The New Palgrave Dictionary of Economics Online

field experiments

John A. List and David Reiley
From The New Palgrave Dictionary of Economics, Second Edition, 2008
Edited by Steven N. Durlauf and Lawrence E. Blume

Abstract

Field experiments have grown significantly in prominence since the 1990s. In this article, we provide a summary of the major types of field experiments, explore their uses, and describe a few examples. We show how field experiments can be used for both positive and normative purposes within economics. We also discuss more generally why data collection is useful in science, and more narrowly discuss the question of generalizability. In this regard, we envision field experiments playing a classic role in helping investigators learn about the behavioural principles that are shared across different domains.

Keywords

charitable giving; field experiments; generalizability; laboratory experiments; matching funds; testing; uniform-price auctions; Vickrey auctions

Article

Field experiments occupy an important middle ground between laboratory experiments and naturally occurring field data. The underlying idea behind most field experiments is to make use of randomization in an environment that captures important characteristics of the real world. Distinct from traditional empirical economics, field experiments provide an advantage by permitting the researcher to create exogenous variation in the variables of interest, allowing us to establish causality rather than mere correlation. In relation to a laboratory experiment, a field experiment potentially gives up some of the control that a laboratory experimenter may have over her environment in exchange for increased realism.

The distinction between the laboratory and the field is much more important in the social sciences and the life sciences than it is in the physical sciences. In physics, for example, it appears that every hydrogen atom behaves exactly alike. Thus, when astronomers find hydrogen's signature wavelengths of light coming from the Andromeda Galaxy, they use this information to infer the quantity of hydrogen present there. By contrast, living creatures are much more complex than atoms and molecules, and they correspondingly behave much more heterogeneously. Despite the use of 'representative consumer' models, we know that not all consumers purchase the same bundle of goods when they face the same prices. With complex, heterogeneous behaviour, it is important to sample populations drawn from many different domains – both in the laboratory and in the field. This permits stronger inference, and one can also provide an important test of generalizability, testing whether laboratory results continue to hold in the chosen field environment.

We find an apt analogy in the study of pharmaceuticals, where randomized experiments scientifically evaluate new drugs to treat human diseases. Laboratory experiments evaluate whether drugs have desirable biochemical effects on tissues and proteins *in vitro*. If a drug appears promising, it is next tested *in vivo* on several species of animals, to see whether it is absorbed by the relevant tissues, whether it produces the desired effects on the body, and whether it produces undesirable side effects. If it remains with significant promise after those tests, it is then tested in human clinical trials to explore efficacy and measure any side effects.

Even after being tested thoroughly in human clinical trials and approved by regulators, a drug may

sometimes reveal new information in large-scale use. For example, *effectiveness* may be different from the *efficacy* measured in clinical trials: if a drug must be taken frequently, for example, patients may not remember to take it as often as they are supposed to or as often as they did in closely supervised clinical trials. Furthermore, rare side effects may show up when the drug is finally exposed to a large population.

Much like this stylized example, in economics there are a number of reasons why insights gained in one environment might not perfectly map to another. Field experiments can lend insights into this question (see also Bohm, 1972; Harrison and List, 2004; Levitt and List, 2007; List, 2007). First, different types of subjects might behave differently; university students in the laboratory might not exhibit the same behaviour as financial traders or shopkeepers. In particular, the people who undertake a given economic activity have selected into that activity and market forces might have changed the composition of players as well; you might expect regular bidders to have more skill and interest in auctions than a randomly selected laboratory subject, for example.

A second reason why a field experiment might differ from a laboratory experiment is that the laboratory environment might not be fully representative of the field environment. For example, a typical donor asked to give money to charity might behave quite differently if asked to participate by choosing how much money to contribute to the public fund in a public-goods game (List, 2007). The charitable-giving context could provide familiar cognitive cues that make the task easier than an unfamiliar laboratory task. Even the mere fact of knowing that one's behaviour is being monitored, recorded, and subsequently scrutinized might alter choices (Orne, 1962).

Perhaps most important is the fact that any theory is an approximation of reality. In the laboratory, experimenters usually impose all the structural modelling assumptions of a theory (induced preferences, trading institutions, order of moves in a game) and examine whether subjects behave as predicted by the model. In a field experiment, one accepts the actual preferences and institutions used in the real world, jointly testing both the structural assumptions (such as the nature of values for a good) and the behavioural assumptions (such as Nash equilibrium).

For example, Vickrey (1961) assumes that in an auction there is a fixed, known number of bidders who have valuations for the good drawn independently from the same (known) probability distribution. He uses these assumptions, along with the assumption of a risk-neutral Nash equilibrium, to derive the 'revenue equivalence' result: that Dutch, English, first-price, and second-price auctions all yield the same expected revenue. However, in the real world the number of bidders might actually vary with the good or the auction rules, and the bidders might not know the probability distribution of values. These exceptions do not mean that the model should be abandoned as 'wrong'; it might well still have predictive power if it is a reasonable approximation to the truth. In a field experiment (such as Lucking-Reiley, 1999, for this example), we approach the real world; we do not take the structural assumptions of a theory for granted.

Such an example raises the natural question related to the actual difference between laboratory and field experiments. Harrison and List (2004) propose six factors that can be used to determine the field context of an experiment: the nature of the subject pool, the nature of the information that the subjects bring to the task, the nature of the commodity, the nature of the task or trading rules applied, the nature of the stakes, and the environment in which the subjects operate. Using these factors, they discuss a broad classification scheme that helps to organize one's thoughts about the factors that might be important when moving from the laboratory to the field.

A first useful departure from laboratory experiments using student subjects is simply to use 'non-standard' subjects, or experimental participants from the market of interest. Harrison and List (2004) adopt the term 'artefactual' field experiment to denote such studies. While one might argue that such studies are not 'field' in any way, for consistency of discussion we denote such experiments as artefactual field experiments for the remainder of this article, since they do depart in a potentially important manner from typical laboratory studies. This type of controlled experiment represents a useful type of exploration beyond traditional laboratory studies.

Moving closer to how naturally occurring data are generated, Harrison and List (2004) denote a 'framed field experiment' as the same as an artefactual field experiment but with field context in the

commodity, task, stakes, or information set of the subjects. This type of experiment is important in the sense that a myriad of factors might influence behaviour, and by progressing slowly towards the environment of ultimate interest one can learn about whether, and to what extent, such factors influence behaviour in a case-by-case basis.

Finally, a 'natural field experiment' is the same as a framed field experiment but where the environment is one where the subjects naturally undertake these tasks and where the subjects do not know that they are participants in an experiment. Such an exercise represents an approach that combines the most attractive elements of the laboratory and naturally occurring data – randomization and realism. In this sense, comparing behaviour across natural and framed field experiments permits crisp insights into whether the experimental proclamation, in and of itself, influences behaviour.

Several examples of each of these types of field experiments are included in List (2006). Importantly for our purposes, each of these field experimental types represents a distinct manner in which to generate data. As List (2006) illustrates, these field experiment types fill an important hole between laboratory experiments and empirical exercises that make use of naturally occurring data. Yet an infrequently discussed question is: why do we bother to collect data in economics, or in any science?

First, we use data to collect enough facts to help construct a theory. Several prominent broader examples illustrate this point. After observing the anatomical and behavioural similarities of reptiles, one may theorize that reptiles are more closely related to each other than they are to mammals on the evolutionary tree. Watson and Crick used data from Rosalind Franklin's X-ray diffraction experiment to construct a theory of the chemical structure of DNA. Careful observations of the motions of the planets in the sky led Kepler to theorize that planets (including Earth) all travel in elliptical orbits around the Sun, and Newton to theorize the inverse-square law of gravitation. After observing with a powerful telescope that the fuzzy patches called 'spiral nebulae' are really made up of many stars, one may theorize that our solar system is itself part of its own galaxy, and the spiral nebulae are external to our Milky Way galaxy. Robert Boyle experimented with different pressures using his vacuum pump in order to infer the inverse relationship between the pressure and the volume of a gas. Rutherford's experiments of shooting charged particles at a piece of gold foil led him to theorize that atoms have massive, positively charged nuclei.

Second, we use data to test theories' predictions. Galileo experimented with balls rolling down inclined planes in order to test his theory that all objects have the same rate of acceleration due to gravity. Pasteur rejected the theory of spontaneous generation with an experiment that showed that microorganisms grow in boiled nutrient broth when exposed to the air, but not when exposed to carefully filtered air. Arthur Eddington measured the bending of starlight by the sun during an eclipse in order to test Einstein's theory of general relativity.

Third, we use data to make measurements of key parameters. On the assumption that the electron is the smallest unit of electric charge, Robert Millikan experimented with tiny, falling droplets of oil to measure the charge of the electron. On the assumption that radioactive carbon-14 decays at a constant rate, archaeologists have been able to provide dates for various ancient artifacts. Similarly, scientists have assumed theory to be true and designed careful measurements of many other parameters, such as the speed of light, the gravitational constant, and various atomic masses.

Field experiments can be a useful tool for each of these purposes. For example, Anderson and Simester (2003) collect facts useful for constructing a theory about consumer reactions to nine-dollar endings on prices. They explore the effects of different price endings by conducting a natural field experiment with a retail catalogue merchant. Randomly selected customers receive one of three catalogue versions that show different prices for the same product. Systematically changing a product's price varies the presence or absence of a nine-dollar price ending. For example, a cotton dress may be offered to all consumers, but at prices of 34, 39, and 44 dollars, respectively, in each catalogue version. They find a positive effect of a nine-dollar price on quantity demanded, large enough that a price of 39 dollars actually produced higher quantities than a price of 34 dollars. Their results reject the theory that consumers turn a price of 34 dollars into 30 dollars by either truncation or rounding. This finding provides empirical evidence on an interesting topic and demonstrates the need for a better theory of how

consumers process price endings.

List and Lucking-Reiley (2000) present an example of a framed field experiment designed to test a theory. The theory of multi-unit auctions predicts that a uniform-price sealed-bid auction will produce bids that are less than fully demand-revealing, because such bids might lower the price paid by the same bidder on another unit. By contrast, the generalized Vickrey auction predicts that bidders will submit bids equal to their values. In the experiment, List and Lucking-Reiley conduct two-person, two-unit auctions for collectible sportscards at a card trading show. The uniform-price auction awards both items to the winning bidder(s) at an amount equal to the third-highest bid (out of four total bids), while the Vickrey auction awards the items to the winning bidder(s) for amounts equal to the bids that they displaced from winning. List and Lucking-Reiley find that, as predicted by the theory of demand reduction, the second-unit bids submitted by each bidder were lower in the uniform-price treatment than in the Vickrey treatment. The first-unit bids were predicted to be equal across treatments, but in the experiment they find that the first-unit bids were anomalously higher in the uniform-price treatment. Subsequent laboratory experiments (see, for example, Engelmann and Grimm, 2003; Porter and Vragov, 2003), have confirmed this finding.

Finally, Karlan and List (2007) is an example of a natural field experiment designed to measure key parameters of a theory. In their study, they explore the effects of 'price' changes on charitable giving by soliciting contributions from more than 50,000 supporters of a liberal organization. They randomize subjects into several different groups to explore whether solicitees respond to upfront monies used as matching funds. They find that simply announcing that a match is available considerably increases the revenue per solicitation – by 19 per cent. In addition, the match offer significantly increases the probability that an individual donates – by 22 per cent. Yet, while the match treatments relative to a control group increase the probability of donating, larger match ratios – 3:1 dollars (that is, 3 dollars match for every 1 dollar donated) and 2:1 dollar – relative to smaller match ratios (1:1 dollar) have no additional impact.

In closing, we believe that field experiments will continue to grow in popularity as scholars continue to take advantage of the settings where economic phenomena present themselves. This growth will lead to fruitful avenues, both theoretical and empirical, but it is clear that regardless of the increase in popularity, the various empirical approaches should be thought of as strong complements, and combining insights from each of the methodologies will permit economists to develop a deeper understanding of our science.

See Also

- experimental economics
- experimental labour economics
- experimental methods in environmental economics
- experimental methods in economics
- experiments and econometrics
- experimental economics, history of

Bibliography

Anderson, E.T. and Simester, D. 2003. Effects of $9 price endings on retail sales: evidence from field experiments. *Quantitative Marketing and Economics* 1, 93–110.

Bohm, P. 1972. Estimating the demand for public goods: an experiment. *European Economic Review* 3, 111–30.

Engelmann, D. and Grimm, V. 2003. Bidding behavior in multi-unit auctions—an experimental investigation and some theoretical insights. Working paper, Centre for Economic Research and Graduate Education, Economic Institute, Prague.

Harrison, G.W. and List, J.A. 2004. Field experiments. *Journal of Economic Literature* 42, 1009–55.

Karlan, D. and List, J.A. 2007. Does price matter in charitable giving? Evidence from a large-scale

natural field experiment. *American Economic Review*, forthcoming.

Levitt, S.D. and List, J.A. 2006. What do laboratory experiments measuring social preferences tell us about the real world? *Journal of Economic Perspectives* 21(2), 153–74.

List, J.A. 2006. Field experiments: a bridge between lab and naturally occurring data. *Advances in Economic Analysis & Policy* 6(2), Article 8. Abstract online. Available at: http://www.bepress.com/bejeap/advances/vol6/iss2/art8, accessed 26 May 2007.

List, J.A. and Lucking-Reiley, D. 2000. Demand reduction in a multi-unit auction: evidence from a sportscard field experiment. *American Economic Review* 90, 961–72.

Lucking-Reiley, D. 1999. Using field experiments to test equivalence between auction formats: magic on the Internet. *American Economic Review* 89, 1063–80.

Orne, M.T. 1962. On the social psychological experiment: with particular reference to demand characteristics and their implications. *American Psychologist* 17, 776–83.

Porter, D. and Vragov, R. 2003. An experimental examination of demand reduction in multi-unit versions of the uniform-price, Vickrey, and English auctions. Working paper, Interdisciplinary Center for Economic Science, George Mason University.

Vickrey, W. 1961. Counterspeculation, auctions, and competitive sealed tenders. *Journal of Finance* 16, 8–37.

How to cite this article