Contents

What is a laboratory experiment

What is a laboratory experiment

Two different meanings in the literature:

- Decision-making investigation in a highly-controlled environment (existence of different treatments not necessary)
- Random assignment by investigator to treatment(s) or control

The first meaning examines/documents behavior and mechanisms; the second seeks to establish causality.

What is a laboratory experiment

Ex. 1: public goods game with 10 rounds

Possible findings:

- participants contribute more than zero
- contributions fall over time
- gender, age, risk tolerance, altruism, etc. predict contribution size (not implying causality)

No treatments assigned, but there was control and measurement of covariates

What is a laboratory experiment

Ex 2: public goods game with 10 rounds, two versions (treatments)

- T1: no possibility of punishment
- T2: possibility of punishment
- comparing results T1 vs. T2 allows us to assess the effect of the possibility of punishment on the contribution size
- possible finding: the possibility of punishment increases the contribution size (causality)

Design considerations: within- or between-subjects treatment assignment? what varies between T1 and T2? what is an appropriate control?

What is a laboratory experiment

- An experiment with randomly-assigned treatments (meaning #2) is not "superior" to a decision-making experiment (meaning #1)
  - they do not necessarily pursue the same goal (document a phenomenon vs. estimate an effect)
  - also, the two types are not mutually exclusive, and in fact their combination is powerful
  - an experiment with randomly-assigned treatments is not exempt from criticism

Deaton's critique to RCTs

"[...] learning about development requires us to investigate mechanisms. Finding out about how people in low-income countries can and do escape from poverty is unlikely to come from the empirical evaluation of actual projects or programs, whether through randomized trials or econometric methods that are designed to extract defensible causal inferences, unless such analysis tries to discover why projects work rather than whether they work." Deaton (2010), Nobel Prize 2015

What is a laboratory experiment

- Deaton's critique can apply to laboratory experiments: finding that a treatment has an effect does not in itself tell us why the treatment has an effect
- Knowing why a treatment has an effect may be relevant to advance knowledge (generate new questions and experiments) and to design more effective policies
- In our example, different follow-up experiments and policies might result depending on whether the possibility of punishment has an effect due to increased monitoring, increased trust in others' contributions, increased stigma of non-contribution, the prospect of losing money, etc.

What is a laboratory experiment

How do we investigate why the possibility of punishment has an effect?

What is a laboratory experiment

How to investigate mechanisms

1. Experiment with multiple treatments
    - T1: no possibility of punishment
    - T2: revealing individual identity and contributions
    - T3: sequential contributions
    - T4: possibility of symbolic punishment with points

What is a laboratory experiment

How to investigate mechanisms

2. Estimate the heterogeneity of the treatment effect over relevant covariates

- T1: no possibility of punishment
- T2: possibility of punishment
- examine the interaction between the treatment effect and beliefs about others' contributions or risk aversion

What is a laboratory experiment

Conclusion

- an experiment as understood in the literature may or may not have treatments
- the use of treatments depends on whether the goal is to draw comparisons and assess a causal effect
- when the goal is not only to estimate a treatment effect but also to understand its mechanisms of action, it is helpful to design multiple treatment variants and/or estimate heterogeneity of effect over relevant covariates

What is a laboratory experiment

In addition, a laboratory experiment in the social sciences is commonly distinguished (Harrison and List 2004) by the following:

- subjects are aware of being subjects, and the experiment occurs in the lab (duh!, although according to Charness, Gneezy, Kuhn 2013 there can be a field experiment in the lab when the lab is the representative context; e.g. examining participants' willingness to report subject overpayment)
- convenience sample (college students)
- moreover, in economics and political science, certain procedures have become standard practice

Procedures and their reasons

Procedures and their reasons

Economics and Political Science have converged toward a series of laboratory experiment procedures that not only are considered good practice, but also often have to be adopted in order to do work in many labs (eg CESS), receive funding, and publish in academic journals. Let's look at these procedures and the reasons for their existence, as we walk through the steps involved in a typical experiment

Procedures and their reasons

1. Recruit participants from a "convenience" pool
   - college students (convenient access by investigators working in the same university)
   - if recruiting also for future experiments, ask students to share their emails and accept to receive invitations to participate in future experiments
   - in this case it is helpful to add email addresses to a specialized subject pool management software (eg ORSEE or Sona Systems) to control communications, participation history, inclusion/exclusion criteria, etc

Procedures and their reasons

2. Invite participants according to inclusion/exclusion criteria

   - typical experiment consists of multiple sessions of 1-2 hr each (eg 5 sessions of 24 participants each = 120 participants)
   - implement sessions under as similar conditions as possible (schedule, experimenter, etc)
   - oversubscribe participants in each session to counter no-shows (pay a "show-up fee" to those not allowed in, and let them re-register to a later session)
   - information to include in invitation: average pay, session length, session schedule, little knowledge required

Procedures and their reasons

3. Obtain informed consent

   - obtain before session begins
   - information to provide: risk of harm and potential benefits; voluntary participation can be withdrawn at any time; privacy and anonymity conditions; no deception involved; potential pay; researcher and ethical committee contact information in case of questions/complaints
   - allow participants to keep a copy of the consent form
   - make sure participants sign the form

Procedures and their reasons

3. Obtain informed consent

   - if the researcher can anticipate that participants might be reluctant/ashamed to decline to participate in public (eg admit to suffering a psychological condition that is an exclusion criterion) then information about such exclusion condition should be provided before participants come to the lab (eg at the invitation stage)
   - example of consent form available on request

Procedures and their reasons

Why no deception?

- to ensure participants are responding to the incentives presented to them
- the participant pool is a public good to other researchers in the lab

Procedures and their reasons

4. Provide instructions at the start of and during the session

   - written (preferably shown on the computer rather than on paper to avoid leak of info) and read aloud to establish common knowledge
   - typically in neutral language (concrete terms such as "employee" may help comprehension but reduce control over what context participants perceive)

Procedures and their reasons

4. Provide instructions at the start of and during the session
   - show examples and/or comprehension quizzes to reduce noise, taking care not to suggest specific behavior, and using quiz results in robustness tests of the results
   - answer participants' questions privately to maintain control and similarity across sessions

Procedures and their reasons

5. Participants make decisions, express opinions, provide information

- typically privately and anonymously unless identity is of research interest
- variables that can prime behavior and are hardly primed themselves (eg gender, race, socioeconomic level) tend to be elicited at the end
- interface preferably computerized to speed up session and data processing, and to increase the range of possible stimuli (typical platforms are zTree, oTree, Qualtrics, pen and paper)

Procedures and their reasons

6. Monetarily incentivized decisions

   - pay not only for participation but also depending on participants' decisions
   - tasks and elicitations are "incentive compatible": incentives are aligned with preferences and it does not pay to misrepresent preferences
   - participants cannot leave with less money than they come in with
   - participants typically receive earnings information at the end of the session to maintain comparability between subjects (and within subjects across rounds) and to avoid wealth effects

Procedures and their reasons

6. Monetarily incentivized decisions

   - often one or a few rounds/participants are chosen at random at the end to count for payment, to reduce costs and to maintain similar incentives across
   - payment in cash (to avoid satiation) at the end of the session
   - average at CESS: 8 USD Santiago, 10 GBP Oxford

Procedures and their reasons

Why incentivize monetarily?

- large part of behavior that interests us is in response to benefits and costs, thus we want subjects to face those
- what one says one would do is not necessarily what one actually does when there is skin in the game

Procedures and their reasons

Why incentivize monetarily?

- however, note that the literature finds mixed evidence on whether and when monetary incentives affect behavior in experiments (Camerer & Hogarth 1999, Bonner et al. 2000): (i) they do not affect average behavior in market situations such as negotiation, exchange, risk taking; (ii) they affect average behavior when greater effort results in better performance; (iii) they generally reduce the variance in behavior
- also worth noting that some successful theories (eg prospect theory) and key economic indicators (GDP, inflation) are constructed at least in part from unincentivized self-reports

Procedures and their reasons

In addition to the procedures discussed, researchers must take some steps prior to conducting an experiment

- ethical training
- ethical approval of experiment
- possibly, pre-registration of the research

Procedures and their reasons

0.1. Ethical training

- the researcher must understand the responsibilities involved in working with human subjects
  - guarantee voluntary and informed participation (exceptions exist)
  - minimize risks and maximize benefits
  - ensure confidentiality and de-identification of data
- the researcher must demonstrate ethical training when applying for funding, seeking to run at labs such as CESS, and submitting manuscript to journals
- ethical training, open and free at USA NIH

Procedures and their reasons

0.2. Ethical approval of experiment

- external committee evaluates ethicality of proposal
- universities generally have an ethical committee
- good to have in mind that approval can take weeks
- example of CESS application available on request

Procedures and their reasons

0.3. Pre-registration

- it is a tool to deal with selective reporting of positive results
  - aka "p-hacking," "data fishing"
  - selective reporting invalidates p-values

Procedures and their reasons

Simplified example of how selective reporting invalidates p-values

- suppose we have a 100-sided die (1-100)
- if rolled once, the prob. of it landing on 96-100 is
  - p(96-100, 1 roll) $= 0.05$
- suppose our exercise is to roll it once and to conclude that it is biased if it lands on 96-100

Procedures and their reasons

Simplified example of how selective reporting invalidates p-values

- if we throw it more than once, but report having thrown it only once
  - p(not 96-100, 1 throw) $= 0.95$
  - p(not 96-100, 2 throws) $= (0.95)^2$
  - p(96-100 at least once, 2 throws) $= 1 - (0.95)^2 = 0.0975$
  - p(96-100 at least once, 3 throws) $= 1 - (0.95)^3 = 0.1423$
- the probability of landing on $96 - 100$ is no longer 0.05
- problem: selective reporting of positive results introduces a greater number of false positives (false results y conclusions) in the literature and body of knowledge

Procedures and their reasons

solution 1: transparent reporting of all experiment procedures and analysis

- but that may make for a cumbersome (to write and to read) manuscript
- but that requires us to trust in the transparency of the reporting researcher

Procedures and their reasons

solution 2: pre-registration

- it is a document describing the experiment design and analysis plan before data collection
- allows separating hypothesis testing from exploratory analysis
- must be registered (unalterable and time-stamped) for its value added to apply
  - *list of registries*
  - *aspredicted.org*
- *Alejandro Ganimian's pre-analysis plan template*

Procedures and their reasons

solution 3: replications

- pushed for by some *lab experimentalists*
- attempts ex-post identification of false positives
- not many (but increasingly more) journals explicitly welcome replications

Advantages of the lab

Advantages of the lab

- high control
  - ensures participants do what one thinks they are doing
  - negligible attrition
  - practically full treatment compliance
  - high internal validity
  - allows measurement of relevant variables that are inaccessible in other settings
- easier replicability, which is a pillar of the scientific process
- relatively simple logistics and quick execution (important to consider if work is for thesis)

Criticisms to the lab

Criticisms to the lab

- participants are subject to unprecedented scrutiny
- abstract and unrealistic context
- participants are not representative of the general population
- results do not generalize (little external validity)

Criticisms to the lab

Criticism: unprecedented scrutiny

- creates "experimenter demand effects"

Response:

- it allows for high control, internal validity, and replicability to a level not easily accessible in the field and online
- no evidence of such effect (Camerer 2015)
- it presupposes that participants know the researcher's hypothesis
- it requires participants to be willing to forgo money to please the experimenter
- many field experiments also involve scrutiny, such as from the implementer NGO

Criticisms to the lab

Criticism: abstract and unrealistic context

Response:

- allows the researcher to isolate the mechanism of interest and "turn off" other channels, boosting internal validity
- monetary incentives gain us realism
- private decisions/interactions via the computer are increasingly common in daily life
- the lab, as any tool, is appropriate for some questions but not for others

Criticisms to the lab

Criticism: college students are a highly selected group

Response:

- true: participants are willing to share their emails with the researcher, add the experiment session to their schedules, come to the lab on time, give informed consent...
- if experimenting on a specific population is important, it may be possible to bring a mobile lab to them

Criticisms to the lab

Criticism: little external validity

"Perhaps the most fundamental question in experimental economics is whether findings from the lab are likely to provide reliable inferences outside the laboratory." Levitt and List (2007)

Criticisms to the lab

Criticism: little external validity

Response:

- the lab does not automatically have little external validity—it is a question to be answered
- some lab settings approximate closely to the external world, eg:
  - effect of providing social information on donation behavior when making a transaction via the computer
  - standardized tests
- clearly it is not the tool to examine certain questions (eg effect of school uniforms on student attendance)

Criticisms to the lab

Criticism: little external validity

Response:

- criticism assumes a "public policy" view of experiments, but in the "scientific" view all evidence can help to advance knowledge and it is equally valid to ask whether the field generalizes to the lab (Camerer 2015)
    - eg psychologists run rarefied experiments on memory, attention, perception, sensation, etc that are perfectly valid and valuable pieces of scientific knowledge
    - similarly we can be interested in understanding economic behavior, rationality, etc. independently of the public policy view

Criticisms to the lab

Criticism: little external validity

Response:

- the field can also suffer from little external validity (results dependent on geography, participants, quality of implementer)
- more measured take: attaining external validity in the lab is important and possible (Kessler y Vesterlund 2015)
- if he wants to generalize his findings, the onus is on the experimenter to demonstrate external validity

Criticisms to the lab

Criticism: little external validity

Response:

- recent effort to combine lab experiments with administrative data to examine external validity
- we'll see an example

Example

Example

Gender differences in willingness to compete

- lab experiment "without treatments"
- results replicated in 15+ lab experiments
- recent effort to combine a lab experiment with administrative data to examine external validity

Example

*Niederle, M. and Vesterlund, L. (2007). "Do women shy away from competition? Do men compete too much?" The Quarterly Journal of Economics 122(3):1067-1101.

*Reuben, E., Sapienza, P., and Zingales, L. (2016). "Taste for competition and the gender gap among young business professionals" Working Paper.

Example

Motivation

- persistent gender gap in labor market outcomes
  - salary gap
  - occupational and hierarchical segregation
- some main explanations in the literature:
  - preferences for flexibility
  - human capital formation
  - discrimination
- more recent explanation from experimental evidence: women are less willing to compete than men
  - reluctant to enter competitive (better paid) careers
  - reluctant or less effective at negotiating raises and promotions

Example

Niederle, M. and Vesterlund, L. (2007)

- goal:
  - document a phenomenon: gender gap in willingness to compete
  - investigate mechanisms driving the gap
- possible mechanisms:
  - attitude toward competition (main hypothesis)
  - ability
  - beliefs over relative performance
  - risk aversion and negative feedback aversion
  - social preferences (unwillingness to impose externalities on others)
- teasing out mechanisms is often *the* problem of experimental design

Example

Experiment:

- participants seated in rows of 4 form a group (2 men and 2 women, but no explicit emphasis on gender)
- participants have 5 min to add as many sets of 5 2-digit random numbers as they can

Example

$$57 + 33 + 18 + 11 + 49 =$$

$$23 + 44 + 31 + 33 + 60 =$$

$$84 + 12 + 78 + 15 + 21 =$$

$$59 + 59 + 27 + 83 + 32 =$$

$$43 + 75 + 11 + 74 + 52 =$$

Example

Experiment:

- participants seated in rows of 4 form a group (2 men and 2 women, but no explicit emphasis on gender)
- each participant has 5 min to add as many sets of 5 2-digit random numbers as he can
- for each sum the participant learns whether answer was correct, but learns nothing about others' performance or relative position until the end of the session

Example

4 parts

- only one chosen at random for payment at the end of the session
- participants know there are 4 parts, but instructions for each part come at the beginning of the corresponding part (to avoid interactions across)

Example

- Part 1: \$0.50 piece rate
- Part 2: highest performer in group gets \$2 per piece, others get \$0
- Part 3: choice between piece rate of \$0.50 or competitive rate \$2 if one beats other members' performance from Part 2
- Part 4: choice between piece rate of \$0.50 or piece rate of \$2 for Part 1 if one outperformed others in Part 1 (no more sums in this part)
- before the end of the session, receive \$1 for correctly guessing own raking in Part 1 and Part 2

Example

Outcome variable of interest:

- proportion of men vs. women choosing to compete in Part 3

So, why so many other parts?

Example

Why so many other parts?

- if the goal were only to document a competitiveness gap, Part 3 would suffice
- but the goal is also to explore why the gap exists (to gain deeper understanding and possibly suggest policies)

Example

Why so many other parts?

- Parts 1 y 2 let us measure task ability under piece rate and tournament
  - if women are less willing to compete in the experiment, is it because of lower ability at the task?
  - also let us determine whether the decision to compete or not maximizes individual earnings (ceteris paribus)

Example

Why so many parts?

- Part 3: Why compete against Part 2 performance?

Example

Why so many parts?

- Part 3: Why compete against Part 2 performance?
  - if women are less willing to compete in the experiment, is it because of greater concern for affecting others' payoffs?
  - to turn off social preferences channel
  - to compete against individuals who also perform under a tournament
  - who would one compete against if others decided not to compete?

Example

Why so many parts?

- Part 4 lets us "subtract" or control for risk aversion and feedback aversion in a choice as similar as possible to the decision to compete, but without the element of having to compete
  - if women are less willing to compete in the experiment, is it because of more risk and/or feedback aversion?
  - subtraction design: if the competitiveness gap appears (or is larger) in Part 3 but not in Part 4, competition per se plays a role in generating the gap

Example

Why so many parts?

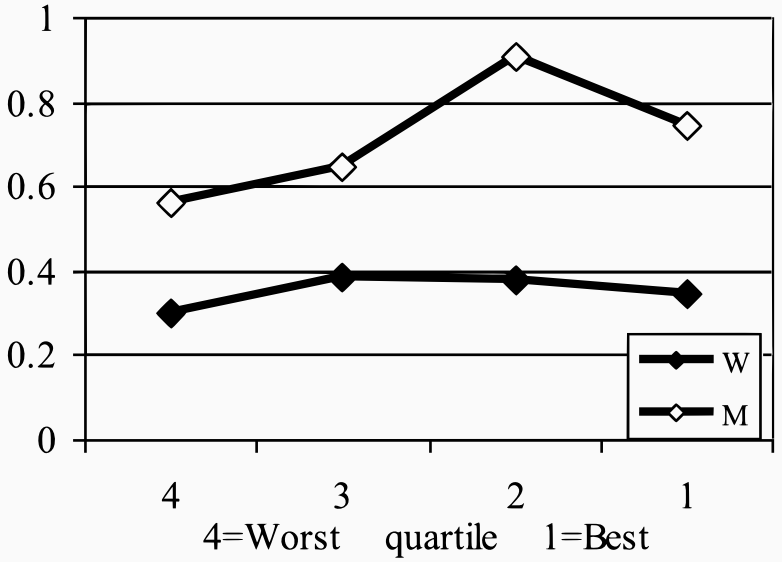
- guessing one's ranking lets us measure beliefs about absolute and relative ability
    - if women are less willing to compete in the experiment, is it because of lower self-confidence?

Example

Results, in brief:

- men and women have equal ability: 10 correct in Part 1 and 12 in Part 2
- 35% of women and 73% of men decide to compete in Part 3
- 25% of women and 55% of men decide to enter their Part 1 score into the tournament in Part 4
- 43% of women and 75% of men believe they are the best in their groups (both genders overconfident, but men more often so)
- gender gap in Part 4, but not in Part 3, is entirely explained by differences in self-confidence

Example



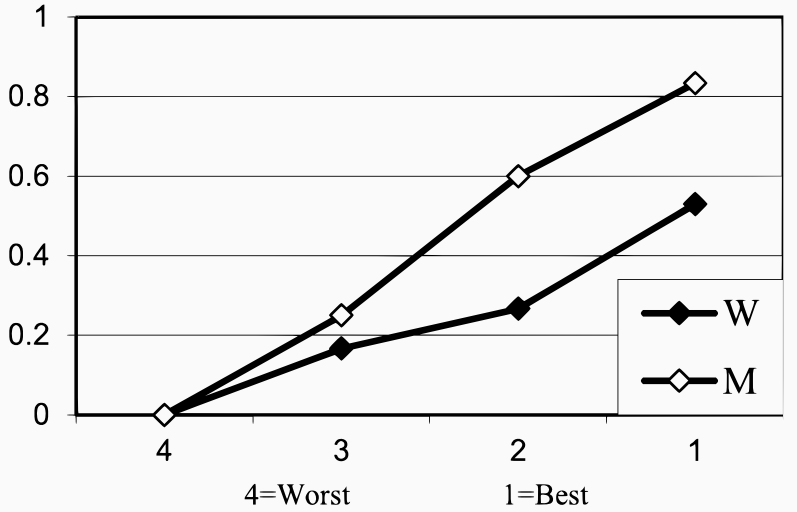


Figure 1: Proportion of participants selecting the tournament: (a) depending on performance quartile, (b) depending on believed performance rank, for women (W) and men (M), separately.

Example

Conclusion:

- women are less willing to compete than men, even though they have the same ability, and beyond the role played by self-confidence, risk tolerance, and feedback aversion

Example

- study published 10 years ago
- result replicated 15+ times in the lab
- but, does the competitiveness gap explain at least part of the labor market outcomes outside the lab?

Example

Reuben, E., Sapienza, P., and Zingales, L. (WP)

Design:

- experiment similar to previous
- participants are entrant Chicago Booth MBA students
- combination of competitiveness results with administrative data:
  - admission (sociodemographics and academic records)
  - graduation (salary and industry type of first job after graduation)
  - survey 7 years after experiment to track labor market progression

Example

Reuben, E., Sapienza, P., and Zingales, L. (WP)

Results:

- women earn 15% less than men at graduation
- women have 8(12)% larger probability of working in lower paid industries at graduation (7 years later)
- participants who choose to compete in the experiment earn $15k more at graduation (controlling for gender)
    - competitiveness explains 10% of the gap
- participants who choose to compete in the experiment are more likely to work in higher paid industries at graduation and 7 years later

Designing a lab experiment

Steps and tips for designing an experiment

- we've discussed how to conduct an experiment
- in other sessions we'll discuss how to program an experiment
- we'll now discuss some steps and tips for designing an experiment, in reference to the competitiveness examples

Steps and tips for designing an experiment

- formulate a research question
  - can it be studied in the lab while maintaining sufficient relevance?
  - the gender gap in labor market outcomes is quite a complex issue and yet the lab can shed some light
- define the outcome variable(s)
  - simple binary decision of whether to compete
- define the treatments, if they exist
  - variants of a mechanism that one wishes to evaluate
  - variants of an experiment that turn off alternative mechanisms
- if there are treatments, should they be administered within or between subjects?

Steps and tips for designing an experiment

- identify control variables helpful in the analysis
- determine number of participants and sessions
  - function of required number of independent observations, lab seating capacity, interactions between subjects, subgroup analysis, etc.
- computerized, or pen and paper?
  - computerizing saves time in session, avoids need for digitizing data, allows to reach larger number of participants, permits interactions between participants, allows presentation of greater range of stimuli
- write instructions
- write a session script

Steps and tips for designing an experiment

- simplify the design when possible — Einstein: "Everything should be made as simple as possible, but no simpler"
  - know in advance how data will be analyzed
- an experimentalist is someone who intervenes the environment to generate his own data
  - the quality of the data is thus the experimentalist's responsibility