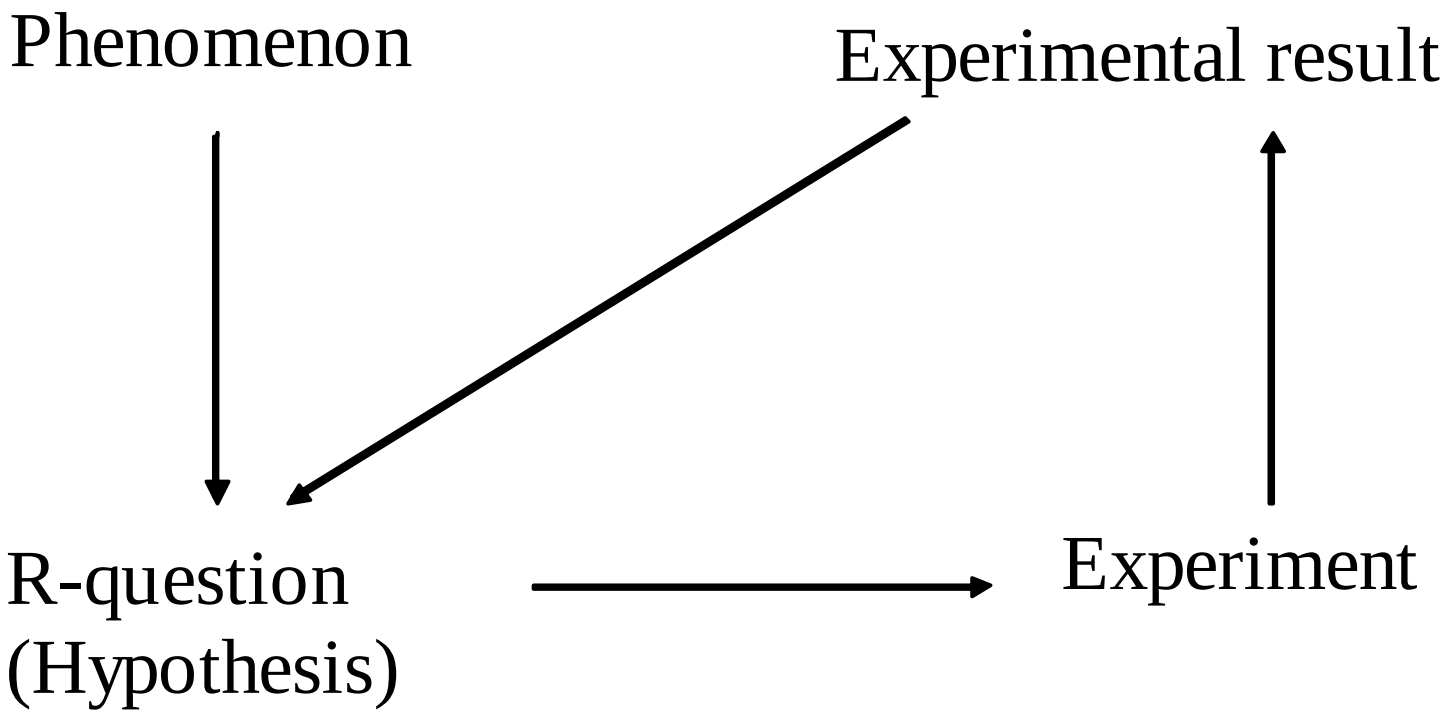
How to do experiments
(a personal view)

Economic phenomenon, hypothesis and experiment



Steps

- Questions
- Design
- Hypotheses
    - Standard
    - Alternative
- Preparing the experiment
    - Instructions
    - Computer program
    - Procedural details
- Running the experiment
- Data analysis
- Writing the paper and presenting the results

What is a good experiment?
Seven questions by Shyam Sunder

1. What is the question that you would like to have answered after the experiment? (Your answer should be a single sentence with a question mark at the end.)
2. What do you know already about the possible answers to the question you have stated above?
3. What are the various possible ways of finding an answer to the question you have stated above? Include both experimental as well as any other methods you know.
4. What are the advantages and disadvantages of using an experiment to find an answer?
5. What are the chances that the answer you get from the experiment will surprise you or others? What are the chances that it will change someone's mind?
6. How would you conduct the experiment? (Write down a design and instructions.)
7. Is your experimental design the simplest possible design to help answer the question you have stated?

Some expressions and technicalities

- **Treatment**: a particular condition of the experiment
  - Often a (main) treatment and a control treatment (or more)
  - Everything else kept equal, only one change
- An **experiment** usually consists of several **sessions**
  - In a session a group of people takes part in the experiment at a particular date and place
- **Subjects** = participants in the experiment

Across- und Within-Subject-Design

- Within-Subject-Design: Subjects participate in more than one treatment
  - Allows individual comparison
  - Control for individual fixed effects
  - More powerful tests
  - Order effect problem
    - In the second treatment subjects have learned something already
    - Solution: reverse order to control for order effects
      - AB / BA Design
- Across-Subject-Design: Subjects participate only in one treatment

- Note: Different designs require different statistical tests
- Example:
- Non-parametric tests for equality of distributions (null hypothesis: same distribution)
  - Matched pairs (within subj.): Wilcoxon test
    - Stata: *signrank x = y*
  - Independent samples (across subj.) Mann Whitney test
    - Stata: *ranksum x, by(treat)*

Wilcoxon Signed Rank Test

- 2 Treatments „Within-Subject"
- H_0: treatments are the same
- Procedure
  - Take differences
  - Sort differences according to size of difference
  - Allocate ranks (smallest Diff. → rank 1)
  - Sum of ranks for positive differences determine T^+
  - Evaluate p-value for T^+ (Table)
    - H_0 can be rejected on 5% level if $p \leq 0.05$
    - H_0 cannot be rejected if $p > 0.05$

Wilcoxon-Mann-Whitney Test

- „Across-Subject"
- Assumption: ordinal data
- H_0: treatments have the same distribution P[X>Y]=1/2
- Procedure
  - Sort observations according to size of observation
  - Allocate ranks (smallest value → rank 1)
  - Calculate sum of ranks to determine W_X
  - Evaluate p-value for W_X(Table)
    - H_0 can be rejected on 5% level if $p \leq 0.05$
    - H_0 cannot be rejected if $p > 0.05$

What are observations?

- Distinction between observation and statistically independent observation
  - Example: 5 Sessions of a market experiment with ten periods and ten trades each
  - 500 price observations
  - Only 5 independent observations (means per session, which in this case is a matching group)
  - Independent because no interaction across matching group
- If there are only few (independent) observations, experimentalists often use non-parametric tests instead of regressions

- Example:
  - 10 matching groups play public goods game, two treatments
  - Decisions: contribution level c
  - Question: Are contribution levels different across treatments?

  Regresssion
    - Stata: *reg c treatdummy, r cl(mg)*
  - Non-parametric test: take averages of the 10 matching groups, attach a treatmentdummy
    - Stata: *ranksum c, by(treatdummy)*

One-Shot vs. repeated observations

- Pro One-Shot
  - Strong incentives for decision
  - No strategic spillovers across periods (particularly important if „true preferences“ are to be elicited
  - Easy to perform and short
- Pro repetitions (“repeated one-shot”)
  - Learning
  - Possible to observe dynamics, e.g., convergence to predicted or behavioral equilibrium
  - More observations

Implementing repeated games

- E.g., partner design: Groups of subjects stay together for more than one period (and know that)
- Finitely repeated game
  - If only selfish types and unique Nash equilibrium in stage game: backward induction gives solution to game (start in last period…).
    - If stage game has multiple Nash equilibria, "anything goes": loss of a clear prediction.
  - If multiple types (e.g., reciprocal and selfish players) many Bayesian equilibria, see Kreps et al. (1982).

- "Infinitely" repeated games
  - Implementation with the help of a termination probability
  - Problem: length of the experiment is endogenous
    - Do you want to throw dice for five hours?
    - What if after the first period the game ends?
    - Different sessions have different lengths

Partner- vs. Stranger Design

- Partner (groups of subjects stay together for several periods)
  - Within a group of partners you have one observation (4 "partner groups" of 4 subjects each vs. one "stranger group" with 16 subjects: 4 vs. 1 independent observations)
  - Allows analysis of strategic considerations
- Stranger (groups are recomposed randomly)
  - Similar behavior/prediction as "one-shot" but more observations
- Perfect stranger: probability of being re-matched with the same person is exactly zero (and subjects know that)

Strategy method

- Strategy method was first used by Reinhard Selten
- Idea: Instead of just playing the game, subjects are asked to indicate an action at each information set
  - i.e., the experimenter really elicits a strategy
- Example: Sequential prisoner's dilemma
  - Second mover is asked: What do you do (defect or cooperate)
    - if first mover cooperates
    - if first mover defects

- Advantages
  - More information about motivation/behavior of players
    - Figure out, e.g., that someone is a reciprocal player, even though first movers always defect
  - Information about how people would play "off equilibrium" or "off action path" (since this is not usually reached, you have no information how they play unless use strategy method)
- Problems
  - Incentives are weaker, since each information set is reached only with probability < 1.
  - Hot vs. cold emotions: People might feel and act differently knowing they have reached a particular information set, compared to potentially reaching it
  - Explaining the SM to subjects is tricky (loss of understanding, control)
  - Lose move structure of game

- Does SM induce a different behavior relative to a situation where a subject responds to the actual move of an opponent?
  - Brandts and Charness (1998 "Hot versus Cold: Sequential Responses and Preference Stability in Experimental Games", Discussion Paper, Universidad Autonoma de Barcelona)
  - Cason and Mui ("Social Influence in the Sequential Dictator Game", Journal of Mathematical Psychology)
  - report evidence indicating that the strategy method does not induce different behavior.
- Moreover: You may use SM in all your treatments, and focus on treatment differences

Role reversal

- Role reversal: Subjects act in different roles, e.g., in the ultimatum game as a proposer and a responder
- Helps to put oneself in the shoes of the other person. If this is what you want to study, fine.

- May not be a good procedure, because you lose information about how people act in a given role.
- On top: potential strategic considerations (e.g., ultimatum game, 2 periods, partners)

Learning trials

- In complicated experiments (e.g., with difficult trading rules in markets) it is a good idea to have subjects try out the rules of the game first, without monetary consequences
- Advantage
  - It guarantees subjects' understanding from the first paid period on
  - Allows answering "new" questions of subjects that arise after learning trials
- However
  - You lose information about the "true" first period
  - People infer uncontrolled things from the learning trials
  - Subjects may send (costless) signals
- Makes most sense if the institutions are really difficult (e.g., in a continuous double auction)
- Maybe it is not necessary to play a full game (e.g., just the complicated part) and maybe it is not necessary to display all information about others' actions
- In any case: if learning trials, then in all treatments

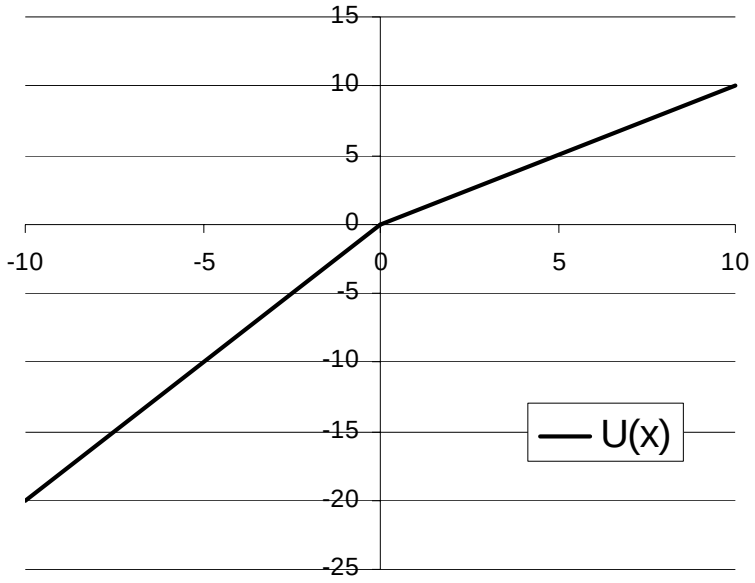
Implementing risky decisions

- Most important: use credible chance moves
  - If many chance moves are necessary:
    - Random device at the computer
  - If only few chance moves and if credibility is easily doubted (e.g., imposing infinitely repeated games)
    - Throwing dice may be better
    - You may have people do that on their own (see FFF, 2002)
- Risk preferences
  - May influence theoretical implications.
  - Possible to control risk preferences with binary lottery method
    - In the experiment Ss earn points
    - Payments depend on winning a lottery
    - Probability of winning the higher the more points a subject has
  - Risk preferences can be neglected under the expected utility paradigm (Rabin, Risk Aversion and Expected-Utility Theory: A Calibration Theorem," Econometrica 68(5), 1281-1292, September 2000)

Losses

- Interesting to study losses
  - o Asymmetry between gains and losses (prospect theory, e.g., Kaheman/Tversky 1992)
- Relative to a given reference standard, people dislike a loss more than they like a gain of equal size
  - o Loss aversion is behaviorally relevant (e.g., Tversky and Kahneman 1992, application Falk/Fehr 2002 in tournaments).

Example:



- Sometimes losses may occur given the nature of the experiment (e.g., in gift-exchange experiments)
- Losses must be credible
  - Instructions: if you make losses these have to be covered
  - Show up fee
  - Ss have to pay in order to keep on going
  - Ss has to stop

Elicitation of beliefs

- Example: Prisoner's dilemma
  - Before Ss make their decisions both players are asked, what they think the other player will do, to cooperate or to defect?
- Beliefs can be very informative to understand their motivation
- Beliefs are of particular importance to check the rationality of decisions
  - Example: Guessing game
- Problems
  - Experimenter-Demand–Effect (you may make people think about stuff they would not have thought about)
    - Directs focus on particular problems, e.g., guessing game!
  - Desire to be consistent: people state beliefs to "match" their actions
  - People have a desire to "justify" actions: someone defects and states the other person will defect also

Pay beliefs?

- Pros
  - Subjects have an incentive to state correct beliefs
- Cons
  - Is costly and – given a budget – goes at the cost of incentives in the decision part
  - Subjects have no incentive to state wrong beliefs anyway
  - Distribution vs. mean
  - Sometimes complicated to explain (e.g., payment dependent on distance measure between true outcome and expected outcome)
  - Can pollute incentives in the experiment if people "hedge" decisions, e.g., in coordination games, see next slide

Beliefs are paid: 2 points for correct belief

Hedge strategy of player 1, e.g.: “I believe 2 plays left,” but actually plays down. In this case 1 earns at least 2 points

Player 2

Player 1	left	right
up	2,2	0,0
down	0,0	2,2

Paper and pencil vs. computerized experiments

- Advantages paper and pencil experiments
  - Flexibility (quickly develop new treatments)
  - Relatively low start up costs
  - Natural environment
    - Not a lab with computers etc. but a classroom
    - Procedures more visible and credible
      - E.g., throw dice in front of people instead of random device
      - Matching of people is easy to recognize (walking around of experimenter)
- Advantages of computerized experiments
  - Better control
    - no communication among subjects
    - less interaction with experimenter
  - Running of experiment much simpler (e.g., markets)
  - Fewer mistakes
  - Automatic data collection

Deception

- Never cheat on subjects, even though it is tempting from a scientific point of view.
- Why?
  - First, there is a moral code among economic experimentalists not to do it.
  - You will never publish a paper and people won‘t like your research.
  - You will lose your reputation towards your subjects: If you lie once they will never believe you in the future. This blurs all incentives.

  - http://www.apa.org/ethics/code2002.html

Hypotheses: Standard Hypothesis

- In almost all experiments you want to have a (set of) prediction(s)/hypotheses
- Traditional assumptions in game theory
  - Rationality
  - Selfishness = money maximizing
  - Both are „common knowledge"
- Determine equilibria
  - Often simple and unique prediction
  - Describes behavior often not very well
- Use the standard prediction as a benchmark

Alternative Hypotheses

- There are many good reasons to question the standard prediction, examples:
  - Bounded rationality
    - Cognitive limits
    - Rules of thumb
    - Heuristics
    - Imitation
  - Social motives
    - Altruism
    - Fairness (reciprocity, inequity aversion)
    - Status preferences
    - Preferences for efficiency
  - Emotions
    - Anger
    - Joy
    - Arousal
  - Etc.

Deriving alternative predictions

- Observations from every day life, intuition
- Previous experimental results (economics, psychology)
- Game theoretic analysis under alternative assumptions
  - Prospect Theory (risk behavior, loss aversion)
  - Fairness theories
  - Statistical game theory, quantal response (errors depend on cost of error)
  - Visceral factor perspective, emotions

Writing instructions

- Simple language
  - Simple, short and unambiguous sentences
  - Use redundancies if issues are complicated
  - Consistent/uniform descriptions and framing
  - Avoid suggestive terms
    - Punishment: negative points
    - Defect: contribute nothing
- Neutral framing pros and cons
  - Concrete framing (goods market, labor market)
    - Easy to understand
    - Problem (?): Associations from real life
  - Abstract framing
    - Avoids every day associations (does it really?)
    - Harder to understand the rules of the game
    - No control about what subjects really think

Writing instructions

- Complete description of the rules of the game
  - Sequence of decisions
  - Interaction
  - Payoff consequences
- Different ways to explain the payoff function
  - Formula
  - Verbal explanation
  - Table
  - Figure
- Control questions
  - Check understanding
  - Knowing who is done with the instructions
  - One should not be suggestive with his examples

Recruiting subjects

- Students
  - o + easy access
  - o + relatively low opportunity costs (low costs of conducting experiments)
  - o + quick learning…
  - o +/- not much experience with the object of interest
  - o + analytical skills, quick understanding of instructions
  - o - selection effect (not representative)
- Non-students
  - o +/- experienced subjects (know institutions etc.)
  - o - Larger variance in learning and understanding
  - o - Potentially high opportunity costs (salience?)

Recruiting: What do you tell people when you invite them?

- It is not
  - a medical experiment
  - Intelligence test
  - Marketing research
- It is an economic experiment
  - Study human behavior
  - Important for understanding economic problems
- Why should you take part?
  - You can earn money (do not mention concrete amounts of money: this creates expectations and may pollute behavior "if I do not earn at least x, I must have been wrong")
  - Learn about an interesting method in the social sciences

Running experiment: Preparation

- Z-tree
- Start experimenter computer
- Start z-Tree at experimenter computer
- Start subject computers (z-Leaf starten)
- Open treatments
  - Welcome (welcome.ztt)
  - Actual Treatments (experiment.ztt)
  - Questionnaires (questionnaire.zqt)
- Start Chatter
- Distribute instructions, shut down screens
- Distribute ID-cards (for random allocation of roles and cubicles in the lab, include additional xxx cards)
- Check money…

Running experiment: Conducting

- Welcome Ss and distribute ID cards
- Communication is forbidden
- Pay show up fee for extra Ss
- Ss read instructions and solve control questions
- Start the Welcome-Treatments
- Check control questions, start screens
- Summarizing instructions answer questions (in privacy)
- Start first treatment
- New instructions, summarize, start new treatment etc.
- Announce questionnaires (sometimes paid)
- Subjects fill in questionnaire (in this time the payment-file is generated)
- Print receipts
- Pay Ss (and listen to what they have to say)

Questionnaires

- Test understanding of experiment
- Infer something about motives
- Credibility of experiment
- Control
  - How many Ss did know each other?
  - Socio economic questions (sex, age, money, city, subject of study etc. etc.)
- Psychological questionnaires (used to construct particular types)

Paying subjects

- Use hypothetical currencies in the experiment
- Show up fee
  - To compensate extra subjects
  - To cover losses in the experiment
- Goal: total payments should cover opportunity costs (typical job)
- Ensure Anonymity when paying

Data analysis: general

- Collect data in systematic way (one master file, which remains unchanged)
- Descriptive statistics
  - Tables
    - Title, clear variable names, round numbers
  - Figures
    - As simple as possible, title, label axes, complete legend, good contrasts, not too many
    - Figures often understood and remembered best
- Test Hypotheses
  - Frequently used:
    - Means (t-Test)
    - Wilcoxon Signed Rank Test
    - Wilcoxon-Mann-Whitney Test
    - Kolmogorov-Smirnov Two Sample Test

Descriptive statistics

- Mean/Median/Variance
  - Subjects
  - Groups of subjects
  - Time series
  - Sessions
  - Treatments
- Distributions (histograms)

Group Projects

- Students can form groups of exactly **5** on their own.
- Must let me know by sending an e-mail with names of 5 group members, by 8:00 Tuesday morning.
- The remaining students will be randomly assigned to groups of 5.
- I will announce these groups in the next lecture.

- Grading: **pass / fail.**
- A group must produce a reasonable, passing project in order to take the final exam.

Group Projects

- Step 1: Choose the hypothesis you want to test.
  - Is there a variation on one of the studies we have discussed in class that you would like to try?
  - Some other economic question that interests you?

- Step 2: Read some of the literature on this topic.
  - If there is an experimental literature
    - How have other experiments been designed?
    - Has someone else answered your question?
    - If you are using game X, cite the first paper to introduce this game.
  - If there is a theoretical or field literature, what are the main findings? Why is it a good idea to do an experiment?

Group Projects

- Step 3: Design an experiment to test your hypothesis.
  - Make it as simple as possible (but no simpler).
  - Use friends/family/strangers as subjects.
  - Try to collect enough data to do some basic statistics
  - Ideally you should pay subjects
    - Be creative: e.g., randomly select one subject to be paid.
    - Payments can be very low (20 cents).

- Step 4: Run the experiment.

Group Projects

- Step 5: Analyze the data.
  - Report means and other descriptive statistics.
  - Test for treatment differences using simple non-parametric tests.
  - Run regressions if appropriate.

Group Projects

- Step 6: Write up the results.
  - **8 page maximum!!**
    - This includes tables and figures, appendices, and references.
    - I stop reading after 8 pages…
    - Later on I will return to the issue of how to write a nice experimental paper.

Groups

1	Simola	Mikael	5	Tilmann	Dreben
1	Toikka	Samuli	5	Bierbrauer	Christoph
1	Markkula	Valtteri	5	Krugljakov	Nikita
1	Schlicht	Anne	5	Karczewski	Philipp
1	Brettschneider	Nils	5	Droschel	Claudia
2	Berger	Johannes	6	Zeghers	Dainis
2	Aussenhofer	Ruth	6	Hebebrand	Jan
2	Schneider	Sonja	6	Honekamp	Ivonne
2	Pinzon	Javier	6	Kubny	Julia
2	Becker	Anke	6	Klinger	Jan

3	Ho	Shin-May
3	Jungherr	Joachim
3	Milarepa	Lee
3	Sallge	Martin
3	Becker	Alexander
4	Bohlmann	Elisabeth
4	Leuermann	Andrea
4	Stauf	Julia
4	Ollig	Julia

7	Engels	Sebastian
7	Arp	Joachim
7	Weisser	Johannes
7	Suebsing	Kanyawan
7	Wostrack	Dominic
8	Stender	Dirk
8	Evers	Bastian
8	Datkiewicz	Sascha
8	Rhoder	maximilian
8	Wilmink	Carsten

9	Daniels	Joerg
9	Sumnikov	Serhiy
9	Deloy	Patrick
9	Stuhler	Jan
9	Seithe	Mirko
10	Weigand	Moritz
10	Kolle	Felix
10	Schauihsland	Alexandra
10	Becker	Gavin
10	Sherwood	martin
11	Argudo	Sara
11	Hild	Robert
11	v. Heussinger	Georg
11	Aretz	Bodo
11	Quack	Daniel