Laboratory Experiments in Economics[1]

Gerardo L. Largoza[2]

ABSTRACT

By its nature, empirical approaches to hypothesis-testing in research areas such as social capital, which includes the study of such elusive 'state-of-mind' phenomena as norms and informal networks, would be difficult unless ways could be found to induce the relevant behavioral patterns in a controlled setting. This is where experiments could potentially provide a breakthrough. This paper aims to address common methodological concerns on the use of controlled laboratory experiments in economics, and to demonstrate how an experimental approach can help pin down important but elusive social / economic phenomena by presenting the design of an experiment to induce and measure prejudice.

Introduction

Controlled laboratory experiments have been used in economics since the 1910s, and for all sorts of purposes: from pedagogical attempts to confirm the existence of well-ordered indifference curves, to 'wind-tunnel' trials for various auction formats in aid of privatization efforts, to the simulation of oligopolistic behavior for use as evidence in anti-trust cases. Despite the conferral of the 2002 Nobel Prize to Vernon L Smith for his work in pioneering the use of lab experiments in economics, it has taken some time for the profession to warm to the use of such methods to study behavior. Friedman (2004) suggests why:

...a discipline becomes experimental when two conditions are met: its theory matures sufficiently to generate laboratory-testable predictions and pioneers develop useful laboratory techniques...Economic theory before the 1960s had little room for laboratory experiments (as) macroeconomics referred to prohibitively large-scale events (while) microeconomics referred mainly to the consequences of competitive equilibrium. Meaningful economic experiments became possible with the emergence of new theories in the 1960s (and) by the early 1970s, many economists began to recognise the potential of experiments to distinguish among the many alternatives.

This author's interest in experimental methods arose out of a 'frustration' with the progress of research in social capital, which includes the study of such elusive 'state-of-mind' phenomena as norms and informal networks. Although interesting datasets exist for such indicators as trust levels (say, from the World Values Survey), they are essentially 'happenstance' and are generally not amenable to the requirements of control. Surveys and focused-group discussions, on the other hand, while useful to some degree, are easily criticized on the grounds of their being 'hypothetical' and thus unreliable[3]. At the same time, interest

---

[1] Presentation made to the faculty of the Economics Department of Soka University, Tokyo 25 August 2004 & the 41st Annual Meeting of the Philippine Economic Society, 4 March 2005, Bangko Sentral ng Pilipinas, Roxas Boulevard Manila

[2] Assistant professor, Economics Department. Comments welcome at largozag@dlsu.edu.ph

[3] Often referred to as the Wallis-Friedman critique.

in the phenomenon of "implicit social contracts" – of the sort de Soto (2004) refers to as critical to the understanding of the informal economy in places like the Philippines – is growing. By its nature, empirical approaches to hypothesis-testing in these research areas would be difficult unless ways could be found to induce the relevant behavioral patterns in controlled setting. This is where experiments could potentially provide a breakthrough.

The objectives of the paper are as follows: (1) to address common methodological concerns on the use of controlled laboratory experiments in economics, and (2) to demonstrate how an experimental approach can help pin down important but elusive social/economic phenomena by presenting the design of an experiment to induce and measure prejudice.

The reader is strongly urged to consider John H Hagel & Alvin E Roth's ***The Handbook of Experimental Economics*** (1995, Princeton University Press), which provides an excellent survey of other uses of lab experiments – in the study of auction theory, imperfect markets, valuation, choice anomalies, etc, and

Friedman, Daniel & Alessandra Cassar (2004) ***Economics Lab: An Intensive Course in Experimental Economics***, Routledge London & New York., and

Friedman, Daniel & Shyam Sunder (1994) ***Experimental Methods: A Primer for Economists***, Cambridge: Cambridge University Press

which provide good reading for the design issues in lab work (control, randomization, factorial design, etc).

What follows are a series of issues designed to bring the reader up to speed with both the methodological foundations of experimentation in economics and some best-practices used in economics laboratories.

Part One

I. Methodological misgivings: How can small-scale lab results be generalized to the larger world?

Most experimenters have experienced resistance to their work coming from colleagues in the mainstream. This springs from a number of methodological misgivings about the use of experiments to study economic behavior, but easily the most cited one refers to the problem of *external validity* – i.e. the issue of how results obtained in artificial laboratory environments can be generalized to the larger world outside the lab.

There are 3 well-known responses to this most common of concerns. The first is a somewhat rhetorical one, given by Vernon Smith, one of the 2002 Nobel Laureates in Economics:

'propositions about the behaviour of individuals and the performance of institutions in laboratory micro-economies should apply also to non-laboratory micro-economies where similar *ceteris paribus* conditions hold' (1928, in Friedman & Cassar 2004)

The point here is that few experimenters would dare make any broad claims about their lab results in the first place. They use lab experiments the same way analytical theorists use abstract math models – to generate insight, in this case, through inductive means (by searching for empirical regularities in the results, or seeing which among competing theories best 'organizes' the data). This inductive process of theorizing has very deep roots in the natural sciences, but – for reasons particular to the historical development of economics in North America (Blaug, 1998) – remains on the margins of economic science. Most training in economic model-building is *deductive* in nature; that is to say, abstract mathematical models are set up and principles are derived purely by logical processes. Statistical analysis with data is then used, more often than not, to confirm these principles.

But this implies that since most laws of economic behavior are themselves cast in general terms (since math models often abstract from culture, gender, power relations etc), one can argue that they ought to apply in the special case of laboratory set-ups using real people dealing with real incentives and receiving real payoffs. Indeed, an honest skeptic would recognize that it is not enough to merely dismiss experimental results because they were generated from artificial lab conditions; it behooves one to specify exactly what elements of a laboratory set up would taint the naturally-occurring behavior of participants in an experiment.

A second response is empirical in nature. Charles Plott (1982) carried out a series of experiments comparing the strategies of student subjects and professional traders and found sufficiently similar equilibrium behavior (although means & variances tended to differ). He went on to suggest that 'the contrast between lab behavior and the real world is often exaggerated'. The main difference seems to be that laboratory processes are **simpler** than naturally occurring ones. But simplicity (parsimony) is a virtue in science (!). Even analytical and econometric models can never hope to capture all the complexity inherent in nature – nature is its own best model, as they say. But experiments provide a useful element of **control**; data can be generated for the specific purposes of the researcher, to a degree not often possible with happenstance data.

Finally, there is Hicks' Lemma (1939), from which we can imply that a person's unobserved preferences can be induced in laboratory experiments if the experimental design bears the properties of **monotonicity**, **salience** & **dominance**.

The proof is as follows:

Suppose the experiment is intended to induce smooth preference functions. The participant is clearly told that scrip made up of red (x) and blue (y) sheets of paper will correspond to payments defined by some table, such that

$$\Delta m = U(x, y)$$

The induced preferences would then be

$$W(x,y) = V(m_0 + U(x,y), z_0 + \Delta z)$$

where (m_0, z_0) is the (unobservable) initial endowment of money and all other non-pecuniary rewards before the experiment, and z represents all the non-pecuniary rewards from the experiment.

Hicks demonstrated that two utility functions (say, for scrip and for monetary rewards) would represent the exact same preferences if their marginal rates of substitution were the same throughout.

$$MRS_w = \frac{W_x}{W_y} = \frac{V_m U_x + V_z \Delta z_x}{V_m U_y + V_z \Delta z_y} = \frac{V_m U_x}{V_m U_y} = \frac{U_x}{U_y} = MRS_u$$

In the above expression, the first and last terms represent marginal rates of substitution for the induced and actual underlying preferences. The second equality implies that to successfully induce accurate preferences, the marginal rate of substitutions must exist and exhibit **monotonicity**; that is to say, there must exist within the experiment a reward medium for which satiation can (ideally) be indefinitely postponed.

The third term in the expression, derived through the use of the chain rule implies the requirement of a property known as **salience** – that is to say, a valid economic experiment must make it clear to participants that whatever rewards they receive will be on account of actions undertaken and decisions made within the experiment.

If the effects of non-pecuniary preferences of participants can be minimized (as in the fourth term), then the (ideal) property of (full) **dominance** is achieved. Well-designed experiments will use a number of procedures to ensure that actions are *dominated by* the participant's desire for the rewards on offer.

The point is not that Hicks' Lemma proves experiments can say something about agent behavior but that well-designed ones (i.e. experiments that feature the important properties of monotonicity,

salience & dominance) can simulate or reproduce unobserved economic behavior.

This, of course, implies questions of degree. Unfortunately, tests for external validity do not yet exist and some authors (Sturm & Weimann 2002, for instance) claim that such tests are analytically impossible to construct.

II. The question of robustness and context

A second methodological issue often raised has to do with the fact that the games participants are often asked to play in lab experiments are just that – games. The artificial setting abstracts from the social context that often animates economic decisions – and this ought to weaken any claims that experimenters make about robustness of their results.

One response would be to point out that **all** theorists – even those that do analytical & econometric modeling – are faced with the same dilemma. *Which factors should one control for when examining economic behavior?* is a question every researcher faces, regardless of the method used. Indeed, the ability to create an artificial environment in a lab often brings with it the temptation to replicate every nuance of the larger world therein. But to do so (or aspire to it) would not be good scientific practice (at least in most sciences with a long experimental tradition), if we are to wield Occam's Razor. Science, after all, values the ability to generate deep insights from the most simple (uncluttered, elegant) principles. Thus, we must gently warn that experiments are not meant to supplant the judicious use of existing theory, if only to identify *the most relevant variables* for control.

Having said this, with a large enough research budget, one can probably find that most important variables in question are amenable to experimental control. If one feels that payoffs are too low to make a difference to professional traders, for instance, increasing them is always a possibility. If certain sample characteristics are sought (say, if you needed 'inequality-averse' people), one may use a battery of

existing psychological tests to screen or group participants. All told, concerns about social context and the complexity of the real world seem to be more related to the **design** of particular experiments rather than a critique of the experimental method itself.

Closely allied to this concern is the comment often made that since most experiments use university students, the possibility that self-selection will taint the results is serious.

There is no denying that the vast majority of papers in experimental economics feature the use of students as participants. Students are frequently used because only they bear some desirable qualities: they are often adept at following written instructions, they have steep learning curves (which reduce the number of trials needed in order for the participants to 'get it'), they have low opportunity costs (which keeps budgets down). In most cases, if there are no special theoretical demands, students ought to be considered 'real' people.

Having said this, there is nothing that prevents researchers from carrying out controlled *field* experiments where 'real' people are invited as subjects. Indeed, the increasing popularity of actual electronic markets makes field experiments a natural extension to the general methods discussed here.

III. Lab experiments versus surveys or similar experimental practices (e.g. psychology)

Lab experiments are an endeavor in generating controlled data sets. It is sometimes argued, however, that surveys, focused group discussions, or indeed experimental practices in psychology are likewise capable of producing data amenable to statistical treatment. Given this, experimenters are often asked what distinguishes lab experiments from these other methods of data generation.

Two things, generally. First, economic experiments must be salient – i.e. there must be a payoff structure that is directly related to a person's actions or decisions in the experiment. Surveys are not salient

(although those that make people choose from goods while allowing them to keep their choices are).

The use of deception is the other. Since experimental psychologists often deal with cognitive processes or states of mind, there is scope for the use of deception in their work. In general, experimental economists frown upon the use of deception as it tends to compromise dominance (although Sturm & Weimann 2002 report work that provides some justification for the selective use of it).

Having said this, economic experimenters share a number of difficulties with their colleagues from different experimental traditions. Chief among them is the problem of how to deal with 'confounded' variables – or those that provide 'joint' and seemingly 'inextricable' explanations of outcomes[4].

It may be said that the approach taken by economic experimenters is similar to the tack taken by most other theorists. First, experimenters must have some intuition about the variables that are confounded; they must be able to identify what these factors might be. There is really no substitute for good theoretical sense, as experienced experimenters point out. Next, they may wish to *directly* control for them. For instance, if gender is seen as an explanatory variable that is confounded with a person's natural tendency toward altruism, one could create male and female groupings in the experiment.

Alternatively, one can rely on the statistical principle that demonstrates how randomization of treatments can separate out confounded variables over the long run. Repetition of the trials and the use of a number of efficient factorial designs can help 'tease out' the isolated impact of variables.

At this point, of course, experimenters are also exhorted to keep experiments as simple as possible. Simplicity encourages salience.

[4] The well-known story of 'aviophiles' and 'luminists' trying to explain rapid plant growth illustrates this problem well. 'Aviophiles' believe bird droppings do the trick while 'luminists' argue that it must be the shade from trees that best explains plant growth. The two explanations are confounded because they often occur jointly.

Interminable repetitions, on the other hand, create 'demand effects' in which subjects try to 'guess' the objectives of the experiment and respond to them, rather than to the incentives (it's said economics majors are notorious for doing this; many experiments in fact treat economics majors as a special class of subject).

IV. Best-practices for achieving monotonicity, salience & dominance

Apart from those mentioned elsewhere in this piece, experimenters may benefit from:

- Test-bedding experiments with a small group before the actual run (to make sure instructions are clear, among other reasons), and conduct a number of 'practice' trials during the experiment itself to get subjects used to the incentive structure;

- Avoiding the use of loaded terms in order to encourage dominance. Use A & B instead of 'cooperate & betray' in PD games, for instance – unless, of course, the whole point of the experiment is to test how the framing of propositions affects behavior;

- Maintaining the privacy of one's own objectives, of the subjects' payoffs, of their & others' decisions – unless the whole point is to study, say, behavior in tournaments;

- Using scrip (if one wishes to abstract from income or scale effects), with a conversion rate known only to you (as long as the *existence* of a conversion rate is made clear to the subjects).

We must stress that, as in most other experimental disciplines, the rules in economic experiments are meant to be tailored to the needs of the researcher. These best-practices are not intended to be inflexible, but are meant to be driven by the objectives of one's work.

As a final point, we bring up the matter of there being a considerable amount of literature in cognitive psychology that presents a critique of the *internal* validity of experiments. Basically, it argues that lab subjects

have varied ways of 'framing' games, and it is impossible for experimenters to know which 'games' their subjects are actually playing. For instance, dictator games may be viewed as confrontational situations in some instances (triggering Nash behavior) or they may be construed as scenarios where some sort of 'other-orientated behavior' is expected (which would then lead to fairness-driven strategies). All this implies that large variations in behavior can result from small changes in the way the game is 'framed'.

There is much truth in this at present, and it represents one of the major research areas in the evolution of experimental theory (most of the best-practices were developed only 30 years ago, it must be noted). Kahneman & Tversky's **Choices, Values & Frames** (2000 Cambridge University Press) provides a compelling account of the issues, but operationally, little can be done at present other than to conduct series of experiments, each featuring small modifications in the games, and then testing for robustness.

Part Two

A Simple Experiment to Induce and Measure Prejudice

Prejudice – or the instinctive preference for economic partners of one's own race – is a phenomenon that is at once relevant and difficult to pin down using traditional methods of data gathering. Happenstance or secondary data often have limited use since they are not generally amenable to control; furthermore, since there is a stigma attached to being 'prejudiced', respondents to surveys or focused-group discussions are likely to suffer from self-awareness, which leads to unreliable results.

An experimental approach not only allows the researcher to overcome these difficulties, but a properly-designed lab set-up can help incorporate some fairly nuanced psychological manifestations of prejudice (distrust, punishment severity, coalition-forming behaviour, dominance) into the dataset.

Such an experiment would proceed thus: participants of various ethnicities (say, Filipinos and Chinese-Filipinos) are invited to the laboratory. The intent of the experiment is couched in only the most general terms, and no effort is made to segregate the participants. Instead, answer sheets are provided and the experimenters 'track' ethnicities via the numbers found on the sheets.

The experiment is made up of a set of well-known one-shot games, designed to elicit different manifestations of prejudicial behavior as participants take turns experiencing partners of same and different ethnicities. These 'partners' are not fellow participants – instead, they are surnames found on the answer sheet that provide clear signals of ethnicity[5]. The participants are informed that 10% of them will be selected to receive the actual amounts they played with in the course of the experiment (to induce salience)

The sequence is as follows: the 'trust game' is played in order to provide participants an opportunity to manifest prejudice that arises out of greater distrust for people outside one's ethnicity (this is indicated on the following page as the 'first decision'). The 'dictator game' (indicated as 'second decision') is designed to analyze coalition-forming behavior; to test whether prejudice is manifested in the way partners are chosen or favored in economic transactions. The 'ultimatum game' ('third decision') is a test of prejudice, as manifested in punishment behavior. The null hypothesis here is that a participant would punish unfair behavior equally, regardless of the partner's ethnicity. Finally, the 'battle-of-the-sexes game' ('fourth decision') is a test of prejudice-driven first-mover aggression. The idea is that prejudiced participants would choose domineering outcomes if their partners of different ethnicities are perceived as being somehow 'weaker'.

Although the games are one-shot, repeated *treatments* are necessary; each participant must experience being partnered with persons of similar and different ethnicities in order to allow the experimenter to gain control over motivations. The outcome is a panel data set made up of values signifying decisions from each of the games

played, all of which are then amenable to a variety of statistical treatments.

INSTRUCTIONS FOR THE EXPERIMENT

Welcome

This is an experiment to investigate decision-making behavior. The instructions are simple. If you follow them carefully, you can earn a considerable amount of money – 10% of the participants will be chosen to receive the actual amounts they declare, for a randomly-chosen round. Earnings will be paid in cash immediately after the experiment. In addition, you receive Php 100 for showing up. The amount has been supplied by the Asia Development Research Forum.

In the experiment you will have to make 4 decisions involving persons whose names are found on the card you have been given. These individuals are not fictional. Your decisions are to be written down on the sheet provided, treated separately, and will not be marked for identification. In other words, you are anonymous.

Do not communicate with the other participants and try to avoid the possibility of others seeing your form.

First decision

Here is Php 1000. Choose an amount to give to the person whose name you find on the card. That amount will then be trebled after which the person will decide how much to return to you.

Second decision.

You have two choices. You may split Php 1000 between yourself and two other people whose names are found on Line 2 of the card (with you deciding how much each gets)

or

You may split Php 800 between yourself and just one other person. You get to decide whom to split with and how much each of you gets.

Third decision

You must split Php 800 with a person whose name is found on Line 3 of the card. You can split it only two ways – either by taking Php 500

[5] Only surnames are used in order to avoid confounding gender signals with ethnicity.

or Php 300. The other person will be making the decision simultaneously.

You will receive the amount you bid only if your bids are complementary.

Fourth decision

The person whose name is found on Line 1 has made a similar *First Decision* and the amount offered to you is found on Line 4. You may choose to accept the partitioning or reject it. If you accept, you will get the amount offered. If you reject it, both of you will receive nothing.

References

Blaug, Mark (1998). 'The Problem of Formalism'. *Challenge Magazine.*

De Soto, Hernando (2005). 'The Mystery of Capital: The Role of Property Rights in Creating Wealth and Alleviating Poverty'. *Sir John Crawford Memorial Lecture*, Consultative Group on International Agricultural Research. Mexico.

Friedman, Daniel & Alessandra Cassar (2004) **Economics Lab: An Intensive Course in Experimental Economics**, Routledge London & New York

Friedman, Daniel & Shyam Sunder (1994) **Experimental Methods: A Primer for Economists**, Cambridge: Cambridge University Press

Hagel, John H & Alvin E Roth (1995) **The Handbook of Experimental Economics**, Princeton University Press

Hicks, John (1939) **Value and Capital**, Oxford: Oxford University Press.

Kahneman, Daniel & Amos Tversy (2000) **Choices, Values & Frames**, Cambridge University Press

Plott, Charles R & Louis L Wilde (1982). *Professional Diagnosis vs Self-Diagnosis: An Experimental Examination of Some Special Features of Markets with Uncertainty.* In Vernon L Smith (ed) **Research in Experimental Economics**. Vol 2 Greenwich, Conn: JAI Press.

Sturm, Bodo & Joachim Weimann (2002). *Experiments in Environmental Economics – a Survey.* Otto-von-Guericke University Magdeburg.