

Experimental Methods in Social Sciences (in particular economics)

Dirk Engelmann

Chances and Challenges of Different Types of Experiments

Laboratory Experiments – Motivation

- Laboratory setting has a number of key advantages
 - high degree of control on conditions
 - no access to other information, internet, etc.
 - secure data gathering
 - computerization
 - highly interactive experiments possible
 - double auctions
 - multi-stage punishment

Laboratory Experiments – Challenges

- Laboratory setting is special in many respects
 - people know they participate in an experiment
 - possible cause for “experimenter demand effect”
 - participants are volunteers
 - selection effects
 - participants are typically students
 - not representative of general population or the group we are interested in (unless we are interested in student behavior)
 - situation is artificial
 - participants may activate very different routines than in their normal life
 - participants have typically no experience with the task
 - but often we may be interested in behavior in tasks people are experienced with
 - stakes are often small
 - hence other concerns, e.g., for following norms may take a large role
 - Levitt and List (JEP 2007) address this in context of social preferences

Class Room Experiments – Motivation

- Running experiments during a (large) lecture can be preferable to an experiment in a computerized laboratory
 - typically a (very) large share participate
 - selection is much less of an issue
 - opportunity costs are low, so it is comparably cheap for large number of observations
 - particularly useful for very short experiments
 - word about experiment cannot spread around between sessions
 - at least if only one lecture is used
- Small class room experiments also useful for explorative studies
 - pilot session
 - testing known results with a selective group (e.g., PhD students)
- “Poor man’s lab”

Class Room Experiments – Challenges and Limitations

- Ensuring control more challenging
 - preventing communication and access to other material
- Pen and paper data collection tedious and possible source for errors
- At most very little interaction possible
- Latter two problems can be resolved by running experiments via mobile apps
- Replicability reduced

Field Experiments – Motivation

- Field experiments do not invite participants, but study their behavior in the environment they usually are in
- Field experiments thus collect information more akin to classical empirical work but preserve the exogenous variation of variables of interest
- Field experiments reduce several problematic aspects of lab experiments
 - possibly more representative (or more relevant) subject pool than student participants
 - no selection of participants into experiment
 - real goods as opposed to fictitious goods with induced valuations
 - tasks and rules subjects are familiar with
 - possibly higher stakes
 - e.g., if we investigate the impact of different information conditions on take-up rates of pension funds
 - natural environment of subjects
 - including them not being aware that they participate in an experiment

Field Experiments – A Taxonomy

- Following Harrison and List (2004), we can classify (field) experiments according to the degree of moving from classical lab study to field
 - conventional lab experiment:
 - standard subject pool (i.e., university student volunteers), abstract framing and imposed set of rules
 - artefactual field experiment:
 - as conventional lab experiment, but with nonstandard subject pool
 - example: standard experiments with more general population, e.g., newspaper readers
 - framed field experiment:
 - as artefactual field experiment, but with field context in good, task or information
 - example: auctions with goods people have experience with, such as List's (2003) experiments auctioning sportscards to sportscard traders

Field Experiments – A Taxonomy

- Classification continued
 - natural field experiment:
 - like framed field experiment (with field context in all aspects), but in natural environment for the task and where subjects do not know they are in an experiment
 - example: manipulating a natural market without informing participants that this is an experiment, e.g., experimenter acts as bidder in market, or experimenter varies conditions in stores, or experimenter calls for donations with variations, or varies marketing strategies
- Classification naturally not always straightforward
 - what about experiments with field context but student subject pool?
- Approach by Harrison and List has thus been criticized by people with experience in field experiment
- Nevertheless, the suggested taxonomy and the debate help by clarifying that there is not “the lab” and “the field”

Field Experiments – Challenges and Limitations

- Field experiments may yield less control than laboratory experiments
 - interaction in natural environment may lead subjects to consult with others
 - some newspaper readers ran experiments in order to enter their choice
- Sometimes we may not know whether the treatment variable has been noticed by the participants
 - hence we might infer that an intervention is ineffective, but possibly we just did not make it transparent
- While selection of participants into experiment is not an issue, randomization of participants into treatments is possibly complicated
 - we typically cannot run several treatments in the same place at the same time
 - but in the field, people who show up at the same place at different times may be very different

Field Experiments – Challenges and Limitations

- Harder to gather data on demographics and other confounds
- We cannot check whether people understood the decision problem they faced
- We cannot ask them about their motivation (sometimes we can, even in the field)
- Organization is often complicated
- Replicability reduced because many conditions cannot be easily reproduced

Lab-in-the-Field Experiments – Motivation

- Lab-in-the-field refers to situations, where a simplified laboratory experiment is reproduced in participants' regular environment
- Typically in context of development economics
- Research questions often center around relationship of measures often elicited in laboratory experiments (e.g., trust) and behavior outside the laboratory (e.g., contributions to local community projects) with aim at developing interventions

Lab-in-the-Field Experiments – Challenges and Limitations

- Lab-in-the-field experiments are similar to small class room experiments and thus share some of their problems
 - pen and paper data collection tedious and possible source for errors
 - complicated interaction not possible (though simple forms work in small groups)
- Specific problems due to subject pool and setting
 - participants often have limited cognitive, language, and numeracy skills
 - reducing sample to most skilled people typically contrary to research objective
 - hence only relatively simple experiments can be run
 - and specific care has to be taken with respect to instructions and procedures
 - language and organizational issues often require collaboration with local NGO or similar
 - interaction with these people may produce strong demand effects
 - comparisons and robustness checks across locations confounded by differences in local partners

Lab-in-the-Field Experiments – Challenges and Limitations

- Above issues lead to reduced replicability
- Lab-in-the-field experiments serve specific purpose, though, and for that are often fine

Internet Experiments – Motivation

- Already in early 2000's researchers ran experiment on the Internet
- Key advantages relate to easy access for participants
 - large sample sizes can be employed
 - very short (or very long) experiments can be efficiently done
 - relatively cheap, in particular for short experiments
 - potentially broader subject pool
- Experiments on MTurk now popular
 - allegedly about 30% of all MTurk tasks are now social science experiments

Internet Experiments – Challenges

- Lack of control
 - not clear what information sources participants access
 - not clear whether they decide alone
 - not clear whether background information on education etc. is reliable
- Potentially very biased subject pool
 - apparently not a major issue on MTurk
- Opportunities for interactive experiments limited
 - several participants have to be active at the same time
 - on MTurk that seems to work

Survey Experiments – Motivation

- Laboratory experiments are partly problematic because students are not representative
 - that problem is often overrated
 - for many questions we do not want representative quantitative estimates
 - but we often want to understand the impact of certain variables
 - and these do not vary enough amongst students
 - age
 - cognitive abilities
 - political opinions
- Same applies to class room experiments but also often to field experiments (sports card dealers)
- Including experiments into (representative) surveys addresses this problem
- Also large sample size possible
- In panel, experimental choices can be linked to other survey responses

Survey Experiments – Challenges

- Only relatively simple experiments can be done
 - no clarifying questions can be asked
 - participants with low cognitive abilities
 - very small time allocation
- Instructions also have to be very clear
- Typically expensive
- Typically truly interactive experiments not possible
 - typically participants not logged on simultaneously
 - panel survey not designed for data transmission
 - applies in particular to panels with data collection through human interviewer (GSOEP)

Levitt, List (2007) Model

- Utility function for i

$$U_i(a, v, n, s) = M_i(a, v, n, s) + W_i(a, v)$$

- W_i wealth payoff, depends on
 - action a and stakes v
- M_i moral payoff, depends on
 - action a
 - financial externalities on others, which depend on v
 - social norms n (e.g., if a is illegal)
 - scrutiny s

Levitt, List (2007) Main Arguments

- Levitt-List discuss following issues that can limit generalizability
 - moral and ethical concerns
 - moral component important for other determinants to have influence
 - scrutiny by others
 - in lab obvious that behavior will be observed and recorded
 - context
 - framing has substantial effect in some experiments
 - suggests that activating different perceived contexts has effect
 - subjects may not play the game the experimenter intends
 - selection
 - students may differ from other people
 - volunteers to experiments may be more pro-social
 - stakes
 - evidence mixed but several studies find more selfish behavior with higher stakes
 - limited choice set
 - outside lab, one can sometimes avoid moral cost by avoiding interaction, exit options often taken in lab if available
 - lab experiments typically cover short time span
 - emotional aspects might matter more in short than long run

Levitt, List (2007) Discussion

- Levitt-List raise a number of important concerns that should be kept in mind when interpreting (lab) experiments
- But as a critique on the usefulness of lab experiments some of their concerns miss the point (see also Camerer, 2011)
 - purpose of lab experiments is not to estimate parameters
 - instead special usefulness of lab is for theory testing, in particular for comparative statics
 - interestingly, implicitly (and to some degree explicitly) Levitt-List acknowledge this
 - nearly all their supporting evidence comes from lab experiments
 - for questions such as effects of stake size or earned vs. windfall profits, lab is ideal test bed

Levitt, List (2007) Discussion

- Small stakes critique is not as convincing as it sounds
 - lots of daily decisions are small-stake, but the aggregate effect can be substantial
 - many people paying 10 cents more for Fairtrade chocolate makes huge difference for producer
 - in particular combination of small-stakes critique and lack of experience critique does not make much sense
 - for how many large-stakes decisions do you have substantial experience?
 - key results are often robust to stake size, in particular ultimatum game
- Direct tests of generalizability often finds support
 - if there are differences, non-students often show stronger social preferences than students
 - Hannan, Kagel, Moser (2002) find more reciprocity by MBA students than undergrad students
 - experiments with representative Danish sample via iLEE typically finds strong social preferences
 - direct tests of selection provide little evidence that it matters (Cleave, Nikiforakis, Slonim 2013)

Levitt, List (2007) Discussion

- Scrutiny present in lab
 - but outside lab, behavior is rarely completely anonymous
 - in contrast experiments nearly always single-blind and often double-blind
 - hence lab experiments might underestimate pro-social behavior, though not pure social preferences
 - double-blindness matters a lot in dictator game, but little in most other games

Generalizability of Laboratory and Field Experiments

- Not always clear that moving away from standard lab experiments increases generalizability (or is even the point of the change)
 - if we are interested in behavior of young children, we should obviously have young children as subjects
 - but this makes the experiment only more “natural” for the issue of behavior of children, but not more general
- Using specific goods yields results more informative for this type of good
 - it might also yield more generalizable results, because subjects act with respect to something familiar
 - but results might also be down to something specific about the good (e.g. measuring discount rates with ice-cream)
 - results may become more accurate for a specific domain, but also more specific

Use of “Natural” Tasks

- How people behave in experiments that correspond to nothing they will ever experience is not important
- Hence one could restrict experiments to subjects who have experience with the task
- Experienced agents can have developed heuristics that
 - serve them well for relevant tasks
 - but that may not travel to abstract versions of the task
- But sometimes we are interested in how people behave in tasks they rarely perform
 - e.g. choosing education, deciding upon pension plans

Control and the Purpose of Lab Experiments

- Increased control (sterility) of lab experiments is suitable for theory testing
 - we can vary exactly what theory predicts to matter
- But lab experiments are not ideal for parameter estimating and hence for predicting behavior in the field due to selection and artificiality
 - but even if quantitative effects are likely to differ in the field, qualitative effects can in general be expected to be more general
- Combination of lab and field useful
 - results from field can suggest crucial aspects left out in the lab, which can then be included in the lab
 - explanations for results in field can be tested with increased control in lab

The Purpose of Lab and Field Experiments

- Plott and Zeiler (2007, footnote 3) "We caution the reader to be aware of common misunderstandings regarding the purposes of experimental research that lead to unhelpful views about the usefulness of laboratory experiments. When choosing an experimental environment (e.g. lab versus field experiment) the purpose of the experiment becomes important. Two different purposes, parameter estimation and theory testing, call for different environments. If the question posed concerns measurement of a parameter, then the field could be the appropriate environment simply because the field might be the only environment in which the parameter confidently resides. For example, the field is appropriate if the intent is to measure the elasticity of market demand for a specified commodity or damages owed due to monopoly power or price-fixing schemes. A second and much different purpose of experiments, theory testing, must be recognized as a substantially different experimental activity. Relative to parameter measurement, theory testing is more naturally directed towards laboratory environments, especially in the case of very general theories, such as endowment effect theory. Theory testing requires that predictions of competing theories be clearly separated so that the theory that best accounts for observed phenomena can be identified convincingly. Thus, depending on the theories, theory testing might require implementation of controls and replications under different sets of controls that are unimplementable in the field or have little resemblance to any field environment."

Framing Effects

- Description and presentation of decision situation (“frame”) can impact behavior
 - famous example: presenting prisoner’s dilemma as “community game” or “wall-street game” affects cooperation rate
- Laboratory experiments in economics often employ an abstract frame (“Choice A” and “Choice B” rather than “Trust” and “Do not trust”)
- Fundamental reason for abstract nature of lab experiments is that this shall guarantee control
 - experimenter controls all actions that subjects can take
 - in theory the experimenter also knows what the subjects know about the situation, because everything relevant is in the instructions
 - subjects do not bring heuristics from outside to lab

Framing Effects

- That abstraction brings control can be an illusion
 - if situation is presented in abstract way, subjects may well bring their own interpretation and we do not know which
 - context might also reduce confusion
 - so giving context might increase control
 - but again, subjects may then add their own associations to the frame
 - say we frame something as interaction between buyer and seller, then subjects may think of their baker or their car dealer and we do not know which
- Useful approach: do both, sessions with abstract frame and those with context frame
- Sometimes, abstract frame is artificial and not helpful
 - typically, whenever I have prices in an experiment, I call them prices

Experimenter Demand Effect

- Experimenter demand refers to the phenomenon that participants might do what they believe the experimenter “wants” them to do
- There are two different aspects to this (Zizzo, 2010)
 - cognitive experimenter demand effect
 - experimenter typically does not define which behavior is appropriate
 - framing, anchoring on information (examples in instructions), choice set, etc., may provide cues what the experimenter considers appropriate
 - social experimenter demand effect
 - in addition to cognitive aspect perceived social pressure
 - perception that experimenter makes some form of judgment on behavior

Experimenter Demand Effect

- Whether experimenter demand effect is an issue depends on how direction of effect relates to direction of hypothesis
 - no problem if demand effect orthogonal to hypothesis
 - e.g., my hypothesis addresses how various treatment conditions affect decision speed but the participants want to show me how altruistic they are
 - e.g., demand effect is clearly identical in all treatments
 - no serious problem if they go in opposite directions, because demand effect might only weaken treatment effect
 - e.g., my hypothesis is that a certain intervention makes people selfish but experimenter demand could make them generous, because they want to impress me
 - problematic if they go in same direction
 - e.g., my hypothesis is that a certain intervention makes people altruistic and some cue triggers the perception that I am interested in altruism
 - e.g., my hypothesis is that giving a hint will improve understanding of a game, but participants may perceive social pressure to follow the hint

Experimenter Demand Effect

- Experimenter demand is also unproblematic if it is in line with the behavior experimenter wants to incentivize
 - if I want to study performance in a general knowledge quiz and pay participants for good performance, then if they also want to show me how smart they are, intrinsic motivation just adds to extrinsic motivation
 - so this is only problematic if it differs systematically across treatments
- If social pressure is the object of the study, it is also not a problem
 - e.g., conformity experiment by Ash (1956)

Measures to Deal with Experimenter Demand Effects

- Having unknown research assistants rather than professor running sessions reduces perceived social pressure
- Not informing research assistants about objectives of study reduces risk that they unintentionally signal appropriate action
 - but then they might guess what the objective is and thus give (unconscious) hints
 - hence researcher loses control about possible direction of demand effect
- Double blindness reduces social pressure
 - but only feasible in relatively simple experiments
- Enlarge choice set to remove focus on particular action
 - e.g., allow for taking in dictator game (Bardsley, 2008)
 - e.g., add alternative activity in experiments on bubbles

Measures to Deal with Experimenter Demand Effects

- (Non-deceptive) obfuscation of objective
 - neutral frame
 - alternative frame (sellers and buyers rather than workers and employers)
 - filler tasks
 - adding alternative cues if cues are necessary for understanding, e.g., examples with different choices if examples are needed
- Between-subject designs typically reduce demand effects because seeing only one condition makes guessing the research hypothesis much more difficult

Literature

- Ash, Solomon E. (1956) "Studies of Independence and Conformity: I. A Minority of One Against a Unanimous Majority", *Psychology Monographs* 70(9), 1–70.
- Bardsley, Nicholas (2008) "Dictator Game Giving: Altruism or Artefact?", *Experimental Economics* 11(1), 122–133.
- Camerer, Colin F. (2011) "The Promise and Success of Lab-Field Generalizability in Experimental Economics: A Critical Reply to Levitt and List", available at SSRN 1977749.
- Cleave, Blair L., Nikiforakis, Nikos, and Slonim, Robert (2013) "Is there Selection Bias in Laboratory Experiments? The Case of Social and Risk Preferences", *Experimental Economics* 16(3), 372–382.
- Hannan, Lynn R., Kagel, John H., and Moser, Donald V. (2002) "Partial Gift Exchange in an Experimental Labor Market: Impact of Subject Population Differences, Productivity Differences, and Effort Requests on Behavior", *Journal of Labor Economics* 20(4), 923–951.

Literature

- Harrison Glenn W. and List, John A. (2004) "Field Experiments", *Journal of Economic Literature* 42, 1009–1055.
- Levitt, Steven D. and List, John A. (2007) "What Do Laboratory Experiments Measuring Social Preferences Reveal About the Real World?", *Journal of Economic Perspectives* 21(2), 153–174.
- List, John A. (2003) "Does Market Experience Eliminate Market Anomalies?", *Quarterly Journal of Economics* 118(1), 41–71.
- Plott, Charles R. and Zeiler, Kathryn (2007) "Exchange Asymmetries Incorrectly Interpreted as Evidence of Endowment Effect Theory and Prospect Theory?", *American Economic Review* 97(4), 1149–1466.
- Zizzo, Daniel J. (2010) "Experimenter Demand Effects in Economic Experiments", *Experimental Economics* (13(1), 75–98.