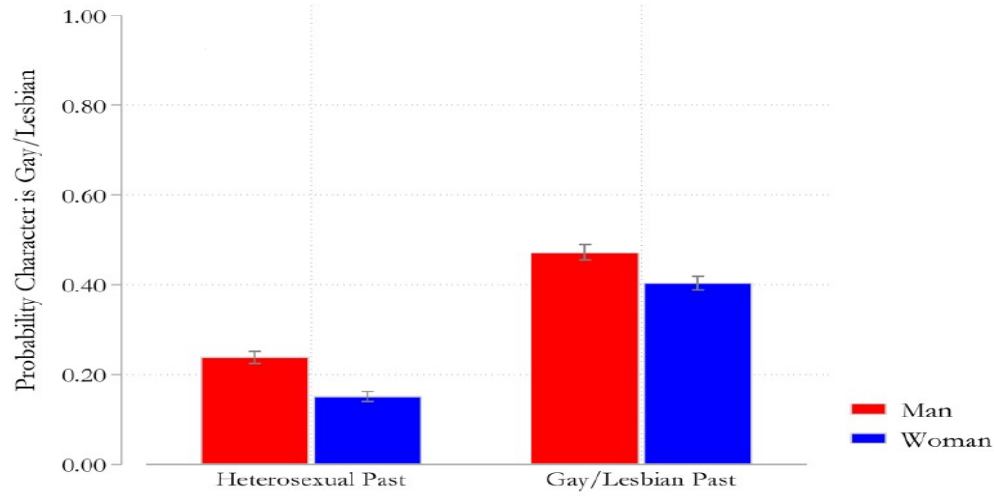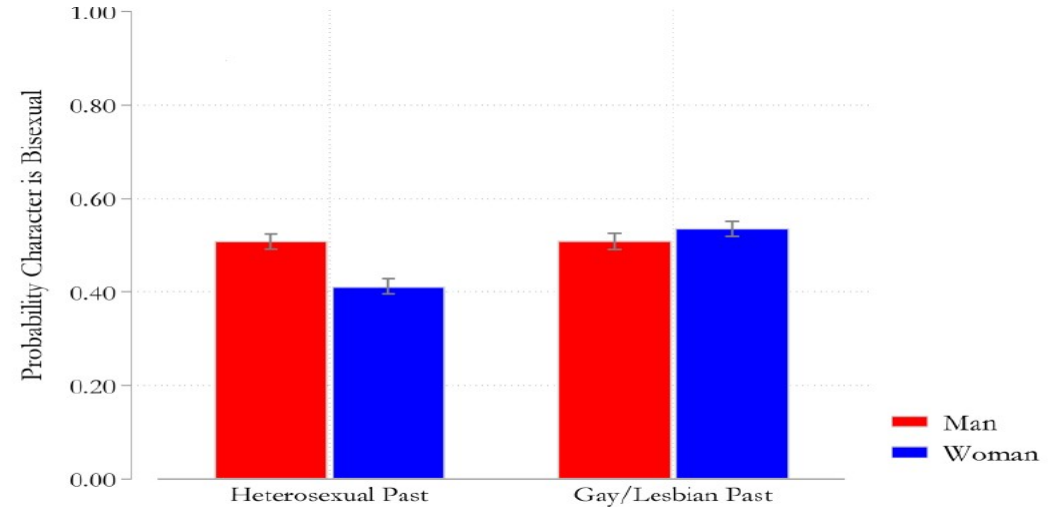- Nationally representative sample
- 2000 participants
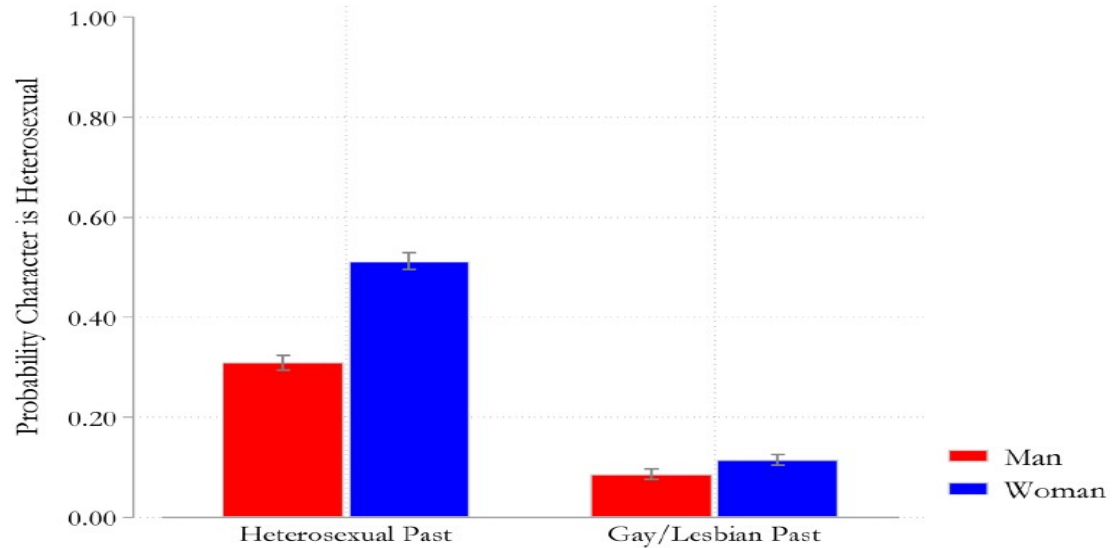
Figure 2. Probability Rating That Vignette Character Is Heterosexual, Study 1

Probability Rating That Vignette Character Is Gay/Lesbian, Study 1

Results continued

1. Probability Rating That Vignette Character Is Bisexual, Study 1

In one table

Table 1. Means for Sexual Orientation Percent Guesses Based on Vignette Characteristics (Top Panel) and Fractional Response Logit Regression Results (Bottom Panel); Study 1 ($N = 1,965$)

	Percent Certainty That Target Character Is Listed Sexual Orientation		
	Heterosexual	Bisexual	Gay/Lesbian
Woman (Heterosexual Past)	51	41	15
Man (Heterosexual Past)	31	51	24
Woman (Lesbian Past)	11	54	40
Man (Gay Past)	9	51	47
	Fractional Response Logit Results		
	Heterosexual	Bisexual	Gay/Lesbian
Woman vs. Man (Heterosexual Past)	.205***	−.101***	−.078***
	(.022)	(.023)	(.017)
Woman vs. Man (Lesbian/ Gay Past)	.029*	.027	−.068**
	(.014)	(.023)	(.023)

Sciences Po students' responses (Spring 2021)

	Heterosexuality	Bisexuality	Homosexuality
Female character with heterosexual past	54.17	74.17	32.73
(s.d. in parentheses)	(15.64)	(9.96)	(18.49)
	N=12	N=12	N=11
Male character with heterosexual past	41.82	79.09	44.55
	(16.01)	(15.14)	(18.64)
	N=11	N=11	N=11
Difference in means	12.35	-4.92	-11.82
P.value	0.08*	0.37	0.15

Sciences Po students' responses (Fall 2021)

	Heterosexuality	Bisexuality	Homosexuality
Female character with heterosexual past	35.00	75.42	37.50
(s.d. in parentheses)	(22.46)	(19.33)	(17.75)
	N=24	N=24	N=24
Male character with heterosexual past	42.14	74.64	38.93
	(19.88)	(19.53)	(15.95)
	N=28	N=28	N=28
Difference in means	-7.14	0.78	-1.43
P.value	0.23	0.89	0.76

Sciences Po students' responses (Fall 2022)

	Heterosexuality	Bisexuality	Homosexuality
Female character with heterosexual past (N=15)	30	83	35
(s.d. in parentheses)	(24)	(20)	(19)
Male character with heterosexual past (N=19)	52	79	36
	(18)	(14)	(16)
Difference in means	-21	3	-1
P.value	0.006	0.6	0.8

Example 2. The making of economists

Pablo Zamith. *The Making of Economists: A Transatlantic Investigation* PhD in progress.
- © Pablo Zamith

- Introduction of randomized questions
  - Randomization of the framing (pretext)
  - Test of 3 different effects:
    - Peer effect, master effect, model effect
  - "Proof of concept" rather than estimation of the magnitude of the true effect
    - If no significant effect➔ the effect might exist but design not powerful enough to show it
    - If significant effect ➔ qualitative proof of its existence but no estimate of true magnitude

Peer effects

MOYENNES DE VOTRE CLASSE
(données provisoires)

DEGRÉ D'ACCORD
23.1%
(tout à fait en désaccord)

Il faudrait réguler les marchés financiers si l'on veut éviter une prochaine crise

	Tout à fait en désaccord	En désaccord	Neutre	D'accord	Tout à fait d'accord	nsp

ASSIGNMENT	N	AGREEMENT
Control	157	57.5% (14.4)
Class average strongly agree	140	60% (20.3)
Class average strongly disagree	165	50.1%* (22.6)
Individual responses will be revealed to the class	141	47.9%* (13.9)

Master effect and model effect

« L'idée selon laquelle le salaire minimum détruit des emplois est toute simple: les travailleurs coûtent plus cher, donc on embauche moins »

**(Gilles Saint-Paul, économiste, Professeur à la Toulouse School of Economics)**

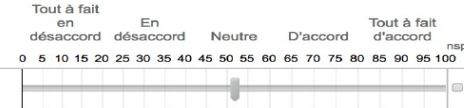Aujourd'hui, pour réussir à créer de nouveaux emplois, il faut baisser le salaire minimum

Tout à fait en désaccord	En désaccord	Neutre	D'accord	Tout à fait d'accord	nsp
0 5 10 15 20 25 30 35 40 45 50 55 60 65 70 75 80 85 90 95 100
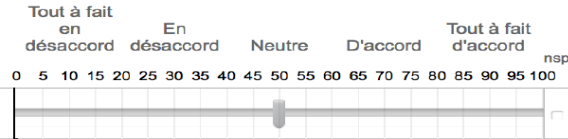
ASSIGNMENT	N	AGREEMENT
Control	121	49.5% (29.4)
Journalist	117	47.2% (21.1)
Nobel Prize	111	53.1%* (25.7)
Imaginary professor	120	50.3% (34.1)
Professor from UT1	119	55.9%* (22.1)

Dans votre cours de micro-économie vous avez vu le schéma suivant (inspiré du manuel de H.R. Varian) :

MARCHÉ DU LOGEMENT
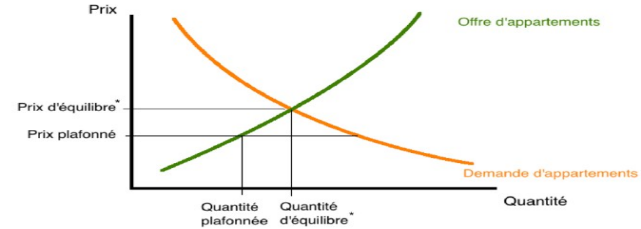
Grâce à ce modèle, on peut conclure qu'un plafonnement des loyers aurait pour effet de limiter l'offre, provoquer une pénurie d'appartements et constitue donc une politique défavorable aux personnes qui cherchent un logement

Tout à fait en désaccord	En désaccord	Neutre	D'accord	Tout à fait d'accord	nsp
0 5 10 15 20 25 30 35 40 45 50 55 60 65 70 75 80 85 90 95 100

ASSIGNMENT	N	AGREEMENT
Without model	300	58.7% (25.6)
With model	322	62.2%* (11.1)

From controlled experiments to natural experiments

- Natural experiments
  - Social and/or natural configuration which produces a (quasi-) random assignment of a population between a treated group and a control group (or treatment A versus treatments B)
  - Ex. Social lotteries. Weather shocks Date of birth. Public policy threshold of inclusion

- Ex. Godechot, 2016. "The Chance of Influence: A Natural Experiment on the Role of Social Capital in Faculty Recruitment", *Social Networks*
  - Part of the recruitment committee at EHESS randomly drawn out of the faculty
  - Enables to test the role of having a contact in the committee on the chances of recruitment

Experimental framework for testing the true causal effect of a given contact

- Experimental framework
  - Treatment: the contact is randomly drawn in the electoral commission
  - Control: the contact, although eligible, is not drawn in the electoral commission
  - Treatment's causal effect: treatment effect – control effect

- If the draw is really random, it is orthogonal (independent of) individual characteristics.
  - No unobserved heterogeneity. No reverse causality. No need to multiply control variables.

- However, the mechanisms through which contacts have an effect can be debated:
  - conscious favoritism, intellectual bias, shared common interests, reduced costs of evaluation, etc.

Main descriptive results

Table 7: Vote share and proposition of candidates by the electoral commission depending on the supervisor's membership of the electoral commission

Candidatures whose PhD advisor is	Mean (s.d.)	N	%	N	Mean (s.d.)	N	%	N	Mean (s.d.)	N	%	N
Randomly drawn member of EC	28.1 % (0.334)	62	34 %	62	31.3 % (0.345)	54	37 %	54	30 % (0.329)	48	38 %	54
Member of EHESS outside electoral commission	22.2 % (0.263)	360	20 %	360	21.7 % (0.261)	371	20 %	373	22 % (0.266)	377	20 %	379
Member of EC as a member of scientific council	30.4 % (0.307)	13	31 %	13	30.4 % (0.307)	13	31 %	13	30.4 % (0.307)	13	31 %	13
Member of EC as member of the bureau	31.2 % (0.323)	18	22 %	18	31.2 % (0.323)	18	22 %	18	31.2 % (0.323)	18	22 %	18
Outside EHESS	17.5 % (0.268)	1741	16 %	1756	17.5 % (0.268)	1741	16 %	1756	17.5 % (0.268)	1741	16 %	1756
All candidatures	18.7 % (0.27)	2194	17 %	2209	18.7 % (0.27)	2194	17 %	2209	18.7 % (0.27)	2194	17 %	2209
Definition of the membership the electoral commission	Drawn as titular or substitute if possible, presence otherwise				Drawn as titular (if possible, presence otherwise)				Presence (if possible, composition otherwise)			

Takeaways and limits

Why randomized experiment?

- Random assignment ensures that that all individual characteristics (both observed and moreover unobserved) have equal chances of being in treated or control groups
  - Several techniques: simple or stratified randomized assignment
- Estimation is not biased anymore by confounding variables (unobserved heterogeneity)

- Considerable simplification of statistical analysis
  - Intensity of treatment effect: indicated thanks to difference (or ratio) of average (or of proportions)
  - Significance: (Student T) test of the significance of the difference means (or of proportions)
- Randomized experiment versus random sample
  - Random sample: establish statistics that will give a true representation of a population➜ external validity
  - Randomized experiment: random assignment of a sample between control and treated ➜ internal validity.

Experiments and their blinds

- Simple blind
  - Participant does not know to which group assigned (treated or placebo)
- Double blind
  - Both patient and experimenter don't know in which group the participant is
- Triple blind
  - Patient, experimenter and statistician don't know
- Preregistration
  - FDA: declaration ex ante of the type of model that will be used

Random experiment in social sciences rarely uses true "placebo" and are rarely blind

- Placebo is a complex thing… It must have the shape, form, etc. than the treatment pill. (Secondary effects ?)
- Often 2 groups in social sciences: one gets the intervention and gets nothing.
- People know in which group they are.
- Attrition is not random, but might depend on treatment and ego's characteristics
- If that's the case: statistical analysis rather on the *intention to treat* rather than *on treatment on treated*

Identifying and interpreting:
Are we really testing what we say we test?

- Translating a theoretical mechanism to experimental manipulation
  - Example:
    - Theoretical hypothesis: Discrimination against homosexual males (Tilcsick 2011)
    - Experimental manipulation: Contrast in reply to applications between
      - treasurer of Gay and lesbian alliance
      - treasurer of Progressive and socialist alliance
- Bertrand, Marianne, and Sendhil Mullainathan. 2004. "Are Emily and Greg More Employable than Lakisha and Jamal? A Field Experiment on Labor Market Discrimination." *American Economic Review* 94:991–1013.
  - Imaginary linked to Emily and Greg (white) vs Lakisha and Jamal (black)
  - Information on race OR on race and class?

Technical limits

- Size of the sample (limit common to any approach based on statistical test)
  - Significance is function of size: $f(\sqrt{n})$
- Heterogeneity
  - RCT treatment effect ➔ average effect
  - Effect may be stronger in subgroups (males versus females, youth versus elderly)
  - Possible to analyze heterogeneity
    - Subgroup analysis.
  - Data mining risk
    - P-hacking => always possible to finds subgroups among which difference is significant
    - FDA: declaration of statistical analysis.

- Spillover effects
  - Subjects are not the only individuals impacted. Externalities (neighbors, kin, networks)➔ unintended effects in return.
- Debatable generalization
  - Internal validity of the gap treated vs control, within a given sample.
    - Often sample of volunteers, non representative (biased)
    - Inquirers are also volunteers and scientists
    - Estimated parameter == > don't hold for full population
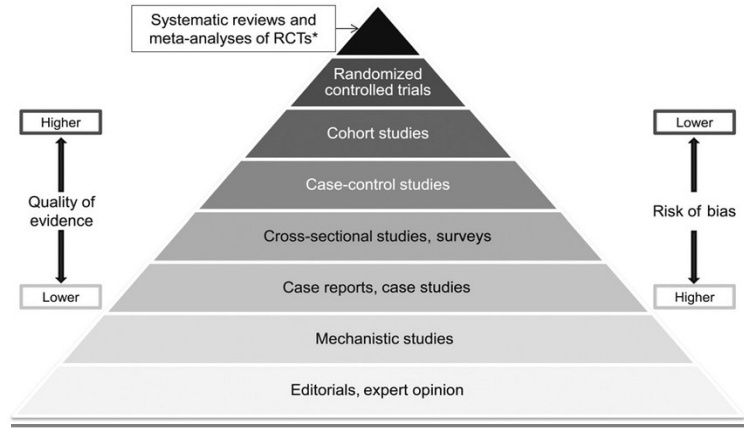 - Even if representative sample... Partial equilibrium validity. Does not correspond to general equilibrium validity

Conceptual limits

- Experiment different from real life
  - Both an advantage (replication, control) and a limit (little relations with true life situations)
  - Experimental conditions may modify the results of the experiment
    - Hawthorne Effect
    - Subjects want to please inquirers
    - Seen as a game
  - Experimental framing does not account for the embeddedness of social life
    - Multiple layers of interpretation
    - Sensitivity to the framing and the wording of the experiment

- Many social objects can not be the object of experiments
- Causal mechanism underlying the treatment's efficacy is often not clear
  - Cf. Lind. Why Lemon works?
- Ethical manipulation
  - Manipulation of subjects become a topics of discussion
    - Milgram experiment
  - Equity concerns between subjects
    - If outcome very different between treatment and control (cf. experimental cure of AIDS)

The golden standard of science and its critiscism

Is there a hierarchy of proof? As medical science believes

Discussable for social sciences

Deaton, Angus, and Nancy Cartwright. 2018. "Understanding and misunderstanding randomized controlled trials." Social Science & Medicine 210 : 2-21.

Deaton, Angus. 2010. "Instruments, randomization, and learning about development." Journal of economic literature 48 (2): 424-55.

- Unscrewing the superiority of RCT
- One method among others
- Interesting
- But with limits: heterogeneity and generalization
- Not "better"

Experiments in short

- RCT experiments has virtues
  - Randomization enable to establish mechanisms without risk of bias
  - Internal validity
  - "Severe test" à la Deborah Mayo (Epistemologist)

- But
  - External validity ?
  - Realism of experiments ??
  - Science of social behavior or science of social behaviors in experiments
  - Limited number of social phenomenon you can use in experiments

- One tool of inquiry among many
  - Good proof
  - Limited scope