The Australian Economic Review, vol. 43, no. 3, pp. 337–45

For the Student

Experimental Economics

Nikos Nikiforakis*
Department of Economics, The University of Melbourne

1. Introduction

Experimental economics is a branch of economics that uses controlled experiments to explore the predictive power of theories, evaluate behavioural assumptions, investigate behavioural regularities and test the implementation of policies. It is one of the fastest growing areas of economics (Oswald 2010).

The development of an experimental methodology in economics is recent relative to other disciplines such as physics and psychology. Although informal experiments were conducted as early as in the eighteenth century (see Bernoulli 1738), the results from the first formal economic experiment appeared in an article by Edward Chamberlin in 1948. After half a century of continuous growth in the number of economic experiments, in 2002, Vernon Smith, a participant in Chamberlin's experiment, was awarded the Nobel Prize 'for establishing laboratory experiments as a tool for empirical economics analysis' (Nobel Announcement 2002).

This article is an introduction to experimental economics aimed primarily at students and scholars with little or no prior knowledge on the topic. The next section discusses why controlled experiments are a valuable tool in economics and proceeds to present lab and field experiments, provide examples of experiments, and address some of the common criticisms regarding laboratory experiments.

2. Why We Need Experiments in Economics

Economics aims to understand how different policies, incentives and institutions affect individual behaviour and social welfare. For this purpose, it is essential to understand the determinants of human decision-making. This, however, is not always simple.

Economists, like other social scientists, examine (naturally occurring) data patterns and try to construct explanations for different phenomena. In order for science to progress, we must be able to compare the different explanations for the phenomena. The complexity of economic life, however, means that many variables are changing simultaneously, making it difficult to distinguish between cause and effect. The result is that we are often left with multiple equally plausible explanations for a given phenomenon.

The need for experiments arises from the difficulty of establishing causality using non-experimental data. In the laboratory, the experimenter can implement an 'exogenous' change to the variable of interest (for example, increase the demand for a good) *keeping all other variables unchanged* (for example, the supply for the good) and measure how the change affects individual behaviour and social welfare (for example, how prices and consumer and producer surplus are affected).

Ensuring that a change is exogenous and all other variables remain unchanged is essential for establishing causality, but difficult (if not impossible) to achieve with non-experimental data. Although the development of sophisticated econometric techniques and the availability of more detailed datasets have improved our ability to analyse non-experimental data, these techniques typically require that additional statistical assumptions be made. Furthermore, non-experimental datasets may not include variables that may be of interest; they may also lack the necessary

* I would like to thank Mick Coelli, Nisvan Erkal and an anonymous referee for helpful comments.

Figure 1 The Experimental Economics Laboratory at the University of Melbourne

level of detail for evaluating microeconomic models.

3. Laboratory and Field Experiments

A feature of economic experiments that distinguishes them from experiments in other disciplines, such as psychology, is that participants make decisions that have real monetary consequences. This feature also distinguishes experimental data from self-reported survey data that are sometimes used for empirical investigation. The presence of monetary rewards gives participants an incentive to truthfully reveal their preferences.

Experimental data are most commonly collected in purpose-built laboratories. These laboratories typically consist of a number of computers that are partitioned so that anonymity is preserved and peer pressure is eliminated (see Figure 1). Participants are usually university students that use the computers to interact with each other and make decisions with real monetary consequences.

The main advantage of laboratory experiments is the control they afford researchers. Unlike in naturally occurring environments, experimenters can easily control among others the information participants have at their disposal, the actions participants can take and whether participants anticipate future interactions with the other participants. The unparalleled control allows researchers to repeatedly study the same environment changing only one variable at a time and measuring its impact.[1] It also permits researchers to replicate the conditions found in other studies and evaluate the robustness of previous findings. Replication is important if we wish to establish beyond reasonable doubt the causal relationship between two variables.

Although the number of laboratory experiments continues to grow, controlled field experiments have also become popular in recent years. Most economists use the term 'field experiment' to refer to controlled experiments that take place in naturally occurring environments, such as within firms. Participants may be students, but often are not. The appeal of field experiments is that participants are observed in their natural environment and may be unaware that they are part of an experiment. However, this comes at a cost as some control is lost in the field. For example, researchers may be unable to establish and control an individual's willingness to pay for a good. Thus, it is much harder to keep 'all other factors' constant in a field experiment, making the replication of a field experiment difficult. Furthermore, there can be limitations to what one can do in the field. For example, a firm may not be willing to cut wages so that researchers can study the reaction of its employees.

Laboratory and field experiments are complementary tools. Whether one should use a laboratory or a field experiment depends on the research question being asked. For example, one may prefer to use a laboratory experiment if he or she wishes to evaluate the predictive power of a particular theory, as the laboratory environment allows him or her to recreate and control the theoretical assumptions. Once the theory has been scrutinised in the laboratory, one may wish to examine its predictive power using a field experiment. If the theoretical predictions are borne out in laboratory and field experiments, then that provides strong support for the predictive power of the theory.

4. Experimental Treatments

Experiments, whether in the lab or in the field, typically consist of a set of treatments. A treatment is a 'completely specified set of procedures which includes instructions, incentives, rules of play, etc.' (Holt 2007).

Typically, experiments consist of at least two treatments—a 'baseline' and a 'main' treatment. The purpose of the baseline (or 'control') treatment is to provide a benchmark for evaluating the effect of our variable of interest (the 'treatment variable'). The main treatment replicates the baseline treatment changing only the treatment variable—the effect of which we are interested in measuring. By comparing outcomes in the baseline and the main treatment, we can evaluate the impact of a change in our treatment variable.

Although the experimental design of a study depends on the type of question one wishes to address, it is common (and advisable) to have at least two treatments. One reason is that participants are not randomly selected to participate in experiments. This means that inferences cannot be always made from a laboratory sample to a more general population because of potential selection bias. For example, if participants in a laboratory experiment tend to be risk-loving, we cannot safely infer whether this will be the case for non-participants too. It may be that the most risk-averse individuals do not sign up for experiments as they dislike the variance in payments. Although there is no evidence for such selection effects (see Cleave, Nikiforakis and Slonim, 2010), in general, if participants are randomly assigned to treatments, then inferences about the effect of a *treatment* may be generalised to non-participants.

5. Examples of Experiments

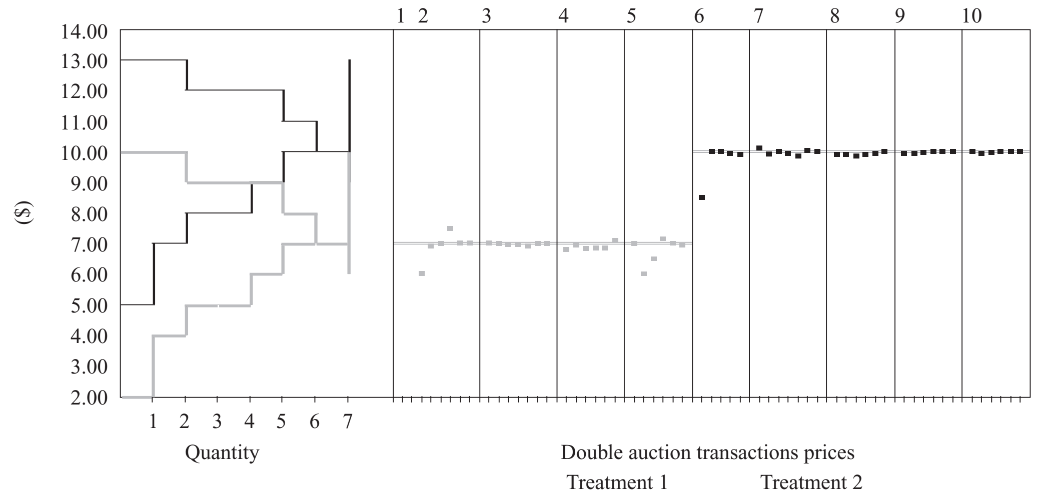
Experiments can be used for various purposes. Because of space constraints, here I will discuss three common reasons for conducting experiments: (i) studying theoretical predictions; (ii) studying behavioural assumptions; and (iii) establishing behavioural regularities.[2]

5.1 Studying Theoretical Predictions

Economic theories are usually evaluated on their generality, plausibility and predictive power. Although evaluating the generality of a theory can be straightforward, evaluating the plausibility and predictive power of a theory requires the collection of data. Theories make predictions about what will happen if a certain variable changes. As mentioned earlier, it can be difficult (if not impossible) to ensure that factors that we are not interested in studying are held constant outside the laboratory. For this reason, laboratory experiments have often been used to evaluate theoretical predictions.

A prominent example of such an experiment is Vernon Smith's (1962) original study of the predictive power of competitive price theory

Figure 2 Transaction Data from a Computerised Double Auction Classroom Experiment at the University of Melbourne



Note: Data for 14 participants.

in an experimental market.[3] According to the theory, in a competitive market, goods will be traded for a price that equates supply and demand.

Participants in Smith's experiment were students. They were divided into buyers and sellers. Each buyer was given a card with a number representing how much he or she valued a fictitious good (v). Each seller was given a card with a number representing his or her cost of production for the fictitious good (c). The buyers' values and the sellers' costs were private information. Any buyer could agree with any seller on a price to trade the fictitious good (p). In order to agree on a price, both buyers and sellers could make public offers. The trading institution used by Smith has become known as the 'double auction'.

The earnings of a buyer in a given trading period were $v - p$ if he or she agreed on a trade and 0 otherwise. Similarly, the earnings of a seller in a given trading period were $p - c$ if he or she agreed on a trade and 0 otherwise. The buyers' values induced a downward-sloping (ladder-like) demand function, while the sellers' costs induced an upward-sloping (ladder-like) supply function. Figure 2 presents a market similar to that induced by Smith that was used in a computerised classroom experiment at the University of Melbourne. The experiment lasted 10 periods with seven buyers and seven sellers. The left panel presents the supply and demand schedules. The equilibrium price is found at the point at which supply and demand functions cross. As can be seen, the equilibrium price in periods 1–5 was \$7 and the equilibrium quantity was 6 or 7 units. In period 6, there was an unannounced shock in the supply and demand, leaving the equilibrium quantity unaffected, but raising the equilibrium price to \$10. The right panel of Figure 2 presents the individual trades and the price at which they occurred.

The data in Figure 2 show that the equilibrium is a remarkably good predictor of outcomes when both buyers and sellers can make public offers. Prices quickly converge to the equilibrium. Even when supply and demand change in period 6 without any notice, causing the equilibrium price to increase sharply, prices of traded goods adjust quickly.

There have been dozens of experiments examining the robustness of these results. For example, Plott and Smith (1978) found that the results are similar (albeit convergence to equilibrium is slightly slower) if only sellers can post prices as is typically the case in retail markets; Holt, Langan and Vilamil (1986) showed that prices converge to equilibrium, even if one side extracts almost all profit from trading.

What is important to emphasise for the purpose of this article is the difficulty of evaluating competitive price theory with non-experimental data as, outside the lab, demand and supply are constantly changing and usually unknown.

5.2 Studying Behavioural Assumptions

Theories rely on multiple assumptions. If a theoretical prediction fails to be borne out in the lab, we need to examine which of the assumptions was incorrect for our experimental sample. One common assumption in many microeconomic models is that individuals are self-regarding. That is, they maximise their utility without any regard for others.

It is difficult to establish in the field whether an individual cares for others or not. Even if we observe someone doing what appears to be a selfless act, in most cases, a sceptic can always come up with an alternative explanation. For example, giving to a beggar may be driven by a desire to impress your friends; helping a colleague may be motivated by a desire to impress your boss, and so on. However, in the lab we can isolate many of the alternative explanations.

The most famous experiment studying whether individuals are other-regarding is the 'ultimatum game' that was first studied by Güth, Schmittberger and Schwarze (1982). The game is simple. One individual (Alice) is given an endowment, say $10. She is then asked to make a take-it-or-leave-it offer to another individual (Bob). If Bob accepts the offer, it is implemented. If Bob rejects the offer, then both earn 0.

What 'should' Alice and Bob do in the ultimatum game? If Alice and Bob know that they are both self-regarding and that they will never interact with each other again (which removes any concerns about strategic behaviour, such as rejecting low offers today to receive higher offers tomorrow), the game has a unique (subgame-perfect) equilibrium: Bob prefers 1 cent to nothing; Alice—who wants to keep as much for herself as possible—will thus offer Bob 1 cent and keep the remaining $9.99 for herself.

The experimental results in the ultimatum game refute this prediction. The average offer is about $4, and offers under $2 are frequently rejected. Interestingly, the likelihood of rejection increases with the size of the offer. Given that participant anonymity in the lab is guaranteed (as is the fact that Alice and Bob will interact only once), the results suggest that many individuals are indeed other-regarding; no alternative explanation seems plausible. Those in the role of Bob prefer to earn nothing than to accept a low offer. The simplicity of Bob's decision suggests that his behaviour cannot be attributed to confusion. Many individuals in the role of Bob adhere to the norm of an equal (50–50) split and are willing to punish norm violators. In fact, the likelihood of rejecting a low offer appears to increase over time.

5.3 Establishing Behavioural Regularities

One may be interested in using laboratory experiments to establish behavioural regularities for three reasons. First, it may be that a theory admits multiple equilibria for a given game. Second, we may need more observations before we know how we can substitute an incorrect behavioural assumption and construct new theoretical models. Third, we may be interested in observing the robustness of results in a range of situations even if our existing theories do not predict any differences across situations.[4]

An equilibrium can be thought of as a prediction of what will happen in a strategic interaction. When a game has multiple equilibria, we may be interested to know if one of the equilibria is more commonly observed in practice than the other. A classic example of the first category is experiments on the 'minimum-effort game'. The game models situations in which group performance depends on the performance of its weakest link and was first studied by Van Huyck, Battalio and Beil (1990). For example, consider the preparation of a burger at a fast-food chain or the preparation of an airplane for take-off. In both cases, the speed of the service (and thus customer satisfaction) depends on the speed of the slowest worker. Assuming there

Table 1 The Minimum-Effort Game
(entries are monetary payoffs)

	Minimum effort chosen by any group member						
Effort chosen by individual	*7*	*6*	*5*	*4*	*3*	*2*	*1*
7	\$1.65	\$1.40	\$1.15	\$0.90	\$0.65	\$0.40	\$0.15
6	–	\$1.50	\$1.25	\$1.00	\$0.75	\$0.50	\$0.25
5	–	–	\$1.35	\$1.10	\$0.85	\$0.60	\$0.35
4	–	–	–	\$1.20	\$0.95	\$0.70	\$0.45
3	–	–	–	–	\$1.05	\$0.80	\$0.55
2	–	–	–	–	–	\$0.90	\$0.65
1	–	–	–	–	–	–	\$0.75

are sufficient incentives, the best outcome for each group member is for all involved to work as hard as possible. However, because working hard is individually costly, an individual would prefer to work less if others do not work hard.

The minimum-effort game can be seen in Table 1. The entries in Table 1 represent the earnings of a given group member as a function of the *minimum* level of effort chosen by the others in the group. For example, if individuals A, B, C and D choose effort levels of 3, 4, 7 and 1, their respective earnings would be 0.55, 0.45, 0.15 and 0.75. A careful examination of Table 1 shows that if individuals are motivated by a desire to maximise their earnings from the experiment, the game has seven equilibria. In particular, if the minimum effort in the group is 1, then the best response is to also choose an effort of 1; if the minimum effort in the group is 2, then the best response is to also choose an effort of 2, and so on.

What makes the minimum-effort game interesting is the tension between the payoff-dominant equilibrium (all choose an effort of 7) and the 'safe' equilibrium that maximises the smallest possible payoff (all choose an effort of 1). Is there an equilibrium that is more salient to subjects?

In laboratory implementations of the minimum-effort game, a number of players must choose simultaneously and without communication a 'level of effort'. Dozens of minimum-effort games have been carried out (see Devetag and Ortmann 2007). The results suggest that which equilibrium is selected depends on a range of factors, such as the cost of coordination failure, the number of available actions, the existence of repeated interaction with the same participants, and the availability of information about others' past actions. However, it is not uncommon for groups to converge to the 'safe' equilibrium, which yields the lowest payoffs for everyone involved. These observations allow us to enrich our behavioural assumptions about how individuals play games with multiple equilibria.

6. Common Criticisms of Laboratory Experiments

Laboratory experiments were originally met with scepticism. Although the scepticism has subsided, it has not disappeared completely. Most experimental economists have had to explain to colleagues at some point in their career why a certain experimental result was interesting. A typical question is: 'Why should we care about the behaviour of *students* who make decisions *under scrutiny* and for *low stakes* in a laboratory experiment? What can we really learn from such experiments about *the real world*?'

These questions are valid. Indeed, you may have asked yourself the same question while reading this article. However, the most important thing to understand is that, while things such as the experimental subject pool may (or may not) affect the ability to generalise experimental results in different domains, they do not appear to pose a *fundamental* problem for laboratory experiments. In fact, as experimental economists say: '*Criticisms regarding laboratory experiments can be addressed with new experiments*.'

6.1 The 'Real World'

Most experimental economists dislike the distinction between the laboratory and the 'real world'. They prefer referring to 'naturally occurring environments' or to the world 'outside the laboratory'. The reason is nicely explained in the following quote by Vernon Smith:

> The laboratory becomes a place where *real people* earn *real money* for making *real decisions* about abstract claims that are just as 'real' as a share of General Motors.
> [Smith 1976, p. 275]

This quote highlights the fact that when people ask what we can learn about the 'real world' from lab data, they are simply thinking of different situations in which individuals with different characteristics to those of the experimental subjects make decisions in environments also with different characteristics. Although it may be true that behaviour may be different in different circumstances, when one identifies the circumstances of interest, they can usually be recreated in the laboratory.

6.2 Subject Pool

Most laboratory experiments use students as participants. The reason is threefold. First, academics have easy access to students. Second, students around the world are fairly homogeneous, which allows researchers to replicate previous experiments more easily. Third, students are smart and can thus understand abstract instructions and complex problems quickly. This isolates one potential explanation for experimental results, namely, that the results are a consequence of participants' confusion.

Nothing prevents researchers from using non-students as participants. If one is interested in learning something about the behaviour of a particular population (say, managers), then it is better to use participants from that population. Or, if one is interested in how people respond to the threat of punishment, then students may be a better sample to start with (for the reasons mentioned above).

Two more things are also important to remember. First, experimenters typically make inferences on the effect of the different treatments and not on individual behaviour. Therefore, unless different populations are expected to react to the treatment manipulation differently, any results obtained with students will also hold for other populations. Second, economic theories are general and do not target a particular population or environment. Thus, they should also predict behaviour inside the laboratory.

6.3 Low Stakes

Another common criticism is that decisions in laboratory experiments involve low stakes. Behaviour, some argue, may be different when stakes are large. Although this is true, it is not a fundamental criticism of laboratory experiments. Nothing prevents researchers from raising the stakes and seeing the impact the change has on behaviour. In fact, some people have done exactly that. Lisa Cameron (1999), for example, took advantage of the purchasing power of the Australian Dollar in Indonesia and ran ultimatum games in which the amount at stake was up to 1 month's wages. She found that the size of the stakes had little effect on behaviour. Other studies have found similar results.

6.4 Scrutiny

Human decisions are the product of a complex process. People may behave very differently in the laboratory, where they feel they are being observed, than in naturally occurring situations. This may seem at first like a fundamental flaw of the experimental methodology, but it is not. Given that experimenters typically make inferences about the effect of a particular treatment, scrutiny is not a problem as long as its extent has been kept constant in the baseline and in the main treatment. Moreover, in the laboratory, one can actually vary the extent of scrutiny in order to study its impact on behaviour. After all, participants are real people who care about scrutiny as much in the lab as in other places. Some experimental economists have developed protocols preventing the experimenter from observing a participant's actions and even his or her final earnings.

7. Final Remarks

This article offers a brief introduction to experimental economics. As most controlled experiments are conducted in laboratories, the article places most emphasis on laboratory experiments.

Laboratory experiments can be thought of as a tool in the economics toolbox. Like any tool in any toolbox, however, laboratory experiments cannot perform all functions by themselves. Data from field experiments and other types of data can be used to address questions that may be difficult to answer in the laboratory (for example, the predictive power of the permanent income hypothesis). They can also be used to add validity to laboratory findings. Nevertheless, laboratory experiments are an essential tool for economists.

Laboratory experiments allow researchers to implement truly exogenous changes in their variables of interest, holding all other variables constant. They also allow economists to explore questions that are difficult (if not impossible) to explore using different methods, as evinced by the examples in Section 5. For these reasons, experiments allow us to learn more than ever before about the determinants of human behaviour.

June 2010

Endnotes

1. The ability to establish a causal relationship between two variables is often referred to as 'internal validity'.

2. Experiments have also been used for other purposes, such as to study consumer behaviour, inform court decisions in anti-trust cases, and test policy implementation. The Victorian Government in Australia frequently uses lab experiments to inform policy (for example, Duke and Gangadharan 2008). For an exhaustive review of the experimental literature, see Plott and Smith (2008).

3. Two other notable examples in this category are experiments testing the predictions of auction theory and contract theory. For a discussion, see Plott and Smith (2008).

4. Because of space constraints, I only give an example from experiments using a game with multiple equilibria. An example of the second and third categories is the public good game with punishment opportunities. See, for example, Fehr and Gächter (2000); Denant-Boemont, Masclet and Noussair (2007); and Nikiforakis (2008, 2010).

References

Bernoulli, D. 1738, 'Specimen theoriae novae de mensura sortis', *Commentarii Academiae Scientiarum Imperialis Petropolitanae*, vol. 5, pp. 175–92. 1954, 'Exposition of a new theory on the measurement of risk', *Econometrica*, vol. 22, pp. 23–36. English translation.

Cameron, L. 1999, 'Raising the stakes in the ultimatum game: Experimental evidence from Indonesia', *Economic Inquiry*, vol. 37, pp. 47–59.

Cleave, B.L., Nikiforakis, N. and Slonim, B. 2010. 'Is there selection bias in laboratory experiments?' *Working Paper Series, Dept. Economics, The University of Melbourne*.

Chamberlin, E. H. 1948, 'An experimental imperfect market', *Journal of Political Economy*, vol. 56, pp. 95–108.

Denant-Boemont, L., Masclet, D. and Noussair, C. 2007, 'Punishment, counterpunishment and sanction enforcement in a social dilemma experiment', *Economic Theory*, vol. 33, pp. 145–67.

Devetag, G. and Ortmann, A. 2007, 'When and why? A critical survey on coordination failure in the laboratory', *Experimental Economics*, vol. 10, pp. 331–44.

Duke, C. and Gangadharan, L. 2008, 'Salinity in water markets: An experimental investigation of the Sunraysia salinity levy, Victoria', *Ecological Economics*, vol. 68, pp. 486–503.

Fehr, E. and Gächter, S. 2000, 'Cooperation and punishment in public goods experiments', *American Economic Review*, vol. 90, pp. 980–94.

Güth, W., Schmittberger, R. and Schwarze, B. 1982, 'An experimental analysis of ultimatum bargaining', *Journal of Economic Behavior and Organization*, vol. 3, pp. 367–88.

Holt, C. 2007, *Markets, Games, and Strategic Behavior*, Pearson Education, New York.

Nikiforakis, N. 2008, 'Punishment and counterpunishment in public good games: Can we really govern ourselves?', *Journal of Public Economics*, vol. 92, pp. 91–112.

Nikiforakis, N. 2010, 'Feedback, punishment and cooperation in public good experiments', *Games and Economic Behavior*, vol. 68, pp. 689–702.

Nobel Announcement 2002, 'Press release: The Sveriges Riksbank Prize in Economic Sciences in Memory of Alfred Nobel', Nobelprize.org, viewed June 2010, <http://nobelprize.org/nobel_prizes/economics/laureates/2002/press.html>.

Oswald, A. 2010, 'Notes on economics and the future of quantitative social science', May, University of Warwick, viewed June 2010, <http://www2.warwick.ac.uk/fac/soc/economics/staff/academic/oswald/maysciencedata2010.pdf>.

Plott, C. R. and Smith, V. L. 2008, *Handbook of Experimental Economics Results*, vol. 1, North-Holland, Amsterdam.

Smith, V. L. 1962, 'An experimental study of competitive market behavior', *Journal of Political Economy*, vol. 70, pp. 111–37.

Smith, V. L. 1976, 'Experimental economics: Induced value theory', *American Economic Review*, vol. 66, no. 2, pp. 274–9.

Van Huyck, J. B., Battalio, R. C. and Beil, R. O. 1990, 'Tacit coordination games, strategic uncertainty, and coordination failure', *American Economic Review*, vol. 80, pp. 234–48.